ABSTRACT
In this supplementary material we give a full description of the treatment of astrophysical processes in our 2014 model of galaxy formation. This model is built on subhalo merger trees constructed from the Millennium and Millennium-II simulations after scaling to represent the first-year Planck cosmology. A set of coupled differential equations allow us to follow the evolution of six baryonic components. Five of these are associated with individual galaxies – a hot gas atmosphere, cold interstellar gas, a reservoir of gas ejected in winds, stars split into bulge, disk and intracluster light components, and central supermassive black holes. The sixth, diffuse primordial gas, is associated with dark matter which is not yet part of any halo. Primordial gas falls with the dark matter onto sufficiently massive halos, where it is shock-heated. The efficiency of radiative cooling then determines whether it is added directly to the cold gas of the central galaxy, or resides for a while in a hot gas atmosphere. Cold interstellar gas forms stars both quiescently and in merger-induced starbursts which also drive the growth of central supermassive black holes. Stellar evolution not only determines the photometric appearance of the final galaxy, but also heats and enriches its gas components, in many cases driving material into the wind reservoir, from which it may later fall back into the galaxy. Accretion of hot gas onto central black holes gives rise to "radio-mode" feedback, regulating condensation of hot gas onto the galaxy. Environmental processes like tidal and ram-pressure stripping and merging affect the gas components of galaxies, as well as the partition of stars between disks, bulges and the intracluster light, a diffuse component built from tidally disrupted systems. Disk and bulge sizes are estimated from simple energy and angular momentum-based arguments.

1 INTRODUCTION
The "Munich" model of galaxy formation is a semi-analytic scheme for simulating the evolution of the galaxy population as a whole and has been continually developed over the last quarter century (White 1989; White & Frenk 1991; Kauffmann et al. 1993, 1999; Springel et al. 2001, 2005). The 2005 completion of the Millennium Simulation enabled implementation of the model on dark matter simulations of high enough resolution to detect the structures associated with the formation of individual galaxies throughout cosmologically relevant volumes. Updates to the baryonic physics have resulted in a series of publicly released galaxy/halo/subhalo catalogues that have been widely used by the community (Croton et al. 2006; De Lucia & Blaizot 2007; Bertone et al. 2007; Guo et al. 2011, 2013). 1 The model of the current paper updates that of Guo et al. (2011), aiming at better representation of the observed build-up over time and of the present star formation activity of the low-mass galaxy population. Guo et al. (2011) itself updated earlier treatments of supernova feedback and of galaxy mergers in order to agree better with observations of dwarf and satellite galaxies. It also introduced detailed tracking of the angular momentum of different galaxy components so that the size evolution of disks and bulges could be followed. Finally, Guo et al. (2013) implemented the procedure of Angulo & White (2010) so that the Millennium Simulation could be used to model evolution in cosmologies other than its native WMAP1 cosmology.

1 See http://www.mpa-garching.mpg.de/millennium
In this Supplementary Material we aim to give a detailed and fully self-contained summary of the treatment of baryonic physics in our current model. Many aspects of this are unchanged since earlier models but repetition of material in a single coherent and complete description seems preferable to referring each model element back to the particular earlier paper where it was first used. We anticipate updating this supplementary material as future versions of our model are released, so that each will have its own full astrophysics and algorithmic summary.

1.1 Dark Matter Simulations

The galaxy formation model of this paper is built on subhalo merger trees describing the evolution of dark matter structures in two large dark matter simulations, the Millennium (Springel et al. 2005) and Millennium-II (Boylan-Kolchin et al. 2009) simulations. Both assume a ΛCDM cosmology with parameters derived by a combined analysis of the 2dFGRS (Colless et al. 2001) and the first-year WMAP data (Spergel et al. 2003). For this work the original cosmology has been scaled, using the Angulo & White (2010) technique, as updated by Angulo & Hibert (2014), to represent the best-fit cosmological parameters derived from the first-year Planck data. The underlying cosmology of the dark matter simulations and thus the galaxy formation model is then: \( \sigma_8 = 0.829, H_0 = 73 \text{ km s}^{-1} \text{ Mpc}^{-1}, \Omega_m = 0.285, \Omega_b = 0.045 \) and \( n = 1.0 \). For this simulation the particle data were stored in 64 and 68 output snapshots, respectively, for the Millennium and Millennium-II with the last 60 overlapping between the two simulations. After rescaling, the last five snapshots of each simulation correspond to the future, and \( z = 0 \) corresponds to the sixth from last of the original snapshots. At each time the data were post-processed in order to produce a friend-of-friends (FOF) group catalogue by joining particles separated by less than 20% of the mean interparticle spacing (Davis et al. 1985). The SUBFIND algorithm (Springel et al. 2001) was then applied to identify all the self-bound substructures in each FOF group. The radius of the FOF group is defined as the radius of the largest sphere centered on the potential minimum which contains an overdensity larger than 200 times the critical value. The group mass is then the total mass within this sphere and other group properties are related by:

\[
M_{200c} = \frac{100}{G} H^2(z) R_{200c}^3 = \frac{V_{200c}^3}{10GH(z)},
\]

where \( H(z) \) is the Hubble constant at redshift \( z \).

Every subhalo in a given snapshot which contains 20 or more bound particles is connected to a unique descendant in the subsequent snapshot and these links are then used to build subhalo merger trees which encode the assembly history of every subhalo identified at \( z = 0 \). These trees are the basis on which the galaxy formation model is constructed (see Springel et al. 2005). They allow us to build much more realistic satellite galaxy populations than would be possible using trees linking the FOF halos themselves. The most massive subhalo in each FOF group is usually much bigger than all the others, and is defined as the “main halo”; the group central galaxy (which we often refer to as a “type 0” galaxy) is located at the minimum of the potential of this main halo. All other bound subhalos contain satellite galaxies at their centres (type 1’s). In addition, our galaxy formation model follows satellites which have already lost their own dark matter subhalos but which are yet to merge with the central galaxy. Such objects are referred to as “type 2” galaxies or “orphan” satellites. Their position and velocity are tied to those of the dark matter particle that was the most bound within their subhalo at the last time that this was identified by SUBFIND with at least 20 particles.

