

# The broadening of Lyman- $\alpha$ forest absorption lines

Antonella Garzilli<sup>1\*</sup>, Tom Theuns<sup>2</sup> and Joop Schaye<sup>3</sup>

<sup>1</sup> Lorentz Institute, Leiden University, Niels Bohrweg 2, Leiden, NL-2333 CA, The Netherlands

<sup>2</sup> Institute for Computational Cosmology, Department of Physics, Durham University, DH1 3LE Durham, UK

<sup>3</sup> Leiden Observatory, Leiden University, P.O. Box 9513, 2300 RA Leiden, The Netherlands

Accepted 2015 February 20. Received 2015 February 19; in original form 2014 December 28

## ABSTRACT

We provide an analytical description of the line broadening of H I absorbers in the Lyman- $\alpha$  forest resulting from Doppler broadening and Jeans smoothing. We demonstrate that our relation captures the dependence of the line-width on column density for narrow lines in  $z \sim 3$  mock spectra remarkably well. Broad lines at a given column density arise when the underlying density structure is more complex, and such clustering is not captured by our model. Our understanding of the line broadening opens the way to a new method to characterise the thermal state of the intergalactic medium and to determine the sizes of the absorbing structures.

**Key words:** cosmology: large scale structure of Universe – quasars: absorption lines – intergalactic medium

## 1 INTRODUCTION

The intergalactic medium (IGM) is detected as intervening absorption in the spectra of background sources (Gunn & Peterson 1965), in the form of a ‘forest’ of H I (Lynds (1971); Weymann et al. (1981), see *e.g.* Rauch (1998) for a review) and He II (Jakobsen et al. 1994; Davidsen et al. 1996) absorption lines. This gas is very highly ionised,  $x = n_{\text{HI}}/n_{\text{H}} \sim 10^{-4}$ , by a pervasive background of ionising radiation produced by galaxies and quasars (*e.g.* Haardt & Madau 1996), with hints of an increasingly neutral fraction,  $x \sim 0.1$ , above redshift  $z \sim 7$  (Mortlock et al. 2011). At all times, it contains the majority of baryons (Fukugita et al. 1998).

The growth of structure in a cold dark matter (CDM) Universe naturally gives rise to a ‘cosmic web’ of voids, sheets and filaments, with absorbers arising when the sight-line intersects higher density regions, as demonstrated in analytic models (Bi & Davidsen 1997; Schaye 2001) and simulations (Cen et al. 1994; Weinberg et al. 1996; Theuns et al. 1998). The temperature of this gas is set by the balance between photo-heating and adiabatic expansion and compression, resulting in a well-defined temperature-density relation,  $T = T_0(\rho/\langle\rho\rangle)^{\gamma-1}$ , where  $\gamma \approx 1$  immediately after reionisation, and  $\gamma \rightarrow 1 + 1/1.7$  asymptotically long after reionisation (Hui & Gnedin 1997; Theuns et al. 1998). Patchy reionisation of either H I or He II may lead to patchy photo-heating (Abel & Haehnelt 1999; McQuinn et al. 2009) and hence spatial fluctuations in the temperature, but these have not (yet) been detected (Theuns et al. 2002; McQuinn et al. 2011).

Feedback from star formation is an essential ingredient in models of galaxy formation and might impact the IGM (Theuns et al. 2001). For example, elements synthesised in stars are detected in the IGM (Cowie et al. 1995), even at low densities (Schaye et al.

2003). If, as is likely, this enrichment is due to galactic winds, then the IGM might not be ideal for making *cosmological* inferences. Fortunately, it appears that the effects of such winds are relatively small (Theuns et al. 2002; McDonald et al. 2005; Viel et al. 2013). However, current simulations still struggle somewhat to reproduce the detected enrichment patterns as a function of density (Aguirre & Schaye 2005; Wiersma et al. 2010, 2011) and hence may underestimate the effects of galaxy formation on the IGM.

Accurate measurements of the evolution of  $T_0$  and  $\gamma$  are interesting, since they could constrain the epoch of H I and He II reionisation (Theuns et al. 2002), but also because the formation of small galaxies is quenched due to a reduction of atomic cooling (Efsthathiou 1992) following reionisation, and because photo-heating evaporates the gas out of shallow dark matter potentials (Okamoto et al. 2008). This is a crucial ingredient in models of galaxy formation, in particular for dwarf galaxies and satellites (Benson et al. 2002; Sawala et al. 2014). The thermal state of the IGM is also an important ‘nuisance’ factor when using the Lyman- $\alpha$  forest to study the nature of the dark matter (Boyarsky et al. 2009), or to determine the amplitude of the UV-background (Bolton et al. 2005).

Schaye et al. (1999) described a method for inferring  $(T_0, \gamma)$  from the line width versus column density ( $b - N_{\text{HI}}$ ) scatter plot of Lyman- $\alpha$  lines, when the Lyman- $\alpha$  forest is fitted as a sum of Voigt profiles for example using VPFIT<sup>1</sup>. This improved and extended the work of Theuns et al. (1998) who had noted that the small  $b$  cut-off of the  $b$ -parameter distribution was sensitive to temperature. Schaye et al. (2000), and Ricotti et al. (2000); Bryan & Machacek (2000) and McDonald et al. (2001) used this to measure the  $T - \rho$  relation at redshifts  $z \sim 2 - 4$ .

<sup>1</sup> VPFIT is developed by R. Carswell and J. Webb, <http://www.ast.cam.ac.uk/ffc/vpfit.html>.

\* E-mail: garzilli@lorentz.leidenuniv.nl

Since then other measurements have been performed over a larger redshift range, some also based on Voigt profile fitting (Rudie et al. 2012), or based on small-scale power spectrum or wavelet analysis (Theuns & Zaroubi 2000; McDonald et al. 2000; Zaldarriaga et al. 2001; Theuns et al. 2002; Viel et al. 2009; Lidz et al. 2010; Garzilli et al. 2012), transmission PDF statistics (Bolton et al. 2008; Viel et al. 2009; Calura et al. 2012; Garzilli et al. 2012) and more recently on curvature statistics (Becker et al. 2011). Current measurements still have large error bars for  $T_0$  and constraints are even poorer for  $\gamma$ . Bolton et al. (2008) and Viel et al. (2009) even argue that  $\gamma < 1$ , which is difficult to understand from a theoretical point of view (McQuinn et al. 2009; Compostella et al. 2013).

The basis for all these methods is the assumption that line widths are relatively directly related to the temperature of the absorbing gas. However, it is in fact well known that - in the simulations - the absorption lines are invariably wider than just the thermal broadening. In addition there is a very large *scatter* in line widths at a given density, in the sense that some lines are *much* wider than the thermal width at a given density.

In this paper we present a new model for measuring and interpreting line widths, which builds on previous work. In our model lines are *always* broader than the thermal width, and we also explain the origin of the large scatter. We describe how the model can be used to infer not just  $T_0$  and  $\gamma$ , but also to constrain the size of the absorbers. This paper is organised as follows. In Section 2 we begin by contrasting the ‘fluctuating Gunn-Peterson’ description of the IGM with one where the absorption is due to a distribution of individual absorbers. We do so in order to establish notation and explain the motivation for our modelling. We discuss the properties of these absorbers, and in particular what sets their line widths. In Section 3 we introduce the simulations and use them to test the basic assumptions of the model. In Section 4 we examine whether the method can in principle be applied to observables. We present a summary of our conclusions in Section 5.

## 2 LYMAN- $\alpha$ SCATTERING IN THE IGM

A Lyman- $\alpha$  photon has a large cross section for exciting a H I atom from the ground state to the electronic  $n = 2$  level. When the atom falls back to the ground state, the photon is in general not re-emitted in the same direction. This scattering of Lyman- $\alpha$  photons out of the sight-line to a source is usually called Lyman- $\alpha$  ‘absorption’ and is quantified by the optical depth  $\tau$ , such that  $\exp(-\tau)$  is the resulting transmission - often referred to as the ‘flux’. We begin by calculating  $\tau$  for the IGM in two very different approximations.

### 2.1 The fluctuating Gunn-Peterson approximation

The Lyman- $\alpha$  optical depth observed at wavelength  $\lambda$  due to a *homogeneous* density distribution of neutral hydrogen with proper number density  $n_{\text{HI}}(z)$  in a cosmological setting, is given by

$$\tau(\lambda) = \int_0^{z_s} \sigma \left( \frac{(1+z)c}{\lambda} \right) n_{\text{HI}}(z) \frac{c dz}{(1+z)H(z)}, \quad (1)$$

(Gunn & Peterson 1965), where  $c$  is the light speed,  $\sigma$  is the Lyman- $\alpha$  cross section as function of frequency  $c/\lambda$ ,  $\lambda_0 \sim 1215.67 \text{ \AA}$  is the laboratory wavelength of the H I  $n = 1 \rightarrow 2$  transition, and  $H(z)$  the Hubble parameter at redshift  $z$  (see *e.g.* Meiksin 2009 for an extensive recent review of the relevant physics). Since the proper length  $dr$  corresponding to a redshift interval  $dz$  is  $dr = c dz / [(1+z)H(z)]$ , Eq. (1) simply states that a column of neutral

hydrogen  $n_{\text{HI}} dr$  contributes  $\sigma n_{\text{HI}} dr$  to the optical depth. Here,  $z_s$  is the redshift of the source.

Structure formation results in the growth of the amplitude of fluctuations in the density field. This causes the transmission as a function of wavelength to be in the form of relatively well defined ‘absorbers’ (Lynds 1971), collectively referred to as the Lyman- $\alpha$  forest (Weymann et al. 1981). At higher  $z$  the absorbers blend into each other such that regions of near zero transmission are interrupted by small clearings (a Lyman- $\alpha$  jungle), whereas at low  $z$  a near unity transmission is occasionally interrupted by a single absorber (a Lyman- $\alpha$  savannah).

