



Mergers In Λ CDM: Uncertainties in Theoretical Predictions and Interpretations of the Merger Rate

Citation

Hopkins, Philip F., Darren Croton, Kevin Bundy, Sadegh Khochfar, Frank van den Bosch, Rachel S. Somerville, Andrew Wetzel, et al. 2010. "MERGERS IN Λ CDM: UNCERTAINTIES IN THEORETICAL PREDICTIONS AND INTERPRETATIONS OF THE MERGER RATE." *The Astrophysical Journal* 724 (2): 915–45. <https://doi.org/10.1088/0004-637x/724/2/915>.

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:41381865>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

MERGERS IN Λ CDM: UNCERTAINTIES IN THEORETICAL PREDICTIONS AND INTERPRETATIONS OF THE MERGER RATE

PHILIP F. HOPKINS¹, DARREN CROTON², KEVIN BUNDY¹, SADEGH KHOCHFAR³, FRANK VAN DEN BOSCH⁴, RACHEL S. SOMERVILLE^{5,6}, ANDREW WETZEL¹, DUSAN KERES⁷, LARS HERNQUIST³, KYLE STEWART⁸, JOSHUA D. YOUNGER⁹, SHY GENEL³, & CHUNG-PEI MA¹

Accepted to ApJ, June, 2010

ABSTRACT

Different theoretical methodologies lead to order-of-magnitude variations in predicted galaxy-galaxy merger rates. We examine how this arises, and quantify the dominant uncertainties. Modeling of dark matter and galaxy inspiral/merger times contribute factor ~ 2 uncertainties. Different estimates of the halo-halo merger rate, the subhalo “destruction” rate, and the halo merger rate with some dynamical friction time delay for galaxy-galaxy mergers, agree to within this factor ~ 2 , provided proper care is taken to define mergers consistently. There are some caveats: if halo/subhalo masses are not appropriately defined the major merger rate can be dramatically suppressed, and in models with “orphan” galaxies and under-resolved subhalos the merger timescale can be severely over-estimated. The dominant differences in galaxy-galaxy merger rates between models owe to the treatment of the baryonic physics. Cosmological hydrodynamic simulations without strong feedback and some older semi-analytic models, with known discrepancies in mass functions, can be biased by large factors (~ 5) in predicted merger rates. However, provided that models yield a reasonable match to the total galaxy mass function, the differences in properties of *central* galaxies are sufficiently small to alone contribute small (factor ~ 1.5) additional systematics to merger rate predictions. But variations in the baryonic physics of *satellite* galaxies in models can also have a dramatic effect on merger rates. The well-known problem of satellite “over-quenching” in most current semi-analytic models (SAMs) – whereby SAM satellite populations are too efficiently stripped of their gas – could lead to order of magnitude under-estimates of merger rates for low-mass, gas-rich galaxies. Models in which the masses of satellites are fixed by observations (or SAMs adjusted to resolve this “over-quenching”) tend to predict higher merger rates, but with factor ~ 2 uncertainties stemming from the uncertainty in those observations. The choice of mass used to defined “major” and “minor” mergers also matters: stellar-stellar major mergers can be more or less abundant than halo-halo major mergers by an order of magnitude. At low masses, most true major mergers (mass ratio defined in terms of their baryonic or dynamical mass) will appear to be minor mergers in their stellar mass ratio - observations and models using just stellar criteria could underestimate major merger rates by factors $\sim 3 - 5$. We discuss the uncertainties in relating any merger rate to spheroidal formation (in observations or theory): in order to achieve better than factor ~ 3 accuracy, it is necessary to account for the distribution of merger orbital parameters, gas fractions, and the full efficiency of merger-induced effects as a function of mass ratio.

Subject headings: galaxies: formation — galaxies: evolution — galaxies: active — cosmology: theory

1. INTRODUCTION

In the now established Λ CDM cosmology, structure grows hierarchically (e.g. White & Rees 1978), making mergers an inescapable element in galaxy formation. The galaxy-galaxy merger rate, as a function of properties such as galaxy mass, redshift, and mass ratio, is a quantity of fundamental interest. It is critical for informing models of the growth and

assembly of galaxies, the distribution of bulge and spheroidal mass, properties of disks such as their thickness, morphologies, and flaring, and the fueling and growth of the most luminous starbursts and infrared systems as well as massive BHs and quasars.

The last few years have seen the emergence of a “concordance” precision Λ CDM cosmology (see e.g. Komatsu et al. 2009, and references therein), with remarkable convergence between different probes of structure formation. Meanwhile, cosmological dark-matter simulations have developed the ability to track large populations of dark matter halos and substructure within halos over cosmological volumes and timescales. Different calculations now yield results for e.g. the dark matter halo mass function that agree at the $\sim 5 - 10\%$ level (see e.g. Reed et al. 2007, and references therein).

Despite these advances, however, there has yet to be a similar convergence in theoretical predictions of the *galaxy* merger rate. Depending on the models used to convert dark matter halo growth histories into galaxy growth histories, different theoretical predictors of the merger rate in the same cosmology can differ by more than an order of magnitude (see e.g. the comparison in Jogee et al. 2008; Bertone & Conselice 2009; López-Sanjuan et al. 2009a). And the differences can-

¹ Department of Astronomy, University of California Berkeley, Berkeley, CA 94720, USA

² Centre for Astrophysics & Supercomputing, Swinburne University of Technology, P.O. Box 218, Hawthorn, VIC 3122, Australia

³ Max Planck Institut für Extraterrestrische Physik, Giessenbachstr., D-85748, Garching, Germany

⁴ Department of Physics and Astronomy, University of Utah, 115 South 1400 East, Salt Lake City, UT 84112-0830, USA

⁵ Space Telescope Science Institute, 3700 San Martin Dr., Baltimore, MD 21218, USA

⁶ Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD 21218, USA

⁷ Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA

⁸ NASA Postdoctoral Program Fellow, Jet Propulsion Laboratory, Pasadena, CA 91109, USA

⁹ Hubble Fellow, Institute for Advanced Study, Einstein Drive, Princeton, NJ 08540, USA

not be trivially attributed to e.g. the known fact that models predict somewhat different absolute galaxy masses and abundances – in terms of mergers *per galaxy*, there is similar variation.

Some of this may owe to the fact that dark matter merger rates are – unlike dark matter mass functions – derivative (instantaneous) quantities and so are more sensitive to choices of definition and methodology in simulations. But much of the difference owes to the fact that the mapping from dark matter merger rates to galaxy merger rates is non-trivial, and requires some relation between the history of a given dark matter halo and a galaxy inside that halo. In models that attempt to predict galaxy formation in an *a priori* sense, from physical prescriptions for gas cooling, star formation, and feedback from stars and accreting BHs, the resulting predicted halo to galaxy “map” and corresponding merger rates can be very sensitive to the input prescriptions. It has been shown that even in otherwise identical models, small changes in the treatment of gas cooling or stellar wind physics can lead to order-of-magnitude changes in the predicted major merger rates: compare the merger rates in Bower et al. (2006) and Font et al. (2008) (our § 4.2.3), or those in de Lucia & Blaizot (2007) and Bertone et al. (2007) (presented in Bertone & Conselice 2009).

These quantitative differences in the number of mergers per galaxy have led to qualitatively different conclusions regarding galaxy formation. Most models, especially those based on empirical halo occupation constraints or direct cosmological hydrodynamic simulation, have found that there are too many mergers in low mass systems to explain the survival and prevalence of disks in the simplest scenario (Granato et al. 2004; Somerville et al. 2001, 2008a; Koda et al. 2009; Stewart et al. 2009a; Khochfar & Silk 2009; Hopkins et al. 2009b; Stewart et al. 2009b; Sommer-Larsen et al. 2003; Abadi et al. 2003; Governato et al. 2007; Robertson et al. 2004; Okamoto et al. 2005; Scannapieco et al. 2008; Kazantzidis et al. 2008, 2009; Purcell et al. 2009).

More directly, *observed* merger rates find that $\sim 5 - 10\%$ of low or intermediate-mass galaxies ($M_* \lesssim 10^{10} M_\odot$ at redshifts $z \sim 0.2 - 1.2$) are in mergers – strongly morphologically disturbed or in “major” similar-mass pairs about to merge (at small scales and small relative velocities) (Bridge et al. 2007, 2010; Kartaltepe et al. 2007; Conselice et al. 2009; Jogee et al. 2009; Lotz et al. 2008b; Lin et al. 2008; Robaina et al. 2010). This is equal to or higher than the predicted rates from the models above. Even without correcting for the short expected duty cycles/lifetimes of this merger phase (see e.g. Lotz et al. 2008a; Conselice et al. 2009), these fractions are *already* higher than the observed fraction of bulge-dominated galaxies at these masses (Allen et al. 2006; Benson et al. 2007; Domínguez-Palmero et al. 2008; Weinzirl et al. 2009). Likewise the number density of IR luminous systems apparently in mergers at high redshifts is so high as to yield tension with the relic bulge mass in the Universe (see e.g. Younger et al. 2007; Genel et al. 2008; Tacconi et al. 2008; Hopkins et al. 2006; Cimatti et al. 2008). Typically, some additional physics (often some form of strong stellar feedback) must be invoked to preserve high gas fractions and suppress the efficiency of bulge formation. Indeed, this is *always* the case in hydrodynamic simulations of galaxy formation. This has led to the development of various feedback models (e.g. Sommer-Larsen et al. 2003; Abadi et al. 2003; Governato et al. 2007; Okamoto et al.

2005; Scannapieco et al. 2008, and references therein) and emphasis on gas-richness as a stabilizing factor (Robertson et al. 2006; Hopkins et al. 2009a; Stewart et al. 2009b; Moster et al. 2010a) that allows disks to survive such mergers. However, some semi-analytic calculations have claimed the opposite: that there are not enough mergers at low masses to explain even the small bulges observed at these masses (Parry et al. 2009).

On the other hand, at $\gtrsim L_*$, most empirical and semi-analytic models predict similar numbers of mergers and agree that massive bulges are primarily formed in major mergers (see e.g. Hopkins et al. 2010; Parry et al. 2009; Cattaneo et al. 2010, and references therein). But some simulations have claimed that there are more minor and fewer major mergers at these high masses (Naab et al. 2007).

As observations of e.g. pair and morphologically disturbed fractions improve, it is rapidly becoming possible to empirically constrain merger rates and fractions at a level better than the apparent order-of-magnitude scatter in predictions (see e.g. Kartaltepe et al. 2007; Lin et al. 2008; Domingue et al. 2009; Bundy et al. 2009; Conselice et al. 2009; Robaina et al. 2009, 2010; Comerford et al. 2009; Lotz et al. 2008b; Bridge et al. 2007; López-Sanjuan et al. 2009b). However, the conversion of such observations to merger rates still requires knowledge of appropriate timescales and contamination from projection effects and/or minor mergers – all of which mean that interpretation of the observations is somewhat dependent on how well or poorly theoretical predictions agree on basic quantities. Moreover, an observed pair or morphologically disturbed fraction is fundamentally *not* invertible at much better than the factor of ~ 2 level, as disturbances and dynamics are non-unique and depend on other quantities such as galaxy gas content; as such, they must be forward-modeled.

Fundamentally, the galaxy-galaxy merger rate depends on two basic quantities: the dark matter halo merger rate and the manner in which galaxies populate halos. However, models arrive at these quantities by various methods. In this paper, we attempt to examine in detail different theoretical models and identify the sources of apparently large variations in galaxy merger rates between the models. We are interested both in defining the “error budget” of theoretical predictions, as well as in identifying the definitions, physics, and prescriptions that lead to substantial differences.

In the follow sections we show that combining empirical halo occupation constraints with dark matter merger histories from high-resolution simulations leads to a relatively small (factor ~ 2) systematic uncertainty in merger rates (§ 2). We then compare various different substructure based calculations for subhalo-subhalo mergers and/or merger time delays (§ 3). We highlight several caveats and potential problems in adopting these merger rates: without proper care, it is possible to obtain apparently order-of-magnitude different dark matter merger rates from the same simulation or analytic calculation.

We reveal that most of the differences between model predictions owe to the treatment of baryons in galaxy formation (§ 4). Specifically, we show that cosmological hydrodynamic merger simulations lacking feedback do not reproduce the observed $M_{\text{gal}}(M_{\text{halo}})$ relation (not just the normalization of this relation, but also its shape) and so do not map robustly between halo-halo and galaxy-galaxy mergers (§ 4.1). Semi-analytic models, on the other hand, perform well for central galaxies, but commonly have well-known problems reproducing observed properties of satellite galaxies; this ac-

counts for most of the discrepancies both between different semi-analytic models and between those models and empirical approaches/calculations (§ 4.2). In § 5, we discuss the importance of consistent and appropriate mass ratio definitions, and the variations between different definitions, as well as the range in the effects of a given merger assuming changes in galaxy gas fractions, orbital parameters, and structural properties. We summarize our conclusions and discuss future tests and improvements in § 6.

Except where otherwise specified, we adopt a “concordance” cosmology with $(\Omega_M, \Omega_\Lambda, h, \sigma_8, n_s) = (0.3, 0.7, 0.7, 0.9, 1.0)$ and a Chabrier (2003) IMF, and appropriately normalize all observations and models shown. The choice of IMF systematically shifts the normalization of stellar masses herein, but does not otherwise change our comparisons.

Throughout, we use the notation M_{gal} to denote the baryonic (stellar+cold gas) mass of galaxies; the stellar, cold gas, and dark matter halo masses are denoted M_* , M_{gas} , and M_{halo} , respectively. When we refer to merger mass ratios, we use the same subscripts to denote the relevant masses used to define a mass ratio (e.g. $\mu_{\text{gal}} = M_{\text{gal},2}/M_{\text{gal},1}$), always taken such that $0 < \mu < 1$ ($M_{\text{gal},1} > M_{\text{gal},2}$).

2. OVERVIEW: COMPARING MODEL MERGER RATES

The galaxy merger rate is fundamentally determined by two quantities: the halo/subhalo merger rate, and the manner in which galaxies populate halos (the halo occupation function). The halo merger rate (the precise definition of which will be discussed in § 3.2) is determined in cosmological simulations; there are two basic choices, however, for the halo occupation function. Semi-analytic models and cosmological hydrodynamic simulations attempt to predict this from first principles. Empirical models based on the halo occupation formalism simply adopt this quantity based on observational constraints.

2.1. Methodologies

Semi-analytic models and direct simulations are a well known and traditional means of modeling galaxy formation. In cosmological hydrodynamic simulations, galaxy formation is simulated from cosmological initial conditions. Cooling is computed self-consistently, but some prescriptions must be adopted for star formation (as a function of local gas properties) and feedback from star formation and stellar evolution as well as black hole (BH) and active galactic nuclei (AGN) fueling, growth and feedback. It is well-known however that, in the absence of detailed prescriptions for both stellar and AGN feedback, such models do not reproduce the observed correlations between galaxy and halo mass – they overpredict the masses of both low and high mass galaxies, by a mass-dependent factor (i.e. predict a different magnitude and shape of $M_{\text{gal}}(M_{\text{halo}})$; see Kereš et al. 2009a; Maller et al. 2006; Oppenheimer et al. 2010; Choi & Nagamine 2009). As a consequence the most successful versions of such models have considered high-resolution “zoom-in” models of individual galaxies in formation, with extensive feedback prescriptions (e.g. Sommer-Larsen et al. 2003; Governato et al. 2007; Okamoto et al. 2005; Scannapieco et al. 2008). Unfortunately, it remains prohibitively expensive to simulate cosmological populations of galaxies (needed for the statistics to quantify galaxy-galaxy merger rates as a function of mass, mass ratio, redshift, and other properties) with the resolution and physical prescriptions necessary to follow these feedback

and formation models, let alone to widely vary these prescriptions. And even many strong-feedback simulations overpredict the baryon conversion efficiency (Guo et al. 2010a; Abadi et al. 2003; Okamoto et al. 2005; Governato et al. 2007; Scannapieco et al. 2008; Piontek & Steinmetz 2009).

Given this limitation, a more common approach to modeling galaxy formation in an *a priori* fashion has been the construction of semi-analytic models. In these models, galaxy formation prescriptions are “painted onto” a dark matter background. Dark matter halo locations, growth histories, merger rates, and other properties are taken from dark-matter only cosmological simulations and/or analytic extended Press-Schechter theory designed to match those simulations. Galaxies are analytically assigned to each halo from the simulation, with their evolution followed in a Monte Carlo fashion according to a simple set of analytic prescriptions, designed to approximate the processes of cooling from halo gas, star formation, and stellar and AGN feedback. Such approaches have the advantage that they are inexpensive, allowing experimentation with a wide variety of formulations for e.g. feedback (with various free parameters), constrained so as to reproduce observable quantities such as the galaxy mass function at $z = 0$. However, the prescriptions on which the results depend are approximations that usually must be applied to a variety of regimes well outside the range where they are observationally or physically (from e.g. direct simulations) calibrated. The results in certain regimes can be unexpected or, in some cases, unphysical. In addition, it is well known that there are many degeneracies between small changes to different prescriptions – there is no unique solution to reproducing e.g. the $z = 0$ stellar mass function, so the models are not guaranteed to be similar even if they are calibrated to the same observations initially (see e.g. Neistein & Weinmann 2010, and references therein).

In order to circumvent these physical uncertainties, models constructed using an empirical approach (the halo occupation formalism) have become popular as a means to predict the evolutionary paths of galaxies, including their clustering evolution, net growth, star formation histories, and merger rates (Zheng et al. 2007; Yan et al. 2003; Conroy et al. 2007; Conroy & Wechsler 2009; Pérez-González et al. 2008a; Tinker et al. 2005; Cooray 2006; Brown et al. 2008; Wetzel et al. 2009; Stewart et al. 2009a; Hopkins et al. 2010; Moster et al. 2010b). By matching the observed correlation functions of galaxies versus stellar mass, redshift, and other properties, one can empirically infer which halos the galaxies populate.¹⁰ Independent observational constraints on halo masses (discussed in detail in § 4) such as weak lensing, kinematic modeling, X-ray gas measurements, and group velocity dispersions yield complementary constraints. Populating a cosmological simulation of halos+subhalos in this manner, one can evolve it forward and determine where and when mergers occur.

¹⁰ We should formally distinguish traditional strict HOD models from subhalo abundance matching models. In the former, one paints galaxies onto host halos with no regard for halo merger histories or dark matter substructure, using clustering as a constraint (see e.g. Yan et al. 2003; Yang et al. 2003; Tinker et al. 2005; Kravtsov et al. 2004; Wechsler et al. 2006). In the latter, subhalo orbits are tracked in simulations and galaxies are assigned strictly to subhalos, but by matching the observed stellar mass function, so that the correlation function is a *prediction* of the model (which agrees very well; see e.g. Conroy et al. 2006; Vale & Ostriker 2006; Lee et al. 2009; Wetzel et al. 2009; Wetzel & White 2010; Moster et al. 2010b; Behroozi et al. 2010). For our purposes however, the two methods give the same mass functions and clustering/substructure, so are treated interchangeably.

Such approaches are not, of course, models for galaxy formation – they populate galaxies in halos according to what is observed, and then (evolving them forward some arbitrarily small amount in time) can be used to predict quantities such as the merger rate (or what the merger rate must be if both dark-matter clustering simulations and the clustering observations are correct). They do not make any explicit statement about how the galaxies formed in the first place, nor the physics that are important. They require, for example, that the ratio of halo to stellar mass be high in low and high-mass systems, but include no information about what feedback processes (believed to be stellar and AGN feedback, respectively) explain this. They are also, of course, limited by the observations: there are uncertainties in those observations corresponding to a range of $M_{\text{gal}}(M_{\text{halo}})$ allowed, and there are redshifts and masses that the observations do not probe (see Moster et al. 2010b; Behroozi et al. 2010, and references therein), as well as degeneracies in fitting the model parameters that may hide critical differences between different sub-populations (for example, those which will or will not merge in a short time (see the discussion in Guo et al. 2010a). Roughly speaking, the observational constraints themselves are robust above $\sim 10^9 M_{\odot}$ at $z = 0$ and $\sim 10^{11} M_{\odot}$ at $z \sim 2$, but rapidly become more uncertain outside of this range.

2.2. Semi-Empirical Models: Sources of Uncertainty

With the advent of dark-matter simulations that can follow substructure at high resolution and tight observational constraints on the stellar mass functions and clustering of galaxies as a function of mass and redshift, semi-empirical approaches have become quite robust. Stewart et al. (2009a) and Hopkins et al. (2010) present predicted merger rates derived from such models, based on a variety of different assumptions and observational constraints. We refer to those papers for model details: the basic methodology is similar. In Hopkins et al. (2010), halo-halo merger histories are taken from the Millenium dark matter simulation (Fakhouri & Ma 2008) (these are updated with the Millenium-II results in Fakhouri et al. 2010, but the results for our purposes are identical), and in Stewart et al. (2009a) from independent dark matter simulations¹¹; each halo is then populated with galaxies using the abundance-matching methodology (although variations and alternative halo occupation approaches are attempted), ensuring (by construction) a match to observed galaxy mass functions and correlation functions at all redshifts. In short time intervals, the model halos are evolved forward; when halos merge, either the subhalos are followed until they are no longer resolved or a “dynamical friction timescale” is assigned as calibrated from high-resolution simulations (Boylan-Kolchin et al. 2008), and at the end of this time the hosted galaxies are assumed to merge. Summing over the entire simulation volume, this gives the predicted galaxy-galaxy merger rate.

The two independent models find predicted merger rates that agree reasonably well: typical variations in the model inputs lead to factor < 2 changes in the predicted merger rates, at least at redshifts $z \lesssim 2$ where the halo occupation function is reasonably well-constrained by observations. From this per-

spective, the advantage of the semi-empirical models is that they do *not* attempt to model galaxy formation – as a result, differences between various predictions must owe either to the halo occupation distribution adopted (i.e. constraints on the galaxy mass-halo mass distribution), or to the “background” dark matter dynamics.

Figure 1 summarizes the comparisons from Hopkins et al. (2010).¹² The different model choices are discussed therein, and we will break each down in detail in what follows. For now, we simply illustrate the different effects on the merger rate.¹³ Specifically, we compare the predicted major merger rate (mergers of mass ratio $\mu_{\text{gal}} > 1/3$, where $\mu_{\text{gal}} \equiv M_{\text{secondary}}/M_{\text{primary}} < 1$ in terms of the galaxy baryonic mass M_{gal}) for intermediate-mass galaxies as a function of redshift. We begin with the model from Hopkins et al. (2010), and consider the effects of the choices below. We take this as our basic HOD or semi-empirical model throughout, frequently comparing with the model from Stewart et al. (2009a) (which uses different sets of empirical HOD constraints, dark matter simulations, and methods of subhalo tracking to identify mergers), but with many of the variations below considered separately (each can be considered its own model).

(1) **Cosmology:** We re-run the Hopkins et al. (2010) model adopting the cosmological parameters from WMAP1 (Spergel et al. 2003), WMAP3 (Spergel et al. 2007) and WMAP5 (Komatsu et al. 2009), as well as a “concordance” model with $(\Omega_{\text{M}}, \Omega_{\Lambda}, h, \sigma_8, n_s) = (0.3, 0.7, 0.7, 0.9, 1.0)$. Because the halo occupation statistics are constrained to reproduce the same stellar mass function, most of the difference between cosmologies (for example, in the predicted halo mass function) is effectively normalized out in this approach.

(2) **Halo-Halo Merger Rates:** The Hopkins et al. (2010) model adopts the halo-halo merger rate fits from Fakhouri et al. (2010), from the Millenium simulation (Springel et al. 2005b). An alternative dark-matter simulation, of comparable resolution, with halo merger rates determined using a different methodology¹¹, is described in Stewart et al. (2009a). Another is found in Gottlöber et al. (2001) (see also Kravtsov et al. 2004; Zentner et al. 2005); they quantify the fit separately to field, group, and cluster environments. We also consider the galaxy-galaxy merger rates directly determined from cosmological hydrodynamic simulations in Maller et al. (2006). It is well-known that such simulations yield galaxies that are overmassive relative to those observed (i.e. predict an HOD in conflict with that observed) – therefore we only use these mergers as “markers” of where the galaxy-galaxy mergers should occur, and (in order to study the stellar mass and mass ratio dependencies) re-populate the galaxies with the “correct” masses according to the HOD.

(3) **Substructure:** Instead of using a halo-halo merger rate with some “delay” applied to model subhalos in the HOD, we can attempt to follow subhalos directly after the halo-halo merger, and define the galaxy-galaxy merger when the subhalos are fully merged/destroyed. We compare the Hopkins et al. (2010) rates to those obtained tracking the halo+subhalo populations in cosmological simulations from Stewart et al. (2009a) (populating subhalos according to the

¹¹ The simulation is run with the ART code (Kravtsov et al. 1997), with distinct halo identification and subhalo tracking algorithms from those used in the Millenium simulations (Kravtsov et al. 2004), and a different subhalo ‘destruction’ criteria based on the fractional mass loss via stripping (subhalos merged when this falls below a fixed fraction of their inflow mass) as opposed to an absolute mass resolution limit.

¹² The merger rates derived from this model, including several of the variations discussed here, can be obtained from the “merger rate calculator” routine publicly available at <http://www.cfa.harvard.edu/~phopkins/mergercalc.html>.

¹³ Henceforth, we take the term “merger rate” to refer to the galaxy-galaxy merger rate, unless otherwise specified.

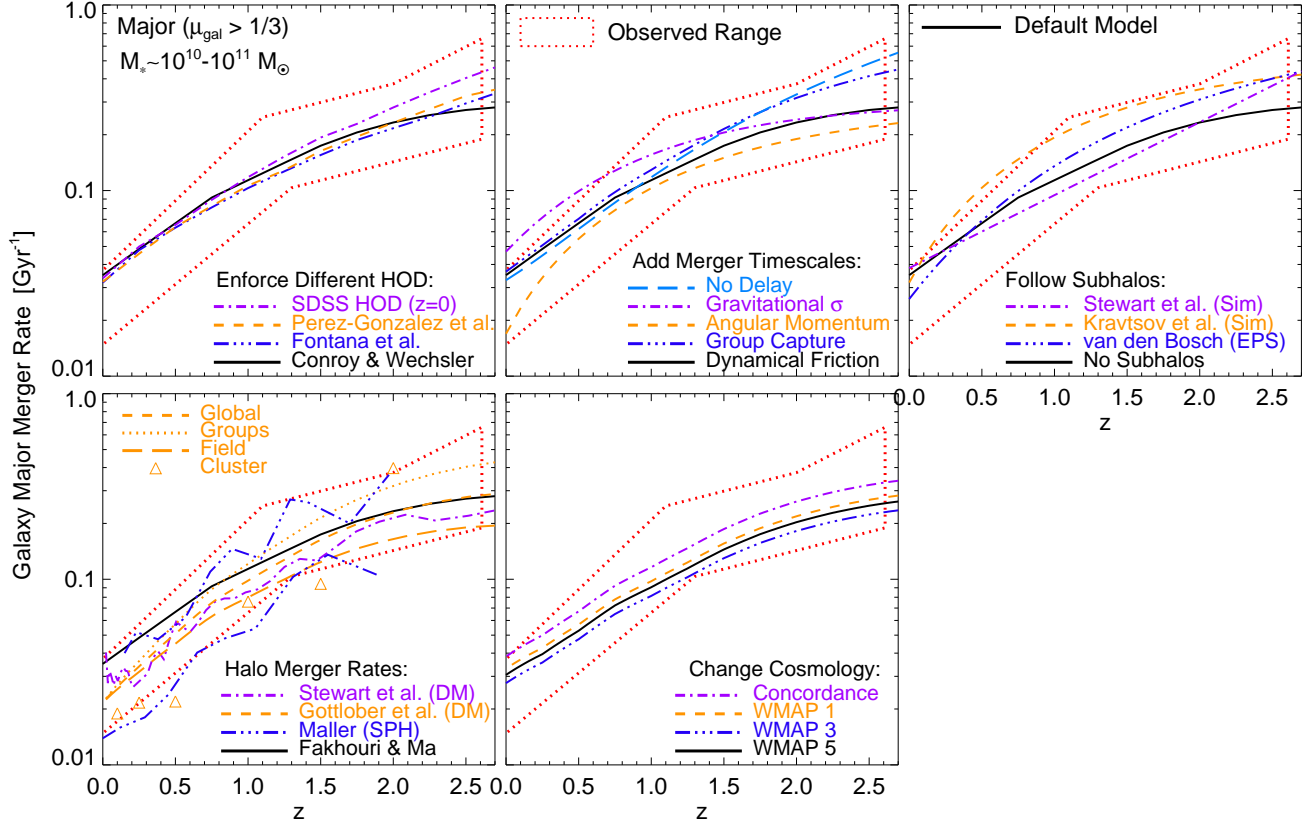


FIG. 1.— Comparison of the galaxy-galaxy major merger rate (mergers per galaxy per Gyr) for $M_{\text{gal}} \sim 10^{10} - 10^{11} M_{\odot}$ galaxies as a function of redshift, from different variations in semi-empirical models – i.e. models in which galaxies are placed in halos strictly according to observational constraints. Black line in each case corresponds to the “default” model from Hopkins et al. (2010). Red dotted range is the range allowed by observations compiled in that paper. *Top Left*: Changing the halo occupation model (observational constraints used to place galaxies in halos); each of the model choices listed is discussed in detail in the text (§ 2.2). The choices shown bracket the range allowed by a number of independent observational constraints. *Top Center*: Changing the merger “timescale” – i.e. timescale used to estimate the delay between a halo-halo merger and the resulting galaxy-galaxy merger (during which the satellite no longer grows). *Top Right*: Directly following subhalos in simulations, and assigning mergers when those subhalos are destroyed/fully merge into the central halo, instead of using the merger “timescale” approach. Results are shown from various methods of following subhalos in simulations and EPS trees. *Bottom Left*: Adopting estimates of the parent halo-halo merger rate, from different dark matter simulations. For the merger rates from Gottlöber et al. (2001), we also show the results using their merger rates in different environments (labeled). We also compare rates using cosmological hydrodynamic simulations to “tag” the halo-halo mergers of interest by identifying their resulting galaxy-galaxy mergers. *Bottom Center*: Changing the cosmological parameters. The allowed variations in these semi-empirical or HOD models lead to factor $\sim 2 - 3$ differences in the predicted merger rate.

same HOD). We also compare with the results using the different subhalo-based methodology described in Hopkins et al. (2008b) (essentially, beginning from the subhalo mass function constructed from cosmological simulations and evolving this forward in short time intervals after populating it, at each time, according to the HOD constraints). We compare two different constructions of the subhalo mass functions: that from cosmological dark-matter only simulations in Kravtsov et al. (2004) (see also Zentner et al. 2005) and that from the extended Press-Schechter formalism coupled to basic prescriptions for subhalo dynamical evolution, described in van den Bosch et al. (2005).

