MNRAS **000**, 000–000 (2020) Preprint 10 November 2020 Compiled using MNRAS LATEX style file v3.0

Galaxy formation with L-GALAXIES: Model description

Mohammadreza Ayromlou,[1]★ Guinevere Kauffmann,[1] Robert M. Yates,[1]
Dylan Nelson,[1,2] Simon D. M. White[1]

[1]*Max Planck Institute for Astrophysics, Karl-Schwarzschild-Str. 1, 85741 Garching bei München, Germany*
[2]*Universität Heidelberg, Zentrum für Astronomie, Institut für theoretische Astrophysik, Albert-Ueberle-Str. 2, 69120 Heidelberg, Germany*

ABSTRACT

In this supplementary material, we give a full description of the treatment of astrophysical processes in our model of galaxy formation, a variant of the Henriques et al. 2020 public release. The most significant modification of our model with respect to Henriques et al. 2020 is the implementation of a novel gas stripping method. Like its predecessors, our 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 model the formation and evolution of baryonic matter. In L-Galaxies, baryonic matter bound to each galaxy/subhalo is divided into seven main components: hot gas, cold gas (partitioned into HI and H_2), stellar disc, bulge stars, halo stars, the supermassive black hole, and ejected material. There is also diffuse primordial gas associated with dark matter, which is not yet part of any halo. Primordial gas falls with the dark matter onto sufficiently massive haloes, 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. The properties of cold interstellar gas are followed in concentric rings where cold gas is partitioned into HI and H_2 and the latter is converted into stars, both quiescently and in merger-induced starbursts which also drive the growth of central supermassive black holes. Stellar evolution is tracked independently in each ring and 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. Galaxy mergers affect the gas components of galaxies, as well as the partition of stars between discs, bulges, and the intracluster light (the halo stars), a diffuse component built from tidally disrupted systems. The radial structure of discs and bulge sizes are estimated from simple energy and angular momentum-based arguments. Environmental processes such as tidal and ram-pressure stripping influence the gas content of galaxies and play a key role in quenching star formation of galaxies in dense environments. These processes are implemented based on local background environment measurements made directly on the particle data of the underlying dark matter only simulations.

Key words: galaxies: formation – galaxies: evolution – large-scale structure of Universe – methods: analytical – methods: numerical

S1 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; Henriques et al. 2015, 2020).[1] The model of the current paper is a variant of the Henriques et al. (2020) model (H20 hereafter), which uses the particle data of the underlying dark-matter-only (DMO)

★ E-mail: ayromlou@mpa-garching.mpg.de

[1] See http://www.mpa-garching.mpg.de/millennium

simulations to improve the L-Galaxies treatment of environmental processes. H20 itself improved on the Henriques et al. (2015) model (H15 hereafter) by adding a radially resolved treatment of formation and evolution of galaxy discs, while H15 updated earlier models through an improved representation of the build-up of the galaxy population over time and of the present star formation activity of the low-mass galaxy population ($8.0 \leqslant \log M_*/\,\mathrm{M}_\odot \leqslant 9.5$). Guo et al. (2011) 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 discs 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.

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.

S2 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):$\sigma_8 = 0.9$, $H_0 = 73\,\mathrm{km\,s^{-1}Mpc^{-1}}$, $\Omega_\Lambda = 0.75$, $\Omega_\mathrm{m} = 0.25$, $\Omega_\mathrm{b} = 0.045$ and $n = 1.0$ For this work the original cosmology has been scaled, using the Angulo & White (2010) technique, as updated by Angulo & Hilbert (2015), to represent the best-fitting 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: $\sigma_8 = 0.829$, $H_0 = 67.3\,\mathrm{km\,s^{-1}Mpc^{-1}}$, $\Omega_\Lambda = 0.685$, $\Omega_\mathrm{m} = 0.315$, $\Omega_\mathrm{b} = 0.0487$ ($f_\mathrm{b} = 0.155$) and $n = 0.96$.

Both the Millennium and Millennium-II simulations trace 2160^3 (~10 billion) particles from $z = 127$ to the present day. The Millennium was carried out in a box of original side $500\,h^{-1}\mathrm{Mpc} = 685\,\mathrm{Mpc}$. After rescaling to the *Planck* cosmology, the box size becomes 714 Mpc, implying a particle mass of $1.43 \times 10^9\,\mathrm{M}_\odot$. The Millennium-II follows a region a fifth the linear size, resulting in 125 times better mass resolution. Combined, the two simulations follow dark matter haloes which host galaxies spanning five orders of magnitude in stellar mass at $z = 0$. 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:

$$M_{200\mathrm{c}} = \frac{100}{G} H^2(z) R^3_{200\mathrm{c}} = \frac{V^3_{200\mathrm{c}}}{10\,G H(z)}, \tag{S1}$$

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 haloes 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 subhaloes 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 subhaloes 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.

S3 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 disc of the central galaxy, or is added to a quasi-static hot atmosphere which accretes more slowly through a cooling flow. The disc of cold gas fuels the formation of stars which eventually die, releasing energy, mass and heavy elements into the surrounding medium. In the present release of our model all processes in galactic discs are treated in spatially resolved rings. SN energy reheats cold disc 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 subhaloes 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 discs 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 $z = 0$ and 2 and abundances as a function of HI mass at $z = 0$).

S4 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, $f_b^{\cos} = 15.5\%$ for the *Planck* cosmology. As haloes 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 haloes 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 haloes with the dark matter (Efstathiou 1992). In order to model this, we use results from Gnedin (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:

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

For haloes with $M_{200c} > M_F$ suppression of the baryon fraction is small, but for haloes with $M_{200c} \ll M_F(z)$ the baryon fraction drops to $(M_{200c}/M_F(z))^3$. We adopt $\alpha = 2$ and take $M_F(z)$ from the numerical results of Okamoto et al. (2008). M_F varies from $\sim 6.5\times10^9\,\mathrm{M}_\odot$ at $z = 0$, to $\sim 10^7\,\mathrm{M}_\odot$ just before reionization starts at $z = 8$.

The total halo mass can decrease with time because of changes in morphology or halo concentration (see De Lucia et al. 2004a for discussions about subhalo mass fluctuations). At the same time, in L-Galaxies the halo baryonic mass within R_{200} remains unchanged by construction, which could cause an increase in the baryon fraction. Following the prescription of Yates et al. (2017), we correct the input halo merger trees to prevent M_{200} from decreasing with time. This accounts for any artificial decrease in M_{200} measured when determining R_{200} based on the assumption of spherical symmetry.

S5 COOLING MODES

Infalling diffuse gas is expected to shock-heat as it joins a halo. At early times and for low-mass haloes 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 disc at essentially the free-fall rate. At later times and for higher mass haloes 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 $\sim 10^{12}\,\mathrm{M}_\odot$ (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 haloes 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:

$$t_{\rm cool}(r) = \frac{3\mu m_{\rm H} k T_{200c}}{2\rho_{\rm hot}(r)\Lambda(T_{\rm hot}, Z_{\rm hot})}, \quad \text{(S3)}$$

where $\mu m_{\rm H}$ is the mean particle mass, k is the Boltzmann constant, $\rho_{\rm hot}(r)$ is the hot gas density and $Z_{\rm hot}$ is the hot gas metallicity. $T_{\rm hot}$ is the temperature of the hot gas which is assumed to be the virial temperature of the halo given by $T_{200c} = 35.9\,(V_{200c}/\mathrm{km\,s^{-1}})^2\,\mathrm{K}$ (for subhaloes we use this temperature as estimated at infall). $\Lambda(T_{\rm hot}, Z_{\rm hot})$ 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:

$$\rho_{\rm hot}(r) = \frac{M_{\rm hot}}{4\pi R_{200c} r^2} \quad \text{(S4)}$$

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

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

where $t_{\rm dyn,h}$ is the halo dynamical time defined as $R_{200c}/V_{200c} = 0.1H(z)^{-1}$ (De Lucia et al. 2004b). The specific choice of coefficient for the dynamical time of the halo is, of course, somewhat arbitrary.

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

$$\dot{M}_{\rm cool} = M_{\rm hot}\frac{r_{\rm cool}}{R_{200c}}\frac{1}{t_{\rm dyn,h}}. \quad \text{(S6)}$$

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

$$\dot{M}_{\rm cool} = \frac{M_{\rm hot}}{t_{\rm dyn,h}}. \quad \text{(S7)}$$

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

S6 RESOLVED PROPERTIES OF DISCS

Semi-analytic models traditionally aim to follow the key baryonic components of galaxies in a global manner. i.e. a single value describes the mass of cold gas in the interstellar medium (ISM), of stars in a disc, or of hot gas in a halo, with no attempt to model the spatial distribution of material within these components. Consequently, there is a relatively simple connection between dark matter and galaxy baryonic properties.

Despite introducing a significant additional layer of complexity, there are advantages to following the internal structure of these

components. One clear benefit is the possibility of comparing directly to kpc-scale observations of properties such as stellar or gas surface density, SFR or metallicity within nearby galaxies. Such comparisons are becoming viable now that dedicated surveys with modern multi-object IFU spectrographs have greatly increased the number of galaxies with spatially resolved data(Bacon et al. 2010; Croom et al. 2012; Sánchez et al. 2012; Bundy et al. 2015).

The ability to track the internal structure of galactic discs is also critical for modelling the transition from atomic to molecular gas. Molecular gas formation is believed to depend on gas density and to occur predominantly in the densest regions near the centres of galaxies. Thus, to follow the formation of H_2, it is necessary to model the surface density distribution of cold gas within discs. This in turn enables implementation of a spatially resolved model for star formation based on H_2 surface density rather than on total ISM gas content.

