Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **492,** 2835–2846 (2020)
Advance Access publication 2020 January 9

doi:10.1093/mnras/staa056

Testing galaxy formation simulations with damped Lyman-α abundance and metallicity evolution

Sultan Hassan [1,2]★† Kristian Finlator [1,3] Romeel Davé [2,4,5]
Christopher W. Churchill[1] and J. Xavier Prochaska[6,7]

[1]*Department of Astronomy, New Mexico State University, Las Cruces, NM 88003, USA*
[2]*Department of Physics and Astronomy, University of the Western Cape, Bellville, Cape Town 7535, South Africa*
[3]*Cosmic Dawn Center (DAWN), Niels Bohr Institute, University of Copenhagen / DTU-Space, Technical University of Denmark*
[4]*Institute for Astronomy, Royal Observatory, Edinburgh EH9 3HJ, UK*
[5]*South African Astronomical Observatory, Observatory, Cape Town 7925, South Africa*
[6]*Department of Astronomy and Astrophysics, UCO/Lick Observatory, University of California, 1156 High Street, Santa Cruz, CA 95064, USA*
[7]*Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU), 5-1-5 Kashiwanoha, Kashiwa 277-8583, Japan*

Accepted 2020 January 7. Received 2020 January 7; in original form 2019 October 15

ABSTRACT

We examine the properties of damped Lyman-α absorbers (DLAs) emerging from a single set of cosmological initial conditions in two state-of-the-art cosmological hydrodynamic simulations: SIMBA and TECHNICOLOR DAWN. The former includes star formation and black hole feedback treatments that yield a good match with low-redshift galaxy properties, while the latter uses multifrequency radiative transfer to model an inhomogeneous ultraviolet background (UVB) self-consistently and is calibrated to match the Thomson scattering optical depth, UVB amplitude, and Ly α forest mean transmission at $z > 5$. Both simulations are in reasonable agreement with the measured stellar mass and star formation rate functions at $z \geq 3$, and both reproduce the observed neutral hydrogen cosmological mass density, $\Omega_{\rm H\,I}(z)$. However, the DLA abundance and metallicity distribution are sensitive to the galactic outflows' feedback and the UVB amplitude. Adopting a strong UVB and/or slow outflows underproduces the observed DLA abundance, but yields broad agreement with the observed DLA metallicity distribution. By contrast, faster outflows eject metals to larger distances, yielding more metal-rich DLAs whose observational selection may be more sensitive to dust bias. The DLA metallicity distribution in models adopting an H_2-regulated star formation recipe includes a tail extending to $[M/H] \ll -3$, lower than any DLA observed to date, owing to curtailed star formation in low-metallicity galaxies. Our results show that DLA observations play an important role in constraining key physical ingredients in galaxy formation models, complementing traditional ensemble statistics such as the stellar mass and star formation rate functions.

Key words: galaxies: evolution – galaxies: formation – intergalactic medium – quasars: absorption lines – cosmology: theory.

1 INTRODUCTION

Strong H I absorbers with column densities above 2×10^{20} cm^{-2}, known as Damped Lyman-α systems (DLAs; Wolfe et al. 1986; Prochaska, Herbert-Fort & Wolfe 2005; Wolfe, Gawiser & Prochaska 2005; Prochaska & Wolfe 2009), are rare and distinct profiles in the spectrum of background quasars. DLAs contain the dominant reservoir of cosmic neutral gas that eventually feeds star formation within galaxies, and hence they provide a unique way to constrain star formation and associated processes in galaxy formation models.g

The nature of DLAs and their relationship to host galaxies have long been a source of controversy. Early works proposed that DLAs are the progenitors of present-day galaxies (e.g. Lanzetta et al. 1991; Wolfe et al. 1995), and Prochaska & Wolfe (1997) favoured a scenario where DLAs are thick, rapidly rotating discs, owing in part to their large internal kinematic velocities, now confirmed in larger samples (Neeleman et al. 2013). Such large

★ E-mail: shassan@nmsu.edu
† Tombaugh Fellow.

© 2020 The Author(s)
Published by Oxford University Press on behalf of the Royal Astronomical Society

Downloaded from https://academic.oup.com/mnras/article/492/2/2835/5698819 by University of Western Cape user on 12 February 2021

discs at early epochs were a challenge to favoured cold dark matter (CDM) cosmologies (Prochaska & Wolfe 1998). Subsequently, clustering studies have broadly constrained the DLA hosting halo mass scale to be $10^{9-11}\,M_{\odot}$ (e.g. Pontzen et al. 2008; Lochhaas et al. 2016; Pérez-Ràfols et al. 2018; Jeon, Besla & Bromm 2019), and virial velocity range of about $50-200\,km\,s^{-1}$ (e.g. Barnes & Haehnelt 2009; Bird et al. 2014). This suggested that DLAs occur in a wide range of galaxies and environments. Observations of DLA metallicities showed that they had relatively low metallicities compared to typical star-forming galaxies at the same epoch along with alpha element ratios comparable to halo stars, with mild evolution and a clear metallicity floor (e.g. Rafelski et al. 2012, 2014). These results provide constraints on the nature of DLAs, but attempts to comprehensively fit all these observations within a single cosmologically situated model remain elusive.

The dynamic nature of neutral hydrogen gas within hierarchically growing galaxies favours the use of hydrodynamic simulations to interpret DLA observations. Early simulations by Haehnelt, Steinmetz & Rauch (1998) countered the notion that DLAs challenge CDM models by showing that irregular protogalactic clumps can equivalently reproduce the observed distribution of DLA velocity widths within a hierarchical context (but see Prochaska & Wolfe 2010). Nagamine, Springel & Hernquist (2004) showed that cosmological simulations can approximately reproduce the abundance of DLAs over cosmic time when star formation feedback is included. However, Berry et al. (2014) used a semi-analytic model to show that it is possible to match DLA abundances, but only in a model where the H I disc is highly extended compared to expectations, and such models still fail at $z > 3$.

More recent simulations have shown that DLA abundances are particularly sensitive to assumptions regarding stellar feedback. Bird et al. (2014) used cosmological simulations with AREPO to demonstrate a degeneracy between the wind speed and UVB amplitude in which both were anticorrelated with the DLA cross-section. Bird et al. (2015) further showed that hierarchical models can statistically reproduce DLA kinematics. Faucher-Giguère et al. (2015) used higher resolution Feedback in Realistic Environments (FIRE) zoom simulations to show that stellar feedback has a large impact on the cross-section of high-column H I gas. Rhodin et al. (2019) used similarly high-resolution simulations to explore the impact parameter distribution of DLAs from host galaxies at lower redshifts, showing that high resolution and efficient star formation feedback are required to match observations. These studies indicate that DLA properties such as their abundances, metallicities, impact parameters, and kinematic spreads provide constraints on processes of star formation and the strength of the local photoionizing flux.

To make progress in understanding DLAs in a cosmological context, ideally one requires simulations with sufficient volume for good statistics, sufficient resolution to fully capture relevant dense gas processes, a model for star formation and associated feedback that is concordant with a wide range of observations, and self-consistent modelling of the photoionizing background including self-shielding. Unfortunately, no simulation currently exists that fully satisfies all these criteria. Nonetheless, recent models have made substantial progress towards this, and by comparing simulations with contrasting implementations of these processes, it is possible to gain more robust insights into the nature of DLAs within hierarchical structure formation models. This is the goal of this paper, the first in a series to explore $z \geq 3$ DLA properties in two state-of-the-art high-resolution cosmological simulations.

In this work, we use the observed DLA abundance and metallicity evolution to test two simulations, namely, SIMBA (Davé et al. 2019), a cosmological hydrodynamic simulation with black hole growth feedback; and TECHNICOLOR DAWN (TD; Finlator et al. 2018), a cosmological hydrodynamic simulations including an on-the-fly multifrequency radiative transfer solver. While the ultraviolet ionizing background (UVB) treatment, feedback effects, and star formation recipes are all quite different, both simulations have been calibrated to reproduce key galaxy observables. For instance, SIMBA reproduces (Davé et al. 2019) galaxy stellar mass function from $z = 0$–6, the stellar mass–star formation rate main sequence, low-z H I and H_2 mass fractions, the mass–metallicity relation at $z \sim 0-2.5$, black hole–galaxy co-evolution (Thomas et al. 2019), and galaxy dust properties (Li, Narayanan & Davé 2019). TD, which focuses more on high redshifts, reproduces the history of reionization, the galaxy stellar mass–star formation rate relation, the abundance of high-z metal absorbers, the UVB amplitude, and the Lyman-α flux power spectrum at $z = 5.4$. Hence both simulations appears to have sub-grid recipes for star formation and other processes that yield agreement with a range of current constraints.

The need for complementary tests of these codes arises from the fact that, although they both reproduce a broad variety of observations of galaxy growth, they do so by modelling star formation and feedback in different ways. TD adopts the Springel & Hernquist (2003) sub-grid multiphase model in which the local star formation rate depends only on gas density, whereas SIMBA employs an H_2-regulated star formation model based on the sub-grid model of Krumholz, McKee & Tumlinson (2009). While both employ kinetic star formation feedback, SIMBA employs scalings taken from FIRE (Anglés-Alcázar et al. 2017a) that have significantly lower outflow speeds but higher mass loading than those assumed in TD which come from tuned constraints (Davé et al. 2013). Finally, SIMBA employs a different and potentially improved hydrodynamics solver relative to that in TD. These differences are expected to impact the predicted circumgalactic medium (CGM) neutral gas reservoirs, and thus DLAs provide complementary, sensitive tests that provide novel constraints on models. Here, we employ the exact same initial conditions for our SIMBA and TD runs down to $z = 3$, therefore allowing us to isolate the differences in DLA predictions purely owing to input physics variations.

This paper is organized as follows. We describe the SIMBA and TD simulations in Section 2. In Section 2.4 we cast sightlines through the simulation volumes and extract catalogues of synthetic DLAs along with their associated low and high ion transitions, including a comparison between the results of identifying DLAs in observed versus theoretical spaces. We compare the predicted gas density distribution, star formation history, and UVB treatments in Section 3. We explore predictions for DLA abundance evolution and the column density distribution in Section 4, and metallicity distribution and evolution in Section 5. We summarize our results in Section 6.

