Rahmati, A, Schaye, J, Crain, RA, Bower, RG, Furlong, M, Schaller, M and Theuns, T

 The distribution of neutral hydrogen around high-redshift galaxies and quasars in the EAGLE simulation

The distribution of neutral hydrogen around high-redshift galaxies and quasars in the EAGLE simulation

Alireza Rahmati,[1,2]★ Joop Schaye,[3] Richard G. Bower,[4] Robert A. Crain,[5,3]
Michelle Furlong,[4] Matthieu Schaller[4] and Tom Theuns[4]

[1]*Institute for Computational Science, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland*
[2]*Max-Planck Institute for Astrophysics, Karl-Schwarzschild-Strasse 1, D-85748 Garching, Germany*
[3]*Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA Leiden, the Netherlands*
[4]*Institute for Computational Cosmology, Department of Physics, University of Durham, South Road, Durham DH1 3LE, UK*
[5]*Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF, UK*

ABSTRACT

The observed high covering fractions of neutral hydrogen (H I) with column densities above $\sim 10^{17}\,\mathrm{cm}^{-2}$ around Lyman-Break Galaxies (LBGs) and bright quasars at redshifts $z \sim 2$–3 has been identified as a challenge for simulations of galaxy formation. We use the Evolution and Assembly of Galaxies and their Environment (EAGLE) cosmological, hydrodynamical simulation, which has been shown to reproduce a wide range of galaxy properties and for which the subgrid feedback was calibrated without considering gas properties, to study the distribution of H I around high-redshift galaxies. We predict the covering fractions of strong H I absorbers ($N_{\mathrm{H\,I}} \gtrsim 10^{17}\,\mathrm{cm}^{-2}$) inside haloes to increase rapidly with redshift but to depend only weakly on halo mass. For massive ($M_{200} \gtrsim 10^{12}\,\mathrm{M_\odot}$) haloes, the covering fraction profiles are nearly scale-invariant and we provide fitting functions that reproduce the simulation results. While efficient feedback is required to increase the H I covering fractions to the high observed values, the distribution of strong absorbers in and around haloes of a fixed mass is insensitive to factor of 2 variations in the strength of the stellar feedback. In contrast, at fixed stellar mass the predicted H I distribution is highly sensitive to the feedback efficiency. The fiducial EAGLE simulation reproduces both the observed global column density distribution function of H I and the observed radial covering fraction profiles of strong H I absorbers around LBGs and bright quasars.

Key words: methods: numerical – galaxies: formation – galaxies: high-redshift – intergalactic medium – quasars: absorption lines.

1 INTRODUCTION

Galaxies need to acquire large quantities of fresh gas from the intergalactic medium (IGM) to sustain their star formation activities through time (e.g. Bauermeister, Blitz & Ma 2010). Simulations predict that a large fraction of the accreting material enters haloes relatively cold and could therefore contain significant amounts of neutral gas (e.g. Fumagalli et al. 2011; van de Voort et al. 2012). The presence of shock-heated gas complicates the journey of the accreting gas on to galaxies. Moreover, energetic feedback from stars and active galactic nuclei (AGN), which regulate the consumption of the accreted gas and launch galactic outflows, affect the dynamics and chemical composition of gas around galaxies. As

a result, the complex distribution of gas around galaxies contains the finger-prints of the aforementioned processes and studying it, mostly by analysing the absorption signature of neutral hydrogen and metals in the spectra of bright background sources, is of great value for understanding galaxies and the physical processes that regulate them.

Observations and simulations show that both the number of H I absorbers and their typical column densities increase closer to galaxies (e.g. Adelberger et al. 2003; Chen & Mulchaey 2009; Rakic et al. 2012; Rahmati & Schaye 2014; Turner et al. 2014). This suggests that absorbers with higher H I column densities are better probes of the gas in the vicinity of galaxies. Cosmological simulations, however, suggest that most strong H I absorbers, such as Lyman Limit Systems (LLSs; with $N_{\mathrm{H\,I}} \gtrsim 10^{17.2}\,\mathrm{cm}^{-2}$) and Damped Lyman α systems (DLAs; with $N_{\mathrm{H\,I}} \gtrsim 10^{20.5}\,\mathrm{cm}^{-2}$), are close to galaxies that are too faint to be easily detectable in current surveys

(Rahmati & Schaye 2014), which is in agreement with the lack of detected counterparts close to most of strong H I absorbers (e.g. Fumagalli et al. 2015). Not knowing the properties of the host galaxy complicates the use of strong H I absorbers to study the relation between galaxies and their environments. This problem can, however, be circumvented by studying the distribution of H I absorbers around easily detectable bright galaxies.

Several modern observational campaigns have adopted this galaxy-centred approach by using quasar absorption lines to systematically investigate the distribution of neutral hydrogen around massive galaxies at different epochs (e.g. Adelberger et al. 2003; Hennawi et al. 2006; Chen & Mulchaey 2009; Rakic et al. 2012; Prochaska, Hennawi & Simcoe 2013a; Tumlinson et al. 2013; Turner et al. 2014). For instance, Rudie et al. (2012) measured the covering fraction of H I around Lyman-Break galaxies (LBGs) at $z \sim 2$ and found that there is an $\sim$30 per cent chance for finding LLSs in the spectra of background quasars that have impact parameters less than $\sim$100 proper kpc (hereafter pkpc). Noting that this impact parameter is comparable to the virial radii of LBGs at $z \sim 2$, this result implies that LLSs have a covering fraction of 30 per cent within the virial radius of LBGs. Prochaska et al. (2013b) showed that the abundance of H I absorbers is significantly enhanced out to several virial radii from bright quasars at $z \sim 2$. They found that there is a more than 60 per cent chance of finding an LLS within $\sim$150 pkpc from a bright quasar (Prochaska et al. 2013a), which is comparable to the typical virial radius of the haloes with $M_{200} \sim 10^{12.5}$ M$_\odot$ that are expected to host the observed quasars at $z \sim 2$.

Motivated by recent observational constraints, several groups used simulations to study the distribution of H I around galaxies (e.g. Faucher-Giguère & Kereš 2011; Fumagalli et al. 2011, 2014; Rakic et al. 2013; Shen et al. 2013; Erkal 2015; Meiksin, Bolton & Tittley 2014; Faucher-Giguère et al. 2015). However, reproducing the relatively large observed covering fractions of strong H I absorption around massive galaxies turned out to be a major challenge (e.g. Fumagalli et al. 2014; Faucher-Giguère et al. 2015).

Previous studies of high column density H I around massive galaxies were based on the analysis of simulations that zoom into only one galaxy (e.g. Faucher-Giguère & Kereš 2011; Shen et al. 2013), or a handful of galaxies spanning a limited range in mass and redshift (e.g. Fumagalli et al. 2011, 2014; Faucher-Giguère et al. 2015). Given the diversity of observed galaxies, one may expect a large intrinsic variation in the spatial distribution of H I from one galaxy to another. Consequently, a large sample of simulated galaxies is required to predict the average distribution of H I and to compare it robustly with observations. Because observational constraints are at present limited to galaxies residing in relatively massive haloes ($M_{200} \gtrsim 10^{12}$ M$_\odot$), which are rare, especially at high redshifts, simulating a large number of them requires large cosmological volumes. Moreover, without cosmological simulations it is not straightforward to check whether the simulations satisfy other important constraints on the cosmic distribution of H I, for instance the global H I column density distribution function (CDDF) and/or the H I cosmic density, which has been reproduced successfully in recent cosmological simulations (e.g. Altay et al. 2011; McQuinn, Oh & Faucher-Giguère 2011; Davé et al. 2013; Rahmati et al. 2013a; Vogelsberger et al. 2014). The aforementioned issues may limit the power of studies that use small numbers of zoom simulations and indicate that cosmological simulations of representative volumes are needed to study the average distribution of H I around large numbers of galaxies.

The strength and implementation of feedback mechanisms are also crucial factors for simulations of the distribution of gas around galaxies (e.g. Faucher-Giguère et al. 2015; Suresh et al. 2015). For instance, galactic winds driven by stellar feedback can change the distribution of H I around galaxies by carrying the cold neutral gas farther away from galaxies and by providing resistance against the accretion of gas. While feedback implementations vary widely, simulations that use strong stellar feedback have been more successful in reproducing the LLS covering fractions observed around LBGs (e.g. Shen et al. 2013; Faucher-Giguère et al. 2015), which are significantly underproduced in simulations with weak feedback (e.g. Faucher-Giguère & Kereš 2011; Fumagalli et al. 2011, 2014). Moreover, feedback from AGN, which is required to form reasonable galaxies in very massive haloes and is missing from most previous studies, may affect the observed H I distribution in and around haloes expected to host quasars at $z \sim 2$. However, several simulations indicate that only the strongest absorbers, which on average reside closer to or even inside galaxies, are significantly affected by feedback (e.g. Theuns et al. 2002; Altay et al. 2013; Bird et al. 2014; Rahmati & Schaye 2014). Hence, the more important effect may be that feedback changes the relation between stellar mass and halo mass and thus the predicted H I distribution at fixed stellar mass (e.g. Rakic et al. 2013).

In this work, we study the H I distribution around galaxies using state-of-the-art cosmological hydrodynamical simulations. For this purpose, we use the Evolution and Assembly of Galaxies and their Environment (EAGLE) simulations (Schaye et al. 2015, hereafter S15). The large cosmological volume of the main EAGLE run (100 cMpc) together with its relatively high resolution for a simulation of this type, allow us to study large numbers of haloes with masses similar to those targeted by recent observations, without compromising the resolution needed to simulate the distribution of relevant H I systems (e.g. LLSs). Efficient stellar and AGN feedback enables the simulation to successfully reproduce a large number of basic observed characteristics of galaxies over wide mass and redshift ranges (S15; Furlong et al. 2015; Schaller et al. 2015; Crain et al. 2015) and S15 already showed that EAGLE also reproduces the observed present-day column density distributions of C IV and O VI. These factors make EAGLE ideal for studying the gas distribution around galaxies.

We combine the EAGLE simulations with the accurate photoionization corrections from Rahmati et al. (2013a), which are based on high-resolution radiative transfer calculations. After showing the success of the simulation in reproducing the observed cosmic distribution of H I, we look at the H I distribution around galaxies from $z = 4$ to 1, bracketing the era during which the cosmic star formation density peaked. We focus on the strong H I absorbers whose high covering fractions were found to be difficult to reproduce by previous simulations (e.g. Faucher-Giguère & Kereš 2011; Fumagalli et al. 2011, 2014; Faucher-Giguère et al. 2015). We predict that strong H I absorbers, such as LLSs and DLAs, have a mean covering fraction within the virial radius that increases rapidly with redshift, but depends only weakly on the halo mass (or star formation rate) at fixed redshift, suggesting that the distribution of absorbing gas around galaxies has a similar shape for different halo masses. Indeed, we show that the covering fraction of LLSs, sub DLAs and DLAs around massive galaxies ($M_{200} \gtrsim 10^{12}$ M$_\odot$) follows profiles with similar shapes but different scalelengths that are tied to the virial radius and redshift.

We construct samples of simulated galaxies by matching the halo masses and redshifts to those used in the observational studies of Rudie et al. (2012) and Prochaska et al. (2013b) for LBGs and

quasars, respectively. Accounting for the uncertainties in the amplitude of the ultra violet background (UVB) photoionization rate, our predictions are in excellent agreement with the observed H I distributions. This shows that cosmological hydrodynamical simulations that are successful in reproducing reasonable galaxy properties, are also capable of predicting gas distributions in agreement with current observations. We conclude that there is no obvious missing ingredient in our general understanding of galaxy formation and evolution required to explain the observed H I distributions around LBGs and bright quasars at $z \sim 2$.

The structure of this paper is as follows. In Section 2, we introduce our cosmological simulations and discuss the photoionization corrections required for obtaining the H I column densities and calculating their distribution around galaxies. We present our predictions for the H I covering fractions and their evolution in Section 3. We compare the predictions with recent observations in Section 4 and discuss the impact of feedback on our results in Section 5. We conclude in Section 6.

2 SIMULATION TECHNIQUES

In this section, we briefly describe the hydrodynamical simulations that we use to predict the H I distributions. We further explain our halo finding method (Section 2.2) and our photoionization correction required for the H I column density calculations (Section 2.3).

2.1 Hydrodynamical simulations

We use the reference simulation of the EAGLE project, described in S15, as our fiducial simulation. The cosmological simulation was performed using a significantly modified and extended version of the smoothed particle hydrodynamics (SPH) code GADGET-3 (last described in Springel 2005). In particular, we use ANARCHY (Dalla Vecchia, in preparation; see also Appendix A of S15) which is an updated hydrodynamics algorithm incorporating the pressure entropy formulation of SPH derived by Hopkins (2013) and the time-step limiter of Durier & Dalla Vecchia (2012) (see Appendix E where the impact of using ANARCHY is discussed). The subgrid physics used in the simulation is based on that of the OWLS project (Schaye et al. 2010) with numerous important improvements. Stellar and AGN feedback are implemented using the stochastic, thermal prescription of Dalla Vecchia & Schaye (2012), without turning off radiative cooling or the hydrodynamics. Galactic winds develop naturally, without pre-determined mass loading factors or velocities. We use a metallicity-dependent subgrid model for star formation together with the pressure-dependent star formation prescription of Schaye & Dalla Vecchia (2008). The feedback from AGN is updated such that the subgrid model for accretion of gas on to black holes accounts for angular momentum (Rosas-Guevara et al. 2013). A metallicity and density dependent stellar feedback efficiency is adopted to account, respectively, for greater thermal losses when the metallicity increases and for residual spurious resolution dependent numerical radiative losses (Dalla Vecchia & Schaye 2012; Crain et al. 2015). The implementation of metal enrichment is similar to that of the OWLS project and is described in Wiersma, Schaye & Smith (2009a). We follow the abundances of 11 elements assuming a Chabrier (2003) initial mass function. These abundances are used for calculating radiative cooling/heating rates, element-by-element and in the presence of the uniform cosmic microwave background and the Haardt & Madau (2001) UVB model (Wiersma et al. 2009b). The simulation is calibrated based on the present day observed galaxy stellar mass function and galaxy sizes, which

are reproduced with unprecedented accuracy for a hydrodynamical simulation (S15; Crain et al. 2015). The same simulation also shows very good agreement with other observed galaxy properties such as the observed galaxy specific star formation rates, passive fractions, Tully–Fisher relation and the distribution of metals in the IGM (S15; Rahmati et al. in preparation), galaxy rotation curves (Schaller et al. 2015), the evolution of the galaxy stellar mass function (Furlong et al. 2015) and the molecular hydrogen content of galaxies (Lagos et al. 2015).

The adopted cosmological parameters are based on the most recent Planck results: $\{\Omega_{\rm m} = 0.307, \Omega_{\rm b} = 0.048\,25, \Omega_{\Lambda} = 0.693, \sigma_8 = 0.8288, n_{\rm s} = 0.9611, h = 0.6777\}$ (Planck Collaboration I 2014). Our reference simulation, *Ref-L100N1504*, has a periodic box of $L = 100$ comoving Mpc (cMpc) and contains 1504^3 dark matter particles with mass $9.7 \times 10^6\,{\rm M}_\odot$ and an equal number of baryonic particles with initial mass $1.81 \times 10^6\,{\rm M}_\odot$. The Plummer-equivalent gravitational softening length is set to $\epsilon_{\rm com} = 2.66$ comoving kpc (ckpc) and is limited to a maximum physical scale of $\epsilon_{\rm prop} = 0.7$ pkpc. We use simulations with different feedback implementations to test the impact of feedback variations on our results in Section 5. Simulations with different box sizes and resolutions are used to study the impact of those factors on our results in Appendix D. Table 1 summarizes the simulations we used in this work.

2.2 Identifying galaxies in EAGLE

We identify galaxies using the Friends-of-Friends (FoF) algorithm to select groups of dark matter particles that are near each other (i.e. FoF haloes), choosing a linking length of $b = 0.2$. In other words, we assume that galaxies reside in dark matter haloes. In the next step, we group gravitationally bound particles of unique structures (subhaloes) using SUBFIND (Springel et al. 2001; Dolag et al. 2009). We identify the centre of each halo/galaxy as the position of the particle with the minimum gravitational potential in that halo. Then we define the virial radius, r_{200}, as the radius within which the average density of the halo equals 200 times the mean density of the Universe at any given redshift. The mass contained within that radius is then defined as the halo mass, M_{200}. The most massive substructure in each halo is defined as the *central* galaxy. The focus of this study is on the distribution of gas around bright high-redshift galaxies (e.g. LBGs and bright quasars), which are in most cases the brightest and most massive objects in their host haloes, we only consider central galaxies in our analysis.

2.3 H I fractions

For an accurate calculation of the simulated H I column densities, the main ionizing processes that shape the distribution of neutral hydrogen must be taken into account. Besides collisional ionization, which is dominant at high temperatures, photoionization by the metagalactic UVB radiation is the main contributor to the bulk of hydrogen ionization on cosmic scales, particularly at $z \gtrsim 1$ (e.g. Rahmati et al. 2013a). On smaller scales and close to sources, local radiation could be the dominant source of photoionization (see Appendix B and Rahmati et al. 2013b).

In this work, we use the UVB model of Haardt & Madau (2001) to account for the mean ionizing radiation field from quasars and galaxies. This UVB model was also used for calculating radiative heating/cooling rates in the hydrodynamical simulations. Moreover, this UVB model has been shown to reproduce results consistent with

Table 1. List of cosmological simulations used in this work. The first four simulations use model ingredients identical to the EAGLE reference simulation of Schaye et al. (2015), while the higher-resolution *Recal-L025N0752* has been re-calibrated to the observed present-day galaxy mass function. Model *NoAGN* does not include AGN feedback, *WeakFB* and *StrongFB* use half and twice as strong stellar feedback compared to the reference simulation, respectively (see Crain et al. 2015). Models *NoFB* and *NoFB-Gadget*, do not include any feedback from stars or AGN. While the former uses ANARCHY (Dalla Vecchia, in preparation) for the hydrodynamical calculations, the latter uses the standard GADGET-3 implementation (Springel 2005), as does model *Ref-Gadget*. From left to right the columns show: simulation identifier; comoving box size; number of particles (there are equally many baryonic and dark matter particles); initial baryonic particle mass; dark matter particle mass; comoving (Plummer-equivalent) gravitational softening; maximum physical softening, and a brief description.