(4) Merger Timescales: Another commonly applied method in determining galaxy-galaxy merger rates as distinct from halo-halo mergers is to assign a simple delay to the galaxy-galaxy merger after the halo-halo merger, based on e.g. the dynamical friction time for the subhalo+galaxy orbit to decay, as calibrated in high-resolution N-body simulations that resolve the galaxy and follow the entire system self-consistently. The Hopkins et al. (2010) model adopts the traditional dynamical friction time, using the calibration from numerical simulations in Boylan-Kolchin et al. (2008). We compare with

the model re-run using the characteristic timescale for pair-pair gravitational capture in group environments, calibrated to simulations following Mamon (2006). We again repeat, considering capture in angular momentum space rather than gravitationally, following Binney & Tremaine (1987), as well as the gravitational cross-section timescale (similar to the group capture timescale) for capture between passages in e.g. loose group or field environments, calibrated from simulations in Krivitsky & Kontorovich (1997).

(5) Halo Occupation: We re-run the Hopkins et al. (2010) model using a different set of halo occupation constraints to determine the galaxy mass-halo mass relations. The default Hopkins et al. (2010) employs the fits to these distributions from Conroy & Wechsler (2009), which adopt a monotonic rank-ordering approach to assign galaxies to halos in a manner constructed to match observed mass functions and clustering (see the compilation in Fig. 1 therein and Bell et al. 2003b; Drory et al. 2005; Borch et al. 2006; Fontana et al. 2006; Panter et al. 2007; Pérez-González et al. 2008b). Gas masses are added to galaxies according to the observed correlation between galaxy gas fractions and stellar mass, from observations spanning $z = 0 - 3$ (Bell & de Jong

2000; McGaugh 2005; Calura et al. 2008; Shapley et al. 2005; Erb et al. 2006; Puech et al. 2008; Mannucci et al. 2009; Cresci et al. 2009; Forster Schreiber et al. 2009; Erb 2008; Mannucci et al. 2009) compiled in Hopkins et al. (2010) (§ 2.2.2 therein). We compare to the results using the same monotonic rank-ordering HOD methodology, but with galaxy stellar mass function constraints from Pérez-González et al. (2008b) or Fontana et al. (2006) (each determined up to redshift $z \sim 4$). We also compare to the results selecting the fit to $M_*(M_{\text{halo}})$ and its scatter for central and satellite galaxies from the observed SDSS clustering at $z = 0$ (Wang et al. 2006); in this case, we simply adopt the $z = 0$ fit at all redshifts – we do not allow for evolution.

We will discuss each of these aspects in detail in what follows. However, it should already be clear in Figure 1 that these choices, while not negligible, introduce only factor ~ 2 uncertainties in merger rates. Order-of-magnitude effects, it seems, must depend either on some particular combination of effects above, or relate to the baryonic physics of galaxy formation, such that the galaxy-halo mappings predicted are significantly different from those empirically determined by HOD models.

2.3. *A Priori Models Compared*

We now consider various *a priori* models for galaxy formation. In these cases, there are of course similar choices to those discussed above, for quantities such as the halo-halo merger rate, merger timescales and/or subhalo tracking, and cosmological parameters. However, the halo occupation distribution is not determined from observations, but is predicted as a consequence of attempting to model galaxy formation in an *a priori* manner based on some set of physical assumptions. That is not to say this is completely unconstrained, however; the models (especially semi-analytic models) are generally adjusted so as to reproduce various observations such as the stellar mass function and large-scale galaxy clustering. As a result, at least certain portions of the model should be similar to the HOD models.

The specific models we will consider throughout the paper are briefly as follows:

(a) Somerville et al. (2008a): A semi-analytic model updating those in Somerville & Primack (1999); Somerville et al. (2001). Halo growth is tracked using the extended Press-Schechter formalism (using a slightly modified version of the method of Somerville & Kolatt (1999), as described in Somerville et al. (2008a), to match the results of N -body simulations), with merging of sub-halos in virialized halos followed with appropriate dynamical friction times (including mass loss and tidal destruction). Cooling is calculated in the typical manner, distinguishing between “cold” and “hot” mode regimes (gas accretes onto central galaxies in a dynamical time in halos with $M_{\text{halo}} \lesssim 10^{12} M_{\odot}$, but forms quasi-static pressure-supported halos at larger masses, as motivated by cosmological simulations; see e.g. Kereš et al. 2005). Disks form assuming conservation of specific angular momentum as described in Somerville et al. (2008b), and star formation follows Kennicutt (1998) with supernova feedback. Black holes and bulges grow in mergers, with feedback associated with bright quasars and low-luminosity AGN, the latter of which suppresses new cooling in the “hot mode” of massive halos. This model adopts the prescription for bulge formation calibrated to simulations in Hopkins et al. (2009a), given the gas fractions and mass ratios of mergers.

Satellite galaxies are assumed to lose their hot gas reservoirs to their new parent halos (and cease to accrete new gas) upon halo-halo merger in the EPS tree, but do not lose the recycled gas from star formation internally in their pre-existing cold disks.

(b) Khochfar & Silk (2009) (see also Weinzierl et al. 2009): This semi-analytic model differs from that of Somerville et al. (2008a) primarily in that there is no “quasar mode” of feedback, and the prescriptions for the low-luminosity “radio mode” of AGN feedback and stellar wind feedback differ somewhat in detail. Galaxy structural properties are followed using energetic arguments and scalings motivated by simulations (Khochfar & Silk 2006). Bulge formation is followed with a more simplified recipe common in most semi-analytic models: the secondary stellar disk is destroyed (added to the bulge), but the primary disk is only violently relaxed in a major merger, and starbursts in mergers are tracked following Cox et al. (2008). The differences stemming from the bulge formation model are discussed in detail in Hopkins et al. (2009b). Satellite galaxies retain their cold and hot gas reservoirs until the merger with the primary central galaxy (but do not accrete new gas).

(c) Bower et al. (2006): This iteration of the Durham Galform SAM revises previous versions (Lacey & Cole 1993; Benson et al. 2003; Baugh et al. 2005) by adopting halo merger trees determined in N -body simulations, specifically the Millenium simulation (Springel et al. 2005b), and tracking substructure within those halos, as well as adding a mode of low-luminosity AGN feedback (a “radio mode”) which suppresses new cooling in “hot mode” halos. Cooling is followed as in the models above, and bulge formation with the prescription later adopted by Khochfar & Silk (2009), with the addition of allowed bulge formation in a strong disk instability mode that dominates at high redshift. This mode sets in when the disk becomes self-gravitating, and in consequence the disk is assumed to largely collapse to a bulge. Satellite galaxies are assumed to lose all of their hot gas to the new parent halo upon first linkage of the parent halos (often several R_{vir}), and gas which galaxies “expel” owing to the assumed stellar wind mode of feedback is considered part of the halo gas and treated similarly.

(d) Font et al. (2008): The most recent iteration of the Durham SAM. The model improves on that of Bower et al. (2006) by including a physically-motivated prescription for the removal of hot gas from subhalos after they are first linked to the main halo group (as a function of e.g. their position in the primary halo and the background gas density; motivated by ram-pressure stripping models). It is otherwise identical. We will show in § 4.2 that this has a large effect on the predicted merger rates.

(e) Croton et al. (2006): Another semi-analytic model based on the output of the Millenium simulation and broadly similar to that in Bower et al. (2006), but with important differences in the prescriptions for stellar feedback and cooling in both central and satellite galaxies (see § 4.2). Accretion in the “cold mode” is delayed by a galaxy dynamical time. Satellite halos are instantly stripped as in Bower et al. (2006); the mass expelled by star formation is contained in a separate reservoir to be recycled, but this too is lost to the parent halo for satellite galaxies. The merger trees are constructed from

the same simulation, but with different algorithms and cuts. The model also includes a much weaker disk instability mode of bulge formation, relative to that in Bower et al. (2006). The implementation of AGN feedback is qualitatively similar between the two models, but quantitatively the Croton et al. (2006) mode is “stronger” and yields more radio power.

(f) de Lucia & Blaizot (2007): The most recent iteration of the Munich SAM, this is a somewhat revised version of the Croton et al. (2006) SAM. The primary modifications include the adoption of a different stellar IMF and revised dust attenuation calculation, and the removal of a dynamical time delay in “cold mode” accretion (making “quenching” more rapid). Also, in this version, dynamical friction times assigned to galaxies before mergers are increased relative to the canonical dynamical friction time by a factor ~ 2 , with the goal of suppressing mergers at the massive end of the galaxy mass function (since the models tend to have difficulty reproducing this).

(g) Maller et al. (2006): The results of direct cosmological hydrodynamic simulations, incorporating cooling and star formation self-consistently, but without feedback of any kind from stars or BHs. Galaxy-galaxy mergers are directly identified, and their mass ratio quantified at the time just before merger. Hydrodynamic simulations with very large dynamic ranges in mass and redshift remain prohibitively expensive, so the comparison is restricted to intermediate and high-mass galaxies at redshifts $z \lesssim 2$. As the authors note, without feedback, the mass function at $z = 0$ is over-produced (and the break is not as strong as observed). They adopt a mean correction for this; essentially, re-normalizing their stellar masses by a systematic factor of ~ 3 .

Figure 2 compares the predicted merger rates from these models. We compare these with observational determinations of the galaxy-galaxy merger rate, from close pairs. Specifically, from the fraction of *major* ($\mu_{\text{gal}} > 1/3$) pairs with small projected separations $r_p < 20 h^{-1} \text{kpc}$ (often with the additional requirement of a line-of-sight velocity separation $< 500 \text{ km s}^{-1}$), and stellar masses in a specific range (labeled). For each mass bin, the pair fractions as a function of redshift can be empirically converted to a merger rate using the merger timescales at each radius. Lotz et al. (2008a) specifically calibrate these timescales for the same projected separation and velocity selection from a detailed study of a large suite of hydrodynamic merger simulations (including a range of galaxy masses, orbital parameters, gas fractions and star formation rates) using mock images obtained by applying realistic radiative transfer models, with the identical observational criteria to classify mock observations of the galaxies at all times and sightlines during their evolution. For this specific pair selection criterion, they find a median merger timescale of $t_{\text{merger}} \approx 0.35 \text{ Gyr}$, with relatively small scatter and very little dependence on simulation parameters ($\pm 0.15 \text{ Gyr}$).¹⁴ We use

¹⁴ The merger timescale from simulations at this radius is somewhat shorter than the time obtained assuming dynamical friction and circular orbits in e.g. an isothermal sphere, as has commonly been done (this is assumed in e.g. both Patton et al. 2002; Kitzbichler & White 2008). This owes to two effects: first, angular momentum loss at these radii is *not* dominated by dynamical friction (at least in the traditional sense of a small mass moving through a smooth, isotropic velocity dispersion background), but rather by exchange in strong resonances between the baryonic components that act much more

this median t_{merger} to convert the observations to a merger rate. Completeness corrections are discussed in the various papers; we also adopt the standard correction from Patton & Atfield (2008), calibrated to high-resolution simulations, for the fraction of systems on early or non-merging passages (to prevent double-counting systems on multiple passages); but this is relatively small (20 – 40%; see also Lotz et al. 2008a). For more detailed discussion of these observations and related constraints from other methods and observations, we refer to Hopkins et al. (2010).

We are not attempting to comprehensively discuss the predicted merger rates here: those rates are presented in other papers in much greater detail.¹⁵ However, we do wish to examine what can affect the predicted merger rate. It is clear in Figure 2 that the variation in predictions between different *a priori* models is significantly larger than that within the semi-empirical models. This motivates our further examination of the differences and similarities in these models.

For clarity, we have limited the number of semi-analytic models shown in Figure 2. However, we have compared with a number of other such models and find similar results (including e.g. Benson et al. 2003; Baugh et al. 2005; Kang et al. 2005; Cattaneo et al. 2006; Monaco et al. 2007; Bertone et al. 2007). The models considered span the representative behavior in these other semi-analytic models as well.

3. THE DARK SIDE

First, we examine the differences in the dark matter side of the merger rate, i.e. effects on the predicted merger rates that are independent of the detailed baryonic physics of galaxy formation.

3.1. Cosmological Parameters

As shown in Figure 1, the assumed cosmology makes little difference for merger rates, so long as models are normalized to produce similar stellar mass functions at each redshift. Present constraints on the cosmological parameters are also quite strong (Komatsu et al. 2009); there is little freedom to change the cosmology in a significant fashion. At high redshifts, the differences as a consequence of e.g. the value of σ_8 become larger, but so do the other sources of uncertainty below – at no point do we find that the uncertainties in cosmological parameters dominate the variations in predicted merger rates. Moreover, most of the models considered here adopt an identical cosmology (for example, all of the semi-analytic models in Figure 1 have been run or have available versions adopting the “concordance” cosmological parameters defined in § 1), and exhibit nearly identical variation between models.

efficiently. Second, by these radii, even initially circular orbits have become highly radial, leading to shorter merger times. Because of these effects, the remaining merger time at this scale depends only weakly on initial conditions or orbital parameters – essentially, these processes have erased most of the “memory” of the original orbital configuration.

¹⁵ For predicted merger rates from the Somerville et al. (2008a) model, see Hopkins et al. (2009b, 2010). For those from Khochfar & Silk (2009), see Weinzirl et al. (2009). From Bower et al. (2006), see Mateus (2008) and Parry et al. (2009). From de Lucia & Blaizot (2007), see Guo & White (2008), Kitzbichler & White (2008), and Bertone et al. (2007). The Maller et al. (2006) rates are discussed therein. The model predictions from de Lucia & Blaizot (2007); Bower et al. (2006); Font et al. (2008) can also be obtained online from the Millennium public database, at <http://www.mpa-garching.mpg.de/millennium/> or [http://icc.dur.ac.uk/index.php?content=Millennium/Millennium\(q](http://icc.dur.ac.uk/index.php?content=Millennium/Millennium(q) shown here from these two models are obtained from the same database).

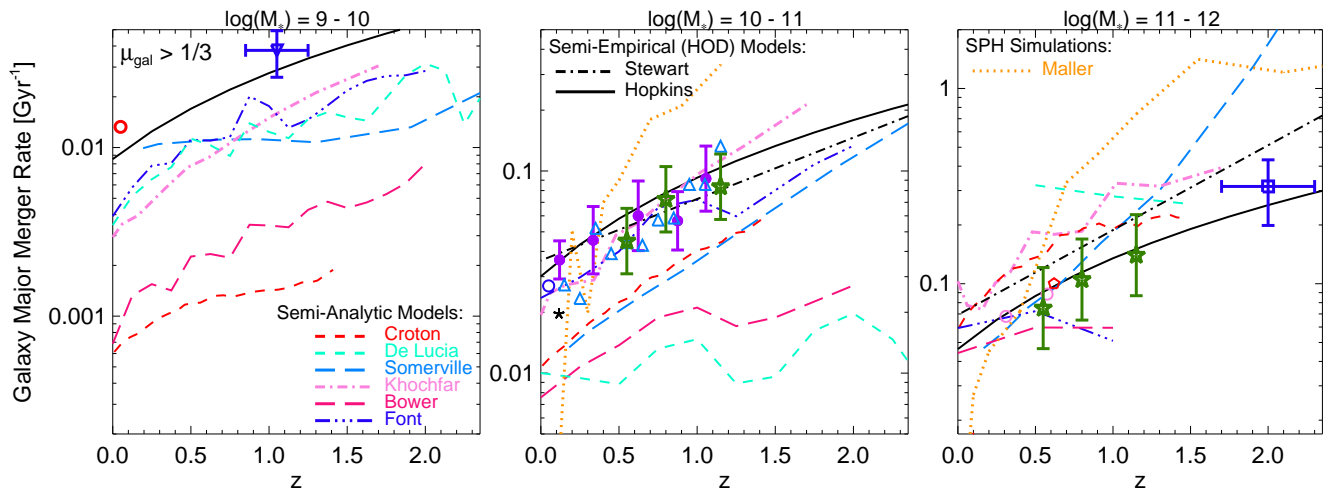


FIG. 2.— As Figure 1, but comparing the HOD-based models with different predictions from *a priori* galaxy formation models, in different galaxy stellar mass intervals ($\log(M_*/M_\odot)$, as labeled). We compare the standard semi-empirical HOD models from Hopkins et al. (2010) and Stewart et al. (2009a) (described in § 2.2) to the direct predictions of cosmological hydrodynamic simulations (with cooling and star formation, but without feedback from either stars or AGN; Maller et al. 2006) and the predictions of various semi-analytic galaxy formation models (Croton et al. 2006; Bower et al. 2006; de Lucia & Blaizot 2007; Khochfar & Silk 2009; Somerville et al. 2008a; Font et al. 2008, described in § 2.3). We compare these predictions with observations in each mass interval, based on close pair counts ($r < 20h^{-1}$ kpc) with calibration of merger timescales specific for each observed separation and selection criteria from large suites of high-resolution galaxy merger simulations (Lotz et al. 2008a). The observations are compiled in Hopkins et al. (2010) from Kartaltepe et al. (2007, blue triangles), Conselice et al. (2009, blue inverted triangles), Lin et al. (2004, 2008, pink circles), Xu et al. (2004, blue circle), De Propriis et al. (2005, black asterisk), Bluck et al. (2009, blue squares), Bundy et al. (2009, green stars), and Bell et al. (2006b,a, red pentagons).

Comparing the halo mass function and halo accretion histories with different cosmological parameters (see e.g. Neistein et al. 2006), the dominant effect is the predicted halo mass function shifting to higher masses with larger σ_8 . However, if we re-normalize the model to enforce the *same* observed galaxy stellar and baryonic mass functions and clustering, then these differences are largely normalized out, at least for variation in σ_8 in the relatively small observationally allowed range (here $0.7 - 0.9$ considered; much larger differences may significantly change dynamical friction times and other higher-order effects). Elahi et al. (2009) show that the quantity of greatest importance for our conclusions, the normalized substructure mass function or (equivalently) dimensionless merger rate (mergers per halo per Hubble time per unit mass ratio) is almost completely independent of cosmological parameters including e.g. the power spectrum shape $n_s \sim 1$ and amplitude over the range of variations here (not until one goes to much larger effective $n_s \sim 3$ does one see this function change shape).

3.2. Halo-Halo Merger Rates

3.2.1. Comparison of Different Simulations and EPS Calculations

Even given the same cosmological parameters, different methods used to determine the halo-halo merger rate can yield different answers. In Figure 3, we plot the dependence of the halo-halo merger rate on halo mass ratio $\mu_{\text{halo}} \equiv M_{\text{halo},2}/M_{\text{halo},1} < 1$ and redshift, from different cosmological simulations and analytic calculations. Specifically, we plot the number of mergers per unit redshift above some μ_{halo} at $z = 0$, and the evolution with redshift for major ($\mu_{\text{halo}} > 1/3$) mergers (evaluated in a narrow mass interval around $10^{12} M_\odot$ halos, in terms of post-merger mass, although all these models find that in these units, the mass dependence is weak).

We consider several different determinations of the halo-halo merger rate. Fakhouri & Ma (2008) calculate the merger

rate in the Millenium simulation in the usual fashion (determining the number of mergers per descendant halo in terms of the halo friends-of-friends group mass and mass ratio $\mu_{\text{halo}} < 1$). These are updated with the results of the higher-resolution Millenium-II simulation in Fakhouri et al. (2010), but the results are for our purposes identical. Genel et al. (2009) analyze the same simulation, but determine the merger rate using another methodology: defining halo masses in a different manner and quoting the merger rate per progenitor halo; they discuss in detail how this leads to subtle but potentially important variations in the merger rate fits. However, converted to the same units (following the methodology in their Appendix), the results are similar. Stewart et al. (2009a) determine the halo-halo merger rate from an independent dark matter simulation, using a slightly revised methodology and different time-step criteria. We also compare these simulation results with an analytic approximation: using the methodology in van den Bosch et al. (2005) to predict the merger rate analytically from a modified extended Press-Schechter formalism (note that these authors consider only the main branch mergers; because we are considering instantaneous merger rates here, this makes no difference, but it will lead to a different integrated number of mergers if not accounted for). Neistein & Dekel (2008) also present Press-Schechter trees modified to better match numerical simulations. Yet another methodology is described in Hopkins et al. (2008b); those authors take the subhalo mass function from cosmological dark matter simulations, and infer the halo-halo merger rate required to maintain this mass function allowing for dynamical evolution of the subhalos (i.e. assume steady-state subhalo mass functions over a narrow redshift interval, given certain initial conditions). We specifically adopt their estimate using the subhalo mass functions presented in Kravtsov et al. (2004) and Zentner et al. (2005). In that paper, however, they consider several other simulation results as well, and find similar results in all cases (Shaw et al. 2006; Springel et al. 2001; Tormen et al. 2004; De Lucia et al. 2004; Gao et al. 2004;

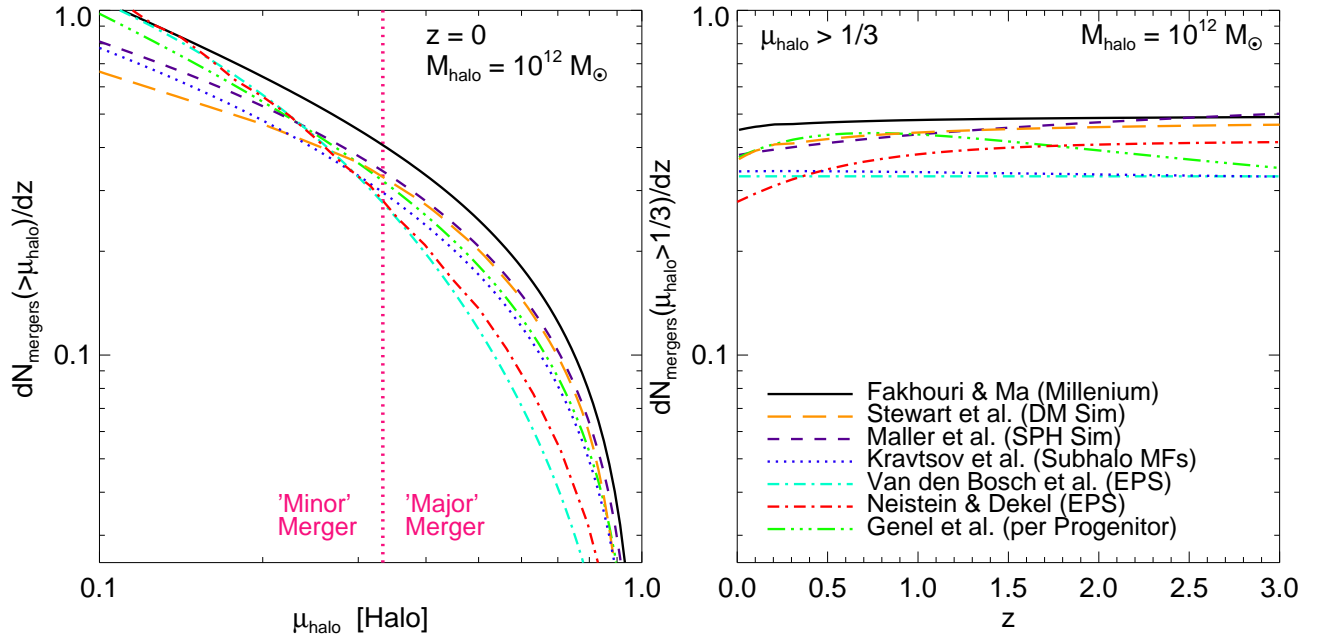


FIG. 3.— *Left*: DM halo-halo merger rates (mergers per halo per unit redshift) as a function of *halo* mass ratio (at $z = 0$), from different simulations and methodologies shown. Fakhouri & Ma (2008) and Genel et al. (2009) extract merger rates from the Millenium and Millenium-II DM-only simulations, with different methodologies. Stewart et al. (2009a) determine rates from independent DM simulations. The Kravtsov et al. (2004) rates are based on the integral constraint from subhalo mass functions, rather than an instantaneous merger rate. Rates from Maller et al. (2006) are from hydrodynamic cosmological simulations (using galaxies as halo tracers). Rates from van den Bosch et al. (2005) and Neistein & Dekel (2008) are constructed with the EPS formalism. *Right*: Corresponding halo major merger rate versus redshift. Most different simulated merger rates agree well (within $\sim 20 - 50\%$ for $\mu > 0.1$), provided definitions are appropriately accounted for. Results are shown for $M_{\text{halo}} = 10^{12} M_{\odot}$ (at the given redshift), but depend very weakly on halo mass.

Nurmi et al. 2006).

The mergers of halo barycenters are the real mergers of interest, and it is possible that the presence of baryons could alter halo mass functions and merger rates at some level. These and other baryonic effects can only be followed in cosmological hydrodynamic simulations. Of course, such models may not reproduce the correct observed galaxy baryonic masses. However, we can still use such simulations as tracers of galaxy-galaxy mergers; using the information about the position and mergers of galaxies, but replacing the galaxy masses according to the observed HOD (essentially using galaxies in the simulation to ‘tag’ the relevant halos merging).¹⁶ This also allows for the possibility that the baryonic material changes the halo-halo merger rate, by e.g. contributing to halo self-gravity, although this is typically estimated to be a $\sim 10 - 20\%$ effect for the major-merger regime of interest (see e.g. Weinberg et al. 2008). Figure 3 shows the results of this procedure given the cosmological smoothed particle hydrodynamics simulations in Maller et al. (2006).

Overall, there is quite good agreement between these different calculations – halo-halo merger rates as determined in cosmological simulations are generally converged, at least for intermediate masses and redshifts $z \lesssim 2 - 3$, at the $\sim 50\%$ level or better. The scatter in different simulation-based determinations grows to a factor ~ 2 at high/low masses.

Furthermore, many of these differences owe to choices of definition that, treated properly, will not enter into the galaxy-galaxy merger rate. For example, Genel et al. (2009) demonstrate in detail how changing the halo mass definition and definition of merger rate in terms of mergers per progenitor (and allowing for $\mu_{\text{halo}} > 1$; i.e. mergers into larger halos) leads

¹⁶ We do this by simply replacing the rank-ordered galaxy masses with those obtained by abundance matching.

to differences in the halo-halo merger rate relative to the fits presented in Fakhouri & Ma (2008) & Fakhouri et al. (2010) from the same simulation, and the differences can be significant (factor ~ 2). However, the merger trees in terms of e.g. halo *barycenters* in both cases are almost identical. In terms of galaxy-galaxy merger rates, the precise halo mass definition, for example, should make no difference, so long as the halo occupation statistics are appropriately renormalized to yield the same galaxy mass function and galaxy clustering versus luminosity (halo mass is simply a label to use in tagging galaxies to halos). Many of the differences in Figure 3, therefore, are actually reduced in terms of their effect on galaxy-galaxy merger rates, so long as sufficient care is taken in correcting for the definitions used.