In the present work, we limit spatial tracking to the stellar and gas discs of galaxies. Following Fu et al. (2013), this is done by dividing discs into a series of concentric annuli or 'rings', with outer edges given by:

$$r_i = 0.01 \times 2.0^i \, h^{-1}\mathrm{kpc} \; (\mathrm{i} = 1, 2...12). \tag{S8}$$

This allows us to follow the accretion of cooling gas, the transition of atomic to molecular gas, the conversion of molecular gas into stars, and the release of energy and enriched material in individual rings within each model galaxy. The radius we adopt for the inner rings is slightly smaller than in Fu et al. (2013) in order to resolve the properties of dwarf galaxies. The fact that the same ring radii are used for all galaxies allows easy addition of each baryonic component ring by ring when two galaxies merge, though this is, of course, a grossly oversimplified version of what actually happens in a galaxy merger.

S7 GAS INFALL INTO GALAXIES

The first step in the modelling of resolved disc properties is the choice of the surface density profile of newly added material that cools from the hot halo. We assume that newly accreted cold gas follows an exponential profile (with a uniform metallicity equal to that of the hot gas):

$$\Sigma_{\mathrm{gas}}(r) = \Sigma^0_{\mathrm{gas}} \exp(-r/r_{\mathrm{infall}}), \tag{S9}$$

where:

$$\Sigma^0_{\mathrm{gas}} = \frac{m_{\mathrm{cool}}}{2\pi r^2_{\mathrm{infall}}}. \tag{S10}$$

Assuming angular momentum is conserved during gas cooling and infall (Mo et al. 1998):

$$r_{\mathrm{infall}} = \frac{j_{\mathrm{halo}}}{2V_{\mathrm{c}}}, \tag{S11}$$

where j_{halo} is the specific angular momentum of the halo and V_{c} is its circular speed.[2] Unlike in Guo et al. (2011) and Henriques et al. (2015), we assume that the angular momenta of infalling material is aligned with that already in the disc. This, combined with having a ring structure with fixed sizes, allows us to directly add new material into pre-existing discs without having to shuffle material during accretion. The scale length of the accreted gas is determined from the spin parameter of the halo of the central galaxy and added to the existing disc. We treat the accretion of cold gas from satellite galaxies during mergers as any other gas accretion event.

For each cooling episode, the newly accreted gas is directly superimposed onto the pre-existing gas profile. Since the disc and halo sizes are smaller at high redshift, so is the scale length of the infalling gas. As a result, the radial extent of the infalling material is larger at later times, causing the disc to grow (see Figure 1 in Fu et al. 2010). This naturally leads to an inside-out growth of discs, as is incorporated in many disc formation models (e.g. Kauffmann 1996; Dalcanton et al. 1997; Avila-Reese et al. 1998; Dutton 2009; Fu et al. 2009; Pilkington et al. 2012).

[2] For an isothermal sphere, V_{c} is independent of radius; we set it equal to the maximum circular velocity of the dark matter halo.

S8 H_2 FORMATION

Fu et al. (2010, 2012, 2013) have tested different prescriptions for the conversion of atomic hydrogen into molecules. In particular, two prescriptions were implemented. The first was the Krumholz et al. (2009) model, in which the H_2 fraction is primarily a function of local cold gas surface density and metallicity. The second was the relation originating from Elmegreen (1989, 1993), Blitz & Rosolowsky (2006) and Obreschkow & Rawlings (2009), in which the $f_{\mathrm{H_2}}$ is a function of the pressure in the ISM.

As shown in Fu et al. (2010, 2012, 2013) galaxy properties are quite insensitive to this choice and depend much more critically on the adopted star formation law. Nevertheless, some differences can be seen for the metal content of cold gas, which better resemble observations when using the Krumholz et al. (2009) prescription. Therefore, this will be our default choice.

The Krumholz et al. (2009) model for H_2 formation calculates an equilibrium H_2 fraction, $f_{\mathrm{H_2}} = \Sigma_{\mathrm{H_2}}/(\Sigma_{\mathrm{H_2}} + \Sigma_{\mathrm{HI}})$, for a spherical cloud with a given dust content and surrounded by a photo-dissociating UV field. Following Fu et al. (2013) this prescription is updated using the McKee & Krumholz (2010) fitting equations with the molecular gas fraction $f_{\mathrm{H_2}}$ given by:

$$f_{\mathrm{H_2}} = \begin{cases} \frac{2(2-s)}{4+s}, & s < 2; \\ 0, & s \geqslant 2. \end{cases} \tag{S12}$$

In this prescription, s is given by:

$$s = \frac{\ln(1 + 0.6\chi + 0.01\chi^2)}{0.6\tau_c}, \tag{S13}$$

in which $\chi = 3.1\,(1 + 3.1Z'^{0.365})/4.1$, $\tau_c = 0.066\,(\Sigma_{\mathrm{comp}}/\mathrm{M_\odot pc^{-2}})\,Z'$, $Z' = Z_{\mathrm{gas}}/Z_\odot$ is the gas-phase metallicity in solar units (with $Z_\odot = 0.0134$, following Asplund et al. 2009) and Σ_{comp} is the gas surface density of the gas cloud. Since the gas surface density in the model is the azimuthally averaged value in each concentric ring, a clumping factor c_{f} is introduced to take into account the fact that the gas in real disc galaxies is not smooth. We introduce an effective gas density Σ_{comp}:

$$\Sigma_{\mathrm{comp}} = c_{\mathrm{f}}\Sigma_{\mathrm{gas}}. \tag{S14}$$

Since there is observational evidence that the gas in metal poor dwarf galaxies is more clumpy than in more metal rich galaxies like our own Milky Way (Lo et al. 1993; Stil & Israel 2002), we adopt a

variable clumping factor that depends on gas-phase metallicity:

$$c_{\rm f} = \begin{cases} 0.01^{-0.7}, & Z' < 0.01; \\ Z'^{-0.7}, & 0.01 \leqslant Z' < 1; \\ 1, & Z' \geqslant 1. \end{cases} \quad \text{(S15)}$$

Fu et al. (2013) found that the Krumholz et al. (2009) prescription can easily yield non-convergent results at very low metallicities. The reason is that molecular cloud formation can only happen after metals have been produced, while metal production requires molecular clouds and star formation. As a result, the H_2 formation rates for galaxies that have recently started forming stars are quite uncertain. We therefore assume that galaxies with $Z' < 0.01$ have $Z' = 0.01$ for the purpose of calculating the clumping factor, which yields $c_{\rm f} \sim 25$, in agreement with the values of 20-30 used in simulations of high-redshift, low-metallicity systems (e.g. Wang et al. 2011; Kuhlen et al. 2012). Between $Z' = 0.1$ and 1 the clumping factor varies from 5 to 1, which agrees with the values suggested for normal galaxies in Krumholz et al. (2009).

S9 H_2 BASED STAR FORMATION LAW

Once a model for the spatially resolved formation of H_2 has been implemented, it is possible to adopt a star formation law in which the amount of stars formed is directly related to the amount of H_2 present in a certain region of a galactic disc. As in Fu et al. (2013), we assume that star formation surface density is proportional to the H_2 surface density (e.g. Leroy et al. 2008; Bigiel et al. 2011; Schruba et al. 2011; Leroy et al. 2013). However, we also include an inverse dependence with dynamical time, such that:

$$\Sigma_{\rm SFR} = \alpha_{\rm H_2} \Sigma_{\rm H_2} / t_{\rm dyn}, \quad \text{(S16)}$$

where $t_{\rm dyn} = R_{\rm cold}/V_{\rm max}$. This ensures that star formation is more efficient at early times where dynamical times are shorter, in accordance with recent observational findings (Scoville et al. 2017; Genzel et al. 2015). In addition, it makes more physical sense to have gas clouds collapsing on a timescale that is related to the dynamical time of the disc rather than a universal constant.

S10 GAS INFLOW

Whenever gas cools into the disc we assume that it retains its angular momentum and that there is no subsequent angular momentum loss during mergers or disc instabilities. In reality, both the differences in angular momentum between disc and infalling material (Lynden-Bell & Pringle 1974), the transfer of angular momentum between the disc and the dark matter halo and galaxy interactions can lead to the radial inflow of material towards the centre of galactic discs. As in Fu et al. (2013), we attempt to model this phenomena by introducing a radial inflow of gas from the outer to the inner disc. This should be interpreted as a simple phenomenological description of the effect; a physically-motivated treatment of this process will be included in future work.

We assume that the rate of change of the gas disc angular momentum is proportional to its angular momentum:

$$\frac{dL_{\rm gas}}{dt} \propto L_{\rm gas}, \quad \text{(S17)}$$

which leads to the velocity scaling with distance from the centre:

$$v_{\rm inflow} = \alpha_v r = \frac{r}{t_v}, \quad \text{(S18)}$$

since $L_{\rm gas} = m_{\rm gas} r_{\rm gas} v_{\rm cir}$.

We do not include α_v in the MCMC sampling of the parameter space since we do not include observations of spatially resolved properties as constraints. Nevertheless, we adopt a larger value than in Fu et al. (2013) ($\alpha_v = 1.0$ instead of $0.7\,{\rm km\,s^{-1}kpc^{-1}}$, corresponding to $t_v \approx 1.1$ instead of 1.4 Gyr) in order to reach a compromise between having extended enough gas profiles for Milky-Way like galaxies and ensuring that gas in the outer rings of very massive galaxies flows to the centre and forms stars within a few Gyrs after cooling has stopped. This is necessary to reproduce the predominantly red colours of these systems at $z = 0$.

