

# Dark-ages Reionization and Galaxy Formation Simulation – XIV. Gas accretion, cooling, and star formation in dwarf galaxies at high redshift

Yuxiang Qin,<sup>1,2★</sup> Alan R. Duffy,<sup>2,3</sup> Simon J. Mutch,<sup>1,2</sup> Gregory B. Poole,<sup>1,3</sup>  
Paul M. Geil,<sup>1</sup> Andrei Mesinger<sup>4</sup> and J. Stuart B. Wyithe<sup>1,2</sup>

<sup>1</sup>*School of Physics, University of Melbourne, Parkville, VIC 3010, Australia*

<sup>2</sup>*ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D)*

<sup>3</sup>*Centre for Astrophysics and Supercomputing, Swinburne University of Technology, PO Box 218, Hawthorn VIC 3122, Australia*

<sup>4</sup>*Scuola Normale Superiore, Piazza dei Cavalieri 7, I-56126 Pisa, Italy*

Accepted 2018 March 19. Received 2018 March 19; in original form 2018 February 11

## ABSTRACT

We study dwarf galaxy formation at high redshift ( $z \geq 5$ ) using a suite of high-resolution, cosmological hydrodynamic simulations and a semi-analytic model (SAM). We focus on gas accretion, cooling, and star formation in this work by isolating the relevant process from reionization and supernova feedback, which will be further discussed in a companion paper. We apply the SAM to halo merger trees constructed from a collisionless  $N$ -body simulation sharing identical initial conditions to the hydrodynamic suite, and calibrate the free parameters against the stellar mass function predicted by the hydrodynamic simulations at  $z = 5$ . By making comparisons of the star formation history and gas components calculated by the two modelling techniques, we find that semi-analytic prescriptions that are commonly adopted in the literature of low-redshift galaxy formation do not accurately represent dwarf galaxy properties in the hydrodynamic simulation at earlier times. We propose three modifications to SAMs that will provide more accurate high-redshift simulations. These include (1) the halo mass and baryon fraction which are overestimated by collisionless  $N$ -body simulations; (2) the star formation efficiency which follows a different cosmic evolutionary path from the hydrodynamic simulation; and (3) the cooling rate which is not well defined for dwarf galaxies at high redshift. Accurate semi-analytic modelling of dwarf galaxy formation informed by detailed hydrodynamical modelling will facilitate reliable semi-analytic predictions over the large volumes needed for the study of reionization.

**Key words:** methods: numerical – galaxies: dwarf – galaxies: formation – galaxies: high-redshift.

## 1 INTRODUCTION

About 150 Myr after the big bang, the Universe entered the Epoch of Reionization (EoR; Planck Collaboration XLVII 2016), when the first stars and galaxies were created and started to photoionize the intergalactic neutral hydrogen. Over the past decade, there have been significant advances in the study of galaxies during the EoR and the evolution of the ionized intergalactic medium (IGM). High-redshift galaxies thought to have sourced reionization have been discovered all the way to  $z \sim 11$  (Oesch et al. 2016), when the Universe was only 400 Myr old. However, the sample of these early galaxies remains small with only  $\sim 1000$  candidates at  $z > 6$  identified using advanced space-based instruments (Bouwens et al. 2015, 2016). Achieving a physical understanding between galaxies

and the progress of reionization is therefore aided by numerical simulations (see e.g. Bertschinger 1998; Baugh 2006; Dolag et al. 2008; Somerville & Davé 2015).

Two of the main numerical approaches to model galaxy formation are hydrodynamic simulations and semi-analytic models (SAMs) applied to dark matter halo merger trees constructed from cosmological  $N$ -body simulations. Since they adopt different approximations, their strengths and requirements are diverse. Hydrodynamic simulations include baryons as well as dark matter, and simulate the complex baryonic physics more directly (e.g. Crain et al. 2009; Schaye et al. 2010; Vogelsberger et al. 2014; Schaye et al. 2015; Feng et al. 2016; Pawlik et al. 2017). However, while galaxies form and evolve in the presence of gravity and hydrodynamical forces in a simulation volume, its large computational expense limits the capability of simultaneously resolving small-scale regions and capturing massive systems. For instance, in order to study reionization and galaxy formation during the EoR, simulations are required to

\* E-mail: [Yuxiang.L.Qin@gmail.com](mailto:Yuxiang.L.Qin@gmail.com)

have a volume size of at least  $10^6 \text{ Mpc}^3$  (Iliev et al. 2014) and a particle resolution that is able to resolve haloes with masses around  $5 \times 10^7 M_\odot$  (i.e.  $>10$  billion particles; Barkana & Loeb 1999). While it remains challenging to explore the relevant parameter space in hydrodynamic simulations, SAMs, with their advantages of low expense on computational resources, become the alternative in this case.

### 1.1 Semi-analytical models

$N$ -body simulations (e.g. Navarro, Frenk & White 1997; Springel et al. 2005; Iliev et al. 2008; Boylan-Kolchin et al. 2009; Klypin, Trujillo-Gomez & Primack 2011; Garrison et al. 2017) solve for the gravitational force on each collisionless dark matter particle but neglect the baryonic physics. In order to implement the galaxy formation physics of baryons into these simulations, halo merger trees are constructed and based on the properties inherited from the merger trees, SAMs parametrize baryonic processes to evolve galaxies, offering an efficient approach to simulate galaxies in a cosmological context (e.g. Cole et al. 2000; Hatton et al. 2003; Baugh et al. 2005; Croton et al. 2006; De Lucia & Blaizot 2007; Somerville et al. 2008; Guo et al. 2011; Henriques et al. 2015).

In general, a SAM first designates haloes as galaxies and endows them with stellar components and several gas reservoirs of varied functionalities. The latter usually include a cold gas disc where star formation occurs and a hot halo of non-star-forming gas. Mass is then manipulated and transferred between these baryonic sectors as a result of varied baryonic processes including accretion, cooling, star formation, and feedback from supernovae or active galactic nuclei (AGN). These processes are implemented using scaling functions of astrophysics that are motivated either from the first principle (e.g. the thermal cooling rate, see Section 2.1.2), from observations [e.g. the Kennicutt–Schmidt (KS) star formation law, see Section 2.1.3], or from more complicated simulations such as hydrodynamic and radiative transfer calculations (e.g. reionization feedback, Sobacchi & Mesinger 2013).

Depending on the scientific goal and applicable range, the complexity of SAMs ranges from the treatment of halo profiles to the implementation of baryonic physics. Taking star formation as an example, it can be modelled in a simplified manner by consuming the entire cold gas disc within a given time-scale (e.g. SAGE, Croton et al. 2006), or as Stevens, Croton & Mutch 2016 demonstrated (in the updated version, DARK SAGE), by

- (i) considering the disc as a combination of a series of gas rings with angular momentum conserved between each annulus and transferred dynamically;
- (ii) using the mid-plane pressure inferred from local observations and theoretical work (e.g. Blitz & Rosolowsky 2006; Krumholz, McKee & Tumlinson 2009) to split the disc into molecules and atoms, and forming stars directly from the giant  $\text{H}_2$  clouds; and
- (iii) including disc instabilities and forming additional components representing stellar bulges, which might possess different ages or dynamics from the stellar disc.

Despite these improvements recently implemented into modern SAMs, some prescriptions are still commonly adopted by many of these models. For instance, in order to initialize the baryonic component of haloes from dark matter only  $N$ -body simulations, the universal baryon fraction is applied to every virialized system with a further suppression due to reionization feedback. These baryons are considered as infalling gas and are usually assumed to share the virial temperature of the host halo as a result of experiencing shock

heating. However, we note that the degeneracy between parameters of SAMs, as well as their time and mass dependencies introduced by these simplified baryonic prescriptions are not well understood. In addition, lack of available observational data means that semi-analytic predictions have generally been tested against observables that are massive and in the nearby Universe (Croton et al. 2006; Guo et al. 2011; Stevens et al. 2016). Their consistency and performance in the low-mass regime or at high redshift remain unclear – a model might be able to reproduce a limited number of observed quantities by either careful calibrations or by introducing more free parameters accompanied by *physical motivations*, but incorrectly predict the other properties that cannot be observed at this stage.

This motivates us to (1) test the performance of SAMs at high redshift, which were originally developed from the literature of low-redshift galaxy formation; and (2) inform more accurate semi-analytic modelling of dwarf galaxies, which will be applied to large-volume simulations to study the EoR in the future. In the absence of detailed observational data, constraints on the properties of SAM galaxies can be supplemented through comparisons against physics-rich hydrodynamic simulations, which are expected to predict galaxies that are more representative than simplified calculations using semi-analytic approaches (e.g. Guo et al. 2016; Mitchell et al. 2017; Stevens et al. 2017a; Côté et al. 2017). In this work, we take the MERAXES SAM (Mutch et al. 2016a) as an example, and apply it to the halo merger trees constructed from a collisionless  $N$ -body simulation sharing identical initial conditions to the *Smaug* suite of hydrodynamic simulations (Duffy et al. 2014). We calibrate the SAM against the *Smaug* hydrodynamic simulations and make direct comparisons of the calculated galaxy properties using the two modelling techniques. Since the hydrodynamic simulations were run with different assumptions of galactic or stellar feedback, the free parameters of the SAM are chosen accordingly to reproduce the predicted galaxy properties from the corresponding hydrodynamic simulation.

This paper focuses on modelling of gas accretion, cooling, and star formation, the implementation of which is common among many modern SAMs (Croton et al. 2006; Somerville et al. 2008; Guo et al. 2011; Henriques et al. 2015). The effect of feedback on star formation (which varies between models) is excluded and will be presented in a companion paper (Qin et al., in preparation). We begin with a brief introduction of the  $N$ -body/hydrodynamic simulations and SAM utilized in this work in Section 2. Then we discuss the consequence of applying SAMs directly to collisionless  $N$ -body simulations and propose modifications accordingly in Section 3. In Section 4, we proceed with the proposed modifications and calibrate the SAM to reproduce the hydrodynamic result. We investigate galaxy properties in detail and make comparisons between the hydrodynamic simulation and the SAM results. Based on these, we propose additional modifications to the SAM to inform better modelling of dwarf galaxies at high redshift. We conclude in Section 5. In this work, we adopt cosmological parameters from WMAP7 ( $\Omega_m, \Omega_b, \Omega_\Lambda, h, \sigma_8, n_s = 0.275, 0.0458, 0.725, 0.702, 0.816, 0.968$ ; Komatsu et al. 2011) in all simulations.

## 2 THE DRAGONS PROJECT

The Dark-ages Reionization And Galaxy Formation Observable from Numerical Simulation (DRAGONS<sup>1</sup>) project employs  $N$ -body simulations (Poole et al. 2016), hydrodynamic simulations (Duffy

<sup>1</sup> <http://dragons.ph.unimelb.edu.au/>

et al. 2014; Qin et al. 2017a), and SAMs (Mutch et al. 2016a; Qin et al. 2017c) to study reionization and galaxy formation at high redshift ( $z \geq 5$ , Geil et al. 2016; Geil et al. 2017; Liu et al. 2016, 2017; Mutch et al. 2016b; Park et al. 2017; Duffy et al. 2017; Qin et al. 2017b). The resulting galaxy catalogues from both the hydrodynamic simulation and SAM with feedback included are in good agreement with observations including the stellar mass and galaxy ultraviolet luminosity functions across cosmic time.

In this work, we focus on high-redshift modelling ( $z \geq 5$ ) of currently unobservable dwarf galaxies ( $M_{\text{vir}} \lesssim 10^{10.5} M_{\odot}$ ) which are believed to be the dominant source of ionizing photons driving reionization (Duffy et al. 2014; Liu et al. 2016). Using simplified models, we isolate prescriptions of gas accretion, cooling, and star formation from reionization and supernova feedback,<sup>2</sup> which will be further discussed in a companion paper. In the following two subsections, we summarize the DRAGONS SAM and hydrodynamic simulation, named MERAXES and *Smaug*, respectively.

## 2.1 MERAXES

MERAXES is based on the *Munich* SAM (Croton et al. 2006; Guo et al. 2011) and optimized to study galaxy formation at high redshift. Based on halo properties read from merger trees, it evolves galaxies using simplified prescriptions, which calculate baryonic infall, cooling, star formation, supernova feedback, metal enrichment, stellar mass recycling, AGN feedback, and mergers. In addition, MERAXES incorporates 21 cm FAST (Mesinger, Furlanetto & Cen 2011), a seminumerical approach to evolve ionization fields, for the purpose of investigating reionization feedback and exploring the IGM state during the EoR. We refer the interested reader to Mutch et al. (2016a, hereafter *Paper-III*) and Qin et al. (2017c) for a detailed description of MERAXES and provide a brief review of the relevant modelling prescriptions in this section.

### 2.1.1 Accretion (and stripping)

To connect  $N$ -body simulations of dark matter with baryonic physics, the first implementation is the accretion of gas. Following convention, a universal baryon fraction ( $f_b = \Omega_b/\Omega_m$ ) with suppression due to reionization heating ( $f_{\text{mod}}$ ) is applied and each halo group (i.e. friend-of-friend, FOF group, see Appendix A1.1), with a total mass of  $M_{\text{vir}}$ , accretes gas from the IGM

$$\Delta m_{\text{hot}} = f_{\text{mod}} f_b M_{\text{vir}} - \sum_{N_{\text{halo}}} m_{\text{baryon}}, \quad (1)$$

where  $m_{\text{baryon}}$  is the total baryonic mass including<sup>3</sup> hot gas ( $m_{\text{hot}}$ ), cold gas ( $m_{\text{cold}}$ ), and stellar mass ( $m_*$ ), and the summation symbol indicates that baryons of all  $N_{\text{halo}}$  haloes including the central and satellite haloes are considered.  $f_{\text{mod}} \leq 1$  represents a baryon fraction modifier accounting for the suppression due to the reionization heating background.

In the case of  $\Delta m_{\text{hot}} > 0$ , all of the mass is assumed to be accreted by the central halo due to its dominant gravitational potential in the entire virialized system, and are stored in its hot gas reservoir. However, when  $\Delta m_{\text{hot}} < 0$ , baryons will be removed from the hot gas reservoir of the central halo in order to maintain the universal baryon

fraction. Note that baryons that have reached the cold gas disc or have formed stars are assumed to be gravitationally bound, and do not get further removed even if  $f_{\text{mod}} f_b M_{\text{vir}} - \sum_{N_{\text{halo}}} (m_* + m_{\text{cold}}) < 0$ .

### 2.1.2 Cooling

Gas falling into a halo is stored in the hot gas reservoir, which is assumed to share the halo virial temperature,  $T_{\text{vir}}$ , due to shock heating. It then cools within the following time-scale

$$t_{\text{cool}}(r) = \frac{3\bar{\mu}m_p k T_{\text{vir}}}{2\rho_{\text{hot}}(r) \Lambda(T_{\text{vir}})}, \quad (2)$$

where  $\bar{\mu} = 0.59$ ,  $m_p$ ,  $k$ ,  $\rho_{\text{hot}}$ , and  $\Lambda$  are the mean molecular weight of fully ionized gas, the mass of a proton, the Boltzmann constant, the hot gas density, and the cooling function (Sutherland & Dopita 1993) that is determined by the temperature.<sup>4</sup> Assuming the hot gas follows a singular isothermal sphere (SIS) profile, we calculate the cooling radius,  $r_{\text{cool}}$ , at which the cooling time is equal to the halo dynamical time,  $t_{\text{dyn}}$ , through

$$r_{\text{cool}} = \sqrt{\frac{m_{\text{hot}} \Lambda(T_{\text{vir}})}{6\pi\bar{\mu}m_p k T_{\text{vir}} V_{\text{vir}}}}, \quad (3)$$

where  $V_{\text{vir}}$  is the virial velocity of the host halo.

