MNRAS 000, 1–13 (2018) Preprint 5 November 2019 Compiled using MNRAS LaTeX style file v3.0

The evolution of the baryon fraction in halos as a cause of scatter in the galaxy stellar mass in the EAGLE simulation

Andrea Kulier[1⋆], Nelson Padilla[1], Joop Schaye[2], Robert A. Crain[3], Matthieu Schaller[4], Richard G. Bower[4], Tom Theuns[4], Enrique Paillas[1]
[1] Instituto de Astrofisica, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, Santiago, Chile
[2] Leiden Observatory, Leiden University, P.O. Box 9513, NL-2300 RA Leiden, The Netherlands
[3] Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF, UK
[4] Institute for Computational Cosmology, Department of Physics, University of Durham, South Road, Durham, DH1 3LE, UK

5 November 2019

ABSTRACT

The EAGLE simulation suite has previously been used to investigate the relationship between the stellar mass of galaxies, M_*, and the properties of dark matter halos, using the hydrodynamical reference simulation combined with a dark matter only (DMO) simulation having identical initial conditions. The stellar masses of central galaxies in halos with $M_{200c} > 10^{11}\mathrm{M}_\odot$ were shown to correlate with the DMO halo maximum circular velocity, with ≈ 0.2 dex of scatter that is uncorrelated with other DMO halo properties. Here we revisit the origin of the scatter in the $M_* - V_{\mathrm{max,DMO}}$ relation in EAGLE at $z = 0.1$. We find that the scatter in M_* correlates with the mean age of the galaxy stellar population such that more massive galaxies at fixed $V_{\mathrm{max,DMO}}$ are younger. The scatter in the stellar mass and mean stellar population age results from variation in the baryonic mass, $M_{\mathrm{bary}} = M_{\mathrm{gas}} + M_*$, of the galaxies' progenitors at fixed halo mass and concentration. At the redshift of peak correlation ($z \approx 1$), the progenitor baryonic mass accounts for 76% of the variance in the $z = 0.1$ $M_* - V_{\mathrm{max,DMO}}$ relation. The scatter in the baryonic mass, in turn, is primarily set by differences in feedback strength and gas accretion over the course of the evolution of each halo.

Key words: galaxies : formation — galaxies : evolution — galaxies : halos

1 INTRODUCTION

Understanding the relationship between galaxies and their host dark matter halos has been a longstanding problem relevant to both galaxy evolution and cosmology. Owing to the difficulty of directly measuring the properties of dark matter halos, it is often necessary to infer them from the observable properties of the galaxies that they host. Therefore, relations between measurable galaxy properties and halo properties have been much sought after.

Hydrodynamical cosmological simulations offer a way to investigate these relationships. However, such simulations were until recently unable to produce large enough samples of galaxies at sufficient resolution to perform statistical studies of galaxy properties. Partly as a result, a variety of methods have been created for the purpose of assigning galaxies to dark matter halos from dark matter-only simulations, which are much less computationally expensive to perform. These include halo occupation distributions (Seljak 2000; Peacock & Smith 2000) and abundance matching (Vale & Ostriker 2004, 2006; Kravtsov et al. 2004). Such models are generally calibrated to reproduce the observed properties of populations of galaxies; e.g., their spatial clustering.

In contrast to their predecessors, recent hydrodynamical cosmological simulations such as EAGLE (Schaye et al. 2015; Crain et al. 2015), Illustris (Vogelsberger et al. 2014a,b; Genel et al. 2014), and Horizon-AGN (Dubois et al. 2016) allow for measurements of galaxy and halo properties for sizeable galaxy populations. Such simulations can be used to study galaxy-halo relations and to inform semi-analytic methods such as those previously mentioned.

One topic that has recently been investigated with the latest hydrodynamical simulations is the correlation between galaxy stellar masses and the properties of their host dark matter halos. This is particularly relevant to abundance matching models, which assign observed samples of galaxies to simulated dark matter halos by assuming a monotonic relation (with some scatter) between galaxy stellar mass or luminosity and a given dark matter halo parameter. Simu-

⋆ E-mail: akulier@astro.puc.cl

lations can be used to identify the most suitable halo property by which to assign galaxy stellar masses to halos. The EAGLE simulation suite has been used for this purpose because it reproduces the galaxy stellar mass function (Schaye et al. 2015), which is reproduced by construction in halo abundance matching models, and because it includes a dark matter-only variant of the main hydrodynamical simulation with identical initial conditions, allowing the identification of "corresponding" host dark matter halos in the dark matter-only simulation.

In particular, Chaves-Montero et al. (2016) and Matthee et al. (2017) both used the set of EAGLE simulations to examine the relationship between the stellar mass of galaxies and the properties of their matched dark matter halos in the dark-matter only simulation. Chaves-Montero et al. (2016) found that the stellar mass of central and satellite galaxies is most tightly correlated with the parameter $V_{\rm relax}$, the maximum circular velocity attained by the host halo in its history while satisfying a relaxation criterion. This parameter had slightly less scatter with the stellar mass than $V_{\rm peak}$, the maximum circular velocity achieved by the halo during its entire history, and $V_{\rm infall}$, the maximum circular velocity of the halo before it becomes a subhalo of a larger halo. Furthermore, the authors found that parameters based on the maximum circular velocity of the halo are more strongly correlated with the galaxy stellar mass than those based on the halo mass. This is in agreement with results from abundance matching fits to observed halo clustering (e.g, Reddick et al. 2013).

Matthee et al. (2017) considered only central galaxies, obtaining results consistent with Chaves-Montero et al. (2016). They found that $V_{\rm max}$ in the dark matter-only simulation correlates better with the stellar mass M_* than the halo mass $M_{\rm 200c}$. However, there was a remaining scatter of ≈ 0.2 dex in the correlation between $V_{\rm max}$ and M_* for their halo sample, defined by a mass cut of $M_{\rm 200c} > 10^{11}{\rm M}_\odot$. Interestingly, they found that the residuals of the $V_{\rm max} - M_*$ relation did not correlate with *any* of several other halo parameters that they considered — including concentration, half-mass formation time, sphericity, triaxiality, spin, and two simple measures of small- and large-scale environment.

In this paper, we investigate the source of the scatter in the relation between $V_{\rm max}$ and M_* for central galaxies. In contrast to Chaves-Montero et al. (2016) and Matthee et al. (2017), we focus on correlations between the scatter and the baryonic properties of galaxies and halos. In Section 2 we describe the EAGLE simulation suite used in our analysis and how we selected our sample of halos. In Section 3 we present our results on the origin of the $V_{\rm max} - M_*$ scatter at $z = 0.1$. Finally, we summarize our conclusions in Section 4.

Throughout this paper we assume the Planck cosmology (Planck Collaboration et al. 2014) adopted in the EAGLE simulation, such that $h = 0.6777$, $\Omega_\Lambda = 0.693$, $\Omega_m = 0.307$, and $\Omega_b = 0.048$.

2 SIMULATIONS AND HALO SAMPLE

2.1 Simulation overview

EAGLE (Schaye et al. 2015; Crain et al. 2015; McAlpine et al. 2016) is a suite of cosmological hydrodynamical simulations, run using a modified version of the N-body smooth particle hydrodynamics (SPH) code GADGET-3 (Springel 2005). The changes to the hydrodynamics solver, referred to as "Anarchy" and described in Schaller et al. (2015a), are based on the formulation of SPH in Hopkins (2013), and include changes to the handling of the viscosity (Cullen & Dehnen 2010), the conduction (Price 2008), the smoothing kernel (Dehnen & Aly 2012), and the time-stepping (Durier & Dalla Vecchia 2012).

The reference EAGLE simulation has a box size of 100 comoving Mpc per side, containing 1504^3 particles each of dark matter and baryons, with a dark matter particle mass of $9.70 \times 10^6{\rm M}_\odot$, and an initial gas (baryon) particle mass of $1.81 \times 10^6{\rm M}_\odot$. The Plummer-equivalent gravitational softening length is 2.66 comoving kpc until $z = 2.8$ and 0.70 proper kpc afterward. The EAGLE suite also includes a second simulation containing only dark matter that has the same total cosmic matter density, resolution, initial conditions, and number of dark matter particles (each with mass $1.15 \times 10^7{\rm M}_\odot$) as the reference simulation.

Subgrid physics in EAGLE includes radiative cooling, photoionization heating, star formation, stellar mass loss, stellar feedback, supermassive black hole accretion and mergers, and AGN feedback. Here we briefly summarize these subgrid prescriptions, which are described in more detail in Schaye et al. (2015).

Radiative cooling and photoionization heating is implemented using the model of Wiersma et al. (2009a). Cooling and heating rates are computed for 11 elements using CLOUDY (Ferland et al. 1998), assuming that the gas is optically thin, in ionization equilibrium, and exposed to the cosmic microwave background and the evolving Haardt & Madau (2001) UV and X-ray background that is imposed instantaneously at $z = 11.5$. Extra energy is also injected at this redshift and at $z = 3.5$ to model HI and HeII reionization respectively.

Gas particles undergo stochastic conversion into star particles using the prescription of Schaye & Dalla Vecchia (2008), which imposes the Kennicutt-Schmidt law (Kennicutt 1998) on the gas. A metallicity-dependent density threshold for gas to become star-forming is used based on Schaye (2004). Star particles are assumed to be simple stellar populations with a Chabrier (2003) initial mass function. The prescriptions for stellar evolution and mass loss from Wiersma et al. (2009b) are used. The fraction of the initial stellar particle mass that is leaving the main sequence at each time step is used in combination with the initial elemental abundances of the star particle to compute the mass that is ejected from the particle due to stellar winds and supernovae.

To model the effect of stellar feedback on the ISM, the stochastic feedback prescription of Dalla Vecchia & Schaye (2012) is used, in which randomly selected gas particles close to a star particle that is losing energy are instantly heated by $10^{7.5}$ K. Each star particle is assumed to lose the total amount of energy produced by type II supernovae in a Chabrier IMF when it reaches an age of 30 Myr. The strength of the feedback in EAGLE is calibrated by adjusting the fraction of this energy that is assumed to heat the nearby gas.

Halos that reach a mass of $10^{10}{\rm M}_\odot/h$ are seeded with black holes of subgrid mass $10^5{\rm M}_\odot/h$ at their centers by converting the most bound gas particle into a "black hole" seed particle (Springel et al. 2005). These particles accrete mass at a rate specified by the minimum of the Eddington rate and the modified Bondi-Hoyle accretion rate from Rosas-Guevara et al. (2016) with $\alpha = 1$. Black hole particles are also able to merge with one another.

AGN feedback is modeled in a stochastic manner similar to stellar feedback, with the energy injection rate proportional to the black hole accretion rate. In contrast to the stellar feedback, adjustment of the fraction of lost energy assumed to heat the gas does not significantly affect the masses of galaxies due to self-regulation (Booth & Schaye 2010).