When the gas is in photo-ionisation equilibrium, the optical depth resulting from a single absorber depends on density and temperature as  $\tau \propto \rho^2 T^{-0.76}$  where  $T^{-0.76}$  is (approximately) the temperature dependence of the hydrogen recombination rate. Taking peculiar velocities and thermal broadening into account as well results in the ‘fluctuating Gunn-Peterson’ approximation (Gnedin & Hui 1998; Croft et al. 1998) for the optical depth,

$$\tau(v) = \sum \int \frac{n_{\text{HI}}}{1+z} \left| \frac{dv'}{dx} \right|^{-1} \sigma_\alpha dv' \quad (2)$$

$$\sigma_\alpha = \sigma_0 \frac{c}{b\sqrt{\pi}} e^{-(v-v_0)^2/b^2}$$

$$\sigma_0 \equiv \left( \frac{\pi e^2}{m_e c} \right) \left( \frac{1}{4\pi\epsilon_0} \right) f_{lu}, \quad (3)$$

$$(4)$$

where  $e$  is the electric charge of an electron,  $m_e$  the electron mass,  $\epsilon_0$  is the electric constant, and  $f_{lu}$  is the upward oscillator strength (usually referred to as the ‘ $f$ ’ factor). The velocity coordinate  $v$  is related to the observed frequency through

$$\nu = \frac{\nu_0}{1+\bar{z}} \left( 1 - \frac{v}{c} \right), \quad (5)$$

where  $\nu_0 = c/\lambda_0$  is the Lyman  $\alpha$  frequency in the lab,  $\bar{z}$  is a mean redshift of interest. The integral is now over the profile of each individual ‘absorber’, the Jacobian  $|dv'/dx|$  takes into account velocity gradients, and the sum is over many absorbers. The Gaussian profile expresses (thermal) broadening and neglects the intrinsic line profile of the Lyman- $\alpha$  transition.

The temperature of the IGM is set by the balance between photo-ionisation heating of the UV-background, the adiabatic compression, and the adiabatic expansion of the universe. This introduces the temperature-density relation which, for densities around the cosmic mean, is approximately a power-law (Hui & Gnedin (1997))

$$T = T_0 \left( \frac{\rho}{\langle \rho \rangle} \right)^{\gamma-1} \equiv T_0 \Delta^{\gamma-1}, \quad (6)$$

where  $T_0$  is the temperature at the mean density  $\langle \rho \rangle$ , and  $\Delta$  the density contrast. Close to reionisation the gas is nearly isothermal,  $\gamma \sim 1$ , whereas asymptotically long after reionisation,  $\gamma \rightarrow 1 + 1/1.7$  (Hui & Gnedin 1997; Theuns et al. 1998). The  $z < 5$  Lyman- $\alpha$  forest can be reasonably well described as being a sum of individual absorbers, and we discuss their properties next.

### 2.2 Lyman- $\alpha$ absorbers

The simplest single absorber, a cloud of uniform density, temperature  $T$ , and with HI column density  $N_{\text{HI}}$ , produces a Gaussian

absorption line when neglecting the intrinsic Lyman- $\alpha$  line profile,

$$\begin{aligned}\tau(v) &= \tau_0 e^{-(v-v_0)^2/b^2} \\ \tau_0 &= \frac{\sigma_0 c N_{\text{HI}}}{\sqrt{\pi} b}.\end{aligned}\quad (7)$$

The line centre is at velocity  $v_0 = c \log(1+z) + v_{\text{pec}}$  for an absorber at redshift  $z$ , with peculiar velocity  $v_{\text{pec}}$ . Equation (7) can be obtained from Eq. (2) in the limit that  $n_{\text{HI}}|du/dx|^{-1}$  has a peak at  $u_0$  with a width in velocity space much less than the thermal width  $b$ , this is the so called narrow-maximum limit in Hui et al. (1997).

Several processes contribute to the broadening  $b$  of the line. The thermal broadening is given by

$$b_T^2 = \frac{2k_B T}{m_H} \quad (8)$$

$$= (12.8 \text{ km s}^{-1})^2 \left( \frac{T_0}{10^4 \text{ K}} \right) \Delta^{\gamma-1}, \quad (9)$$

with  $m_H$  the mass of the hydrogen atom and  $k_B$  Boltzmann's constant. The second line applies to gas on a temperature-density relation of the form  $T = T_0 \Delta^{\gamma-1}$  from Eq. (6).

A more realistic absorber should have at least *some* density structure. Assuming the line centre corresponds to a maximum  $\rho_0$  in density, a Taylor-expansion in velocity space yields to lowest order

$$\log(\rho(v)/\rho_0) = -(v-v_0)^2/b_\rho^2. \quad (10)$$

Real Lyman- $\alpha$  absorbers may not be strongly peaked, but note that what matters is not the density profile of the absorber, but its *neutral* density. Since  $n_{\text{HI}} \propto \rho^2$  for isothermal gas in ionisation equilibrium, peaks in neutral gas density are better defined than the maxima in density - as we will show using numerical simulations in the next section.

The absorption profile of this more realistic absorber is then a convolution of the Gaussians of Eqs. (7) and (10), *i.e.* another Gaussian, with width

$$b^2 = b_T^2 + b_\rho^2. \quad (11)$$

What sets the velocity extent  $b_\rho$ ? Absorbers at modest density contrast are expanding at a rate somewhat smaller than the Hubble rate, suggesting that  $b_\rho \sim \lambda H$ , with  $\lambda$  the physical extent of the structure along the line of sight. Based on the papers of Schaye (2001) and Gnedin & Hui (1998), we expect the physical extent of the absorbers to be of the order the Jeans length  $\lambda_J$ , defined in Eq. (15) below (see also Miralda-Escude & Rees 1993). In the model of Schaye (2001) absorbers are assumed to be locally in near-hydrostatic equilibrium, with column density related to density by

$$N_{\text{HI}} \sim n_{\text{HI}} \lambda_J, \quad (12)$$

and hence 'size'  $\lambda_J$ . Because the dynamical times of low-density absorbers are long, the structures are still expanding, and are not exactly in hydrostatic equilibrium but are lagging behind. This is encapsulated by the 'filtering scale'  $\lambda_F = 2\pi/k_F$  of Gnedin & Hui (1998) which depends on the full history of the Jeans length over cosmic time and hence depends on the thermal history, not just the instantaneous Jeans length. The physical 'extent' of the absorber in this approximation is then  $\sim 1/k_F$ , therefore its velocity extent is  $\sim H(z)/(k_F(1+z)) \approx \lambda_J H(z)/(2\pi)$ , where  $\lambda_J$  is now the physical - as opposed to co-moving - extent of the absorber. This

then motivates us to write

$$b_\rho = f_J \frac{\lambda_J H(z)}{2\pi}. \quad (13)$$

The factor  $f_J$  parametrises the time-dependent Jeans smoothing of the gas density profiles, and we expect it to be of order unity.

For gas on a power-law temperature-density, the values of the sound speed and Jeans length depend on the density contrast  $\Delta$  as

$$c_s^2 = \left( \frac{5k_B T}{3\mu m_H} \right) \quad (14)$$

$$= (15.2 \text{ km s}^{-1})^2 \left( \frac{T_0}{10^4 \text{ K}} \right) \Delta^{\gamma-1}$$

$$\lambda_J^2 \equiv \frac{c_s^2 \pi}{G \rho_t}$$

$$= \frac{(5/3)\pi k_B T_0}{\mu m_H G \rho_t(z=0)(1+z)^3} \Delta^{\gamma-2}$$

$$\approx (260 \text{ kpc})^2 \left( \frac{T_0}{10^4 \text{ K}} \right) \Delta^{\gamma-2}$$

$$\times \left( \frac{\Omega_m}{0.307} \right)^{-1} \left( \frac{h}{0.6777} \right)^{-2} \left( \frac{1+z}{4} \right)^{-3}, \quad (15)$$

where  $\mu$  is the mean molecular weight and  $\rho_t = \rho_{\text{dm}} + \rho_b$  is the total density, dark matter plus baryons,  $G$  is the gravitational constant,  $\Omega_m$  is the matter density. We write the Hubble parameter at redshift  $z$  as  $H(z)$ , with  $H_0 = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}$  its value today. The numerical values assume  $\mu = 0.59$ , that dark matter and gas move together, and for reference we use the Planck Collaboration et al. (2013) values for the cosmological parameters. Numerically, the broadening due to Jeans smoothing is thus of order

$$b_\rho^2 = f_J^2 \frac{10 k_B T}{9\mu m_H \Delta} \frac{H^2(z)}{H_0^2 \Omega_m (1+z)^3} \quad (16)$$

$$\approx (11.1 \text{ km s}^{-1})^2 \left( \frac{f_J}{0.88} \right)^2 \left( \frac{T_0}{10^4 \text{ K}} \right) \Delta^{\gamma-2},$$

independent of redshift in the high-redshift approximation  $H(z) \propto (1+z)^{3/2}$ . The model of Hui & Rutledge (1999) uses a very similar reasoning to infer line widths;  $b_\rho$  is usually referred to as the 'Hubble broadening'. We demonstrate below that a value of  $f_J \sim 0.88$  describes our numerical simulations well at  $z = 3$ .

The total broadening,  $b$ , of the absorber is then

$$\begin{aligned}b^2 &= b_T^2 \left[ 1 + f_J^2 \frac{5}{9\mu} \Delta^{-1} \right] \\ &\approx b_T^2 \left[ 1 + 0.75 \left( \frac{f_J}{0.88} \right)^2 \Delta^{-1} \right].\end{aligned}\quad (17)$$

The line-widths of weak absorbers ( $\Delta \lesssim 1$ ) are increasingly dominated by Hubble broadening,  $b^2 \sim b_T^2 \Delta^{-1} \propto \Delta^{\gamma-2}$ , and hence the filtering scales determines their widths. Hubble expansion plays a smaller role for stronger lines, with line widths ultimately set by thermal broadening alone. Equation (17) smoothly interpolates between filtering and pure thermal broadening. However, we note that for typical  $z \sim 3$  absorbers with  $\Delta \sim 3$ , Hubble broadening still plays a significant role, increasing  $b$  by almost 70 per cent compared to thermal broadening alone. It might be possible to measure the extent of absorbers in real space using close quasar pairs (see Rorai et al. (2013); the impact of Jeans length on quasar pairs was also considered in Peebles et al. (2010)). The extra broadening discussed here is a longitudinal measure of the Jeans length, as opposed to the transverse Jeans length probed with pairs.

The Gaussian shape of the absorber directly relates column

density and width,

$$b^2 = -\frac{2\tau^2}{\tau\tau'' - \tau'^2} = -\frac{2\tau_0}{\tau_0''} \quad (18)$$

$$N_{\text{HI}} = \frac{\sqrt{\pi}}{\sigma_0 c} b\tau_0, \quad (19)$$

where  $\tau' \equiv d\tau/dv$ ,  $\tau'' \equiv d^2\tau/dv^2$ , and the subscript 0 indicates that the quantity is to be evaluated at the (local) maximum of the optical depth. We will use these relations below. Equations (18-19) follow directly from Eqs. (7).

