The thermal history of the intergalactic medium

Joop Schaye, Tom Theuns, Michael Rauch, George Efstathiou and Wallace L.W. Sargent

Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK
Max-Planck-Institut für Astrophysik, Postfach 1523, 85740 Garching, Germany
European Southern Observatory, Karl-Schwarzschild-Str. 2, 85748 Garching, Germany
Astronomy Department, California Institute of Technology, Pasadena, CA 91125, USA

21 December 1999

ABSTRACT
At redshifts $z \approx 2$, most of the baryons reside in the smooth intergalactic medium which is responsible for the low column density Ly$\alpha$ forest. This photoheated gas follows a tight temperature-density relation which introduces a cut-off in the distribution of widths of the Ly$\alpha$ absorption lines ($b$-parameters) as a function of column density. We have measured this cut-off in a sample of nine high resolution, high signal-to-noise quasar spectra, and determined the thermal evolution of the intergalactic medium in the redshift range $2.0-4.5$. At redshift $z \sim 3$, the temperature at the mean density shows a prominent peak and there is evidence that the gas becomes nearly isothermal. We interpret this as evidence for the reionization of He$\text{II}$. We also find that current models of the ionizing background from quasars predict too much absorption at $z \approx 4$, suggesting that galaxies might contribute significantly to the UV-background at these redshifts.

Key words: cosmology: miscellaneous – intergalactic medium – quasars: absorption lines

1 INTRODUCTION

According to the standard big bang model, the primordial hydrogen and helium comprising the intergalactic medium (IGM) was hot and highly ionized at early times. As the universe expanded, the hot plasma cooled adiabatically, becoming almost completely neutral at a redshift of $z \sim 10^3$. The IGM remained neutral until the first stars and quasars began to produce ionizing photons. Eventually, the ionizing radiation became intense enough to reionize hydrogen and later, because of its higher ionization potential, to fully reionize helium. Since the thermal evolution of the IGM depends strongly on its reionization history, it can be used as a probe of the end of the ‘dark ages’ of cosmic history, when the first stars and quasars were formed (Miralda-Escudé & Rees 1994; Hui & Gnedin 1997; Haehnelt & Steinmetz 1998).

The absence of Gunn-Peterson absorption (Gunn & Peterson 1965) in quasar spectra, i.e. the complete absorption of quasar light blueward of the H I and He$\text{II}$ Ly$\alpha$ wavelengths requires that hydrogen must have been highly ionized by $z \sim 5$ (Schneider, Schmidt & Gunn 1991; Songaila et al. 1999) and helium by $z \sim 2.5$ (Davidsen, Kriss & Zheng 1996). Measurements of the He$\text{II}$ Ly$\alpha$ opacity suggest that helium may have reionized around $z \sim 3$ (Heap et al. 2000; Reimers et al. 1997; Jakobsen et al. 1994; Davidsen et al. 1996; Anderson et al. 1999). This would fit in with evidence for a hardening of the UV background around this time, as derived from the ratio of Si$\text{IV}$/C$\text{IV}$ in high redshift quasar absorption lines (Songaila & Cowie 1996; Songaila 1998), although both the observational result and its interpretation are still controversial (Boksenberg, Sargent & Rauch 1998; Giroux & Shull 1997).

The resonant Ly$\alpha$ absorption by residual low levels of neutral hydrogen along the line of sight to a quasar produces a forest of absorption lines. Although many of the basic observational facts about the Ly$\alpha$ forest at high redshift ($z \sim 2-5$) had been established before the 10 m telescope era, the advent of the Keck telescope has lead to much larger data samples at much higher signal-to-noise ratio than hitherto available (e.g. Hu et al. 1995; Lu et al. 1996; Kirkman & Tytler 1997). The observational progress has been matched on the theoretical side by cosmological hydro-simulations (e.g. Cen et al. 1994; Zhang, Anninos & Norman 1995; Petitjean, Mücket & Kates 1995; Hernquist et al. 1996; Miralda-Escudé et al. 1996) which together with the new data are now beginning to yield significant quantitative cosmological constraints (see Rauch 1998 for a review).

These simulations show that the low column density
(N10^{4.5} \text{ cm}^{-2}) absorption lines arise in a smoothly varying IGM of low density contrast (\delta), which contains most of the baryons in the universe. Since the overdensity is only mildly non-linear, the physical processes governing this medium are well understood and relatively easy to model. On large scales the dynamics are determined by gravity, while on small scales gas pressure is important. Since shock heating is unimportant for the low-density gas, the interplay between photoionization heating and adiabatic cooling due to the expansion of the universe results in a tight temperature-density relation, which is well described by a power-law for densities around the cosmic mean, T = T_0(\rho/\bar{\rho})^{\gamma - 1} (Hui & Gnedin 1997). This relation is generally referred to as the ‘equation of state’ (even though the true equation of state is that of an ideal gas).

For models with abrupt reionization, the IGM becomes nearly isothermal (\gamma \approx 1) at the redshift of reionization. After reionization, the temperature at the mean density (T_0) decreases while the slope (\gamma - 1) increases because higher density regions undergo increased photoheating and expand less rapidly. Eventually, the imprints of the reionization history are washed out and the equation of state approaches an asymptotic state, \gamma = 1.62, T_0 \propto \left[\frac{\Omega_b h^2}{\sqrt{\Omega_m h^2}}\right]^{1/1.7} (Miralda-Escudé & Rees 1994; Hui & Gnedin 1997; Theuns et al. 1998). However, the timescale for recombination cooling in the low density IGM is never small compared to the age of the universe for z \leq 20 and inverse Compton cooling of free electrons off the cosmic microwave background is only efficient for z < 5. Consequently, unless both hydrogen and helium were fully reionized at redshifts considerably higher than this, the gas will have retained some memory of when and how it was reionized.

A standard way of analyzing Ly\alpha forest spectra is to decompose them into a set of distinct absorption lines, assumed to have Voigt profiles (e.g. Carswell et al. 1991). Various broadening mechanisms, like Hubble broadening (i.e. the differential Hubble flow across the absorber), peculiar and thermal velocities contribute to the line widths (Meiksin 1994; Hui & Rutledge 1999; Theuns, Schaye & Haehnelt 2000). Recently it was shown (Theuns et al. 1998; Bryan et al. 1999) that high resolution simulations of the standard cold dark model, using the ionizing background computed by Haardt & Madau (1996, hereafter HM) produce a larger fraction of narrow lines than observed. Because of the difficulty in disentangling the various broadening mechanisms, different authors have proposed different solutions to this problem. Theuns et al. (1998) suggested that the gas temperature in the simulations was too low, while Bryan et al. (1999) argued that the amplitude of primordial fluctuations was too high.

For a given reionization history, the temperature in the simulations could be increased by increasing the baryon density and the age of the universe (Theuns et al. 1999), including Compton heating by the hard X-ray background (Madau & Efstathiou 1999) and possibly photo-electric heating by dust grains (Nath, Sethi & Shchekinov 1999). Alternatively, Abel & Haehnelt (1999) pointed out that the photoheating rates around the time of reionization are significantly underestimated when the gas is assumed to be optically thin, as is generally done in cosmological simulations. Clearly, knowledge of the evolution of the equation of state would enable us to discriminate between the various possibilities.

Even though various mechanisms contribute to the broadening of the absorption lines, there exists a lower limit to the line-width, set by the temperature of the gas. Because the physical density of the IGM correlates strongly with the column density of the absorption lines, this results in a cut-off in the distribution of line widths (b-parameters) as a function of column density, which traces the equation of state of the gas (Schaye et al. 1999, hereafter STLE; Ricotti, Gnedin & Shull 2000; Bryan & Machacek 2000). Thus by measuring the Ly\alpha line width as a function of column density, b(N), we can infer the equation of state of the IGM.