Simulation	L (cMpc)	N	m_b ($M_\odot$)	m_{dm} ($M_\odot$)	ϵ_{com} (ckpc)	ϵ_{prop} (pkpc)	Remarks
Ref-L100N1504	**100**	**2×1504^3**	**1.81×10^6**	**9.7×10^6**	**2.66**	**0.7**	**Ref. stellar and ref. AGN feedback**
Ref-L050N0752	50	2×752^3	1.81×10^6	9.70×10^6	2.66	0.70	,,
Ref-L025N0376	25	2×376^3	1.81×10^6	9.70×10^6	2.66	0.70	,,
Ref-L025N0752	25	2×752^3	2.26×10^5	1.21×10^6	1.33	0.35	,,
Recal-L025N0752	25	2×752^3	2.26×10^5	1.21×10^6	1.33	0.35	Recalibrated stellar and AGN feedback
NoAGN	25	2×376^3	1.81×10^6	9.70×10^6	2.66	0.70	Ref. stellar and no AGN feedback
WeakFB	25	2×376^3	1.81×10^6	9.70×10^6	2.66	0.70	Weak stellar and ref. AGN feedback
StrongFB	25	2×376^3	1.81×10^6	9.70×10^6	2.66	0.70	Strong stellar and ref. AGN feedback
NoFB	25	2×376^3	1.81×10^6	9.70×10^6	2.66	0.7	No feedback using ANARCHY SPH
NoFB-Gadget	25	2×376^3	1.81×10^6	9.70×10^6	2.66	0.70	No feedback using Gadget SPH
Ref-Gadget	25	2×376^3	1.81×10^6	9.70×10^6	2.66	0.70	*Ref-L025N0376* using Gadget SPH

the observed H I CDDF (Rahmati et al. 2013a) and $z \sim 3$ metal absorption lines (Aguirre et al. 2008). However, we emphasis that both observational constraints and model predictions for the amplitude (Γ_{UVB}) and spectral shape of the photoionizing background are uncertain by a factor of a few (e.g. Faucher-Giguère et al. 2008, 2009; Haardt & Madau 2012; Becker & Bolton 2013), which could change the H I column density distribution. Where appropriate, we consider the impact of those uncertainties on our results by varying the UVB model in our calculations.

At very low H I column densities, where the gas is highly ionized the gas satisfies the so-called optically-thin limit. As the H I column density increases and its corresponding optical depth gets close to unity, photon absorptions become more important and eventually the gas can shield itself against the incoming flux of ionizing photons. To account for this self-shielding, we use the same approach as we adopted in Rahmati & Schaye (2014). Namely, we use the fitting function presented in Rahmati et al. (2013a) for calculating the photoionization rate and hence the ionization state of hydrogen atoms. This fitting function accurately reproduces the result from radiative transfer simulations of the UVB and recombination radiation in cosmological density fields using TRAPHIC (Pawlik & Schaye 2008, 2011; Raičević et al. 2014). One can characterize the UVB at any given redshift by the hydrogen photoionization rate in optically-thin limit, $\Gamma_{UVB, z}$, and the effective hydrogen absorption cross-section, $\bar{\sigma}_{\nu_{H_I}}$. Then the fitting function gives the effective photoionization rate, Γ_{Phot}, as a function of density:

$$\frac{\Gamma_{Phot}}{\Gamma_{UVB,z}} = 0.98 \left[1 + \left(\frac{n_H}{n_{H,SSh}} \right)^{1.64} \right]^{-2.28}$$
$$+ 0.02 \left[1 + \frac{n_H}{n_{H,SSh}} \right]^{-0.84}, \quad (1)$$

where n_H is the hydrogen number density and $n_{H, SSh}$ is the self-shielding density threshold predicted by the analytic model of

Schaye (2001a)

$$n_{H,SSh} = 6.73 \times 10^{-3} \text{ cm}^{-3} \left(\frac{\bar{\sigma}_{\nu_{H_I}}}{2.49 \times 10^{-18} \text{ cm}^{-2}} \right)^{-2/3}$$
$$\times \left(\frac{\Gamma_{UVB,z}}{10^{-12} \text{ s}^{-1}} \right)^{2/3}. \quad (2)$$

Combining equations (1) and (2) allows us to calculate the equilibrium hydrogen neutral fraction of each SPH particle in our hydrodynamical simulations after calculating its collisional ionization rate, which depends on the temperature, and its optically-thin recombination rate (i.e. Case A)[1] which depend on both the density and temperature (see appendix A2 of Rahmati et al. 2013a). Since the temperature of star-forming gas in our simulations is defined by a polytropic equation of state that is used to limit the Jeans mass, and therefore is not physical, we set the temperature of the interstellar medium (ISM) particles to $T_{ISM} = 10^4$ K which is the typical temperature of the warm-neutral ISM.

Noting that the covering fraction of extremely high H I column densities are negligible, we chose not to account for the formation of molecular hydrogen which is expected to be dominant only at $N_{H_I} \gtrsim 10^{22}$ cm^{-2} (Schaye 2001b; Krumholz et al. 2009; Rahmati et al. 2013a). We also neglect the impact of radiation from local sources, which is thought to become increasingly important for very high H I column densities and very close to galaxies (Schaye 2006; Rahmati et al. 2013b). Noting that the distribution of LLSs around the virial radii of galaxies is not strongly affected by local radiation (see Appendix B and Rahmati et al. 2013b; Shen et al. 2013; Rahmati & Schaye 2014) we postpone a treatment of local radiation, which is potentially important but requires complex radiative transfer simulations, to future work.

To calculate H I column densities, we use SPH interpolation and project the H I content of desired regions, which can range from the

---

[1] Note that the impact of recombination radiation is accounted for by the fitting function and there is no need to use the Case B recombination rate.

full simulation box to only a small volume around a galaxy, on to a 2D grid. We found that using a grid with $10\,000^2 = 10^8$ pixels for projecting the full box of the *Ref-L100N1504* simulation results in converged H I covering fractions for $N_{\rm H\,I} \lesssim 10^{22}\,{\rm cm}^{-2}$, which is the range of H I column densities we study in this work. We use 16 slices with equal widths for calculating the H I column densities in the full 100 Mpc simulation box. This enables us to calculate H I column densities as low as $N_{\rm H\,I} \sim 10^{15}\,{\rm cm}^{-2}$ without being affected by projection effects. Moreover, this choice enables us to calculate the covering fraction of H I around simulated galaxies in analogy with what is done observationally, where absorbers are considered to be around a galaxy only if their line-of-sight (LOS) velocity differences from that of the galaxy do not exceed a fixed value.[2] Splitting the full 100 cMpc simulation box into 16 slices results in a velocity width few times smaller than the typical velocity cuts used in observational studies (e.g. $\Delta V \approx 400\,{\rm km\,s}^{-1}$ at $z \sim 2$ for each slice). Adding together appropriate number of slices allows us to efficiently calculate the distribution of H I around galaxies by using velocity cuts comparable to what is typically used in observational studies (e.g. Rudie et al. 2012; Prochaska et al. 2013a,b).

In previous theoretical studies, the covering fraction of H I was usually measured by considering a finite region around galaxies which is often normalized to the virial radius of each galaxy (e.g. Faucher-Giguère & Kereš 2011; Fumagalli et al. 2011, 2014; Shen et al. 2013; Faucher-Giguère et al. 2015). The covering fractions measured using this method only account for the gas that is much closer to galaxies along the LOS compared to what is measured observationally. Observations take into account a much longer path-length along the LOS compared to the virial radii of galaxies and are therefore not directly comparable to the covering fractions reported in previous theoretical studies. However, considering a finite region normalized to the virial radius of each galaxy is a sensible choice to study the distribution of gas that has a close physical connection to galaxies. For this reason, and also to facilitate comparison between our results and previous theoretical work, we not only mimic the observed velocity cuts, but also measure the covering fraction of H I systems around galaxies by considering only absorbers that are within $2 \times r_{200}$ from each galaxy.

3 RESULTS

3.1 Cosmic distribution of H I

Before studying the distribution of H I around simulated galaxies, we need to test whether the EAGLE simulation reproduces the observed cosmic distribution of strong H I systems. If the simulation fails to satisfy the existing constraints on the cosmic abundance of H I absorbers, then it is hard to trust the predictions drawn from it about the distribution of H I close to galaxies. While performing this consistency check is not straightforward for studies that use zoom simulations, using a cosmological simulation enables us to do so.

The cosmic H I distribution is often quantified as the H I CDDF. The H I CDDF, $f(N_{\rm H\,I}, z)$, is conventionally defined as the number of absorbers per unit column density, $dN_{\rm H\,I}$, per unit absorption length, $dX = dz\,(H_0/H(z))(1 + z)^2$, and is measured observationally by searching for H I absorbers in the spectra of background quasars (e.g. Kim et al. 2002, 2013; Péroux et al. 2005; O'Meara et al. 2007, 2013; Noterdaeme et al. 2009, 2012; Prochaska & Wolfe
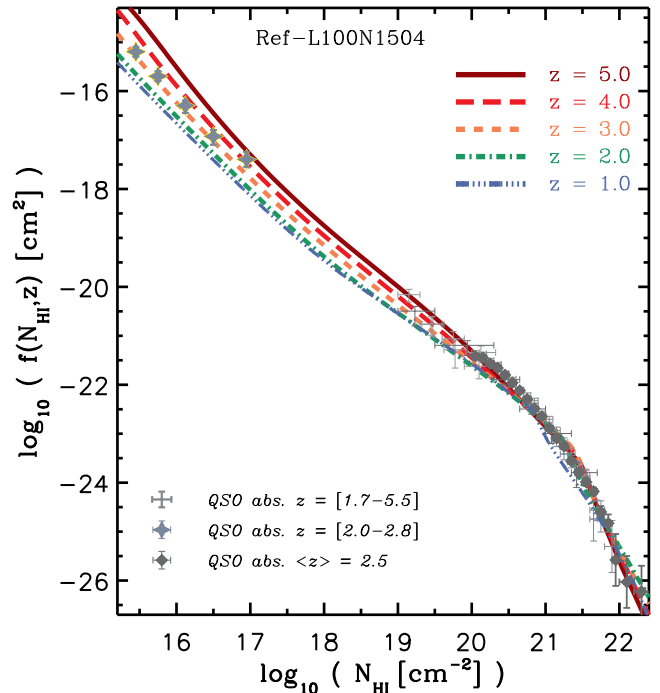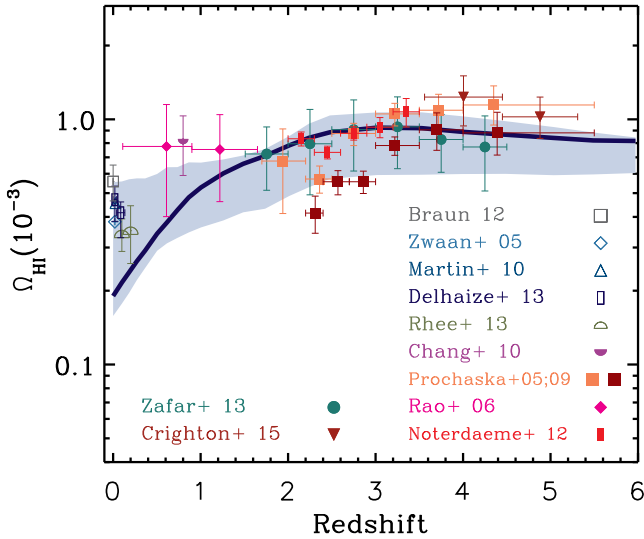

Figure 1. CDDF of neutral gas at different redshifts for the *Ref-L100N1504* EAGLE simulation. The data points represent a compilation of various quasar absorption line observations at high redshifts (i.e. $z = [1.7, 5.5]$) taken from Péroux et al. (2005) with $z = [1.8, 3.5]$, O'Meara et al. (2007) with $z = [1.7, 4.5]$, Noterdaeme et al. (2009) with $z = [2.2, 5.5]$ and Prochaska & Wolfe (2009) with $z = [2.2, 5.5]$. The grey diamonds at $N_{\rm H\,I} > 10^{20}\,{\rm cm}^{-2}$ represent the most recent constraints on the high end of the H I CDDF which are taken from Noterdaeme et al. (2012) with $\langle z \rangle = 2.5$. The grey star-shaped data points at $N_{\rm H\,I} < 10^{17}\,{\rm cm}^{-2}$ are taken from Rudie et al. (2013) with $z = [2.0, 2.8]$. The simulation results are in good agreement with the observations.

2009; Prochaska, Worseck & O'Meara 2009; Rudie et al. 2013; Zafar et al. 2013). In Fig. 1, we compare the H I CDDF predicted by the *Ref-L100N1504* simulation with a compilation of observational results.[3] The predicted H I CDDF is shown using different colours and line styles for different redshifts ranging from $z = 1$ to 5. For comparison, the grey data points show the observed H I CDDF for $N_{\rm H\,I} \lesssim 10^{17}\,{\rm cm}^{-2}$ at $z \sim 2$–3 from Rudie et al. (2013), and at higher $N_{\rm H\,I}$ a compilation containing various observations spanning the redshift range of $1.7 < z < 5.5$ (Péroux et al. 2005; O'Meara et al. 2007; Noterdaeme et al. 2009; Prochaska & Wolfe 2009). The most recent measurements of the H I CDDF at very high $N_{\rm H\,I}$ and an average redshift of $\langle z \rangle = 2.5$ from Noterdaeme et al. (2012) are also shown using dark grey diamonds.

Fig. 1 shows that there is good agreement between the predicted H I CDDFs and observations for strong H I absorbers ($N_{\rm H\,I} \gtrsim 10^{19}\,{\rm cm}^{-2}$), similar to what was found for OWLS (Altay et al. 2011; Rahmati et al. 2013a). The weak evolution of the high end of the H I CDDF that we reported for OWLS in Rahmati et al. (2013a) is also evident. However, we note that the measurements of Rudie et al. (2013) for the CDDF are slightly underproduced by our simulation which indicates the need to use a lower hydrogen photoionization rate at $z = 2.5$ compared to what our fiducial UVB model implies

---

[2] The typical velocity difference cut often used in observational studies is $> \pm 1000\,{\rm km\,s}^{-1}$ (e.g. Rudie et al. 2012; Prochaska et al. 2013a,b).

[3] We apply appropriate corrections for the different cosmological parameters adopted in different studies, converting all results to the Planck cosmology used in EAGLE.

Figure 2. Cosmic density of H I as a function of redshift in the *Ref-L100N1504* EAGLE simulation (solid curve). The shaded area around the curve indicates the range covered by all the simulations listed in Table 1 (expect models with no feedback). The data points represent a compilation of various quasar absorption line observations taken from Rao, Turnshek & Nestor (2006) with $z = [0.11–1.65]$, Prochaska, Herbert-Fort & Wolfe (2005), Prochaska & Wolfe (2009) with $z = [2.2, 5.5]$, Noterdaeme et al. (2012) with $z = [2.0, 3.5]$, Zafar et al. (2013) with $z = [1.5, 5.0]$ and Crighton et al. (2015). The low-redshift compilation of data is based on 21-cm emission studies of Zwaan et al. (2005), Martin et al. (2010), Braun (2012) and Delhaize et al. (2013) at $z \sim 0$, stacked 21-cm emission studies of Rhee et al. (2013) at $z \sim 0.1$–0.2 and 21-cm intensity mapping of Chang et al. (2010) at $z \approx 0.8$.

(see Appendix A). It should also be noted that because we do not correct the simulation for H$_2$, the agreement at $N_{\mathrm{H\,I}} \gtrsim 10^{22}\ \mathrm{cm}^{-2}$ may be fortuitous.

The cosmic H I density, $\Omega_{\mathrm{H\,I}}$, which is defined as the mean H I density divided by the critical density, ρ_{crit}, can be calculated either by estimating the observed H I mass density through H I 21-cm emission at low redshifts or by integrating the H I CDDF of absorbers at high redshifts:

$$\Omega_{\mathrm{H\,I}} = \frac{H_0 m_{\mathrm{H}}}{c \rho_{\mathrm{crit}}} \int_0^\infty N_{\mathrm{H\,I}} f(N_{\mathrm{H\,I}}, z) \mathrm{d} N_{\mathrm{H\,I}}, \tag{3}$$

where $H_0 = 100\ h\ \mathrm{km\ s}^{-1}\ \mathrm{Mpc}^{-1}$ is the Hubble constant, m_{H} is the mass of a hydrogen atom, c is the speed of light and $\rho_{\mathrm{crit}} = 1.89 \times 10^{-29} h^2 \mathrm{g\ cm}^{-3}$.

The predicted evolution of the cosmic H I density for the *Ref-L100N1504* EAGLE simulation is shown with the solid curve in Fig. 2. The shaded area around the curve shows the range covered by the other feedback enabled simulations we use in this work (see Table 1). For comparison, a compilation of various observational measurements are overplotted with different symbols. The high-z ($z \gtrsim 1$) measurements of the $\Omega_{\mathrm{H\,I}}$ are often based on the observed abundance of DLAs[4] (e.g. Prochaska et al. 2005; Rao et al. 2006; Prochaska & Wolfe 2009; Noterdaeme et al. 2012; Zafar et al. 2013). The low-redshift measurements are generally based on measuring the H I mass using 21-cm emission and often involve adopting non-

trivial assumptions about the H I gas fraction of the full galaxy population to derive the cosmic H I density (e.g. Zwaan et al. 2005; Chang et al. 2010; Martin et al. 2010; Delhaize et al. 2013; Rhee et al. 2013). The comparison between the predicted and observed $\Omega_{\mathrm{H\,I}}$ shows good agreement, particularly at the redshifts of interest here, $z > 1$. While the cosmic H I density remains nearly constant from $z \sim 6$ to $z \sim 2$–3, it declines towards lower redshifts. This decline, which is also evident in observed evolution of the $\Omega_{\mathrm{H\,I}}$, is not reproduced in some previous theoretical studies (e.g. Davé et al. 2013; Lagos et al. 2014). The cosmic H I density in the *Ref-L100N1504* simulation seems to drop faster than observed, while the range of predictions obtained by varying the resolution and/or box-size of the simulation (shaded region around the solid curve in Fig. 2) is fully consistent with the observational measurements at $z \approx 0$. Moreover, as we mentioned above, it is important to note that measurements of $\Omega_{\mathrm{H\,I}}$ at low redshift often involve strong assumptions about the H I mass fractions of all galaxies and are therefore not as robust as the direct measurements at higher redshifts.

Having shown that the observed cosmic distribution of H I is reproduced reasonably well, we can use the simulation with more confidence to study the H I distribution around galaxies.

3.2 Covering fraction of LLSs inside haloes

Examples of the distribution of H I around simulated galaxies at $z = 3$ are shown in Fig. 3. In this figure, the coloured maps show the H I column density distributions in $1 \times 1\ \mathrm{pMpc}^2$ regions centred on galaxies with $M_{200} = 10^{12}$–$10^{13}\ \mathrm{M}_\odot$. The virial radius, r_{200}, of each galaxy is shown with a blue circle and each galaxy map is shown using three different orthogonal projections. A significant fraction of the area within the virial radii of massive galaxies is covered with LLSs (i.e. $N_{\mathrm{H\,I}} > 10^{17.2}\ \mathrm{cm}^{-2}$) which have highly inhomogeneous distributions with often form filamentary structures. As a result, the fraction of the area covered by LLSs inside the virial radius, which is indicated in the top-right corner of each panel, varies depending on the point of view and from one galaxy to another. However, the average LLSs covering fraction does not depend strongly on the halo mass. On the other hand, as Fig. 4 shows, the typical LLSs covering fraction decreases significantly from $z = 4$ to 2.

