Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **452,** 217–234 (2015) doi:10.1093/mnras/stv1182

The neutral hydrogen cosmological mass density at $z = 5$

Neil H. M. Crighton,[1⋆] Michael T. Murphy,[1] J. Xavier Prochaska,[2] Gábor Worseck,[3] Marc Rafelski,[4] George D. Becker,[5] Sara L. Ellison,[6] Michele Fumagalli,[7,8] Sebastian Lopez,[9] Avery Meiksin[10] and John M. O'Meara[11]

[1]*Centre for Astrophysics and Supercomputing, Swinburne University of Technology, Hawthorn, VIC 3122, Australia*
[2]*Department of Astronomy and Astrophysics, UCO/Lick Observatory, University of California, 1156 High Street, Santa Cruz, CA 95064, USA*
[3]*Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany*
[4]*Infrared Processing and Analysis Center, Caltech, Pasadena, CA 91125, USA*
[5]*Space Telescope Science Institute, 3700 San Martin Dr, Baltimore, MD 21218, USA*
[6]*Department of Physics and Astronomy, University of Victoria, Victoria, BC V8P 1A1, Canada*
[7]*Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK*
[8]*Carnegie Observatories, 813 Santa Barbara Street, Pasadena, CA 91101, USA*
[9]*Departamento de Astronomía, Universidad de Chile, Casilla 36-D, Santiago, Chile*
[10]*Scottish Universities Physics Alliance, Institute for Astronomy, University of Edinburgh, Blackford Hill, Edinburgh EH9 3HJ, UK*
[11]*Department of Chemistry and Physics, Saint Michael's College, One Winooski Park, Colchester, VT 05439, USA*

Accepted 2015 May 21. Received 2015 May 21; in original form 2015 February 3

ABSTRACT

We present the largest homogeneous survey of $z > 4.4$ damped Lyα systems (DLAs) using the spectra of 163 QSOs that comprise the Giant Gemini GMOS (GGG) survey. With this survey we make the most precise high-redshift measurement of the cosmological mass density of neutral hydrogen, $\Omega_{\rm H\,I}$. At such high redshift, important systematic uncertainties in the identification of DLAs are produced by strong intergalactic medium absorption and QSO continuum placement. These can cause spurious DLA detections, result in real DLAs being missed or bias the inferred DLA column density distribution. We correct for these effects using a combination of mock and higher resolution spectra, and show that for the GGG DLA sample the uncertainties introduced are smaller than the statistical errors on $\Omega_{\rm H\,I}$. We find $\Omega_{\rm H\,I} = 0.98^{+0.20}_{-0.18} \times 10^{-3}$ at $\langle z \rangle = 4.9$, assuming a 20 per cent contribution from lower column density systems below the DLA threshold. By comparing to literature measurements at lower redshifts, we show that $\Omega_{\rm H\,I}$ can be described by the functional form $\Omega_{\rm H\,I}(z) \propto (1+z)^{0.4}$. This gradual decrease from $z = 5$ to 0 is consistent with the bulk of H I gas being a transitory phase fuelling star formation, which is continually replenished by more highly ionized gas from the intergalactic medium and from recycled galactic winds.

Key words: quasars: absorption lines – cosmological parameters.

1 INTRODUCTION

The neutral hydrogen mass density of the Universe, $\Omega_{\rm H\,I}$, is an important cosmological observable. It determines the precision with which cosmological parameters can be constrained by observations of the H I intensity power spectrum (e.g. Barkana & Loeb 2007; Chang et al. 2008; Wyithe & Loeb 2008; Padmanabhan, Choudhury & Refregier 2015), and we expect its evolution to be linked to the cosmic star formation history. The main contributor to $\Omega_{\rm H\,I}$ is high column density, predominantly neutral gas clouds (e.g. O'Meara et al. 2007; Zafar et al. 2013), self-shielded from ionizing radiation and therefore likely fuel for future star formation (e.g. Wolfe, Gawiser & Prochaska 2005). Thus, tracing the evolution of $\Omega_{\rm H\,I}$ from the end of reionization, through the epoch of the cosmic star formation peak at $z \sim 2$ to the present day, is of central importance to our understanding of galaxy formation. It also provides an excellent integral constraint against which theoretical models of galaxy formation can be tested.