The feedback scheme used by EAGLE is able to approximately reproduce the local galaxy stellar mass function; some differences near the "knee" of the distribution cause the EAGLE stellar mass density to be $\approx 20\%$ lower than that inferred from observations. The feedback parameters have been calibrated so as to additionally reproduce the distribution of present-day galaxy sizes (Crain et al. 2015). EAGLE has been found to reproduce, without further parameter calibration, a number of other observed features of the population of galaxies, such as the $z = 0$ Tully-Fisher relation, specific star formation rates, rotation curves, colors, and the evolution of the galaxy stellar mass function and galaxy sizes (Schaye et al. 2015; Furlong et al. 2015; Schaller et al. 2015a; Trayford et al. 2016; Furlong et al. 2017).

2.2 Halo/galaxy sample and properties

Halos in EAGLE are identified by applying a friends-of-friends (FoF) algorithm with a linking length of $b = 0.2$ times the mean interparticle separation to the distribution of dark matter particles (Davis et al. 1985). Other particles types (gas, stars, and black holes) are assigned to the FoF halo of the nearest dark matter particle. The SUBFIND (Springel et al. 2001; Dolag et al. 2009) algorithm is then used to identify local overdensities of all particles types within FoF halos — referred to as subhalos. SUBFIND assigns to each subhalo only those particles that are gravitationally bound to it, with no overlap in particles between distinct subhalos. When we refer to "galaxies", we are referring to the baryonic particles associated with each subhalo. The subhalo in each FoF halo that contains the most bound particle is defined to be the central subhalo, and all others are defined as satellites. The location of the most bound particle is also used to define the center of the FoF halo, around which mean spherical overdensities are calculated to obtain halo masses such as M_{200c}, the mass inside the radius within which the mean overdensity is 200 times the critical density of the Universe.

The FoF and SUBFIND algorithms are run at a series of 29 simulation snapshots from $z = 20$ to $z = 0$, with the time between snapshots increasing from ≈ 0.1 Gyr at the beginning of the simulation to ≈ 1 Gyr at the end. Galaxy and halo catalogs as well as particle data from EAGLE have been made publicly available (McAlpine et al. 2016).

We use the method described in Schaller et al. (2015b) to match halos from the reference hydrodynamic simulation to those from the dark matter-only (DMO) simulation, and the reader is referred to that paper for details. To summarize, the reference and DMO EAGLE simulations have identical initial conditions save for the fact that the DMO simulation has slightly more massive dark matter particles to account for the mass in baryons present in the reference simulation. Each particle is tagged with a unique identifier where two particles with the same identifier in the two simulations have the same initial conditions. We define two subhalos in the reference simulation and the DMO simulation to correspond to one another if they share at least half of their 50 most bound particles.

We take as our primary sample in the reference simulation one identical to that of Matthee et al. (2017): central galaxies with redshift $z = 0.1$ and host halo mass $M_{200c} > 10^{11} \mathrm{M}_{\odot}$, resulting in a sample of 9929 galaxies and their host halos. We successfully match 9774 of these halos (98.4%) in the DMO simulation. However, we discard the halos whose matches in the DMO simulation are satellite subhalos rather than centrals, leaving 9543 halos (96.1% of our original sample).

In our analysis, we consider the properties of the progenitors of our galaxy sample in order to determine the origin of the scatter in their stellar masses. Merger trees have been created from the EAGLE simulation snapshots using a modified version (Qu et al. 2017) of the D-TREES algorithm (Jiang et al. 2014). D-TREES links subhalos to their descendants by considering the $N_{\rm link}$ most bound particles and identifying the subhalo that contains the majority of these particles in the next time snapshot. For EAGLE, $N_{\rm link}$ is set to be $\min(100, \max(0.1 N_{\rm subhalo}, 10))$, where $N_{\rm subhalo}$ is the total number of particles in the subhalo. Each subhalo is assigned only a single descendant, but a subhalo may have multiple progenitors. Each subhalo with at least one progenitor has a single "main progenitor", defined as the progenitor that has the largest mass summed across all earlier outputs, as suggested by De Lucia & Blaizot (2007) to avoid swapping of the main progenitor during major mergers. In some cases, galaxies can disappear in a snapshot and reappear at a later time; because of this, descendants are identified up to 5 snapshots later.

Essentially all (99.9%) of the galaxies in our $z = 0.1$ main sample have at least one progenitor up to $z = 4$, although in this paper we mainly concern ourselves with $z \leq 2$. We investigate the correlations between the properties of the central galaxies/subhalos and their FoF host halos at $z = 0.1$ and the properties of their progenitors at each prior timestep. We do this using the properties of the main progenitor subhalo and its FoF host halo, as well as the combined properties of all the progenitor subhalos.

The main progenitor of a central galaxy/subhalo will not necessarily be a central. There are two possible causes for this. The first is due to the definition of the main progenitor as being the one with the most massive total mass history. Because the central subhalo of a FoF halo is taken to be the one containing the most bound particle, this definition of the main progenitor means that the main progenitor of a central may not be the "central" galaxy during a roughly equal-mass merger. The second cause, which tends to affect less massive subhalos, is that halos can be "flybys": they can enter the physical space associated with a more massive FoF halo and become temporarily assigned to it as a subhalo, but later re-emerge as a separate halo. The physical state of flybys can be complex, and their bound gas mass especially can change rapidly while they are in the process of interacting with the more massive halo. In some cases, a FoF halo containing multiple subhalos can enter a more massive FoF halo and lose some of its subhalos. To avoid these complexities, we exclude non-central progenitors from our sample.

When comparing the properties of subhalos to those of their main progenitors at each timestep, we exclude from the sample those subhalos whose main progenitors are not central subhalos *at that particular timestep*. If the subhalo's main progenitor is a central subhalo at earlier or later timesteps, then we include the subhalo and main progenitor in our sample at those timesteps. At each timestep with $z \leq 2$, $> 95\%$ of the main progenitors are centrals, and $> 92\%$ are centrals for $2 \leq z \leq 4$.

When we examine the combined properties of all the progenitor subhalos rather than only those of the main progenitors, we perform a similar exclusion if there exists *any* progenitor at a given timestep that is a satellite but whose corresponding central is *not* also a progenitor. Due to this criterion, only those subhalos with a progenitor that is a "true" flyby are excluded, and not those with a progenitor that is undergoing an equal-mass merger. We do not place a mass cut on the progenitors other than that they must contain a non-zero mass in either stars or gas. This means that subhalos at $z = 0.1$ have a large number of progenitors at high redshifts, and as a result, a larger fraction have at least one progenitor that is a flyby than have a main progenitor that is not a central. At $z = 2.24$, we exclude the largest fraction of our sample, with only 83% of the descendant galaxies included. Toward both higher and lower redshifts, the fraction of galaxies retained in our sample increases, with 91% in the sample at $z = 1$ and 90% at $z = 4$. This peak at intermediate redshifts is due to the redshift evolution of the rate of flybys, which is different from that of the rate of mergers (see e.g. Sinha & Holley-Bockelmann 2012). Because of the larger fraction of galaxies excluded here, we comment during the presentation of our results in §3.2 on the impact of excluding suhalos with flyby progenitors.

We also note that the FoF halos hosting the galaxy progenitors at each timestep may contain flyby subhalos that are not present in the FoF halo hosting the $z = 0.1$ descendant. We do not correct for this as we expect these subhalos to generally constitute little of the total mass of the FoF halo, but they will contribute some scatter to the correlation between progenitor and descendant host halo properties.

For the main progenitors of the galaxies in our sample, we match the subhalos to the corresponding subhalos in the DMO simulation, in the same manner as for our $z = 0.1$ sample. We do this at a subset of redshift snapshots: $z = 0.27, 0.50, 0.74, 1.00, 1.50$ and 2.00. At $z = 2.00$, the main progenitor host halo masses are typically $\sim 1/4$ of the mass of the host halos of the descendants, but with a very large scatter; 99.7% of the main progenitors have host halo masses above $10^{10}\mathrm{M}_\odot$, which contain over 1000 particles. Once the non-central progenitors have been excluded at each redshift as described above, we are able to match $97 - 99\%$ of the progenitors to the DMO simulation at the selected redshifts. Of the successfully matched progenitor subhalos, 99% are centrals in the DMO simulation. Because we exclude non-central progenitors, our progenitor samples differ slightly depending on whether we consider progenitors in the reference simulation or their matches in the DMO simulation; this has negligible effect on our results.

We use as galaxy stellar masses the total stellar mass assigned to each galaxy's subhalo by SUBFIND, which includes some diffuse stellar mass that is similar to "intracluster light". This differs from the definition in Matthee et al. (2017), who used only the stellar mass within 30 kpc, although they found that their analysis would be nearly identical if they had used the total stellar mass because the two masses are only significantly different in very massive halos.

As a measure of the age of each galaxy's stellar population, we use the initial-mass-weighted mean stellar age. This is the mean age of the star particles belonging to a galaxy weighted by their initial mass—the mass of each star particle at the moment it formed from a gas particle, before it has lost mass due to stellar winds and supernovae (see §2.1).

We also examine the baryonic masses (stars and gas) of halos in EAGLE. (We do not include black hole particles, as they are a minuscule fraction of the total baryonic mass in each halo.) Because we have restricted our sample to only include main progenitors that are centrals, we can consider, in general, all the baryonic mass in the FoF halo hosting the progenitor to be potentially collapsing onto it. To define the baryonic mass of a FoF halo we take the sum of the masses of all the gas and stellar particles in all the SUBFIND subhalos assigned to the halo. We include both cold and hot phase gas particles.

Throughout the results section, we refer to dark matter halo properties from the DMO simulation using the subscript "DMO", whereas those without this subscript are taken from the reference simulation. $M_{200\mathrm{c}}$ refers to the mass within the radius within which the mean overdensity is 200 times the critical density, and M_{dark} is used to refer to the total mass in dark matter particles assigned to an FoF halo.

We use as a proxy for the NFW halo concentration parameter $c = R_{200}/R_s$ the ratio V_{max}/V_{200} (Prada et al. 2012). Here V_{max} is the maximum circular velocity and $V_{200} = (GM_{200}/R_{200})^{1/2}$. We note, however, that because the maximum circular velocity of each central subhalo is computed by SUBFIND, it does not include the mass contribution of any other subhalos inside the FoF halo; as a result, in a minority of cases (4% of our sample), $V_{\mathrm{max,DMO}}/V_{200\mathrm{c,DMO}} < 1$.