Based on the ratio of the cooling radius to the virial radius,  $R_{\text{vir}}$ , we calculate the cooling rate and consider the following two scenarios:

- (i)  $r_{\text{cool}} < R_{\text{vir}}$  (*static hot halo*): cooling is assumed in thermal equilibrium and the cooling rate is determined by the continuity equation;
- (ii)  $r_{\text{cool}} \geq R_{\text{vir}}$  (*rapid cooling*): cooling is treated as free-fall collapse and all of the hot gas component becomes cold within the dynamical time-scale,  $t_{\text{dyn}}$ .

Therefore, the cooling rate can be calculated by

$$\dot{m}_{\text{cool}} = \frac{m_{\text{hot}}}{t_{\text{dyn}}} \times \min\left(1, \frac{r_{\text{cool}}}{R_{\text{vir}}}\right). \quad (4)$$

### 2.1.3 Star formation

Cooling leads to the build-up of a cold gas reservoir, which is assumed to form a rotationally supported disc, where stars form, with an exponential surface density profile. When the cold gas mass,  $m_{\text{cold}}$ , exceeds a critical value,  $m_{\text{crit}}$ , we use a phenomenological relation suggested by observations (Kennicutt 1998) to calculate the star formation rate (SFR)<sup>5</sup>

$$\dot{m}_* = \alpha_{\text{sf}} \times \frac{\max(0, m_{\text{cold}} - m_{\text{crit}})}{t_{\text{dyn, disc}}}, \quad (5)$$

where  $t_{\text{dyn, disc}} \equiv 1.5\sqrt{2}\lambda t_{\text{dyn}}$  is the dynamical time of the cold gas disc inferred by the halo spin parameter (Bullock et al. 2001),  $\lambda$ , with an additional assumption that the specific angular momentum is conserved between the cold gas disc and the host halo (Mo, Mao

<sup>2</sup> AGN feedback, which is expected to have an insignificant impact to the formation of dwarf galaxies at high redshift, is not considered here.

<sup>3</sup> Gas that has been ejected from galaxies due to supernova feedback is not included in this work. We refer the interested readers to *Paper-III* for more details about feedback and gas stripping prescriptions.

<sup>4</sup> Metals are predominately created by massive stars that have reached type II supernovae. With supernova feedback included, metals are released into the cold gas disc, which are further transferred between various gas reservoirs through heating, cooling, and reincorporation. However, these are not considered in this work and  $\Lambda$  only accounts for primordial elements.

<sup>5</sup> The second channel of star formation is through mergers, which does not have a significant impact in our model.

& White 1998). A free parameter,  $\alpha_{\text{sf}}$ , is introduced for the purpose of adjusting the star formation efficiency.

We note that, in order to be consistent with the hydrodynamic simulations introduced in the next subsection, the newly formed stars are assumed to follow an initial mass function in the mass range of  $0.1\text{--}120 M_{\odot}$  with the following form (Chabrier 2003)

$$\phi(m) = \begin{cases} 0.85m^{-1} \exp\left[-\frac{(\log_{10}m - \log_{10}0.079)^2}{2 \times 0.69^2}\right], & \text{if } 0.1M_{\odot} \leq m < 1M_{\odot}; \\ 0.24m^{-2.3}, & \text{if } 1 \leq m \leq 120M_{\odot}. \end{cases} \quad (6)$$

## 2.2 Smaug

*Smaug* is a series of high-resolution hydrodynamic simulations (Duffy et al. 2014, Qin et al. 2017a) which were run with a modified GADGET-2  $N$ -body/hydrodynamics code (Springel 2005). These models follow the OverWhelmingly Large Simulations project (OWLS; Schaye et al. 2010), in which the full hydrodynamic simulation that involves star formation (Schaye & Dalla Vecchia 2008), radiative cooling (Wiersma, Schaye & Smith 2009a), supernova feedback (Dalla Vecchia & Schaye 2008, 2012), and reionization feedback (Haardt & Madau 2001), has been shown to successfully reproduce observed galaxy properties at low redshift including the cosmic star formation history (Schaye et al. 2010) and the stellar mass function at  $z < 2$  (Haas et al. 2013). In addition, Katsianis et al. (2017) recently showed that the Evolution and Assembly of GaLaxies and their Environments (EAGLE) simulations, an extension of OWLS, are also in agreement with the observed star formation rate function across cosmic time ( $z \sim 0\text{--}8$ ). These achievements motivate us to make use of the hydrodynamic simulations and to investigate the validity of semi-analytic prescriptions at high redshift where observational constraints are unavailable.

### 2.2.1 Galaxy physics

In this work, the simulation from *Smaug* comprises  $512^3$  baryon and  $512^3$  dark matter particles within a cube of comoving side of  $10 h^{-1}$  Mpc. This equates to a mass resolution of  $4.7(0.9) \times 10^5 h^{-1} M_{\odot}$  per dark matter (gas) particle. The time interval between two outputs is around 11 Myr, and the initial conditions were generated with the GRAFIC package (Bertschinger 2001) at  $z = 199$  using the Zel'dovich approximation (Zeldovich 1970). We summarize the subgrid physics adopted in *Smaug* in terms of cooling and star formation for comparison with the semi-analytic prescriptions (see Section 2.1), and refer the interested reader to Schaye et al. (2010) and Duffy et al. (2014) for more details.

(i) *Cooling* (Wiersma et al. 2009a) consists of only primordial elements and is pre-tabulated using the CLOUDY package (Ferland et al. 1998). The cooling rate is calculated in the presence of the cosmic microwave background (CMB), accounting for both free-free scattering between gas particles and Compton scattering between gas particles and CMB photons.

(ii) *Star formation* (Schaye & Dalla Vecchia 2008) occurs by stochastically converting gas particles to star particles in the ISM. In practice, the local density around a gas particle,  $\rho_{\text{g}}$ , can be inferred by smoothing the particle mass,  $m_{\text{g}}$ , within a certain volume. When  $\rho_{\text{g}}$  exceeds a critical value, the gas particle is assumed to be multiphase ISM state with pressure  $P \propto \rho^{\gamma_{\text{eff}}}$ , where  $\gamma_{\text{eff}} = \frac{4}{3}$  represents the effective ratio of specific heats. Note that resolving

star-forming regions in a cosmological hydrodynamic simulation is computationally unrealistic. Therefore, the gas depletion time,  $t_{\text{g}}$ , in the hydrodynamic simulation is calculated based on the observed KS (Kennicutt 1998) star formation law, and the SFR is given by  $\dot{m}_{*} = m_{\text{g}}/t_{\text{g}}$ . We note that although the star formation recipes adopted in the SAM and hydrodynamic simulation are both developed from the KS law, there are discrepancies between the two numerical approaches, which are further discussed in Section 4.1.

### 2.2.2 Simulations

The two simulations used in this work are

(1) *DMONLY*, a collisionless  $N$ -body simulation performed using the same initial conditions from the full simulation but neglecting hydrodynamic forces from the baryonic component. It consists of  $512^3$  collisionless particles with masses of  $5.7 \times 10^5 h^{-1} M_{\odot}$  within a cube of comoving side of  $10 h^{-1}$  Mpc. In order to apply the MERAXES SAM and compare the result with *Smaug*, *DMONLY* is used to construct dark matter halo merger trees (See Appendix A);

(2) *NOSN\_NOZCOOL\_NoRe*, a toy model with cooling in the absence of metal line emissions and reionization heating. It ignores feedback from exploding supernovae and heating due to reionization, and will be used to compare with the SAM for the purpose of investigating gas accretion, cooling, and star formation without impacts from any feedback. The top left panel of Fig. A1 presents an example of the gas density distribution at  $z = 5$  in the *NOSN\_NOZCOOL\_NoRe Smaug* simulation.

### 2.2.3 Cold gas (star-forming gas)

In the SAM, cold gas is considered as potentially star-forming gas (see Section 2.1.3). In order to facilitate direct comparisons of the gas reservoir with the semi-analytic calculation, we specify, for each galaxy in the hydrodynamic simulation, a star-forming gas component. This reservoir comprises all potential star-forming particles of the galaxy, making it comparable to the assumption adopted in the SAM.

As mentioned before, a gas particle located in a cold dense region is considered as multiphase and can potentially form stars. More specifically in *Smaug*, this requires the particle to satisfy *A AND B AND (C OR D)*, a combination of the following four criteria (Schaye et al. 2010):

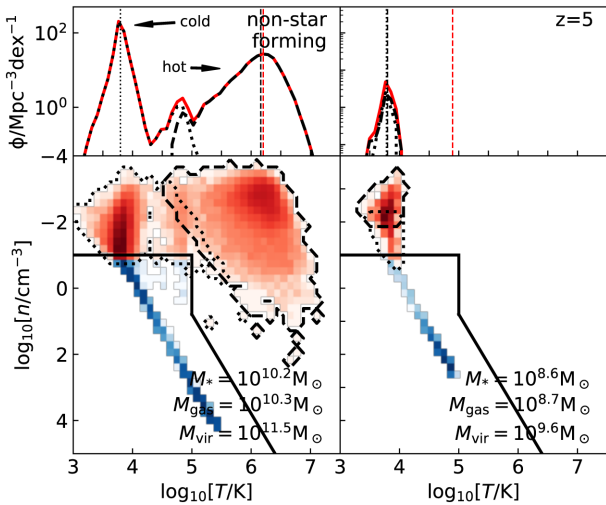
(i) a high physical density, the threshold of which is a fixed value ( $0.1 \text{ cm}^{-3}$ ) and informs the formation of a multiphase ISM through gravitational instability;

(ii) a high comoving density, the threshold of which is 57.7 times the cosmic mean and prevents spurious star formation in less dense region at high redshift;

(iii) *C*: a high physical density and a low temperature, the thresholds of which are informed from the equation of state of the unresolved warm dense ISM, and can recover the KS star formation law;

(iv) a low temperature, the threshold of which is a constant ( $10^5 \text{ K}$ ) and is proposed to capture cold ISM in less dense regions.

The bottom panels of Fig. 1 show the gas density–temperature phase diagram of two haloes at  $z = 5$  in the *NOSN\_NOZCOOL\_NoRe Smaug* simulation for illustration. Particles located in the lower left region are identified as star-forming gas; otherwise they are non-star-forming gas.



**Figure 1.** Bottom panels: gas density–temperature phase diagram of two haloes at  $z = 5$  in the *NOSN\_NOZCOOL\_NoRe Smaug* simulation. The stellar mass, gas mass, and halo mass are shown on the lower right corner. The piecewise function shown with black solid lines indicates the separation of gas particles with different phases – particles inside (outside) the lower left region are identified as star-forming (non-star-forming) gas and are indicated with the blue (red) colour. Using the *KMeans* machine-learning clustering algorithm, the non-star-forming gas particles are further divided into two groups, marked by the dashed (hotter) and dotted (colder) contours, respectively. Top panels: number densities of all (red solid line), hotter (black dashed line), and colder (black dotted line) non-star-forming gas of the two haloes as functions of temperature. The vertical red dashed, black dashed, and black dotted lines represent the FOF halo virial temperature and the median temperatures of the two non-star-forming gas groups, respectively.

### 2.2.4 Hot gas

On the other hand, hot gas in the SAM represents a reservoir where the temperature is as high as the halo virial temperature as a result of shock heating during gas accretion (see Section 2.1.2). Gas within the hot gas reservoir cannot form stars in the SAM. Therefore, in order to define a hot gas component for a galaxy in hydrodynamic simulations, which can be compared with the SAM, we first exclude all potential star-forming particles from the galaxy. We plot the temperature distribution of non-star-forming gas particles of the two haloes in the *NOSN\_NOZCOOL\_NoRe Smaug* simulation in the top panels of Fig. 1. In the more massive halo, three populations<sup>6</sup> can be observed, corresponding to, from right to left, (1) infall hot gas; (2) infall gas that has been through cooling, and the cooling rate decreases significantly when temperature<sup>7</sup> drops to a few  $10^4$  K (Wiersma et al. 2009a); and (3) cold gas which is about to reach the multiphase ISM disc and become dense enough to form stars. However, we can only identify the group of less dense (non-star-forming) cold gas in the less massive halo.

Based on the density–temperature phase diagram of the non-star-forming gas particles of a galaxy, the hot gas mass and temperature of the galaxy can be determined. In practice, depending on the

number of non-star-forming gas particles ( $N_{\text{nonSF}}$ ) that a galaxy comprises,

(i) when  $N_{\text{nonSF}} = 0$ , the hot gas mass of the galaxy is assigned 0;

(ii) when  $N_{\text{nonSF}} = 1$ , the hot gas mass and temperature of the galaxy are assigned the properties of the gas particle ( $m_{\text{nonSF}}$  and  $T_{\text{nonSF}}$ ) if  $T_{\text{nonSF}} > T_{\text{crit}}$ , otherwise, the hot gas mass of the galaxy is assigned 0;

(iii) when  $N_{\text{nonSF}} > 1$ , we use the *scikit-learn* PYTHON package (Pedregosa et al. 2011) – *KMeans*, a clustering algorithm commonly used for machine learning. We consider two features (density and temperature) and separate the non-star-forming gas particles into two groups (see the dashed and dotted contours in the bottom panels of Fig. 1).

(a) If the median temperatures of the two groups differ by more than  $\Delta_{T,\text{crit}}$  in logarithm, the hot gas mass and temperature of the galaxy are assigned the median temperature and the total mass of the hotter group (the dashed contour in the left bottom panel of Fig. 1).

(b) If the median temperatures of the two groups differ less than  $\Delta_{T,\text{crit}}$  in logarithm, the two groups of the non-star-forming gas particles are considered as one cluster. In this case, the hot gas mass and temperature of the galaxy are assigned the total mass and the median temperature ( $\bar{T}_{\text{nonSF}}$ ) of all non-star-forming gas particles if  $\bar{T}_{\text{nonSF}} > T_{\text{crit}}$ , otherwise, the hot gas mass of the galaxy is assigned 0.

We adopt  $T_{\text{crit}} = 10^5$  K and  $\Delta_{T,\text{crit}} = 0.5$  in this work and we note that, for haloes with efficient cooling,<sup>8</sup> the determination of hot gas is not sensitive to the temperature thresholds (i.e.  $T_{\text{crit}}$  and  $\Delta_{T,\text{crit}}$ ; see Section 4.2 and Appendix B).

When the properties of hot gas are defined, we also characterize the remaining particles and identify them as the cold non-star-forming gas component, whose temperature and mass are represented by the median temperature and total mass of the remaining particles.