2 SIMULATIONS

We here briefly describe SIMBA and TD and refer the reader to Davé et al. (2019), Davé, Thompson & Hopkins (2016), and Finlator et al. (2018) for more detailed information of the physics implemented in these simulations.

2.1 SIMBA

The SIMBA model was introduced in Davé et al. (2019). SIMBA is a follow-up to the MUFASA (Davé et al. 2016) cosmological galaxy formation simulation using GIZMO's meshless finite mass

Downloaded from https://academic.oup.com/mnras/article/492/2/2835/5698819 by University of Western Cape user on 12 February 2021

hydrodynamics solver (Hopkins 2015, 2017). Radiative cooling and photoionization heating are implemented using the updated GRACKLE-3.1 library (Smith et al. 2017). A spatially uniform ionizing background from Haardt & Madau (2012) is assumed, in which self-shielding is accounted for following Rahmati et al. (2013). The chemical enrichment model tracks nine elements (C, N, O, Ne, Mg, Si, S, Ca, Fe) arising from Type II supernovae (SNe), Type Ia SNe, and Asymptotic Giant Branch (AGB) stars. Type Ia SNe and AGB wind heating are also included. The star formation-driven galactic winds are kinetically launched and decoupled into hot and cold phase winds, and the winds are metal-loaded owing to local enrichment from supernovae, with overall metal mass being conserved. The mass rate entering galactic outflows is modelled with a broken power law following Anglés-Alcázar et al. (2017b). The quasi-linear scaling of wind velocity with escape velocity from Muratov et al. (2015) is adopted. Exact equations are summarized in Table 1. SIMBA further implements black hole growth via a torque-limited accretion model (Anglés-Alcázar et al. 2017a), which is a unique feature of SIMBA, plus Bondi accretion from hot gas. Feedback from Active Galactic Nuclei (AGN) is implemented in two modes, a radiative mode at high accretion rates that follows constraints from observed ionized AGN outflows (e.g. Fabian 2012; Perna et al. 2017), and a jet mode at low accretion rates that ejects material at velocities approaching 10^4 km s^{-1}. SIMBA also accounts for the X-ray AGN feedback in the surrounding gas following Choi et al. (2012). Dust production and destruction are modelled on-the-fly, leading to predictions of dust abundance that broadly agree with a variety of constraints over cosmic time (Li et al. 2019).

Given the importance of the star formation recipe in setting the gas content in the ISM, we describe this in somewhat more detail. SIMBA employs a molecular gas-based prescription following Krumholz et al. (2009, hereafter KMT) to form stars. KMT is a physically motivated recipe to model star formation as seen in local disc galaxies, where a strong correlation is seen between SFR and molecular gas content (e.g. Leroy et al. 2008). In the KMT model, the H_2 mass fraction f_{H_2} depends on the metallicity and local column density as follows:

$$f_{H_2} = 1 - 0.75\frac{s}{1+0.25s}\,, \tag{1}$$

where

$$s = \frac{\ln(1+0.6\chi+0.01\chi^2)}{0.0396\,Z\,(\Sigma/\mathrm{M_\odot\,pc^{-2}})}\,, \tag{2}$$

where Z is the metallicity in solar units, χ is a function of metallicity given in KMT, and Σ is the column density. A resolution-varying clumping factor is also implemented. This improvement over the constant value in the original KMT model, which was calibrated on ∼kpc scale, enables higher resolution calculations to adopt a higher threshold density for H_2 formation. Finally, ISM gas is pressurized to keep the Jeans mass resolved at all densities, resulting in an equation of state where $T \propto \rho^{1/3}$ at the highest densities.

Because the KMT model predicts a steep dependence of star formation efficiency on metallicity and was calibrated on a Milky Way-like ISM, the extrapolation to low-metallicity situations is highly uncertain. As such, SIMBA implements a metallicity floor for Z; that is, we set $Z = \mathrm{MAX}(Z, Z_{\rm floor})$ in the KMT equations. Note that this does not impact the overall metallicity of the gas or the metal cooling rate; it is only a floor applied when using the KMT formulae. Importantly, this adjustment allows star formation to occur in primordial gas, where a literal adoption of the KMT model would prevent star formation outright because metals are required for molecules to form. The fiducial value is $\log Z_{\rm floor} = -2$, but we will evaluate how sensitive DLA properties are to this choice by comparing models in which $\log Z_{\rm floor} = [-1, -2, -3]$.

Given f_{H_2}, the SFR of an individual gas element is given by a Schmidt (1959) Law as follows:

$$\frac{\mathrm{d}M_*}{\mathrm{d}t} = \epsilon_* \frac{\rho\, f_{H_2}}{t_{\rm dyn}}\,, \tag{3}$$

where ρ is the gas density, $t_{\rm dyn} = 1/\sqrt{G\rho}$ is the local dynamical time, and the star formation efficiency ϵ_* is set to 0.02 (Kennicutt 1998). Star formation is only allowed in the dense gas phase (ISM gas) above a hydrogen number density $n_{\rm H} \geq 0.13\,\mathrm{cm^{-3}}$, though in practice the H_2 fraction forces star formation to occur at much higher densities ($n_{\rm H} \gg 1\,\mathrm{cm^{-3}}$). SIMBA assumes a Chabrier (2003) initial mass function (IMF) throughout.

2.2 TECHNICOLOUR DAWN (TD)

The TD simulations (Finlator et al. 2018), are a suite of cosmological radiative hydrodynamic transfer simulations. Hydrodynamics are modelled using a density-independent formulation of smoothed particle hydrodynamics (SPH; Hopkins 2013). Radiative cooling is implemented following Katz, Weinberg & Hernquist (1996), although the ionization states of H and He are tracked using a non-equilibrium solver. The chemical enrichment model tracks ten elements (C, O, Si, Fe, N, Ne, Mg, S, Ca, Ti), accounting for enrichment from Type II SNe, Type Ia SNe, and asymptotic giant branch (AGB) stars. Galactic outflows and wind scaling are modelled similarly as in SIMBA, with different normalization factors as quoted in Table 1. In contrast to SIMBA TD does not assume two-phase outflows, but assumes all outflows are at the ISM temperature. Photoionization feedback and reionization are treated via a self-consistent, inhomogeneous, multifrequency UVB, using a moment-based radiative transfer approach on Cartesian grid (Finlator, Özel & Davé 2009). The emissivity from galaxies is computed directly from the star formation rate of star-forming gas particles, whereas the quasar contribution depends on redshift following Manti et al. (2017) and on energy using the Lusso et al. (2015) continuum slope.

For modelling star formation, TD adopts the sub-grid multiphase model developed by Springel & Hernquist (2003), in which the star-forming gas (ISM) is composed of cold clouds embedded within an ambient hot medium, following McKee & Ostriker (1977). These cold clouds represent the prime repository for star formation. The mass exchange between these phases occurs through star formation, evaporation from supernovae, and cloud growth due to cooling. The SFR here depends on the gas density as follows:

$$\frac{\mathrm{d}\rho_*}{\mathrm{d}t} = (1-\beta)\frac{\rho_c}{t_*}\,, \tag{4}$$

where $\rho_{\rm c}$ is the cloud density, t_* is characteristic time-scale to covert the $\rho_{\rm c}$ into stellar density ρ_*, and β is the mass fraction of stars that explodes as supernovae. Similar equation can be written for the hot medium accounting for the mass-loss from clouds and supernovae energies. This model also follows the Schmidt law (Schmidt 1959) where the SFR is proportional to $n_{\rm H}^{1.5}$, with the t_* tuned to match Kennicutt relation (Kennicutt 1998). Similarly, star formation only applied for dense gas with a hydrogen number density of $n_{\rm H} \geq 0.13\,\mathrm{cm^{-3}}$. TD assumes the Kroupa (2001) IMF, slightly different than SIMBA.

The major differences between the two simulations are summarized in Table 1.

Downloaded from https://academic.oup.com/mnras/article/492/2/2835/5698819 by University of Western Cape user on 12 February 2021

Table 1. Major differences between SIMBA and TD.

	SIMBA	TD
Star formation	H_2-regulated (Krumholz et al. 2009)	Hybrid multiphase (Springel & Hernquist 2003)
Radiative transfer	No	Yes
UVB	homogeneous, spectrum from Haardt & Madau (2012)	inhomogeneous, multifrequency (Finlator et al. 2009)
Galactic outflows	Yes	Yes
Mass loading factor $\eta(M_*)$	$= \begin{cases} 9\left(\frac{M_*}{M_0}\right)^{-0.317} & if\, M_* < M_0 \\ 9\left(\frac{M_*}{M_0}\right)^{-0.761} & if\, M_* > M_0 \end{cases}$,	–
	where $M_0 = 5.2 \times 10^9 M_\odot$ (Anglés-Alcázar et al. 2017b).	$= 2.3\left(\frac{M_*}{10^{10} M_\odot}\right)^{-0.35}$, (Muratov et al. 2015).
	This is about $\sim 2\times$ higher at $M_* < M_0$, and about $\sim 2\times$	–
	lower at $M_* > M_0$ than TD's adopted outflows rate	–
	(Muratov et al. 2015). Due to the small volume ($\sim$ 22 Mpc),	–
	$\eta(M_*)$ is higher for SIMBA at all relevant masses.	–
Wind velocity, v_w	$= 0.85\, v_{\rm circ}^{1.12} + \Delta v(0.25 R_{\rm vir})$, (Based on Muratov et al. 2015)	$= 2.4\, v_{\rm circ}^{1.12} + \Delta v(0.25 R_{\rm vir})$, (Based on Muratov et al. 2015). This is about ~ 3 times faster than SIMBA's wind velocity.
Black holes	Yes	No

2.3 Simulation runs

To establish a consistent comparison between these simulations, we use identical initial condition of $15\,h^{-1}$ Mpc with 2×640^3 dark matter and gas particles each, to run SIMBA and TD assuming the following cosmology: (Ω_m, Ω_Λ, Ω_b, H_0, X_H) = (0.3089, 0.6911, 0.0486, 67.74, 0.751), using a parallel version of N-GenIC[1] to initialize our simulations at $z = 199$. We then identify galaxies and haloes using a friends-of-friends group finder within 3D for TD and 6D for SIMBA. In this study, we restrict our analysis to the redshift range of $z = 3$–5 in d$z = 0.5$, where a sizeable population of DLAs have been observed, in order to test these simulations against observable DLA properties. It is worth mentioning that our simulations' box size ($15\,h^{-1}$ Mpc) is large enough to capture a representative population of DLAs although it may miss those that are associated with the most massive DLA-hosting haloes. For a concrete reference point, our mass resolution is about one order of magnitude higher than the fiducial runs considered in Bird et al. (2014), who modelled a $25\,h^{-1}$ Mpc volume using 512^3 gas resolution elements. Meanwhile, our cosmological volume is about 22 per cent as large. Low-mass galaxies and haloes and their contributions to DLAs are therefore more completely accounted for, while haloes with circular velocities exceeding $\approx$120 km s^{-1} are under-represented. Bird et al. (2014) found that the DLA hosting haloes' velocity distribution peaks at $\sim$100 km s^{-1}. Assuming this is correct, it suggests that our simulations may indeed be systematically deficient in DLAs associated with the most massive DLA-hosting haloes. We leave exploring the relationship between the DLAs kinematics and their hosting haloes/galaxies properties for a follow-up work.