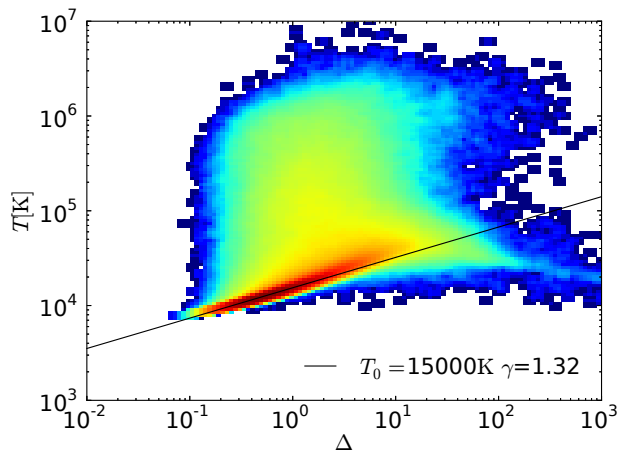
### 3 BROADENING OF LYMAN- $\alpha$ ABSORBERS IN SIMULATIONS

#### 3.1 Mock spectra from simulations

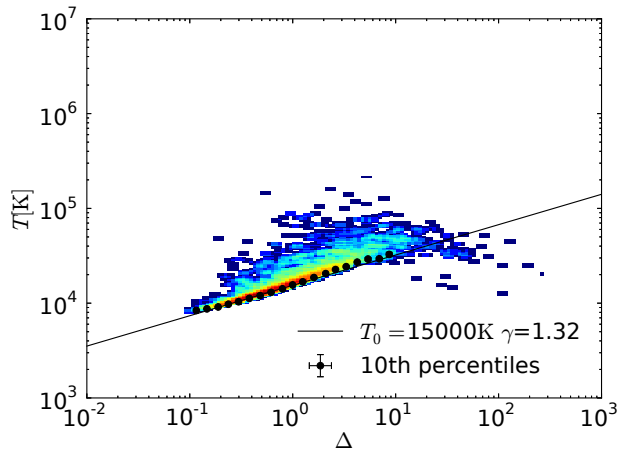
We use the REFERENCE model of the OWLS suite (Schaye et al. 2010) to test our model for line broadening. Briefly, the OWLS suite is a set of cosmological hydrodynamical simulations performed with the GADGET-3 code, an improved version of GADGET-2 last described by Springel (2005), with subgrid models for unresolved galaxy formation physics, in particular star formation, and feedback by and enrichment from stars, as described in Schaye & Dalla Vecchia (2008), Dalla Vecchia & Schaye (2008) and Wiersma et al. (2009), respectively. Although efficient feedback from star formation results in outflows from galaxies in this simulation, such winds do not have a significant impact on the Lyman- $\alpha$  forest for reasons discussed in Theuns et al. (2002), see also Viel et al. (2013). Most relevant for this paper is that we simulate a  $\Lambda$  cold dark matter model using  $512^3$  SPH particles that represent the gas, and an equal number of dark matter particles. The computational volume is a periodic box of co-moving size  $25h^{-1}$  Mpc on a side. The masses of the gas particles ( $\sim 2 \times 10^6 h^{-1} M_\odot$ ) are sufficiently low to obtain numerical convergence for the line widths (Theuns et al. 1998). Photo-heating of the gas in the presence of an optically thin UV/X-ray background from Haardt & Madau (2001) is implemented using interpolation tables from Wiersma et al. (2009).

Photo-heating, adiabatic processes, shocks from structure formation and feedback from galactic winds and radiative cooling produce the temperature-density relation shown in Fig. 1a at redshift  $z = 3$ . Most of the gas is on a tight temperature-density relation,  $T = T_0 \Delta^{\gamma-1}$ , with  $T_0 \approx 15000$  K and  $\gamma \approx 1.32$  for  $\Delta \lesssim 5$ , and this gas is the main reservoir responsible for producing the Lyman- $\alpha$  forest. There is gas at significantly higher temperatures at these densities, which results from structure formation and feedback. The increased importance of radiative cooling decreases the amount of hot gas at higher densities, and an equilibrium sequence where radiative cooling and photo-heating balance results in the appearance of significant amounts of gas at  $T \sim 10^4$  K. This higher density gas results in Lyman-limit and damped Lyman- $\alpha$  absorbers (e.g. Altay et al. 2011). Self-shielding is important for this higher-density gas, and the impact of feedback becomes more important (Altay et al. 2013).

Given a snapshot of the simulation at some redshift (we use  $z = 3$  in this paper), we generate mock Lyman- $\alpha$  spectra using the method described in Theuns et al. (1998) but using the interpolation tables of Wiersma et al. (2009) to relate total to neutral hydrogen density. We generate these spectra at very high resolution using pixels of width  $\sim 2 \text{ km s}^{-1}$ , much narrower than any of the absorption features that appear. Importantly, we neither add noise nor instrumental broadening to mimic observed spectra - limiting our



(a)

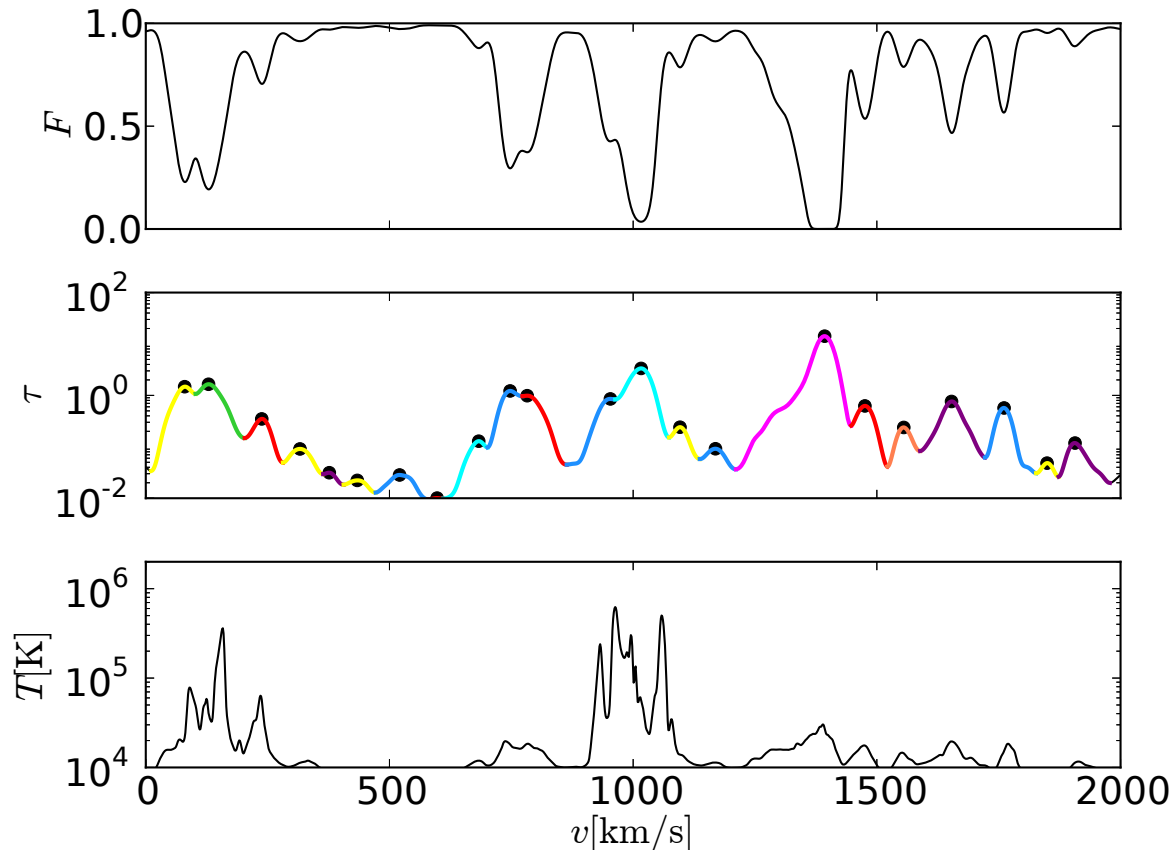


(b)

**Figure 1.** Mass-weighted (top panel) and optical depth weighted (bottom panel) temperature-density relation in the REFERENCE OWLS run at  $z = 3$ . Values in the bottom panel are evaluated at the locations of maxima in the optical depth. Colours are a measure of the density of points in these plots, and  $\Delta = \rho/(\rho_0)$ . In the lower panel the black dots indicate the median temperature in bins of  $\Delta$  with error bars computed by bootstrap resampling. The black solid line in both panels show the power-law  $T = T_0 \Delta^{\gamma-1}$  that best fits the black dots for  $\Delta < 3$ . There is a much larger scatter in  $T$  vs  $\Delta$  in the top panel than in the bottom panel, but this shocked gas contributes little to the Lyman- $\alpha$  forest. The lower-envelope of points in the  $T - \Delta$  relation in the bottom panel mirrors the temperature-density relation of the lower-envelope in the top panel, demonstrating that the optical-depth weighted temperature at the optical depth maxima of absorption lines follows very closely the  $T - \Delta$  relation of the photo-heated IGM.

analysis to comparing our model to idealised observations. Each sight line is parallel to a coordinate axis of the simulation volume, and we use velocity pixels of size  $v = H(z)L/[(1+z)N]$ , where  $L = 25h^{-1}$  Mpc is the co-moving box size, and  $N$  the number of pixels. We calculate  $\tau(v)$ ,  $T(v)$  and  $\rho(v)$ , where  $T$  and  $\rho$  are the *optical depth weighted* temperature and density, respectively (see Schaye et al. 1999). Unless explicitly stated, we will always use optical depth weighted quantities for variables measured from spectra.

In Fig. 1b we plot values of optical depth weighted temperature versus density, for pixels that correspond to local maxima in



**Figure 2.** A stretch of a simulated spectrum at redshift  $z \sim 3$ . From top to bottom we show the transmission  $F \equiv (-\tau)$ , optical depth  $\tau$ , and optical depth weighted temperature,  $T$ , as function of velocity  $v$ . Individual lines are regions of the spectra between two minima in optical depth, they are coloured using different colours, with a filled circle at the maximum  $\tau$  for each line. Note that  $T$  is not constant over a line.

the optical depth from these mock spectra. Most points lie close to the lower-envelope of the points in the scatter plot, which tracks the same  $T - \Delta$  relation as the lower envelope in Fig. 1a. The scatter around this relation is much less in Fig. 1b, because the shocked gas that causes the scatter to higher  $T$  in Fig. 1a contributes little to the optical depth. At higher  $\Delta \gtrsim 10$ ,  $T$  decreases with increasing  $\Delta$ , reflecting the effects of radiative cooling also apparent in Fig. 1a.