In the top panels of Fig. 1, we show the temperature function of the two groups separately and indicate their median temperatures with black lines with line styles corresponding to the two contours in the bottom panels of Fig. 1. We also indicate the virial temperature<sup>9</sup> of the system using vertical red dashed lines. We see that the median temperature of the hotter group is in agreement with the virial temperature of the more massive halo, suggesting the infall gas particles of massive haloes have been through virial shocks and are heated to the virial temperature. However, less massive galaxies do not comprise any hot gas component. We will further discuss it in Section 4.2.

## 3 MODIFICATIONS OF HALO MASSES AND BARYON FRACTIONS

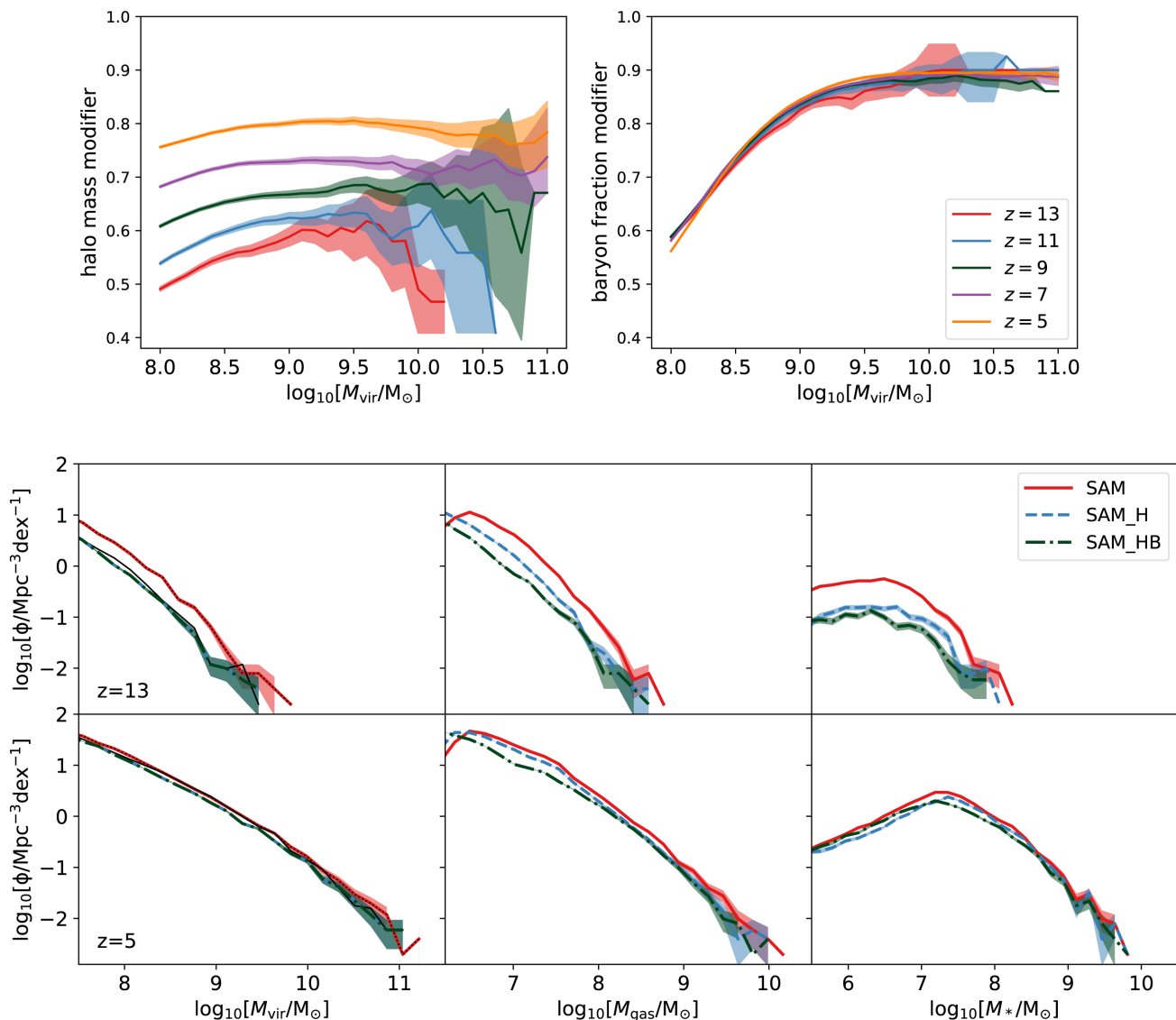
Before performing any comparisons, there is one modification regarding the suppressed growth of dark matter haloes that needs to be addressed. Hydrostatic pressure keeps gas from collapsing into shallow potential wells of low-mass haloes, which consequently decreases the gravitational potential within these haloes

<sup>6</sup> Supernova feedback and reionization are not included in *NOSN\_NOZCOOL\_NoRe*, which heat the ISM and alter the gas temperature–phase diagram as demonstrated in the companion paper.

<sup>7</sup> In collisional ionization equilibrium, hydrogen is ionized at this temperature ( $\sim 5 \times 10^4$  K), which, however, is low for helium ionization and appears as a trough on the cooling function curve of primordial gas.

<sup>8</sup> Molecular cooling is not implemented in this work. Therefore only haloes that are more massive than the atomic cooling limit possess efficient cooling.

<sup>9</sup> The virial temperature is calculated using the FOF group properties (see Section 2.1.1).



**Figure 2.** Top panels: halo mass modifier and baryon fraction modifier as functions of halo mass in *N-body* simulation at  $z = 13$ – $5$ , with solid lines and shaded regions representing the mean values and uncertainties showing 95 per cent confidence intervals around the mean using 100 000 bootstrap resamples. Note that only the means are implemented as modifiers in this work. Bottom panels: comparisons of the virial mass, gas mass, and stellar mass functions (from left to right) at  $z = 5$  and  $13$  between three SAM *NOSN\_NOZCOOL\_NoRe* runs with fixed parameters but different implementations of modifiers: (1) no modifiers (*SAM*); (2) halo mass modifier only (*SAM\_H*); and (3) two modifiers including halo mass and baryon fraction (*SAM\_HB*). Shaded regions represent the  $1\sigma$  Poisson uncertainties. For comparison, the halo mass functions predicted by *NOSN\_NOZCOOL\_NoRe Smaug* hydrodynamic simulation and the *N-body* simulation are indicated by the thin black solid and dotted lines, respectively.

and slows their growths compared to a collisionless universe. This has been illustrated using comparisons between hydrodynamic simulations and *N-body* simulations of dark matter only (Sawala et al. 2013; Schaller et al. 2015; Velliscig et al. 2014), which show that the inclusion of baryons significantly reduces halo masses<sup>10</sup> (e.g.  $\sim 20$  per cent for haloes around  $10^{11} M_{\odot}$  at  $z \sim 0$ ).

We investigated the same baryonic effect in dwarf galaxies at high redshift ( $z > 5$ ) in Qin et al. (2017a, hereafter Paper-VIII). We found that the reduction of mass can be up to a factor of 2 and that the fraction of baryons in haloes with masses between  $10^7$  and  $10^9 M_{\odot}$ , which host dwarf galaxies, never exceeds 90 per cent of the cosmic mean ( $\Omega_b/\Omega_m$ ) during reionization. Thus, applying

SAMs to halo merger trees generated from a collisionless *N-body* simulation and assuming the baryon fraction of every virialized system is  $\Omega_b/\Omega_m$  overestimates halo and baryon masses of dwarf galaxies at high redshift, and suggests the necessity of incorporating modifications to SAMs. We proposed two modifiers using a simplified hydrodynamic simulation in Paper-VIII – *ADIAB*, where gas only cools adiabatically with no stellar or galactic physics involved. Note that the halo mass modifier was calculated as the mass ratio of matched haloes (see matching between galaxies in Appendix A) between *ADIAB* and *DMONLY*, while the baryon fraction modifier is the baryon fraction of these haloes in units of  $\Omega_b/\Omega_m$  calculated from the *ADIAB* simulation. In this work, which aims to inform a better semi-analytic description of high-redshift dwarf galaxies, we make use of these two modifiers, which are shown in the top panels of Fig. 2 as functions of halo mass at  $z = 13$ – $5$ . In practice, the halo

<sup>10</sup> We refer to the halo mass as the total mass of dark matter and baryons.

mass modifier is included when halo properties are read from the merger trees and the baryon fraction modifier is embedded in  $f_{\text{mod}}$  (see equation 1).

However, we also point out the fact that these modifiers do not offer a consistent modification<sup>11</sup> accounting for all aspects of astrophysical processes. In [Paper-VIII](#), these have been shown to provide additional  $\sim 10$  per cent alterations to halo mass due to radiative cooling, star formation and supernova feedback. Reionization, on the other hand, provides a more dramatic suppression, but can be incorporated through more efficient reionization feedback, which will be further discussed in the companion paper.

In order to demonstrate the effect of incorporating the modifiers within semi-analytic galaxy formation modelling, we apply `MERAXES` with different implementations. We first apply the SAM with both the halo mass and baryon fraction modifiers implemented and calibrate the free parameters in the `NOSN_NOZCOOL_NoRe` regime where feedback is not included (`SAM_HB`). The detailed calibration strategy is addressed in the next section. Next with all free parameters remaining the same, we then apply the SAM with only the halo mass modifier included (`SAM_H`) and without any modifiers (`SAM`). We show the halo mass, gas mass, and stellar mass functions at  $z = 5$  and 13 of the three SAM results in the bottom panels of [Fig. 2](#). The halo mass functions predicted by the *Smaug* simulations are also included in the left subpanels. Note that matching of individual systems (see [Appendix A2](#)) is not performed in this section. We see that compared to the hydrodynamic simulations (black solid lines), the collisionless  $N$ -body simulation (black dotted lines) as well as the uncorrected SAM overestimate the virial mass function, especially at higher redshifts. However, the halo mass modifier results in a SAM virial mass function that is in better agreement with the hydrodynamic calculation. This leads to a reduction as much as a factor of 2 in of the halo number density at  $z \sim 12$  for a given mass, and since the growth of the system becomes suppressed, it harbours fewer baryons (see [Section 2.1.1](#)) and results in suppressed mass functions of gas and stars. In addition, the baryon fraction modifier due to hydrostatic pressure causes a further reduction of baryons for a given halo mass, which subsequently causes less massive stellar masses.

To summarize, the mass ratio modifier can be interpreted as providing slower evolution of haloes when considering the hydrostatic pressure from baryons, while the baryon fraction modification accounts for the fact that the cosmic mean,  $\Omega_b/\Omega_m$ , cannot be achieved for haloes hosting dwarf galaxies – gas accretion is less efficient.

#### 4 COMPARISON BETWEEN DWARF GALAXIES IN SAM AND HYDRODYNAMIC SIMULATIONS

In this section, we present the comparison of accretion and cooling between `MERAXES` and *Smaug*. We note that in this section, if not otherwise stated,

(1) for the purpose of performing fair comparisons of the gas reservoir, we define star-forming gas, non-star-forming gas, and hot gas in hydrodynamic simulations following the method introduced in [Sections 2.2.3 and 2.2.4](#);

(2) for the purpose of facilitating direct comparisons, we match each individual galaxy in the `MERAXES` and *Smaug* outputs using the method presented in [Appendix A2](#). The free parameters in the SAM, which are summarized in [Table 1](#), are calibrated

**Table 1.** A summary of different SAM results with values of the main parameters used and modified in this work.

Name	Halo mass modifier (Section 3)	Baryon fraction modifier (Section 3, equation 1)	Star formation efficiency – $\alpha_{\text{sf}}$ (Section 4.1, equation 5)	Maximum cooling factor – $\kappa_{\text{cool}}$ (Section 4.3, equation 14)
<i>SAM</i>	×	×	0.05	1
<i>SAM_H</i>	✓	×	0.05	1
<i>SAM_HB</i>	✓	✓	0.05	1
<i>SAM_HBS</i>	✓	✓	$0.05 \times \left(\frac{1+z}{6}\right)^{-1.3}$	1
<i>SAM_HBSC</i>	✓	✓	$0.05 \times \left(\frac{1+z}{6}\right)^{-1.3}$	$\min\left(5, \frac{1+z}{6}\right)$

to reproduce the  $z = 5$  stellar mass function predicted by the `NOSN_NOZCOOL_NoRe` hydrodynamic simulation;

(3) for the purpose of minimizing the impact from mismatch and central-satellites switching (see [Appendix A1](#)), in the SAM, central galaxies with  $M_{\text{vir}} < 10^{7.5} M_{\odot}$  as well as all satellites<sup>12</sup> are excluded from the final matched galaxy sample;

(4) for the purpose of providing more consistent halo properties between `MERAXES` and *Smaug*, the two modifiers (i.e. halo mass and baryon fraction; see `SAM_HB` in [Section 3](#)) have been implemented in the SAM.

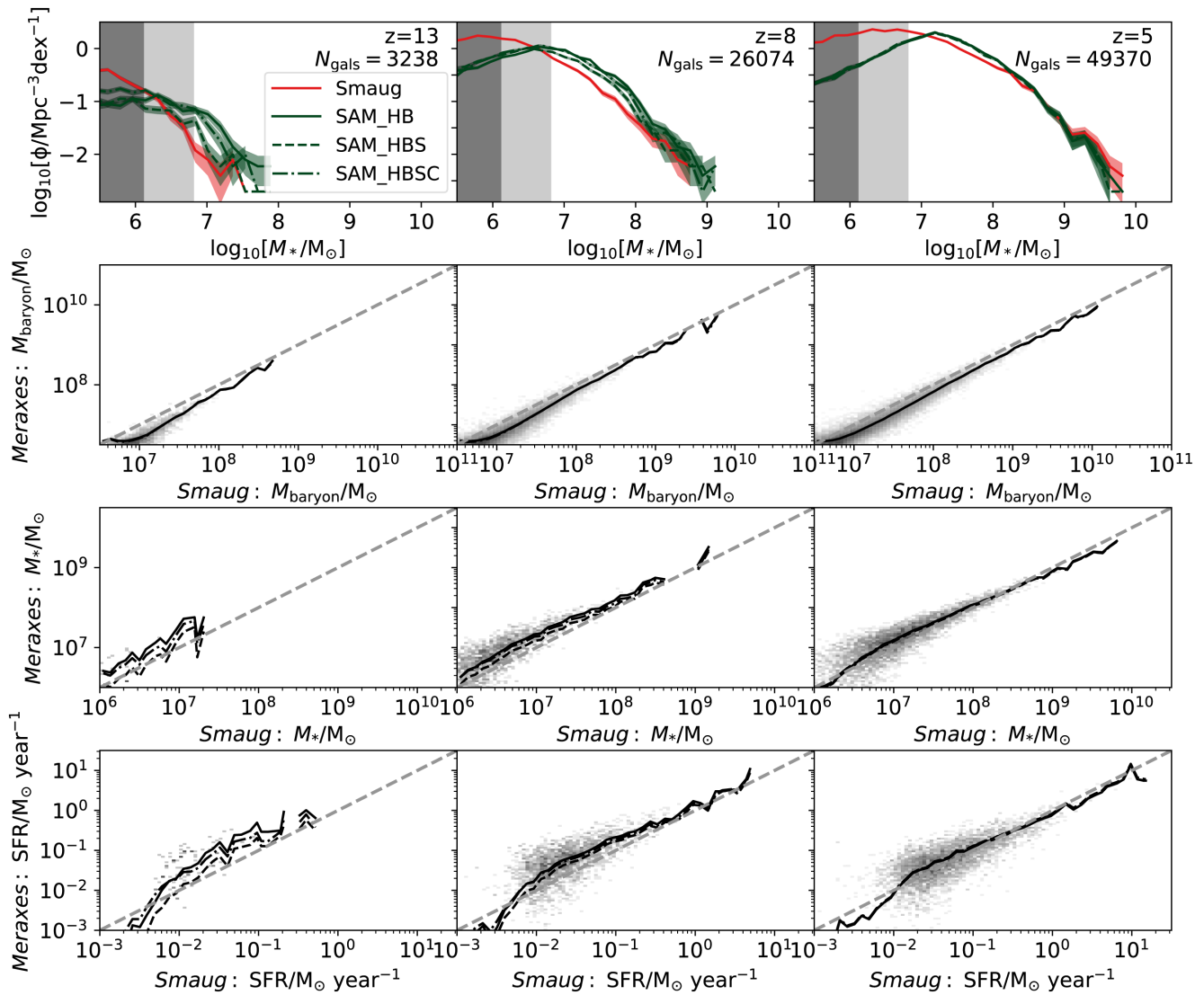
The first row of [Fig. 3](#) presents the stellar mass function of matched galaxies between `SAM_HB` and *Smaug* at  $z = 13$ –5. We see that the two  $z = 5$  mass functions are in agreement at  $M_* > 10^7 M_{\odot}$ , below which the models are limited by resolution and cooling mechanism. This indicates that the two modelled galaxy catalogues comprise similar stellar components at  $z = 5$  in a cosmological context. However, the results diverge towards higher redshifts – `SAM_HB` produces more massive galaxies than *Smaug*. This highlights the issue that stellar build-up proceeds faster in the SAM. As a consequence of the high-redshift modelling, the SAM produces more ionizing photons<sup>13</sup> at earlier times compared to the hydrodynamic calculation, overestimating the contribution of high-redshift dwarf galaxies to reionization ([Liu et al. 2016](#)).