2.4 The simulated DLA sample

From each simulation, we generate mock spectra following the recipe developed in Finlator et al. (2018), which we review briefly here. We pass an oblique long sightline through each simulation volume, from $z = 3$ to $z = 5$ in intervals $\Delta z = 0.5$, wrapping at the simulations boundaries, till a velocity width of 4×10^7 km s^{-1} is achieved. Choosing a large velocity width is necessary to ensure detection of a representative catalogue of synthetic DLAs. We next smooth the simulated density, temperature, metallicity, and velocity field on to the sightline with a pixel size of 2.67 km s^{-1}. The absorption spectra are generated using Voigt profiles. We use Haardt & Madau (2012) photoionization background within SIMBA whereas TD uses its own inhomogeneous photoionization background. We then compute the optical depth following Theuns et al. (1998). We use an 8 km s^{-1} full-width at half-maximum (FWHM) to smooth the generated spectra with a Gaussian filter. This results in a survey of 3 pixels per resolution element. We finally add a Gaussian noise (signal-to-noise, SNR) of 20 km s^{-1} per pixel. These values of the pixel and FWHM resolutions are similar to those values used in observations (e.g. Neeleman et al. 2013).

We identify DLAs as H I absorbers whose column density satisfies $N_{\rm H\,I} \geq 10^{20.3}$ cm^{-2}. As DLAs are on the damping regime of the curve of growth, the column density relates directly to the equivalent width $W \propto \sqrt{N_{\rm H\,I}}$, in which the DLA threshold translates into an equivalent width threshold of 9.17 Å in the rest frame. Our procedure to identify DLAs is as follows: we first scan our long sightline of 4×10^7 km s^{-1} and consider any pixels where the flux is more than 3σ below the continuum to indicate a significant absorption feature. We split the sightline into sequences of successive absorption features, assuming that each sequence represents one absorber. We then compute the equivalent width for all these features, and regard an absorption feature as a DLA candidate if its equivalent width is equal or above the threshold. To be more conservative, we consider a smaller equivalent width value of 9.0 Å as a cut-off to capture any possible DLA candidate.

Having identified DLA candidates, we consider two methods to compute each candidate's H I column density. The first is to fit a Voigt profile (VP) to its synthetic spectrum (observational method), while the second is to use the intrinsic column density as generated by the simulation (theoretical method). For the Voigt profile fit, we use the PYIGM,[2] a python package for the analysis of the intergalactic medium and the circumgalactic medium. Some examples for the VP fit using PYIGM are shown in Fig. 1. For the intrinsic column density, we compute the total H I column density

[1] https://wwwmpa.mpa-garching.mpg.de/gadget/

[2] https://github.com/pyigm/pyigm. The package performs by-hand fit for the input DLA profile.

Downloaded from https://academic.oup.com/mnras/article/492/2/2835/5698819 by University of Western Cape user on 12 February 2021

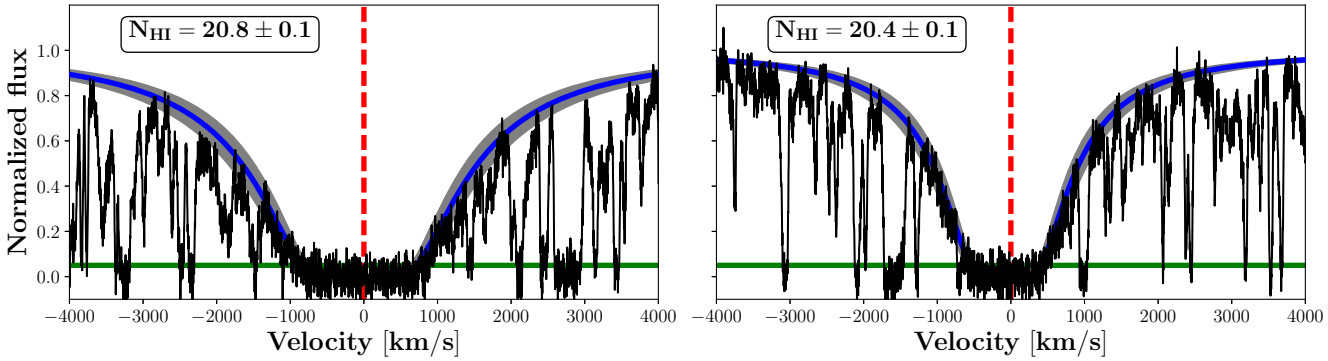

Figure 1. Examples of two DLA spectra with their VP fit (solid blue) using the PYIGM package. The green and red lines represent the uncertainty and velocity centroid, respectively. The column density values are quoted in each plot.

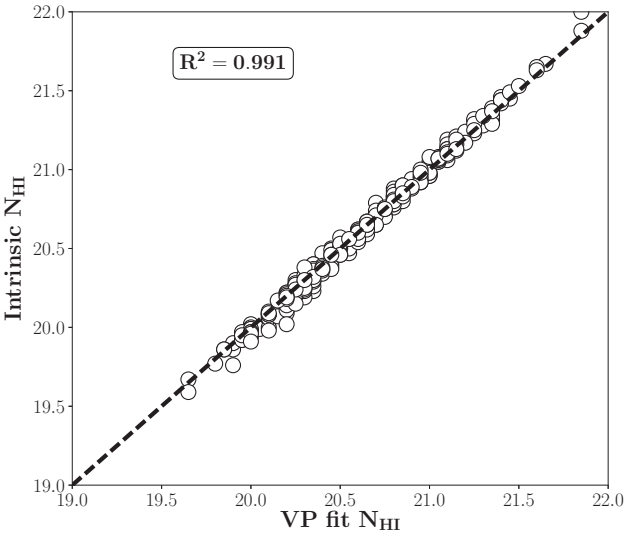

Figure 2. The DLAs column density comparison at $z = 3$ between values determined using a Voigt Profile (VP) fitter versus the intrinsic value from the simulation, which is computed as the total column density from pixels along the DLA profile within 500 km s^{-1} window. All DLAs are clustered along the dashed line (perfect match), which is an evidence that there is no bias towards using either method, with a coefficient of determination R2 > 99 per cent. Hence, we proceed using the intrinsic column density value at all redshifts.

along the DLA profile within a window of 500 km s^{-1} about the centroid (i.e. highest column density pixel).[3] We now compare the outcome of the two methods on all $z = 3$ simulated DLA profiles in Fig. 2. We notice that all DLAs are tightly clustered along the dashed line, which indicates a perfect match between the two methods, with a coefficient of determination R2 > 99 per cent. This shows that there isn't a bias towards using any method, and hence we proceed to use the intrinsic H I column density method at all redshifts.

[3]The choice of the window size does not affect the resulting column, since most of pixels along the DLA profile have a column density of 10^{11-13} cm^{-2}, which is about 9 order or magnitude lower the threshold. The 500 km s^{-1} choice is made appropriately according to the maximum observed velocity width of DLAs.

3 SIMULATION COMPARISON

We begin by comparing SIMBA and TD in terms of their dense gas distribution, star formation, and UVB treatment, in order to set the stage for understanding the physics driving DLA properties.

3.1 ISM density distribution

Fig. 3 shows gas density distribution in SIMBA and TD at the highest and lowest redshifts considered in this study ($z = 5, 3$). In the top panel, we compare the hydrogen number density n_H distribution in the two simulations. While both simulations have started from the same initial condition, we see a big difference in the n_H distribution, particularly at the dense end (i.e. the shelf-shielding regime). In the top axis, we show a rough estimate of the H I column density corresponding to the hydrogen number density, following Schaye (2001a) as:

$$N_{\rm H\,I} \sim 1.6 \times 10^{21} {\rm cm}^{-2}\, n_{\rm H}^{1/2}\, T_4^{1/2} \left(\frac{f_g}{0.17}\right)^{1/2}, \quad (5)$$

where f_g is the universal baryon fractions of absorbers $\frac{\Omega_b}{\Omega_m}$, and the temperature $T_4 \equiv T/10^4$ is set to unity, ignoring collision ionization. This equation provides an approximate estimate of the H I column density at high densities and high neutral fraction values. The vertical dashed lines shows the DLA column density threshold ($N_{\rm H\,I} = 2 \times 10^{20}$ cm^{-2}). There is clearly more gas in TD than SIMBA above the DLA threshold and below $N_{\rm H\,I} \sim 10^{21}$ cm^{-2}, in the hydrogen number density range $0.01-1$ cm^{-3}, where the majority of DLA population exists. This might be due to the lower outflow rate (about $\sim 2\times$ less than SIMBA; see Table 1) that TD applies at the intermediate stellar masses as well as the TD's faster winds (about $\sim 3\times$ higher than SIMBA). The density distribution in SIMBA extends to 10^3 cm^{-3}, which is roughly two orders of magnitude higher than the highest gas density in TD. This could reflect either SIMBA's low outflows speed (about $3\times$ less than TD; see Table 1). Alternatively, it could reflect SIMBA's tendency to suppress star formation in low-metallicity gas due to the H_2-regulated star formation model.