S11 SUPERNOVA FEEDBACK

Stars release large amounts of mass and energy in the latter stages of their evolution, both through supernovae and through stellar winds. In previous versions of our model we assumed that most metals and energy were released by short lived massive stars and adopted the instantaneous recycling approximation: the full yield of metals and energy were released immediately after any episode of star formation. With the incorporation of a spatially resolved chemical enrichment model we can now track in detail the release of energy and metals by stars of different mass at the appropriate time after their formation.

At any given time, energy is released by the different stellar populations in each galactic ring and reheats part of its cold interstellar medium into the hot gas atmosphere. 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 immediately 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:

$$\Delta E_{\rm SN} = \epsilon_{\rm halo} \times \Delta M_{\star,R} \eta_{\rm SN} E_{\rm SN}, \quad \text{(S19)}$$

where $\Delta M_{\star,R}$ is the mass returned to the ISM via stellar winds and SNe (as opposed to the mass of stars formed used in Henriques et al. 2015), $\eta_{\rm SN}$ is the number of supernovae expected per solar mass of stars returned to the ISM ($0.0149\,{\rm M_\odot}^{-1}$, assuming a universal Chabrier 2003 IMF), $E_{\rm SN}$ is the energy released by each supernova (10^{51} erg) and $\epsilon_{\rm halo}$ is a free parameter given by:

$$\epsilon_{\rm halo} = \eta_{\rm eject} \times \left[0.5 + \left(\frac{V_{\rm max}}{V_{\rm eject}} \right)^{-\beta_{\rm eject}} \right]. \quad \text{(S20)}$$

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 returned to the ISM:

$$\Delta M_{{\rm reheat},i} = \epsilon_{\rm disc} \Delta M_{\star,R_i}, \quad \text{(S21)}$$

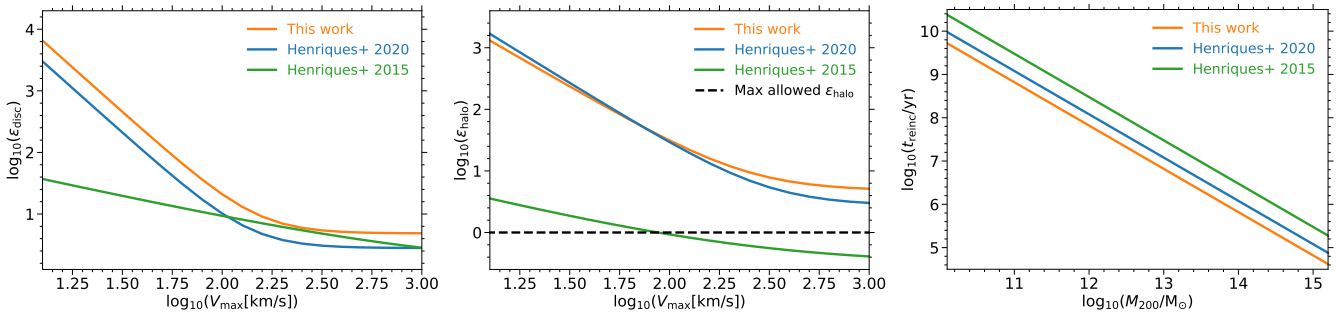

Figure S1. Illustration of the dependencies of SN feedback related parameters on halo properties in our model (orange), Henriques et al. (2015) (green), and Henriques et al. (2020) (blue). The left panel shows the disc reheating efficiency $\epsilon_{\rm disc}$ as a function of maximum circular velocity $V_{\rm max}$. This relates to the mass-loading factor for material reheated into the hot gas component. The middle panel shows the halo ejection efficiency $\epsilon_{\rm halo}$ as a function of $V_{\rm max}$. This is the fraction of the available SN energy which is used in reheating disc gas and in ejecting hot gas from the halo. We note that the maximum value allowed for this quantity is one. The right panel shows the reincorporation time-scale $t_{\rm reinc}$ as a function of halo virial mass M_{200}.

where the second efficiency is:

$$\epsilon_{\rm disc} = \epsilon_{\rm reheat} \times \left[0.5 + \left(\frac{V_{\rm max}}{V_{\rm reheat}}\right)^{-\beta_{\rm reheat}}\right] \tag{S22}$$

and $\Delta M_{{\rm reheat},i}$ and $\Delta M_{\star,R_i}$ are computed locally for each ring. The dependence of $\epsilon_{\rm disc}$ on $V_{\rm max}$ is shown in the left panel of Fig. S1. This reheating is assumed to require energy $\Delta E_{\rm reheat,i} = \frac{1}{2}\Delta M_{{\rm reheat},i} V^2_{\rm 200c}$. If $\Delta E_{{\rm reheat},i} > \Delta E_{\rm SN} \times \Delta M_{\star,R_i}/\Delta M_{\star,R}$, the reheated mass in each ring is assumed to saturate at $\Delta M_{{\rm reheat},i} = \Delta E_{\rm SN} \times \Delta M_{\star,R_i}/\Delta M_{\star,R}/\left(\frac{1}{2}V^2_{\rm 200c}\right)$.

Any remaining SN energy is added up and used to eject a mass $\Delta M_{\rm eject}$ of hot gas into an external reservoir:

$$\frac{1}{2}\Delta M_{\rm eject} V^2_{\rm 200c} = \Delta E_{\rm SN} - \Delta E_{\rm reheat}, \tag{S23}$$

where $\Delta E_{\rm reheat}$ is the sum over all the rings.

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. As clarified in Yates et al. (submitted), the mass-loading factors for reheating in L-Galaxies 2020 (i.e. $\Delta M_{\rm reheat}/\Delta M_\star$) are not exactly the same as the $\epsilon_{\rm disc}$ parameter defined by Eq. S22 and shown in the left panel of Fig. S1. This is because the model now allows (a) mass to be returned by stars at different times, depending on their mass and metallicity, and (b) some fraction of this returned mass to be deposited directly into the hot gas compnent (see Section S14). Therefore, mass-loading factors for reheating are typically lower than $\epsilon_{\rm disc}$, lying in the region of 1-10, as often seen in observations and simulations.

Ejection of disc 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 middle panel of Fig. S1 shows $\epsilon_{\rm halo}$, the fraction of the available energy that is used in feedback processes, as a function of $V_{\rm max}$.

It is interesting to note that the best-fit parameters resulting from the MCMC sampling effectively represent a model with significantly fewer degrees of freedom than originally allowed. While three parameters were used to describe each of the functional forms showed in the left and middle panels of Fig. S1, for our best fit model they can be represented only using three parameters in total. In practice $\epsilon_{\rm disc}$ can be described by an inverse linear dependence with $V_{\rm max}$ (2 parameters) while all the available SN energy is used (1 parameter).

S11.1 SN feedback in satellite galaxies

The details of gas reheating and ejection just described, accurately represent the impact of SN feedback in isolated galaxies and in galaxies at the centre of an FoF group (type 0's). For satellites at the centre of a subhalo (type 1's), or orphan satellites with no dark matter, hot or ejected gas (type 2's), the impact of environment must be taken into account. When gas is reheated into the hot phase of a type 1 galaxy, either its own cold gas or gas originated at a type 2 satellite, a fraction is immediately removed due to tidal stripping. In type 2's, reheating will move their own cold gas into the hot phase of their direct central companion, either a type 1 galaxy at the centre of a subhalo or a type 0 galaxy at the centre of the main halo, from which the left-over energy will eject material.

When calculating the amount of energy available from SN and the reheating efficiency, eqs. S20 and S22, we always use the $V_{\rm max}$ of the halo from where the gas will be moved, using the value at infall for satellites. The value for the virial temperature at which reheating saturates and the escape velocity of haloes at which gas is ejected are always taken from the halo where the gas will end up.

S12 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 ($8.0 \leqslant \log M_*/{\rm M}_\odot \leqslant 9.5$) too early, leading to an overabundance of these objects at $z \geqslant 1$ (Fontanot et al. 2009; Henriques et al. 2011; Guo et al. 2011; Weinmann et al. 2012; Lu et al. 2014; 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 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:

$$\dot{M}_{\rm ejec} = -\frac{M_{\rm ejec}}{t_{\rm reinc}}, \tag{S24}$$

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

$$t_{\rm reinc} = \gamma_{\rm reinc} \frac{10^{10}\,{\rm M}_\odot}{M_{\rm 200c}}, \tag{S25}$$

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.

Reincorporation times for Henriques et al. (2015) and our new model are shown as a function of virial velocity and redshift in the lower right panel of Fig. S1. Note that the redshift dependence simply results from the relation between $M_{\rm 200c}$ and $V_{\rm 200c}$ (eq. S1). In practice gas ejected in winds from low-mass haloes will never be reincorporated unless they become part of a more massive system, while gas returns immediately in the most massive haloes. This implementation agrees qualitatively with the behaviour seen by Oppenheimer & Davé (2008) and Oppenheimer et al. (2010) in their numerical simulations.

S13 DETAILED STAR FORMATION HISTORIES

As in Henriques et al. (2015), the current version of our model stores star formation and metal enrichment histories using the algorithm described in Shamshiri et al. (2015). This aspect is essential for the implementation of the Yates et al. (2013) chemical enrichment model in order to follow the time-dependent release of mass and energy by dying stars. In addition, this makes it possible to compute luminosities, colours and spectral properties in post-processing using any stellar population synthesis model.

An important difference relevant for the present version of our model is that since the detailed chemical enrichment scheme is implemented at the level of the spatially resolved rings in galactic discs, the recorded star formation and metal enrichment histories must also be recorded separately for each ring. We store these history bins in each ring with a time resolution that degrades for older stellar populations. The most recent activity is always stored with the maximum resolution, which is set to be equal to a single substep of the main timestep of the semi-analytic model. As the computation progresses, older bins are merged together logarithmically. Shamshiri et al. (2015) showed that with each timestep split into ~20 substeps, one can recover the UV luminosities of galaxies in post-processing with less than 10% scatter for more than 90% of the galaxies at any given time (and with much lower scatter at longer wavelengths).