Here, we measure the mean absorption and the b(N) cut-off in nine high resolution, high S/N quasar spectra, spanning the redshift range 2.0–4.5. We use hydrodynamic simulations to calibrate the relations between the parameters of the b(N) cut-off and the equation of state. By applying these relations to the observations, we are able to measure the evolution of the equation of state over the observed redshift range. We find that the thermal evolution of the IGM is drastically different from that predicted by current models. The temperature peaks at z \sim 3, which, together with supporting evidence from measurements of the He\alpha opacity and the Si\alpha/\alpha ratios, we interpret as evidence for the second reionization of helium (He\alpha \rightarrow He\beta).

This paper is organized as follows. In sections 2 and 3 we describe the observations and the simulations respectively. We discuss the difference between evolution of the b-distribution and evolution of the temperature in section 4. In section 5 we briefly describe our method for measuring the equation of state, before we present our results in section 6. Systematic errors are discussed in section 7. Finally, we discuss and summarize the main results in sections 8 and 9. Appendix A1 contains a detailed description of our methods.

2 OBSERVATIONS

We analyzed a sample of nine quasar spectra, spanning the redshift range z_{\text{obs}} = 2.14–4.55 (Table 1). The spectra of Q1100–264 and APM 08279+5255 were kindly provided by R. Carswell and S. Ellison respectively. All spectra were taken with the high-resolution spectrograph (HIRES) on the Keck telescope, except the spectrum of Q1100–264, which was taken with the UCL echelle spectrograph of the Anglo Australian Telescope. Details on the data and reduction procedures, as well as the continuum fitting, can be found in Carswell et al. (1991) for Q1100–264, Ellison et al. (1999) for APM 08279+5255 and Barlow & Sargent (1997) and Rauch et al. (1997) for the others. The nominal velocity resolution (FWHM) was 8 km s\(^{-1}\) for Q1100–264 and 6.6 km s\(^{-1}\) for the others and the data were rebinned onto 0.04 A pixels on a linear wavelength scale. The signal to noise ratio per pixel is typically about 50, except for Q1100–264 for which it is about 20.

In order to avoid confusion with the Ly\beta forest, only the regions of a spectrum between the quasars Ly\beta and Ly\alpha emission lines were considered. In addition, spectral regions...
close to the quasar (typically 5–6 $h^{-1}$ Mpc, but 13 $h^{-1}$ Mpc for APM 08279+5255 and 9 $h^{-1}$ Mpc for Q1100–264) were omitted to avoid proximity effects. Regions contaminated by metals and damped Lyα lines were removed. The absorption features in the remaining spectral regions were fitted with Voigt profiles using the same automated version of VPFIT (Carswell et al. 1987) as was used for the simulated spectra. Using a fully automatic fitting program invariably results in a few ‘bad fits’. However, given that there is no unique way of decomposing intrinsically non-Voigt absorption lines into a set of discrete Voigt profiles, it is essential to apply the same algorithm to simulated and observed spectra. Since ‘bad fits’ will also occur in the synthetic spectra, we have made no attempt to correct them.

The Lyα forest of a single quasar spans a considerable redshift range ($\Delta z \sim 0.5$). In order to minimize the effects of redshift evolution and S/N variation across a single spectrum, we divided each Lyα forest spectrum in two parts of equal length. STLE showed that their algorithm for measuring the cut-off of the $b(N)$ distribution is relatively insensitive to the number of absorption lines, the statistical variance is almost the same for e.g. 150 and 300 lines. Hence little information is lost in analyzing narrow redshift bins if the absorption line density is high. The two halves of the spectra were analyzed separately and each was compared with its own set of simulated spectra (see section 3). For the two lowest redshift quasars the number of absorption lines is too small to split the data in half. Hence each quasar, except for Q1100–264 and Q2343+123, provides two nearly independent data sets.

The complete set of absorption line samples is listed in Table 2. The median, minimum and maximum redshifts of the absorption lines used to determine the $b(N)$ cut-off are listed in columns 2–4. The minimum column density considered was $10^{12.5} \text{cm}^{-2}$ for all samples, the maximum column densities are listed in column 5 and the total number of absorption lines in this column density range in column 6 (only lines for which VPFIT gives relative errors in both the $b$-parameter and column density less than 0.25 are considered). See appendix A1.1 for a discussion of our choice of parameters.

<table>
<thead>
<tr>
<th>Sample</th>
<th>z</th>
<th>$z_{\text{min}}$</th>
<th>$z_{\text{max}}$</th>
<th>$\log N_{\text{max}}$</th>
<th># lines</th>
<th>$\log N_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1100</td>
<td>1.96</td>
<td>1.85</td>
<td>2.09</td>
<td>14.0</td>
<td>44</td>
<td>13.0</td>
</tr>
<tr>
<td>2343</td>
<td>2.29</td>
<td>2.23</td>
<td>2.38</td>
<td>14.2</td>
<td>48</td>
<td>13.0</td>
</tr>
<tr>
<td>1422a</td>
<td>2.21</td>
<td>2.10</td>
<td>2.33</td>
<td>14.5</td>
<td>87</td>
<td>13.0</td>
</tr>
<tr>
<td>1422b</td>
<td>2.50</td>
<td>2.33</td>
<td>2.63</td>
<td>14.5</td>
<td>90</td>
<td>13.0</td>
</tr>
<tr>
<td>1107a</td>
<td>2.59</td>
<td>2.48</td>
<td>2.71</td>
<td>14.5</td>
<td>76</td>
<td>13.0</td>
</tr>
<tr>
<td>1107b</td>
<td>2.84</td>
<td>2.71</td>
<td>2.95</td>
<td>14.5</td>
<td>91</td>
<td>13.0</td>
</tr>
<tr>
<td>1425a</td>
<td>2.66</td>
<td>2.55</td>
<td>2.88</td>
<td>14.5</td>
<td>94</td>
<td>13.0</td>
</tr>
<tr>
<td>1425b</td>
<td>3.00</td>
<td>2.89</td>
<td>3.14</td>
<td>14.5</td>
<td>118</td>
<td>13.0</td>
</tr>
<tr>
<td>1422a</td>
<td>3.08</td>
<td>2.91</td>
<td>3.22</td>
<td>14.5</td>
<td>145</td>
<td>13.0</td>
</tr>
<tr>
<td>1422b</td>
<td>3.37</td>
<td>3.22</td>
<td>3.53</td>
<td>14.5</td>
<td>152</td>
<td>13.0</td>
</tr>
<tr>
<td>0827a</td>
<td>3.23</td>
<td>3.15</td>
<td>3.42</td>
<td>14.5</td>
<td>105</td>
<td>13.4</td>
</tr>
<tr>
<td>0827b</td>
<td>3.55</td>
<td>3.43</td>
<td>3.70</td>
<td>14.5</td>
<td>106</td>
<td>13.4</td>
</tr>
<tr>
<td>0000a</td>
<td>3.72</td>
<td>3.61</td>
<td>3.81</td>
<td>14.8</td>
<td>77</td>
<td>13.5</td>
</tr>
<tr>
<td>0000b</td>
<td>3.91</td>
<td>3.81</td>
<td>4.01</td>
<td>14.8</td>
<td>97</td>
<td>13.5</td>
</tr>
<tr>
<td>2237a</td>
<td>3.84</td>
<td>3.69</td>
<td>4.02</td>
<td>14.8</td>
<td>140</td>
<td>13.5</td>
</tr>
<tr>
<td>2237b</td>
<td>4.31</td>
<td>4.15</td>
<td>4.43</td>
<td>14.8</td>
<td>127</td>
<td>13.5</td>
</tr>
</tbody>
</table>

are clear differences between the samples. We will show in section 4 that even a non-evolving $b$-distribution would imply a strong thermal evolution. Several samples contain a few lines that fall far below the cut-off. These lines, which have no significant effect on the measured cut-off, are most likely blends or unidentified metal lines.

