

Formation of disc galaxies in preheated media: a preventative feedback model

Yu Lu,^{1★} H. J. Mo² and Risa H. Wechsler¹

¹*Kavli Institute for Particle Astrophysics and Cosmology, Physics Department, and SLAC National Accelerator Laboratory, Stanford University, Stanford, CA 94305, USA*

²*Department of Astronomy, University of Massachusetts, Amherst, MA 01003-9305, USA*

Accepted 2014 October 15. Received 2014 October 7; in original form 2014 February 9

ABSTRACT

We develop a semi-analytic galaxy formation model that implements a self-consistent treatment for the hot halo gas and the assembly of central discs. We use the model to explore a preventative feedback scenario, in which the circumhalo medium is assumed to be preheated up to a certain entropy level by early starbursts or other processes, and contrast it with an ejective feedback scenario in which baryons are first accreted into dark matter haloes and subsequently ejected by feedback. When the preheating entropy is comparable to the halo virial entropy, baryon accretion can be reduced and delayed. The accreted medium establishes an extended gaseous halo, and the cooling gas that forms the central disc has a higher specific angular momentum than the halo material. The preventative feedback model can reproduce remarkably well a number of observational scaling relations, including the cold baryon mass fraction–halo mass relations, star formation histories, disc size–stellar mass relation and its evolution, and the evolution of the number density of low-mass galaxies. In contrast, the ejective feedback model fails to reproduce these observational trends. These results show that the properties of disc galaxies are closely tied to the thermal state of hot halo gas and perhaps also the circumhalo medium, which suggests that observations of the disc properties and circumgalactic hot/warm medium may jointly provide important constraints on galaxy formation.

Key words: galaxies: evolution – galaxies: formation – galaxies: stellar content – galaxies: structure.

1 INTRODUCTION

Galaxy formation remains one of the most challenging problems in astrophysics largely because galaxy formation involves many complicated baryonic processes, which are still poorly understood (e.g. Fall 2002; Mo, van den Bosch & White 2010; Silk & Mamon 2012). The early theories, pioneered by Binney (1977), Rees & Ostriker (1977), Silk (1977), and White & Rees (1978) used a cooling rate argument to place an upper limit on the masses of galaxies. Soon after that, people realized that some form of feedback has to be involved in galaxy formation to reduce the star formation efficiency, especially in low-mass systems (e.g. Dekel & Silk 1986). In the conventional picture of galaxy formation, all baryons are first accreted into dark matter haloes and a substantial fraction is subsequently heated and ejected out of the halo by feedback processes. Various sophisticated models based on this general assumption have been built and explored extensively (Bower et al. 2006; Croton et al. 2006; Somerville et al. 2008a; Lu et al. 2011). However, no galaxy

formation model has been able to reproduce the full range of the most important observational constraints with satisfaction. To date, models either poorly reproduce observational data or are based on physically implausible assumptions (see e.g. Mutch, Poole & Croton 2013; Lu et al. 2014b). It is important, at this point, to explore models beyond the standard assumption of full baryon accretion and ejective feedback. In this paper, we propose a new model for a preventative feedback process based on an assumption that the circumgalactic medium is preheated to a certain level of entropy by early feedback processes. We use a new semi-analytic model (SAM) to explore the basic consequences of the assumption of preheating on galaxy formation and contrast the model with the conventional assumption of full baryon accretion and ejective feedback in low-mass haloes.

A great wealth of observational data have become available to test models of galaxy formation. With multiwavelength imaging and redshift surveys, such as the SDSS (York et al. 2000) and GAMA (Driver et al. 2011) surveys, the luminosity/stellar mass functions of the local galaxy population can now be measured to high accuracy (e.g. Blanton et al. 2005; Moustakas et al. 2013). Surveys in radio bands, such as HIPASS (Meyer et al. 2004; Zwaan et al. 2004) and

★ E-mail: luyu@stanford.edu

ALFALFA (Giovannelli et al. 2005a,b), have also made it possible to estimate the cold gas mass function of local galaxies, providing a complete census of the cold baryonic contents of present-day galaxies. Combined with the halo mass function predicted by the current Λ cold dark matter (CDM) model, these data can be used to statistically establish connections between galaxies and their host dark matter haloes (Yang, Mo & van den Bosch 2003; Conroy, Wechsler & Kravtsov 2006; Behroozi, Conroy & Wechsler 2010; Yang et al. 2012, 2013; Behroozi, Wechsler & Conroy 2013c; Moster, Naab & White 2013; Reddick et al. 2013; Lu et al. 2014a). This type of study has produced important results for the luminosity/stellar mass–halo mass relation, the cold gas mass–halo mass relation, and the star formation and stellar mass assembly histories (MAHs) in haloes of different masses. An important characteristic trend is that the fraction of cold baryons (stars and cold gas) decreases rapidly with decreasing halo mass for haloes with virial masses $M_{\text{vir}} < 10^{12} M_{\odot}$ (Yang et al. 2003; van den Bosch, Abel & Hernquist 2003; Yang, Mo & van den Bosch 2008; Behroozi et al. 2010, 2013c; Papastergis et al. 2012), which provides stringent constraints on the star formation and assembly processes modelled in SAMs (Bower et al. 2006; Croton et al. 2006; Somerville et al. 2008a; Lu et al. 2012, 2014b) and in hydrodynamical simulations (e.g. Davé et al. 2013; Hopkins et al. 2014).

Another important observational constraint, whose constraining power has not yet been widely exploited, comes from the sizes of galaxies. With the advent of deep surveys from both ground-based and space telescopes, the galaxy luminosity/stellar mass functions have been estimated to $z \sim 8$ (e.g. Bradley et al. 2012; Finkelstein et al. 2012; Oesch et al. 2012; Yan et al. 2012; McLure et al. 2013), allowing us to study the star formation and assembly histories of galaxies over the cosmic time. In addition, high-quality imaging data from the *Hubble Space Telescope* also permits studies of the evolution of galaxy morphology out to $z \sim 2$ (e.g. Trujillo et al. 2006; Cassata et al. 2013; Patel et al. 2013a,b; van der Wel et al. 2014).

Based on the conventional assumption that all baryons are accreted into dark matter haloes, models predict that the cooling efficiency in low-mass haloes is always much higher than the star formation efficiency in observed galaxies (Thoul & Weinberg 1996; Silk 1997; Kennicutt 1998). Assuming strong feedback, a number of models have successfully reproduced the galaxy mass function or luminosity function observed in the local universe (Bower et al. 2006; Croton et al. 2006; Somerville et al. 2008a; Lu et al. 2014). However, simultaneously reproducing the observed number density of galaxies at different redshifts, the baryon mass–halo mass relation, and the size–mass relation remains an extremely challenging problem in galaxy formation (see e.g. discussion in Weinmann et al. 2012; Lu et al. 2014b). The formation histories of galaxies, especially those with masses equal to or lower than that of the Milky Way, are also hard to reproduce. Observed low-mass field galaxies are typically star forming, maintaining a specific star formation rate (sSFR) $\approx 0.5 \text{ Gyr}^{-1}$ all the way to the present day (Moustakas et al. 2013), while star formation feedback in theoretical models is typically more effective in lower mass haloes, resulting in the predicted SFRs in faint galaxies that are too low (Wang, Weinmann & Neistein 2012; Weinmann et al. 2012; Lu et al. 2014b). In addition, low-mass haloes are predicted to form stars very efficiently at high redshift, producing too many low-mass galaxies at $z > 2$ to match observations (Weinmann et al. 2012; Lu et al. 2014b).

The extended size of observed disc galaxies has been another long-standing challenge for modelling galaxy formation. Early hydrodynamical simulations showed that the cooling gas forming the

central galaxy in a halo has too low angular momentum to produce an extended disc (e.g. Katz & Gunn 1991; Navarro & Benz 1991; Steinmetz & Navarro 1999). This angular momentum ‘catastrophe’ has been attributed partially to numerical limitations, and partially to uncertainties in modelling baryonic processes such as feedback (Fall 2002; Maller & Dekel 2002). Analytic and semi-analytic models normally assume that the baryons are initially mixed with the dark matter and share the same specific angular momentum, j . Even when the baryons cool and decouple from the dark matter to collapse on a disc, the material assembling the disc still retains the same j as the dark matter (e.g. Fall & Efstathiou 1980; Dalcanton, Spergel & Summers 1997; Mo, Mao & White 1998; Somerville et al. 2008b). Assuming the baryonic j is conserved, such models can reproduce both the zero-point and the slope of the observed spiral-galaxy $j^* - M^*$ relation (Dutton & van den Bosch 2012; Romanowsky & Fall 2012). However, more detailed analyses have demonstrated that this assumption does not solve the ‘angular momentum catastrophe’, because the material that forms the central disc tends to cool from the high-density central region of the halo, where the specific angular momentum is low, and, hence, produces discs that are too compact (Dutton & van den Bosch 2012). Indeed, using analytical models and hydrodynamical simulations, Dutton (2009) and Brook et al. (2011) showed that gas with low specific angular momentum has to be removed by outflows in order to reproduce the exponential surface density profiles of low-mass discs with the correct characteristic sizes.

All of these issues could be centred on our understanding of feedback processes, which is one of the biggest uncertainties in modelling galaxy formation. Most models investigated so far assume that feedback is ejective, in the sense that supernova (SN) explosions associated with stellar evolution ejects cold gas from the disc, thereby reducing subsequent star formation. However, it is unclear how effective such feedback is in reality. Detailed simulations capable of resolving the interface of the SN-driven super-bubbles and the interstellar medium (ISM) have shown that the feedback is inefficient in driving gas out of a galaxy because of the rapid development of Rayleigh–Taylor instabilities (Mac Low & Ferrara 1999; Krumholz & Thompson 2012). Yet, a high fraction of SN energy is required to be coupled with the ISM in order to explain the faint end of galaxy luminosity function, the low-mass end of the $H\text{ I}$ mass function (Lu et al. 2014b), and the evolution of the stellar mass function (Mutch et al. 2013). Moreover, Henriques et al. (2013) found that, even if SN is effective in ejecting gas from galaxies, the ejected gas is required to follow a particular schedule to reincorporate into the halo later in order to simultaneously match the observed galaxy luminosity functions at multiple redshifts. Recent hydrodynamical simulations suggest that additional energy from radiation pressure associated with massive stars may be able to provide sufficient amounts of energy to reproduce the observed low baryon mass fractions in low-mass haloes as well as a number of other important scaling relations (Brook et al. 2012; Davé et al. 2013; Stinson et al. 2013; Hopkins et al. 2014; Kannan et al. 2014). However, how the feedback governs the budget of the baryonic matter in these simulations is still unclear. Observationally, star-forming galaxies at $z \lesssim 2$ only show evidence of outflows with a mass-loading factor ~ 2 (Bouché et al. 2012; Newman et al. 2012), which does not seem to be sufficient to explain the low baryon mass fraction in low-mass haloes (Papastergis et al. 2012; Lu et al. 2014b).

In this paper, we propose a new galaxy formation model that implements a preventative scenario of feedback. We assume that the intergalactic gas is heated to some finite entropy before it is accreted into dark matter haloes, motivated by processes that have

been suggested in the literature, such as preheating by SN/active galactic nucleus (AGN) winds (Mo & Mao 2002, 2004), by gravitational pancaking (Mo et al. 2005; Lu & Mo 2007), by blazar heating (Pfrommer, Chang & Broderick 2012), and by intergalactic turbulence (Zhu, Feng & Fang 2011). The enhanced entropy affects galaxy formation in two ways. First, the baryon fraction that can collapse into low-mass haloes is strongly reduced (Mo & Mao 2002; Oh & Benson 2003; van den Bosch et al. 2003; McCarthy et al. 2004; Scannapieco & Oh 2004; Lu & Mo 2007). Secondly, when the entropy of the pre-collapsed gas is higher than would be generated by accretion shocks, the halo gas is expected to develop an extended density distribution (Mo & Miralda-Escude 1996; Mo & Mao 2002; Maller & Bullock 2004; Kaufmann et al. 2009; Fang, Bullock & Boylan-Kolchin 2013a), thereby affecting where and when halo gas cools to fuel the central disc. Using a SAM built upon simulated halo accretion histories, we examine in detail the disc size evolution, mass assembly, and star formation histories of central galaxies hosted by haloes with masses $M_{\text{vir}} < 10^{12} M_{\odot}$ at the present time in the preheating scenario. We also compare the results with those predicted by an ejective feedback model and with observational data.

The paper is organized as follows. In Section 2, we introduce the physics of how a preheated circumhalo medium affects galaxy formation and describe the implementation of the model. We show the predictions of a typical ejective feedback model and our preventative feedback model, and compare the model predictions with observations in Section 3. In Section 4, we summarize our results and discuss their implications. We also describe the detailed implementations of model recipes for reionization, star formation in galaxy discs, and stellar mass loss due to stellar evolution in the appendices. Throughout the paper, we use a Λ CDM cosmology with $\Omega_{\text{M},0} = 0.27$, $\Omega_{\Lambda,0} = 0.73$, $\Omega_{\text{B},0} = 0.044$, $h = 0.70$, $n = 0.95$, and $\sigma_8 = 0.82$.

2 THE MODEL

To study the impact of preheated circumhalo gas on the formation of disc galaxies, we develop a new SAM, which follows realistic halo MAHs and includes the most important physical processes that are generally implemented in galaxy formation models. Our model consists of the following parts: 1) halo mass accretion histories and density profiles; 2) gas accretion and distribution in dark matter

haloes; 3) radiative cooling of halo gas and formation of galaxy discs; 4) star formation and SN feedback in galaxy discs. The model follows these processes and makes predictions for the evolution of central galaxies from an early time to the present day. More importantly, the prescriptions include the key physics that allows us to model galaxy formation based on different assumptions for the entropy level of the circumhalo gas. The reason for us to develop this completely new model, instead of using our existing model (Lu et al. 2014, 2014b) is because the present study focuses on experimenting with new physics rather than exploring the parameter space, which is the main task for our previously published models. In the following, we describe the model and highlight the model prescriptions that are newly implemented for this paper.

