4

Predictions for the relation between strong HI absorbers and galaxies at redshift 3

We combine cosmological, hydrodynamical simulations with accurate radiative transfer corrections to investigate the relation between strong HI absorbers ($N_{\rm HI} > 10^{16}\,{\rm cm}^{-2}$) and galaxies at redshift $z = 3$. We find a strong anti-correlation between the column density and the impact parameter that connects the absorber to the nearest galaxy. The median impact parameters for Lyman Limit (LL) and Damped Lymanα (DLA) systems are ~ 10 and ~ 1 proper kpc, respectively. If normalized to the size of the halo of the nearest central galaxy, the median impact parameters for LL and DLA systems become ~ 1 and ~ 0.1 virial radii, respectively. At a given HI column density the impact parameter increases with the mass of the closest galaxy, in agreement with observations. We predict most strong HI absorbers to be most closely associated with extremely low-mass galaxies, $M_\star \lesssim 10^8\,M_\odot$. We also find a correlation between the column density of absorbers and the mass of the nearest galaxy. This correlation is most pronounced for DLAs with $N_{\rm HI} > 10^{21}\,{\rm cm}^{-2}$ which are typically close to galaxies with $M_\star \gtrsim 10^9\,M_\odot$. Similar correlations exist between column density and other properties of the associated galaxies such as their star formation rates, halo masses and HI content. The galaxies nearest to HI absorbers are typically far too faint to be detectable with current instrumentation, which is consistent with the high rate of (often unpublished) non-detections in observational searches for the galaxy counterparts of strong HI absorbers. Moreover, we predict that the detected nearby galaxies are typically not the galaxies that are most closely associated with the absorbers, thus causing the impact parameters, star formation rates and stellar masses of the counterparts to be biased high.

Alireza Rahmati, Joop Schaye

to be submitted

4.1 Introduction

Studies of high-redshift galaxy are often based on the light emitted by stars and hot/ionized gas. This limits the observations to the small fraction of galaxies that are bright enough to be detected in emission. Given the large number density of faint galaxies, one may speculate that most high-redshift galaxies are missing from observational studies. The analysis of absorption features in the spectra of background QSOs, provides an alternative probe of the distribution of matter at high redshifts. The large distances that separate most absorbers from their background QSOs make it unlikely that there is a physical connection between them. This opens up a window to study an unbiased sample of matter that resides between us and background QSOs. The rare strong HI Lyα absorbers which are easily recognizable in the spectra of background QSOs due to their Lyα damping wings, for which they are called Damped Lyman-α (DLA) systems[1], are of particular interest. DLAs are likely to be representative of the cold gas in, or close to, the interstellar medium (ISM) in high-redshift galaxies (Wolfe et al., 1986). Because of this, DLAs provide a unique opportunity to define an absorption-selected galaxy sample and to study the ISM, particularly at early stages of galaxy formation, and they have therefore been studied intensely since their discovery (see Wolfe et al., 2005 for a review).

Based on the observed velocity width of metal lines associated with DLAs, it was initially suggested that large, massive galactic disks are responsible for the observed DLAs at $z \sim 3$ (Prochaska & Wolfe, 1997, 1998). However, it has been shown that (collections of) smaller systems are also capable of having high velocity dispersions as a result of infall of material during structure formation (Haehnelt et al., 1998) or galactic winds (McDonald & Miralda-Escudé, 1999; Schaye, 2001a). Nevertheless, reproducing the observed velocity width distribution remains a challenge for hydrodynamical simulations (e.g., Razoumov et al., 2006; Pontzen et al., 2008).

Some recent studies suggest that at $z \sim 2-3$, a large fraction of strong HI absorbers like Lyman Limit Systems (LLS; $N_{\rm HI} > 10^{17}\,{\rm cm}^{-2}$) and DLAs are primarily associated with galaxies similar to Lyman-Break Galaxies (e.g., Steidel et al., 2010; Rudie et al., 2012; Font-Ribera et al., 2012), which have stellar and total halo masses $\sim 10^{10}$ and $\sim 10^{12}$ $M_\odot$, respectively. If such massive galaxies were indeed the prime hosts of strong HI absorbers, then many of the galaxy counterparts of strong absorbers should be detectable with current surveys. However, observations that aim to find galaxies close to DLAs often result in non-detections (e.g., Foltz et al., 1986; Smith et al., 1989; Lowenthal et al., 1995; Bunker et al., 1999; Prochaska et al., 2002; Kulkarni et al., 2006; Rahmani et al., 2010; Bouché et al., 2012) or find galaxies that are at unexpectedly large impact parameters from DLAs (e.g., Yanny et al., 1990; Teplitz et al., 1998; Mannucci et al., 1998). Those findings suggest that strong HI

[1]The official column density limit of a DLA is somewhat arbitrarily defined to be $N_{\rm HI} > 10^{20.3}\,{\rm cm}^{-2}$.

systems such as DLAs are more closely associated to low-mass galaxies which are too faint to be observable with the detection thresholds of the current studies.

Because observational studies are limited by the small number of known DLAs and are missing low-mass galaxies, we have to resort to cosmological simulations to help us understand the link between DLAs and galaxies. Many studies have used simulations to investigate the nature of strong HI absorbers and particularly DLAs (e.g., Gardner et al., 1997, 2001; Haehnelt et al., 1998; Nagamine et al., 2004; Razoumov et al., 2006; Pontzen et al., 2008; Tescari et al., 2009; Fumagalli et al., 2011; Cen, 2012; van de Voort et al., 2012a). To maximize the numerical resolution required for accurate modelling of the high HI column densities, most previous studies have used small simulation boxes or zoomed simulations. Those studies often try to compensate for the lack of a full cosmological distribution of absorbers by combining the results from their small-scale simulations with analytic halo mass functions to predict the properties of the DLA population (e.g., Gardner et al., 1997, 2001) or to determine what kinds of galaxies dominate the cosmic DLA distribution (e.g., Pontzen et al., 2008). This approach requires some preconceptions about the types of environments that can give rise to DLA absorbers and cannot easily account for the large scatter in the distribution of absorbers in halos with similar properties. Moreover, zoom simulations cannot fully capture the possibility of absorber-galaxy pairs being found outside of the considered regions. As a result, the statistical properties found using zoomed simulations may be biased. Finally, the impact of finite detection thresholds on the observed relation between strong HI absorbers and galaxies cannot be studied with simulations that do not contain a representative sample of HI absorbers and galaxies.

In this work, we use cosmological hydrodynamical simulations that contain a representative sample of the full distribution of strong HI systems (Rahmati et al., 2013a). Similar to what is done observationally, we connect each absorber to its nearest galaxy. A significant improvement in this work is the use of photoionization corrections that are based on accurate radiative transfer simulations and that account for both the uniform ultraviolet background (UVB) radiation and recombination radiation. In addition we show that our main conclusions are insensitive to the inclusion of local sources and to variations in the subgrid physics. The ionization corrections we use have been shown to reproduce the observed HI column density distribution function over a wide range of redshifts (Rahmati et al., 2013a).

We predict correlations between the column density of strong HI absorbers, their impact parameters, and the properties of the associated galaxies. While the fraction of HI absorbers that are linked to relatively massive galaxies increases with HI column density, most LLS and DLAs are closely associated with very low-mass galaxies, with typically $M_\star \lesssim 10^8\ M_\odot$, that are generally undetectable with current instruments. We show that our predictions are nevertheless in good agreement with existing observations, including those of Rudie et al. (2012) who found that a large fraction of strong HI absorbers at $z \sim 3$ are within 300 proper

kpc radius from massive Lyman-Break galaxies.

The structure of this chapter is as follows. In §4.2 we discuss our numerical simulations and ionization calculations for obtaining the HI column densities and describe our method for connecting HI systems to their host galaxies. We present our results in §4.3 and compare them with observations. In this section we also investigate how the distribution of HI absorbers varies with the properties of their host galaxies. Finally, we conclude in §4.4.

4.2 Simulation techniques

In this section we describe different parts of our simulations. We briefly explain the details of the hydrodynamical simulations that are post-processed to get the HI distribution by accounting for various ionization processes. Then we explain our halo finding method, our HI column density calculations, and the procedure we use to connect HI absorbers to their host halos.

4.2.1 Hydrodynamical simulations

We use cosmological simulations performed using a significantly modified and extended version of the smoothed particle hydrodynamics (SPH) code GADGET-3 (last described in Springel, 2005). The simulations are part of the Overwhelmingly Large Simulations (OWLS) described in Schaye et al. (2010). For our reference model, we use a subgrid pressure-dependent star formation prescription of Schaye & Dalla Vecchia (2008) which reproduces the observed Kennicutt-Schmidt law. The chemodynamics is based on the model of Wiersma et al. (2009b) which follows the abundances of eleven elements assuming a Chabrier (2003) IMF. 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., 2009a). Galactic winds driven by star formation are modeled using a kinetic feedback recipe that assumes 40% of the kinetic energy generated by Type II SNe is injected as outflows with initial velocity of 600 kms^{-1} and with a mass loading factor $\eta = 2$ (Dalla Vecchia & Schaye, 2008). To bracket the impact of feedback, we also consider simulations with different feedback and sub-grid models. We found that our results are not sensitive to the variations in feedback and sub-grid physics (see Appendix B).

We adopt cosmological parameters consistent with the WMAP year 7 results: $\{\Omega_\mathrm{m} = 0.272,\ \Omega_\mathrm{b} = 0.0455,\ \Omega_\Lambda = 0.728,\ \sigma_8 = 0.81,\ n_\mathrm{s} = 0.967,\ h = 0.704\}$ (Komatsu et al., 2011). Our reference simulation has a periodic box of $L = 25$ comoving h^{-1}Mpc and contains 512^3 dark matter particles with mass $6.3 \times 10^6\ h^{-1}\mathrm{M}_\odot$ and an equal number of baryons with initial mass $1.4 \times 10^6\ h^{-1}\mathrm{M}_\odot$. The Plummer equivalent gravitational softening length is set to $\epsilon_\mathrm{com} = 1.95\ h^{-1}$kpc and is limited to a minimum physical scale of

$\epsilon_{\rm prop} = 0.5\ h^{-1}$kpc. In addition to our reference simulation explained above, we use simulations with different resolutions and box-sizes to investigate numerical effects (see Appendix D).

4.2.2 Finding galaxies

For identifying individual galaxies in our cosmological simulations, we assume that galaxies are bound to their dark matter haloes. This assumption implies that any given baryonic particle belongs to its closest dark matter halo. We use the Friends-of-Friends (FoF) algorithm to identify groups of dark matter particles that are near each other (i.e., FoF haloes), using a linking length of $b = 0.2$. Then, we use SUBFIND (Dolag et al., 2009) to connect gravitationally bound particles as part of unique structures (halos) and to identify the center of each halo/galaxy as the position of the most bound particle in that halo. We take the radius within which the average density of a given halo reaches 200 times the mean density of the Universe at a given redshift, R_{200}, as the size of that halo. The galaxy that sits in the center of each halo is considered as a *central* galaxy and all the other gravitationally bound structures in that FoF halo are considered as *satellite* galaxies. Note that we do not require satellite galaxies to be within the R_{200} of their central galaxy.

In our analysis, we use all the simulated galaxies that have star formation rates SFR $> 4 \times 10^{-3}$ M$_\odot$ yr^{-1}. By using this SFR threshold, more than 99% of our selected galaxies are resolved with > 100 resolution elements (i.e., dark matter particles and/or baryonic particles). We test the impact of different SFR thresholds on our results, which provides useful insights for observational studies with finite detection threshold (see §4.3).

4.2.3 Finding strong HI absorbers

The first step in identifying HI absorbers in the simulations is to accurately calculate the hydrogen neutral fractions. To accomplish this, the main ionization processes that shape the distribution of neutral hydrogen must be accounted for. In this context, photoionization by the metagalactic UVB radiation is the main contributor to hydrogen ionization at $z \gtrsim 1$ while collisional ionization becomes more important at lower redshifts (Rahmati et al., 2013a). Although the photoionization from local stellar radiation is the dominant source of ionization at high HI column densities (e.g., Rahmati et al., 2013b), our tests show that it does not have a significant impact on the results we present in this work (see Appendix C).

We use the UVB model of Haardt & Madau (2001) to account for the large-scale photoionization effect of quasars and galaxies. The same UVB model is used for calculating heating/cooling in our hydrodynamical simulations. It has been shown that this UVB model is consistent with metal absorption lines at

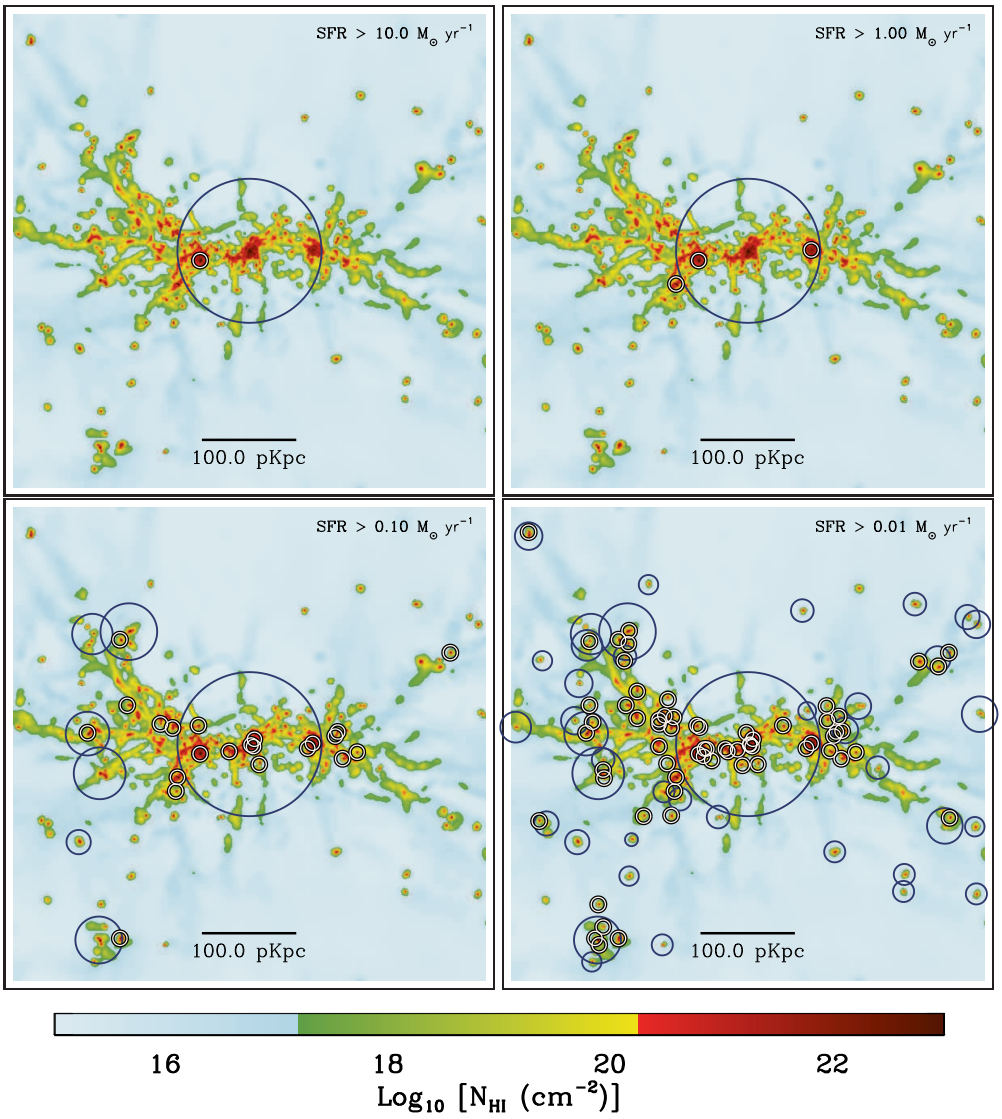

Figure 4.1: The simulated HI column density distribution around a massive galaxy with $M_\star = 10^{10}$ $M_\odot$ and SFR $= 29$ $M_\odot$ yr^{-1} at $z = 3$. In each panel, galaxies with different SFRs are at the center of circles. The size of dark circles indicates the virial radius of the central galaxies (R_{200}) while the small white circles show satellite galaxies. From top-left to bottom-right, panels show galaxies with SFR > 10 $M_\odot$ yr^{-1}, SFR > 1 $M_\odot$ yr^{-1}, SFR > 0.1 $M_\odot$ yr^{-1} and SFR > 0.01 $M_\odot$ yr^{-1}, respectively. As the SFR threshold decreases, more galaxies are detected and the typical impact parameter between galaxies and absorbers decreases.