### Table 1. Quasar spectra used

<table>
<thead>
<tr>
<th>QSO</th>
<th>$z_{\text{em}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1100–264</td>
<td>2.14</td>
</tr>
<tr>
<td>Q2343+123</td>
<td>2.52</td>
</tr>
<tr>
<td>Q1424+293</td>
<td>2.67</td>
</tr>
<tr>
<td>Q1107+485</td>
<td>3.00</td>
</tr>
<tr>
<td>Q1425+604</td>
<td>3.20</td>
</tr>
<tr>
<td>Q1422+230</td>
<td>3.62</td>
</tr>
<tr>
<td>APM 08279+5255</td>
<td>3.91</td>
</tr>
<tr>
<td>Q0000–262</td>
<td>4.11</td>
</tr>
<tr>
<td>Q2237–061</td>
<td>4.55</td>
</tr>
</tbody>
</table>

### Table 2. Observed absorption line samples.

<table>
<thead>
<tr>
<th>Sample</th>
<th>$z$</th>
<th>$z_{\text{min}}$</th>
<th>$z_{\text{max}}$</th>
<th>$\log N_{\text{max}}$</th>
<th># lines</th>
<th>$\log N_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1100</td>
<td>1.96</td>
<td>1.85</td>
<td>2.09</td>
<td>14.0</td>
<td>44</td>
<td>13.0</td>
</tr>
<tr>
<td>Q2343</td>
<td>2.29</td>
<td>2.23</td>
<td>2.38</td>
<td>14.2</td>
<td>48</td>
<td>13.0</td>
</tr>
<tr>
<td>Q1422a</td>
<td>2.21</td>
<td>2.10</td>
<td>2.33</td>
<td>14.5</td>
<td>87</td>
<td>13.0</td>
</tr>
<tr>
<td>Q1422b</td>
<td>2.50</td>
<td>2.33</td>
<td>2.63</td>
<td>14.5</td>
<td>90</td>
<td>13.0</td>
</tr>
<tr>
<td>Q1107a</td>
<td>2.59</td>
<td>2.48</td>
<td>2.71</td>
<td>14.5</td>
<td>76</td>
<td>13.0</td>
</tr>
<tr>
<td>Q1107b</td>
<td>2.84</td>
<td>2.71</td>
<td>2.95</td>
<td>14.5</td>
<td>91</td>
<td>13.0</td>
</tr>
<tr>
<td>Q1425a</td>
<td>2.66</td>
<td>2.55</td>
<td>2.88</td>
<td>14.5</td>
<td>94</td>
<td>13.0</td>
</tr>
<tr>
<td>Q1425b</td>
<td>3.00</td>
<td>2.89</td>
<td>3.14</td>
<td>14.5</td>
<td>118</td>
<td>13.0</td>
</tr>
<tr>
<td>Q1422a</td>
<td>3.08</td>
<td>2.91</td>
<td>3.22</td>
<td>14.5</td>
<td>145</td>
<td>13.0</td>
</tr>
<tr>
<td>Q1422b</td>
<td>3.37</td>
<td>3.22</td>
<td>3.53</td>
<td>14.5</td>
<td>152</td>
<td>13.0</td>
</tr>
<tr>
<td>Q0827a</td>
<td>3.23</td>
<td>3.15</td>
<td>3.42</td>
<td>14.5</td>
<td>105</td>
<td>13.4</td>
</tr>
<tr>
<td>Q0827b</td>
<td>3.55</td>
<td>3.43</td>
<td>3.70</td>
<td>14.5</td>
<td>106</td>
<td>13.4</td>
</tr>
<tr>
<td>Q0000a</td>
<td>3.72</td>
<td>3.61</td>
<td>3.81</td>
<td>14.8</td>
<td>77</td>
<td>13.5</td>
</tr>
<tr>
<td>Q0000b</td>
<td>3.91</td>
<td>3.81</td>
<td>4.01</td>
<td>14.8</td>
<td>97</td>
<td>13.5</td>
</tr>
<tr>
<td>Q2237a</td>
<td>3.84</td>
<td>3.69</td>
<td>4.02</td>
<td>14.8</td>
<td>140</td>
<td>13.5</td>
</tr>
<tr>
<td>Q2237b</td>
<td>4.31</td>
<td>4.15</td>
<td>4.43</td>
<td>14.8</td>
<td>127</td>
<td>13.5</td>
</tr>
</tbody>
</table>

### 3 SIMULATIONS

In order to calibrate the relation between the parameters of the $b(N)$ cut-off and the effective equation of state, we have simulated eight variants of the currently favoured flat, scale invariant, cosmological constant dominated cold dark matter model, which vary only in their heating rates (Table 3).

The calibration was repeated for each of the observed samples of absorption lines listed in Table 2. Synthetic spectra were computed along 1200 random lines of sight through the simulation box at the nearest redshift output ($\Delta z = 0.25$). The background flux was rescaled such that the mean effective optical depth in the simulated spectra matches that of the observed sample. Each spectrum was convolved with a Gaussian with full width at half maximum (FWHM) identical to that of the observations and resampled onto pixels of the same size. The noise properties of the observed spectrum were computed as a function of flux and imposed on the simulated spectra. The resulting spectra were continuum fitted as described in Theuns et al. (1998). Finally, Voigt profiles were fitted using the same automated version of VPFIT as we used for the observed spectra. We will refer to the sample of lines drawn from the synthetic spectra of simulation X, designed to mimic the observed spectrum Y as as model X-y, e.g. model L1-1442a.

All models have a total matter density $\Omega_m = 0.3$, vac-
Figure 1. $b(N)$-Distributions for the observed samples listed in Table 2. Crosses indicate positions of absorption lines, errors are not displayed. Only lines that are used for the determination of the cut-off are shown, i.e. lines for which VPFIT gives relative errors in both $b$ and $N$ smaller than 25 per cent. Solid lines are the measured cut-offs. Vertical dashed lines indicate the maximum column density used when fitting the cut-off. Horizontal dashed lines are identical and correspond to the cut-off of sample 1425b. Note that evolution of the $b(N)$ cut-off cannot be interpreted as evolution of the equation of state in any straightforward way, because the density-column density relation changes with redshift (see section 4). Furthermore, the cut-offs in different panels can only be compared directly if they contain a similar number of absorption lines.

Table 3. Simulations used for calibrating the relation between the $b(N)$ cut-off and the effective equation of state.

<table>
<thead>
<tr>
<th>Model</th>
<th>$\epsilon_{He}$</th>
<th>$H_X$</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>L0.3</td>
<td>1/3</td>
<td>0</td>
<td>low $T_0$</td>
</tr>
<tr>
<td>L1</td>
<td>1</td>
<td>0</td>
<td>reference model</td>
</tr>
<tr>
<td>L2</td>
<td>2</td>
<td>0</td>
<td>high $T_0$</td>
</tr>
<tr>
<td>L3</td>
<td>3</td>
<td>0</td>
<td>very high $T_0$</td>
</tr>
<tr>
<td>Lx</td>
<td>1</td>
<td>1</td>
<td>low $\gamma$, high $T_0$</td>
</tr>
<tr>
<td>Lx2.5</td>
<td>1</td>
<td>5/2</td>
<td>very low $\gamma$, high $T_0$</td>
</tr>
<tr>
<td>Lx5</td>
<td>1</td>
<td>5</td>
<td>very low $\gamma$, very high $T_0$</td>
</tr>
<tr>
<td>L1e</td>
<td>1</td>
<td>0</td>
<td>high $\gamma$</td>
</tr>
</tbody>
</table>

These simulations follow the evolution of a periodic, cubic region of the universe and are performed with a modified version of HYDRA (Couchman, Thomas & Pearce 1995), which uses smooth particle hydrodynamics (Lucy 1970; Gingold & Monaghan 1977). The simulations use $64^3$ gas particles and $64^3$ cold dark matter particles in a box of of comoving size 3.85 Mpc, so the particle mass is $1.14 \times 10^6 M_\odot$ for the gas and $6.51 \times 10^6 M_\odot$ for the dark matter.