2.1 Formation history and structure of dark matter haloes

Our model follows the main-branch MAH of dark matter haloes. Any secondary progenitors and satellite galaxies associated with subhaloes are ignored. For haloes hosting a galaxy with a stellar mass comparable to or lower than that of the Milky Way, this is a good approximation because the stellar mass in these haloes is assembled mainly through *in situ* star formation in the main-branch progenitor rather than through mergers (Yang et al. 2012, 2013; Behroozi et al. 2013c; Lu et al. 2014a). We adopt realistic MAHs extracted directly from an N -body cosmological simulation, the Bolshoi simulation (Klypin, Trujillo-Gomez & Primack 2011). Dark matter haloes are identified at 180 time steps from $z = 14$ to 0 using the ROCKSTAR halo finder (Behroozi, Wechsler & Wu 2013a), and halo merger trees are constructed by linking a halo with all its progenitors using the Consistent Tree method (Behroozi et al. 2013b). The halo MAHs provide enough mass and time resolution to study the formation of central disc galaxies in the halo mass range of 10^{10} – $10^{12} M_{\odot}$. For this paper, we randomly select a large number of haloes with masses and redshifts that are relevant to our study. Fig. 1 shows two random subsets of the MAHs for haloes with a present-day virial mass in the range of $10^{11.0}$ – $10^{11.2}$ or $10^{12.0}$ – $10^{12.2} M_{\odot}$. As one can see, the simulated MAHs for a given final halo mass have quite large variations. For comparison, we also show a smoothed, average accretion history based on the fitting formula proposed by McBride, Fakhouri & Ma (2009),

$$M_{\text{vir}}(z) = M_{\text{vir},0} (1+z)^{\alpha} \exp(-\eta z), \quad (1)$$

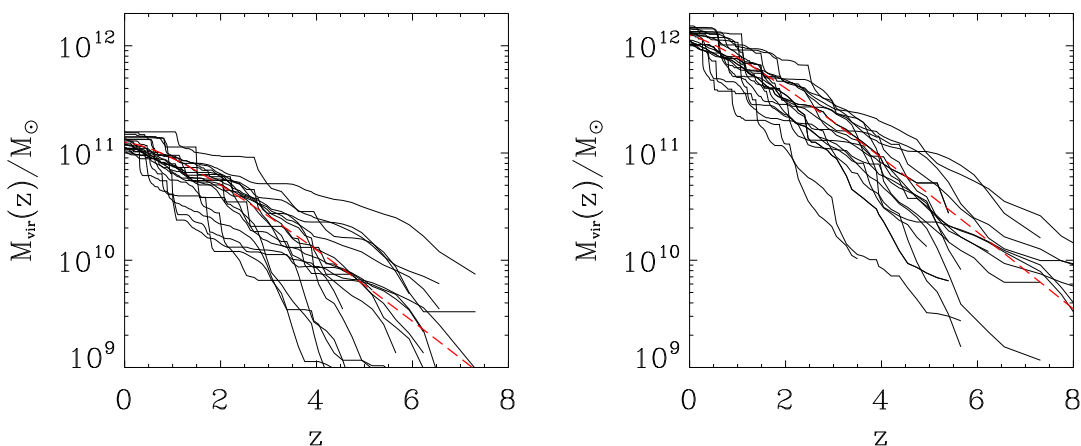


Figure 1. MAHs of dark matter haloes with a final ($z = 0$) virial mass in the range of 10^{11} – $10^{11.2} M_{\odot}$ (left) and in the range of 10^{12} – $10^{12.2} M_{\odot}$ (right). The black lines show 20 MAHs randomly selected from the Bolshoi simulation. The red dashed lines are a smoothed version of the MAHs using the McBride model (equation 1) with $\alpha = 0.8$, $\eta = 0.9$ for $M_{\text{vir},0} = 10^{11.1} M_{\odot}$ haloes, and $\alpha = 0.6$, $\eta = 0.9$ for $M_{\text{vir},0} = 10^{12.1} M_{\odot}$ haloes.

where the normalization $M_{\text{vir},0}$ is the halo mass at $z = 0$ and α and η are parameters determining the shape of the MAH. The values of the two parameters adopted in the plot are $\alpha = 0.6$ and $\eta = 0.9$ for $M_{\text{vir},0} \approx 10^{12.1} M_{\odot}$ haloes, and $\alpha = 0.8$ and $\eta = 0.9$ for $M_{\text{vir},0} \approx 10^{11.1} M_{\odot}$ haloes.

At any redshift, the density distribution within a dark matter halo is assumed to be spherically symmetric and to follow an NFW profile (Navarro, Frenk & White 1996):

$$\rho(r) = \frac{4\rho_s}{(r/r_s)(1+r/r_s)^2}, \quad (2)$$

where ρ_s and r_s are the characteristic density and radius of the halo. The halo concentration is defined as the ratio of the virial radius to the characteristic radius as $c = r_{\text{vir}}/r_s$. Cosmological simulations show that the halo concentration depends on both halo mass and redshift at which the halo is identified (e.g. Bullock et al. 2001; Eke, Navarro & Steinmetz 2001; Zhao et al. 2003, 2009; Macciò et al. 2007). Here, we adopt the recent simulation result of Prada et al. (2012) to compute c for a halo with a given mass at a given redshift. In this model, the concentration of a halo increases with time. For a present-day $M_{\text{vir}} = 10^{12} M_{\odot}$ halo, the typical concentration is $c \approx 10$, and $c \approx 7$ at $z = 2$.

2.2 Baryon accretion of dark matter haloes

As a dark matter halo grows, a certain amount of baryonic mass is expected to follow dark matter to collapse into the halo. One of the most basic differences between our preventative model and normal ejective models is that in our preventative model, only a reduced fraction of baryonic matter can collapse into a halo, while a nearly cosmic baryon fraction of baryonic matter can collapse in the ejective feedback model. In general, we write the baryon accretion rate in terms of the halo mass accretion rate as

$$\dot{M}_{\text{acc}} = f_b \dot{M}_{\text{vir}}, \quad (3)$$

where the coefficient $f_b = f_{\text{acc}} f_{b,0}$, with $f_{b,0} \approx 0.17$ the cosmic baryon mass fraction and f_{acc} a parameter to be determined by relevant baryonic processes. Following other SAMs, we include the effect of reionization on the baryon accretion fraction by adopting a model proposed by Gnedin (2000) and Kravtsov, Gnedin & Klypin (2004, see Appendix A for details). For haloes relevant to our study, this effect is small.

In our preventative model, we explore the impact on galaxy formation if the circumhalo media is preheated to an entropy level higher than that due to the photoionization heating alone. In this model, we assume that the baryons are preheated to have an entropy S . Lu & Mo (2007) simulated the accretion process of the preheated gas on to growing dark matter haloes and found that the baryon mass fraction in a halo scales with the preheating entropy S and the virial entropy of halo, S_{vir} , as

$$f_b = \frac{1}{\left[1 + \left(\frac{S/S_{\text{vir}}}{0.8}\right)^3\right]^{1/2}}, \quad (4)$$

where the virial entropy is defined as

$$S_{\text{vir}} = \frac{T_{\text{vir}}}{n_{\text{vir}}^{2/3}}, \quad (5)$$

with T_{vir} the virial temperature of the halo, and n_{vir} the mean gas particle number density of a virialized halo assuming the cosmic baryon fraction $f_{b,0}$. Thus, if the preheating entropy is much lower than the virial entropy of the halo, the baryonic matter follows the

dark matter to collapse into the halo, resulting in $f_b \approx f_{b,0}$. If, on the other hand, the preheating entropy is much higher than the halo virial entropy, then the baryon mass fraction $f_b \propto (S_{\text{vir}}/S)^{3/2}$. For a constant S , this gives $f_b \propto S_{\text{vir}}^{3/2} \propto M_{\text{vir}}$ at a given redshift.

The model predicts the initial fraction of baryons that collapse into a halo using equation (4). We trace the evolution of the hot halo gas using

$$M_{\text{hot}} = f_b M_{\text{vir}} - (M_* + M_{\text{cold}} + M_{\text{eject}}), \quad (6)$$

where M_* , M_{cold} , and M_{eject} are the baryon masses in stars, cold gas, and ejected material that are associated with the halo, respectively.

2.3 Configuration of the hot halo gas

We model the structure of the gaseous halo under the assumption that the hot gas accreted into a halo is in hydrostatic equilibrium with the gravitational potential of the host halo. The hot gas distribution is then solved with an assumed equation of state for an ideal gas:

$$p = A \rho^\gamma, \quad (7)$$

where A is called the *adiabat*, which is a constant if the gas experiences an adiabatic process, and $\gamma = 5/3$ for a monoatomic gas. The adiabat of an ideal gas can be written as

$$A = \frac{kT}{\mu m_p \rho^{\gamma-1}}, \quad (8)$$

where k is the Boltzmann constant, μ is the mean molecular weight, m_p is the mass of proton, and T is the temperature of the gas. The adiabat is related to the specific entropy of the gas,

$$A = \frac{k}{(\mu m_p)^\gamma} S, \quad (9)$$

where the specific entropy, S , is defined as

$$S = \frac{T}{n^{2/3}}, \quad (10)$$

with n being the number density of gas particles: $n = \rho/(\mu m_p)$.

Analytical models and hydrodynamic simulations of halo formation neglecting radiative cooling have shown that the radial profile of the specific entropy of the hot gas generated by accretion shocks roughly follows a power-law function of radius with a power index $\beta \sim 1.1$ (e.g. Tozzi & Norman 2001; Voit, Kay & Bryan 2005; Lu & Mo 2007). When radiative cooling is included, the entropy profile can be modified. Using a simplified model, Bryan (2000) argued that because rapid cooling always happens in low-entropy gas, and the high-entropy gas will expand adiabatically to occupy the volume of the halo, resulting in a flatter entropy profile. However, Tang et al. (2009) found that cooling in Milky Way-sized haloes is a runaway process in their simulations, and the resulting entropy profile remains a power law. When feedback is taken into account, the entropy profile can be modified further (Voit & Donahue 2005; McCarthy et al. 2010). Hydrodynamical simulations show that, when feedback is included, the entropy profile of the halo gas becomes shallower than that predicted in adiabatic simulations (e.g. Crain et al. 2010). While observing the hot gas in low-mass haloes is difficult, the entropy of hot haloes of giant ellipticals shows a similar power-law profile with flattening in the centre due to heating of feedback from central galaxies (Werner, Allen & Simionescu 2012). In this paper, we adopt a generic model where the entropy profile of the halo gas is assumed to be a power law,

$$S(r) = S_0 \left(\frac{r}{r_{\text{vir}}}\right)^\beta, \quad (11)$$

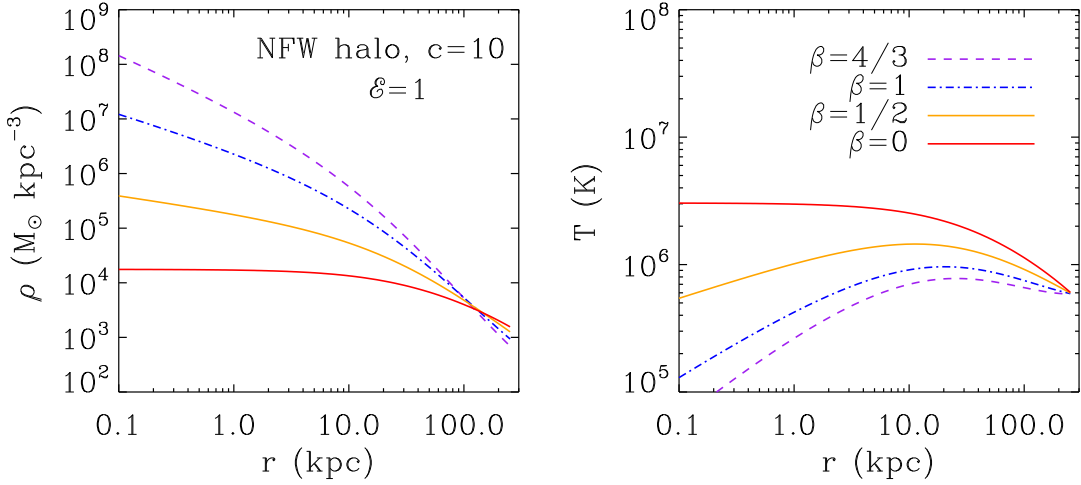


Figure 2. Density profiles (left) and temperature profiles (right) of the hot halo gas in hydrostatic equilibrium in a $10^{12} M_{\odot}$ NFW halo with $c = 10$. The lines with different colours denote models with different logarithmic slopes for the entropy radial profile. In all the models, the total gas mass is fixed to be $0.17 M_{\text{vir}}$ and the gas entropy at the virial radius is normalized so that the gas temperature at the virial radius equals to the virial temperature of the halo (i.e. $\mathcal{E} = 1$). Note that the preheating model we explore in this paper corresponds to the $\beta = 0$ case.

or

$$A(r) = A_0 \left(\frac{r}{r_{\text{vir}}} \right)^{\beta}, \quad (12)$$

where S_0 and A_0 are the gas entropy and adiabat at the halo virial radius r_{vir} , respectively, and β determines the slope of the profile. Thus, setting up $\beta = 0$ gives rise to an isentropic gas profile (e.g. Mo & Miralda-Escude 1996; Mo & Mao 2002; Maller & Bullock 2004).

To obtain the gas density profile, we solve the following hydrostatic equilibrium equation,

$$\frac{1}{\rho} \frac{dp}{dr} = -G \frac{M(<r)}{r^2}, \quad (13)$$

where $M(<r)$ is the gravitational mass within a radius r . Assuming the gravitational mass to be dominated by the dark matter, we obtain $M(<r)$ from the halo mass profile. Substituting equations (7) and (12) into equation (13), we have

$$\frac{d\rho^{\gamma-1}}{dr} + \frac{\gamma-1}{\gamma} \left[\beta \rho^{\gamma-1} \frac{1}{r} + G \frac{M(<r)}{r^2} \frac{1}{A_0} \left(\frac{r}{r_{\text{vir}}} \right)^{-\beta} \right] = 0. \quad (14)$$

For convenience, we change the variables to $x = r/r_{\text{vir}}$, $y = \frac{\mu m_p A_0}{k T_{\text{vir}}} \rho(r)^{\gamma-1}$, where the virial temperature of the halo $T_{\text{vir}} \equiv \mu m_p G M_{\text{vir}} / 2k$. The equation to be solved is then

$$\frac{dy}{dx} + \frac{\gamma-1}{\gamma} \left[\beta \frac{y}{x} + 2 \frac{m(x)}{x^{\beta+2}} \right] = 0, \quad (15)$$

where $m(x)$ is the halo mass profile normalized by the virial mass. For a NFW profile given by equation (2), the normalized mass profile is

$$m_{\text{NFW}}(x) = \frac{\ln(1+xc) - \frac{xc}{1+xc}}{\ln(1+c) - \frac{c}{1+c}}. \quad (16)$$

We numerically solve the one-dimensional differential equation for the gas density profile using the Runge–Kutta Dormand–Prince method (Dormand & Prince 1980). The boundary condition is chosen so that the total gas mass enclosed by the virial radius equals the total hot gas of the halo at the time in question. Once the density

profile is obtained, the corresponding temperature profile can be calculated based on the assumed entropy profile as

$$T(r) = A_0 \left(\frac{r}{r_{\text{vir}}} \right)^{\beta} \frac{\mu m_p \rho(r)^{\gamma-1}}{k}. \quad (17)$$

In some special cases, the density profile can be solved analytically. For example, the density profile of an isentropic gas in an NFW halo is

$$\rho = \rho_{\text{vir}} \left\{ 1 + \frac{4}{5} \mathcal{E} \frac{c}{\ln(1+c) - \frac{c}{1+c}} \left[\frac{\ln(1+cx)}{cx} - \frac{\ln(1+c)}{c} \right] \right\}^{3/2}, \quad (18)$$

where \mathcal{E} is defined as the ratio of the halo virial temperature and the gas temperature at the virial radius, $\mathcal{E} \equiv T_{\text{vir}}/T(r=r_{\text{vir}})$. This solution is similar to that of Maller & Bullock (2004) but generalized for an arbitrary gas entropy value. Similarly, if a dark matter halo has a singular isothermal density profile, $\rho_{\text{DM}} \propto r^{-2}$, the hydrostatic equilibrium gas density profile can be written as $\rho \propto r^{-(3/2)\beta}$ for $\beta \neq 0$. For the special case where $S \propto r^{4/3}$, the gas density profile is then parallel to the dark matter profile as $\rho \propto r^{-2}$ and the temperature profile is constant. If $\beta = 0$, the gas density profile is

$$\rho(r) = \rho_{\text{vir}} \left[1 - \frac{4}{5} \mathcal{E} \ln \frac{r}{r_{\text{vir}}} \right]^{3/2}, \quad (19)$$

which is equivalent to the solution obtained by Mo & Miralda-Escude (1996).