We investigate the properties of individual galaxies in the two models, and show the comparisons of the total baryon mass, stellar mass, and SFR predicted by the hydrodynamic simulation and SAM in the bottom three rows of [Fig. 3](#). The grey 2D histograms and black solid lines represent the distribution and mean of all matched galaxies, respectively. We see that the overall baryon mass is underestimated by the SAM, with the offset becoming larger at the low-mass end. This is because the `NOSN_NOZCOOL_NoRe` *Smaug* simulation predicts larger halo masses and baryon fractions than `ADIAB` (where only adiabatic cooling of gas is included; see [Section 3](#)) due to cooling and star formation ([Paper-VIII](#)). Therefore, modifying the halo mass and baryon fraction using the `ADIAB` *Smaug* simulation overestimates the baryonic effect when feedback is not included. In addition, we see that both stellar mass and SFR are similar between the two models at  $z = 5$ , which are, however,

<sup>12</sup> Satellite haloes in the SAM do not receive any fresh gas, due to the assumption that their central haloes dominate the gravitational potential, which might not be the case in hydrodynamic simulations.

<sup>13</sup> Assuming other relevant factors including the escape fraction remain the same.

<sup>11</sup> Modifications on subgroup properties are not considered.



**Figure 3.** First row: the stellar mass functions at  $z \sim 13$ –5 in the *SAM\_HB* and *Smaug NOSN\_NOZCOOL\_NoRe* results. Shaded regions represent the  $1\sigma$  Poisson uncertainties. The number of matched galaxies,  $N_{\text{gals}}$  is indicated in the top right corner of each panel in the first row. The *MERAXES* results with a redshift-dependent star formation efficiency (*SAM\_HBS*) and redshift-dependent star formation efficiency and cooling rate (*SAM\_HBSC*) are shown with green dashed and dashed–dotted lines, respectively. The two grey regions mark the approximate resolution of the simulation, which correspond to 10 and 50 stellar particles, respectively. Second row: comparison of the baryon mass between *SAM\_HB* and *Smaug*. The grey 2D histograms indicate the distribution of the entire sample and the black solid lines represent the mean. The grey dashed lines indicate the quantities are identical in the two results. Bottom two rows: comparisons of the stellar mass and SFR between *Smaug* and *SAM\_HB*. The comparisons between *SAM\_HBS*, *SAM\_HBSC* and *Smaug* are indicated with dashed and dashed–dotted lines, respectively.

underestimated in the SAM result when the stellar mass approaches  $10^6 M_{\odot}$  or SFR reaches  $10^{-2} M_{\odot} \text{yr}^{-1}$ . These objects are close to or below the atomic cooling threshold, where cooling becomes inefficient. For these objects, while a fraction of gas particles in the hydrodynamic simulation are still dense enough to form stars, star formation is quenched in the SAM due to the lack of replenishment of cold gas (see Section 2.1.3). This causes an underestimation of the number of ionizing photons at the high redshifts when the first galaxies start to ionize the IGM.

We next discuss the involved prescriptions that might induce such discrepancies in the history of star formation. We note that in the *NOSN\_NOZCOOL\_NoRe* case when feedback is not implemented, there are only two processes that might cause the different stellar growth histories between SAMs and hydrodynamic simula-

tions: (1) star formation efficiency; and (2) gas fuelling and cooling efficiencies.<sup>14</sup>

#### 4.1 Star formation efficiency

We note again that since star formation is not resolved in cosmological simulations, both the hydrodynamic simulation and SAM

<sup>14</sup> In this work, we keep the original cooling functions of primordial elements adopted in *Paper-III* and Duffy et al. (2010). The difference in the two cooling curves (i.e. *MAPPING II* for the SAM, Sutherland & Dopita 1993 and *CLOUDY* for the hydrodynamic simulations, Ferland et al. 1998) is minor and does not have a significant impact on our results.

start from an empirical relation between SFR and density – the KS law (Kennicutt 1998), which proposes that a galaxy is able to form stars when its surface density ( $\Sigma_g$ ) exceeds a critical value ( $\Sigma_{\text{crit}} \sim \text{a few } M_{\odot} \text{pc}^{-2}$ ), and the surface SFR density ( $\dot{\Sigma}_*$ ) can be estimated by

$$\frac{\dot{\Sigma}_*}{(2.5 \pm 0.7) \times 10^{-4} M_{\odot} \text{yr}^{-1} \text{pc}^{-2}} = \left( \frac{\Sigma_g}{1 M_{\odot} \text{kpc}^{-2}} \right)^n, \quad (7)$$

where  $n = 1.4 \pm 0.15$  is suggested by observations (Kennicutt 1998). We have shown that cooling is not well modelled in the SAM when the virial temperature is lower than the atomic cooling threshold, which leads to significant underestimations of the number of low stellar mass and SFR objects in Fig. 3. However, this cannot explain the discrepancy of star formation history for more massive objects. In this subsection, we further probe the star formation laws utilized in MERAXES and *Smaug*.

#### 4.1.1 Star formation in hydrodynamic simulations

In the hydrodynamic simulation, local 3D densities,  $\rho_g$  can be calculated through smoothing particle mass,  $m_g$  within a certain volume. *Smaug* converts the critical surface density to three dimensions ( $\rho_{\text{crit}}$ ) assuming a self-gravitating disc. The gas on the disc is determined as multiphase ISM with an effective ratio of specific heats,  $\gamma_{\text{eff}} = 4/3$ . Then the pressure of the disc can be derived from the equation of state. When the local pressure of a gas particle with a relatively low temperature ( $< 10^5 \text{K}$ ) exceeds the critical value (see Section 2.2.3), the particle is considered as a potential star-forming region. This infers an SFR of (see more details in Schaye & Dalla Vecchia 2008)

$$\dot{m}_{*,\text{hydro}} \equiv \frac{m_g}{t_g} \equiv m_g \frac{\dot{\Sigma}_*}{\Sigma_g} \propto m_g \Sigma_g^{(n-1)} \propto m_g \rho_g^{0.5(n-1)\gamma_{\text{eff}}}, \quad (8)$$

where  $t_g \equiv \Sigma_g / \dot{\Sigma}_*$  is the gas depletion time-scale, the third step makes use of equation (7) and the last step assumes that  $\Sigma_g$  is of the order of the Jeans column density in the case of self-gravitating disc. With a further assumption that the disc follows an isothermal exponential surface density profile (Schaye 2004)

$$\rho_g = \frac{GM_{\text{disc}}^2}{12\pi c_s^2 R_{\text{disc}}^4}, \quad (9)$$

where  $M_{\text{disc}} \propto M_{\text{vir}}$ ,  $R_{\text{disc}} \propto R_{\text{vir}}$ , and  $c_s$  are the mass, radius, and sound speed of the disc, respectively, we find that for a given halo mass  $M_{\text{vir}}$ ,  $\rho_g \propto (1+z)^4$ , and hence

$$\dot{m}_{*,\text{hydro}} \propto (1+z)^{2\gamma_{\text{eff}}(n-1)} \propto (1+z)^{1.1}. \quad (10)$$

#### 4.1.2 Star formation in SAMs

On the other hand, the average SFR of a galaxy in the SAM is calculated using equation (5). For a given halo mass and with  $m_{\text{cold}} \propto M_{\text{vir}}$ , the redshift dependency of  $\dot{m}_{*,\text{SAM}}$  is approximately

$$\dot{m}_{*,\text{SAM}} = \alpha_{\text{sf}} \times \frac{\max(0, m_{\text{cold}} - m_{\text{crit}})}{t_{\text{dyn,disc}}} \propto (1+z)^{1.5}. \quad (11)$$

Note that without supernova feedback in the SAM, the fraction of galaxies that comprise a massive cold gas reservoir (i.e.  $m_{\text{cold}} > m_{\text{crit}}$ ) increases towards higher redshifts and in more massive objects. This suggests that the index can be larger than 1.5 on average, especially for less massive haloes.

#### 4.1.3 Redshift-dependent star formation efficiency

Comparing equations (10) and (11), we find that although both numerical calculations of star formation are derived from the KS law, they possess different evolutionary histories for a given halo mass. This is essentially due to the different assumption of gas depletion time-scale adopted in the two modelling approaches, which is chosen to be the disc dynamical time in the SAM, while in the hydrodynamic simulation is inferred from observations. In order to be consistent with the hydrodynamic simulation, a suppressed star formation efficiency of  $\alpha_{\text{sf}} \propto (1+z)^{-m}$  towards higher redshifts is required in the SAM. We next further discuss the value of  $m$ .

(i) The scaling indices in equations (10) and (11) assume a disc profile for star-forming gas and a negligible cold gas threshold of star formation (i.e.  $m_{\text{cold}} \gg m_{\text{crit}}$ ) in the SAM, respectively. These infer  $m \sim 0.4$ .

(ii) However, at higher redshifts, galaxies have larger velocity dispersions, indicating increased turbulence and thickened discs (Newman et al. 2012; Price et al. 2016). Meanwhile, simulations suggest that mergers which happen frequently at high redshift (Poole et al. 2016) can also thicken discs (Moster et al. 2010, 2012). Therefore, the aforementioned assumption of a self-gravitating disc might not be valid at high redshift. If we assume an SIS profile for star-forming gas,  $\rho_g$  scales  $(1+z)^3$ , which leads to a larger value of  $m \sim 0.7$ .

(iii) Furthermore, during the experiment we find that, in the SAM, galaxies are more likely to have an insufficient cold gas reservoir (i.e.  $m_{\text{cold}} < m_{\text{crit}}$ ) at lower redshifts. This requires a more suppressed star formation efficiency at higher redshift, and in practice,  $m = 1.3$  is adopted in this work.

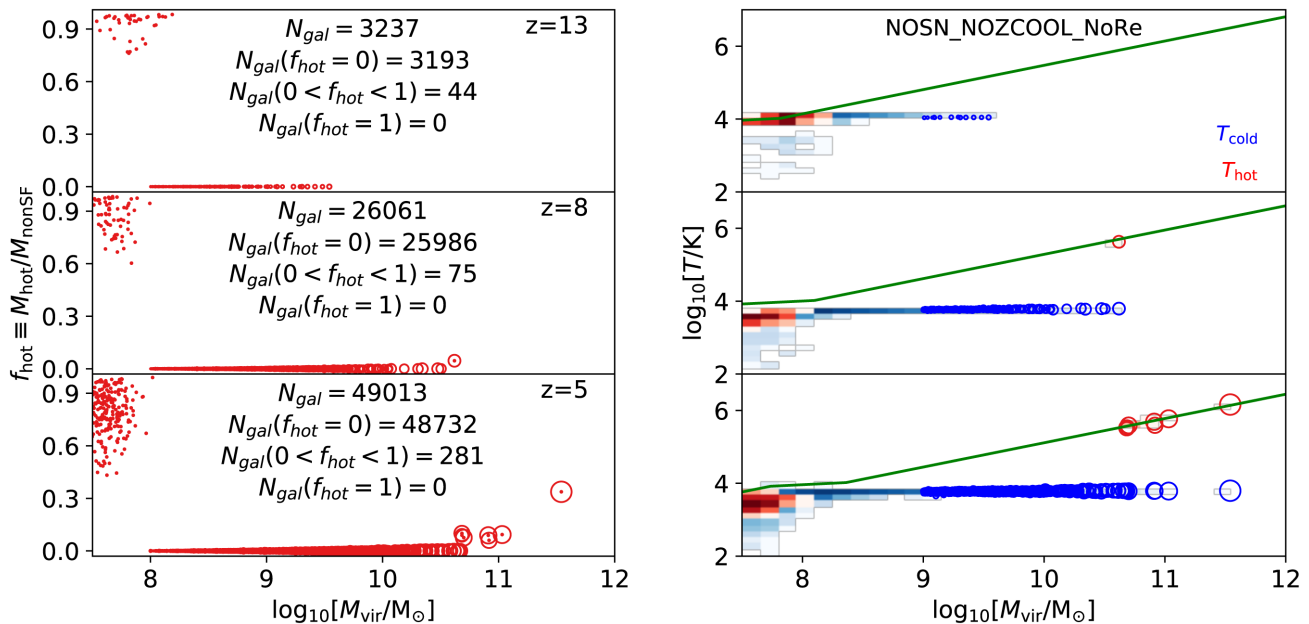
We note that since the star-forming gas profile varies with different implementations of physics including feedback,  $m$  is introduced as a free parameter modulating the global star formation history.

In Fig. 3, the semi-analytic result using the redshift-dependent star formation efficiency (*SAM\_HBS*) is presented using dashed lines. We see that suppressing star formation at higher redshifts results in a better agreement of the stellar mass function, stellar mass, and SFR with the hydrodynamic calculations. Note that in the absence of feedback, star formation is not limited by the gas reservoir of galaxies in the SAM (see equation 5). Therefore, updating the star formation efficiency does not have a significant impact on the gas component discussed in the next subsections.

