Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

Mon. Not. R. Astron. Soc. **425,** 1640–1663 (2012) doi:10.1111/j.1365-2966.2012.21545.x

Absorption signatures of warm-hot gas at low redshift: broad H I Lyα absorbers

Thorsten Tepper-García,[1⋆] Philipp Richter,[1] Joop Schaye,[2] C. M. Booth,[2,3,4,5] Claudio Dalla Vecchia[2,6] and Tom Theuns[7,8]

[1]*Institut für Physik und Astronomy, Universität Potsdam, Karl-Liebknecht-Str. 24/25, 14476 Potsdam, Germany*
[2]*Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, the Netherlands*
[3]*Department of Astronomy & Astrophysics, The University of Chicago, Chicago, IL 60637, USA*
[4]*Kavli Institute for Cosmological Physics, The University of Chicago, Chicago, IL 60637, USA*
[5]*The Enrico Fermi Institute, The University of Chicago, Chicago, IL 60637, USA*
[6]*Max-Planck-Institut für Extraterrestrische Physik, Giessenbachstraße 1, 85748 Garching, Germany*
[7]*Institute for Computational Cosmology, Department of Physics, University of Durham, South Road, Durham DH1 3LE*
[8]*Department of Physics, University of Antwerp, Groenenborgerlaan 171, B-2020 Antwerpen, Belgium*

Accepted 2012 June 14. Received 2012 May 14; in original form 2012 January 26

ABSTRACT

We investigate the physical state of H I absorbing gas at low redshift ($z = 0.25$) using a subset of cosmological, hydrodynamic simulations from the OverWhelmingly Large Simulations project, focusing in particular on broad ($b_{\rm H\,I} \geq 40\,{\rm km\,s^{-1}}$) H I Ly$\alpha$ absorbers (BLAs), which are believed to originate in shock-heated gas in the warm-hot intergalactic medium (WHIM). Our fiducial model, which includes radiative cooling by heavy elements and feedback by supernovae and active galactic nuclei, predicts that by $z = 0.25$ nearly 60 per cent of the gas mass ends up at densities and temperatures characteristic of the WHIM and we find that half of this fraction is due to outflows. The standard H I observables (distribution of H I column densities $N_{\rm H\,I}$, distribution of Doppler parameters $b_{\rm H\,I}$, $b_{\rm H\,I}$–$N_{\rm H\,I}$ correlation) and the BLA line number density predicted by our simulations are in remarkably good agreement with observations.

BLAs arise in gas that is hotter, more highly ionized and more enriched than the gas giving rise to typical Lyα forest absorbers. The majority of the BLAs arise in warm-hot [$\log(T/{\rm K}) \sim 5$] gas at low ($\log\Delta < 1.5$) overdensities. On average, thermal broadening accounts for at least 60 per cent of the BLA linewidth, which in turn can be used as a rough indicator of the thermal state of the gas. Detectable BLAs account for only a small fraction of the true baryon content of the WHIM at low redshift. In order to detect the bulk of the mass in this gas phase, a sensitivity at least one order of magnitude better than achieved by current ultraviolet spectrographs is required. We argue that BLAs mostly trace gas that has been shock heated and enriched by outflows and that they therefore provide an important window on a poorly understood feedback process.

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

1 INTRODUCTION

The analysis of intervening H I Lyα absorption in the spectra of distant quasars (QSOs) has become an extremely powerful tool to study the spatial distribution of the diffuse intergalactic medium (IGM) that follows the large-scale distribution of cosmological filaments, and to constrain the baryon content of the IGM as a function of redshift. At redshifts $z > 3$, more than 95 per cent of the baryonic matter resides in the form of photoionized, diffuse gas giving rise to the 'Lyα forest' in the spectra of distant QSOs (e.g. Rauch, Haehnelt & Steinmetz 1997). As a consequence of expansion, the Lyα forest thins out, and at $z \approx 0$ the contribution of the Lyα forest to the total baryon budget has decreased to ∼20 per cent (e.g. Penton, Stocke & Shull 2004; Lehner et al. 2007). At the same time, the formation of galactic structures and the gravitational heating of the IGM by

⋆E-mail: tepper@astro.physik.uni-potsdam.de

collapsing large-scale filaments lead to a gradually increasing amount of shock-heated intergalactic gas at temperatures $T \gtrsim 10^5$ K, which is referred to as the warm-hot intergalactic medium (WHIM, Cen & Ostriker 1999; Theuns et al. 1998; Davé et al. 2001; Bertone, Schaye & Dolag 2008).

Since collisional ionization determines the ionization state of the shock-heated IGM, the neutral gas fraction in the WHIM is significantly lower, by at least one order of magnitude, than in the photoionized IGM of the same density (e.g. Richter, Paerels & Kaastra 2008). Because of this very small neutral hydrogen fraction in the WHIM, most of the recent observational campaigns to study warm-hot intergalactic gas at low redshift have concentrated on intervening absorption by highly ionized metals in ultraviolet (UV) spectra of bright QSOs. In particular, five times ionized oxygen (O VI) has been used extensively to trace shock-heated intergalactic gas at low redshift and to constrain the baryon content of the WHIM (e.g. Tripp, Savage & Jenkins 2000; Richter et al. 2004; Danforth et al. 2006; Danforth & Shull 2008; Thom & Chen 2008b; Tripp et al. 2008; Danforth, Stocke & Shull 2010). However, because O VI predominantly traces metal-enriched gas in a critical (in terms of ionization balance) temperature regime at $T \approx 3 \times 10^5$ K, and because the metals may well be poorly mixed on small scales (Schaye, Carswell & Kim 2007), the interpretation of intervening O VI absorbers is still controversial (e.g. Oppenheimer & Davé 2009; Smith et al. 2011; Tepper-García et al. 2011). In particular, it is not yet clear whether O VI absorbers predominantly arise in photoionized (e.g. Thom & Chen 2008a) or collisionally ionized gas (e.g. Danforth & Shull 2008), or in complex absorbing structures with cool gas intermingled with warm-hot gas (Tripp et al. 2008).

An alternative to highly ionized metals as tracers of warm-hot gas is offered by H I absorption. Due to the low neutral hydrogen fraction expected from collisional ionization at temperatures $T \gtrsim 10^5$ K, Lyα absorption from shock-heated WHIM filaments is expected to be very weak. In addition, H I absorption lines arising in gas at temperatures $T > 10^5$ K are expected to be relatively broad because of the effect of thermal broadening. Such broad ($b_{\rm H\,I} \geq 40\,{\rm km\,s^{-1}}$) and shallow ($[N_{\rm H\,I}\, b_{\rm H\,I}] \sim 10^{11}$ cm^{-2} km^{-1}s or $\tau_0({\rm H\,I}) \sim 0.1$) Ly$\alpha$ absorption features, the so-called broad ($b_{\rm H\,I} \geq 40\,{\rm km\,s^{-1}}$) H I Ly$\alpha$ absorbers (BLAs; Richter et al. 2006a), are hence difficult to identify in the UV spectra of QSOs because of the limited signal-to-noise ratio (S/N) and the low resolution of spectral data obtained with current space-based UV spectrographs.

In spite of being observationally challenging, directly detecting the small amounts of neutral hydrogen in the WHIM in absorption is a feasible task. The first systematic studies of BLAs at low redshift have been conducted using high-resolution *Hubble Space Telescope* (*HST*)/Space Telescope Imaging Spectrograph (STIS) spectra of bright QSOs (Richter et al. 2004; Sembach et al. 2004; Richter et al. 2006a; Williger et al. 2006; Lehner et al. 2007; Danforth et al. 2010). These studies indicate that BLAs may indeed account for a substantial fraction of the baryons in the WHIM at $z \approx 0$. They also show, however, that identification and interpretation of broad spectral features in UV spectra with limited data quality is afflicted with large systematic uncertainties. In particular, the effects of non-thermal broadening and unresolved velocity structure in the lines lead to the occurrence of broad spectral features that do not necessarily arise in gas at high temperatures. The Cosmic Origins Spectrograph (COS; Green et al. 2012), a new UV spectrograph which has recently been installed on *HST*, is expected to substantially increase the number of BLA candidates at low redshift. Due to the limited spectral resolution of COS ($\sim$17 km s^{-1}), the systematic uncertainties in identifying thermally broadened H I lines in the WHIM temperature range will nevertheless remain.

To investigate the physical properties and spectral signatures of BLAs at low redshift, Richter, Fang & Bryan (2006b) have studied broad H I absorption features using a cosmological simulation based on a grid-based adaptive mesh refinement method (Norman & Bryan 1999). Their simulation reproduces the observed BLA number density and supports the idea that BLAs trace (at least in a statistical sense) a substantial fraction of shock-heated gas in the WHIM at temperatures $T \sim 10^5$–10^6 K. However, since this (early) simulation ignored several important physical processes that are expected to affect the thermal state of this gas phase (i.e. energetic feedback, radiative heating and cooling by hydrogen and metals), it is important to reassess the frequency and physical properties of BLAs using state-of-the-art cosmological simulations with more realistic gas physics.

In this paper, we present a systematic study of BLAs at low redshift based on a set of cosmological simulations from the OverWhelmingly Large Simulations (OWLS) project (Schaye et al. 2010). This work complements our previous study on intervening O VI absorbers and their relation to the WHIM based on a slightly different set of OWLS simulations (Tepper-García et al. 2011, henceforth Paper I). The main features of the simulations we use are briefly described in Section 2. As we have done in Paper I for the case of low-redshift O VI absorbers, we compare the predictions from our fiducial model to a set of standard H I observables, and discuss various physical properties of the general H I absorber population in Section 3. Given the dependence of the WHIM mass fraction predicted by simulations on the particular implementation of the relevant physical processes reported in the past (e.g. Cen & Ostriker 2006), we investigate the impact of different physical models on the thermal state of the various gas phases in our simulations in Section 4. In this section, we also present and discuss the results on the physical properties of the absorbing gas traced by BLAs. Finally, we summarize our main findings in Section 5. In the appendices, we include: a full description of our fitting algorithm (Appendix A); a detailed calculation of the observability of H I absorbing gas in terms of optical depth as a function of density and temperature (Appendix B); a discussion of the convergence of our results with respect to the adopted physical model (Appendix C), and with respect to the adopted resolution and simulation box size (Appendix D).

2 SIMULATIONS

The simulations used in this work are part of a large set of cosmological simulations that together comprise the OWLS project, described in detail in Schaye et al. (2010, and references therein). Briefly, the simulations were performed with a significantly extended version of the *N*-Body, Tree-PM, smoothed particle hydrodynamics (SPH) code GADGET III – which is a modified version of GADGET II (last described in Springel 2005) – a Lagrangian code used to calculate gravitational and hydrodynamic forces on a system of particles. The initial conditions were generated from an initial glass-like state (White 1996) with CMBFAST (version 4.1; Seljak & Zaldarriaga 1996) and evolved to redshift $z = 127$ using the Zel'dovich (1970) approximation.

The reference model, dubbed *REF*, in the OWLS framework adopts a flat ΛCDM cosmology characterized by the set of parameters $\{\Omega_{\rm m}, \Omega_{\rm b}, \Omega_\Lambda, \sigma_8, n_{\rm s}, h\} = \{0.238, 0.0418, 0.762, 0.74, 0.95, 0.73\}$ as derived from the *Wilkinson Microwave*

Anisotropy Probe (*WMAP*) 3-year data[1] (Spergel et al. 2007). This model includes star formation following Schaye & Dalla Vecchia (2008), metal production and timed release of mass and heavy elements by intermediate-mass stars, that is, asymptotic giant branch stars and Type Ia supernovae (SNeIa), and by core-collapse SNe (SNeIIe) as described by Wiersma et al. (2009b). It further incorporates kinetic feedback by SNeIIe based on the method of Dalla Vecchia & Schaye (2008), as well as thermal feedback by SNeIa (Wiersma et al. 2009b). Radiative cooling by hydrogen, helium and heavy elements is included following the method of Wiersma, Schaye & Smith (2009a). The ionization balance for each SPH particle is computed as a function of redshift, density and temperature using precomputed tables obtained with the photoionization (PI) package CLOUDY (version 07.02.00 of the code last described by Ferland et al. 1998), assuming the gas to be optically thin and exposed to the Haardt & Madau (2001) model for the X-ray/UV background radiation from galaxies and QSOs. It is worth noting that a simulation run that adopts the *REF* model, although with a slightly different set of values for the cosmological parameters (from *WMAP7*), has been shown to reproduce the H I absorption observed at $z = 3$ in great detail (Altay et al. 2011).

Along with *REF*, we consider three further models from the OWLS suite referred to as *NOSN_NOZCOOL*, *NOZCOOL* and *AGN*. All these models differ from the reference model in one or more respects. *NOSN_NOZCOOL* neglects kinetic feedback by SNeIIe, and the calculation of radiative cooling assumes primordial abundances. It is the most simple model in terms of input physics, and it is similar (and hence useful for comparison) to the simulation used by Richter et al. (2006b). The model *NOZCOOL* assumes primordial abundances when computing radiative cooling, and the model *AGN* includes feedback by active galactic nuclei (AGNs) based on the model of black hole growth developed by Booth & Schaye (2009, see also Springel et al. 2005).

All these simulations were run in a cubic box of $100\,h^{-1}$ comoving Mpc on a side, containing 512^3 dark matter (DM) particles and equally many baryonic particles. The initial mass resolution is $4.1 \times 10^8\,h^{-1}\,\mathrm{M}_\odot$ (DM) and $8.7 \times 10^7\,h^{-1}\,\mathrm{M}_\odot$ (baryonic). The gravitational softening is set to $8\,h^{-1}$ comoving kpc and is fixed at $2\,h^{-1}$ proper kpc below $z = 3$.

In this study, we choose *AGN* as our fiducial model since it is the most complete model in terms of input physics. In addition to reproducing various standard H I statistics (see Appendix C), this model has been shown to reproduce: the observed mass density in black holes at $z = 0$; the black hole scaling relations (Booth & Schaye 2009) and their evolution (Booth & Schaye 2011); the observed optical and X-ray properties, stellar mass fractions, star formation rates (SFRs), stellar age distributions and the thermodynamic profiles of groups of galaxies (McCarthy et al. 2010); and the steep decline in the cosmic SFR below $z = 2$ (Schaye et al. 2010; van de Voort et al. 2011). Note that, while the H I statistics predicted by the *AGN* model are very similar to the predictions of the other models considered here (see Appendix C), there are notable differences in the temperatures of the gas traced by BLAs (see Fig. 11 shown later). We will address this point in more detail in Section 4.4. Table 1 briefly summarizes the relevant features of the models described above. For a more detailed description of these (and other) models that are part of the OWLS project, see Schaye et al. (2010).

[1] These parameter values are largely consistent with the *WMAP* 7-year results (Jarosik et al. 2011), the largest difference being the value of σ_8, which is 2σ lower in the *WMAP* 3-year data than allowed by the *WMAP* 7-year data.

Table 1. Overview of the simulations used in this study. All model variations are relative to the model *REF*.

Model	Description
NOSN_NOZCOOL	Neglects SN energy feedback and cooling assumes primordial abundances
NOZCOOL	Cooling assumes primordial abundances
REF	OWLS reference model (see text for details)
AGN	Includes feedback by AGNs (fiducial model)

3 THE GENERAL H I ABSORBER POPULATION

In this section, we test the predictions of our fiducial model (*AGN*) against observations using a set of well-measured H I observables: the H I column density distribution function (CDDF), the distribution of H I linewidths, and the correlation between H I column density and linewidth.

3.1 Synthetic spectra

For a meaningful comparison to existing data, we generate 5000 random sightlines (1000 at five redshifts spanning the range $0 \leq z \leq 0.5$ with step $\mathrm{d}z = 0.125$) through the simulation box covering a total redshift path $\Delta z = 189$, corresponding to an absorption path length $\Delta\chi = 275$.

We use the package SPECWIZARD written by Joop Schaye, Craig M. Booth and Tom Theuns to generate a synthetic spectrum for each sightline containing absorption by H I Lyα only. Briefly, we draw a random physical sightline across the simulation box of size L, which is simply defined as the line between two given points on (and perpendicular to) opposite faces of the box, and the collection of SPH particles with projected distances to this line smaller than their smoothing length. Next, we calculate the ionization balance for each SPH particle as a function of redshift, density and temperature, which we do using precomputed tables obtained with the PI package CLOUDY (version 07.02 of the code last described by Ferland et al. 1998), assuming the gas to be optically thin and exposed to the Haardt & Madau (2001) model for the X-ray/UV background radiation from galaxies and QSOs. We divide the physical sightline into $N_{\mathrm{pix}} = [a(z)\,L/h]\,/\Delta x$ pixels of constant width Δx, where h and $a(z)$ are the Hubble constant in units of $100\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ and the expansion factor at the box's redshift z, respectively, and compute the smoothed ion density n_{ion}, the ion density weighted gas temperature and the ion density weighted peculiar velocity at each pixel. Proper distance bins of width Δx along the sightline are transformed into velocity bins of width $\Delta v = H(z)\Delta x$, where $H(z)$ is the Hubble parameter at redshift z; ion number densities are transformed into ion column densities via $N_{\mathrm{ion}} = n_{\mathrm{ion}}\Delta x$, and gas temperatures into Doppler parameters using $b_T = \sqrt{2kT/m_{\mathrm{ion}}}$, where k is Boltzmann's constant and m_{ion} is the ion's mass. The H I optical depth $\tau(v)$ at each pixel is computed assuming a thermal (i.e. Gaussian) profile, taking peculiar velocities into account, as described by Theuns et al. (1998, their appendix 4). Finally, the optical depth spectrum is transformed into a continuum normalized flux via $F(v) = \exp[-\tau(v)]$.

We convolve our spectra with a Gaussian line spread function with full width at half-maximum $= 7$ $\mathrm{km\,s^{-1}}$ and resample our

spectra on to 3.5 km s^{-1} pixels. We add Gaussian noise to each spectrum assuming a flux-dependent root mean square amplitude given by (S/N)$^{-1}F(v)$, where S/N is the adopted signal-to-noise ratio. We assume a *minimum*, that is, flux-independent noise level $\sigma_{\rm min} = 10^{-4}$. This implies that our algorithm will underestimate the true column density of absorption features with a flux of the order of (or lower than) $\sigma_{\rm min}$, which corresponds to a logarithmic central optical depth $\log \tau_0 \sim 1$ (see Appendix A). Our choice of a perhaps unrealistically low value for $\sigma_{\rm min}$ thus allows us to reduce the gap between the true and the fitted column densities of saturated lines.

We generate three sets of spectra, adopting S/N = 10, 30 and 50. The spectra with S/N = 10 and 30 thus closely match the properties of the large sample thus far obtained with *HST*/STIS; these will be used in Sections 3.2, 3.3 and 3.4 to test the predicted H I observables against observations; the synthetic spectra with S/N = 50 are intended to investigate the physical properties of the H I absorbing gas following a statistical approach, in the remaining sections of this paper.

Fitting of our 5000 synthetic spectra using the procedure described in Appendix A yields a total of 93 430, 66 705 and 28 649 components for S/N = 50, 30 and 10, respectively. The resulting line number densities and their corresponding Poisson uncertainties are (dN/dz) = 494 ± 22 (S/N = 50), 353 ± 19 (S/N = 30) and 152 ± 12 (S/N = 10). For reference, the sample of 341 Lyα absorbers at $z \lesssim 0.4$ identified in seven *FUSE*+STIS spectra with average S/N $\gtrsim$ 10 by Lehner et al. (2007) along an unblocked redshift path Δz = 2.064 yields (dN/dz) = 165 ± 13 at S/N ≈ 10, which agrees (within the Poisson uncertainties) with our result at a similar S/N.

3.2 Column density distribution function

In Fig. 1, we show the H I CDDF, $f(N_{\rm H\,I})$, obtained from our spectra with S/N = 10 (red) and S/N = 30 (blue) spanning the redshift range [0, 0.5], together with results from different observations at similar redshifts using spectra with comparable (average) S/N values. Assuming that the H I CDDF can be parametrized in the form of a single power law, $f(N_{\rm H\,I}) \propto {N_{\rm H\,I}}^{-\beta}$, we find $\beta = 1.916 \pm 0.044$ (1.917 ± 0.032) for S/N = 10 (30) for logarithmic column densities in the range [13.0, 15.2] ([12.5, 15.2]). The lower limit in $\log N_{\rm H\,I}$ approximately corresponds in each case to the completeness limit as given by equation (A3), while the upper limit roughly indicates the column density above which our fitting algorithm underestimates the true H I column density due to the minimum noise level adopted (see Appendix A).

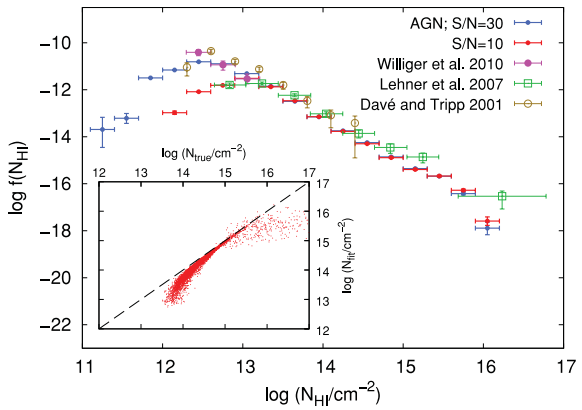

Figure 1. H I CDDF, $f(N_{\rm H\,I})$, obtained from observations and from 5000 simulated spectra with different S/N values spanning the redshift range [0, 0.5]. The error bars along the y-axis show Poisson uncertainties computed using the tables by Gehrels (1986). Assuming $f(N_{\rm H\,I}) \propto {N_{\rm H\,I}}^{-\beta}$ for logarithmic column densities in the range [13.0, 15.2] ([12.5, 15.2]), we find $\beta = 1.916 \pm 0.044$ (1.917 ± 0.032) for S/N = 10 (30). The inset shows a comparison between the true (x-axis) and the fitted (y-axis) H I column densities integrated along each individual sightline. For clarity, only the result for S/N = 10 is shown. The dashed line corresponds to a perfect match (see text for discussion).