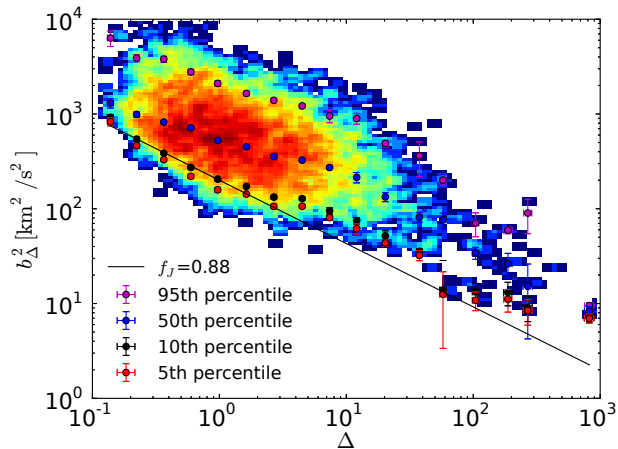
### 3.2 Identifying individual absorbers

We identify individual absorbers (or ‘lines’) with regions of the spectrum between two minima in the optical depth, dissecting the whole spectrum in a set of unique stretches. An example spectrum is shown in Fig. 2, with different lines coloured differently, and with diamonds at the locations of local maxima in  $\tau$ . Unsurprisingly, the temperature varies across a line, and hence assigning a single temperature (or density) to a line is necessarily somewhat ambiguous. We could decide to associate to a line the maximum, the (weighted) mean or the value corresponding to the maximum of the line for any given physical quantity. In the following, where not specified otherwise, we will associate to the line the value of the physical quantity that corresponds to the optical depth weighted value evaluated at the local maximum in optical depth. This method for identifying lines is not directly applicable to observed spectra in the presence of noise. Such spectra should be smoothed to avoid incor-

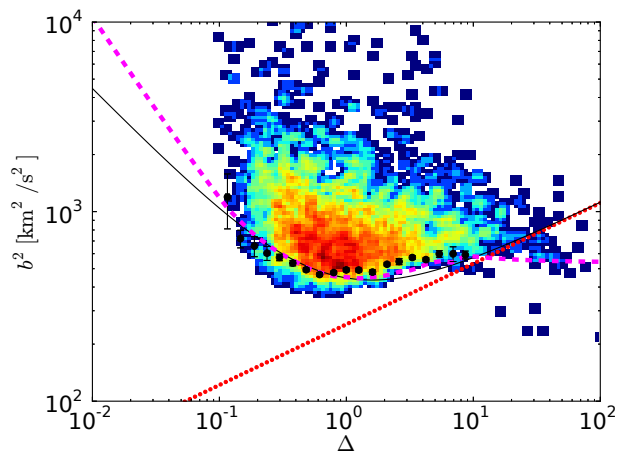
rectly identifying noise features with small lines, however this may lead to missing true weak absorbers. Preliminary analysis shows that fitting a Gaussian over a set of contiguous pixels to calculate the derivatives that appear in Eq. (18) is promising, we will report on this in a future paper.

As demonstrated in Fig. 1, the (real-space) temperature-density relation of the IGM in the simulation translates approximately to a power-law (optical depth weighted) temperature-density relation for the lines generated in mock spectra. Most lines follow  $T = T_0 \propto \Delta^{\gamma-1}$  up to  $\Delta \lesssim 3$ , with some scatter to significantly higher  $T$ , but none scatter significantly *below* this line. However for  $\Delta \gtrsim 3$ , the median temperature does fall below the power law fit: this is because this gas cools radiatively, and this results in a decrease in  $T$  with higher  $\Delta$  for the real-space temperature-density relation as well. Note further that the  $T - \Delta$  relation is only approximately a power-law, with evidence for some curvature even over a relatively narrow  $\Delta$  range. Finally, note that in the absence of noise, we see that the majority of lines have  $0.1 \lesssim \Delta \lesssim 10$ .

The optical depth profile of several of the lines in Fig. 2 appear reasonably Gaussian in shape near their maxima, and we will investigate to what extent their widths follow Eq. (17). However it is also clear that some lines blend with nearby lines, and we will show below that this ‘clustering’ impacts how well we can infer the underlying  $T - \Delta$  relation from spectra - even in the absence of noise.

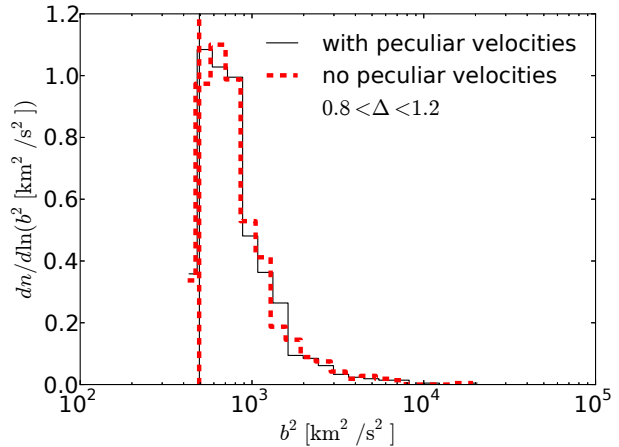


**Figure 3.** Velocity extent of absorbers in real space  $b_\Delta$  versus density contrast  $\Delta$  for lines identified in mock spectra without peculiar velocities. Colours indicate density of points in the plot. The 5th, 10th, 90th and 95th percentiles of  $b_\Delta$  in bins of  $\Delta$  are shown as symbols with bootstrap errors. Line widths *decrease* with increasing density, opposite to the case of thermal broadening. This behaviour is captured by the solid black line, the broadening  $b_\rho$  from Eq. (13) for  $f_J = 0.88$ , which fits the lower percentiles of the  $b_\Delta - \Delta$  relation well in shape and amplitude for  $\Delta \lesssim 3$ .



**Figure 4.** Total broadening  $b^2 \equiv -2\tau/\tau''$  of absorbers at line centres as a function of density contrast  $\Delta$  for lines identified in mock spectra. Colours indicate density of points in the plot. *Black dots* with bootstrap error bars denote the location of the 10th lower percentile of  $b$  in bins of  $\Delta$ . The *dotted red* line is the thermal width from Eq. 9 for the temperature-density relation from Fig. 1a; lines are invariably wider than  $b_T$  except at  $\Delta \gtrsim 10$ . The *solid black* line is the model of Eq. (11) with  $f_J = 0.88$ , for the same  $T - \Delta$  power-law relation. It describes the black dots very well up to  $\Delta \sim 3$ , above which the simulation's  $T - \Delta$  relation is not well fit by a power-law. The *dashed magenta* line is again the model of Eq. (11) but now using the measured  $T - \Delta$  relation rather than a power-law fit: it captures the down-turn in  $b$  for  $\Delta \gtrsim 3$  due to the onset of radiative cooling.

We suggested that line broadening has a contribution  $b_\rho$  due to the spatial extent of the absorber as described by Eq. (13). To test this directly, we should plot  $b_\rho$  versus  $\Delta$ . However, maxima in  $\tau$  do not correspond directly to maxima in  $\Delta$  because of peculiar velocities. Therefore, to unambiguously identify a line with a density structure, we generate mock spectra setting peculiar ve-

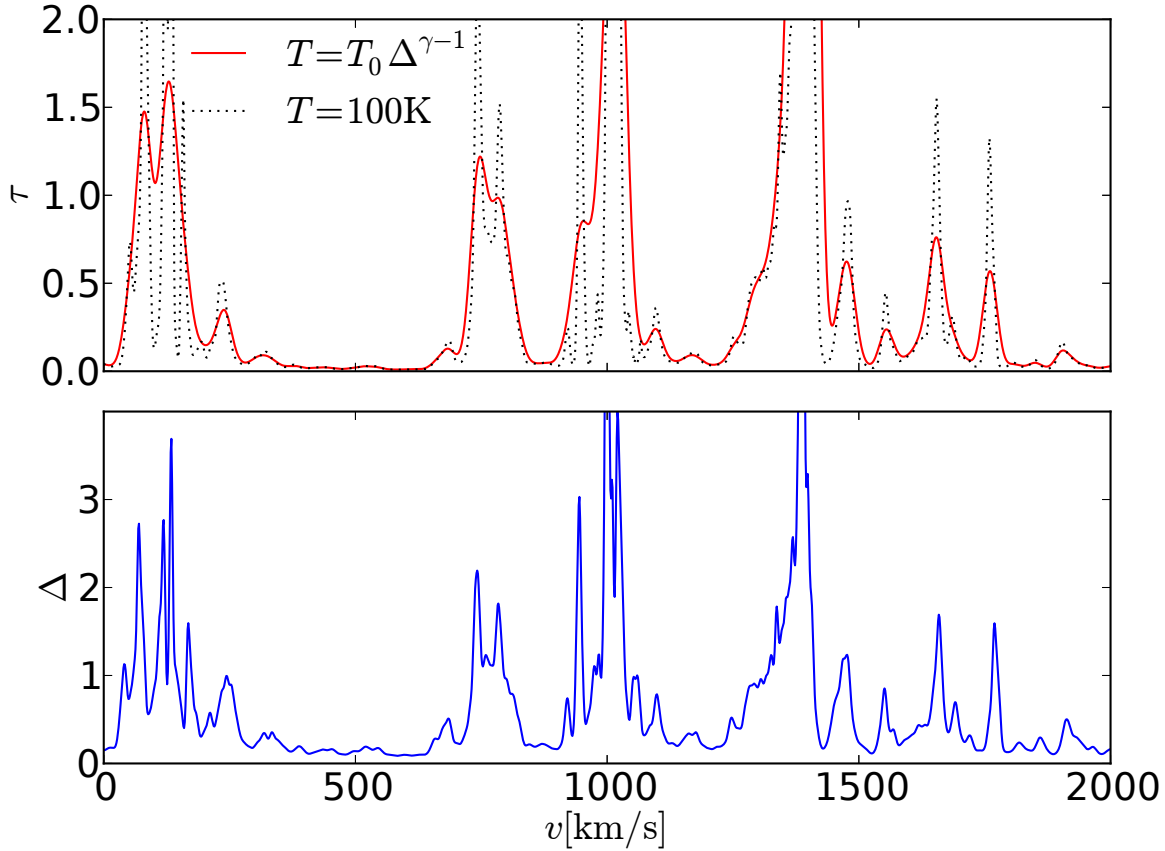


**Figure 5.** Histogram of line broadening  $b^2 = -2\tau/\tau''$  for simulations with and without peculiar velocities, shown as *solid black* and *dashed red* lines, respectively. The 10th percentiles of narrowest lines are indicated by vertical lines. The impact of peculiar velocities on the distribution of  $b$  values is small.

locities to zero. We later show that this has very little effect on the line broadening (see also Fig. 5). For each line identified in these mock spectra, we can now unambiguously identify the corresponding density stretch that gives rise to that line. However, the shape of the density structure of lines is often complex and it is not immediately apparent what extent to associate with a given line. We decided to use the following prescription. For each line we determine its start and end velocity,  $v_1$  and  $v_2$ , the location ( $v_{\max}$ ) and height ( $\Delta_{\max}$ ) of the maximum density, and the integral  $I \equiv \int_{v_1}^{v_2} \Delta(v) dv$ . We now determine the value of  $b_\Delta$  of the Gaussian profile,  $G(\Delta) = \Delta_{\max} \exp(-(v - v_{\max})^2/b_\Delta^2)$ , for which  $\int_{v_1}^{v_2} G(v) dv = I$ . In other words, we define the extent of the absorber,  $b_\Delta$ , as the width of the Gaussian that has the same integral and maximum value as the line itself. If the line were a Gaussian, then  $b_\Delta$  would simply be its width. Further details on the fitting procedure can be found in Appendix B. We expect that  $b_\Delta \sim b_\rho$  from Eq. (13).