$z \sim 3$ (Aguirre et al., 2008) and the observed HI column density distribution function and its evolution in a wide range of redshifts (Rahmati et al., 2013a).

We characterize the UVB by its optically thin hydrogen photoionization rate, $\Gamma_{\rm UVB}$, and the effective hydrogen absorption cross-section, $\bar{\sigma}_{\nu_{\rm HI}}$ (see equations 3 and 4 and Table 2 in Rahmati et al., 2013a). In self-shielded regions, $\Gamma_{\rm UVB}$ is attenuated to an effective total photoionization rate, $\Gamma_{\rm Phot}$, which is decreasing with density. In Rahmati et al. (2013a) we performed radiative transfer simulations of the UVB and recombination radiation in cosmological density fields using TRAPHIC (Pawlik & Schaye, 2008, 2011). We showed that the effective photoionization rate at all densities can be accurately reproduced by the following fitting function:

$$\frac{\Gamma_{\rm Phot}}{\Gamma_{\rm UVB}} = 0.98 \left[1+\left(\frac{n_{\rm H}}{n_{\rm H,SSh}}\right)^{1.64}\right]^{-2.28} + 0.02 \left[1+\frac{n_{\rm H}}{n_{\rm H,SSh}}\right]^{-0.84}, \qquad (4.1)$$

where $n_{\rm H}$ is the hydrogen number density and $n_{\rm H,SSh}$ is the self-shielding density threshold given by

$$n_{\rm H,SSh} = 6.73\times10^{-3}\,{\rm cm}^{-3} \left(\frac{\bar{\sigma}_{\nu_{\rm HI}}}{2.49\times10^{-18}\,{\rm cm}^2}\right)^{-2/3} \left(\frac{\Gamma_{\rm UVB}}{10^{-12}\,{\rm s}^{-1}}\right)^{2/3}. \qquad (4.2)$$

We use the photoionization rate from equations (4.1) and (4.2) together with the hydrogen number density and temperature of each SPH particle in our hydrodynamical simulations to calculate the equilibrium hydrogen neutral fraction of that particle in post-processing (see Appendix A2 in Rahmati et al., 2013a). It is also worth noting that in our hydrodynamical simulations, ISM gas particles (which all have densities $n_{\rm H} > 0.1\ {\rm cm}^{-3}$) follow a polytropic equation of state that defines their temperatures. Since these temperatures are not physical and only measure the imposed pressure (Schaye & Dalla Vecchia, 2008), in our calculations we set the temperature of the ISM particles to $T_{\rm ISM} = 10^4$ K, the typical temperature of the warm-neutral phase of the ISM.

At very high HI column densities, where the gas density and the optical depth for H_2-dissociating radiation is high, hydrogen is expected to be mainly molecular. This process has been suggested as an explanation for the observed cut-off in the abundance of absorbers at $N_{\rm HI} \gtrsim 10^{22}\,{\rm cm}^{-2}$ (Schaye, 2001c; Krumholz et al., 2009; Prochaska & Wolfe, 2009). It has been also shown that accounting for H_2 formation can produce a good agreement between cosmological simulations and observations of the HI column density distribution function (Altay et al., 2011; Rahmati et al., 2013a). To test the impact of H_2 formation on the spatial distribution of HI absorbers, we adopted the observationally inferred pressure law of Blitz & Rosolowsky (2006) to compute the H_2 fractions in post-processing (see Appendix A in Rahmati et al., 2013b). Once the H_2 fractions have been calculated, we exclude the molecular hydrogen from the total neutral gas for calculating the HI column densities. We note that the adopted empirical relation for calculating the H_2 fractions is calibrated based on observation of

low redshift galaxies and may not be accurate in very low metallicity regimes. However, due to the tight mass-metallicity-SFR relation observed out to $z \gtrsim 3$ (Mannucci et al., 2010; Lara-López et al., 2010), exceptionally low metallicities are not very likely to be typical at $z = 3$, the redshift on which we focus here.

We calculate HI column densities by projecting the HI content of the simulation box along each axis onto a grid with 10000^2 pixels[2], using SPH interpolation. While the projection may merge distinct systems along the line of sight, this is not expected to affect very high HI column density systems that are rare in the relatively small simulation boxes that we use. We tested the impact of projection effects by performing projections using multiple slices instead of the full box. In fact, our numerical experiments show that at $z = 3$ and for simulations with box sizes comparable to that of our simulation, the effect of projection starts to appear only at $N_{\rm HI} < 10^{16}\,{\rm cm}^{-3}$. Since the focus of our study is to characterize the properties of strong HI absorbers with $N_{\rm HI} \gtrsim 10^{17}\,{\rm cm}^{-3}$, our results are not sensitive to the above mentioned projection effect. The top-left panel of Figure 4.1 shows an example for the distribution of the HI column density around a galaxy with stellar mass of $M_\star = 10^{10}\,M_\odot$ in our simulation at $z = 3$.

In addition to HI column densities, we calculate the HI-weighted velocity along each line of sight (LOS), $\langle V_{\rm LOS}\rangle_{\rm HI}$ and use it to constrain the position of the strongest absorber along the projection direction. That this is the same procedure used in observations to associate absorbers and galaxies. Compared to the actual LOS velocity of absorbers, calculating the HI-weighted velocity is less expensive but using it as a proxy for the position of the strongest absorber along each line of sight might be prone to projection effects. However, we found that both the $\langle V_{\rm LOS}\rangle_{\rm HI}$ velocity and the actual velocity of absorber along the LOS produce nearly identical results. For calculating $\langle V_{\rm LOS}\rangle_{\rm HI}$ we use the combination of local peculiar velocity of each SPH particle, $\overrightarrow{V}_{\rm peculiar}$, and its Hubble velocity, $\overrightarrow{V}_{\rm Hubble}$:

$$\langle V_{\rm LOS}\rangle_{\rm HI} = \left\langle (\overrightarrow{V}_{\rm peculiar} + \overrightarrow{V}_{\rm Hubble}) \,.\, \hat{x}_{\rm LOS} \right\rangle_{\rm HI}, \tag{4.3}$$

where $\hat{x}_{\rm LOS}$ is the unit vector along the projection direction. The Hubble velocity of each SPH particle is given by the position vector of that particle, $\overrightarrow{X}$, and the Hubble parameter, $H(z)$,

$$\overrightarrow{V}_{\rm Hubble} = H(z)\ \overrightarrow{X}. \tag{4.4}$$

We take the Cartesian coordinates of the simulation box as the reference frame. As we describe in the next section, we use the same reference frame to calculate the velocities of galaxies along the projection direction. The choice of the origin is not important since we use the velocity differences between HI absorbers and galaxies for our analysis and not the absolute velocities.

[2]Using 10000^2 cells produces converged results. The corresponding cell size is similar to the minimum smoothing length of SPH particles at $z = 3$ in our simulation.

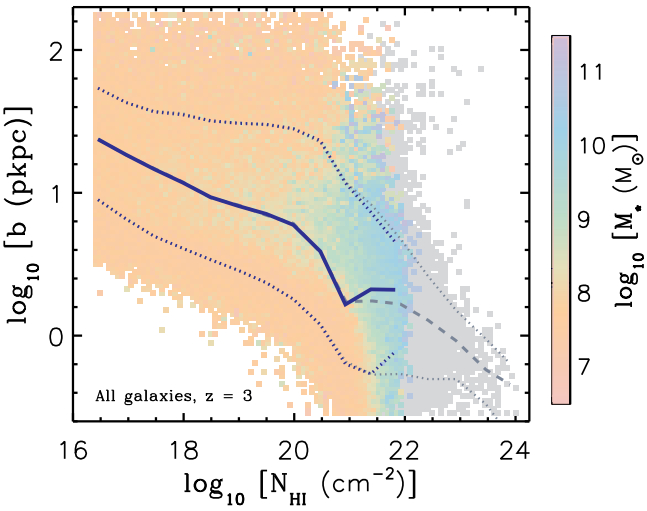

Figure 4.2: The predicted impact parameter of absorbers (in proper kpc) as a function of HI column density at $z = 3$. The color of each cell (in the 2D grid) shows the median stellar mass of the galaxies associated with the HI absorbers in that cell. The median impact parameter as a function of $N_{\rm HI}$ is shown using blue solid curve while the dotted curves indicate the $15\% - 85\%$ percentiles. The gray cells show the region where the H_2 formation drains the atomic gas. The gray dashed and dotted curves show the median impact parameter and the $15\% - 85\%$ percentiles as a function of $N_{\rm HI}$ for a fully atomic HI where the conversion of high pressure gas into H_2 is neglected.

4.2.4 Connecting HI absorbers to galaxies

An example of how galaxies and HI absorbers are distributed in our simulation is shown in Figure 4.1. The colored map, which is repeated in all four panels, shows the HI column density distribution in a 500^2 proper kpc^2 region which is centered on a galaxy with $M_\star = 10^{10}$ $M_\odot$. Galaxies are shown with circles while the star formation rate cut for illustrating galaxies is decreasing from SFR > 10 $M_\odot$ yr^{-1} in the top-left panel to SFR > 0.01 $M_\odot$ yr^{-1} in the bottom-right panel. The dark circles, with sizes proportional to the virial radius of galaxies, are centered on the central galaxies and the white small circles show satellite galaxies. Galaxies with different SFRs are shown in the bottom panels of Figure 4.1. As this figure shows, galaxies and LLSs and DLAs (that are shown using green and red colors, respectively) are strongly correlated. In addition, it seems that the HI column density of absorbers is increasing as they get closer to the center of the galaxies. For a quantitative study, a well defined connection between absorbers and galaxies must be established. This connection can be made in two ways: by linking any given absorber to its closest galaxy (i.e., absorber-centered) or by finding absorbers that are closest to a given galaxy (i.e.,

galaxy-centered). In the present work, we use the absorber-centered matching to connect the simulated HI absorbers to their neighboring galaxies. This approach is particularly efficient for associating rare strong HI absorbers to galaxies. The galaxy-centered approach, on the other hand, is more suited for studying the properties of absorbers around certain classes of galaxies. As we discuss in §4.3.6, these two approaches are closely related, but not identical.

The projected distances between HI absorbers and galaxies, together with their LOS velocity differences, can be used to associate them with each other. We use this method for a direct comparison between simulations and observational studies that employ the same approach. First, we calculate the velocity of each simulated galaxy along the LOS by adding its peculiar and Hubble velocities along the projection direction (see equation 4.3). Then, for every simulated absorber we define a galaxy counterpart that has the shortest projected distance (i.e., the impact parameter) among galaxies with the LOS velocity differences less than a chosen maximum value, $\Delta V_{\rm LOS,\,max}$, with respect to the LOS velocity of the absorber, $\langle V_{\rm LOS} \rangle_{\rm HI}$. With this approach, each galaxy can be connected to more than one absorber, but each absorber is connected to one and only one galaxy.

We note that the difference between the LOS velocities of absorbers and galaxies includes not only the distance between the absorbers and galaxies along the LOS, but also their relative peculiar velocities along the LOS. Therefore, choosing values of $\Delta V_{\rm LOS,\,max}$ that are less than the expected peculiar velocities around galaxies results in unphysical associations between HI absorbers and neighboring galaxies. We know that accretion of the gas into halos together with galactic outflows produces typical peculiar velocities of a few hundreds of kilometers per second. Similar velocity differences have been observed between the LOS velocity of absorbers and their host galaxies (e.g., Fynbo et al., 1999; Rakic et al., 2012; Rudie et al., 2012) in addition to being common in our simulations (van de Voort & Schaye, 2012b). For this reason, we chose $\Delta V_{\rm LOS,\,max} = 300$ km s^{-1}, which is consistent with recent observations (Rudie et al., 2012). However, as we show in Appendix A, our results are not sensitive to this particular choice, for $\Delta V_{\rm LOS,\,max} \gtrsim 100$ km s^{-1}.

4.3 Results and discussion

Using the procedure described in the previous section, we match HI column density systems that have $N_{\rm HI} > 3 \times 10^{16}\,{\rm cm}^{-2}$ to the galaxies with non-zero SFRs in our simulations (i.e., $\approx 2 \times 10^6$ strong HI absorbers and more than 10000 galaxies for every projection). In the following we use this matching to study the relative spatial distribution of galaxies and HI absorbers around them.

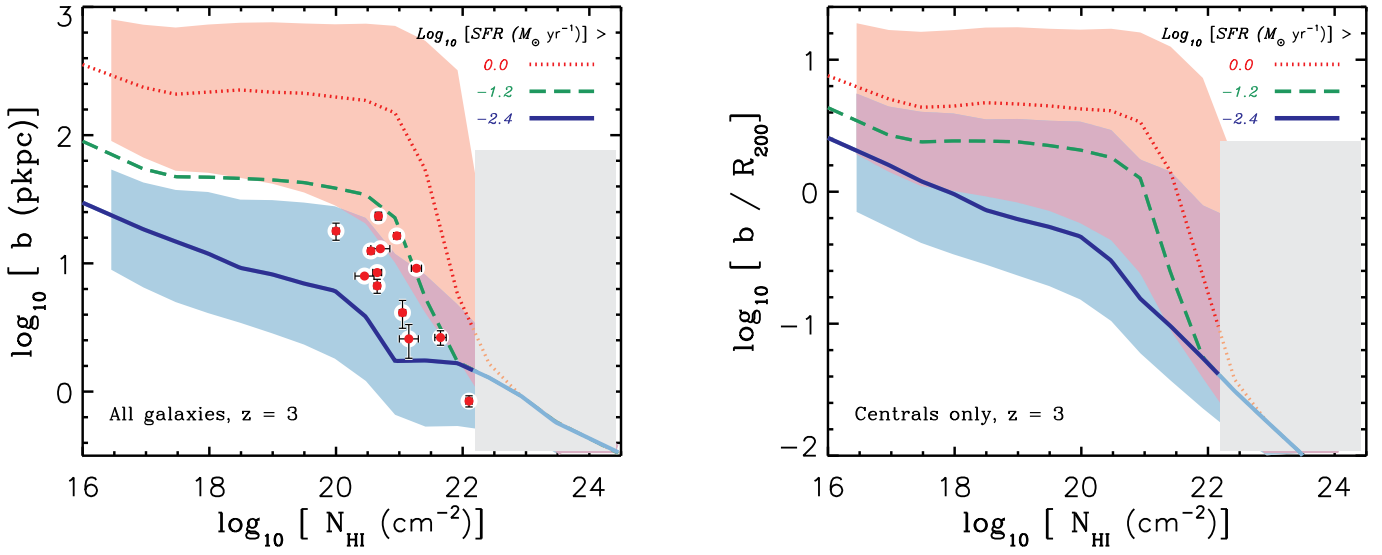

Figure 4.3: Left: Predicted median impact parameters vs. $N_{\rm HI}$ for different SFR thresholds at $z = 3$. Right: Median impact parameters normalized to the virial radii (R_{200}) as a function of $N_{\rm HI}$. Since satellite galaxies reside in the halo of central galaxies, they do not have defined virial radius. Therefore, the result in the right panel is based on matching HI absorbers to central galaxies only. In both panels, the red dotted, green dashed and blue solid curves show the SFR thresholds of 1, 0.06 and 0.004 $M_\odot$ yr^{-1}, respectively. The shaded areas around the blue solid and red dotted curves show the $15\% - 85\%$ percentiles. Red data points in the left panel show a compilation of DLAs with observed galaxy counterparts described in Table 4.1. Because of a very efficient conversion of hydrogen atoms into molecules, absorbers with $N_{\rm HI} \gtrsim 10^{22}\,{\rm cm}^{-2}$ (indicated by the gray areas) are not expected to exist.