The slope we obtain is in fairly good agreement with the slope measured from different observations. For $z \lesssim 0.4$ Lehner et al. (2007, their table 7) measure a range of values β = 1.52–1.92 for absorbers in selected column density intervals between $\log(N_{\rm H\,I}/{\rm cm}^{-2}) = 13.2$ and 16.5, and linewidths $b_{\rm H\,I} \leq 40$ km s^{-1} or $b_{\rm H\,I} \leq 80$ km s^{-1}. If we extend the fitted column density range to $\log(N_{\rm H\,I}/{\rm cm}^{-2}) = 16.5$, we find $\beta = 1.90 \pm 0.06$ and 1.95 ± 0.06 for S/N = 10 and 30, respectively. Williger et al. (2010) use a subsample from the Lehner et al. (2007) data and their own data at $\log(N_{\rm H\,I}/{\rm cm}^{-2}) \leq 12.3$, and find $\beta = 1.79 \pm 0.1$. Davé et al. (2001) measure $\beta = 2.04 \pm 0.23$ for absorbers with column densities $\log(N_{\rm H\,I}/{\rm cm}^{-2}) \geq 12.9$ at a median redshift $\widetilde{z} = 0.17$. Note, however, that a significantly shallower slope is found by Penton et al. (2004) who identify 109 Lyα absorbers at $z < 0.069$ along 15 STIS spectra with S/N $\gtrsim$ 20, and measure $\beta = 1.65 \pm 0.07$ for logarithmic H I column densities in the range [12.5, 14.5].

The amplitude of the H I CDDF resulting from the analysis of our synthetic spectra adopting different S/N is also in remarkable agreement with the observations. Note that the amplitude comes out naturally from our simulation, that is, the H I CDDF has not been normalized to match the data in any way (even though that could have been justified because of uncertainties in the intensity of the UV background). At column densities $\log(N_{\rm H\,I}/{\rm cm}^{-2}) \lesssim 15$, our predicted amplitude agrees well with the data all the way down to the lowest column densities measured, $\log(N_{\rm H\,I}/{\rm cm}^{-2}) = 12.3$. At $\log(N_{\rm H\,I}/{\rm cm}^{-2}) > 15$, the amplitude of our predicted H I CDDF appears slightly lower (or its slope is steeper) than the result by Lehner et al. (2007). Note, however, that their data point at highest measured column density bin has a rather large uncertainty. On the other hand, it is very likely that our choice of fitting parameters leads us to underestimate the amplitude of the predicted H I CDDF at $\log(N_{\rm H\,I}/{\rm cm}^{-2}) \gtrsim 14.5$ by underestimating the true column density of *saturated* lines, as explained in Appendix A. A comparison between the true and the fitted H I column densities integrated along each sightline reveals that our fitting procedure indeed yields integrated H I column densities which are systematically lower than the true total column density, in particular for $\log(N_{\rm H\,I}/{\rm cm}^{-2}) \gtrsim 15$ (see the inset in Fig. 1). This could explain the difference between our predicted H I CDDF and the result by Lehner et al. (2007) at the high-$N_{\rm H\,I}$ end.

3.3 Linewidth distribution

Fig. 2 shows the distribution of Doppler parameters, $b_{\rm H\,I}$, obtained from our synthetic spectra for S/N = 10 and 30 spanning the redshift range [0, 0.5], together with the linewidth distributions obtained from data with comparable S/N values and redshifts by Lehner et al. (2007, green data points) and Danforth & Shull (2008, orange data points). The median values of our predicted distributions are $\widetilde{b_{\rm H\,I}} \approx$ 30.4, 29.8 and 29.4 km s^{-1} for S/N = 10, 30 and 50 (not shown), respectively. All of these agree well with the median value found by Heap et al. (2002), $\widetilde{b_{\rm H\,I}} = 27$ km s^{-1}, by Shull et al. (2000), $\widetilde{b_{\rm H\,I}} = 28$ km s^{-1}, and with the median value $\widetilde{b_{\rm H\,I}} = 31$ km s^{-1} for the full Lehner et al. (2007) sample. Note that all of these values are significantly larger than the median value $\widetilde{b_{\rm H\,I}} = 21$ km s^{-1}

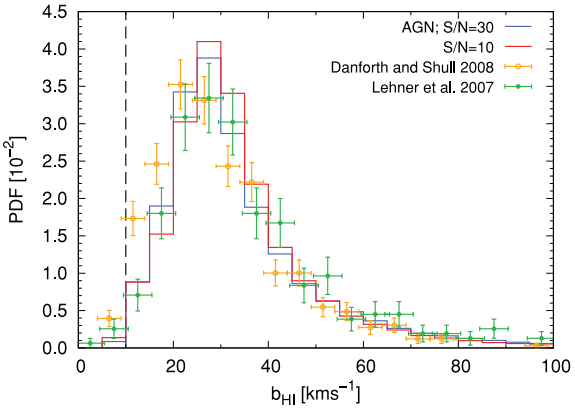

Figure 2. Distribution of $b_{\rm H\,I}$ values from observations and from 5000 simulated spectra with S/N = 10 (red), and S/N = 30 (blue), spanning the redshift range [0, 0.5]. Data points from Lehner et al. (2007, green) and Danforth & Shull (2008, orange) with y-error bars showing Poisson uncertainties. The dashed vertical line indicates our adopted minimum allowed b value. The distributions from both data and simulations have been binned using $\Delta b_{\rm H\,I} = 5\,{\rm km\,s^{-1}}$.

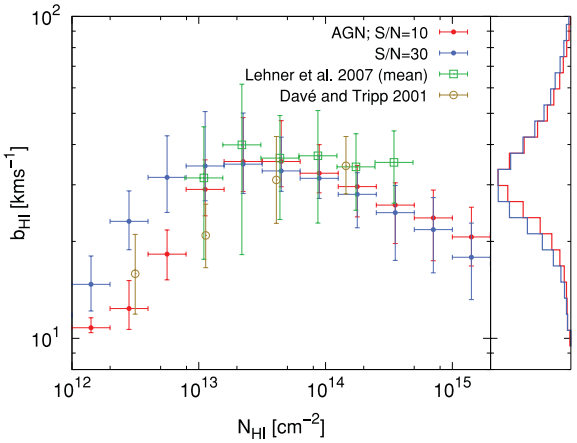

Figure 3. $b_{\rm H\,I}$–$N_{\rm H\,I}$ distribution obtained from 5000 simulated spectra with S/N = 10 (red), and S/N = 30 (blue), spanning the redshift range [0, 0.5], and measurements by Lehner et al. (2007, green data points from their table 3 for $b_{\rm H\,I} > 0\,{\rm km\,s^{-1}}$) and Davé & Tripp (2001, olive data points). The points show the median $b_{\rm H\,I}$ value in each bin of size $\Delta\log(N_{\rm H\,I}/{\rm cm^{-2}}) = 0.3$, while the error bars parallel to the y-axis correspond to 25 and 75 percentiles. Note that the red (blue) histogram on the right-hand subpanel corresponds to the red (blue) histogram in Fig. 2, but with a different binning and on a different scale.

measured by Davé & Tripp (2001). Our simulation shows a lower fraction of broad ($b_{\rm H\,I} > 40\,{\rm km\,s^{-1}}$) absorbers when compared to the Lehner et al. (2007) b-value distribution, but our results compare well to the linewidth distribution from Danforth & Shull (2008).

The predicted median $b_{\rm H\,I}$ values indicate that a lower S/N value systematically shifts the linewidth distribution to slightly larger values. Yet, the number of components with $b_{\rm H\,I} \geq 40\,{\rm km\,s^{-1}}$ relative to the total number of components identified in each case decreases from ∼26 to ∼23 per cent when the adopted S/N value decreases from 50 to 10. Here, two competing mechanisms are at work: on the one hand, a low S/N value results in a stronger blending of narrow components into (artificial) broad features, while, on the other hand, since broader lines are shallower (at a given column density), and thus more difficult to detect at low S/N, the number of broad components detected decreases with decreasing S/N. Compared to a higher S/N value, the net effect of a low S/N value is to yield a smaller number (both relative and absolute) of broad absorption features (at a given resolution and sensitivity).

3.4 The $b_{\rm H\,I}$–$N_{\rm H\,I}$ distribution

Finally, we compare the $b_{\rm H\,I}$–$N_{\rm H\,I}$ distribution obtained from our simulated spectra with S/N = 30 and 10 to two different sets of observations used for the comparison of our predicted H I CDDF and the linewidth distribution discussed in the last sections. To this end, we bin the lines from observations and from our synthetic spectra in $N_{\rm H\,I}$ using $\Delta\log(N_{\rm H\,I}/{\rm cm^{-2}}) = 0.3$, and compute the median $b_{\rm H\,I}$ value, and 25/75 percentiles in each bin. The result is shown in Fig. 3. The $b_{\rm H\,I}$–$N_{\rm H\,I}$ distribution from our simulated spectra matches the observations well within the uncertainties. Even the drop in $b_{\rm H\,I}$ observed at low $N_{\rm H\,I}$ in Davé & Tripp (2001) is well reproduced by our simulation. Note that lines with $\log(N_{\rm H\,I}/{\rm cm^{-2}}) < 13.4$ identified in spectra with S/N = 30 generally have larger widths. This is a consequence of the fact that, at a fixed column density, lines with a given width are shallower with respect to narrower lines, and they can only be detected if the S/N is high enough.

Summarizing, we conclude that the H I observables predicted by our fiducial model are in excellent agreement with observations. This agreement may be surprising in view of the uncertainty in the input physics used in our simulation. However, in Appendix C we show that these results are quite robust against the model variations with respect to our fiducial model considered here (see Section 4.1). We now proceed with the analysis of the physical conditions in low-z H I absorbers.

3.5 Physical state of the H I absorbing gas

In this section, we present and discuss the physical properties of the gas detected via H I absorption in our fiducial model (*AGN*; see Table 1). The method we use is similar to the method described in Paper I, in which we used optical depth weighted quantities. Briefly, to compute the desired H I optical depth weighted quantity (e.g. density) associated with a given absorption line, we first compute the optical depth weighted density in redshift space along the sightline as in Schaye et al. (1999). Next, we compute the average of the optical depth weighted density over the line profile, weighted again by the optical depth in each pixel and assign this last weighted average to the line. In concordance with Paper I, in the following we shall denote quantities weighted by H I optical depth by adding a corresponding subscript; thus, for example, the H I optical depth weighted temperature is denoted by $T_{\rm H\,I}$. We refer the reader to section 5.1 of Paper I for a more detailed description about our method for computing optical depth weighted quantities.

For simplicity, we obtain a new line sample of H I absorbers identified in synthetic spectra with S/N = 50 generated from 5000 sightlines across a simulation box at a single redshift,[2] $z = 0.25$, spanning a total redshift path $\Delta z = 187.5$, which corresponds to an absorption path length $\Delta\chi = 270$. These spectra have been fitted following the method described in Appendix A.

[2] Note that our chosen redshift is slightly higher than the median redshift of most H I absorption-line studies at low redshift (e.g. $\widetilde{z} \approx 0.17$ in Lehner et al. 2007). Although *some* evolution does take place from $z = 0.5$ to 0.0, we do not expect the choice of this particular redshift to affect our conclusions in any significant way.

Table 2. Line number density, (dN/dz), total baryon content in H I, $\Omega_{\rm H\,I}$, and total baryon content in the gas traced by H I, $\Omega_{\rm b}$(H I), related to simple H I absorbers identified in 5000 spectra with S/N = 50, 30 and 10 at z = 0.25.

	S/N = 50	S/N = 30	S/N = 10
Components (related to S/N = 50)	1	0.72	0.31
Simple absorbers (related to total)	0.47	0.53	0.65
(dN/dz)[a]	460 ± 22	332 ± 18	144 ± 12
$\Omega_{\rm H\,I}$ [10^{-7}][b]	1.20	1.15	1.12
$\Omega_{\rm b}$(H I)/ $\Omega_{\rm b}$	0.57	0.47	0.29

[a]Quoted uncertainties are purely Poissonian. For comparison, Lehner et al. (2007) obtain (dN/dz) = 165 ± 13 at S/N ≈ 10.
[b]Total baryon content in H I obtained by adding the column densities of all identified H I components. The true total baryon content in H I along the fitted sightlines at $z = 0.25$ is $\Omega_{\rm H\,I} = 2.11 \times 10^{-5}$.

We restrict our analysis to 'simple', that is, single-component, absorbers, unless stated otherwise. We define an absorber i as 'simple' if the velocity distance from its centre to any other component j along the same sightline satisfies $\Delta v > 2\sigma_b$, where $\sigma_b^2 \equiv 0.5[b_{\rm H\,I}^2(i) + b_{\rm H\,I}^2(j)]$. Absorption lines that do not satisfy this condition are referred to as 'complex'.

Table 2 contains various statistical and physical quantities resulting from the analysis of this new line sample, such as the relative number of identified components, the relative number of simple absorbers, the line number density, (dN/dz), the total baryon content[3] in H I, $\Omega_{\rm H\,I}$, and the total baryon content in the gas traced by H I, $\Omega_{\rm b}$(H I) (see also Section 4.5). Note that the statistical and physical properties of this new sample are very similar to the corresponding properties of the line sample discussed in Sections 3.2–3.4.

3.5.1 *Physical density and absorber strength*

As previously noted by several studies (e.g. Davé et al. 1999; Schaye et al. 1999), there exists a tight correlation between the H I column density, $N_{\rm H\,I}$, and overdensity,[4] $\Delta \equiv \rho_{\rm b}/\langle\rho_{\rm b}\rangle$, of the absorbing gas usually parametrized in the form of a power law, $\Delta/\Delta_0 = (N_{\rm H\,I}/N_0)^a$. Due to variations in the (local) ionizing radiation field, the influence of other heating mechanisms (shocks), and other factors such as the geometry of the absorbing structures, etc., this relation has an intrinsic scatter, which decreases with increasing redshift (Davé et al. 1999).

The relation between overdensity and H I column density for the diffuse IGM has been derived analytically by Schaye (2001), who assuming local hydrostatic equilibrium[5] and optically-thin gas finds

$$\log\Delta \propto \frac{2}{3+\alpha(1-2b)}\left[\log N_{\rm H\,I} - \frac{9}{2}\log(1+z)\right] . \quad (1)$$

In the above equation, α is the slope of the temperature–density relation, $T = T_0\,\Delta^{\alpha}$, which results from the balance between photoheating and adiabatic cooling (Hui & Gnedin 1997), and $-b$ is the power of the temperature in the expression for the H I recombination rate coefficient which behaves as $\propto T^{-b}$. If reionization of the IGM takes place at sufficiently high redshifts, its imprints on the thermal state of the IGM are eventually washed out, and the slope of temperature–density relation is expected to reach an asymptotic limit determined by the temperature dependence of the H I recombination rate. More specifically, at low redshift $\alpha \to 1/(1+b)$. We find[6] $b = 0.755 \pm 0.001$ in the temperature range $[10^3, 5\times10^4]$ K, and hence $\alpha \to 0.570$. Inserting this value into equation (1) gives $\Delta \propto N_{\rm H\,I}^{0.738}\,(1+z)^{-3.36}$. Thus, the value of the amplitude Δ_0 in the Δ–$N_{\rm H\,I}$ relation decreases with redshift, implying that absorbers of a given column density trace gas at higher overdensities at lower redshift.

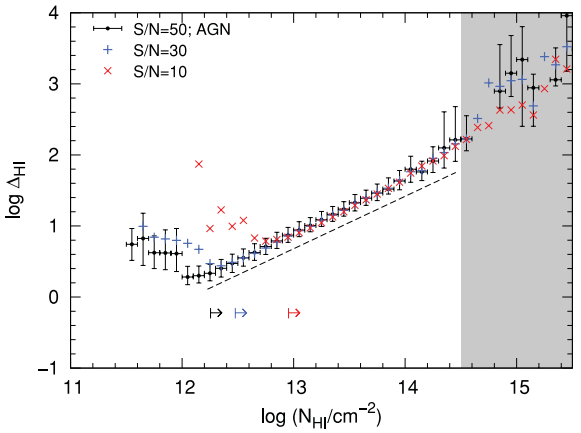

Figure 4. $\Delta_{\rm H\,I}$–$N_{\rm H\,I}$ relation for simple H I absorbers identified in spectra with different S/N values at $z = 0.25$ for. The points show the median overdensity in each bin of size $\Delta\log(N_{\rm H\,I}/{\rm cm}^{-2}) = 0.1$. The error bars along the y-direction correspond to 25 and 75 percentiles in each bin. For clarity, only the error bars for the S/N = 50 result are shown, but they are similar for S/N = 30 and 10. The horizontal arrows indicate the formal completeness limit for each adopted S/N value as given by equation (A3). Note the deviation of the $\Delta_{\rm H\,I}$–$N_{\rm H\,I}$ relation from a single power law for column densities below the formal completeness limit (at a given S/N) and in the column density range for which the H I Lyα line generally saturates (shaded area). A power law with the theoretical expected slope 0.738 (but with arbitrary amplitude; Schaye 2001) has been included to guide the eye (black dashed line).

In Fig. 4, we show the $\Delta_{\rm H\,I}$–$N_{\rm H\,I}$ relation resulting from our fiducial model for all simple H I absorbers identified in spectra with different S/N values at $z = 0.25$. The turnover in overdensity at column densities below the sensitivity limit (equation A3) for each adopted S/N is caused by errors in the measured $N_{\rm H\,I}$ below this limit. Note also the deviation of the $\Delta_{\rm H\,I}$–$N_{\rm H\,I}$ relation from a single power law at column densities $\log(N_{\rm H\,I}/{\rm cm}^{-2}) \gtrsim 14.5$ (indicated by the shaded area), which corresponds approximately to the column density for which the Lyα line saturates. This is in part due to the inability of our algorithm to properly fit saturated lines.

Performing a least-squares, error-weighted fit to the S/N = 50 result for column densities above the corresponding sensitivity limit [$\log(N_{\rm H\,I}/{\rm cm}^{-2}) = 12.3$; see equation A3] and restricted to $\log(N_{\rm H\,I}/{\rm cm}^{-2}) \leq 14.5$, we find $a = 0.786 \pm 0.010$ and $\Delta_0 = 48.3$, normalized to $N_0 = 10^{14}\,{\rm cm}^{-2}$. The resulting slope for S/N = 30 (10) is $a = 0.798 \pm 0.010$ (0.847 ± 0.014) in the logarithmic column density range [12.5, 14.5] ([13.0, 14.5]), where the lower

[3] The total baryon content in H I is computed via

$$\Omega_{\rm H\,I} = \frac{m_{\rm H}}{\rho_{\rm c}}\left(\frac{c}{H_0}\sum_{i=1}^{N_{\rm LOS}}\Delta\chi_i\right)^{-1}\sum_{i=1}^{N_{\rm LOS}}\sum_{j=1}^{N_{\rm abs}}(N_{\rm H1})_{ij} .$$

[4] The mean baryonic density in our model is $\langle\rho_{\rm b}\rangle = 4.18 \times 10^{-31}(h/0.73)^2(1+z)^3\,{\rm g\,cm}^{-3}$.
[5] The assumption of 'local hydrostatic equilibrium' implies that the size of a self-gravitating gas cloud is of the order of the local Jeans length.
[6] We compute the recombination rate coefficient for recombination case A numerically, and fit a power law in the given temperature range.

column density limit is given by equation (A3). A power law with the theoretical expected slope 0.738 and an arbitrary amplitude has been included in this figure for reference (black dashed line).

For comparison, Davé et al. (2010, their equation 3) find $a = 0.741 \pm 0.003$ and $\Delta_0 = 38.9$ at $z = 0.25$, for absorbers arising in gas with temperatures $\log(T/\mathrm{K}) < 4.5$ in their simulation. If we restrict our sample to single-component absorbers with $\log(T_{\rm H\,I}/\mathrm{K}) < 4.5$, we find $a = 0.786 \pm 0.014$ and $\Delta_0 = 45.1$. We note that we do not rescale the amplitude of the UV background in our simulation, while Davé et al. (2010) adjust its amplitude by a factor of 3/2 to bring their predicted evolution of the H I optical depth into better agreement with observations.

3.5.2 *Gas temperature and linewidth*

A matter of interest is to which extent the measured H I linewidth can be used to estimate the temperature of the H I absorbing gas. We explore this by comparing the H I thermal linewidth, $b_T = 12.9\sqrt{T_{\rm H\,I}/10^4\,\mathrm{K}}$, computed from the optical depth weighted gas temperature, $T_{\rm H\,I}$, to the total H I linewidth, $b_{\rm H\,I}$, as a function of the line strength as given by the optical depth at the line centre, τ_0. The optical depth at the line centre is computed using the H I column density and the H I linewidth inferred from a Voigt profile fit to the line (see Appendix A). We bin the ratio $(b_T/b_{\rm H\,I})$ in τ_0, and plot in Fig. 5 the median value and the 25/75 percentiles in each bin as a function of τ_0, for all single-component absorbers identified in our synthetic spectra with different S/N values at $z = 0.25$.