Considerable effort has also gone into tuning EPS-predicted merger rates to better correspond to the results of N-body simulations (see e.g. Zentner 2007; Neistein & Dekel 2008). With these improvements, many current-generation EPS trees differ from the simulation results they are tuned to by as little as $\sim 10 - 20\%$. For detailed comparison of EPS and simulation merger rates, we refer to Genel et al. (2009). Clearly, this is not the dominant source of uncertainty in merger rates, as the differences between various counts of mergers from the same simulation differ by larger amounts, depending on their exact methodology.

3.2.2. Some Caveats in Adopting Halo-Halo Merger Rates

We do not mean to imply that choices of definition are entirely negligible. Depending on the application, a halo merger rate defined in terms of e.g. progenitor galaxies or descendant galaxies is more appropriate (the former being an estimate of the probability that a given halo *will have* a merger; the latter being an estimate of the probability that a given halo *did have*

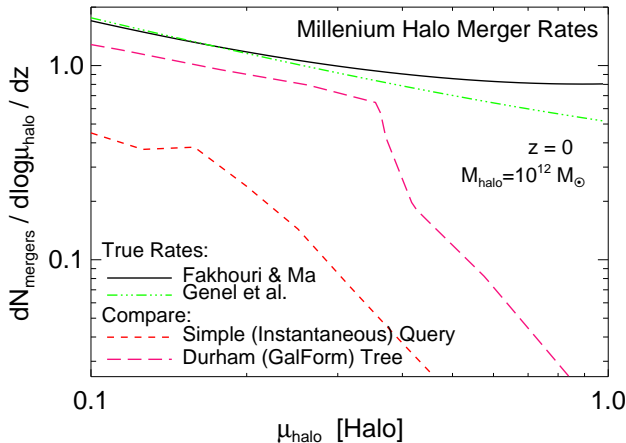


FIG. 4.— As Figure 3, showing the differential number of halo-halo mergers per log mass ratio (at $z = 0$) from the Millenium DM-only simulation. The Fakhouri & Ma (2008) and Genel et al. (2009) rates are intended to represent the halo-halo merger rate. We compare with the instantaneous rate taken directly from the Durham semi-analytic model halo trees (Harker et al. 2006), or from a simple query of the simulation database of the Munich halo trees using the instantaneous mass ratios just before the secondary is “lost” in the tree. Despite this being the same set of trees, these “instantaneous” mass ratio definitions lead to an apparent large suppression of major halo-halo mergers. This is an artificial effect of the definitions used; we show the differences to illustrate the importance of accounting for different halo mass ratio definitions in constructing the halo-merger rate.

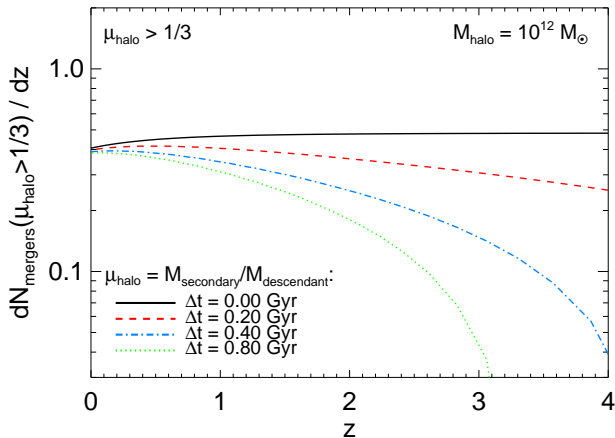


FIG. 5.— As Figure 3, comparing the halo merger rate as a function of redshift from the Fakhouri & Ma (2008) trees. We compare the merger rate using the minimum available time spacing ($\Delta t = 0.00$ Gyr, solid; in fact the spacing is finite but the halos are tracked on-the-fly in the simulation, so it is small for our purposes), to the rate inferred from comparing snapshots only with the given time spacing, if the mass ratio is defined by comparison of subsequent snapshots. Large timestepping means major mergers appear minor or are lost. If a quantity depends super-linearly on halo mass (as e.g. galaxy mass does at low masses), then the apparent suppression will be enhanced.

a merger). And there are means of constructing halo catalogs that could give quite different halo-halo merger rates, despite being ultimately a reflection of the same merger tree. The application of any halo merger rate requires careful accounting for consequences of such definitions.

Figure 4 illustrates this. We compare the halo-halo merger rates constructed from the Millenium simulation by Fakhouri & Ma (2008); Fakhouri et al. (2010) and Genel et al. (2009) to those constructed from the same simulation for use in the Durham semi-analytic models (Bower et al. 2006; Font et al. 2008), according to the method described in

Harker et al. (2006). There is an abrupt cutoff in the number of major halo-halo mergers in the latter tree. The differences are analyzed in detail in the Appendix of Fakhouri & Ma (2008), but they largely owe to two post-processing cuts made after the original halo+subhalo tree construction: subhalos are removed from their parent friends-of-friends (FOF) identified halo if their centers are outside twice the half-mass radius, or if they have retained at least 75% of their mass at the time they were last an independent halo. Subhalos (and their progenitors) that are removed from groups in this way are discarded from the trees.

The intention of these cuts is to reduce the number of spurious linkings from the FOF group finder, and both cuts above probably do so. However, they also suppress the number of halo-halo major mergers in the tree for three reasons. First, the discarded subhalos change the shape of the halo mass function – some high-mass halos lose a non-negligible fraction of their mass and, on the exponential tail of the mass function, this can have large effects (since small changes in mass correspond to large changes in number densities). Second, the secondary (sub)halo in a major merger is (fractionally) more robust to stripping than a small secondary in a minor merger, so major mergers are preferentially affected by the mass retention threshold. And third, in an equal mass merger, by definition the halos first overlap when the secondary center is at $\sim 2R_{\text{vir}}$ from the primary (in fact given the FOF algorithm used for linking, the distance can be $\sim 3 - 5R_{\text{vir}}$), so many of its subhalos will be instantaneously split off (being outside twice the half-mass radius at this instant) and thus not contribute to the secondary halo mass at the moment of merger (so the instantaneous mass ratio suddenly “dips” just before the merger).

This is not necessarily problematic for galaxy merger rates. The trees so constructed are designed for tracking the galaxy population in a specific semi-analytic model; they are *not* designed to represent the global halo-halo merger rate. The effects above mainly pertain to definitions; they mostly apply to the early stages of subhalo-halo mergers, not necessarily the final, relaxed stages when a galaxy merger will occur. As such they only indirectly affect galaxies, for example, via the small changes to the total halo mass introduced and corresponding gas mass assumed available for cooling. So the apparent difference in halo-halo merger rates translates much less directly in terms of galaxy-galaxy merger rates (basically, the baryonic physics act as a buffer between the halo-halo merger rate and galaxy-galaxy merger rate). But clearly, simply adopting the halo trees without accounting for how this processing is different from what is typically done (where e.g. halo mass ratio is the maximum pre-interaction halo mass ratio, for example) could lead to an apparent deficit of mergers (where the same mergers are, in fact, present).

This is related to another illustration in Figure 4, showing the result of adopting what might seem to be the most obvious definition of halo-halo mergers given a merger tree. Using a pair of closely spaced snapshots, and a descendant halo in the later snapshot, we simply take all the halos+subhalos in the earlier snapshot and define the mass ratio as the ratio of their instantaneous mass in that snapshot (the last where they are independent) to that of the main (most massive) progenitor. Specifically, we apply this algorithm to the standard Millenium simulation MPA merger trees. Mergers are again dramatically suppressed; this was noted in Bundy et al. (2007). The issue is that the definition of “halo-halo merger” is not trivial: in the merger tree used to compute this particular ex-

ample, halos “survive” in the tree so long as they can still be recognized as distinct subhalos/substructure/density peaks. In effect, this query is measuring the merger rate in terms of the instantaneous mass of a secondary subhalo just before it is no longer identifiable as a distinct subhalo/structure in the simulation (typically well after the two original halos have overlapped and become part of the same FOF group). By this time the smaller halo will have been stripped significantly, and with infinite resolution this particular way of defining mergers would make all mergers have a secondary mass $\rightarrow 0$. The correct definition of secondary mass is clearly some maximum mass or mass before interaction/joining into the same FOF group.¹⁷

Another, somewhat related caveat, regards the time-stepping of simulation outputs. We illustrate one of the possible dangers in Figure 5. Obviously, for a process to be well-resolved in time, simulation or model outputs must be spaced with some Δt which is much less than the characteristic evolutionary timescale of interest. For galaxy halos, this is the Hubble time at each z . Halos grow with a characteristic doubling time that is of order the Hubble time (only weakly dependent on mass); also, at each redshift, the time between mergers is of order the Hubble time; and the timescale from dynamical friction for a major merger to complete is a factor of $\sim 1/10 - 1/5$ of the Hubble time. In short, $\Delta t \ll t_H(z)$ is required, or in terms of redshift, $\Delta z \ll 1$ at all z where resolved dynamics are desired.

Without such resolution, mergers will be entirely missed – for example, a system with initial mass $m_1 = m$ could experience a merger with mass $m_2 = 3m$, then the product (mass = $4m$) merge with a mass $m_3 = 12m$ system. The central system has experienced two consecutive major mergers ($\mu = 1/3$) in this case. However, if the timestepping is large, the outputs of the simulation will only indicate that at some initial time, halos of mass $m_1 = m$, $m_2 = 3m$, and $m_3 = 12m$ were separate, then at a later time they were all together, and the most common assumptions applied in reconstructing merger histories will assume each merged onto the “primary” (m_3) separately, giving two minor mergers (a 1:12 and 1:4 merger, as opposed to the correct 1:3 and 1:3 merger). Considering the merger history along all branches of a given halo tree, the resulting uncertainty in merger rates can grow quickly if the time-stepping used is too large.

If one adopts a “merger timescale” to follow the galaxies in their last stages of infall after the subhalo is unresolved, then information is obviously lost without narrow timesteps – the merger time will necessarily be based on the last resolved dynamics, at large radii, which tends to lead to a significant over-estimate of the remaining “merger time” (see the discussion in § 3.3.3). This and the above can be significant concerns at high redshifts in semi-analytic models that interpolate between individual snapshots of the dark matter background at various times.

Moreover, if merger rates or mass ratios are defined in terms of the “descendant” mass – i.e. if the mass ratio is defined by the mass of a progenitor m_i (in one output) to the mass of the final halo m_f (in the next output), as is common in

many analyses, or if the halo mass bin for which the merger is “counted” (in e.g. asking the merger rate at a given mass) is m_f – then there can be a significant bias without narrow time-stepping because the halo will grow by a non-trivial amount in mass between the two timesteps (not just from the merger itself).

Figure 5 illustrates this. We extract the halo-halo merger rate from the Millenium simulation using a simple query – for every halo in a given timestep, we ask whether it had a progenitor (other than its primary/main-branch progenitor) in the previous timestep with mass $> 1/4$ the final mass (corresponding to a $> 1/3$ merger). Mass here, to avoid the other problems above, is defined as the maximum pre-accretion mass. But we vary the length of that timestep. If the timestep is large, then at redshifts $z > 2$, the apparent merger rate obtained by such a method plummets, even though it is well-known that the halo-halo merger rate continues to rise indefinitely with redshift. This is because the time-spacing of simulation outputs becomes large ($\Delta z \ll 1$ is no longer satisfied) and halos grow non-trivially between timesteps, so major mergers may represent only a small fraction of the mass of the “final” halo after a large time (in units of the Hubble time) passes.

3.3. Substructure: Tracking Subhalos and Their Dynamics

3.3.1. The Subhalo Merger Rate

Of course, a halo-halo merger is not a galaxy-galaxy merger; there will be some finite time where the galaxy is part of a satellite/substructure system before the final galaxy-galaxy coalescence. To follow this, simulations have made great improvements in tracking not just primary halos but also substructure in those halos, identifying subhalos on the fly (see e.g. Springel et al. 2001). Figure 6 plots the merger rates of such subhalos, measured from cosmological dark matter simulations: the subhalo “merges” when it is effectively destroyed in the simulation (it is mixed/stripped to the point where no subhalo can be identified by the algorithm, at least above the simulation mass resolution limit; we discuss some aspects of these limits below).¹⁸ The subhalo is continuously losing mass to stripping after the initial halo-halo merger (and by definition its mass is effectively zero when it “merges”), so the mass ratio is most appropriately defined in terms of the ratio of the *maximum* secondary halo mass (maximum mass it had when it was its own halo before any onset of stripping) to the instantaneous primary mass (technically primary minus subhalo mass, since we define an equal mass merger as 1:1, or $\mu_{\text{halo}} < 1$).

We show the results of this determination from several different sources: the subhalo trees from the Millenium simulation (the “MPAhalo” catalog); the comparable but independent dark matter simulations in Stewart et al. (2009a); and from a third set of simulations presented in Wetzel et al. (2009). We also consider subhalo tracking using a slightly different methodology, beginning with e.g. the subhalo mass function (rather than the differential subhalo merger rate) inside a radius $< R_{\text{vir}}$, as determined from cosmological simulations in Kravtsov et al. (2004), and evolving these subhalos forward from this already small radius according to

¹⁷ These issues are described in detail at <http://www.astro.utoronto.ca/~bundy/millennium/> (K. Bundy). There, the authors both outline the definitions that lead to this artificial merger rate suppression and present a modified query of the Millenium database (courtesy of V. Springel & S. White) that defines halo-halo mergers in terms of the more proper pre-accretion secondary mass, that does not suffer from this problem.

¹⁸ Implicit in Figure 6 is the assumption that all subhalos merge with the central one, not with each other. This is not strictly true, but simulations find it effects merger rates at only the $\sim 10 - 20\%$ level, much less than the differences between different models shown (Wetzel et al. 2009; Angulo et al. 2009).

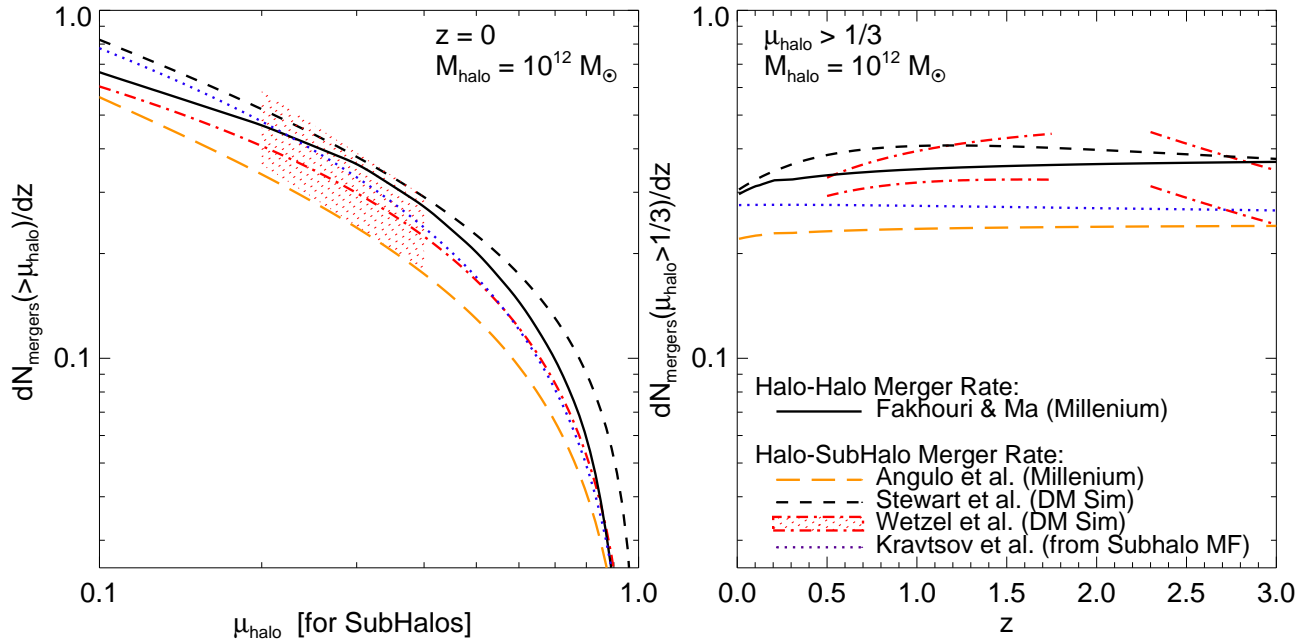


FIG. 6.— As Figure 3, but comparing halo-halo merger rates (from Fakhouri & Ma 2008) to subhalo-halo merger rates. Subhalo mergers are defined when subhalos are fully merged or destroyed with $\mu_{\text{halo}} = M_{\text{subhalo}}/M_{\text{halo}}$; where M_{halo} is the instantaneous total primary halo mass, and $M_{\text{subhalo}} = M_{\text{infall}}$ is the subhalo’s infall/maximum pre-accretion (pre-stripping) halo mass. We compare with the results from the subhalo trees in the Millenium simulation (the same simulation as Fakhouri & Ma 2008), taken from Angulo et al. (2009); Kitzbichler & White (2008); with the subhalo trees constructed from the independent DM simulation in Stewart et al. (2009a); from the subhalo mass functions determined in Kravtsov et al. (2004) with the methodology outlined in Hopkins et al. (2008b); and with the subhalo mergers defined in another different manner in Wetzel et al. (2009). For the Wetzel et al. (2009) results, shaded range (left) shows typical statistical and systematic uncertainties; double lines (right) bracket this range, they are broken as a function of redshift because of the different volume simulations used to probe different redshifts. Despite the fact that subhalos can survive a significant time after halo-halo merger, the rates are similar within a factor ~ 2 to the halo-halo merger rates; different subhalo identification methodologies also yield similar results to within the same factor.

the results of high-resolution galaxy-galaxy merger simulations used to calibrate e.g. dynamical friction times (see Hopkins et al. 2008b,a).

Despite the fact that these are all independent simulations, with somewhat different algorithms used to identify and track subhalos, the agreement is reasonably good: differences in the major subhalo merger rate are at the factor < 2 level. Moreover, the predicted subhalo major merger rate is quite similar to the predicted halo-halo major merger rate; depending on the calculation, subhalo major mergers appear to be between a systematic factor of ~ 2 lower than, or equal to, the halo-halo major merger rate.

3.3.2. The Halo Merger Rate with a “Merger Delay”

Another common approach, especially given low-resolution cosmological simulations or analytic (e.g. extended Press-Schechter) approaches where subhalo tracking is not possible, is to assign some “merger timescale.” This is a delay between the initial halo-halo merger and the final galaxy-galaxy merger, approximating the time for the decay of the secondary orbit and final resonant tidal interaction and coalescence of the two galaxies. Typically some variant of the standard Chandrasekhar dynamical friction time is employed, but this does not have to be the case: in fact, the angular momentum loss and orbital decay in the final stages of the merger are governed by resonant tidal processes, not dynamical friction against the smooth background. Because of these uncertainties, a number of authors have attempted to calibrate these timescales using high resolution simulations of galaxy-galaxy mergers and halo-subhalo or group encounters: we compare such calibrations in Figure 7.

We consider: (a) The directly extracted time required for

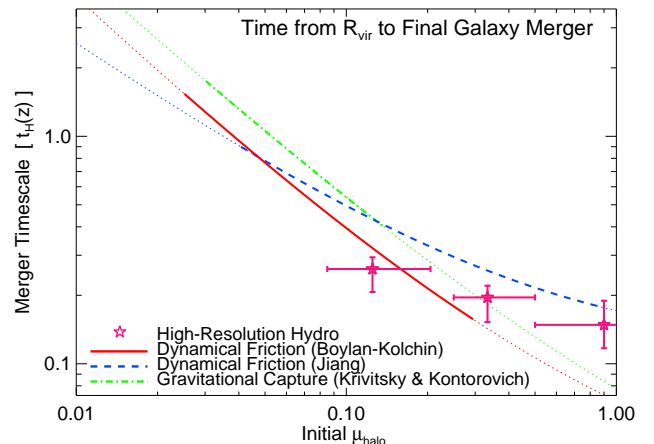


FIG. 7.— Comparison of different calibrations of the “merger timescale” (average time between halo-halo merger defined at crossing of R_{vir} and central galaxy-galaxy merger) in units of the Hubble time as a function of initial mass ratio. Each comes from high-resolution N-body simulations: dissipationless binary galaxy+halo merger simulations surveying orbital parameters from Boylan-Kolchin et al. (2008) (thick line corresponds to the range explicitly simulated; thin dotted line to extrapolation beyond this range), hydrodynamic binary galaxy+halo mergers (Cox et al. 2006; Younger et al. 2008); lower-resolution hydrodynamic cosmological simulations (Jiang et al. 2008), and dissipationless gravitational/group capture cross sections for galaxies in groups (Krivitsky & Kontorovich 1997). All yield a similar factor ≈ 2 scatter in timescales owing to cosmological distributions of orbital parameters (lines are for median circularity and other orbital parameters in cosmological simulations). Different estimators agree well where calibrated. Major-merger timescales are $\ll t_{\text{Hubble}}$; as a result, subhalo major merger rates are not very different from halo major merger rates.

the final merger in a galaxy-galaxy hydrodynamic (SPH)

merger simulation (of disk+bulge+halo systems), including star formation, multi-component galaxies, and gas physics (Cox et al. 2006; Younger et al. 2008; Hopkins et al. 2009a; Lotz et al. 2008a). We date the final merger by the coalescence of the two galactic nuclei. Error bars show the variation in timescales corresponding to sampling the full (isotropic) range of relative initial disk inclinations and orbital parameters (prograde mergers being a factor ~ 2 more rapid to coalesce than retrograde mergers, owing to the enhanced tidal responses dissipating angular momentum). **(b)** Calibration of a revised dynamical friction formula from gas-free (dissipationless) bulge+halo merger simulations in Boylan-Kolchin et al. (2008). These authors also survey a wide set of orbital parameters: we show the result for median and $\pm 1\sigma$ range of orbital parameters measured in cosmological simulations (Benson 2005; Khochfar & Burkert 2006). We show the full fit, but note that the simulations used span $\mu_{\text{halo}} = \mu_{\text{gal}} = 0.025 - 0.3$, over which range they agree with the SPH simulations. **(c)** A similar fit, for cosmological SPH simulations Jiang et al. (2008) & Jiang et al. (2010), calibrated from $\mu_{\text{halo}} = 0.04 - 1.0$. The cosmological nature of the simulations allows full sampling of representative orbits, with self-consistent galaxies; however the lower resolution suppresses resonant effects in major mergers, contributing to a slightly longer merger time. **(d)** The characteristic timescale for pair-pair gravitational capture in loose group or field environments, calibrated from simulations in Krivitsky & Kontorovich (1997) (see also Mamon 2006). Although dynamical friction can be a reasonable approximation to these scalings, such capture cross-sections are technically more appropriate for collisions in e.g. small groups or field environments, or major mergers, where the idea of a long-lived, slow inspiral is not appropriate. An alternative calibration in angular-momentum space gives similar results (see Binney & Tremaine 1987).

These timescales are discussed in greater detail in Hopkins et al. (2010) and Hopkins et al. (2008b). For our purposes they all agree reasonably well over the range where they are calibrated – to a factor better than ~ 2 . Where they are extrapolated, of course, they should be regarded with some caution.

Figure 8 shows the results of combining these merger timescales with halo-halo merger rates. In detail, each halo-halo merger is assigned a “merger time” based on the curves in Figure 7, appropriate for the redshift of the initial halo-halo merger and the halo-halo mass ratio at that moment, assuming the secondary begins at R_{vir} (which is what the fitted merger times are calibrated to use). The final merger is then delayed by this interval, during which the primary halo continues to grow and accrete. The mass ratio of the final merger is defined as the ratio of the maximum (pre-accretion) secondary halo to the primary halo mass at this time of final merger (as is done for subhalo merger rates). We consider both the fitted merger times from Boylan-Kolchin et al. (2008) and Jiang et al. (2008), which bracket the other calculations in Figure 7.

To see how merger rates would be affected by very long merger timescales, we have also considered the Boylan-Kolchin et al. (2008) time multiplied by a constant factor of 4 (for major mergers, roughly, the Jiang et al. (2008) time multiplied by a factor of ~ 2 , or $\sim 0.5 t_{\text{Hubble}}(z)$). This represents a fairly extreme case. We compare to the subhalo and halo-halo merger rates, as a function of mass ratio and redshift.

The merger rates calculated with this approach agree to

within a factor ~ 2 with those determined from a direct tracking of subhalos in simulations: it appears that evolution on the timescales adopted for “typical” orbital parameters in simulations, shown in Figure 7, is a good approximation to the full dynamics. The redshift evolution is slightly different: high redshift mergers (when halo growth is more rapid) have their mass ratios somewhat more suppressed, but in terms of mergers per unit redshift the evolution is still weak and within the range of various subhalo calculations.¹⁹ The differences between each different merger time calibration are small (for major mergers): it is only when we artificially increase the merger timescale by a factor of several that a significant difference appears, in the sense that high-redshift mergers are strongly suppressed. This emphasizes that, although existing calibrations from high-resolution simulations agree well and yield little difference when used properly, adopting a significantly longer merger time can substantially suppress the merger rate, especially at high redshifts.

3.3.3. Caveats: Subhalo Identification and The Application of Dynamical Friction to Orphans

Despite the reassuring agreement seen above, it should be noted that merger rates are sensitive at a non-trivial level to issues of subhalo identification: in particular the choice of algorithm and criteria used to determine subhalo “membership,” and the time, mass, and force resolution of the simulation. These effects have been examined in a number of simulations (see e.g. Klypin et al. 1999; Springel et al. 2001; Shaw et al. 2007; Wetzel et al. 2009; Wetzel & White 2010; Giocoli et al. 2010, and references therein). We note here a couple of possible pathological regimes.

First, it is possible in some particular circumstances that subhalos could survive “too long” in the simulations shown above, depending on the definition of the time of merger. This can happen, for example, because the contribution of baryonic galaxies to speeding up the merger is neglected. Including baryons keeps subhalos more tightly bound and changes the mass distribution, shortening the dynamical friction time (see Weinberg et al. 2008). Moreover, the baryonic components actually control the final merger, allowing for resonant interactions once the secondary is within a radius that contains a mass less than or of order its own mass (at which point the Chandrasekhar dynamical friction approximation is no longer valid). This makes the merger time within $\sim 20 - 50\text{kpc}$ shorter by a factor ≈ 2 , relative to that estimated without baryonic galaxies (see e.g. Lotz et al. 2008a, 2010b,a), and can even lead to the merger and/or destruction of the baryonic galaxies *before* the subhalo (their bound dark matter) is significantly gravitationally disrupted or stripped (see e.g. D’Onghia et al. 2009). The lack of very high resolution and accurate *ab initio* formation of realistic disks (in e.g. their scale heights and sizes) in full cosmological simulations contributes to the fact that the major merger timescales in the calibration of Jiang et al. (2008) are significantly longer than those in high-resolution hydrodynamic simulations of individual mergers of disk galaxies. And in the opposite, infinite-resolution limit of dark-matter simulations, a subhalo

¹⁹ If satellite-central mergers were simply a delayed version of halo-halo mergers, keeping mass ratio fixed in infall, then the merger rate distribution versus mass ratio would have the same shape as that of halo-halo mergers but with a redshift lag – and because merger rates increase with redshift, it would be higher than the immediate halo-halo merger rate. But differential mass evolution in this time leads to fewer major mergers. Together these drive the effects seen in Figure 8 (see Wetzel et al. 2009).