In Fig. 3 we plot  $b_\Delta^2$  versus  $\Delta$  for all lines identified in mock spectra generated while ignoring peculiar velocities. As before, colour encodes the number density of lines in this plot, and 5th, 10th, 50th and 95th percentiles are plotted as symbols with bootstrap error bars. Values of  $b_\Delta^2$  *increase* with decreasing  $\Delta$  - the opposite of what is expected from thermal broadening. The solid line indicates  $b_\rho^2$  from Eq. (13) with  $f_J = 0.88$  applied to gas with the  $T - \Delta$  relation  $T = 15000 \text{ K } \Delta^{1.3-1}$  which, as we showed earlier, fits the actual temperature-density relation of the simulation well for  $0.1 \lesssim \Delta \lesssim 3$ . Fig. 3 shows that both the 5th and 10th percentiles follow the solid line well over the same  $\Delta$  range. This implies that  $b_\rho$ , with  $f_J = 0.88$ , is indeed a good estimate of the width of the absorbers in real space.

Given that our estimate of the broadening of the density profile works well, we plot the total line broadening (including peculiar velocities) versus density contrast, and compare it to the model of Eq. (17) (Fig. 4). As before, the colour represents the density of points in the plot. We compute  $b$  from the curvature of the optical depth profile  $\tau(v)$  using Eq. (18); fitting a Gaussian to the pixels close to the maximum yields very similar values of  $b$ . Values of  $b - \Delta$  for individual lines show a larger scatter in  $b$  at given  $\Delta$ , but



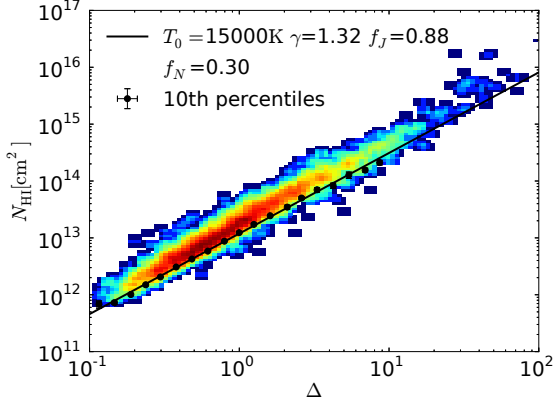
**Figure 6.** Optical depth  $\tau$  (*top panel*) and corresponding density contrast  $\Delta$  (*bottom panel*) for mock spectra for the simulation with the original temperature-density relation (*solid red line*) and for the case where  $T = 100$  K everywhere (*dotted black line*). Clustering of density peaks in real space makes some of the absorption lines much wider than the estimate of  $b$  from Eq. (17), for example the wider lines around  $v \sim 800$  km s $^{-1}$  and  $v \sim 1000$  km s $^{-1}$ . Thermal broadening and Hubble broadening together smooth the individual peaks seen in the underlying density field into one wide line, whose width cannot be described accurately by eq. 17.

there is a well-defined lower envelope below which few lines appear. This lower envelope is traced by the 10th percentile of  $b$  values in narrow bins of  $\Delta$ , plotted using black symbols with bootstrap errors. The red dashed line indicates the thermal broadening,  $b_T$ , from Eq. (9), using the power-law fit to the temperature-density relation measured in the simulation. For  $\Delta \lesssim 10$ , the measured broadening along the cut-off is significantly wider than  $b_T$ , with the discrepancy between the two increasing with decreasing  $\Delta$ . The solid black line shows  $b_T^2 + b_\rho^2$  as a function of  $\Delta$  from Eq. (17), again using the best-fitting power-law temperature-density relation, and adopting the value  $f_J = 0.88$  found above. It captures accurately the dependence of the broadening on  $\Delta$  for the lower envelope of the absorption lines traced by the 10th percentile for  $\Delta \lesssim 3$ , in particular reproducing the measured *upturn* in  $b^2$  for  $\Delta \lesssim 1$ . Because the solid black line assumes that the relation between temperature and density is a power law, it does not describe the simulated  $b - \Delta$  relation well above  $\Delta \sim 3$ . The dashed black line uses the median value of the temperature at a given density for absorption lines when calculating the Jeans smoothing term  $b_\rho$  in Eq.(13), rather than a power-law. It does much better in capturing the downturn in  $b - \Delta$  at  $\Delta \gtrsim 10$ . We conclude from this plot that the model in which the line broadening is a combination of thermal and Jeans smoothing, Eq. (17) describes the lower-envelope in the  $b - \Delta$  plane well. Note that deviations of the true temperature-density relation from

a power-law, and Jeans smoothing, cause the  $T - \Delta$  relation to be non-monotonic, with a minimum value of  $b$  for  $\Delta \sim 1$  in this particular simulation at this particular redshift.

Peculiar velocities are not the cause of the appearance of lines that are much *wider* than the broadening  $b$  computed from Eq. (17). We compare the distribution of  $b$  values for lines identified in mock spectra with and without peculiar velocities in Fig. 5. The peculiar velocities do not have a large effect on the broadening, if anything they make lines slightly *narrower*, see also Theuns et al. (2000).

Our analytic expression for broadening, Eq. (17), provides only a lower limit on the width of a collection of lines in the Lyman  $\alpha$  forest. The additional broadening can be partially attributed to the underlying density structure in the absorber (Hui & Rutledge 1999), for example in terms of the angle under which the sightline intersects the filament that corresponds to the absorber. However, another important contribution that is not easily captured by linear theory is the *clustering* of density peaks. We illustrate this in Fig. 6, comparing a mock spectrum with the original  $T - \Delta$  relation with the same spectrum now computed assuming  $T = 100$  K everywhere. We see that some individual absorption lines identified in the solid red spectrum have several underlying density maxima. If the temperature of the gas is low, as in the case of the black dotted spectrum, then these individual density maxima give rise to their own absorption lines. But if the gas is sufficiently hot, then thermal



**Figure 7.** The neutral hydrogen column density of absorption lines versus density contrast for the REFERENCE model at  $z = 3$ ; colours are a measure of the density of points in the plot. *Black dots* denote the 10th percentiles of column density in logarithmic bins of  $\Delta$ . The *solid black line* corresponds to the model of Eq. (19), using the value of  $\Gamma$  appropriate for the REFERENCE model with values of  $T_0$ ,  $\gamma$  and  $f_J$  taken from Fig. 4, and for  $f_N = 0.3$ . This value of  $f_N$  is the result of best-fitting Eq. (19) to the 10th percentiles, for  $\Delta \leq 0.3$ . This provides an excellent fit to the results from the simulation over 4 orders of magnitude in column density.

and Hubble broadening can merge multiple individual components into one much wider line. The width of that absorber is not described well by our model. We conclude that it is the clustering of density maxima that gives rise to the lines that are much wider than expression  $b$  from Eq. (17). Another way to demonstrate this is to count the number of individual density peaks for each absorber: we find that absorbers with widths close to that from Eq. (17) typically have only one density maximum, whereas those that are significantly wider correspond to several underlying peaks.

#### 4 INFERRING THE TEMPERATURE-DENSITY RELATION FROM OBSERVABLES

Our description of line broadening so far used  $b$  and  $\Delta$  to relate temperature to density, yet density contrast  $\Delta$  is not directly measurable. Here we investigate whether we can use the column density  $N_{\text{HI}}$  of the line instead. Following the reasoning of Schaye (2001), we make the *Ansatz* that the column density, density, and extent of the line are related by

$$N_{\text{HI}} \sim n_{\text{HI}} f_J \lambda_J. \quad (20)$$

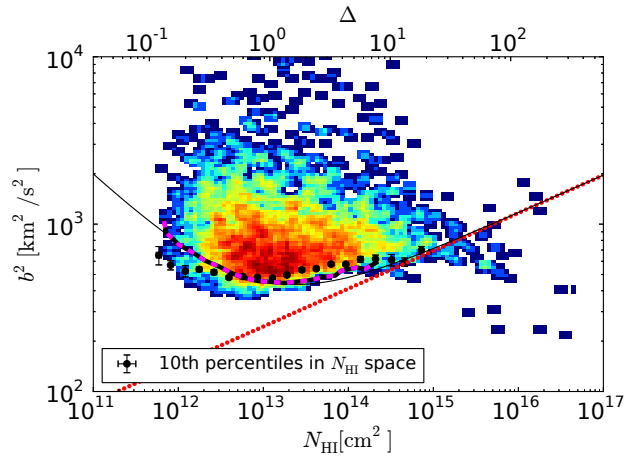
For highly-ionised gas in photo-ionisation equilibrium

$$\begin{aligned} \frac{n_{\text{HI}}}{n_{\text{H}}} &= \alpha_{\text{HII}} n_e \Gamma^{-1} \\ &\approx 5.6 \times 10^{-6} \left( \frac{n_{\text{H}}}{1.2 \cdot 10^{-5} \text{ cm}^{-3}} \right) \\ &\times \left( \frac{T}{10^4 \text{ K}} \right)^{-0.76} \left( \frac{\Gamma}{10^{-12} \text{ s}^{-1}} \right)^{-1}, \end{aligned} \quad (21)$$

where

$$\begin{aligned} \alpha_{\text{HII}} &= \alpha_0 \left( \frac{T}{10^4 \text{ K}} \right)^{-0.76} \\ \alpha_0 &= 4 \times 10^{-13} \text{ cm}^3 \text{ s}^{-1}, \end{aligned} \quad (22)$$

is the value of the case-A recombination coefficient as a function of temperature,  $n_e$  is the electron number density,  $n_{\text{H}}$  is the total hy-