Fig. 2 shows the density and temperature profiles for the hot gas in hydrostatic equilibrium in an NFW halo of $M_{\text{vir}} = 10^{12} M_{\odot}$ predicted by assuming different entropy profiles, $\beta = 4/3, 1, 1/2$, and 0, as indicated in the figure. For all the models, the normalization of the entropy profile is chosen so that $\mathcal{E} = 1$, i.e. the temperature at the virial radius equals the halo virial temperature. The figure shows that hot halo gas with a shallower slope for the entropy profile has more extended gas distribution.

In the preventative model that we consider in this paper, when the baryonic matter is preheated to an entropy level equal to or higher than the virial entropy of the halo, the gas accretes into the halo adiabatically. Therefore, the hot halo gas will conserve its entropy and will have a flat entropy profile, corresponding to an isentropic

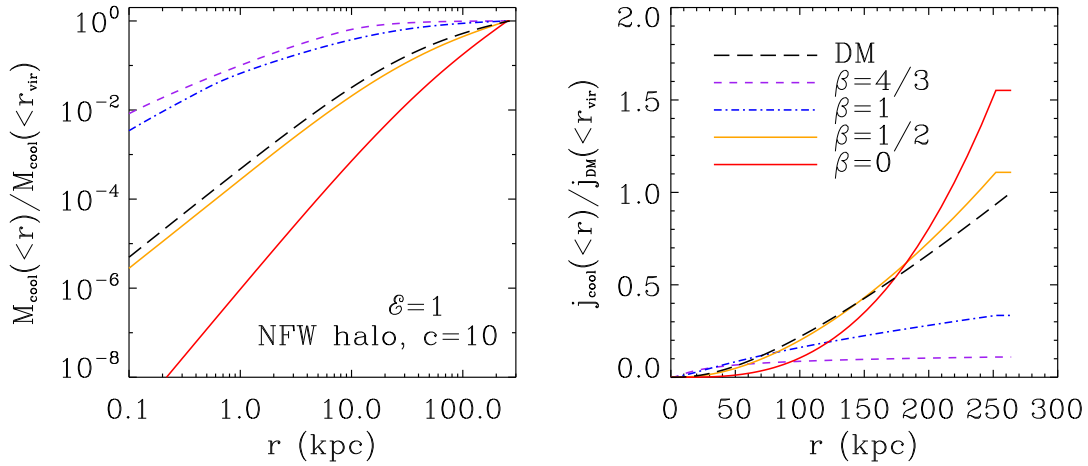


Figure 3. Left: the cooling gas mass profile predicted by models with different entropy radial profiles. When β increases, the entropy profile has a steeper entropy profile. $\beta = 0$ corresponds to an isentropic configuration, which is assumed in the preheating model we explore in the present paper. The black dashed line shows the dark matter mass profile of the **NFW** halo with $c = 10$. Right: the specific angular momentum profile of the cooled baryonic matter. The black dashed line shows the specific angular momentum profile of dark matter as a function of radius. Flatter entropy profiles have more extended density distribution and result in more cooling from large radii and higher specific angular momentum.

gas distribution ($\beta = 0$). In contrast, if the baryonic matter is cold and has an entropy much lower than the virial entropy of the halo, the collapse of the gas is expected to induce a virial shock, which produces a steep entropy profile. We will explore the impact of the different entropy profiles as a consequence of different assumptions for the entropy of the circumhalo matter on the growth of central discs.

2.4 Radiative cooling of hot halo gas

In many SAMs of galaxy formation, cooling time is compared with a pre-defined time-scale (e.g. the dynamical time of the system) to determine whether the gas can cool: no gas can cool if the cooling time is longer than the time-scale chosen. In reality, however, there must be fluctuations in the density and temperature of the hot gas so that part of the hot gas can still cool, developing a multiphase medium, even if the overall cooling time is long. Moreover, when the halo potential has been established in the fast-accretion phase, the inner hot gaseous halo can cool continuously without depending on the newly accreted gas on the halo outskirts. To catch the realistic cooling process, we consider a modified model in which gas is allowed to cool slowly even if the cooling time-scale is long.

In practice, we first choose the values of A_0 and β to set up the gas distribution, with the gas density profile normalized so that the total hot gas mass matches M_{hot} predicted by the model. We follow the gas density profile with 200 shells equally spaced in $\log r$ from $r = 0.03$ to 300 kpc, with the gas density and temperature in each shell given by the solution of the hydrostatic equilibrium equation. We then compute the cooling time-scale for each shell, using

$$\tau_{\text{cool}}(r) = \frac{3}{2} \frac{\mu m_p k T}{\rho \Lambda(T, Z)}, \quad (20)$$

where $\Lambda(T, Z)$ is the cooling function adopted from Sutherland & Dopita (1993). If the cooling time-scale is shorter than the free-fall time of the shell, $\tau_{\text{ff}} = r_i/V_c$, where r_i is the radius of the shell and V_c is the circular velocity of the halo, the gas mass in the shell is assumed to cool and be accreted on to the central galaxy in a free-fall time. If, on the other hand, the cooling time-scale is longer than the free-fall time, the gas in the shell is assumed to cool and be accreted over the cooling time-scale. In both cases, the accretion

rate of the cooling gas on to the central disc from a mass shell, i , is given by

$$\dot{m}_{i,\text{cool}} = \frac{m_{i,\text{hot}}}{\max(\tau_{i,\text{cool}}, \tau_{i,\text{ff}})}, \quad (21)$$

where $m_{i,\text{hot}}$ is the hot gas mass of the mass shell. The total gas accretion rate is just a summation of the cooling rates from all radii,

$$\dot{M}_{\text{cool}} = \sum_i \dot{m}_{i,\text{cool}}, \quad (22)$$

where the summation is over all the shells that are enclosed by the virial radius of the halo.

The left-hand panel of Fig. 3 shows the cooling gas mass as a function of the radius where the gas is cooling from for different hot gas profiles in an **NFW** halo. For comparison, the black dashed line shows the halo mass profile, $M(<r)$. As one can see, a steeper entropy profile predicts a more concentrated cooling gas mass distribution; the isentropic case with $\beta = 0$ predicts that the cooled gas has a much more extended distribution than the dark matter distribution. The figure illustrates that the thermal state of the circumhalo medium has a clear physical consequence on where the halo gas can cool and eventually contribute mass and angular momentum to the disc.

To estimate the total angular momentum of the cooling baryonic matter at a given time, we assume that the specific angular momentum of the halo material has a radial profile given by

$$j(r) \propto r^\alpha, \quad (23)$$

with $\alpha = 1.1$ (see Bullock et al. 2001). The baryons cooling from a given radius are assumed to have the same specific angular momentum as the dark matter at the same radius. Thus, the specific angular momentum of the cooling baryonic matter relative to that of the entire halo is

$$\frac{j_{\text{cool}}}{j_{\text{DM}}} = \frac{\int j(r) dm_{\text{cool}}(r)}{\int dm_{\text{cool}}(r)} \frac{\int dm_{\text{DM}}(r)}{\int j(r) dm_{\text{DM}}(r)}. \quad (24)$$

Since the cooling gas from the hot halo gas with a shallower entropy profile comes preferentially from larger radii where specific angular momentum is higher, the cooling gas is expected to have a

larger specific angular momentum than the halo. We compute the accumulative angular momentum of the cooling baryonic matter and normalized it by the same quantity of the dark matter. The right-hand panel of Fig. 3 shows the normalized specific angular momentum profile as a function of radius for halo gas with different entropy profiles. For a steep entropy profile, cooling is dominated by inner halo where angular momentum is relatively low, and so the resulting $j_{\text{cool}}/j_{\text{DM}}$ ratio is much smaller than unity. It has been well appreciated that disc galaxies need to retain most of their primordial specific angular momentum to match the observed j – M_* relation (Dutton & van den Bosch 2012; Romanowsky & Fall 2012). Our calculation demonstrates that the halo gas with a steep entropy profile unavoidably produces discs that are too concentrated to match observations. In contrast, when the entropy profile is flat, the baryonic matter can cool from the outer part of a halo where the specific angular momentum is larger, and the contribution of low angular momentum gas is reduced. For the isentropic case ($\beta = 0$), which is expected in the preheating model, we find $j_{\text{cool}}/j_{\text{DM}} \sim 1.5$, which can result in a disc with a characteristic radius about five times as large as that formed in a steep entropy profile with $\beta = 1$.

2.5 Formation of central disc galaxies

Here we focus on modelling the characteristic size of disc galaxies, but do not attempt to derive the origin of the exponential surface density profile. For this reason, we adopt a *weak* form of the angular momentum conservation, namely the total angular momentum of the cooling halo gas is conserved but the detailed distribution of angular momentum of individual mass elements is ignored (Fall 2002). As demonstrated in Dutton (2009), if each gas element conserves its angular momentum as it settles into the disc, the angular momentum distribution predicted by cosmological simulations would produce a disc with material with too low and high angular momentum to match an exponential disc. Specifically, we assume that, in every time step, the newly accreted cold gas has an exponential radial profile with a total angular momentum the same as that of the accreted gas. The exponential scale radius is determined using the Mo et al. (1998, hereafter MMW) model,

$$r_d = \frac{\lambda}{\sqrt{2}} f_j f_c^{-1/2} f_R r_{\text{vir}}, \quad (25)$$

where λ is the spin parameter of the gas, $f_j \equiv (J_{\text{cool}}/m_{\text{cool}})/(J_{\text{DM}}/M_{\text{vir}}) = j_{\text{cool}}/j_{\text{DM}}$ is the ratio of the specific angular momentum of the newly accreted gas and the halo, computed with equation (24), r_{vir} is the virial radius of the halo at the time in question, f_c is a term depending on the halo density profile, and the term f_R corrects for the gravitational effect of the disc and the host halo that result in a non-flat rotation curve. We adopt equations (23) and (29) in MMW to compute f_c and f_R , respectively. When computing f_R , we use rotation curves that take into account both the halo and the existing disc. The effect of contraction of the dark matter halo due to disc formation (Blumenthal et al. 1986; Gnedin et al. 2004; Choi et al. 2006) is ignored. We assume $\lambda = 0.035$, which is the median value of dark matter haloes in cosmological simulations (Bullock et al. 2001; Macciò et al. 2007), and ignore the scatter in the halo spin parameter.

We assume star formation starts when the local surface density of the cold gas on the disc is higher than a certain threshold. We adopt the molecular star formation model proposed by Krumholz, McKee & Tumlinson (2009b). The implementation of this model can be found in Appendix B. The model predicts a surface density of star formation of a disc galaxy as a function of radius, $\Sigma_{\text{SFR}}(R)$.

The formed stars return a fraction of mass back into the ISM as they evolve over time. In our model, we implement a time-dependent mass return model, which is described in Appendix C. The model allows us to trace the mass return from star formation in the past over the entire evolution of a model galaxy. We treat the mass return as a local process, so that the surface density of the mass return rate as a function of radius, $\Sigma_{\text{re}}(R)$, depends on the star formation history of the annulus at the radius R .

The cold gas in the central disc can be also affected by star formation feedback. Here, we model the effect of such feedback by taking into account possible ejection of cold gas from the galactic disc through outflows. The outflow rate is assumed to be proportional to the SFR with an efficiency α_{LD} , known as the mass-loading factor. Thus, the surface density of the outflow rate of each annulus is written as

$$\dot{\Sigma}_{\text{of}}(R) = \alpha_{\text{LD}} \Sigma_{\text{SFR}}(R). \quad (26)$$

Since the time-scale of feedback (due to massive stars) is shorter than the time-scale of mass-loss due to stellar evolution, we treat the feedback as an instantaneous process following star formation. Including the outflow, the evolution of cold disc gas is then given by

$$\dot{\Sigma}_{\text{gas}}(R) = \dot{\Sigma}_{\text{cool}}(R) - \Sigma_{\text{SFR}}(R) - \dot{\Sigma}_{\text{of}}(R) + \dot{\Sigma}_{\text{re}}(R). \quad (27)$$

For the ejective feedback model, we maximize the effect of the ejection by assuming the ejected mass leaves the halo and is never reincorporated back into the halo. This is an extreme assumption, but we will show that even with such an extreme assumption, the ejective feedback model still tends to overpredict the cold baryon mass in low-mass haloes.

In summary, for a given hot halo, whose structure and evolution are modelled in Sections 2.2 and 2.3, the value of $\dot{\Sigma}_{\text{cool}}$ at a given time is obtained from the cooling model described in Section 2.4 in combination with the disc model described in Section 2.5. Once $\dot{\Sigma}_{\text{cool}}$ is predicted and a model for α_{LD} is adopted, equations (B1) in Appendix B, (C4) in Appendix C and (26) can be combined to solve for the surface density profiles of different disc mass components: stars, total cold gas, molecular gas, and atomic gas. All these together provide a complete prescription to follow the formation and evolution of a galaxy disc in an evolving dark matter halo.

3 MODEL PREDICTIONS

Using the model described in the previous section, we make predictions for the properties of galaxy discs and their redshift evolution. In this paper, we consider two distinctively different models to demonstrate the impact of preheating on the formation and evolution of discs in haloes with a mass similar to or lower than that of the Milky Way.

(i) Model-EJ: an ejective feedback model. No preheating is assumed, i.e. the circumhalo gas has an initial entropy that is much lower than the level produced by virial shocks ($S \ll S_{\text{vir}}$). Nearly the cosmic baryon fraction of baryonic matter is accreted into haloes and is shock heated.