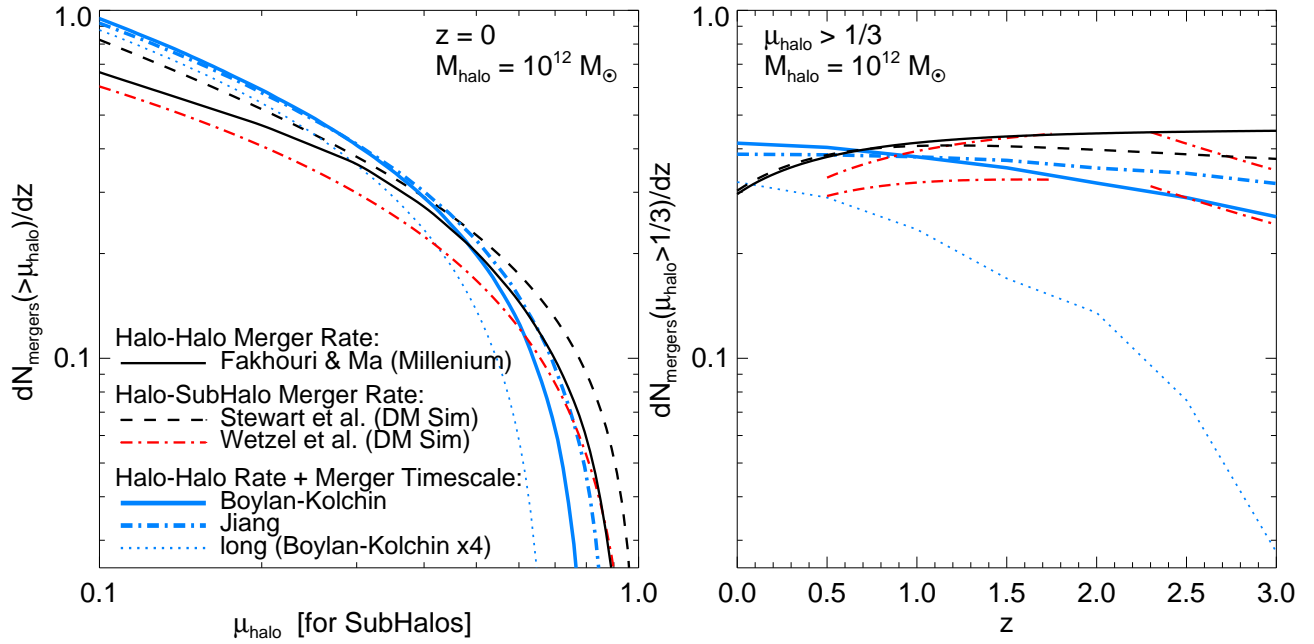


FIG. 8.— Comparison of halo and subhalo merger rates from Figure 6, to the merger rate determined by taking an “initial” halo-halo merger rate (that from Fakhouri & Ma 2008) and combining with the merger timescales from Figure 7 (i.e. including a “delay” between the time of halo-halo merger in the simulation and the final recording of the merger or assumed subhalo-halo merger, given by the different merger timescale fits in Figure 7). The timescales of Boylan-Kolchin et al. (2008) and Jiang et al. (2008) bracket the major-merger timescales therein. The difference is relatively small. We also compare the result given by artificially increasing the merger time by a large factor ($= 4$) relative to the calibration of Boylan-Kolchin et al. (2008); in this case the merger timescale becomes so long ($\sim t_{\text{Hubble}}$) that high-redshift mergers are substantially suppressed.

might never be completely disrupted (there always being a few tightly bound particles), even when the contained galaxies would have long ago merged.

Correcting for these effects is difficult. Subhalos could be identified as “merged” if they have lost some large fraction ($\gtrsim 90 - 99\%$; for calibrations see Wetzal & White 2010) of their mass (as opposed to simply reaching some resolution limit), but in low-resolution simulations this may prematurely “destroy” the subhalo. One could attempt to account for the resonant effects of baryons by merging any systems that have a passage within a close radius r to another galaxy, such that the enclosed mass $M(< r)$ is comparable to the subhalo/secondary mass (really the secondary galaxy mass plus tightly bound dark matter within $\sim a$ couple R_e) – within any such radius the resonant and baryonic interactions will be efficient and dominate over the dark matter dynamics and simple dynamical friction is not applicable.

A more common problem, especially in cosmological simulations (which by necessity have less than ideal resolution), is that subhalos are lost below the resolution limit “too early.” Identifying a subhalo as a distinct bound substructure is demanding in terms of resolution, and at some point subhalos will fall below these limits as they are stripped. Lack of inclusion of baryons (which would keep the subhalo more tightly bound), lower force resolution, or conservative subhalo membership criteria will all accelerate this. Subhalos in the Millenium simulation, for example, typically fall below the resolution limit within $\sim 0.5 - 1 R_{\text{vir}}$ (at low masses in particular). Wang et al. (2006) show that if one simply took only the still-resolved subhalos in this simulation at any given time, it would not be possible to reproduce the clustering/halo occupation of observed galaxies on small scales ($\lesssim 100$ kpc), or that of subhalos in higher force-resolution simulations. Similar conclusions are reached in Wetzal & White (2010).

To compensate for this, many models built on halo+subhalo trees invoke a population of “orphan” galaxies. This is the same principle as the “merger timescale” above: the models follow subhalos in the cosmological simulation as far as possible, and then assign a “merger time” to the galaxy based on the last recorded subhalo properties in the last timestep where the subhalo could be identified. Usually, the assigned timescale follows the Chandrasekhar dynamical friction formula,

$$t_{\text{merger}} \approx 1.17 \frac{V_{\text{vir}} r_{\text{sat}}^2}{G m_{\text{sat}} \ln(1 + M_{\text{vir}}/m_{\text{sat}})} \quad (1)$$

$$= (1.05 \langle \epsilon \rangle^{0.47} + 0.6) \frac{R_{\text{vir}}}{V_{\text{vir}}} \frac{M_{\text{vir}}/m_{\text{sat}}}{\ln(1 + M_{\text{vir}}/m_{\text{sat}})} \left\langle \frac{R_c}{R_{\text{vir}}} \right\rangle^{1/2}$$

from Binney & Tremaine (1987), with the second equality coming from the calibration in Jiang et al. (2008) assuming initial $r_{\text{sat}} = R_{\text{vir}}$ (their Equation 8; $\epsilon = j/j_c$ is the orbital eccentricity with $\langle \epsilon \rangle = 0.5$, and R_c/R_{vir} is related to the initial pericentric passage distance, both taken from the distributions they measure). Fits to high-resolution simulations with somewhat more freedom in the scalings give

$$t_{\text{merger}} = 0.56 \left(\frac{R_{\text{vir}}}{V_{\text{vir}}} \right) \frac{(M_{\text{vir}}/m_{\text{sat}})^{1.3}}{\ln(1 + M_{\text{vir}}/m_{\text{sat}})} \left(\frac{R_{\text{sat}}}{R_{\text{vir}}} \right) \quad (2)$$

from Boylan-Kolchin et al. (2008) (again adopting a the median initial circularity $j/j_c = 0.5$), where m_{sat} and r_{sat} are the satellite mass and location at the last time its subhalo could be identified, and V_{vir} and M_{vir} are the virial velocity and mass of the primary (minus the subhalo/secondary itself). Similar scalings are obtained from analytic considerations (at least in the small $m_{\text{sat}}/M_{\text{vir}}$ limit), accounting for subhalo mass loss and orbital parameter evolution (see Taylor & Babul 2004).

These particular formulations, discussed above (see Figure 7), have been tested in simulations and shown to be a

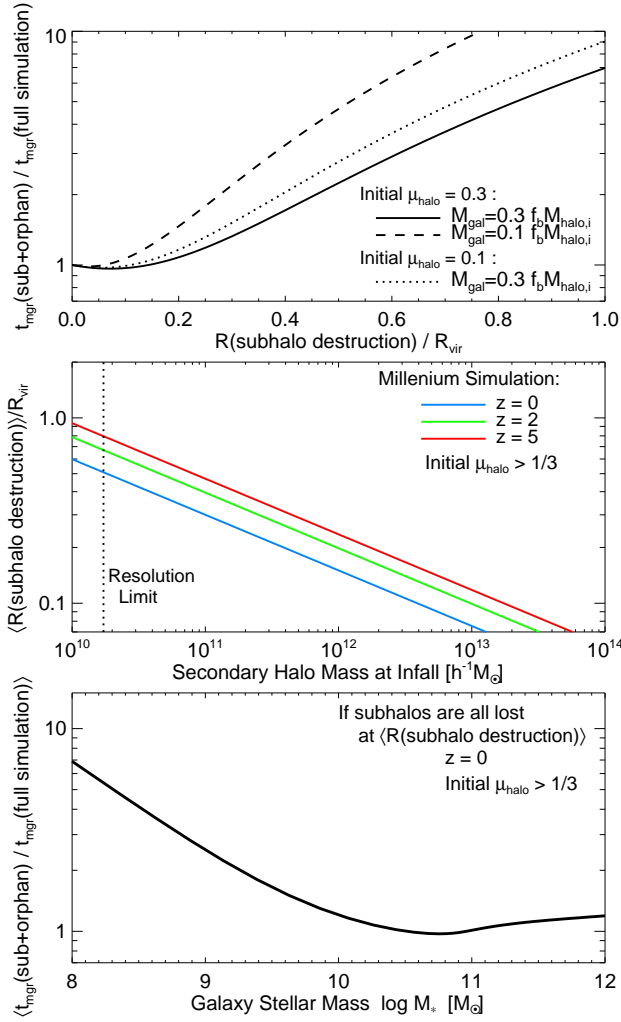


FIG. 9.— Caveats of using a subhalo+merger time (“orphan”) calculation of the total merger timescale. *Top*: Ratio of the “subhalo plus orphan” timescale $t_{\text{mgr}}(\text{sub} + \text{orphan})$ to the true merger timescale $t_{\text{mgr}}(\text{full simulation})$ from R_{vir} to final merger. The true timescale $t_{\text{mgr}}(\text{full simulation})$ is given by the calibrations in Figure 7 – simulations that follow galaxies fully from R_{vir} to $R \rightarrow 0$. The subhalo+orphan timescale comes from following a subhalo self-consistently to some $R(\text{subhalo destruction})$ where the subhalo is lost (falls below some resolution limit) – a timescale given by the same calibrations – then re-applying the merger time formula based on the instantaneous $R(\text{subhalo destruction})$ and remaining galaxy+subhalo mass. The ratio of this timescale to the true merger time is shown as a function of $R(\text{subhalo destruction})/R_{\text{vir}}$, the initial halo-halo mass ratio, and galaxy mass. If $R(\text{subhalo destruction})/R_{\text{vir}} \gtrsim 0.2$, the total merger time can be significantly over-estimated. *Middle*: Average $R(\text{subhalo destruction})/R_{\text{vir}}$ in the Millenium simulation fall below the mass resolution limit, as a function of the secondary/satellite halo infall mass and redshift (S. Genel; private communication). Small satellites, which are close to the simulation resolution limit, naturally fall below the resolution limits at relatively large R in any simulation. *Bottom*: Effect on the “typical” inferred merger time for 1 : 3 mergers, assuming subhalos are lost at the radii of the $z = 0$ curve at *middle* and given the methodology shown at *top*, as a function of galaxy mass (using the $M_{\text{gal}} - M_{\text{halo}}$ relation from Wang et al. 2006). If subhalos are only marginally resolved, caution should be used in applying merger times based on instantaneous subhalo properties; better calibration is needed.

reasonable approximation to the *total* merger time from initial halo-halo contact at $r_{\text{sat}} = 1 - 2R_{\text{vir}}$, with m_{sat} defined as the maximum (pre-stripping) mass of the secondary (this is the input when used to calculate the merger time for a halo-halo merger, as in § 3.3.2). However, these formulae have *not* been calibrated for application to a satellite at a later time, after the subhalo has self-consistently evolved

and is heavily stripped. Assume, for example, that a subhalo with initial (pre-stripping) mass m_s self-consistently decays in orbit from an initial radius $R_i \sim R_{\text{vir}}$ according to the calibration from Boylan-Kolchin et al. (2008) (Equation 2 above), but at some point – when the subhalo is at a radius $r'_s \equiv R[\text{subhalo destruction}]$ and time $t = t_{\text{Eqn. 2}}(R_i \rightarrow r'_s | m = m_s)$ – can no longer be resolved. At this point the subhalo mass is m'_s . Now assume that the remaining merger time is then assigned according to the same Equation 2, with these revised input values, i.e. a remaining $t = t_{\text{Eqn. 2}}(r'_s \rightarrow 0 | m = m'_s)$.

The *total* merger time is then simply the time from R_{vir} to r'_s plus the assigned remaining time. But we also know what the total merger time should be from the initial R_{vir} radius to $r = 0$ (i.e. without breaking into sub-steps in this manner) – this is in fact what the formulae used above are explicitly calibrated from high-resolution simulations to reproduce (with the inputs of initial R_{vir} and m_s). Comparing the two, we obtain the ratio of the total merger time estimated from this two-stage breakdown, to the total merger time from R_{vir} to $r = 0$ directly from the simulations. This is

$$\begin{aligned} & \frac{t_{\text{mgr}}(\text{sub} + \text{orphan})}{t_{\text{mgr}}(\text{full simulation})} \\ &= \frac{t_{\text{Eqn. 2}}(R_{\text{vir}} \rightarrow r'_s | m = m_s) + t_{\text{Eqn. 2}}(r'_s \rightarrow 0 | m = m'_s)}{t_{\text{Eqn. 2}}(R_{\text{vir}} \rightarrow 0 | m = m_s)} \\ &= 1 + \left\{ \frac{m_s^{1.3} \ln[1 + M_{\text{vir}}/m_s]}{m_s'^{1.3} \ln[1 + M_{\text{vir}}/m'_s]} \left(\frac{r'_s}{R_{\text{vir}}} \right) - 1 \right\} \left(\frac{r'_s}{R_{\text{vir}}} \right) \quad (3) \end{aligned}$$

But m_s/M_{vir} is just the initial halo-halo mass ratio (μ_s). The final mass m'_s , evaluated instantaneously when the subhalo is nearly or entirely stripped, will by definition just be the galaxy mass $M_{\text{gal}} = \eta f_b M_{\text{halo}}$, where f_b is the Universal baryon fraction, M_{halo} is the *pre-accretion* subhalo mass, and η is some star formation efficiency ($\eta = 0.3$ around $\sim L_*$ where star formation is most efficient and drops rapidly at lower or higher masses). If we include the mass of the subhalo at the last time it was resolved, we add to this roughly the simulation resolution limit; but for intermediate to high mass galaxies and the resolution limits of simulations of interest, this is similar ($\sim 10^9 - 10^{10} M_{\odot}$). For typical $\sim L_*$ galaxies and the parameters of e.g. the Millenium simulation, with an initial 1:3 halo-halo merger, this yields $t(\text{sub} + \text{orphan})/t(\text{full simulation}) \approx 1 + (6.5x - 1)x$, where $x = r'_s/R_{\text{vir}}$ is the fraction of the virial radius where the subhalo is no longer resolved.

Figure 9 illustrates this as a function of initial mass ratio, galaxy mass/resolution threshold, and the radius where the subhalo becomes unresolved. In short, if subhalos can be self-consistently followed to small radii $< 0.1 - 0.2R_{\text{vir}}$, where the remaining merger time is short, there is no issue. However, if subhalos fall below the resolution limit at large radii, then the merger time can be over-estimated by a large factor. We compare the typical radii in the Millenium simulation where subhalos of a given initial (pre-accretion) mass become unresolved (after being accreted into a larger halo), at several redshifts (S. Genel, private communication). Unsurprisingly, when the subhalo initial mass is close to the resolution limit, subhalos are lost at $\sim R_{\text{vir}}$. Given the $M_{\text{gal}}(M_{\text{halo}})$ relation and these median radii as a function of mass, the Figure illustrates how this affects the expected merger time for satellites of a given mass: at high masses, subhalos are massive and can be followed self-consistently, at low masses however, subhalos are lost quickly and the application of Equations 1-2 to the “orphan” tends to over-estimate merger timescales. This will

have a large effect on predicted galaxy mass assembly at high redshifts.

The issue is that fitting formulae such as Equations 1 & 2 are not necessarily calibrated for these multi-tiered applications. This is not to say that dynamical friction is fundamentally invalid – orbits of small satellites can be well-approximated by application of an instantaneous dynamical friction drag force and tidal stripping (Quinn & Goodman 1986; Hernquist & Weinberg 1989; Velazquez & White 1999; Font et al. 2001; Villalobos & Helmi 2008; Benson et al. 2004). (Although for major mergers, the underpinning assumptions of dynamical friction are not strictly valid). But this is not as simple as a single timescale, and implicitly assumes a high-resolution limit – they are not necessarily designed to account for a resolution “floor” self-consistently. Moreover, many estimates of the merger time average out or ignore the detailed orbital information; often, cases where the subhalo is quickly dissociated/stripped correspond to highly radial orbits, and simply re-applying an orbit-averaged formula is not ideal. If a system has just had a radial passage, for example, it will most likely be “caught” in a snapshot at apocenter – applying the dynamical friction time from this point, assuming the orbit is circular (as in most models) will in fact over-estimate the remaining merger time by a factor of ≈ 6 .

4. THE LIGHT SIDE: HOW GALAXIES OCCUPY HALOS

The issues discussed thus far pertain to all models, whether they are halo occupation, semi-analytic, or simulation-based. Although we have seen that some of the aspects of the treatment of dark matter mergers can be important, we know from our comparison in § 2 and the discussion above that most of the order-of-magnitude differences in merger rates must primarily owe to the treatment of baryons. Once we know how halos/subhalos behave, then the other necessary ingredient to predict galaxy merger rates is some prescription for how galaxies populate those halos (in other words, the “halo occupation” statistics). We therefore investigate how differences in these halo occupation aspects of the models give rise to different merger rates.

4.1. Halo Occupation Statistics of Central Galaxies

4.1.1. At Low-Redshift

With some complete *a priori* knowledge of subhalo evolution and mergers, the galaxy-galaxy merger rate, as a function of galaxy mass and mass ratio, can be determined knowing the distribution of galaxy masses in each halo. To lowest order, this is given by the function $M_{\text{gal}}(M_{\text{halo}}, z)$ and its scatter. Of course, this could depend on other variables such as environment, but as discussed below, observations indicate that such additional parameters have much smaller systematic effects than the relevant uncertainties in $M_{\text{gal}}(M_{\text{halo}}, z)$.

For simplicity, and to isolate the important physics, we first consider only *central* galaxies (i.e. those where M_{halo} is the parent halo, not satellites in subhalos), at $z = 0$. Figure 10 shows the median $M_{\text{gal}}(M_{\text{halo}})$, as a function of M_{halo} , at $z = 0$ (we could also compare the scatter in this quantity, but all the models we consider predict a similar, relatively small $\lesssim 0.15 - 0.2$ dex scatter; see e.g. Yang et al. 2008; Moster et al. 2010b; Behroozi et al. 2010; Wetzel & White 2010; More et al. 2009; Guo et al. 2010a). Of course, since every satellite was a central galaxy in some halo of mass M_{halo} before being accreted, the halo occupation function and $M_{\text{gal}}(M_{\text{halo}})$ function of satellites should be identical (barring

strong satellite-specific physics and/or strong redshift evolution in the $M_{\text{gal}}(M_{\text{halo}})$ function since the time the satellite was accreted), so long as M_{halo} for satellites is defined as the maximum (pre-accretion) mass of the satellite subhalo (i.e. maximum mass when the satellite was central in its own halo). We will free this assumption and discuss the implications in § 4.2 below. It is worth noting that the vast majority of the galaxy number density and mass density at all galaxy masses is in central galaxies: this is the quantity that is in practice most strongly constrained by a model’s fit to the galaxy stellar mass function or the clustering amplitude (large-scale bias) of galaxies as a function of mass/luminosity.

Given some function $M_{\text{gal}}(M_{\text{halo}})$, the galaxy-galaxy merger rate is simply this function convolved with the population of halo-halo mergers, with the appropriate accounting for delays before the final subhalo merger/destruction or the “merger timescale” from e.g. dynamical friction. Since we aim to isolate the specific effects of differences in the distribution $M_{\text{gal}}(M_{\text{halo}})$, we show in Figure 10 the resulting merger rate after convolving each presented halo occupation function with the *same* halo merger rate (in this case that from Fakhouri & Ma 2008). A different choice of halo merger rates or subhalo treatment will systematically change the predicted rates along the lines shown in § 3, but will not change how the galaxy-galaxy merger rate systematically depends on HOD quantities.

We compare the predicted $M_{\text{gal}}(M_{\text{halo}})$ distributions from different theoretical models with several observational estimates. First, fitting to the correlation function of galaxies with a given baryonic or stellar mass²⁰ yields a characteristic host halo mass for each galaxy mass bin – we show the results of this exercise performed on SDSS galaxies at $z = 0$ from Wang et al. (2006). We also show the empirical “abundance matching” or “rank ordering” results: it has been shown that good fits to halo occupation statistics (e.g. group counts, correlation functions as a function of galaxy mass and redshift, etc.) over a range of redshifts $z = 0 - 4$ are obtained by simply rank-ordering all galaxies and halos+subhalos in a given volume and assigning one to another in a monotonic one-to-one manner (see e.g. Conroy et al. 2006; Vale & Ostriker 2006). Here we perform this exercise using the redshift-dependent stellar mass functions from Fontana et al. (2006) (showing the result at $z = 0$); but as discussed in § 2.2 other choices of stellar mass function make little difference. Recent studies (Guo et al. 2009; Moster et al. 2010b; Tinker & Wetzel 2010; Behroozi et al. 2010) obtain similar results (they are statistically indistinguishable from the other observational results shown in Figure 10). We compare these results to independent estimates of the $M_{\text{gal}}(M_{\text{halo}})$ distribution, where halo masses are determined from weak lensing studies in Mandelbaum et al. (2006). Other independent measurements give nearly identical constraints: these include estimates of halo masses in low-mass spirals as a function of mass from rotation curve fitting (see e.g. Persic & Salucci

²⁰ Because it is closer to the relevant parameter in e.g. merger simulations, and because it better highlights the relevant systematic effects, we will show our comparisons here usually in terms of the galaxy baryonic mass M_{gal} . However, our qualitative statements in this section apply to stellar mass M_* as well. Where necessary, we will convert between observed stellar and baryonic masses using the fits to the observed M_{gas}/M_* relation of galaxies as a function of redshift from $z = 0 - 3$ given in Stewart et al. (2009b); this is consistent with a large number of direct observations over this redshift range (Bell & de Jong 2000; McGaugh 2005; Calura et al. 2008; Shapley et al. 2005; Erb et al. 2006; Puech et al. 2010; Mannucci et al. 2009; Cresci et al. 2009; Forster Schreiber et al. 2009; Erb 2008).

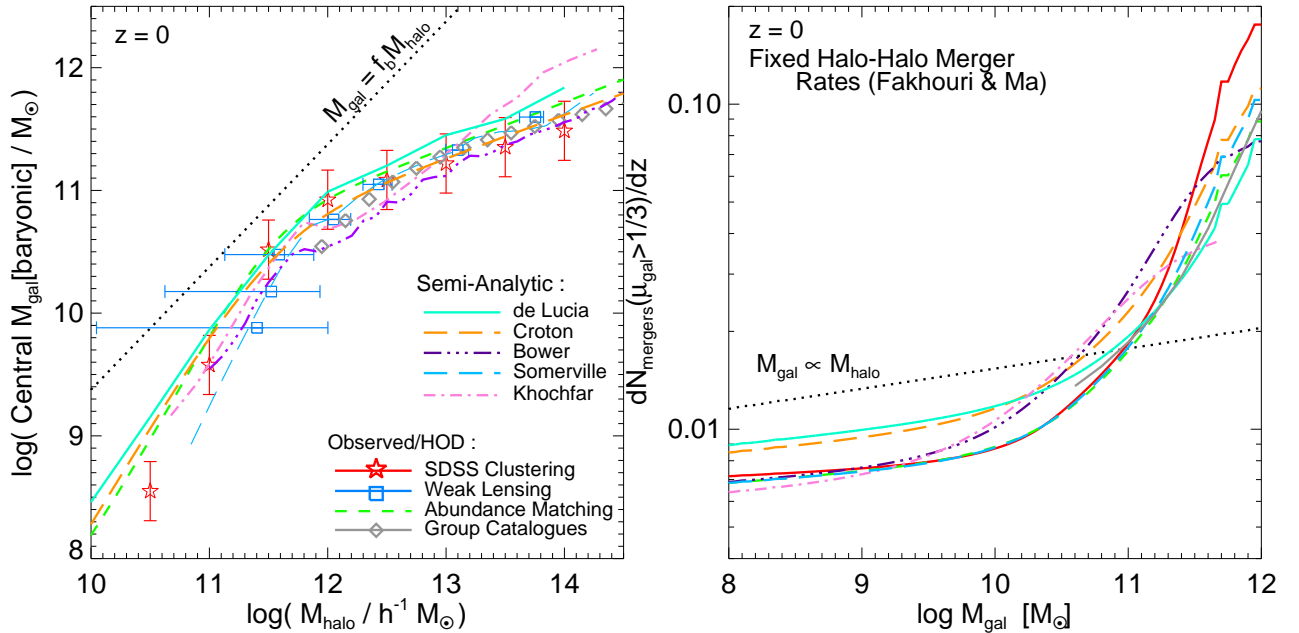


FIG. 10.— *Left*: Halo occupation statistics (mean galaxy baryonic mass as a function of host halo mass) from observations of weak lensing (Mandelbaum et al. 2006), fits to halo occupation models from clustering data in Wang et al. (2006), using the monotonic abundance matching method in Conroy et al. (2006), or from group catalogs with satellite abundances and kinematics (Yang et al. 2008). We compare with different semi-analytic model predictions, and the simplest efficient star formation model ($M_{\text{gal}} = f_b M_{\text{halo}}$; similar to hydrodynamic simulations without feedback). *Right*: Corresponding merger rates (at $z = 0$) as a function of M_{gal} (in ± 0.5 dex bins of M_{gal}), combining these HODs with the dark matter merger rates from Fakhouri & Ma (2008). We hold all other properties fixed. For now, we consider just the *central* galaxy HOD, and assume satellites obey the same relation (where M_{halo} is the maximum pre-accretion mass for satellites). Different observational constraints yield small (factor ~ 1.5 near $\sim L_*$) differences at $z = 0$; SAMs (being constrained to reproduce the stellar mass function) agree within factors $\sim 1.5 - 2$. However, simply taking halo-halo merger rates or $M_{\text{gal}} = f_b M_{\text{halo}}$, the merger rate predicted is different by factors of $\sim 3 - 10$.

1988, 1990; Persic et al. 1996b,a; Bell & de Jong 2000, 2001; Borriello & Salucci 2001; Borriello et al. 2003; Shankar et al. 2006; Avila-Reese et al. 2008) and estimates of the halo masses of massive groups and clusters from e.g. X-ray gas and group kinematics, as a function of central galaxy mass in massive brightest group or brightest cluster galaxies (BGGs/BCGs; see Eke et al. 2004; Yang et al. 2005a; Brough et al. 2006; van den Bosch et al. 2007, and references therein).

In semi-empirical or halo occupation-based models of galaxy-galaxy merger rates, the required $M_{\text{gal}}(M_{\text{halo}})$ distribution is simply adopted from these or similar empirical constraints, and enforced on all the galaxies in the model. Such models are therefore different only insofar as different empirical constraints yield different estimates of $M_{\text{gal}}(M_{\text{halo}})$. We find that the effects of these differences are relatively small (factor < 2 near $\sim L_*$, at $z \lesssim 2$); the level of variation within such models owing to observational uncertainties is discussed in detail in Hopkins et al. (2010) and also Stewart et al. (2009a), and summarized in § 2.2. Figure 10 also clearly illustrates that the different observational estimates of the $M_{\text{gal}}(M_{\text{halo}})$ distribution yield similar merger rates; the uncertainties are largest at very low mass (where observations are difficult and knowing the true gas fractions of galaxies, which are gas-dominated at these masses, is critical) and very high mass (where statistics are few and, given the stellar and halo mass function shapes, the merger rate will be most sensitive to small changes in the HOD).

Now compare these results with those from the $M_{\text{gal}}(M_{\text{halo}})$ distribution predicted by various *a priori* models. The semi-analytic models, for the most part, match the observational constraints quite well. This should not be entirely surpris-

ing: a major constraint used in building those models is that they reproduce the observed $z = 0$ galaxy luminosity and stellar mass functions. Such a result (an important achievement in and of itself, we stress) implicitly guarantees a reasonable match to the observations in Figure 10. Of the specific choices shown, the Bower et al. (2006) model appears to deviate from the observations to a slightly greater extent, around $\sim L_*$. The manner of this discrepancy reflects a common tendency for models to not reproduce quite as sharp a break in the stellar mass function as is observed. In a very narrow stellar mass bin, this could lead to non-trivial differences in the predicted merger rate, but averaged over a reasonable baseline ($\sim 0.5 - 1$ dex in stellar mass), the *shape* of the predicted HOD is nevertheless in sufficiently good agreement with the observational constraints that it yields merger rates converged at the $\sim 50\%$ level. It seems clear that, insofar as they succeed in reproducing the observed stellar mass functions, the differences between merger rates in semi-analytic models owing to differences in the predicted $M_{\text{gal}}(M_{\text{halo}})$ distribution of *central* galaxies at low redshifts are no larger than the differences that result when one attempts to directly adopt the $M_{\text{gal}}(M_{\text{halo}})$ distribution from observational constraints.