4.3.1 Spatial distribution of HI absorbers

After connecting absorbers and galaxies, one can measure the typical projected distances (i.e., impact parameters, b) separating them. The predicted distribution of impact parameters as a function of HI column density is shown in Figure 4.2 for our simulation at $z = 3$. Each cell in this figure shows the position of HI absorbers in the $b - N_{\rm HI}$ space. The color of each cell indicates the median stellar mass of galaxies that are associated with the absorbers in that cell (see the colorbar on the right-hand side). The distribution of cells only shows the range of the $b - N_{\rm HI}$ space that is spanned by the HI absorbers in our simulation at $z = 3$. To show how the impact parameter of absorbers is distributed at any given $N_{\rm HI}$, we plot the median impact parameter as a function of $N_{\rm HI}$ using the blue solid curve and the $15\% - 85\%$ percentiles using the blue dotted curves. This result shows that our simulation predicts a strong anti-correlation between the HI column density of absorbers and their impact parameters. While the weak Lyman Limit Systems (LLSs) with $N_{\rm HI} \approx 10^{17}\,{\rm cm}^{-2}$ have typical impact parameters $b \approx 30$ proper kpc, the impact parameter decreases with increasing the HI column density such that strong DLAs with $N_{\rm HI} > 10^{21}\,{\rm cm}^{-2}$ are typically within a few proper kpc from the center of their neighboring galaxies. The increase in the impact parameter of HI absorbers with decreasing HI column density is in agreement with observations (Moller & Warren, 1998; Christensen et al., 2007; Monier et al., 2009; Rao et al., 2011; Péroux et al., 2011; Krogager et al., 2012) and consistent with previous theoretical studies (Gardner et al., 2001; Pontzen et al., 2008).

Despite the strong anti-correlation between the impact parameter and $N_{\rm HI}$, there is a large scatter around the median impact parameter at any given HI column density, as dotted curves in Figure 4.2 show. Since galaxies actively exchange material with their surroundings through accretion and outflows, the HI distribution around them has a very complex geometry (see the top-left panel of Figure 4.1 for an example). This complexity is a major contributor to the scatter in the impact parameters. In addition, part of this scatter is due to the fact that, at any given $N_{\rm HI}$, there is a large number of host galaxies with different sizes that contribute to the total distribution of absorbers. This is also consistent with the color gradients in Figure 4.2: at any given impact parameter, the mass of galaxies that are linked to HI absorber is increasing with their $N_{\rm HI}$ and at any given HI column density (particularly for DLAs), the mass of galaxy counterparts is increasing with the impact parameter of the HI absorbers. We will discuss this further in §4.3.5.

To show the impact of H_2 formation on the distribution of HI absorbers, in Figure 4.2 we also show the regions where the HI gas is fully converted into molecules using gray cells (see §4.2.3 for the details of H_2 calculation). The median impact parameters and $15\% - 85\%$ percentiles for the fully atomic gas (i.e., no HI to H_2 conversion) are shown with the gray dashed and gray dotted curves, respectively. The comparison between the colored and gray areas (and

curves) in Figure 4.2 shows that H_2 formation only affects HI column densities $N_{HI} > 10^{22}\,cm^{-2}$. This is consistent with the sharp cut-off in the observed HI column density distribution at $N_{HI} > 10^{22}\,cm^{-2}$ as shown in Rahmati et al. (2013a) (see also Altay et al., 2011; Erkal et al., 2012). The formation of H_2 thus only drains the atomic gas at very high HI column densities and does not significantly affect the impact parameters of the HI absorbers with $N_{HI} < 10^{22}\,cm^{-2}$.

4.3.2 The effect of a finite detection threshold

As seen from the colors in Figure 4.2, our simulation predicts that most strong HI absorbers with $10^{17} < N_{HI} \lesssim 10^{21}\,cm^{-2}$ are closely associated with low-mass galaxies, with typical stellar masses of $M_\star \sim 10^8\ M_\odot$. The typical SFR for those galaxies is $\sim 10^{-1}\ M_\odot\ yr^{-1}$. On the other hand, the typically accessible sensitivity of observations only allows the detection of galaxies that have SFR $\gtrsim 1 - 10\ M_\odot\ yr^{-1}$ (at $z \approx 3$)[3]. Because of this relatively high detection threshold, most galaxy counterparts are not detectable and the chance of observing galaxies that host LLSs and DLAs is slim. In fact, this could be the main reason why observational surveys that are aiming to find galaxies close to DLAs, often result in non-detections (e.g., Foltz et al., 1986; Smith et al., 1989; Lowenthal et al., 1995; Bunker et al., 1999; Prochaska et al., 2002; Kulkarni et al., 2006; Rahmani et al., 2010; Bouché et al., 2012). Moreover, the finite and relatively low sensitivity of observational surveys might cause biases in the measured typical impact parameter of absorbers by mis-associating them to the closest detectable galaxy in their vicinity, instead of their real hosts that are likely to fall below the detection limit.

Galaxies in our simulation are resolved down to SFRs that are much lower than the typical limited detection threshold of observations. Therefore, we are able to analyze the impact of varying the detection limit on the impact parameter of strong HI absorbers. Figure 4.1 shows the distribution of the HI column densities and positions of galaxies in a simulated region of size 500 proper kpc around a randomly selected massive galaxy at $z = 3$. The top-left panel of Figure 4.1 shows the distribution of HI column density and galaxies that have SFR $> 10\ M_\odot\ yr^{-1}$. With this detection threshold, only the central galaxy (shown with the dark circle whose size is proportional to the virial radius of the central galaxy) and one of its satellites (shown with the small white circle) are detectable. Other panels in this figure show that as the SFR threshold for detecting galaxies decreases, more galaxies show up in the field, which decreases the typical impact parameter.

The impact of varying detection threshold on the impact parameter of strong HI absorbers (i.e., $N_{HI} \gtrsim 10^{17}\,cm^{-2}$) is shown more quantitatively in the left panel of Figure 4.3. Different curves show the median impact parameter as a

[3]For a Chabrier IMF, a star formation rate of $1\ M_\odot\ yr^{-1}$ corresponds to a Lyα luminosity of $7.3 \times 10^{41}\ erg\,s^{-1}$, which translates into an observed flux of $\approx 3 \times 10^{-17}\ erg\,s^{-1}\,cm^{-2}$ and $\approx 1 \times 10^{-17}\ erg\,s^{-1}\,cm^{-2}$ at redshifts $z = 2$ and $z = 3$, respectively.

Table 4.1: A compilation of the confirmed DLA-galaxy pairs from the literature. The columns from left to right show respectively: the ID of each DLA, their redshifts, the HI column densities, the impact parameters in arc seconds, the impact parameters in proper kpc, the star formation rates of the host galaxies and finally the references from which these values are extracted. We note that the star formation rate estimates are often based on Lyα emissions, which provides a lower limit for the SFR since this is difficult to account for the dust reddening and incomplete flux measurements. The SFR estimates that are corrected for the dust reddening are indicated with bold number.

ID	z_{DLA}	$\log N_{HI}$ (cm^{-2})	b (arcsec)	b_p (pkpc)	SFR ($M_\odot\ yr^{-1}$)	$M_\star$ ($10^9\ M_\odot$)	Reference
Q2206-1958	1.92	20.65	0.99	8.44	3	-	[1]
Q0151+048A	1.93	20.45	0.93	7.93	71	-	[3] & [4]
PKS 0458-02	2.04	21.65	0.31	2.63	6	-	[2]
Q1135-0010	2.21	22.10	0.10	0.84	25	-	[6]
Q0338-0005	2.22	21.05	0.49	4.12	-	-	[5]
Q2243-60	2.33	20.67	2.28	23.37	**36**	-	[7]
Q2222-0946	2.35	20.65	0.8	6.67	**13**	2	[8] & [9]
Q0918+1636	2.58	20.96	2.0	16.38	**27**	12.6	[10] & [5] & [15]
Q0139-0824	2.67	20.70	1.6	13.01	-	-	[11]
J073149+285449	2.69	20.55	1.54	12.50	**12**	-	[12]
PKS 0528-250	2.81	21.27	1.14	9.15	17	-	[1]
2233.9+1318	3.15	20.00	2.3	17.91	20	-	[13]
Q0953+47	3.40	21.15	0.34	2.58	-	-	[14]

[1]- Möller et al. (2002); [2]- Möller et al. (2004); [3]- Moller & Warren (1998); [4]- Fynbo et al. (1999); [5]- Krogager et al. (2012); [6]- Noterdaeme et al. (2012); [7]- Bouché et al. (2012); [8]- Fynbo et al. (2010); [9]- Krogager et al. (2013); [10]- Fynbo et al. (2011); [11]- Wolfe et al. (2008); [12]- Fumagalli et al. (2010); [13]- Djorgovski et al. (1996); [14]- Prochaska et al. (2003); [15]- Fynbo et al. (2013)

function of $N_{\rm HI}$ assuming different SFR detection thresholds. The blue solid curve assumes a SFR detection threshold identical to that of Figure 4.2, where all galaxies with SFR $> 4 \times 10^{-3}$ M$_\odot$ yr^{-1} are considered as galaxy counterparts. The green dashed and red dotted curves, which respectively indicate SFR thresholds of $> 6 \times 10^{-2}$ M$_\odot$ yr^{-1} and > 1 M$_\odot$ yr^{-1}, show the impact of increasing the SFR threshold on the resulting impact parameter distribution. The shaded areas around the blue solid and red dotted curves show the $15\% - 85\%$ percentiles, and the overlap region between the two shaded areas is shown in purple. The gray area at $N_{\rm HI} > 10^{22}\,{\rm cm}^{-2}$ shows the region affected by the formation of H_2. The comparison between the three curves in the left panel of Figure 4.3 shows that the impact parameters increase with the detection threshold. Almost all absorbers that are associated to galaxies using the highest SFR threshold (i.e., red dotted curve) find fainter galaxies close the them (i.e., blue solid curve). Moreover, the anti-correlation between the impact parameter and $N_{\rm HI}$ is sensitive to the detection threshold. As the green dashed and red dotted curves show, for detection threshold SFR $\gtrsim 10^{-1}$ M$_\odot$ yr^{-1} the strong anti-correlation between the impact parameter and $N_{\rm HI}$ becomes insignificant at $N_{\rm HI} < 10^{20}\,{\rm cm}^{-2}$. The main reason for this is that most galaxy counterparts that are detectable with relatively low sensitivities (i.e., high detection thresholds) are not physically connected to the strong HI absorbers. As a result, the probability distribution of the impact parameters for those systems is controlled by the average projected distribution of the detectable galaxies. As the red dotted curve in the left panel of Figure 4.3 shows, our simulation predicts that with a detection threshold of SFR > 1 M$_\odot$ yr^{-1} the typical impact parameters between strong HI absorbers and their nearest galaxies vary from several tens of kpc to a few hundred kpc. This result is in excellent agreement with the measured impact parameters between DLAs and galaxies in observational surveys that used similar detection thresholds at $z \approx 2 - 3$ (not shown in this plot but see e.g., Teplitz et al., 1998; Mannucci et al., 1998). It is also worth noting that the anti-correlation between b and $N_{\rm HI}$ remains in place for absorbers with highe HI column densities (i.e., $N_{\rm HI} > 10^{21}\,{\rm cm}^{-2}$), even for a relatively high detection threshold like SFR > 1 M$_\odot$ yr^{-1}. However, the impact parameters of those systems are increasingly over-estimated as the galaxy detection threshold is increased.

We show the observed impact parameters of a compilation of confirmed DLA-galaxy pairs using red symbols with error-bars in the left panel of Figure 4.3 (see Table 4.1 for more details). Given the large scatter in the $b - N_{\rm HI}$ relation, there is a broad agreement between the observations and simulations. We note, however, that the observed impact parameters are generally above the blue solid curve. Before interpreting this systematic difference, it is important to keep in mind that there are many more non-detections of close galaxy-DLA pairs in the literature compared to those that are detected (e.g., Foltz et al., 1986; Smith et al., 1989; Lowenthal et al., 1995; Bunker et al., 1999; Prochaska et al., 2002; Kulkarni et al., 2006; Rahmani et al., 2010; Bouché et al., 2012). Therefore,

the existing observational sample is most likely unrepresentative and cannot be used for statistical analysis. Moreover, relatively high SFR detection thresholds used in the observations strongly bias the measured impact parameters towards larger values. In other words, it not easy to rule out the presence of fainter and therefore non-detected galaxies at smaller impact parameters, for DLAs that have observed bright galaxies around them. Finally, we note that SFRs of the observed galaxies that are associated with DLAs are typically $\gtrsim 10\ \mathrm{M}_{\odot}\ \mathrm{yr}^{-1}$ (see Table 4.1), which implies that those galaxies are massive and therefore large systems. As we show in §4.3.5, the observed impact parameters of those DLA-galaxy pairs are consistent with the distribution of impact parameters around similar galaxies in our simulation.

4.3.3 Distribution of HI absorbers relative to halos

It is useful to compare the impact parameters that connect absorbers to their neighboring galaxies with the size of their halos. We therefore define the normalized impact parameter of absorbers as the ratio of the impact parameter and the virial radius (R_{200}) of the host galaxies. Since the virial radius is well defined only for central galaxies, we only consider those objects when associating absorbers to galaxies. As shown by the blue solid curve in the right panel of Figure 4.3, strong HI absorbers tend to be located closer to the center of haloes as their N_{HI} increases. Most LLSs are found within one virial radius of their host galaxies and the majority of DLAs is found within a few tenths of viral radius from their associated central galaxies. There is, however, a large scatter in the normalized impact parameters at given N_{HI} which is shown by the shaded area around blue solid and red dotted curves in the right panel of Figure 4.3. As implied by this figure, a non-negligible fraction of DLAs are expected at impact parameters comparable to, or even larger than the virial radius of their associated central galaxies. This is, in part, due to neglecting satellite galaxies in matching absorbers and galaxies (because they do not have a well defined virial radius) which effectively associates the strong HI absorbers that are near satellites to their closest central galaxies[4]. For the same reason, the median impact parameter at a given N_{HI} increases by leaving the satellite galaxies out of the analysis (not shown). Moreover, the complex and highly structured distribution of strong HI absorbers which often extends to distances beyond the virial radius of central galaxies (see Figure 4.1) also contributes to the large scatter around the median normalized impact parameter at a given N_{HI}.