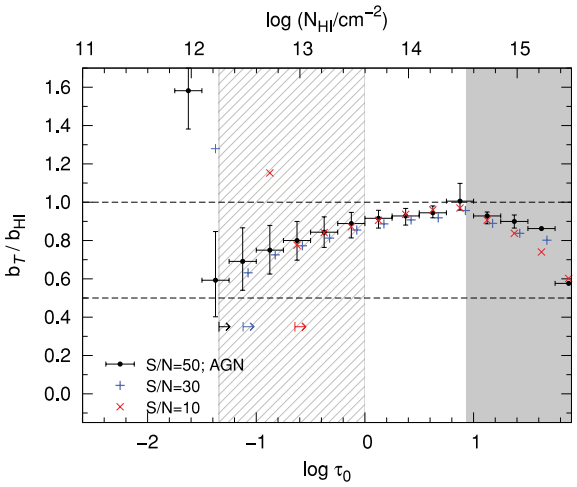

Figure 5. Ratio of H I line thermal width, $(b_T/\mathrm{km\,s^{-1}}) = 12.9\sqrt{T_{\rm H\,I}/10^4\,\mathrm{K}}$, to total linewidth, $b_{\rm H\,I}$, as a function of the optical depth at line centre, τ_0, inferred from Voigt profile fits, for simple H I absorbers identified in spectra at $z = 0.25$ adopting different S/N values. The top axis indicates the corresponding H I column density assuming $b_{\rm H\,I} = 30\,\mathrm{km\,s^{-1}}$, which corresponds to the median value of the Doppler parameter distribution (see Section 3.3). The points show the median $(b_T/b_{\rm H\,I})$ value in each bin of size $\Delta\log\tau_0 = 0.25$ indicated by the error bars parallel to the x-axis; the lower and upper error bars parallel to the y-axis correspond to the 25 and 75 percentiles in each bin, respectively. For clarity, only the error bars for the S/N = 50 result are shown, but they are similar for S/N = 30 and 10. Note that the values corresponding to S/N = 30 (blue crosses) have been slightly shifted for display purposes. The horizontal arrows indicate the formal completeness limit for each adopted S/N value as given by equation (A3). The dashed horizontal lines enclose the range $0.5b_{\rm H\,I} \le b_T \le b_{\rm H\,I}$, and have been included to guide the eye. The grey, shaded area indicates the central optical depth (or column density) range for which the H I Lyα line generally saturates. The hatched area indicates the typical range in τ_0 for BLAs, $-1.34 < \log\tau_0 < 0$, detected in spectra with S/N = 50 (see Section 4.2.1).

We see that thermal broadening becomes increasingly important with increasing line strength (or H I column density), and it contributes with at least 50 per cent to the total linewidth, that is, $b_T \gtrsim 0.5b_{\rm H\,I}$, irrespective of the line strength and the adopted S/N value. The temperature of gas giving rise to absorption lines with central optical depths in the range $1 < \tau_0 < 8.5$ (corresponding to strong lines) on average contributes with at least 90 per cent to the total linewidth, that is, $b_T \gtrsim 0.9b_{\rm H\,I}$. The lower $(b_T/b_{\rm H\,I})$ value shown by highly saturated lines, that is, lines with $\tau_0 > 8.5$ (grey, shaded area) is most probably due to the uncertainty in the fit parameters of such lines. These results suggest, in view of the tight $\Delta_{\rm H\,I}$–$N_{\rm H\,I}$ relation discussed above, that absorption arising in low-density gas is subject to more significant non-thermal (i.e. Hubble) broadening than gas at higher density. This is consistent with the idea that low-density, unbound gaseous structures are subject to the universal expansion, while gas at higher densities residing closer to galaxies may have detached from the overall expansion. We find (not shown) that the linewidth correlates well with the gas temperature for $\log(N_{\rm H\,I}/\mathrm{cm^{-2}}) \gtrsim 13$, but that it is a poor indicator of the thermal state of the gas for lower column densities.

Fig. 5 shows also that lines with central optical depths corresponding to H I column densities below the formal completeness limit for each adopted S/N value (indicated by the arrows) as given by equation (A3) have, on average, $b_T > b_{\rm H\,I}$, which is unphysical. These lines correspond to absorption by gas at high temperatures, which gives rise to very shallow and extremely broad features that are (incorrectly) fitted with several components, thus yielding linewidths that are narrower than allowed by the gas temperature.

In the central optical depth range typical for BLAs detected in spectra with S/N = 50, $-1.34 < \log\tau_0 < 0$ (see Section 4.2.1) indicated by the hatched area, the contribution of thermal broadening to the linewidth amounts to 60–90 per cent. If non-thermal processes (e.g. turbulence) contribute to the line broadening in such a way that the total linewidth is given by $b_{\rm H\,I}^2 = b_T^2 + b_{\rm nt}^2$, where $b_{\rm nt}$ is the non-thermal broadening (as would be the case for a purely Gaussian turbulence field), then the ratio of non-thermal broadening to total linewidth can be important, even though the thermal contribution is substantial. Take as an example the maximum, average thermal broadening to total line ratio for BLAs $b_T/b_{\rm H\,I} = 0.9$; this value together with $b_{\rm H\,I}^2 = b_T^2 + b_{\rm nt}^2$ implies $b_{\rm nt}/b_{\rm H\,I} = 0.4$.

3.5.3 *The* $(N_{\rm H\,I},\ b_{\rm H\,I})$ *plane*

A deeper insight into the physical state of gas giving rise to H I absorption identified in real QSO spectra can be gained by studying the relation between selected physical quantities and the line observables, $N_{\rm H\,I}$ and $b_{\rm H\,I}$, simultaneously. We have followed such an approach in Paper I in order to study the physical conditions of O VI absorbing gas, and now apply it to study the gas traced via H I absorption. We focus on four quantities: gas temperature, $T_{\rm H\,I}$, neutral hydrogen fraction, $(n_{\rm H\,I}/n_{\rm H})_{\rm H\,I}$, total hydrogen column density, $N_{\rm H}$, and gas metallicity, Z. Note that gas temperature and metallicity are 'true' optical depth weighted quantities, while total hydrogen column density and ionization fraction are 'derived' quantities. For instance, the neutral fraction is computed using the optical depth weighted hydrogen particle density, $(n_{\rm H})_{\rm H\,I}$, and the optical depth weighted gas temperature is computed using precomputed tables obtained with the PI package CLOUDY (version 07.02 of the code last described by Ferland et al. 1998), as described in Section 2.

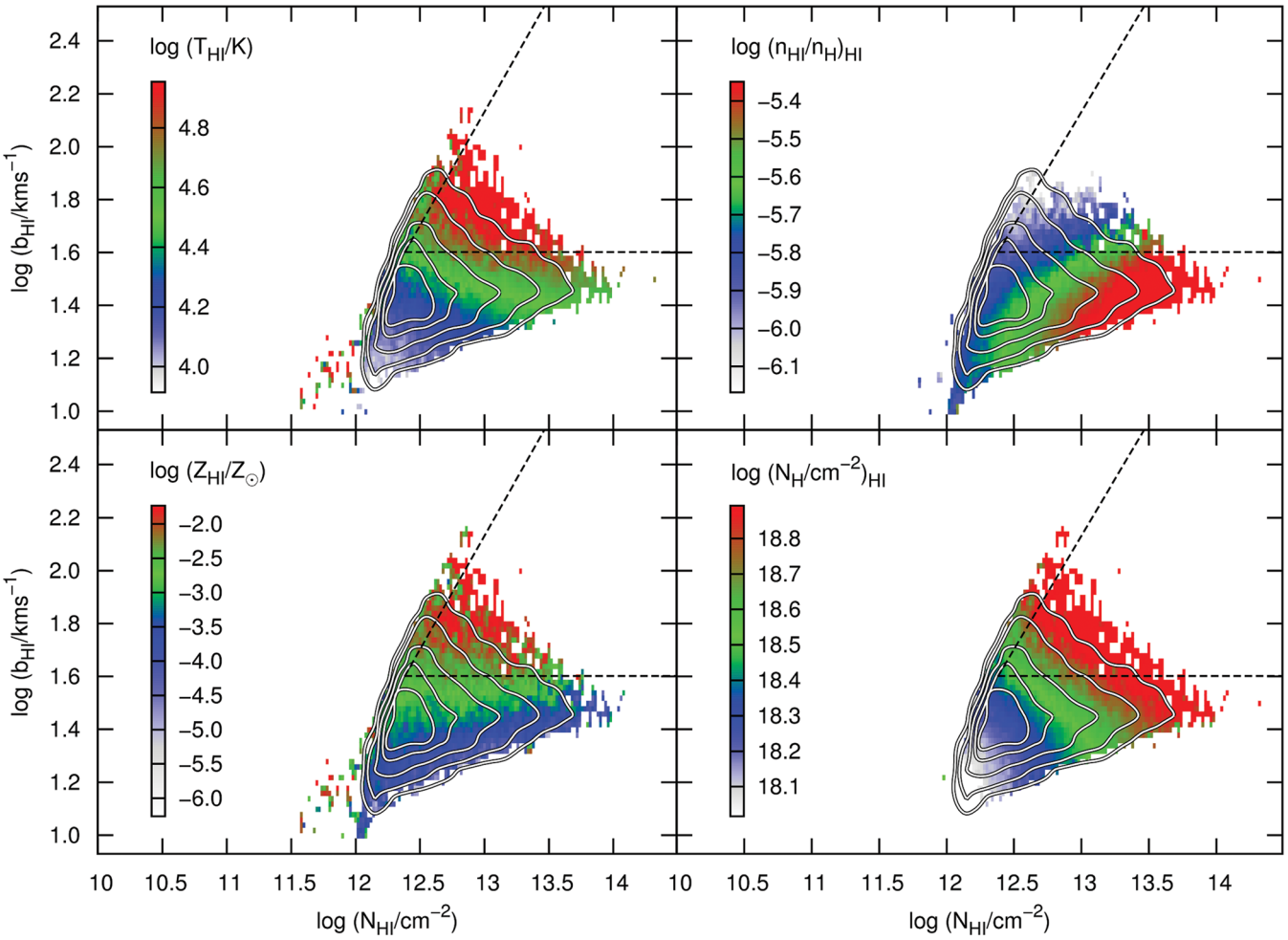

Figure 6. Temperature (top left-hand panel), neutral hydrogen fraction (top right-hand panel), metallicity (bottom left-hand panel) and total hydrogen column density $N_{\rm H}$ (bottom right-hand panel) of the gas at $z = 0.25$ traced by simple H I absorbers identified in spectra with S/N = 50, as a function of the line parameters $N_{\rm H\,I}$ and $b_{\rm H\,I}$. The colour coding shows the median value in each cell of the corresponding physical quantity. The white contours show the distribution (by number) of the sample of simple H I absorbers, and they are identical in all panels. These contours enclose, starting from the innermost, 20, 40, 60, 80 and 90 per cent of the total number of single-component H I absorbers. The horizontal and diagonal dashed lines indicate the usual BLA selection criteria $b_{\rm H\,I} \geq 40\,{\rm km\,s^{-1}}$ and $(N_{\rm H\,I}/b_{\rm H\,I}) = 3 \times 10^{12}\,{\rm cm^{-2}\,km^{-1}\,s\,(S/N)^{-1}}$ (e.g. Richter et al. 2006a), respectively. Absorbers within this boundaries typically trace gas at $\log(T/{\rm K}) \gtrsim 4.7$, have total hydrogen column densities $18 \lesssim \log(N_{\rm H}/{\rm cm^{-2}}) \lesssim 18.9$ and low ionization fractions $\log(n_{\rm H\,I}/n_{\rm H})_{\rm H\,I} \lesssim -5.5$. Also, the (local) metallicity of the gas traced by these absorbers is typically $\log(Z_{\rm H\,I}/Z_{\odot}) \gtrsim -2.5$.

For each of these physical quantities we proceed as follows. First, we compute the desired physical quantity, for example, $T_{\rm H\,I}$, for each of the simple absorbers in our S/N = 50 line sample. We then divide the ($N_{\rm H\,I}$, $b_{\rm H\,I}$) plane into cells, and compute the median value for the desired quantity using the values of all absorbers with ($N_{\rm H\,I}$, $b_{\rm H\,I}$) values in that cell. Fig. 6 displays the result for temperature (top left-hand panel), neutral hydrogen fraction (top right-hand panel), metallicity (bottom left-hand panel) and total hydrogen column density (bottom right-hand panel). The colour code indicates the median value of the corresponding physical quantity. For reference, we include contours (white solid curves) showing the distribution by number of the simple H I absorbers on the ($N_{\rm H\,I}$, $b_{\rm H\,I}$) plane. These contours enclose, starting from the innermost, 20, 40, 60, 80 and 90 per cent of the total number of H I components. The dashed horizontal and diagonal lines at the top right-hand corner of each panel indicate, respectively, the BLA selection criteria $b_{\rm H\,I} \geq 40\,{\rm km\,s^{-1}}$ and $(N_{\rm H\,I}/b_{\rm H\,I}) = 3 \times 10^{12}\,{\rm cm^{-2}\,km^{-1}\,s}$ $({\rm S/N})^{-1}$ (Richter et al. 2006a) adopting S/N = 50.

There are several interesting features in this figure. First, all four physical quantities appear to have a relatively simple dependence on $N_{\rm H\,I}$ and $b_{\rm H\,I}$. The temperature of the gas (top left-hand panel), for example, shows a positive correlation with the linewidth, which appears to be tighter for absorbers at a given $N_{\rm H\,I}$ (range), in agreement with the results presented in Section 3.5.2. In this respect, note the population of narrow ($b_{\rm H\,I} \sim 10\,{\rm km\,s^{-1}}$), low column density ($N_{\rm H\,I} < 10^{12}$ cm^{-2}) absorbers with high [$\log(T_{\rm H\,I}/{\rm K}) > 4.8$] optical depth weighted temperatures. These correspond to the absorbers with column densities below the formal completeness limit and with $b_T > b_{\rm H\,I}$, previously discussed.

The neutral hydrogen fraction (top right-hand panel) increases with $N_{\rm H\,I}$, but strongly decreases with $b_{\rm H\,I}$. This can be interpreted as a temperature dependence, given the positive correlation between $T_{\rm H\,I}$ and $b_{\rm H\,I}$. Correspondingly, the total hydrogen column density (bottom right-hand panel) increases with both $N_{\rm H\,I}$ and $b_{\rm H\,I}$. The bottom left-hand panel shows that the (local) metallicity of the gas is strongly correlated with the linewidth. Given the correlation between gas temperature and linewidth shown in the top

left-hand panel, this suggests that there is a correlation between gas temperature and (local) gas metallicity. This correlation is very likely a consequence of strong feedback. Indeed, high-temperature, high-metallicity absorbers could be tracing shock-heated, enriched outflows in the surroundings of galaxies that have not had enough time to mix with the surrounding gas and to cool down, whereas low-temperature, low-metallicity absorbers could be tracing both gas that has not yet been impacted by outflows and wind material that has been ejected at redshifts high enough for it to cool down and to dilute its metal content in the ambient gas.

The BLA selection regime defined by the dashed lines in each panel reveals a population of H I absorbers tracing highly ionized [$\log(n_{\rm H\,I}/n_{\rm H})_{\rm H\,I} \sim -6$] gas with median temperatures $\log(T_{\rm H\,I}/{\rm K}) \gtrsim 4.7$, median (local) metallicities $\log(Z_{\rm H\,I}/{\rm Z}_{\odot}) \gtrsim -2.5$ and total hydrogen column densities $\log(N_{\rm H}/{\rm cm}^{-2}) \approx 18.7$, which is almost an order of magnitude higher than the total hydrogen column density of typical Lyα forest absorbers (see also Fig. 12). According to our previous interpretation of the $T_{\rm H\,I}$–$Z_{\rm H\,I}$ correlation, these results suggest that BLAs may be tracing galactic outflows. We will come back to this point in Section 4.4.2.

The results presented in this section indicate that the H I column density of unsaturated absorbers is a reliable tracer of the underlying physical density of the gas giving rise to the detected H I Lyα absorption. Moreover, the temperature of the absorbing gas may be roughly estimated from the measured linewidth, as suggested by the average contribution of thermal broadening to the total linewidth of these absorbers. Finally, H I absorbers subject to the commonly adopted BLA selection criteria trace gas which appears to be physically distinct from the gas traced by typical Lyα forest absorbers.

4 THE WARM-HOT DIFFUSE GAS

In the next sections, we explore in detail the effect of feedback and metal line cooling on the physical state (i.e. density and temperature) of the gas in our simulations. Also, we investigate the H I absorption signatures of warm-hot diffuse gas, and the physical properties of the gas traced by BLAs identified in synthetic QSO absorption spectra. For this purpose, we use the H I sample from our fiducial model *AGN* presented in Section 3.5, and generate similar samples for all the other model runs. Comparisons between model predictions and observations (whenever possible) are done exclusively for our fiducial run.

4.1 Model dependence of the predicted warm-hot gas mass

On supergalactic scales, two mechanisms are able to shock heat intergalactic gas to temperatures $T \gtrsim 5 \times 10^4$ K: (i) galactic outflows driven by SNII explosions and by AGN activity; (ii) accretion shocks caused by infall on to the potential wells of DM haloes. We have selected four model runs from the OWLS project, *NOSN_NOZCOOL*, *NOZCOOL*, *REF* and *AGN*, to investigate the effect of each of these mechanisms on the properties of diffuse gas and its imprints on simulated absorption spectra. Note that these models have already been described in Section 2 (see also Table 1).

We are interested in the predicted distribution of gas mass among the various (gas) phases – in particular the warm-hot diffuse phase – defined in Table 3. We adopt a temperature threshold at $\log(T/{\rm K}) = 4.7$ (or $T \approx 5 \times 10^4$ K) and a density threshold at $\Delta = 10^2$ to distinguish these gas phases. The density threshold has been chosen so as to roughly separate unbound gas from collapsed structures (at $z = 0.25$). The temperature threshold is motivated by the bimodality in the gas mass distribution predicted by the models considered here (see below). Note that our value is somewhat below the 'canonical' but to some extent arbitrary $T = 10^5$ K commonly adopted to distinguish cool from warm-hot intergalactic gas (but see Wiersma et al. 2010).

Table 3. Definition of the various gas phases we consider in terms of temperature and density thresholds. For example, we define the *warm-hot*, *diffuse* gas (WHIM) to have temperatures $T \geq 5 \times 10^4$ K and overdensities $\Delta \leq 10^2$.

Gas phase	Overdensity (Δ)	Temperature [T (K)]
Cool	–	$<5 \times 10^4$
Warm-hot	–	$\geq 5 \times 10^4$
Diffuse	$\leq 10^2$	–
Condensed	$>10^2$	–
Star forming	$\gtrsim 3 \times 10^5$	(EoS)[a]

[a]We consider 'star forming' the gas with physical densities that exceed our adopted star formation threshold $n_{\rm H}^* = 0.1\,{\rm cm}^{-3}$ and which is allowed to form stars. The temperature of this gas phase is set by an imposed equation of state (EoS) of the form $P \propto \rho^{4/3}$. This gas phase can be thought of as the ISM. Note that, although this gas is cold, it is not included in the gas phase defined as 'cool'.

The distribution of gas mass as a function of temperature and (over)density predicted by the different models is presented in Fig. 7. The coloured areas show the cumulative gas mass (in per cent) indicated by the colour bar to the right-hand side. The vertical (horizontal) solid line in each panel indicates the density (temperature) threshold separating the various phases. Star-forming gas, which is defined as gas with physical densities $n_{\rm H}^* \geq 0.1\,{\rm cm}^{-3}$ (or[7] $\Delta \approx 3 \times 10^5$ at $z = 0.25$), is shown to the right-hand side of the blue, vertical dashed line in each panel. The percentage included in each separate region indicates the baryonic mass in the corresponding (gas) phase relative to the total baryonic mass. In particular, the number (orange) at the top left-hand corner of each panel gives the mass fraction of gas with $T \geq 5 \times 10^4$ K and $\Delta \leq 10^2$, that is, warm-hot diffuse gas. The starred percentage indicates in each case the baryonic mass confined in stars. Note that our adopted temperature threshold seems appropriate to separate cool, photoionized from shock-heated gas; the gas mass distribution at $\Delta \leq 10^2$ is clearly bimodal, with two phases having significant gas mass fractions above and below $\log(T/{\rm K}) = 4.7$. The dotted contours indicate the neutral hydrogen fraction, $(n_{\rm H\,I}/n_{\rm H})$, as a function of density and temperature at $z = 0.25$; it has been computed using precomputed CLOUDY tables as described in Section 2. The logarithmic $(n_{\rm H\,I}/n_{\rm H})$ value is indicated next to the corresponding contour in the top left-hand panel.

The sequence of models given by moving clockwise from the top left-hand panel is essentially a sequence of increasing feedback strength (and model complexity). The mass fraction in warm-hot diffuse gas in the model *NOSN_NOZCOOL* (top left-hand panel) indicates that by $z = 0.25$ roughly 30 per cent of the gas mass is shock heated to temperatures $\log(T/{\rm K}) > 4.7$ by gravity alone. In the absence of any feedback on galactic scales, a large fraction of the gas that is accreted via gravitational infall at higher redshifts is able to cool and fuel star formation, with nearly 15 per cent of the gas mass ending up in stars by $z = 0.25$. The cool, photoionized diffuse gas at $\log(T/{\rm K}) \lesssim 4.7$ and $\Delta \leq 10^2$ in this model contains roughly 40 per cent of the total gas mass.

[7] The relation between hydrogen particle density, $n_{\rm H}$, and (baryonic) overdensity, Δ, is given by equation (B7) in Appendix B.