The resulting bin structure is shown as a function of time in Fig. S2. As can be seen, there is a relatively high resolution for stellar populations with ages less than ~1 Gyr. Typically, these stellar populations are tracked with more than 5 bins with the highest resolution timesteps varying from a few Myr at high redshifts to just over 10 Myr at low redshifts. Populations with ages more than 1 Gyr are represented by ~ half a dozen bins at $z = 0$.

S14 DETAILED CHEMICAL ENRICHMENT

With the present update of L-Galaxies, we also incorporate a galactic chemical enrichment (GCE) model that tracks how much enriched material is returned to the interstellar medium and circumgalactic medium by each stellar population at any given time. This GCE scheme, introduced by Yates et al. (2013), directly follows the delayed enrichment of eleven individual chemical elements (H, He, C, N, O, Ne, Mg, Si, S, Ca, and Fe) produced by SNe-Ia, SNe-II, and winds from AGB stars, adopting mass and metallicity-dependent stellar yields and lifetimes. This scheme also includes a reformulation of the associated SN feedback, so that energy and heavy elements are released into the ISM and CGM when stars die, rather than when they are born (i.e. we eliminate the instantaneous recycling approximation). In addition, in the present work, all the material and energy ejected into the ISM by stellar populations is spatially resolved by being placed in the same ring in which the star or supernova occurred.

The bulk of the GCE set-up used in this new version of L-Galaxies is the same as described in Yates et al. (2013). In brief, the total ejection rate of chemical element X by a simple stellar population (SSP) at time t is given by,

$$e_{\rm X}(t) = \int_{M_L}^{M_U} M_{\rm X}(M, Z_0)\, \psi(t - \tau_{\rm M})\, \phi(M)\, {\rm dM}\ , \tag{S26}$$

where M is the initial mass of a star, $\tau_{\rm M}$ is its lifetime, M_L is the lowest mass of star to eject material at time t (i.e. one with a lifetime of $\tau_{\rm M} = t$), M_U is the maximum star mass considered ($120\,{\rm M}_\odot$ in this work), and $M_{\rm X}$ is the mass of element X ejected per star, which depends on the initial mass M and initial metallicity Z_0. This ejecta mass comprises both the yield, $y_{\rm X}$, and the mass of element X that passes through a star unprocessed before being ejected. Finally, $\psi(t - \tau_{\rm M})$ is the SFR at a star's birth, and $\phi(M)\,{\rm d}M$ is the number of stars in the mass range $M \mapsto M + {\rm d}M$ per unit mass of star formation. For more details, we refer the reader to section 4 of Yates et al. (2013).

Mass- and metallicity-dependent yields are taken from Marigo (2001) for AGB stars, from Thielemann et al. (2003) for SNe-Ia (not metallicity dependent), and from Portinari et al. (1998) for SNe-II. Mass- and metallicity-dependent stellar lifetimes are also taken from the calculations of Portinari et al. (1998). A Chabrier (2003) IMF with a constant high-mass-end slope of $\alpha_{\rm IMF} = -2.3$ is always assumed. When modelling the lifetimes of SNe-Ia, a constant power-law delay-time distribution (DTD) with a slope of -1.12 is assumed, following Maoz et al. (2012).

Two modifications to the parameters governing GCE have been made here, in comparison to those chosen by Yates et al. (2013). First, the fraction of objects between 3 and $16\,{\rm M}_\odot$ in each stellar population that are assumed to form SN-Ia progenitors has been increased from $A_{\rm old} = 0.028$ to $A = 0.04$ (for our chosen IMF, this is equivalent to a fraction of *all* objects in a stellar population that are assumed to form SN-Ia progenitors of $A' = 0.00154$). This new value is still well within the range inferred from observations of the SN-Ia rate, which suggest $0.024 \lesssim A \lesssim 0.081$ (Maoz et al. 2012).

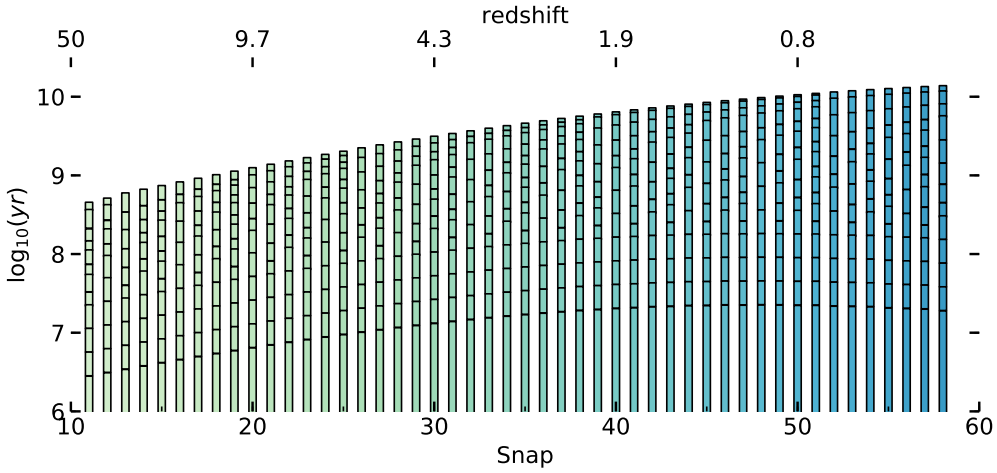

Figure S2. The binning structure in which star formation histories are recorded at different snapshots/redshifts. The most recent bins, at the bottom of the plot, have the best resolution, with a median age corresponding to a single internal time-step for the semi-analytic calculation (20 times smaller than the time between snapshots). Older stellar populations are binned together resulting in widths ranging from a few to ten million years for recent bins to a few Gyrs for the oldest stars.

Second, the amount of direct enrichment of the hot circumgalactic medium (CGM) by supernovae has been modified. Previously, all supernovae exploding in the stellar disc were assumed to directly enrich the ISM. This material was then fully mixed with the ambient cold gas, before being reheated into the CGM, or expelled from the dark matter halo in galactic winds. In our new version of L-Galaxies, we instead allow 30% of the ejecta from both SNe-II and SNe-Ia to be directly dumped into the CGM, constituting a metal-rich wind which is then allowed to cool and re-accrete onto the galaxy along with the ambient hot gas. More sophisticated prescriptions for the direct enrichment of the CGM will be investigated in future work.

S15 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 ($\log M_*/\mathrm{M}_\odot \geqslant 10.5$) 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 haloes. 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).

S15.1 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,

$$\Delta M_{\mathrm{BH,Q}} = \frac{f_{\mathrm{BH}}(M_{\mathrm{sat}}/M_{\mathrm{cen}})\, M_{\mathrm{cold}}}{1 + (V_{\mathrm{BH}}/V_{\mathrm{200c}})^2}, \tag{S27}$$

where M_{cen} and M_{sat} are the total baryon masses of the central galaxy and the satellite which merges with it, M_{cold} is their total cold gas mass, V_{200c} is the virial velocity of the central halo and f_{BH} and V_{BH} are two adjustable parameters which control the fraction of the available cold gas that is accreted and the virial velocity at which the efficiency saturates. The mass of the black hole at the centre of the final merged galaxy is thus taken to be $M_{\mathrm{BH,f}} = M_{\mathrm{BH,1}} + M_{\mathrm{BH,2}} + \Delta M_{\mathrm{BH,Q}}$ where the subscripts 1 and 2 denote the masses of the progenitor black holes.

Mass accretion during mergers is the main channel of black hole growth in our model. The fact that bulges and black holes are formed in related processes results in a tight relation between black hole and bulge masses.

S15.2 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. Following Henriques et al. (2015) we used a modified version of the original Croton et al. (2006) model for radio mode feedback:

$$\dot{M}_{\mathrm{BH}} = k_{\mathrm{AGN}} \left(\frac{M_{\mathrm{hot}}}{10^{11}\,\mathrm{M}_\odot}\right) \left(\frac{M_{\mathrm{BH}}}{10^{8}\,\mathrm{M}_\odot}\right). \tag{S28}$$

This formula is equivalent to that of Croton et al. (2006) divided by a factor of $H(z)/H_0$, so accretion is enhanced at lower redshifts. A comparison for the efficiency of AGN growth and feedback between the current model and that of Henriques et al. (2015) is shown in Fig. S3.

In our new model, as in its predecessors, the mass growth of

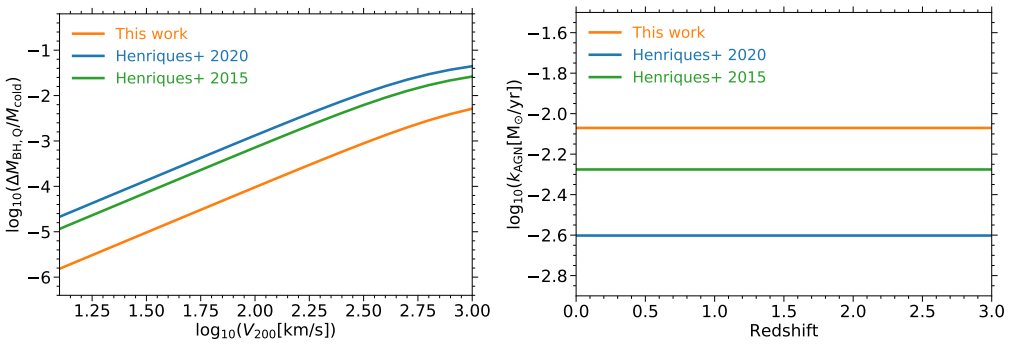

Figure S3. The scalings of the processes controlling black hole growth and AGN feedback in our model (orange), Henriques et al. (2015) (green), and Henriques et al. (2020) (blue). The left panel shows the maximum fraction of cold gas accreted (for a major merger of equal mass galaxies) onto central black holes during mergers (quasar accretion) as a function of virial velocity (eq. S27). 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. S28) as a function of redshift.