There is also an anti-correlation between the normalized impact parameter of absorbers and their HI column density, which is steeper than the relation we see between the absolute impact parameters and N_{HI} (shown in the right panel of Figure 4.3). This difference is most pronounced at $N_{\mathrm{HI}} \gtrsim 10^{21}\,\mathrm{cm}^{-2}$ where

[4]Note that satellite galaxies (and hence their associated absorbers) are not necessarily within the virial radius of their host galaxies, due to the non-spherical distribution of FoF structures. This is shown in Figure 4.1.

the impact parameter flattens with increasing $N_{\rm HI}$ but the normalized impact parameters still decrease steeply with $N_{\rm HI}$. This trend can be explained by the contribution of very massive (and hence large) galaxies becoming increasingly more dominant at very high HI column densities (see §4.3.5). Therefore, there are two competing trends: on the one hand, systems with higher HI column densities tend to be closer to the center of their neighboring (host) haloes. On the other hand, the size of the haloes that are linked to those absorbers becomes larger with increasing column density. These two effects neutralize each other at very high HI column densities, forming a less steep $b - N_{\rm HI}$ relation at $N_{\rm HI} > 10^{21}\,{\rm cm}^{-2}$ compared to that of the lower column densities. By normalizing the impact parameter of absorbers to the virial radii of their associated haloes, the latter effect (i.e., the increasing size of the dominant haloes with increasing $N_{\rm HI}$) is canceled out.

The trends we discussed earlier for the effect of varying the SFR threshold on the impact parameter of absorbers hold qualitatively for the normalized impact parameters as well. As the green dashed and red dotted curves in the right panel of Figure 4.3 show, increasing the SFR threshold for galaxies that are considered in matching absorbers to galaxies, results in larger normalized impact parameters. Despite the qualitatively similar trends, the differences between the normalized impact parameters for different SFR thresholds are smaller than the differences between absolute impact parameters. At $10^{17} < N_{\rm HI} < 10^{21}\,{\rm cm}^{-2}$, both the green dashed and red dotted curves in the right panel of Figure 4.3 are nearly flat. This implies that absorbers at those HI column densities are several virial radii away from their closest central galaxies (with SFR $> 6 \times 10^{-2}\,{\rm M}_\odot\,{\rm yr}^{-1}$ and SFR $> 1\,{\rm M}_\odot\,{\rm yr}^{-1}$, respectively). This supports our earlier statement that the association between bulk of strong HI absorbers and galaxies that have relatively high SFRs is not physical (particularly at $N_{\rm HI} < 10^{21}\,{\rm cm}^{-2}$), if fainter galaxies be omitted.

4.3.4 Resolution limit in simulations

The simulated HI distribution and its connection to galaxies are both sensitive to the high resolution adopted in our simulations. The HI column density distribution function is converged for LLSs and most DLAs at the resolution we use in this work (Rahmati et al., 2013a), but this does not rule out that their distribution relative to galaxies is also resolution independent. The relative position of galaxies that are identified as bound structures in a cosmological simulation is mainly determined by the distribution of cosmological overdensities, and therefore, is not expected to be highly resolution dependent. However, by increasing the resolution, more and more low-mass galaxies start to be identified which might affect the typical distances between strong HI absorbers and galaxies. Because of this, it is important to adopt a lower limit for galaxies that are taken into account in our analysis to make sure our findings are not driven primarily by the resolution we use. At the same time, this lower limit should be such that it

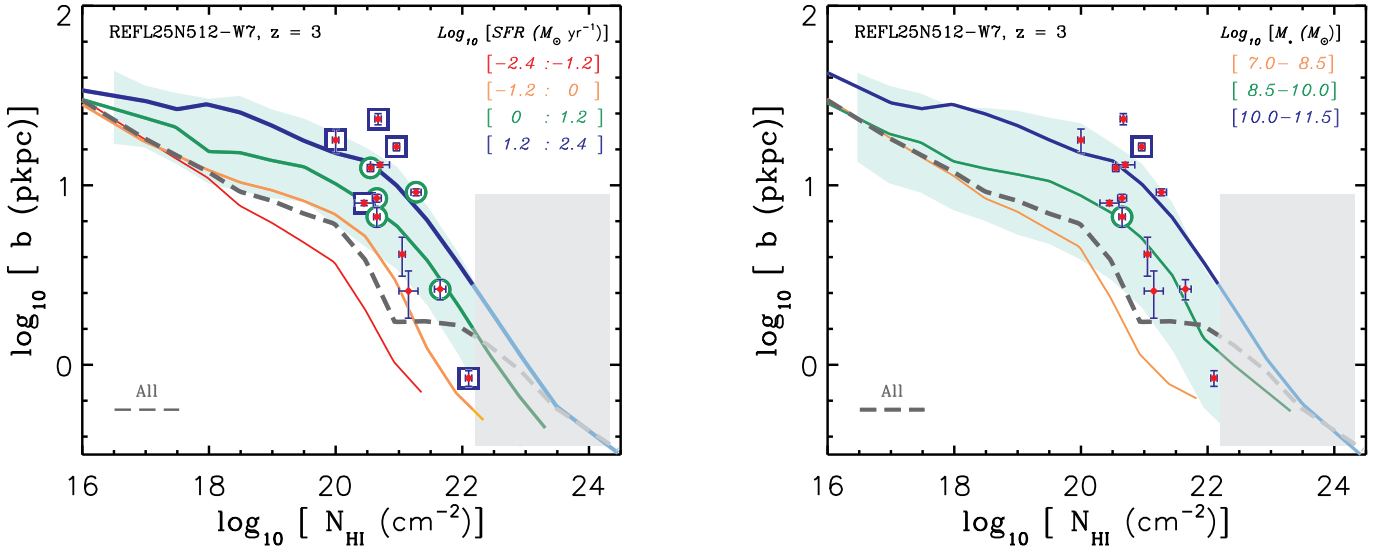

Figure 4.4: The predicted median impact parameters for subsets of HI absorbers associated with galaxies in different star formation rate bins (left) and stellar mass bins (right) as a function of $N_{\rm HI}$, at $z = 3$. The shaded area around the solid green curves in the left (right) panel shows the $15\% - 85\%$ percentile in the distribution of absorbers that are linked to galaxies with $1 < {\rm SFR} < 16$ ${\rm M}_\odot$ ${\rm yr}^{-1}$ ($2 \times 10^8 < {\rm M}_\star < 10^{10}$ ${\rm M}_\odot$). The dashed curves in both panels show the median impact parameter of all absorbers as a function of $N_{\rm HI}$. The data points show the observed impact parameters for the confirmed DLA-galaxy pairs. The colored circles and squares around the data points show the SFR/mass bin to which they belong. Note that the squares and circles, which show the two bins with the highest values (of SFR or mass) respectively, are in agreement with our results. Because of a very efficient conversion of hydrogen atoms into molecules, absorbers with $N_{\rm HI} \gtrsim 10^{22}\,{\rm cm}^{-2}$ (indicated with the gray areas) are not expected to exist.

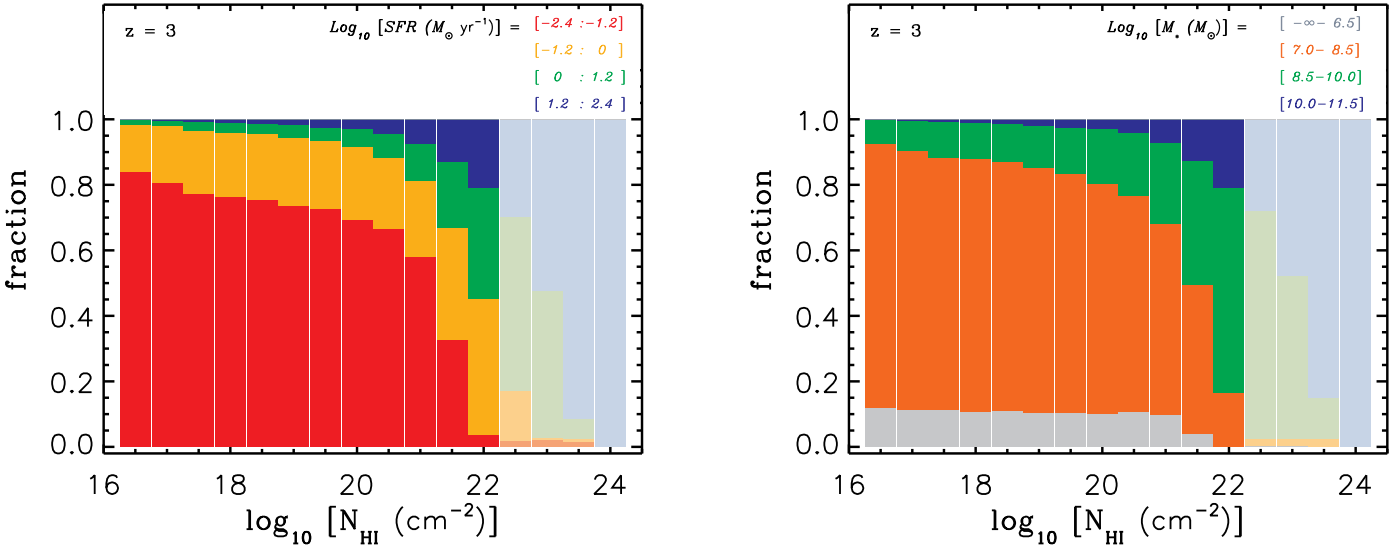

Figure 4.5: The top panels show the predicted median impact parameters for subsets of HI absorbers associated with galaxies in different star formation rate bins (left) and stellar mass bins (right) as a function of $N_{\rm HI}$, at $z = 3$. The fraction of absorbers that are associated with galaxies in different star formation rate bins (left) and stellar mass bins (right) to the total number of strong HI absorbers as a function of $N_{\rm HI}$, at $z = 3$. The star formation rate bins and stellar mass bins are identical to that of Figure 4.4. Because of a very efficient conversion of hydrogen atoms into molecules, absorbers with $N_{\rm HI} \gtrsim 10^{22}\,{\rm cm}^{-2}$ (indicated with the gray areas) are not expected to exist.

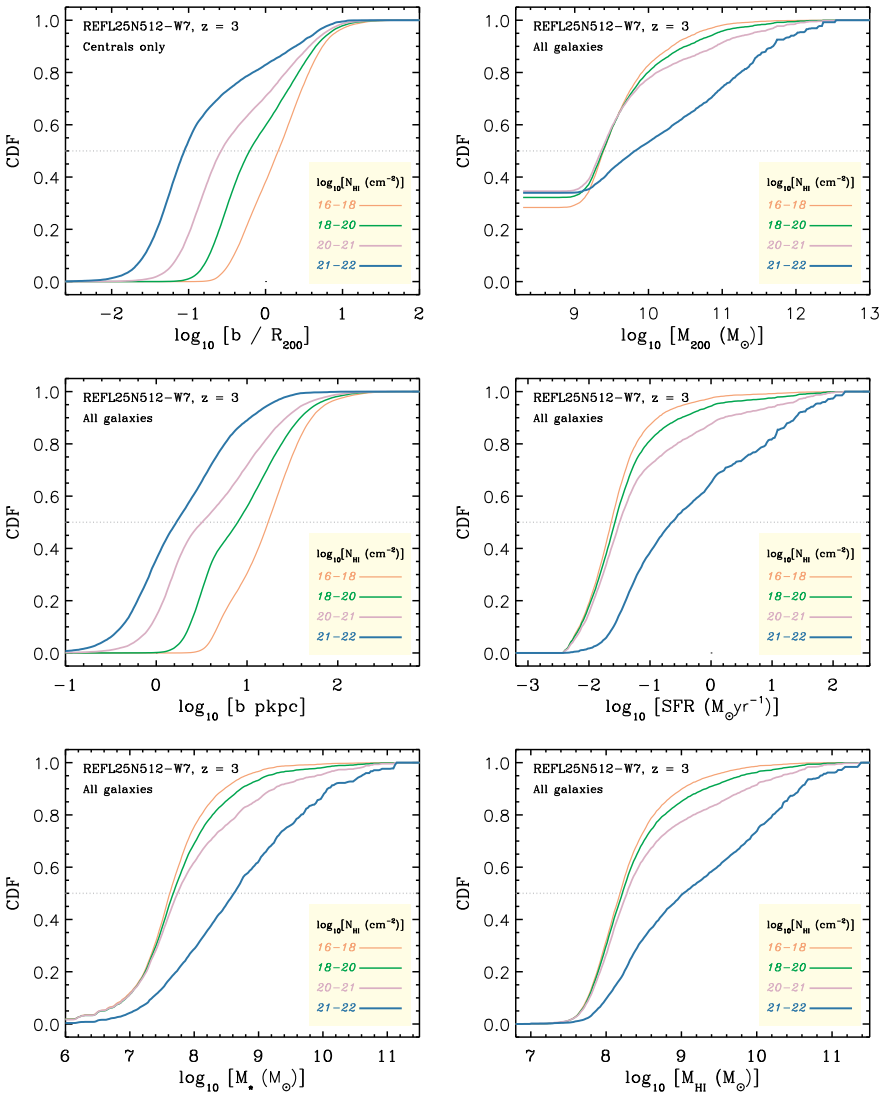

Figure 4.6: The top-left panel shows the predicted cumulative distribution of normalized impact parameter of absorbers in different $N_{\rm HI}$ bins at $z = 3$, where only central galaxies are taken into account. The other five panels show the cumulative distribution of different properties where all galaxies (with SFR > 0.004 $\rm M_\odot$ $\rm yr^{-1}$) are taken into account. The top-right panel shows the cumulative distribution of halo masses and other panels from middle-left to bottom right respectively show the cumulative distribution of impact parameters, star formation rates (of associated galaxies), stellar masses and HI masses. Except for strong DLAs with $10^{21} < N_{\rm HI} < 10^{22}\,\rm cm^{-2}$, bulk of HI absorbers reside close to galaxies with similar properties (i.e., virial mass, stellar mass, gas mass and star formation rate). However, their typical (normalized) impact parameters are significantly different.

allows for the inclusion of as many galaxies as possible, without throwing away useful information.

We use SFR as a quantity to limit galaxies that are considered in our analysis, which is the same quantity we used in the previous section as a proxy for the observational detectability of galaxies. The advantage of using the SFR of galaxies to identify them is the better sampling it provides compared to other reasonable quantities such as the stellar mass (see the discussion in §4.2 in Rahmati et al., 2013b). We adopt SFR $> 4 \times 10^{-3}$ $M_\odot$ yr^{-1} as a threshold and include only galaxies that have SFRs above this threshold in the analysis. Most bound (sub) structures (i.e., $\approx$ 98%) that are found in our simulation have SFRs above this threshold.

It is reasonable to expect that an increase in the number density of galaxies would decrease the typical distance between absorbers and galaxies. This would mean that the cumulative number density of galaxies in the simulation is the key factor that changes the typical distances between strong HI absorbers and galaxies. We show in Appendix D that this is indeed the case, and the $b - N_{\rm HI}$ relation is converged with resolution if the cumulative number density of galaxies that are considered is the same. Unlike the position of galaxies, the lowest SFRs in simulations are sensitive to the resolution and tend to decrease with increasing the resolution for a fixed number density. As a result, we do not expect that our result to remain the same if the resolution of our simulation increases for the same adopted SFR threshold (i.e., SFR $> 4 \times 10^{-3}$ $M_\odot$ yr^{-1}). Instead, we expect our results to be insensitive to the resolution if the adopted threshold be such that the cumulative number density of galaxies be identical to the corresponding value in our simulation. We note that the number density of galaxies in our simulation that satisfy the adopted criterion is 0.5 galaxy per comoving Mpc^3 (i.e., equivalent to 31.5 galaxy per proper Mpc^3).