To quantify the distribution of H I around galaxies, the covering fraction of LLSs within the virial radius, $f_{<r_{200}}$, is defined as the probability of finding systems with $N_{\mathrm{H\,I}} > 10^{17.2}\ \mathrm{cm}^{-2}$ with impact parameters smaller than the virial radius, r_{200} and with LOS distances from the galaxy shorter than a specific value comparable to the virial radius. Equivalently, $f_{<r_{200}}$ can be defined as the fraction of the surface area within $R < r_{200}$ that is covered by LLSs after projecting the gas distribution within a specific LOS distance from the galaxy on to a 2D plane. We calculate this quantity for each galaxy by projecting the H I within a slice with $4 \times r_{200}$ thickness centred on the galaxies redshift[5] and repeating the same calculation for projections along three different orthogonal directions. Then we calculate $f_{<r_{200}}$ by measuring the fraction of the surface contained within the r_{200} that is covered by LLSs.

The predicted $f_{<r_{200}}$ for different redshifts is shown in Fig. 5 as a function of halo mass, M_{200}, in the left-hand panel, and as a function of specific star formation rate, $\dot{M}_\star/M_\star$, in the right-hand panel. Each solid curve and the shaded area around it show, respectively, the

---

[4] Note that due to the shape of the H I CDDF, the cosmic H I density is very sensitive to the abundance of DLAs and that the contribution of lower $N_{\mathrm{H\,I}}$ systems is negligible.

[5] We used this definition to be consistent with previous studies (e.g. Fumagalli et al. 2014). We note, however, that choosing thinner or thicker slices does not change our results noticeably.
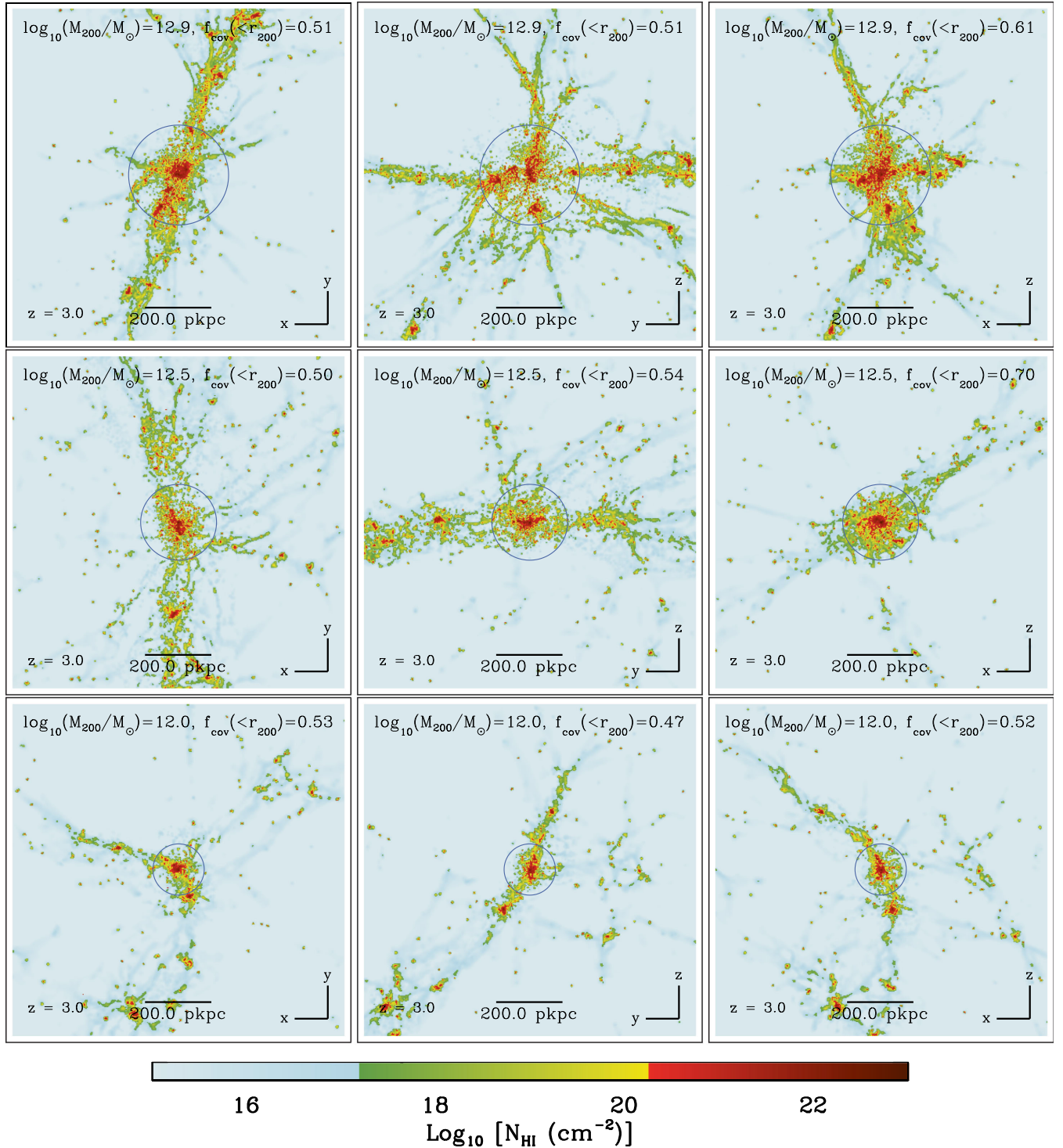
Figure 3. The simulated H I column density distribution around randomly selected massive galaxies at $z = 3$. Top, middle and bottom rows show galaxies with $M_{200} \approx 10^{12}$, $\approx 10^{12.5}$ and $\approx 10^{13}$ M$_{\odot}$, respectively. The columns in each row show a single galaxy as seen from three different orthogonal angles. Blue circles are centred on galaxies and show the virial radii (i.e. r_{200}). Each panel shows a 1×1 pMpc^2 region with the same projected depth. The covering fraction of LLSs (i.e. $N_{H I} > 10^{17.2}$ cm^{-2}) with impact parameters less than r_{200} is indicated in the top-right of each panel. At $z = 3$, LLSs (green regions) form filamentary structures and their distribution varies strongly from galaxy to galaxy, and with the viewing angle. The typical covering fraction of LLSs within r_{200} does not vary strongly with halo mass at a given redshift.

median covering fraction and the corresponding 15–85 percentiles for the *Ref-L100N1504* simulation. In this figure, red, orange, green and blue curves show $z = 4$, 3, 2 and 1, respectively. For massive haloes with $M_{200} \gtrsim 10^{12}$ M$_{\odot}$, $f_{<r_{200}}$ does not depend strongly on

halo mass. For less massive haloes (i.e. $M_{200} \lesssim 10^{12}$ M$_{\odot}$) on the other hand, the covering fraction of LLSs increases more strongly (though still relatively weakly) with halo mass and the slope of this relation increases with redshift.
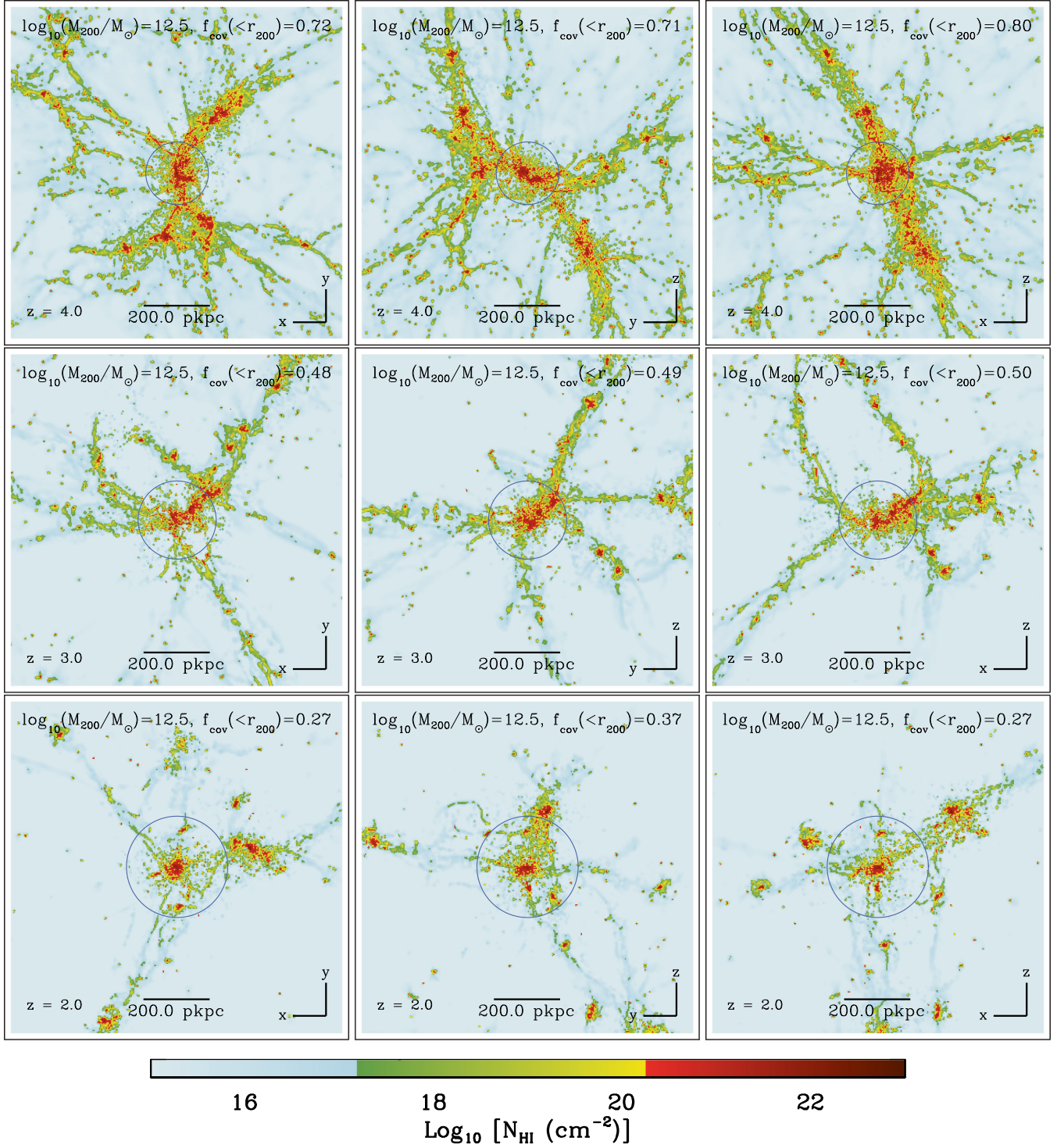
Figure 4. The same as Fig. 3 but for galaxies with $M_{200} \approx 10^{12.5} \, M_{\odot}$ at redshifts $z = 4, 3$ and 2 (from top to bottom, respectively). The columns in each row show a single galaxy as seen from three different orthogonal angles. LLSs (green regions) form filamentary structures at all redshifts, which become less prominent by decreasing redshift. The typical covering fraction of LLSs within r_{200} evolves rapidly with redshift.

While the dependence of $f_{<r_{200}}$ on halo mass is rather weak, the covering fractions increase strongly with redshift. This result is consistent with haloes containing increasingly higher gas fractions at higher redshifts as a result of increasing rates of cold accretion and the higher mean density of the Universe. As we will show in Section 3.3, this weak halo mass dependence enables us to charac-

terize the distribution of absorbers over a wide range of halo mass as a function of redshift and radius relative to the virial radius. The strong correlation between the covering fraction of LLSs and redshift has important consequences for the interpretation of observations because the observed sample often contain galaxies with a wide range of redshifts. For instance, the average probability of
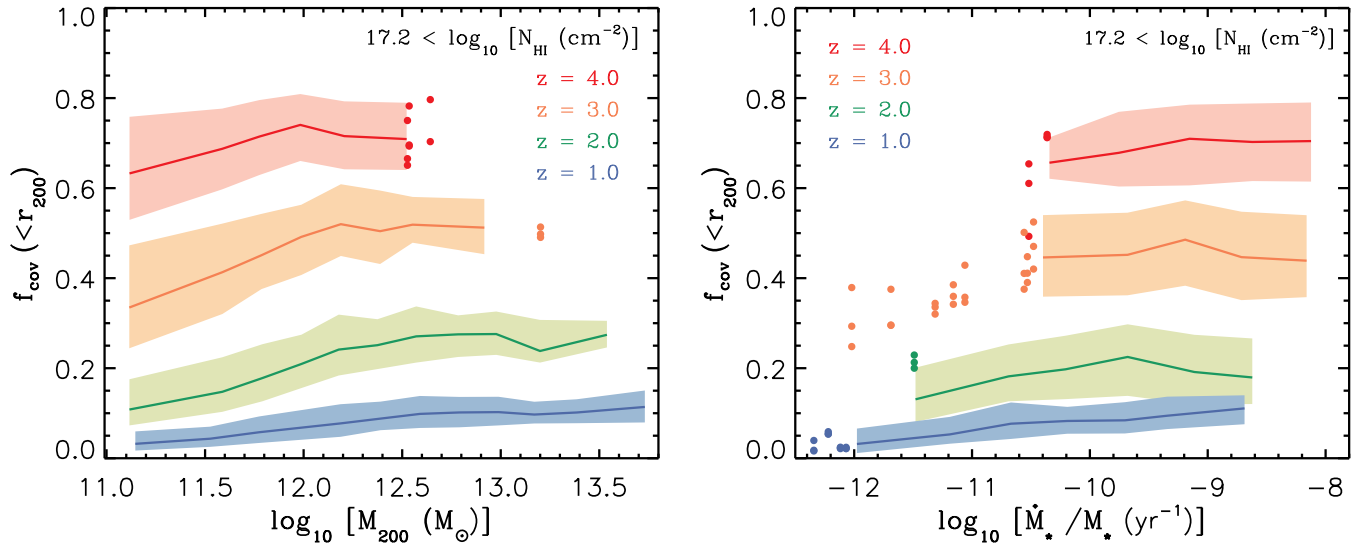
Figure 5. Cumulative covering fraction of LLSs within r_{200} as a function of halo mass, M_{200} (left) and specific star formation rate, $\dot{M}_{\star}/M_{\star}$ (right). Curves from top to bottom show redshifts $z = 4$ (red), 3 (orange), 2 (green) and 1 (blue). Solid curves show the median covering fractions for the *Ref-L100N1504* simulation and the shaded areas around them indicate the 15–85 percentiles. Galaxies are shown with individual data points instead of curves in bins that contain fewer than 10 galaxies. About 35 000 galaxies (each in three different orientations) were used to make this figure. The covering fraction increases strongly with redshift for all halo masses. It also increases with halo mass but this dependence is weak for massive objects. There is no strong correlation between the covering fraction of LLSs and the specific star formation rate.

finding LLSs within a given distance from galaxies does not represent the covering fraction of LLSs at the typical (e.g. mean) redshift of the galaxy sample, because higher redshift galaxies have larger contributions to the average covering fraction of the population. As we will discuss in Section 4, together with other biases like the wide range of halo masses represented by observational samples, this issue can explain the large covering fractions derived from some observational samples (e.g. Prochaska et al. 2013b).

The right-hand panel of Fig. 5 shows that $f_{<r_{200}}$ does not depend strongly on the specific star formation rate of galaxies.[6] Only galaxies with $M_{200} > 10^{11.5}\,\mathrm{M_{\odot}}$ are shown in the right-hand panel of Fig. 5, but our experiments show that a narrow range of halo masses only strengthens the independence of $f_{<r_{200}}$ from the specific star formation rate. This trend thus suggests that the covering fraction of LLSs does not depend strongly on the transient variations in the star formation activity of galaxies, and is set by their average star formation activity and the large-scale distribution of gas around them.

As the shaded areas around the median curves in Fig. 5 illustrate, there is significant scatter in the predicted covering fraction at any given mass and specific star formation rate. This scatter is larger at higher redshifts where the typical covering fractions are also larger. We note that the covering fraction of a single simulated galaxy can change from one projection axis to another by a factor close to the typical scatter for its mass range (see Fig. 3), which is consistent with what Fumagalli et al. (2014) found. This, together with the lack of strong dependence between the covering fraction and specific star formation rate, suggests that most of the scatter shown in Fig. 5 can be attributed to the highly inhomogeneous and filamentary distribution of H I around galaxies.