**Figure 8.** The line-width column density  $b^2 - N_{\text{HI}}$  relation in the REFERENCE OWLS model at redshift  $z = 3$ , with broadening determined from Eq. (18) and column density from Eq. (19). The top axis gives the value of the density contrast that corresponds to  $N_{\text{HI}}$  from Eq. (23). Colours denote the density of points in this plot, *black dots* denote the 10th percentile of  $b^2$  in bins of column density with bootstrap error bars. The *dotted red line* is the thermal broadening  $b_T^2$  from Eq. (9) using Eq. (23) to relate column density to density contrast. The *solid black line* is our model for broadening of Eq. (25) using the values  $T_0$  and  $\gamma$  from Fig. 1a,  $f_J = 0.88$  from Fig. 3 and  $f_N = 0.3$  from Fig. 7. The *black line connecting magenta dashes* is Eq. (25) using the  $T-\Delta$  relation measured from the simulation, rather than a power-law fit. The analytic formula describes the minimum line width at a given column density accurately, and works even better if the actual  $T-\Delta$  relation is used rather than a power-law fit.

drogen number density, and  $\Gamma$  the photo-ionisation rate. Combining these with the expression of the Jeans length from Eq. (15) yields

$$\begin{aligned} N_{\text{HI}} &= \frac{3\sqrt{10}}{64\pi} f_N f_J (2-Y)(1-Y) \frac{\Omega_b^2 \alpha_0}{\Omega_m^{1/2} \Gamma} \frac{H_0^3}{(G m_{\text{H}})^2} \\ &\times \left( \frac{k_{\text{B}} 10^4 \text{ K}}{\mu m_{\text{H}}} \right)^{1/2} \left( \frac{T}{10^4 \text{ K}} \right)^{-0.26} (1+z)^{9/2} \Delta^{3/2} \\ &\equiv N_0 \frac{f_N}{0.3} \frac{f_J}{0.88} \left( \frac{T}{10^4 \text{ K}} \right)^{-0.26} \Delta^{3/2}, \end{aligned} \quad (23)$$

where numerically

$$\begin{aligned} N_0 &= 1.4 \times 10^{13} \text{ cm}^{-2} \left( \frac{\Gamma}{10^{-12} \text{ s}^{-1}} \right) \left( \frac{1+z}{4} \right)^{9/2} \\ &\times \left( \frac{\Omega_b}{0.04825} \right)^2 \left( \frac{\Omega_m}{0.307} \right)^{-1/2} \left( \frac{h}{0.6777} \right)^3. \end{aligned} \quad (24)$$

Here,  $f_N$  is the proportionality factor implicit in Eq. (20) which we show below is  $\sim 0.3$ , and the numerical value for  $N_0$  is for very highly ionised ( $\mu = 0.59$ ) primordial gas with the Planck Collaboration et al. (2013) cosmological parameters, and  $Y$  is the primordial helium abundance. For a power-law temperature-density relation,  $T = T_0 \Delta^{\gamma-1}$ , the hydrogen column density depends only weakly on temperature,  $N_{\text{HI}} \propto T_0^{-0.26}$ , and it scales with density contrast as  $\Delta^{1.76-0.26\gamma}$ . We compare the results from Eq. (23) to our simulation in Fig. 7, where we find the best fit of the model to the 10th percentile distribution of the  $N_{\text{HI}}$  versus  $\Delta$ , the justification of this choice is in Appendix A. The model fits very accurately the results from the simulation over nearly four order of magnitude in column density, as also shown in Rahmati et al. (2013) and Tepper-García et al. (2012).



Finally, we combine Eq. (23) with Eq. (17) to obtain a relation between the two observable quantities, namely the column density  $N_{\text{HI}}$  and the lower-envelope of line widths at a given column density,  $b$ , in terms of underlying temperature-density relation of the gas, as

$$\begin{aligned} b_T^2 &= \frac{2k_B T_0}{m} N^{(\gamma-1)/(1.76-0.26\gamma)} \\ b^2 &= b_T^2 \left[ 1 + 0.75 \left( \frac{f_J}{0.88} \right) N^{-1/(1.76-0.26\gamma)} \right] \\ N &\equiv \left( \frac{N_{\text{HI}}}{N_0} \right) \left( \frac{T_0}{10^4 \text{K}} \right)^{0.26} \left( \frac{f_N}{0.3} \right)^{-1} \left( \frac{f_J}{0.88} \right)^{-1}. \end{aligned} \quad (25)$$

This relation is compared to the  $b - N_{\text{HI}}$  distribution of the simulation in Fig. 8. Our model for line broadening does well in predicting the minimum line width as a function of column density, in particular predicting correctly that line width is not a monotonic function of column density. At low  $\Delta$ , line widths increase as Jeans smoothing becomes more important, whereas at high  $\Delta$ , line widths decrease due to radiative cooling. The underlying  $T - \Delta$  relation of the simulation is only approximately a power law: using the relation from the simulation directly in Eq. (23) improves the agreement between model and simulation further, in particular it might be possible to measure the temperature of denser gas,  $\Delta \sim 10$ . We illustrate the sensitivity of the  $b - N_{\text{HI}}$  relation to small changes in  $T_0$  and  $\gamma$ , as well as the fitting parameters  $f_J$  and  $f_N$ , in Appendix C.

Our model overestimates line broadening of very weak lines,  $N_{\text{HI}} \lesssim 10^{12.5} \text{ cm}^{-2}$ , yet it correctly predicts line widths at low density contrast (see Fig. 4). In Appendix A we show that this is because line width biases how well a line of given column can be associated with a given density contrast. In practise these lines are too weak to be detected in observed spectra.

## 5 CONCLUSIONS

We have presented an analytic description of the broadening of lines in the Lyman- $\alpha$  forest, in terms of the temperature, Jeans smoothing and density of the absorbers. Identifying individual absorbers in a spectrum as spectral stretches between two minima in the optical depth, we calculated the line width,  $b$ , from the second derivative of the optical depth at the local maximum in  $\tau$ ,  $b^2 \equiv -2\tau/\tau''$  (Eq.18). We argued that the thermal width of a line,  $b_T$ , combined with its width due to Jeans smoothing,  $b_\rho$ , sets a lower-limit on the total line width,  $b^2 \gtrsim b_T^2 + b_\rho^2$ . We derived an expression for  $b_\rho$  that depends on the temperature  $T$  and the density contrast  $\Delta$  of the absorber, which smoothly interpolates between the filtering length interpretation of line-widths from Hui & Gnedin (1997) at low  $\Delta$ , and the near hydrostatic equilibrium model of Schaye (2001) at higher  $\Delta$ . This model gives a good description of the lower values of  $b$  at given  $\Delta$  (Fig. 4). We also discussed the origin of the lines that are significantly wider than the lower envelope, arguing that they result from the clustering of several individual density maxima that cannot be distinguished in the spectrum due to thermal broadening and Jeans smoothing. Such clustering is not captured by our quasi-linear model.

Based on the model of Schaye (2001), we write an expression relating column density to density contrast. Combining this with our model for line broadening yields an analytic expression for the lower envelope of  $b$  as a function of column density,  $N_{\text{HI}}$

(Eq. 25). This model describes the lower envelope much better than one based on thermal broadening alone (Fig. 8).

In our interpretation of  $b$ , line widths of weaker lines are dominated by the filtering length, which in principle depends on the *thermal history* of the IGM and not just on its instantaneous state. If this is true, then not taking this into account may affect the interpretation of the  $b - N_{\text{HI}}$  cut-off in terms of the underlying temperature-density relation of the gas.

Our analytic description of the line broadening opens the way to a new method for determining the IGM thermal state, and in particular for measuring the ‘longitudinal’ size of the absorbers (the ‘filtering length’), that is complementary to what can be measured from close QSO pairs. This is particularly interesting, because for the first time we have a method to estimate the sizes of the clouds from a single line of sight.

## 6 ACKNOWLEDGEMENT

AG thanks the Astrophysical Sector in SISSA for providing computing facilities during the realization of this paper, Carlos Frenk and Carlton Baugh for supporting her visit at ICC in Durham University, Matteo Viel for supporting her visit at Osservatorio Astronomico di Trieste, and Uroš Seljak for appointing her as a postdoctoral fellow at Ewha Womans University. This work has been partially done during her permanence at the Institute for Early Universe at Ewha Womans University, and it was supported by the WCU grant no. R32-10130 and the Research fund no. 1-2008-2935-001-2 by Ewha Womans University. AG thanks Jeroen Franse for the English revision of part of the manuscript. She also thanks Fedele Lizzi for being her mentor over many years. JS acknowledges support from the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007-2013)/ERC Grant agreement 278594-GasAroundGalaxies. This work was supported by the Science and Technology Facilities Council [grant number ST/F001166/1], and by the Interuniversity Attraction Poles Programme initiated by the Belgian Science Policy Office ([AP P7/08 CHARM]), and used the DiRAC Data Centric system at Durham University, operated by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). This equipment was funded by BIS National E-infrastructure capital grant ST/K00042X/1, STFC capital grant ST/H008519/1, and STFC DiRAC Operations grant ST/K003267/1 and Durham University. DiRAC is part of the National E-Infrastructure. The data used in the work is available through collaboration with the authors.