1.2 Overview of the galaxy formation physics

Our model for galaxy formation starts by assigning a cosmic abundance of baryons to each collapsed dark matter halo. Subsequent growth brings its fair share of baryons in the form of primordial diffuse gas which shock-heats and then either cools immediately onto the disk of the central galaxy, or is added to a quasi-static hot atmosphere which accretes more slowly through a cooling flow. The disk of cold gas fuels the formation of stars which eventually die, releasing energy, mass and heavy elements into the surrounding medium. This energy reheats cold disk gas, injecting it into a hot atmosphere, which may itself also be ejected into an external reservoir to be reincorporated only at some much later time. Black holes are assumed to grow primarily through the accretion of cold gas during mergers, but also through quiescent accretion from the hot atmosphere, which releases energy which can counteract the cooling flow. This form of feedback eventually curtails star formation in the most massive systems. A number of environmental processes act on satellites as soon as they cross the virial radius of their host. Tidal forces are assumed to remove hot gas, cold gas and stars while hot gas is also removed by ram-pressure stripping. These processes gradually quench star formation, particularly in satellites orbiting within more massive systems. As dark matter subhalos merge, so do their associated galaxies, although with some delay. Once a subhalo is fully disrupted, its galaxy spirals into the central galaxy, merging after a dynamical friction time and creating a bulge and a burst of star formation. Bulges also form through secular processes whenever disks become sufficiently massive to be dynamically unstable. Finally, the light emitted from stellar populations of different ages is computed via population synthesis models and dust extinction corrections are applied. The uncertain efficiencies and scalings characterising all these physical processes are simultaneously determined by using MCMC techniques to fit a set of calibration observations (in this paper, abundances and passive fractions as a function of stellar mass at a variety of redshifts).
1.3 Infall and reionization

Following the standard White & Frenk (1991) approach we assume that each collapsed dark matter structure will, at every time, have a mass of associated baryons given by the cosmic mean baryon fraction, $f_{\text{cos}} = 15.5\%$ for the PLANCK cosmology. As halos grow, we assume that matter that was not previously part of any object is added in these same proportions, with the baryons in the form of diffuse primordial gas which shock-heats on accretion, thereafter either cooling again immediately or being added to a quasi-static hot atmosphere.

For sufficiently low-mass halos and over a large part of cosmic history this simple picture needs modification, since photo-heating by the UV background field raises the temperature of diffuse intergalactic gas to the point where pressure effects prevent it from accreting onto halos with the dark matter (Efstathiou 1992). In order to model this, we use results from Gneden (2000) who defines a filtering halo mass, $M_F(z)$, below which the baryonic fraction is reduced with respect to the universal value according to:

$$f_b(z, M_{200c}) = f_{\text{cos}} \left(1 + \left(2^{n/3} - 1\right) \left[M_{200c} / M_F(z)\right]^{2/3} \right)^{-\alpha/3}.$$  \hfill (S2)

For halos with $M_{200c} > M_F$ suppression of the baryon fraction is small, but for halos with $M_{200c} \ll M_F(z)$ the baryon fraction drops to $(M_{200c} / M_F(z))^{2/3}$. We adopt $\alpha = 2$ and take $M_F(z)$ from the numerical results of Okamoto et al. (2008). $M_F$ varies from $\sim 6.5 \times 10^9 M_\odot$ at $z = 0$, to $\sim 10^7 M_\odot$ just before reionization starts at $z = 8$.

1.4 Cooling modes

Infalling diffuse gas is expected to shock-heat as it joins a halo. At early times and for low-mass halos the accretion shock happens close to the central object and the post-shock cooling time is short enough that new material settles onto the cold gas disk at essentially the free-fall rate. At later times and for higher mass halos the accretion shock moves away from the central object, settling at approximately the virial radius, while the post-shock cooling time exceeds the halo sound crossing time. The shocked heated gas then forms a quasi-static hot atmosphere from which it can gradually accrete to the centre via a cooling flow. The halo mass separating these two regimes is $\sim 10^{12} M_\odot$ (White & Rees 1978; White & Frenk 1991; Forcada-Miro & White 1997; Birnboim & Dekel 2003). In a realistic, fully three-dimensional situation a hot quasi-static atmosphere can coexist with cold inflowing gas streams in halos near the transition mass (Kereš et al. 2005; Nelson et al. 2013) but the overall rate of accretion onto the central object remains similar to that given by the formulae below (Benson et al. 2001; Yoshida et al. 2002).

Following the formulation of White & Frenk (1991) and Springel et al. (2001), we assume that, in the quasi-static regime, gas cools from a hot atmosphere where its distribution is isothermal. The cooling time is then given by the ratio between the thermal energy of the gas and its cooling rate per unit volume:

$$t_{\text{cool}}(r) = \frac{3 \mu m H k T_{\text{gas}}}{2 \rho_{\text{hot}}(r) \Lambda(T_{\text{hot}}, Z_{\text{hot}})}.$$

where $\mu m_H$ is the mean particle mass, $k$ is the Boltzmann constant, $\rho_{\text{hot}}(r)$ is the hot gas density and $Z_{\text{hot}}$ is the hot gas metallicity. $T_{\text{hot}}$ is the temperature of the hot gas which is assumed to be the virial temperature of the halo given by $T_{200c} = 35.9 (V_{200c}/\text{km s}\^{-1})^2$ K (for subhalos we use this temperature as estimated at infall). $\Lambda(T_{\text{hot}}, Z_{\text{hot}})$ is the equilibrium cooling function for collisional processes which depends both on the metallicity and temperature of the gas but ignores radiative ionization effects (Sutherland & Dopita 1993). The hot gas density as a function of radius for a simple isothermal model is given by:

$$\rho_{\text{hot}}(r) = \frac{M_{\text{hot}}}{4\pi R_{\text{cool}}^3 \rho_{\text{hot}}(0)^2},$$

and assuming that the cooling radius is where the cooling time equals the halo dynamical time:

$$r_{\text{cool}} = \left[ \frac{t_{\text{dyn}, h} M_{\text{hot}} \Lambda(T_{\text{hot}}, Z_{\text{hot}})}{6 \pi \mu m H k T_{200c} R_{200c}} \right]^{1/2}.$$  \hfill (S5)

where $t_{\text{dyn}, h}$ is the halo dynamical time defined as $R_{\text{vir}} / V_{\text{vir}} = 0.1 H(z)^{-1}$ (De Lucia et al. 2004). The specific choice of coefficient for the dynamical time of the halo is, of course, somewhat arbitrary.

When $r_{\text{cool}} < R_{200c}$, we assume that the halo is in the cooling flow regime with gas cooling from the quasi-static hot atmosphere at a rate:

$$M_{\text{cool}} = M_{\text{hot}} \frac{r_{\text{cool}}}{R_{200c}} \frac{1}{t_{\text{dyn}, h}}.$$  \hfill (S6)

When $r_{\text{cool}} > R_{200c}$, the halo is in the rapid infall regime and material accretes onto the central object in free fall, thus on the halo dynamical time:

$$M_{\text{cool}} = \frac{M_{\text{hot}}}{t_{\text{dyn}, h}}.$$  \hfill (S7)

This particular formula for rapid infall was introduced in Guo et al. (2011) in order to ensure a smooth transition between the two regimes.

1.5 Disk formation and angular momentum

As primordial material accretes onto a halo, its dark matter and baryonic components are expected to have similar specific angular momenta. Some of this gas is subsequently added to the central galaxy, and its remaining angular momentum then determines the radius at which it settles within the galactic disk. We follow these processes using the simple model introduced by Guo et al. (2011). The properties of the cold gas and stellar disks are calculated separately and their time evolution is modelled explicitly. The angular momentum of the cold gas disk changes as a result of star formation and of gas accretion through cooling and minor merger events:

$$\Delta \bar{J}_{\text{gas}} = \delta \bar{J}_{\text{gas, cooling}} + \delta \bar{J}_{\text{gas,SF}} + \delta \bar{J}_{\text{gas,acc}}.$$  \hfill (S8)
$\mathbf{\delta \vec{J}}$ which is removed from the gas disk by star formation events, (the mean specific angular momentum of the cold gas disk)

stellar disks, respectively, and the rotation velocity of both disks is approximated by the maximum circular velocity of their host halo, $V_{\text{max}}$. This simple picture needs modification if baryons have a significant impact on the inner structure of their dark matter halos. We refer the reader to Guo et al. (2011) for further discussion and for a comparison between predicted and observed disk sizes.