3 RESULTS

3.1 Stellar mass scatter at $z = 0.1$

In Matthee et al. (2017) it was found that the stellar mass, M_*, of central galaxies correlated well with the maximum circular velocity of their matched DMO halos, $V_{\mathrm{max,DMO}}$. The authors investigated whether the residual scatter in this relation correlated with any other DMO halo properties, including concentration and assembly time, finding that it did not. Here we attempt to identify the origin of this scatter by considering correlations with baryonic galaxy properties. We find that the scatter in M_* *does* correlate with the mean age of the stellar population of the galaxy. This can be seen in the top panel of Figure 1, which plots the mean stellar mass in fine bins of $V_{\mathrm{max,DMO}}$, split by the median galaxy stellar population age in each bin. The thickness of the lines shows the error on the mean — the scatter in M_* for galaxies above and below the median age is significant, but there is a clear offset in their mean M_*, such that galaxies with younger stellar populations have higher stellar masses at fixed $V_{\mathrm{max,DMO}}$.

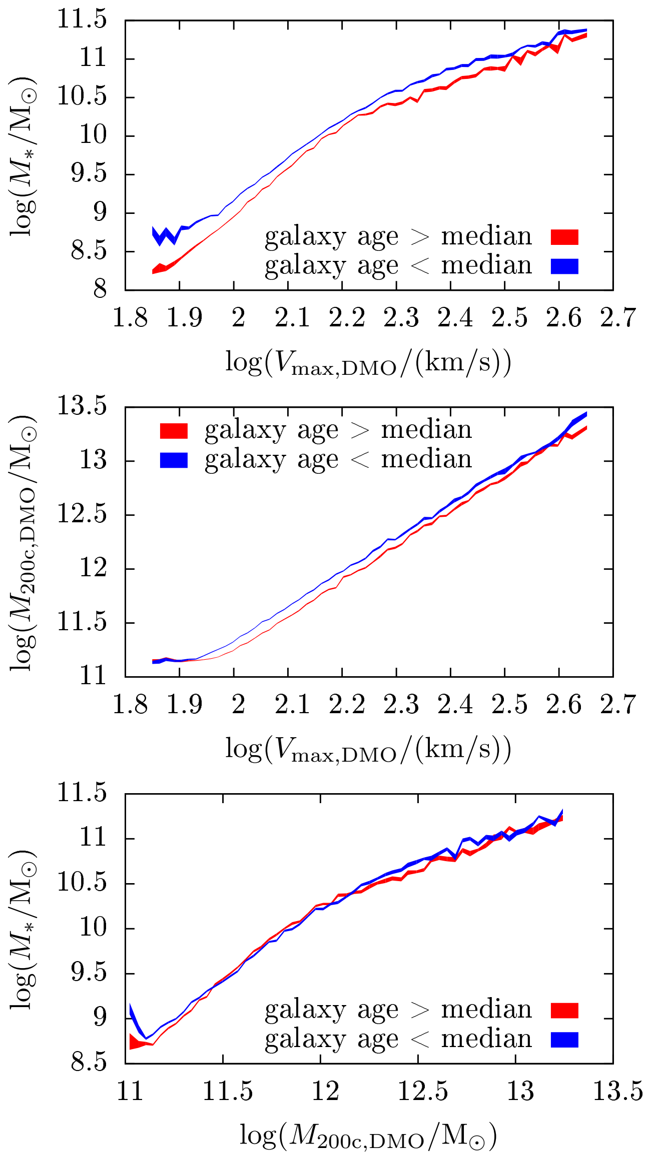

Figure 1. *Top Panel:* The relationship between the stellar mass, M_*, of central galaxies, and the maximum circular velocity of the matched dark matter halo in the dark-matter only simulation (see text), $V_{\mathrm{max,DMO}}$. In each of 90 fine bins in $V_{\mathrm{max,DMO}}$, the red line shows the mean M_* of galaxies above the median stellar population age in the bin, while the blue line is the same for galaxies below the median. The thickness of the lines represents the error on the mean M_* in each bin. Galaxies with older stellar population ages have lower stellar masses, on average, at fixed $V_{\mathrm{max,DMO}}$. *Middle Panel:* Same as the top panel, but showing the DMO halo mass $M_{\mathrm{200c,DMO}}$ on the vertical axis rather than the central stellar mass of the galaxy. Central galaxies with older stellar population ages are associated with less massive (i.e., more concentrated) halos at fixed $V_{\mathrm{max,DMO}}$. This is a reflection of the influence of halo assembly time, which is highly positively correlated with halo concentration, on the age of the central galaxy. *Bottom Panel:* The mean central galaxy stellar mass M_* as a function of the DMO halo mass, $M_{\mathrm{200c,DMO}}$, again split by the median galaxy stellar population age in each bin. There is little correlation between M_* and galaxy age at fixed halo mass.

The middle panel shows the same bins in $V_{\mathrm{max,DMO}}$, again split by the median stellar age in each bin, but now versus the halo mass of each galaxy's matched DMO halo, $M_{\mathrm{200c,DMO}}$. The halo mass is related to the halo concentration at fixed $V_{\mathrm{max,DMO}}$, such that less massive halos have higher concentrations (indeed, for a perfect NFW halo profile, V_{max} is simply an increasing function of M_{200c} and concentration). A higher halo concentration is highly correlated with an earlier halo formation time (Wechsler et al. 2002), implying that halos with lower $M_{\mathrm{200c,DMO}}$ at fixed $V_{\mathrm{max,DMO}}$ have earlier assembly times.

In the middle panel, we see that galaxies with younger stellar populations have more massive (less concentrated, later-forming) halos at fixed $V_{\mathrm{max,DMO}}$. This implies a positive correlation between halo age and galaxy age at fixed $V_{\mathrm{max,DMO}}$, as might be expected. However, in Matthee et al. (2017), it was found that there is no correlation between M_* and concentration or halo formation time at fixed $V_{\mathrm{max,DMO}}$. Thus, the age difference seen in the middle panel of Figure 1 has no correlation with the stellar mass of the galaxy, and is uncorrelated with the trend in the top panel.

The bottom panel shows the relation between halo mass and stellar mass — i.e. the stellar-halo mass relation — split by galaxy stellar population age. The trend seen here is a combination of the trends seen in the top two panels. At fixed $M_{\mathrm{200c,DMO}}$, halos have a range of values of $V_{\mathrm{max,DMO}}$. Those with higher $V_{\mathrm{max,DMO}}$ have on average central galaxies with higher M_*; furthermore, the galaxies are older on average, as seen in the middle panel. If these were the only trends present, there would be a positive correlation between galaxy stellar mass and stellar population age at fixed $M_{\mathrm{200c,DMO}}$. However, there is an additional inverse correlation between M_* and stellar population age at fixed $V_{\mathrm{max,DMO}}$, as seen in the top panel. The combination of these two opposing trends results in a lack of significant correlation between galaxy stellar mass and stellar population age at fixed $M_{\mathrm{200c,DMO}}$.

We now understand how M_* varies as a function of halo mass and concentration, which are the two "most important" halo parameters with which most other halo parameters are highly correlated (Jeeson-Daniel et al. 2011; Skibba & Macciò 2011; Wong & Taylor 2012). Therefore, we wish to remove the mean dependence of M_* and other galaxy properties on the halo mass and concentration and consider the correlations between deviations from the mean. The manner in which we do this is demonstrated in Figure 2. The leftmost panels plot $c_{\mathrm{DMO}} \equiv V_{\mathrm{max,DMO}}/V_{\mathrm{200c,DMO}}$, a proxy for the halo concentration (see §2.2), versus the DMO halo mass $M_{\mathrm{200c,DMO}}$. Each halo is color-coded by the value of one of its baryonic properties — from top to bottom: central galaxy stellar mass M_*, central galaxy mean stellar population age, and the sum of the bound stellar and gas mass in the halo (including all substructure), referred to as M_{bary}. From these plots various mean trends are evident: the stellar mass follows lines of constant $V_{\mathrm{max,DMO}}$, M_{bary} correlates primarily with $M_{\mathrm{200c,DMO}}$, and stellar population age traces a more complex increasing function of both halo mass and concentration.

We compute the mean dependence of each parameter on $M_{\mathrm{200c,DMO}}$ and c_{DMO} by fitting a bivariate smoothing spline in log-space. We do not find that varying the smoothing parameters has a large effect on our results, and simply