Although $f_{<r_{200}}$ is widely used to quantify the distribution of H I around galaxies, both in theoretical and observational studies, one should note that the virial radius is not a directly observable quantity. Moreover, as mentioned above, the virial radius of a sample of galaxies with a wide range of different characteristics (e.g. mass, redshift) is not well defined. For this reason, we opted not to compare the covering fractions shown in Fig. 5 with those reported in observational studies (e.g. Rudie et al. 2012; Prochaska et al. 2013a,b). Instead, we compare to observations after matching the redshift and halo mass distribution of observed and simulated samples in Section 4.
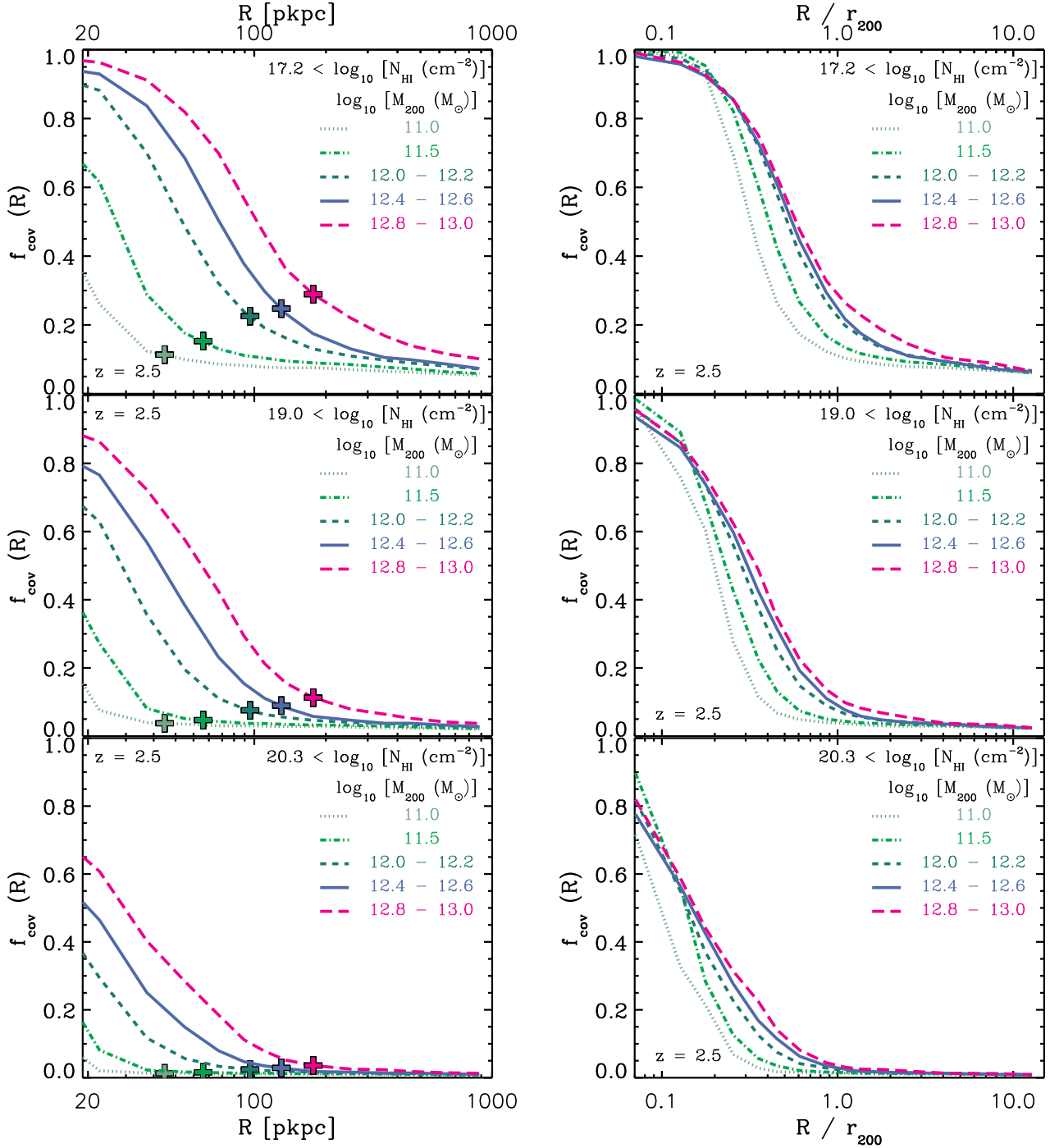
3.3 Covering fraction profiles

In the previous section, we studied the cumulative covering fraction of LLSs within the virial radius (i.e. $f_{<r_{200}}$). However, more information is embedded in the differential covering fraction profile of H I absorbers with different column densities as a function of impact parameter from galaxies. We define the differential covering fraction in a given impact parameter bin as

$$f_{\mathrm{cov}}(R) \equiv f_{\mathrm{cov}}(R_i < R < R_{i+1}) \equiv \frac{A_{\mathrm{abs}}\,|_{R_i}^{R_{i+1}}}{\pi\,(R_{i+1}^2 - R_i^2)}, \qquad (4)$$

where A_{abs} is the area covered by absorbers (e.g. LLSs) within a radial bin defined by the impact parameters R_i and R_{i+1}, assuming that $R_{i+1} > R_i > 0$. Note that throughout this work, we reserve $f_{\mathrm{cov}}(R)$ the differential covering fraction and $f_{\mathrm{cov}}(<R)$ for the cumulative covering fraction (e.g. Figs 5 and 10). For instance, we denote the cumulative covering fraction within an impact parameter equal to the virial radius by $f_{<r_{200}} = f_{\mathrm{cov}}(0 < R < r_{200})$ as shown in Fig. 5.

The top-left panel in Fig. 6 shows the predicted mean differential covering fraction of LLSs as a function of impact parameter for five different halo mass bins in the *Ref-L100N1504* simulation at $z = 2.5$. In analogy with observational studies (e.g. Prochaska et al. 2013a,b),

---

[6] Note that we neglect the impact of local sources on $f_{<r_{200}}$. If local sources were to change the $f_{<r_{200}}$ significantly, then they could introduce a dependence on the specific star formation rate.

Figure 6. Profiles showing the mean differential covering fraction of LLSs (top), sub DLAs (middle) and DLAs (bottom) as a function of impact parameter (left) and normalized impact parameter (right) for different halo mass bins in the *Ref-L100N1504* EAGLE simulation at $z = 2.5$ using $\Delta V = 3150 \, \text{km} \, \text{s}^{-1}$. In the left-hand panel, the virial radius corresponding to each halo mass bin is indicated with a cross. As the left-hand panels show, all covering fractions depend strongly on halo mass at fixed impact parameters, particularly very close to galaxies. The right-hand panels show, however, that only a weak mass dependence remains in the covering fraction profiles of haloes with $M_{200} \gtrsim 10^{12} \, \text{M}_\odot$, after normalizing the impact parameters to the virial radii. This suggests that the shape of the covering fraction profiles of LLSs, sub DLAs and DLAs is similar in all haloes with $M_{200} \gtrsim 10^{12} \, \text{M}_\odot$ and the mass dependence of the covering fractions at fixed physical impact parameters stems mainly from the differences in the halo sizes.

around each galaxy a velocity window of $\Delta V = 3150 \, \text{km} \, \text{s}^{-1}$ (i.e. the allowed velocity difference between absorbers and galaxies is $\leq \pm 1575 \, \text{km} \, \text{s}^{-1}$) is adopted for calculating its covering fraction, but we note that increasing or decreasing the allowed velocity width by a factor of a few does not change the results for $R < r_{200}$ (see Fig. C1).

The five mass bins shown in the top-left panel of Fig. 6 have similar LLSs covering fractions at the outermost impact parameters, but they vary strongly with halo mass close to galaxies. As the middle-left and bottom-left panels in Fig. 6 show, the same qualitative trend holds for sub DLAs (i.e. $N_{\text{H\,I}} > 10^{19} \, \text{cm}^{-2}$) and DLAs. However, by

increasing the H I column density of absorbers, the covering fraction at fixed impact parameters decreases. Despite the sensitivity of the covering fractions to the halo mass, it seems that the shapes of the curves are very similar for halo masses $M_{200} \gtrsim 10^{12}\,\mathrm{M}_\odot$ which suggests that they can be matched by a re-scaling to account for differences in the virial radii of the haloes.

To show that the covering fraction profiles are nearly scale invariant, we normalize the impact parameters to the virial radii of the haloes in the panels of the right-hand side of Fig. 6. The good agreement between the three curves that represent $M_{200} \gtrsim 10^{12}\,\mathrm{M}_\odot$ haloes shows that the covering fraction profiles of LLSs, sub DLAs and DLAs around those haloes are self-similar with a characteristic scalelength very close to the virial radius. This is the main reason behind the very weak dependence of the total LLS covering fraction within r_{200} ($f_{<r_{200}}$) and halo mass for galaxies with $M_{200} \gtrsim 10^{12}\,\mathrm{M}_\odot$ (see Fig. 5). As the curves for the two lowest mass bins in the right-hand panels of Fig. 6 show, the scale-invariance of the covering fraction profiles breaks down for galaxies with $M_{200} < 10^{12}\,\mathrm{M}_\odot$, where the total covering fraction of absorbers within r_{200} also slowly decreases with decreasing halo mass (see Fig. 5).

The scale-invariance of the distribution of strong H I absorbers around massive haloes allows us to calculate the typical normalized covering fraction profiles that characterize the distribution of LLSs, sub DLAs and DLAs around haloes with $M_{200} \gtrsim 10^{12}\,\mathrm{M}_\odot$ at any given redshift.[7] Based on the strong evolution of the total H I covering fraction inside haloes (see Fig. 5), it is expected that the normalized covering fraction profiles also evolve strongly with redshift. To illustrate this, we show the normalized differential covering fraction for LLSs, sub DLAs and DLAs in, respectively, the top, middle and bottom panels of Fig. 7. The different curves in each panel indicate different redshifts where long-dashed (red), dashed (orange), dot–dashed (green) and dotted curves show $z = 4$, 3, 2 and 1, respectively. Note that a LOS velocity window of width $\Delta V = 3000\,\mathrm{km\,s^{-1}}$ is adopted for calculating the covering fractions shown in this figure. To illustrate the typical intrinsic scatter in the covering fraction profiles, the 15–85 percentiles of the covering fractions at $z = 4$ are indicated by the shaded areas around the long-dashed red curves. The normalized covering fraction profiles of all strong H I absorbers indeed evolve strongly.

Defining $x \equiv R/r_{200}$ as the normalized impact parameter, the covering fraction of H I absorbers around galaxies with a given virial radius, r_{200}, at redshift z can be fitted with the following function:

$$f_{\mathrm{cov}}(x, z) = 1 - \frac{1}{1 + \left(\frac{L_z}{x}\right)^\alpha} + C\left[\frac{1}{1 + \left(\frac{L_z}{x}\right)^3}\right] 10^{\frac{z-4}{3}}, \qquad (5)$$

where L_z is a redshift-dependent characteristic length scale that is given by

$$L_z = A\,B^z, \qquad (6)$$

and A, B, C and α are free parameters that vary for LLSs, sub DLAs and DLAs. Based on this fitting function, the covering fraction approaches unity very close to galaxies where $x \ll L_z$ and at very high redshifts where $z \to \infty$. However, note that the latter is the case only if $B < 10^{1/6}$. Far from galaxies where $x \gg L_z$, on the other hand, the covering fraction approaches the asymptotic value of $C\,10^{\frac{z-4}{3}}$.

---

[7] The *Ref-L100N1504* EAGLE simulation contains 95, 345, 824 and 1436 haloes with $M_{200} \gtrsim 10^{12}\,\mathrm{M}_\odot$ at redshifts $z = 4$, 3, 2 and 1, respectively.
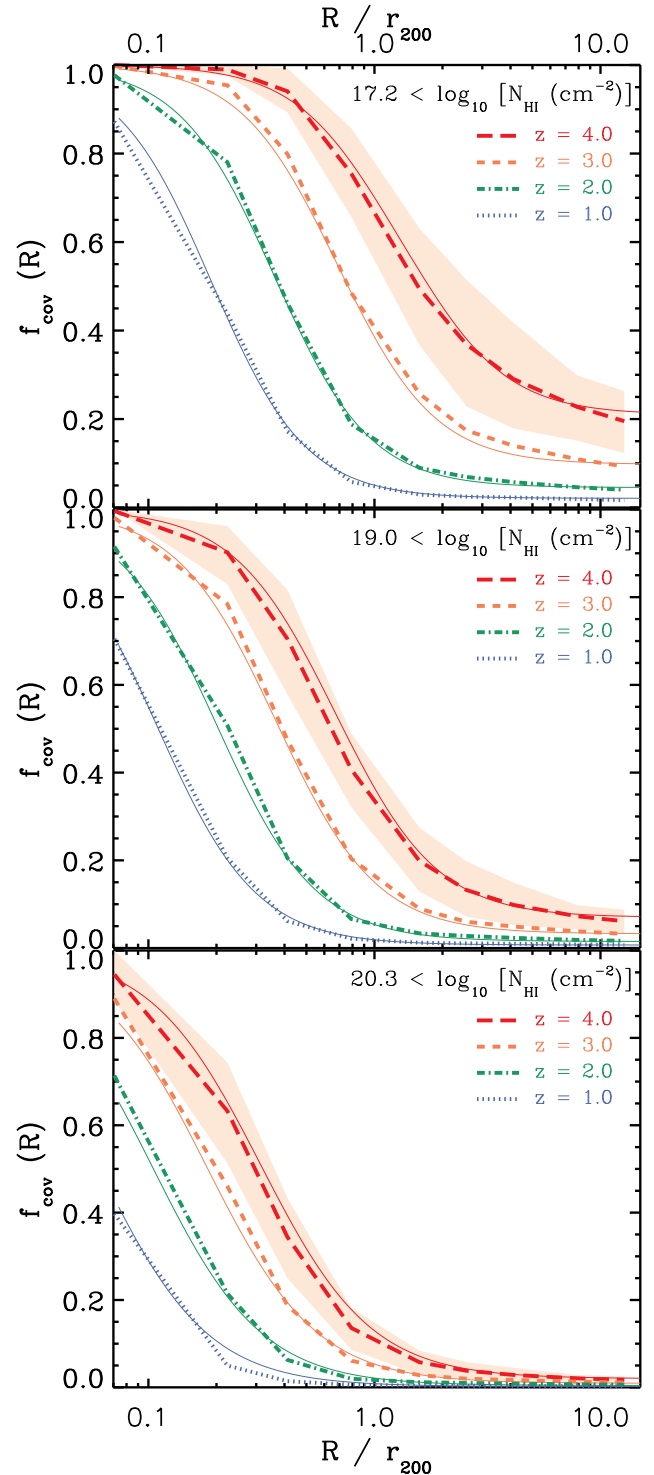
Figure 7. Normalized differential covering fraction profiles of H I absorbers around galaxies with $M_{200} \geq 10^{12}\,\mathrm{M}_\odot$ in the *Ref-L100N1504* EAGLE simulation at different redshifts using $\Delta V = 3000\,\mathrm{km\,s^{-1}}$. The top, middle and bottom panels show the covering fractions of LLSs, sub DLAs and DLAs, respectively. The red (long-dashed), orange (dashed), green (dot–dashed) and blue (dotted) curves show the results at $z = 4$, 3, 2 and 1, respectively, while the thin solid curves show the fitting function given by equations (5) and (6). The covering fractions for stronger H I absorbers are lower and their profiles have shorter characteristic scale lengths. The covering fractions of all absorbers drop rapidly with decreasing redshift at all impact parameters.

Table 2. The best-fitting values for the free parameters of the fitting function, equations (5) and (6), for the predicted normalized covering fraction of LLSs, sub DLAs and DLAs for haloes with $M_{200} \geq 10^{12}\,\mathrm{M}_\odot$ at redshifts $z \lesssim 4$. Note that parameter C is sensitive to the chosen velocity width and is reported for two velocity width $\Delta V = 3000$ and $1500\,\mathrm{km\,s^{-1}}$. The performance of the fitting function (thin solid curves) is shown in Fig. 7 for $\Delta V = 3000\,\mathrm{km\,s^{-1}}$.

Absorbers	A	B	C $\Delta V = 3000\,\mathrm{km\,s^{-1}}$	C $\Delta V = 1500\,\mathrm{km\,s^{-1}}$	α
LLS ($N_{\mathrm{H\,I}} > 10^{17.2}\,\mathrm{cm^2}$)	0.100	1.90	0.21	0.15	2.1
sDLA ($N_{\mathrm{H\,I}} > 10^{19.0}\,\mathrm{cm^2}$)	0.060	1.83	0.07	0.05	2.0
DLA ($N_{\mathrm{H\,I}} > 10^{20.3}\,\mathrm{cm^2}$)	0.035	1.73	0.02	0.015	1.8

Table 2 lists the best-fitting values for the free parameters. As shown in Appendix C, the covering fraction of absorbers at relatively large impact parameters ($R > r_{200}$) is sensitive to the size of the LOS velocity window that is used for associating H I absorbers and galaxies. As a result, parameter C in the fitting function of equation (5) is sensitive to the adopted velocity width. Therefore, we report two sets of best-fitting values of C where each set corresponds to either $\Delta V = 3000\,\mathrm{km\,s^{-1}}$ or $\Delta V = 1500\,\mathrm{km\,s^{-1}}$.

As the thin solid curves in Fig. 7 show, the predicted normalized covering fraction profiles are all closely matched by the values obtained from equation (5) for appropriate choices of the free parameters that are listed in Table 2. We note that the differences between the fitting formula and the simulation are much smaller than the typical intrinsic scatter in the covering fractions, which are shown in Fig. 7 for $z = 4$ by the shaded areas around the long-dashed red curves. Note that variations in the assumed UVB radiation can change the covering fraction of LLSs. For instance, reducing the UVB photoionization rate by a factor of 3 results in LLS covering fractions that are higher by ~ 0.1. However, such moderate changes in the UVB do not significantly affect the distribution of highly self-shielded stronger absorbers such as sub DLAs and DLAs.[8]

The values of the free parameters that control the empirical fitting function that we introduced above are physically meaningful. For instance, L_z could be regarded as a typical projected distance between absorbers and galaxies. Taking the values of A and B from Table 2, the implied typical projected distances between LLSs, sub DLAs and DLAs and their host galaxies at $z \approx 3$ are, respectively, $\approx r_{200}$, $\approx 0.5 r_{200}$ and $\approx 0.2 r_{200}$, which is in excellent agreement with the predictions of Rahmati & Schaye (2014) from the OWLS simulations (see the right-hand panel of their fig. 3). Moreover, equation (6) implies that the typical normalized impact parameter of strong H I absorbers increases exponentially with redshift. The best-fitting values for B, which controls the rate of this change, imply that the impact parameters of LLSs evolve slightly faster than those of DLAs. Moreover, the higher α value for LLSs indicates that their covering fractions drop faster than those of DLAs with increasing normalized impact parameters.

The rightmost term in equation (5) is related to the contribution of the background absorbers outside of the halo. Given the steep H I CDDF (see Fig. 1), there are many more LLSs than DLAs. Given the fixed LOS velocity cut imposed to obtain the covering fraction of absorbers, one would expect the ratio between the covering fraction of absorbers at large virial radii to follow the ratio between their cosmic abundances. As the best-fitting values for the C parameter

imply, this is indeed the case (e.g. LLSs are ~ 10 times more frequent than DLAs at all redshifts).

We note that local sources of ionizing radiation, which we have ignored, may cause the fitting function to underpredict the covering fraction of strong H I absorbers close to galaxies (see Appendix B). We emphasize that we only considered the redshift range $1 \leq z \leq 4$ when deriving our fitting function. While the same function produces a reasonable match to the simulated normalized H I covering fractions at $z = 5$ for $R \lesssim r_{200}$, it predicts LLS covering fractions that are ≈ 10 per cent too high for larger impact parameters. This difference is smaller than (but comparable to) the intrinsic scatter (i.e. 15–85 percentiles) in the simulated normalized profiles.

4 COMPARISON WITH OBSERVATIONS

Recent observations found large covering fractions of LLSs close to massive haloes at $z \sim 2$ (Rudie et al. 2012; Prochaska et al. 2013a,b) which implies the existence of a large reservoir of neutral and hence relatively cold gas around massive $z \sim 2$ galaxies. However, recent simulations of $\sim 10^{12.5}\,\mathrm{M}_\odot$ haloes at $z \sim 2$ by Fumagalli et al. (2014) and Faucher-Giguère et al. (2015) resulted in LLSs covering fractions that are much smaller than observed by Prochaska et al. (2013b). This discrepancy may indicate that our current theoretical understanding of galaxy formation and evolution is inadequate.