With these assumptions, the surface density profiles of the gas and stellar disks are given by:

\[ \Sigma_{\text{gas}}(R) = \Sigma_{\text{gas},0} \exp(-R/R_{\text{gas}}) \]  

and

\[ \Sigma_{\star}(R) = \Sigma_{\star,0} \exp(-R/R_{\star}) \]

where $\Sigma_{\text{gas},0} = M_{\text{gas}}/2\pi R_{\text{gas}}^2$ and $\Sigma_{\star,0} = M_{\star}/2\pi R_{\star}^2$ are the central surface densities of the cold gas and stellar disks.

1.6 Star formation

As noted in the last section, stars are assumed to form from cold gas within the disk of each galaxy. The star formation rate is taken to be:

\[ M_{\star} = \alpha_{\text{SF}} \left( \frac{M_{\text{gas}} - M_{\text{crit}}}{t_{\text{dyn,disk}}} \right) \]

where $M_{\text{gas}}$ is again the total mass of cold gas, $t_{\text{dyn,disk}} = R_{\star}/V_{\text{max}}$ is the dynamical time of the disk, and $M_{\text{crit}}$ is a threshold mass (see below). From the total mass of stars formed, $M_{\star}$, we assume that a fraction $R_{\text{ret}}$ is associated with massive, short-lived, stars and is immediately returned to the cold gas. $R_{\text{ret}} = 0.43$ is determined from the Chabrier (2003) initial mass function. Thus the stellar mass of the disk is augmented by $\delta M_{\star} = (1 - R_{\text{ret}})M_{\star}\delta t$ and the cold disk mass is reduced by the same amount.

Applying the arguments of Kauffmann (1996) we set
the threshold mass for star formation, $M_{\text{crit}}$, to be:

$$M_{\text{crit}} = M_{\text{crit},0} \left( \frac{V_{\text{esc}}}{200 \text{ km s}^{-1}} \right) \left( \frac{R_{\text{gas}}}{10 \text{ kpc}} \right). \quad (S15)$$

Since Kauffmann et al. (1999) all versions of the Munich model have adopted $M_{\text{crit},0} = 3.8 \times 10^8 M_\odot \text{pc}^{-2}$ which still appears tenable in comparison with some recent observations in the Milky Way (Lada et al. 2010; Heiderman et al. 2010). However, this and other work (Bigiel et al. 2008; Leroy et al. 2008) suggest that star formation should be linked explicitly to a molecular gas component rather than to the total amount of cold gas. Recently, Fu et al. (2012, 2013) introduced a detailed prescription for the evolution of atomic and molecular components into the Munich semi-analytic model, allowing star formation to be connected directly to molecular content. This is clearly more realistic than Eq. (S15) and will be incorporated in future large-scale modelling efforts, but here we simply allow the star formation threshold $M_{\text{crit}}$ to be a free parameter in our MCMC sampling, recognising that our previous fixed value was poorly justified. The new preferred value is about a factor of two smaller, mainly in order to slow the quenching of satellites by allowing them to use up a larger fraction of the cold gas with which they fall into their host. Fig. S1 compares our new model’s predictions for the atomic gas mass function (left panel) and for the atomic gas over luminosity ratios (right panel) in HI mass-limited bins. The reduced threshold still results in reasonable gas properties but there is a significant deficit of cold gas around the knee of the mass function for atomic gas. In this respect, the earlier model of Guo et al. (2013) does significantly better than the current model, presumably because of its higher star formation threshold and lower star formation efficiency.

Stars can also form whenever two galaxies merge since their cold gas components are strongly disturbed, typically initiating a starburst and feeding some cold gas into the central black hole. This and all other merger-related processes are described in Section 1.12.

### 1.7 Supernova feedback

Massive stars are relatively short-lived. Consequently, soon after an episode of star formation, a large number of them explode as supernovae, strongly clustered both in space and time. The collective energy released by these supernovae and by the stellar winds which precede them is injected into surrounding gas, both cold and hot. As a result, some of the cold interstellar medium is reheated to join the hot gas atmosphere, and this atmosphere itself is also heated, compensating for its cooling and causing some of it to flow out of the galaxy in a wind. This feedback process is a critical aspect of galaxy formation and has long been identified as the main agent controlling its overall efficiency (Larson 1974; White & Rees 1978; Dekel & Silk 1986). As a result, detailed modelling is required if a simulation is to produce a realistic galaxy population. Our specific feedback model is controlled by two main efficiencies, each with three adjustable parameters. One efficiency sets the fraction of the “SN” energy which is available to drive long-term changes in the thermodynamic state of the galaxy’s gas components (rather than being lost to cooling radiation), while the other controls the fraction of this energy which is used to reheat cold gas and inject it into the hot gas atmosphere, the remainder being used to heat this atmosphere directly. Heating of the hot atmosphere results in ejection of “wind” material to an external reservoir from which it may or may not be reincorporated at a later time, depending on the mass of the host system.

The energy effectively available to the gas components from supernovae and stellar winds is taken to be:

$$\Delta E_{\text{SN}} = \epsilon_{\text{halo}} \times \frac{1}{2} \Delta M V_{\text{SN}}^2. \quad (S16)$$

where $\frac{1}{2}V_{\text{SN}}^2$ is the mean energy injected per unit mass of stars formed (we take $V_{\text{SN}} = 630 \text{ km s}^{-1}$) and the efficiency is

$$\epsilon_{\text{halo}} = \eta \times \left[ 0.5 + \left( \frac{V_{\text{max}}}{V_{\text{jet}}} \right)^{-\beta_1} \right]. \quad (S17)$$

The mass of cold gas reheated by star formation and added to the hot atmosphere is assumed to be directly proportional to the amount of stars formed:

$$\Delta M_{\text{reheat}} = \epsilon_{\text{disk}} \Delta M_*, \quad (S18)$$

where the second efficiency is

$$\epsilon_{\text{disk}} = \epsilon \times \left[ 0.5 + \left( \frac{V_{\text{max}}}{V_{\text{reheat}}} \right)^{-\beta_1} \right]. \quad (S19)$$

This reheating is assumed to require energy $\Delta E_{\text{reheat}} = \frac{1}{2} \Delta M_{\text{reheat}} V_{200}^2$. If $\Delta E_{\text{reheat}} > \Delta E_{\text{SN}}$, the reheated mass is assumed to saturate at $\Delta M_{\text{reheat}} = \Delta E_{\text{SN}} / \left( \frac{1}{2} V_{200}^2 \right)$. Otherwise, the remaining SN energy is used to eject a mass $\Delta M_{\text{eject}}$ of hot gas into an external reservoir, where

$$\frac{1}{2} \Delta M_{\text{eject}} V_{200}^2 = \Delta E_{\text{SN}} - \Delta E_{\text{reheat}}. \quad (S20)$$

There is now considerable observational evidence for ejection of interstellar gas due to star formation activity (Shapley et al. 2003; Rupke et al. 2005; Weiner et al. 2009; Martin et al. 2012; Rubin et al. 2013). While the overall impact of such processes is still debated, observations of rapidly star-forming systems tend to favour mass-loading factors (the ratio of reheated mass to the mass of stars formed) between 1 and 10. The mass-loading factors preferred by our MCMC chains are shown as a function of virial velocity in the top left panel of Fig. S2 and seem similar or somewhat larger than observed.