## 4.2 Gas fuelling

#### 4.2.1 Hot gas in hydrodynamic simulations

We define hot gas fraction as the mass ratio of the hot gas (see Section 2.2.4) to the non-star-forming gas (see Section 2.2.3;  $f_{\text{hot}} \equiv M_{\text{hot}}/M_{\text{nonSF}}$ ), and show its correlation with halo mass in *Smaug* at  $z = 13\text{--}5$  in the left-hand panel of Fig. 4. We highlight galaxies above the atomic cooling threshold with red circles (circle size representing the stellar mass) and we see that these galaxies in general comprise little or no hot gas. We show the numbers of all galaxies, galaxies with  $f_{\text{hot}} = 0$ ,  $0 < f_{\text{hot}} < 1$  and  $f_{\text{hot}} = 1$  in each subpanel, and indicate galaxies with  $f_{\text{hot}} \neq 0$  with red points. We see that the majority of galaxies at high redshift do not possess any hot gas, and their non-star-forming gas particles are identified as one group – the cold non-star-forming gas. However, due to the lack of molecular cooling, galaxies with  $M_{\text{vir}} \lesssim 10^8 M_{\odot}$  do comprise two components. In the right-hand panel of Fig. 4,



**Figure 4.** Left-hand panel: the fraction of hot gas in the non-star-forming gas ( $f_{\text{hot}} \equiv M_{\text{hot}}/M_{\text{nonSF}}$ ) from the *NOSN\_NOZCOOL\_NoRe Smaug* simulation at  $z = 13$ – $5$  as a function of the halo mass ( $M_{\text{vir}}$ ). Galaxies with  $f_{\text{hot}} \neq 0$  are indicated with red dots. Galaxies with  $M_{\text{vir}} > 10^8 M_{\odot}$  are emphasized with red circles with circle size representing the stellar mass ( $M_{\star}$ ). In this mass range, galaxies comprise no or little hot gas. The numbers of all galaxies, galaxies with  $f_{\text{hot}} = 0$ ,  $0 < f_{\text{hot}} < 1$  and  $f_{\text{hot}} = 1$  are shown in each subpanel. Right-hand panel: gas temperature as a function of  $M_{\text{vir}}$ . The red and blue 2D histograms represent hot and cold non-star-forming gas in the *Smaug* simulation (see Section 2.2.4), while the green solid line indicates the correlation between the virial temperature and halo mass. Galaxies with  $M_{\text{vir}} > 10^9 M_{\odot}$  are also indicated by the red (hot gas) and blue (cold non-star-forming gas) circles. Fewer than 10 galaxies in our simulation comprise hot gas at  $z = 5$ , and their hot gas has been shock heated to the virial temperature.

we show the 2D histograms of the hot (red) and cold (blue) non-star-forming gas temperatures as functions of halo mass in *Smaug*. We see that for galaxies with  $M_{\text{vir}} \lesssim 10^8 M_{\odot}$ , their hot and cold non-star-forming gas particles<sup>15</sup> possess median temperatures  $> 10^3$  and  $< 10^3$  K, respectively. This broad distribution on the density–temperature phase diagram is from adiabatic cooling of the gas as the universe expands, which is also indicated by the decreasing median temperature at lower redshifts of these less massive objects.

#### 4.2.2 Hot gas in SAMs

Gas in the IGM falls into the gravitational potential of a halo and fuels star formation in the host galaxy (see Section 2.1.1). Although the hydrodynamic simulation and SAM both assume ionization equilibrium for the accreted gas, the temperature profile varies significantly. While gas particles cover a broad range of temperatures from  $\sim 10^3$  to  $10^8$  K (Wiersma et al. 2009b) in the hydrodynamic simulation, (hot) gas initially accreted by a galaxy in the SAM is assumed to be shock heated to the virial temperature (see Section 2.1.2). Hot gas then cools through thermal radiation and builds a cold gas disc where stars can form (see Section 2.1.3). In the *NOSN\_NOZCOOL\_NoRe* regime where metals cooling is not included, temperature alone (assuming the helium fraction is a constant) determines the cooling rate (see equation 2) and therefore alters star formation in each galaxy. In order to test the validity of this

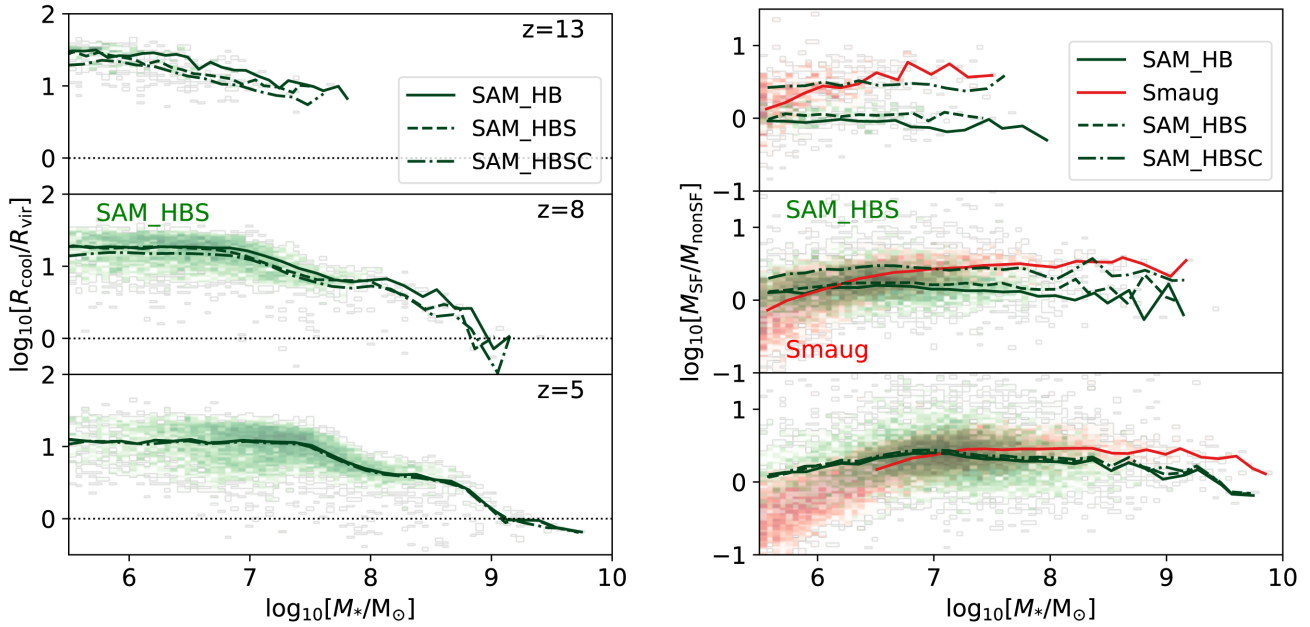
simplified gas infall prescription for modelling of dwarf galaxies, we compare the temperature of non-star-forming gas in the hydrodynamic simulation to the halo virial temperature. In the right-hand panel of Fig. 4, we highlight galaxies with  $M_{\text{vir}} > 10^9 M_{\odot}$  in *Smaug* with circles and show the correlation between virial temperature and halo mass with green solid lines. We see that  $\lesssim 10$  galaxies with  $M_{\text{vir}} > 10^9 M_{\odot}$  in the *Smaug* simulation comprise hot gas at  $z = 5$ , and their hot gas temperatures are close to the virial temperatures, suggesting that these galaxies have been heated through shocks. However, the non-star-forming gas of most galaxies in *Smaug* is considered as cold with  $T \lesssim 10^4$  K, which is much lower than the virial temperature and becomes more common at higher redshifts. This is consistent with the prediction of EAGLE simulations (Schaye et al. 2015), where the mass ratio of hot<sup>16</sup> to all baryons is less than 1 per cent on average for haloes around  $10^{10} M_{\odot}$  (Correa et al. 2017), with the fraction decreasing towards higher redshifts and in less massive haloes.

#### 4.2.3 Gas accretion in cold mode

We note that in the SAM, cold and hot gas, where stars can and cannot form, are defined geometrically. They represent a star-forming disc and the outer region, following an exponential and an SIS profile, respectively. However, we show that full virialization of infall gas might be problematic for high-redshift modelling of dwarf galaxy formation. In these less massive systems, virial shocks might not form and the infall gas could then only reach the virial temperature when it arrives at the star-forming disc (Birnboim & Dekel

<sup>15</sup> Note that with a smaller threshold of the temperature offset between hot and cold non-star-forming gas ( $T_{\text{crit}}$ ; see Section 2.2.4), more particles that belong to haloes with  $M_{\text{vir}} \lesssim 10^8 M_{\odot}$  are identified as hot. However, the temperature thresholds (i.e.  $T_{\text{crit}}$  and  $\Delta T_{\text{crit}}$ ) have less impact on more massive haloes, and their  $f_{\text{hot}}$  remains zero or small (see Appendix B).

<sup>16</sup> In Correa et al. (2017), hot gas is defined as particles with cooling time longer than the dynamical time.



**Figure 5.** Left-hand panel: the ratio of the cooling radius to the virial radius as a function of stellar mass ( $M_*$ ) from the SAM\_HB (solid line), SAM\_HBS (dashed line) and SAM\_HBSC (dash-dotted line) results at  $z = 13$ –5. Lines represent the median while the 2D histogram shows the distribution in SAM\_HBS. Right-hand panel: the mass ratio of star-forming gas to non-star-forming gas ( $M_{\text{SF}}/M_{\text{nonSF}}$ ) as a function of  $M_*$ . The red and green 2D histograms represent galaxies in the Smaug and SAM\_HBS results, with the medians shown with red solid and green dashed lines, respectively. The median  $M_{\text{SF}}/M_{\text{nonSF}}$  from SAM\_HB and SAM\_HBSC are shown with green solid and dashed–dotted lines for comparison.

2003; Cattaneo et al. 2017). Instead, these cold and less dense non-star-forming gas particles of dwarf galaxies in Smaug represent a cold accretion mode (Kereš et al. 2005, 2009; Benson & Bower 2011; Correa et al. 2017). Therefore, in the SAM, instead of being fully virialized, the infall gas of dwarf galaxies should be separated into hot and cold components, which reach the disc and contribute to star formation on different time-scales (i.e. thermal cooling time,  $t_{\text{cool}}$  in equation 2, and dynamical time,  $t_{\text{dyn}}$  in equation 4).

In addition, since the non-star-forming gas particles of the majority of high-redshift dwarf galaxies are *not* hot in the hydrodynamic simulation, we advocate that the terminology of hot gas becomes misleading in the redshift and mass ranges discussed in this work.

### 4.3 Rapid cooling

The gas infall and cooling prescription adopted for the SAM distinguishes two regimes – a static hot halo and rapid cooling gas (see Section 2.1.2). In the second regime, gas cools on to the disc at the dynamical time-scale, which has the properties of cold accretion if one combines the process of gas infall and rapid cooling. Therefore, altering the rapid cooling rate is an alternative to incorporating cold accretion<sup>17</sup> directly (Cattaneo et al. 2017).

The relative importance of the two cooling modes is determined by the ratio of the cooling radius to the virial radius ( $R_{\text{cool}}/R_{\text{vir}}$ ), which is shown in the left-hand panel of Fig. 5. We see that galaxies with  $M_* < 10^9 M_{\odot}$  are mostly in the rapid cooling regime with  $R_{\text{cool}}/R_{\text{vir}} > 1$ . The determination of static hot halo and rapid cooling regimes has been discussed in Croton et al. (2006). Comparing with hydrodynamic simulations, they found  $M_{\text{vir}} \sim 10^{11} M_{\odot}$  separates the two regimes, which is approximately independent with redshift

up to  $z \sim 6$  (see also Correa et al. 2017). This is in agreement with our finding<sup>18</sup> of  $M_* \sim 10^9 M_{\odot}$ .

The balance of instantaneous cooling rate between the two numerical approaches can be inferred from the evolution of the mass ratio of star-forming to non-star-forming gas (see Appendix C for more information of galaxy properties in the SAM). This is shown in the right-hand panel of Fig. 5. We see that while the SAM\_HBS result is in agreement with the hydrodynamic calculation at low redshift,  $M_{\text{SF}}/M_{\text{nonSF}}$  in the SAM is underestimated at higher redshifts. We have shown that in the SAM, dwarf galaxies at high redshift are identified as being in the rapid cooling regime. Therefore, in order to understand the different transition rates between non-star-forming reservoir and star-forming gas in MERAXES and Smaug, we next discuss the assumed SIS profile of hot gas reservoir in the SAM.

#### 4.3.1 Gas profile

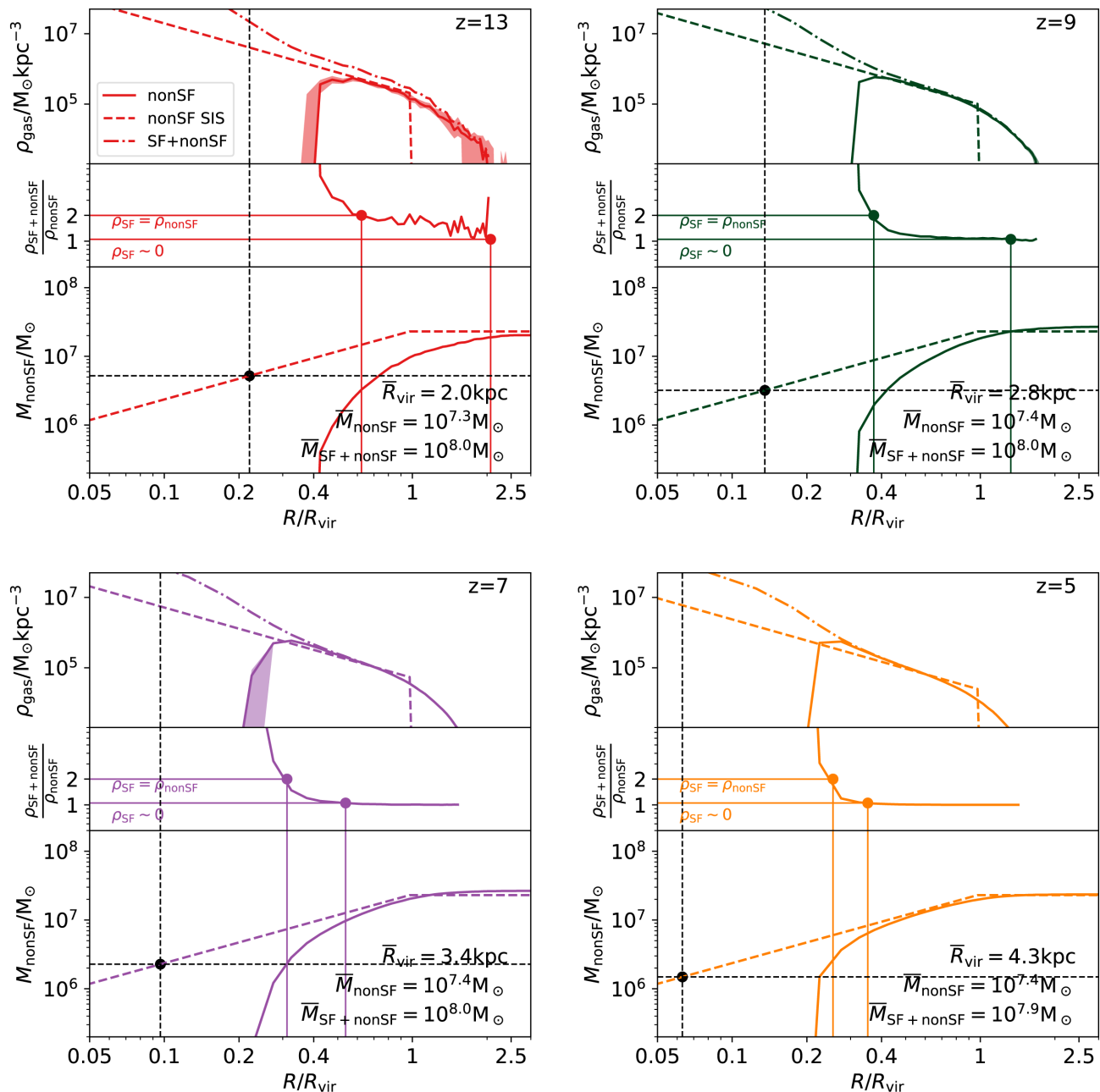
We present the median radial profile of the non-star-forming gas of galaxies with  $M_* = 10^{7 \pm 0.5} M_{\odot}$  in Smaug at  $z = 13$ –5 in Fig. 6. We see that the density of non-star-forming gas drops at the inner region (where particles become star forming) when compared to an SIS mass profile that follows (Croton et al. 2006)

$$\mathcal{M}_{\text{nonSF}}(r) = M_{\text{nonSF}} \times \min\left(\frac{r}{R_{\text{vir}}}, 1\right), \quad (12)$$

the non-star-forming gas particles (1) have a large dispersion up to  $\gtrsim 2R_{\text{vir}}$ ; and (2) are less concentrated at higher redshifts. The SAM assumes the SIS profile for hot gas. Therefore, in the rapid cooling

<sup>17</sup> We leave the task of implementing cold accretion in a future project when molecular cooling is included.