The EAGLE simulation is ideal to revisit this problem since its volume is sufficiently large to contain a large number of haloes with $M_{200} \gtrsim 10^{12}\,\mathrm{M}_\odot$ at $z \sim 2$–3 without compromising the resolution needed to reproduce the observed cosmic distribution of H I (see Section 3.1). Because of the large intrinsic scatter in the covering fractions, a large sample of simulated galaxies is required to constrain the average covering fraction at any given mass. Moreover, thanks to efficient stellar and AGN feedback, EAGLE reproduces the basic observed characteristics of galaxies over wide ranges of mass and redshift (S15; Furlong et al. 2015). In addition, using a large cosmological simulation enables us to calculate the H I covering fractions for a LOS path length comparable to what is typically used in observational studies.

In the following, we compare EAGLEs predictions with the observational constraints on the distribution of H I around massive star-forming galaxies (Rudie et al. 2012) and quasars (Prochaska et al. 2013b) at $z \sim 2$.

4.1 H I distribution around quasars at $z \sim 2$

Prochaska et al. (2013a,b) observed the distribution of H I around bright quasars at $z \sim 2$. Clustering analysis indicate that those quasars which typically reside in massive haloes with $M_{200} \sim 10^{12.5}\,\mathrm{M}_\odot$ (White et al. 2012). Using background quasars to probe the distribution of gas in absorption around a large

---

[8] It is important to keep in mind that very close to galaxies ($R \ll r_{200}$) the predicted covering fractions may be overestimates because we neglect local sources of ionizing radiation.
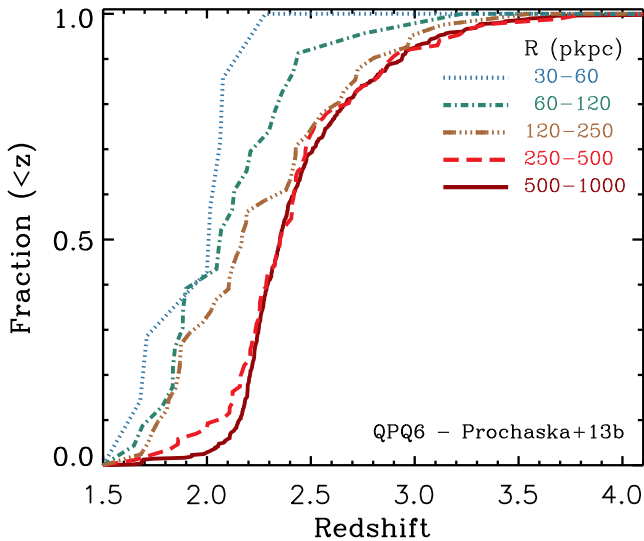
Figure 8. Cumulative redshift distribution of the foreground quasars from the QPQ6 sample (Prochaska et al. 2013b). Different curves show the redshift distribution of foreground quasars in different impact parameter bins, identical to those used in Fig. 9. Most quasars are at large impact parameters and have a wide range of redshifts, extending up to $z \gtrsim 4$. More than 30 per cent of quasars with $R \gtrsim 100$ pkpc have redshifts $z \gtrsim 2.5$. Given the rapid increase of the H I covering fraction with redshift that we found in this work, those quasars can strongly bias the estimated covering fraction of LLSs, particularly at large impact parameters. Moreover, for a magnitude-limited survey, higher-redshift quasars are likely to reside in more massive haloes which may further bias the estimated covering fractions at fixed impact parameters.

sample of foreground quasars (hereafter the QPQ6 sample), Prochaska et al. (2013b) measured the covering fraction of LLSs in different impact parameter bins. Adopting a fixed typical halo mass of $10^{12.5}$ M$_\odot$ at $z \sim 2$, and consequently a fixed virial radius of 160 kpc for all quasars in the QPQ6 sample, Prochaska et al. (2013a) concluded that more than $\gtrsim 60$ per cent of the area within the virial radii of haloes with masses around $10^{12.5}$ M$_\odot$ at $z \sim 2$ is covered by LLSs. For their covering fraction measurement, Prochaska et al. (2013a) adopted a velocity width of $\Delta V = 3000$ km s^{-1} to associate absorbers to quasars.

For a proper comparison between our simulation and the observations of Prochaska et al. (2013a,b), it is important to note that the foreground quasars in the QPQ6 sample have a relatively wide range of bolometric luminosities. Since quasars are variable, a one-to-one relation between quasar luminosity and halo mass is not expected. However, as a rough estimate, if we assume a positive correlation between the halo mass of galaxies and the bolometric luminosity of the quasars, we could conclude that quasars in the QPQ6 sample represent a relatively wide range of halo masses some of which could well exceed the typical halo mass of $M_{200} \sim 10^{12.5}$ M$_\odot$. If the host haloes of the observed quasars indeed have a range of masses, then the adopted fixed virial radius of 160 pkpc for calculating the total covering fraction of LLSs inside the virial radii of haloes with $M_{200} \approx 10^{12.5}$ M$_\odot$ at $z \approx 2$ could result in an overestimate of the true covering fraction due to the non-negligible contribution of haloes with $M_{200} > 10^{12.5}$ M$_\odot$ in the QPQ6 sample.

It is also important to note that the foreground quasars in the QPQ6 sample are not all at the same redshift. In addition, as Fig. 8 shows, there is a systematic trend between the typical redshift of foreground quasars and the impact parameter at which their gas content is measured. As a result, the typical redshift of quasars

for the smallest impact parameters (~ 10–100 pkpc) is $z \approx 2$ but it increases to $z \approx 2.5$ for impact parameters ~ 1 pMpc. The high-z tail of the distribution for quasars with impact parameters $R \gtrsim 100$ pkpc is quite extended and more than 30 per cent of them have $z > 2.5$. Given our finding that the H I distribution around galaxies at $z \sim 2$–3 evolves rapidly, this systematic bias should be taken into account when comparing simulations with observations.

In addition, the H I distribution is sensitive to the intensity of the UVB radiation. However, observational constraints on the intensity of the UVB at $2 \lesssim z \lesssim 6$ are model dependent and uncertain by a factor of a few (e.g. Bajtlik, Duncan & Ostriker 1988; Rauch et al. 1997; Bolton et al. 2005; Faucher-Giguère et al. 2008; Calverley et al. 2011; Becker & Bolton 2013). UVB models are also uncertain due to the various assumptions they need to adopt (e.g. the escape fraction of ionizing photons into the IGM, mean-free-path of ionizing photons, abundance of faint sources) and differ from each other by a factor of a few (e.g. Haardt & Madau 2001; Faucher-Giguère et al. 2009; Haardt & Madau 2012). While the intensity of our fiducial UVB model (i.e. Haardt & Madau 2001) is well within the range of the most recent estimates (e.g. Becker & Bolton 2013), intensities lower by up to a factor of ~ 3 are consistent with some observations/models at $z \sim 2$–3 (e.g. Faucher-Giguère et al. 2008, 2009; Haardt & Madau 2012), and would further improve the agreement between EAGLE and the observed H I column density distribution below $N_{\rm H\,I} \approx 10^{17}$ cm^{-2} (see Fig. 1).

It is necessary to take the aforementioned considerations into account when comparing simulations and observations. To do this, we calculate the covering fraction of LLSs around simulated galaxies with $M_{200} > 10^{12.5}$ M$_\odot$ in the *Ref-L100N1504* simulation at $z = 2.2$ and 3, resulting in 116 and 39 haloes,[9] respectively, with a median mass of $M_{200} = 10^{12.6}$ M$_\odot$.[10] The two selected redshifts bracket the range of redshifts that is represented by the QPQ6 sample. As shown in Appendix A, varying the UVB model changes the resulting H I covering fraction. It is therefore important to include also the uncertainties in the amplitude of the UVB photoionization rate. To do this, we calculate the H I distributions using both our fiducial UVB model of Haardt & Madau (2001) and the Haardt & Madau (2012) UVB model. Noting that at $z \approx 3$ the latter yields a hydrogen photoionization rate ≈ 3 times weaker than for our fiducial UVB model, we consider the $z = 3$ H I covering fractions calculated using the Haardt & Madau (2012) model upper limits on the predictions. Given the steep evolution in the H I covering fractions from $z = 3$ to 2, we use the simulation results at $z = 2$ that use our fiducial Haardt & Madau (2001) UVB model as lower limits for the predictions. Then, we calculate the covering fraction of LLSs, sub DLAs and DLAs for each halo using impact parameter bins identical to the analysis of Prochaska et al. (2013b). We use a LOS velocity window of $\Delta V = 3000$ and 3400 km s^{-1} around each galaxy at $z = 2.2$

---

[9] Although the number of simulated haloes we use is less than the observed number (155 simulated haloes versus 646 observed quasars), we use $\sim 10^5$ sight-lines per simulated object to calculate the covering fraction profiles. In other words, we use $\sim 10^7$ sight-lines to calculate the predicted H I distributions that are shown in Fig. 9 while only ≈ 600 observed sight-lines are used in Prochaska et al. (2013b).

[10] We note that due to steepness of the mass function around the halo mass ranges of interest for our analysis, most selected haloes have masses close to the lower mass limit we imposed in selecting them. Using a similar argument, a small fraction of the observed quasars is expected to be in haloes with masses far from the mean halo mass implied from the clustering measurements.
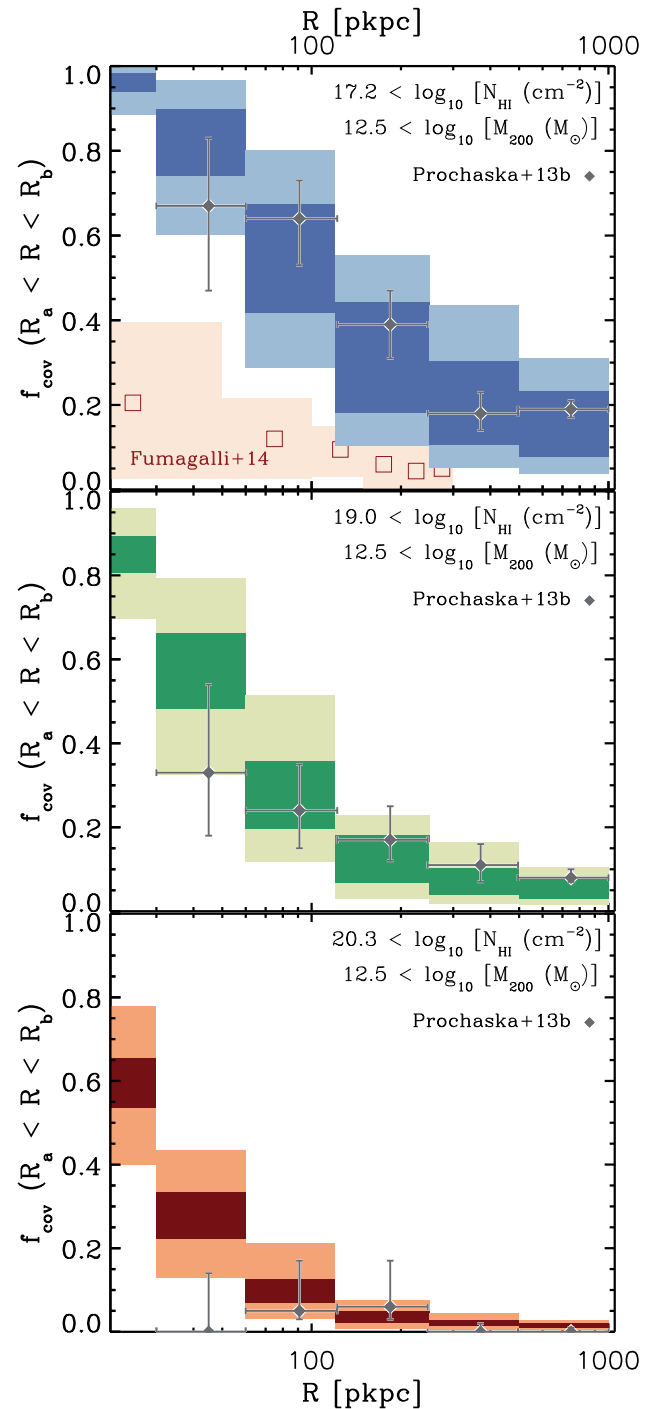
and $z = 3$ for calculating the covering fractions to mimic closely what is done observationally.[11]

The predicted covering fractions of H I absorbers are shown in Fig. 9 for LLSs (top panel), sub DLAs (middle panel) and DLAs (bottom panel). The upper and lower edges of the dark-coloured areas in each panel show the lower and upper limits of our predicted mean covering fractions obtained by applying the fiducial UVB model to $z = 2.2$ haloes and a three times weaker UVB model to $z = 3$ haloes, respectively. The shaded areas around the dark regions, which are shown using light colours, indicate the regions enclosed between the 15 percentiles of the lower limit for covering fraction (i.e. at $z = 2.2$ and using the fiducial UVB model) and the 85 percentile of the upper limit for the covering fraction (i.e. at $z = 3$ and using the weaker UVB model) in each impact parameter bin. In other words, the dark regions show how much variation is expected in the predicted covering fractions due to systematic effects caused by the redshift distribution of the quasar sample and the photoionization rate of the UVB, and the light-coloured areas around the dark regions show the predicted 1σ scatter (15–85 percentiles) around the mean due to object-to-object variations in the covering fraction within the sample of simulated haloes.

Grey diamonds with error bars show the observations of Prochaska et al. (2013b) where the horizontal error bars show the impact parameter bins and the vertical error bars show only the 1σ statistical uncertainty. Comparing the observed data points with the predicted results shows overall good agreement for absorbers with different H I column densities. The agreement is particularly good for larger impact parameters (>60 pkpc) despite the fact that the observational error bars are smallest there owing to the larger number of quasar pairs.
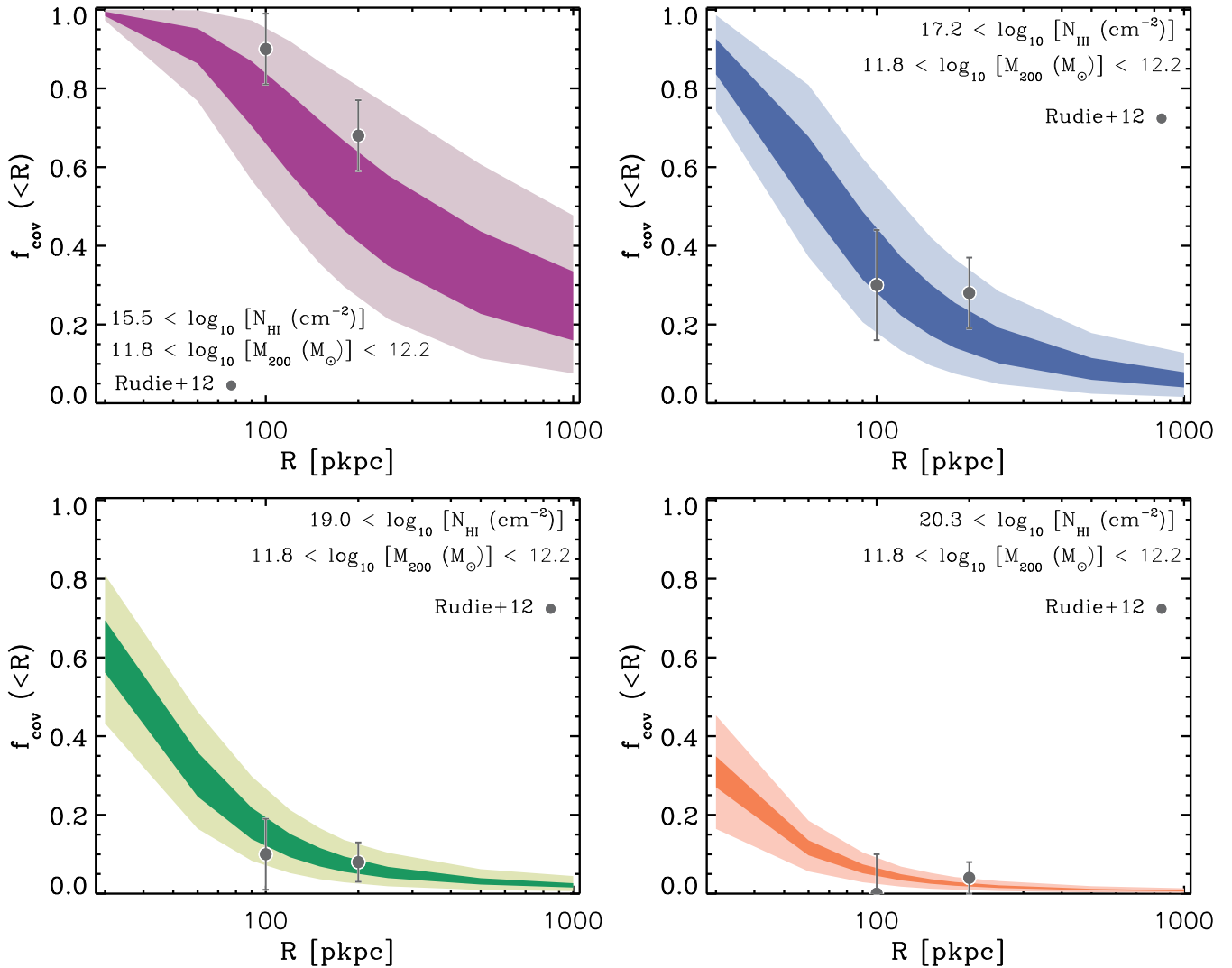
For impact parameters in the range 30–60 pkpc we appear to predict too high covering fractions and for DLAs this discrepancy is marginally significant. However, given that only six of the nearly 600 observed quasar pairs fall in this bin, one may question the robustness of the error estimates. There could also be biases. For example quasars covered by DLA absorption may be missing from the bright sample because of obscuration. The theoretical prediction is also most uncertain at the smallest impact parameters. For example, radiation from local stars thought to be the dominant source of hydrogen photoionization close to galaxies (e.g. Schaye 2006; Rahmati et al. 2013b) and would reduce the abundance of H I (see Appendix B). The presence of bright quasars will strengthen this effect. Quantifying the impact of local radiation on our results would require detailed radiative transfer simulations that also account for the duty cycle of quasars.

The top panel of Fig. 9 shows the simulation predictions from Fumagalli et al. (2014) using open squares and light-red shaded areas which, respectively, show the mean and 1σ scatter for covering fraction of LLSs around five simulated galaxies with halo masses $M_{200} \approx 10^{12.2}\,\mathrm{M_\odot}$ at $z = 2$. Their LLS covering fractions are significantly lower than both our predictions and observations. There are several potential explanations for this difference. Fumagalli et al. (2014) analysed the H I distribution at lower redshift and with lower masses than the objects in the QPQ6 sample. In addition,

[11] Since we use slices with fixed comoving lengths to mimic the velocity windows around galaxies, the width of the resulting velocity window becomes redshift dependent, but remains close enough to the value used in the observational analysis.