Ejection of disk gas into the hot atmosphere has relatively little impact when the latter has a short cooling time, since this effectively drives a galactic fountain in which the material soon returns and becomes available for star formation again. Ejection of gas from the hot phase to an external reservoir has substantially stronger long-term effects, however, since such wind ejecta are unavailable for star formation for much longer periods. The top right panel of Fig. S2 shows $\epsilon_{\text{halo}}$, the fraction of the available energy that is used in feedback processes, as a function of virial velocity, while the bottom left panel shows $\Delta M_{\text{eject}} / \Delta M_*$, the ratio of the mass of gas ejected in a wind from the galaxy/halo system to the mass of stars formed. For the parameters preferred by our MCMC chains, the available energy is used with high efficiency in low-mass systems, and winds are able to eject material from the halos of all galaxies with virial velocity less than about 200 km s$^{-1}$.
Illustration of the dependences of SN feedback on halo properties. The top left panel shows the disk reheating efficiency $\epsilon_{\text{disk}}$ as a function of maximum circular velocity $V_{\text{max}}$. Often referred to as the mass-loading factor, this is the ratio of the star formation rate to the rate at which ISM material is heated and injected into the hot halo. The top right panel shows the halo ejection efficiency $\epsilon_{\text{halo}}$ as a function of $V_{\text{max}}$. This is the fraction of the available SN energy which is used in reheating disk gas and in ejecting hot gas from the halo. The bottom left panel gives $\Delta M_{\text{eject}}/\Delta M_{\star}$, the ratio between the hot gas mass ejected to an external reservoir and the cold gas mass which is turned into stars. The bottom right panel shows the reincorporation timescale $t_{\text{reinc}}$ as a function of halo virial velocity $V_{\text{200c}}$ and of redshift (note that the redshift evolution comes solely from the evolution in the relation between $V_{\text{200c}}$ and $M_{\text{200c}}$). In each panel dotted lines refer to the G11-WMAP7 model, dashed lines to the Henriques et al. (2013) model and solid lines to our new model with its best-fit parameter values. The blue shaded regions give the 2σ range allowed by our MCMC sampling. Colours in the bottom right panel indicate redshift as shown by the label.

1.8 Reincorporation of gas ejected in winds

A number of recent papers have argued that most published semi-analytic models and cosmological hydrodynamics simulations form low-mass galaxies too early, leading to an over-abundance of lower mass galaxies ($M_{\star} \sim 10^{10} M_\odot$) at $z \geq 1$ (Fontanot et al. 2009; Henriques et al. 2011; Guo et al. 2011; Weinmann et al. 2012; Lu et al. 2013; Genel et al. 2014; Vogelsberger et al. 2014). In the context of the Munich galaxy formation model, the MCMC analysis of Henriques et al. (2013) concluded that this can only be corrected by coupling strong winds in low-mass galaxies with long reincorporation times for the ejecta. This results in slower growth at early times followed by a stronger build-up between $z = 2$ and $z = 0$ as the ejecta finally fall in again.

In the current work we adopt the implementation of Henriques et al. (2013). The mass of gas returned to the hot gas halo from the ejecta reservoir is taken to be:

$$M_{\text{eject}} = \frac{M_{\text{eject}}}{t_{\text{reinc}}}$$  \hspace{1cm} (S21)

where the reincorporation time scales inversely with the mass of the host halo,

$$t_{\text{reinc}} = \gamma \frac{10^{10} M_\odot}{M_{\text{200c}}}$$  \hspace{1cm} (S22)

rather than with the ratio of its dynamical time and circular
velocity, as in Guo et al. (2011). Note that a key aspect of this phenomenological model is that diffuse gas is not available for cooling onto the central galaxy as long as it remains in the external reservoir. The precise location of this reservoir is unspecified, and the gas may not leave the halo entirely. Rather, its entropy may simply be raised above the level assumed by our simple “isothermal” model, in which case the reincorporation time-scales should be interpreted as the time needed to cool to the point where the gas can again be considered part of our standard cooling flow.

The differences between the new reincorporation times and those adopted in Guo et al. (2011) are shown as a function of virial velocity and redshift in the lower right panel of Fig. S2. Note that the redshift dependence in our new model simply results from the relation between $M_{200c}$ and $V_{200c}$ (Eq. S1). In practice gas ejected in winds from low-mass halos will never be reincorporated unless they become part of a more massive system, while gas returns immediately in the most massive halos. This implementation agrees qualitatively with the behaviour seen by Oppenheimer & Davé (2008) and Oppenheimer et al. (2010) in their numerical simulations.

1.9 Metal enrichment

When stars die, they release newly formed heavy elements into the surrounding medium in addition to mass and energy. In the current work, we follow the total mass of metals only, assuming that each solar mass of stars produces a mass $y$ of heavy metals, with this “yield” treated as a free parameter in the MCMC. The newly formed metals are mixed instantaneously into the cold gas, and thereafter follow it through the various baryonic components of the galaxy, thus enriching the hot gas atmosphere and future generations of stars. In recent work, Yates et al. (2013) and De Lucia et al. (2014) have introduced two different implementations of chemical enrichment into the Munich model. These follow in detail the return of individual elements as stellar populations age, and include metallicity-dependences both in the yields and in population evolution modelling. We expect to incorporate such effects in future large-scale population models, but they are ignored in the model presented here.

The metallicities predicted by the current model are illustrated in Fig. S3, where the theoretical stellar mass-stellar metallicity relation at $z = 0$ is compared with observations. The current model and those of Henriques et al. (2013) and Guo et al. (2013) all show similar stellar metallicities, despite significant changes in the treatment of wind ejecta. The $z = 3$ predictions of the current model (and also of the earlier ones) show chemical enrichment to happen very early. The same behaviour is found for the metallicity of the cold gas. This appears to disagree with observation (Maiolino et al. 2008) so further work is clearly needed on this point.

1.10 Black hole related processes

In our model, the energy released by supernovae and stellar winds has a dramatic effect on low-mass galaxies, but is unable to reduce cooling onto massive systems to the very low rates inferred from their observed stellar masses and star formation rates. We follow Croton et al. (2006) in assuming that feedback from central supermassive black holes is the agent that terminates galaxy growth in massive halos. Black holes are taken to form and to grow when cold gas is driven to the centre of merging systems. In addition, pre-existing black holes merge as soon as their host galaxies do. This “quasar mode” growth is the main channel by which black holes gain mass in our model, but we do not associate it with any feedback beyond that from the strong starbursts which accompany gas-rich mergers. Black holes are also allowed to accrete gas from the hot gas atmospheres of their galaxies, however, and this is assumed to generate jets and bubbles which produce “radio mode” feedback, suppressing cooling onto the galaxy and so eliminating the supply of cold gas and quenching star formation. The relative importance of these two modes to black hole growth is shown as a function of time and galaxy mass in Fig. 3 of Croton et al. (2006).