At redshift <0.3, H I 21 cm emission can be used to measure $\Omega_{\rm H\,I}$ either directly or by stacking analyses (e.g. Zwaan et al. 2005; Martin et al. 2010). At higher redshifts, where emission is too weak to be detected with current facilities, $\Omega_{\rm H\,I}$ can instead be inferred from the incidence rate of damped Lyα systems (DLAs, defined as absorption systems with $N_{\rm H\,I} \geq 20.3$ cm^{-2}), which trace the bulk of neutral gas in the Universe (Prochaska, Herbert-Fort & Wolfe 2005). These systems are detected in absorption in the spectra of background QSOs, and their characteristic damping

⋆ E-mail: neilcrighton@gmail.com

wings allow column densities to be measured even at low spectral resolution.

Early DLA surveys at $2 < z < 4$, which were typically comprised of a few hundred QSOs and assumed a cosmological deceleration parameter $q_0 = 0.5$ or 0, suggested that the gas mass density in DLAs may have been sufficient to produce most of the stars seen in the local Universe (Lanzetta et al. 1991; Wolfe et al. 1995; Storrie-Lombardi, McMahon & Irwin 1996). However, a change to a modern concordance cosmology revealed that DLAs at $z \sim 3$ contain <50 per cent of the present-day mass density in stars (e.g. Storrie-Lombardi & Wolfe 2000; Péroux et al. 2005, see also Section 5.2). In addition, recent DLA surveys at $2 < z < 4$ using more than 10 000 QSOs assembled from the Sloan Digital Sky Survey (SDSS; Prochaska & Herbert-Fort 2004; Prochaska et al. 2005; Noterdaeme et al. 2009, 2012; Prochaska & Wolfe 2009) have shown that there is very little evolution in the H I mass density from $z = 3$ to the present day. This is starkly at odds with the strong evolution in the star formation rate over the same period (e.g. Madau & Dickinson 2014). One view is that H I represents a transitory phase fuelling star formation (e.g. Prochaska et al. 2005; Davé et al. 2013), which is continually replenished by more highly ionized gas from either the intergalactic medium (IGM) or recycled galactic outflows.

While it is important to constrain $\Omega_{\rm H\,I}$ across the whole of cosmic history, it is of particular interest at the highest redshifts. Rafelski et al. (2014) report a decrease in the metal mass density in DLAs from $z = 5$ to 4.5, hinting at an abrupt change in the enrichment of H I gas past $z = 5$. This may be caused by a change in the population of objects containing neutral hydrogen, which could be accompanied by a similarly abrupt evolution in $\Omega_{\rm H\,I}$. Moreover, since massive stars in galaxies are believed to have reionized the Universe (e.g. Bouwens et al. 2012), it is important to track the evolution of the fuel for star formation up to the epoch of reionization. However, it is a challenge to assemble the large sample of high-redshift QSO spectra necessary for a $z > 4.5$ DLA survey. The decline in the QSO space density at $z > 3$ means that relatively few redshift >4.4 QSOs were observed by the SDSS, and those that were typically have too low a signal-to-noise (S/N) to reliably identify DLAs. For example, Rafelski et al. (2012, 2014) find a misidentification rate of 26 per cent for DLA candidates from SDSS DR5 at $z > 4$ and of 97 per cent for candidates from DR9 at $z > 4.7$. For this reason, smaller DLA surveys have been performed at higher redshift, often using higher resolution spectra to make robust identifications of DLAs. Péroux et al. (2003), Guimarães et al. (2009) and Songaila & Cowie (2010) have all presented measurements of $\Omega_{\rm H\,I}$ at $z > 4.5$. Songaila & Cowie (2010, hereafter S10) give a cumulative result including data from all these previous studies, and this represents the highest redshift measurement of $\Omega_{\rm H\,I}$ to date. They use a sample of 19 QSOs with emission redshifts >4.5, and their measurement hints at a possible downturn in $\Omega_{\rm H\,I}$ at $z \geq 4$, but the uncertainties from sample variance at $z > 4.3$ are large.