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; Bîrzan et al. 2004). The energy input rate is taken to be

$$\dot{E}_{\rm radio} = \eta \dot{M}_{\rm BH} c^2, \tag{S29}$$

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 disc, resulting in an effective cooling rate given by

$$\dot{M}_{\rm cool,eff} = \max\left[\dot{M}_{\rm cool} - 2\dot{E}_{\rm radio}/V^2_{\rm 200c},\ 0\right]. \tag{S30}$$

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 as a purely phenomenological representation of some process which acts to prevent the cooling of gas onto massive central galaxies without requiring additional star formation.

In addition to the change of eq. S28, a feature that we recently discovered in our dark matter merger trees motivated another adjustment to our AGN radio mode feedback. For one massive dark matter structure in the Millennium Simulation, a low-mass satellite subhalo is at some point converted into the main subhalo of the FOF group. As a result, a very small galaxy with a low-mass black hole suddenly acquires $> 10^{14}\,{\rm M}_\odot$ of hot gas. This leads to a short episode of catastrophic cooling that increases the stellar mass of the central galaxy by as much as two orders of magnitude.

In order to correct this numerical artefact we assume that the AGN energy left over after offsetting the cooling of hot gas in satellite galaxies can be used to offset cooling in the hot gas of their hosts (for satellites within R_{200}). As a result, when the low-mass galaxy suddenly becomes the central object of a cluster, the energy released from black holes in other satellite galaxies is enough to suppress excessive cooling from the hot atmosphere onto the central object.

S16 ENVIRONMENTAL PROCESSES

The growth of structure in a ΛCDM universe affects galaxies by tidal forces, 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.

Early versions of L-Galaxies completely stripped hot gas out of satellite galaxies once they fell within R_{200}. Modern versions, since Guo et al. (2011), have gradual hot gas stripping, and therefore satellite galaxies are able to retain a fraction of their hot gas. The (Henriques et al. 2015, 2020) versions of the model implement tidal stripping for satellite galaxies within the halo boundary, R_{200}, and limit ram-pressure stripping to satellites within R_{200} of massive clusters with $M_{200} > M_{\rm r.p.}$, where $\log_{10}(M_{\rm r.p.}/{\rm M}_\odot) = 14.7$ in H20. This ram-pressure stripping threshold, $M_{\rm r.p.}$, was a free parameter in the H15 and H20 model calibrations, and the resulting value was found necessary in order to avoid having too many low-mass, red galaxies. We note that although this results in good agreement with the observed quenched fraction, the approach is merely a numerical fix and is not physical. Following Ayromlou et al. (2019), we completely remove this mass threshold and extend ram-pressure stripping to all galaxies in the simulations. We also extend tidal stripping to all satellite galaxies, both within and beyond the halo virial radius.

S16.1 Measuring the properties of the local background environment of galaxies

In order to model ram-pressure we must quantify the local environmental properties of galaxies. To do so we follow the method introduced by Ayromlou et al. (2019) which measures the Local Background Environment (LBE) of each galaxy directly from the particle data of the simulation. This background is defined within a spherical shell surrounding the galaxy and its subhalo. We choose the background shell's radii to exclude the galaxy and its subhalo, and keep the LBE sufficiently local. We choose the inner radius, $r_{\rm in} = 1.25\, r_{\rm subhalo}$ and the outer radius, $r_{\rm out} = 2\, r_{\rm subhalo}$, where $r_{\rm subhalo}$ is the subhalo radius and is defined as the distance between the most bound and the most distant subhalo particles. In order to have proper statistics, we set a minimum of $n_{\rm min} = 30$ for the number of particles within the shell, which results in a statistical error smaller than $1/\sqrt{n_{\rm min}} \sim 20\%$. If there are fewer than $n_{\rm min}$ particles within the shell, we allow the outer radius to extend as needed. The

shell density, $\rho_{\rm shell}$ is the total mass within the shell divided by its volume. The mean density of the shell, $\vec{v}_{\rm shell}$, is calculated by averaging over the velocities of all the shell particles.

We choose the inner radius of the background shell larger than the subhalo size to remove all the detected (via SUBFIND, Springel et al. 2001) bound subhalo particles. However, the background shell might still contain subhalo-associated particles which are not detected as a part of the subhalo by the subhalo finder algorithm. These particles contaminate the background shell and cause misleading values for its density and velocity, and in Ayromlou et al. (2019) we devised a Gaussian mixture method to remove this contribution. In this paper we use the decontaminated, pure LBE properties to measure ram-pressure for every simulation galaxy.

Our final step is to convert the LBE properties taken from DMO simulations to local estimates of the properties of the gas, needed to estimate ram-pressure. For the LBE velocity, we assume that gas and dark matter follow each other and, therefore, $\vec{v}_{\rm LBE,gas} = \vec{v}_{\rm LBE}$. To derive the gas density we take slightly different approaches for central and satellite galaxies. For central galaxies, we multiply the total LBE density, $\rho_{\rm LBE}$, by the cosmic baryon fraction $\Omega_{\rm b}$, while for satellites we multiply it by $f_{\rm hotgas}$, which is the hot gas fraction of the parent host FOF halo. The details of the method are explained in §3 of Ayromlou et al. (2019).

S16.2 Ram-pressure stripping

If the ram-pressure force from a subhalo's environment becomes stronger than the subhalo's self-gravity on its gas, the gas beyond a given scale, called the stripping radius $R_{\rm rp}$, will be stripped out of the subhalo. The stripping radius is defined as the radius where these two forces are equal. We calculate the ram-pressure using the formula derived by Gunn & Gott (1972)

$$P_{\rm rp} = \rho_{\rm LBE,gas}\, v^2_{\rm gal,LBE}\,, \tag{S31}$$

where $\rho_{\rm LBE,gas}$ and $v_{\rm gal,LBE}$ are the gas density of the galaxy's LBE, and the velocity of the galaxy relative to its local environment, respectively. Both are measured directly using the simulation particle data as described in §S16.1. In contrast, L-GALAXIES models prior to this work have adopted $\rho_{\rm LBE,gas} \propto r^{-2}$ and $v_{\rm gal,LBE}$ equal to the virial velocity (V_{200}) of the satellite's host halo, with no ram-pressure force for central galaxies nor satellites beyond R_{200}.

For the gravitational restoring force we adopt the approach suggested by McCarthy et al. (2008) as extended in Ayromlou et al. (2019). This approach was implemented in L-Galaxies by Guo et al. (2011). However, it was limited to satellite galaxies within the halo R_{200}. The gravitational restoring force per unit area is

$$F_{\rm g}(r) = g_{\rm max}(r)\, \rho^{\rm proj}_{\rm hotgas}(r)\,, \tag{S32}$$

where $g_{\rm max}(r)$ is the maximum gravitational acceleration of the subhalo on its hot gas and $\rho^{\rm proj}_{\rm hotgas}(r)$ is the subhalo's projected hot gas density. Considering a spherically symmetric isothermal profile with $\rho_{\rm hotgas} \propto r^{-2}$ for the hot gas, $g_{\rm max}(r)$ equals

$$g_{\rm max}(r) = \frac{G M_{\rm subhalo}(r)}{2r^2}, \tag{S33}$$

where $M_{\rm subhalo}(r)$ is the subhalo mass within the radius r. Taking a similar isothermal $\rho_{\rm subhalo} \propto r^{-2}$ profile for the subhalo, we can calculate the subhalo mass at any given scale as

$$M_{\rm subhalo}(r) = M_{\rm g}\frac{r}{R_{\rm g}}, \tag{S34}$$

where $M_{\rm g}$ is the mass within $R_{\rm g}$ and both are known variables which we use to estimate $M_{\rm subhalo}(r)$. In earlier L-GALAXIES versions, $R_{\rm g}$ and $M_{\rm g}$ were taken to be R_{200} and M_{200} at infall time for satellite galaxies. Although these are reasonable estimates, they represent the satellite subhalo at an earlier time and ignore its evolution. Therefore, for satellite galaxies we take $R_{\rm g}$ and $M_{\rm g}$ to be the half mass radius, $R_{\rm halfmass}$, and the total mass within $R_{\rm halfmass}$ radius ($M_{\rm halfmass}$) at every redshift, thus representing the subhalo's current properties rather than its properties at infall. For central galaxies, we take $R_{\rm g} = R_{200}$ and $M_{\rm g} = M_{200}$.

Assuming an isothermal ($\rho_{\rm hotgas} \propto r^{-2}$) profile for the subhalo's hot gas, with the boundary condition of having total mass of $M_{\rm hotgas}$ within the hot gas radius, $R_{\rm hotgas}$, the 2D projected hot gas density is given as

$$\rho^{\rm proj}_{\rm hotgas}(r) = \frac{M_{\rm hotgas}}{2\pi R_{\rm hotgas} r}. \tag{S35}$$

Finally, the ram-pressure radius, the scale on which the ram-pressure equals the gravitational restoring force per unit area, is calculated to be

$$R_{\rm rp} = \left(\frac{G M_{\rm g} M_{\rm hotgas}}{4\pi R_{\rm g} R_{\rm hotgas} \rho_{\rm LBE,hotgas} v^2_{\rm gal,LBE}}\right)^{1/2}. \tag{S36}$$

We apply the above formula to all the subhaloes uniformly.