1.10.1 Quasar mode - black hole growth

Whenever two galaxies merge, their cold gas components are strongly disturbed and a significant fraction is driven into the inner regions where it may form a black hole or be accreted onto a pre-existing black hole. When both galaxies contain a pre-existing black hole, these are expected to merge during this highly dynamic phase of evolution.

The amount of gas accreted in the quasar mode is taken to depend on the properties of the two merging galaxies as,

$$
\Delta M_{\text{BH},Q} = f_{\text{BH}}(M_{\text{sat}}/M_{\text{cen}})M_{\text{cold}}/ \left(1 + (V_{\text{BH}}/V_{200c})^2\right),
$$

where $M_{\text{cen}}$ and $M_{\text{sat}}$ are the total baryon masses of the central galaxy and the satellite which merges with it, $M_{\text{cold}}$ is their total cold gas mass, $V_{200c}$ is the virial velocity of the central halo and $f_{\text{BH}}$ and $V_{\text{BH}}$ are two adjustable parameters which control the fraction of the available cold gas.
Figure S4. The scalings of the processes controlling black hole growth and AGN feedback. The left panel shows the maximum fraction of cold gas accreted (for a major merger of equal mass galaxies) onto central black holes during mergers (quasar accretion) as a function of virial velocity (Eq. S23). The right panel shows the ratio of hot gas accretion rate to the product of hot gas and black hole masses (i.e. the coefficient in Eq. S24) as a function of redshift. The additional scaling with $V_{200c}/M_{200c}$ results in the redshift variation seen in models prior to this work. Accretion in this mode is assumed to suppress cooling in massive systems. In both panels the best-fit and allowed $\pm 2\sigma$ regions for the current model are shown as solid blue lines and light blue regions. The scalings adopted in Henriques et al. (2013) are shown as a dashed blue lines and the Guo et al. (2013) scalings as dotted blue lines.

1.10.2 Radio mode - feedback

We assume that central supermassive black holes continually accrete gas from the hot gas atmosphere of their host galaxies, and that this produces “radio mode feedback” which injects energy into the hot atmosphere. Recent changes to our model have increased the amount of hot gas available to cool onto massive systems at late times, and as a result we find that the original Croton et al. (2006) model for radio-mode feedback is unable to suppress star formation sufficiently just above the knee of the galaxy stellar mass function. An MCMC analysis shows that this cannot be solved simply by changing parameters in the original formulation, but that acceptable results can be obtained by assuming the accretion rate to be given by

$$\dot{M}_{\text{BH},\text{eff}} = \dot{M}_{\text{BH}}^\text{AGN} \left( \frac{M_{\text{hot}}}{10^{12} M_\odot} \right) \left( \frac{M_{\text{BH}}}{10^8 M_\odot} \right). \quad (S24)$$

This formula is equivalent to that of Croton et al. (2006) divided by a factor of $H(z)$, so accretion is enhanced at lower redshifts. The differences in the treatment of AGN growth and feedback between the current model and those of Henriques et al. (2013) and Guo et al. (2013) are shown in Fig. S4. Note that in our new model, as in its predecessors, the mass growth of black holes through the radio mode is negligible in comparison with quasar mode accretion.

This form of growth is, however, important in that it is assumed to produce relativistic jets which deposit energy into the hot gas halo in analogy with the hot bubbles seen in galaxy clusters (McNamara & Nulsen 2007; Birzan et al. 2004). The energy input rate is taken to be

$$\dot{E}_{\text{radio}} = \eta \dot{M}_{\text{BH}} c^2, \quad (S25)$$

where $\eta = 0.1$ is an efficiency parameter and $c$ is the speed of light. This energy then suppresses cooling from the hot gas to the cold disk, resulting in an effective cooling rate given by

$$\dot{M}_{\text{cool,eff}} = \max \left[ \dot{M}_{\text{cool}} - 2 \frac{\dot{E}_{\text{radio}}}{V_{200c}^2}, 0 \right]. \quad (S26)$$

We assume that elimination of the cooling flow also cuts off the supply of gas to the black hole, so that heating of the hot atmosphere beyond this point is not possible.

Despite growing observational and theoretical evidence for the interaction of black holes with their gaseous environment, we still lack an established theory for this process. The equations given here, like those of Croton et al. (2006), should be regarded a purely phenomenological representation of some process which acts to prevent the cooling of gas onto massive central galaxies without requiring additional star formation. The comparison with observation presented in the main paper and in Paper II suggest that our current assumptions result in quenching of star formation in intermediate and high-mass galaxies approximately as required by the data.

1.11 Environmental processes

The growth of structure in a ΛCDM universe affects galaxies as they and their halos fall into larger systems and are influenced by tides, by hydrodynamical forces from the hot gas through which they move, and by encounters with other galaxies. Such environmental effects remove material and
modify the structure and evolution of the galaxies, in some cases leading to their complete disruption. Several such processes are incorporated in our modelling and their treatment here follows that of Guo et al. (2011) closely. However, environmental effects appear overestimated in the earlier model, which predicts a significantly higher fraction of quenched satellite galaxies than is observed, particularly in intermediate mass halos (e.g. Wang et al. 2012, 2014). We address this problem here by suppressing ram-pressure stripping in such systems.

### 1.11.1 Tidal and ram-pressure stripping

As soon a halo falls into a larger system its mass growth reverses as tidal forces begin to remove dark matter (e.g. De Lucia et al. 2004). In the Guo et al. (2011) model, this implies that no new baryonic material is added to the system and its hot gas atmosphere is stripped away in proportion to its dark matter mass,

$$ M_{\text{hot,infall}}(R_{\text{tidal}}) / M_{\text{hot,infall}} = M_{\text{DM,infall}} / M_{\text{DM,infall}}, $$  \hspace{1cm} (S27)

where the limiting radius is given by a simple “isothermal” model,

$$ R_{\text{tidal}} = \left( \frac{M_{\text{DM,infall}}}{M_{\text{DM,infall}}} \right) R_{\text{DM,infall}}. $$  \hspace{1cm} (S28)

In these equations, $M_{\text{DM,infall}}, R_{\text{DM,infall}}$, and $M_{\text{hot,infall}}$ are $M_{200c}, R_{200c}$, and the hot gas mass of the halo just prior to infall, and $M_{\text{DM}}$ and $M_{\text{hot}}$ are the current masses of these two components. By construction, tidal stripping will have removed all hot gas once the subhalo is disrupted and the galaxy becomes an orphan.