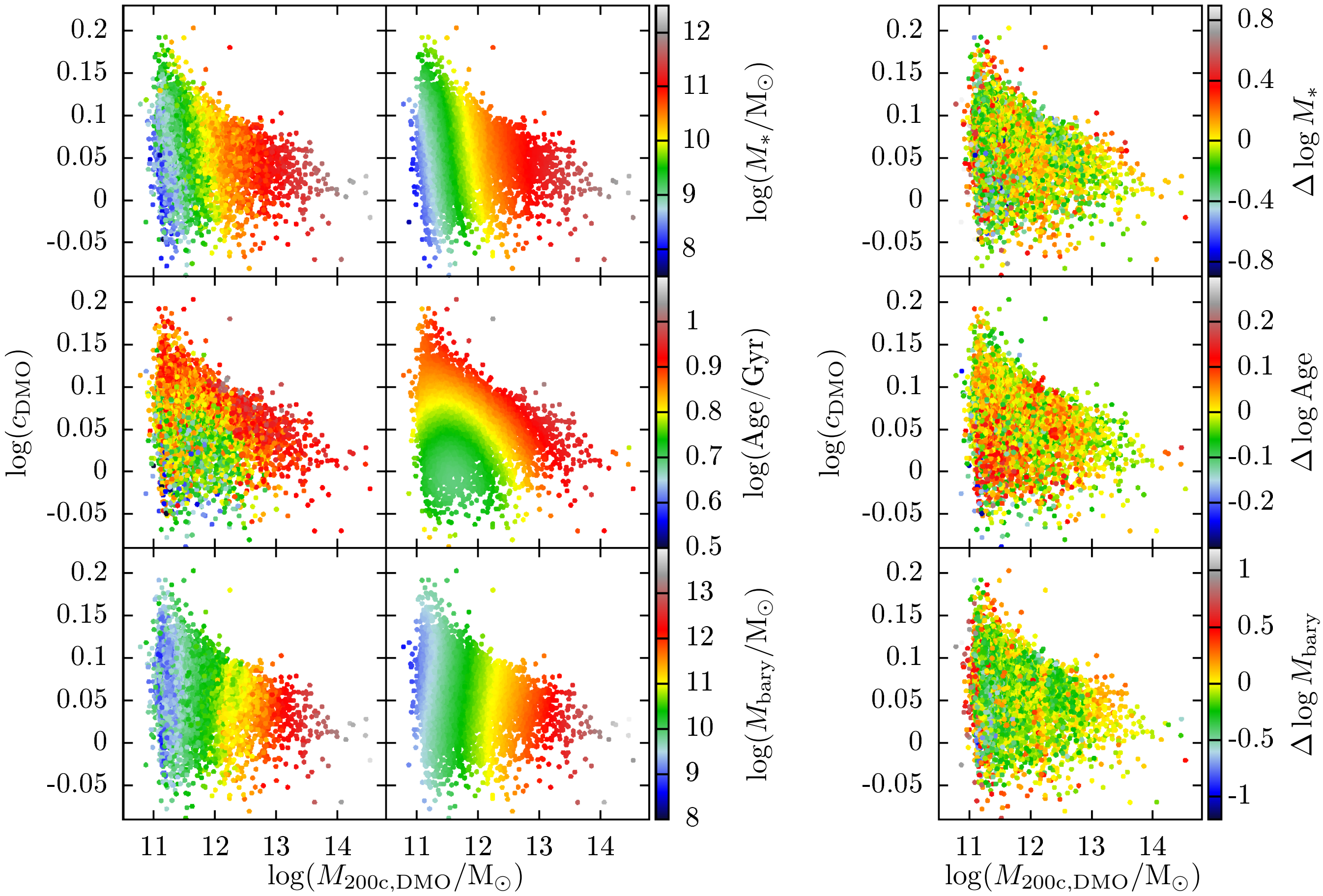

Figure 2. Galaxy/halo properties as a function of $M_{\rm 200c,DMO}$ and $c_{\rm DMO} \equiv V_{\rm max,DMO}/V_{\rm 200c,DMO}$ of the matched halo in the DMO simulation (see text). *Leftmost panels:* Points are colored by the following properties, from top to bottom: central galaxy stellar mass, central galaxy mean stellar population age, and total bound baryonic mass (gas plus stars) within the halo (including substructure). *Middle panels:* Same as the leftmost panels, but now smoothed via a smoothing spline to obtain the mean relation as a function of $M_{\rm 200c,DMO}$ and $c_{\rm DMO}$. *Rightmost panels:* The difference of the leftmost and middle panels, showing the scatter in each galaxy/halo property, denoted by "Δ" (see also Eqn. 1).

subtracting a mean in bins of $\log(M_{\rm 200c,DMO})$ and $\log(c_{\rm DMO})$ produces consistent results. These mean relations are shown in the middle set of panels in Figure 2. We then define the deviation from this mean for M_* as

$$\Delta \log M_* \equiv \log(M_*) - \overline{\log(M_*)}(\log(M_{\rm 200c,DMO}), \log(c_{\rm DMO})) \quad (1)$$

and similarly for the other galaxy/halo parameters. The deviations from the mean produced by subtracting the middle panels from the leftmost panels of Figure 2 is shown in the rightmost panels.

In Figure 3, we plot the deviation of the central galaxy stellar population age from the mean relation, $\Delta \log \mathrm{Age}$, versus $\Delta \log M_*$, confirming that there is a negative correlation (Spearman correlation coefficient $R_s = -0.56$) between the two as could be inferred from Figure 1. In the bottom panel of Figure 3, we plot $\Delta \log M_*$ versus $\Delta \log M_{\rm gas}$, where the latter is computed using the total gas contained in the host halo, including any that is bound to substructure. There is a weak positive correlation ($R_s = 0.30$) between $\Delta \log M_*$ and $\Delta \log M_{\rm gas}$, such that halos whose central galaxies have above-average stellar masses also tend to have a slight excess of gas relative to similar halos. Interestingly, this implies that such halos tend to contain a higher overall baryonic mass relative to other halos of the same mass and concentration.

3.2 Correlation of stellar mass scatter with progenitor properties

To understand the origin of the scatter in the $V_{\rm max} - M_*$ relation at $z = 0.1$, we attempt to correlate the scatter to the properties of the progenitors of the galaxies. The selection of the progenitors and the cuts made to our sample are described in §2.2. As for the stellar population age and gas mass above, we examine differences in the progenitor properties after removing the mean dependence on the halo mass and concentration of the descendant $z = 0.1$ halos, denoting this with a "Δ" in front of the property. In this way we examine the variation in the growth histories of galaxies and halos with the same present-day properties and how this affects the stellar mass of their central galaxies.

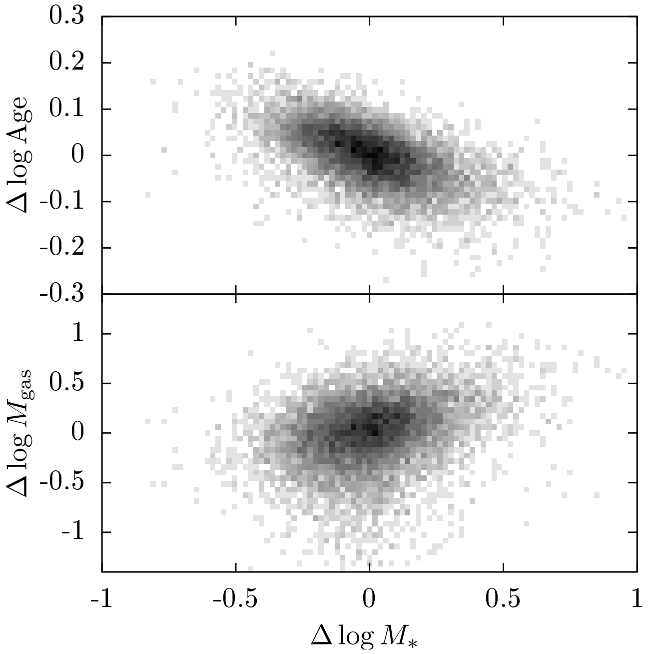

Figure 3. *Top Panel:* The deviation from the mean value at fixed $M_{\rm 200c,DMO}$ and $c_{\rm DMO}$ of the stellar mass ($\Delta \log M_*$) versus the deviation from the mean stellar population age ($\Delta \log {\rm Age}$). (See Eqn. 1 and text for details.) The darkness of the shade represents the log-density of points in each bin. *Bottom Panel:* $\Delta \log M_*$ versus the deviation from the mean of the total gas mass inside the galaxy's host halo, $\Delta \log M_{\rm gas}$.

We consider the properties of the main progenitor branch (defined in §2.2), including the stellar mass of the main progenitor galaxy, the total baryonic mass[1] within the halo hosting said galaxy, and the halo mass of the corresponding DMO halo. We also look at the sum of the stellar, baryonic, and dark matter masses of all the progenitor subhalos of each $z = 0.1$ galaxy/subhalo at different redshifts.

We denote the baryonic mass of the FoF halo hosting the main progenitor galaxy as $M_{\rm bary}$ and the sum of the baryonic masses of all the progenitor subhalos as $\Sigma M_{\rm bary}$. Similarly, M_* refers to the stellar mass of the main progenitor and ΣM_* to the sum of the stellar masses of all the progenitors. We match the main progenitor subhalos at selected redshifts to the corresponding subhalos in the DMO simulation, as described in §2.2, and refer to the mass of the host FoF halo as $M_{\rm 200c,DMO}$. We do not attempt to match the full sample of all progenitor subhalos because many are low-mass and it is more difficult to obtain accurate matches between the two simulations for low-mass subhalos. We do utilize the sum of the $M_{\rm 200c}$ halo masses from the reference simulation, minus the baryonic component, denoting this as $\Sigma(M_{\rm 200c} - M_{\rm bary})$.

In the top row of panels in Figure 4, we show a comparison of $\Delta \log M_*$ at $z = 0.1$ to $\Delta \log M_*$ of the main progenitor galaxy at $z = 0.5, 1.0,$ and 2.0 (all computed relative to $M_{\rm 200c,DMO}$ and $c_{\rm DMO}$ of the descendant at $z = 0.1$). Unsurprisingly, those galaxies with atypically high stellar masses at $z = 0.1$ tend to also have progenitors with high stellar masses. The correlation decreases with increasing redshift: the Spearman correlation coefficient is $R_s = 0.86$ at $z = 0.5$, 0.61 at $z = 1.0$, and 0.26 at $z = 2.0$. The points are color-coded by $\Delta \log {\rm Age}$ at $z = 0.1$, which follows a diagonal trend in the top panels because it is correlated with the mass of stars formed between the redshift of that panel and $z = 0.1$.

It is interesting to compare the top panels of Figure 4 to the bottom ones, which show $\Delta \log \Sigma M_{\rm bary}$ computed for the same redshifts as the top panels. Here we see that $\Delta \log M_*$ at $z = 0.1$ is positively correlated with $\Delta \log M_{\rm bary}$ at each redshift, with $R_s = 0.75$ at $z = 0.5$, 0.85 at $z = 1.0$, and 0.68 at $z = 2.0$. Unlike for the stellar mass in the top panels, the correlation strengthens between $z = 0.5$ and $z = 1.0$, and for $z \gtrsim 1$ the correlation between $\Delta \log M_*(z = 0.1)$ and $\Delta \log \Sigma M_{\rm bary}$ is stronger than than that between $\Delta \log M_*(z = 0.1)$ and $\Delta \log M_*$. Although the stellar mass of the progenitors is part of $M_{\rm bary}$, the correlation between $\Delta \log M_*(z = 0.1)$ and $\Delta \log \Sigma M_{\rm bary}$ at higher redshifts is mainly driven by the gas mass, as will be shown below.

In the bottom panels of Figure 4, it is also apparent that for $z \lesssim 1$, the mean stellar age of the galaxy at $z = 0.1$ is negatively correlated with $\Delta \log \Sigma M_{\rm bary}$. This reveals the origin of the negative correlation between stellar mass and mean stellar population age at $z = 0.1$. It is possible for two sets of halo progenitors at $z \sim 1$ with different total baryonic masses to evolve into halos with the same $M_{\rm 200c,DMO}$ and $c_{\rm DMO}$ at $z = 0.1$; however, due to their different initial baryonic masses, they will experience different amounts of star formation at $z < 1$ and the one with higher initial baryonic mass will tend to have a younger, more massive central galaxy.

The relationship between $\Delta \log M_*(z = 0.1)$ and progenitor properties is revealed in greater detail in Figure 5. Here we show the fraction of the variance of $\Delta \log M_*$ at $z = 0.1$ that can be accounted for by the different progenitor properties as a function of redshift. This is done by fitting a line to the relationship between each progenitor property and $\Delta \log M_*(z = 0.1)$, defined by $f(x) = ax$ (the intercept is taken to be zero because all properties are normalized by removing the mean at fixed $M_{\rm 200c,DMO}$ and $c_{\rm DMO}$). The fractional contribution to the variance is $[{\rm Var}(\Delta \log M_{*,z=0.1}) - {\rm Var}(\Delta \log M_{*,z=0.1} - ax)]/{\rm Var}(\Delta \log M_{*,z=0.1})$. Here ${\rm Var}(\Delta \log M_{*,z=0.1})$ varies slightly for the different redshift points due to the different sample cuts at each point (see §2.2) but is always $\approx (0.19\ {\rm dex})^2$.