Our reference model, L1, is photoionized and photoheated by the UV-background from quasars as computed...
by HM, using the optically thin limit. Models L0.3, L2 and L3 are identical, except that we have multiplied the helium photoheating rates (column \( \epsilon_{\text{He}} \) in Table 3) by factors of 1/3, 2 and 3 respectively (keeping the ionization rates constant). The effective helium photoheating rate may be higher than computed in the optically thin limit because of radiative transfer effects (Abel & Haehnelt 1999). Model Lx is identical to model L1, except that we have included Compton heating by the hard X-ray background as computed by Madau & Efstathiou (1999) (column \( \rho_{\text{H}} \) in Table 3). For a highly ionized plasma, the energy input per particle from Compton scattering of free electrons is independent of the density. Hence, Compton scattering tends to flatten the effective equation of state. We have used this fact to artificially construct models with low values of \( \gamma \), by multiplying the X-ray heating rates by (unrealistic) factors of 2.5 and 5 for models Lx2.5 and Lx5 respectively. Finally, model L1e is identical to model L1, except that we have set the ionization and heating rates for H I and He I for redshifts between 6 and 10 equal to those at \( z = 6 \). In this model H I and He I ionize early (at \( z = 10 \)), which drives \( \gamma \) to larger values.

In addition to the models listed in Table 3, we have performed some simulations to investigate possible systematic effects. We simulated model L1 twice with a lower resolution (54\(^3\) and 44\(^3\) particles) and model L3 in a larger box, but with the same resolution (5\(h^{-1}\) Mpc and 128\(^3\) instead of 2.5\(h^{-1}\) Mpc and 64\(^3\)). Model L1 was simulated twice more with a lower normalization of the initial power spectrum (\(\sigma_8 = 0.65\) and 0.4 instead of 0.9).

\section{Evolution of the b-Distribution vs Thermal Evolution}

Williger et al. (1994) and Lu et al. (1996) found that the \(b\)-parameters at \( z \sim 4 \) are smaller than at \( z = 2-3 \). Kim et al. (1997) showed that the increase in the line widths with decreasing redshift continues over the range \( z = 3.5 \) to 2.1. It is tempting to interpret these results as evidence for an increase in the temperature \( T_0 \) with decreasing redshift. However, we will show in this section that the \( b\)-values are smaller at higher redshift even for models in which \( T_0 \) is higher, as is the case for models in which the universe is fully reionized by \( z = 4 \).

As pointed out by STLE, any statistic that is sensitive to the temperature of the absorbing gas, will in general depend on both the normalization, \( T_0 \), and the slope, \( \gamma \), of the equation of state. This is because temperature is a function of density and the absorbing gas is in general not all at the mean density of the universe. After reionization, \( T_0 \) will decrease and \( \gamma \) will increase with time. Consequently, the evolution of the temperature at a given overdensity can be very different from the evolution of \( T_0 \). This is illustrated in Fig. 2, where the temperature \( T_b \) at a density contrast \( \delta \equiv \rho/\bar{\rho} - 1 \) is plotted as a function of redshift for model L1. Even though the temperature at the mean density decreases with time (solid line), the temperature at a density contrast as little as 2 remains almost constant.

The general expansion of the universe ensures that the column density corresponding to a fixed overdensity is a strongly increasing function of redshift. In fact, most of the evolution of the Ly\(\alpha \) forest can be understood in terms of the resulting scaling of the optical depth (e.g. Miralda-Escudé et al. 1996; Davé et al. 1999; Machacek et al. 1999). When interpreting the evolution of the \( b\)-distribution, one therefore has to keep in mind that: (a) at fixed column density, absorption lines at higher redshift will correspond to absorbers of smaller overdensities; (b) the evolution of the temperature at a fixed overdensity depends on the evolution of both \( T_0 \) and \( \gamma \). Together these effects can conspire to make the \( b\)-parameters smaller at higher redshift, even when the temperature \( T_0 \) is higher. Fig. 3 shows that this will happen for models in which the IGM is fully reionized at the observed redshifts (\( z \geq 4 \)).

The discussion in this section shows that in order to derive the evolution of \( T_0 \) using a statistic that is sensitive
to the temperature of the absorbing gas, one needs to determine the evolution of: (1) the temperature of the gas; (2) the overdensity of the gas and (3) the slope of the equation of state. We will see that the uncertainty in γ is the limiting factor.

5 MEASURING THE EQUATION OF STATE

STLE demonstrated that the observed cut-off in the distribution of b-parameters as a function of column density can be used to measure the equation of state of the IGM. In particular, they showed that the b(N) cut-off can be fitted by a power-law, \( b = a N_0 (N/N_0)^{\epsilon - 1} \), whose parameters \( \log b N_0 \) and \( \Gamma - 1 \) are proportional (for a suitable choice of \( N_0 \)) to the parameters of the underlying equation of state, \( \log T_0 \) and \( \gamma - 1 \), respectively. We will first calibrate these relations using the simulations and then use them to convert the observed cut-offs into measurements of the equation of state.

For each observed sample of absorption lines (Table 2) we go through the following procedure. First mock spectra are generated from the 8 simulations listed in Table 3. The synthetic spectra are processed to give them the same characteristics (mean absorption, resolution, pixel size and noise properties) as the corresponding observed spectra. These are then fitted with Voigt profiles using the same automated fitting package that was used for the observations. For each of the eight simulated sets of absorption lines, the b(N) cut-off is fitted\(^\dagger\) for a subset of n lines taken at random from the complete set, where n is the number of lines in the observed sample. Probability distributions for the parameters of the cut-off are generated by repeating this last step 1000 times. The medians of the resulting distributions are then used as our best estimates of the cut-offs in the synthetic spectra.

We then use these 8 simulations to calibrate the relations between the parameters of the b(N) cut-off, \( b b_0 \) and \( \Gamma \) and the parameters of the equation of state, \( (T_{b(N)}, \gamma) \). We measure the density contrast corresponding to the pivot column density, \( N_0 \), by using the fact \( \log T_{b(N_0)} \propto \log b N_0 \) (STLE), essentially because both thermal broadening and Jeans smoothing scale as the square root of the temperature. Finally, we use these relations to convert the observed b(N) cut-off into a measurement of the equation of state.

The various steps in this process, as well as the error analysis, are described in detail in appendix A1.

6 RESULTS

6.1 Mean absorption

Fig. 4 shows the effective optical depth, \( \tau_{\text{eff}} \equiv - \ln\langle F \rangle \), of the observed samples as a function of (decreasing) redshift. The scatter is surprisingly small, considering that most data points represent just half of a Lyα forest spectrum. Note that Rauch et al. (1997) studied the opacity of the forest using seven out of nine of the quasars from this sample. They found a slightly less rapid increase with redshift, because they rebinned the data into three redshift bins, centered on \( z = 2, 3 \) and 4.

In an Einstein-de Sitter universe, with a constant photoionizing background and no growth of structure, the mean optical depth varies as \( \tau \propto (1 + z)^{\alpha} \), which corresponds to the slope of the dashed line. The good agreement with the data is somewhat artificial, since we have plotted the effective optical depth whereas the theoretical relation holds for the mean optical depth, which is not observable. Although the two should be similar at low redshift, the mean optical depth will be much larger than the effective optical depth at high redshift, where there are many saturated pixels. From the simulated spectra we can estimate the necessary correction; it increases from about a factor 3 for sample 1100 to more than an order of magnitude for 2237. Hence this effect alone can account for the flattening of \( \tau_{\text{eff}} \) at high redshift. Continuum fitting, which becomes progressively more difficult at higher redshifts, will also tend to decrease the apparent absorption at high redshift.

The dot-dashed line indicates the evolution of the effective optical depth for model L1, which uses the HM ionizing background. The model seems to slightly underpredict the absorption, except for redshifts \( z \leq 4 \), where the optical depth is significantly higher than observed. In ionization equilibrium, which holds after reionization, the neutral density scales as \( n_{\text{HI}} \propto T^{-0.7} \). Jeans smoothing causes the dependence of the optical depth on the temperature to be somewhat weaker than this: when the temperature is higher, less gas is locked up in the high density regions giving rise to saturated absorption lines. Hence the discrepancy at \( z \sim 2.5–3.5 \) would be considerably more pronounced for a model with a higher temperature than L1, which would be in better agreement with the observations as we shall see shortly. From the simulations, we estimate that \( \log \tau_{\text{eff}} \) could be lower by as much as 0.1 dex for a model with the correct temperature.