Hot gas can also be stripped by ram-pressure effects which are followed starting when the satellite first falls within the virial radius of its host. At a certain distance $R_{e,p}$ from the centre of the satellite, self-gravity is approximately balanced by ram pressure,

$$ \rho_{\text{sat}}(R_{e,p}) V_{\text{sat}}^2 = \rho_{\text{par}}(R) V_{\text{orbit}}^2, $$  \hspace{1cm} (S29)

where $\rho_{\text{sat}}(R_{e,p})$ is the hot gas density of the satellite at radius $R_{e,p}$, $V_{\text{sat}}$ is the virial velocity of the subhalo at infall (which we assume to be constant as the subhalo orbits around the main halo), $\rho_{\text{par}}(R)$ is the hot gas density of the parent dark matter halo at the distance $R$ of the satellite from the centre of its potential well, and $V_{\text{orbit}}$ is the orbital velocity of the satellite, which we approximate as the virial circular velocity of the main halo. The densities here are again estimated from the total mass and limiting radius of the relevant component according to an “isothermal” model, $\rho \propto r^{-2}$. Finally, the radius of the hot gas component is taken to be the smaller of $R_{e,p}$ and $R_{\text{tidal}}$.

In the current work we apply this ram-pressure model only in halos above a threshold mass ($M_{\text{min}}$) which we introduce as a free parameter which observational constraints then require to be $\lesssim 10^{14} M_\odot$. Combined with our lower threshold for star formation, this changes reduces the excess of passive satellites found in the Guo et al. (2011) and Henriques et al. (2013) models, while remaining consistent with observation of ram-pressure stripping phenomena in rich clusters.

Finally we note that ram-pressure effects on the cold gas component are not included in our model. Such effects are expected (e.g. Bekki 2014) and are indeed observed in high density regions (e.g. Crowl et al. 2005; Fumagalli et al. 2014) but they require more extreme conditions than the effects considered in this section.

#### 1.11.2 Tidal disruption of galaxies

Our implementation of the tidal disruption of the stellar and cold gas components of galaxies is unchanged from Guo et al. (2011). Since both components are considerably more concentrated than the dark matter, we consider disruption only for galaxies that have already lost their dark matter and hot gas components. For such orphans, the baryonic (cold gas + stellar mass) density within the half-mass radius is compared to the dark matter density of the main halo within the pericentre of the satellite’s orbit. If the latter is larger, i.e.

$$ \frac{M_{\text{DM, halo}}(R_{\text{peri}})}{R_{\text{peri}}^3} \geq \rho_{\text{DM, halo}} > \rho_{\text{sat}} \equiv \frac{M_{\text{hot}}}{R_{\text{hot, half}}^3}, $$  \hspace{1cm} (S30)

the satellite is completely disrupted, its stars are added to the intracluster light (ICL) and its cold gas is added to the hot gas atmosphere of the central galaxy. The galaxy’s half-mass radius is calculated from those of the cold gas and stellar disks and the bulge (assuming exponential surface density profiles for the first two and a surface density scaling with $r^{1/4}$ for the latter), while its orbital pericentre is calculated as

$$ \left( \frac{R}{R_{\text{peri}}} \right)^2 = \frac{\ln R/R_{\text{peri}} + \frac{1}{2} (V/V_{200c})^2}{\frac{1}{2} (V_{\text{orb}}/V_{200c})^2}, $$  \hspace{1cm} (S31)

assuming conservation of energy and angular momentum and a singular isothermal potential for the orbit, $\phi(R) = V_{200c}^2 \ln R$. In these equations, $R$ is the current distance of the satellite from halo centre, and $V$ and $V_{\text{orb}}$ are the total
and tangential velocities of the satellite with respect to halo centre (see Section 1.12.1 for a description on how these are determined for orphans). We tested that this condition for complete disruption of satellites gives very similar answers to the more detailed implementation of gradual stripping proposed by Henriques & Thomas (2010) (See Contini et al. (2014) for a more extensive comparison of different implementations of tidal disruption).

1.11.3 SN feedback in orphan galaxies

For orphan galaxies environmental effects are particularly dramatic. Since our implementation of tidal stripping of hot gas is directly connected to the stripping of dark matter, once galaxies lose their halo, they also have no hot gas left. From this point on, we also assume that any cold gas re-heated by star formation activity leaves the galaxy and is added to the hot gas atmosphere of the main halo. This can lead to rapid depletion of any remaining cold gas.

1.12 Mergers and bulge formation

1.12.1 Positions and velocities of orphans

Once a satellite subhalo is disrupted, its central galaxy becomes an orphan and its position and velocity are linked to those of the dark matter particle which was most strongly bound within the subhalo just prior to its disruption. As soon as a disruption event occurs, this particle is identified and a merging clock is started, based on an estimate of how long the satellite will take to spiral into the central object due to dynamical friction. This time is computed using the Binney & Tremaine (1987) formula:

$$t_{\text{friction}} = \alpha_{\text{friction}} \frac{V_{\text{200c}} \, r_{\text{sat}}}{GM_{\text{sat}} \ln \Lambda}, \quad (S32)$$

where $M_{\text{sat}}$ is the total mass of the satellite (dark and baryonic), $\ln \Lambda = \ln(1 + M_{\text{200c}}/M_{\text{sat}})$ is the Coulomb logarithm and $\alpha_{\text{friction}} = 2.4$ is a parameter originally set by De Lucia & Blaizot (2007) to match the bright end of the $z = 0$ luminosity functions. This value was later shown to be consistent with inferences from direct numerical simulation (Boylan-Kolchin et al. 2008; De Lucia et al. 2010) but should still be considered poorly known. The Millennium-II simulation is able to resolve subhalos which have been stripped to masses below that of their central galaxy. In such cases we turn on the merging clock as soon as the subhalo mass drops below the stellar mass in the galaxy.

Following Guo et al. (2011) we model the decay of the satellite’s orbit due to dynamical friction by placing the orphan galaxy not at the current position of the particle with which it is identified, but at a position whose (vector) offset from the central galaxy is reduced from that of the particle by a factor of $(1 - \Delta t/t_{\text{friction}})$ where $\Delta t$ is the time since the dynamical friction clock was started. The (vector) velocity of the orphan galaxy is set equal to that of the tagged particle. This time dependence is based on a simple model for satellite with “isothermal” density structure spiralling to the centre of an isothermal host on a circular orbit. When $\Delta t = t_{\text{friction}}$ the orphan merges with the central galaxy.

$$a_{\text{SF,burst}}(m_{\text{sat}}/m_{\text{central}})^{S32}$$

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure6}
\caption{The fraction of cold gas converted to stars in merger-triggered starbursts as a function of the mass ratio between satellite and central galaxies. The best-fit and $\pm 2\sigma$ regions for the current model are shown as a solid blue line and a light blue region respectively. The scaling adopted in Henriques et al. (2013) and Guo et al. (2013) is shown as a dashed blue line.}
\end{figure}

1.12.2 Merger-triggered star formation

When a satellite finally merges with the object at the centre of the main halo, the outcome is different for major and minor mergers. We define a major merger to be one in which the total baryonic mass of the less massive galaxy exceeds a fraction $R_{\text{merge}}$ of that of the more massive galaxy. We treat $R_{\text{merge}}$ as a free parameter in our MCMC analysis, finding it to be strongly constrained by our calibrating observations with a best-fit value $\sim 0.4$, close to previous choices. In a major merger, the disks of the two progenitors are destroyed and all their stars become part of the bulge of the descendant, along with any stars formed during the merger. In a minor merger, the disk of the larger progenitor survives and accretes the cold gas component of the smaller galaxy, while its bulge accretes all the stars of the victim. Stars formed during the merger stay in the disk of the descendant. In both types of merger, cold gas is fed to the central black hole according to the formulæ of Section 1.10.1.