Here we measure $\Omega_{\rm H\,I}$ as traced by DLAs at $3.5 < z < 5.4$ using a homogeneous sample of 163 QSOs with emission redshifts between 4.4 and 5.4. This represents an increase in redshift path of a factor of 8 over S10 at $z > 4.5$. Identifying DLAs becomes increasingly difficult at higher redshift, as H I absorption from the highly ionized IGM becomes more severe, and blending with strong systems below the DLA threshold can cause misidentification of DLAs. Therefore, we carefully check for systematic misidentifications in our sample using both mock spectra and higher resolution spectra of DLA candidates. More than 70 per cent of our DLA candidates (and >85 per cent at $z > 4.5$) have been observed at higher resolution (Rafelski et al. 2012, 2014), allowing us to confirm their $N_{\rm H\,I}$ despite the increased IGM blending at high redshift.

This paper is structured as follows. In Section 2, we describe the QSO spectra used for the analysis. Section 3 describes the formalism used to derive $\Omega_{\rm H\,I}$ from our observations, and Section 4 describes our method for measuring the DLA incidence rate, accounting for systematic effects. Section 5 describes our main result, a measurement of the neutral hydrogen mass density at $z = 5$, and discusses its implications. Section 6 summarizes our conclusions. We assume a flat Λ cold dark matter cosmology, with $H_0 = 70\,{\rm km\,s^{-1}\,Mpc^{-1}}$, $\Omega_{\rm m,\,0} = 0.3$ and $\Omega_{\Lambda,\,0} = 0.7$. All distances are comoving unless stated otherwise. The data and code used for this paper are available at https://github.com/nhmc/GGG_DLA.

2 DATA

Our main data sample consists of Gemini Multi-Object Spectrograph (GMOS) spectra for the 163 QSOs which comprise the Giant Gemini GMOS (GGG) survey (Worseck et al. 2014). The QSOs were taken from the SDSS and all have emission redshifts $4.4 < z < 5.4$. At these emission redshifts, the QSO sightlines are likely unbiased regarding the number density of DLAs, unlike sightlines with $2.7 < z_{\rm em} < 3.6$ (Prochaska, Worseck & O'Meara 2009; Worseck & Prochaska 2011; Fumagalli et al. 2013). We also use a smaller sample of 59 QSOs with higher resolution spectra, listed in Table 1. In contrast to the GGG sample, most of these QSOs were targeted because of a known DLA candidate towards the QSO. One of these higher resolution spectra was taken with the Magellan Echellette Spectrograph on the Magellan Clay Telescope (Jorgenson, Murphy & Thompson 2013) and the remainder were taken with Echellette Spectrograph and Imager (ESI) on the Keck II Telescope (Rafelski et al. 2012, 2014). 39 of these QSOs are also in the GGG sample, and the remaining 20 have a similar emission redshift to the GGG QSOs. We use these higher resolution spectra to assess the reliability of our DLA identifications and to estimate the importance of systematic effects, but they are not included in the statistical sample used to measure $\Omega_{\rm H\,I}$. Fig. 1 shows the QSO emission redshift distribution for our sample and the redshift path, $g(z)$, where DLAs can be detected in comparison to previous high-redshift surveys. We define

$$g(z) = \sum H(z_i^{\rm max} - z)H(z - z_i^{\rm min}), \quad (1)$$

where H is the Heaviside step function, and $z_i^{\rm min}$ and $z_i^{\rm max}$ are redshift limits for detecting DLAs in each QSO spectrum (e.g. Zafar et al. 2013).

For a detailed description of the GGG spectra and the procedure used to reduce them, see Worseck et al. (2014). In brief, they were observed with the Gemini Multi-Object Spectrometers on the Gemini telescopes, yielding a typical S/N $\sim$ 20 per 1.85 Å pixel in the Lyα forest at a resolution of $\sim$5.5 Å (full width at half-maximum, FWHM). The spectral coverage was tuned to be roughly constant in the quasar rest frame (typically 850–1450 Å). The high-resolution ESI spectra we use[1] have a typical S/N of 15 per $10\,{\rm km\,s^{-1}}$ pixel and a resolution FWHM of $31\,{\rm km\,s^{-1}}$ (see Table 1). The single MagE spectrum has a similar S/N but a resolution of $56\,{\rm km\,s^{-1}}$.

[1] The reduced spectra are available at http://www.rafelski.com/data/DLA/hizesi

Table 1. Higher resolution spectra used in our analysis. Columns list the QSO name, RA and Dec. (J2000), emission redshift, the instrument used to take the spectrum, the median S/N per pixel over rest-frame wavelengths 1240–1280 Åand whether the QSO is in the GGG sample.