As discussed at length in Hopkins et al. (2010), the HOD shape is non-trivial, and this leads to a significant galaxy-mass dependence of the galaxy-galaxy major merger rate, even though the halo-halo merger rate depends only weakly on halo mass. If the galaxy formation efficiency were to be independent of halo mass, so that $M_{\text{gal}} \propto M_{\text{halo}}$, then galaxy-galaxy mergers would be a direct, trivial reflection of halo-halo mergers (regardless of the absolute value of that constant of proportionality). However, at low masses, the dependence of M_{gal} on M_{halo} is steep, $M_{\text{gal}} \propto M_{\text{halo}}^{1.5-2}$. So a “typical” 1:3 halo-halo

merger is a 1:9 galaxy-galaxy merger, and the galaxy-galaxy major merger rate is correspondingly suppressed (relative to the halo-halo merger rate). At high masses, the dependence is shallow, $M_{\text{gal}} \propto M_{\text{halo}}^{0.3-0.5}$, so a typical 1:9 halo-halo merger is a 1:3 galaxy-galaxy merger, and the galaxy-galaxy merger rate is enhanced.

This is important for model merger rates in hydrodynamic galaxy formation simulations: it is well-established that in cosmological hydrodynamic simulations (at least those without detailed prescriptions for both stellar and AGN feedback), galaxy formation is efficient, with $M_{\text{gal}} \sim f_b M_{\text{halo}}$ (see e.g. Springel & Hernquist 2003b; Maller et al. 2006; Naab et al. 2007, and references therein). Although considerable progress is being made (see e.g. Governato et al. 2007, 2009; Sijacki et al. 2007; Scannapieco et al. 2008; Di Matteo et al. 2008; Croft et al. 2009), no cosmological hydrodynamic simulation has yet been successful at reproducing the $z=0$ stellar mass function and HOD; and simulations without feedback tend to produce something much closer to efficient cooling/galaxy formation at all masses, with $M_{\text{gal}} \propto M_{\text{halo}}$ (or at least closer to this limit than to the $M_{\text{gal}}(M_{\text{halo}})$ distribution observed). As a consequence, such a simulation will predict a galaxy merger history that simply reflects the halo growth history: fewer major (and more minor) mergers, with a factor of several bias in the predicted merger rate. This effect is especially important at masses $\gtrsim L_*$; a cosmological hydrodynamic simulation that does not correctly reproduce galaxy “quenching” and match the observed bright end of the galaxy stellar mass function can under-predict the number of major mergers by factors of $\sim 5-10$.

In fact, these effects have been seen in such simulations (Maller et al. 2006; Naab et al. 2007). Our comparison here demonstrates that indeed this is expected in a simulation of this nature and reflects halo growth. However, if we take these same simulations, and simply re-populate all of the galaxies according to the observed HOD constraints (i.e. take the simulation but replace the galaxies just before merger with ones of the “correct” mass according to the observational constraints), then we obtain the results from Figure 1. These results agree to within a factor ~ 2 with the semi-empirical models and observed merger rates.

4.1.2. High-Redshift Results

Thus far we have only considered results at $z=0$. Figure 11 compares the $M_{\text{gal}}(M_{\text{halo}})$ distribution as a function of redshift, from different observational estimates and semi-analytic models. The agreement is reasonable around $\sim L_*$, but less so at higher and lower masses. This reflects well-known aspects of the comparison between such models and observed galaxy stellar mass functions – the SAMs tend to predict more low-mass and fewer high-mass galaxies at high redshifts (Fontana et al. 2006; Ilbert et al. 2010; Fontanot et al. 2009; Marchesini et al. 2009). Of course, the uncertainties in SAMs grow at high redshift, but so do observational uncertainties. Low-mass galaxies are subject to completeness concerns and selection effects; high-mass galaxies may have non-trivial corrections from extended intra-group light and/or uncertainties in their photometrically inferred stellar masses and photometric redshifts (Maraston 2005; Maraston et al. 2006), and there may be evolution in the stellar initial mass function (Hopkins & Beacom 2006; van Dokkum 2008; Davé 2008).

The growing differences and uncertainties in $M_{\text{gal}}(M_{\text{halo}})$ highlight the need for further study and observational constraints. Nevertheless, the resulting merger rates, integrated

over a reasonably broad stellar mass interval, are still similar to within a factor ~ 2 . This owes to the fact that, at a given stellar mass M_* , the merger rates do not depend on the absolute value of $M_{\text{gal}}(M_{\text{halo}})$, but rather the *shape* of this function.

4.2. Halo Occupation Statistics of Satellite Galaxies

We now generalize our previous comparison to allow for *different* $M_{\text{gal}}(M_{\text{halo}})$ distributions in satellite or central galaxies. In other words, whereas before, a satellite galaxy’s properties were set by its maximum pre-accretion halo mass (i.e. its state when it was last a central galaxy), we now free that assumption.

4.2.1. The Central-Satellite Difference and its Effect on Merger Rates

Figure 12 shows the relevant comparison. Specifically, we compare the median $M_{\text{gal}}(M_{\text{halo}})$ relation, for both central and satellite galaxies. As always, we define M_{halo} for satellites as the maximum M_{halo} before their accretion, i.e. their “in-fall” mass M_{infall} . We show two representative models: first, the HOD constraints from Conroy et al. (2006), inferred from the observational constraints, in which the two $M_{\text{gal}}(M_{\text{halo}})$ relations (that for satellites and centrals) are nearly *identical* (with this pre-accretion definition of M_{halo} ; see Gao et al. 2004; Nagai & Kravtsov 2005). Indeed, that the two distributions can be nearly the same is one of the major findings of such empirical HOD models. To be specific, the observational input in this method typically constrains the stellar mass $M_{*,\text{sat}}/M_{*,\text{cen}}$, not the baryonic mass $M_{\text{gal,sat}}/M_{\text{gal,cen}}$ – we *assume* here similar $f_{\text{gas}}(M_*)$ for the *cold*, bound, disk gas component. However we will show below (Figure 13) that observations show no difference in $f_{\text{gas}}(M_*)$ for the specific (major-merger candidate) systems of interest here. Further exploration of these distinctions is important, however (e.g. Neistein et al., in prep) – for now, this should more conservatively be considered an assumption in the semi-empirical models.

Observationally, the large-scale clustering of galaxies (and, to first order, the stellar mass function) constrain primarily central galaxies. However, the small-scale galaxy correlation function (the “one halo” term) probes satellite populations, and independent constraints are available from e.g. group counting statistics and satellite kinematics. As noted in § 4.1, most halo occupation models find good agreement with these observations by assuming that satellites/subhalos are described by the *same* M_{halo} distribution as central galaxies, so long as the subhalo mass M_{halo} used to assign the subhalo galaxies (satellites) their masses is the maximum pre-accretion mass, as opposed to their instantaneous (stripped) subhalo mass (see e.g. Conroy et al. 2006; Zheng et al. 2007; Conroy et al. 2007; Vale & Ostriker 2006; Conroy & Wechsler 2009; Pérez-González et al. 2008a; Stewart et al. 2009a; Hopkins et al. 2010). This includes all satellites in the model, even those about to merge (at very small scales $\lesssim 20-100$ kpc), and results in excellent agreement with the observed number of pairs on those scales as a function of galaxy mass and redshift (Conroy et al. 2006; Bell et al. 2006a; Stewart et al. 2009a).

The top panel of Figure 12 shows this result: $M_{\text{gal}}(M_{\text{halo}})$ is identical for both central and satellite galaxies, defined in these terms. But although this is a typical assumption in halo occupation models, the observational constraints may not necessarily require identical satellite and central properties at fixed infall mass. Therefore, in terms of the

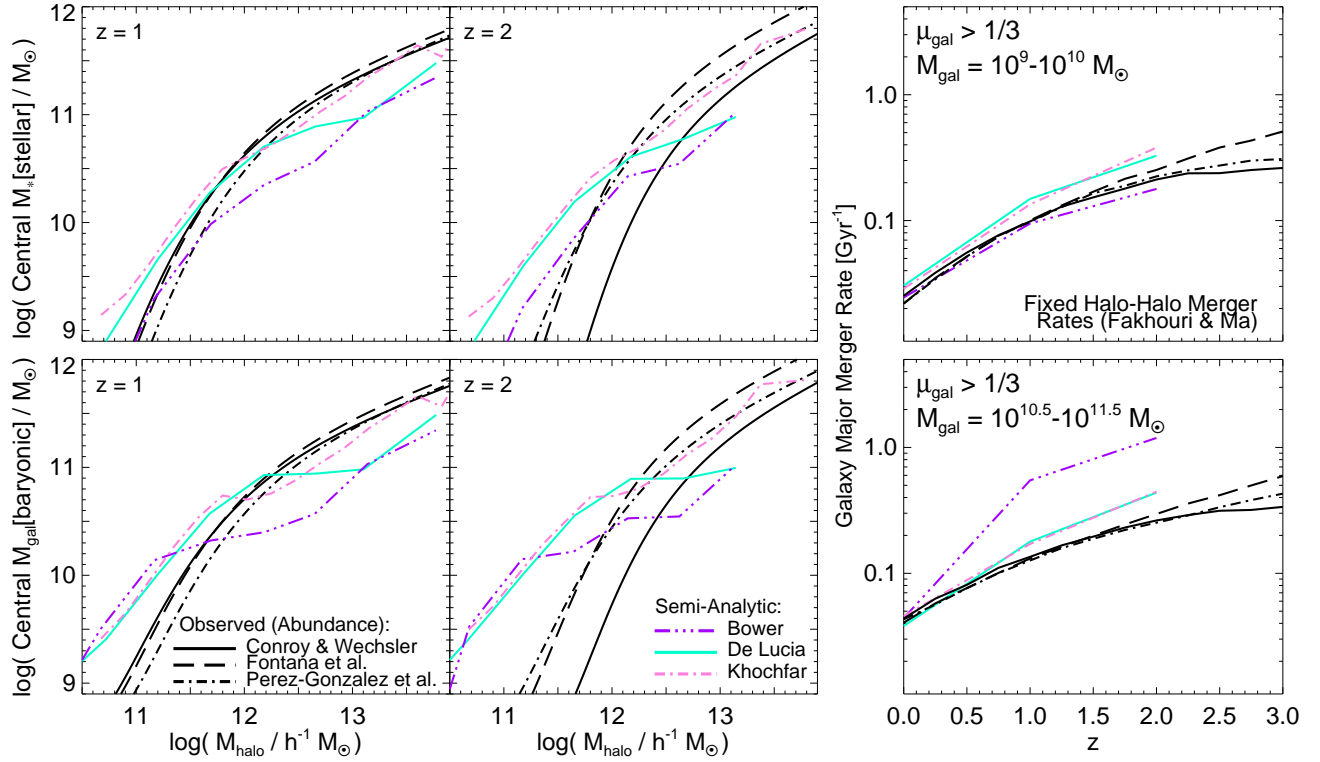


FIG. 11.— As Figure 10, but comparing empirical HOD constraints and *a priori* model predictions at different redshifts. *Top Left*: Stellar HOD (median central stellar mass versus M_{halo}) at $z = 1$ (left) and $z = 2$ (right). We compare the empirical HOD from abundance-matching, using stellar mass functions/HOD fits from Conroy & Wechsler (2009); Fontana et al. (2006) and Pérez-González et al. (2008b). We contrast predictions from the Bower et al. (2006), de Lucia & Blaizot (2007), and Khochfar & Silk (2009) SAMs. *Bottom Left*: Same, in terms of baryonic mass. *Right*: Resulting galaxy-galaxy major merger rate as a function of redshift, in a given baryonic mass bin, from the different HOD curves. In general, where the *shape* of the HOD is similar (regardless of normalization), the merger rates are in agreement.

ratio $M_{\text{satellite}}/M_{\text{central}}$ – the typical satellite versus central galaxy mass at fixed halo infall mass, the top panel shows the allowed scatter or systematic offset that could be tolerated within observational errors, given the constraints presented in Conroy et al. (2006), Conroy & Wechsler (2009) and Stewart et al. (2009a). The authors find that less than ~ 0.2 dex in scatter or systematic offset of satellite populations from central populations is acceptable, before agreement with the observed clustering on small scales breaks down (see also Moster et al. 2010b; Guo et al. 2009; Tinker & Wetzel 2010; Pasquali et al. 2010; Weinmann et al. 2009b).

The conclusion is not strongly redshift-dependent at least for redshifts $z = 0 - 4$ where clustering and stellar mass function measurements are available (and observations of other populations even at $z \sim 6$ suggest it remains true there as well; White et al. 2008; Shankar 2009; Shen et al. 2010). This has also been found in a number of other studies (Zheng et al. 2005, 2007; Yang et al. 2008; Tinker et al. 2010; More et al. 2009; Moster et al. 2010b).

For example, Wang et al. (2006) arrive at similar conclusions with independent data sets, simulations, and different fitting methodologies. They find reasonable agreement (especially at low masses, which is the regime of particular interest here) assuming that centrals and satellites obey an identical M_{halo} distribution. They do obtain a marginal improvement for high-mass galaxies if the preferred satellite $M_{\text{gal}}(M_{\text{halo}})$ is offset from the central $M_{\text{gal}}(M_{\text{halo}})$ by a small amount, shown in Figure 12; satellites might be systematically less massive at the same halo infall mass by a factor ~ 1.2 around $\sim L_*$ and ~ 2 at $L \gg L_*$. But differences in the fitted *scatter* actually

“cancel out” much of the predicted offset, and the statistical difference at low masses is small.²¹

Interestingly, cosmological hydrodynamic simulations predict a nearly identical $M_{\text{gal}}(M_{\text{halo}})$ for satellite and central galaxies, in agreement with the HOD model fits (including the assumption that the $f_{\text{gas}}(M_*)$ relation is unchanged for satellites and centrals, and therefore that there is no difference either in $M_*(M_{\text{halo}})$ or $M_{\text{gal}}(M_{\text{halo}})$). The reasons for this are discussed below (§ 4.2.3).

We compare these empirical HOD constraints with the predictions of the semi-analytic model from de Lucia & Blaizot (2007) (the results are similar for other SAMs shown, we show this as a representative case). Since we are interested in effects on mergers, we specifically isolate populations in the “orphan” stage, i.e. for which the “merger clock” in the model has been initialized and for which the remaining time until merger is less than the Hubble time.²² In this model, and in many other SAMs, there is actually a significant difference between the two $M_{\text{gal}}(M_{\text{halo}})$ relations (especially at

²¹ Note that the simulations used in Wang et al. (2006) require the inclusion of an “orphan” (unresolved subhalo but un-merged) population in order to match the observed small-scale clustering: repeating the analysis with subhalo populations in some other simulations (e.g. Stewart et al. 2009a) that require a smaller orphan population yields a smaller satellite-central offset.

²² This is not exactly the same as the satellite population in the halo occupation models constrained as described above. However, in the HOD models, the satellite mass as a function of infall mass is by definition preserved at smaller radii, and comparing the systems constrained by observed clustering at small radii $< R_{\text{vir}}$ (let alone $\ll 100$ kpc), which are near-merger, they are almost all in the orphan population in these particular SAMs. Regardless, the selection from the SAMs does not qualitatively change our conclusions.

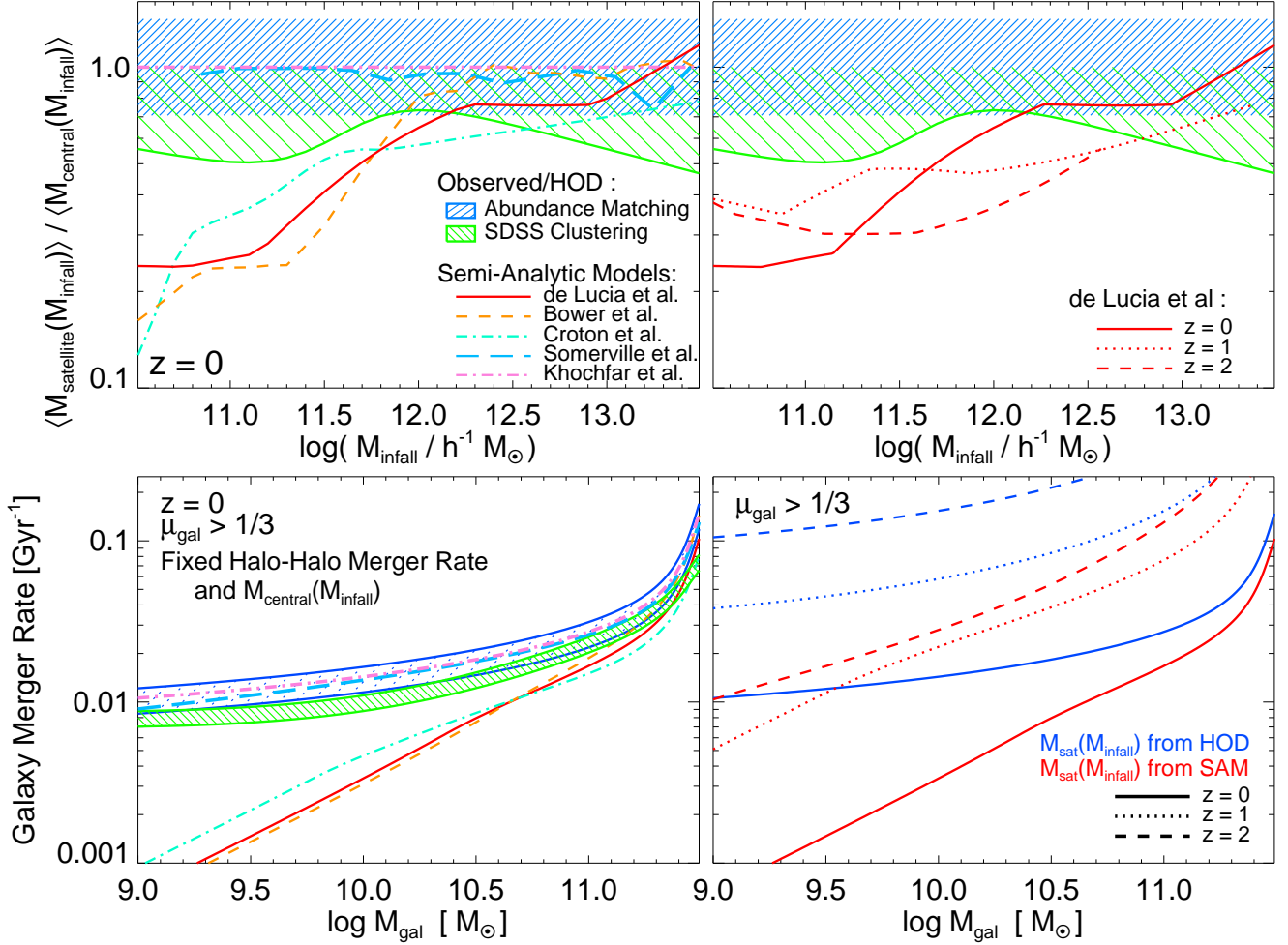


FIG. 12.— Possible effects of allowing the *satellite* HOD to differ from the *central* HOD by a significant amount. *Top Left*: Average satellite versus average central galaxy mass (total baryonic mass, not stellar mass) at fixed maximum or infall halo mass (for the centrals, just M_{halo} , for the satellites, their maximum pre-accretion mass – i.e. halo mass when they were themselves last a central galaxy). A model where satellites were totally un-stripped would have this = 1 exactly. Blue shaded range shows the assumed range of systematic offsets in the abundance-matching HOD models matched to abundance and clustering versus *stellar* mass following Conroy & Wechsler (2009); the same $f_{\text{gas}} - M_*$ relation is applied to all systems (i.e. no gas stripping allowed). By assumption, the typical HOD model has no offset. Green range show the same, but fitting both $M_* - M_{\text{halo}}$ relations for satellites and centrals separately, to local SDSS clustering measurements from Wang et al. (2006). We compare the predictions from various SAMs (labeled), at $z=0$. Since we are interested in satellites “about to merge,” we restrict the SAM ‘satellites’ to those in the ‘orphan’ stage with a remaining merger time $< t_{\text{Hubble}}(z)/2$ (by assumption, this gives the same result in the HOD models, but is not directly constrained by observations); differences are much smaller when all satellites are included (or when stellar, not baryonic, mass is plotted). *Top Right*: Same, comparing the prediction from de Lucia & Blaizot (2007) at three redshifts. *Bottom Left*: Resulting ($z=0$) galaxy-galaxy major merger rate (mergers per galaxy per Gyr, for galaxies with mass $> M_{\text{gal}}$), from convolving halo-halo merger rates with the above HOD for both centrals and satellites. In each, the halo-halo merger rate is held fixed (Fakhouri & Ma 2008), as is the central galaxy $M_{\text{gal}}(M_{\text{halo}})$ relation (from abundance matching following Conroy & Wechsler 2009); differences stem from the predicted difference in satellite-central properties at fixed infall mass. Linestyles correspond to the models above. *Bottom Right*: Same, comparing just the abundance-matching prediction (blue lines) and de Lucia & Blaizot (2007) model at $z=0, 1, 2$. In low mass, gas-rich galaxies, the difference between models with all gas retained (the HOD case above; an upper limit) and satellite gas stripping can translate to significant differences in predicted *baryonic* major merger rates.

low masses), owing to satellite-specific physics that lower $M_{\text{gal}}(M_{\text{halo}})$ for satellites relative to its value at the time of infall/accretion. Semi-analytic models tend to predict that low-mass satellites (especially those about to merge – the population of interest) have substantially lower baryonic masses than central galaxies with the same halo mass at the time of the satellite accretion (“infall mass”). The baryonic mass difference is about a factor of $\sim 2 - 3$ at the lowest masses, where disks are gas-dominated.²³ We consider the difference in the predicted distribution ($M_{\text{gal}}(M_{\text{halo}})$) for satellites and centrals

²³ Note that this distinction does not necessarily appear in the stellar mass as well (S. White, private communication), but we focus on baryonic masses here since it more appropriately focuses on the relevant dynamical major merger events (but see § 5.1 below).

at higher redshifts ($z=1$ and $z=2$) as well, and find that the gap between centrals and satellites, in the models, extends to higher masses at higher redshifts.

Given this, we can estimate how large an effect the different satellite-versus-central distributions will have on merger rates. As above, we can compare the galaxy-galaxy merger rate on an even footing by adopting a fixed halo-halo merger rate (here that in Fakhouri & Ma 2008), and populating each halo according to the $M_{\text{gal}}(M_{\text{halo}})$ relation. Unlike in the previous section, however, we do not assume that both galaxies obey the same $M_{\text{gal}}(M_{\text{halo}})$ relation. Rather, the primary and secondary subhalos separately follow the relations for central and satellite galaxies, shown here.

Recall that differences between semi-analytic and semi-

empirical estimates of the $M_{\text{gal}}(M_{\text{halo}})$ distribution for *central* galaxies have relatively little effect on merger rates. Thus, the variations in merger rates reflected in Figure 12 almost entirely owe to the *difference* (or lack thereof) between the $M_{\text{gal}}(M_{\text{halo}})$ of satellites and that of centrals. This is in contrast to the HOD-based model in which the $M_{\text{gal}}(M_{\text{halo}})$ is nearly the same for centrals and satellites, leading to the same curves predicted in § 4.1. In the SAM, the median $M_{\text{gal}}(M_{\text{halo}})$ can be significantly lower for satellites at the same infall/pre-accretion mass (especially at low masses). Thus even for a merger of exactly equal-mass halos (which, on average, must have equal-mass galaxies just before the halo-halo merger, i.e. at infall), the median expectation in the SAM is that the “secondary” (whichever happens to have slightly less initial mass) will be only some fraction of the mass of the primary by the time the actual galaxy-galaxy merger occurs, if it were initially gas-rich. The number of major mergers can therefore be substantially suppressed.

It is clear how a larger discrepancy between $M_{\text{gal}}(M_{\text{halo}})$ of satellites versus centrals at low masses leads to a stronger suppression of the merger rate. Likewise, at high redshifts, if the gap between central and satellite masses grows, the major merger rate will be further suppressed. At high masses, though, and in regimes where galaxies are gas-poor, both SAM and HOD constraints find a similar result where the $M_{\text{gal}}(M_{\text{halo}})$ distributions of satellites and centrals are nearly identical – as a result, their predicted merger rates are not dramatically different.

The results from the Bower et al. (2006) and Croton et al. (2006) semi-analytic models are qualitatively similar. In contrast, the semi-analytic models of Somerville et al. (2008a) and Khochfar & Silk (2009) predict little discrepancy between the $M_{\text{gal}}(M_{\text{halo}})$ relations of satellites and centrals. This is because they do not include any mechanism for the satellite galaxies to lose a significant mass in gas before their final merger; we discuss this further below.

4.2.2. *Effective Mass Loss in Satellites: What Drives These Differences?*

Satellite-specific physics in the semi-analytic models are clearly important for the predicted merger rates at low masses. Why do some models predict different $M_{\text{gal}}(M_{\text{halo}})$ distributions for satellites versus central galaxies?

In general, the predicted differences in $M_{\text{gal}}(M_{\text{halo}})$ between satellites and centrals relate to a well-known tendency for semi-analytic models to predict a satellite population that is “over-quenched.” Satellites in these models tend to be too red, too gas-poor, too low-mass, and insufficiently star-forming relative to observed satellite galaxies. Figure 13 briefly illustrates this effect: we show first the observed fraction of passively evolving (“quenched”) central and satellite galaxies, as a function of galaxy stellar mass and halo mass. We compare these observations to the predicted passive fractions in semi-analytic models.

The observational results are taken from Kimm et al. (2009), from their analysis of the Yang et al. (2005b, 2008) SDSS ($z = 0$) galaxy catalogs, with the passive population identified by the observed bimodal distribution of specific star formation rates at each mass. The observations show that, at fixed stellar and halo mass, the fraction of satellite galaxies that have “quenched” is not much different from that of central galaxies. At low stellar masses, in particular, *most* satellites are still star-forming (even those satellites in high-mass halos). Similar results have been found by other analyses

and in a wide range of different surveys, over a range of redshifts from $z = 0 - 1$ (see e.g. Gerke et al. 2007; Haines et al. 2007; Weinmann et al. 2006a,b, 2009a; Wang et al. 2007; Yang et al. 2008; Guo et al. 2009; van den Bosch et al. 2008). This agrees well with the expectation from semi-empirical models and HOD constraints (independently determined via the observed small-scale clustering of galaxies), that (at least massive) satellites should have properties that reflect their halo when it was last a central, but otherwise differ little from central galaxies.

This is also reflected in galaxy gas content, shown in Figure 13. We show the observed gas fraction distribution of central and satellite galaxies (separately), as a function of stellar mass. We specifically show compilations of observed disk gas fractions as a function of their stellar mass from Bell & de Jong (2000), McGaugh (2005), and Kannappan (2004). These observations span both satellite and central galaxies. The sample of Kannappan (2004), being simply mass-selected, is dominated by central galaxies at all masses. But from the other samples we can separate both central (chiefly massive field) galaxies and satellite galaxies, in particular the Ursa Major Cluster sample of Tully et al. (1996). This is ideal for our comparison, as it represents a rich, but not extremely massive (and thus unusual) group with a group velocity dispersion of $\sim 150 \text{ km s}^{-1}$ – this is typical of the host halos of central/primary galaxies with $M_{\text{gal}} \sim 10^{10} - 10^{11} M_{\odot}$. We stress that we are *not* trying to compare with galaxies in e.g. massive clusters, where the gas-rich galaxies tend to be much less massive than their cluster host – stripping and other processes may be efficient in such galaxies, but those would not represent major mergers in *any* model. This is also seen in many of the color or SFR-based studies above; for example, Weinmann et al. (2009a); Guo et al. (2009); Wang et al. (2009); van den Bosch et al. (2008); Kang & van den Bosch (2008) find that only in the most extreme (rare) environments could there be significant gas depletion in satellites – over a wide range of satellite and halo masses, they find that there is at most very minor depletion (for major-merger candidates) and, in particular, that the amount of depletion does not increase or decrease with halo mass in any significant manner. This is true even in the central regions of groups, i.e. systems “about to merge,” for which the predicted satellite-central gas mass difference in some models is maximized (see references above). The important result observed is, for galaxies with similar M_* in similar-mass halos, which have some possibility of being major merger pairs, there is not much difference in gas fractions between central and satellite systems.

We compare both of these diagnostics – the “quenched” fraction and gas mass fractions in central and satellite galaxies – to the predictions of the semi-analytic models. The differences are immediately apparent. The SAMs predict that essentially all satellites are quenched, especially those in halos above the critical “quenching mass” of $\sim 10^{12} M_{\odot}$. The average predicted gas fraction of satellites at low mass is much less than that of equivalent central galaxies, in contrast to the observations.

For clarity, we show in Figure 13 just the predictions from the de Lucia & Blaizot (2007) model, but Kimm et al. (2009) and others have demonstrated that this satellite “overquenching” tendency is quite general in recent SAMs, including e.g. those of Croton et al. (2006), Monaco et al. (2007), Bower et al. (2006), Kang et al. (2005), Cattaneo et al. (2006), and Bertone et al. (2007).