(ii) Model-PR: a preventative feedback model assuming preheating, i.e. the circumhalo gas is preheated to have an initial entropy that is higher than or comparable to the halo virial entropy ($S \gtrsim S_{\text{vir}}$). Ejection is completely switched off at $z < 2.5$.

In Model-EJ, because the gas is cold when it is accreted into a dark matter halo, a virial shock will heat the gas up to the virial temperature when it is incorporated into the virial radius of the halo.

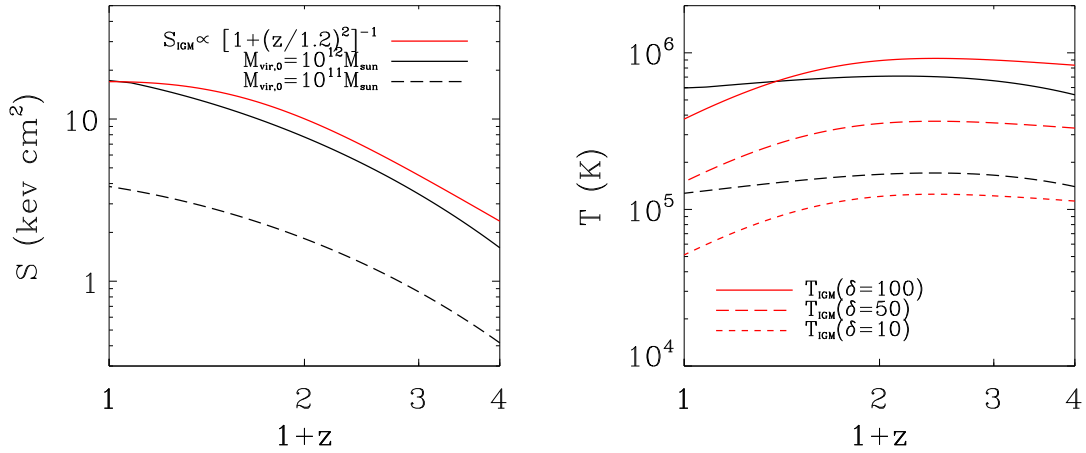


Figure 4. The red line in the left-hand panel shows the history of the assumed preheating entropy as a function of redshift. The right-hand panel shows the corresponding temperature of the preheated medium with different overdensities. The solid red line shows the corresponding temperature if the gas has an overdensity $\delta = 100$. The long- and short-dashed lines show the corresponding temperature if the gas has densities equal to 50 and 10 of the mean density of the universe, respectively. The virial entropy and temperature of dark matter haloes with final ($z = 0$) masses 10^{12} and $10^{11} M_{\odot}$ are shown by the black solid and dashed lines, respectively, in the panels for comparison. The preheating entropy we assume is similar to the virial entropy of present-day $10^{12} M_{\odot}$ haloes, and about five times higher than that of present-day $10^{11} M_{\odot}$ haloes.

For simplicity, we assume that the post-shock gas has an entropy profile with $\beta = 1.1$ in equation (12), consistent with adiabatic hydrodynamical simulations and theoretical models for virial shock heating (e.g. Tozzi & Norman 2001; Voit et al. 2005; Lu & Mo 2007). In this model, we follow the commonly adopted energy-driven wind model, which assumes that the mass-loading factor of star formation feedback is proportional to V_c^{-2} , where V_c is the halo circular velocity at any given time. Specifically, we assume $\alpha_{\text{LD}} = 2 \left(\frac{200 \text{ km s}^{-1}}{V_c} \right)^2$. We note that this choice of outflow mass-loading factor is typical for many existing SAMs (e.g. Benson & Madau 2003; De Lucia, Kauffmann & White 2004; Somerville et al. 2008a) and, as we will show later, the parameters we adopt produce a model that closely matches the baryon mass–halo mass ratio for Milky Way-sized galaxies.

In Model-PR, we assume that the circumhalo medium is generally preheated before it collapses into haloes, and the entropy level increases with time. As a simple model we take

$$S = S_0 \left(\frac{M_{\text{vir},0}}{10^{12} M_{\odot}} \right)^{\mu} \frac{1}{1 + (z/z_c)^{\nu}}, \quad (28)$$

where S_0 is an amplitude, z_c a characteristic redshift, and μ and ν control the halo mass and redshift dependence, respectively. The form is chosen to capture the entropy of the Inter-Galactic Medium (IGM) generated by various processes discussed in the Introduction. The entropy is expected to build up over cosmic time, and to increase with halo mass because at a given time, haloes with higher masses are biased towards higher density regions where star formation is more active. We have tried varying the parameters, and found that setting $S_0 = 17 \text{ Kev cm}^2$, $z_c = 1.2$, $\mu = 0.2$, and $\nu = 2$ matches the observational data remarkably well. Fig. 4 shows the entropy and temperature histories of this preheating model. If the preheating entropy is equal to or higher than the virial entropy of the haloes, no strong accretion shock is expected as the gas accretes into dark matter haloes. The hot halo gas is then expected to have a flat entropy profile, corresponding to an isentropic gas distribution ($\beta = 0$). If, on the other hand, the virial entropy is higher than the preheating entropy, we neglect the effect of preheating and the model is equivalent to Model-EJ.

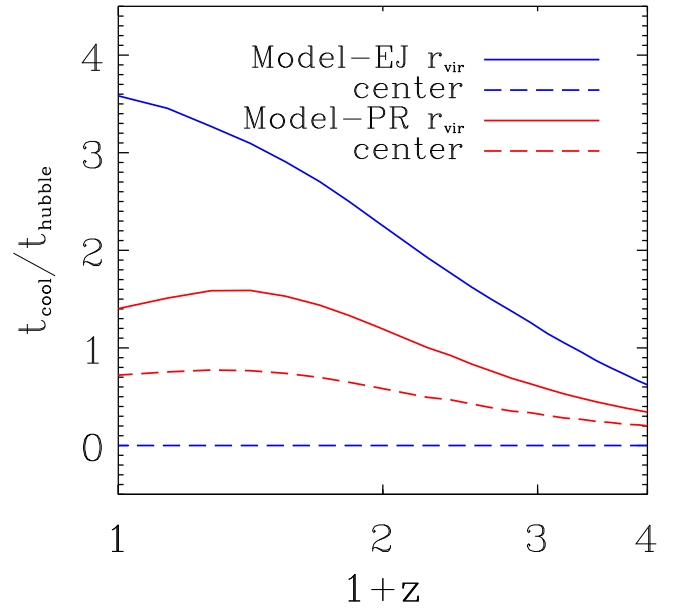


Figure 5. The cooling time-scale of a halo with a final mass of $10^{12} M_{\odot}$ at the virial radius (solid line) and at the halo centre ($r = 0.03 \text{ kpc}$, dashed line) predicted by Model-EJ (blue) and Model-PR (red) as a function of redshift. The cooling time is normalized by the Hubble time at the corresponding redshift z .

To understand the impact of preheating on the cooling time-scale of the halo gas, we take a smooth halo mass accretion history for a halo with final mass $M_{\text{vir},0} = 10^{12} M_{\odot}$ at $z = 0$ and compute the cooling time-scale of the halo gas at the virial radius and the halo centre ($r = 0.03 \text{ kpc}$) using Model-EJ and Model-PR. In the calculation, we switch off cooling and star formation just to show the cooling time-scale given by the original gas distribution predicted by the two models under different assumptions of the entropy. The results are shown in Fig. 5. In Model-EJ, the cooling time-scale is not only a strong function of radius, owing to the steep gas density profile, but also depends strongly on redshift. The cooling time-scale is always much shorter than the Hubble time at the

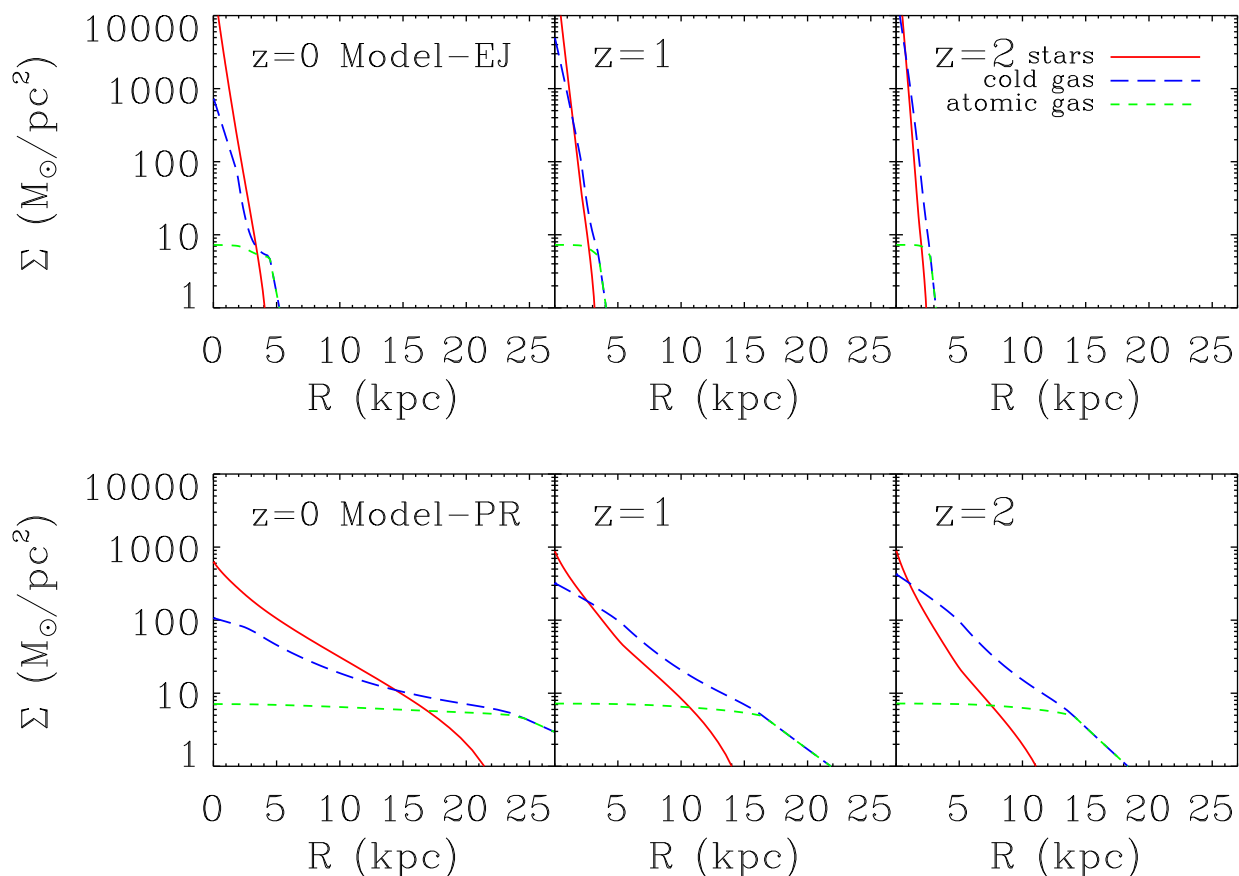


Figure 6. The surface density profiles of the cold baryonic matter of the central disc at $z = 0, 1$, and 2 predicted by Model-EJ (upper panels) and Model-PR (lower panels). In each panel, the red solid line shows the stellar mass surface density of the central disc as a function radius; the blue long-dashed line is the total cold (atomic plus molecular) gas mass surface density profile; the green short-dashed line is for atomic gas only. The preheating model produces significantly more extended discs.

halo centre. At the virial radius, the cooling time-scale is shorter than the Hubble time at high redshift when the mass of the progenitor is low, but the cooling time becomes several times longer than the Hubble time at low redshift as the virial temperature increases and gas density decreases. The situation is very different in Model-PR, where the cooling time-scale in units of Hubble time depends only weakly on radius and redshift. The cooling time-scale of the halo gas stays roughly in the range from 0.5 to 1.5 times the Hubble time over a large cosmic time since $z = 2$. Thus, the effect of preheating is to create a situation where the cooling time of the halo gas in low-mass haloes is comparable to the Hubble time.

In the preheating model, no outflow is assumed at low redshifts when the SFR is reduced by preheating. At high redshifts before preheating, a constant mass-loading factor with no halo mass dependence is assumed. Specifically, we write the loading factor as

$$\alpha_{\text{LD}}(z) = \frac{\alpha_{\text{LD,inf}}}{2} \left[1 + \text{erf} \left(\frac{z - z_c}{\Delta z_c} \right) \right], \quad (29)$$

and take $\alpha_{\text{LD,inf}} = 5$, $z_c = 2.5$, and $\Delta z_c = 1$. This early outflow only affects star formation at high redshift ($z \gtrsim 2.5$), which is not the focus of this paper.

3.1 The growth of discs

Using the two models described above, we predict how a disc builds up. Fig. 6 shows the predicted surface density profiles of the stellar mass, cold gas mass, and atomic gas at $z = 0, 1$, and 2 for a halo with

a final mass $M_{\text{vir},0} = 10^{12} M_{\odot}$. Results are shown for both Model-EJ (upper panels) and Model-PR (lower panels). In both cases, the stellar mass surface densities follow roughly an exponential profile with a density enhancement at the centre ($r \lesssim 3$ kpc). The atomic gas has a flat distribution and dominates the total cold (atomic plus molecular) gas in the outer disc, while the total gas distribution is concentrated and dominated by the molecular gas in the inner region of the disc. Furthermore, the atomic gas disc is more extended than the stellar disc at all redshifts because the outer gaseous disc has surface densities too low to form molecular gas and stars. At any given redshift, Model-PR predicts a more extended disc than Model-EJ, because the hot halo gas can cool from a more extended volume in Model-PR, and, hence, the disc has larger angular momentum.