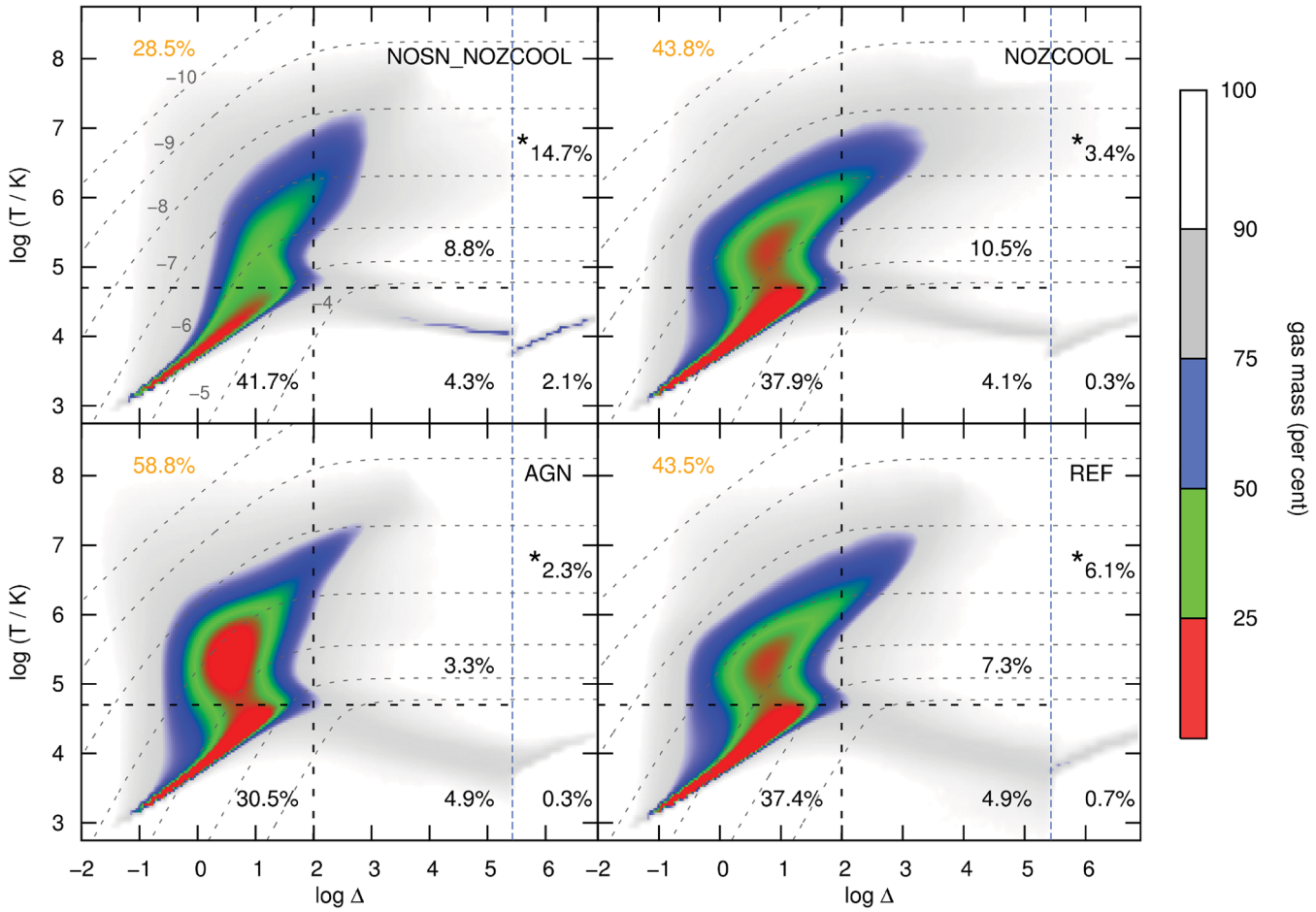

Figure 7. Distribution of gas mass over various phases predicted by different models, indicated on the top right-hand corner of each panel (see Table 1). The vertical (horizontal) dotted line indicates the density (temperature) threshold at $\Delta = 10^2$ ($T = 5 \times 10^4$ K) which separates unbound (cool) from collapsed (warm-hot) gas. The region to the right-hand side of the blue dashed line shows the high-density, star-forming gas with physical densities $n_{\rm H}^* \geq 0.1\,{\rm cm}^{-3}$ (or $\Delta \gtrsim 3 \times 10^5$ at $z = 0.25$). The coloured areas show the cumulative gas mass (in per cent) indicated by the colour bar on the right-hand side. The percentages in each panel show the baryon mass fraction in the corresponding phase defined by the temperature/density thresholds. We have highlighted the baryon mass fraction in the diffuse, warm-hot gas in orange. The starred percentage indicates in each case the baryonic mass fraction in stars. The dotted contours, which are identical in all panels, give the neutral hydrogen fraction ($n_{\rm H\,I}/n_{\rm H}$) of gas in ionization equilibrium as a function of Δ and T; the logarithmic ($n_{\rm H\,I}/n_{\rm H}$) value is indicated next to the corresponding contour in the top left-hand panel. The left-hand panels show the effect of our two most extreme scenarios, that is, including both feedback by SNeII and AGNs with respect to neglecting feedback altogether; the right-hand panels show the effect of neglecting radiative cooling by heavy elements. Clearly, feedback by SNeII (and AGNs) heats a significant amount of gas above temperatures $T = 5 \times 10^4$ K, with the WHIM fraction increasing from 28.5 per cent (top left-hand panel) to 58.8 per cent (bottom left-hand panel). Interestingly, the IGM fraction only changes by ∼10 per cent, indicating that feedback shifts a large fraction of the ISM from haloes into intergalactic space.

Moving on to the top right-hand panel (model *NOZCOOL*), we see that nearly 45 per cent of the baryonic mass in the simulation is in the form of warm-hot diffuse gas as a consequence of the kinetic energy released by SN explosions. This corresponds to an absolute increase in mass of 15 per cent in this gas phase compared to *NOSN_NOZCOOL*. At the same time, the mass fraction in the cool diffuse IGM predicted by *NOZCOOL* decreases with respect to *NOSN_NOZCOOL*, but by a far smaller amount (∼4 per cent in absolute terms). Thus, about 10 per cent of the gas mass that ends up in the WHIM by $z = 0.25$ must be removed from a gas phase other than the cool diffuse IGM. The significantly lower mass in stars in the *NOSN_NOZCOOL* model compared to the *NOZCOOL* model strongly suggests that at higher redshifts SN feedback shock heats and blows a significant fraction of the interstellar medium (ISM) out of haloes, which ends up in the WHIM by $z = 0.25$.

Including radiative cooling by heavy elements (model *REF*; bottom right-hand panel) has a negligible effect on the WHIM and the cool diffuse IGM, suggesting that the metal line cooling in these gas phases is inefficient, either because their density is low, or the metals they contain are not yet well mixed, or perhaps because the level of enrichment is low, or a combination of them all. Interestingly, the *REF* model predicts a much higher mass fraction in stars and star-forming gas with respect to *NOZCOOL*, in spite of including SN feedback. The corresponding decrease in mass in warm-hot gas at high densities (which can be considered as the intragroup and intracluster medium, hereafter ICM) suggests that some of the gas in this phase is accreted and fuels star formation. However, the exact evolutionary path of the gas in T–Δ phase space might be more complex than this.

Perhaps the most remarkable result is the fact that feedback from AGNs has a very strong impact on the thermal state of the diffuse gas. Indeed, a comparison of the bottom panels shows that an additional ∼15 per cent of the total gas mass in the simulation is shock heated to temperatures above $\log(T/{\rm K}) = 4.7$ and pushed

into regions of low density ($\Delta < 10^2$), such that by $z = 0.25$ nearly 60 per cent of the gas mass ends up in the WHIM. A comparison of the mass distributions among the different phases predicted by *REF* and *AGN* suggests that half of the additional WHIM mass, that is, ∼7 per cent, at $z = 0.25$ is removed at higher redshifts mainly from the ISM (thus reducing the mass in stars at $z = 0.25$ by a factor of ∼3), and from the ICM. The remaining ∼7 per cent of the WHIM mass apparently comes from the IGM. Comparison of the gas mass fractions in the warm-hot diffuse phase between the models *NOZCOOL* and *AGN* suggests that SNe and AGNs contribute with roughly a similar amount to the baryon content of the WHIM. Equally important, the gas mass in this gas phase predicted by the models *NOSN_NOZCOOL* (∼30 per cent) and *AGN* (∼60 per cent) indicates that (strong) feedback (both by SNe and by AGNs) may be able to shock heat an amount of gas comparable to the gas shock heated via gravitational infall. These results thus indicate that it is crucial to understand feedback processes on supergalactic scales before being able to make any reliable predictions about the baryon content of warm-hot gas in the Universe.

Consider finally the hydrogen neutral fraction ($n_{\rm H\,I}/n_{\rm H}$) indicated by the dotted contours. The logarithmic value of ($n_{\rm H\,I}/n_{\rm H}$) is indicated next to the corresponding contour in the top left-hand panel, and they are identical in all the other panels. At a fixed temperature, the neutral hydrogen fraction increases with density, since the ionization state of the gas is dominated by PI. However, at sufficiently high densities, that is, $\Delta \gtrsim 10^2$ at $z = 0.25$, collisional ionization dominates and the neutral hydrogen fraction depends only on the gas temperature, resulting in contours running parallel to the Δ-axis. In either case, the neutral hydrogen fraction steeply decreases with temperature at all densities (see also Richter et al. 2008; Danforth et al. 2010). As a consequence, the gas at densities and temperatures characteristic of the warm-hot diffuse gas is expected to be highly ionized. In particular, the model *AGN* predicts that the vast majority of the gas in the WHIM contains a neutral hydrogen fraction ($n_{\rm H\,I}/n_{\rm H}$) $\lesssim 10^{-6}$. This has important implications for the detectability of this gas phase via H I absorption which will be discussed in detail in Section 4.2.1.

4.2 Observability of (warm-hot) gas using (broad) H I absorption

In this section, we explore to what extent the actual gas mass (distribution) in the cool and warm-hot diffuse phases is traced by the H I detected in absorption. In particular, we investigate the thermal state of the gas traced by absorbers selected in terms of their linewidth, albeit only on a statistical basis. At the same time, we invert the approach and inquire about the spectral signatures (and physical properties) of H I absorbers arising in warm-hot gas. Even though we are interested primarily in broad absorbers, we include narrow H I Lyα absorbers (NLAs) in our analysis as well. This allows for a more robust interpretation of our results. We define the following classes of H I absorbers in terms of their spectral and/or physical properties:

(i) *NLA:* H I absorber components with Doppler parameters $b_{\rm H\,I} < 40\,{\rm km\,s^{-1}}$. We adopt the notation introduced by Lehner et al. (2007).

(ii) *BLA:* H I absorber components with Doppler parameters $b_{\rm H\,I} \geq 40\,{\rm km\,s^{-1}}$ that satisfy the sensitivity limit introduced by Richter et al. (2006a) and adopted in other studies (Danforth et al. 2010; Williger et al. 2010):

$$\left(\frac{N_{\rm H\,I}/{\rm cm^{-2}}}{b_{\rm H\,I}/{\rm km\,s^{-1}}}\right) \geq 3 \times 10^{12}\,({\rm S/N})^{-1}\ .$$

This limit is equivalent to an H I Lyα central optical depth

$$\tau_0 \geq 2.27\,({\rm S/N})^{-1}\ . \qquad (2)$$

We feel that using a detection limit in terms of τ_0 is more intuitive than the limit in terms of ($N_{\rm H\,I}/b_{\rm H\,I}$), in particular for small values of τ_0, since in this limit $\tau(v) \approx 1 - F(v)$. Henceforth, we will use equation (2) instead of the limit in terms of ($N_{\rm H\,I}/b_{\rm H\,I}$) as our second BLA selection criterion; also, all corresponding results will be expressed in terms of τ_0 rather than ($N_{\rm H\,I}/b_{\rm H\,I}$).

(iii) *Hot BLA:* BLAs with optical depth weighted temperatures $T_{\rm H\,I} \geq 5 \times 10^4$ K. This class is defined in order to isolate BLAs genuinely tracing warm-hot gas. Note that our adopted temperature threshold is lower than the actual temperature implied by Doppler parameters $b_{\rm H\,I} \geq 40\,{\rm km\,s^{-1}}$ assuming pure thermal broadening.[8] This is, however, not an issue since, as we have shown in Section 3.5.2, the linewidth of BLAs is never entirely set by thermal broadening.

Table 4. Number fraction (in per cent) of simple absorbers relative to the total number of H I components in each class identified in synthetic spectra at $z = 0.25$ with S/N = 50 for different models.

	NOSN_NOZCOOL	*NOZCOOL*	*REF*	*AGN*
All[a]	43	44	44	47
NLA[b]	37	37	37	38
BLA	4.5	5.5	5.6	7.1
Hot BLA	1.0	1.7	1.7	3.1

[a]Note that about half of the identified absorbers are single component, irrespective of the model.
[b]Note that the fraction of NLAs is similar for all models, suggesting that the gas traced by these absorbers is not significantly affected by feedback.

As in previous sections, all the results presented here refer to *simple* absorbers as defined in Section 3.5, unless stated otherwise. Table 4 gives an overview of the statistics of simple absorbers in each of the classes defined above, for all four models considered here. An example of such a simple (broad) H I absorber is shown in Fig. 8. The top panel shows the spectrum (black) and corresponding fit (red) centred at an H I Lyα line with $b_{\rm H\,I} = 63\,{\rm km\,s^{-1}}$ and $N_{\rm H\,I} \sim 10^{13}$ cm^{-2}. The next panel shows the smoothed overdensity (black) and the H I optical depth weighted overdensity (red dashed) along the sightline. Note the relative simple density structure of the absorbing gas. The smoothed (black) and H I optical depth weighted (red dashed) temperatures are shown in the third panel (from the top). Clearly, the gas giving rise to the BLA shown in the top panel has a temperature $T \gtrsim 10^5$ K. As a consequence of the high temperature, the absorbing gas is highly ionized, with a total hydrogen column density $N_{\rm H} \approx 4 \times 10^{19}$ cm^{-2}, and thus it represents a significant baryon reservoir. The slight offset between the BLA and the density peak of the absorbing gas is due to its (small) peculiar velocity along the sightline (bottom panel).

[8] The thermal width of an H I line arising in gas at $T \gtrsim 5 \times 10^4$ K is $b_{\rm H\,I} \gtrsim 30\,{\rm km\,s^{-1}}$.

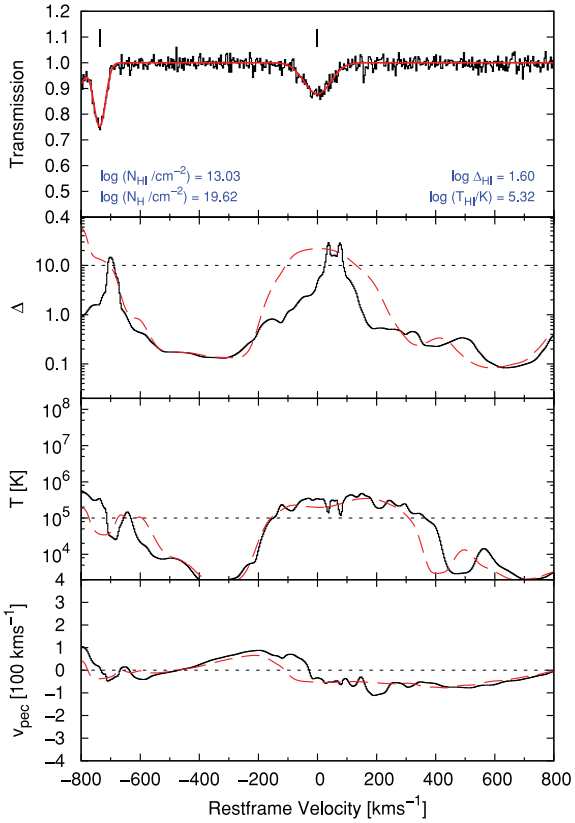

Figure 8. Example of a sightline through a simulation box at $z = 0.25$ from a run adopting our fiducial model (*AGN*). The top panel shows a synthetic spectrum generated by SPECWIZARD using S/N = 50 (black) and the corresponding fit (red) centred on a simple BLA with $b_{\rm HI} \approx 63\,{\rm km\,s^{-1}}$, and showing an additional narrow H I Lyα absorption line (vertical marks flag the line centres). The subsequent panels show, from the top to bottom, the gas overdensity, gas temperature and peculiar velocity along the sightline. The black solid lines indicate smoothed quantities; the red dashed lines indicate H I optical depth weighted quantities. The dotted horizontal line in the second panel (from the top) indicates the detection limit at S/N = 50 (see equation A3) expressed in terms of Δ using the $N_{\rm HI}(\Delta)$ relation discussed in Section 3.5.1. The values in the lower left-hand and lower right-hand corner of the top panel give, respectively, the H I column density and the H I optical depth weighted total hydrogen column density, and the overdensity and the temperature of the gas traced by the BLA. Note that $\log(T/{\rm K}) = 5.32$ corresponds to a thermal width $b_T = 59.0\,{\rm km\,s^{-1}}$, implying that the width of this BLA is dominated by thermal broadening.

4.2.1 *Spectral sensitivity*

Before we investigate the physical and statistical properties of BLAs in our simulations, we need to assess how setting a fixed sensitivity limit as given by equation (2) may bias the detection of warm-hot gas. Under rather simple assumptions, it is possible to model the H I Lyα central optical depth, τ_0, of H I absorbing gas as a function of its temperature and density. This allows one to put constraints on the physical state of the gas phase traced using H I absorption, given a set of instrumental limitations that lead to a minimum detection (or sensitivity) limit. Conversely, following this approach it is possible to estimate the sensitivity needed to detect gas at a given temperature and density. We describe the basic assumptions and give a detailed calculation of our model in Appendix B. In particular, we show that it depends on the assumed size of the absorbing structure. With no better estimate at hand, we assume the absorbers to have a linear size[9] of the order of the local Jeans length (Schaye 2001, see also equation B5). Note that we have already showed in Section 3.5.1 that this assumption can account for the $N_{\rm HI}(\Delta)$ relation predicted by the simulations. Also, our model neglects non-thermal broadening, implying that all sensitivities in terms of τ_0 given henceforth are strict lower limits.

We now investigate the relation between the gas mass distribution in our simulations and the gas mass detected in H I absorption. For each of the BLAs in the line sample obtained for each model considered here, we estimate $\Delta_{\rm HI}$ and $T_{\rm HI}$, and plot the resulting distribution on the $T_{\rm HI}$–$\Delta_{\rm HI}$ plane. The result is shown in Fig. 9. The coloured areas indicate the cumulative number fraction (in per cent) of BLAs at a given density and temperature. The grey solid contours correspond in each case to the gas mass distribution shown in Fig. 7. Note that for the contours the axes indicate the actual gas overdensity and gas temperature. We plot in each panel a series of black dashed contours which indicate the central optical depth, τ_0, as a function of $\Delta_{\rm HI}$ and $T_{\rm HI}$, as given by equation (B6). The corresponding contour values are indicated next to each curve only in the top left-hand panel, but they are identical for all the other panels. Note in particular the thick dashed contour (magenta) which indicates our adopted sensitivity limit as given by equation (2) for S/N = 50, that is, $\log\tau_0 = -1.34$ [or $\log(N_{\rm HI}/b_{\rm HI}) = 10.8$].

One notable feature in this figure is the bimodality of the distribution of the gas traced by broad H I absorbers, irrespective of the model. We see in each case a population of BLAs tracing gas at low temperatures ($T_{\rm HI} < 5 \times 10^4$ K) and overdensities $\log\Delta_{\rm HI} < 0.5$, and a second population tracing warm-hot gas at $T_{\rm HI} > 5 \times 10^4$ K and overdensities $\log\Delta_{\rm HI} > 0.5$. Note, however, that the amplitude of the distribution varies from model to model. Comparing the grey contours to the coloured distribution we see clearly that the H I detected in absorption traces only a fraction of the gas mass in the simulations. In particular, the model *AGN* (bottom left-hand panel) shows a large fraction of gas mass at $10^5 \lesssim T_{\rm HI} \lesssim 3 \times 10^5$ K and $\log\Delta_{\rm HI} \sim 0.5$ which is not detected in H I absorption. The same is true for the models *NOZCOOL* and *REF*, although at slightly different temperatures and overdensities. Consideration of the thick dashed contour reveals that this is a selection effect, that is, the gas is simply not detectable at our adopted sensitivity limit. As discussed above, this comes about because the gas at such high temperatures and relative low densities is extremely ionized, with neutral hydrogen fractions $\log(n_{\rm HI}/n_{\rm H}) \lesssim -6$, and its absorption simply falls below our adopted detection threshold (cf. Fig. B1).

Thus, in our fiducial model BLAs detected in spectra with S/N = 50 typically have $-2 < \log\tau_0 < 0$, while the bulk of the gas mass at $10^5 \lesssim T_{\rm HI} \lesssim 3 \times 10^5$ K and $\log\Delta_{\rm HI} \sim 0.5$ is predicted to give rise to absorption with $\log\tau_0 < -2$. The fact that we do detect in absorption some of the gas at temperatures and densities which correspond to sensitivities slightly smaller than our adopted value (i.e. to the left-hand side of the thick dashed contour) reflects the simplicity of the assumptions that go into modelling the absorption strength in terms of Δ and T. Nevertheless, the expected and actual detections are fairly consistent with each other. Based on this, we estimate that in order to detect most of the baryonic mass in the WHIM using thermally broadened H I absorption, spectra with very high S/N are required that allow detection at the $\log\tau_0 \sim -2$ level, which is roughly an order of magnitude lower than the typical

[9] Note that the simulation runs we use do resolve the Jeans length, in particular at the relatively low densities considered here.