Figure 9. Predicted and observed differential H I covering fractions around quasars at $z \approx 2$. The data points with error bars show the observations of Prochaska et al. (2013b) for a sample of quasar at $\langle z \rangle = 2.3$. Predicted mean covering fractions for haloes with $M_{200} \geq 10^{12.5}\,\mathrm{M_\odot}$ in the *Ref-L100N1504* EAGLE simulation are shown with dark-coloured regions which indicate the systematic uncertainty in the mean due to uncertainties in the hydrogen photoionization rate of the UVB and the redshift range of the observed quasars (see the text). The light shaded areas indicate the 15–85 percentiles for the scatter due to object-to-object variation. The top, middle and bottom panels show the results for LLSs, sub DLAs and DLAs, respectively. The squares in the top panel show predictions from Fumagalli et al. (2014). The observed covering fractions agree with the EAGLE predictions.

Figure 10. Cumulative covering fraction of H I systems with different column densities inside a given impact parameter, *R*, as a function of impact parameter for LBGs. From top-left to the bottom-right, panels show the covering fraction of H I systems with $N_{\rm H\,I} > 10^{15.5}$ cm^{-2}, $N_{\rm H\,I} > 10^{17.2}$ cm^{-2}, $N_{\rm H\,I} > 10^{19.0}$ cm^{-2} and $N_{\rm H\,I} > 10^{20.3}$ cm^{-2}, respectively, and with impact parameters <*R*. The data points with error bars show the measurements from Rudie et al. (2012) for a sample of LBGs with $M_{200} \approx 10^{12}$ M$_\odot$ at ⟨*z*⟩ = 2.4 and 2.3 for the inner and outer impact parameter bins, respectively. Predicted mean covering fractions for haloes with $10^{11.8} < M_{200} < 10^{12.2}$ M$_\odot$ in the *Ref-L100N1504* EAGLE simulation are shown with dark-coloured regions which indicate the systematic uncertainty in the mean due to uncertainties in the background ionizing radiation and the redshift range of the observed LBGs (see the text). The light shaded areas indicate the 15–85 percentiles for the scatter due to object-to-object variation. The predictions agree well with the observations.

the simulations analysed by Fumagalli et al. (2014) did not include the efficient feedback that, as we show in Section 5, is required to obtain reasonable stellar masses and H I covering fractions for the haloes they considered. Furthermore, because they used zoom simulations, Fumagalli et al. (2014) only considered the distribution of absorbers with small LOS separations from the galaxies ($R \sim r_{200}$) when calculating covering fractions. In contrast, observations used a LOS velocity window of $\Delta V = 3000$ km s^{-1} which translates into distances much larger than the virial radii of the relevant haloes. While this difference does not affect the covering fractions at impact parameters $R \ll r_{200}$, it results in large differences at $R \gtrsim r_{200}$ (e.g. up to $\gtrsim 100$ per cent difference in the covering fraction of LLSs at $R \sim 200$–1000 pkpc for $M_{200} \approx 10^{12.5}$ M$_\odot$ haloes at $z = 2.5$, as shown in Fig. C1).

4.2 H I distribution around LBGs at $z \sim 2$

In addition to quasars, there are observational constraints on the H I distribution around star-forming galaxies at $z \sim 2$ (e.g. Rakic et al. 2012; Rudie et al. 2012; Turner et al. 2014). Here we compare with the data from Rudie et al. (2012), who used spectra of background quasars behind a sample of LBGs at $z \approx 2$–2.5 to measure the covering fraction of H I close to LBGs, which have halo masses $M_{200} \sim 10^{12}$ M$_\odot$ (Adelberger et al. 2005; Trainor & Steidel 2012; Rakic et al. 2013). By considering all absorbers that are within a LOS velocity window of $\Delta V = 1400$ km s^{-1} around each galaxy, Rudie et al. (2012) calculated the average covering fraction of H I systems with four different absorption strengths, and within impact parameters 100 pkpc and 200 pkpc for samples of 10 and 25 galaxies, respectively. Note that the typical virial radius of those galaxies

is ≈ 90 pkpc, the chosen impact parameters are close to one and two virial radii.

To compare the EAGLE predictions with the observations of Rudie et al. (2012), we select all simulated galaxies with $10^{11.8} < M_{200} < 10^{12.2}$ M$_\odot$ at $z = 2.2$ and $z = 2.5$ in the *Ref-L100N1504* simulation, which results in nearly a thousand haloes at each redshift. This redshift range closely matches the redshift distribution of the galaxies used by Rudie et al. (2012). We then calculate the H I covering fractions by adopting velocity windows of $\Delta V = 1350$ and 1312 km s^{-1} around galaxies at $z = 2.5$ and 2.2, respectively. To account for uncertainties in the strength of the UVB radiation, we adopt the same approach as in the previous section and recalculate the H I distributions after reducing the H I photoionization rate of our fiducial model (Haardt & Madau 2001) by a factor of 3 (i.e. to $\Gamma_{\rm UVB} = 7 \times 10^{-13}$ s^{-1} at $z = 2.5$). To bracket the redshift range of observed galaxies and the range of possible UVB photoionization rates, we take the covering fraction results based on our fiducial UVB at $z = 2.2$ as the lower limit and the lower UVB model at $z = 2.5$ as the upper limit. Fig. 10 shows the predicted cumulative covering fraction profiles where dark coloured regions show the area between the mean covering fractions of the two bracketing cases, and the light-coloured areas around them show the range between the 15 percentiles of our lower limit and the 85 percentiles of the upper limit, indicating the object-to-object variations in the covering fraction within the sample of simulated haloes. Counter clockwise from the top-right, panels show, respectively, the cumulative covering fraction profiles of systems with $N_{\rm H I} > 10^{15.5}$ cm^{-2}, LLSs ($N_{\rm H I} > 10^{17.2}$ cm^{-2}, sub DLAs ($N_{\rm H I} > 10^{19}$ cm^{-2}) and DLAs ($N_{\rm H I} > 10^{20.3}$ cm^{-2}). Note that the quantity shown on the vertical axis, $f_{\rm cov}(<R)$, is different from the previous plots and indicates the total covering fraction of systems with impact parameters smaller than R. In each panel, the measurements of Rudie et al. (2012) at impact parameters $R = 100$ pkpc and $R = 200$ pkpc are shown as filled circles and the error bars show the statistical 1σ errors. Note that these errors cannot be compared directly with the intrinsic 1σ scatter due to object-to-object variation (15–85 percentiles) in the predicted values (shown by the light-coloured areas).

As Fig. 10 shows, the predicted covering fractions agree very well with the observed values from Rudie et al. (2012).

5 IMPACT OF FEEDBACK

There are two ways in which feedback can affect the comparison between the simulated and observed distribution of H I around galaxies. First, feedback can change the distribution of gas around individual galaxies. Secondly, feedback can change the stellar mass of galaxies that reside in haloes of a fixed virial mass. This will affect comparisons with observations of the gas around galaxies of a fixed stellar mass, since the H I covering fractions are sensitive to halo mass. In this section, we investigate the effect of feedback by comparing simulations that use different feedback implementations, a box size $L = 25$ cMpc, and our default resolution (i.e. $N = 2 \times 376^3$ particles for this box size; see Table 1).

Fig. 11 shows the impact of feedback on the average differential covering factor of LLSs around simulated haloes with $10^{11.8} < M_{200} < 10^{12.2}$ M$_\odot$ at $z = 2.2$ (i.e. galaxies similar to LBGs; Adelberger et al. 2005; Trainor & Steidel 2012; Rakic et al. 2013). Shown are the reference model (*Ref-L025N0376*, blue solid), a model for which both stellar and AGN feedback are turned off (*NOFB*, black dotted), a simulation without AGN feedback (*NoAGN*, purple dot–dashed), and models with, respectively, half and twice the amount of stellar feedback as *Ref* (*WeakFB*, green
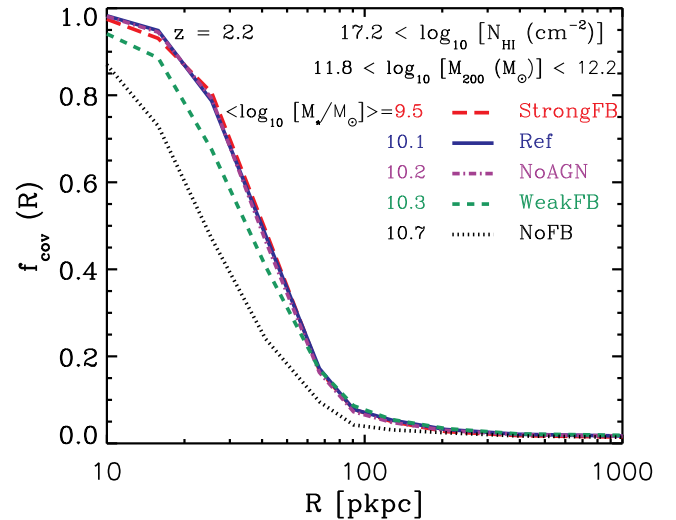

Figure 11. Differential covering fraction of LLSs around (and within $\Delta V = 1294$ km s^{-1} from) haloes with mass $10^{11.8} < M_{200} < 10^{12.2}$ M$_\odot$ at $z = 2.2$ as a function of impact parameter for simulations with different amounts of feedback. The solid blue curve shows the $L = 25$ cMpc reference model, *Ref-L025N0376*. Also shown are simulations using the same box size, resolution, and initial conditions but without any feedback (*NoFB*; black dotted), without AGN feedback (*NoAGN*; purple dot–dashed), with half as strong stellar feedback (*WeakFB*, green short-dashed) and with twice as strong stellar feedback (*StrongFB*, red long-dashed) as model *Ref*. For each model, the typical stellar mass of the central galaxies in the haloes is indicated in the legend. While the covering fraction is substantially lower in the absence of any feedback, all the models with feedback predict similar H I distributions even though the stellar masses vary greatly.

short dashed; *StrongFB*, red long dashed). See Crain et al. (2015) for more information on these simulations. The typical stellar mass of the central galaxies in the chosen haloes is shown next to the name of each simulation. Varying the strength of stellar feedback by a factor of 2 has a dramatic effect on the stellar mass, which increases by nearly an order of magnitude from *StrongFB* to *WeakFB*. However, the effect on the H I distribution is small, with stronger feedback yielding slightly higher covering fractions in the inner haloes. On the other hand, turning off feedback altogether does lead to a large (up to a factor of 2) reduction in the covering fraction of LLSs. The similarity of our *NoFB* results to those of Fumagalli et al. (2014) suggests that the reason why they found low H I covering fractions may be that stellar feedback is highly inefficient in their simulations.

As Fig. 11 shows, the stellar mass does not change significantly between *NoAGN* and *Ref*. This indicates that AGN feedback does not have a very strong impact on galaxies with the halo mass range chosen for this figure ($10^{11.8} < M_{200} < 10^{12.2}$ M$_\odot$ at $z = 2.2$). However, AGN feedback does become important for more massive galaxies, e.g. boosting $f_{<r_{200}}$ by ≈ 20 per cent for galaxies with $M_{200} \sim 10^{12.5}$ M$_\odot$ at $z = 2.2$.

Because of the sensitivity of stellar mass to feedback, the LLS covering fraction is sensitive to feedback if the stellar mass is held fixed, as would be appropriate when comparing to observations of the gas around galaxies selected by stellar mass. This is shown in Fig. 12, where we compare the mean differential covering fraction of LLSs around galaxies with stellar masses in the range $10^{9.8} < M_\star < 10^{10.2}$ M$_\odot$ at $z = 2.2$ (i.e. galaxies similar to LBGs; Shapley et al. 2005). The curves correspond the models presented in the previous figure except that *StrongFB* is missing
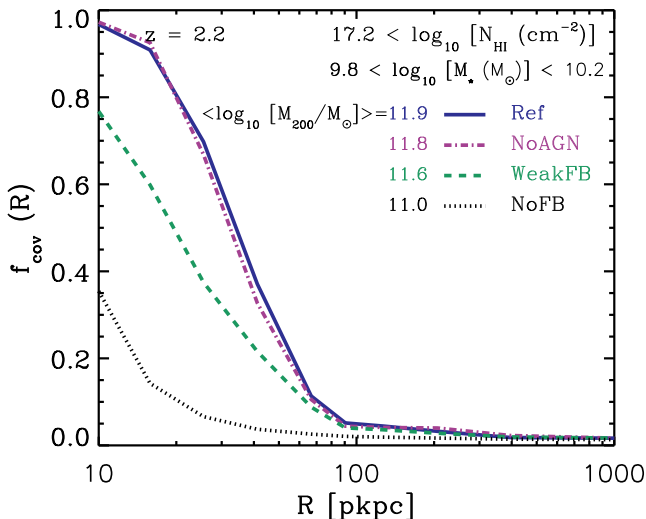
Figure 12. The same as Fig. 11 but for galaxies with stellar mass $10^{9.8} < M_\star < 10^{10.2}$ M$_\odot$. The typical halo masses corresponding to each stellar mass are indicated in the legend. The distribution of H I is highly sensitive to the strength of the feedback if the galaxies are selected by stellar mass, with more efficient feedback yielding higher covering fractions.

because at $z = 2.2$ this simulation does not contain any galaxy with $M_\star > 10^{9.8}$ M$_\odot$. The typical masses of the haloes in which galaxies reside are shown next to the model names. It is evident that the covering fraction increases rapidly with the efficiency of the feedback. Hence, the distribution of H I around galaxies selected by stellar mass, is sensitive probe of their halo mass (see also Kim & Croft 2008; Rakic et al. 2013).

Comparing Figs 11 and 12, we see that in contrast to the other models, the covering fraction predicted by model *Ref* is nearly the same for the samples selected by halo and stellar mass. This suggests that (only) this model reproduces the stellar mass–halo mass relation of LBGs, which is consistent with the finding of (Furlong et al. 2015) that model *Ref* agrees with the observed galaxy stellar mass function at these redshifts. This means that when we compared the simulations results with observations in Section 4.1 and Section 4.2, we could have chosen to match the stellar masses of the simulated galaxies to the observed values instead of matching their halo masses, without obtaining different results.

We note that the sensitivity of the distribution of H I to feedback depends somewhat on the column density. Stronger absorbers, e.g. DLAs, are slightly more sensitive to the feedback efficiency, consistent with previous studies (Theuns et al. 2002; Altay et al. 2013; Rahmati & Schaye 2014).

We conclude that the inclusion of a relatively efficient stellar feedback is necessary to increase the covering fraction of LLSs around LBGs to the observed values (see Section 4.2). At fixed halo mass, but not at fixed stellar mass, the results are insensitive to the precise efficiency of stellar feedback. AGN feedback on the other hand, has a mass-dependent impact and helps to boost the covering fraction LLS around bright quasars to the observed values (see Section 4.1).

6 SUMMARY AND CONCLUSIONS

The observed high covering fractions of strong H I absorbers around high-redshift galaxies and quasars has been identified as a challenge for simulations of galaxy formation (e.g. Fumagalli et al. 2014; Faucher-Giguère et al. 2015). It is therefore important to

test whether the EAGLE cosmological, hydrodynamical simulation, which, thanks to the implemented efficient stellar and AGN feedback, reproduces a large number of observed galaxy properties over wide ranges of mass and redshift, can also reproduce the H I observations.

We combined the EAGLE simulation with photoionization corrections based on radiative transfer simulations of the UVB and recombination radiation, to study the distribution of relatively high H I column densities (i.e. LLSs; $N_{\rm HI} \gtrsim 10^{17}$ cm^{-2}). Because the main EAGLE simulation uses a 100 cMpc box size, it includes a statistically representative sample of galaxies, with a relatively high resolution for a simulation of this kind.

We first demonstrated that EAGLE reproduces the observed column density distribution of strong H I absorbers (i.e. LLSs and DLAs) at $z = 1–5$ and yields an evolution of the cosmic H I density that is in agreement with observations. We then analysed the H I distribution around galaxies from $z = 4$ to 1, bracketing the era during which the cosmic star formation rate peaked. We found that the mean covering fraction of LLSs within the virial radius of galaxies, $f_{<r_{200}}$, evolves strongly from ~70 per cent at $z = 4$ to $\lesssim 10$ per cent at $z = 1$. However, the LLS covering fraction depends only weakly on halo mass at a fixed redshift, particularly for $M_{200} \gtrsim 10^{12}$ M$_\odot$. We also showed that $f_{<r_{200}}$ is insensitive to the specific star formation rate, which suggests that the distribution of LLSs is regulated on time-scales that are longer than the typical time-scale for episodic fluctuations in the star formation rate.

At a fixed impact parameter from galaxies, the covering fractions of LLSs, sub DLAs and DLAs increase rapidly with halo mass and redshift. However, for a fixed redshift and after normalizing the impact parameters to the halo virial radii, the covering fraction profiles of strong H I absorbers around galaxies with $M_{200} \gtrsim 10^{12}$ M$_\odot$ depend only weakly on halo mass. The covering fraction profiles of strong H I absorbers in and around massive haloes are thus nearly scale-invariant and have characteristic lengths similar to the virial radius. Exploiting this result, we presented a fitting function that reproduces the covering fraction profiles for each class of strong H I absorbers (i.e. LLSs, sub DLAs & DLAs) around haloes with $M_{200} \gtrsim 10^{12}$ M$_\odot$ at different redshifts.

For a given halo mass and redshift, there is a significant intrinsic scatter around the mean LLS covering fraction, which is related to the complex geometry of the gas distribution around galaxies. This relatively large scatter limits the predictive power of studies that use small numbers of galaxies to model the distribution of H I.

We compared our predictions with measurements of the covering fraction profiles of strong H I absorbers around LBGs and bright quasars at $z \sim 2–3$ from Rudie et al. (2012) and Prochaska et al. (2013b), finding agreement. This success may be due to our use of a cosmological simulation that includes the efficient stellar and AGN feedback that is required to produce galaxies with properties close to those of the observed population at different epochs. In addition, instead of choosing a fixed mass, redshift and virial radius for calculating the H I distributions, we found it to be important to match the observations more closely. We matched not only the redshift range and halo masses, but also the LOS velocity interval used in observations for finding the absorbers. Moreover, we compared our predictions with the observed covering fractions at the impact parameters that are probed by the observations instead of normalizing them to the virial radii which is problematic since the observed quasars span a range of redshifts and, presumably, halo masses.