To quantify the structural evolution of disc galaxies, we randomly select 200 haloes with masses ranging from 10^{10} to $10^{12} M_{\odot}$ from the Bolshoi simulation volume at each of the three redshifts, $z = 0, 1$, and 2 , down-weighting low-mass haloes with a selection probability $p \propto M_{\text{vir}}^{1.3}$ to avoid having too many low-mass haloes that are similar in the predicted quantities. We then apply Model-EJ and Model-PR to the MAHs of these haloes to predict the half stellar mass radius, defined as the radius within which half of the disc stellar mass is contained, as a function of stellar mass. The model predictions are shown in Fig. 7 in comparison with existing observational data. For the observational data, we only show the mean relation between the half-mass radius and stellar mass, as the scatter in the relation is not a focus of the model. The observational data are from Dutton et al. (2007), for local galaxies, and from Dutton et al. (2011) and

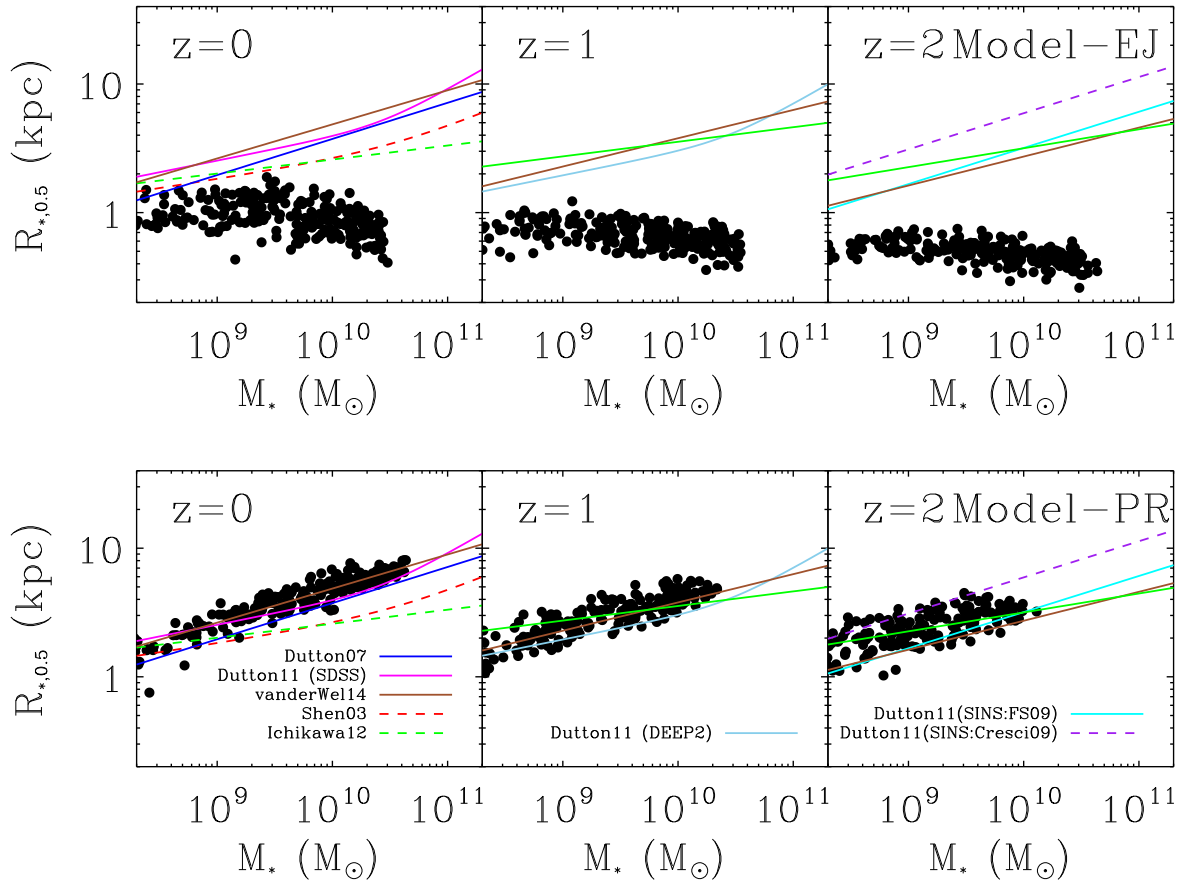


Figure 7. Half-mass radius of the stellar disc as a function of stellar mass at $z = 0, 1$, and 2 . The black dots are model predictions using a set of 200 randomly selected MAHs from a cosmological simulation at each redshift. All haloes have final masses ranging from 10^{10} to $10^{12} M_{\odot}$ at each redshift. The upper row shows the predictions of Model-EJ, and the lower row shows Model-PR. The lines are compilations of observational data. The blue line is the result of Dutton et al. (2007) for local galaxies. The brown lines are the results of van der Wel et al. (2014). The light blue line in the $z = 1$ panel is the result of Dutton et al. (2011) for DEEP2 Galaxy Redshift Survey data. The cyan line in the $z = 2$ panel is the result of Dutton et al. (2011) based on the results of Förster Schreiber et al. (2009) derived from the Spectroscopic Imaging Survey (SINS) data. The red dashed line and the green dashed lines are measure measurements by Shen et al. (2003) and Ichikawa, Kajisawa & Akhlaghi (2012), respectively. The purple dashed line shows the sizes of SINS samples measured by Cresci et al. (2009).

van der Wel et al. (2014), for galaxies in a range of redshift from $z = 0$ to 2 . In addition, we also show the results of Shen et al. (2003) and Ichikawa et al. (2012), which may underestimate the disc sizes because of the use of circular profiles, and the results of Cresci et al. (2009), which have been shown by Dutton et al. (2011) to overestimate the sizes.

The upper panels show the predictions of Model-EJ. We see clearly that this model significantly underpredicts the half-mass radius, especially at high- z and for high-mass galaxies. In contrast, the predictions of Model-PR, shown in the lower panels, are in excellent agreement with the observational data.

Fig. 8 shows the average half-mass radii of the stellar disc as functions of redshift for haloes with a present-day virial mass in a narrow range of $12 \leq \log(M_{\text{vir},0}/M_{\odot}) < 12.2$ predicted by the two models; 300 randomly selected halo mass accretion histories are used to make this prediction. The error bars are 1σ scatter among the model galaxies. The recent observational result of van Dokkum et al. (2013) for the evolution of the progenitors of the present-day Milky Way-sized galaxies since $z \sim 2.5$ is shown in Fig. 8 as filled circles. We note that the sizes of Milky Way-like galaxies in the van Dokkum et al. (2013) result are roughly two times smaller than those of disc galaxies in Dutton et al. (2007), Dutton et al. (2011), and van der Wel et al. (2014). The difference arises mainly because the van Dokkum et al. sample include all

galaxies with the same stellar mass as the Milky Way, including both star-forming disc galaxies and quiescent spheroidal galaxies, but other results shown in Fig. 7 are only for disc galaxies. In addition, van Dokkum et al. (2013) measured circularized radii of stacked images, whereas other results either measure major axis radii or corrected for inclination. Knowing these issues, we find that the prediction of Model-PR is in general agreement with the observational data with an overprediction of the size by about a factor of 2–3 compared to the observational estimate for Milky Way-sized galaxies at late times ($z < 1$). This overprediction is not surprising because we only attempt to model disc galaxies, and ignore angular momentum loss, which is expected in spheroidal galaxies formed mainly through mergers (e.g. Shen et al. 2003). This overprediction leaves room for angular momentum loss possibly happening in reality. In contrast, Model-EJ predicts disc sizes that are too small to match the observation for all redshifts. If the angular momentum is lost during the assembly of the disc, the size of the disc is expected to be even smaller. In terms of evolution, Model-PR predicts that the half-mass radius increases by a factor of 2 since $z \sim 2$, but the van Dokkum et al. (2013) result suggests a rather weak evolution. The galaxy samples used to represent progenitors of present-day Milky Way galaxies in van Dokkum et al. (2013) are selected based on a constant number density. According to the study of Behroozi et al. (2013d), the constant number density

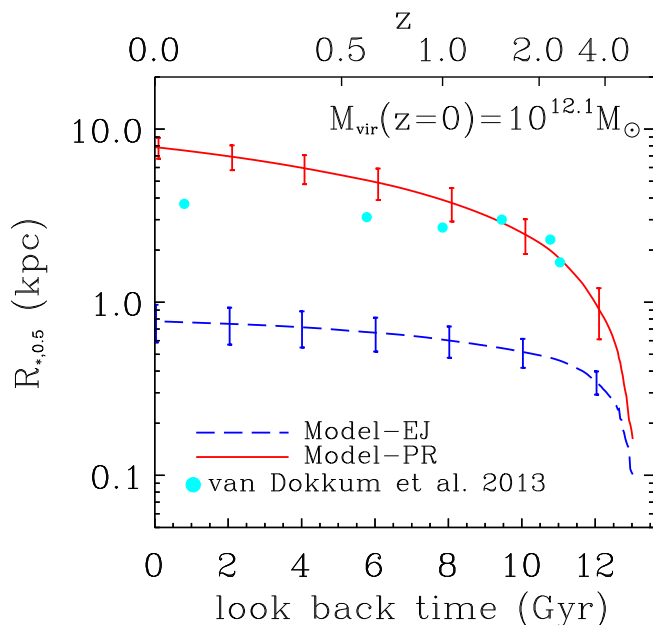


Figure 8. The evolution of the half stellar mass radius of the central galaxy of a halo with final mass $10^{12.1} M_{\odot}$ at $z = 0$. The blue line is the prediction of Model-EJ, and the red line is that of Model-PR. The filled circles denote the observational result of van Dokkum et al. (2013) for the evolution of the progenitors of the present-day Milky Way-sized galaxies since $z \sim 2.5$.

selection could overestimate the mass of the progenitors at higher redshifts. Given the current uncertainties in selecting progenitors in observations, the predicted evolution trend for disc sizes remains to be tested by more accurate observational estimates.

3.2 The growth of disc stellar mass

In this subsection, we examine how the stellar mass of discs grows in haloes with final masses of about $10^{11.1}$ and $10^{12.1} M_{\odot}$ in both Model-EJ and Model-PR. We again use 300 randomly se-

lected simulation MAHs for each of the two final halo mass bins, $11 \leq \log(M_{\text{vir},0}/M_{\odot}) < 11.2$ and $12 \leq \log(M_{\text{vir},0}/M_{\odot}) < 12.2$, and apply the two models to these MAHs to make predictions. Fig. 9 shows the stellar mass of central galaxies as a function of time (redshift), with the left-hand panel for the $\sim 10^{11.1} M_{\odot}$ haloes and the right-hand panel for the $\sim 10^{12.1} M_{\odot}$ haloes. The solid lines are the median stellar masses of the MAHs, and the error bars show the 50 percent of the distribution around the median. The predictions of Model-EJ and Model-PR are shown as the blue and red lines, respectively. The predictions show that stars form earlier in Model-EJ than in Model-PR, and the difference is larger for the lower mass haloes. For $M_{\text{vir},0} = 10^{11.1} M_{\odot}$ haloes, Model-EJ predicts a stellar mass at $z = 2$ about five times higher than Model-PR. Even for $10^{12.1} M_{\odot}$ haloes, the prediction of Model-EJ at $z = 2$ is about two times as high as that of Model-PR at the same redshift. We compare our model predictions with results obtained by Behroozi et al. (2013c) and Lu et al. (2014a) using observationally constrained empirical models. The high value of M_{\star} at $z \geq 2$ for $M_{\text{vir},0} = 10^{11.1} M_{\odot}$ haloes obtained by Lu et al. (2014a) is due to a boost of star formation in low-mass haloes at high- z as they inferred to match the faint-end luminosity function of present-day clusters of galaxies. We do not attempt to capture this behaviour in our model because the total mass of stars formed in this mode is only a small fraction ($\sim 1/10$) of the total final stellar mass, and it happens at high redshift ($z \gtrsim 2$). As one can see, Model-EJ overpredicts the stellar mass for $M_{\text{vir},0} = 10^{11.1} M_{\odot}$ haloes over the entire redshift range, while Model-PR matches the empirical results remarkably well given the simplicity of the model. For $M_{\text{vir},0} = 10^{12.1} M_{\odot}$ haloes, Model-EJ matches the empirical results well at $z < 1$ but still overpredicts M_{\star} at higher z . In contrast, Model-PR matches the empirical results reasonably well at $z > 1$ but overpredicts M_{\star} at very low redshift ($z < 0.5$).

These discrepancies between the model predictions and the empirical results can also be seen in the SFR histories shown in Fig. 10. For $M_{\text{vir},0} = 10^{11.1} M_{\odot}$ haloes, Model-EJ predicts an SFR history that is peaked at $z \approx 2$, in contrast to the rather flat histories predicted by Model-PR and obtained from the empirical models. For the case of $M_{\text{vir},0} = 10^{12.1} M_{\odot}$, although Model-EJ reproduces the

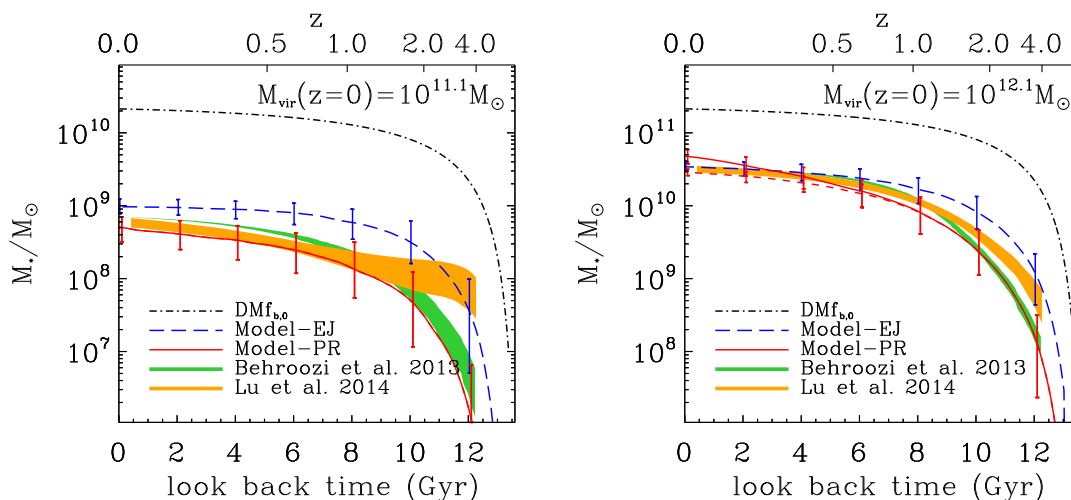


Figure 9. The evolution of stellar mass for haloes with final mass $\sim 10^{11.1} M_{\odot}$ (left) and $\sim 10^{12.1} M_{\odot}$ (right) at $z = 0$. The blue line denotes the prediction of Model-EJ averaged over 200 simulation halo MAHs, the solid red line denotes the same prediction with preheating, Model-PR. The dashed red line denotes a modified model based on the preheating model (see text). The error bars show the standard deviation of the mean expected from a sample of 100 galaxies. For comparison the black line shows the smoothed halo MAH using equation (1) for the corresponding halo mass multiplied by the universal baryon fraction $f_{b,0}$ as a function of time (or z). The colour bands show the central galaxy stellar mass histories obtained from the empirical models of Behroozi et al. (2013c) and Lu et al. (2014a).