The red line with solid circular points in Figure 5 shows the fraction of the variance of $\Delta \log M_*(z = 0.1)$ accounted for by $\Delta \log \Sigma M_{\rm bary}$ at each redshift — the quantity that was plotted along the x-axis in the bottom panels of Figure 4. The correlation between $\Delta \log M_*(z = 0.1)$ and $\Delta \log \Sigma M_{\rm bary}$ peaks at $z \approx 1.1$, where the baryonic mass of

[1] We compute the baryonic mass of the main progenitor host as the sum of the gas and stellar masses bound to each subhalo in the FoF halo that hosts the main progenitor galaxy. However, this halo may contain subhalos that do not merge with the central galaxy by $z = 0.1$ and are thus not its progenitors. In practice, this is a minor difference because the gas of satellite subhalos is generally stripped quickly upon entering a FoF halo and is reassigned to the central subhalo, and also because the satellite galaxies that take a long time to merge with the central tend to have low masses.

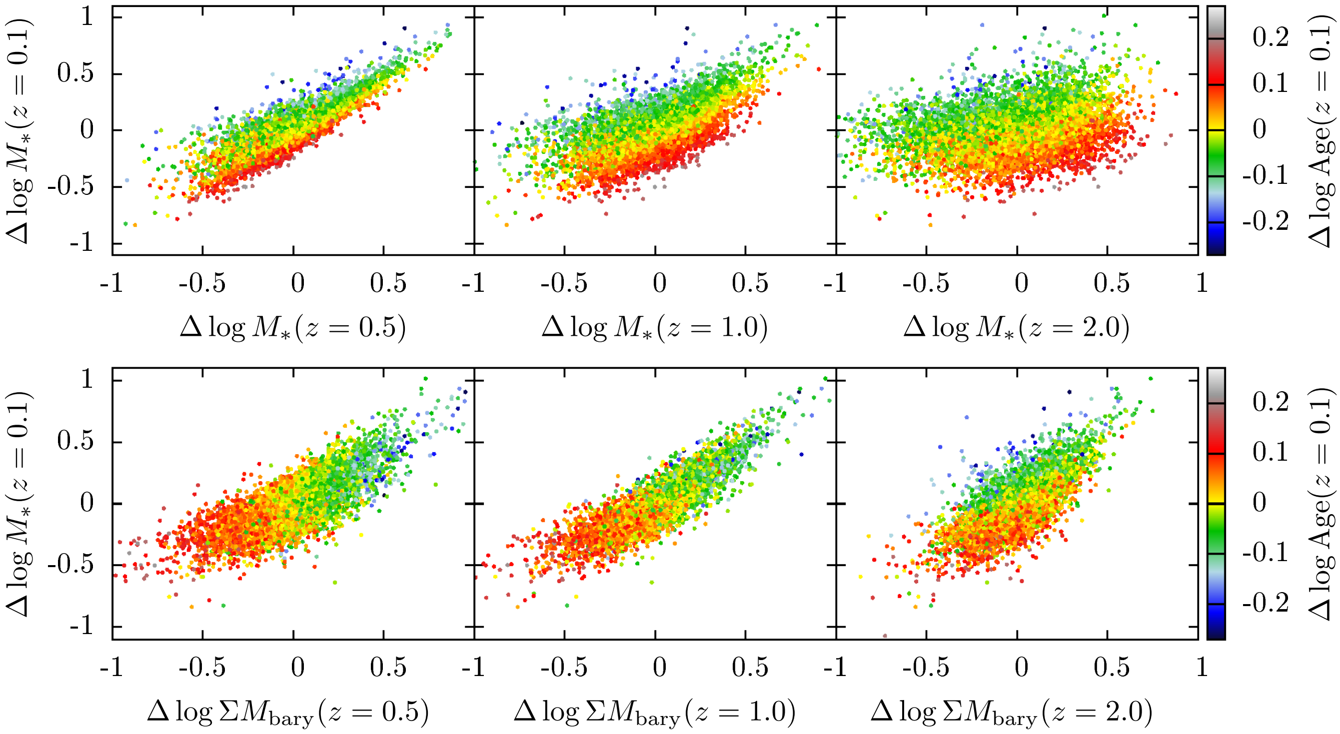

Figure 4. As in Figure 3, the deviation of various galaxy and halo properties from the mean at fixed $z = 0.1$ $M_{\rm 200c,DMO}$ and $c_{\rm DMO}$ (see Eqn. 1 and text of §3.1 for more details). The top panels show $\Delta \log M_*$ for the $z = 0.1$ galaxy sample versus $\Delta \log M_*$ of their main progenitor galaxies at $z = 0.5$ (left), 1.0 (middle), and 2.0 (right). Points are colorered by $\Delta \log \mathrm{Age}$ at $z = 0.1$, where the Age refers to the stellar population age of each galaxy. The bottom panels show the same, but for $\Delta \log M_*(z = 0.1)$ versus $\Delta \log \Sigma M_{\rm bary}$ at $z = 0.5, 1.0$, and 2.0, where $\Sigma M_{\rm bary}$ is the sum of the stellar and gas masses of all the progenitors of each galaxy. For $z \gtrsim 1$, $\Sigma M_{\rm bary}$ of the progenitor halos is a better predictor of $\Delta \log M_*(z = 0.1)$ than $\Delta \log M_*$ of the main progenitor galaxy.

the progenitors accounts for 76% of the variance of $\Delta \log M_*$ at $z = 0.1$. For comparison, we show as the teal line with triangular points $\Delta \log \Sigma M_*$, where ΣM_* is the sum of the stellar masses of the progenitor galaxies at each redshift. (Note that this is different from what is plotted in the top panels of Figure 4, which shows only the stellar mass of the main progenitor galaxy). For $z \gtrsim 0.8$, the total baryonic mass accounts for a larger fraction of the scatter in M_* at $z = 0.1$ than ΣM_*. This indicates that the gas reservoir available for star formation is the major factor determining the eventual stellar mass of the central galaxy in a halo.

The green line is the same as the red line but including only the baryonic content of the host halo of the main progenitor galaxy. As noted previously, at each timestep we have excluded from consideration those progenitors that are temporarily a satellite within a larger halo, ergo all the halos considered are those that have the main progenitor as their central subhalo. The baryon content within the host halo of the main progenitor galaxy (which is also generally the most massive progenitor halo) accounts for 68% of the variance of $\Delta \log M_*(z = 0.1)$ at $z \approx 0.95$, meaning that the properties of the main progenitor halo alone account for the majority (89%) of the variance that is accounted for by all the progenitors.

The above results are for our full galaxy/halo sample; however, due to our chosen lower halo mass cut of $10^{11}\mathrm{M}_\odot$ and the steepness of the halo mass function, the typical halo in our sample has fairly low mass, so it does not gain a significant fraction of its mass from mergers. For higher-mass subsamples of our main sample, the peak of the correlation between $\Delta \log M_{\rm bary}$ and $\Delta \log M_*(z = 0.1)$ becomes broader in redshift and shallower in the fraction of the M_* variance it accounts for at a single redshift. For halos with $10^{11.5}\mathrm{M}_\odot < M_{\rm 200c,DMO} < 10^{12.0}\mathrm{M}_\odot$, the combined baryonic masses of all the progenitors account for 74% of the variance at $z = 0.1$ at the redshift of peak correlation, for $10^{12.0}\mathrm{M}_\odot < M_{\rm 200c,DMO} < 10^{12.5}\mathrm{M}_\odot$ it is 70%, and for $M_{\rm 200c,DMO} > 10^{12.5}\mathrm{M}_\odot$, it is 55%. Interestingly, the redshift of peak correlation does not vary significantly for different halo mass ranges, likely due to the fact that higher-mass halos are assembled from multiple lower-mass halos.

On the other hand, the redshift of peak correlation between $\Delta \log M_*(z = 0.1)$ and $\Delta \log M_{\rm bary}$ of the main progenitor halo does vary with the halo mass range, owing to the later assembly time for higher-mass halos. For halos with $10^{11.5}\mathrm{M}_\odot < M_{\rm 200c,DMO} < 10^{12.0}\mathrm{M}_\odot$, the redshift of peak correlation for the main progenitor is $z \approx 0.95$, close to the value for the full halo sample, and the M_* variance at $z = 0.1$ accounted for at this redshift is 64%. For halos with $10^{12.0}\mathrm{M}_\odot < M_{\rm 200c,DMO} < 10^{12.5}\mathrm{M}_\odot$, the redshift of peak correlation is $z \approx 0.55$ and the fraction of the variance

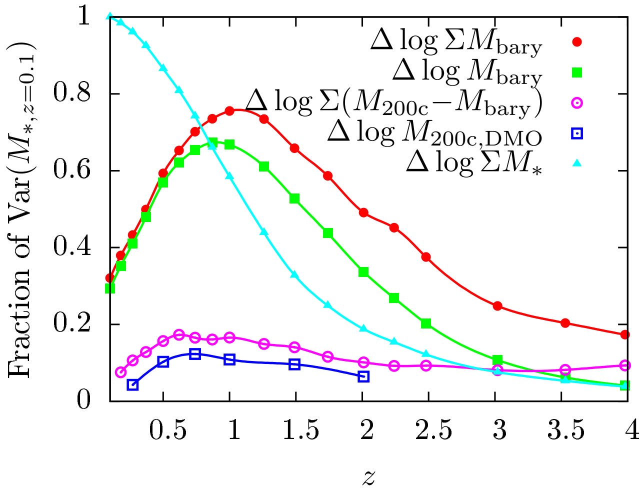

Figure 5. The fraction of the variance of $\Delta \log M_*$ for central galaxies with $M_{\rm 200c} > 10^{11}{\rm M}_\odot$ at $z = 0.1$ that can be accounted for by the scatter in various properties of their progenitors as a function of progenitor redshift. (See Eqn. 1 and §3.1 for an explanation of the notation.) The teal curve with triangular points corresponds to $\Delta \log \Sigma M_*$, where ΣM_* is the sum of the stellar masses of all the progenitors of each $z = 0.1$ galaxy. The red line with solid circular points shows $\Delta \log \Sigma M_{\rm bary}$, where $\Sigma M_{\rm bary}$ is the sum of the baryonic masses ($M_{\rm gas} + M_*$) of the progenitors at each redshift. The green line with solid square points shows $\Delta \log M_{\rm bary}$, where $M_{\rm bary}$ is the baryonic mass within the host halo of the main progenitor galaxy. The magenta line with open circular points corresponds to $\Delta \log \Sigma(M_{\rm 200c} - M_{\rm bary})$, where $\Sigma(M_{\rm 200c} - M_{\rm bary})$ is the sum of the total halo masses of each galaxy's progenitors, minus the mass of their baryonic components. The blue curve with open square points corresponds to $\Delta \log M_{\rm 200c,DMO}$, where $M_{\rm 200c,DMO}$ is the mass of the DMO halo corresponding to the host halo of the main progenitor galaxy in the reference simulation.