Noting that earlier studies showed that local sources of ionizing radiation mainly affect H I absorbers that are very close to galaxies (e.g. Schaye 2006; Rahmati et al. 2013b), it is likely that they would

only significantly change the H I covering fractions for $R \ll r_{200}$. Therefore, we do not expect the neglect of local sources in the present study to change our main findings, although it may explain our overprediction of the H I covering fractions at the smallest observed impact parameters from bright quasars (see Fig. 9). Accounting for the impact of local sources of radiation on the distribution of H I absorbers requires complex modelling of sources in addition to accurate radiative transfer. We postpone such analysis to future work.

We also found the assumed strength of the UVB radiation to be important. While our fiducial UVB model, Haardt & Madau (2001), produces results that are in reasonable agreement with the observations, we found that using a model with a three times weaker UVB, similar to Haardt & Madau (2012) and Faucher-Giguère et al. (2009), improves the agreement with the observed column density distribution of LLSs and weaker absorbers at $z \approx 2.5$. We note, however, that differences in the UVB intensity cannot explain the discrepancy between previous simulations and both EAGLE and the observations, since the earlier models used hydrogen photoionization rates that are close to this lower value (e.g. Fumagalli et al. 2014; Faucher-Giguère et al. 2015; Suresh et al. 2015).

We tested the impact of feedback on our results by comparing EAGLE models that do not include stellar and/or AGN feedback and models that use a factor of two stronger or weaker stellar feedback. The impact of AGN feedback on the H I covering fraction become stronger with increasing halo mass and helps to boost the LLS covering fraction around bright quasars. While efficient stellar feedback is required to increase the H I covering fractions to the observed values, varying its efficiency by a factor of 2 does not change the results significantly at fixed halo mass. This suggests that the H I distribution around galaxies is mainly determined by the cosmic supply of neutral hydrogen into haloes. This conclusion is consistent with the lack of a strong correlation between LLS covering fractions and specific star formation rates, the rapid evolution in the H I distribution around galaxies, the scale invariance of H I distribution in and around massive haloes and the filamentary structure of H I systems around galaxies.

However, at fixed stellar mass the H I covering fractions are highly sensitive to the efficiency of the feedback. Of the EAGLE models analysed here, only the reference model matches the observations of LLSs around LBGs both when the galaxies are selected by the halo mass and by the stellar mass corresponding to the observed galaxies. This confirms that the fiducial EAGLE model reproduces the relation between stellar mass and halo mass for these galaxies.

We have shown that a careful comparison with the observed covering fraction of strong H I absorbers, matching the mass and redshift distribution of the observed galaxies and quasars as well as the allowed velocity differences between absorbers and galaxies, results in agreement between EAGLE and the data. We conclude that these observations therefore do not point to a problem in our general understanding of galaxy formation. Note that EAGLE was not calibrated by considering gas properties and its success in reproducing the H I distribution around galaxies was by no means guaranteed.

REFERENCES

Adelberger K. L., Steidel C. C., Shapley A. E., Pettini M., 2003, ApJ, 584, 45

Adelberger K. L., Shapley A. E., Steidel C. C., Pettini M., Erb D. K., Reddy N. A., 2005, ApJ, 629, 636

Aguirre A., Dow-Hygelund C., Schaye J., Theuns T., 2008, ApJ, 689, 851

Altay G., Theuns T., Schaye J., Crighton N. H. M., Dalla Vecchia C., 2011, ApJ, 737, L37

Altay G., Theuns T., Schaye J., Booth C. M., Dalla Vecchia C., 2013, MNRAS, 2469

Bajtlik S., Duncan R. C., Ostriker J. P., 1988, ApJ, 327, 570

Bauermeister A., Blitz L., Ma C.-P., 2010, ApJ, 717, 323

Becker G. D., Bolton J. S., 2013, MNRAS, 436, 1023

Bird S., Vogelsberger M., Haehnelt M., Sijacki D., Genel S., Torrey P., Springel V., Hernquist L., 2014, MNRAS, 445, 2313

Bolton J. S., Haehnelt M. G., Viel M., Springel V., 2005, MNRAS, 357, 1178

Braun R., 2012, ApJ, 749, 87

Calverley A. P., Becker G. D., Haehnelt M. G., Bolton J. S., 2011, MNRAS, 412, 2543

Chabrier G., 2003, PASP, 115, 763

Chang T.-C., Pen U.-L., Bandura K., Peterson J. B., 2010, Nature, 466, 463

Chen H.-W., Mulchaey J. S., 2009, ApJ, 701, 1219

Crain R. A. et al., 2015, MNRAS, 450, 1937

Crighton N. H. M. et al., 2015, preprint (arXiv:1506.02037)

Dalla Vecchia C., Schaye J., 2012, MNRAS, 426, 140

Davé R., Katz N., Oppenheimer B. D., Kollmeier J. A., Weinberg D. H., 2013, MNRAS, 434, 2645

Delhaize J., Meyer M. J., Staveley-Smith L., Boyle B. J., 2013, MNRAS, 433, 1398

Dolag K., Borgani S., Murante G., Springel V., 2009, MNRAS, 399, 497

Durier F., Dalla Vecchia C., 2012, MNRAS, 419, 465

Erkal D., 2015, MNRAS, 451, 5423

Faucher-Giguère C.-A., Kereš D., 2011, MNRAS, 412, L118

Faucher-Giguère C.-A., Lidz A., Hernquist L., Zaldarriaga M., 2008, ApJ, 688, 85

Faucher-Giguère C.-A., Lidz A., Zaldarriaga M., Hernquist L., 2009, ApJ, 703, 1416

Faucher-Giguère C.-A., Hopkins P. F., Kereš D., Muratov A. L., Quataert E., Murray N., 2015, MNRAS, 449, 987

Fumagalli M., Prochaska J. X., Kasen D., Dekel A., Ceverino D., Primack J. R., 2011, MNRAS, 418, 1796

Fumagalli M., Hennawi J. F., Prochaska J. X., Kasen D., Dekel A., Ceverino D., Primack J., 2014, ApJ, 780, 74

Fumagalli M., O'Meara J. M., Prochaska J. X., Rafelski M., Kanekar N., 2015, MNRAS, 446, 3178

Furlong M. et al., 2015, MNRAS, 450, 4486

Haardt F., Madau P., 2001, in Neumann D. M., Tran J. T. V., eds, Clusters of Galaxies and the High Redshift Universe Observed in X-rays

Haardt F., Madau P., 2012, ApJ, 746, 125

Hennawi J. F. et al., 2006, ApJ, 651, 61

Hopkins P. F., 2013, MNRAS, 428, 2840

Kim Y.-R., Croft R. A. C., 2008, MNRAS, 387, 377

Kim T.-S., Carswell R. F., Cristiani S., D'Odorico S., Giallongo E., 2002, MNRAS, 335, 555

Kim T.-S., Partl A. M., Carswell R. F., Muller V., 2013, A&A, 552, A77

Krumholz M. R., Ellison S. L., Prochaska J. X., Tumlinson J., 2009, ApJ, 701, L12

Lagos C. D. P., Baugh C. M., Zwaan M. A., Lacey C. G., Gonzalez-Perez V., Power C., Swinbank A. M., van Kampen E., 2014, MNRAS, 440, 920

Lagos C. d. P. et al., 2015, preprint (arXiv:1503.04807)

McQuinn M., Oh S. P., Faucher-Giguère C.-A., 2011, ApJ, 743, 82

Martin A. M., Papastergis E., Giovanelli R., Haynes M. P., Springob C. M., Stierwalt S., 2010, ApJ, 723, 1359

Meiksin A., Bolton J. S., Tittley E. R., 2014, MNRAS, 445, 2462

Noterdaeme P., Petitjean P., Ledoux C., Srianand R., 2009, A&A, 505, 1087

Noterdaeme P. et al., 2012, A&A, 547, L1

O'Meara J. M., Prochaska J. X., Burles S., Prochter G., Bernstein R. A., Burgess K. M., 2007, ApJ, 656, 666

O'Meara J. M., Prochaska J. X., Worseck G., Chen H.-W., Madau P., 2013, ApJ, 765, 137

Pawlik A. H., Schaye J., 2008, MNRAS, 389, 651

Pawlik A. H., Schaye J., 2011, MNRAS, 412, 1943

Péroux C., Dessauges-Zavadsky M., D'Odorico S., Sun Kim T., McMahon R. G., 2005, MNRAS, 363, 479

Planck Collaboration I, 2014, A&A, 571, A1

Prochaska J. X., Wolfe A. M., 2009, ApJ, 696, 1543

Prochaska J. X., Herbert-Fort S., Wolfe A. M., 2005, ApJ, 635, 123

Prochaska J. X., Worseck G., O'Meara J. M., 2009, ApJ, 705, L113

Prochaska J. X., Hennawi J. F., Simcoe R. A., 2013a, ApJ, 762, L19

Prochaska J. X. et al., 2013b, ApJ, 776, 136

Rahmati A., Schaye J., 2014, MNRAS, 438, 529

Rahmati A., Pawlik A. H., Raičević M., Schaye J., 2013a, MNRAS, 430, 2427

Rahmati A., Schaye J., Pawlik A. H., Raičević M., 2013b, MNRAS, 431, 2261

Raičević M., Pawlik A. H., Schaye J., Rahmati A., 2014, MNRAS, 437, 2816

Rakic O., Schaye J., Steidel C. C., Rudie G. C., 2012, ApJ, 751, 94

Rakic O., Schaye J., Steidel C. C., Booth C. M., Dalla Vecchia C., Rudie G. C., 2013, MNRAS, 433, 3103

Rao S. M., Turnshek D. A., Nestor D. B., 2006, ApJ, 636, 610

Rauch M. et al., 1997, ApJ, 489, 7

Rhee J., Zwaan M. A., Briggs F. H., Chengalur J. N., Lah P., Oosterloo T., van der Hulst T., 2013, MNRAS, 435, 2693

Rosas-Guevara Y. M. et al., 2013, preprint (arXiv:1312.0598)

Rudie G. C. et al., 2012, ApJ, 750, 67

Rudie G. C., Steidel C. C., Shapley A. E., Pettini M., 2013, ApJ, 769, 146

Schaller M. et al., 2015, MNRAS, 451, 1247

Schaye J., 2001a, ApJ, 559, 507

Schaye J., 2001b, ApJ, 562, L95

Schaye J., 2006, ApJ, 643, 59

Schaye J., Dalla Vecchia C., 2008, MNRAS, 383, 1210

Schaye J., Dalla Vecchia C. et al., 2010, MNRAS, 402, 1536

Schaye J. et al., 2015, MNRAS, 446, 521 (S15)

Shapley A. E., Steidel C. C., Erb D. K., Reddy N. A., Adelberger K. L., Pettini M., Barmby P., Huang J., 2005, ApJ, 626, 698

Shen S., Madau P., Guedes J., Mayer L., Prochaska J. X., Wadsley J., 2013, ApJ, 765, 89

Springel V., 2005, MNRAS, 364, 1105

Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726

Suresh J., Bird S., Vogelsberger M., Genel S., Torrey P., Sijacki D., Springel V., Hernquist L., 2015, MNRAS, 448, 895

Theuns T., Viel M., Kay S., Schaye J., Carswell R. F., Tzanavaris P., 2002, ApJ, 578, L5

Trainor R. F., Steidel C. C., 2012, ApJ, 752, 39

Tumlinson J. et al., 2013, ApJ, 777, 59

Turner M. L., Schaye J., Steidel C. C., Rudie G. C., Strom A. L., 2014, MNRAS, 445, 794

van de Voort F., Schaye J., Altay G., Theuns T., 2012, MNRAS, 421, 2809

Vogelsberger M. et al., 2014, Nature, 509, 177

White M. et al., 2012, MNRAS, 424, 933

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

Zafar T., Péroux C., Popping A., Milliard B., Deharveng J.-M., Frank S., 2013, A&A, 556, AA141

Zwaan M. A., Meyer M. J., Staveley-Smith L., Webster R. L., 2005, MNRAS, 359, L30

APPENDIX A: IMPACT OF VARYING THE UVB

Self-shielding against the ionizing background radiation starts at $N_{\rm H\,I} \gtrsim 10^{18}$ cm^{-2}. Systems with lower column densities may not be dominated by neutral hydrogen. As a result, the abundance of those systems is not only sensitive to the distribution of hydrogen, but also to the intensity of the UVB radiation. Observational constraints on the intensity of the UVB at $2 \lesssim z \lesssim 6$ are model dependent and uncertain within a factor of a few (e.g. Bajtlik et al. 1988; Rauch et al. 1997; Bolton et al. 2005; Faucher-Giguère et al. 2008; Calverley et al. 2011; Becker & Bolton 2013). Models for the UVB are also uncertain due to the assumptions they need to adopt (e.g. the escape fraction of ionizing photons) and differ from each other by a factor of a few (e.g. Haardt & Madau 2001; Faucher-Giguère et al. 2009; Haardt & Madau 2012). This in turn makes the predicted abundance and distribution of H I absorbers with $N_{\rm H\,I} \lesssim 10^{18}$ cm^{-2} uncertain.

The impact of varying the UVB photoionization rate on the H I CDDF at $z = 2.5$ is shown in Fig. A1. The dashed red curve shows the result using our fiducial UVB model (Haardt & Madau 2001). The long-dashed green curve shows the H I CDDF calculated by assuming a constant UVB photoionization rate at all densities (i.e. no self-shielding) which, as mentioned above, starts to deviate from the reference result at $N_{\rm H\,I} \sim 10^{18}$ cm^{-2}. The solid blue curve shows the result of reducing the amplitude of the hydrogen photoionization rate by a factor of 3 from that of our fiducial UVB model, which improved the agreement with the observed abundance of H I absorbers with column-densities $N_{\rm H\,I} < 10^{17}$ cm^{-2} from Rudie et al. (2013). As Fig. A2 illustrates, this weaker UVB model produces somewhat higher covering fractions for LLSs and weaker absorbers, which is also in better agreement with the observations (see Fig. 10).

APPENDIX B: IMPACT OF LOCAL STELLAR RADIATION

In Rahmati et al. (2013b) we post-processed a cosmological hydrodynamic simulation with full radiative transfer of the ionizing background, recombination radiation and local stellar radiation. Here, we use those simulations to estimate the impact of local ionizing radiation from stars on the distribution of H I around galaxies. The
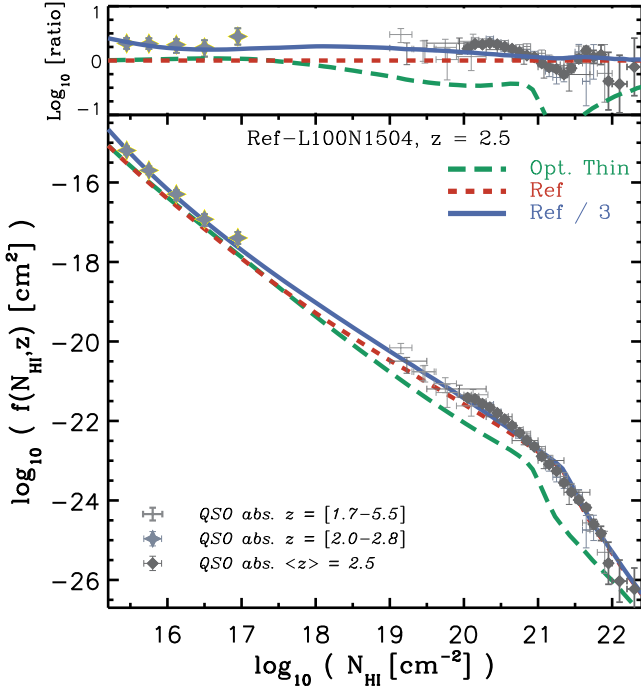
Figure A1. CDDF of neutral gas at $z = 2.5$ for the EAGLE *Ref-L100N1504* simulation and different UVB models. The observational data points are identical to those shown in Fig. 1. The dashed red curve shows our prediction using the fiducial UVB model, i.e. Haardt & Madau (2001) while the solid blue curve shows the result of using a three times smaller hydrogen photoionization rate. The long-dashed green curve shows the result of using the fiducial UVB model without any self-shielding (i.e. the optically-thin limit). The ratios between the two CDDFs and that of the fiducial model are shown in the top panel. The observational measurements, in particular the grey star-shaped data points at $N_{\mathrm{H I}} < 10^{17}$ cm^{-2} taken from Rudie et al. (2013) with $\langle z \rangle \approx 2.4$, are in better agreement with the model with a three times weaker UVB radiation.

underlying cosmological simulation uses the OWLS reference sub-grid feedback model (see Schaye et al. 2010) in a periodic box with $L = 6.25$ cMpc, using cosmological parameters consistent with Wilkinson Microwave Anisotropy Probe 7 year results and a resolution similar to that of the reference model in the present work (see Rahmati et al. 2013b for more details).

In Rahmati et al. (2013b), we showed that photoionization by local stellar radiation becomes more important than, or comparable with, ionization by the UVB in regions less than r_{200} away from galaxies.

To illustrate the impact on the H I covering fractions, we show the differential covering fraction of LLSs around galaxies with $M_{200} \approx 10^{11} \, \mathrm{M_\odot}$ at $z = 2$ in Fig. B1, once without including the local stellar radiation (i.e. in the presence of the UVB and recombination radiation; red solid curve) and once after including it (blue dashed curve). The impact of local stellar radiation on the covering fraction of LLSs is strongest very close to galaxies, the reduction of the covering factor due to local sources increases from ≈ 10 per cent at r_{200} to ≈ 20 per cent at $R \sim 0.1 \, r_{200}$. For DLAs (not shown) the reduction due to local sources varies from ≈ 10 per cent at r_{200} to ≈ 60 per cent at $0.1 \, r_{200}$.