Nonetheless, we found that the $b - N_{\rm HI}$ relation for various SFR threshold is converged if SFR $> 6 \times 10^{-2}$ $M_\odot$ yr^{-1}. In other words, the total number density of galaxies with SFR $> 6 \times 10^{-2}$ $M_\odot$ yr^{-1} is not expected to change by increasing the resolution of our simulation. This is another way of stating that for galaxies that satisfy the above mentioned condition, the SFR is converged with respect to the resolution.

4.3.5 Correlations between absorbers and various properties of their associated galaxies

The gas content of galaxies is correlated with their other properties like stellar mass, size and star formation rate. This implies also correlations between the abundance and distribution of strong HI absorbers and properties of galaxies that are associated with them. For instance, as mentioned earlier, at a given impact parameter, the typical stellar mass of host galaxies increases with increasing HI column density. Similarly, at a fixed $N_{\rm HI}$, the typical host stellar mass increases with increasing the impact parameter (see Figure 4.2). Moreover,

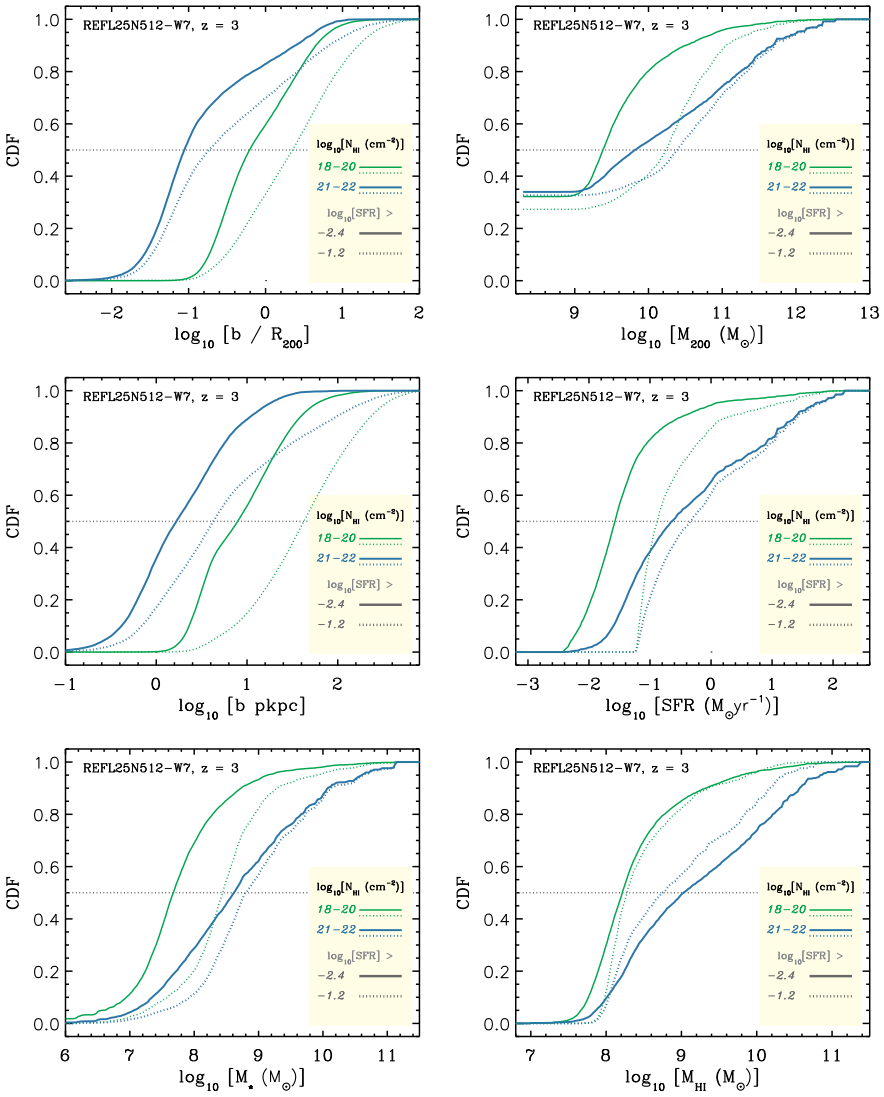

Figure 4.7: The impact of SFR threshold on the cumulative distributions that are shown in Figure 4.6. In each panel the blue and green solid curves show the cumulative distribution of LLSs (i.e., $10^{18} < N_{\rm HI} < 10^{20}\,{\rm cm}^{-2}$) and strong DLAs (i.e., $10^{21} < N_{\rm HI} < 10^{22}\,{\rm cm}^{-2}$), respectively. While the solid curves, that are identical to those shown in Figure 4.6, indicate the SFR threshold of 4×10^{-3} $M_\odot$ yr^{-1}, the dotted curves show the result obtained by imposing a SFR threshold of 6×10^{-2} $M_\odot$ yr^{-1}. Panels from top-left to bottom-right show the cumulative distribution of normalized impact parameters, halo masses, impact parameters, SFRs, stellar masses and HI masses, respectively. The detection of galaxies with SFRs as low as SFR $\sim 10^{-2} M_\odot$ yr^{-1} future instruments such as MUSE.

the LLSs are closely linked to low-mass galaxies with typical stellar masses of $M_\star \sim 10^8\ M_\odot$, in contrast to strong DLAs which tend to be associated with more massive galaxies.

The above mentioned trends are shown more clearly in Figure 4.4. These plots illustrate the $b - N_{\rm HI}$ relation for subsets of HI absorbers that are linked to galaxies with different star formation rates and stellar masses (respectively shown in the left and right panels). The colored curves in the left (right) panel which show different bins of SFR (stellar mass), increasing by going from red to blue indicate that the impact parameter of absorbers is increasing with the SFR and mass of their associated galaxies. Note that there is a relatively large scatter around the median impact parameter of each subset of absorbers, as shown with the shaded area around the green curves. This result explains the color gradient that we see in Figure 4.2 and is in good agreement with observations that are shown with the red diamonds in Figure 4.4. The observed DLA-galaxy pairs with reliable SFR or mass estimate are shown with the green circles and blue cubes, respectively for the highest two bins. Note that the curves are close to the observational results (see also Table 4.1).

We also show the median impact parameters of all absorbers using the dashed curve in Figure 4.4. The comparison between this curve and colored solid curves confirms our earlier conclusion (based on the color gradient in Figure 4.2) that the impact parameter of absorbers with higher HI column densities indicates that they are more likely to be around more massive galaxies with higher SFRs. To illustrate this more clearly, we show also the fraction of HI absorbers, at any given column density, that are associated to galaxies with a particular property. The result of this analysis for different SFR and mass bins is shown in the left and right panels of Figure 4.5, respectively. These results clearly show that the fraction of HI absorbers associated with massive galaxies (which also have high SFRs), is decreasing rapidly with decreasing their HI column densities. Also, it is clear from the left panel of Figure 4.5 that most absorbers with $N_{\rm HI} < 10^{21}\,{\rm cm}^{-2}$ are closely linked to galaxies with SFR $< 6 \times 10^{-2}\ M_\odot\ {\rm yr}^{-1}$, or equivalently $M_\star < 10^8\ M_\odot$. While there are only about $20 - 30\%$ of those systems associated with more massive galaxies with higher SFRs (i.e., $M_\star > 10^8\ M_\odot$ and SFR $> 6 \times 10^{-2}\ M_\odot\ {\rm yr}^{-1}$), the same galaxies are associated with a large fraction of strong DLAs with $N_{\rm HI} > 10^{21}\,{\rm cm}^{-2}$. We reiterate that these results are not changing by increasing the resolution and only the lowest SFRs are expected to be reduced at higher resolutions (shown with the red regions in the left panel of Figure 4.5), which is not expected to affect the fraction of absorbers that are associated with well-resolve galaxies that have $M_\star > 10^8\ M_\odot$ and SFR $> 6 \times 10^{-2}\ M_\odot\ {\rm yr}^{-1}$.

We note that our results are in agreement with Tescari et al. (2009) and van de Voort et al. (2012a). Namely, those authors found that the larges fraction of HI absorbers with $N_{\rm HI} < 10^{21}\,{\rm cm}^{-2}$ are in very low-mass haloes with $M_{200} < 10^{10}\ M_\odot$ (i.e., $M_\star \lesssim 10^8\ M_\odot$; see the orange regions in the right panel of Figure 4.5). Moreover, van de Voort et al. (2012a) found that at higher HI

column densities, the contribution of more massive haloes with $M_{200} > 10^{11}\,M_\odot$ (i.e., $M_\star \gtrsim 10^9\,M_\odot$; see the green regions in the right panel of Figure 4.5) rapidly increases.

To investigate further the distribution of galaxy properties for absorbers with different HI column densities, we split absorbers in different $N_{\rm HI}$ bins. Then, for galaxies that are associated with the absorbers in that bin, we construct the cumulative distribution of different properties. The result of this exercise is shown in Figure 4.6. The top-left panel of this figure shows the cumulative distribution of normalized impact parameters, where in contrary to the other five panels, only central galaxies are taken into account (because satellite galaxies do not have well defined viral radius required for the normalization). The top-right panel shows the cumulative distribution of halo masses. Since satellite galaxies are in the halo of their centrals, they do not have well defined halo masses and as the starting point of the cumulative distribution functions in the top-right panel indicates, $\approx 30\%$ of absorbers are associated to satellite galaxies. The other four panels from middle-left to the bottom-right respectively show the cumulative distribution of impact parameters, star formation rates, stellar masses and HI masses.

Comparing the cumulative distribution of (normalized) impact parameters with other four panels in this figure indicates that the HI column density of absorbers is mostly sensitive to their projected distance from their associated galaxies (i.e., the impact parameter), compared to other properties, such as the stellar mass, SFR, HI mass or halo size of their associated galaxies. As shown in the top-left panel, while more than 50% of strong DLAs are within $R \lesssim 0.1\,R_{200}$, most weak LLSs with $N_{\rm HI} \lesssim 10^{18}\,{\rm cm}^{-2}$ are likely to be beyond the virial radius of their host galaxies.

Another result that is shown in Figure 4.6 is that the fraction of absorbers that are linked to more massive galaxies is increasing by increasing their HI column densities. The same trend is visible for the SFRs, halo masses and HI masses. However, as the distinction between strong DLAs $N_{\rm HI} > 10^{22}\,{\rm cm}^{-2}$ (blue solid curves) and lower HI column densities show, galaxies associated with strong DLAs have distinct distributions in SFR, HI mass, stellar mass and halo mass. For instance, the median SFR of galaxies that are associated with other strong HI absorbers is ≈ 10 times lower than the same quantity for galaxies associated with strong DLAs. Similar trend also holds for HI masses, stellar and halo masses. Comparing the median properties (i.e., at CDF = 0.5) shows that only $\approx 10\%$ of strong DLAs are linked to galaxies that are typically associated with weaker HI absorbers. This suggests that strong DLAs with $N_{\rm HI} > 10^{21}\,{\rm cm}^{-2}$ are preferentially linked to the most massive galaxies while other HI absorbers are distributed among more abundant galaxies with lower masses.

The middle-right panel of Figure 4.6 is particularly useful to understand the difficulty of finding the observed galaxy counterparts of strong HI absorbers. The distribution of SFRs shown in this plot can be used to predict the chance by which a galaxy counterpart can be detected above a cer-

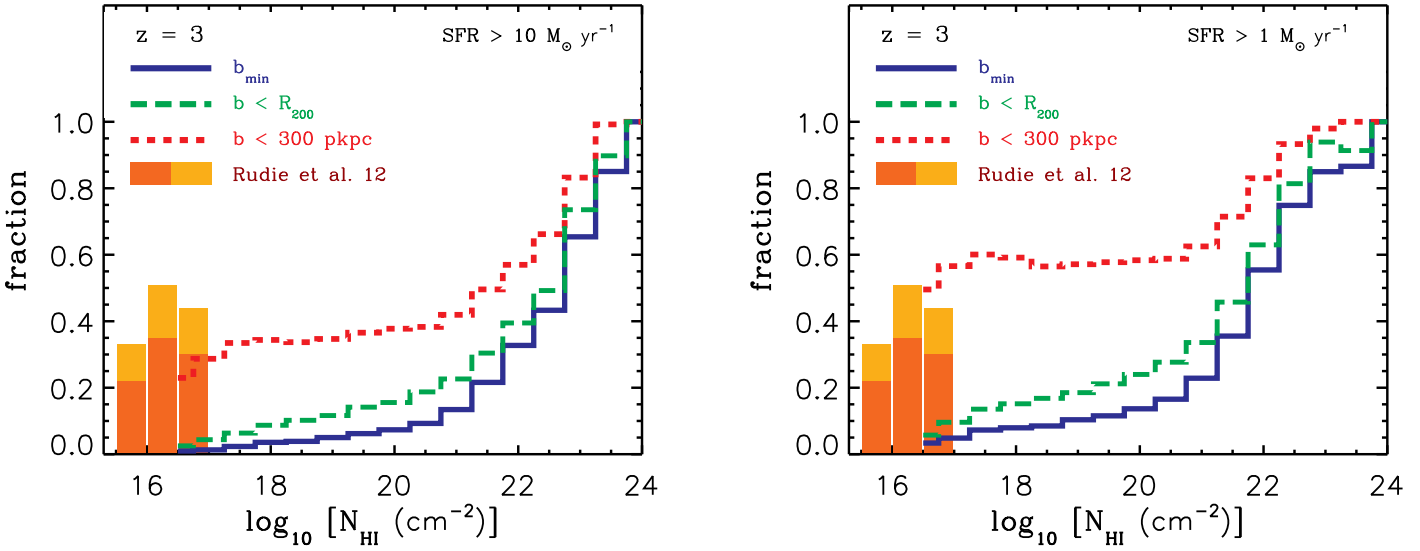

Figure 4.8: Fraction of strong HI systems that are associated to galaxies with SFR $> 10\ M_\odot\ yr^{-1}$ (left) and SFR $> 1\ M_\odot\ yr^{-1}$ (right) in our simulation at $z = 3$. Curves with different line styles and colors show results that are obtained using different methods for associating HI absorbers to galaxies that are above the imposed SFR threshold: blue solid curves are obtained by associating HI absorbers to their closest galaxies, where all the simulated galaxies are taken into account (see Figures 4.4 and 4.5); green long-dashed curves show the fraction of absorbers that reside within the virial radius of the selected galaxies, and finally, red short-dashed curves show the fraction of absorbers that are within 300 proper kpc from the selected galaxies. For comparison, the colored bars show Rudie et al. (2012) findings for fraction of absorbers within 300 proper kpc from galaxies with similar SFRs that are shown here. The dark orange part of the bar is the observed fractions and the light orange part shows the correction for missing galaxies in their spectroscopic sample. Predicted fractions are in excellent agreement with observations and show a large fraction of strong HI absorbers are less than 300 proper kpc away from galaxies with SFR $> 1\ M_\odot\ yr^{-1}$. However, most of those systems are far beyond the virial radius of those galaxies and are associated with less massive objects with lower SFRs which are below the detection limit of observations.

tain detection threshold. For instance, if the detection threshold is equivalent to SFR $= 1$ $M_\odot$ yr^{-1}, the chance of detecting the galaxy counterpart of a strong DLA with $N_{HI} \approx 10^{22}$ cm^{-2} is 3 in 10. In other words, using this detection threshold results in non-detection rate of 70%. For weak DLAs (i.e., $10^{20} < N_{HI} < 10^{21}$ cm^{-2}) the non-detection rate would be even higher at 90%. This explains the large number of non-detections in observational studies (e.g., Foltz et al., 1986; Smith et al., 1989; Lowenthal et al., 1995; Bunker et al., 1999; Prochaska et al., 2002; Kulkarni et al., 2006; Rahmani et al., 2010; Bouché et al., 2012).