\( \dagger \) The cut-off is fitted using bootstrap resampling, as described in appendix A1.2.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{image.png}
\caption{The observed effective optical depth as a function of redshift. The dashed line has a slope of 4.5, which corresponds to the evolution of the mean optical depth in an Einstein-de Sitter universe with no evolution apart from the cosmological expansion. The dot-dashed line is for our reference model, L1, which uses the HM ionizing background. Horizontal error bars indicate redshift intervals used, vertical error bars are 1σ errors as determined from a bootstrap analysis using chunks of 4 Å.}
\end{figure}
6.2 Thermal evolution

The evolution of the inferred temperature at the mean density of the IGM is plotted in Fig. 5. At redshift $z \sim 3.5$, the temperature increases sharply until it peaks around $z \sim 3$, after which it decreases again. This behavior is drastically different from that predicted by simple models of the photoionization of the IGM by the integrated UV flux from quasars. For comparison, the dashed curve shows our reference model, L1, which uses the HM ionizing background. In this model, the universe is fully ionized by $z \sim 4.5$ and the temperature of the IGM declines slowly as the Universe expands. Such a model cannot account for the abrupt increase in temperature of the IGM at $z \sim 3$ (the reduced $\chi^2$ for the solid curve is 6.9). The redshift of the temperature change in the IGM, and its abruptness, suggests that it may be associated with the photoionization of singly ionized helium ($\text{He}^\text{II}$) to doubly ionized helium ($\text{He}^\text{III}$).

Fig. 6 shows the measured evolution of $\gamma$, the slope of the equation of state. Although most of the data points have large errors, the results are again clearly inconsistent with model L1 (solid curve, reduced $\chi^2 = 3.6$). From $z \sim 3.5$ the observed slope decreases until it becomes close to isothermal at $z \sim 3$, after which it increases again. If reionization of He\text{II} happens locally on a timescale that is short compared to the recombination timescale, which for He\text{III} is of the order of the age of the universe at $z \sim 3$, then the energy density injected by photoionization will be proportional to the gas density. Consequently, the temperature increase will be independent of density and the equation of state of the IGM will become more isothermal. The change in the slope of the equation of state at $z \sim 3$ is thus physically consistent with our interpretation of the temperature jump at the same redshift.

Globally, reionization may take some time to complete, which would give rise to large spatial fluctuations in the temperature. Because we measure the temperature-density relation from the lower cut-off of the b(N)-distribution, our results should be regarded as lower limits to the average temperature. Absorption lines arising in a local, hot ionization bubble would not necessarily raise the observed cut-off in the b(N)-distribution.

The errors in Figs 5 and 6 can be directly traced back to the corresponding b(N)-distributions (Fig. 1). Take for example 0827a, which has an extremely low value of $\gamma$, with a very large error. This is clearly due to the gap in the b-distribution at log $N \sim 12.5$–13.0. The lack of lines in that region could be a statistical fluctuation, or it could be an indication of large variations in the temperature of the IGM.

Our results are in qualitative agreement with those reported recently by Ricotti et al. (2000). They used a method which relies on the assumption that only thermal broadening contributes to the line widths of the absorption lines at the peak of the b-distribution and used approximate simulation techniques to determine the density - column density relation. By applying their method to published lists of Voigt profile fits they found that at $z \sim 3$, $\gamma$ is smaller than would be expected if the reionization of helium had been completed at high redshift.
7 SYSTEMATICS

In this section we will investigate whether there are any systematic effects that could affect our results.

7.1 Numerical resolution

The $b$-distribution has been shown to be very sensitive to numerical resolution (Theuns et al. 1998; Bryan et al. 1999). It is therefore important to check that the lower limit to the line widths in our simulations is not set by the numerical resolution. We have resimulated our reference model, L1, which is the second coldest model, twice more at a lower resolution. The resolution was decreased by decreasing the number of particles from $2 \times 64^3$ to $2 \times 54^3$ and $2 \times 44^3$ respectively, while keeping the size of the simulation box constant. The resulting probability distributions for the intercept and the slope of the $b(N)$ cut-off are plotted in Fig. 7 for our lowest and highest redshift samples. Only for the lowest resolution simulation do the differences become noticeable. The intercept increases slightly and the slope becomes slightly shallower, indicating that the lines at the low column density end are not resolved. We conclude that the simulations used for this work have sufficient resolution.

7.2 Simulation box size

In order to investigate the effect of the simulation box size, we resimulated model L3 using a larger box, but with the same resolution ($5 h^{-1}$ Mpc and $128^3$ instead of $2.5 h^{-1}$ Mpc and $64^3$). The effect of increasing the size of the simulation box is more difficult to determine than the effect of numerical resolution. When the box size is increased, the gas becomes slightly hotter ($T_0$ increases by a few per cent), presumably because shock heating is more effective due to the larger infall velocities. Although this does not affect the equation of state derived from the observations, it does complicate the interpretation of the effect of the box size on the relation between the cut-off and the equation of state. Taking this correction into account, we find that using the larger simulation box results in higher values of $T_0$. The effect is negligible at $z \sim 4$ and increases to $\Delta \log T_0 \approx 0.05$ ($\Delta T_0/T_0 \approx 0.12$) at $z \sim 2$. The change in the derived value of $T_0$ is very small ($0.05$). Although $T_0$ and $\gamma$ may change a bit more if we increase the box size further, the small difference between the two box sizes indicates that the effect of the box size is insignificant compared to the statistical errors.

7.3 Cosmology

STLE showed that the relation between the cut-off in the $b(N)$ distribution and the equation of state is independent of the assumed cosmology. However, the initial power spectra of the models investigated by STLE were all normalized to match the observed abundance of galaxy clusters at $z = 0$. Bryan & Machacek (2000) claimed that the $b$-distribution depends strongly on the amplitude of the power spectrum, as predicted by the model of Hui & Rutledge (1999). However, Theuns et al. (2000) found the dependence to be very weak, provided that the line fitting is done using an algorithm, like VPFIT, that attempts to deblend absorption lines into a set of thermally broadened components. Although the absorption features do become broader for models with less small-scale power, the curvature in the line centers does not change much. Consequently, the total number of Voigt profile components used by VPFIT will generally increase, but the fits to the line centers will change very little.

In order to quantify the effect of decreasing the amount of small-scale power, we resimulated our reference model L1 twice using a lower normalization of the initial power spectrum. We find that normalizing to $\sigma_8 = 0.65$ instead of $\sigma_8 = 0.9$, changes the derived values of $T_0$ by less than 3 per cent for all redshifts. The change in $\gamma$ is negligible, except around $z \sim 3$, where it is higher by about 0.1. For the extreme case of $\sigma_8 = 0.4$, the derived values of $T_0$ are about 15 per cent lower and $\gamma$ differs by about 0.15. We conclude that the effect of the uncertainty in the normalization of the primordial power spectrum is very small.

7.4 Continuum fitting and the mean absorption

Errors in the continuum fit of the observed spectra, will lead to errors in the effective optical depth. Underestimating the observed continuum will decrease the measured effective optical depth. Decreasing the mean absorption in the simulations will increase the density corresponding to a given column density. Hence the slope of the cut-off will remain unchanged, but the intercept will increase, although the effect is small (STLE). Increasing the intercepts of the calibrating simulations will decrease the derived temperature $T_0$. The higher derived density contrast will work in the same direction (provided that $\gamma > 1$), resulting in a lower $T_0$.

The measured effective optical depth for sample 2343 seems to be relatively high compared to the other samples (Fig. 4). We tried scaling the synthetic spectra to the effective optical depth corresponding to the dashed line in Fig. 4 at the redshift of 2343, which is about 30 per cent lower than the measured value. This resulted in an increase of log $T_0$ by $0.04 (\Delta T_0/T_0 \approx 0.09)$, while leaving $\gamma$ unchanged.

In addition to errors in the continuum fit of the observed spectra, the continuum fitting of the synthetic spectra could also lead to systematic errors. We checked this by repeating the analysis of the simulations corresponding to our highest redshift sample, 2237b, but this time without continuum fitting the synthetic spectra. In this case the derived value of $T_0$ would be 9 per cent lower, while the value of $\gamma$ would be higher by 0.07. Hence systematic effects in the continuum fitting of the observed and the synthetic spectra are unlikely to be important.