S16.3 Tidal stripping

Unlike ram-pressure that strips gas from subhaloes, tidal stripping is able to strip both dark and baryonic matter out of subhaloes and galaxies. As the total mass change of satellite subhaloes is known from our DMO simulations, we use it to estimate the amount of the stripped gas. We assume that the fraction of gas lost is the same as the fraction of dark matter lost

$$\frac{M_{\rm hot}(R_{\rm tidal})}{M_{\rm hot,infall}} = \frac{M_{\rm DM}}{M_{\rm DM,infall}}, \tag{S37}$$

where $M_{\rm DM,infall}$ and $M_{\rm hot,infall}$ are the satellite's virial mass and hot gas mass at infall. In addition, $M_{\rm DM}$ and $M_{\rm hot}(R_{\rm tidal})$ are the satellite's mass and hot gas mass after tidal stripping. The only unknown variable in Eq. S37 is $M_{\rm hot}(R_{\rm tidal})$. Assuming an isothermal profile ($\rho \propto r^{-2}$), we calculate the tidal stripping radius beyond which the gas is stripped

$$R_{\rm tidal} = \frac{M_{\rm DM}}{M_{\rm DM,infall}} R_{\rm hot,infall}\,, \tag{S38}$$

where $R_{\rm hot,infall}$ is the satellite's hot gas radius at infall time. We note that as orphan galaxies do not have an identified subhalo, their subhalo mass is set to zero. Eq. S37 then implies that they will lose all their hot gas as well. As a result, orphan galaxies are entirely empty of hot gas due to tidal stripping.

S16.4 Stripping implementation

Satellite galaxies are subject to both tidal and ram-pressure stripping. Therefore, we take the stripping radius to be the smaller of the tidal stripping radius and the ram-pressure stripping radius. For central galaxies, the stripping radius is equal to the ram-pressure stripping radius, as we do not consider them subject to tidal stripping. After stripping, we assume the density profile of the remaining gas remains isothermal with $\rho \propto r^{-2}$. We note that ram-pressure stripping is the dominant effect for most galaxies.

In addition to hot gas stripping we also strip the ejected reservoirs of galaxies. The fraction of gas stripped from the ejected reservoir of a galaxy is assumed to be the same as the fraction of its stripped hot gas. Stripping of material that has been expelled from the galaxy by feedback processes and is assumed to lie at least temporarily in this ejecta reservoir plays a critical role in the evolution of galaxies.

For every galaxy, if $R_{\rm strip} < R_{\rm hotgas}$, all the gas beyond $R_{\rm strip}$ is removed. For satellite galaxies within the halo boundary, R_{200}, the stripped gas goes directly to the hot gas component of their host halo. For central galaxies and satellites beyond the halo R_{200}, the stripped gas does not go to any halo immediately, but we keep track of it through time. If the galaxy subsequently falls into R_{200} of a halo, then the stripped gas is added to the host halo's hot gas with the condition that the host halo's baryon fraction does not exceed the cosmic value. After stripping, the new hot gas radius of a satellite will be the minimum of its former hot gas radius and $R_{\rm strip}$. On the other hand, as long as a galaxy is categorised as a central galaxy, since it accretes hot gas from its environment, its hot gas radius is set to its FOF R_{200}.

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). In future work, we will use the local background environment of galaxies (§S16.1) to handle cold gas stripping as well.

S16.5 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.

$$\frac{M_{\rm DM,halo}(R_{\rm peri})}{R_{\rm peri}^3} \equiv \rho_{\rm DM,halo} > \rho_{\rm sat} \equiv \frac{M_{\rm sat}}{R_{\rm sat,half}^3}, \qquad \text{(S39)}$$

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 discs 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

$$\left(\frac{R}{R_{\rm peri}}\right)^2 = \frac{\ln R/R_{\rm peri} + \frac{1}{2}(V/V_{\rm 200c})^2}{\frac{1}{2}(V_t/V_{\rm 200c})^2}, \qquad \text{(S40)}$$

assuming conservation of energy and angular momentum and a singular isothermal potential for the orbit, $\phi(R) = V_{\rm 200c}^2 \ln R$. In these equations, R is the current distance of the satellite from halo centre, and V and V_t are the total and tangential velocities of the satellite with respect to halo centre (see Section S17.1 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).

S16.6 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. As described in Section S11.1, from this point on, we also assume that any cold gas reheated 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.

S17 MERGERS AND BULGE FORMATION

S17.1 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:

$$t_{\rm friction} = \alpha_{\rm friction} \frac{V_{\rm 200c} r_{\rm sat}^2}{G M_{\rm sat} \ln\Lambda}, \qquad \text{(S41)}$$

where $M_{\rm sat}$ is the total mass of the satellite (dark and baryonic), $\ln\Lambda = \ln(1 + {\rm M_{200c}}/{\rm M_{sat}})$ is the Coulomb logarithm and $\alpha_{\rm frction}$ is a free parameter in our model calibration and is originally set by De Lucia & Blaizot (2007) to match the bright end of the luminosity functions at $z = 0$. Although this value was later compared to direct numerical simulation (Boylan-Kolchin et al. 2008; De Lucia et al. 2010), it should still be considered poorly known. The Millennium-II simulation is able to resolve subhaloes 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 $(1 - \Delta t/t_{\rm friction})$ where Δt 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 a satellite with "isothermal" density structure spiralling to the centre of an isothermal host on a circular orbit. When $\Delta t = t_{\rm friction}$ the orphan merges with the central galaxy.

S17.2 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 $R_{\rm merge}$ of that of the more massive galaxy. In a major merger, the discs of the two progenitors are destroyed and all their stars become part of of the bulge of the descendent, along with any stars formed during the merger. In a minor merger, the disc of the larger progenitor survives and accretes the cold gas component of the smaller galaxy (with the proportion allocated to each ring being the same as for the accretion of gas cooling from the hot halo), while its bulge

accretes all the stars of the victim. Stars formed during the merger stay in the disc of the descendent. In both types of merger, cold gas is fed to the central black hole according to the formulae of Section S15.1. $R_{\rm merge}$ is fixed and not included in the MCMC analysis since we find some tension between the values required for model predictions to reproduce observed colours and observed morphologies. Our adopted value of 0.1, slightly compromises the agreement with the observed red fraction of galaxies at $z = 1$, but ensures that the morphologies of massive galaxies in the model are closer to observation.

The stellar mass formed during a merger in each ring is modelled using the "collisional starburst" formulation of Somerville et al. (2001):

$$M_{\star,\rm burst} = \alpha_{\rm SF,burst} \left(\frac{M_1}{M_2}\right)^{\beta_{\rm SF,burst}} M_{\rm cold}, \tag{S42}$$

where $M_1 < M_2$ are the baryonic masses of the two galaxies, and $M_{\rm cold}$ is the cold gas mass in that ring. The $\alpha_{\rm SF,burst}$ and $\beta_{\rm SF,burst}$ 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-fitting model the fraction of cold gas converted to stars in merger-related bursts is relatively close to what was previously assumed.

S17.3 Bulge Formation in mergers and disc instabilities

In our model, bulges can form through major and minor mergers and through the buckling instability of discs. After a major merger, all stars are considered part of the new bulge, but the remnant of a minor merger retains the stellar disc of the larger progenitor and its bulge gains only the stars from the smaller progenitor. We assume that whenever a new bulge is formed during mergers its mass distribution will follow a Jaffe (1983) profile with a scale length described below. In disc instabilities, the material moved to the bulge retains its radial distance to the centre of the galaxy.

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:

$$C\frac{GM^2_{\rm new,bulge}}{R_{\rm new,bulge}} = C\frac{GM_1^2}{R_1} + C\frac{GM_2^2}{R_2} + \alpha_{\rm inter}\frac{GM_1M_2}{R_1+R_2}, \tag{S43}$$

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. C parametrizes the binding energy of the galaxy and $\alpha_{\rm inter}$ parametrizes the effective interaction energy deposited in the stellar components. As in Guo et al. (2011) we adopt $C = 0.5$ and $\alpha_{\rm inter} = 0.5$. For major mergers, M_1 and M_2 are the sum of the mass of stars and of the cold gas converted into stars for the two progenitors, and R_1 and R_2 are the corresponding half-mass radii. For minor mergers, M1 and R1 are the mass and half-mass radius of the bulge of the major progenitor, and M_2 and R_2 are the stellar mass and the half-stellar-mass radius of the minor progenitor, respectively.

We assume that the material that forms a new bulge after a merger is instantly mixed and has a characteristic length given by equation S43 above and a Jaffe (1983) profile:

$$\rho_{\rm bulge}(r) = \frac{m_{\rm bulge}}{4\pi r_{\rm b}^3}\left(\frac{r}{r_{\rm b}}\right)^{-2}\left(1+\frac{r}{r_{\rm b}}\right)^{-2}. \tag{S44}$$

For the material transferred between the discs of the two merging galaxies, we treat the accreted gas from the satellite in the same way as gas cooling from the halo. The scale length of the accreted gas is determined from the specific angular momentum and $V_{\rm max}$ of the halo of the central galaxy using Eq. S11.

Another important channel of bulge growth is secular evolution through disc 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 disc dominates the gravitational effects of the bulge and halo. As a criterion for disc instability, we follow Guo et al. (2011) in adopting

$$V_{\rm max} < \sqrt{\frac{GM_{\star,\rm d}}{3R_{\star,\rm d}}}, \tag{S45}$$

where $M_{\star,\rm d}$ and $R_{\star,\rm d}$ are the stellar mass and exponential scale-length of the stellar disc and $V_{\rm max}$ is the maximum circular velocity of the dark matter halo hosting the disc.