QSO name	z_{em}	Origin	S/N	GGG?
SDSS J000749.16+004119.6	4.78	ESI	9.7	No
SDSS J001115.23+144601.8	4.97	MAGE	31.4	Yes
SDSS J005421.42−010921.6	5.02	ESI	16.7	No
SDSS J021043.16−001818.4	4.77	ESI	7.9	Yes
SDSS J023137.65−072854.4	5.42	ESI	26.2	Yes
SDSS J033119.66−074143.1	4.73	ESI	17.8	Yes
SDSS J075618.10+410409.0	5.06	ESI	10.8	No
SDSS J075907.57+180054.7	4.82	ESI	18.3	Yes
SDSS J081333.30+350811.0	4.92	ESI	16.5	No
SDSS J082454.02+130217.0	5.21	ESI	22.5	Yes
SDSS J083122.60+404623.0	4.89	ESI	19.6	No
SDSS J083429.40+214025.0	4.50	ESI	21.0	No
SDSS J083920.53+352459.3	4.78	ESI	13.9	Yes
SDSS J095707.67+061059.5	5.18	ESI	18.1	No
SDSS J100449.58+404553.9	4.87	ESI	11.4	No
SDSS J100416.12+434739.0	4.87	ESI	19.7	Yes
SDSS J101336.30+424027.0	5.04	ESI	22.7	No
SDSS J102833.46+074618.9	5.15	ESI	12.4	No
SDSS J104242.40+310713.0	4.69	ESI	24.8	No
SDSS J105445.43+163337.4	5.15	ESI	21.5	Yes
SDSS J110045.23+112239.1	4.73	ESI	23.4	Yes
SDSS J110134.36+053133.8	5.04	ESI	22.3	Yes
SDSS J113246.50+120901.6	5.18	ESI	32.7	Yes
SDSS J114657.79+403708.6	5.00	ESI	25.7	Yes
SDSS J120036.72+461850.2	4.74	ESI	19.0	Yes
SDSS J120110.31+211758.5	4.58	ESI	31.8	Yes
SDSS J120207.78+323538.8	5.30	ESI	25.6	Yes
SDSS J120441.73−002149.6	5.09	ESI	15.6	Yes
SDSS J122042.00+444218.0	4.66	ESI	11.3	No
SDSS J122146.42+444528.0	5.20	ESI	15.5	Yes
SDSS J123333.47+062234.2	5.30	ESI	14.1	Yes
SDSS J124515.46+382247.5	4.96	ESI	16.4	Yes
SDSS J125353.35+104603.1	4.92	ESI	23.8	Yes
SDSS J130215.71+550553.5	4.46	ESI	24.0	Yes
SDSS J131234.08+230716.3	4.96	ESI	19.2	Yes
SDSS J133412.56+122020.7	5.13	ESI	10.8	Yes
SDSS J134040.24+281328.1	5.35	ESI	23.0	Yes
SDSS J134015.03+392630.7	5.05	ESI	17.7	Yes
SDSS J141209.96+062406.9	4.41	ESI	25.6	Yes
SDSS J141839.99+314244.0	4.85	ESI	15.3	No
SDSS J142103.83+343332.0	4.96	ESI	24.1	No
SDSS J143751.82+232313.3	5.32	ESI	24.4	Yes
SDSS J143835.95+431459.2	4.69	ESI	27.0	Yes
SDSS J144352.94+060533.1	4.89	ESI	5.6	No
SDSS J144331.17+272436.7	4.42	ESI	24.4	Yes
SDSS J151320.89+105807.3	4.62	ESI	8.3	Yes
SDSS J152345.69+334759.3	5.33	ESI	8.6	No
SDSS J153459.75+132701.4	5.04	ESI	4.9	Yes
SDSS J153627.09+143717.1	4.88	ESI	8.2	No
SDSS J160734.22+160417.4	4.79	ESI	17.9	Yes
SDSS J161425.13+464028.9	5.31	ESI	13.1	Yes
SDSS J162626.50+275132.4	5.26	ESI	33.9	Yes
SDSS J162629.19+285857.5	5.04	ESI	12.0	Yes
SDSS J165436.80+222733.0	4.68	ESI	33.6	No
SDSS J165902.12+270935.1	5.32	ESI	23.1	Yes
SDSS J173744.87+582829.6	4.91	ESI	12.9	Yes
SDSS J221644.00+001348.0	5.01	ESI	8.2	No
SDSS J225246.43+142525.8	4.88	ESI	14.8	Yes
SDSS J231216.40+010051.4	5.07	ESI	4.8	No