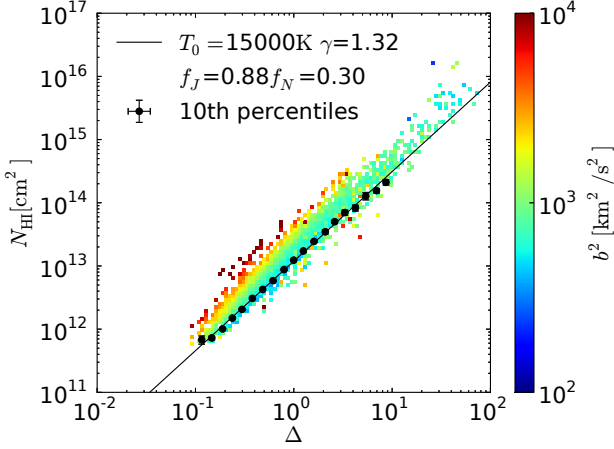
## REFERENCES

- Abel T., Haehnelt M. G., 1999, *ApJ*, 520, L13
- Aguirre A., Schaye J., 2005, in Williams P., Shu C.-G., Menard B., eds, *IAU Colloq. 199: Probing Galaxies through Quasar Absorption Lines* Observational tests of intergalactic enrichment models. pp 289–294
- Altay G., Theuns T., Schaye J., Booth C. M., Dalla Vecchia C., 2013, *MNRAS*, 436, 2689
- Altay G., Theuns T., Schaye J., Crighton N. H. M., Dalla Vecchia C., 2011, *ApJ*, 737, L37
- Becker G. D., Bolton J. S., Haehnelt M. G., Sargent W. L. W., 2011, *MNRAS*, 410, 1096

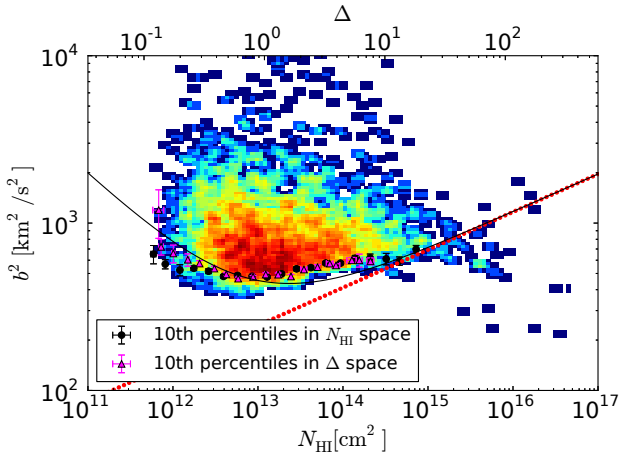
- Benson A. J., Lacey C. G., Baugh C. M., Cole S., Frenk C. S., 2002, *MNRAS*, 333, 156
- Bi H., Davidsen A. F., 1997, *ApJ*, 479, 523
- Bolton J. S., Haehnelt M. G., Viel M., Springel V., 2005, *MNRAS*, 357, 1178
- Bolton J. S., Viel M., Kim T.-S., Haehnelt M. G., Carswell R. F., 2008, *MNRAS*, 386, 1131
- Boyarsky A., Lesgourgues J., Ruchayskiy O., Viel M., 2009, *J. Cosmology Astropart. Phys.*, 5, 12
- Bryan G. L., Machacek M. E., 2000, *ApJ*, 534, 57
- Calura F., Tescari E., D’Odorico V., Viel M., Cristiani S., Kim T.-S., Bolton J. S., 2012, *MNRAS*, 422, 3019
- Cen R., Miralda-Escudé J., Ostriker J. P., Rauch M., 1994, *ApJ*, 437, L9
- Compostella M., Cantalupo S., Porciani C., 2013, *MNRAS*
- Cowie L. L., Songaila A., Kim T.-S., Hu E. M., 1995, *AJ*, 109, 1522
- Croft R. A. C., Weinberg D. H., Katz N., Hernquist L., 1998, *ApJ*, 495, 44
- Dalla Vecchia C., Schaye J., 2008, *MNRAS*, 387, 1431
- Davidsen A. F., Kriss G. A., Zheng W., 1996, *Nature*, 380, 47
- Efstathiou G., 1992, *MNRAS*, 256, 43P
- Fukugita M., Hogan C. J., Peebles P. J. E., 1998, *ApJ*, 503, 518
- Garzilli A., Bolton J. S., Kim T.-S., Leach S., Viel M., 2012, *MNRAS*, 424, 1723
- Gnedin N. Y., Hui L., 1998, *MNRAS*, 296, 44
- Gunn J. E., Peterson B. A., 1965, *ApJ*, 142, 1633
- Haardt F., Madau P., 1996, *ApJ*, 461, 20
- 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 Modelling the UV/X-ray cosmic background with CUBA*
- Hui L., Gnedin N. Y., 1997, *MNRAS*, 292, 27
- Hui L., Gnedin N. Y., Zhang Y., 1997, *ApJ*, 486, 599
- Hui L., Rutledge R. E., 1999, *ApJ*, 517, 541
- Jakobsen P., Boksenberg A., Deharveng J. M., Greenfield P., Jedrzejewski R., Paresce F., 1994, *Nature*, 370, 35
- Lidz A., Faucher-Giguère C.-A., Dall’Aglío A., McQuinn M., Fechner C., Zaldarriaga M., Hernquist L., Dutta S., 2010, *ApJ*, 718, 199
- Lynds R., 1971, *ApJ*, 164, L73
- McDonald P., Miralda-Escudé J., Rauch M., Sargent W. L. W., Barlow T. A., Cen R., 2001, *ApJ*, 562, 52
- McDonald P., Miralda-Escudé J., Rauch M., Sargent W. L. W., Barlow T. A., Cen R., Ostriker J. P., 2000, *ApJ*, 543, 1
- McDonald P., Seljak U., Cen R., Bode P., Ostriker J. P., 2005, *MNRAS*, 360, 1471
- McQuinn M., Hernquist L., Lidz A., Zaldarriaga M., 2011, *MNRAS*, 415, 977
- McQuinn M., Lidz A., Zaldarriaga M., Hernquist L., Hopkins P. F., Dutta S., Faucher-Giguère C.-A., 2009, *ApJ*, 694, 842
- Meiksin A. A., 2009, *Reviews of Modern Physics*, 81, 1405
- Miralda-Escudé J., Rees M. J., 1993, *MNRAS*, 260, 617
- Mortlock D. J., Warren S. J., Venemans B. P., Patel M., Hewett P. C., McMahon R. G., Simpson C., Theuns T., González-Solares E. A., Adamson A., Dye S., Hambly N. C., Hirst P., Irwin M. J., Kuiper E., Lawrence A., Röttgering H. J. A., 2011, *Nature*, 474, 616
- Okamoto T., Gao L., Theuns T., 2008, *MNRAS*, 390, 920
- Peebles M. S., Weinberg D. H., Davé R., Fardal M. A., Katz N., 2010, *MNRAS*, 404, 1295
- Planck Collaboration Ade P. A. R., Aghanim N., Armitage-Caplan C., Arnaud M., Ashdown M., Atrio-Barandela F., Aumont J., Baccigalupi C., Banday A. J., et al. 2013, *ArXiv e-prints*
- Rahmati A., Pawlik A. H., Raičević M., Schaye J., 2013, *MNRAS*, 430, 2427
- Rauch M., 1998, *ARA&A*, 36, 267
- Ricotti M., Gnedin N. Y., Shull J. M., 2000, *ApJ*, 534, 41
- Rorai A., Hennawi J. F., White M., 2013, *ApJ*, 775, 81
- Rudie G. C., Steidel C. C., Pettini M., 2012, *ApJ*, 757, L30
- Sawala T., Frenk C. S., Fattahi A., Navarro J. F., Bower R. G., Crain R. A., Dalla Vecchia C., Furlong M., Jenkins A., McCarthy I. G., Qu Y., Schaller M., Schaye J., Theuns T., 2014, *ArXiv e-prints*
- Schaye J., 2001, *ApJ*, 559, 507
- Schaye J., Aguirre A., Kim T.-S., Theuns T., Rauch M., Sargent W. L. W., 2003, *ApJ*, 596, 768
- Schaye J., Dalla Vecchia C., 2008, *MNRAS*, 383, 1210
- Schaye J., Dalla Vecchia C., Booth C. M., Wiersma R. P. C., Theuns T., Haas M. R., Bertone S., Duffy A. R., McCarthy I. G., van de Voort F., 2010, *MNRAS*, 402, 1536
- Schaye J., Theuns T., Leonard A., Efstathiou G., 1999, *MNRAS*, 310, 57
- Schaye J., Theuns T., Rauch M., Efstathiou G., Sargent W. L. W., 2000, *MNRAS*, 318, 817
- Springel V., 2005, *MNRAS*, 364, 1105
- Tepper-García T., Richter P., Schaye J., Booth C. M., Dalla Vecchia C., Theuns T., 2012, *MNRAS*, 425, 1640
- Theuns T., Leonard A., Efstathiou G., Pearce F. R., Thomas P. A., 1998, *MNRAS*, 301, 478
- Theuns T., Mo H. J., Schaye J., 2001, *MNRAS*, 321, 450
- Theuns T., Schaye J., Haehnelt M. G., 2000, *MNRAS*, 315, 600
- Theuns T., Schaye J., Zaroubi S., Kim T.-S., Tzanavaris P., Carswell B., 2002, *ApJ*, 567, L103
- Theuns T., Viel M., Kay S., Schaye J., Carswell R. F., Tzanavaris P., 2002, *ApJ*, 578, L5
- Theuns T., Zaroubi S., 2000, *MNRAS*, 317, 989
- Theuns T., Zaroubi S., Kim T.-S., Tzanavaris P., Carswell R. F., 2002, *MNRAS*, 332, 367
- Viel M., Bolton J. S., Haehnelt M. G., 2009, *MNRAS*, 399, L39
- Viel M., Schaye J., Booth C. M., 2013, *MNRAS*, 429, 1734
- Weinberg D. H., Hernquist L., Katz N. S., Miralda-Escudé J., 1996, in Bremer M. N., Malcolm N., eds, *Cold Gas at High Redshift Vol. 206 of Astrophysics and Space Science Library, Small Scale Structure and High Redshift HI*. p. 93
- Weymann R. J., Carswell R. F., Smith M. G., 1981, *ARA&A*, 19, 41
- Wiersma R. P. C., Schaye J., Dalla Vecchia C., Booth C. M., Theuns T., Aguirre A., 2010, *MNRAS*, 409, 132
- Wiersma R. P. C., Schaye J., Smith B. D., 2009, *MNRAS*, 393, 99
- Wiersma R. P. C., Schaye J., Theuns T., 2011, *MNRAS*, 415, 353
- Wiersma R. P. C., Schaye J., Theuns T., Dalla Vecchia C., Tornatore L., 2009, *MNRAS*, 399, 574
- Zaldarriaga M., Hui L., Tegmark M., 2001, *ApJ*, 557, 519

## APPENDIX A: BIASES IN THE $N_{\text{HI}}-\Delta$ RELATION

In Fig. A1 we plot  $\Delta-N_{\text{HI}}$  for lines, with  $N_{\text{HI}}$  computed from Eq. (19), and individual lines coloured with the value of the broadening  $b$ , calculated from Eq. (18). Line width biases the value of column density, in the sense that the analytic model of Eq. (23) (black line in the figure) yields *lower* values of  $N_{\text{HI}}$  than Eq. (19) for broad lines. The effect of this bias on the  $N_{\text{HI}}-b$  relation is illustrated in Fig. A2. Our model for line broadening (black line)



**Figure A1.** Density contrast - column density  $\Delta$ - $N_{\text{HI}}$  relation for lines from the REFERENCE model, coloured according to their line width,  $b$ ; *black dots* denote the 10th percentile lowest value of column-density at given density contrast. Line width  $b$  and column density  $N_{\text{HI}}$  are calculated from Eqs. (18-19), respectively. The *solid black line* is the analytic relation from Eq. (23). At given  $\Delta$ , broader lines have *higher* column density, illustrating that line width biases the column density associated with a given density contrast.