8 DISCUSSION

The high ionization potential of He II requires ionizing sources with a hard spectrum, e.g. quasars. The dashed lines in Figs. 5 and 6 are for a simulation which uses a uniform UV background from quasars and galaxies, constructed to fit the observations. This model, for which stellar sources ionize HI and He I by $z \sim 5$ and quasars ionize He II at $z \sim 3.2$, has a much softer spectrum at high redshift. The good match to the observations (reduced $\chi^2 = 0.24$ for log $T_0$...
The thermal history of the IGM

0.00  0.02  0.04  0.06
P

Fig. 7. The effect of numerical resolution on the b(N) cut-off. Shown are the probability distributions for the intercept (left) and slope (right) of the cut-off in the b(N)-distributions of models L1-Q1100 (z = 1.97) (top) and L1-Q2237b (z = 4.27) (bottom), using 2 × 64^3 (solid), 2 × 54^3 (dashed) and 2 × 44^3 (dot-dashed) particles respectively. The small difference between the intermediate and high resolution simulations indicates that the latter has converged. The intercept was measured at a column density of 10^{14.0} cm^{-2} for L1-Q2237b and 10^{13.4} cm^{-2} for L1-Q1100.

and 1.38 for γ) suggests that at high redshift the contribution from quasars to the ionizing background at 4 Rydberg has been overestimated by HM. Before reionization, when the gas is optically thick to ionizing photons, the energy per photoionization is higher than in the optically thin limit (Abel & Haehnelt 1999). We have approximated this effect in this simulation by enhancing the photoheating rates during reionization, so raising the temperature of the IGM.

A quantitative comparison requires more realistic simulations of reionization, that include radiative transfer effects. In such models, reionization will occur inhomogeneously, going to completion first in the voids and progressing to higher density regions as the intensity of the photoionizing background rises (Miralda-Escudé, Haehnelt & Rees 2000, Gnedin 2000). The spectrum of the ionization front will initially be hard because the cross-section for ionization decreases rapidly with photon energy (Abel & Haehnelt 1999). By the time the high density regions become ionized, the mean free path will be much larger and the spectrum may be closer to the intrinsic spectrum of the sources. This effect, which could result in an equation of state with a negative slope (γ < 1), will be counterbalanced by the fact that recombinations are more important in the higher density regions, leading to increased photoheating.

Although the sharp rise of the temperature in our crude model is suggestive, the form of the peak is clearly not very well constrained by the data. Locally (e.g. in our small simulation box), the reionization has to occur on a timescale short compared to the recombination timescale for HeII, which is less than half a Hubble time for gas at the mean density at redshift three, in order for γ to decrease. However, as discussed above, radiative transfer effects may also help to decrease γ, which would allow for a more gradual temperature increase. Furthermore, globally reionization may take some time to complete, which would also lead to a more gradual rise in the derived temperature.

Although the reionization of helium may not be the only process which can explain the derived peak in the temperature, it appears to be the only process that can simultaneously account for the observed decrease in γ. Galactic winds for example, would tend to be less important in the low density regions, and would therefore result in an increase in the slope of the equation of state. Another possible explanation is a hardening of the ionizing background at z ∼ 3, as might be expected because of the increase in the number of quasars. If helium had already been ionized, this would still raise the temperature somewhat, but the effect would be stronger in the high density regions where the gas recombines faster. Hence this would also lead to an increase in γ, contrary to what is observed.

There are two other lines of evidence for late reionization of HeII. The first are direct measurements of the optical depth from HeII Lyα absorption. Although these observations are difficult because they need to be done in the UV and because most sightlines are intersected by a Lyman limit system which absorbs all the flux at short wavelengths, HeII Lyα absorption has so far been observed in four quasars (Jakobsen et al. 1994; Davidsen et al. 1996; Reimers et al.
These observations already provide strong evidence for a drop in the mean absorption from $z \sim 3.0$ to 2.5. The Far Ultraviolet Spectroscopic Explorer, launched in June 1999, may soon provide better constraints on the evolution of the mean He II over the interval $z = 2-3$.

The second piece of evidence concerns a change in the spectral shape of the ionizing background. As He II is ionized, the mean free path of hard UV photons will increase and the spectrum of the UV background will become harder. Songaila & Cowie (1996) and Songaila (1998) have reported a rapid increase with decreasing redshift of the Si IV/C IV ratio at $z \sim 3$, which they interpreted as evidence for a sudden reionization of He II. However, Boksenberg et al. (1998) found only a gradual change with redshift. The interpretation of this metal line ratio is complicated because local stellar radiation is likely to be important (Giroux & Shull 1997).

It should be kept in mind that these three different types of observations probe different physical structures. Our results apply to density fluctuations around the cosmic mean, the effective optical depth depends mostly on the neutral fraction in the voids and the metal line ratios probe the high density peaks. These structures will probably not be ionized simultaneously. After the ionization front breaks through the haloes surrounding the source of He II ionizing photons, e.g. a quasar, it will propagate quickly into the voids. The filaments, where the recombination rate is much higher, will get ionized more slowly, starting from the outside (Miralda-Escudé et al. 2000; Gnedin 2000). The IGM will still be thick to Lyo photons when only a small neutral fraction remains in the voids. Hence a drop in the He II Lyo optical depth at $z \sim 3$ would suggest that the reionization of the voids, which of course cover most of the volume, is complete (Miralda-Escudé 1998).

Detailed modeling, probably in the form of large hydrodynamical simulations, which include radiative transfer, is required to see whether the various observational constraints can be fit into a consistent picture. However, the ingredients necessary to explain our discovery of a peak in the temperature of the IGM at $z \sim 3$ are clear even from our crude model: a softer background at high redshift to delay helium reionization and enhanced heating rates compared to the optically thin limit. Once the evolution of helium heating is understood, the measurements of the temperature at higher redshifts can be used to constrain the epoch of hydrogen reionization.

Finally, we would like to note that because of their hard spectrum, quasars tend to ionize helium shortly after hydrogen, although the delay depends on the clumpiness of the IGM (Madau, Haardt & Rees 1998). It may therefore be difficult to postpone the reionization of helium until $z \sim 3$, if quasars were responsible for reionizing hydrogen at $z > 5$. Hence the mounting evidence for helium reionization at $z \sim 3$ suggests that hydrogen was reionized by stars.

9 SUMMARY

The low column density Ly$\alpha$ absorption lines in the spectra of high redshift ($z \sim 2-5$) quasars arise in the smoothly fluctuating, photoionized intergalactic medium (IGM), which contains most of the baryons. For densities around the cosmic mean, shock heating is not important and most of the gas follows a tight temperature-density relation. This effective ‘equation of state’ is well described by a power-law, $T = T_0(1 + \delta)^{-1}$. If the gas is ionized on a timescale short compared to the recombination timescale, then it will become nearly isothermal. When ionization equilibrium is established, higher density regions undergo more recombinations and therefore increased photoheating, leading to an increase in $\gamma$. Since the low-density gas can only cool adiabatically over a Hubble time, it retains some memory of when and how it was ionized.

A standard way of analyzing Ly$\alpha$ forest spectra is by decomposing them into a set of Voigt profiles. Although different mechanisms contribute to the broadening of the absorption lines, the temperature of the gas provides a lower limit to the line widths ($b$-parameters). Because the column density ($N$) is tightly correlated with the density of the absorber, the cut-off in the $b(N)$ distribution traces the equation of state.

Care has to be taken in inferring the evolution of the temperature of the IGM from the evolution of the observed line widths. The minimum line width depends on the temperature of the corresponding absorbers. If these are not at the mean density, then their thermal evolution will be different from that of $T_0$. For slightly overdense absorbers, the increase in $\gamma$ with time can compensate the decrease in $T_0$, resulting in an almost constant, or even increasing temperature. Furthermore, the expansion of the universe ensures that the overdensity corresponding to a fixed column density is a strongly decreasing function of redshift. Hence the $b$-parameters at a given column density are expected to be smaller at higher redshift, even if the temperature is higher.