<sup>18</sup> Although we are now focusing on a simplified scenario where supernova and metals are not included, it is also true when the feedback is implemented as we will show in the companion paper.



**Figure 6.** Gas profiles of galaxies with  $M_* \sim 10^{7 \pm 0.5} M_\odot$  at  $z = 13$ –5 in the *NOSN\_NOZCOOL\_NoRe Smaug* simulation (coloured lines). In each panel, top: the median radial density profiles of all gas (i.e. star-forming and non-star-forming gas; dashed–dotted line) and the non-star-forming gas (solid line). Lines with shaded regions represent the median and 95 per cent confidence intervals around the median using 100 000 bootstrap re-samples of the non-star-forming gas profile. The median SIS profile assumed in the SAM is calculated using the same amount of non-star-forming gas, and is indicated with the coloured dashed line; middle: the ratio of the density profiles of all gas to non-star-forming gas; bottom: the median radial mass profile of non-star-forming gas (solid line) compared to the SIS profile (dashed line). The radius, within which the SIS gas is able to reach the centre through free-fall after one time-step in the SAM, and the enclosed SIS gas mass are indicated with thin vertical and horizontal dashed lines, respectively. The radii, where the star-forming gas is as dense as the non-star-forming gas (i.e.  $\rho_{\text{SF}} = \rho_{\text{nonSF}}$ ) or becomes deficient (i.e.  $\rho_{\text{SF}} \sim 0$ ), are indicated with thin solid lines. The median virial radius, masses of non-star-forming, and all gas mass are shown in the bottom right corner of each panel.

regime (see equation 4), the cooling mass during one time-step ( $\Delta t$ ) can be calculated through

$$M_{\text{cool}} = M_{\text{nonSF}} \times \frac{\Delta t}{t_{\text{dyn}}} \equiv \mathcal{M}_{\text{nonSF}} \left( r = R_{\text{vir}} \times \frac{\Delta t}{t_{\text{dyn}}} \right). \quad (13)$$

The last step implies there is a radius, within which gas is able to reach the *centre* through free-fall. This is indicated with ver-

tical dashed lines in Fig. 6 at different redshifts. For comparison, we also present the density profile of all gas of the galaxies with  $M_* = 10^{7 \pm 0.5} M_\odot$  in Fig. 6 (coloured dashed–dotted lines), and indicate the radii, at which the star-forming gas is as dense as the non-star-forming gas (i.e.  $\rho_{\text{SF}} = \rho_{\text{nonSF}}$ ) or becomes deficient (i.e.  $\rho_{\text{SF}} \sim 0$ ), using vertical solid lines. We see that the star-forming gas also possesses a large dispersion. At

high redshift ( $z > 9$ ),  $r(\rho_{\text{SF}} = 0)$  can be larger than the virial radius.

#### 4.3.2 Transition radius of gas reservoir

We see that the SIS profile significantly overestimates the non-star-forming gas density in the inner region compared to the simulation result. Therefore, with the assumption that gas can only transform from non-star-forming to star-forming (or hot to cold in the SAM) gas when collapsing into the centre, the SIS profile leads to an overestimation of the inward collapse rate (see the horizontal dashed line in Fig. 6). However, due to the large radial region of star-forming gas observed in Fig. 6, this assumption (i.e. transition radius is  $\sim 0$ ) might not be accurate. Gas particles in dwarf galaxies at  $z > 9$  can be triggered as star-forming regions as far as the virial radius. The combination of corresponding effects between the overestimated collapse rate from the assumed SIS profile and the underestimated transition radius of gas reservoir is the primary cause of the discrepancy in  $M_{\text{SF}}/M_{\text{nonSF}}$  between MERAXES and *Smaug*.

#### 4.3.3 Redshift-dependent maximum cooling factor

Whilst adopting the correct gas profile and finding the accurate transition radius of gas reservoir can help explain the underestimation of  $M_{\text{SF}}/M_{\text{nonSF}}$  in the SAM at high redshift, an analytic solution is not well defined. The large dispersion in the gas profile is mainly due to frequent mergers at these redshifts, which can be affected by numerical configurations, such as particle resolution, as well as by physics implementations including feedback. In order to take the underestimated cooling rate of dwarf galaxies at high redshift into account, we propose an alteration to equation (4) as follows

$$\dot{m}_{\text{cool}} = \frac{m_{\text{hot}}}{t_{\text{dyn}}} \times \min \left( \kappa_{\text{cool}}, \frac{r_{\text{cool}}}{R_{\text{vir}}} \right), \quad (14)$$

and allow the free parameter,  $\kappa_{\text{cool}}$  representing the maximum cooling factor to exceed unity at high redshift<sup>19</sup>. We show the result of  $M_{\text{SF}}/M_{\text{nonSF}}$  calculated with  $\kappa_{\text{cool}} = \min \left( 5, \frac{1+z}{6} \right)$  (*SAM\_HBSC*) in Fig. 5, and we see that with higher cooling rates, the semi-analytic calculation becomes more consistent with the hydrodynamic result at high redshift. However, the stellar mass function becomes higher at  $z = 13$  compared to the *Smaug* result (see Fig. 3), suggesting the necessity of stronger suppression of the star formation efficiency (i.e.  $m > 1.3$ ) due to the relatively larger disc mass at higher redshifts in the SAM (see Section 4.1.2).

Lastly, we point out that without a self-consistent radiative transfer calculation, atomic cooling, and gas temperature might not be properly simulated in the hydrodynamic simulation either, which alters the importance of thermal radiation cooling. For instance, with local sources (of each gas particle) of ionizing radiation neglected, the cooling rate is potentially overestimated (Schaye et al. 2015), and since self-shielding cannot be properly captured in our hydrodynamic simulations, the cooling rate might be underestimated in dense regions (McQuinn, Peng Oh & Faucher-Giguère 2011). Moreover, supernova feedback and reionization are expected to regulate galaxy formation through altering the gas component, which will be the topic of a forthcoming paper.

<sup>19</sup> The other interpretation of  $\kappa_{\text{cool}}$  is to modulate the time-scale (i.e.  $t_{\text{inflow}} \equiv \kappa_{\text{cool}}^{-1} t_{\text{dyn}}$ ) of gas inflow from the circumgalactic medium to the ISM. See more in the companion paper.

## 5 CONCLUSIONS

While most assumptions adopted for semi-analytic galaxy formation models, including the gas reservoirs (i.e. the density profile, temperature, and transition), are in good agreement with hydrodynamic simulations for Milky Way size objects (Guo et al. 2016; Stevens et al. 2017b), they become less accurate for less massive galaxies at high redshift. In this work, we propose modifications to SAMs based on the comparison of dwarf galaxy properties calculated by the MERAXES SAM and the *Smaug* high-resolution hydrodynamic simulation. We focus on gas accretion, cooling, and star formation at  $z \geq 5$ , and consider scenarios in the absence of reionization and supernova feedback. We summarize the modifications below:

(1) The parent cosmological simulation of a SAM usually includes only collisionless particles, where baryonic physics is neglected. In making comparisons between  $N$ -body and hydrodynamic simulations starting from identical initial conditions, we previously showed (Paper-VIII) that dwarf galaxy host halo masses are significantly overestimated when hydrostatic pressure is not considered, and that the fraction of baryons accreted by dwarf galaxies cannot reach the level of cosmic mean. While inclusion of halo masses directly from  $N$ -body dark matter only simulation and assumption of a universal baryon fraction are standard features of SAMs, the impact of these assumptions becomes significant at high redshift and small scale. We have considered implementations to modify the relevant properties (i.e. the halo mass and baryon fraction) in SAMs (see Section 3).

(2) We find that the star formation prescription in the SAM that is based on the consumption of a cold gas reservoir does not represent the evolutionary path followed by gas in the hydrodynamic simulation, while both of the two modelling approaches start from the observational relation between surface density and SFR (Kennicutt 1998). We find that this results from variation in the calculated depletion time-scale of the gas reservoir. We address this by modifying the efficiency that modulates star formation in the SAM with a redshift dependency (i.e. replacing the constant  $\alpha_{\text{sf}}$  to  $\alpha_{\text{sf}}(z)$ )

$$\dot{m}_* = \alpha_{\text{sf}}(z) \times \frac{m_{\text{cold}} - m_{\text{crit}}}{t_{\text{dyn, disc}}}. \quad (15)$$

In this work, we adopt  $\alpha_{\text{sf}}(z) = 0.05 \times [(1+z)/6]^{-1.3}$ , which allows the model to follow the *NOSN\_NOZCOOL\_NoRe* hydrodynamic simulation in the *Smaug* suite (see Section 4.1).

(3) The majority of dwarf galaxies at high redshift are in the rapid cooling regime, where the infalling gas cannot form stable shocks or remain in hydrostatic equilibrium. This represents a cold accretion mode, which is not well modelled with the assumption that the hot gas reservoir follows the SIS profile and falls into the centre within the dynamical time. We find, in the hydrodynamic simulation, that gas in high-redshift dwarf galaxies can form stars as far as the virial radius and that the SIS profile overestimates the density in the inner regions of these low-mass objects. In order to take this into account, we propose of modulation of the cooling prescription with a redshift-dependent collapse rate (i.e. changing the cooling rate upper limit from  $m_{\text{hot}} t_{\text{dyn}}^{-1}$  to  $\kappa_{\text{cool}}(z) m_{\text{hot}} t_{\text{dyn}}^{-1}$ )

$$\dot{m}_{\text{cool}} = \frac{m_{\text{hot}}}{t_{\text{dyn}}} \times \min \left[ \kappa_{\text{cool}}(z), \frac{r_{\text{cool}}}{R_{\text{vir}}} \right], \quad (16)$$

where  $\kappa_{\text{cool}}(z) = \min [5, (1+z)/6]$  leads to an agreement with the hydrodynamic calculation on the cold gas mass evolution (see Section 4.3).

Furthermore, we point out that the terminology of *hot gas* and *cold gas* becomes misleading when applying SAMs to the formation of high-redshift dwarf galaxies. The hot gas representing non-star-forming gas within a galaxy does not experience full virialization and possess a median temperature that is much lower than the virial temperature of the host halo (Section 4.2). In a future paper (Qin et al., in preparation), we will discuss star formation in the presence of feedback from reionization and supernovae. We will compare the SAM calculation with the most complete hydrodynamic simulation in the *Smaug* suite that considers reionization as an instantaneous heating background (Haardt & Madau 2001) and distributes supernova energy in thermal form (Dalla Vecchia & Schaye 2012). We will also consider an additional star formation prescription to the one shown in this work, in which molecular gas directly drives the star formation history (Lagos et al. 2011). The goal of these two papers will be more complete and faithful recreation of hydrodynamic simulations at high redshift, which will further leverage the observations that we can make in this early history of our Universe.

## ACKNOWLEDGEMENTS

This research was supported by the Victorian Life Sciences Computation Initiative, grant ref. UOM0005, on its Peak Computing Facility hosted at the University of Melbourne, an initiative of the Victorian Government, Australia. Part of this work was performed on the gSTAR national facility at Swinburne University of Technology. gSTAR is funded by Swinburne and the Australian Governments Education Investment Fund. This research was conducted by the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), through project number CE170100013. This work was supported by the Flagship Allocation Scheme of the NCI National Facility at the ANU, generous allocations of time through the iVEC Partner Share and Australian Supercomputer Time Allocation Committee. AM acknowledges support from the European Research Council under the European Union’s Horizon 2020 research and innovation program (Grant No. 638809 – AIDA).