In the middle panel, we show the 50th percentile of neutral $f_{\rm H\,I}$ and molecular $f_{\rm H_2}$ fractions as a function of hydrogen number density by the solid and dashed lines, respectively. Both simulations indicate a sharp increase in $f_{\rm H\,I}$ at gas densities near 0.01 cm^{-3}, which is due to self-shielding (Rahmati et al. 2013). At densities higher than 10 cm^{-3}, the neutral fraction $f_{\rm H\,I}$ in SIMBA drops suddenly, marking the transition to molecular gas as shown by the rapid increase of $f_{\rm H_2}$. This is a distinct feature of SIMBA, which is a

Downloaded from https://academic.oup.com/mnras/article/492/2/2835/5698819 by University of Western Cape user on 12 February 2021

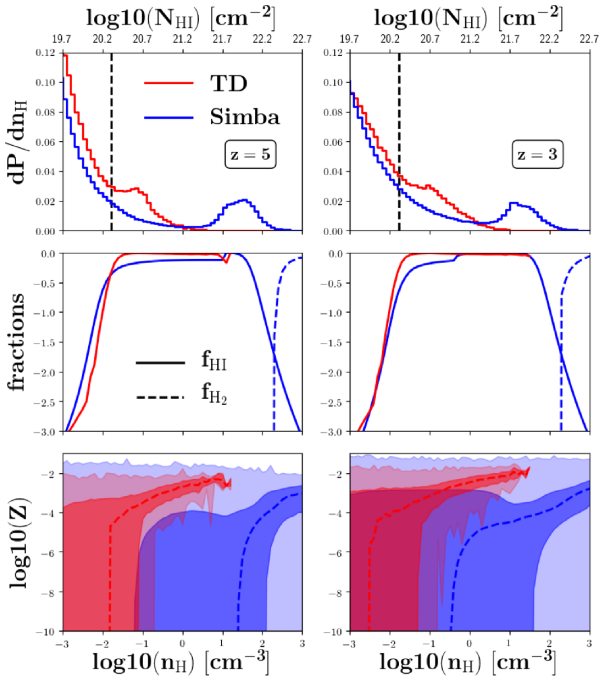

Figure 3. Density distribution in SIMBA (blue) and TD (red) at $z = 5$ (left-hand panel) and $z = 3$ (right-hand panel). Top axes are labelled with H I column density computed from H number density using equation (5). Top: hydrogen number density distribution. TD has more gas at $0.01 < n_H < 1\,cm^{-3}$ than SIMBA where the majority of DLAs are expected to form. This is shown by the dashed vertical line, corresponding to the DLA column density threshold. The density distribution in SIMBA extends to much higher densities $\sim 10^3\,cm^{-3}$ that are two orders of magnitude higher than the highest in TD, due to the lower wind speed and H_2-regulated SF model. Middle: the 50th percentile of neutral and molecular hydrogen fractions as a function of hydrogen number density. Both simulations show a sharp increase in $f_{H\,I}$ (solid lines) at $0.01\,cm^{-3}$. The $f_{H\,I}$ in SIMBA decreases rapidly at $10\,cm^{-3}$ where most of the gas is being converted to molecular as indicated by the increase of f_{H_2} (dashed line). TD does not form molecular hydrogen. Bottom: total metallicity as a function of hydrogen number density. The dark shaded areas encompass the 15.9 per cent and 84.1 per cent (i.e. $\sim$1-σ level) of the metallicity distribution whereas the dashed lines represent the 50th percentile. The faster winds in TD eject metals to larger distances and less denser regions than in SIMBA, indicating TD's ability to form more metal-rich DLAs.

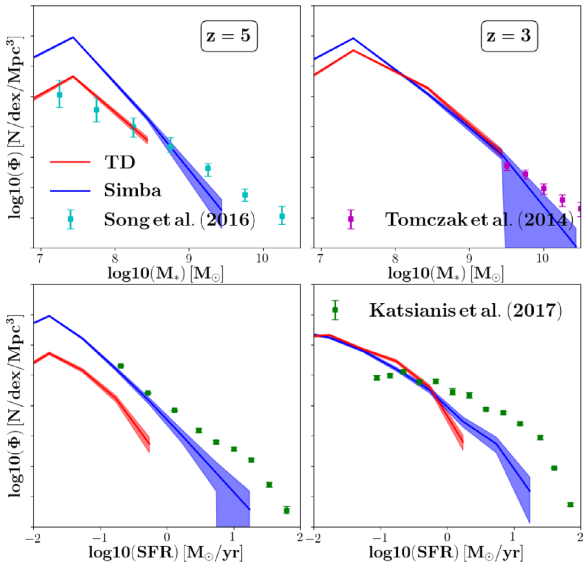

Figure 4. Comparison between SIMBA (blue) and TD (red) against the measured stellar mass functions (top) and SFR functions (bottom) at $z = 5$ (left-hand panel) and $z = 3$ (right-hand panel). The shaded areas reflect the Poisson errors. At $z = 3$, both simulations, regardless of the adopted star formation model and feedback effects, have relatively similar stellar mass and SFR functions at $M_* \leq 10^{9.5}\,M_\odot$ and $SFR \leq 1\,M_\odot\,yr^{-1}$. The stellar mass function in SIMBA extends to larger masses up to $\sim 10^{10.5}\,M_\odot$ that are one order of magnitude larger than the highest stellar mass resolved in TD, owing to the low wind speed in SIMBA and the H_2-regulated SF model. At higher redshift $z = 5$, SIMBA produces stellar mass and SFR functions that are $\sim\times 3$ higher than those of TD, and the ability of SIMBA to resolve larger stellar systems is still seen in this epoch. This indicates that TD applies stronger feedback ($\sim 3\times$ faster wind speed than SIMBA) that suppresses SFR at small stellar masses as compared with SIMBA. The agreement between our simulations at $z = 3$ is largely due to TD's weaker UVB effects as depicted in Fig. 5. In general, our simulations both are within $\sim$1–3σ level of measurements.

consequence of implementing the H_2-regulated recipe to form stars. This feature is less important in TD, where the hybrid multiphase model triggers star formation based solely on the total gas density. The total metallicity as a function of gas density in our simulations is depicted in the bottom panel, in which the effect of wind speed is prominent. The dark shaded areas encompass the 1-σ level about the 50th percentile that is represented by the dashed lines. The faster winds in TD eject metals to larger distances and less denser regions than in SIMBA. This indicates that TD would be able to form more metal-rich DLAs than SIMBA.

3.2 Star formation and stellar mass growth

Before examining the DLA population, we would like to ensure that both TD and SIMBA are producing a reasonable population of high-z galaxies. To assess the galaxy population, we compare their stellar mass and star formation rate functions at $z = 5$ and $z = 3$ to each other as well as to observations. The shaded areas reflect the Poisson errors. Differences in stellar growth are most sensitive to variations in feedback prescriptions. Fig. 4 shows the stellar mass (top panels) and SFR (bottom) functions in our simulations at $z = 5$ and $z = 3$. At higher redshifts, SIMBA forms more stars than TD, resulting in higher stellar mass and SFR functions by a factor of $\sim$3. Even though the outflow mass loading factor is $\sim\times 2$ higher in SIMBA than TD which should reduce stellar growth, the wind velocities are $\sim\times 3$ lower, which evidently yields significantly more wind recycling at early times that counters this and enables rapid early growth.

By $z = 3$, both models are in fairly good agreement. The main difference is that, owing to its rapid early growth, SIMBA produces some fairly large galaxies with $M* \gtrsim 10^{10}\,M_\odot$ with $SFR \gtrsim 10\,M_\odot\,yr^{-1}$, whereas TD does not. As such, the cold gas content that gives rise to DLAs in these two models can be robustly compared, as there has not been a strong difference in the amount of cold gas converted into stars. On the other hand, Pawlik & Schaye (2009) (and later confirmed by Finlator, Davé & Özel 2011) have previously shown that outflows and the UVB couple non-linearly to suppress inflows into haloes at all masses, which in turn suppresses star formation. Put differently, the UVB's impact on galaxy growth is stronger in the presence of outflows. It is reasonable to suppose that, if the UVB thus suppresses inflows, then it also suppresses the DLA abundance; we will return to this point below. At this epoch

Downloaded from https://academic.oup.com/mnras/article/492/2/2835/5698819 by University of Western Cape user on 12 February 2021

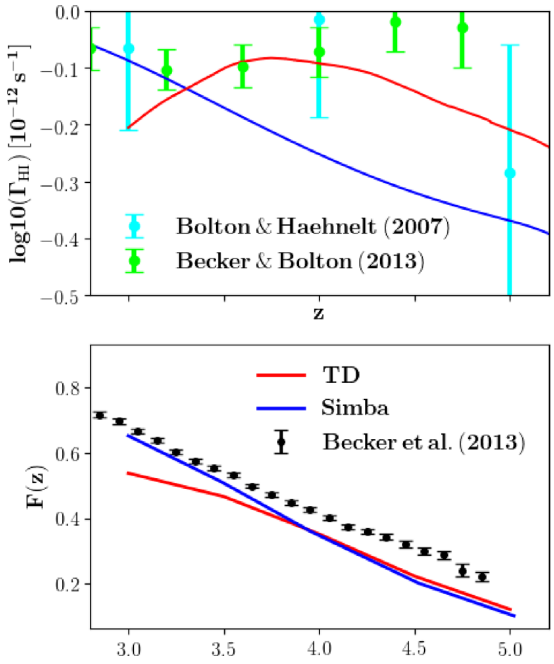

Figure 5. Top: H I photoionization rate evolution over redshift. TD models a spatially inhomogeneous UVB, whereas SIMBA adopts the homogeneous Haardt & Madau (2012) UVB. At $z > 3.5$, TD has about $\sim$1.5 $\times$ higher photoionization rate than SIMBA, and yet the mean transmitted flux is the same in both simulations. This might also be due to the TD's high photon escape fraction at these epochs. Bottom: Mean transmitted flux in Ly α forest comparison between SIMBA (blue), TD (red), and the measurements by Becker et al. (2013). Both simulation have lower transmitted flux than observed, indicating that the UVB is weak, temperature is low, and the opacity is high. TD systematically underproduces the Haardt & Madau (2012) predicted mean transmission for z<4, which partially contributes to the DLA over abundance at these epochs.