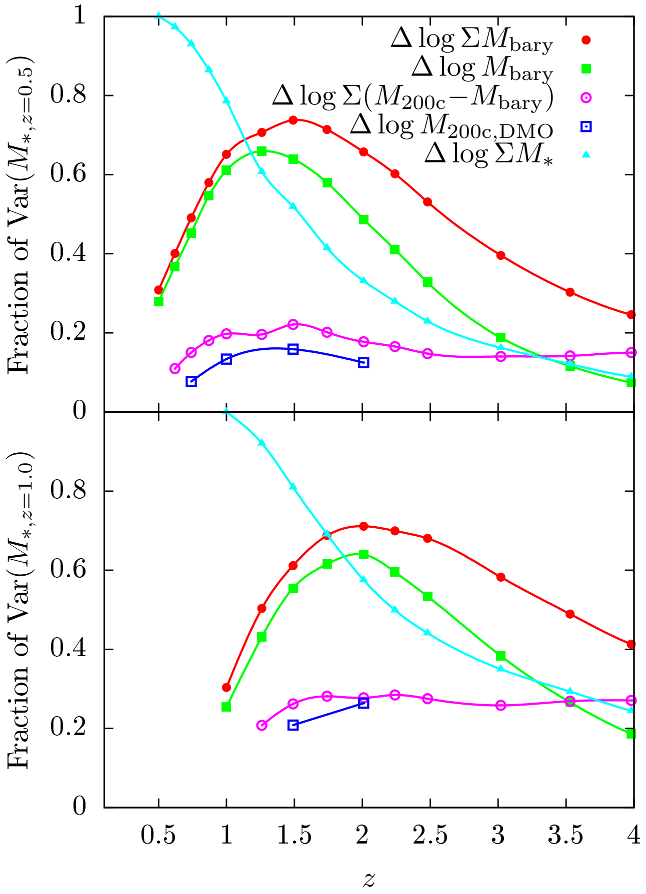

Figure 6. Same as Figure 5, but at $z = 0.5$ (top panel) and $z = 1.0$ (bottom panel), with all "Δ" values calculated with respect to the DMO halo properties at those redshifts. The galaxy samples at both redshifts consist of central galaxies whose host halos have $M_{\rm 200c} > 10^{11}{\rm M}_\odot$.

accounted for is 60%, and for $M_{\rm 200c,DMO} > 10^{12.5}{\rm M}_\odot$, the values are $z \approx 0.45$ and 44%.

As noted previously, when calculating the correlation between $\Delta \log M_*$ and $\Delta \log \Sigma M_{\rm bary}$, we exclude at each redshift the galaxies for which any of the progenitor subhalos are "flybys" within a non-progenitor halo, excluding a maximum of 17% of the sample at $z = 2.24$ (see §2.2). If we compute $\Sigma M_{\rm bary}$ for sets of progenitors that include such subhalos, such that we simply use in the sum the baryonic mass bound to the flyby subhalo while it is in the larger halo, we find that the variance of $\Delta \log M_*(z = 0.1)$ accounted for at the redshift of peak correlation drops from 76% to 68%. This is because flyby subhalos are generally extreme outliers in $M_{\rm bary}$, most likely because their baryon fraction changes quickly while they pass through the larger halo.

The scatter in the baryonic masses of the progenitors of $z = 0.1$ galaxies results from a combination of scatter in the progenitor halo masses and scatter in the baryon mass fraction of the halo ($M_{\rm bary}/M_{\rm 200c}$). The contribution to the variance of $\Delta \log M_*(z = 0.1)$ by scatter in the main progenitor DMO halo mass, $M_{\rm 200c,DMO}$, is shown as the blue line with open square points. The magenta line with open circular points shows the contribution to the M_* variance by the sum of the $M_{\rm 200c}$ halo masses from the reference simulation, minus their baryonic component, denoted $\Sigma(M_{\rm 200c} - M_{\rm bary})$. Both curves show that the contribution to the variance of $\Delta \log M_*(z = 0.1)$ from differing progenitor halo masses is quite low, implying that the majority of the scatter in $\Delta \log M_*$ is the result of scatter in the baryon fraction within galaxy progenitor halos.

To check whether the correlation between $\Delta \log M_*$ and $\Delta \log M_{\rm bary}$ is specific to low redshifts, we recreate Figure 5 for samples of central galaxies at $z = 0.5$ and $z = 1.0$ and their progenitors. Specifically, we select all central galaxies at these two redshifts whose host halos have $M_{\rm 200c} > 10^{11}{\rm M}_\odot$, and match the host halos to the corresponding halos in the DMO simulation. This results in samples of 10241 and 10556 galaxies for $z = 0.5$ and 1.0, respectively. We then recompute all the properties shown in Figure 5 relative to $M_{\rm 200c,DMO}$ and $c_{\rm DMO}$ of the $z = 0.5$ and $z = 1.0$ samples. For both samples the variance of $\Delta \log M_*$ is $\approx (0.18~{\rm dex})^2$ for the full sample.

The results are shown in Figure 6, which uses the same symbols as Figure 5. The qualitative similarity between the trends in the two figures implies that most of the scatter in M_* is produced by scatter in the baryonic masses of progenitor halos at all redshifts up to at least $z = 1$. The redshift

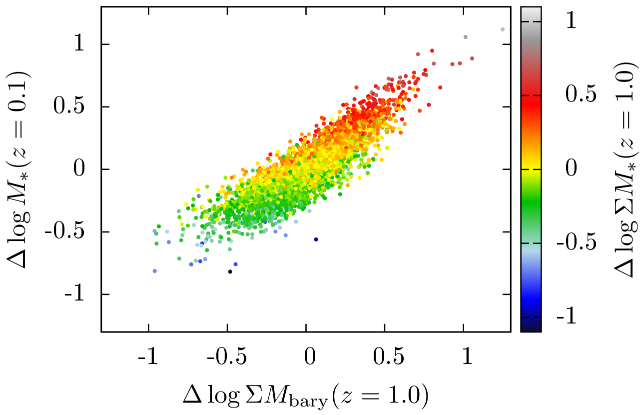

Figure 7. Same as the bottom centre panel of Figure 4, except now color-coded by $\Delta\log \Sigma M_*(z=1.0)$, where ΣM_* is the sum of the stellar masses of each galaxy's progenitors at $z=1.0$. In addition to the positive correlation between descendant stellar mass and progenitor baryonic mass at fixed $M_{\rm 200c,DMO}$ and $c_{\rm DMO}$, those progenitors with a higher ratio of stars to gas have descendants with higher stellar masses.

of peak correlation between $\Delta\log M_*$ and $\Delta\log M_{\rm bary}$ or $\Delta\log\Sigma M_{\rm bary}$ is essentially shifted by the redshift difference between the samples of galaxies. The fraction of the variance of $\Delta\log M_*$ accounted for at the peaks of the curves is 73% for the sample of galaxies at $z=0.5$ and 69% for that at $z=1.0$. The contribution to the scatter in $\Delta\log M_*$ from $\Delta\log M_{\rm 200c,DMO}$ and $\Delta\log\Sigma(M_{\rm 200c}-M_{\rm bary})$ appears to be larger for higher-redshift galaxy samples, although the scatter in the baryon fraction of the progenitors remains the dominant factor.

As shown above, scatter in the baryonic mass of progenitors produces most of the scatter in the $z=0.1$ $M_*-V_{\rm max}$ relation. However, the stellar mass of the $z=0.1$ descendants also depends somewhat on $\Delta\log M_*$ of the progenitors independently of its correlation with the baryonic mass. Figure 7 shows $\Delta\log M_*(z=0.1)$ versus $\Delta\log\Sigma M_{\rm bary}(z=1.0)$, colored by $\Delta\log\Sigma M_*(z=1.0)$. The progenitor stellar and gas masses at $z=1$ together account for a total of 86% of the variance of $\Delta\log M_*$ at $z=0.1$. For the galaxy samples at $z=0.5$ and $z=1.0$ the number is 83% for both samples.

It is important to note that all the correlations described above are calculated at only a single redshift of the simulation. Since gas physics is continuous in time, one would expect the baryonic mass in different snapshots to make independent contributions to the variance in M_*. For example, in Figure 7 we showed that there is an independent correlation between $\Delta\log\Sigma M_*(z=1.0)$ and $\Delta\log M_*(z=0.1)$ at fixed $\Delta\log\Sigma M_{\rm bary}(z=1.0)$; however, $\Delta\log\Sigma M_*(z=1.0)$ is itself highly correlated with the baryonic masses of the galaxies' progenitors at $z=2.0$, as shown in the lower panel of Figure 6. Thus the scatter of the $z=0.1$ stellar mass in EAGLE can be almost entirely accounted for by the evolution of the baryonic content within the progenitor halos.

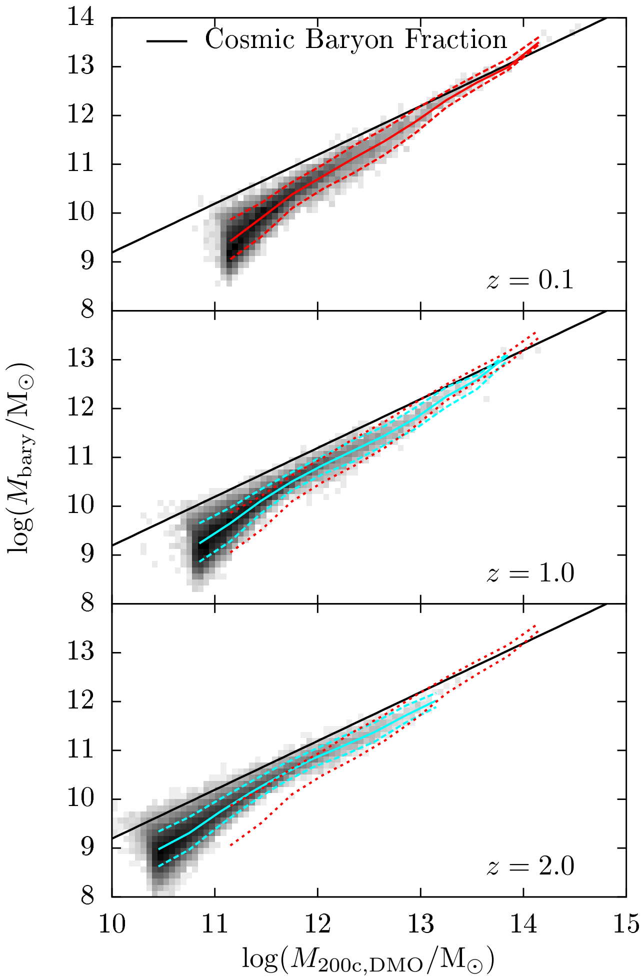

Figure 8. The total baryonic mass in each halo, including substructure, versus the matched DMO halo mass $M_{\rm 200c,DMO}$. The different panels show this relationship at three redshifts: top $z=0.1$, middle $z=1.0$, and bottom $z=2.0$. The halo mass limits for the bottom two panels are chosen to approximately encompass the masses of the main progenitors of the halos in the top panel. The shading represents the log-density of halos in each bin. The solid black line shows the baryonic mass expected if all halos contained the cosmic fraction of baryons. In the top panel, the red solid line represents the median $M_{\rm bary}$ as a function of $M_{\rm 200c,DMO}$, while the dashed lines demarcate the bottom and top deciles. In the lower two panels, the median and deciles are represented by cyan lines, and the deciles from the top panel are reproduced with red dotted lines for comparison. We are able to see that, for halos with $M_{\rm 200c,DMO}\lesssim 10^{13}{\rm M}_\odot$, the mean $M_{\rm bary}$ at fixed values of $M_{\rm 200c,DMO}$ decreases with time, and its scatter increases.