3 FORMALISM

Our aim is to measure the cosmic H I mass density at $3.5 < z < 5.4$. The bulk of the neutral gas at $2 < z < 5$ is in DLAs, with an ∼15 per cent contribution from sub-DLAs (which have $10^{19} < N_{\rm H\,I}/({\rm cm}^{-2}) < 10^{20.3}$) and more highly ionized Lyman limit and Lyα forest absorbers with $N_{\rm H\,I} < 10^{19}\,{\rm cm}^{-2}$ (Péroux et al. 2005; Prochaska et al. 2005; O'Meara et al. 2007; Zafar et al. 2013). There are several ways to express the comoving mass density of neutral hydrogen used in the literature. For measurements at low redshift using radio emission, authors typically quote $\Omega_{\rm H\,I}$, which is the mass of neutral hydrogen alone, excluding any mass in molecules and helium. For DLA absorption studies, authors generally quote the gas mass in DLAs, $\Omega_g^{\rm DLA}$ (sometimes the g subscript is omitted) including a factor μ to account for helium. Prochaska et al. (2005) advocate using the quantity $\Omega_g^{\rm neut}$, which is the mass in predominantly neutral gas, which can be different from $\Omega_g^{\rm DLA}$. In this work, we quote the mass density from H I alone, $\Omega_{\rm H\,I}$, and exclude any mass contribution from helium or molecules. Due to contamination and the low resolution of the GMOS spectra, we only measure H I in DLAs, $\Omega_{\rm H\,I}^{\rm DLA}$. To convert to $\Omega_{\rm H\,I}$, we apply a correction derived from measurements of lower $N_{\rm H\,I}$ systems in previous work.

We measure $\Omega_{\rm H\,I}^{\rm DLA}$ by counting the incidence rate of DLAs in the spectra and measuring $N_{\rm H\,I}$ from their strong damping wings. Below is a summary of the formalism used to derive $\Omega_{\rm H\,I}^{\rm DLA}$ from the DLA incidence rate. See section 4.1 of Prochaska et al. (2005) and the review by Wolfe et al. (2005) for a more detailed description.

The number of DLAs in the intervals $(N_{\rm H\,I}, N_{\rm H\,I} + {\rm d}N_{\rm H\,I})$ and $(X,\ X\ +\ {\rm d}X)$ is defined as the frequency distribution, $f_{\rm DLA}(N_{\rm H\,I}, X){\rm d}N_{\rm H\,I}{\rm d}X$. Here X is the 'absorption distance', defined such that a non-evolving population has a constant absorption frequency:

$$ {\rm d}X \equiv \frac{H_0}{H(z)}(1+z)^2\,{\rm d}z, \tag{2} $$

where H is the Hubble parameter. The DLA incidence rate is then

$$ \ell_{\rm DLA}(X)\,{\rm d}X = \int_{N_{\rm H\,I,min}}^{\infty} f_{\rm DLA}(N_{\rm H\,I}, X)\,{\rm d}N_{\rm H\,I}\,{\rm d}X. \tag{3} $$

It is related to the comoving number density of DLAs, $n_{\rm DLA}(X)$, and the proper absorption cross-section, $A(X)$, by

$$ \ell_{\rm DLA}(X) = \frac{c}{H_0} n_{\rm DLA}(X) A(X). \tag{4} $$

Since DLAs are mostly neutral, the H I mass per DLA is $m_{\rm H} N_{\rm H\,I} A(X)$, where $m_{\rm H}$ is the hydrogen atom mass. Combining this with equation (4) gives

$$ \begin{aligned} \Omega_{\rm H\,I}^{\rm DLA}(X){\rm d}X &= \frac{H_0}{c}\frac{m_{\rm H}}{\rho_{\rm crit,0}} \int_{N_{\rm H\,I,min}}^{\infty} N_{\rm H\,I} f_{\rm DLA}(N_{\rm H\,I}, X){\rm d}N_{\rm H\,I}{\rm d}X \\ &= \frac{8\pi G}{3H_0}\frac{m_{\rm H}}{c} \int_{N_{\rm H\,I,min}}^{\infty} N_{\rm H\,I} f_{\rm DLA}(N_{\rm H\,I}, X){\rm d}N_{\rm H\,I}{\rm d}X. \end{aligned} \tag{5} $$

$N_{\rm H\,I,min} = 10^{20.3}\,{\rm cm}^{-2}$, so this expression does not include the contribution from lower $N_{\rm H\,I}$ systems to $\Omega_{\rm H\,I}$. We discuss how we include this contribution in Section 3.2.