## REFERENCES

Barkana R., Loeb A., 1999, *ApJ*, 523, 54  
 Baugh C. M., 2006, *Rep. Prog. Phys.*, 69, 3101  
 Baugh C. M., Lacey C. G., Frenk C. S., Granato G. L., Silva L., Bressan A., Benson A. J., Cole S., 2005, *MNRAS*, 356, 1191  
 Benson A. J., Bower R., 2011, *MNRAS*, 410, 2653  
 Bertschinger E., 1998, *ARA&A*, 36, 599  
 Bertschinger E., 2001, *ApJS*, 137, 1  
 Birnboim Y., Dekel A., 2003, *MNRAS*, 345, 349  
 Blitz L., Rosolowsky E., 2006, *ApJ*, 650, 933  
 Bouwens R. J. et al., 2015, *ApJ*, 803, 34  
 Bouwens R. J. et al., 2016, *ApJ*, 830, 67  
 Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., Lemson G., 2009, *MNRAS*, 398, 1150  
 Bullock J. S., Dekel A., Kolatt T. S., Kravtsov A. V., Klypin A. A., Porciani C., Primack J. R., 2001, *ApJ*, 555, 240  
 Cattaneo A. et al., 2017, *MNRAS*, 471, 1401  
 Chabrier G., 2003, *PASP*, 115, 763  
 Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, *MNRAS*, 319, 168  
 Correa C. A., Schaye J., Wyithe J. S. B., Duffy A. R., Theuns T., Crain R. A., Bower R., 2017, *MNRAS*, 473, 538  
 Côté B., Silvia D., O’Shea B. W., Smith B., Wise J. H., 2017, *ApJ*, preprint (arXiv:1710.06442)  
 Crain R. A. et al., 2009, *MNRAS*, 399, 1773  
 Croton D. J. et al., 2006, *MNRAS*, 365, 11

Dalla Vecchia C., Schaye J., 2008, *MNRAS*, 387, 1431  
 Dalla Vecchia C., Schaye J., 2012, *MNRAS*, 426, 140  
 De Lucia G., Blaizot J., 2007, *MNRAS*, 375, 2  
 Dolag K., Borgani S., Schindler S., Diaferio A., Bykov A. M., 2008, *Space Sci. Rev.*, 134, 229  
 Duffy A. R., Schaye J., Kay S. T., Dalla Vecchia C., Battye R. A., Booth C., 2010, *MNRAS*, 405, 2161  
 Duffy A. R., Wyithe J. S. B., Mutch S. J., Poole G. B., 2014, *MNRAS*, 443, 3435  
 Duffy A. R., Mutch S. J., Poole G. B., Geil P. M., Kim H.-S., Mesinger A., Wyithe J. S. B., 2017, *MNRAS*, 470, 3300  
 Feng Y., Di-Matteo T., Croft R. A., Bird S., Battaglia N., Wilkins S., 2016, *MNRAS*, 455, 2778  
 Ferland G., Korista K., Verner D., Ferguson J., Kingdon J., Verner E., 1998, *PASP*, 110, 761  
 Garrison L. H., Eisenstein D. J., Ferrer D., Tinker J. L., Pinto P. A., Weinberg D. H., 2017, *ApJS*, preprint (arXiv:1712.05768)  
 Geil P. M., Mutch S. J., Poole G. B., Angel P. W., Duffy A. R., Mesinger A., Wyithe J. S. B., 2016, *MNRAS*, 462, 804  
 Geil P. M., Mutch S. J., Poole G. B., Duffy A. R., Mesinger A., Wyithe J. S. B., 2017, *MNRAS*, 472, 1324  
 Guo Q. et al., 2011, *MNRAS*, 413, 101  
 Guo Q. et al., 2016, *MNRAS*, 461, 3457  
 Haardt F., Madau P., 2001, in Neumann D. M., Tran J. T. V., eds, XX1st Moriond Astrophys. Meeting, Clusters of Galaxies and the High Redshift Universe Observed in X-rays. Editions Frontieres, Paris, p. 64  
 Haas M. R., Schaye J., Booth C. M., Vecchia C. D., Springel V., Theuns T., Wiersma R. P. C., 2013, *MNRAS*, 435, 2931  
 Hatton S., Devriendt J. E. G., Ninin S., Bouchet F. R., Guiderdoni B., Vibert D., 2003, *MNRAS*, 343, 75  
 Henriques B. M., White S. D., Thomas P. A., Angulo R., Guo Q., Lemson G., Springel V., Overzier R., 2015, *MNRAS*, 451, 2663  
 Iliev I. T., Shapiró P. R., Mellema G., Merz H., Pen U.-L., 2008, Proceedings of the TeraGrid 2008 Conference. Las Vegas, USA, p. 31  
 Iliev I. T., Mellema G., Ahn K., Shapiró P. R., Mao Y., Pen U.-L., 2014, *MNRAS*, 439, 725  
 Katsianis A., Tescari E., Blanc G., Sargent M., 2017, *MNRAS*, 464, 4977  
 Kennicutt R. C., Jr, 1998, *ApJ*, 498, 541  
 Kereš D., Katz N., Weinberg D. H., Davé R., 2005, *MNRAS*, 363, 2  
 Kereš D., Katz N., Fardal M., Davé R., Weinberg D. H., 2009, *MNRAS*, 395, 160  
 Klypin A. A., Trujillo-Gomez S., Primack J., 2011, *ApJ*, 740, 102  
 Komatsu E. et al., 2011, *ApJS*, 192, 18  
 Krumholz M. R., McKee C. F., Tumlinson J., 2009, *ApJ*, 699, 850  
 Lagos C. D. P., Baugh C. M., Lacey C. G., Benson A. J., Kim H.-S., Power C., 2011, *MNRAS*, 418, 1649  
 Liu C., Mutch S. J., Angel P. W., Duffy A. R., Geil P. M., Poole G. B., Mesinger A., Wyithe J. S. B., 2016, *MNRAS*, 462, 235  
 Liu C., Mutch S. J., Poole G. B., Angel P. W., Duffy A. R., Geil P. M., Mesinger A., Wyithe J. S. B., 2017, *MNRAS*, 465, 3134  
 McQuinn M., Peng Oh S., Faucher-Giguère C.-A., 2011, *ApJ*, 743, 82  
 Mesinger A., Furlanetto S., Cen R., 2011, *MNRAS*, 411, 955  
 Mitchell P. D. et al., 2017, *MNRAS*, 474, 492  
 Mo H. J., Mao S., White S. D. M., 1998, *MNRAS*, 295, 319  
 Moster B. P., Macci A. V., Somerville R. S., Johansson P. H., Naab T., 2010, *MNRAS*, 403, 1009  
 Moster B. P., Macciò A. V., Somerville R. S., Naab T., Cox T. J., 2012, *MNRAS*, 423, 2045  
 Mutch S. J., Geil P. M., Poole G. B., Angel P. W., Duffy A. R., Mesinger A., Wyithe J. S. B., 2016a, *MNRAS*, 462, 250 (Paper-III)  
 Mutch S. J. et al., 2016b, *MNRAS*, 463, 3556  
 Navarro J. F., Frenk C. S., White S. D. M., 1997, *ApJ*, 490, 493  
 Newman S. F. et al., 2012, *ApJ*, 761, 43  
 Oesch P. A. et al., 2016, *ApJ*, 819, 129  
 Park J. et al., 2017, *MNRAS*, 472, 1995  
 Pawlik A. H., Rahmati A., Schaye J., Jeon M., Vecchia C. D., 2017, *MNRAS*, 466, 960  
 Pedregosa F. et al., 2011, *J. Mach. Learn. Res.*, 12, 2825

- Planck Collaboration XLVII, 2016, *A&A*, 596, A108
- Poole G. B., Angel P. W., Mutch S. J., Power C., Duffy A. R., Geil P. M., Mesinger A., Wyithe S. B., 2016, *MNRAS*, 459, 3025
- Poole G. B., Mutch S. J., Croton D. J., Wyithe S., 2017, *MNRAS*, 472, 3659
- Price S. H. et al., 2016, *ApJ*, 819, 80
- Qin Y., Duffy A. R., Mutch S. J., Poole G. B., Geil P. M., Angel P. W., Mesinger A., Wyithe J. S. B., 2017a, *MNRAS*, 467, 1678 (Paper-VIII)
- Qin Y., Mutch S. J., Duffy A. R., Geil P. M., Poole G. B., Mesinger A., Wyithe J. S. B., 2017b, *MNRAS*, 471, 4345
- Qin Y. et al., 2017c, *MNRAS*, 472, 2009
- Sawala T., Frenk C. S., Crain R. a., Jenkins A., Schaye J., Theuns T., Zavala J., 2013, *MNRAS*, 431, 1366
- Schaller M. et al., 2015, *MNRAS*, 451, 1247
- Schaye J., 2004, *ApJ*, 609, 667
- Schaye J., Dalla Vecchia C., 2008, *MNRAS*, 383, 1210
- Schaye J. et al., 2010, *MNRAS*, 402, 1536
- Schaye J. et al., 2015, *MNRAS*, 446, 521
- Sobacchi E., Mesinger A., 2013, *MNRAS*, 432L51
- Somerville R. S., Davé R., 2015, *ARA&A*, 53, 51
- Somerville R. S., Hopkins P. F., Cox T. J., Robertson B. E., Hernquist L., 2008, *MNRAS*, 391, 481
- Springel V., 2005, *MNRAS*, 364, 1105
- Springel V. et al., 2005, *Nature*, 435, 629
- Springel V. et al., 2008, *MNRAS*, 391, 1685
- Stevens A. R. H., Croton D. J., Mutch S. J., 2016, *MNRAS*, 461, 859
- Stevens A. R. H., del P Lagos C., Contreras S., Croton D. J., Padilla N. D., Schaller M., Schaye J., Theuns T., 2017a, *MNRAS*, 467, 2066
- Stevens A. R. H., Lagos C. d. P., Contreras S., Croton D. J., Padilla N. D., Schaller M., Schaye J., Theuns T., 2017b, *MNRAS*, 467, 2066
- Sutherland R. S., Dopita M. a., 1993, *ApJS*, 88, 253
- Velliscig M., van Daalen M. P., Schaye J., McCarthy I. G., Cacciato M., Le Brun A. M. C., Vecchia C. D., 2014, *MNRAS*, 442, 2641
- Vogelsberger M. et al., 2014, *Nature*, 509, 177
- Wiersma R. P. C., Schaye J., Smith B. D., 2009a, *MNRAS*, 393, 99
- Wiersma R. P. C., Schaye J., Theuns T., Dalla Vecchia C., Tornatore L., 2009b, *MNRAS*, 399, 574
- Zeldovich Y. B., 1970, *A&A*, 5, 84

## APPENDIX A: MATCHED GALAXIES BETWEEN MERAXES AND SMAUG

In this work, we make direct comparisons of the semi-analytic and hydrodynamic galaxy properties calculated by MERAXES and *Smaug*, respectively. We connect galaxies produced by MERAXES to their host haloes in the  $N$ -body simulation, and then match galaxies in MERAXES with the corresponding hydrodynamic simulation. We describe the method in detail for matching galaxies between MERAXES and *Smaug* in this section.

### A1 Running MERAXES on dark matter halo merger trees from *Smaug*

The same algorithm of matching haloes between snapshots is adopted to match between simulations. Therefore, we first discuss the construction of halo merger trees in the DRAGONS framework (Poole et al. 2016; Poole et al. 2017).

#### A1.1 Dark matter halo merger trees

Dark matter (and baryon) particles evolve (co-evolve) in *Smaug*, and FOF groups and subgroups (or haloes), to which particles belong are determined using a standard FOF halo finder and subgroup finder (SUBFIND, Springel et al. 2008) in post-processing. More specifically, the halo finder identifies FOF groups using a standard linking length of 0.2 and then treats each group as a sphere with centre located at

the position of its most bounded particle (*MBP*). Then by tracking halo particles in consecutive snapshots, each halo can be linked to its progenitors and form a merger tree. Each merger tree can be traced towards higher redshifts until no progenitors can be identified. This process is applied to all haloes at one snapshot, in order to horizontally construct the merger trees, which allows the SAM to evolve all galaxies at each snapshot and calculate reionization feedback self-consistently. We note that the *Smaug* trees provide 103 snapshots between  $z = 50$  and 5 with a time interval of  $\sim 11.3$  Myr. This high temporal resolution is considered a major improvement of the DRAGONS project (Paper-III), which is able to resolve the dynamical time of the galactic disc at high redshift (see equation 5).

#### A1.2 Matching haloes

We next give a brief overview of the GBPTREES algorithm<sup>20</sup> used to build dark matter halo merger trees, which is essential for understanding the matching process. We refer the interested reader to Poole et al. (2017) for more detailed descriptions and numerical tests.

For each halo, SUBFIND sorts particles according to their kinetic and potential energies, with, in general, more bounded particles ranking higher. More specifically, we consider a halo,  $\mathcal{H}^{(i)}$  at snapshot  $i$ , which consists of  $n^{(i)}$  particles radially sorted in order of decreasing potential energies with ranks of  $r^{(i)} = 1, 2, \dots, n^{(i)}$ . At snapshot  $j \neq i$ , the particles of  $\mathcal{H}^{(i)}$  can be located within several haloes  $\{\mathcal{H}_k^{(j)}; k = 1, 2, \dots, n_k\}$ , each of which is considered as a matching candidate (i.e. progenitor or descendant) and has  $n_k^{(ij)}$  shared particles with  $\mathcal{H}^{(i)}$ . For each candidate, a pseudo-radial moment is defined as

$$S_k^{(ij)}(m) = \sum_{l=1}^{n_k^{(ij)}} \left[ r_l^{(i)} \right]^m, \quad (\text{A1})$$

where  $r_l^{(i)}$  is the rank of particle  $l$  of halo  $\mathcal{H}_k^{(j)}$  in halo  $\mathcal{H}^{(i)}$ . Therefore, <sup>21</sup>  $\sum_k S_k^{(ij)} \leq S_{\max}^{(ij)} \equiv 1^m + 2^m + \dots + n^{(i)m}$ . We define the goodness of a matching candidate as

$$\Delta f_k^{(ij)} \equiv f_k^{(ij)} \Big|_{m=-1} - f_k^{(ij)} \Big|_{m=0},$$

where

$$f_k^{(ij)}(m) = \frac{S_k^{(ij)}(m)}{S_{\max}^{(ij)}(m)}. \quad (\text{A2})$$

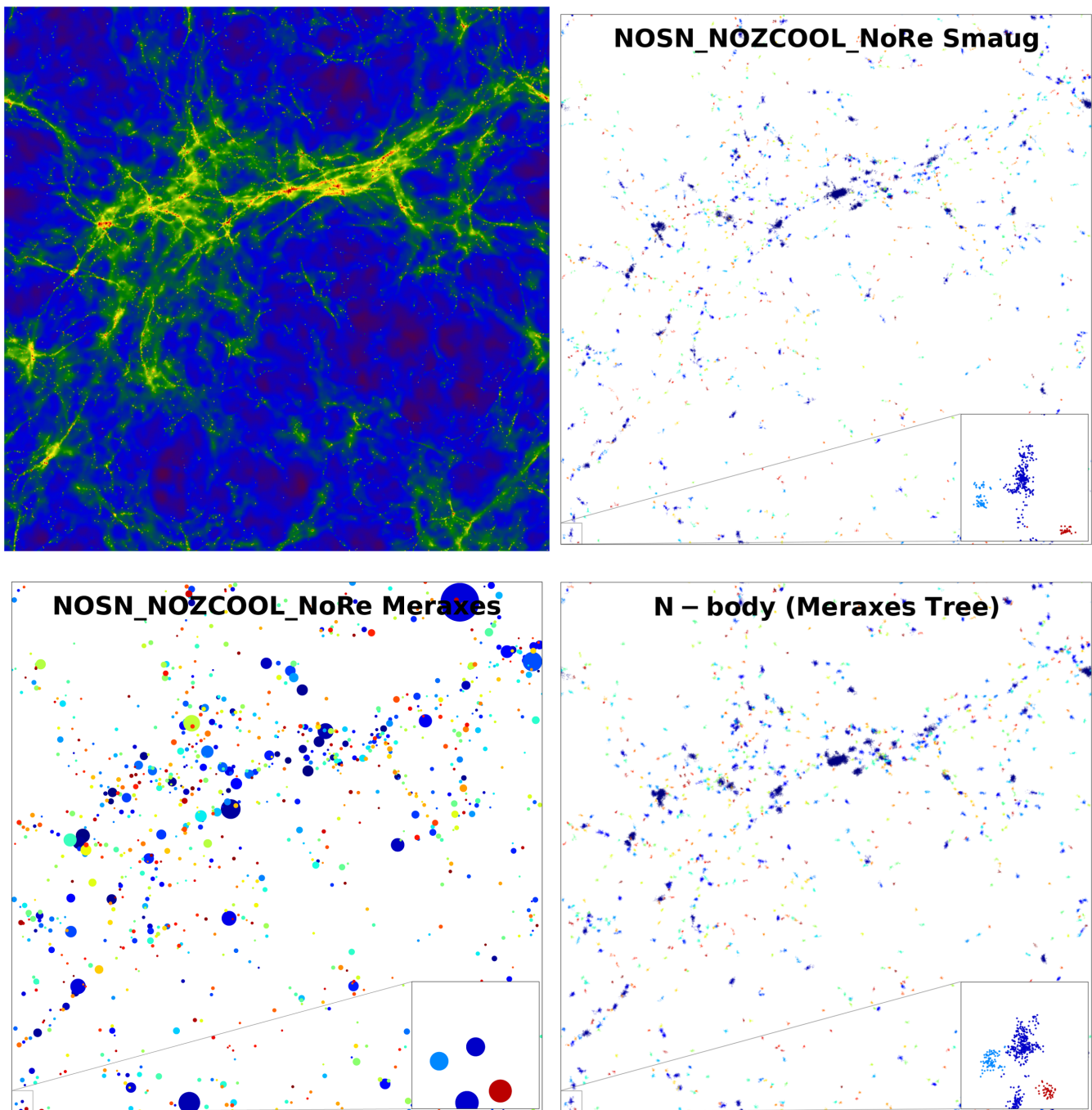
$\mathcal{H}^{(i)}$  and  $\mathcal{H}_k^{(j)}$  are considered as a good match when  $\Delta f_k^{(ij)} > -0.2$  (Poole et al. 2017) Amongst all good matches, the best match is identified by maximizing the statistic of  $S_k^{(ij)}(m=-1)$  during the process of scanning for good matches forwards and backwards over 16 snapshot<sup>22</sup>  $-|j-i| = 1, 2, \dots, 16$ .