As mentioned before, with the typical detection threshold of observations at $z \sim 3$, only galaxies that have SFR $> 1 - 10$ $M_\odot$ yr^{-1} can be identified. On the other hand, galaxies are resolved to much lower SFRs in our simulation and it is not straightforward to compare current observations and the results shown in Figure 4.6. With the advent of future surveys like MUSE (Bacon et al., 2010), the accessible Lyα detection thresholds would be pushed to lower SFRs which allows the identification of galaxies with SFRs as low as SFR $\sim 10^{-2} - 10^{-1}$ $M_\odot$ yr^{-1} at $z \sim 3$. These deep observations can be used to identify faint galaxies associated with strong HI absorbers and analyze the cumulative distribution of their properties. However, the results would still depend on the accessible detection threshold and might be different from what is shown in Figure 4.6. To address this issue, in Figure 4.7 we compare the cumulative distributions for two different detection thresholds. Panels in this figure are identical to Figure 4.6, but show the result for only two HI column density bins. The green and blue curves, respectively, represent the HI absorbers (or galaxies associated to them) that have $10^{18} < N_{HI} < 10^{20}$ cm^{-2} (i.e., LLSs) and $10^{21} < N_{HI} < 10^{22}$ cm^{-2} (i.e., strong DLAs). The solid curves show our fiducial detection threshold of SFR $> 4 \times 10^{-3}$ $M_\odot$ yr^{-1}, which has been used in Figure 4.6. The dotted curves indicate a higher detection threshold of SFR $> 6 \times 10^{-2}$ $M_\odot$ yr^{-1}, which is comparable to what would be accessible using deep MUSE observations. As the difference between the solid and dotted curves show, the predicted/observed distributions are sensitive to the detection threshold. In other words, the bias introduced by the limited detection threshold should be taken into account for interpreting/modeling the observed distributions.

4.3.6 Are most strong HI absorbers at $z \sim 3$ around Lyman-Break galaxies?

In this work, we adopted the absorber-centered point of view in which each absorber is associated with its closest galaxy. This approach is particularly efficient to establish a relationship between rare absorbers and their galaxy counterparts. As we showed in the previous sections, our simulation predicts that for most HI absorbers, nearby galaxies are low-mass objects (i.e., $M_\star \approx 10^8$ $M_\odot$) with low SFRs. We showed that this prediction is consistent with the impact parameters of observed DLA-galaxy counterparts, and more importantly, the high incident

rate of finding no detectable galaxy close to DLAs. There is, however, an alternative approach which is to search for HI absorbers around galaxies (i.e., galaxy-centered point of view) (e.g., Steidel et al., 1995; Adelberger et al., 2003, 2005; Hennawi & Prochaska, 2007; Steidel et al., 2010; Rakic et al., 2012; Rudie et al., 2012; Prochaska et al., 2013). While the two approaches are complementary, their results might seem inconclusive. For instance, using galaxy-centered approach, Rudie et al. (2012) found that at $z \approx 2-3$ most Lyman Limit absorbers (i.e., $N_{\rm HI} \approx 10^{17}\,{\rm cm}^{-2}$) are within an impact parameter of 300 proper kpc, and within 300 km/s LOS velocity difference with respect to rest-frame UV-selected star-forming galaxies (see their Figure 30). Given that the typical galaxy mass in their sample is $M_\star \approx 10^{10}\,M_\odot$, one might conclude that their result is in conflict with our finding that most strong HI absorbers (i.e., LLSs and DLAs) are closely associated with galaxies with $M_\star \approx 10^{8}\,M_\odot$.

To understand the source of this discrepancy, first we note that the 300 proper kpc transverse distance and the 300 km/s LOS velocity difference that are adopted by Rudie et al. (2012), are, respectively, 300 and 50 per cent larger. than the virial radius and the circular velocity which is expected for the host haloes of their galaxies (Rakic et al., 2012, 2013). In other words, the region Rudie et al. (2012) define as the "circumgalactic medium" lies well beyond the virial radius of the haloes that are thought to host their galaxies.

In addition, it is important to note that galaxies are strongly clustered and low-mass galaxies prefer to live around more massive galaxies. This implies that most strong HI absorbers are also likely to be found close to massive galaxies, while they actually belong to their low-mass parent galaxies. As a result, searching for HI absorbers within a reasonably large radius around massive, and hence easily observable galaxies, recovers a large fraction of the existing strong HI absorbers. This effect can be seen in the bottom-right panel of Figure 4.1, which shows the HI column density distribution around a galaxy with $M_\star = 10^{10}\,M_\odot$. While the maximum projected distance between the galaxy and HI absorbers that are shown around it is less than 300 proper kpc, most of those absorbers have low-mass galaxies very close to them.

The aforementioned arguments are illustrated more quantitatively in Figure 4.8, which shows the fraction of strong HI absorbers in our simulation at $z = 3$ that are in the vicinity of galaxies with star formation rates comparable to that of LBG galaxies at the similar redshifts[5]. The red dashed curve in the left (right) panels shows the fraction of HI absorbers that are within the impact parameter of 300 proper kpc from galaxies with SFR $> 10\,M_\odot\,{\rm yr}^{-1}$ (SFR $> 1\,M_\odot\,{\rm yr}^{-1}$), as a function of $N_{\rm HI}$. This result is in excellent agreement with the fractions Rudie et al. (2012) found (shown with the colored bars) and predicts that the fraction of absorbers that are within 300 proper kpc from these galaxies is roughly constant for $10^{16} < N_{\rm HI} < 10^{21}\,{\rm cm}^{-2}$. However, only a small fraction of LLSs that are within the impact parameter of 300 proper kpc from those

[5]The star formation rate of galaxies used in Rudie et al. (2012) is between several to a few hundreds of $M_\odot\,{\rm yr}^{-1}$.

galaxies are also within their virial radii (shown with the green long-dashed curves). As the blue solid curves show, even smaller fractions of LLSs remain associated to those galaxies if we account for galaxies with lower SFRs, which are typically closer to LLSs but are too faint to be observed.

4.4 Summary and conclusions

We have used cosmological simulations that have been post-processed using accurate radiative transfer corrections that account for photoionization by the UVB and recombination radiation to investigate the relation between strong HI absorbers (i.e., LLSs and DLAs) and galaxies at $z = 3$. The simulation we used for our study has been shown to reproduce the observed HI column density distribution function accurately (Altay et al., 2011; Rahmati et al., 2013a). After identifying sight-lines with high HI column densities (i.e., $N_{\rm HI} > 3 \times 10^{16}$ cm^{-2}) and calculating the line-of-sight velocity of absorbers, we used a procedure similar to that used in observational studies to associate absorbers with nearby galaxies. Namely, we associated each strong HI absorber to the galaxy which has the shortest transverse distance to the absorber and a line-of-sight velocity difference within ± 300 km s^{-1}. Associating all strong HI absorbers in the periodic simulation box to galaxies close to them allowed us to predict statistical trends between the strength of the absorbers and the distance to, and properties of, the galaxies.

Among the various dependencies we studied in this work, we found that the anti-correlation between the HI column density of absorbers and the transverse distance that connects them to their neighbouring galaxies (i.e., the impact parameter) to be the strongest. While LLSs have impact parameters $\gtrsim 10$ proper kpc, DLAs are typically within a few proper kpc from the nearest galaxies. Relative to the virial radius of the halo that hosts the nearest central galaxy, LLSs have typical impact parameters $\gtrsim 1$ R$_{200}$, while DLAs are typically ~ 10 times closer to the center of the nearest halo (i.e. $\lesssim 0.1$ R$_{200}$). The predicted strong anti-correlation between the impact parameter of strong HI absorbers and their HI column densities is in agreement with observations and previous work. We also found a relatively large scatter around the median impact parameter of absorbers, at a given $N_{\rm HI}$, due to the complex geometry of gas distribution around galaxies, and also the variation in the size and gas content of the galaxies that are producing the absorbers.

We predict that most strong HI absorbers are closely associated with relatively low-mass galaxies, M$_\star \lesssim 10^8$ M$_\odot$, but that the fraction of strong HI absorbers that are linked to more massive galaxies increases rapidly with the HI column density. This correlation between column density and galaxy mass is particularly pronounced for strong DLAs, $N_{\rm HI} > 10^{21}$ cm^{-2}, the majority of which are associated with galaxies with M$_\star \gtrsim 10^9$ M$_\odot$. We analyzed different properties of galaxies that are linked to strong HI absorbers with different

HI column densities and found similar trends as we found for stellar mass: most strong HI absorbers are closely associated with galaxies that have relatively low halo masses, low SFRs and low HI masses, but strong DLAs (i.e., $N_{\rm HI} > 10^{21}\,{\rm cm}^{-2}$) are typically linked to more massive galaxies with significantly higher halo masses, SFRs and HI masses.

By analyzing subsets of strong HI absorbers for which the associated galaxies have specific properties, we found that observationally confirmed DLA-galaxy pairs that have measured mass or SFR, have impact parameters that are in good agreement with our predictions. We stress, however, that the majority of DLAs are predicted to be more closely associated with galaxies that are at smaller impact parameters, but are too faint to be detected with current surveys. Hence, the masses and impact parameters of the observed galaxy counterparts of DLAs are both biased high. This is consistent with the large number of non-detections in observational campaigns that searched for galaxies close to DLAs (e.g., Foltz et al., 1986; Smith et al., 1989; Lowenthal et al., 1995; Bunker et al., 1999; Prochaska et al., 2002; Kulkarni et al., 2006; Rahmani et al., 2010; Bouché et al., 2012).

In order to facilitate the comparison between cosmological simulations and observations, we provides statistics on DLA-galaxy pairs for a few different SFR thresholds. However, a proper comparison requires observational studies aiming to find galaxies close to DLAs, to report their detection limit, the maximum allowed velocity separation, and either impact parameter of the nearest detectable galaxy or, in the case of non-detection (which must always be reported), the maximum impact parameter that has been searched. For the few studies that report such information (e.g., Teplitz et al., 1998; Mannucci et al., 1998), we found a good agreement with our simulation.

Interestingly, some recent observational studies indicate that strong HI absorbers at $z \sim 2-3$ are associated with surprisingly massive galaxies. In particular, Rudie et al. (2012) studied the distribution of HI absorbers around a sample of rest-frame UV-selected Lyman-Break galaxies (LBG) with typical masses of $M_\star > 10^{10}\,M_\odot$ at $z \sim 2-3$ and found that nearly half of absorbers with $10^{16} < N_{\rm HI} < 10^{17}\,{\rm cm}^{-2}$ reside within a line-of-sight velocity difference of 300 km s^{-1} and a transverse seperation of 300 proper kpc from a LBG, a region they labelled the circumgalactic medium. This result appears to contradict our finding that most strong HI absorbers are associated with galaxies with $M_\star \lesssim 10^8\,M_\odot$. We demonstrated, however, that even though the absorbers are physically most closely associated with low-mass galaxies, these galaxies cluster strongly around galaxies as massive as LBGs, which is sufficient to reproduce the observations of Rudie et al. (2012). Moreover, we noted that the clustering does not even need to be that strong: Since 300 proper kpc and 300 km s^{-1} exceed the virial radius and the circular velocity of the haloes thought to host LBGs, by more than 300 and 50 per cent, respectively (e.g., Rakic et al., 2012, 2013), nearly all the volume of the "circumgalactic medium" lies beyond the virial radius if we employ the definition of Rudie et al. (2012).

Future deep observational surveys using new instruments (e.g., MUSE; Bacon et al., 2010) will improve this situation by detecting fainter galaxies. However, missing faint galaxies is a generic feature for any survey that has a finite detection limit and that takes an absorber-centered point of view. The incompleteness problem can be overcome by taking a galaxy-centered point of view, but this approach is inefficient for rare absorbers such as the interesting strong HI systems we studies here. Moreover, while galaxy-centered surveys can measure the statistical distribution of absorbers (such as their covering factor), we still need to avoid interpreting the selected galaxy as the counterpart to any absorber that is detected. Absorber-centered surveys will probably remain the most efficient way to build up large numbers of galaxy-DLA paris. Even with modest detection limits, such surveys provide highly valuable constraints on the relation between absorbers and galaxies, provided all non-detections are reported and that the detection limits and the maximum possible impact parameter are clearly specified.

Acknowledgments

We would like to thank A. Pawlik, X. Prochaska, M. Raičevič and M. Shirazi for useful discussion, reading an earlier version of this chapter and providing us with comments that improved the text. The simulations presented here were run on the Cosmology Machine at the Institute for Computational Cosmology in Durham (which is part of the DiRAC Facility jointly funded by STFC, the Large Facilities Capital Fund of BIS, and Durham University) as part of the Virgo Consortium research programme. This work was sponsored with financial support from the Netherlands Organization for Scientific Research (NWO), also through a VIDI grant and an NWO open competition grant. We also benefited from funding from NOVA, from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement 278594-GasAroundGalaxies and from the Marie Curie Training Network CosmoComp (PITN-GA-2009-238356).