As noted above, we are re-stating a well-known result; the

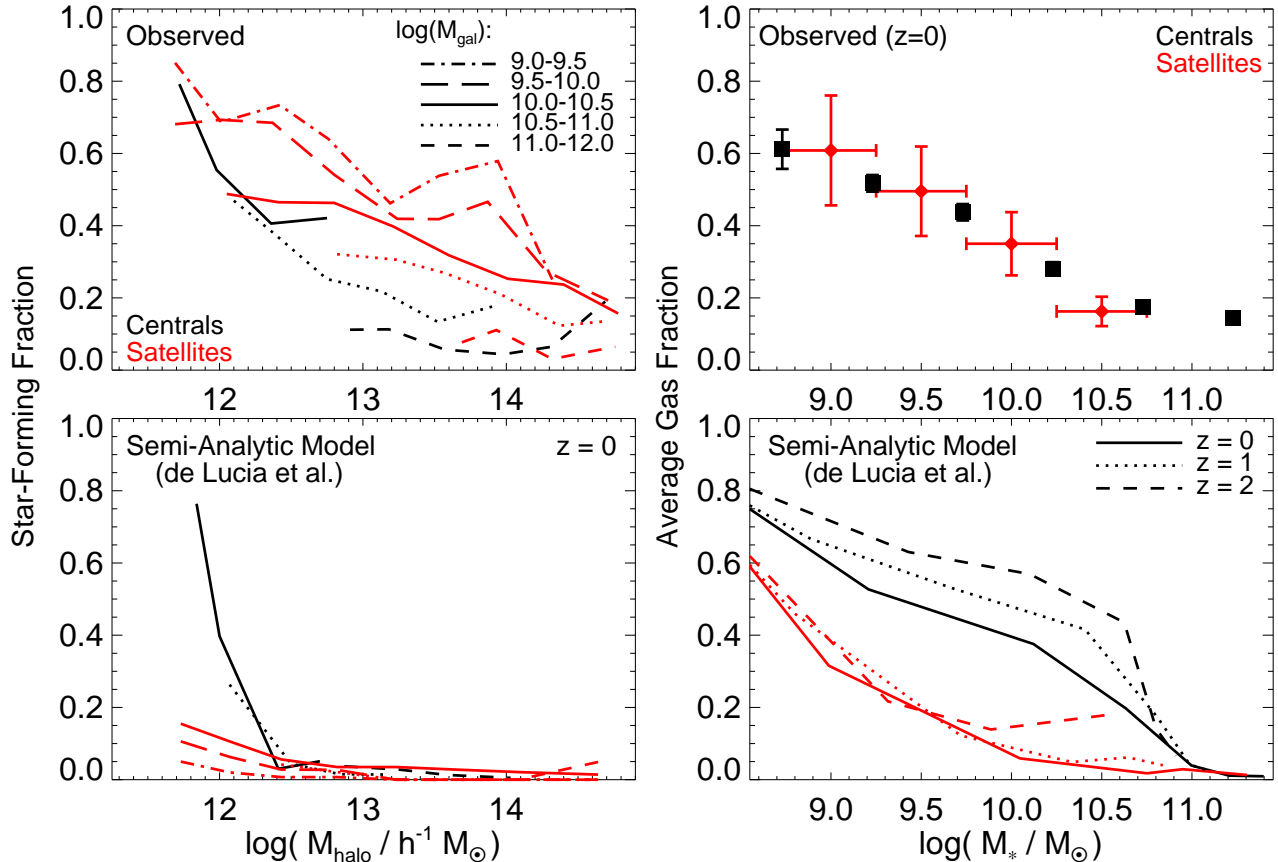


FIG. 13.— Illustrations of the effect seen in Figure 12, where low-mass satellites in many SAMs have suppressed baryonic masses. *Left*: Fraction of satellite and central galaxies that are star-forming (at $z=0$), as a bivariate function of galaxy stellar and halo mass. *Top Left*: Observed distributions from Kimm et al. (2009). Low-mass satellites have large star-forming fractions, similar to central galaxies. *Bottom Left*: Prediction from the de Lucia & Blaizot (2007) model. Satellites tend to be uniformly quenched. *Right*: Average gas fractions (and scatter in gas fractions) versus galaxy mass, for satellites and centrals. *Top Right*: Observed distributions at $z=0$ for field centrals (Kannappan 2004), and satellites in “normal” groups (not massive clusters, where satellite properties are significantly different; Bell & de Jong 2000). Barring extreme (rare) environments or very minor mass ratios, satellites are not strongly gas-depleted. *Bottom Right*: Prediction from the SAM for satellites and centrals. Model satellites are strongly gas-depleted. These are general trends in a number of semi-analytic models.

same physical deficiencies are reflected in a number of other measures: in semi-analytic models, satellite mass functions tend to be under-massive, the fraction of blue satellites is lower than that observed, and the observed small-scale clustering of star-forming or blue galaxies tends to be poorly reproduced (see references above and Cooper et al. 2006, 2007; Park et al. 2007; Blanton & Berlind 2007; Li et al. 2006; Coil et al. 2008).

What, in these models, drives the “overquenching” of satellites? In most, once a system becomes a satellite, its halo gas reservoir is immediately added to that of its “parent” halo, and (as a consequence) no new cooling occurs onto the secondary. But if this removal of un-cooled, pristine halo gas were the only effect, the consequences would not be so pronounced. In all of these models, strong stellar feedback is assumed to affect star-forming galaxies – generally some mass loss of order several times the star formation rate is ejected from the disk, per unit stars formed. Given such strong mass ejection, in order to maintain observed (high) star formation rates in disk galaxies, the ejected gas must be recycled rapidly – either processed through the halo or some (for numerical reasons) designated “reservoir” and returned quickly to the galaxy. But when a system is a “satellite” in these models (even if the mass ratio is close to 1:1), the mass is simply lost to the par-

ent halo. The gas mass is therefore quickly depleted and the satellite quenches as it has ejected most of its gas.²⁴

This can be a very large effect. A system with $\sim 50\%$ gas will lose $\sim 1/2$ its baryonic mass in just a short time after accretion. Thus an initial 1:2 major merger will become a 1:4 minor merger (or a more minor merger, if the parent “steals” this ejected mass). As a consequence, there will be almost no very *gas-rich* major mergers.

This problem is further exacerbated in many of these models by the assumption of instantaneous recycling from stellar evolution. When some stars form in the new satellite, the mass loss of $\sim 50\%$ that will occur over their lifetime is immediately turned into gas, which can then suffer the consequences above. A more gradual mass loss over actual stellar lifetimes would, obviously, yield somewhat less rapid an evolution (especially for major mergers, where the merger timescale might be only a Gyr).

Since the consequences of this gas loss will clearly be stronger for systems with higher gas content, low-mass galaxies are preferentially affected. This gives rise to the mass-dependent trend in the discrepancy between $M_{\text{gal}}(M_{\text{halo}})$ pre-

²⁴ This is despite the fact that many of the models do not nominally include an additional explicit “ram-pressure stripping prescription” for the cold disk gas.

dicted for satellite and central galaxies by such models (shown in Figure 12), where low-mass satellites have their masses more suppressed relative to centrals. Likewise, since at high redshifts all galaxies are more gas-rich, the discrepancy grows in the models with redshift. This, in turn, suppresses predicted merger rates at low masses and high redshifts.

Differences between different semi-analytic models largely depend on exactly what prescription is used for stellar feedback (controlling how efficiently the system self-quenches) as well as e.g. when and where this occurs and how it is connected to the lifetimes/stripping/identification of subhalos/satellites. For example, in some models based on cosmological dark matter merger trees (e.g. de Lucia & Blaizot 2007; Bower et al. 2006), the system is labeled a “satellite” when the halos are first linked by the friends-of-friends (FOF) halo finder algorithm, which occurs for major pairs at anywhere from $\sim 2 - 5 R_{\text{vir}}$. In such a case, these effects can “strip” the secondary even before it enters the primary halo. In contrast, simulations and observations find that, if any stripping occurs, it often only does so in the central, most dense and high-pressure regions of massive halos (Weinmann et al. 2006a,b; Wang et al. 2007; Yang et al. 2009; Weinmann et al. 2009a; van den Bosch et al. 2008; Kang & van den Bosch 2008; McCarthy et al. 2008). In fact, cosmological hydrodynamic simulations show that satellite galaxies tend to keep growing and accreting *new* material (both onto their subhalo and into the galaxy) until they are in the final stages of orbiting a relaxed massive system near its center (Simha et al. 2009). Observations likewise see no significant difference between satellites in the outer regions of even massive groups, relative to centrals in similar-mass subhalos. The tendency of satellites to continue accreting in simulations is stronger at higher redshifts, and observational constraints from $z = 0 - 2$ require that the gas-rich, star-forming satellite fraction remain large over this entire interval (Tinker & Wetzel 2010), enhancing the discrepancy with these models at high redshift.

Unsurprisingly, at high masses, where gas fractions tend to be small ($\sim 10 - 20\%$), these effects are minimized. Thus the $M_{\text{gal}}(M_{\text{halo}})$ distributions for satellites and centrals are similar, and there is little difference in predicted merger rates at these masses.

4.2.3. Towards Resolution: What Physics for Satellites?

It is important to recall that the vast majority of mergers by both number and mass density are in the field (essentially single halo-halo mergers) or small/loose group environments (analogous to e.g. the Local Group). The observations indicate that satellite-specific physics appear to be important only for systems with the smallest mass ratios $\mu_{\text{gal}} \lesssim 0.01 - 0.05$ in these environments (specifically see e.g. Haines et al. 2007; Weinmann et al. 2009a; Kimm et al. 2009), and then only in the final stages and/or close passages, *not* just at the moment of crossing R_{vir} (references above and Yang et al. 2009; van den Bosch et al. 2008; Kang & van den Bosch 2008). Even if stripping were extremely efficient in dense environments (e.g. for clusters, or for dwarf galaxies in massive halos such as the Local Group), such that it suppressed all such major mergers, it would have almost no effect on the *overall* merger rate.

In contrast, much of the intuition used to design the prescriptions above is built up from satellite-specific studies based on massive clusters with dense, high-pressure ICM gas. But such systems contain only a small fraction of

the bulge mass density of the Universe ($\lesssim 5\%$; see e.g. Carollo et al. 1998; Kormendy & Kennicutt 2004; Allen et al. 2006; Ball et al. 2006; Driver et al. 2007; Gadotti 2009).

Indeed, with the possible exception of dwarf galaxies in massive halos (D’Onghia et al. 2009; Wang et al. 2009), the observations favor a “gentle” treatment of satellites. Considering a volume-limited sample of satellites (i.e. one not dominated by extreme environments) that might become major mergers (i.e. are not so different in mass from their centrals), the observed specific star formation rates, colors, and gas fractions can be well-fitted (to first order) by a model in which there is *no* removal of the hot or cold satellite gas (including the “ejecta” or recycled material from stellar winds and supernovae). The only constraint is that the satellite halo stops growing, so e.g. the total baryonic mass supply is “frozen in” at the value it had when the subhalo was accreted.

Such a simple model leads to a color and passive/quenched galaxy distribution as a bivariate function of galaxy mass and halo mass similar to that of centrals (with just slightly more “quenching” in low mass satellites), in better agreement with observations (Weinmann et al. 2006a,b; van den Bosch et al. 2008; Kang & van den Bosch 2008; Kimm et al. 2009), and similar gas fractions (as the subhalo retains ejected stellar wind feedback), as observed (Bell & de Jong 2000; McGaugh 2005; Kannappan 2004). This also naturally leads to the observed result that satellite quenched fractions are a much stronger function of the satellite stellar mass, and weaker function of e.g. the local density or halo mass, than currently predicted (Weinmann et al. 2006b; Haines et al. 2007), at both $z = 0$ and high redshifts $z \sim 1 - 1.5$ (Gerke et al. 2007). And the observed weak dependence of red satellite fraction on distance from halo center (Weinmann et al. 2006a) is similarly obtained. The observationally favored gradual decline of satellite star formation rates relative to those that such systems had when they were central galaxies (see e.g. Wang et al. 2007) arises naturally in an effectively “closed-box” model, given the observed star formation rate-gas density scaling laws (Kennicutt 1998), as systems deplete their original baryon reserve without new halo growth.

High-resolution hydrodynamic simulations appear to give consistent results. Gas-rich systems are in the “cold accretion” regime: their halos are dense, with a cooling time shorter than the local dynamical time, so gas outflows may be efficiently halted by the dense ISM and more gas in new accretion is brought into the galaxy in cold filaments on the dynamical time. Especially in this regime, these cold flows dominate accretion – the idea of spherical accretion from a quasi-static halo accounts for only a small amount of residual cooling in the most massive halos (essentially, this is the old “cooling flow” problem in massive clusters; but it contributes negligibly to the global cold gas budget or star formation rate density). These cold filaments are sufficiently dense to survive disruption by the diffuse parent halo gas when systems become satellites; the simulations find that even relatively small satellites continue to hold onto their halo gas.

In fact, the simulated satellites continue to accrete new material, and grow by cooling, well into their parent halo virial radius and for $\sim \text{Gyr}$ after become satellites (see e.g. Simha et al. 2009). In this regime, satellite growth in the simulations is indistinguishable from central galaxy growth, for the same halo/infall mass. At high redshift, the simulations suggest that even the “frozen in” models may quench satellites too severely; subhalos in the simulations continue to accrete new halo gas at a rate equal to the central galaxy well

into the halo. In effect, it is as if the system does not “know” it is a satellite until at very small radii from the primary center, just before the final galaxy-galaxy merger (Simha et al. 2009; Kereš et al. 2009b,a).

Recall also that we are ultimately most interested in the case of major mergers: in these cases, it is clear that the distinction of “satellite” versus “central” is somewhat semantic. Physically, the two are mutually gravitating (one is not moving in the halo or potential of the other; rather the two mutually coalesce in a single halo/potential) and (by definition) have similar halo properties (and therefore similar halo gas masses, virial temperatures, etc.). Simple physical considerations (and simulations) imply that there should be no significant difference in the cooling/mass loss of one system relative to the other (let alone ram-pressure stripping of one and not the other).

Of course, it is easier to point out these differences than to design physical prescriptions that effectively model them – let alone prescriptions that are robust across the entire regime of interest (from dwarf galaxies and/or small satellites in massive clusters to equal-mass pairs in field or Local Group-analogue environments). And as noted above, it is not enough to simply retain pristine halo gas associated with the satellite – ejected gas from stellar feedback should be recycled in some manner appropriate to how it is implemented in the different models.

Considerable progress is being made in this area. For example, Font et al. (2008) updates the Bower et al. (2006) model, with the primary change being a more detailed satellite cooling model. Specifically, the models are otherwise identical, but instead of removing subhalo gas instantly (as in Bower et al. 2006), the revised model strips the gas from the assumed extended subhalo gas halo according to the model in McCarthy et al. (2008), as a function of e.g. the local gas density (itself a function of radius from the primary, for assumed isothermal sphere gas profiles). There are still some sources of uncertainty in e.g. the exact wind treatment and efficiency of subhalo gas stripping (and the prescription adopted may not be applicable if most accretion comes via cold flows), but the agreement with the star-forming fractions of satellite galaxies is substantially improved.

Figure 14 compares the merger rates predicted by this revised model to those from the previous Bower et al. (2006) iteration of the model (with strong satellite quenching). We also compare the rates determined from HOD-based approaches, and observational constraints. At high masses where systems are gas-poor and not star-forming, all the models (and observations) agree well. At low masses, where galaxies are gas-rich, however, there is an order-of-magnitude difference in the predicted merger rates. The revised Font et al. (2008) yields merger rates more similar to the HOD-based models, as expected, since the primary adjustment to the model is intended to bring it into better agreement with the observed HOD statistics used as input in the HOD-based models. This yields significantly improved agreement with observational estimates of the merger rates. There are clearly remaining differences; however, they are at the factor $\sim 2 - 3$ level, typical as we have shown of the other differences between the models.

This is reinforced by comparison of Figures 2 & 12. Besides the Font et al. (2008) model discussed above, the two semi-analytic models that predict merger rates in best agreement with the halo occupation models, over the entire galaxy mass range, are those from Khochfar & Silk (2009) and Somerville et al. (2008a). These models do not include any

mechanism for cold gas (including that affected by stellar feedback) to be removed from the satellite before a merger; the Somerville et al. (2008a) model allows for some stripping of hot gas, but only inside of the parent virial radius where the timescale for a major merger is short and so there is little effect on the satellite mass. In Figure 2, we showed that these models are precisely those which predict little or no discrepancy between the $M_{\text{gal}}(M_{\text{halo}})$ relation of satellites and centrals. This does not mean that there are not other, perhaps very significant, differences between the models (nor does it necessarily mean that the same answer is being obtained for the correct reasons); but it does imply that, controlling for differences in predicted satellite properties removes a significant factor driving different predictions for merger rates.

5. THE OTHER SIDE: DEFINITIONS OF MASS AND EFFECTS OF MERGERS ON MORPHOLOGY

5.1. The Definition of Mass Ratio

Throughout this paper, we have focused on the “merger rate,” with respect to mass ratios defined in a specific manner. Observationally, of course, some of these are more or less accessible (the galaxy-galaxy luminosity or stellar mass ratios being most so, the total dark matter halo-halo mass ratios least so). But *physically*, the same “number” of mergers will have very different implications for galaxy growth, morphologies, bulge formation, and star formation, depending on the mass ratio involved – and the merger rate is *not* the same for different definitions of the mass ratio.

Figure 15 illustrates some of these caveats. First, consider the merger rate itself, within the same model, but given different definitions of merger mass ratio. There are three obvious choices of mass ratio, commonly used: galaxy-galaxy stellar mass ratio ($\mu_* = M_{*,2}/M_{*,1}$), galaxy-galaxy baryonic (stellar+cold gas) mass ratio ($\mu_{\text{gal}} = M_{\text{gal},2}/M_{\text{gal},1}$), and halo-halo mass ratio ($\mu_{\text{halo}} = M_{\text{halo},2}/M_{\text{halo},1}$). We have shown how the shape of the $M_{\text{gal}} - M_{\text{halo}}$ relation means that the number of major mergers in terms of μ_{gal} can be quite different from the number of mergers in terms of μ_{halo} . But since galaxy gas fractions are not uniformly constant, μ_{gal} and μ_* will also not be the same.

In Figure 15, we plot the galaxy major merger rate (major mergers per galaxy per Gyr) at $z = 0$, as a function of galaxy stellar mass. We do so for a single model – the “default” model in Hopkins et al. (2010) using the halo-halo merger rates from Fakhouri & Ma (2008); Fakhouri et al. (2010) and abundance-matching HOD method from Conroy & Wechsler (2009). But the results are qualitatively the same regardless of model. We show the major merger rate for different definitions of a “major” merger: a merger with halo-halo mass ratio $\mu_{\text{halo}} > 1/3$, with galaxy stellar-stellar mass ratio $\mu_* > 1/3$, and with galaxy baryonic-baryonic mass ratio $\mu_{\text{gal}} > 1/3$. The halo-halo rate is only weakly dependent on mass, a result well-known from cosmological simulations (see § 3.2.1). The stellar-stellar rate, on the other hand, depends strongly on mass – this owes to the shape of $M_*(M_{\text{halo}})$. At low masses, galaxy stellar mass is a steep function of halo mass, suppressing major mergers ($M_* \propto M_{\text{halo}}^2$, so a 1:3 halo-halo merger is just a 1:9 stellar-stellar merger); at high masses, stellar mass is a shallow function of halo mass, enhancing major mergers ($M_* \propto M_{\text{halo}}^{0.5}$, so a 1:9 halo-halo merger is a 1:3 stellar-stellar merger). The effect is very large – an order of magnitude at the extremes of the mass function – and cannot be neglected. The baryonic-baryonic rate is quite similar to the stellar-stellar rate at high masses (not surprising, since high-

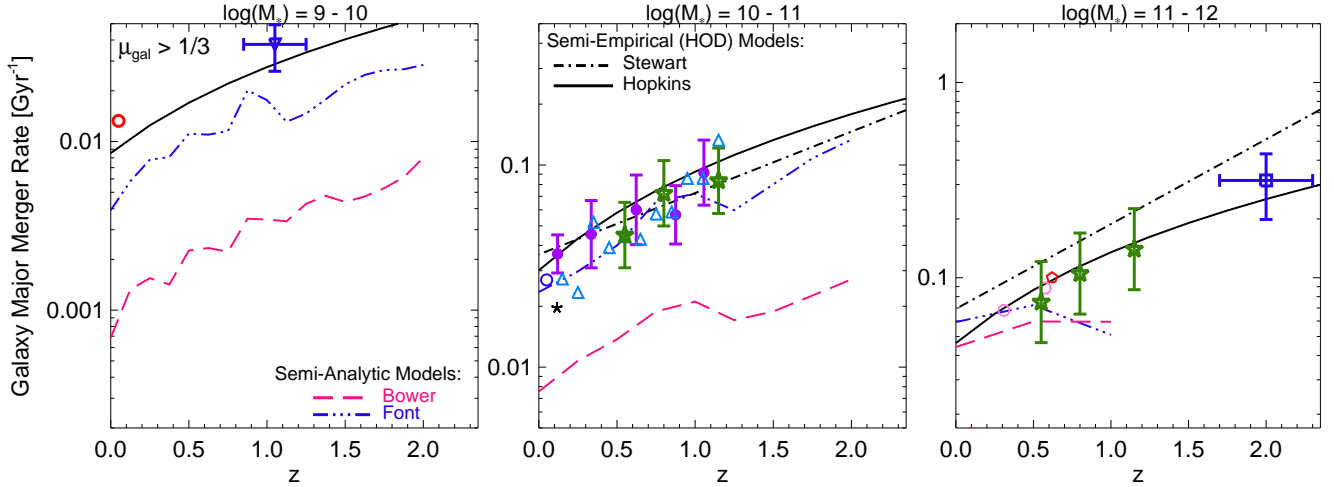


FIG. 14.— Comparison of the predicted major galaxy merger rates (as Figure 2) in two similar semi-analytic models, Bower et al. (2006) and Font et al. (2008). The SAMs are otherwise identical, but Font et al. (2008) implement a more physical, less severe satellite stripping/quenching model, improving the agreement with e.g. the observations in Figures 12-13. We compare the predictions to estimates directly adopting the observational HOD constraints, and to observed merger rates. Rates at high M_{gal} , where cooling/gas fractions are negligible, are unchanged. Rates at low masses are significantly higher in the revised model, as expected from Figure 12, and agree more closely with direct HOD models.

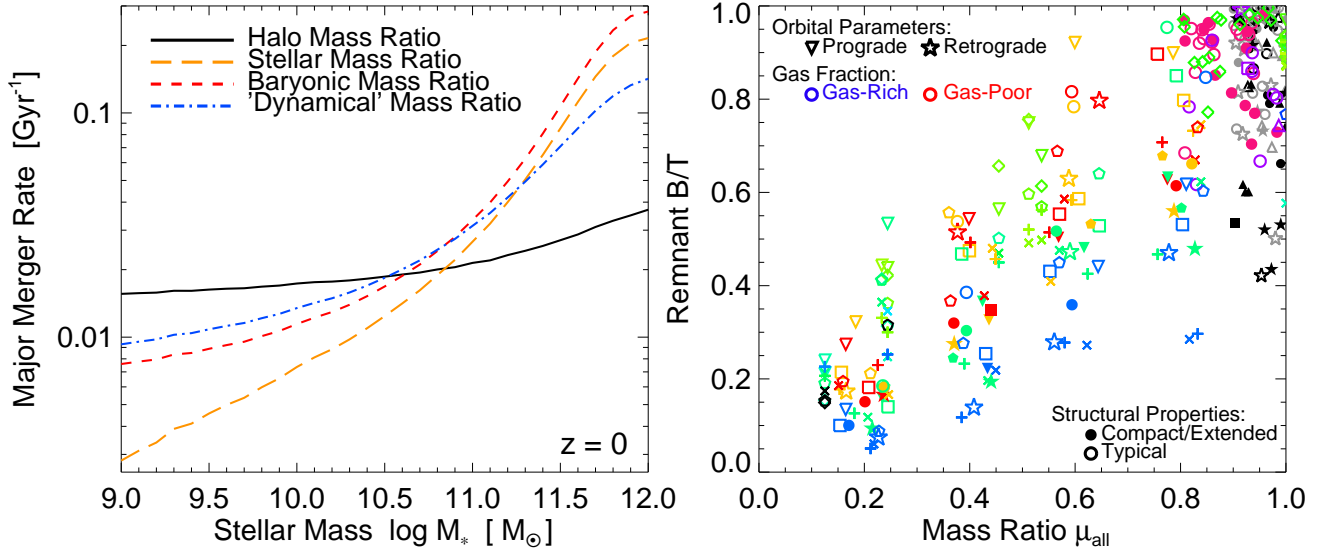


FIG. 15.— *Left*: Importance of the definition of “mass ratio” for merger rates. We show the major ($\mu > 1/3$) merger rate per galaxy at $z = 0$, for galaxies at a given stellar mass, in terms of the merger halo-halo mass ratio, galaxy stellar-stellar mass ratio, baryonic-baryonic (stellar+cold disk gas) mass ratio, and “dynamical” mass ratio (gravitational mass, i.e. baryons plus the tightly bound, non-stripped dark matter inside the baryonic radii; see Equation 4). The different ratios behave very differently as a function of mass. Halo mass ratios are most easily predicted, but very different from galaxy mass ratios. The baryonic or dynamical mass ratio is most relevant for e.g. morphological disturbances and bulge formation, but difficult to directly observe. Stellar mass ratios are accessible in e.g. pair samples, but under-estimate the rate of “damaging” major mergers at low masses by factors ~ 3 . *Right*: Impact of mergers in high-resolution hydrodynamic simulations from Hopkins et al. (2009a). We show the remnant bulge-to-total mass ratio after a merger with mass ratio μ ; to remove ambiguity in the merger mass ratio and rate, all systems have just one merger, with the systems initialized so that $\mu = \mu_* = \mu_{\text{halo}} = \mu_{\text{gal}} = \mu_{\text{dyn}}$. We show simulations with different orbital parameters (symbol types), gas fractions (color, from $f_{\text{gas}} = 0$ in red to $f_{\text{gas}} = 1$ in black), and structural parameters (variation in disk sizes, initial B/T , and halo concentrations in solid). Other parameter variations include surveys of stellar wind models (magenta), black hole feedback (violet), and gas equations of state (grey). At fixed, perfectly-defined μ , there is factor of $\sim 2 - 3$ variation in the amount of bulge formed by mergers owing to these parameters – neglecting their full distributions will inherently limit any prediction of the “impact” of mergers. Merger impact is also clearly continuous across the major/minor ($\mu = 1/3$) distinction.

mass galaxies are uniformly gas-poor). At low masses, however, it lies somewhere between the stellar-stellar and halo-halo rate; this is because low-mass galaxies have increasing cold disk gas masses, even while their stellar mass falls rapidly with halo mass. Thus the total baryonic mass is not as steep a function of halo mass as is the stellar mass, and major mergers are less strongly suppressed (at low masses, $M_{\text{gal}} \propto M_{\text{halo}}^{1.0-1.5}$, so e.g. a 1:2 halo-halo merger remains a major baryonic-baryonic merger; see Avila-Reese et al. 2008;

Bell et al. 2003a; McGaugh 2005; Stark et al. 2009). The differences are again very non-trivial – at low masses, the baryonic-baryonic merger rate is a factor of 3 larger than the stellar-stellar merger rate. For further discussion of these distinctions, we refer to Stewart (2009).

These various ratios not only relate differently to observable quantities; they also have a number of physical relationships to the impact of galaxy mergers. Using halo-halo mergers as a proxy for galaxy-galaxy mergers is clearly not a good

approximation in either models or observations, except just at $\sim L_*$ where the mass function turns over. Physically, the halo-halo mass ratio is also not the most important quantity for e.g. bulge formation. This is because, by the time the final merger/galaxy coalescence takes place, most of the secondary halo mass will already have been stripped and/or mixed with the primary halo mass. In general, the outer parts of the halo (which contain much of the mass), being extended and low-density relative to the baryonic galaxy, are not able to induce significant perturbations or torques on that galaxy.

The stellar-stellar mass ratio has the advantage of being observable, as well as identifying at least a portion of the tightly bound material in the galaxies that can actually induce structural perturbations and bulge formation. However, at low masses or high redshifts, where gas fractions are non-negligible, stellar mass alone clearly misses a very important quantity – the amount of cold, rotationally supported gas in galaxy disks. Observations indicate that in these regimes the cold gas is gravitationally *dominant* over the stars in a large fraction of the galaxy population. From a gravitational perspective, in terms of the perturbation induced, it makes no difference whether the same merging disk is 100% gas or 100% stars (although the end effects will be different and gas may be distributed differently from stars; see Hopkins et al. 2009a), even though one would be a stellar-stellar major merger and the other would not appear in any observed merger catalog.