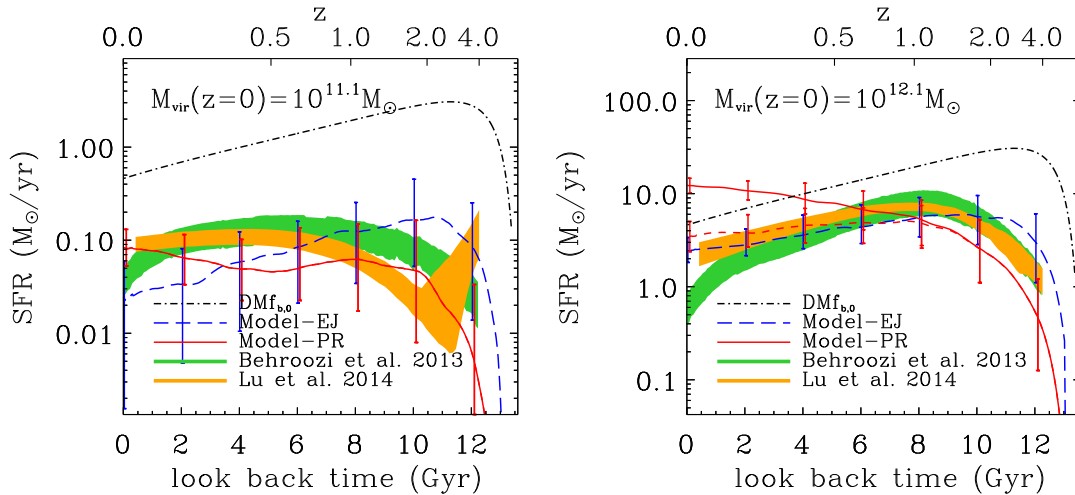


Figure 10. The SFR histories of haloes with final mass $10^{11.1} M_{\odot}$ (left) and $10^{12.1} M_{\odot}$ (right) at $z = 0$. The blue line is the prediction of Model-EJ, the solid red line is prediction of Model-PR, and the dashed red line is the prediction of a modified model based on the preheating model, which will be discussed in Section 3.5. The error bars are the expected standard deviation of the mean with 100 galaxy samples. The black line denotes the halo mass accretion rate characterized by equation (1) multiplied by the universal baryon fraction f_b as a function of z . The colour bands show the central galaxy stellar mass histories obtained by Behroozi et al. (2013c) and Lu et al. (2014a).

decreasing trend for SFR at late time as seen in the empirical results, it predicts a broad peak for the star formation history at $z \sim 3$, which is much earlier than that in the empirical results. Overall, the ejective feedback implemented in Model-EJ is more effective in suppressing star formation in low-mass haloes and at late times. This trend makes it difficult for the model to match the SFR histories derived from the empirical models. In contrast, Model-PR generally predicts a rising SFR history for both mass bins. It matches the empirical results well for the lower mass bin, and early times ($z > 1$) for the $10^{12.1} M_{\odot}$ haloes. At late times, however, the model produces a slowly rising SFR history. If the observed decreasing SFR for the $10^{12.1} M_{\odot}$ haloes at low- z is real, it suggests that some of the galaxies in this halo mass bin must be undergoing some quenching process at late times and this process occurs in relatively massive galaxies but not low-mass ones. We will come back to a possible implication of this behaviour in Section 3.5.

3.3 Cold baryon mass fractions in dark matter haloes

We use 200 randomly chosen halo MAHs with final mass ranging from 10^{10} to $10^{12} M_{\odot}$ at $z = 0$ to make predictions for the final cold gas mass and stellar mass of the central galaxies hosted by those haloes. In the left-hand panels of Fig. 11, we show the predictions for the total cold baryon mass (stellar mass plus atomic cold gas) fraction as a function of halo mass and compare the model predictions with the result of Papastergis et al. (2012) for the same quantities obtained from abundance matching. In the right-hand panels, we show the stellar mass fraction (stellar mass to halo mass ratio) as a function of halo mass at $z = 0$ and compare the predictions with the results of Behroozi et al. (2013c) and Lu et al. (2014a).

Since the feedback parameters in Model-EJ are tuned to match the observed cold gas mass and stellar mass for $10^{12} M_{\odot}$ haloes, it is not surprising that the predictions of the model at the high-mass end are in agreement with the data. However, the model predicts a much shallower slope of the cold baryon mass–halo mass relations. Compared to the result of Papastergis et al. (2012) for halo masses lower than $10^{11} M_{\odot}$, Model-EJ overpredicts the cold baryon mass fractions for such haloes, even though this model invokes a strong

feedback outflow. The upper-right panel further shows that the predicted stellar masses for these haloes are too high. The reason is that the only way to reduce the cold baryon mass in the ejective model is to form more stars in order to produce more outflow. For the particular model we adopt here, the mass-loading factor is as large as 11 for haloes with mass $10^{11} M_{\odot}$ at $z = 0$. The result demonstrates that in order to further reduce the cold baryon mass fraction in low-mass haloes, the model assuming full baryon accretion would require an even stronger ejection and an even larger mass-loading factor for low-mass galaxies than what is assumed here. In contrast, Model-PR nicely reproduces the observed trends in both the cold baryon mass and stellar mass fractions over the entire halo mass range. The entropy assumed for the circumhalo gas naturally produces a relation in which the baryon mass fraction in a halo is proportional to the halo mass, as shown in Lu & Mo (2007). These results demonstrate that accurate measurements for the stellar mass and cold baryon mass fraction in haloes with mass lower than $10^{11} M_{\odot}$ can provide stringent constraints on feedback models.

3.4 Evolution of the number density of low-mass galaxies

We apply both Model-EJ and Model-PR to mass accretion histories of haloes with masses in the range of 10^{10} – $10^{14} M_{\odot}$ at $z = 0, 0.5, 1, 2, 3, 4, 5$, and 6, and select predicted galaxies with stellar masses in the range of $9.27 < \log M_{\star} / M_{\odot} < 9.77$ at each redshift. Fig. 12 shows the predicted number density of galaxies in this stellar mass range as a function of redshift in comparison with observational data. The pink band shows the data compiled in Weinmann et al. (2012), and the points with error bars are our own compilation from various recent observations: Baldry, Glazebrook & Driver (2008) and Baldry et al. (2012) for local galaxies; Moustakas et al. (2013) for galaxies out to $z = 1$; Santini et al. (2012) and Tomczak et al. (2014) for galaxies out to $z \approx 3$; and Lee et al. (2012), González et al. (2011) and Stark et al. (2013) for galaxies at higher redshifts. The Santini et al. results give a higher number density of galaxies in this stellar mass range than the data sets adopted in Weinmann et al. In particular, the data of Santini et al. show a very steep

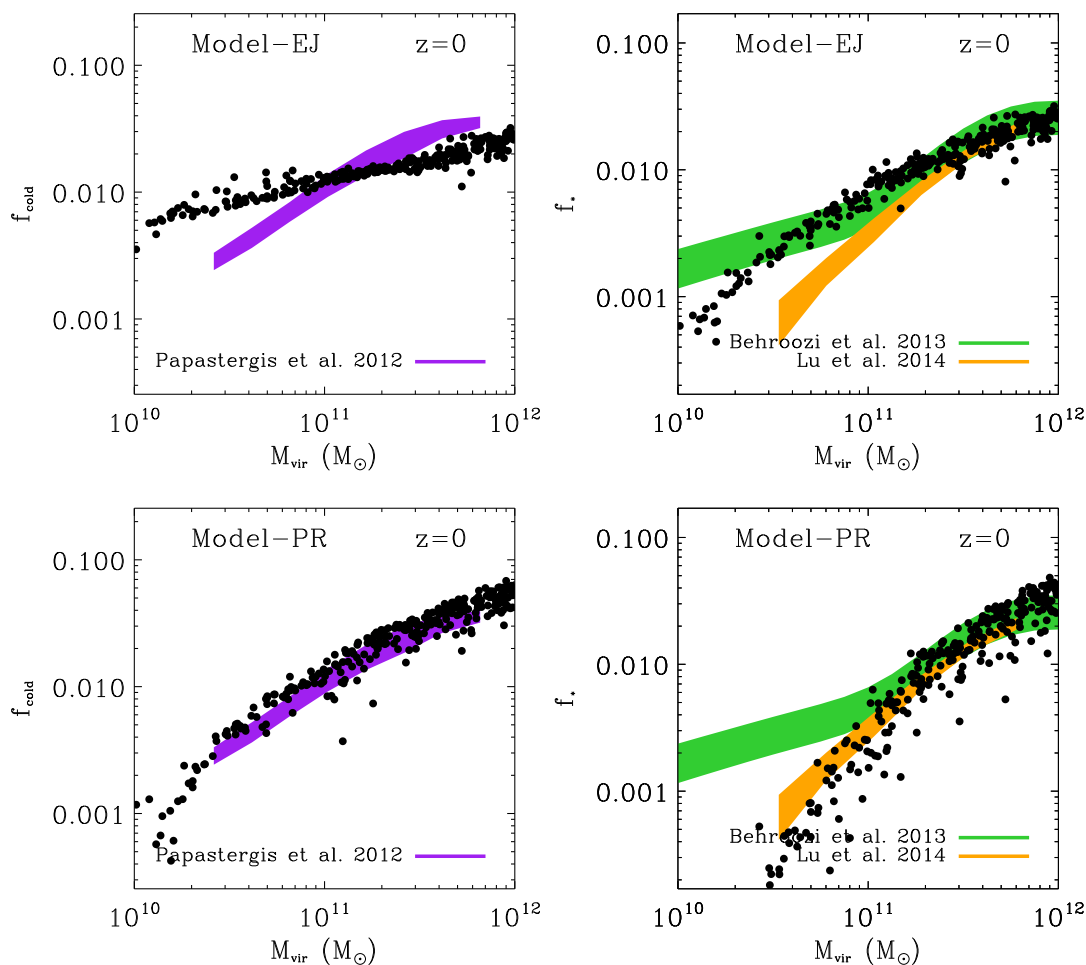


Figure 11. The baryon mass fractions as a function of halo mass for central galaxies at $z = 0$. The black dots show the model predictions for 200 randomly selected realistic halo accretion histories. The left-hand column shows the ratio between the cold baryon mass (stellar mass plus atomic cold gas mass) and halo mass. The predictions are compared with the results of Papastergis et al. (2012). The right-hand column shows the stellar mass fraction, M_*/M_{vir} , predicted by the models and comparison with the results of Behroozi et al. (2013c) and Lu et al. (2014a) obtained from observational data (the differences between the two are due to different treatments of the faint-end slope of the stellar mass function, which is not well constrained observationally). The upper row panels are for Model-EJ, and the lower row panels are for Model-PR.

low-mass end slope at $z \approx 2$, leading to an exceptionally high number density at this redshift. The number density evolution predicted by Model-EJ is very similar to the predictions of the SAMs and hydrodynamical simulations shown in Weinmann et al. (2012) and that of the SAM in Lu et al. (2014b) which is also based on the assumption of strong outflow. On the other hand, Model-PR predicts a much lower number density of galaxies in this mass range at high redshifts (by half a dex at $z \geq 2$). The prediction of Model-PR agrees better with data compiled in Weinmann et al. (2012) than Model-EJ, and it agrees with the observational points we compiled remarkably well in the entire redshift range, except the jump at $z \approx 2$ in the Santini et al. (2012) data. The difference between the model predictions and the discrepancy among the observational data clearly demonstrate the importance of more accurate determinations of the number density of low-mass galaxies in the redshift range of 1–4. We expect that final analysis of the CANDELS data (Grogin et al. 2011; Koekemoer et al. 2011) will improve these constraints in the near future.

Recently, Kannan et al. (2014) showed that simulations with early stellar feedback, which prevents gas around star-forming galaxy from cooling, together with subsequent SN ejections, can reproduce the number density of low-mass galaxies at $z > 2$. Our result

is consistent with this in that prevention of gas accretion of low-mass galaxies by some early processes is needed to match the observational data. Our models demonstrate how gas flows assumed in different scenarios can affect the stellar and cold gas contents of galaxies.

3.5 Late quenching of star formation in high-mass galaxies

Observations have shown that the sSFR, which is defined as the SFR divided by the stellar mass, generally decreases with increasing stellar mass (e.g. Noeske et al. 2007; Salim et al. 2007). For local galaxies with a stellar mass as high as $10^{10} M_{\odot}$, about 70 per cent have an sSFR below a certain threshold, and hence are classified as quiescent galaxies (Moustakas et al. 2013). As the analyses of Lu et al. (2014a) and Behroozi et al. (2013c) represent the average behaviour of star formation histories of the entire observed galaxy population, the quiescent population is included in their results. The models we have considered so far are for star-forming galaxies, without taking into account any process that may quench star formation in high-mass galaxies. Recent observations have indicated that the quenching of star formation in galaxies are closely related to the stellar mass surface density in the inner 1 kpc^2 of galaxies (Cheung

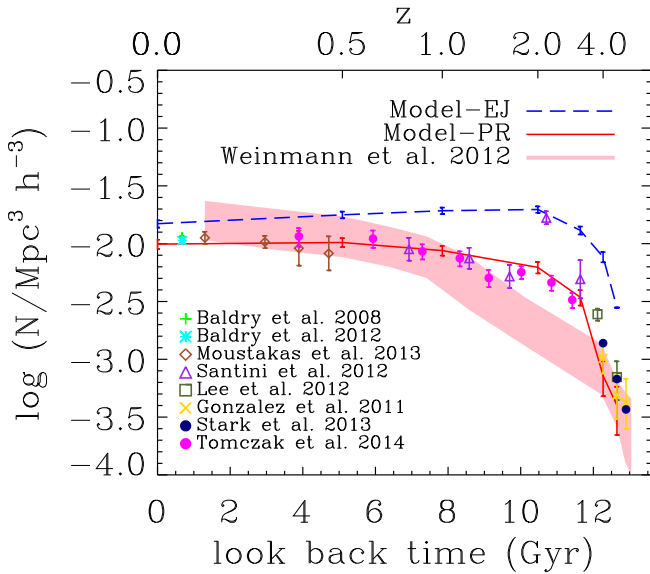


Figure 12. The evolution of the number density of galaxies with stellar mass in the range $9.27 < \log M_*/M_\odot < 9.77$. The pink band shows the compilation of data by Weinmann et al. (2012). The colour points with error bars are data from recent observational results. The sources of the data are listed in the figure.

et al. 2012; Fang et al. 2013b): galaxies with a central stellar mass surface density in the central 1 kpc^2 area lower than $\sim 10^9 M_\odot \text{ kpc}^{-2}$ are dominated by star-forming galaxies, while those with a central stellar mass surface density higher than this are quenched. This quenching is probably related to the bulge component, which is not explicitly modelled in this paper. To mimic the quenching at the high-mass end for the preheating model, we include a recipe assuming that the mass-loading factor depends on the stellar mass as $\alpha_{\text{LD}} = \left(\frac{M_*}{2.5 \times 10^{10} M_\odot} \right)^2$. This feedback only affects high-mass galaxies whose central stellar mass density is high. Although the underlying physics is not well understood, this phenomenological model is able to capture the observational trend that quenching is more pronounced in galaxies with higher stellar mass and higher central stellar surface density. With the disc size model adopted here, in which the characteristic disc size scales with stellar mass as $r_d \sim M_*^{0.3}$, the central stellar mass surface density goes with the stellar mass as $\Sigma_{*,0} \sim M_*^{0.4}$. Thus the mass-loading factor assumed above scales with the central stellar mass surface density roughly as $\Sigma_{*,0}^5$, implying that the feedback strength increases rapidly with increasing central stellar mass surface density. We note, however, to fully understand the ‘central stellar surface density quenching’, one needs to build a model that includes bulge formation (Porter et al. 2014). As shown in Fig. 10 with the dashed lines, this model effectively reduces the SFR in high-mass galaxies since $z \sim 1$, while the SFR history for low-mass galaxies is not affected at all. With such a quenching, the preheating model reproduces well the decreasing trend of the SFR at late times for Milky Way-sized galaxies, but all other predictions are not affected significantly (and hence are not shown for the sake of clarity). We stress that this stellar mass-dependent quenching recipe does not fix the problem of Model-EJ, because with strong outflow the model already predicts a declining SFR history for Milky Way-sized galaxies, and including this quenching recipe makes the decline even more rapid and hence more inconsistent with the data.