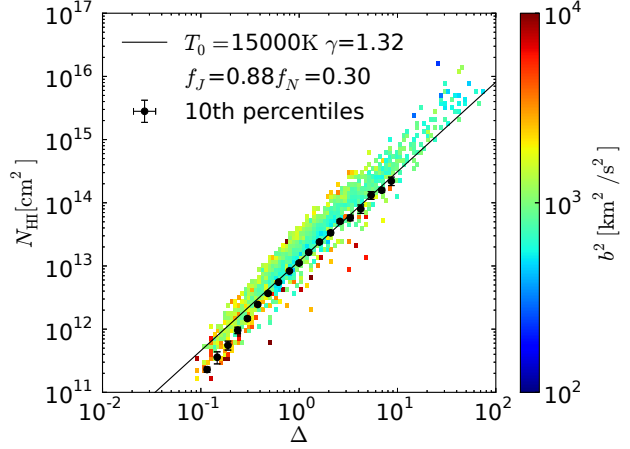
**Figure A2.** Column-density line width  $N_{\text{HI}}$ - $b$  relation for the REFERENCE model, colours are a measure of the density of points in the plot; the *black line* is the model illustrated in Fig. 8; the *dotted red line* is the thermal width. *Black dots* denote the 10th percentile lowest values of  $b$  for bins in column density, *magenta triangles* denote the 10th percentile lowest values of  $b$  for bins in density contrast  $\Delta$ , using Eq. (23) to convert  $\Delta$  to  $N_{\text{HI}}$ . The bias that line width introduces in this relation causes the black dots and red dots to diverge at low  $N_{\text{HI}}$ .

works better when lines are characterised by their density contrast (red dots) than by column density (black dots), especially for weak lines. This is because relation Eq. (23) is fitted to narrower lines.

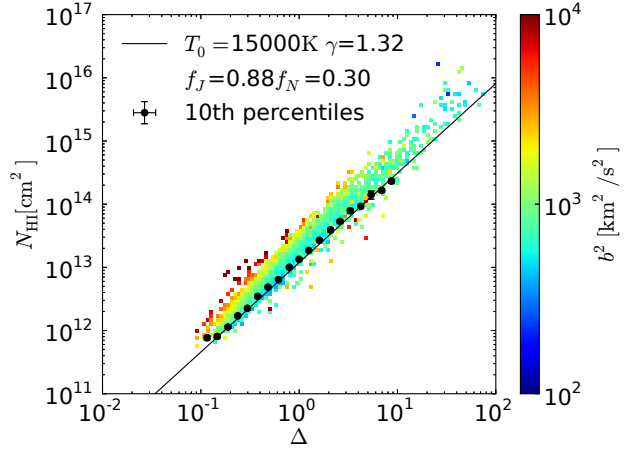
Here we also show that Eq. (19) is equivalent to computing  $N_{\text{HI}}$  with the integral of the optical depth, as it is often done in the literature. In Fig. A3 we show  $\Delta$ - $N_{\text{HI}}$ , with  $N_{\text{HI}}$  computed from

$$N_{\text{HI}} = \frac{1}{\sigma_0 c} \int_{v_1}^{v_2} \tau dv, \quad (\text{A1})$$

where  $v_1$  and  $v_2$  are the extremes of the line. If we compare Fig. A3



**Figure A3.** Same as Fig. A1, but with  $N_{\text{HI}}$  computed from Eq. (A1). For the low-density lines the neutral hydrogen column density deviates from Eq. (23) due to the blending of the faint lines.



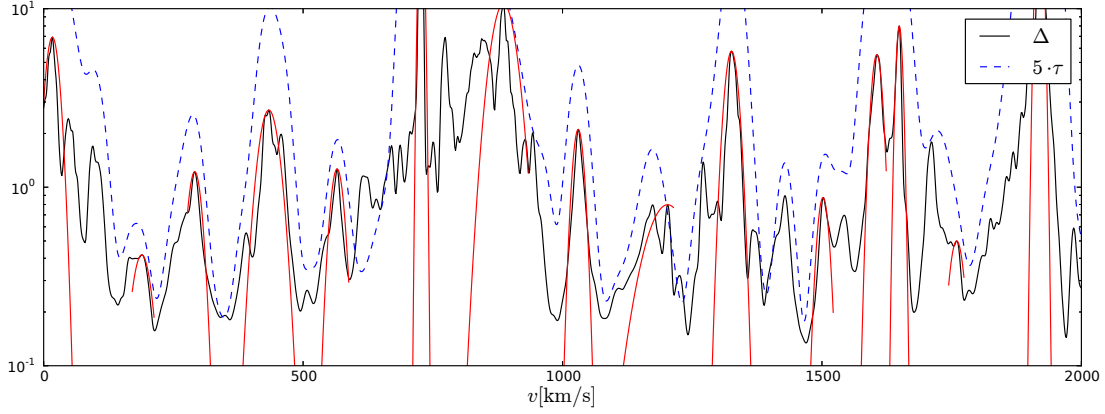
**Figure A4.** Same as Fig. A1 and A3, but with  $N_{\text{HI}}$  computed from Eq. (A2). The distribution  $\Delta$ - $N_{\text{HI}}$  looks similar to the one shown in Fig. A1.

with Fig. A1, we can see that Eq. (A1) underestimates the neutral hydrogen column density for the low-density lines.

When lines are blended, the column density computed from Eq. (A1) may become smaller and smaller as a shorter velocity interval is associated with the line (i.e. when  $v_1 \approx v_2$ ). We can estimate the contribution of these wings, and compute a corrected column-density

$$N_{\text{HI}} = \frac{1}{\sigma_0 c} \int_{-\infty}^{\infty} \tau dv, \quad (\text{A2})$$

by extrapolating a Gaussian fit to the line. Such an extrapolated density is closer to what is done in Voigt-profile fitting. Using Eq. (A2) results in a  $b - N_{\text{HI}}$  distribution which is very similar from that obtained in Fig. A1, as shown in Fig. A4.



**Figure B1.** Density profile of absorbers in the absence of peculiar velocities (*black line*) for a mock spectrum calculated from the REFERENCE model; *dashed blue line* is the scaled optical depth for comparison. *Red lines* are Gaussian profiles from Eq. (10), where the width is such that the integral of the density profile of the Gaussian is the same as the one of the multi-peaked absorber. The widths of the red Gaussians are often a good approximation to the extent of an absorber.

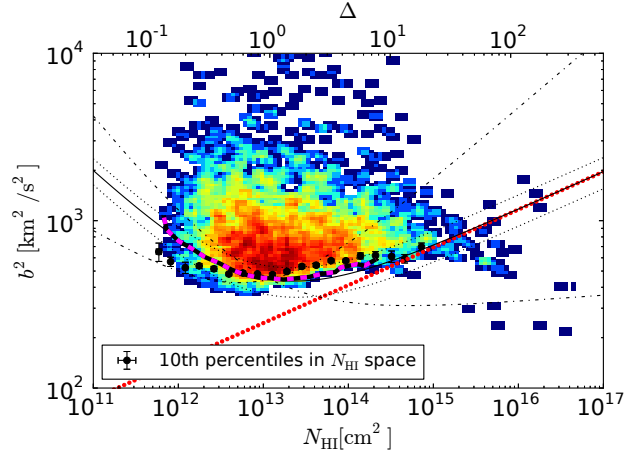
## APPENDIX B: DETERMINATION OF $f_J$

In this Appendix we illustrate that our estimate for the extent of absorbers in real space indeed is a good approximation for their actual size. We begin by neglecting peculiar velocities (i.e. generate mock spectra *after zeroing* all peculiar velocities), so that we can unambiguously identify the real-space density structure that gives rise to a given line; we demonstrated in the main text that peculiar velocities have little impact on line-widths. Often individual absorption lines correspond to more than one density peak. Such clustering of peaks sometimes gives rise to very wide lines, as we argued in the text.

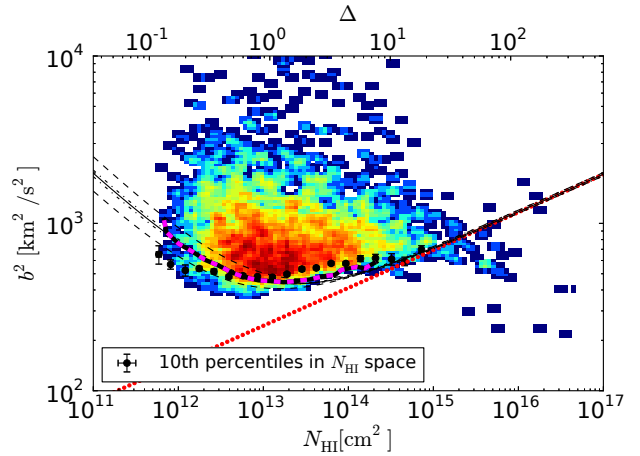
In Fig. B1 we show how we have computed the broadening  $b_\Delta$  in Fig. 3 from the density profile of the multi-peaked absorber. The red lines are Gaussian profiles as in Eq. (10), with  $v_0$  and  $\rho_0$  respectively the velocity and the density of the local maximum in optical depth. The width  $b_\Delta$  is such that the integral of the Gaussian density profile is the same as the integral of the density profile of the absorbers. Even though there is often no unique way of associating a ‘size’ with such complex absorbers, we believe  $b_\Delta$  is often a good approximation of the extent over which the density contrast of an absorber is significant.

## APPENDIX C: DEGENERACY OF PARAMETERS IN THE LINE BROADENING RELATION

For completeness we illustrate the sensitivity of Eq. (17) to changes in the  $T - \Delta$  relation (varying  $T_0$  and  $\gamma$ ; Fig.C1), and varying the fitting parameters  $f_J$  and  $f_N$  (Fig. C2)



**Figure C1.** Same as figure 8. The dashed (dashed-dot) lines are the model from Eq. 25 with  $T_0$  ( $\gamma$ ) changed by  $\pm 20\%$ .



**Figure C2.** Same as figure 8. The dashed (dashed-dot) lines are the model from Eq. 25 with  $f_J$  ( $f_N$ ) changed by  $\pm 10\%$ .