$z = 3$, TD is able to form more stars, similar to SIMBA, due to TD's weaker UVB (see Fig. 5).

Comparing the models to observations, there are clearly some discrepancies, and neither model agrees well with all the data. The most reliable observations among those shown here are the Tomczak et al. (2014) stellar mass functions, as they are obtained from rest-optical observations. The Song et al. (2016) results are derived from rest-UV data, from which getting a stellar mass can be sensitive to many uncertainties regarding stellar populations and IMF. Similarly, the Katsianis et al. (2017) results come from rest-UV data which are sensitive to extinction in large galaxies, and stellar populations in small galaxies. In general, both models fall short in predicting the abundance of the highest SFR and M_* galaxies. This is in part explainable by the small volume; in the case of SIMBA, the $100h^{-1}$ Mpc run presented in Davé et al. (2019) agrees well with the stellar mass functions out to $z = 6$. However, it is unclear that simple cosmic variance explains all the discrepancy. Given the various uncertainties in both the observations and the simulations, we will not draw strong conclusions regarding any discrepancies, but rather defer a more careful comparison of the galaxy population in observational space to future work.

3.3 Ultraviolet ionizing background (UVB) treatment

As described earlier, TD self-consistently implements a radiative transfer to model a multifrequency spatially inhomogeneous UVB, while SIMBA includes no radiative transfer routine to generate its own UVB field but rather uses the homogeneous Haardt & Madau (2012) background. The UVB spatial distribution and strength affects the neutral fraction along a sightline and hence the Ly α transmitted flux. We explore the differences in the UVB between these two simulations by examining the H I photoionization rate $\Gamma_{\rm H\,I}(z)$, and its resulting impact on the mean transmitted flux in the Ly α forest. While this does not significantly impact DLAs since they come from dense self-shielded gas, it provides an interesting comparison between the self-consistently generated $\Gamma_{\rm H\,I}$ in TD versus that in Haardt & Madau (2012).

Fig. 5, top panel, shows the H I photoionization rate $\Gamma_{\rm H\,I}$ as a function of redshift in SIMBA (blue) and TD (red) against inferred measurements from the Ly α forest by Bolton & Haehnelt (2007) and Becker & Bolton (2013), shown in cyan and green, respectively.

Both simulations have a $\Gamma_{\rm H\,I}$ that is fairly consistent with observations, given the large uncertainties at high redshifts. SIMBA $\Gamma_{\rm H\,I}$ taken from Haardt & Madau (2012) shows a steady increase from $z = 5 \rightarrow 3$. The self-consistent modelling of RT in TD predicts a higher $\Gamma_{\rm H\,I}$ by a factor of $\sim\times$1.5 versus SIMBA at $z \gtrsim 4$. This might be due to TD's high photon escape fraction at these epochs, compared to what is assumed in Haardt & Madau (2012). At $z \lesssim 3.5$, TD predicts that $\Gamma_{\rm H\,I}$ turns over. This may be because of a lack of high mass galaxies in the small volume to self-consistently generate ionizing photons, or else the small contribution from AGN which begin to be an important contributor to $\Gamma_{\rm H\,I}$ at these epochs.

Fig. 5, bottom panel, shows the mean transmitted flux in the Ly α forest TD and SIMBA at these redshifts as the red and blue lines, respectively, versus observations as compiled by Becker et al. (2013). Following Becker et al. (2013), we define the IGM as pixels with column density $N_{\rm H\,I} < 10^{19}$ cm^{-2}, to compute the mean transmitted flux from only diffuse and high-ionized absorbers.[4] Broadly, the mean transmitted flux is most sensitive to marginally saturated lines, i.e. $N_{\rm H\,I} \sim 10^{14}$ cm^{-2}, since above this column density the lines enter the logarithmic portion of the curve of growth.

The mean transmitted flux increases with time, as the Universe expands and its density drops. The rate of increase is similar in both models, and is comparable to observations. However, we see that both simulations somewhat underproduce the mean Ly α transmission at all redshifts, which suggests the UVB in both simulation is slightly too weak. It has been noted previously that the Haardt & Madau (2012) UVB underproduces mean transmission (BosmUnman et al. 2018; Finlator et al. 2018). Gnedin, Becker & Fan (2017) find qualitatively similar results within a different radiative hydrodynamic simulation. The top panel suggests that the predicted UVB in TD is consistent with observations, so it is not entirely clear why the mean transmitted flux is different. For SIMBA, the low photoionization rate directly translates to too little transmission by a similar factor.

At $z < 4$, the mean transmitted flux starts to increase in SIMBA and becomes in a good agreement with measurements at $z = 3$. Meanwhile, by $z = 3$, TD under-predicts the observed mean transmission by a factor of $\sim$1.5, indicating that the UVB is too weak and opacity is too high. This may play into the DLA statistics at some level.

[4]We have checked that including the DLA profiles does not change the resulting mean transmitted flux, since DLAs are rare.

Downloaded from https://academic.oup.com/mnras/article/492/2/2835/5698819 by University of Western Cape user on 12 February 2021

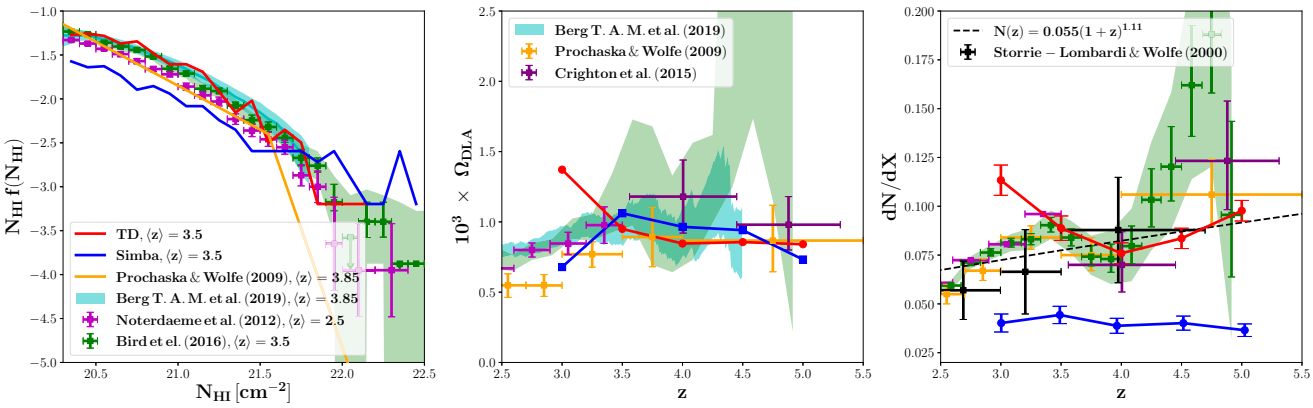

Figure 6. Comparison between SIMBA and TD simulations with measurements by Berg et al. (2019, cyan), Bird, Garnett & Ho (2017, green), Noterdaeme et al. (2012, magenta), Prochaska & Wolfe (2009, orange), Crighton et al. (2015, purple), and Storrie-Lombardi & Wolfe (2000, black), in terms of the column density function distribution (left-hand panel), the H I density evolution (middle), and the abundance evolution of DLAs (right-hand panel). Error bars and shaded areas represent the 1-σ and 2-σ levels of measurements, respectively. Mean redshift of each sample is quoted in the legend. SIMBA underpredicts the observed abundance by a factor of 2, although uncertainty increases with redshift. TD produces a consistent abundance with observations, albeit DLAs are overproduced at $z = 3$ by a factor of 2. Similar trends are seen in the column density distribution function, except SIMBA resolves higher column density systems at $\geq 10^{22}$ cm^{-2}, which drives the increase seen for the DLAs H I density at the mean redshift. DLAs underproduction are partially due to the difference in the outflows feedback strength and perhaps the star formation recipe.

4 DLA ABUNDANCE

In this section, we test the viability of our simulations to reproduce the DLA observations for abundance evolution, the column density distribution, and neutral density evolution.

4.1 DLA Abundance evolution

The DLA abundance ($\frac{\mathrm{d}N}{\mathrm{d}X}$) is the number of DLAs identified in the simulation volume at redshift z for a survey width Δv corresponding to an absorption length dX, alternatively called, the line density of DLAs per comoving absorption length dX. The DLA abundance is mainly driven by the abundance of the low column density systems, since the high column density systems are quite rare.

To compute the dX for our mock survey at a redshift z, we first convert the survey velocity width ($\Delta v = 4 \times 10^7$ km s^{-1}) into the redshift interval using the relation: $\mathrm{d}z = \Delta v(1+z)/c$, where c is the speed of light. The absorption length dX is then computed as follows:

$$\frac{\mathrm{d}X}{\mathrm{d}z} = (1+z)^2 \frac{H_0}{H(z)} = (1+z)^2/\sqrt{\Omega_m(1+z)^3 + \Omega_\Lambda}. \quad (6)$$

We now compare the $\frac{\mathrm{d}N}{\mathrm{d}X}$ from SIMBA (blue errorbars) and TD (red errorbars) with observations in the right-hand panel of Fig. 6. The black errorbars and their dashed black line fit are the early measurements by Storrie-Lombardi & Wolfe (2000), orange errorbars are complied by Prochaska & Wolfe (2009) using the SDSS DR5, magenta errorbars are measurements by Noterdaeme et al. (2012), green errorbars are the measurements by Bird et al. (2017) using SDSS DR12 survey (Garnett et al. 2017), purple errors are measurements by Crighton et al. (2015) using the Giant Gemini GMOS survey, and cyan errors are the most recent measurements reported by Berg et al. (2019) using $XQ-100$ survey. We find that SIMBA underpredicts the observed DLA abundance at $z>3$ approximately by a factor of 2, but still within $1-3\sigma$ level of the observations, particularly at $z > 4$ where the observational uncertainty is large. The TD simulation, on the other hand, is more consistent with the current DLA abundance estimates at $z \geq 3.5$. Below this redshift, the TD simulation starts to overproduce DLAs as seen at $z = 3$. This overproduction of DLAs might correlate with the weak UVB and the high IGM opacity that TD predicts at $z = 3$, as seen in Fig. 5. This effect has previously been found in Bird et al. (2014), where they conclude that an increased UVB amplitude reduces the DLA cross-section, and suppresses the DLA abundance.