4 CONCLUSIONS AND DISCUSSION

In this paper, we have developed a semi-analytic galaxy formation model with a self-consistent treatment for the hot halo gas configuration and disc formation. The model follows realistic halo mass accretion histories extracted from a cosmological N -body simulation, and makes predictions for the structure of baryonic matter in different phases for central disc galaxies hosted by haloes with mass similar to or lower than the Milky Way galaxy. We contrast many predictions made by two models based on different assumptions for the thermal state of the circumhalo medium and how the medium is accreted into dark matter haloes to establish a gaseous halo. The models certainly have uncertainties in various components governing the baryonic processes, and we are aware of degeneracy between model parameters. Here, we only attempt to demonstrate very basic behaviours of the model based on two distinct assumptions. We defer comprehensive analysis in the parameter space to a future paper using our established method (Lu et al. 2011).

One of the models makes the conventional assumption that the accretion of gas by haloes is from a cold medium. In this model, baryons collapse with dark matter and accretion shocks heat the gas to form a gaseous halo with a steep power-law entropy profile. This model thus predicts that cooling is inside-out. In general the discs from such models contain too much material with low angular momentum, resulting in galaxy discs that are too compact. Although, with a fine tuning of the feedback mass-loading factor, the model is able to reproduce the baryon mass fraction of Milky Way-sized haloes at the present day, it still overpredicts the baryon mass fraction in low-mass haloes, even if the chosen outflow is much stronger than what is observed. The model tends to suppress star formation too much at late times and in low-mass haloes, which results in predicted star formation histories which differ significantly from current observations. It also predicts too many low-mass galaxies at high- z . All these results confirm recent findings of similar problems in this class of models by other investigators (e.g. Dutton & van den Bosch 2012; Weinmann et al. 2012; Lu et al. 2014b). On the other hand, however, a number of recent hydrodynamical simulations have shown that early star formation feedback sourced by the radiation pressure of massive stars, combined with subsequent feedback from SNe, is able to suppress star formation in low-mass haloes and to reproduce a number of observed scaling relations of the galaxy population (Brook et al. 2012; Davé et al. 2013; Hopkins et al. 2014; Stinson et al. 2013; Kannan et al. 2014). Based on the results obtained in the present paper, we suggest that feedback in the hydrodynamical simulations may not be entirely effective, but preventative effects may play an important role. It is, therefore, important to investigate in detail how the baryon flows are actually affected by these feedback processes in the simulations.

The new model proposed here assumes that the gas to be accreted by haloes has a finite entropy gradually increasing with time and reaching $\sim 15 \text{ Kev cm}^2$ at $z = 0$. The thermal pressure of the gas with the entropy is high enough to prevent a large fraction of baryons from collapsing into low-mass haloes, and the collapsed gas has a flat entropy radial profile in the halo, which prevents cooling from the very central region of the halo. In this model, gas cooling from large radii brings higher angular momentum, resulting in a central disc extended enough to match the observed galaxy size–stellar mass relation over a large redshift range. This model also reproduces the observed relations between stellar mass and halo mass, between cold baryon (stellar plus cold gas) mass and halo mass, and between disc size and stellar mass, as well as the observed evolution of the number density of low-mass galaxies, without invoking

excessive outflows at late times. It should be pointed out, however, that outflows are allowed at high- z in the preventative model considered here, because the entropy is assumed to be important only at $z < 2$. It might even be required, because early outflows may provide the sources for the preheating. As demonstrated by Mo & Mao (2004) and Pontzen & Governato (2012), such ejective feedback at high- z may also help remove the central cusps in dark matter haloes.

In addition to the predictions we have compared with observational data, the model we propose in this paper has a number of other observational implications which may be used to test the basic picture. The preventative model implies that a large fraction of baryons has never collapsed into low-mass haloes. Furthermore, because all (low-mass) haloes started with relatively low baryon fraction after preheating, no strong feedback is needed to expel baryonic gas from haloes at late time. Thus, the preheating model predicts that any outflows associated with low-mass disc galaxies at low- z are weak. We note, however, that if feedback from star formation/AGN activities is responsible for the entropy generation, strong outflows are still expected at high- z . Metals can be tracers, although maybe strongly biased, of inflow and outflow. Metallicity measurements of low-mass galaxies may provide useful constraints to the model (e.g. Zahid et al. 2012). Moreover, the preheated media at the present time implied by our model may be detected by quasar absorption line systems of highly ionized elements, such as C IV and O VI, and may indeed be the ‘missing’ baryons we are looking for (e.g. Yao et al. 2008; Gupta et al. 2012). Finally, our model suggests that the morphologies of galaxies are closely tied to the thermal state of hot halo gas and even the circumhalo medium, as it determines the angular momentum of the baryons that assemble the disc galaxies. Such a framework may be used to connect observed disc properties with underlying dark matter haloes, and thereby infer the spatial distribution of the pre-collapsed baryonic matter (e.g. Kassin et al. 2012).

The model we explore in this paper is based on a hypothesis of preheated circumhalo medium. Although the general picture we capture in this model is motivated by physical consideration of various plausible early feedback processes, the origin of the preheating entropy and detailed physics of how the preventative feedback works remain to be solved. Because of the unknown physics, the parameters adopted here are largely ad hoc, and the uncertainties of those parameters and their impact on the model predictions cannot be meaningfully discussed. The remaining question is, of course, what is the origin of preheating? There are a number of suggestions in the literature, as described in the Introduction, but none of them has been investigated in detail. These different scenarios are expected to make different predictions for the level of preheating, as well as for its halo mass and redshift dependences. The consequences of these differences can be explored in more detail using a full SAM, such as the one developed in Lu et al. (2011). Furthermore, the uncertainties of the baryon processes including the preheating process can be constrained by observational data using the inference approach developed by Lu et al. (2011, 2012, 2014b). We will come back to this in a forthcoming paper.

ACKNOWLEDGEMENTS

The authors thank Tom Abel, Eric Bell, Andrew Benson, James Bullock, Edmond Cheung, Aaron Dutton, Sandra Faber, Mike Fall, Zhankui Lu, Ari Maller, Joel Primack, Rachel Somerville, Joop Schaye, and Frank van den Bosch for useful discussions. YL and RHW received partial support from HST-AR-12838, provided by NASA through a grant from the Space Telescope Science

Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555. HJM acknowledges support from NSF AST-1109354. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1053575.

REFERENCES

- Baldry I. K., Glazebrook K., Driver S. P., 2008, *MNRAS*, 388, 945
 Baldry I. K. et al., 2012, *MNRAS*, 421, 621
 Behroozi P. S., Conroy C., Wechsler R. H., 2010, *ApJ*, 717, 379
 Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013a, *ApJ*, 762, 109
 Behroozi P. S., Wechsler R. H., Wu H.-Y., Busha M. T., Klypin A. A., Primack J. R., 2013b, *ApJ*, 763, 18
 Behroozi P. S., Wechsler R. H., Conroy C., 2013c, *ApJ*, 770, 57
 Behroozi P. S., Marchesini D., Wechsler R. H., Muzzin A., Papovich C., Stefanon M., 2013d, *ApJ*, 777, L10
 Benson A. J., Madau P., 2003, *MNRAS*, 344, 835
 Binney J., 1977, *ApJ*, 215, 483
 Blanton M. R., Lupton R. H., Schlegel D. J., Strauss M. A., Brinkmann J., Fukugita M., Loveday J., 2005, *ApJ*, 631, 208
 Blumenthal G. R., Faber S. M., Flores R., Primack J. R., 1986, *ApJ*, 301, 27
 Bouché N., Hohensee W., Vargas R., Kacprzak G. G., Martin C. L., Cooke J., Churchill C. W., 2012, *MNRAS*, 426, 801
 Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, *MNRAS*, 370, 645
 Bradley L. D. et al., 2012, *ApJ*, 760, 108
 Brook C. B. et al., 2011, *MNRAS*, 415, 1051
 Brook C. B., Stinson G., Gibson B. K., Wadsley J., Quinn T., 2012, *MNRAS*, 424, 1275
 Bruzual G., Charlot S., 2003, *MNRAS*, 344, 1000
 Bryan G. L., 2000, *ApJ*, 544, L1
 Bullock J. S., Dekel A., Kolatt T. S., Kravtsov A. V., Klypin A. A., Porciani C., Primack J. R., 2001, *ApJ*, 555, 240
 Cassata P. et al., 2013, *ApJ*, 775, 106
 Chabrier G., 2003, *PASP*, 115, 763
 Cheung E. et al., 2012, *ApJ*, 760, 131
 Choi J., Lu Y., Mo H. J., Weinberg M. D., 2006, *MNRAS*, 372, 1869
 Conroy C., Wechsler R. H., Kravtsov A. V., 2006, *ApJ*, 647, 201
 Crain R. A., McCarthy I. G., Frenk C. S., Theuns T., Schaye J., 2010, *MNRAS*, 407, 1403
 Cresci G. et al., 2009, *ApJ*, 697, 115
 Croton D. J. et al., 2006, *MNRAS*, 365, 11
 Dalcanton J. J., Spergel D. N., Summers F. J., 1997, *ApJ*, 482, 659
 Davé R., Katz N., Oppenheimer B. D., Kollmeier J. A., Weinberg D. H., 2013, *MNRAS*, 434, 2645
 De Lucia G., Kauffmann G., White S. D. M., 2004, *MNRAS*, 349, 1101
 Dekel A., Silk J., 1986, *ApJ*, 303, 39
 Dormand J., Prince P., 1980, *J. Comput. Appl. Math.*, 6, 19
 Driver S. P. et al., 2011, *MNRAS*, 413, 971
 Dutton A. A., 2009, *MNRAS*, 396, 121
 Dutton A. A., van den Bosch F. C., 2012, *MNRAS*, 421, 608
 Dutton A. A., van den Bosch F. C., Dekel A., Courteau S., 2007, *ApJ*, 654, 27
 Dutton A. A. et al., 2011, *MNRAS*, 410, 1660
 Eke V. R., Navarro J. F., Steinmetz M., 2001, *ApJ*, 554, 114
 Fall S. M., 2002, in Da Costa G. S., Sadler E. M., Jerjen H., eds, *ASP Conf. Ser. Vol. 273, The Dynamics, Structure & History of Galaxies: A Workshop in Honour of Professor Ken Freeman*, Astron. Soc. Pac., San Francisco, p. 289
 Fall S. M., Efstathiou G., 1980, *MNRAS*, 193, 189
 Fang T., Bullock J., Boylan-Kolchin M., 2013a, *ApJ*, 762, 20
 Fang J. J., Faber S. M., Koo D. C., Dekel A., 2013b, *ApJ*, 776, 63
 Finkelstein S. L. et al., 2012, *ApJ*, 756, 164
 Förster Schreiber N. M. et al., 2009, *ApJ*, 706, 1364
 Giovanelli R. et al., 2005a, *AJ*, 130, 2598