Thus, attempts to estimate the “impact” of mergers (e.g. the amount of bulge they might form), or even the total number of mergers a galaxy experiences, based on observed merger fractions identified with stellar mass or luminosity cuts, will under-estimate the true impact of mergers by a factor of at least ~ 3 at low masses ($M_* \lesssim 10^{10} M_\odot$). More directly sampling the total baryonic content of these galaxies is necessary to overcome this limitation; in the meantime, empirical results should be corrected for this distinction or at least recognize its very large importance at low masses. And gas fraction is critical for understanding the consequences of a given merger; the dynamics of mergers are fundamentally different and bulge formation less efficient in gas-dominated, as opposed to stellar-dominated, systems (see Hopkins et al. 2009a). At high masses, $> 10^{11} M_\odot$, at least at low redshifts $z < 1$, such a procedure does not produce a result very different from the baryonic-baryonic merger rate.

But even the baryonic mass ratio is not really representative of the gravitational mass that induces perturbations and forms bulges, heats disks, and scatters stars in mergers. What is actually desired is the surviving, tightly bound total gravitational mass of material that is involved in the last couple of passages in the final merger. In high-resolution simulations of a large number of mergers, spanning a large parameter space, Cox et al. (2008) find that this can be *roughly* approximated by the sum of the galaxy stellar mass, the cold disk mass, and the dark matter mass that is tightly bound to the baryons and thus will be able to resist stripping – approximately that inside a radius $\sim 2 - 4 R_e$ (where R_e is the effective radius of the baryonic galaxy) or $\sim 1 r_s$ (where $r_s = R_{\text{vir}}/c$ is the NFW scale radius of the halo, with c the halo concentration). Note that since this dark matter retention at small radii is largely a baryonic effect, it will not appear in collisionless simulations.

We therefore define a “dynamical” mass M_{dyn} , and corresponding mass ratio μ_{dyn} :

$$\begin{aligned} M_{\text{dyn}} &\equiv M_* + M_{\text{gas}} + M_{\text{DM}}(< 3R_e) \\ \mu_{\text{dyn}} &\equiv M_{\text{dyn},2}/M_{\text{dyn},1} \end{aligned} \quad (4)$$

This is a significantly better approximation to what is most important for bulge formation and the dynamical properties of galaxies than any of the mass ratios discussed above. Figure 15 shows the corresponding major merger rate, as a function of primary stellar mass, where a major merger is defined by $\mu_{\text{dyn}} > 1/3$ (to calculate the dark matter content inside $3R_e$, we assume the halos follow NFW profiles with a concentration-halo mass relation from Bullock et al. (2001), and that the galaxy R_e follows the relation as a function of stellar mass from Shen et al. (2003)). Roughly, the result is similar to that for baryonic mass ratio, but it is slightly closer to the halo-halo merger rate at both high and low masses (unsurprisingly, since it includes baryons plus a halo contribution).

Of course, this mass is difficult to extract observationally. However, it is readily accessible in most semi-empirical and semi-analytic models. Adopting such a definition for mass ratio, as opposed to the simpler definition of baryonic mass ratio, is important for predictions at low and high masses. In fact, many such models still adopt either just the stellar-stellar mass ratio or the halo-halo mass ratio as their proxy for the “impact” of a merger – this is not only clearly not the most physical quantity, but as Figure 15 illustrates, it can lead to an order-of-magnitude over or under-estimate of the consequences of mergers.

5.2. The Importance of Other Parameters

Even with perfect knowledge of the merger rate and distribution of mass ratios, however, we stress that the impact of mergers *cannot* be determined at better than factor ~ 2 accuracy, absent a large amount of additional information. Figure 15 illustrates this. We consider a large suite of high-resolution, hydrodynamic galaxy merger simulations. They include realistic progenitor galaxy models (gas+stellar disk+bulge+halo), cooling, star formation, and feedback from star formation and BH growth in a multi-phase ISM (Springel & Hernquist 2003a; Di Matteo et al. 2005). These are presented and discussed in detail in Hopkins et al. (2009a); for now, we use them to illustrate a couple of simple points. Although it may not be the most representative for galaxies at extreme masses, we wish to remove the already-discussed ambiguities in merger rates and mass ratio definitions. Therefore, every galaxy undergoes exactly one isolated merger with the simulations all initialized such that the “mass ratio” is the *same* by any of the definitions above, i.e. $\mu = \mu_{\text{halo}} = \mu_* = \mu_{\text{gal}} = \mu_{\text{dyn}}$. We run the simulations until they are fully relaxed, and then quantify the bulge-to-total mass ratio B/T in the remnant stars – the property most often desired in relation to mergers. We consider mass ratios $\mu = 0.1 - 1$, and vary a large number of other parameters in the simulations, including galaxy gas fractions, orbital parameters, feedback prescriptions, and structural properties of the initial disks, halos, and bulges.

First, it is very clear that there is no special “division” at the “major” merger threshold of $\mu = 1/3$. In many models, it is simply assumed that major mergers completely destroy disks, while minor mergers leave them intact. Likewise, in many observational studies, simple assumptions such as this are used to relate the number of major mergers to some “expected” number of bulges. But in fact, to lowest order, bulge formation is continuous, with the fraction of the initial disk destroyed (remnant B/T) scaling as $B/T \sim \mu$. This is discussed in greater detail in Hopkins et al. (2009a), with similar results found in a number of studies (Bournaud et al. 2005;

Johansson et al. 2009b; Younger et al. 2008), and some of the consequences are discussed in Hopkins et al. (2009b, 2010). Such a simple assumption regarding the efficiency of mergers can lead to systematic factors of $\sim 2 - 10$ differences in the total bulge mass formed (depending on e.g. the galaxy mass), and clearly skews the distribution of mergers impacting bulge formation.

Also clear in Figure 15 is a large scatter in remnant B/T at fixed μ , despite the fact that the mass ratio and number of mergers are *perfectly* well-known and defined for these simulations. This stems from the other parameters varied in the simulations. Most important are the orbital parameters and gas fractions. We consider a wide variety of different orbital parameters (denoted by different symbols). Because stellar scattering and the gaseous angular momentum loss that drives dissipation, starbursts, and the formation of the bulge from the galaxy gas are dominated by resonant processes, orbital parameters have a huge impact (Barnes & Hernquist 1991, 1996). It is well-known that prograde mergers tend to be much more destructive (form much larger bulges) than retrograde mergers, and we see that in Figure 15. We also see a continuum of behavior in between, as has been noted in many numerical studies (Hernquist 1989; Hernquist & Mihos 1995; Naab & Burkert 2003; Cox et al. 2008; Younger et al. 2008; Bournaud et al. 2005; Hopkins et al. 2009a; Johansson et al. 2009b,a). The effect is large, with e.g. a typical 1:3 merger leading to $B/T \approx 0.5$ in the most prograde mergers and just $B/T \approx 0.25$ in corresponding retrograde mergers. In Hopkins et al. (2009a), it is noted that gas loses angular momentum and falls to the galaxy center, contributing to a nuclear starburst and bulge mass, within an initial critical radius $R \propto 1/(1 - \psi \cos \theta)$, where $\psi \sim 0.5 - 0.6$ depends on structural properties and θ is the initial disk inclination ($\theta = 0$ is prograde, $\theta = \pi$ retrograde) – i.e. angular momentum loss is efficient out to a radius $\approx 3 - 4$ times larger in prograde cases, relative to retrograde cases.

Likewise, the gas fraction has a similar effect – the most gas-poor mergers lead to very efficient bulge formation while the most gas-rich mergers lead to very little bulge formation even in extremely major mergers. Figure 15 shows a large number of $\mu = 0.5 - 0.8$ cases with only $B/T \sim 0.2 - 0.3$, provided that the gas fractions at the time of the merger are high ($f_{\text{gas}} \sim 0.6 - 0.8$). Again, this has been seen in a number of works (Springel & Hernquist 2005; Robertson et al. 2006; Governato et al. 2009; Richard et al. 2010; Johansson et al. 2009b,a). In Hopkins et al. (2009a), these simulations are used to show that, to lowest order, the efficiency of gas angular momentum loss scales with a factor $\propto (1 - f_{\text{gas}})$ – i.e. in a merger with 50% gas, the ultimate gas angular momentum loss per unit mass is (on average) half as efficient as in a merger with just 10% gas.

Other parameters lead to further scatter in B/T at fixed μ . Varying galaxy structural parameters, for example, can stabilize or destabilize systems (Springel et al. 2005a). The presence of initial bulges in a certain mass and scale radius range suppresses gas inflows and disk heating on first passages (Mihos & Hernquist 1994, 1996). Allowing gaseous disks to be more extended than stellar disks leads to less efficient angular momentum loss in the gas (because that angular momentum loss is usually dominated by transfer from the gas to the stars). Changing the prescriptions for star formation, stellar feedback, pressurization of the multi-phase interstellar medium, and cooling will lead to more or less efficient gas compression and star formation in the earlier phases of the

merger – this changes the absolute amount and spatial distribution of the gas at the time of the final merger, and thus the efficiency with which gas is channeled into a central starburst and bulge. In sufficiently minor mergers, tidal stripping of the external portions of the halo can effect some of the gas and stars, and this can be accelerated by ram-pressure effects if we model a halo with a very dense, high-pressure hot gas background (analogous to e.g. massive clusters). There is even the possibility, especially in very low- μ mergers, of complete tidal disruption of the secondary. This may be important for the overall mass budget in the most massive galaxies (although that depends on the precise separation between ‘galaxy’ and intra-cluster or intra-group light, which is poorly defined; see e.g. Dolag et al. 2009).

It is not our purpose here to address these effects in detail; we refer to Hopkins et al. (2009a) for a more thorough discussion. However, we wish to stress that if any physical *consequence* of mergers is desired, then even perfect knowledge of the merger rate and mass ratio distribution is only sufficient for an order-of-magnitude estimate of the impact of mergers. Further knowledge of e.g. the distribution of merger orbital parameters and gas fractions is necessary to improve estimates of the impact of mergers at the factor $\sim 2 - 3$ level. And to improve to an accuracy much beyond a factor ~ 2 , knowledge of the much more detailed structural properties of galaxies and highly uncertain effects of feedback would be required.

6. SUMMARY AND DISCUSSION: THE UNCERTAINTY BUDGET

A simple comparison of different predictions of the galaxy-galaxy merger rate from e.g. different semi-analytic models and simulations demonstrate that the predictions vary by an order of magnitude (Figure 2). We have attempted to survey the sources of uncertainty and systematic differences between various theoretical attempts to predict the galaxy merger rate, to identify what drives these differences and address how progress can be made.

6.1. Dark Matter and Dynamics

All models depend similarly on the “background” dark matter merger rate. However, with proper caution in adopting definitions, and modern convergence in high-resolution cosmological simulations, the uncertainties in this rate can be reduced to the factor $\sim 2 - 3$ level. In other words, if how galaxies populate halos is appropriately fixed, there is relatively little uncertainty in the merger rate owing to the dynamics of the dark matter (Figure 1). The relevant quantities considered here include:

Cosmology: Nominally, changing the cosmology within the statistical uncertainty in modern observational constraints (Komatsu et al. 2009) leads to factor ~ 1.5 differences in the merger rate. However, most of this owes to changes in the mass function which can be normalized out – if e.g. the galaxy population is normalized so as to match the observed stellar mass function and large-scale bias, then the resulting differences in the merger rate are negligible compared to the other sources of uncertainty below (Figure 1).

Halo-Halo Merger Rate Determinations: Again, nominally different halo-halo merger rate determinations differ by factors of ~ 2 . Some of this stems from e.g. the inherent ambiguity in masses and extents of halos. However, much owes to the definitions adopted when attempting to fit/quantify the instantaneous merger rate “function.” If these definitions are properly accounted for, or if halo merger trees are used di-

rectly to track galaxies (rather than defining halo merger rates at all), the real uncertainties are small, a factor ~ 1.5 . Effects of baryons contribute similarly small ($\sim 10 - 20\%$) uncertainties in the overall halo-halo merger rate (Figure 3). However, care is needed with definitions – defining halo mergers simply in terms of “instantaneous” mass ratios, and/or applying cuts which can be useful for constructing *galaxy* merger trees, or using timestep-sensitive mass ratio definitions can lead to an apparent (artificial) suppression of major halo-halo mergers by an order-of-magnitude or more (Figures 4-5). This requires careful consideration in simulation-based semi-analytic models.

Subhalo versus Halo Merger Rates: Defining the merger rate not when halos first merge into e.g. a larger friends-of-friends group, but when subhalos merge/are destroyed into the central (primary) group subhalo, may be more representative of galaxy-galaxy mergers. Doing so, however, introduces additional uncertainties owing to e.g. how subhalos are identified. It is also important, in this case, that the subhalo “mass” (for merger mass ratio purposes) be defined as the “in-fall” or maximum pre-accretion/pre-stripping mass (i.e. maximum mass before the system became a subhalo), otherwise the mass of all systems $\rightarrow 0$ at merger, by definition. Different calculations from various simulations, with different subhalo identification/destruction criteria, yield rates converged to within a factor ~ 2 (Figure 6). Using hydrodynamic simulations to “tag” mergers (using the galaxies therein as the ideal subhalo/halo “tracers”) yields consistent results.

Merger Timescales: In models without subhalo resolution or with Press-Schechter merger trees, galaxy-galaxy mergers are often assumed to be “delayed” with respect to the halo-halo merger by a timescale given by e.g. the dynamical friction time (this approximates the subhalo evolution). We show that calibrations of such timescales, from high-resolution galaxy merger simulations, are relatively well-converged (Figure 7). Adopting one of these calibrations, we show that the implied merger rate agrees well with that obtained from full tracking of subhalos, with similar factor ~ 2 uncertainties (Figure 8). For *major* mergers, both methods agree well with the simple halo-halo merger rate – this is because the timescale for such a merger to complete ($\ll t_{\text{Hubble}}$) is short compared to the time between such mergers ($\sim t_{\text{Hubble}}$).

However, we show that adopting an artificially long merger timescale, or adopting e.g. the simple Chandrasekhar timescale with a normalization higher than the specific numerical calibrations here, can suppress high-redshift mergers by factors $\sim 5 - 10$. We also caution that these calibrations are designed for the time from halo-halo merger – they are *not* constructed for application in “hybrid” models, where subhalos are followed in simulations down to some resolution limit, and then a merger time is applied to the residual “orphan” based on its instantaneous radius and mass. We show that in such cases, if the system is lost to resolution at large radius $> 0.2 R_{\text{vir}}$, the calibrated formulae are not self-consistent, and can lead to over-estimates of the total merger timescale by factors of $\sim 2 - 8$ (Figure 9).

6.2. Baryonic Physics

Controlling for some of the caveats and definitions above, the dominant uncertainties in predicted merger rates owe to variations in how galaxies populate halos. The differences can be identified with the shape of the assumed halo occupation distribution (HOD) – essentially, the distribution of galaxy

masses in given (sub)halo masses, $M_{\text{gal}}(M_{\text{halo}})$. Some model for this is necessary to translate halo-halo mergers (or halo-subhalo mergers) into galaxy-galaxy mergers. We therefore consider how these physics are modeled and how they lead to variations in the merger rate, in three classes of models with very different approaches towards modeling the galaxy-halo connection:

Semi-Empirical (HOD-based) Models: In semi-empirical models the HOD is adopted explicitly from observational constraints. As such, it is subject to the attendant uncertainties and limited to the dynamic range observed. At low redshifts and galaxy masses within factors of several around $\sim L_*$, the constraints are tight and different methods agree well – resulting uncertainties in the galaxy-galaxy merger rate (holding halo-halo merger rates *fixed*) are a factor ~ 1.5 (Figures 1 & 10). The uncertainties grow to a factor ~ 2 at the lowest and highest masses, and $z \sim 1 - 2$. Above $z \sim 2$, the uncertainties grow very rapidly: there are not sufficient observational constraints on the HOD to make strong statements about galaxy merger rates. Predicted merger rates and pair counts from such models agree well with direct observations over a stellar mass range $\sim 10^{9.5} - 10^{11.5} M_{\odot}$ and redshifts $z = 0 - 2$ (see Figures above and Stewart et al. 2009a; Hopkins et al. 2010). At this level, it appears, there is no tension between Λ CDM merger rates and merger/pair counts. To the extent that other models disagree significantly with the observed merger fractions, it should owe to issues in the model baryonic physics leading to a $M_{\text{gal}}(M_{\text{halo}})$ distribution different from that observed.

Cosmological Hydrodynamic Simulations: In hydrodynamic simulations, the $M_{\text{gal}}(M_{\text{halo}})$ distribution is predicted in an *a priori* manner based on the cooling, star formation, and feedback models implemented in the simulation. Unfortunately, it is not yet possible to run large-volume cosmological hydrodynamic simulations (needed to quantify the galaxy-galaxy merger rate) with the spatial and mass resolution and detailed prescriptions for feedback from stars and black holes that it is becoming clear are necessary to form “realistic” galaxies. Simulations available that do not include feedback yield poor agreement with the observed HOD and stellar mass function, predicting a relationship closer to the efficient star formation limit ($M_{\text{gal}} = f_b M_{\text{halo}}$). Relative to what the merger rates would be, taking the same dynamics and merger locations but re-populating galaxies with masses chosen to fit the observed HOD (stellar mass and clustering), these simulations tend to over-predict merger rates at low masses and high redshifts, and under-predict rates at high masses and low redshifts by factors $\sim 3 - 5$, as well as over-predicting the relative importance of minor versus major mergers at all masses (Figure 10). Because it is the *shape* of the $M_{\text{gal}} - M_{\text{halo}}$ relation that is most important, simply re-normalizing predicted masses by a uniform factor will not correct for these effects. Re-normalizing all masses to their “correct” masses given some observed HOD is an improvement, but care is still needed, since the incorrect masses and morphologies will affect quantities such as the dynamical friction time.

Semi-Analytic Models: In semi-analytic models, the difficulties and expense of simulations are replaced by use of analytic prescriptions, given some background dark matter population, to predict ultimate galaxy properties. Such models are adjusted to give good agreement with the galaxy stellar mass function (and clustering) at $z = 0$; as such, some agreement with the $M_{\text{gal}}(M_{\text{halo}})$ distribution of *central* galaxies (which dominate the stellar mass function at all masses) is implicit

itly guaranteed. Indeed, we find that adopting the predicted SAM HODs for central galaxies, instead of the empirically determined HOD, yields no systematic difference and scatter within the same factor ~ 2 allowed by different observational constraints (Figures 10). At high redshifts, the uncertainties in both grow. Some SAMs yield growing discrepancies relative to semi-empirical models, directly related to issues such as e.g. the known tendency of SAMs to over-predict the abundance of low-mass galaxies at high redshift, but these are still within the factor $\sim 3 - 5$ level at $z < 3$ (Figure 11).

However, the HOD of *satellite* galaxies in semi-analytic models is extremely sensitive to prescriptions for cooling, stellar feedback, and halo mixing/stripping in satellites. Moreover, satellite masses are not strongly constrained by the stellar mass function, so there is no implicit guarantee/check that these are correct, and more detailed comparison with e.g. observed group catalogs and small-scale clustering must be used to calibrate the models.

In detail, it is well-known that most SAMs have difficulty reproducing the observed properties of satellite galaxy populations: they tend to predict satellite galaxies that are under-massive and over-quenched, relative to observations (see Figures 12 & 13 and e.g. Weinmann et al. 2006a,b; Wang et al. 2007; Kimm et al. 2009, and references therein). This can occur even in models where the small-scale clustering of satellites is over-predicted, relative to observations (see e.g. Guo et al. 2010b, who find both effects, although the clustering discrepancy is probably due to the large value of σ_8 used in the simulation). In most of the models, satellites lose their entire halo gas reservoir at the moment they become such – even in major mergers (and even when the moment of “becoming” a satellite occurs at several times the primary virial radius). Moreover, the combination of simple stellar wind feedback and cooling models often leads to the satellites *also* losing almost all of their cold/disk gas reservoir.

As such, at low masses where gas fractions are important, initially major and even equal-mass mergers can easily become minor mergers in the model, potentially suppressing the predicted merger rate by a large factor. Models which yield more “over-quenched” satellite populations tend to yield correspondingly smaller merger rates (Figures 12 & 14). Correcting for these differences, for example by enforcing that satellite galaxies obey a similar HOD to central galaxies, or by adjusting the prescriptions for cooling onto satellite galaxies such that they better reproduce the observed color and star formation rate distributions of satellites, leads to larger merger rates that converge with the merger rates predicted from semi-empirical models (Figure 14). Whether or not this is necessary for matching e.g. close pair counts and small-scale clustering is unclear (see e.g. the comparison in Kitzbichler & White 2008; Guo et al. 2010b), but it demonstrates an important uncertainty in predictions of merger rates.

6.3. Impact of Mergers on Galaxies

Of course, even with perfect knowledge of the merger rate and distribution of mass ratios under some definition, it is not trivial to relate this to either observable or physical quantities such as merger fractions or the amount of bulge mass formed in mergers. We consider how two basic aspects of this relate to uncertainties in merger rates and their consequences:

Mass Ratio Definitions: As outlined in Stewart (2009), a halo-halo major merger is not necessarily a galaxy-galaxy major merger, and vice versa. But there are also several means of defining galaxy-galaxy mass ratio, including e.g. the stellar-

stellar mass ratio μ_* and baryon (stellar mass plus cold gas within the galaxy disk)-baryon mass ratio μ_{gal} . At high (low) masses, the merger rate in terms of stellar-stellar mass ratio is enhanced (suppressed) by an order of magnitude relative to the halo-halo merger rate (Figure 15). In comparing e.g. model predictions to stellar mass ratio-selected pair samples, clearly the stellar mass ratio is most applicable. However, in comparing to luminosity-ratio selected pair samples, or to morphological studies, where the total baryonic mass is what matters, the merger rate in terms of the baryon-baryon mass ratio is more relevant. At high masses, this behaves similarly to the merger rate in terms of the stellar-stellar mass ratio. At low masses, however, galaxies are increasingly gas rich – thus many mergers with minor stellar-stellar mass ratios in fact have major baryon-baryon mass ratios, and the baryon-baryon major merger rate is a factor ~ 3 higher than the stellar-stellar merger rate. This may, in fact, explain at least part of the well-established fact that morphology-inferred merger rates at low masses tend to be systematically higher than pair count-inferred merger rates (see e.g. López-Sanjuan et al. 2009a,b, and references therein).

Although often a superior definition, even the baryonic mass ratio misses the tightly-bound dark matter that is important for the dynamics of final merger, which will not be stripped because it is within the baryonic radii. We therefore propose a new “dynamical” mass and mass ratio which approximates the most important quantity in high-resolution simulations, and includes both baryons and some dark matter. Merger rates in terms of this quantity behave similarly to baryonic major mergers, but with somewhat ($\sim 20 - 50\%$) higher (lower) rates at low (high) masses (Figure 15).

If one wishes to infer the amount of bulge or other “damage” done by major mergers but uses *either* the stellar-stellar mass ratio or halo-halo mass ratio, the estimate can easily be systematically incorrect at the factor $\sim 3 - 10$ level. This is critical for observational studies seeking to infer the amount of bulge formed by major mergers, using a stellar-mass-ratio selected sample (see e.g. Bundy et al. 2009); at low masses, there might be ~ 3 times as many “damaging” mergers as given by this statistic. Also, many analytic models simply use the stellar (or even halo) mass ratio as a criterion for determining the impact of a merger – the systematic errors introduced by this assumption can be *larger* than those from halo mis-identification, complete ignorance of subhalos, assignment of incorrect merger timescales, or discrepancies between model and observed stellar mass functions. Yet despite the considerable literature on those problems, there has been relatively little focus on the adoption of better mass ratio proxies.

Other Parameters and Merger ‘Impact’: Even when controlling for the above uncertainties, we show that at perfectly well-defined merger rates and merger mass ratios, there is a large variation in the physical effects of mergers, owing to other parameters such as the merger dynamics and orbit, merger gas fractions, and initial structural properties of the merging systems (Figure 15). At the same merger mass ratio, a prograde orbit can build twice as much bulge as a retrograde orbit, and will lead to much more dramatic tidal features and distortions observable in morphology-selected samples. Also, the merger timescale will be significantly different at moderate/small radii < 100 kpc, so pair-selected samples will also see biased distributions. A gas-poor major merger will typically violently relax the entire stellar disk and funnel the gas entirely into a starburst that builds central bulge mass, but a very gas-rich merger will experience very inefficient angular

momentum transfer, suppressing the burst mass by a factor $\sim (1 - f_{\text{gas}})$, a factor $\sim 3 - 5$ at low masses and high redshifts. More subtle properties lead to scatter and offsets at a smaller, but non-negligible level. Many of these are discussed in detail in Hopkins et al. (2009a).

From a purely empirical perspective, without knowledge of these properties, it is difficult to assess the impact of mergers (in terms of e.g. the bulge mass formed) at better than the factor $\sim 2 - 3$ level. From the theoretical perspective, neglecting quantities such as orbital parameters and, especially, gas fractions, in forward-modeling merger remnants again will introduce systematic uncertainties that are larger than any of the uncertainties from modeling the dark matter distribution, subhalos, merger timescales, and the like. Many of these are outlined in Hopkins et al. (2009b); at high masses, the systematic uncertainties are less severe, but at low masses, where systems tend to be gas-rich, they can be factors $\sim 5 - 10$ in the total bulge mass formed.

Another important point is that mergers are continuous – there is no special division at the traditional 1:3 distinction between “major” and “minor” mergers. In many models, and in many observational assessments of merger “effects,” it is assumed that major mergers are completely destructive, while minor mergers do no damage. This simple assumption introduces systematic factor ~ 2 errors in the total merger “damage budget” and total bulge formation efficiency (see Hopkins et al. 2010); the effect in terms of skewing which mergers do “more” or “less” for bulge formation can obviously be severe.

6.4. Observational Constraints and Outlook

Testing these models and determining the true merger rate in a robust manner will of course ultimately depend on observations. Continued observations of the merger rate, bearing in mind the caveats above, are of obvious importance. In improving such estimates, calibration of specific samples

to high-resolution N -body simulations, *specifically with mock observations matched to the exact selection and methodology adopted*, will be critical (see e.g. Lotz et al. 2008a). And even with such a calibration, the discussion above makes it clear why the relation between observed merger rates and bulge buildup (depending on a number of secondary properties not directly observed), in detail, must be *forward-modeled*.

Tighter observational constraints on the halo occupation distribution – from e.g. group catalogs, kinematics, weak lensing, and clustering – in particular at low masses and at high redshifts, will directly improve the semi-empirical models, and put strong constraints on the a priori galaxy formation models in the areas that have greatest effect on predicted merger rates. As we have shown, constraints on satellite populations specifically will be valuable. However, the satellite populations of particular interest are not the historically well-studied extreme cases of e.g. dwarfs in the local group or low-mass galaxies in Virgo and massive clusters (from which much of our intuition regarding dynamical friction, stripping, and satellite gas exhaustion comes). The case of interest for merger rates is that of major mergers (i.e. near equal-mass galaxies) in field or loose group environments (i.e. systems analogous to the local group, but where either the Milky Way or Andromeda is the “satellite” of interest). Also, at low stellar masses and high redshifts, large uncertainties remain in galaxy gas masses, and these matter as much or more relative to the stellar mass in assessing the “impact” of mergers.

We thank Simon White, Gabriella de Lucia, Owen Parry, Carlos Frenk, Andrew Benson, Sharda Joge, Thorsten Naab, Eyal Neistein, Simone Weinmann, Volker Springel, Martin White, Joanne Cohn, Carrie Bridge, Jennifer Lotz, T. J. Cox, and Eliot Quataert for helpful discussions throughout the development of this work. Support for PFH was provided by the Miller Institute for Basic Research in Science, University of California Berkeley.