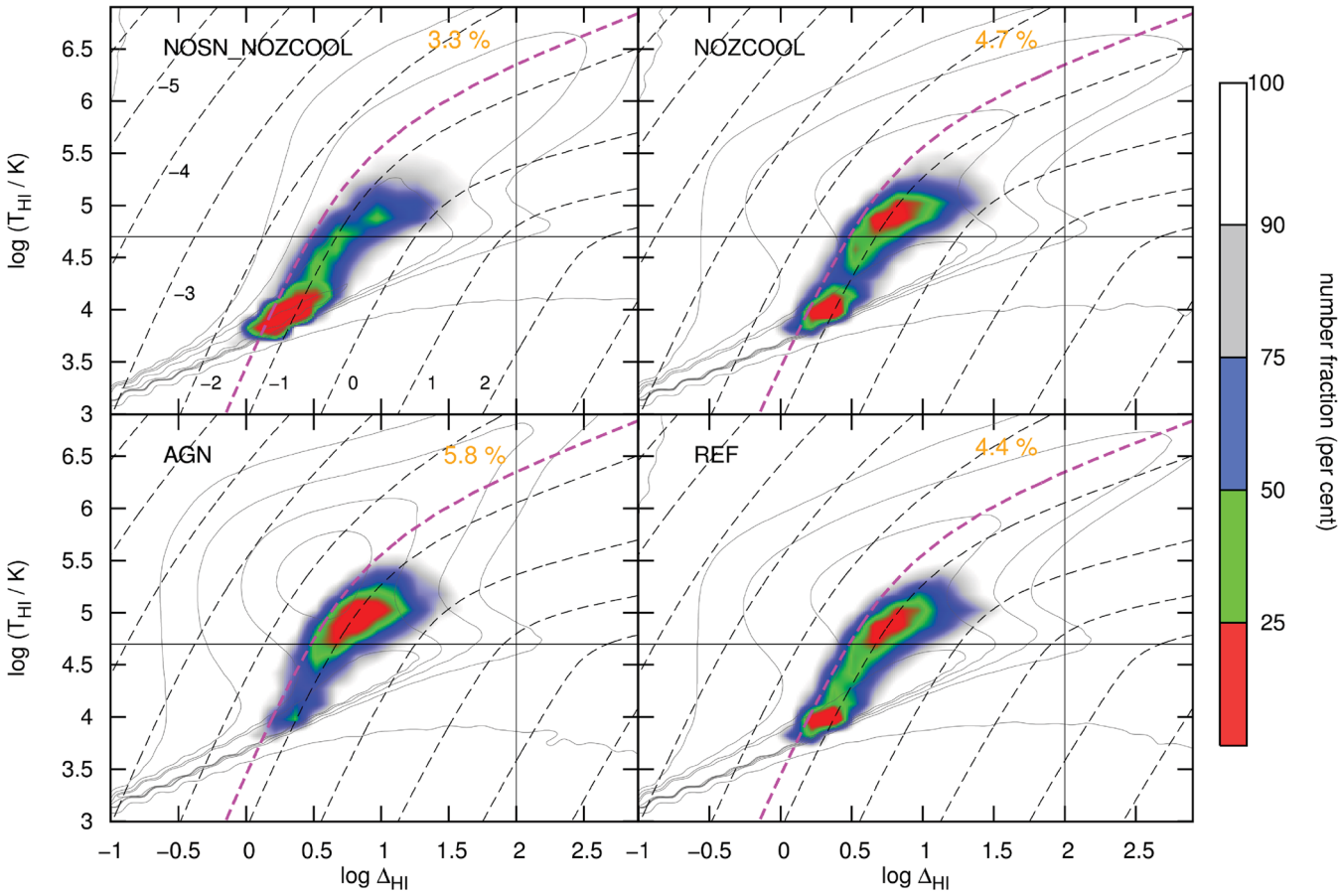

Figure 9. Distribution of temperature and overdensities of the gas traced by simple BLAs (coloured regions) identified in spectra with S/N = 50 at $z = 0.25$ obtained from various models. In each panel, the coloured regions show the cumulative number fraction (in per cent) of BLAs. For instance, the red region contains 25 per cent of the total number of BLAs, while the green and red regions together contain half of them. The grey contours correspond to the gas mass distributions shown in Fig. 7. Note the change in scale on both the *x*- and *y*-axes with respect to Fig. 7. The vertical (horizontal) solid line in each panel indicates the density (temperature) threshold adopted to distinguish the different gas phases (see Table 3). The dashed contours, which are identical in all panels, indicate the central optical depth, τ_0, as a function of (over)density and temperature (see Appendix B, equation B6). For clarity, the corresponding logarithmic value of τ_0 is included next to each curve only in the top left-hand panel. The magenta dashed contour indicates our adopted sensitivity limit for S/N = 50, $\log \tau_0 = -1.34$ (equation 2). The percentage (orange) in each panel indicates the fraction of baryonic mass of the gas traced by simple BLAs with $T_{\rm H\,I} \geq 5 \times 10^4$ K, that is, simple hot BLAs. Reassuringly, the overwhelmingly majority of the identified BLAs lie below the thick magenta contour, irrespective of the model. However, the large fraction of mass in gas at $T \gtrsim 10^5$ K and $\Delta \lesssim 10$ (in particular for the model *AGN*) is not observable at our adopted sensitivity. The dashed contours indicate that at least $\log \tau_0 \sim -2$ is required. BLAs selected in terms of their width and equation (2) thus only trace the low-temperature regime of the warm-hot diffuse gas, since the gas at higher temperatures is not observable, probably due to its high ionization degree (see Fig. B1). Note that the estimated baryon content of the absorbing gas in each case corresponds to ∼10 per cent of the actual baryonic mass in this phase for the corresponding model (cf. Fig. 7).

sensitivities adopted in BLA studies (Richter et al. 2006a; Danforth et al. 2010; Williger et al. 2010).

4.3 BLA number density

Since our adopted sensitivity limit matches the value used to identify BLA candidates in real QSO spectra, we may directly compare the predicted and observed line frequencies. Applying the selection criteria described above [i.e. $b_{\rm H\,I} \geq 40\,{\rm km\,s^{-1}}$ and $\tau_0 \geq 2.27({\rm S/N})^{-1}$] to our *AGN* H I line sample obtained from spectra with S/N = 50 at $z = 0.25$ results in 6120 BLA candidates, which corresponds to a line number density $({\rm d}N/{\rm d}z)_{\rm BLA} = 33 \pm 6$, where the quoted uncertainty is pure Poissonian. The number densities of BLAs identified in spectra with S/N = 30 and 10 are given in Table 5. For completeness, we also include in this table the corresponding numbers

Table 5. Line number density, (d*N*/d*z*), and corresponding Poisson uncertainties for BLAs and NLAs identified in spectra with various S/N values from our fiducial run at $z = 0.25$.

	S/N = 50	S/N = 30	S/N = 10
All[a]	460 ± 21	332 ± 18	144 ± 12
Simple	214 ± 15	175 ± 13	94 ± 10
NLA (simple)	173 ± 13	140 ± 12	73 ± 9[b]
NLA (all)	343 ± 19	248 ± 16	110 ± 10[b]
BLA (simple)	33 ± 6	28 ± 5	15 ± 4
BLA (all)	95 ± 10	66 ± 8	25 ± 5

[a]This corresponds to simple and complex H I absorbers taken together.
[b]For reference, Lehner et al. (2007) find a mean $({\rm d}N/{\rm d}z)_{\rm NLA} = 66 \pm 17$ over seven sightlines for *all* NLAs in their data with an average S/N ≈ 10, discarding lines with associated Voigt-profile parameter errors larger than 40 per cent.

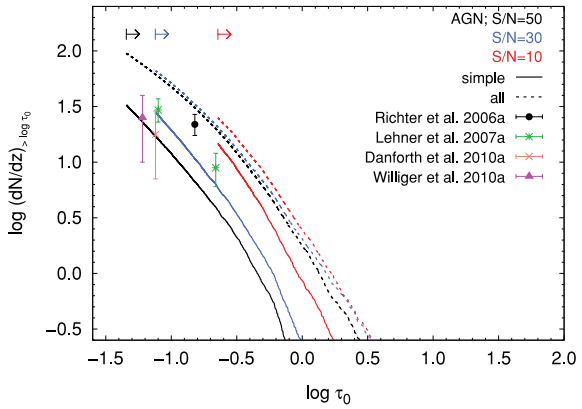

Figure 10. Cumulative number density of BLAs as a function of central optical depth, τ_0. Measurements available in the literature are plotted along with the predictions of our fiducial run *AGN* at $z = 0.25$ (see text for details). The solid lines show the result for simple BLAs identified in 5000 spectra with S/N = 50 (black), 30 (blue) and 10 (red). We include our results for all (i.e. simple *and* complex) BLA candidates as well (dashed lines). The arrows indicate the BLA sensitivity limit in terms of τ_0 for the corresponding S/N value as given by equation (2).

for NLAs. Note that the absolute number of BLAs decreases while the relative number of simple BLAs increases with decreasing S/N. This is due to a low S/N causing line blending and increasing the linewidths, as discussed in Section 3.3 (see also Richter et al. 2006a, their fig. 2).

In Fig. 10, we compare the cumulative BLA number density as a function of τ_0 predicted by our fiducial run obtained from spectra with various S/N values to the available observational results obtained from QSO spectra at comparable redshifts and with similar (average) S/N. The arrows indicate the BLA sensitivity limit in terms of τ_0 for the corresponding S/N value as given by equation (2). The predictions for simple BLAs are indicated by the solid lines. Since our definition of simple absorbers is somewhat arbitrary, we include the corresponding predictions for all, that is, simple and complex, BLA candidates as well (dashed lines). These two sets of lines thus span our predicted line frequency range for S/N $\in$ [10, 50].

It is noteworthy that our predictions are broadly consistent with the observed range of BLA number densities. For example, our prediction for S/N = 30 (blue solid) agrees with the results by Lehner et al. (2007, green data points). These authors find in their data with an average S/N $\approx$ 15, discarding lines with associated line parameter errors larger than 40 per cent, a fraction of single-component BLAs close to 30 per cent and a mean (averaged over seven sightlines) line number density $(\mathrm{d}N/\mathrm{d}z)_{\rm BLA} = 30 \pm 7$ at $\log\tau_0 \gtrsim -1.10$ [equivalent to $\log(N_{\rm H\,I}/b_{\rm H\,I}) \gtrsim 11.02$], and $(\mathrm{d}N/\mathrm{d}z)_{\rm BLA} = 9 \pm 3$ at $\log\tau_0 \gtrsim -0.66$ [or $\log(N_{\rm H\,I}/b_{\rm H\,I}) \gtrsim 11.46$]. Both our S/N = 50 (black) and 30 (blue) predictions match the result by Danforth et al. (2010, orange data point), who find $(\mathrm{d}N/\mathrm{d}z)_{\rm BLA} = 18 \pm 11$ at $\log\tau_0 \gtrsim -1.12$ [$\log(N_{\rm H\,I}/b_{\rm H\,I}) \gtrsim 11.0$] along seven sightlines at $z \lesssim 0.5$, spanning a total redshift path $\Delta z = 2.193$, with S/N $\geq$ 5. Their sample consists of 15 single-component BLAs and 48 BLAs with uncertain velocity structure. Williger et al. (2010, magenta data point) estimate their detection limit at $\log\tau_0 \gtrsim -1.22$ [or $\log(N_{\rm H\,I}/b_{\rm H\,I}) \gtrsim 10.9$] and obtain $(\mathrm{d}N/\mathrm{d}z)_{\rm BLA} = 25 \pm 15$ and 8.5 ± 8.5 for their full and single-component BLA samples, respectively, along a single sightline ($\Delta z = 0.117$) and corresponding spectrum with S/N = 20–30. Finally, Richter et al. (2006a, black data point) measure $(\mathrm{d}N/\mathrm{d}z)_{\rm BLA} = 22$ using their reliable sample of single-component BLAs, detected along four sightlines at $z \lesssim 0.4$ ($\Delta z = 0.928$), at a sensitivity $\log\tau_0 \gtrsim -0.82$ [or $\log(N_{\rm H\,I}/b_{\rm H\,I}) \gtrsim 11.3$], in spectra with an average S/N $\approx$ 15.

4.4 Physical properties of broad H I absorbers

In Section 3, we discussed the relation between the physical conditions of the gas traced by typical H I absorbers and their line observables ($N_{\rm H\,I}$, $b_{\rm H\,I}$) using our fiducial model. We now focus on the physical conditions of the gas giving rise to H I absorbers identified as simple BLAs in our synthetic spectra at $z = 0.25$; these correspond to single-component H I absorbers falling within the region defined by the polygon in Fig. 6 which is defined through two criteria: (i) the linewidth satisfies $b_{\rm H\,I} \geq 40\,\mathrm{km\,s^{-1}}$; and (ii) the central optical depth obeys equation (2), that is, $\tau_0 \geq 2.27(\mathrm{S/N})^{-1}$.

Given the importance of broad H I absorbers as potential WHIM tracers, and the dependence of the predicted WHIM mass fraction on the adopted physical model, we next explore the relation between the measured linewidth and the temperature of the absorbing gas using the models introduced in Section 2. We further use our fiducial model (*AGN*) to investigate the ionization state, neutral hydrogen fraction and metallicity of the gas traced by different classes of H I absorbers. Finally, we estimate the baryon content of the gas traced by BLAs using all models, but compare our results to observations using only the model *AGN*.

4.4.1 Temperature distribution of (broad) lines

In Fig. 11, we show the distribution of temperatures of the gas traced by both BLAs (red histograms) and NLAs (blue histograms) identified in our spectra with S/N = 50 at $z = 0.25$ obtained from different models. The dashed, vertical line in each panel indicates the temperature threshold adopted to separate cool from warm-hot gas (see Table 3).

In all models, the temperature distribution of the gas traced by NLAs shows that lines with Doppler parameters $b_{\rm H\,I} < 40\,\mathrm{km\,s^{-1}}$ preferentially arise in gas at temperatures $T_{\rm H\,I} \lesssim 5 \times 10^4\,\mathrm{K}$ with a peak at $T_{\rm H\,I} \sim 10^4\,\mathrm{K}$, as expected from their width. The few lines which are narrower than allowed by the temperature of the absorbing gas (i.e. the section of the blue histogram to the right-hand side of the vertical dashed line) are mostly weak lines that fall below the formal completeness limit, as discussed in Section 3.5.2. However, some of these lines are real detections, which suggest that gas at different temperatures overlaps in velocity space (due to redshift-space distortions), and some of it may even overlap in position space, indicating the existence of multiphase absorbing structures. In models with feedback (*NOZCOOL*, *REF*, *AGN*), the temperature distribution of the gas traced by NLAs is broader than in the model without feedback (*NOSN_NOZCOOL*), and the fraction of lines with 'unphysical' widths is higher. This can be explained as follows. Outflows driven by SNe and AGNs follow the path of least resistance in space, thus escaping into the voids while leaving the cooler, denser filaments intact (Theuns et al. 2002). With increasing feedback strength, the cross-section of such high-temperature outflows increases as well, and so does the chance for a random sightline to intersect both cool, dense filaments and shock-heated material, with their corresponding absorption overlapping from time to time along the spectrum.

Note that the NLA temperature distributions in the models *NOZCOOL* and *REF* are very similar to each other, and the same is true for the corresponding BLA temperature distributions. Moreover, the fraction of hot BLAs is only slightly lower in *REF* than in

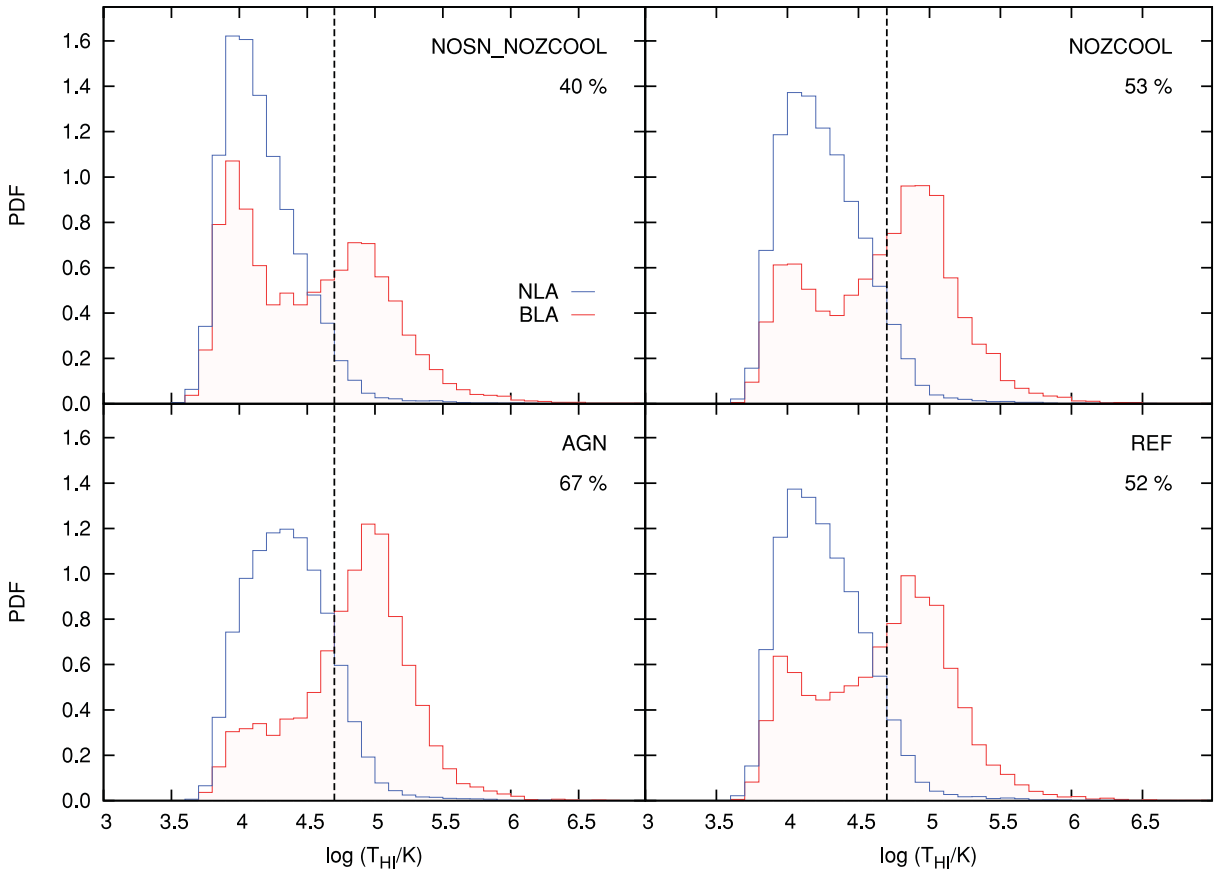

Figure 11. Temperature distribution of the gas traced by simple H I absorbers identified in spectra at $z = 0.25$ with S/N = 50 obtained from different models. The blue histograms show in each case the distribution for the gas traced by NLAs (i.e. $b_{\rm H\,I} < 40\,{\rm km\,s^{-1}}$), while the red histograms indicate the temperature distribution of the gas traced by BLAs. The percentage shown in each panel indicates the number fraction of BLAs arising in gas at $T_{\rm H\,I} \geq 5 \times 10^4$ K, that is, hot BLAs. Note the correspondence between the increasing mass fraction of warm-hot diffuse gas in Fig. 7 and the increasing fraction of BLAs tracing warm-hot gas shown here. The distribution predicted by our fiducial run (*AGN*) suggests that, statistically, two out of three BLA candidates trace gas at $T \gtrsim 5 \times 10^4$ K.

NOZCOOL. This indicates that metal line cooling is of secondary importance in setting the thermal state of the gas phase traced by (broad) H I absorbers.

In contrast to NLAs, broad H I lines trace gas around two different temperatures, $T_{\rm H\,I} \sim 10^4$ K and $T_{\rm H\,I} \sim 10^5$ K, irrespective of the model (red histograms). Clearly, BLAs arising in gas at low temperatures must be subject to substantial non-thermal broadening, such as bulk flows and/or Hubble broadening.[10] Since the linewidth of an H I absorber with $b_{\rm H\,I} = 40\,{\rm km\,s^{-1}}$ arising in gas at $T \sim 10^4$ K is completely dominated by non-thermal broadening, its linear size (assuming the linewidth is entirely due to Hubble broadening) must be ∼500 kpc, which is consistent with the Jeans length of a filament with a mean density $n_{\rm H} \sim 10^{-6}\,{\rm cm^{-3}}$ and temperature $T \sim 10^4$ K (equation B5).

Interestingly, the fraction of BLAs tracing gas at low ($T_{\rm H\,I} \sim 10^4$ K) and high ($T_{\rm H\,I} \sim 10^5$ K) temperatures is very different in each model, as can be judged qualitatively by the shape of the corresponding histograms and quantitatively by the percentage included in each panel which gives the fraction of hot BLAs, that is, BLAs tracing gas with $T \geq 5 \times 10^4$ K. In fact, the ratio of BLAs tracing cool gas to those tracing warm-hot gas appears to be very sensitive to feedback strength. In principle, this could be used to constrain feedback models observationally. The caveat is that a statistically significant sample of confirmed BLAs would be required for which the gas temperature can be measured reliably. In our fiducial model (*AGN*), which includes the strongest feedback, the majority of the BLAs trace gas at $T_{\rm H\,I} \sim 10^5$ K, with little contamination by non-thermally broadened lines. In fact, two out of three BLA candidates arise in gas at temperatures $T \geq 5 \times 10^4$ K.

The bimodal character of the gas temperature distributions for BLAs predicted by the model *NOSN_NOZCOOL* is also consistent with previous results. Richter et al. (2006b) find in a simulation that included a model similar to our *NOSN_NOZCOOL* that ∼30 per cent of the BLAs trace gas at $T < 2 \times 10^4$ K, and a significant fraction trace gas at $T_{\rm H\,I} \gtrsim 5 \times 10^4$ K. The quantitative difference between theirs and our result for *NOSN_NOZCOOL* is probably caused by the difference in the simulation methods used. Using a linear model to fit the thermal to total linewidth, Richter et al. (2006b) find that $b_T/b_{\rm H\,I} = 0.91$ (no error quoted). We note that we do not find such a tight correlation between b_T and $b_{\rm H\,I}$, but if we perform a linear fit between these two quantities, we find[11] $b_T/b_{\rm H\,I} = 0.641 \pm 0.004$. This is consistent with the result presented

[10] Note that our simulations lack the resolution to capture small-scale turbulence within the gas.

[11] For reference, the corresponding results for S/N = 30 and 10 are $b_T/b_{\rm H\,I} = 0.622 \pm 0.04$ and 0.556 ± 0.006, respectively.

in Section 3.5.2 that thermal broadening on average contributes with (at least) 60 per cent to the total linewidth of BLAs.

Lehner et al. (2007) argue that broad H I lines may trace both cool and warm-hot gas, but that the majority of BLAs trace gas at $T \sim 10^5$–10^6 K if their width is dominated by thermal broadening. As mentioned in the previous paragraph, thermal broadening accounts for a significant fraction to the total linewidth of single-component, broad H I absorbers. Thus, our simulations are consistent with the result inferred from observations that these absorbers do preferentially trace gas at high temperatures, at least in models with (some type of) feedback.

In summary, we find that in the absence of feedback BLA samples are contaminated by a large fraction of non-thermally broadened lines. Conversely, the fraction of broad H I absorption lines tracing gas at temperatures $T_{\rm H\,I} \sim 10^5$ K increases when feedback is included. For instance, our fiducial model predicts that, in a statistical significant sample, 67 per cent of the BLAs trace gas at $T_{\rm H\,I} \gtrsim 5 \times 10^4$ K. Our results thus strongly support the idea that reliable BLAs detected in real absorption spectra are genuine tracers of gas at such high temperatures.