The stellar mass formed during a merger is modelled using the “collisional starburst” formulation of Somerville et al. (2001):

$$M_{\text{bul}} = \alpha_{\text{SF,burst}} \left( \frac{M_1}{M_2} \right)^{\beta_{\text{SF,burst}}} M_{\text{cold}}, \quad (S33)$$

where $M_1 < M_2$ are the baryonic masses of the two galaxies, and $M_{\text{cold}}$ is their total cold gas mass. The $\alpha_{\text{SF,burst}}$ and $\beta_{\text{SF,burst}}$ parameters were originally fixed to reflect the results of the Mihos & Hernquist (1996) simulations, but in the current work they are left free and are allowed to vary in our MCMC analysis. Despite this, in our best-fit model the fraction of cold gas converted to stars in merger-related bursts is relatively close to what was previously assumed. Fig. S6 compares this quantity for our current model (solid blue line and light blue regions) to that assumed in the mod-
els of Henriques et al. (2013) and Guo et al. (2013) (dashed blue line).

1.12.3 Bulge Formation

In our model bulges can form through major and minor mergers and through the buckling instability of disks. After a major merger, all stars are considered part of the new bulge, but the remnant of a minor merger retains the stellar disk of the larger progenitor and its bulge gains only the stars from the smaller progenitor. Following Guo et al. (2011), we use energy conservation and the virial theorem to compute the change in sizes in both minor and major mergers:

\[
\frac{GM_{\text{new bulge}}^2}{R_{\text{new bulge}}} = \frac{GM_1^2}{R_1} + \frac{GM_2^2}{R_2} + 2\alpha_{\text{inter}} GM_1 M_2 \left( \frac{1}{R_1} + \frac{1}{R_2} \right),
\]

(S34)

where the left-hand side represents the binding energy of the final bulge, the first two terms of the right-hand side represent the binding energies of the progenitor stellar systems (the radii in these three terms are taken to be the half-mass radii of the corresponding stellar systems) and the last term is the binding energy of the relative orbit of the two progenitors at the time of merger. The coefficient \(\alpha_{\text{inter}}\) quantifies the binding energy invested in this orbit relative to that in the individual systems. Guo et al. (2011) set \(\alpha_{\text{inter}} = 0.5\) and show that this leads to bulge sizes in reasonable agreement with SDSS data. When either of the progenitors is a composite disk+bulge system, its half-mass radius is calculated assuming an exponential disk and an \(r^{1/4}\)-law bulge.

Another important channel of bulge growth is secular evolution through disk instabilities. These dynamical instabilities occur through the formation of bars which then buckle. They transport material inwards to the bulge and they occur in galaxies where self-gravity of the disk dominates the gravitational effects of the bulge and halo. As a criterion for disk instability, we follow Guo et al. (2011) in adopting

\[
V_{\text{max}} < \sqrt{\frac{GM_{\star,d}}{3R_{\star,d}}},
\]

(S35)

where \(M_{\star,d}\) and \(R_{\star,d}\) are the stellar mass and exponential scale-length of the stellar disk and \(V_{\text{max}}\) is the maximum circular velocity of the host dark matter halo hosting the disk. When the instability criterion of Eq. S35 is met, we transfer sufficient stellar mass from the disk to the bulge to make the disk marginally stable again. Following Guo et al. (2011) we assume that this mass \(\delta M_{\star}\) is transferred from the innermost part of the disk. Thus, the half-mass radius of the material to be added to the bulge, \(R_{\star}\), is related to \(\delta M_{\star}\) through

\[
\delta M_{\star} = 2\pi \Sigma_{\text{eq}} R_{\star,d} (R_{\star,d} - (R_0 + R_{\star,d}) \exp(-R_0/R_{\star,d})),
\]

(S36)

where \(\Sigma_{\text{eq}}\) and \(R_{\star,d}\) are the central surface density and exponential scale length of the unstable disk. We neglect the angular momentum of the transferred material, so the disk’s angular momentum is unchanged, resulting in an increase in \(R_{\star,d}\) to compensate for the mass lost.

If the galaxy already has a spheroidal component, the newly created bulge material is assumed to merge with the existing bulge according to Eq. (S34) where we now take \(\alpha_{\text{inter}} = 2\) to account for the fact that the inner disk and the initial bulge are concentric and have no relative motion.

For further discussion of this model for bulge growth and for comparisons with observational data we refer the reader to Guo et al. (2011).

1.13 Stellar populations synthesis

Stellar population synthesis models are a crucial part of galaxy formation synthesis as they link the masses, ages and metallicities predicted for stars to the observable emission at various wavelengths. We use Maraston (2005) as our default stellar population synthesis model, but we have checked that the publicly released but still unpublished Charlot & Bruzual (2007) code leads to very similar results for all the properties we consider. Somewhat different predictions are obtained with the earlier Bruzual & Charlot (2003) code because of the weaker emission it assumes for the TP-AGB stage of evolution of intermediate age stars. Recent work by a number of authors suggests that the more recent models work in better agreement with observed near-infrared emission from bright galaxies at \(z \gtrsim 2\) (Henriques et al. 2011, 2012; Tonini et al. 2009, 2010; Fontanot & Monaco 2010; Tonini et al. 2011; Gonzalez-Perez et al. 2014). For the Munich galaxy formation model, in particular, Henriques et al. (2011) and Henriques et al. (2012) showed that Maraston (2005) or Charlot & Bruzual (2007) populations give stellar mass and K-band luminosity functions for which the massive/bright end agrees with observation from \(z \approx 3\) to \(z = 0\). Nevertheless, as part of our model release we will, for comparison purposes, also include luminosities computed using Bruzual & Charlot (2003) stellar populations.

1.14 Dust Model

Actively star-forming galaxies are known to be rich in dust. This can have a dramatic effect on their emitted spectrum since dust significantly absorbs optical/UV light while having a much milder effect at longer wavelengths. As a result, dust-dominated galaxies will generally have red colours even if they are strongly star-forming. We follow De Lucia & Blaizot (2007) in considering dust extinction separately for the diffuse interstellar medium (ISM) (following Devriendt et al. 1999) and for the molecular clouds in which stars form (following Charlot & Fall 2000)). The optical depth of dust as a function of wavelength is computed separately for each component and then a slab geometry is assumed in order to compute the total extinction of the relevant populations. We do not at present attempt to compute the detailed properties of the dust particles or the re-emission of the absorbed light.