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., et al. 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, ApJL, 737, L37
Altay, G., Theuns, T., Schaye, J., Booth, C. M., & Dalla Vecchia, C. 2013, arXiv:1307.6879
Bacon, R., Accardo, M., Adjali, L., et al. 2010, Proceedings of the SPIE, 7735
Blitz, L., & Rosolowsky, E. 2006, ApJ, 650, 933

Booth, C. M., & Schaye, J. 2009, MNRAS, 398, 53
Bouché, N., Murphy, M. T., Péroux, C., et al. 2012, MNRAS, 419, 2
Bunker, A. J., Warren, S. J., Clements, D. L., Williger, G. M., & Hewett, P. C. 1999, MNRAS, 309, 875
Cen, R. 2012, ApJ, 748, 121
Chabrier, G. 2003, PASP, 115, 763
Christensen, L., Wisotzki, L., Roth, M. M., et al. 2007, A&A, 468, 587
Dalla Vecchia, C., & Schaye, J. 2008, MNRAS, 387, 1431
Di Matteo, T., Springel, V., & Hernquist, L. 2005, Nature, 433, 604
Djorgovski, S. G., Pahre, M. A., Bechtold, J., & Elston, R. 1996, Nature, 382, 234
Dolag, K., Borgani, S., Murante, G., & Springel, V. 2009, MNRAS, 399, 497
Erkal, D., Gnedin, N. Y., & Kravtsov, A. V. 2012, ApJ, 761, 54
Foltz, C. B., Chaffee, F. H., Jr., & Weymann, R. J. 1986, AJ, 92, 247
Font-Ribera, A., Miralda-Escudé, J., Arnau, E., et al. 2012, JCAP, 11, 59
Fumagalli, M., O'Meara, J. M., Prochaska, J. X., & Kanekar, N. 2010, MNRAS, 408, 362
Fumagalli, M., Prochaska, J. X., Kasen, D., et al. 2011, MNRAS, 418, 1796
Fynbo, J. U., Møller, P., & Warren, S. J. 1999, MNRAS, 305, 849
Fynbo, J. P. U., Laursen, P., Ledoux, C., et al. 2010, MNRAS, 408, 2128
Fynbo, J. P. U., Ledoux, C., Noterdaeme, P., et al. 2011, MNRAS, 413, 2481
Fynbo, J. P. U., Geier, S., Christensen, L., et al. 2013, arXiv:1306.2940
Gardner, J. P., Katz, N., Hernquist, L., & Weinberg, D. H. 1997, ApJ, 484, 31
Gardner, J. P., Katz, N., Hernquist, L., & Weinberg, D. H. 2001, ApJ, 559, 131
Haardt F., Madau P., 2001, in Clusters of Galaxies and the High Redshift Universe Observed in X-rays, Neumann D. M., Tran J. T. V., eds.
Haas, M. R., Schaye, J., Booth, C. M., et al. 2012, arXiv:1211.1021
Haehnelt, M. G., Steinmetz, M., & Rauch, M. 1998, ApJ, 495, 647
Hennawi, J. F., & Prochaska, J. X. 2007, ApJ, 655, 735
Krogager, J.-K., Fynbo, J. P. U., Møller, P., et al. 2012, MNRAS, 424, L1
Krogager, J.-K., Fynbo, J. P. U., Ledoux, C., et al. 2013, arXiv:1304.4231
Komatsu, E., et al. 2011, ApJS, 192, 18
Krumholz, M. R., Ellison, S. L., Prochaska, J. X., & Tumlinson, J. 2009, ApJL, 701, L12
Kulkarni, V. P., Woodgate, B. E., York, D. G., et al. 2006, ApJ, 636, 30
Lara-López, M. A., Cepa, J., Bongiovanni, A., et al. 2010, A&A, 521, L53
Lowenthal, J. D., Hogan, C. J., Green, R. F., et al. 1995, ApJ, 451, 484
Mannucci, F., Thompson, D., Beckwith, S. V. W., & Williger, G. M. 1998, ApJL, 501, L11
Mannucci, F., Cresci, G., Maiolino, R., Marconi, A., & Gnerucci, A. 2010, MNRAS, 408, 2115
McDonald, P., & Miralda-Escudé, J. 1999, ApJ, 519, 486
Moller, P., & Warren, S. J. 1998, MNRAS, 299, 661
Möller, P., Warren, S. J., Fall, S. M., Fynbo, J. U., & Jakobsen, P. 2002, ApJ, 574, 51

Möller, P., Fynbo, J. P. U., & Fall, S. M. 2004, A&A, 422, L33
Monier, E. M., Turnshek, D. A., & Rao, S. 2009, MNRAS, 397, 943
Nagamine, K., Springel, V., & Hernquist, L. 2004, MNRAS, 348, 421
Noterdaeme, P., Laursen, P., Petitjean, P., et al. 2012, A&A, 540, A63
Pawlik, A. H., & Schaye, J. 2008, MNRAS, 389, 651
Pawlik, A. H., & Schaye, J. 2011, MNRAS, 412, 1943
Péroux, C., Bouché, N., Kulkarni, V. P., York, D. G., & Vladilo, G. 2011, MNRAS, 410, 2237
Pontzen, A., Governato, F., Pettini, M., et al. 2008, MNRAS, 390, 1349
Prochaska, J. X., & Wolfe, A. M. 1997, ApJ, 487, 73
Prochaska, J. X., & Wolfe, A. M. 1998, ApJ, 507, 113
Prochaska, J. X., Gawiser, E., Wolfe, A. M., et al. 2002, AJ, 123, 2206
Prochaska, J. X., Gawiser, E., Wolfe, A. M., Cooke, J., & Gelino, D. 2003, ApJS, 147, 227
Prochaska, J. X., & Wolfe, A. M. 2009, ApJ, 696, 1543
Prochaska, J. X., Hennawi, J. F., & Simcoe, R. A. 2013, ApJL, 762, L19
Rahmani, H., Srianand, R., Noterdaeme, P., & Petitjean, P. 2010, MNRAS, 409, L59
Rahmati, A., Pawlik, A. H., Raičevič, M., & Schaye, J. 2013a, MNRAS, 430, 2427
Rahmati, A., Schaye, J., Pawlik, A. H., & Raičevič, M. 2013b, MNRAS, 431, 2261
Rakic, O., Schaye, J., Steidel, C. C., et al. 2013, arXiv:1306.1563
Rakic, O., Schaye, J., Steidel, C. C., & Rudie, G. C. 2012, ApJ, 751, 94
Rao, S. M., Belfort-Mihalyi, M., Turnshek, D. A., et al. 2011, MNRAS, 416, 1215
Razoumov, A. O., Norman, M. L., Prochaska, J. X., & Wolfe, A. M. 2006, ApJ, 645, 55
Rudie, G. C., Steidel, C. C., Trainor, R. F., et al. 2012, ApJ, 750, 67
Schaye, J. 2001a, ApJL, 559, L1
Schaye, J. 2001c, ApJL, 562, L95
Schaye, J., & Dalla Vecchia, C. 2008, MNRAS, 383, 1210
Schaye, J., Dalla Vecchia, C., Booth, C. M., et al. 2010, MNRAS, 402, 1536
Smith, H. E., Cohen, R. D., Burns, J. E., Moore, D. J., & Uchida, B. A. 1989, ApJ, 347, 87
Springel, V. 2005, MNRAS, 364, 1105
Steidel, C. C., Pettini, M., & Hamilton, D. 1995, AJ, 110, 2519
Steidel, C. C., Erb, D. K., Shapley, A. E., et al. 2010, ApJ, 717, 289
Tescari, E., Viel, M., Tornatore, L., & Borgani, S. 2009, MNRAS, 397, 411
Teplitz, H. I., Malkan, M., & McLean, I. S. 1998, ApJ, 506, 519
van de Voort, F., Schaye, J., Altay, G., & Theuns, T. 2012a, MNRAS, 421, 2809
van de Voort, F., & Schaye, J. 2012b, MNRAS, 423, 2991
Wiersma, R. P. C., Schaye, J., Theuns, T., Dalla Vecchia, C., & Tornatore, L. 2009a, MNRAS, 399, 574
Wiersma, R. P. C., Schaye, J., & Smith, B. D. 2009b, MNRAS, 393, 99
Wolfe, A. M., Turnshek, D. A., Smith, H. E., & Cohen, R. D. 1986, ApJS, 61, 249
Wolfe, A. M., Gawiser, E., & Prochaska, J. X. 2005, ARA&A, 43, 861

Wolfe, A. M., Prochaska, J. X., Jorgenson, R. A., & Rafelski, M. 2008, ApJ, 681, 881
Yanny, B., York, D. G., & Williams, T. B. 1990, ApJ, 351, 377

Appendix A: Choosing the maximum allowed LOS Velocity difference

As discussed in §4.2.4, for associating the HI absorbers to galaxies that are close to them we take into account their relative LOS velocities. This allows us to take out systems that appear to be close in projection but are apart by large physical distances. In analogy to observational studies, we calculate LOS velocities of HI absorbers and galaxies. If the difference between these two LOS velocities be larger than a minimum value, then we do not associate them as counterparts even though they are very close in projection. One might think that the above mentioned allowed minimum velocity difference between HI absorbers and galaxies, should be as small as possible to minimize the projection effect. In reality, however, the peculiar velocity of the HI gas around galaxies is not negligible and is controlled by complex processes like accretion and outflows. If the allowed LOS velocity differences are too small, peculiar velocity of HI absorbers around galaxies would be miss-interpreted as large LOS sight distances. As shown in Figure 4.9, our exercise show that the median impact parameter of absorbers as a function of $N_{\rm HI}$ is converged for maximum LOS velocity difference of $\Delta V_{\rm LOS,\ max} > 100$ km/s and the scatter around the median impact parameters is converged for $\Delta V_{\rm LOS,\ max} > 300$ km/s. Therefore, we adopt $\Delta V_{\rm LOS,\ max} = 300$ km/s which is also consistent with recent observational studies (e.g., Rakic et al., 2012; Rudie et al., 2012).

Appendix B: Impact of feedback

The evolution of gas and stars is determined by complex baryonic interactions that are often modeled in cosmological simulations by combining various physically motivated and empirical ingredients. In this context, different feedback mechanisms can change both the distribution of gas around galaxies with a given mass and the abundance of galaxies with different masses (e.g., van de Voort & Schaye, 2012b; Haas et al., 2012). As a result, the strength and details of various feedback mechanisms change the distribution of the HI absorbers (Altay et al., 2013) and may also affect the relative distribution of HI absorbers and galaxies. To quantify the impact of feedback on our results, we compare the relation we found in our reference model between the impact parameter of absorbers and their $N_{\rm HI}$ to the same relation in similar simulations that use different feedback prescriptions. Figure 4.10 shows this comparison between the

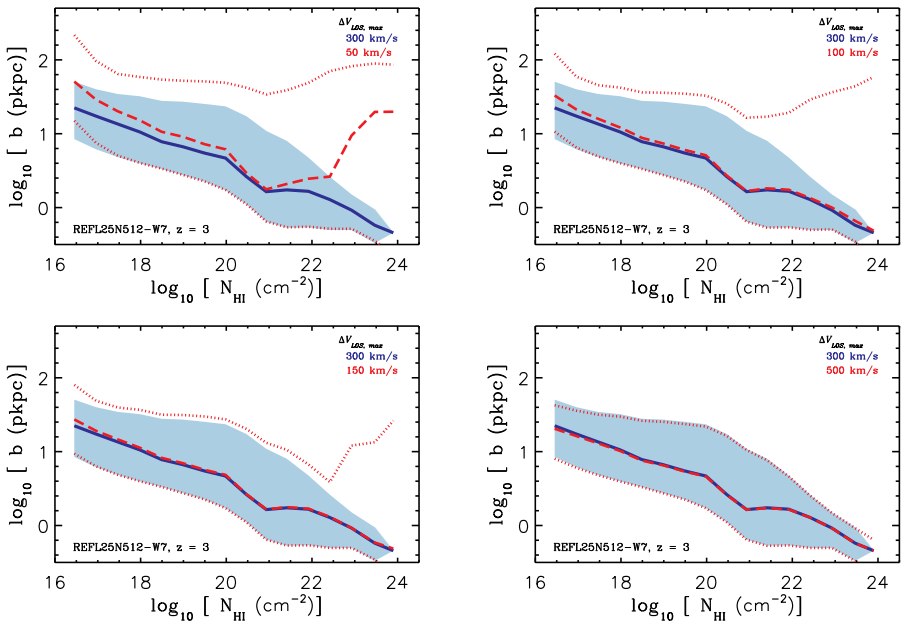

Figure 4.9: The impact of changing the maximum allowed LOS velocity difference between HI absorbers and galaxies to associate them with each other. The blue solid curve in all panels shows the median impact parameter of absorbers in our simulation at $z = 3$ where our fiducial value of $\Delta V_{\rm LOS,\,max} = 300$ km/s is adopted and the shaded area around the blue solid curve shows the $15\% - 85\%$ percentiles. In each panel, the same result but using a different $\Delta V_{\rm LOS,\,max}$ is shown by red dashed and dotted curves which, respectively, show the median and the $15\% - 85\%$ percentiles. The $\Delta V_{\rm LOS,\,max}$ that is compared with our fiducial choice of 300 km/s is varying from 50 km/s from the top-left panel to 500 km/s in the bottom-right panel. While the median impact parameters are converged for $\Delta V_{\rm LOS,\,max} > 100$ km/s, both the median impact parameters and scatter around it are converged for $\Delta V_{\rm LOS,\,max} > 300$ km/s.

reference simulation at $z = 3$ and two bracketing models that use AGN feedback and do not have SNe feedback and metal cooling. The solid green curve in the left panel of Figure 4.10 shows the reference model while the red dashed curve and blue dot-dashed curves respectively indicate the simulation with AGN and the simulation without SNe feedback and metal cooling (NOSN_NOZCOOL). The relation between the normalized impact parameter and $N_{\rm HI}$ is shown in the right panel of Figure 4.10, where only central galaxies are taken into account for the matching process.

The absorbers with $N_{\rm HI} \lesssim 10^{21}\,{\rm cm}^{-2}$ in the AGN simulation have typical impact parameters which are larger than in the reference model. At very high HI column densities however, the two models are similar except that there are fewer very high $N_{\rm HI}$ systems in the AGN simulation. The same trend is seen

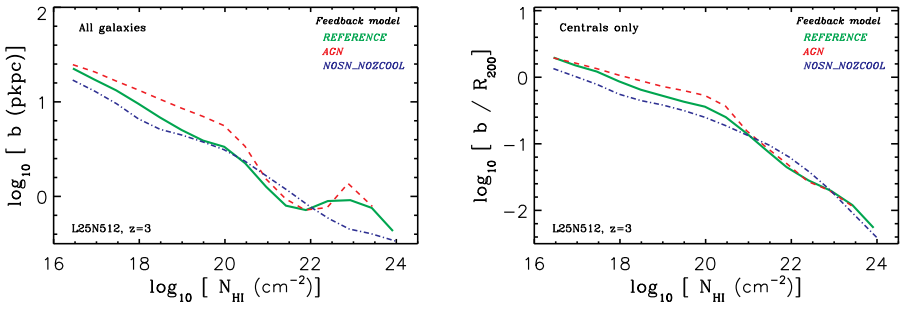

Figure 4.10: The (normalized) impact parameter as a function of $N_{\rm HI}$ at $z = 3$ for simulations with different feedback models is shown in the (right) left panel. The green solid, red dashed and blue dot-dashed curves respectively show the reference simulation, the impact of adding AGN feedback and the result of turning off SNe feedback and metal cooling. While the impact parameters are sensitive to the adopted feedback prescription, the differences are much smaller than the intrinsic scatter caused by the complex geometry of gas distribution around galaxies with a wide range of sizes.

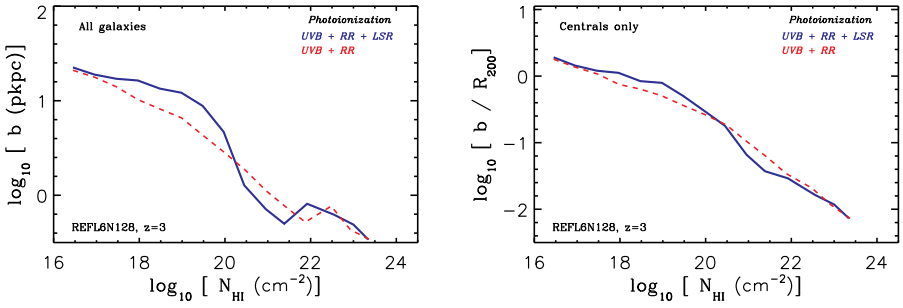

Figure 4.11: The (normalized) impact parameter as a function of $N_{\rm HI}$ at $z = 3$ for the *REFL06N128* simulation with different photoionization models is shown in the (right) left panel. The red dashed curve shows the result when only the UVB and recombination radiations (RR) are present while the blue solid curve indicates the result of including local stellar radiation (LSR). The data points in the left panel are observations. Photoionization from local stellar radiation reduces the impact parameter of DLAs by up to 50% and increases the typical impact parameter of Lyman Limit systems by similar amount.

in the normalized impact parameters and is consistent with previous studies implying that AGN feedback pushes the gas away from the center of haloes by preventing the rapid cooling of the accreted gas and moderates the formation of the most massive galaxies which are the main contributors to very high HI column densities (e.g., Di Matteo et al., 2005; Booth & Schaye, 2009; Schaye et al., 2010; van de Voort & Schaye, 2012b; Haas et al., 2012). We also found that the contribution of galaxies with different stellar masses and SFRs as a function of $N_{\rm HI}$ in the AGN simulation is very similar to the reference model (not shown). The only different between the two simulations in this context is that the number of strong HI absorbers in very massive galaxies (with $M_\star \gtrsim 10^{10}$ $M_\odot$; SFR $\gtrsim$ 1 $M_\odot$ yr^{-1}) becomes smaller while the contribution of less massive galaxies(with $M_\star \lesssim 10^{9}$ $M_\odot$; SFR $\lesssim 10^{-1}$ $M_\odot$ yr^{-1}) remains intact.