Due to the low resolution of the GMOS spectra, confusion from the strong Lyα forest absorption at $z > 4$, uncertainty in the continuum level and systematics affecting sky subtraction, the measured frequency of DLAs, $f_{\rm meas}(N_{\rm H\,I})$, may differ from the true $f_{\rm DLA}$. Therefore, we introduce a correction factor $k(N_{\rm H\,I})$ such that

$$ f_{\rm DLA}(N_{\rm H\,I}) = f_{\rm meas}(N_{\rm H\,I}) k(N_{\rm H\,I}). \tag{6} $$

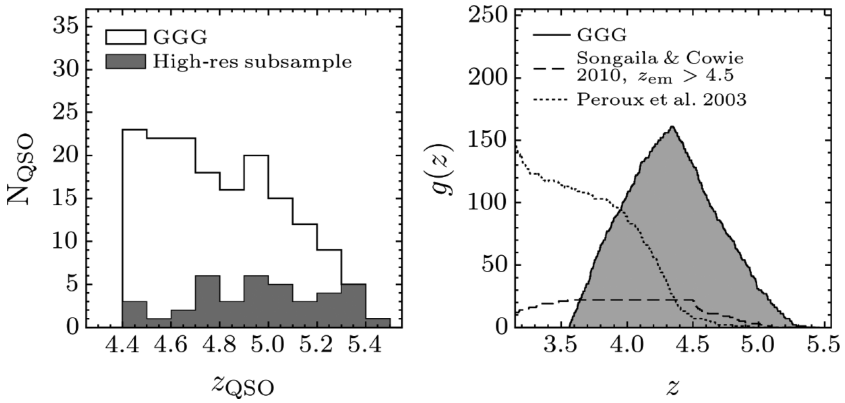

Figure 1. The left-hand panel shows the emission redshift distribution for QSOs in the low-resolution GGG sample (open histogram), and for the subsample of these QSOs targeted with higher resolution spectra. The right-hand panel shows the redshift path, $g(z)$, for detecting DLAs for the GGG sample. $g(z)$ is defined as the number of QSOs where a DLA can be detected as a function of DLA redshift. For comparison, $g(z)$ for previous high-redshift DLA surveys are also shown: for Péroux et al. (2003) and for $z > 4.5$ QSOs from S10. We do not show $g(z)$ for the SDSS DLA surveys (e.g. Noterdaeme et al. 2012). Their $g(z)$ formally extends to $z > 4$, but Prochaska et al. (2005) warn that this high-redshift sensitivity should be viewed conservatively and Noterdaeme et al. (2012) do not include DLAs with $z > 3.5$ in their statistical sample.

$k(N_{\rm H\,I})$ is the result of at least two effects. First, some systems flagged as DLAs will actually be spurious (false positives), and some real DLAs will be missed (false negatives). We estimate $k(N_{\rm H\,I})$ in the following way. Let $N_{\rm cand}$ be the number of DLA candidates flagged in our QSO survey. $N_{\rm cand,\,true}$ of these candidates will be real DLAs, and the remainder will be spurious. If $N_{\rm true}$ is the true number of DLAs in the spectra, then we can denote the fraction of DLA candidates which are not spurious as $k_{\rm real} = N_{\rm cand,\,true}/N_{\rm cand}$, and the fraction of true DLAs that are correctly identified as $k_{\rm found} = N_{\rm cand,\,true}/N_{\rm true}$. This gives

$$f_{\rm DLA}(N_{\rm H\,I}) = f_{\rm meas}(N_{\rm H\,I})\frac{k_{\rm real}}{k_{\rm found}} \quad (7)$$

and thus $k(N_{\rm H\,I}) = k_{\rm real}/k_{\rm found}$. In the following sections, we describe how we measure $f_{\rm meas}$, and how high-resolution and mock spectra are used to estimate $k_{\rm real}$ and $k_{\rm found}$.