4.4.2 Neutral hydrogen fraction, total hydrogen column density and metallicity

In Fig. 12, we show the distribution of neutral hydrogen fraction (left-hand panel), total hydrogen column density (middle panel) and (local) metallicity (right-hand panel) of the gas traced by NLAs (blue), BLAs (red) and hot BLAs (orange) identified in spectra with S/N = 50 obtained from our fiducial model. The vertical dashed line indicates in each case the corresponding median value.

In general terms, the physical properties of the gas traced by BLAs and hot BLAs show similar distributions and comparable median values, but they are somewhat different from the corresponding properties of the gas traced by NLAs. For example, the median neutral hydrogen fraction of the gas traced by NLAs is $\log(n_{\rm H\,I}/n_{\rm H}) \sim -5.5$, which is slightly higher than the neutral hydrogen fraction of the gas traced by (hot) BLAs, $\log(n_{\rm H\,I}/n_{\rm H}) \sim -6$. This is expected since, as we have shown previously, the temperature of gas giving rise to broad H I absorption is, on average, higher than the temperature of gas giving rise to NLAs. Furthermore, the median total hydrogen column density of gas detected via (hot) BLAs is $N_{\rm H} \sim 6 \times 10^{18}$ cm^{-2}, which is several times larger than the median total hydrogen column density of the gas traced by NLAs, $N_{\rm H} \sim 2 \times 10^{18}$ cm^{-2}, and its distribution extends out to significantly larger values, $N_{\rm H} \sim 10^{20}$ cm^{-2}, as compared to $N_{\rm H} \sim 10^{19}$ cm^{-2}. This is due to several factors. First, as shown in Figs B1 and 9, high-temperature gas detected via H I absorption at a fixed sensitivity necessarily has a higher density with respect to gas at lower temperatures. Also, a higher density implies an average higher $N_{\rm H\,I}$ as a consequence of the $\Delta_{\rm H\,I}$–$N_{\rm H\,I}$ correlation. Finally, gas at high temperature has a lower neutral hydrogen fraction, which in turn yields higher total hydrogen column densities for a given $N_{\rm H\,I}$. The high(er) total hydrogen density of the gas traced by broad H I absorbers implies that its baryon content is considerable. We will come back to this point in more detail in Section 4.5.

Quite interesting is the difference between the gas metallicity distributions. While the metallicity of the gas traced by NLAs shows a broad distribution with a tail extending to very low values and a median $Z \sim 10^{-3}\,{\rm Z}_\odot$, the metallicity distribution of the gas traced by broad absorbers is narrow, with most values falling in the range $(0.001, 1)\,{\rm Z}_\odot$, centred around $Z \sim 10^{-2}\,{\rm Z}_\odot$. On average, the metallicity of the gas traced by (hot) BLAs exceeds the metallicity of the gas traced by NLAs by an order of magnitude.

These results together indicate that broad H I absorbers trace gas that is physically distinct from the gas traced by NLAs, as already mentioned in Section 3.5.3. In particular, the relatively high level of enrichment is inconsistent with the idea that BLAs trace primordial gas that is sinking along filaments towards the centre of high-density regions, as commonly assumed. Rather, our results suggest that broad H I absorbers may be tracing recent (or ongoing) galactic outflows, and/or gravitationally shock-heated gas that has been enriched by galactic ejecta at early epochs.

4.5 Baryon content of H I absorbing gas

In this section, we briefly investigate the dependency of the predicted baryon fraction of the gas traced by H I absorbers on the adopted physical model. As we have done in Paper I for O VI absorbers, we estimate the baryon fraction, that is, the baryon density relative to the critical density, $\rho_{\rm c}$, in H I absorbers using

$$\Omega_{\rm b}({\rm H\,I}) = \frac{m_{\rm H}}{\rho_{\rm c}} \left(\frac{c}{H_0} \sum_{i=1}^{N_{\rm LOS}} \Delta\chi_i \right)^{-1} \sum_{i=1}^{N_{\rm LOS}} \sum_{j=1}^{N_{\rm abs}} \frac{N_{\rm H\,I}}{(X_{\rm H})_{\rm H\,I}\ (n_{\rm H\,I}/n_{\rm H})_{\rm H\,I}}, \quad (3)$$

where $m_{\rm H}$ is the hydrogen mass, and $X_{\rm H}$ and $(n_{\rm H\,I}/n_{\rm H})_{\rm H\,I}$ are the optical depth weighted hydrogen mass fraction and neutral hydrogen fraction, respectively. Note that $N_{\rm H\,I}$, $(X_{\rm H})_{\rm H\,I}$ and $(n_{\rm H\,I}/n_{\rm H})_{\rm H\,I}$ are computed for *individual* absorbing components along each sightline, but we have omitted the running indices for simplicity.

The top panel of Fig. 13 shows the baryonic mass fractions, $\Omega_{\rm b}({\rm H\,I})/\Omega_{\rm b}$, predicted by various models in different types of absorbers: NLAs (blue squares), BLAs (red filled circles) and hot BLAs (orange open circles), where each of these classes has been subdivided into simple (solid lines) and complex (dotted lines)

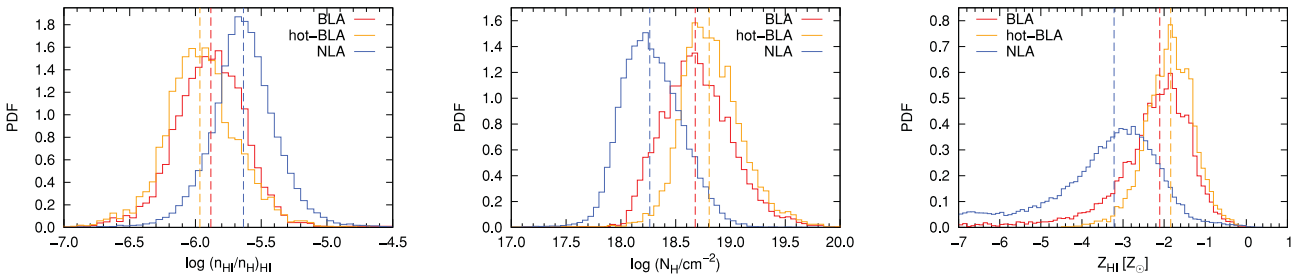

Figure 12. Distribution of various physical quantities characterizing the gas, giving rise to single-component absorption features identified as NLA (blue), BLA (red) and hot BLA (orange) in spectra at $z = 0.25$ with S/N = 50 obtained from our *AGN* model. Left-hand panel: neutral hydrogen fraction; middle panel: total hydrogen column density; right-hand panel: metallicity relative to the solar value ($\rm Z_\odot = 0.0127$). The vertical dashed lines in each panel indicate the median values of the corresponding distributions. On average, NLAs and (hot) BLAs appear to trace different gas phases.

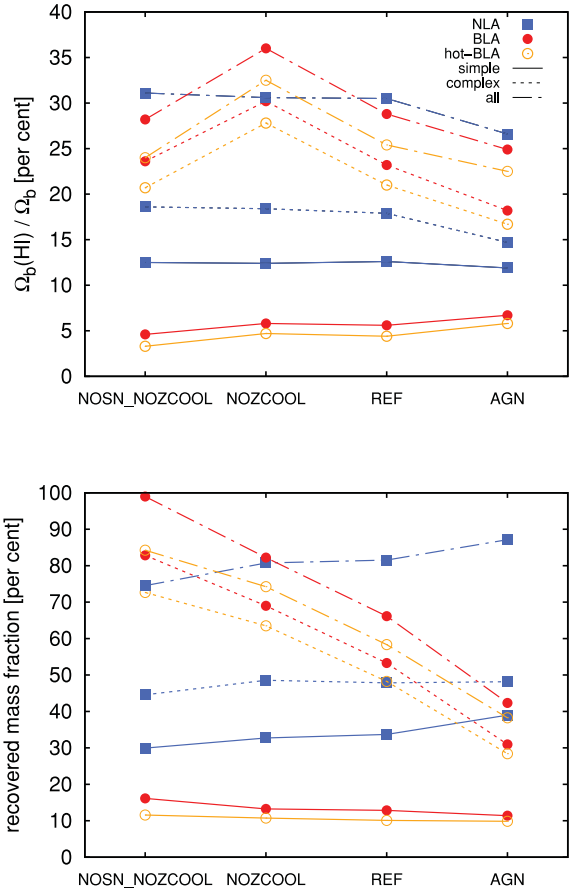

Figure 13. Top panel: baryon content (in per cent) relative to the cosmic value $\Omega_b = 0.0418$ of the gas traced by H I absorbers identified in synthetic spectra with S/N = 50 for different model runs. The H I absorber sample has been dissected into NLAs (blue squares), BLAs (red filled circles) and hot BLAs (orange open circles); these classes have in turn been divided into single-component absorbers (solid lines) and complex systems (dotted lines). The black dot–dashed lines indicate in each case the result for all (i.e. simple and complex) absorbers of a given class. Bottom panel: gas mass traced by a given H I absorber type relative to the actual gas mass in the phase expected to be traced by that particular type (see text for details). The symbols, lines and colours are as in the top panel.

absorbers (see Sections 3.5 and 4.2). Note that we consider the baryonic mass fraction of the gas traced by both simple and complex absorbers, since our adopted criterium to define simple absorbers is somehow arbitrary. The net baryon fractions of simple and complex absorbers (of a given class) taken together are indicated by the dot–dashed lines.

The baryonic mass fraction in NLAs (simple or complex) is very similar in all models, being only slightly lower in our fiducial run *AGN*. This is consistent with the lower gas mass fraction in the cool diffuse gas (which is expected to be traced by NLAs) in this model compared to all other models (see Fig. 7). The small difference in Ω_b(NLA)/Ω_b between all the models indicates that feedback has a negligible impact on the gas phase typically traced by NLAs. This in turn is consistent with the fact that the predicted H I statistics (which are dominated by these absorbers) are rather insensitive to variations in the feedback model (see Appendix C).

The baryonic mass fraction of the gas traced by simple BLAs is relatively low, and varies significantly between the models, from ∼3 per cent (*NOSN_NOZCOOL*) to ∼7 per cent (*AGN*). The baryonic mass fractions in complex BLAs are much higher than the baryonic mass fractions in simple BLAs, and they are also very different in each model. For instance, the baryon fraction of complex BLAs in the model *NOZCOOL* is higher (∼5 per cent) than in the model *NOSN_NOZCOOL*; this is consistent with the fact that SN feedback significantly increases the mass in warm-hot diffuse gas, as has been shown previously (see Fig. 7), together with the idea that BLAs preferentially trace this gas phase. In contrast, the baryon fraction in complex BLAs predicted by the models *REF* and *AGN* is lower (by ∼5 and ∼10 per cent, respectively) compared to predictions of the model *NOZCOOL*.

The lower baryon fraction in complex BLAs predicted by the model *REF* with respect to *NOZCOOL* can be explained as follows. Complex absorbers trace kinematically disturbed gas, most probably SN-driven outflows. These ejecta carry heavy elements with them, which allow a significant fraction of the gas to cool down radiatively, thus reducing the number of thermally broadened lines and their net baryonic mass. However, the lower baryonic mass content of complex BLAs in the model *AGN* with respect to all other models is in contrast with the actual total mass fraction in the warm-hot phase predicted by this model, which is higher compared to all other models (see the bottom left-hand panel of Fig. 7). The discrepancy between the predicted mass fraction in the warm-hot phase and the baryon content of the gas traced by BLAs in the *AGN* model can be understood as a consequence of the limited sensitivity. As already discussed, AGN feedback shifts a significant fraction of gas into the warm-hot phase; however, most of this mass ends up at temperatures and densities which lead to an H I fraction and corresponding absorption signal that is beyond our adopted detection limit (see Fig. 9).

Note that the baryon fractions traced by hot BLAs are only sightly lower than for BLAs, irrespective of the model. This is important because it implies that the contamination of the BLA sample by non-thermally broadened lines does not significantly affect the inferred baryon fraction of the WHIM. In other words, the baryon fraction in warm-hot gas is dominated by the absorbers arising in gas at the highest temperatures. This is a direct consequence of the steep decline of the hydrogen neutral fraction with temperature.

The bottom panel of Fig. 13 shows the *recovered mass fraction* (in per cent) of a given gas phase. This quantity is defined as the total baryonic mass in a given absorber class relative to the actual gas mass in the phase (Fig. 7) *expected* to be traced by that particular absorber class. So, for example, the gas mass recovered from hot BLAs is given in each model as the total baryonic mass in these absorbers (orange filled squares in the top panel of Fig. 13) divided by the gas mass in the warm-hot diffuse phase (orange percentages at the top left-hand corners of the panels of Fig. 7). The gas mass recovered from NLAs is correspondingly given as the total baryonic mass in these absorbers (blue squares in the top panel of Fig. 13) divided by the gas mass in the cool diffuse phase (percentages indicated at the bottom left-hand corners of the different panels of Fig. 7).

Apparently, none of the absorber classes traces the total gas mass in its corresponding (expected) phase, with an exception perhaps of the full (i.e. simple and complex) BLA sample (red filled circles) in the model *NOSN_NOZCOOL*. The gas mass fraction traced by both simple and complex NLAs is very similar in all models, and taken together these absorbers trace between ∼70 per cent (*NOSN_NOZCOOL*) and ∼90 per cent (*AGN*) of the gas mass in cool diffuse gas. This again is consistent with our previous

statement that the gas in this phase is left almost intact by feedback mechanisms such as SN-driven winds and AGN outflows.

Note that simple (hot) BLAs trace roughly 10–15 per cent of the true baryonic mass in warm-hot gas, irrespective of the model. This suggests that baryonic mass estimates based on this type of absorbers are robust. In contrast, the gas mass traced by *complex* (hot) BLAs is very different in each model. As already mentioned above, in the model *NOSN_NOZCOOL* the full BLA sample traces practically all the mass contained in warm-hot gas, with complex BLAs contributing more than 80 per cent to the recovered gas mass. This suggests that the bulk of gas shock heated by gravity (which is the only possible heating mechanism in the model *NOSN_NOZCOOL*) can be fully accounted for using (simple and complex) BLAs, at our adopted sensitivity. The recovered WHIM mass is, however, systematically lower in models that include SN and AGN feedback, and metal line cooling.

Taking the result from our fiducial run *AGN* at face value, we estimate the total baryon content in the gas traced by H I in our simulation at $z = 0.25$ to be $\Omega_{\rm b}({\rm H\,I})/\Omega_{\rm b} = 0.57$ (S/N = 50), 0.48 (S/N = 30) and 0.29 (S/N = 10). The last two values are in remarkable agreement with the results from observations at comparable sensitivity. Assuming a simple ionization model, Penton et al. (2004) measure[12] $\Omega_{\rm b}({\rm H\,I})/\Omega_{\rm b} = 0.31 \pm 0.04$ at $z \approx 0$ for absorbers with column densities $12.5 \leq \log(N_{\rm H\,I}/{\rm cm}^{-2}) \leq 17.5$ and $b_{\rm H\,I} \lesssim 100\,{\rm km\,s}^{-1}$. Similarly, assuming the gas to be isothermal and photoionized, Lehner et al. (2007) obtain[13] $\Omega({\rm H\,I})/\Omega_{\rm b} = 0.40$ from their data with an average S/N $\approx$ 15 and for $12.4 \leq \log(N_{\rm H\,I}/{\rm cm}^{-2}) \leq 16.5$ and $b_{\rm H\,I} \lesssim 150\,{\rm km\,s}^{-1}$.

4.5.1 Baryon content of warm-hot gas at low z

Estimates of baryonic mass contained in the WHIM based on broad H I absorbers detected in real QSO spectra are very uncertain, even with a reliable sample of BLA candidates at hand, since they are highly sensitive to the ionization state of the absorbing gas (see equation 3), which is probably dominated by collisions between ions and electrons in the plasma. In this case, the neutral hydrogen fraction is a steeply decreasing function of temperature, and an accurate estimate of the WHIM baryon content thus relies on a precise measurement of the temperature of the absorbing gas. As we have shown in Section 3.5.2, temperature estimates from the linewidth of broad H I absorbers may yield values that are uncertain by, at least, factors of a few.

The first attempt to measure the baryon content of the WHIM using BLAs was undertaken by Richter et al. (2004), who found $\Omega_{\rm b}({\rm BLA}) \leq 3.2 \times 10^{-3}\,(h/0.73)^{-1}$ assuming collisional ionization equilibrium (CIE), which represents less than 8 per cent of the cosmic baryon budget. In a follow-up study, Richter et al. (2006a) measured $\Omega_{\rm b}({\rm BLA}) \geq 2.6 \times 10^{-3}\,(h/0.73)^{-1}$, corresponding to at least 6 per cent of the baryons in the Universe. These authors also assumed CIE, but recognized the potential importance of PI in determining the ionization state of the WHIM, and concluded that their baryon content measurement could be underestimated by 15–50 per cent.

Using a significantly larger sample than previous studies, Lehner et al. (2007) report $\Omega_{\rm b}({\rm BLA})/\Omega_{\rm b} = 0.08$ assuming the gas to be in CIE for absorbers with $13.2 \leq \log(N_{\rm H\,I}/{\rm cm}^{-2}) \leq 16.5$ and $40 < b_{\rm H\,I} \lesssim 150\,{\rm km\,s}^{-1}$. Using the same sample and assuming a hybrid model (including PI and collisional ionization) to compute the neutral hydrogen fraction, these authors find $\Omega_{\rm b}({\rm BLA})/\Omega_{\rm b} = 0.20$. Both estimates are based on a series of assumptions. First, in order to account for the possible contamination of their sample with lines broadened by unresolved velocity structure or any other non-thermal mechanism, these authors randomly discard one-third of the BLAs in their sample. Moreover, they assume the thermal width to be 90 per cent of the observed linewidth, based on the results from previous simulations by Richter et al. (2006b). If, instead, the linewidth is dominated by thermal broadening, they get $\Omega_{\rm b}({\rm BLA})/\Omega_{\rm b} = 0.13$ (CIE) and $\Omega_{\rm b}({\rm BLA})/\Omega_{\rm b} = 0.32$ (PI+CIE).

In a more recent study,[14] Danforth et al. (2010) report $\Omega_{\rm b}({\rm BLA}) = 6.0^{+1.1}_{-0.8} \times 10^{-3}\ (h/0.73)^{-1}$, equivalent to $\Omega_{\rm b}({\rm BLA})/\Omega_{\rm b} = 0.14^{+0.03}_{-0.02}$. These authors analyse in detail the systematic uncertainties that afflict their (and others') baryon estimates, such as unresolved velocity structure, sample completeness, ionization corrections, and the assumed relation between linewidth and gas temperature, and find their estimate to vary between $\Omega_{\rm b}({\rm BLA}) = 2.3 \times 10^{-3}\,(h/0.73)^{-1}$ and $15.2 \times 10^{-3}\,(h/0.73)^{-1}$, that is, between ∼6 and ∼36 per cent of the cosmic baryon budget. Clearly, there is still a high uncertainty in the estimate of the baryonic mass traced by observed BLAs.

The results from our fiducial model are as follows. If we take simple and complex BLAs together, we find that they trace ∼25 (S/N = 50), ∼20 (S/N = 30) and ∼10 (S/N = 10) per cent of the total baryon budget in our simulation. For comparison, BLAs (simple and complex) that arise in gas at $T \geq 5 \times 10^4$ K yield ∼24 (S/N = 50), ∼18 (S/N = 30) and ∼9 (S/N = 10) per cent, which are very close to the values obtained from the whole BLA sample. If we restrict the BLAs to be single component, the resulting baryon fractions in these absorbers are ∼7 (S/N = 50), ∼6 (S/N = 30) and ∼5 (S/N = 10) per cent. This confirms that contamination of the BLA sample by non-thermally broadened lines does not significantly affect the inferred baryon fraction of the WHIM. Incidentally, this suggests that there should be little overlap between the estimates of $\Omega_{\rm b}$(NLA) and $\Omega_{\rm b}$(BLA).

One final important remark. We have demonstrated that the broad H I absorbers trace only a fraction of the total mass in the WHIM phase. Thus, even in the case that one could accurately estimate the baryon fraction in a given sample of absorbers, there is still a large gap between the observed and true mass contained in this gas phase. Although our simulation suggests that the baryonic masses estimated from observations represent 1/10–1/3 of the true baryonic mass in the WHIM, it is not clear how to use the measured baryonic mass to infer the true, total amount of baryons in this gas phase.

5 SUMMARY

In this paper, we have used a set of cosmological simulations from the OWLS project (Schaye et al. 2010) to study the physical conditions of the gas traced by BLAs with low and moderate column densities [$\log(N_{\rm H\,I}/{\rm cm}^{-2}) \lesssim 15$] observed in QSO spectra. We have chosen the *AGN* model of the OWLS suite to test the predictions of our simulations against a set of H I observables. We have investigated the impact of metal line cooling, kinetic feedback by SN

[12] Relative to $\Omega_{\rm b} = 0.0418$, rather than $\Omega_{\rm b} = 0.047$ assumed by Penton et al. (2004), and rescaled to $h = 0.73$.

[13] Relative to $\Omega_{\rm b} = 0.0418$, rather than $\Omega_{\rm b} = 0.044$, and rescaled to $h = 0.73$ rather than $h = 0.7$.

[14] We note that there is an error in the computation of the total absorption path length in Danforth et al. (2010, their table 1), which might have affected their reported estimate of the baryon content in BLAs (C. Danforth, private communication).

explosions and feedback by AGNs on the distribution of the gas mass over different phases such as the photoionized IGM and the shock-heated WHIM. Finally, we have explored the relation between the physical state and the baryon content of these gas phases and both NLAs and BLAs.

Our detailed results can be summarized as follows:

(i) Accretion shocks due to gravitational infall into the potential wells of DM haloes heat ∼30 per cent of the total gas mass to temperatures $T \geq 5 \times 10^4$ K by $z = 0.25$ (Section 4.1).