Although these simulations have the same mean transmitted flux at $z \geq 4$, still SIMBA underproduces DLAs by a factor of 2, as compared with TD and observations, which implies that the UVB treatment cannot solely explain the differences seen in DLA abundance evolution. However, the remarkable differences seen in the ISM density distribution in Fig. 3 as well as the stellar mass functions in Fig. 4 at $z = 5$, all together indicate that the implemented star formation recipes and feedback effects mainly contribute to the under/overproduction of DLAs as seen in Fig. 6. This can be explained by the difference in the outflows strength. Comparing with SIMBA, it appears that reducing the outflows rate by a factor of 2 as well as boosting the wind speed by a factor or 3 (see Table 1) both suppress the SFR at $z = 5$, and then induces more DLAs in TD. This effect has been previously noted in Faucher-Giguère et al. (2015), where stronger outflows feedback were found to suppress SFR, and enhance the H I covering fractions.

4.2 DLA column density distribution

The column density distribution function (CDDF) is defined as the number of DLAs per unit column density (d$N_{\mathrm{H\,I}}$) per unit comoving absorption length (dX). We compare the CDDF from SIMBA and TD simulations with measurements by Berg et al. (2019, cyan), Bird et al. (2017, green), Noterdaeme et al. (2012, magenta), Crighton et al. (2015, purple), and Prochaska & Wolfe (2009, orange) in the left-hand panel of Fig. 6. For consistent comparison, we only consider DLAs from simulations at $z = 3$, 3.5, 4.0, with $\langle z \rangle = 3.5$ that is nearly equal to the mean redshift of these various measurements as quoted in the legend.

Downloaded from https://academic.oup.com/mnras/article/492/2/2835/5698819 by University of Western Cape user on 12 February 2021

Here we see similar trends as with the $\frac{dN}{dX}$ panel. At $N_{HI} < 10^{22}\,cm^{-2}$, TD produces a consistent CDDF with observations, whereas SIMBA is lower by a factor of 2. This result was anticipated by Fig. 3, where stronger outflows (as in TD) boost the hydrogen number density PDF at DLA column densities. In contrast, suppressed outflows (as in SIMBA) leave more gas in high column density systems ($>10^{22}\,cm^{-2}$). This increase at high column densities is too small to affect the overall abundance ($\frac{dN}{dX}$), which is dominated by low-column systems. Nonetheless, it boosts the CDDF appreciably at higher columns.

Our simulations as well as some observations (Bird et al. 2017; Berg et al. 2019) show no turn over for the CDDF at high column density end at about $N_{HI} = 10^{21.5}\,cm^{-2}$, which was initially suggested by Schaye (2001b), and later predicted by Altay et al. (2011) and Bird et al. (2014), to occur due to molecular hydrogen transition that is responsible to set the maximum H I column density, and hence steepening the CDDF. This turn over was also previously measured, for example, by Prochaska & Wolfe (2009), in which a double power law was used to fit the CDDF as shown by the orange solid line. Unlike TD, SIMBA explicitly includes molecular formation to model star formation via the H_2-regulated SFR recipe (see Fig. 3), and yet show no turn over at the high column density. It remains interesting to test whether the observed turn over at high column densities appears in larger simulation volumes that capture more massive haloes and more high column density DLAs, whose observational selection can be affected by a dust bias (Krogager et al. 2019).

4.3 Neutral density evolution in DLAs

We define total H I density in DLAs, Ω_{DLA} as follows:

$$\Omega_{DLA} = \frac{m_p\, H_0}{c\, \rho_c} \frac{\sum_i N_{HI,i}}{dX}, \quad (7)$$

where m_p is the proton mass and ρ_c is the critical density. Unlike the $\frac{dN}{dX}$, the Ω_{DLA} is weighted towards high column density systems. We now compare the Ω_{DLA} evolution over redshift from our simulations with measurements by Prochaska & Wolfe (2009, orange), Noterdaeme et al. (2012, magenta), Crighton et al. (2015, purple), Bird et al. (2017, green), and Berg et al. (2019, cyan) in the middle panel of Fig. 6. We here find that both simulations are consistent with measurements, except TD overpredicts the H I density in DLAs at $z = 3$, which is consistent with overproduction of DLAs at this redshift (see the right-hand panel of Fig. 6). While it underestimates the CDDF and $\frac{dN}{dX}$ by a factor of $\sim$2, SIMBA is in a good agreement with Ω_{DLA} observations due to the simulation ability to resolve more high column density systems as seen in the CDDF panel. Note that if we use the same column density cut-off when computing Ω_{DLA} for TD and SIMBA, particularly if we use the maximum column density obtained in TD as a cut-off for both simulations, SIMBA then underpredicts the Ω_{DLA} by a factor or 2 at these redshifts. The fact that both simulations have relatively similar H I density evolution over redshift and yet SIMBA underproduces DLAs is suggestive that DLA cross-section in SIMBA is lower than in TD. We leave exploring the DLAs connection to their hosting haloes/galaxies properties to a follow-up work.

4.4 Effect of UVB on Ly α transmitted flux and DLA abundance

In order to determine whether the overproduction of DLAs in TD at $z = 3$ can be attributed entirely to the weak UVB, we rescaled the simulated UVB in post-processing and re-generated our synthetic DLA catalogue under the assumption that the gas is in ionization equilibrium with the adjusted UVB. The gas temperature was left unchanged, and the effect of self-shielding was re-computed for consistency with the new UVB. Fig. 7 summarizes the results of this experiment. This figure shows the mean transmitted flux in Ly α forest as function of DLA abundance for TD (red squares) for different UVB scale factors as quoted next to each point. SIMBA's prediction at $z = 3$, without scaling the Haardt & Madau (2012) UVB, is shown by the blue square. Error bars reflect the Poisson uncertainty. The dark and light green shaded boxes show the corresponding 1-σ and 2-σ of observations by Becker et al. (2013) and Bird et al. (2017), respectively.

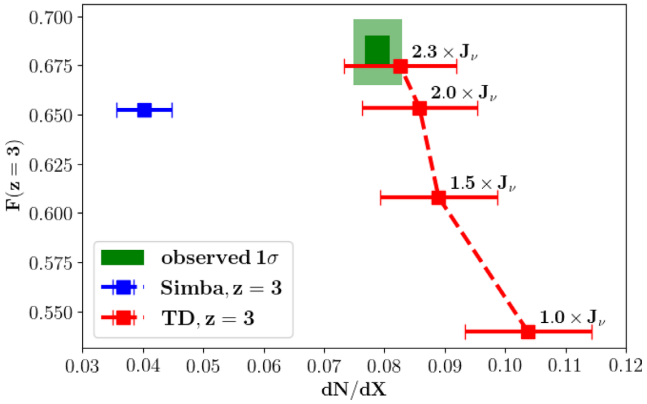

Figure 7. Mean transmitted flux in Ly α forest from TD (red square) as a function of DLA abundance at $z = 3$, for different UVB scale factors as quoted next to each point. The output from SIMBA at $z = 3$, without scaling the Haardt & Madau (2012) UVB, is shown by the blue square. Error bars reflect Poisson uncertainty. The dark and light green shaded box correspond to 1-σ and 2-σ uncertainties from Becker et al. (2013) on the y-axis and Bird et al. (2017) on the x-axis, respectively. The UVB is rescaled in post-processing assuming ionization equilibrium. A stronger UVB boosts the Ly α transmitted flux and reduces the DLA abundance. TD reproduces both observations at z = 3 when we boost the predicted UVB amplitude by a factor of 2.3×. In contrast, scaling the UVB in SIMBA would only improve the model prediction for either the Ly α transmitted flux or DLA abundance. Stronger outflows feedback is clearly needed in the case of SIMBA in order to approach the green box.

In qualitative agreement with Bird et al. (2014), we find that scaling our simulated UVB amplitude up by a factor of 2.3 brings both TD's predicted mean transmission in the Ly α forest and DLA abundance into agreement with observations. In contrast, while adopting a weaker UVB than the Haardt & Madau (2012) model might alleviate the discrepancy between SIMBA and DLA abundance observations, it would exacerbate the discrepancy between the SIMBA predictions and the Ly α transmitted flux measurements. This indicates that stronger outflows feedback is needed for SIMBA to reproduce both measurements. These comparisons support previous suggestions that the DLA abundance is sensitive both to the efficiency with which galaxies eject gas into their CGM, and to the UVB amplitude.