1.14.1 Extinction by the ISM

To estimate the extinction due to the general ISM, the optical depth of diffuse dust in galactic disks is assumed to vary
with wavelength as:
\[ \tau_{\lambda}^{\text{ISM}} = \left( \frac{A_{\lambda}}{A_{\text{H}}} \right) (1 + z)^{-1} \left( \frac{Z_{\text{gas}}}{Z_{\odot}} \right)^s \times \left( \frac{\langle N_H \rangle}{2.1 \times 10^{21} \text{ atoms cm}^{-2}} \right), \]  
(S37)
where \( \langle N_H \rangle \) represents the mean column density of hydrogen and is given by:
\[ \langle N_H \rangle = \frac{M_{\text{cold}}}{1.4 m_p c \sigma_{R_{\text{gas,d}}}} \text{atoms cm}^{-2}, \]  
(S38)
where \( R_{\text{gas,d}} \) is the cold gas disk scale-length and \( a = 1.68 \) in order for \( \langle N_H \rangle \) to represent the mass-weighted average column density of an exponential disk. Following the results in Guiderdoni & Rocca-Volmerange (1987), the extinction curve in Eq. (S37) depends on the gas metallicity and is based on an interpolation between the Solar Neighbourhood and the Large and Small Magellanic Clouds, \( s = 1.35 \) for \( \lambda < 2000 \text{ Å} \) and \( s = 1.6 \) for \( \lambda > 2000 \text{ Å} \). The extinction curve for solar metallicity, \( (A_{\lambda}/A_{\text{H}})_{\odot} \), is taken from Mathis et al. (1983).

The redshift dependence in Eq. (S37) is significantly stronger than in previous versions of our model \((1+z)^{-0.5}\) in Kitzbichler & White (2007) and \((1+z)^{-0.4}\) in Guo & White (2009). The dependence implies that for the same amount of cold gas and the same metal abundance, there is less dust at high redshift. The motivation comes both from observations (Steidel et al. 2004; Quadri et al. 2008) and from the assumption that dust is produced by relatively long-lived stars. However, it may also be that this redshift dependence has to be introduced as a phenomenological compensation for the excessively early build-up of the metal content in model galaxies shown in Fig. S3. In practice we include it simply to ensure that high-z galaxies are almost dust free, as inferred from their observed UV slopes (Bouwens et al. 2012). As will be shown in Clay et al. (2014) this produces luminosity functions of Lyman-break galaxies at \( z > 5 \) compatible with HST data.

### 1.14.2 Extinction by molecular clouds

The second source of extinction affects only young stars and comes from the molecular clouds where they are formed. Following Charlot & Fall (2000), our model assumes that such extinction affects stars younger than the lifetime of stellar birth clouds (taken to be \( 10^7 \) years). The relevant optical depth is taken to be
\[ \tau_{\lambda}^{\text{BC}} = \tau_{\lambda}^{\text{ISM}} \left( \frac{1}{\mu - 1} \right) \left( \frac{\lambda}{5500 \text{ Å}} \right)^{-0.7} \]  
(S39)
where \( \mu \) is given by a random Gaussian deviate with mean 0.3 and standard deviation 0.2, truncated at 0.1 and 1.

### 1.14.3 Overall extinction curve

In order to get the final overall extinction, every galaxy is assigned an inclination given by the angle between the disk angular momentum and the z-direction of the simulation box, and a “slab” geometry is assumed for the disk in the dust. Therefore, for each component, the extinction in magnitudes is written as
\[ A_{\lambda} = -2.5 \log \left( \frac{1 - \exp^{-\tau_{\lambda} \text{ sec } \theta}}{\tau_{\lambda} \text{ sec } \theta} \right) \]  
(S40)
where \( \theta \) is the angle of inclination of the galaxy relative to the line-of-sight and \( \tau_{\lambda} \) corresponds to either \( \tau_{\lambda}^{\text{ISM}} \) or \( \tau_{\lambda}^{\text{BC}} \). Young stars are affected by both extinction components while older stars are affected only by the diffuse ISM component.
1.15 Monte Carlo Markov Chains

In order to sample the full multidimensional parameter space of our model we use MCMC techniques. This enables exploration of the allowed regions when the model is constrained by a broad variety of calibrating observations, which may be of different types and correspond to different redshifts. The same scheme allows us to assess the merits of different implementations of critical astrophysical processes. We use a version of the Metropolis–Hastings method (Metropolis et al. 1953; Hastings 1970); a full description of the algorithm can be found in Section 3 of Henriques et al. (2009). A full MCMC chain requires evaluation of many tens of thousands of models and it is not computationally feasible to build all of these models for the full Millennium or Millennium-II simulation. We therefore use sampling techniques to construct a representative subset of subhalo merger trees on which the galaxy formation model is evaluated during the MCMC procedure (details are given in Appendix 2 of Henriques et al. 2013). Once the best-fit model has been identified, it can be implemented on the full volumes of the two simulations.

Fig. S7 shows marginalised 1D posterior distributions for our model parameters when the model is constrained by observational data on the abundance and passive fraction of galaxies as a function of stellar mass from $z = 3$ down to $z = 0$. Vertical solid blue lines correspond to the parameter values of the best-fit model (taken to be the one for which the MCMC chain found the highest likelihood) and these are also presented in Table S1. Shaded blue regions show the central 95% confidence region of each marginalised posterior distribution and the boundaries of the corresponding parameter interval are also given in Table S1. Interestingly, although the best-fit model has parameters which lie within these regions in almost all cases, this is not true for $k_{\text{AGN}}$ and $v_{\text{cheat}}$. The allowed parameter range is quite narrow in all cases, showing that these observations are sufficient to specify our model completely without major degeneracies.

Fig. S7 also shows the parameter values corresponding to the best-fit models of Henriques et al. (2013) (dashed vertical blue lines) and Guo et al. (2013) (dotted vertical blue lines). Despite changes in cosmology and in several aspects of the astrophysical modelling, the efficiencies of most processes are very similar in the different versions of the model (see also Figs S2, S4 and S6). This indicates that parameters which were not previously included in the MCMC sampling (all of them in the case of Guo et al. (2013)! were, in fact, well constrained by less rigorous comparison to observations. The exceptions are the cold gas density threshold for star formation, which now has a significantly lower value, and the ram-pressure stripping threshold, which was zero in the earlier models. Both changes are required to predict the correct evolution of the fraction of passive galaxies as a function of stellar mass, which simply could not be explained by the previous models. Indeed, if we were to carry out an MCMC analysis of the Henriques et al. (2013) or the Guo et al. (2013) model using this observational constraint, we would find a very low maximum likelihood value. As a result, we do not need to integrate over the posterior distributions to perform Bayesian model selection – the current model is the only one of the three which can come close to representing the observational distributions used as constraints.

REFERENCES

Charlot S., Bruzual G., 2007, provided to the community but not published
Henriques et al.

Figure S7. Shaded blue regions show the 1D, normalised posterior distributions of our model parameters when the model is constrained by observations of the abundance and passive fraction of galaxies as a function of stellar mass from $z = 3$ to $z = 0$. Straight lines represent values corresponding to our overall best-fit model (solid blue lines) and to those of Henriques et al. (2013) (dashed blue lines) and Guo et al. (2013) (dotted red lines). The $x$-axis is plotted logarithmically in all cases.

Guiderdoni B., Rocca-Volmerange B., 1987, Astronomy and Astrophysics Supplement Series, 186, 1
Hastings W. K., 1970, Biometrika, 57, 97
Vogelsberger M., Genel S., Springel V., et al., 2014, Nat., 509, 177