(ii) Feedback by SNe and AGNs each removes a similar amount of gas from the ISM in haloes at early epochs and displaces it to the warm-hot diffuse phase, increasing its total mass fraction by another ∼30 per cent by $z = 0.25$ (Section 4.1); in other words, roughly half of the gas mass predicted to be in the WHIM at low redshift (∼60 per cent) has been heated by accretion shocks while the other half is due to strong feedback.

(iii) The predictions from our simulations are in excellent agreement with standard H I observables (H I CDDF, linewidth distribution, $b_{\rm H\,I}$–$N_{\rm H\,I}$ correlation; Sections 3.2–3.4); these observables are rather insensitive to feedback and/or metal line cooling (Appendix C).

(iv) The line number density of NLAs ($b_{\rm H\,I} \geq 40\,{\rm km\,s^{-1}}$) and BLAs ($b_{\rm H\,I} \geq 40\,{\rm km\,s^{-1}}$) predicted by our fiducial run *AGN* is in broad agreement with the corresponding observed line frequencies (Section 4.3).

(v) The density of the H I absorbing gas shows a tight correlation with the H I column density, which agrees well with the analytic prediction of Schaye (2001; Section 3.5.1); this implies that our simulations are consistent with the assumption that typical H I absorbers are self-gravitating clouds in hydrostatic equilibrium with linear sizes of the order of the local Jeans length.

(vi) The temperature of the H I absorbing gas correlates well with the H I linewidth for $\log(N_{\rm H\,I}/{\rm cm^{-2}}) \gtrsim 13$, but it is a poor indicator of the thermal state of the gas for lower column densities; thermal broadening contributes, on average, with at least 60 per cent to the linewidth of BLAs (Section 3.5.2).

(vii) The overwhelming majority of NLAs are found to trace gas at $T \sim 10^4$ K; their number, temperature distribution and baryon content are very similar in models with/without feedback, thus strongly suggesting that feedback has a negligible impact on the cool, diffuse gas (i.e. the IGM; Section 4.4).

(viii) BLAs trace gas both at $T \sim 10^4$ K and at $T \sim 10^5$ K; our fiducial model, which includes feedback by both SNe and AGNs, predicts that two out of three BLAs arise in gas at $T \gtrsim 5 \times 10^4$ K; the number ratio of thermally to non-thermally broadened H I absorbers is very sensitive to (the adopted) feedback (model), and could in principle be used as an indicator of feedback strength (Sections 4.2 and 4.4).

(ix) The ionization state, the total hydrogen content and the level of enrichment of the gas traced by BLAs indicate that these absorbers arise in gas that is physically distinct from the gas traced by NLAs; we argue that BLAs mostly trace gas that has been recently shock heated and enriched by outflows (Sections 3.5.3 and 4.4).

(x) While models including SN and AGN feedback predict a higher fraction of gas mass to be in the warm-hot diffuse phase, the baryon fraction of the gas inferred from BLAs in these models is lower compared to a model without feedback; the reason is that much of the mass is displaced to temperatures and densities for which the H I fraction is too low for the gas to be detectable at the adopted sensitivity (Section 4.5).

(xi) The baryon fraction of the gas traced by both NLAs and BLAs predicted by our fiducial model shows broad agreement with corresponding measurements from observations (Section 4.5).

(xii) Baryonic mass estimates using simple BLAs are robust; in contrast, the gas mass traced by complex BLAs is very sensitive to the adopted (feedback) model (Section 4.5).

(xiii) Our fiducial model predicts that roughly 6 per cent of the total gas mass is traced by single-component BLAs, which represents about 10 per cent of the total WHIM mass in this model; if the restriction that the absorbers be single component is dropped, then around 25 per cent of the total gas mass (40 per cent of the WHIM mass) can be recovered from the detected broad H I absorption (Section 4.5).

(xiv) Although some of the gas mass with temperatures $T \geq 5 \times 10^4$ K and densities $\Delta \lesssim 10^2$ (i.e. the diffuse warm-hot phase) can be traced using BLAs, a significant fraction remains undetected as a consequence of a minimum (instrumental) sensitivity limit. Detection of the bulk of warm-hot gas requires a sensitivity (in terms of the H I central optical depth) of $\log \tau_0 \lesssim -2$ (Sections 4.2 and 4.5).

ACKNOWLEDGMENTS

The simulations presented here were run on Stella, the LOFAR Blue Gene/L system in Groningen and on the Cosmology Machine at the Institute for Computational Cosmology in Durham as part of the Virgo Consortium research programme. This work was sponsored by the National Computing Facilities Foundation (NCF) for the use of supercomputer facilities, with financial support from the Netherlands Organization for Scientific Research (NWO), an NWO VIDI grant, the Marie Curie Initial Training Network CosmoComp (PITN-GA-2009-238356), and the Deutsche Forschungsgemeinschaft (DFG) through Grant DFG-GZ: Ri 1124/5-1.

REFERENCES

Altay G., Theuns T., Schaye J., Crighton N. H. M., Dalla Vecchia C., 2011, ApJ, 737, L37
Bertone S., Schaye J., Dolag K., 2008, Space Sci. Rev., 134, 295
Booth C. M., Schaye J., 2009, MNRAS, 398, 53
Booth C. M., Schaye J., 2011, MNRAS, 413, 1158
Cen R., Ostriker J. P., 1999, ApJ, 514, 1
Cen R., Ostriker J. P., 2006, ApJ, 650, 560
Dalla Vecchia C., Schaye J., 2008, MNRAS, 387, 1431
Danforth C. W., Shull J. M., 2008, ApJ, 679, 194
Danforth C. W., Shull J. M., Rosenberg J. L., Stocke J. T., 2006, ApJ, 640, 716
Danforth C. W., Stocke J. T., Shull J. M., 2010, ApJ, 710, 613
Davé R., Tripp T. M., 2001, ApJ, 553, 528
Davé R., Hernquist L., Weinberg D. H., Katz N., 1997, ApJ, 477, 21
Davé R., Hernquist L., Katz N., Weinberg D. H., 1999, ApJ, 511, 521
Davé R. et al., 2001, ApJ, 552, 473
Davé R., Oppenheimer B. D., Katz N., Kollmeier J. A., Weinberg D. H., 2010, MNRAS, 408, 2051
Ferland G. J., Korista K. T., Verner D. A., Ferguson J. W., Kingdon J. B., Verner E. M., 1998, PASP, 110, 761
Gehrels N., 1986, ApJ, 303, 336
Green J. C. et al., 2012, ApJ, 744, 60
Haardt F., Madau P., 2001, in Neumann D. M., Tran J. T. V., eds, Clusters of Galaxies and the High Redshift Universe Observed in X-rays. CEA, Saclay, p. 64
Heap S. R., Williger G. M., Davé R., Weymann R. J., Jenkins E. B., Tripp T. M., 2002, in Mulchaey J. S., Stocke J. T., eds, ASP Conf. Ser. Vol. 254, Extragalactic Gas at Low Redshift. Astron. Soc. Pac., San Francisco, p. 63

Hui L., Gnedin N. Y., 1997, MNRAS, 292, 27
Jarosik N. et al., 2011, ApJS, 192, 14
Lanzetta K. M., Turnshek D. A., Wolfe A. M., 1987, ApJ, 322, 739
Lehner N., Savage B. D., Richter P., Sembach K. R., Tripp T. M., Wakker B. P., 2007, ApJ, 658, 680
Levenberg K., 1944, Q. Appl. Math., 2, 164
McCarthy I. G. et al., 2010, MNRAS, 406, 822
Marquardt D., 1963, J. Soc. Ind. Appl. Math., 11, 431
Norman M. L., Bryan G. L., 1999, in Miyama S. M., Tomisaka K., Hanawa T., eds, Numerical Astrophysics. Kluwer Academic Publishers, Boston, MA, p. 19
Oppenheimer B. D., Davé R., 2009, MNRAS, 395, 1875
Penton S. V., Stocke J. T., Shull J. M., 2004, ApJS, 152, 29
Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 1992, Numerical Recipes in Fortran 77, 2nd edn. Cambridge Univ. Press, Cambridge
Rauch M., Haehnelt M. G., Steinmetz M., 1997, ApJ, 481, 601
Richter P., Savage B. D., Tripp T. M., Sembach K. R., 2004, ApJS, 153, 165
Richter P., Savage B. D., Sembach K. R., Tripp T. M., 2006a, A&A, 445, 827
Richter P., Fang T., Bryan G. L., 2006b, A&A, 451, 767
Richter P., Paerels F. B. S., Kaastra J. S., 2008, Space Sci. Rev., 134, 25
Schaye J., 2001, ApJ, 559, 507
Schaye J., Dalla Vecchia C., 2008, MNRAS, 383, 1210
Schaye J., Theuns T., Leonard A., Efstathiou G., 1999, MNRAS, 310, 57
Schaye J., Carswell R. F., Kim T., 2007, MNRAS, 379, 1169
Schaye J. et al., 2010, MNRAS, 402, 1536
Seljak U., Zaldarriaga M., 1996, ApJ, 469, 437
Sembach K. R., Tripp T. M., Savage B. D., Richter P., 2004, ApJS, 155, 351
Shull J. M. et al., 2000, ApJ, 538, L13
Smith B. D., Hallman E. J., Shull J. M., O'Shea B. W., 2011, ApJ, 731, 6
Spergel D. N. et al., 2007, ApJS, 170, 377
Springel V., 2005, MNRAS, 364, 1105
Springel V. et al., 2005, Nat, 435, 629
Tepper-García T., 2006, MNRAS, 369, 2025
Tepper-García T., Richter P., Schaye J., Booth C. M., Dalla Vecchia C., Theuns T., Wiersma R. P. C., 2011, MNRAS, 413, 190 (Paper I)
Theuns T., Leonard A., Efstathiou G., Pearce F. R., Thomas P. A., 1998, MNRAS, 301, 478
Theuns T., Viel M., Kay S., Schaye J., Carswell R. F., Tzanavaris P., 2002, ApJ, 578, L5
Thom C., Chen H.-W., 2008a, ApJS, 179, 37
Thom C., Chen H.-W., 2008b, ApJ, 683, 22
Tripp T. M., Savage B. D., Jenkins E. B., 2000, ApJ, 534, L1
Tripp T. M., Sembach K. R., Bowen D. V., Savage B. D., Jenkins E. B., Lehner N., Richter P., 2008, ApJS, 177, 39
van de Voort F., Schaye J., Altay G., Theuns T., 2011, MNRAS, 421, 2809
White S. D. M., 1996, in Schaeffer R., Silk J., Spiro M., Zinn-Justin J., eds, Cosmology and Large Scale Structure Formation and Evolution of Galaxies. Elsevier Scientific, Amsterdam, the Netherlands, p. 349
Wiersma R. P. C., Schaye J., Smith B. D., 2009a, MNRAS, 393, 99
Wiersma R. P. C., Schaye J., Theuns T., Dalla Vecchia C., Tornatore L., 2009b, MNRAS, 399, 574
Wiersma R. P. C., Schaye J., Dalla Vecchia C., Booth C. M., Theuns T., Aguirre A., 2010, MNRAS, 409, 132
Williger G. M., Heap S. R., Weymann R. J., Davé R., Ellingson E., Carswell R. F., Tripp T. M., Jenkins E. B., 2006, ApJ, 636, 631
Williger G. M. et al., 2010, MNRAS, 405, 1736
Zel'dovich Ya. B., 1970, A&A, 5, 84

APPENDIX A: LINE FITTING

We fit our spectra using a significantly modified version of AUTOVP (Davé et al. 1997), assuming each absorption component to be described by a Voigt profile given by the analytic approximation of Tepper-García (2006). AUTOVP identifies absorption features using the equivalent width significance criterion of Lanzetta, Turnshek & Wolfe (1987). The spectrum is scanned using a window of width n pixels in search of regions with significant absorption. A region is considered significant in absorption if its equivalent width satisfies $W \geq N\sigma_W$, where N is the significance level, and σ_W is the uncertainty in the equivalent width, integrated over n pixels, given by

$$\sigma_W \approx \sqrt{n}\left(\frac{\Delta v}{c}\lambda_0\right)(1+z)(\mathrm{S/N})^{-1}\,. \quad \text{(A1)}$$

Here, Δv is the pixel width, S/N is the adopted signal-to-noise ratio, z is the (central) redshift of the absorption feature, λ_0 is the rest-frame wavelength of the transition considered (e.g. H I Lyα), and c is the speed of light. We adopt $N = 7$ and $n = 25$ (corresponding to approximately 88 km s^{-1} for our chosen resolution), and $\Delta v = 3.5\,\mathrm{km\,s^{-1}}$. With these values, the significance value translates into a *rest-frame* equivalent width

$$W_\mathrm{r} \approx 500\,(\mathrm{S/N})^{-1}\,\mathrm{m\AA}\,. \quad \text{(A2)}$$

Note that our adopted window width does not affect the parameters of the fitted line(s) in any way.

If fitted by a single component, this implies that our line sample is formally complete down to H I column densities[15]

$$N_\mathrm{H\,I} \approx 9\times10^{13}\,(\mathrm{S/N})^{-1}\,\mathrm{cm^{-2}}\,. \quad \text{(A3)}$$

For Doppler parameters in the range $b_\mathrm{H\,I} \geq 40\,\mathrm{km\,s^{-1}}$, characteristic of BLAs, the above corresponds to a sensitivity limit in terms of absorption strength

$$\left(\frac{N_\mathrm{H\,I}/\mathrm{cm^{-2}}}{b_\mathrm{H\,I}/\,\mathrm{km\,s^{-1}}}\right) \lesssim 2.3\times10^{12}\,(\mathrm{S/N})^{-1}\,, \quad \text{(A4)}$$

which is equivalent to an H I Lyα optical depth at the line centre[16] $\tau_0 \gtrsim 1.74(\mathrm{S/N})^{-1}$. Note that the values implied by the above equation are below the value commonly adopted for the identification of BLA candidates in real QSO spectra (see equation 2).

A spectrum is fitted in two steps. In the first step, an absorption component is fitted at the pixel with the minimum flux in each detection region, starting with the region with the overall minimum flux. The column density, $N_\mathrm{H\,I}$, and the Doppler parameter, $b_\mathrm{H\,I}$, of the line are both iteratively reduced by a factor of 0.99 starting from large values (e.g. $N_\mathrm{H\,I} = 10^{20}\,\mathrm{cm^{-2}}$ and $b_\mathrm{H\,I} = 300\,\mathrm{km\,s^{-1}}$; see below) until the flux at that pixel is within 2σ below the actual flux level, that is, in the range $[F-2\sigma, F]$, where σ is the local noise. Further components are added and their parameters correspondingly adjusted, taking all previous fitted lines into account, until the residual flux (i.e. the difference between actual flux and model flux) across the detection region is below 2σ. This procedure is repeated for all detected regions. In a second step, the line parameters (velocity centroid v_0, column density $N_\mathrm{H\,I}$, and Doppler parameter $b_\mathrm{H\,I}$) of all lines are simultaneously adjusted using the Levenberg–Marquardt algorithm (Levenberg 1944; Marquardt 1963) as implemented in Press et al. (1992) until the reduced χ^2 value, that is, the χ^2 value divided by the degrees of freedom, is below $\chi^2_\mathrm{bad} \equiv 1.2$. If convergence is not achieved, the sightline is discarded. We note that the fraction of discarded sightlines is vanishingly small, and it amounts to 1/5000 for our spectra at $z = 0.25$ and none for our spectra in the range $0 \leq z \leq 0.5$.

[15] The quoted value is valid only for absorption lines on the linear part of the curve of growth, which is the case for the majority of the components identified in our synthetic spectra. Also note that we do detect lines with column densities (and rest-frame equivalent widths) smaller than the quoted values, since a detected region can be fitted by more than one component.
[16] The relation between line strength and central optical depth for the Lyα line is given by equation (B3) in Appendix B.

Since we do not take higher order H I Lyman transitions into account, saturated H I Lyα lines deserve special attention. A pixel is considered saturated if the corresponding flux is below 2σ. For (S/N) $\gtrsim$ 10, this implies that the flux is of the order of, or lower than, $2\sigma_{\rm min}$ (where $\sigma_{\rm min} = 10^{-4}$), which is equivalent to an H I Lyα central optical depth $\tau_0 \approx 8.52$ or an H I column density $\log(N_{\rm H\,I}/{\rm cm}^{-2}) \approx 13.1 + \log(b_{\rm H\,I}/{\rm km\,s}^{-1})$ (see equation B3). Assuming a Doppler parameter $b_{\rm H\,I} = 30\,{\rm km\,s}^{-1}$ (which approximately corresponds to the median b value of our line sample, see Section 3.3), this corresponds to $\log(N_{\rm H\,I}/{\rm cm}^{-2}) \approx 14.5$. In order to prevent our algorithm from severely underestimating the true column density of such saturated lines, and at the same time to avoid fitting lines with unrealistically large column densities along a given sightline, during the second fitting step we limit the column density of an individual absorption line[17] to five times this value, for example, $\log(N_{\rm H\,I}/{\rm cm}^{-2})_{\rm max} \approx 15.2$ for $b_{\rm H\,I} = 30\,{\rm km\,s}^{-1}$. Note that this value is not a strict limit but may still vary (in particular, it can be larger) depending on the actual b value of the saturated line. As a consequence, we highly underestimate (by up to two orders of magnitude; see Table 2) the actual baryon content in H I along all fitted sightlines, which is dominated by high H I column density gas. Yet, the properties of the H I absorbers, in particular the broad H I absorbers that are relevant for this study, are not affected, since these are dominated by the low column density population with $\log(N_{\rm H\,I}/{\rm cm}^{-2}) \lesssim 14.5$.

During the second fitting step, we impose a minimum linewidth of $b_{\rm H\,I} = 10\,{\rm km\,s}^{-1}$, corresponding to $T \approx 6000$ K assuming pure thermal broadening. Lines narrower than this are discarded during the fitting process unless the new χ^2 value increases above $1.2\chi^2_{\rm bad}$. We note that this cut-off does not appreciably affect the resulting b-value distribution (see Fig. 2). Indeed, observations indicate that NLAs (i.e. $b_{\rm H\,I} \lesssim 15\,{\rm km\,s}^{-1}$) at low redshift are scarce (Lehner et al. 2007).

Since our synthetic spectra are continuum normalized by construction, we do not fit a continuum prior to line identification. We limit the linewidth to a maximum value of $b_{\rm H\,I} = 300\,{\rm km\,s}^{-1}$, although larger values are allowed if doing so reduces the χ^2 value below $1.2\chi^2_{\rm bad}$. As a consequence, we find a small number of very broad ($b_{\rm H\,I} > 200\,{\rm km\,s}^{-1}$), very shallow absorption features, which become less numerous with decreasing S/N. Since most of these features are real, though scarce, we do not discard them but include them in our resulting line sample.

Finally, any candidate lines with formal relative errors in $N_{\rm H\,I}$ or $b_{\rm H\,I}$ larger than 50 per cent are sequentially discarded unless the χ^2 value increases above $1.2\chi^2_{\rm bad}$. Note that the final formal errors are typically much smaller than this, around 10 per cent for both $N_{\rm H\,I}$ and $b_{\rm H\,I}$.

APPENDIX B: OBSERVABILITY OF H I ABSORBING GAS

The neutral hydrogen column density, $N_{\rm H\,I}$, is given by

$$N_{\rm H\,I} = N_{\rm H} f_{\rm H\,I}\,,$$

where $N_{\rm H}$ and $f_{\rm H\,I} \equiv (n_{\rm H\,I}/n_{\rm H})$ are the total hydrogen column density and neutral hydrogen fraction, respectively. The total hydrogen column density can be written using the hydrogen particle density $n_{\rm H}$ as

$$N_{\rm H} = \int_0^L n_{\rm H}\,{\rm d}l = \overline{n_{\rm H}} L\,,$$

where L is the physical, linear extension of the absorbing structure along the sightline, and $\overline{n_{\rm H}}$ is the average hydrogen particle density. In the following, we will write $n_{\rm H} \equiv \overline{n_{\rm H}}$, but the reader should keep the (slight) difference in mind.

The width of an H I absorbing line as measured by the Doppler parameter $b_{\rm H\,I}$ may be modelled as

$$b^2_{\rm H\,I} = b^2_T + b^2_{\rm H} + b^2_{\rm nt}\,.$$

The thermal width, that is, the broadening due to the temperature T of the absorbing gas is given by

$$\left(b_T/{\rm km\,s}^{-1}\right) = 12.9\sqrt{T/10^4\,{\rm K}}\,,$$

and the Hubble broadening by (see e.g. Schaye 2001)

$$b_H \sim \frac{1}{2} H(z) L\,,$$

where $H(z)$ is the Hubble parameter (at the appropriate epoch), expressed as

$$H(z) \equiv h(z) \times 10^2\ {\rm km\,s}^{-1}\,{\rm Mpc}^{-1}\,, \tag{B1}$$

with

$$h(z) = h_0[\Omega_{\rm m}(1+z)^3 + \Omega_\Lambda]^{1/2}\,. \tag{B2}$$

We adopt the cosmological parameters $\{\Omega_{\rm m}, \Omega_\Lambda, h_0\} = \{0.238, 0.762, 0.73\}$ as derived from the *WMAP* 3-year data, and find, for example, $h(z = 0.25) = 0.81$.