When the instability criterion of eq. S45 is met, we transfer sufficient stellar mass from the disc to the bulge to make the disc marginally stable again. Mass from the inner parts of the disc is transferred first, moving outwards in rings, until all the necessary material has been transferred and the disc is stable again. If this material forms a new bulge, it is assumed to have a half-mass radius equal to the outer radius of the region. If the galaxy already has a spheroidal component, the newly created bulge material is assumed to merge with the existing bulge according to eq. (S43), but adopting $\alpha_{\rm inter} = 2$. The larger value of $\alpha_{\rm inter}$ accounts for the larger interaction energy when having two concentric components with no relative motion.

We note that the present model neglects the contribution of the cold gas component to the stability of discs. This is an over-simplification and in reality instabilities in the cold gas are likely to contribute to the stability of stellar discs and themselves drive star bursts and black hole growth. A more detailed recipe for the treatment of this process has recently been incorporated in the Henriques et al. (2015) version of our model by Irodotou et al. (2018). The authors found the new treatment to significantly improve the properties of instability driven, intermediate mass bulges, and we plan to incorporate this work in future public releases.

S18 STELLAR POPULATIONS SYNTHESIS

Stellar population synthesis models are a crucial part of galaxy formation theory 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 are in better agreement with observed near-infrared emission from bright galaxies at $z \geqslant 2$ (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 $z = 3$ to 0. Nevertheless, as part of our model release we will, for comparison purposes, also include luminosities computed using Bruzual & Charlot (2003) stellar populations.

S19 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 birth clouds within 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.

S19.1 Extinction by the ISM

The optical depth of diffuse dust in galactic discs is assumed to vary with wavelength as

$$\tau_\lambda^{ISM} = (1+z)^{-1} \left(\frac{A_\lambda}{A_V}\right)_{Z_\odot} \left(\frac{Z_{\rm gas}}{Z_\odot}\right)^s \times \left(\frac{\langle N_H \rangle}{2.1\times10^{21}\,{\rm atoms\,cm^{-2}}}\right), \quad \text{(S46)}$$

where $\langle N_H \rangle$ represents the mean column density of hydrogen and is given by

$$\langle N_H \rangle = \frac{M_{\rm cold}}{1.4\, m_p \pi (a R_{\rm gas,d})^2}. \quad \text{(S47)}$$

Here $R_{\rm gas,d}$ is the cold gas disc scale-length, 1.4 accounts for the presence of helium and $a = 1.68$ in order for $\langle N_H \rangle$ to represent the mass-weighted average column density of an exponential disc. Following the results in Guiderdoni & Rocca-Volmerange (1987), the extinction curve in eq. (S46) 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 $\lambda < 2000$ Å and $s = 1.6$ for $\lambda > 2000$ Å. The extinction curve for solar metallicity, $(A_\lambda/A_V)_{Z_\odot}$, is taken from Mathis et al. (1983).

The redshift dependence in eq. (S46) 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. In practice we include it merely as a representation of the observed extinction behaviour (Bouwens et al. 2014), while the physical reason for discrepancy with the simple model remains to be established. As was shown in Clay et al. (2015), this produces luminosity functions and extinction estimates for Lyman-break galaxies at $z > 5$ compatible with HST data.

S19.2 Extinction by molecular clouds

This second source of extinction affects only young stars. 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

$$\tau_\lambda^{BC} = \tau_\lambda^{\rm ISM} \left(\frac{1}{\mu} - 1\right) \left(\frac{\lambda}{5500}\right)^{-0.7}, \quad \text{(S48)}$$

where μ is given by a random Gaussian deviate with mean 0.3 and standard deviation 0.2, truncated at 0.1 and 1.

S19.3 Overall extinction curve

In order to get the final overall extinction, every galaxy is assigned an inclination given by the angle between the angular momentum of their dark matter halo and the z-direction of the simulation box, and a "slab" geometry is assumed for the dust in the diffuse ISM. For sources that are uniformly distributed within the disc then the mean absorption coefficient is

$$A_\lambda^{\rm ISM} = -2.5 \log_{10} \left(\frac{1 - \exp^{-\tau_\lambda^{\rm ISM} \sec\theta}}{\tau_\lambda^{\rm ISM} \sec\theta}\right), \quad \text{(S49)}$$

where θ is the angle of inclination of the galaxy relative to the line-of-sight. Emission from young stars embedded within birth clouds is subject to an additional extinction of

$$A_\lambda^{\rm BC} = -2.5 \log_{10} \left(\exp^{-\tau_\lambda^{\rm BC}}\right). \quad \text{(S50)}$$

S20 MODEL CALIBRATION: MONTE CARLO MARKOV CHAINS

Like all galaxy formation models and simulations, L-Galaxies has a number of parameters (e.g. the star formation efficiency) that need to be fit. 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-fitting model has been identified, it can be implemented on the full volumes of the two simulations.

We use six independent observational constraints: the stellar mass function and the fraction of quenched galaxies, each at

$z = 0, 1, 2$. Choosing proper observational constraints and weighting them to converge to an acceptable fit can be a tricky task. For instance, weighting all observational constraint equally, would result in a rather bad fit for the stellar mass function at $z = 0$, especially for galaxies with $10 < \log_{10}(M_\star/\mathrm{M}_\odot) < 11$. Ultimately, exploring different weightings for different datasets, we find it best to give the highest weight to the observational constraints at $z = 0$. Furthermore, at $z = 0$ itself, we give the stellar mass function a higher weight than the quenched fraction. To properly fit the $z = 0$ stellar mass function for $M^\star$ galaxies, we give an additional weight to the stellar mass function at $10 < \log_{10}(M_\star/\mathrm{M}_\odot) < 11$. At $z > 0$, observational constraints are weighted equally.

We run the MCMC for several tens of thousands of steps, i.e. we execute our model with different free parameters tens of thousands of times. During the calibration, we use L-Galaxies run on the Millennium simulation for galaxies with $\log_{10}(M_\star/\mathrm{M}_\odot) > 9$ and L-Galaxies run on Millennium-II for lower stellar masses. This stellar mass transition value is chosen following H20 and also by monitoring the approximate stellar mass where the two runs converge for a few smaller runs of the model (see also Guo et al. 2011; Henriques et al. 2020). Our best fit parameters in comparison with previous models are given in Table 3 and Fig. 2 of the main paper.