- Giovanelli R. et al., 2005b, *AJ*, 130, 2613
- Gnedin N. Y., 2000, *ApJ*, 542, 535
- Gnedin O. Y., Kravtsov A. V., Klypin A. A., Nagai D., 2004, *ApJ*, 616, 16
- González V., Labbé I., Bouwens R. J., Illingworth G., Franx M., Kriek M., 2011, *ApJ*, 735, L34
- Grogin N. A. et al., 2011, *ApJS*, 197, 35
- Gupta A., Mathur S., Krongold Y., Nicastro F., Galeazzi M., 2012, *ApJ*, 756, L8
- Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R. E., Guo Q., Lemson G., Springel V., 2013, *MNRAS*, 431, 3373
- Hopkins P. F., Keres D., Onorbe J., Faucher-Giguere C.-A., Quataert E., Murray N., Bullock J. S., 2014, *MNRAS*, 445, 581
- Ichikawa T., Kajisawa M., Akhlaghi M., 2012, *MNRAS*, 422, 1014
- Jungwiert B., Combes F., Palouš J., 2001, *A&A*, 376, 85
- Kannan R., Stinson G. S., Macciò A. V., Brook C., Weinmann S. M., Wadsley J., Couchman H. M. P., 2014, *MNRAS*, 437, 3529
- Kassin S. A., Devriendt J., Fall S. M., de Jong R. S., Allgood B., Primack J. R., 2012, *MNRAS*, 424, 502
- Katz N., Gunn J. E., 1991, *ApJ*, 377, 365
- Kaufmann T., Bullock J. S., Maller A. H., Fang T., Wadsley J., 2009, *MNRAS*, 396, 191
- Kennicutt J. R. C., 1998, *ARA&A*, 36, 189
- Klypin A. A., Trujillo-Gomez S., Primack J., 2011, *ApJ*, 740, 102
- Koekemoer A. M. et al., 2011, *ApJS*, 197, 36
- Kravtsov A. V., Gnedin O. Y., Klypin A. A., 2004, *ApJ*, 609, 482
- Krumholz M. R., Thompson T. A., 2012, *ApJ*, 760, 155
- Krumholz M. R., McKee C. F., Tumlinson J., 2009a, *ApJ*, 693, 216
- Krumholz M. R., McKee C. F., Tumlinson J., 2009b, *ApJ*, 699, 850
- Lee K.-S. et al., 2012, *ApJ*, 752, 66
- Lu Y., Mo H. J., 2007, *MNRAS*, 377, 617
- Lu Y., Mo H. J., Weinberg M. D., Katz N., 2011, *MNRAS*, 416, 1949
- Lu Y., Mo H. J., Katz N., Weinberg M. D., 2012, *MNRAS*, 421, 1779
- Lu Y. et al., 2014, *ApJ*, 795, 123
- Lu Z., Mo H. J., Lu Y., Katz N., Weinberg M. D., van den Bosch F. C., Yang X., 2014a, *MNRAS*, 439, 1294
- Lu Y., Mo H. J., Lu Z., Katz N., Weinberg M. D., 2014b, *MNRAS*, 443, 1252
- Mac Low M., Ferrara A., 1999, *ApJ*, 513, 142
- McBride J., Fakhouri O., Ma C.-P., 2009, *MNRAS*, 398, 1858
- McCarthy I. G., Balogh M. L., Babul A., Poole G. B., Horner D. J., 2004, *ApJ*, 613, 811
- McCarthy I. G. et al., 2010, *MNRAS*, 406, 822
- Macciò A. V., Dutton A. A., van den Bosch F. C., Moore B., Potter D., Stadel J., 2007, *MNRAS*, 378, 55
- McKee C. F., Krumholz M. R., 2010, *ApJ*, 709, 308
- McLure R. J. et al., 2013, *MNRAS*, 432, 2696
- Maller A. H., Bullock J. S., 2004, *MNRAS*, 355, 694
- Maller A. H., Dekel A., 2002, *MNRAS*, 335, 487
- Meyer M. J. et al., 2004, *MNRAS*, 350, 1195
- Mo H. J., Mao S., 2002, *MNRAS*, 333, 768
- Mo H. J., Mao S., 2004, *MNRAS*, 353, 829
- Mo H. J., Miralda-Escudé J., 1996, *ApJ*, 469, 589
- Mo H. J., Mao S., White S. D. M., 1998, *MNRAS*, 295, 319 (MMW)
- Mo H. J., Yang X., van den Bosch F. C., Katz N., 2005, *MNRAS*, 363, 1155
- Mo H. J., van den Bosch F., White S. D. M., 2010, *Galaxy Formation and Evolution*, 1st edn. Cambridge Univ. Press, Cambridge
- Moster B. P., Naab T., White S. D. M., 2013, *MNRAS*, 428, 3121
- Moustakas J. et al., 2013, *ApJ*, 767, 50
- Mutch S. J., Poole G. B., Croton D. J., 2013, *MNRAS*, 428, 2001
- Navarro J. F., Benz W., 1991, *ApJ*, 380, 320
- Navarro J. F., Frenk C. S., White S. D. M., 1996, *ApJ*, 462, 563 (NFW)
- Newman S. F. et al., 2012, *ApJ*, 761, 43
- Noeske K. G. et al., 2007, *ApJ*, 660, L43
- Oesch P. A. et al., 2012, *ApJ*, 759, 135
- Oh S. P., Benson A. J., 2003, *MNRAS*, 342, 664
- Papastergis E., Cattaneo A., Huang S., Giovanelli R., Haynes M. P., 2012, *ApJ*, 759, 138
- Patel S. G. et al., 2013a, *ApJ*, 766, 15
- Patel S. G. et al., 2013b, *ApJ*, 778, 115
- Pfrommer C., Chang P., Broderick A. E., 2012, *ApJ*, 752, 24
- Pontzen A., Governato F., 2012, *MNRAS*, 421, 3464
- Porter L. A., Somerville R. S., Primack J. R., Johansson P. H., 2014, *MNRAS*, 444, 942
- Prada F., Klypin A. A., Cuesta A. J., Betancort-Rijo J. E., Primack J., 2012, *MNRAS*, 423, 3018
- Reddick R. M., Wechsler R. H., Tinker J. L., Behroozi P. S., 2013, *ApJ*, 771, 30
- Rees M. J., Ostriker J. P., 1977, *MNRAS*, 179, 541
- Romanowsky A. J., Fall S. M., 2012, *ApJS*, 203, 17
- Salim S. et al., 2007, *ApJS*, 173, 267
- Santini P. et al., 2012, *A&A*, 538, A33
- Scannapieco E., Oh S. P., 2004, *ApJ*, 608, 62
- Shen S., Mo H. J., White S. D. M., Blanton M. R., Kauffmann G., Voges W., Brinkmann J., Csabai I., 2003, *MNRAS*, 343, 978
- Silk J., 1977, *ApJ*, 211, 638
- Silk J., 1997, *ApJ*, 481, 703
- Silk J., Mamon G. A., 2012, *Res. Astron. Astrophys.*, 12, 917
- Somerville R. S., Hopkins P. F., Cox T. J., Robertson B. E., Hernquist L., 2008a, *MNRAS*, 391, 481
- Somerville R. S. et al., 2008b, *ApJ*, 672, 776
- Spergel D. N. et al., 2007, *ApJS*, 170, 377
- Stark D. P., Schenker M. A., Ellis R., Robertson B., McLure R., Dunlop J., 2013, *ApJ*, 763, 129
- Steinmetz M., Navarro J. F., 1999, *ApJ*, 513, 555
- Stinson G. S., Brook C., Macciò A. V., Wadsley J., Quinn T. R., Couchman H. M. P., 2013, *MNRAS*, 428, 129
- Sutherland R. S., Dopita M. A., 1993, *ApJS*, 88, 253
- Tang S., Wang Q. D., Lu Y., Mo H. J., 2009, *MNRAS*, 392, 77
- Thoul A. A., Weinberg D. H., 1996, *ApJ*, 465, 608
- Tomczak A. R. et al., 2014, *ApJ*, 783, 85
- Tozzi P., Norman C., 2001, *ApJ*, 546, 63
- Trujillo I. et al., 2006, *ApJ*, 650, 18
- van den Bosch F. C., Abel T., Hernquist L., 2003, *MNRAS*, 346, 177
- van der Wel A. et al., 2014, *ApJ*, 788, 28
- van Dokkum P. G. et al., 2013, *ApJ*, 771, L35
- Voit G. M., Donahue M., 2005, *ApJ*, 634, 955
- Voit G. M., Kay S. T., Bryan G. L., 2005, *MNRAS*, 364, 909
- Wang L., Weinmann S. M., Neistein E., 2012, *MNRAS*, 421, 3450
- Weinmann S. M., Pasquali A., Oppenheimer B. D., Finlator K., Mendel J. T., Crain R. A., Macciò A. V., 2012, *MNRAS*, 426, 2797
- Werner N., Allen S. W., Simionescu A., 2012, *MNRAS*, 425, 2731
- White S. D. M., Rees M. J., 1978, *MNRAS*, 183, 341
- Yan H. et al., 2012, *ApJ*, 761, 177
- Yang X., Mo H. J., van den Bosch F. C., 2003, *MNRAS*, 339, 1057
- Yang X., Mo H. J., van den Bosch F. C., 2008, *ApJ*, 676, 248
- Yang X., Mo H. J., van den Bosch F. C., Zhang Y., Han J., 2012, *ApJ*, 752, 41
- Yang X., Mo H. J., van den Bosch F. C., Bonaca A., Li S., Lu Y., Lu Y., Lu Z., 2013, *ApJ*, 770, 115
- Yao Y., Nowak M. A., Wang Q. D., Schulz N. S., Canizares C. R., 2008, *ApJ*, 672, L21
- York D. G. et al., 2000, *AJ*, 120, 1579
- Zahid H. J., Dima G. I., Kewley L. J., Erb D. K., Davé R., 2012, *ApJ*, 757, 54
- Zhao D. H., Mo H. J., Jing Y. P., Börner G., 2003, *MNRAS*, 339, 12
- Zhao D. H., Jing Y. P., Mo H. J., Börner G., 2009, *ApJ*, 707, 354
- Zhu W., Feng L.-L., Fang L.-Z., 2011, *MNRAS*, 415, 1093
- Zwaan M. A. et al., 2004, *MNRAS*, 350, 1210

APPENDIX A: IMPLEMENTATION OF THE MODEL FOR REIONIZATION

We model the effect of photoionization heating by the UV background following Gnedin (2000) who showed that the fraction of baryons that can collapse into a halo of a given mass can be

described in terms of a ‘filtering mass’, M_F . The baryon fraction in haloes with masses lower than M_F is reduced relative to the universal fraction according to

$$f_b(z, M_{\text{vir}}) = \frac{f_{b,0}}{[1 + 0.26M_F(z)/M_{\text{vir}}]^3}, \quad (\text{A1})$$

where M_{vir} is the halo virial mass. The filtering mass depends on the reionization history of the Universe and is redshift-dependent. Kravtsov et al. (2004) provided fitting formulae for the filtering mass according to both the redshift at which the first H II regions begin to overlap (z_{overlap}) and the redshift at which most of the medium is reionized (z_{reion}). We make use of the fitting functions (B2) and (B3) in the appendix of Kravtsov et al. (2004) to compute the initial fraction of baryons, f_b , as a function of halo mass and redshift. In this paper, we assume $z_{\text{overlap}} = 11$ and $z_{\text{reion}} = 10$, as suggested by *WMAP* results (e.g. Spergel et al. 2007).

APPENDIX B: IMPLEMENTATION OF THE STAR FORMATION MODEL

To predict star formation in a cold gas disc, we adopt the model developed by Krumholz et al. (2009b), which assumes that the star formation efficiency depends on the local total (atomic and molecular hydrogen) gas surface density, Σ_{cold} , and the molecular gas fraction, f_{H_2} . Specifically, the SFR surface density is given by

$$\Sigma_{\text{SFR}} = \begin{cases} \epsilon f_{\text{H}_2} \Sigma_{\text{cold}} (\Sigma_{\text{cold}} / \Sigma_0)^{-0.33}, & \Sigma_{\text{cold}} < \Sigma_0; \\ \epsilon f_{\text{H}_2} \Sigma_{\text{cold}} (\Sigma_{\text{cold}} / \Sigma_0)^{0.33}, & \Sigma_{\text{cold}} \geq \Sigma_0, \end{cases} \quad (\text{B1})$$

where $\epsilon = 0.39 \text{ Gyr}^{-1}$ and $\Sigma_0 = 85 M_{\odot} \text{ pc}^{-2}$. For an annulus with a surface density of cold gas Σ_{cold} , we compute the molecular fraction, f_{H_2} , using the model of Krumholz, McKee & Tumlinson (2009a):

$$f_{\text{H}_2} = \begin{cases} 1 - \frac{3}{4} \left(\frac{s}{1+0.25\xi} \right) & \text{if } \xi < 2; \\ 0 & \text{if } \xi \geq 2, \end{cases} \quad (\text{B2})$$

where

$$\xi = \frac{\ln(1 + 0.6\chi + 0.01\chi^2)}{0.6\tau_c}, \quad (\text{B3})$$

in which

$$\chi = 3.1 \left(\frac{1 + 3.1Z_0^{0.365}}{4.1} \right), \quad (\text{B4})$$

and

$$\tau_c = 0.066CZ_0\Sigma_0, \quad (\text{B5})$$

with Z_0 the metallicity of the gas in units of solar metallicity, Z_{\odot} , and C a clumpiness factor that accounts for smoothing of the surface density on scales larger than that of a single molecular complex (see also McKee & Krumholz 2010). As suggested in Krumholz et al. (2009b), $C \sim 5$ when Σ_{cold} is measured on $\sim 1 \text{ kpc}$ scales. In our model we take $C = 5$.

APPENDIX C: TIME-DEPENDENT STELLAR MASS LOSS MODEL

In a time interval Δt , the stellar mass at a given radius in the disc can change due to star formation and the mass-loss of formed stars,

$$\dot{\Sigma}_*(R, t)\Delta t = \Sigma_{\text{SFR}}(R, t)\Delta t - \dot{\Sigma}_{\text{re}}(R, t)\Delta t. \quad (\text{C1})$$

Here, $\dot{\Sigma}_{\text{re}}(R, t)\Delta t$ is the surface density of the stellar mass that formed in the past but returned into the ISM in the time interval Δt

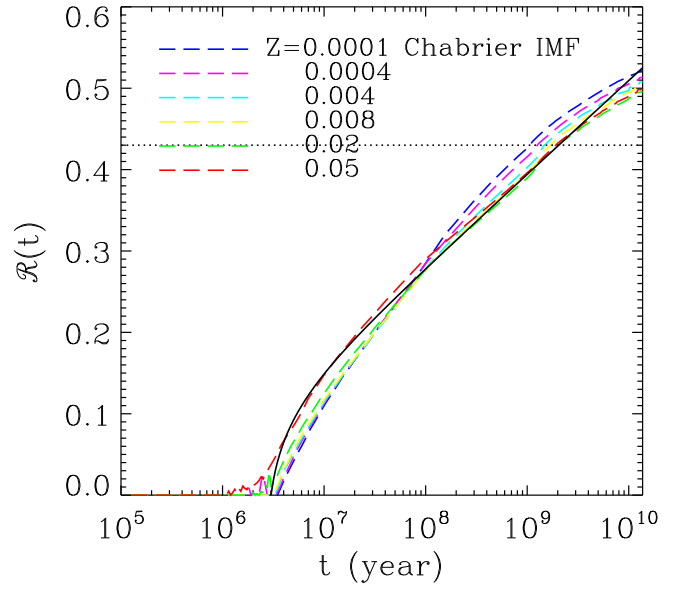


Figure C1. The fraction of stellar mass loss due to the evolution of a simple stellar population. The dashed lines are predicted by the BC03 model for a Chabrier (2003) IMF with different metallicities as indicated in the panel. The black solid line is the model we adopt to mimic the stellar mass loss for a Chabrier IMF (see equation C4). The dotted line denotes the typical value adopted for instantaneous recycling approximation for the IMF.

at time t , and its rate, $\dot{\Sigma}_{\text{re}}$, is determined by the star formation history and the initial mass function (IMF). We write the mass return of the stellar mass formed at radius R with an SFR $\Sigma_{\text{SFR}}(R, t')$ at an early epoch t' as

$$\dot{\Sigma}_{\text{re}}(R, t, t')\Delta t = [\mathcal{R}(t - t' + 0.5\Delta t) - \mathcal{R}(t - t' - 0.5\Delta t)] \Sigma_{\text{SFR}}(R, t'), \quad (\text{C2})$$

where $\mathcal{R}(t)$ is the returned fraction determined by the IMF and stellar evolution. We use the following fitting formula proposed by Jungwiert, Combes & Palouš (2001) to describe $\mathcal{R}(t)$:

$$\mathcal{R}(t) = \begin{cases} 0 & t < t_0; \\ c_0 \ln \left(\frac{t-t_0}{\tau} + 1 \right) & t \geq t_0. \end{cases} \quad (\text{C3})$$

The parameters, $c_0 = 0.05$, $t_0 = 3 \times 10^6 \text{ yr}$, and $\tau = 3.76 \times 10^5 \text{ yr}$ are obtained for a Chabrier (2003) IMF. In Fig. C1 we compare this fitting function (the solid curve) to the return fractions obtained from the BC03 model (Bruzual & Charlot 2003) with the same IMF for different metallicities (dashed curves). Clearly the above formula is a good approximation for the return fraction over a large range of metallicity. Finally we can write

$$\dot{\Sigma}_{\text{re}}(R, t)\Delta t = \int_0^t [\mathcal{R}(t - t' + 0.5\Delta t) - \mathcal{R}(t - t' - 0.5\Delta t)] \Sigma_{\text{SFR}}(R, t') dt' \quad (\text{C4})$$

This model allows us to trace the mass return over a long time-scale that is relevant for stellar evolution over the entire history of a galaxy.