5 DLA METALLICITY

In this section, we use the observed DLA metallicity as a probe to the star formation models implemented in our simulations. We define the DLA metallicity [M/H] as the N_{HI}-weighted metallicity within a window of 500 km s^{-1} about the DLA centroid. To compute

Downloaded from https://academic.oup.com/mnras/article/492/2/2835/5698819 by University of Western Cape user on 12 February 2021

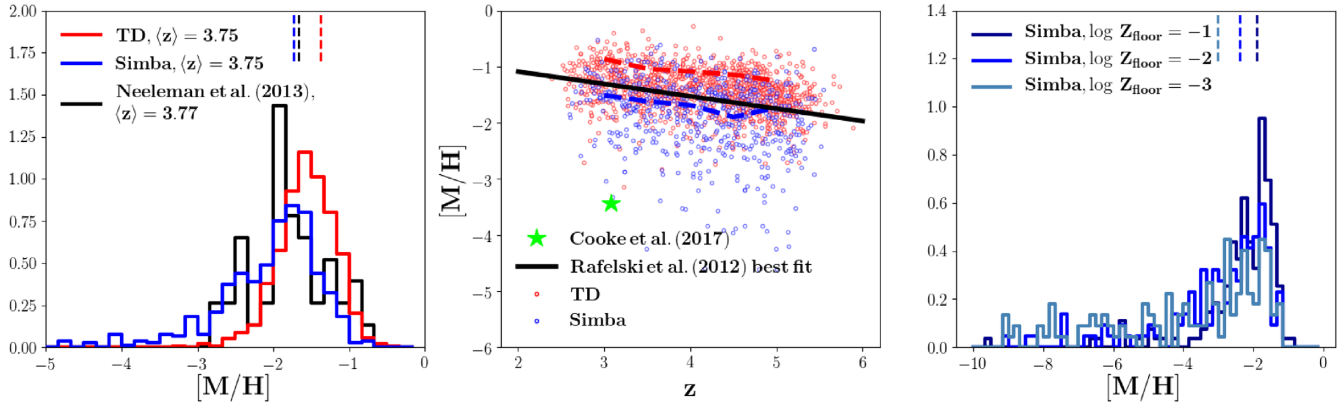

Figure 8. Comparison between the predicted and the observed metallicity distribution and evolution. Left-hand panel: the metallicity distribution from SIMBA (blue) and TD (red) as compared with Neeleman et al. (2013) data. Middle: the metallicity evolution over redshift from our simulations compared with Rafelski et al. (2012). Right-hand panel: the effect of chaning $Z_{\rm floor}$ in the H_2-regulated SFR model in SIMBA on the metallicity distributions. Both simulations agree with the observed metallicity evolution, predicting a negative correlation between metallicity and redshift. SIMBA predicts the existence of extremely low metallcitiy DLAs that partially disappears at high $Z_{\rm floor}$ values. SIMBA provides an opportunity to study the nature of the low metallicity DLA systems, whereas TD enables studying the nature of dusty DLAs at high columns.

the metallicity in pixels along the sightline, we consider all metal species (C, N, O, Ne, Mg, Si, S, Ca, Fe) in SIMBA and consider only four metals (C, O, Si, Mg) in TD that are normalized differently to solar by their corresponding fractions of 0.0134 and 0.00958, respectively. However, these four metals already represent more than 70 per cent out of total metallicity tracked in TD, and it has been shown that with these four metals, TD reproduces the DLA metallicity measurements at z$\sim$ 5 (Finlator et al. 2018).

We compare the predicted and observed DLA metallicity distribution and evolution over redshift in Fig. 8. We begin our discussion here with the metallicity evolution over redshift as seen in the middle panel, where DLAs from TD and SIMBA are shown by red and blue circles, respectively, and the dashed lines show their corresponding $N_{H\,I}$-weighted mean metallicity out of all DLAs at each redshift, following Rafelski et al. (2012). Only for display purpose, we have added a scatter drawn from Gaussian distribution of zero mean and 0.25 standard deviation around each DLA redshift. Best-fitting line from measurements by Rafelski et al. (2012) is represented by the black solid line which indicates that the DLA metallicity decreases with increasing redshift. The most-poor DLA metallicity reported by Cooke, Pettini & Steidel (2017) at z$\sim$ 3 is shown by the green lime star. We here see that both simulations agree with the measured evolution and their corresponding mean metallicity (solid) lines have a similar negative slope to the measurements. TD has a higher running median amplitude than SIMBA by a factor of 1.3. SIMBA predicts the existence of DLAs with extremely low-metallicities, that are much lower than that of the most metal-poor DLA observed to-date (Cooke et al. 2017), as shown by blue circles below the green lime star. This also clearly appears in the metallicity PDF (left-hand and right-hand panels), indicating that SIMBA provides an opportunity to study the nature of the extremely low metallicity DLA systems such as that of Cooke et al. (2017). The formation of these metal-poor DLAs in SIMBA might be due to the weaker winds that induce a rapid drop of metallicities at higher densities as seen in Fig. 3.

In the left-hand panel, we show the metallicity distribution from SIMBA (blue) and TD (red) as compared with Neeleman et al. (2013) data. To establish a proper comparison between our simulations and observations, it is important here to match the mean redshift between samples, due to the metallicity evolution with redshift (Prochaska et al. 2003). We exclude DLAs with z $\leq$ 3 and the only DLA at z$\sim$5 from Neeleman et al. (2013) observational sample, resulting in mean redshift of $\langle z\rangle = 3.77$. From our simulations we exclude $z = 5$ DLAs to obtain a mean redshift of $\langle z\rangle = 3.75$. The minimum and maximum metallicity in Neeleman et al. (2013) sample here are [M/H] $=-2.56$ and -0.64, respectively. We see similar trends here that SIMBA has more low-metallicity DLAs than TD and the measurements. Specifically, SIMBA and TD have about $\sim$ 13 per cent (58) and 0.3 per cent (4) out of all DLAs that have metallicities lower than the observed minimum. Interestingly, TD overproduces and SIMBA underproduces the very high metallicity systems. Vertical tickmarks show the median metallicity for each sample. SIMBA has a median metallicity of -1.74 that is more consistent with the observed median of -1.67 than the median predicted by the TD which is -1.38. This is probably due to the fact that the underproduction of high metallicity systems in SIMBA balances the existence of the extremely low-metallicity systems, resulting in a median that is consistent with observations. Given the metallicity evolution, we expect that our simulations will be able to produce more high metallicity systems at lower redshifts when more massive haloes are formed. Assuming that the number of low-metallicity systems in both simulations is small, SIMBA is in a good agreement with the observed metallicity distribution, whereas TD is skewed more towards high metallicity systems. This is largely due to the feedback effects as seen in the last panel in Fig. 3. The $\sim$3 $\times$ faster wind speed adopted by TD pushes metals to larger distances, where most of DLAs form ($n_{\rm H} = 0.01 - 1\,{\rm cm}^{-3}$) and further contribute to their metal enrichment.

The existence of the very low-metallicity systems in SIMBA might be a consequence of the star formation model. To investigate whether the H_2-regulated star formation model implemented in SIMBA pushes DLAs to have very low-metallicity values, we vary the metallicity floor ($Z_{\rm floor}$) which is the initial seed metallicity necessary to switch on star formation in this model. Since the used SIMBA run for these comparisons adopts $\log Z_{\rm floor} = -2$, we run $12.5\,h^{-1}$ Mpc volume of SIMBA with 2×128^3 dark matter and gas particles each, with three different values of the metallicity floor $\log Z_{\rm floor} = -1, -2$, and -3. We now show the impact of changing $Z_{\rm floor}$ on the DLA metallicites in the right-hand panel of Fig. 8. The dark-blue, blue, and steel-blue are the runs with $\log Z_{\rm floor}$

Downloaded from https://academic.oup.com/mnras/article/492/2/2835/5698819 by University of Western Cape user on 12 February 2021

$= -1$, -2, and -3, respectively, and the vertical tickmarks show the corresponding median metallicity for each run. It is evident that the DLA metallicities overall increase with increasing $Z_{\rm floor}$ as shown by the median vertical tockmarks, each of which increases by approximately 0.5 dex for one order-of-magnitude increase in $Z_{\rm floor}$. The low-metallicity DLAs start to disappear at high values of $Z_{\rm floor} = 0.1$, but not completely. This indicates that the feedback effects partially contribute to the low metallicity in DLAs. The increase in $Z_{\rm floor}$ slightly increases the DLA abundance, and we find that a very high value for the $Z_{\rm floor} > 1$ is required to match the observed DLA abundance. We note that changing the $Z_{\rm floor}$ value has a minimal impact on the stellar mass function. However, a value of $\log Z_{\rm floor} = -2$ is already high since $\log Z_{\rm floor} = -3$ is the commonly used value in these molecular hydrogen SFR models (e.g. Kuhlen, Madau & Krumholz 2013) as motivated by Wise et al. (2012) numerical simulations that follow the transition from Population III to Population II star formation. This comparison indicates that the tail of extremely low-metallicity DLAs in SIMBA is at least partly an artefact of the H_2-regulated SFR model. Alternatively, if extremely low-Z DLAs exist, then they may be the ancestors of ultrafaint dwarf galaxies (Cooke et al. 2017), in which case SIMBA enables study of their kinematics and nature. We leave these questions for future work.

6 CONCLUSIONS

We have examined the DLA properties in two state-of-the-art cosmological hydrodynamic simulations: SIMBA and TECHNICOLOR DAWN. Starting from the same initial conditions, our two simulations were each run down to $z = 3$. We have generated mock DLA profiles and their associated metal lines in the redshift range $z = 3-5$. The simulations adopt different recipes to form stars, implement galactic feedback, and treat the UVB as summarized in Table 1.

Our two key findings are summarized as follows:

(i) SIMBA underpredicts the observed DLA abundance by a factor of $\sim$2, whereas TD is more consistent with the measurements (see right-hand panel in Fig. 6), particularly when post-processing corrections to the UVB amplitude are taken into account (see Fig. 7). This underproduction of DLAs is largely due to the SIMBA's weak feedback effects as compared to TD (see Table 1), which in turn boosts the star formation (see left-hand panel in Fig. 4) and suppresses the DLA incidence rate. Similar trends are seen in the column density distribution function (see left-hand panel in Fig. 6), except SIMBA resolves much higher column density DLAs than TD, as SIMBA continues to suppress feedback in massive galaxies. This results in a good agreement for both simulations with the observed H I density (see middle panel in Fig. 6).

(ii) SIMBA is more consistent with the observed DLA metallicity distribution, whereas TD is skewed towards high metallicity systems (see left-hand panel in Fig. 8). SIMBA further predicts a population of DLAs with metallicities much lower than any observed to date (e.g. Cooke et al. 2017). This population is sensitive to the details of the H_2-regulated SFR model (see right-hand panel of Fig. 8). Both simulations agree with observed slope of DLA metallicity evolution with redshift (see middle panel in Fig. 8).