Finally, we note that local AGN could potentially also have a large impact. This is, however, even more uncertain because the ionizing radiation may be anisotropic and variable in time.
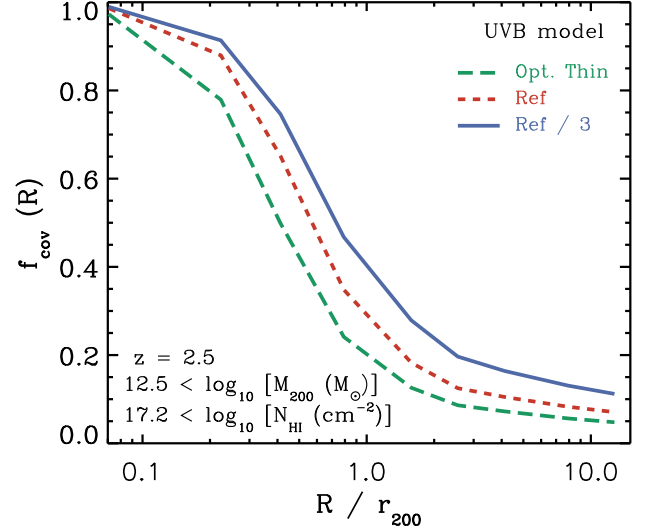
Figure A2. Differential covering fraction of LLSs around galaxies with $M_{200} > 10^{12} \, \mathrm{M_\odot}$ at $z = 2.5$ using different UVB models. The short-dashed orange curve shows the result of using our fiducial UVB model (Haardt & Madau 2001). The solid blue curve shows the result for a UVB model in which the hydrogen photoionization rate is reduced by a factor of 3 compared to our fiducial UVB model and is consistent with the (Haardt & Madau 2012) model. The long-dashed green curve shows the result of not taking into account self-shielding. For calculating the covering fractions only absorbers within a LOS velocity window of $\Delta V = 3150$ km s^{-1} around galaxies are taken into account. The difference between different UVB models affect the distribution of LLS around galaxies significantly and is an important source of uncertainty in the predicted covering fraction profiles.
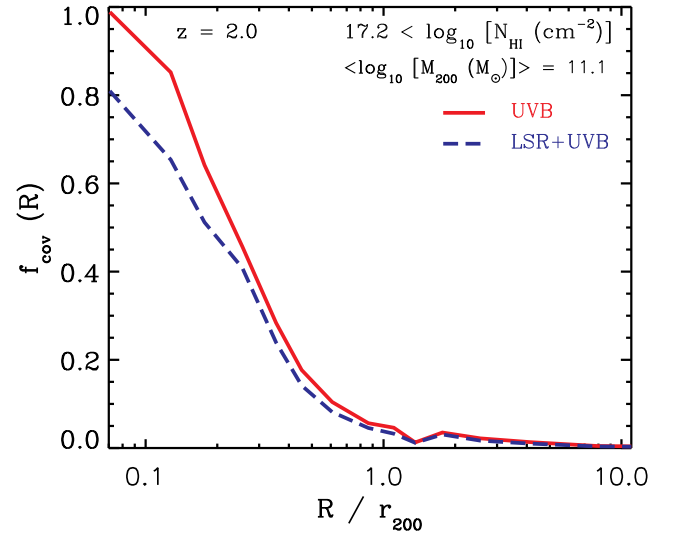
Figure B1. Differential covering fraction of LLSs around galaxies with $M_{200} \approx 10^{11} \, \mathrm{M_\odot}$ at $z = 2$ in the full radiative transfer simulations of Rahmati et al. (2013b) with and without local stellar radiation shown by blue dashed and red solid curves, respectively. Local ionizing radiation from young stars reduces the covering fraction of LLSs close to galaxies by ≈ 20 per cent while at larger impact parameters, $R \gtrsim r_{200}$, the impact of local stellar radiation on the LLS covering fraction becomes smaller.
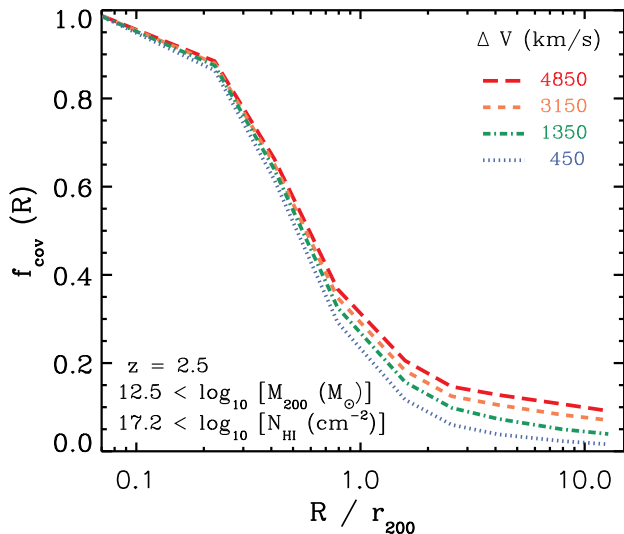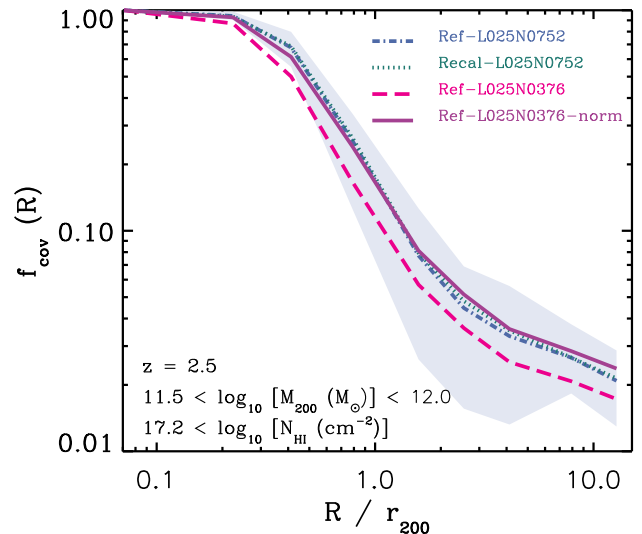
Figure C1. The differential covering fraction of LLSs as a function of normalized impact parameter for different LOS velocity differences between absorbers and galaxies. For impact parameter $R \gtrsim r_{200}$ the covering fraction increases strongly when the velocity cut is increased. Note that the velocity window corresponding to the virial radii of haloes with $M_{200} > 10^{12}$ M$_\odot$ at this redshift is $\lesssim 100$ km s^{-1}

APPENDIX C: ALLOWED LOS VELOCITY DIFFERENCE BETWEEN ABSORBERS AND GALAXIES

As discussed in Section 2.3, when we associate H I absorbers with galaxies we take into account their relative LOS velocities. The typical velocity differences used in observational studies are $\Delta V \sim \pm 1000$ km s^{-1}, which corresponds to cosmic scales that are much larger than what has been used in previous theoretical work based on zoom simulations (e.g. Faucher-Giguère & Kereš 2011; Fumagalli et al. 2011, 2014; Shen et al. 2013; Faucher-Giguère et al. 2015). Fig. C1 shows how LLS covering fraction predictions change by varying the size velocity window that is searched for absorbers around galaxies with $M_{200} > 10^{12}$ M$_\odot$ at $z = 2.5$. The dashed orange curve corresponds to the value we used to calculate the H I covering fraction profiles for comparison with the observations of Prochaska et al. (2013b) who sued a velocity window of $\Delta V = 3000$ km s^{-1} around galaxies. While increasing the size of the allowed velocity window does not affect the result for $R \ll r_{200}$, it increases the LLSs covering fraction for $R \gtrsim r_{200}$. Note that the size of the smallest velocity window we showed in this figure, $\Delta V = 450$ km s^{-1}, is still ≈ 5 times larger than the velocity window that corresponds to the common choice in previous theoretical studies that considered the regions confined within the virial radii of haloes with M_{200} 10^{12} M$_\odot$ at $z = 2.5$.

APPENDIX D: NUMERICAL CONVERGENCE TESTS

To study the impact of numerical resolution on our results, first we use two different cosmological simulations that have identical box sizes of 25 comoving Mpc, but different resolutions. The first simulation, *Ref-L025N0376*, has a resolution that is identical to that of our fiducial simulation (i.e. *Ref-L100N1504*) and the second simulation, *Ref-L025N0752*, has an identical box-size and sub-grid

Figure D1. Differential covering fraction of LLSs around haloes with $10^{11.5} < M_{200} < 10^{12}$ M$_\odot$ at $z = 2.5$ as a function of normalized impact parameter for simulations with different resolutions. The dashed magenta curve shows the *Ref-L025N0376* simulation, which is identical to our fiducial simulation (i.e. *Ref-L100N1504*) except for having a smaller box size of 25 comoving Mpc. The dot–dashed blue curve shows a simulation with eight times higher mass resolution and the shaded area around it shows the 15– 85 percentiles for the *Ref-L025N0752* simulation. The dotted green curve, which is almost identical to the dot–dashed curve, shows a high-resolution simulation recalibrate to achieve weak convergence, i.e. reproducing galaxies with properties very similar to those in the fiducial resolution. The long-dashed purple curve shows the result from the *Ref-L025N0376* simulation but after modifying the UVB such that the global CDDF of LLSs becomes identical to that of the *Ref-L025N0752* simulation. For calculating the covering fractions only absorbers within a LOS velocity window of $\Delta V = 3150$ km s^{-1} around galaxies are taken into account. While a higher resolution results in larger numbers of LLSs and therefore a higher LLS covering fractions, the difference is small compared to the intrinsic scatter of the covering fraction and other uncertainties like the intensity of the UVB radiation. Indeed, as the long-dashed curve shows, if the simulations with different resolutions are normalized to have the same cosmological distribution of LLSs, the covering fraction of LLS around galaxies becomes insensitive to the resolution.

physics but eight times better mass resolution. As Fig. D1 shows, the LLS covering fraction around haloes with $10^{11.5} < M_{200} < 10^{12}$ M$_\odot$ at $z = 2.5$ is not fully converged with resolution and increases by increasing the resolution of the simulation. The dotted green curves shows the results in the *Recal-L025N0752* simulation, for which the feedback implementation is recalibrated to reproduce similar galaxy properties to those found in the *Ref-L025N0376*. As the figure shows, recalibrating hardly has any impact on the covering fraction of LLSs around galaxies (for details of the *Recal-L025N0752* model see S15). However, as the shaded area around the solid blue curve shows, the intrinsic scatter around the mean LLS covering fraction is larger than the difference between the results at different resolutions. Moreover, other uncertainties, such as the amplitude of the UVB radiation, have larger effects on the LLS covering fractions than the resolution effect. In fact, for any resolution, the UVB model should be recalibrate such that the cosmic distribution of H I absorbers is well reproduced. As the long-dashed curve in Fig. D1 shows, this would reduce the differences in the covering fractions of simulations that have different resolutions.
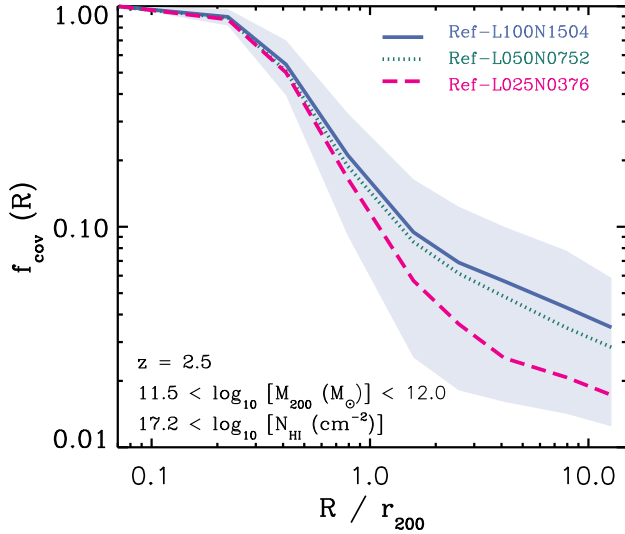
Figure D2. Differential covering fraction of LLSs around haloes with $10^{11.5} < M_{200} < 10^{12}\,\mathrm{M_\odot}$ at $z = 2.5$ as a function of normalized impact parameter for simulations with different box sizes. The solid blue curve shows the *Ref-L100N1504* simulation. Dotted green and dashed magenta curves show simulations with box sizes of 50 and 25 comoving Mpc, respectively. The shaded area around the solid curve shows the 15–85 percentiles for the *Ref-L100N1504* simulation. For calculating the covering fractions only absorbers within a LOS velocity window of $\Delta V = 1294\ \mathrm{km\,s^{-1}}$ around galaxies are taken into account. The LLS covering fraction within $R \gtrsim r_{200}$ increases with the size of the simulation box but the simulations with $L_{\mathrm{box}} \gtrsim 50$ cMpc are nearly converged.

The box-size sensitivity of the LLS covering fraction around haloes with $10^{11.5} < M_{200} < 10^{12}\,\mathrm{M_\odot}$ at $z = 2.5$ is shown in Fig. D2 where the *Ref-L100N1504* simulation is compared with simulations with identical resolution but factors of 2 and 4 smaller box sizes, the *Ref-L050N0752* and *Ref-L025N0376* simulations shown with green dotted and red dashed curves, respectively. The LLSs covering fraction at $R \gtrsim r_{200}$ increases with increasing the simulation box-size but converges for $L_{\mathrm{box}} \gtrsim 50$ cMpc. This highlights the importance of having a large cosmological box for successfully simulating the enhanced distribution of LLSs out to large impact parameters.

APPENDIX E: HYDRODYNAMICS

As mentioned in Section 2, in EAGLE we used ANARCHY (Dalla Vecchio in prep) for hydrodynamics instead of using the default hydrodynamics implementation of GADGET-3. ANARCHY uses the SPH formulation derived by Hopkins (2013) in addition to modified artificial viscosity switch and time step limiters (see appendix A in S15 for more details).

The difference caused in the H I distribution by using ANARCHY instead of the default GADGET-3 hydrodynamics is shown in Fig. E1 for a halo with $M_{200} = 10^{12.3}\,\mathrm{M_\odot}$ at $z = 2.2$ and in the absence of any feedback. The left-hand and right-hand columns show the result obtained using ANARCHY and GADGET-3 , respectively. The H I distribution looks smoother in results obtained by ANARCHY and the resulting LLS covering fractions, $f_{<r_{200}}$, are slightly higher than those in GADGET-3 . This trend can be seen in the differential LLS covering fraction profiles of galaxies with similar masses, as shown in Fig. E3. However, the typical differences in the covering frac-
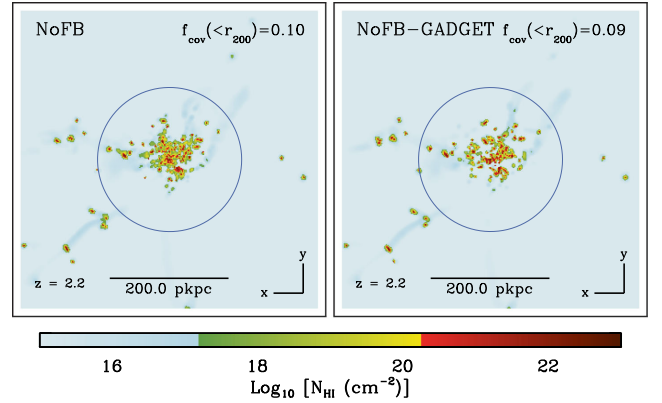

Figure E1. The impact of various hydrodynamics formalism on the distribution of H I around a simulated galaxy in a halo with $M_{200} = 10^{12.3}\,\mathrm{M_\odot}$ at $z = 2.2$ and in the absence of feedback. The left-hand column shows the H I distributions around the halo calculated using ANARCHY, our fiducial hydrodynamics implementation and the right-hand column shows the same but using the standard GADGET-3 implementation. The virial radius of the halo is indicated with the blue circle centred on the halo. The size of the region is 500×500 pkpc. The LLS covering fraction, $f_{<r_{200}}$, is indicated on the top-right corner of each panel. While the covering fraction of LLSs, $f_{<r_{200}}$ is slightly larger in the results obtained using ANARCHY, the difference is smaller than the variations expected due to orientations of galaxies, object to object variations and the impact of feedback.
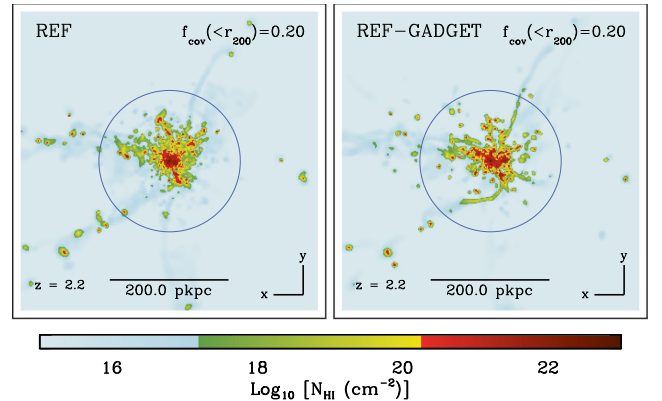
Figure E2. Similar to Fig. E1 but in the presence of stellar and AGN feedback. Despite producing slightly different H I distributions, both SPH implementations result in very similar LLS covering fractions. The impact of different hydrodynamics implementations on the H I covering fractions, which was small in the absence of feedback, becomes even smaller in its presence.

tions are smaller than the variations expected due to orientations of galaxies, object to object variations and the impact of feedback.

In the presence of stellar and AGN feedback, the H I distribution produced by ANARCHY and GADGET-3 look slightly different as shown in Figs E2 and E3. The LLS covering fractions, however, are almost identical. We conclude that the impact of different hydrodynamics implementations on the H I covering fractions, which was small in the absence of feedback, becomes even smaller in its presence.
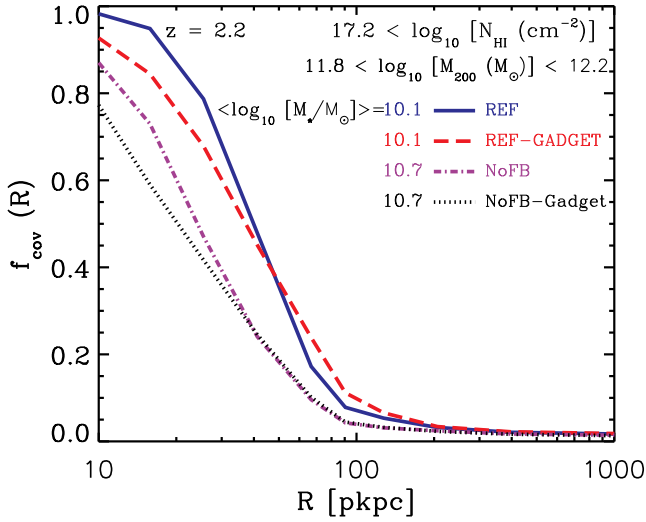
Figure E3. Differential covering fraction of LLSs around *haloes* with $10^{11.8} < M_{200} < 10^{12.2}$ M$_\odot$ at $z = 2.2$ as a function of impact parameter for simulations with different hydrodynamics and feedback implementations. The solid blue curve shows the *REF-L025N0376* simulation which uses the fiducial feedback implementation and ANARCHY hydrodynamics implementation. Long dashed red curve show the result of using standard GADGET-3 SPH implementation in the presence of our fiducial feedback. The black dotted and purple dot–dashed curves show simulations without any feedback which use ANARCHY and GADGET-3 , respectively. The median stellar mass corresponding to haloes in each model is indicated on the left-hand side of the relevant name. For calculating the covering fractions only absorbers within a LOS velocity window of $\Delta V = 1294$ km s^{-1} around galaxies are taken into account. The difference in the LLSs distributions caused by varying the hydrodynamics is much smaller than the difference caused by feedback, and the typical scatter due to galaxy-to-galaxy variations.

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