3.3 Evolution of the baryonic mass scatter

As shown in the previous section, most of the scatter in the $z=0.1$ $M_*-V_{\rm max}$ relation is the result of scatter in the baryon fraction of the halos hosting the galaxies' progenitors. This raises the question of what determines the baryon fraction.

In EAGLE, the baryonic mass within halos is primarily dependent on the halo mass. Figure 8 shows the evolution of the distribution of baryonic masses as a function

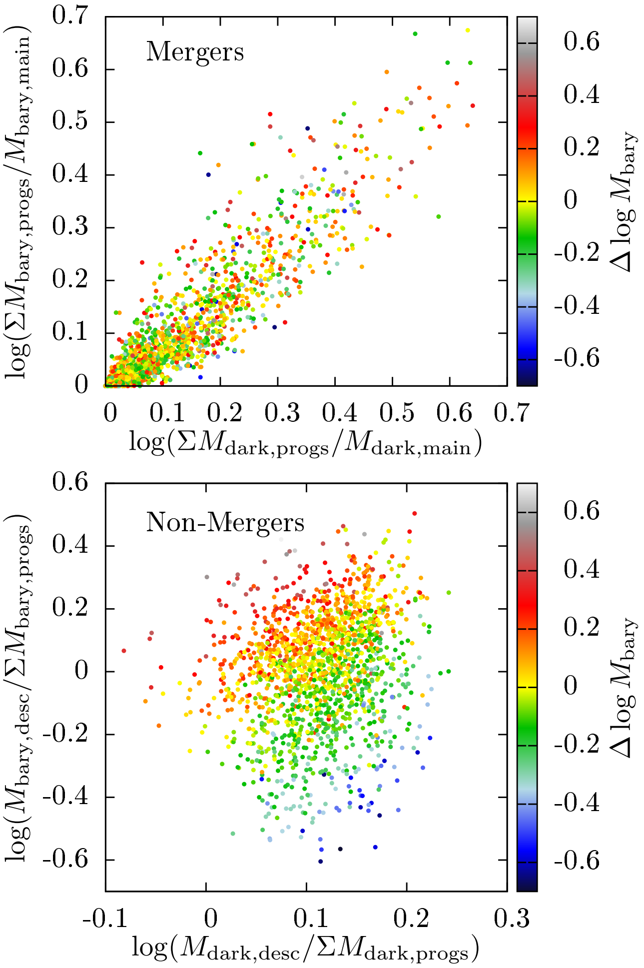

Figure 9. The influence of mergers and non-merger processes such as accretion and feedback on the evolution of $\Delta \log M_{\rm bary}$, the deviation of the baryonic mass of each halo relative to the mean at fixed $M_{\rm 200c,DMO}$ and $c_{\rm DMO}$ (see Eqn. 1). The halo sample comprises central subhalos at $z = 1$ whose main progenitors at $z = 2$ are within 0.02 dex of the mean $M_{\rm bary}$ as a function of $M_{\rm 200c,DMO}$ and $c_{\rm DMO}$ at $z = 2$. The color bar indicates $\Delta \log M_{\rm bary}$ of the descendants at $z = 1$ computed relative to their halo properties at this redshift, showing that $\Delta \log M_{\rm bary}$ has scattered significantly to both larger and smaller values. *Top Panel:* The growth in dark matter mass from mergers between $z = 2$ and $z = 1$ versus the growth in baryonic mass from mergers. The growth due to mergers is defined as the ratio of the sum of all the progenitor masses to the mass of only the main progenitor. The dark matter mass is the total mass in dark matter assigned to each FoF halo in the reference simulation. The growth in dark matter and baryonic mass resulting from mergers correlates poorly with the final $\Delta \log M_{\rm bary}$ of the halo. *Bottom Panel:* Same as the top panel, but for the change in mass not due to mergers (i.e. due to accretion and feedback). The change in mass not due to mergers is defined as the ratio of the mass of the descendant at $z = 1.0$ to the sum of all the progenitor masses at $z = 2.0$. The change in mass not due to mergers shows a far better correlation with $\Delta \log M_{\rm bary}$ of the descendant, implying that feedback and gas accretion are the dominant contributors to the evolution of the baryon fraction.

of $M_{\rm 200c,DMO}$. (The results using $M_{\rm 200c}$ from the reference simulation are very similar.) The sample comprises halos with $M_{\rm 200c} > 10^{11}{\rm M}_\odot$ for $z = 0.1$, $M_{\rm 200c} > 10^{10.7}{\rm M}_\odot$ for $z = 1.0$, and $M_{\rm 200c} > 10^{10.3}{\rm M}_\odot$ for $z = 2.0$. The masses of the latter two redshifts are chosen to approximately encompass the masses of the halos hosting the main progenitors of the $z = 0.1$ sample. The darkness of the shading is proportional to the log of the number of halos in each bin. The solid black line in each panel shows the baryonic mass that would be expected if each halo contained the cosmic baryon fraction time $M_{\rm 200c}$.

At $z = 0.1$, the median value of $M_{\rm bary}$ as a function of $M_{\rm 200c,DMO}$ is represented by a solid red line, and the top and bottom deciles are shown with red dashed lines. For high-mass halos ($M_{\rm 200c,DMO} \gtrsim 10^{13}{\rm M}_\odot$), which are very low in number in EAGLE, the baryon fraction is close to the cosmic value. However, for lower-mass halos, the mean baryon fraction is significantly lower.

In the lower two panels, the median value of $M_{\rm bary}$ is shown with a solid cyan line, and the top and bottom deciles are represented by dashed cyan lines. The deciles at $z = 0.1$ are replicated as red dotted lines. By comparing the top and bottom deciles at $z = 0.1$ to those at $z = 1.0$ and $z = 2.0$, we see that for halos with $M_{\rm 200c,DMO} \lesssim 10^{13}{\rm M}_\odot$, the mean baryon fraction at fixed $M_{\rm 200c,DMO}$ decreases with cosmic time and the scatter in the baryon fraction increases.

Halos in EAGLE undergo continuous evolution in the value of their baryonic mass relative to their halo mass, so $\Delta \log M_{\rm bary}$ at low redshift ($z \approx 0$) is uncorrelated with that at high redshift ($z \gtrsim 4$). Evolution in $\Delta \log M_{\rm bary}$ results from change in both the dark matter mass and the baryonic mass of a halo, as well as the mean evolution of the sample of halos. To determine the primary mechanism that sets the value of $\Delta \log M_{\rm bary}$, we wish to compare the evolution of this value for each halo to the change in the halo's dark matter and baryonic mass resulting from different physical processes — specifically, halo mergers versus non-merger processes such as accretion and feedback.

We select a sample consisting of halos at $z = 1.0$ with $M_{\rm 200c} > 5 \times 10^{10}{\rm M}_\odot$ whose main progenitors at $z = 2.0$ have a baryonic mass within 0.02 dex of the mean value for their $M_{\rm 200c,DMO}$ and $c_{\rm DMO}$. Stated differently, the $z = 2.0$ main progenitors have $|\Delta \log M_{\rm bary}| < 0.02$ relative to their $z =$ 2.0 halo properties. We then compute $\Delta \log M_{\rm bary}(z = 1.0)$, the deviation of $M_{\rm bary}$ from the mean at fixed $M_{\rm 200c,DMO}$ and $c_{\rm DMO}$ at $z = 1.0$. For the descendant halos at $z = 1.0$, the standard deviation of $\Delta \log M_{\rm bary}$ has increased to 0.19 dex, due to evolution in the baryonic and dark matter masses of each halo since $z = 2.0$.

In order to consistently track the co-evolution of the dark matter and baryonic masses, we use the total dark matter mass assigned to each FoF halo in the reference simulation, denoted $M_{\rm dark}$. The baryonic mass $M_{\rm bary}$ is the bound mass in baryons contained in all the subhalos of each FoF halo, as defined previously. We note that due to the influence of baryonic physics, there are differences between the evolution of the FoF halo mass in the reference simulation and that of the corresponding halo in the DMO simulation. For the mass range of the descendant sample considered here, the scatter between the FoF halo mass in the reference and DMO simulations is < 0.06 dex and decreases sharply with halo mass. We use $M_{\rm dark}$ rather than the dark matter mass within M_{200c} because the former is more reflective of the accretion of dark matter onto the halo.

In Figure 9 we show the change in halo dark matter and baryonic mass between $z = 2.0$ and $z = 1.0$ compared to $\Delta \log M_{\rm bary}(z = 1.0)$ for each halo. The color of each point represents $\Delta \log M_{\rm bary}$ at $z = 1.0$, which has evolved from a value of ≈ 0 at $z = 2.0$. The top panel of Figure 9 shows the mass growth due to mergers, which we approximate as the ratio of the sum of the masses of all the progenitors at $z = 2.0$ to the mass of the main progenitor[2]: $\Sigma M_{\rm progs}/M_{\rm main}$. Because the set of all progenitors includes the main progenitor, the mass change due to mergers is positive by definition. The vertical axis shows the growth in the baryonic mass and the horizontal axis shows the growth in dark matter mass.

The growth in baryonic mass from mergers tends to follow the growth in dark matter mass. Due to the low typical mass of halos in our sample, the majority do not gain a large amount of mass via mergers. However, $\Delta \log M_{\rm bary}$ of the descendant halo at $z = 1.0$ is effectively uncorrelated with the mass growth from mergers, even for those halos that experience a significant amount of such growth. This suggests that mergers are not the primary cause of change in $\Delta \log M_{\rm bary}$ over time.