Our comparisons are entirely limited to the simulation resolution and dynamic range. More DLAs are usually found in higher resolution set-up, such as in zoom-in simulations (see Rhodin et al. 2019). The unique aspect in this study is the intrinsic difference between these simulations in the star formation models and the inhomogeneous UVB treatment. This work sets the stage for more interesting inquiries on the use of DLAs to constrain galaxy formation models. Future inquiries will include:

(i) Exploring DLA kinematics in relation to the hosing properties between both simulations.

(ii) Studying the nature of metal-poor DLAs in connection with ultrafaint dwarf galaxies at high redshift.

(iii) Stuyding the nature of dusty DLAs at high column densities, and the effect of dust bias in DLA selection.

Our results have already shown how DLA observations can play a key role to constraining the star formation recipes and feedback effects in galaxy formation models.

ACKNOWLEDGEMENTS

The authors acknowledge helpful discussions with Joseph Burchett, John Chisholm, David Chih-Yuen Koo, and Neal Katz. We particularly thank the referee, Simeon Bird, for his constructive comments which have improved the paper quality significantly. Simulations and analysis were performed at UWC's PUMBAA, IDIA/ILIFU cloud computing facilities and NMSU's DISCOVERY supercomputer. This work also used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562, and computational resources (Bridges) provided through the allocation AST190003P. RD acknowledges support from the Wolfson Research Merit Award program of the U.K. Royal Society. This work used the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility. The equipment was funded by BEIS capital funding via STFC capital grants ST/P002293/1, ST/R002371/1 and ST/S002502/1, Durham University and STFC operations grant ST/R000832/1. DiRAC is part of the National e-Infrastructure

REFERENCES

Altay G., Theuns T., Schaye J., Crighton N. H. M., Dalla Vecchia C., 2011, ApJ, 737,
Anglés-Alcázar D., Davé R., Faucher-Giguère C.-A., Özel F., Hopkins P. F., 2017a, MNRAS, 464, 2840
Anglés-Alcázar D., Faucher-Giguère C.-A., Kereš D., Hopkins P. F., Quataert E., Murray N., 2017b, MNRAS, 470, 4698
Barnes L. A., Haehnelt M. G., 2009, MNRAS, 397, 511
Becker G. D., Bolton J. S., 2013, MNRAS, 436, 1023
Becker G. D., Hewett P. C., Worseck G., Prochaska J. X., 2013, MNRAS, 430, 2067
Berg T. A. M. et al., 2019, MNRAS, 488, 4356
Berry M., Somerville R. S., Haas M. R., Gawiser E., Maller A., Popping G., Trager S. C., 2014, MNRAS, 441, 939
Bird S., Vogelsberger M., Haehnelt M., Sijacki D., Genel S., Torrey P., Springel V., Hernquist L., 2014, MNRAS, 445, 2313
Bird S., Haehnelt M., Neeleman M., Genel S., Vogelsberger M., Hernquist L., 2015, MNRAS, 447, 1834
Bird S., Garnett R., Ho S., 2017, MNRAS, 466, 2111
Bolton J. S., Haehnelt M. G., 2007, MNRAS, 382, 325
Bosman S. E. I., Fan X., Jiang L., Reed S., Matsuoka Y., Becker G., Haehnelt M., 2018, MNRAS, 479, 1055
Chabrier G., 2003, PASP, 115, 763
Choi E., Ostriker J. P., Naab T., Johansson P. H., 2012, ApJ, 754, 125
Cooke R. J., Pettini M., Steidel C. C., 2017, MNRAS, 467, 802
Crighton N. H. M. et al., 2015, MNRAS, 452, 217
Davé R., Katz N., Oppenheimer B. D., Kollmeier J. A., Weinberg D. H., 2013, MNRAS, 434, 2645
Davé R., Thompson R., Hopkins P. F., 2016, MNRAS, 462, 3265

Downloaded from https://academic.oup.com/mnras/article/492/2/2835/5698819 by University of Western Cape user on 12 February 2021

Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, MNRAS, 486, 2827
Fabian A. C., 2012, ARA&A, 50, 455
Faucher-Giguère C.-A., Hopkins P. F., Kereš D., Muratov A. L., Quataert E., Murray N., 2015, MNRAS, 449, 987
Finlator K., Özel F., Davé R., 2009, MNRAS, 393, 1090
Finlator K., Davé R., Özel F., 2011, ApJ, 743, 169
Finlator K., Keating L., Oppenheimer B. D., Davé R., Zackrisson E., 2018, MNRAS, 480, 2628
Garnett R., Ho S., Bird S., Schneider J., 2017, MNRAS, 472, 1850
Gnedin N. Y., Becker G. D., Fan X., 2017, ApJ, 841, 26
Haardt F., Madau P., 2012, ApJ, 746, 125
Haehnelt M. G., Steinmetz M., Rauch M., 1998, ApJ, 495, 647
Hopkins P. F., 2013, MNRAS, 428, 2840
Hopkins P. F., 2015, MNRAS, 450, 53
Hopkins P. F., 2017, preprint (arXiv:1712.01294)
Jeon M., Besla G., Bromm V., 2019, ApJ, 878, 98.
Katsianis A. et al., 2017, MNRAS, 472, 919
Katz N., Weinberg D. H., Hernquist L., 1996, ApJS, 105, 19
Kennicutt R. C., Jr., 1998, ApJ, 498, 541
Krogager J.-K., Fynbo J. P. U., Møller P., Noterdaeme P., Heintz K. E., Pettini M., 2019, MNRAS, 486, 4377
Kroupa P., 2001, MNRAS, 322, 231
Krumholz M. R., McKee C. F., Tumlinson J., 2009, ApJ, 693, 216
Kuhlen M., Madau P., Krumholz M. R., 2013, ApJ, 776, 34
Lanzetta K. M., Wolfe A. M., Turnshek D. A., Lu L., McMahon R. G., Hazard C., 1991, ApJS, 77, 1
Leroy A. K., Walter F., Brinks E., Bigiel F., de Blok W. J. G., Madore B., Thornley M. D., 2008, AJ, 136, 2782
Li Q., Narayanan D., Davé R., 2019, MNRAS, 490, 1425
Lochhaas C. et al., 2016, MNRAS, 461, 4353
Lusso E., Worseck G., Hennawi J. F., Prochaska J. X., Vignali C., Stern J., O'Meara J. M., 2015, MNRAS, 449, 4204
Manti S., Gallerani S., Ferrara A., Greig B., Feruglio C., 2017, MNRAS, 466, 1160
McKee C. F., Ostriker J. P., 1977, ApJ, 218, 148
Muratov A. L., Kereš D., Faucher-Giguère C.-A., Hopkins P. F., Quataert E., Murray N., 2015, MNRAS, 454, 2691
Nagamine K., Springel V., Hernquist L., 2004, MNRAS, 348, 435
Neeleman M., Wolfe A. M., Prochaska J. X., Rafelski M., 2013, ApJ, 769, 54
Noterdaeme P. et al., 2012, A&A, 547, L1
Pawlik A. H., Schaye J., 2009, MNRAS, 396, L46
Pérez-Ràfols I., Miralda-Escudé J., Arinyo-i-Prats A., Font-Ribera A., Mas-Ribas L., 2018, MNRAS, 480, 4702
Perna M., Lanzuisi G., Brusa M., Mignoli M., Cresci G., 2017, A&A, 603, A99
Pontzen A. et al., 2008, MNRAS, 390, 1349
Prochaska J. X., Wolfe A. M., 1997, ApJ, 487, 73
Prochaska J. X., Wolfe A. M., 1998, ApJ, 507, 113
Prochaska J. X., Wolfe A. M., 2009, ApJ, 696, 1543
Prochaska J. X., Wolfe A. M., 2010, preprint (arXiv:1009.3960)
Prochaska J. X., Gawiser E., Wolfe A. M., Castro S., Djorgovski S. G., 2003, ApJ, 595, L9
Prochaska J. X., Herbert-Fort S., Wolfe A. M., 2005, ApJ, 635, 123
Rafelski M., Wolfe A. M., Prochaska J. X., Neeleman M., Mendez A. J., 2012, ApJ, 755, 89
Rafelski M., Neeleman M., Fumagalli M., Wolfe A. M., Prochaska J. X., 2014, ApJ, 782, L29
Rahmati A., Pawlik A. H., Raicevic M., Schaye J., 2013, MNRAS, 430, 2427
Rhodin N. H. P., Agertz O., Christensen L., Renaud F., Fynbo J. P. U., 2019, MNRAS, 488, 3634
Schaye J., 2001a, ApJ, 559, 507
Schaye J., 2001b, ApJ, 562, L95
Schmidt M., 1959, ApJ, 129, 243
Smith B. D. et al., 2017, MNRAS, 466, 2217
Song M. et al., 2016, ApJ, 825, 5
Springel V., Hernquist L., 2003, MNRAS, 339, 289
Storrie-Lombardi L. J., Wolfe A. M., 2000, ApJ, 543, 552
Theuns T., Leonard A., Efstathiou G., Pearce F. R., Thomas P. A., 1998, MNRAS, 301, 478
Thomas N., Davé R., Anglés-Alcázar D., Jarvis M., 2019, MNRAS, 487, 5764
Tomczak A. R. et al., 2014, ApJ, 783, 85
Wise J. H., Turk M. J., Norman M. L., Abel T., 2012, ApJ, 745, 50
Wolfe A. M., Turnshek D. A., Smith H. E., Cohen R. D., 1986, ApJS, 61, 249
Wolfe A. M., Lanzetta K. M., Foltz C. B., Chaffee F. H., 1995, ApJ, 454, 698
Wolfe A. M., Gawiser E., Prochaska J. X., 2005, ARA&A, 43, 861

This paper has been typeset from a TEX/LATEX file prepared by the author.

Downloaded from https://academic.oup.com/mnras/article/492/2/2835/5698819 by University of Western Cape user on 12 February 2021