We analyzed nine high-resolution, high signal-to-noise ratio quasar spectra, spanning the redshift range 2.0–4.5. In order to avoid confusion with the Ly$\beta$ forest, only the regions of a spectrum between the quasar’s Ly$\beta$ and Ly$\alpha$ emission lines were considered. In addition, spectral regions close to the quasar were omitted to avoid proximity effects. Regions thought to be contaminated by metals and damped Ly$\alpha$ lines were removed. In order to minimize the effects of redshift evolution and S/N variation across a single spectrum, we divided each Ly$\alpha$ forest spectrum (except the two at lowest redshifts) in two parts of equal length. Voigt profiles were fitted to the absorption lines using an automated version of VPFIT.

At redshift $z \sim 4$, the effective optical depth is lower than predicted by models of the ionizing background from quasars. Although this discrepancy can be alleviated by decreasing the baryon density in the models, this would lead to a strong disagreement at low redshift. Hence, the most likely explanation is that galaxies contribute significantly to the hydrogen ionizing background. We find some evidence that the mean absorption at $z \sim 3$ is higher than predicted, which suggests that the ionizing background from quasars has been overestimated.

We fitted power-laws, $b = b_0 (N/N_0)^{\frac{1}{2}}$, to the cut-offs of the $b(N)$ distributions of the samples of observed absorption lines. For each sample of observed absorption lines, the relations between the parameters of the $b(N)$ cut-off and the equation of state were calibrated using a set of synthetic spectra extracted from hydrodynamical simulations.
The background flux in the simulations was rescaled to give the mock spectra the same mean absorption as the corresponding observed spectra. The spectra were then processed to give them identical characteristics (resolution, pixel size, noise properties) as the observations. Voigt profiles were fitted to the absorption lines using the same automatic fitting package as was used for the observations, which is a prerequisite for a fair comparison, because the decomposition of spectra into Voigt profiles is non-unique. We have checked possible systematic errors arising from the finite numerical resolution, the finite size of the simulation box, the amount of small-scale power and continuum fitting errors in both synthetic and observed spectra. In all cases, the effects were small compared to the statistical errors.

The measured thermal evolution differs drastically from the scenario predicted by current models of the ionizing background from quasars, in which helium is fully reionized by \( z \sim 4.5 \). At redshift \( z \sim 3.5 \) the temperature increases until it peaks around \( z \sim 3 \), after which it decreases again (Fig. 5). The slope of the equation of state reaches a minimum at \( z \sim 3 \), where it becomes close to isothermal (Fig. 6). These results suggest that the universe was reheated from \( z \sim 3.5-3.0 \), which we interpret as reheating associated with the reionization of He\( ^{\text{II}} \).

Although other processes, like galactic winds or a hardening of the UV-background, might also cause the temperature to increase, they would result in an increase of \( \gamma \), contrary to what is observed. Furthermore, measurements of the He\( ^{\text{II}} \) Ly\( \alpha \) opacity and the ratio Si\( ^{\text{III}} \)/C\( ^{\text{IV}} \) also suggest that He\( ^{\text{II}} \) was reionized at \( z \sim 3 \). Since quasars tend to ionize helium shortly after reionization, the mounting evidence for late helium reionization suggests that hydrogen was reionized by stars.

Detailed modeling, probably in the form of large simulations which include radiative transfer, is required to see whether the different observational constraints can be fit into a consistent physical picture. However, two necessary ingredients are already apparent: enhanced heating rates compared to the optically thin limit during reionization in order to match the height of the peak in the temperature and a soft background at \( z > 3 \) in order to delay the reionization of helium.

ACKNOWLEDGMENTS

We would like to thank Bob Carswell for letting us use the HIRES spectra of Q1100–264 and for helping us with VP-FIT. We are also grateful to Sara Ellison for letting us use the HIRES spectra of APM 08279+5255. We thank Martin Rees, Lai Hui, Jordi Miralda-Escudé and Martin Haehnelt for stimulating discussions. JS thanks the Isaac Newton Trust, St. John’s College and PPARC for support and GE thanks PPARC for the award of a senior fellowship. Research was conducted in cooperation with Silicon Graphics/Cray Research utilising the Origin 2000 supercomputer COSMOS, which is a UK-CCC facility supported by HEFCE and PPARC. This work has been supported by the TMR network on ‘The Formation and Evolution of Galaxies’, funded by the European Commission.

REFERENCES

APPENDIX A1: MEASURING THE EQUATION OF STATE

This appendix contains a detailed discussion of the various steps in our method for measuring the effective equation of state, which was outlined in section 5. Section A1.1 discusses the parameters that determine which absorption lines are used to determine the cut-off in the \( b(N) \) distribution, \( b = b_{N_0}(N/N_0)^{-1} \). Our choice of the pivot column density, \( N_0 \), is also discussed in this section. In section A1.2 we describe our method for measuring the parameters of the \( b(N) \) cut-off. Our method for measuring the density and temperature corresponding to the pivot column density is discussed in section A1.3. Finally, section A1.4 describes how \( \gamma \) is measured and how the above results are combined to yield \( T_0 \).

A1.1 Choice of parameters

The cut-off in the \( b(N) \)-distribution is not absolute because of line blending and contamination from unidentified metal lines. STLE developed an iterative procedure for fitting a power-law to the cut-off that is insensitive to these narrow lines which fall below the edge in the \( b(N) \)-distribution. We will use the same algorithm with one minor modification: we only consider absorption lines for which VPFIT gives relative errors smaller than 25 per cent in both the \( b \)-parameter and column density, whereas STLE allowed errors of up to 50 per cent. This more conservative criterion has been applied to extend the method to high redshifts where line blending is more important. The formal errors given by VPFIT are a measure of the degeneracy of the fit and consequently lines in blends tend to have large errors. By removing all lines with errors greater than 25 per cent we exclude the heavily blended lines and this significantly sharpens the cut-off.

We find that blends dominate at column densities lower than \( N = 10^{12.5}\text{cm}^{-2} \) and therefore take this value as the lower limit of the column density interval over which we measure the cut-off. Our choice of upper limit depends on redshift (column 5 of Table 2) and was determined by the following considerations: (a) the cut-off can be measured more accurately if the column density interval is larger; (b) we only want to measure the cut-off for column densities that correspond to the density range for which the gas follows a power-law temperature-density relation.

STLE showed that the slope of the cut-off, \( \Gamma - 1 \), is proportional to the slope of the equation of state, \( \gamma - 1 \), and that the intercept of the cut-off, \( \log b_{N_0} \), is linearly related to \( \log T_b(N_0) \) (essentially because thermal broadening and Jeans smoothing both scale as \( T_b^{-1/2} \)), where \( \delta(N_0) \) is the density contrast corresponding to the pivot column density \( N_0 \). Hence, if \( N_0 \) is set equal to the column density corresponding to the mean density, \( \log T_0 \) can be measured directly. In general, this will not be the case and we have to compute \( T_0 \) from the measured values of \( T_b \), \( \gamma \) and \( \delta \) (equation A1).

If \( \log N_0 \) does not correspond to the mean \( \log N \) of the lines near the cut-off, then the errors in the measured intercept and slope of the cut-off are correlated (this is just the general correlation one gets for a linear fit if the average abscissa value is nonzero). Hence there is a trade-off between measuring the temperature at a density close to the mean, at the cost of larger errors in the measured \( b(N) \) cut-off, or computing \( T_0 \) from a more precise measurement of the temperature at some higher overdensity and the measured values of \( \delta(N_0) \) and \( \gamma \), in which case the error in the latter will dominate the uncertainty in \( T_0 \). It is important to keep in mind that in both cases the errors in \( T_0 \) and \( \gamma \) will not be independent. We find that in general the error in \( T_0 \) is somewhat smaller when \( N_0 \) is chosen to be near the mean density. Our choice of \( \log N_0 \), which is not particularly critical, is listed in column 7 of Table 2.