The lower panel of Figure 9 shows the change in dark matter and baryonic mass due to non-merger processes, i.e. gas loss due to feedback and accretion of dark matter and/or gas. The mass change due to non-merger processes is approximated as the ratio of the mass of the $z = 1.0$ descendant to the sum of the masses of all its progenitors at $z = 2.0$: $M_{\rm desc}/\Sigma M_{\rm progs}$. The dark matter mass of the descendant is generally larger than the total dark matter mass of the progenitors, but in some cases it can be smaller, perhaps because of ejection of matter during mergers. The baryonic mass of the descendant, on the other hand, is frequently smaller than the sum of the baryonic masses of its progenitors. This indicates that feedback plays a very important role in changing the baryonic mass.

Furthermore, in contrast to the top panel, $\Delta \log M_{\rm bary}$ of the $z = 1.0$ descendant halos correlates clearly with the mass change caused by mechanisms other than mergers. The majority of the evolution in $\Delta \log M_{\rm bary}$ is attributable to change in the baryonic mass at fixed values of accreted dark matter mass. We conclude that the evolution of $\Delta \log M_{\rm bary}$ over time is mainly due to inflow and outflow of gas via feedback and smooth accretion, rather than mergers.

4 DISCUSSION AND CONCLUSIONS

The EAGLE cosmological hydrodynamical simulation was previously used in Matthee et al. (2017) and Chaves-Montero et al. (2016) to investigate the relationship between stellar mass M_* and dark matter halo properties from the dark matter-only (DMO) run of EAGLE, so as to determine the best parameter to use in halo abundance matching. Both found that for central galaxies, the maximum circular velocity of the corresponding DMO halo, $V_{\rm max,DMO}$, correlates better with the stellar mass than the DMO halo mass does, and that this relationship has a mass-dependent scatter that is ≈ 0.2 dex for halos with $M_{200c} > 10^{11}{\rm M}_\odot$ at $z = 0.1$. Matthee et al. (2017) investigated whether the scatter in M_* correlates with any other DMO halo properties, such as the halo half-mass assembly time, sphericity, spin, triaxiality, and environment, but found no additional correlations.

In this paper, we have examined the source of the scatter in M_* at fixed $V_{\rm max,DMO}$ for central galaxies by considering different baryonic (rather than dark matter) properties correlated with the scatter. We used the same sample of central galaxies as Matthee et al. (2017), and the corresponding host halos from the DMO run of EAGLE. Our main conclusion is that the scatter in M_* at fixed $V_{\rm max,DMO}$ can be traced primarily to the scatter in the baryon fraction of the host halos of the galaxy progenitors.

In EAGLE, the baryonic mass of halos correlates primarily with the halo mass. At high redshifts, the initial conditions are such that all halos have approximately the cosmic ratio of baryons to dark matter. However, the mean baryonic mass at fixed halo mass for halos with $M_{200c} \lesssim 10^{13}{\rm M}_\odot$ (which constitute the majority of our halo sample) decreases with cosmic time, and the scatter in the baryonic mass at fixed halo mass increases, as shown in Figure 8.

The star formation rate of a halo's central galaxy depends on the central gas density, such that for an equal gas reservoir, a halo with a higher central density will produce more stars. Furthermore, a higher density implies a higher binding energy and hence less efficient feedback for a fixed rate of energy injection. In addition, more concentrated halos tend to form earlier, allowing more time for star formation to take place. For these reasons, the stellar mass formed at fixed halo mass is higher for halos with higher concentrations, resulting in the stellar mass being better correlated with $V_{\rm max,DMO}$ than $M_{\rm 200c,DMO}$. However, as described above, the baryon content of halos of the same halo mass and concentration has a substantial scatter. As a result, two halos with similar assembly histories but different baryonic mass fractions can produce descendant halos with the same halo mass and concentration but significantly different stellar mass content. We calculate the correlation of the scatter in the central stellar mass at fixed DMO halo mass and concentration with the scatter in the baryonic mass of the galaxy progenitors.

The strongest correlation between the scatter in $z = 0.1$ stellar mass and the scatter in the main progenitor baryonic mass is achieved at $z \approx 0.95$, where it is able to account for 68% of the variance in the $z = 0.1$ $M_* - V_{\rm max,DMO}$ relation for halos with $M_{200c} > 10^{11}{\rm M}_\odot$ (Figure 5). The correlation with the sum of the baryonic masses of all the progenitors is slightly better, peaking for progenitors at $z \approx 1.1$, which account for 76% of the variance in the $z = 0.1$ $M_* - V_{\rm max,DMO}$ relation. Similar trends are seen in Figure 6 for samples of central galaxies at $z = 0.5$ and $z = 1.0$ having halo masses greater than $10^{11}{\rm M}_\odot$, with the location of the peak correlation shifted to $z \approx 1.5$ and $z \approx 2.0$, respectively.

The scatter in the baryonic mass within halos also produces an inverse correlation between the central galaxy's stellar mass and stellar population age at fixed DMO halo mass and concentration, shown in the top panel of Figure

[2] This is an approximation because any mass accreted onto (or lost from) the non-main progenitors after $z = 2$ but before they merge with the main progenitor will not be considered mass change from mergers but rather from non-mergers (second panel of Figure 9).

3. The halos with more massive central galaxies at $z = 0.1$ are those that had a larger amount of recent star formation due to their larger baryon reservoir, causing their central galaxies to be more massive and younger.

Finally, we determined that non-merger processes, such as gas accretion and feedback, are what primarily set the baryonic mass within halos. The complex and stochastic nature of feedback likely explains the lack of significant correlation with the DMO halo properties examined in Matthee et al. (2017). In a companion paper (Kulier et al. 2018, in prep), we describe in detail the origin of variations in feedback strength for different halo mass ranges and timescales and its correlates.

ACKNOWLEDGEMENTS

The authors would like to thank Jorryt Matthee for very useful comments on the first draft of the paper. This work was supported by the Netherlands Organisation for Scientific Research (NWO), through VICI grant 639.043.409. AK acknowledges support from CONICYT-Chile grant FONDECYT Postdoctorado 3160574.

REFERENCES

Booth C. M., Schaye J., 2010, MNRAS, 405, L1
Chabrier G., 2003, PASP, 115, 763
Chaves-Montero J., Angulo R. E., Schaye J., Schaller M., Crain R. A., Furlong M., Theuns T., 2016, MNRAS, 460, 3100
Crain R. A. et al., 2015, MNRAS, 450, 1937
Cullen L., Dehnen W., 2010, MNRAS, 408, 669
Dalla Vecchia C., Schaye J., 2012, MNRAS, 426, 140
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
De Lucia G., Blaizot J., 2007, MNRAS, 375, 2
Dehnen W., Aly H., 2012, MNRAS, 425, 1068
Dolag K., Borgani S., Murante G., Springel V., 2009, MNRAS, 399, 497
Dubois Y., Peirani S., Pichon C., Devriendt J., Gavazzi R., Welker C., Volonteri M., 2016, MNRAS, 463, 3948
Durier F., Dalla Vecchia C., 2012, MNRAS, 419, 465
Ferland G. J., Korista K. T., Verner D. A., Ferguson J. W., Kingdon J. B., Verner E. M., 1998, PASP, 110, 761
Furlong M. et al., 2017, MNRAS, 465, 722
Furlong M. et al., 2015, MNRAS, 450, 4486
Genel S. et al., 2014, MNRAS, 445, 175
Haardt F., Madau P., 2001, in Clusters of Galaxies and the High Redshift Universe Observed in X-rays, Neumann D. M., Tran J. T. V., eds.
Hopkins P. F., 2013, MNRAS, 428, 2840
Jeeson-Daniel A., Dalla Vecchia C., Haas M. R., Schaye J., 2011, MNRAS, 415, L69
Jiang L., Helly J. C., Cole S., Frenk C. S., 2014, MNRAS, 440, 2115
Kennicutt, Jr. R. C., 1998, ApJ, 498, 541
Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, ApJ, 609, 35
Matthee J., Schaye J., Crain R. A., Schaller M., Bower R., Theuns T., 2017, MNRAS, 465, 2381
McAlpine S. et al., 2016, Astronomy and Computing, 15, 72
Peacock J. A., Smith R. E., 2000, MNRAS, 318, 1144
Planck Collaboration et al., 2014, A&A, 571, A16
Prada F., Klypin A. A., Cuesta A. J., Betancort-Rijo J. E., Primack J., 2012, MNRAS, 423, 3018
Price D. J., 2008, Journal of Computational Physics, 227, 10040
Qu Y. et al., 2017, MNRAS, 464, 1659
Reddick R. M., Wechsler R. H., Tinker J. L., Behroozi P. S., 2013, ApJ, 771, 30
Rosas-Guevara Y., Bower R. G., Schaye J., McAlpine S., Dalla Vecchia C., Frenk C. S., Schaller M., Theuns T., 2016, MNRAS, 462, 190
Schaller M., Dalla Vecchia C., Schaye J., Bower R. G., Theuns T., Crain R. A., Furlong M., McCarthy I. G., 2015a, MNRAS, 454, 2277
Schaller M. et al., 2015b, MNRAS, 451, 1247
Schaye J., 2004, ApJ, 609, 667
Schaye J. et al., 2015, MNRAS, 446, 521
Schaye J., Dalla Vecchia C., 2008, MNRAS, 383, 1210
Seljak U., 2000, MNRAS, 318, 203
Sinha M., Holley-Bockelmann K., 2012, ApJ, 751, 17
Skibba R. A., Macciò A. V., 2011, MNRAS, 416, 2388
Springel V., 2005, MNRAS, 364, 1105
Springel V., Di Matteo T., Hernquist L., 2005, MNRAS, 361, 776
Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726
Trayford J. W., Theuns T., Bower R. G., Crain R. A., Lagos C. d. P., Schaller M., Schaye J., 2016, MNRAS, 460, 3925
Vale A., Ostriker J. P., 2004, MNRAS, 353, 189
Vale A., Ostriker J. P., 2006, MNRAS, 371, 1173
Vogelsberger M. et al., 2014a, Nature, 509, 177
Vogelsberger M. et al., 2014b, MNRAS, 444, 1518
Wechsler R. H., Bullock J. S., Primack J. R., Kravtsov A. V., Dekel A., 2002, ApJ, 568, 52
Wiersma R. P. C., Schaye J., Smith B. D., 2009a, MNRAS, 393, 99
Wiersma R. P. C., Schaye J., Theuns T., Dalla Vecchia C., Tornatore L., 2009b, MNRAS, 399, 574
Wong A. W. C., Taylor J. E., 2012, ApJ, 757, 102