The remaining term, $b_{\rm nt}$, in the expression for $b_{\rm H\,I}$ includes all other forms of non-thermal broadening and is less straightforward to model. It may include turbulence within the absorbing gas, peculiar motions of the absorbing structures, etc. Assuming that these are negligible compared to the thermal and Hubble components, the linewidth can be approximated by

$$\left(b_{\rm H\,I}/{\rm km\,s}^{-1}\right)^2 \approx 166\left(T/10^4\,{\rm K}\right) + \frac{1}{4}10^4\left[h(z)L/{\rm Mpc}\right]^2\,.$$

Putting all the above equations together and simplifying, we find that the H I Lyα absorption strength of the gas is given by

$$\left(\frac{N_{\rm H\,I}/{\rm cm}^{-2}}{b_{\rm H\,I}/{\rm km\,s}^{-1}}\right) = \frac{6.17\times10^{12} h(z)^{-1}(n_{\rm H}/10^{-5}{\rm cm}^{-3})(f_{\rm H\,I}/10^{-5})}{\sqrt{6.64\times10^{-2}(T/10^4\,{\rm K})[h(z)L/{\rm Mpc}]^{-2}+1}}\,.$$

The central optical depth of the H I Lyα transition can be expressed in terms of $(N_{\rm H\,I}/b_{\rm H\,I})$ as

$$\tau_0 = \frac{\sqrt{\pi}e^2}{m_{\rm e}c} f_{{\rm Ly}\alpha}\lambda_{{\rm Ly}\alpha}\left(\frac{N_{\rm H\,I}}{b_{\rm H\,I}}\right) = 7.56\times10^{-13}\left(\frac{N_{\rm H\,I}/{\rm cm}^{-2}}{b_{\rm H\,I}/{\rm km\,s}^{-1}}\right)\,. \tag{B3}$$

Using the above equations we obtain an expression for τ_0 in terms of $n_{\rm H}$, T and L:

$$\tau_0 = \frac{4.66 h(z)^{-1}(n_{\rm H}/10^{-5}\ {\rm cm}^{-3})(f_{\rm H\,I}/10^{-5})}{\sqrt{6.64\times10^{-2}(T/10^4\,{\rm K})[h(z)L/{\rm Mpc}]^{-2}+1}}\,. \tag{B4}$$

Note that $f_{\rm H\,I} = f_{\rm H\,I}(n_{\rm H}, T, z)$, where the z dependence comes about through the redshift dependence of the ionization background included in the calculation of $f_{\rm H\,I}$.

If we assume that the absorbers have linear sizes of the order of the local Jeans length (Schaye 2001)

$$L_{\rm J} = 0.169\,{\rm Mpc}\,(n_{\rm H}/10^{-5}\ {\rm cm}^{-3})^{-1/2} \times (T/10^4\,{\rm K})^{1/2}\,(f_{\rm g}/0.168)^{1/2}\,, \tag{B5}$$

[17] This particular value chosen is arbitrary, but has been found to give satisfactory results. Note that our algorithm may still underestimate the true column density of heavily saturated lines.

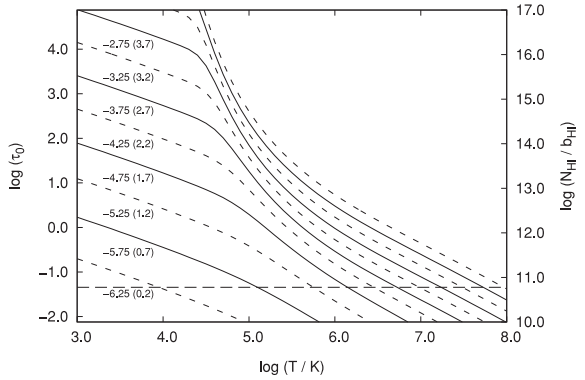

Figure B1. H I Lyα central optical depth, τ_0, as given by equation (B6), required to detect gas with a given temperature for a range of densities at $z = 0.25$. The value along each curve indicates the corresponding logarithmic hydrogen particle density, $\log(n_{\rm H}/{\rm cm}^{-3})$. The number in parentheses indicates the corresponding logarithmic overdensity. For reference, we include the alternative y-axis which shows the corresponding value of the H I Lyα absorption strength at the line centre, $\log(N_{\rm H\,I}/b_{\rm H\,I})$, in units of ${\rm cm}^{-2}\,{\rm km}^{-1}\,{\rm s}$ (see equation B3). The dashed horizontal line indicates the adopted BLA detection threshold given by equation (2).

we get

$$\tau_0 = \frac{4.66h(z)^{-1}\,(n_{\rm H}/10^{-5}\ {\rm cm}^{-3})\,(f_{\rm H\,I}/10^{-5})}{\sqrt{2.32(n_{\rm H}/10^{-5}\ {\rm cm}^{-3})\,h(z)^{-2}+1}}\,, \qquad \text{(B6)}$$

where we have assumed the fraction of mass in gas to be close to its universal value[18] $f_{\rm g} \equiv \Omega_{\rm b}/\Omega_{\rm m} = 0.168$. Note that the above equation does no longer depend explicitly on the temperature, but it does depend *implicitly* on it through the dependence on $f_{\rm H\,I}$.

Using

$$n_{\rm H} = \frac{\langle\rho_{\rm b}\rangle}{m_{\rm H}} X_{\rm H}(1+z)^3\Delta \approx 1.88\times10^{-7}\,{\rm cm}^{-3}\left(\frac{X_{\rm H}}{0.752}\right)(1+z)^3\Delta$$

the above equations can all be expressed in terms of the overdensity Δ as well.

Fig. B1 shows the H I Lyα central optical depth, τ_0, as a function of gas temperature for a range of densities typical of intergalactic gas, as given by equation (B6). The alternative y-axis shows the corresponding values for $\log(N_{\rm H\,I}/b_{\rm H\,I})$ in units of ${\rm cm}^{-2}\,{\rm km}^{-1}\,{\rm s}$. The value along each curve indicates the assumed logarithmic hydrogen particle density, $\log(n_{\rm H}/{\rm cm}^{-3})$, and the value in parentheses indicates the corresponding logarithmic overdensity, $\log\Delta$, at $z = 0.25$. The horizontal dashed line indicates our adopted sensitivity limit as given by equation (2). This figure demonstrates that the detectability of H I absorbing gas at a given density drops sharply with temperature. For example, the detection of gas with $\log(n_{\rm H}/{\rm cm}^{-3}) = -4.75$ (which corresponds to $\Delta \sim 50$ at $z = 0.25$) and $\log(T/{\rm K}) \sim 6$ requires a minimum sensitivity which is two orders of magnitude higher than the one required to detect gas with the same density and $\log(T/{\rm K}) \sim 5$. The reason behind the strong dependence of $\log(N_{\rm H\,I}/b_{\rm H\,I})$ on temperature is that, at a given density, τ_0 is completely dominated by the neutral fraction, $f_{\rm H\,I}$.

[18] The total mass density parameter is $\Omega_{\rm m} = \Omega_{\rm b} + \Omega_{\rm c}$. The most recent measurements of the baryonic and DM density parameters yield, respectively, $\Omega_{\rm b} = 0.0449 \pm 0.0028$ and $\Omega_{\rm c} = 0.222 \pm 0.026$ (Jarosik et al. 2011).

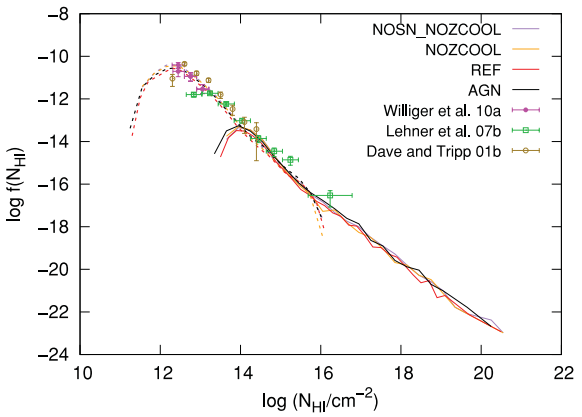

Figure C1. H I CDDF for different model runs in a box $L = 100\,h^{-1}$ Mpc per side at $z = 0.25$ using 5000 random sightlines. The dashed lines show the distribution of column densities for individual absorption lines obtained from fitting the synthetic spectra adopting S/N = 50. The solid lines show the distribution of *integrated* $N_{\rm H\,I}$ along each sightline. Note that the adopted S/N value is higher than the S/N of the data, and the latter are only included for reference. The distribution of column densities for individual absorption components extends to much lower values since a spectrum (corresponding to a single physical sightline) is generally fitted with more than one component.

APPENDIX C: CONVERGENCE WITH RESPECT TO THE PHYSICAL MODEL

In this section, we demonstrate that the predicted H I observables are robust with respect to variations of the adopted physical model.

Fig. C1 shows the H I CDDF for different models run in a box $L = 100\,h^{-1}$ Mpc per side at $z = 0.25$, using 5000 random sightlines. The dashed lines show the distribution of column densities for individual absorption components obtained from fitting the synthetic spectra adopting S/N = 50, as described in Section 3; the solid lines show the distribution of the *integrated* H I column density along each sightline. The latter are important in order to remove the uncertainty introduced in the H I CDDF by our fitting algorithm. We include various data sets in this figure for reference, but note that our adopted S/N value is higher than the S/N of the data. The distribution of column densities for *individual* absorption components extends over a range of much lower values since a spectrum (corresponding to a single physical sightline) is generally fitted with more than one component. Also, as mentioned in Section 3, our version of AUTOVP tends to fit (highly) saturated lines with more than one component, thus generally yielding lower $N_{\rm H\,I}$ values for each individual component with respect to the $N_{\rm H\,I}$ value obtained by integrating over the corresponding absorption feature.

Fig. C2 shows the distribution of Doppler parameters of individual components identified in 5000 spectra at $z = 0.25$ with S/N = 50 adopting different physical models run in a box $L = 100\,h^{-1}$ Mpc per side. Note that feedback (both by SNe and by AGNs) leads to a slightly larger number of broad lines, although the effect is not large.

Finally, Fig. C3 shows the corresponding $b_{\rm H\,I}$–$N_{\rm H\,I}$ distribution of individual components. The $b_{\rm H\,I}$–$N_{\rm H\,I}$ distribution for each model has been obtained as described in Section 3.4. For each model, the corresponding median values at each bin are connected by continuous lines. For simplicity, only the result for model *AGN* displays x- and y-error bars indicating, respectively, the bin size and 25/75 percentiles, but note that the bin size is identical and the scatter similar for all other models. Although the models differ slightly from each

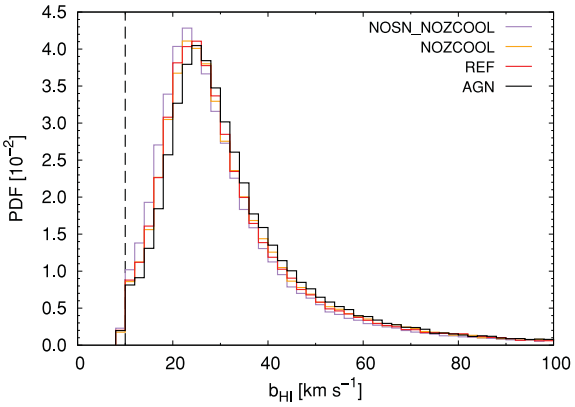

Figure C2. Distribution of Doppler parameters of individual components identified in 5000 spectra at $z = 0.25$ with S/N = 50 adopting different physical models in an $L = 100\,h^{-1}$ Mpc box. The dashed vertical line indicates the minimum Doppler parameter during the fitting process (see Section 3). Note that feedback (both by SNe and by AGNs) leads to a slightly larger number of broad lines, although the effect is small.

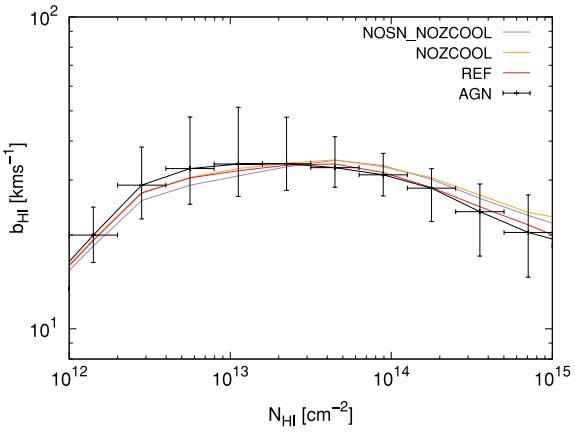

Figure C3. $b_{\rm H\,I}$–$N_{\rm H\,I}$ distribution of individual components identified in 5000 spectra at $z = 0.25$ with S/N = 50 adopting different physical models in a box $L = 100\,h^{-1}$ Mpc per side. For clarity, errors bars (corresponding to the bin size along the x-axis and to the 25/75 percentiles along the y-axis) are shown only for the AGN result. We note that the scatter for the other models is similar.

other, the various $b_{\rm H\,I}$–$N_{\rm H\,I}$ distributions are fully consistent with each other.

Clearly, feedback by SNe and AGNs does not significantly affect the observed properties (H I column density distribution, Doppler parameter distribution, $b_{\rm H\,I}$–$N_{\rm H\,I}$ correlation) of the gaseous structures giving rise to H I absorption. This result is consistent with previous results by Theuns et al. (2002) who, however, did not consider AGN feedback or metal line cooling.

APPENDIX D: NUMERICAL CONVERGENCE

In this section, we address the convergence of our results with respect to varying the box size and the mass and spatial resolution. To this end, we compare the H I CDDF and Doppler parameter distribution of individual components identified in 5000 spectra obtained from simulation runs with different box sizes and resolutions, which all adopt model *REF* at $z = 0.25$. Note that the use of this particular model does not affect our results, since we have shown in Appendix C that both the H I CDDF and the distribution of Doppler parameters are insensitive to the adopted model. In the following, the simulation runs we use are denoted by *LxxxNyyy*, where *xxx* corresponds to the linear size of the cubic box in h^{-1} Mpc, and *yyy* corresponds to the number of (DM, baryonic) particles per side.

To investigate the convergence with box size, we use the simulation runs *L*025*N*128, *L*050*N*256 and *L*100*N*512, which all have the same mass and spatial resolution. The convergence with resolution is investigated using simulations run in a box $L = 50\,h^{-1}$ Mpc per side, and varying the (DM, baryonic) particle number; more specifically, we use the runs *L*050*N*128, *L*050*N*256 and *L*050*N*512, whose mass (spatial) resolution varies in factors of 8 (2). The choice of this particular box size is arbitrary but justified, since our results are converged with respect to box size, as we will show next.

Fig. D1 shows the H I CDDF and the distribution of Doppler parameters using 5000 sightlines obtained from simulation runs with different box size but fixed mass and spatial resolution. The top panel shows the distribution obtained from the H I column density integrated along each individual sightline; the middle panel shows the distribution obtained from the H I column density of each component identified in the corresponding spectra with S/N = 50. The H I CDDF of integrated column densities is fully converged for column densities $N_{\rm H\,I} < 10^{17}{\rm cm}^{-2}$. At higher column densities, we do not expect a good convergence, since these column densities correspond to the optically-thick regime, while our calculations assume optically-thin gas. The H I CDDF of individual components is fully converged at column densities $N_{\rm H\,I} < 10^{15}$ cm^{-2}, which is the relevant column density range for this study. The difference between the various H I CDDFs of individual components in the range $N_{\rm H\,I} > 10^{15}$ cm^{-2} is due to the inability of our fitting algorithm to accurately determine the column density of saturated H I absorption features. Finally, the bottom panel demonstrates that the Doppler parameter distribution in the range $10 \leq b_{\rm H\,I} \leq 100\,{\rm km\,s}^{-1}$ is fully converged with respect to box size.

Fig. D2 shows the H I CDDF and the distribution of Doppler parameters using 5000 sightlines obtained from simulations run in a box $L = 50\,h^{-1}$ Mpc per side with different mass and spatial resolution. Both the distribution of H I column densities integrated along each sightline (top panel) and the distribution of H I column densities of individual components identified in the corresponding spectra (middle panel) appear to be fully converged with respect to resolution at column densities $N_{\rm H\,I} < 10^{15}$ cm^{-2}. At higher column densities, neither distribution is fully converged, although the difference between the two highest resolution runs, *L*050*N*256 and *L*050*N*512, is very small for the integrated column densities. In the case of the H I CDDFs for individual components at $N_{\rm H\,I} > 10^{15}$ cm^{-2}, this is again in part due to the difficulty in determining the true column density of saturated H I absorption features.

The Doppler parameter distribution shown in the bottom panel of Fig. D2 indicates that the resolution of the *L*050*N*128 run is not high enough. The distribution in the *L*050*N*256 run, which has the same resolution as our fiducial run, does not show full convergence at Doppler parameters in the BLA regime ($b_{\rm H\,I} \geq 40\,{\rm km\,s}^{-1}$) with respect to the higher resolution run, *L*050*N*512, although the difference is small (see also Theuns et al. 1998).

In summary, our results are robust with respect to varying the size of the simulation box, and our adopted resolution is high enough to guarantee the convergence of our results in the range of column densities relevant for this study. However, the distribution of Doppler parameters is slightly sensitive to the adopted resolution in the range of interest for BLAs, that is, for Doppler parameters $b_{\rm H\,I} \geq 40\,{\rm km\,s}^{-1}$. Thus, some caution is advised when interpreting results based on or making predictions for this observable.

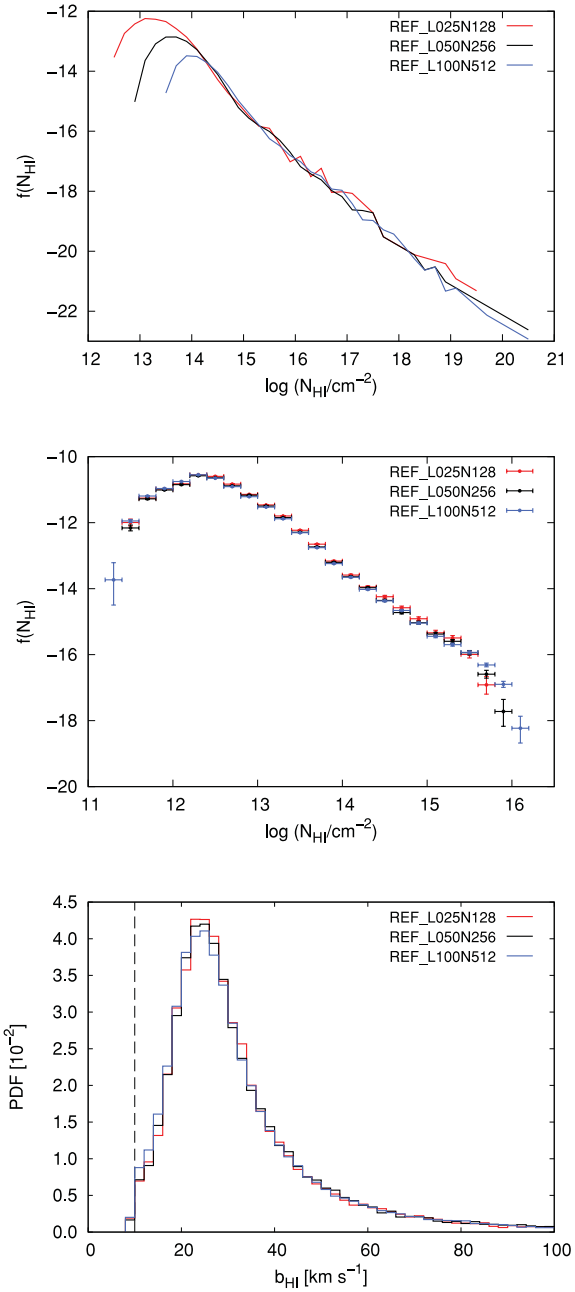

Figure D1. Numerical convergence with respect to box size at a fixed mass and spatial resolution using 5000 sightlines and corresponding spectra obtained from a simulation that adopts model *REF* in an $L = 100\,h^{-1}$ Mpc box per side at $z = 0.25$. Top panel: H I CDDF obtained from the H I column density integrated along each sightline. Middle panel: H I CDDF of the H I absorbing components identified in the corresponding spectra with S/N = 50. Note the difference in the plotted range between the top and middle panels. Bottom panel: distribution of Doppler parameters. The dashed vertical line indicates the minimum Doppler parameter allowed during the fitting process (see Appendix A).

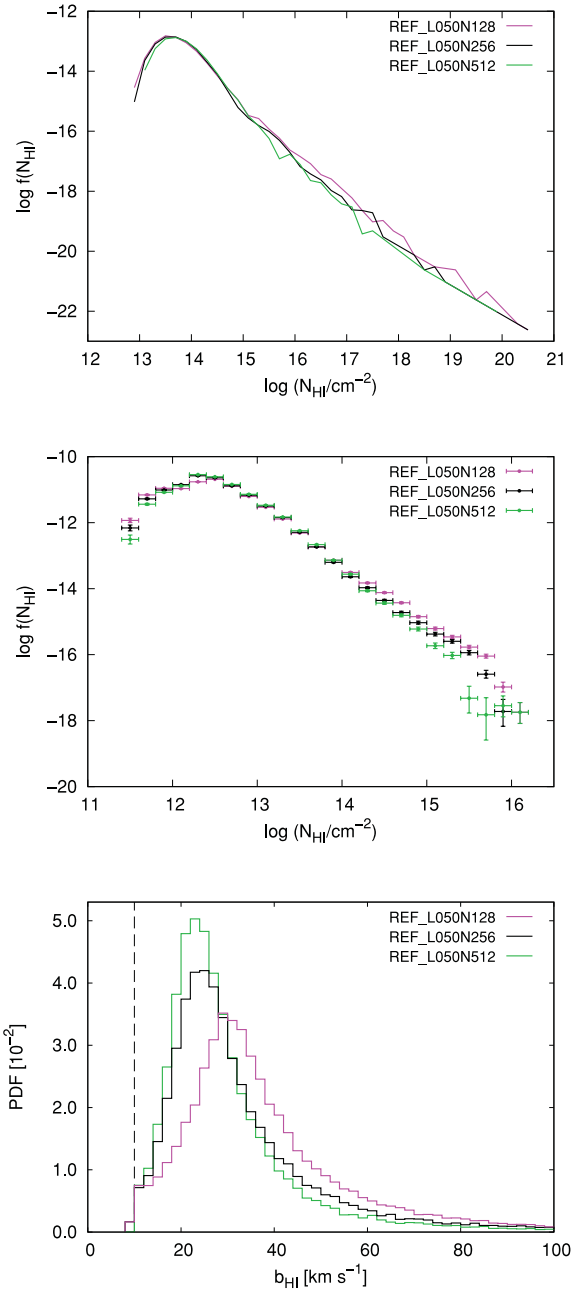

Figure D2. Same as Fig. D1 for the numerical convergence with respect to mass and spatial resolution at a fixed box size.

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