A1.2 Measuring \( b_{N_0} \) and \( \Gamma \) for the observations

The cut-off in the \( b(N) \)-distribution is fitted using the iterative procedure developed by STLE. Since the number of absorption lines in each observed sample is small, the statistical variance in the measured cut-off is considerable. We reduce the statistical variance by bootstrap resampling (drawing \( n \) lines at random from the complete set of \( n \) lines, with replacement) of the observed set of absorption lines, as proposed by STLE. Probability distributions for the parameters of the cut-off are generated by fitting the cut-off of 1000 bootstrap realizations. The medians are used as the best estimates of the true parameters.

Because there will invariably be realizations with many duplicate data points, the resulting probability distributions will have a non-Gaussian tail and will therefore tend to overestimate the true errors in the median. We determined the relation between the errors estimated from the bootstrap analysis and the true errors using Monte Carlo simulations. The cut-off was fitted using the bootstrap analysis for 1000 random realizations of model L1-1100, i.e. we generated 1000 probability distributions for the parameters of the \( b(N) \) cut-off. The resulting medians are normally distributed and we took the average medians as the true values of the parameters of the cut-off. We then determined for what fraction of the 1000 iterations the true value was within a given apparent (i.e. determined from the bootstrap probability distribution) confidence limit. This fraction is then the true confidence limit corresponding to the given apparent confidence limit. The procedure was repeated for model L1-
237b, which has a higher redshift (4.1 vs 1.0) and many more lines (127 vs 44).

The results are shown in Fig. A1. The four solid curves correspond to the intercept and the slope of the cut-offs of the two models. As expected, the confidence limits from the bootstrap probability distributions overestimate the errors (true confidence limits are higher than bootstrap confidence limits). All curves follow a similar track, even though the two models are very different. The 50 per cent confidence limit from the bootstrap analysis is in reality a 68 per cent confidence limit, i.e. 68 per cent of the times the true value will be within the 50 per cent confidence limit determined from the bootstrap analysis. We will therefore take the 50 per cent confidence limits from the bootstrap analysis as our 1 sigma errors.

A1.3 Measuring $T_{\delta(N_0)}$ and $\delta(N_0)$

Fig. A2 illustrates our method for measuring $T_0$ from the intercept of the $b(N)$ cut-off, $\log b_{14.0}$. In the left panel the intercept of the cut-off, measured at $N_0 = 10^{14.0}\text{ cm}^{-2}$ in this example, is plotted as a function of $\log T_0$ for the simulations of sample 1422a (filled circles). As expected, the relation between the intercept, $\log b_{14.0}$, and $\log T_0$ is not very tight. The large scatter arises because $\log b_{14.0}$ is not proportional to $\log T_0$, but to $\log T_{\delta(N_0)}$, where $\delta(N_0)$ is the density contrast corresponding to the pivot column density $N_0 = 10^{14.0}\text{ cm}^{-2}$. Indeed, if we plot $\log b_{14.0}$ as a function of $\log T_{\delta=1.6}$ (right panel), the scatter becomes very small, implying that $\delta(N_0) \approx 1.6$.

Changing the value of $\delta$ shifts the data points in the plot horizontally, the amount depending on the value of $\gamma$ in the simulation corresponding to the data point. Because we do not know a priori what the density contrast corresponding to $N_0$ is, we vary $\delta$ and see for which value the scatter is minimal. The inset in the right panel shows the total $\chi^2$ of the linear least squares fit to the data points as a function of the density contrast. For this example the scatter is minimal for $\delta = 1.6$, which is the density contrast used in the right panel.

In principle, the density - column density relation can be determined directly from the simulations. It is however not clear how to define the density corresponding to an absorption line. Typically, peculiar and thermal velocities will cause gas at a range of densities to contribute to the optical depth at a particular point in redshift space. Furthermore, each absorption line spans a finite redshift range. STLE used the optically depth weighted density and temperature to investigate the relation between the parameters of the Voigt profile fits and the physical state of the absorbing gas. Using the optically depth weighted density - column density relation, we find that the column density $N = 10^{14.0}\text{ cm}^{-2}$ does indeed correspond to a density contrast of about 1.6.

The dot-dashed lines in Fig. A2 show the $b$-value expected for pure thermal broadening, $b = \sqrt{2k_B T/m_p}$, where $k_B$ is the Boltzmann constant and $m_p$ is the proton mass. The relation between $\log b_{14.0}$ and $\log T_{\delta=1.6}$ lies slightly above the pure thermal broadening line, but has the same slope. This implies that the widths at the cut-off are dominated by thermal broadening, with a small additional component that also scales as $T^{1/2}$. We identify this last component as the differential Hubble flow across the absorber, whose size is set by the Jeans smoothing scale and does indeed depend on the square root of the temperature. We consistently found that minimizing the scatter in the $\log b_{14.0}$ - $\log T_0$ relation by varying $\delta$ results in a relation that has a slope of about 0.5 and that the value of $\delta$ found agrees well with direct measurements of the density contrast corresponding to the column density $N_0$. We therefore use this procedure to estimate $\delta$ and $T_0$ corresponding to an observed $\log b_{14.0}$ and conservatively estimate the error in the density contrast to be $\sigma(\log(1 + \delta)) = 0.15$.

Although for this work the error from the determination of the density - column density relation does not contribute significantly to the total error in $T_0$, it may well become the limiting factor when a larger sample of quasars is used. If very high precision is required, it will probably be necessary to determine the equation of state by matching the observed $b(N)$ cut-off directly to those from simulations with the correct thermal evolution, thereby bypassing the need to determine the density corresponding to the pivot column density $N_0$.

After having measured the thermal evolution of the IGM, we ran a simulation designed to match the observations (dashed lines in Figs. 5 and 6). The open squares in Figs. A2 and A3 correspond to this simulation. The difference between the evolution in the calibrating simulations and the true evolution is particularly large at $z \sim 3$ (c.f. Fig. 5), the redshift of model 1422a. The fact that the open square follows the same relation as the other data points, therefore shows that our results would have been the same if our calibrating simulations had had the correct thermal evolution. In particular, it confirms that a model whose equation of state matches the one determined from the observations using the methods described in this section, does indeed have the observed $b(N)$ cut-off.

A1.4 Measuring $T_0$ and $\gamma$

After having converted the observed intercept into a measurement of the temperature at the density contrast $\delta$, we
Figure A2. The intercept of the $b(N)$ cut-off as a function of the temperature at mean density (left panel) and at a density contrast of $\delta = 1.6$ (right panel). Filled circles are the simulations corresponding to the observed sample 1422a. The open square is for a simulation whose thermal evolution matches the observations (dashed line in Fig. 5). The dashed line is the least-squares fit for the filled circles. The dot-dashed line indicates the $b$-value corresponding to pure thermal broadening: $b = (2k_B T/m_p)^{1/2}$. The inset in the right panel shows how the total $\chi^2$ of the fit varies as a function of $\delta$; $\chi^2$ is minimum for $\delta = 1.6$, which is the density contrast corresponding to the column density of the intercept ($10^{14.0} \text{ cm}^{-2}$ in this example).

Figure A3. The slope of the $b(N)$ cut-off as a function of the slope of the equation of state. Filled circles are the simulations corresponding to the observed sample 1422a. The open square is for a simulation whose thermal evolution matches the observations (dashed line in Fig. 6). The dashed line is the least-squares fit for the filled circles.

from the simulations. To these errors we add in quadrature the residual scatter of the data points around the linear fit. The error in $\log T_0$ is then given by,

$$\Delta^2(\log T_0) = \Delta^2(\log T_0) + [\Delta(\gamma) \log(1 + \delta)]^2 + [(\gamma - 1) \Delta \log(1 + \delta)]^2,$$

where we have taken $\Delta \log(1 + \delta) = 0.15$ (cf. section A1.3).

Each measurement of $\log b_{N_0}$ and $\Gamma$ comes with associated errors, which are determined as described in section A1.2. These errors can be converted directly into errors in $\log T_0$ and $\gamma$ respectively, using the linear relations between the cut-off and the equation of state, determined from the simulations.

$$\log T_0 = \log T_0 - (\gamma - 1) \log(1 + \delta).$$

(A1)