REFERENCES

Angulo R. E., Hilbert S., 2015, MNRAS, 448, 364
Angulo R. E., White S. D. M., 2010, MNRAS, 405, 143
Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, ARA&A, 47, 481
Avila-Reese V., Firmani C., Hernández X., 1998, ApJ, 505, 37
Ayromlou M., Nelson D., Yates R. M., Kauffmann G., White S. D. M., 2019, MNRAS, 487, 4313
Bacon R., Accardo M., Adjali L., et al., 2010, in Ground-based and Airborne Instrumentation for Astronomy III. p. 773508, doi:10.1117/12.856027
Bekki K., 2014, MNRAS, 438, 444
Benson A. J., Pearce F. R., Frenk C. S., Baugh C. M., Jenkins A., 2001, MNRAS, 320, 261
Bertone S., De Lucia G., Thomas P. A., 2007, MNRAS, 379, 1143
Bigiel F., Leroy A. K., Walter F., Brinks E., et al., 2011, ApJ, 730, L13
Binney J., Tremaine S., 1987, Galactic dynamics. Princeton, NJ, Princeton University Press, 1987, 747 p.
Birnboim Y., Dekel A., 2003, MNRAS, 345, 349
Bîrzan L., Rafferty D. A., McNamara B. R., Wise M. W., Nulsen P. E. J., 2004, ApJ, 607, 800
Blitz L., Rosolowsky E., 2006, ApJ, 650, 933
Bouwens R. J., et al., 2014, ApJ, 793, 115
Boylan-Kolchin M., Ma C.-P., Quataert E., 2008, MNRAS, 383, 93
Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., Lemson G., 2009, MNRAS, 398, 1150
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Bundy K., Bershady M. A., Law D. R., et al., 2015, ApJ, 798, 7
Chabrier G., 2003, PASP, 115, 763
Charlot S., Bruzual G., 2007, provided to the community but not published
Charlot S., Fall S. M., 2000, ApJ, 539, 718
Clay S. J., Thomas P. A., Wilkins S. M., Henriques B. M. B., 2015, MNRAS, 451, 2692
Colless M., Dalton G., Maddox S., et al., 2001, MNRAS, 328, 1039
Contini E., De Lucia G., Villalobos Á., Borgani S., 2014, MNRAS, 437, 3787
Croom S. M., Lawrence J. S., Bland-Hawthorn J., et al., 2012, MNRAS, 421, 872
Croton D. J., et al., 2006, MNRAS, 365, 11
Crowl H. H., Kenney J. D. P., van Gorkom J. H., Vollmer B., 2005, AJ, 130, 65
Dalcanton J. J., Spergel D. N., Summers F. J., 1997, ApJ, 482, 659
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
De Lucia G., Blaizot J., 2007, MNRAS, 375, 2
De Lucia G., Kauffmann G., Springel V., White S. D. M., Lanzoni B., Stoehr F., Tormen G., Yoshida N., 2004a, MNRAS, 348, 333
De Lucia G., Kauffmann G., White S. D. M., 2004b, MNRAS, 349, 1101
De Lucia G., Boylan-Kolchin M., Benson A. J., Fontanot F., Monaco P., 2010, MNRAS, 406, 1533
Dekel A., Silk J., 1986, ApJ, 303, 39
Devriendt J. E. G., Guiderdoni B., Sadat R., 1999, Astronomy and Astrophysics Supplement Series, 350, 381
Dutton A. A., 2009, MNRAS, 396, 121
Efstathiou G., 1992, MNRAS, 256, 43P
Elmegreen B. G., 1989, ApJ, 338, 178
Elmegreen B. G., 1993, ApJ, 411, 170
Fontanot F., Monaco P., 2010, MNRAS, 405, 705
Fontanot F., Somerville R. S., Silva L., Monaco P., Skibba R., 2009, MNRAS, 392, 553
Forcada-Miro M. I., White S. D. M., 1997, arXiv e-prints,
Fu J., Hou J. L., Yin J., Chang R. X., 2009, ApJ, 696, 668
Fu J., Guo Q., Kauffmann G., Krumholz M. R., 2010, MNRAS, 409, 515
Fu J., Kauffmann G., Li C., Guo Q., 2012, MNRAS, 424, 2701
Fu J., et al., 2013, MNRAS, 434, 1531
Fumagalli M., Fossati M., Hau G. K. T., Gavazzi G., Bower R., Sun M., Boselli A., 2014, ArXiv e-prints, 1407.7527,
Genel S., et al., 2014, preprint (arXiv:1405.3749),
Genzel R., Tacconi L. J., Lutz D., et al., 2015, ApJ, 800, 20
Gnedin N. Y., 2000, ApJ, 542, 535
Gonzalez-Perez V., Lacey C. G., Baugh C. M., Lagos C. D. P., Helly J., Campbell D. J. R., Mitchell P. D., 2014, MNRAS,
Guiderdoni B., Rocca-Volmerange B., 1987, Astronomy and Astrophysics Supplement Series, 186, 1
Gunn J. E., Gott J. Richard I., 1972, ApJ, 176, 1
Guo Q., White S. D. M., 2009, MNRAS, 396, 39
Guo Q., et al., 2011, MNRAS, 413, 101
Guo Q., White S., Angulo R. E., Henriques B., Lemson G., Boylan-Kolchin M., Thomas P., Short C., 2013, MNRAS, p. 135
Hastings W. K., 1970, Biometrika, 57, 97
Henriques B. M. B., Thomas P. A., 2010, MNRAS, 403, 768
Henriques B. M. B., Thomas P. A., Oliver S., Roseboom I., 2009, MNRAS, 396, 535
Henriques B., Maraston C., Monaco P., Fontanot F., Menci N., De Lucia G., Tonini C., 2011, MNRAS, 415, 3571
Henriques B. M. B., White S. D. M., Lemson G., Thomas P. A., Guo Q., Marleau G.-D., Overzier R. A., 2012, MNRAS, 421, 2904
Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R. E., Guo Q., Lemson G., Springel V., 2013, MNRAS, 431, 3373
Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R., Guo Q., Lemson G., Springel V., Overzier R., 2015, MNRAS, 451, 2663
Henriques B. M. B., Yates R. M., Fu J., Guo Q., Kauffmann G., Srisawat C., Thomas P. A., White S. D. M., 2020, MNRAS, 491, 5795
Irodotou D., Thomas P. A., Henriques B. M., Sargent M. T., 2018, arXiv e-prints,
Jaffe W., 1983, MNRAS, 202, 995
Kauffmann G., 1996, MNRAS, 281, 475
Kauffmann G., White S. D. M., Guiderdoni B., 1993, MNRAS, 264, 201
Kauffmann G., Colberg J., Diaferio A., White S., 1999, MNRAS, 303, 188
Kereš D., Katz N., Weinberg D. H., Davé R., 2005, MNRAS, 363, 2
Kitzbichler M. G., White S. D. M., 2007, MNRAS, 376, 2
Krumholz M. R., McKee C. F., Tumlinson J., 2009, ApJ, 693, 216
Kuhlen M., Krumholz M. R., Madau P., Smith B. D., Wise J., 2012, ApJ, 749, 36
Larson R. B., 1974, MNRAS, 169, 229
Leroy A. K., Walter F., Brinks E., Bigiel F., de Blok W. J. G., Madore B., Thornley M. D., 2008, AJ, 136, 2782
Leroy A. K., Walter F., Sandstrom K., et al., 2013, AJ, 146, 19
Lo K. Y., Sargent W. L. W., Young K., 1993, AJ, 106, 507
Lu Y., et al., 2014, ApJ, 795, 123
Lynden-Bell D., Pringle J. E., 1974, MNRAS, 168, 603

Maoz D., Mannucci F., Brandt T. D., 2012, MNRAS, 426, 3282
Maraston C., 2005, MNRAS, 362, 799
Marigo P., 2001, A&A, 370, 194
Martin C. L., Shapley A. E., Coil A. L., Kornei K. A., Bundy K., Weiner B. J., Noeske K. G., Schiminovich D., 2012, ApJ, 760, 127
Mathis J. S., Mezger P. G., Panagia N., 1983, Astronomy and Astrophysics Supplement Series, 128, 212
McCarthy I. G., Frenk C. S., Font A. S., Lacey C. G., Bower R. G., Mitchell N. L., Balogh M. L., Theuns T., 2008, MNRAS, 383, 593
McKee C. F., Krumholz M. R., 2010, ApJ, 709, 308
McNamara B. R., Nulsen P. E. J., 2007, ARA&A, 45, 117
Metropolis N., Rosenbluth A., Rosenbluth M., Teller A., Teller E., 1953, J. Chem. Phys., 21, 1087
Mihos J. C., Hernquist L., 1996, ApJ, 464, 641
Mo H. J., Mao S., White S. D. M., 1998, MNRAS, 295, 319
Nelson D., Vogelsberger M., Genel S., Sijacki D., Kereš D., Springel V., Hernquist L., 2013, MNRAS, 429, 3353
Obreschkow D., Rawlings S., 2009, MNRAS, 394, 1857
Okamoto T., Gao L., Theuns T., 2008, MNRAS, 390, 920
Oppenheimer B. D., Davé R., 2008, MNRAS, 387, 577
Oppenheimer B. D., Davé R., Kereš D., Fardal M., Katz N., Kollmeier J. A., Weinberg D. H., 2010, MNRAS, 406, 2325
Pilkington K., Few C. G., Gibson B. K., et al., 2012, Astronomy and Astrophysics Supplement Series, 540, A56
Portinari L., Chiosi C., Bressan A., 1998, A&A, 334, 505
Quadri R. F., Williams R. J., Lee K.-S., Franx M., van Dokkum P., Brammer G. B., 2008, ApJ, 685, L1
Rubin K. H. R., Prochaska J. X., Koo D. C., Phillips A. C., Martin C. L., Winstrom L. O., 2013, preprint (ArXiv:1307.1476),
Rupke D. S., Veilleux S., Sanders D. B., 2005, ApJ Supp., 160, 115
Sánchez S. F., Kennicutt R. C., Gil de Paz A., et al., 2012, A&A, 538, A8
Schruba A., et al., 2011, AJ, 142, 37
Scoville N., Lee N., Vanden Bout P., et al., 2017, ApJ, 837, 150
Shamshiri S., Thomas P. A., Henriques B. M., Tojeiro R., Lemson G., Oliver S. J., Wilkins S., 2015, MNRAS, 451, 2681
Shapley A. E., Steidel C. C., Pettini M., Adelberger K. L., 2003, ApJ, 588, 65
Somerville R. S., Primack J. R., Faber S. M., 2001, MNRAS, 320, 504
Spergel D. N., Verde L., Peiris H. V., et al., 2003, ApJ Supp., 148, 175
Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726
Springel V., White S. D. M., Jenkins A., et al., 2005, Nat., 435, 629
Steidel C. C., Shapley A. E., Pettini M., Adelberger K. L., Erb D. K., Reddy N. A., Hunt M. P., 2004, ApJ, 604, 534
Stil J. M., Israel F. P., 2002, A&A, 389, 29
Sutherland R. S., Dopita M. A., 1993, ApJ Supp., 88, 253
Thielemann F.-K., et al., 2003, in Hillebrandt W., Leibundgut B., eds, From Twilight to Highlight: The Physics of Supernovae. p. 331, doi:10.1007/10828549_46
Tonini C., Maraston C., Devriendt J., Thomas D., Silk J., 2009, MNRAS, 396, L36
Tonini C., Maraston C., Thomas D., Devriendt J., Silk J., 2010, MNRAS, 403, 1749
Tonini C., Maraston C., Ziegler B., Böhm A., Thomas D., Devriendt J., Silk J., 2011, MNRAS, pp 714–+
Vogelsberger M., et al., 2014, Nat., 509, 177
Wang W., Lu J., Luo Z., Yang Z., Hua H., Chen Z., eds, 2011, Galaxy Evolution: Infrared to Millimeter Wavelength Perspective Astronomical Society of the Pacific Conference Series Vol. 446
Weiner B. J., et al., 2009, ApJ, 692, 187
Weinmann S. M., Pasquali A., Oppenheimer B. D., Finlator K., Mendel J. T., Crain R. A., Macciò A. V., 2012, MNRAS, 426, 2797
White S. D. M., 1989, in Frenk C. S., sEllis R. S., Shanks T., Heavens A. R., Peacock J. A., eds, NATO ASIC Proc. 264: The Epoch of Galaxy Formation. p. 15
White S. D. M., Frenk C. S., 1991, ApJ, 379, 52
White S. D. M., Rees M. J., 1978, MNRAS, 183, 341
Yates R. M., Henriques B., Thomas P. A., Kauffmann G., Johansson J., White S. D. M., 2013, MNRAS, 435, 3500
Yates R. M., Thomas P. A., Henriques B. M. B., 2017, MNRAS, 464, 3169
Yoshida N., Stoehr F., Springel V., White S. D. M., 2002, MNRAS, 335, 762