REFERENCES

- Abadi, M. G., Navarro, J. F., Steinmetz, M., & Eke, V. R. 2003, *ApJ*, 597, 21
- Allen, P. D., Driver, S. P., Graham, A. W., Cameron, E., Liske, J., & de Propris, R. 2006, *MNRAS*, 371, 2
- Angulo, R. E., Lacey, C. G., Baugh, C. M., & Frenk, C. S. 2009, *MNRAS*, 399, 983
- Avila-Reese, V., Zavala, J., Firmani, C., & Hernández-Toledo, H. M. 2008, *AJ*, 136, 1340
- Ball, N. M., Loveday, J., Brunner, R. J., Baldry, I. K., & Brinkmann, J. 2006, *MNRAS*, 373, 845
- Barnes, J. E., & Hernquist, L. 1996, *ApJ*, 471, 115
- Barnes, J. E., & Hernquist, L. E. 1991, *ApJ*, 370, L65
- Baugh, C. M., Lacey, C. G., Frenk, C. S., Granato, G. L., Silva, L., Bressan, A., Benson, A. J., & Cole, S. 2005, *MNRAS*, 356, 1191
- Behroozi, P. S., Conroy, C., & Wechsler, R. H. 2010, *ApJ*, 717, 379
- Bell, E. F., & de Jong, R. S. 2000, *MNRAS*, 312, 497
- , 2001, *ApJ*, 550, 212
- Bell, E. F., McIntosh, D. H., Katz, N., & Weinberg, M. D. 2003a, *ApJ*, 585, L117
- , 2003b, *ApJS*, 149, 289
- Bell, E. F., Phleps, S., Somerville, R. S., Wolf, C., Borch, A., & Meisenheimer, K. 2006a, *ApJ*, 652, 270
- Bell, E. F., et al. 2006b, *ApJ*, 640, 241
- Benson, A. J. 2005, *MNRAS*, 358, 551
- Benson, A. J., Bower, R. G., Frenk, C. S., Lacey, C. G., Baugh, C. M., & Cole, S. 2003, *ApJ*, 599, 38
- Benson, A. J., Džanović, D., Frenk, C. S., & Sharples, R. 2007, *MNRAS*, 379, 841
- Benson, A. J., Lacey, C. G., Frenk, C. S., Baugh, C. M., & Cole, S. 2004, *MNRAS*, 351, 1215
- Bertone, S., & Conselice, C. J. 2009, *MNRAS*, 396, 2345
- Bertone, S., De Lucia, G., & Thomas, P. A. 2007, *MNRAS*, 379, 1143
- Binney, J., & Tremaine, S. 1987, *Galactic dynamics* (Princeton, NJ, Princeton University Press, 1987)
- Blanton, M. R., & Berlind, A. A. 2007, *ApJ*, 664, 791
- Bluck, A. F. L., Conselice, C. J., Bouwens, R. J., Daddi, E., Dickinson, M., Papovich, C., & Yan, H. 2009, *MNRAS*, 394, L51
- Borch, A., et al. 2006, *A&A*, 453, 869
- Borriello, A., & Salucci, P. 2001, *MNRAS*, 323, 285
- Borriello, A., Salucci, P., & Danese, L. 2003, *MNRAS*, 341, 1109
- Bournaud, F., Jog, C. J., & Combes, F. 2005, *A&A*, 437, 69
- 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
- Bridge, C. R., Carlberg, R. G., & Sullivan, M. 2010, *ApJ*, 709, 1067
- Bridge, C. R., et al. 2007, *ApJ*, 659, 931
- Brough, S., Forbes, D. A., Kilborn, V. A., & Couch, W. 2006, *MNRAS*, 370, 1223
- Brown, M. J. I., et al. 2008, *ApJ*, 682, 937
- Bullock, J. S., et al. 2001, *MNRAS*, 321, 559
- Bundy, K., Fukugita, M., Ellis, R. S., Targett, T. A., Belli, S., & Kodama, T. 2009, *ApJ*, 697, 1369
- Bundy, K., Treu, T., & Ellis, R. S. 2007, *ApJ*, 665, L5
- Calura, F., Jimenez, R., Panter, B., Matteucci, F., & Heavens, A. F. 2008, *ApJ*, 682, 252
- Carollo, C. M., Stiavelli, M., & Mack, J. 1998, *AJ*, 116, 68
- Cattaneo, A., Dekel, A., Devriendt, J., Guiderdoni, B., & Blaizot, J. 2006, *MNRAS*, 370, 1651
- Cattaneo, A., Mamon, G. A., Warnick, K., & Knebe, A. 2010, *MNRAS*, in press, arXiv:1002.3257
- Chabrier, G. 2003, *PASP*, 115, 763
- Choi, J., & Nagamine, K. 2009, *MNRAS*, 393, 1595
- Cimatti, A., et al. 2008, *A&A*, 482, 21

- Coil, A. L., et al. 2008, *ApJ*, 672, 153
- Comerford, J. M., et al. 2009, *ApJ*, 698, 956
- Conroy, C., & Wechsler, R. H. 2009, *ApJ*, 696, 620
- Conroy, C., Wechsler, R. H., & Kravtsov, A. V. 2006, *ApJ*, 647, 201
- Conroy, C., et al. 2007, *ApJ*, 654, 153
- Conselice, C. J., Yang, C., & Bluck, A. F. L. 2009, *MNRAS*, 394, 1956
- Cooper, M. C., et al. 2006, *MNRAS*, 370, 198
- , 2007, *MNRAS*, 376, 1445
- Cooray, A. 2006, *MNRAS*, 365, 842
- Cox, T. J., Dutta, S. N., Di Matteo, T., Hernquist, L., Hopkins, P. F., Robertson, B., & Springel, V. 2006, *ApJ*, 650, 791
- Cox, T. J., Jonsson, P., Somerville, R. S., Primack, J. R., & Dekel, A. 2008, *MNRAS*, 384, 386
- Cresci, G., et al. 2009, *ApJ*, 697, 115
- Croft, R. A. C., Di Matteo, T., Springel, V., & Hernquist, L. 2009, *MNRAS*, 400, 43
- Croton, D. J., et al. 2006, *MNRAS*, 365, 11
- Davé, R. 2008, *MNRAS*, 385, 147
- de Lucia, G., & Blaizot, J. 2007, *MNRAS*, 375, 2
- De Lucia, G., Kauffmann, G., Springel, V., White, S. D. M., Lanzoni, B., Stoehr, F., Tormen, G., & Yoshida, N. 2004, *MNRAS*, 348, 333
- De Propriis, R., Liske, J., Driver, S. P., Allen, P. D., & Cross, N. J. G. 2005, *AJ*, 130, 1516
- Di Matteo, T., Colberg, J., Springel, V., Hernquist, L., & Sijacki, D. 2008, *ApJ*, 676, 33
- Di Matteo, T., Springel, V., & Hernquist, L. 2005, *Nature*, 433, 604
- Dolag, K., Murante, G., & Borgani, S. 2009, *ArXiv e-prints*, in press, arXiv:0911.1129
- Domingue, D. L., Xu, C. K., Jarrett, T. H., & Cheng, Y. 2009, *ApJ*, in press, arXiv:0901.4545
- Domínguez-Palmero, L., Balcells, M., Erwin, P., Prieto, M., Cristóbal-Hornillos, D., Eliche-Moral, M. C., & Guzmán, R. 2008, *A&A*, 488, 1167
- D’Onghia, E., Besla, G., Cox, T. J., & Hernquist, L. 2009, *Nature*, 460, 605
- Driver, S. P., Allen, P. D., Liske, J., & Graham, A. W. 2007, *ApJ*, 657, L85
- Drory, N., Salvato, M., Gabasch, A., Bender, R., Hopp, U., Feulner, G., & Pannella, M. 2005, *ApJ*, 619, L131
- Eke, V. R., et al. 2004, *MNRAS*, 355, 769
- Elahi, P. J., Thacker, R. J., Widrow, L. M., & Scannapieco, E. 2009, *MNRAS*, 395, 1950
- Erb, D. K. 2008, *ApJ*, 674, 151
- Erb, D. K., Steidel, C. C., Shapley, A. E., Pettini, M., Reddy, N. A., & Adelberger, K. L. 2006, *ApJ*, 646, 107
- Fakhouri, O., Ma, C., & Boylan-Kolchin, M. 2010, *MNRAS*, 406, 2267
- Fakhouri, O., & Ma, C.-P. 2008, *MNRAS*, 386, 577
- Font, A. S., Navarro, J. F., Stadel, J., & Quinn, T. 2001, *ApJ*, 563, L1
- Font, A. S., et al. 2008, *MNRAS*, 389, 1619
- Fontana, A., et al. 2006, *A&A*, 459, 745
- Fontanot, F., De Lucia, G., Monaco, P., Somerville, R. S., & Santini, P. 2009, *MNRAS*, 397, 1776
- Forster Schreiber, N. M., et al. 2009, *ApJ*, 706, 1364
- Gadotti, D. A. 2009, *MNRAS*, 393, 1531
- Gao, L., White, S. D. M., Jenkins, A., Stoehr, F., & Springel, V. 2004, *MNRAS*, 355, 819
- Genel, S., Genel, R., Bouché, N., Naab, T., & Sternberg, A. 2009, *ApJ*, 701, 2002
- Genel, S., et al. 2008, *ApJ*, 688, 789
- Gerke, B. F., et al. 2007, *MNRAS*, 376, 1425
- Giocoli, C., Tormen, G., Sheth, R. K., & van den Bosch, F. C. 2010, *MNRAS*, 404, 502
- Gottlöber, S., Klypin, A., & Kravtsov, A. V. 2001, *ApJ*, 546, 223
- Governato, F., Willman, B., Mayer, L., Brooks, A., Stinson, G., Valenzuela, O., Wadsley, J., & Quinn, T. 2007, *MNRAS*, 374, 1479
- Governato, F., et al. 2009, *MNRAS*, 398, 312
- Granato, G. L., De Zotti, G., Silva, L., Bressan, A., & Danese, L. 2004, *ApJ*, 600, 580
- Guo, Q., White, S., Li, C., & Boylan-Kolchin, M. 2010a, *MNRAS*, 404, 1111
- Guo, Q., & White, S. D. M. 2008, *MNRAS*, 384, 2
- Guo, Q., et al. 2010b, *MNRAS*, in press, arXiv:1006.0106
- Guo, Y., et al. 2009, *MNRAS*, 1107
- Haines, C. P., Gargiulo, A., La Barbera, F., Mercurio, A., Merluzzi, P., & Busarello, G. 2007, *MNRAS*, 381, 7
- Harker, G., Cole, S., Helly, J., Frenk, C., & Jenkins, A. 2006, *MNRAS*, 367, 1039
- Hernquist, L. 1989, *Nature*, 340, 687
- Hernquist, L., & Mihos, J. C. 1995, *ApJ*, 448, 41
- Hernquist, L., & Weinberg, M. D. 1989, *MNRAS*, 238, 407
- Hopkins, A. M., & Beacom, J. F. 2006, *ApJ*, 651, 142
- Hopkins, P. F., Cox, T. J., Kereš, D., & Hernquist, L. 2008a, *ApJS*, 175, 390
- Hopkins, P. F., Cox, T. J., Younger, J. D., & Hernquist, L. 2009a, *ApJ*, 691, 1168
- Hopkins, P. F., Hernquist, L., Cox, T. J., & Kereš, D. 2008b, *ApJS*, 175, 356
- Hopkins, P. F., Somerville, R. S., Hernquist, L., Cox, T. J., Robertson, B., & Li, Y. 2006, *ApJ*, 652, 864
- Hopkins, P. F., et al. 2009b, *MNRAS*, 397, 802
- , 2010, *ApJ*, 715, 202
- Ilbert, O., et al. 2010, *ApJ*, 709, 644
- Jiang, C. Y., Jing, Y. P., Faltenbacher, A., Lin, W. P., & Li, C. 2008, *ApJ*, 675, 1095
- Jiang, C. Y., Jing, Y. P., & Lin, W. P. 2010, *A&A*, 510, A60+
- Jogee, S., et al. 2008, in *Astronomical Society of the Pacific Conference Series*, Vol. 396, *Formation and Evolution of Galaxy Disks*, *Astronomical Society of the Pacific Conference Series*, ed. J. G. Funes & E. M. Corsini, 337+–
- Jogee, S., et al. 2009, *ApJ*, 697, 1971
- Johansson, P. H., Burkert, A., & Naab, T. 2009a, *ApJ*, 707, L184
- Johansson, P. H., Naab, T., & Burkert, A. 2009b, *ApJ*, 690, 802
- Kang, X., Jing, Y. P., Mo, H. J., & Börner, G. 2005, *ApJ*, 631, 21
- Kang, X., & van den Bosch, F. C. 2008, *ApJ*, 676, L101
- Kannappan, S. J. 2004, *ApJ*, 611, L89
- Kartaltepe, J. S., et al. 2007, *ApJS*, 172, 320
- Kazantzidis, S., Bullock, J. S., Zentner, A. R., Kravtsov, A. V., & Moustakas, L. A. 2008, *ApJ*, 688, 254
- Kazantzidis, S., Zentner, A. R., Kravtsov, A. V., Bullock, J. S., & Debattista, V. P. 2009, *ApJ*, 700, 1896
- Kennicutt, Jr., R. C. 1998, *ApJ*, 498, 541
- Kereš, D., Katz, N., Davé, R., Fardal, M., & Weinberg, D. H. 2009a, *MNRAS*, 396, 2332
- Kereš, D., Katz, N., Fardal, M., Davé, R., & Weinberg, D. H. 2009b, *MNRAS*, 395, 160
- Kereš, D., Katz, N., Weinberg, D. H., & Davé, R. 2005, *MNRAS*, 363, 2
- Khochfar, S., & Burkert, A. 2006, *A&A*, 445, 403
- Khochfar, S., & Silk, J. 2006, *ApJ*, 648, L21
- , 2009, *MNRAS*, 397, 506
- Kimm, T., et al. 2009, *MNRAS*, 394, 1131
- Kitzbichler, M. G., & White, S. D. M. 2008, *MNRAS*, 391, 1489
- Klypin, A., Gottlöber, S., Kravtsov, A. V., & Khokhlov, A. M. 1999, *ApJ*, 516, 530
- Koda, J., Milosavljević, M., & Shapiro, P. R. 2009, *ApJ*, 696, 254
- Komatsu, E., et al. 2009, *ApJS*, 180, 330
- Kormendy, J., & Kennicutt, Jr., R. C. 2004, *ARA&A*, 42, 603
- Kravtsov, A. V., Berlind, A. A., Wechsler, R. H., Klypin, A. A., Gottlöber, S., Allgood, B., & Primack, J. R. 2004, *ApJ*, 609, 35
- Kravtsov, A. V., Klypin, A. A., & Khokhlov, A. M. 1997, *ApJS*, 111, 73
- Krivitsky, D. S., & Kontorovich, V. M. 1997, *A&A*, 327, 921
- Lacey, C., & Cole, S. 1993, *MNRAS*, 262, 627
- Lee, K., Giallisco, M., Conroy, C., Wechsler, R. H., Ferguson, H. C., Somerville, R. S., Dickinson, M. E., & Urry, C. M. 2009, *ApJ*, 695, 368
- Li, C., Kauffmann, G., Jing, Y. P., White, S. D. M., Börner, G., & Cheng, F. Z. 2006, *MNRAS*, 368, 21
- Lin, L., et al. 2004, *ApJ*, 617, L9
- , 2008, *ApJ*, 681, 232
- López-Sanjuan, C., Balcells, M., Pérez-González, P. G., Barro, G., García-Dabó, C. E., Gallego, J., & Zamorano, J. 2009a, *A&A*, 501, 505
- López-Sanjuan, C., et al. 2009b, *ApJ*, 694, 643
- Lotz, J. M., Jonsson, P., Cox, T. J., & Primack, J. R. 2008a, *MNRAS*, 391, 1137
- , 2010a, *MNRAS*, 404, 590
- , 2010b, *MNRAS*, 404, 575
- Lotz, J. M., et al. 2008b, *ApJ*, 672, 177
- Maller, A. H., Katz, N., Kereš, D., Davé, R., & Weinberg, D. H. 2006, *ApJ*, 647, 763
- Mamon, G. A. 2006, in *Groups of Galaxies in the Nearby Universe*; Springer-Verlag, Berlin-Heidelberg, ed. I. Saviane, V. Ivanov, & J. Borissova
- Mandelbaum, R., Seljak, U., Kauffmann, G., Hirata, C. M., & Brinkmann, J. 2006, *MNRAS*, 368, 715
- Mannucci, F., et al. 2009, *MNRAS*, 398, 1915
- Maraston, C. 2005, *MNRAS*, 362, 799
- Maraston, C., Daddi, E., Renzini, A., Cimatti, A., Dickinson, M., Papovich, C., Pasquali, A., & Pirzkal, N. 2006, *ApJ*, 652, 85
- Marchesini, D., van Dokkum, P. G., Förster Schreiber, N. M., Franx, M., Labbé, I., & Wuyts, S. 2009, *ApJ*, 701, 1765
- Mateus, A. 2008, *MNRAS*, in press, arXiv:0802.2720 [astro-ph]
- McCarthy, I. G., Frenk, C. S., Font, A. S., Lacey, C. G., Bower, R. G., Mitchell, N. L., Balogh, M. L., & Theuns, T. 2008, *MNRAS*, 383, 593
- McGaugh, S. S. 2005, *ApJ*, 632, 859
- Mihos, J. C., & Hernquist, L. 1994, *ApJ*, 431, L9

- , 1996, *ApJ*, 464, 641
- Monaco, P., Fontanot, F., & Taffoni, G. 2007, *MNRAS*, 375, 1189
- More, S., van den Bosch, F. C., Cacciato, M., Mo, H. J., Yang, X., & Li, R. 2009, *MNRAS*, 392, 801
- Moster, B. P., Macciò, A. V., Somerville, R. S., Johansson, P. H., & Naab, T. 2010a, *MNRAS*, 403, 1009
- Moster, B. P., Somerville, R. S., Maulbetsch, C., van den Bosch, F. C., Macciò, A. V., Naab, T., & Oser, L. 2010b, *ApJ*, 710, 903
- Naab, T., & Burkert, A. 2003, *ApJ*, 597, 893
- Naab, T., Johansson, P. H., Ostriker, J. P., & Efstathiou, G. 2007, *ApJ*, 658, 710
- Nagai, D., & Kravtsov, A. V. 2005, *ApJ*, 618, 557
- Neistein, E., & Dekel, A. 2008, *Monthly Notices of the Royal Astronomical Society*, 383, 615
- Neistein, E., van den Bosch, F. C., & Dekel, A. 2006, *MNRAS*, 372, 933
- Neistein, E., & Weinmann, S. M. 2010, *MNRAS*, 405, 2717
- Nurmi, P., Heinämäki, P., Saar, E., Einasto, M., Holopainen, J., Martínez, V. J., & Einasto, J. 2006, *A&A*, in press [astro-ph/0611941]
- Okamoto, T., Eke, V. R., Frenk, C. S., & Jenkins, A. 2005, *MNRAS*, 363, 1299
- Oppenheimer, B. D., Davé, R., Kereš, D., Fardal, M., Katz, N., Kollmeier, J. A., & Weinberg, D. H. 2010, *MNRAS*, 406, 2325
- Panther, B., Jimenez, R., Heavens, A. F., & Charlot, S. 2007, *MNRAS*, 378, 1550
- Park, C., Choi, Y.-Y., Vogeley, M. S., Gott, J. R. I., & Blanton, M. R. 2007, *ApJ*, 658, 898
- Parry, O. H., Eke, V. R., & Frenk, C. S. 2009, *MNRAS*, 396, 1972
- Pasquali, A., Gallazzi, A., Fontanot, F., van den Bosch, F. C., De Lucia, G., Mo, H. J., & Yang, X. 2010, *MNRAS*, 1157
- Patton, D. R., & Atfield, J. E. 2008, *ApJ*, 685, 235
- Patton, D. R., et al. 2002, *ApJ*, 565, 208
- Pérez-González, P. G., Trujillo, I., Barro, G., Gallego, J., Zamorano, J., & Conselice, C. J. 2008a, *ApJ*, 687, 50
- Pérez-González, P. G., et al. 2008b, *ApJ*, 675, 234
- Persic, M., & Salucci, P. 1988, *MNRAS*, 234, 131
- , 1990, *MNRAS*, 245, 577
- Persic, M., Salucci, P., & Stel, F. 1996a, *Astrophysical Letters Communications*, 33, 205
- , 1996b, *MNRAS*, 281, 27
- Piontek, F., & Steinmetz, M. 2009, *MNRAS*, in press, arXiv:0909.4167
- Puech, M., Hammer, F., Flores, H., Delgado-Serrano, R., Rodrigues, M., & Yang, Y. 2010, *A&A*, 510, A68+
- Puech, M., et al. 2008, *A&A*, 484, 173
- Purcell, C. W., Kazantzidis, S., & Bullock, J. S. 2009, *ApJ*, 694, L98
- Quinn, P. J., & Goodman, J. 1986, *ApJ*, 309, 472
- Reed, D. S., Bower, R., Frenk, C. S., Jenkins, A., & Theuns, T. 2007, *MNRAS*, 374, 2
- Richard, S., Brook, C. B., Martel, H., Kawata, D., Gibson, B. K., & Sanchez-Blazquez, P. 2010, *MNRAS*, 402, 1489
- Robaina, A. R., Bell, E. F., van der Wel, A., Somerville, R. S., Skelton, R. E., McIntosh, D. H., Meisenheimer, K., & Wolf, C. 2010, *ApJ*, 719, 844
- Robaina, A. R., et al. 2009, *ApJ*, 704, 324
- Robertson, B., Bullock, J. S., Cox, T. J., Di Matteo, T., Hernquist, L., Springel, V., & Yoshida, N. 2006, *ApJ*, 645, 986
- Robertson, B., Yoshida, N., Springel, V., & Hernquist, L. 2004, *ApJ*, 606, 32
- Scannapieco, C., Tissera, P. B., White, S. D. M., & Springel, V. 2008, *MNRAS*, 389, 1137
- Shankar, F. 2009, *New Astronomy Review*, 53, 57
- Shankar, F., Lapi, A., Salucci, P., De Zotti, G., & Danese, L. 2006, *ApJ*, 643, 14
- Shapley, A. E., Coil, A. L., Ma, C.-P., & Bundy, K. 2005, *ApJ*, 635, 1006
- Shaw, L. D., Weller, J., Ostriker, J. P., & Bode, P. 2006, *ApJ*, 646, 815
- , 2007, *ApJ*, 659, 1082
- Shen, S., Mo, H. J., White, S. D. M., Blanton, M. R., Kauffmann, G., Voges, W., Brinkmann, J., & Csabai, I. 2003, *MNRAS*, 343, 978
- Shen, Y., et al. 2010, *ApJ*, 719, 1693
- Sijacki, D., Springel, V., di Matteo, T., & Hernquist, L. 2007, *MNRAS*, 380, 877
- Simha, V., Weinberg, D. H., Davé, R., Gnedin, O. Y., Katz, N., & Kereš, D. 2009, *MNRAS*, 399, 650
- Somerville, R. S., Hopkins, P. F., Cox, T. J., Robertson, B. E., & Hernquist, L. 2008a, *MNRAS*, 391, 481
- Somerville, R. S., & Kolatt, T. S. 1999, *MNRAS*, 305, 1
- Somerville, R. S., & Primack, J. R. 1999, *MNRAS*, 310, 1087
- Somerville, R. S., Primack, J. R., & Faber, S. M. 2001, *MNRAS*, 320, 504
- Somerville, R. S., et al. 2008b, *ApJ*, 672, 776
- Sommer-Larsen, J., Götz, M., & Portinari, L. 2003, *ApJ*, 596, 47
- Spergel, D. N., et al. 2003, *ApJS*, 148, 175
- , 2007, *ApJS*, 170, 377
- Springel, V., Di Matteo, T., & Hernquist, L. 2005a, *MNRAS*, 361, 776
- Springel, V., & Hernquist, L. 2003a, *MNRAS*, 339, 289
- , 2003b, *MNRAS*, 339, 312
- , 2005, *ApJ*, 622, L9
- Springel, V., White, S. D. M., Tormen, G., & Kauffmann, G. 2001, *MNRAS*, 328, 726
- Springel, V., et al. 2005b, *Nature*, 435, 629
- Stark, D. V., McGaugh, S. S., & Swaters, R. A. 2009, *AJ*, 138, 392
- Stewart, K. R. 2009, in *Astronomical Society of the Pacific Conference Series*, Vol. Jogee, S. and Hao, L. and Blanc, G. and Marinova, I., *Astronomical Society of the Pacific Conference Series*, ed. S. Jogee, I. Marinova, L. Hao & G. A. Blanc, 243+
- Stewart, K. R., Bullock, J. S., Barton, E. J., & Wechsler, R. H. 2009a, *ApJ*, 702, 1005
- Stewart, K. R., Bullock, J. S., Wechsler, R. H., & Maller, A. H. 2009b, *ApJ*, 702, 307
- Tacconi, L. J., et al. 2008, *ApJ*, 680, 246
- Taylor, J. E., & Babul, A. 2004, *MNRAS*, 348, 811
- Tinker, J. L., Wechsler, R. H., & Zheng, Z. 2010, *ApJ*, 709, 67
- Tinker, J. L., Weinberg, D. H., Zheng, Z., & Zehavi, I. 2005, *ApJ*, 631, 41
- Tinker, J. L., & Wetzel, A. R. 2010, *ApJ*, 719, 88
- Tormen, G., Moscardini, L., & Yoshida, N. 2004, *MNRAS*, 350, 1397
- Tully, R. B., Verheijen, M. A. W., Pierce, M. J., Huang, J.-S., & Wainscoat, R. J. 1996, *AJ*, 112, 2471
- Vale, A., & Ostriker, J. P. 2006, *MNRAS*, 371, 1173
- van den Bosch, F. C., Aquino, D., Yang, X., Mo, H. J., Pasquali, A., McIntosh, D. H., Weinmann, S. M., & Kang, X. 2008, *MNRAS*, 387, 79
- van den Bosch, F. C., Tormen, G., & Giocoli, C. 2005, *MNRAS*, 359, 1029
- van den Bosch, F. C., et al. 2007, *MNRAS*, 376, 841
- van Dokkum, P. G. 2008, *ApJ*, 674, 29
- Velazquez, H., & White, S. D. M. 1999, *MNRAS*, 304, 254
- Villalobos, Á., & Helmi, A. 2008, *MNRAS*, 391, 1806
- 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
- Wang, Y., Yang, X., Mo, H. J., van den Bosch, F. C., Katz, N., Pasquali, A., McIntosh, D. H., & Weinmann, S. M. 2009, *ApJ*, 697, 247
- Wechsler, R. H., Zentner, A. R., Bullock, J. S., Kravtsov, A. V., & Allgood, B. 2006, *ApJ*, 652, 71
- Weinberg, D. H., Colombi, S., Davé, R., & Katz, N. 2008, *ApJ*, 678, 6
- Weinmann, S. M., Kauffmann, G., van den Bosch, F. C., Pasquali, A., McIntosh, D. H., Mo, H., Yang, X., & Guo, Y. 2009a, *MNRAS*, 394, 1213
- Weinmann, S. M., Kauffmann, G., von der Linden, A., & De Lucia, G. 2009b, *MNRAS*, in press, arXiv:0912.2741
- Weinmann, S. M., van den Bosch, F. C., Yang, X., & Mo, H. J. 2006a, *MNRAS*, 366, 2
- Weinmann, S. M., van den Bosch, F. C., Yang, X., Mo, H. J., Croton, D. J., & Moore, B. 2006b, *MNRAS*, 372, 1161
- Weinzirl, T., Jogee, S., Khochfar, S., Burkert, A., & Kormendy, J. 2009, *ApJ*, 696, 411
- Wetzel, A. R., Cohn, J. D., & White, M. 2009, *MNRAS*, 395, 1376
- Wetzel, A. R., & White, M. 2010, *MNRAS*, 403, 1072
- White, M., Martini, P., & Cohn, J. D. 2008, *MNRAS*, 390, 1179
- White, S. D. M., & Rees, M. J. 1978, *MNRAS*, 183, 341
- Xu, C. K., Sun, Y. C., & He, X. T. 2004, *ApJ*, 603, L73
- Yan, R., Madgwick, D. S., & White, M. 2003, *ApJ*, 598, 848
- Yang, X., Mo, H. J., Jing, Y. P., & van den Bosch, F. C. 2005a, *MNRAS*, 358, 217
- Yang, X., Mo, H. J., & van den Bosch, F. C. 2003, *MNRAS*, 339, 1057
- , 2008, *ApJ*, 676, 248
- , 2009, *ApJ*, 693, 830
- Yang, X., Mo, H. J., van den Bosch, F. C., & Jing, Y. P. 2005b, *MNRAS*, 356, 1293
- Younger, J. D., Hopkins, P. F., Cox, T. J., & Hernquist, L. 2008, *ApJ*, 686, 815
- Younger, J. D., et al. 2007, *ApJ*, 671, 1531
- Zentner, A. R. 2007, *International Journal of Modern Physics D*, 16, 763
- Zentner, A. R., Berlind, A. A., Bullock, J. S., Kravtsov, A. V., & Wechsler, R. H. 2005, *ApJ*, 624, 505
- Zheng, Z., Coil, A. L., & Zehavi, I. 2007, *ApJ*, 667, 760
- Zheng, Z., et al. 2005, *ApJ*, 633, 791