In contrast, turning off the SNe feedback and metal cooling reduces the typical impact parameters of HI absorbers at almost all HI column densities. This can be understood by noting that the absence of metal cooling reduces the amount of cold gas available for accretion into galaxies, forcing the dense gas to be found closer to the center of galaxies. Moreover, the lack of SNe feedback allows the gas to falls more regularly along well defined filaments without being redistributed by SNe driven outflows. The lack of SNe feedback also allows the formation of dens cores in the center of lower mass galaxies which are interrupted easier by SNe explosions. As a result, the contribution of smaller galaxies (with $M_\star \lesssim 10^{9}$ $M_\odot$; SFR < 1 $M_\odot$ yr^{-1}) to the abundance of high $N_{\rm HI}$ systems increases. This results in smaller impact parameter at high HI column densities compared to the reference simulation, while leave their normalized impact parameters nearly identical.

All the above mentioned differences in the impact parameters of strong HI absorbers due to variations in feedback are much smaller than the intrinsic scatter in the expected impact parameter at a given $N_{\rm HI}$ which is caused by the complex geometry of gas distribution around galaxies with different sizes. The strong anti-correlation between the impact parameter of absorbers and their HI column densities is present and remains similar despite large variations in feedback mechanisms. Therefore, we conclude that feedback variation has a minor impact on our main results, although observations agree better with models that have stronger feedbacks.

Appendix C: Impact of local stellar radiation

After the reionization of the Universe at $z \sim 6$, ionizing background radiation produced by the radiation coming from all sources in the Universe keeps hydrogen atoms mostly ionized. While on large scale, e.g., in the intergalactic medium, this radiation is uniform, it becomes highly non-uniform close to radiation sources. In particular, as was shown in Rahmati et al. (2013b), for absorbers with $N_{\rm HI} \gtrsim 10^{17}\,{\rm cm}^{-2}$ the radiation from local stellar radiation becomes import-

ant. Since we study the distribution of such strong HI absorbers, it is important to investigate the impact of local stellar radiation on the results of this work, where we neglect it.

The impact of local stellar radiation on the spatial distribution of strong HI absorbers for the *REFL06N128* simulation at $z = 3$ is shown in Figure 4.11. The blue solid curve shows the result of radiative transfer calculation by taking into account the photoionization of the UVB, recombination radiation and local stellar radiation as explained in Rahmati et al. (2013b), while the red dashed curve indicates the result of including only the photoionization of the UVB and recombination radiation in the same simulation. As can be seen by comparing the two curves in the left panel of Figure 4.11, local stellar radiation can change the impact parameters of strong HI absorbers by up to 50%. The median impact parameters of DLAs with $N_{\rm HI} \sim 10^{21}\,{\rm cm}^{-2}$ is reduced because their HI column densities decreases due to additional photoionization by local stellar radiation. For Lyman Limit systems on the other hand, the ionization from local stellar radiation mainly affect systems that are closer to the galaxies compared to absorbers that are at larger distances. This results in an increase in the median impact parameter of absorbers at a given $N_{\rm HI}$ by decreasing the HI column density of absorbers at shorter impact parameters. At lower HI column densities (i.e., $N_{\rm HI} \lesssim 10^{17}\,{\rm cm}^{-2}$) where the effect of local stellar radiation is negligible (Rahmati et al., 2013b), the impact parameters remain intact [6]. We conclude that local stellar radiation is not changing the results we present in this work.

Appendix D: Resolution tests

We use three different cosmological simulations that have identical box sizes, but different resolutions. These simulations are part of the OWLS project (Schaye et al., 2010) and have cosmological parameters that are consistent with WMAP year-3 values (i.e., $\{\Omega_{\rm m} = 0.238,\ \Omega_{\rm b} = 0.0418,\ \Omega_{\Lambda} = 0.762,\ \sigma_8 = 0.74,\ n_{\rm s} = 0.951,\ h = 0.73\}$), slightly different from our fiducial cosmology in this work. The simulation with the highest resolution (*REFL25N512*) has identical mass resolution and box size compared to the simulation we use in this work and the other two simulations have 8 times (*REFL25N256*) and 64 time (*REFL25N128*) lower resolutions (see Table 4.2 for more details).

As we showed in Rahmati et al. (2013a), the HI column density distribution function is converged for LLS and most DLAs at the resolution that we use in this work. The position of galaxies in cosmological simulations is determined by the distribution of overdensities and therefore, is not expected to be highly sensitive to the resolution. This is not true for the number of galaxies that are

[6] We note that using a small simulation box results in underproducing the strong HI absorbers due to missing very massive galaxies (Rahmati et al., 2013a). Given that the contribution of very massive galaxies to the total abundance of absorbers increases with increasing the $N_{\rm HI}$ (see Figure 4.4 and 4.5), missing them in the small simulation box allows the smaller galaxies to be the main DLA counterparts and hence decreases the typical impact parameters of strong HI absorbers.

Table 4.2: List of cosmological simulations used in this work. The detailed description of the model ingredients are discussed in Schaye et al. (2010). From left to right the columns show: simulation identifier; comoving box size; number of dark matter particles (there are equally many baryonic particles); initial baryonic particle mass; dark matter particle mass; comoving (Plummer-equivalent) gravitational softening; maximum physical softening; final redshift; remarks about the used model, cosmology and the use of explicit radiative transfer calculations instead of fitting function for the HI calculations (RT).

Simulation	L (h^{-1}Mpc)	N	$m_{\rm b}$ ($h^{-1}{\rm M}_\odot$)	$m_{\rm dm}$ ($h^{-1}{\rm M}_\odot$)	$\epsilon_{\rm com}$ (h^{-1}kpc)	$\epsilon_{\rm prop}$ (h^{-1}kpc)	$z_{\rm end}$	Model
REFL06N128	6.00	128^3	1.4×10^6	6.3×10^6	1.95	0.50	0	REF, WMAP7 cosmology, RT
REFL25N512-W7	25.00	512^3	1.4×10^6	6.3×10^6	1.95	0.50	2	REF, WMAP7 cosmology
REFL25N512	25.00	512^3	1.4×10^6	6.3×10^6	1.95	0.50	1	REF
AGN	25.00	512^3	1.4×10^6	6.3×10^6	1.95	0.50	2	with AGN
NOSN_NOZCOOL	25.00	512^3	1.4×10^6	6.3×10^6	1.95	0.50	2	w/o SN, w/o metal cooling
REFL25N256	25.00	256^3	1.1×10^7	5.1×10^7	3.91	1.00	2	REF
REFL25N128	25.00	128^3	8.7×10^7	4.1×10^8	7.81	2.00	0	REF

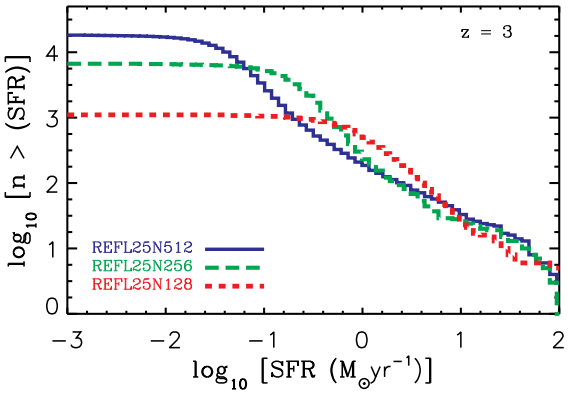

Figure 4.12: Cumulative number of galaxies that are resolved in simulations at $z = 3$ with different resolutions as a function of their SFR. Blue solid, green long-dashed and red dashed curves show the *REFL25N512*, *REFL25N256* and *REFL25N128* simulations, respectively.

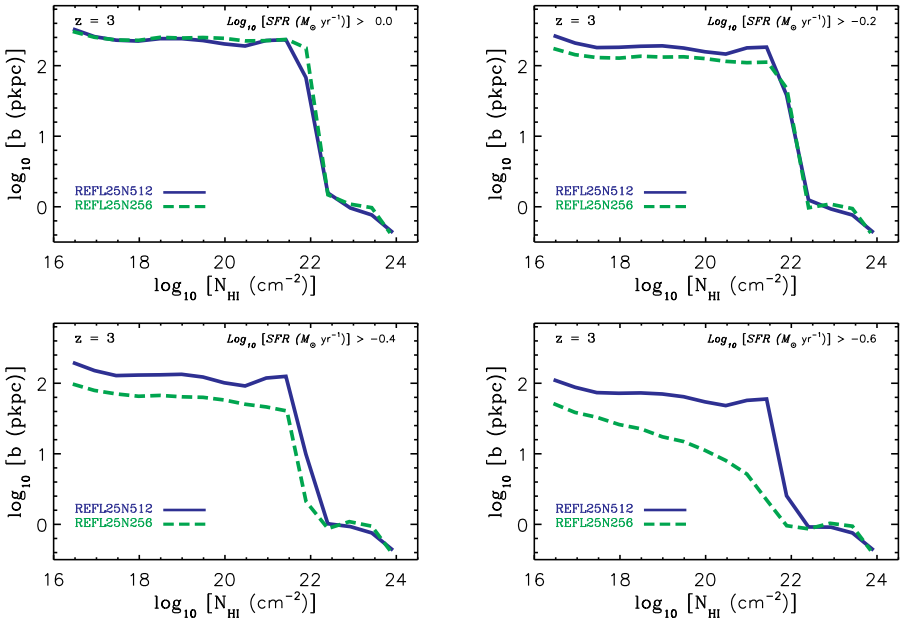

Figure 4.13: The resolution dependence of $b - N_{\rm HI}$ relation at different SFR thresholds. The blue solid and green dashed curves represent the *REFL25N512* and the *REFL25N256* simulations, respectively. Panels show the impact parameter of absorbers as a function $N_{\rm HI}$ by including only galaxies that have SFR above a certain threshold. The SFR thresholds from top-left to bottom-right are 1, 0.63, 0.4 and 0.2 $M_\odot$ yr^{-1}, respectively. The results are similar in the two simulations for SFR > 0.4 $M_\odot$ yr^{-1}.

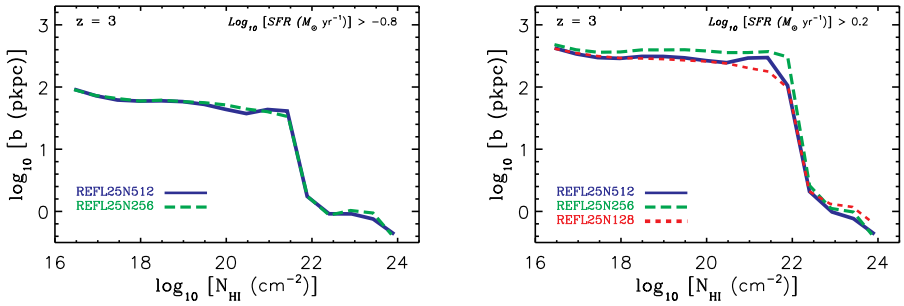

Figure 4.14: Impact parameter of HI absorbers as a function of their $N_{\rm HI}$ for different SFR thresholds. In the left panel, the blue solid curve shows the result for the *REFL25N512* simulation where only galaxies with SFR > 0.16 M$_\odot$ yr^{-1} are taken into account. The green long-dashed curve shows the result for the *REFL25N256* simulation where the SFR threshold is such that the cumulative number density of galaxies is matched to that of galaxies with SFR > 0.16 M$_\odot$ yr^{-1} in the *REFL25N512* simulation. In the right panel, the blue solid curve shows the $b - N_{\rm HI}$ relation for the *REFL25N512* simulation where only galaxies with SFR > 1.6 M$_\odot$ yr^{-1} are taken into account. The green long-dashed and red dashed curves respectively show the result in the *REFL25N256* and *REFL25N128* simulations where the SFR thresholds are chosen such that the total number density of galaxies that are taken into account is matched to that of galaxies with SFR > 1.6 M$_\odot$ yr^{-1} in the *REFL25N512* simulation. This shows that for a fixed total number density of galaxies the relation between impact parameters and the HI column density of absorbers is independent of the resolution.

resolved in simulations and as the resolution increases, the number of structures that are resolved in a simulation also increases. This is shown in Figure 4.12 which shows the cumulative distribution of number of objects that are identified as bound structures in our simulations[7]. Comparing the blue solid and green long-dashed curves in this figure shows that the number of galaxies that have SFR > 1 M$_\odot$ yr^{-1} is identical in the two simulations with highest resolutions. Together with the converged HI column density distribution function, this result implies that the relation between the impact parameter of HI absorbers and galaxies with SFR $\gtrsim 1$ M$_\odot$ yr^{-1} is also converged in the *REFL25N256* simulation. Indeed, as Figure 4.13 shows, this is the case for SFR > 0.4 M$_\odot$ yr^{-1}. This suggests that the $b - N_{\rm HI}$ relation is converged in the *REFL25N512* simulation for SFR $> 6 \times 10^{-2}$ M$_\odot$ yr^{-1}. By increasing the number density of galaxies in a simulation, one would expect a decrease in the average distance between HI absorbers and galaxies. This suggests that the distribution of HI absorbers relative to galaxies to be primarily sensitive to the cumulative number density of

[7]We note that contrary to the *REFL25N512* simulation in which cosmological parameters are based on WMAP year-3 values, our reference simulation in this work is based on WMAP year-7 cosmological parameters. As a result its cumulative distribution of galaxies flattens at lower SFR compared to what is shown by the blue solid curve in Figure 4.12.

all galaxies. As a result, one might expect to retrieve the same relation between impact parameters and $N_{\rm HI}$ of absorbers if the total number density of galaxies be the same in simulations with different resolution. As Figure 4.14 shows, this is indeed true in our simulations where in each panel we choose different SFR thresholds depending on the resolution to match the total number(density) of galaxies above the SFR threshold for all the simulations. This result suggests that if one keeps the total number density of galaxies we use in our study, increasing the resolution is not expected to change the $b - N_{\rm HI}$ relation and other conclusions we derived in this work. In this work we used the relatively low SFR threshold of SFR $> 4 \times 10^{-3}$ M$_\odot$ yr^{-1} to include as many bound (sub) structure as possible in our analysis. The total number density of galaxies that are selected in our reference simulation (i.e., *REFL25N512-W7*) by this criterium is 0.5 galaxy per comoving Mpc^3 (i.e., equivalent to 31.5 galaxy per proper Mpc^3). We note, however, that the SFR threshold that correspond to this total number density at higher resolutions is expected to be lower than 4×10^{-3} M$_\odot$ yr^{-1}. This can be seen in Figure 4.12 which shows that at a fixed cumulative number density, the SFR of galaxies in the *REFL25N256* simulation that have SFR < 1 M$_\odot$ yr^{-1} decreases by increasing the resolution.