A consequence of this algorithm is that matching is performed by tracking halo cores instead of the majority of halo particles. This central weighting algorithm allows us to build more realistic halo merger trees, distributed on which galaxies form and evolve within the central regions of their host haloes. However, pathologies can also arise in this algorithm and require special treatments. We

<sup>20</sup> <https://github.com/gbpoole/gbpCode>

<sup>21</sup>  $\sum_k S_k^{(ij)}$  is smaller than  $S_{\max}^{(ij)}$  if some of the particles are in unresolved haloes.

<sup>22</sup> When  $i$  is close to 0 or the total number of snapshots,  $|j-i| < 16$ , leading to a worse matching result. This usually affects the performance of the halo merger trees when the simulation reaches the end.

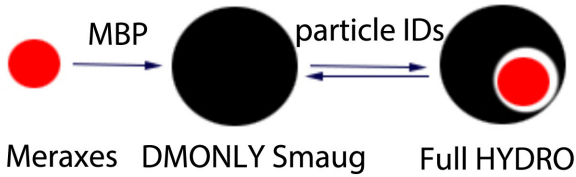


**Figure A1.** The top left panel shows the gas density distribution of the entire *NOSN\_NOZCOOL\_NoRe Smaug* simulation volume at  $z = 5$ . The remaining plots show an example of matched galaxies/haloes between three simulations (i.e. *NOSN\_NOZCOOL\_NoRe MERAXES*, bottom left; the *N*-body simulation, bottom right; and *NOSN\_NOZCOOL\_NoRe Smaug*, top right). The bottom left shows the distribution of the first matched 1000 galaxies in *MERAXES* with circle size representing stellar mass. Varied colours indicate different objects for better visualization. The bottom right shows the distribution of the dark matter particles within the host haloes of the aforementioned 1000 galaxies in the *N*-body simulation. The top right panel shows how the particles of those 1000 haloes in the bottom right panel are distributed in the *NOSN\_NOZCOOL\_NoRe Smaug* simulation. The zoom-in boxes on the bottom right indicate four matching candidates.

briefly review two consequences of central weighting while more dedicated tests and analysis are presented in Poole et al. (2017).

(i) When a halo orbits around a more massive object for a period of time without a subsequent merger event, it might be identified as a fraction of the larger system by *SUBFIND*, leading to a temporary merger event and causing numerical confusions. Under this circumstance and when the two haloes split, the descendant of the tem-

porarily merged system is determined by the dominant object in the central regions before the separation. However, a side effect arises as well that the remaining halo is identified as a newly formed halo without any progenitors. Note that losing progenitors in the merger trees leads to a reset of the hosted galaxy and thus discards its entire evolutionary history, which is catastrophic for semi-analytic galaxy formation models. The same impact also arises when the descendant cannot be identified due to tidal disruptions or numerical



**Figure A2.** Illustration of matching between galaxies in *MERAXES* and full-hydrodynamic simulation *Smaug* outputs. The red and black circles indicate galaxies and their host haloes, respectively. We link galaxies in *MERAXES* with their host haloes in the *N*-body simulation through the most bonded particle, *MBP*. Then haloes between *N*-body and full-hydrodynamic simulations are matched using dark matter particle IDs.

noise. Therefore, *GBTREES* is optimized to capture these and link the progenitor or locate the descendant with matching dedicated over consecutive and multiple snapshots.

(ii) When a central halo, which is usually considered as the most massive halo in an *FOF* group and dominates the gravitational potential of the entire system, is interacting with its satellites, particles are frequently exchanged between central and satellites. This can potentially lead to numerical confusion that the dominant halo switches between different substructures of the *FOF* group, while their cores remain nearly intact. In this case and because matching performed by following the core restricts galaxies in the centres of their host haloes, pathologies might arise that satellite galaxies receive unrealistic increments of mass on their hosts, leading to massive infall from the IGM and dramatic star formation when cooling is efficient.

## A2 Match between *MERAXES* and *Smaug*

*MERAXES* reports the *MBP* of the host halo for each galaxy, which bridges *SAM* galaxies and dark matter haloes in *DMONLY*. In order to further connect haloes in the collisionless *N*-body simulation with simulations including baryonic physics (i.e. *NOSN\_NOZCOOL\_NoRe*), we match haloes between simulations using the same matching strategy described in Appendix A1. Instead of matching 33 snapshots in one simulation, we match haloes from two simulations (e.g. *NOSN\_NOZCOOL\_NoRe* and *DMONLY*) using their dark matter particles at each snapshot. This is achievable because all the *Smaug* simulations including *DMONLY* start from the identical cosmological initial conditions and their dark matter particles possess the same IDs over different simulations. However, with information from only two snapshots, mismatches and pathologies occur in particular when reaching the resolution limit. Therefore, final matched sample is limited to haloes which are matched bidirectionally. Fig. A2 illustrates how galaxies are finally matched between *MERAXES* and *Smaug* with this two-step manner that *MBPs* connect galaxies from *MERAXES* to their host haloes which are matched with *Smaug* galaxies through dark matter particles. In this work, only central galaxies with virial masses exceeding the resolution threshold of  $10^{7.5} M_{\odot}$  in the *SAM* results are considered, and in order to minimize the impact from central-satellites switching mentioned in Appendix A1, satellite galaxies identified in the *SAM* are excluded from the final sample as well.

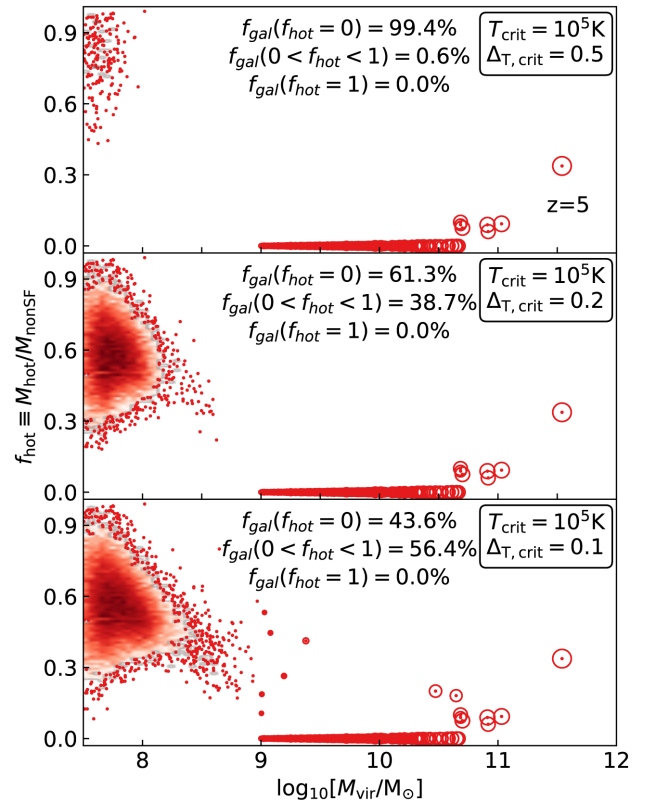
The right-hand panels of Fig. A1 project the spatial distributions of dark matter particles in the *NOSN\_NOZCOOL\_NoRe* and *N*-body simulations from *Smaug* on to the *xy*-plane, which illustrates how well the matching procedure works. In Fig. A1, the bottom left panel shows the distribution of the first matched 1000 galaxies<sup>23</sup> in

<sup>23</sup> The position of galaxy in *MERAXES* is inherited from its host halo in the *N*-body simulation.

*MERAXES* (also without supernova feedback and metal cooling) with circle size representing stellar mass and different colours for better visualization. The bottom right panel shows the distribution of dark matter particles of the corresponding host haloes in the *N*-body simulation, while the top right panel shows how these particles are distributed in the *NOSN\_NOZCOOL\_NoRe Smaug* simulation. The bottom right of each panel zooms into four objects for better illustration. In the *NOSN\_NOZCOOL\_NoRe Smaug* simulation, there are 195 562 galaxies, in which 129 043 galaxies ( $\sim 66$  per cent) are matched with *MERAXES*.

## APPENDIX B: HOT AND COLD NON-STAR-FORMING GAS

In this work, we use a machine learning algorithm to split non-star-forming gas particles into hot and cold components. The two groups are considered as one when the offset between their median temperatures is small (i.e. less than  $\Delta T_{\text{crit}}$  in logarithm). Fig. B1 shows the fraction of hot gas in the non-star-forming gas ( $f_{\text{hot}} \equiv M_{\text{hot}}/M_{\text{nonSF}}$ ) at  $z = 5$  with different choices of  $\Delta T_{\text{crit}}$  and we see that a smaller  $\Delta T_{\text{crit}}$  introduces more galaxies with  $f_{\text{hot}} > 0$ . However, this only affects galaxies with inefficient cooling (i.e. below the atomic cooling thresholds), while  $f_{\text{hot}}$  of more massive galaxies remain zero or small. On the other hand, when non-star-forming gas particles of a galaxy are considered as one group, which is common for galaxies with efficient cooling, whether they are determined as



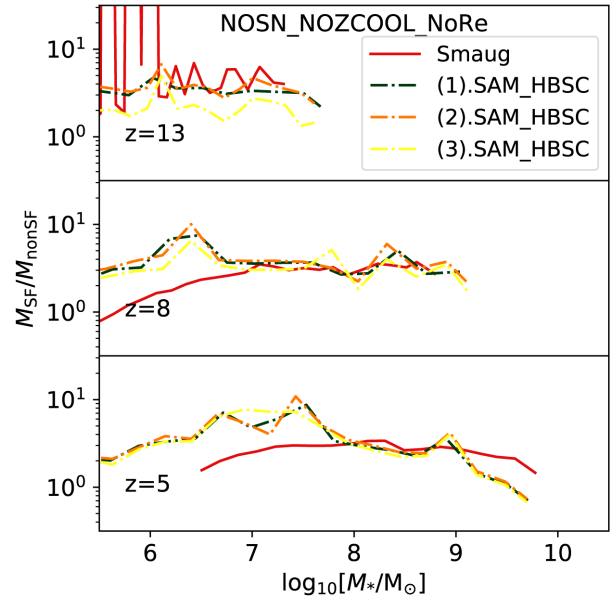
**Figure B1.** The fraction of hot gas in the non-star-forming gas ( $f_{\text{hot}} \equiv M_{\text{hot}}/M_{\text{nonSF}}$ ) from the *NOSN\_NOZCOOL\_NoRe Smaug* simulation at  $z = 5$  as a function of the halo mass ( $M_{\text{vir}}$ ). From top to bottom,  $\Delta T_{\text{crit}}$  is 0.5, 0.2, and 0.1. Galaxies with  $f_{\text{hot}} \neq 0$  are indicated with red contours (logarithm scale) and dots at low-dense regions. Galaxies with  $M_{\text{vir}} > 10^9 M_{\odot}$  are emphasized with red circles with circle size representing the stellar mass ( $M_{*}$ ). The fraction of galaxies with  $f_{\text{hot}} = 0$ ,  $0 < f_{\text{hot}} < 1$  and  $f_{\text{hot}} = 1$  are shown in each subpanel.

hot or cold (i.e.  $f_{\text{hot}} = 1$  or 0) is based on their median temperature. We vary the minimum temperature of particles being identified as hot ( $T_{\text{crit}}$ ) from  $10^5$  to  $10^4$  K and we find that  $f_{\text{hot}}$  is insensitive to  $T_{\text{crit}}$ . Based on these, we note that the choice of  $\Delta T_{\text{crit}}$  and  $T_{\text{crit}}$  does not have a significant impact to our conclusion that high-redshift dwarf galaxies accrete gas in cold mode (see Section 4.2).

### APPENDIX C: SEMI-ANALYTIC PROPERTIES

A numerical experiment evolves a galaxy by calculating physics processes in a pre-defined order. For instance, `MERAXES` in sequence calculates reionization, baryonic infall, cooling, star formation, and the impact of supernova feedback. Thanks to the exquisite time-step of calculation (not output) of hydrodynamic simulations, their outputs can be considered as instantaneous galaxy properties at a given redshift. However, this is not the case for the SAM used in this work, which records properties that either have (e.g. stellar mass and gas mass) or have not been processed (e.g. host halo properties) or are average values (e.g. SFR) during one time-step. For instance, in the right-hand panel of Fig. 5, we show  $M_{\text{SF}}/M_{\text{nonSF}}$ , which is the mass ratio of cold to hot gas in the SAM. Three moments of calculation can be adopted:

- (1) after star formation, when both hot gas and cold gas have been consumed. In this case, gas mass is underestimated due to the lack of replenishment;
- (2) before star formation and after cooling, when hot gas has transitioned to cold gas. In this case, stellar mass and hot gas are underestimated while cold gas is overestimated;
- (3) before cooling and after baryon infall, when hot gas has been refurbished but cold gas still represents the remaining gas reservoir from the previous snapshot or  $\sim 11$  Myr ago. In this case, stellar mass and cold gas are underestimated, while hot gas is overestimated.



**Figure C1.** The mass ratio of star-forming gas to non-star-forming gas as a function of stellar mass at  $z = 13-5$  from `MERAXES` and *Smaug* `NOSN_NOZCOOL_NoRe` results. Three `MERAXES` results with redshift-dependent star formation efficiency and cooling rate (*SAM\_HBSC*) are shown with green dashed for calculations performed at different moments.

We note that case (1) is adopted in the main context, and we show the difference in Fig. C1. We see that because of the high cadence ( $\sim 11$  Myr) of the SAM, the choice of calculation does not have a significant impact to our results.

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.