3.1 Other systematic effects contributing to $k(N_{\rm H\,I})$

In measuring $k(N_{\rm H\,I})$, we explicitly take into account the rate of spurious DLAs (false positives) and missed DLAs (false negatives). There are several other systematic effects which could also contribute to $k(N_{\rm H\,I})$, which we discuss here.

The first of these is any uncertainty in the $N_{\rm H\,I}$ measurements. If there are large uncertainties in $N_{\rm H\,I}$, or systematic offsets in the $N_{\rm H\,I}$ estimated from the spectra as a function of $N_{\rm H\,I}$, this may change the inferred $f(N_{\rm H\,I})$. However, in Section 4.3 we show that the $N_{\rm H\,I}$ error from the GMOS spectra (0.2 dex) does not have a detectable systematic bias, and Section 5 shows that any errors it introduces to $\Omega_{\rm H\,I}$ are negligible compared to other uncertainties. A related effect is for $N_{\rm H\,I}$ measurements at the DLA threshold of $N_{\rm H\,I}$=$10^{20.3}$ cm^{-2}, where the more numerous lower column density systems may be counted as DLAs through $N_{\rm H\,I}$ uncertainties. This bias is a net source of false positives, and so should be taken into account by our procedure for estimating $k_{\rm real}$.

A second possibility is the presence of dust in DLAs. If DLAs contain large amounts of dust, they are able to extinguish the light from a background QSO, removing these sightlines from our survey. In this case, we would measure a lower incidence of high-metallicity, high-$N_{\rm H\,I}$ DLAs, which presumably contain the most dust. However, several studies have shown that most DLAs are not associated with significant amounts of dust (e.g. Murphy & Liske 2004; Vladilo, Prochaska & Wolfe 2008), and DLAs towards radio-selected QSOs, which are insensitive to the presence of dust, have a similar $N_{\rm H\,I}$ distribution to those in optically selected QSOs (Ellison et al. 2001; Jorgenson et al. 2006). Pontzen & Pettini (2009) find that the cosmic H I mass density may be underestimated by 3–23 per cent at $z \sim 3$ due to selection biases from dust. We do not include this relatively small effect in our analysis, but note where its inclusion would affect our conclusions.

Gravitational lensing may also introduce a bias. DLA host galaxies may lens background QSOs, making them more likely to be found in our survey. This would result in brighter QSOs being more likely to show foreground DLA absorption compared to fainter QSOs. At $z \sim 3$, Murphy & Liske (2004) found evidence at the $\sim 2\sigma$ level that DLAs tend to be found towards brighter QSOs. Prochaska et al. (2005) found a higher incidence rate of high-$N_{\rm H\,I}$ DLAs towards brighter QSOs compared to fainter QSOs over a redshift range 2–4.5, which resulted in a significant (>95 per cent) difference in $\Omega_{\rm H\,I}$ between the two samples. They attributed this effect to gravitational lensing. We confirm that this effect is also present in our sample (which has some overlap with the Prochaska et al. sample): there is a 25 ± 15 per cent higher incidence rate of DLAs towards QSOs with z-band magnitude $\leq$19.2 compared to QSOs with $z > 19.2$ mag. DLAs towards bright QSOs also tend to have high $N_{\rm H\,I}$, resulting in a 30 per cent increase in $\Omega_{\rm H\,I}$ for the brighter compared to the fainter QSO sample. The significance of the excesses we measure is modest (1.7σ), and a Kolmogorov–Smirnov test between the $N_{\rm H\,I}$ distributions towards $z \leq 19.2$ and $z > 19.2$ mag quasars yields $D = 0.3$ and a probability of 22 per cent that the two samples are drawn from the same underlying distribution. Therefore, while this difference hints at a selection effect related to the background QSO brightness, we cannot yet rule out a simple statistical fluctuation. We further discuss how this possible bias may affect our $\Omega_{\rm H\,I}$ measurement in Section 5.1.1.

3.2 Conversion from $\Omega_{\rm H\,I}^{\rm DLA}$ to $\Omega_{\rm H\,I}$

Previous absorption studies have shown that the dominant contribution to $\Omega_{\rm H\,I}$ is from DLAs. Lower column density systems also contribute an appreciable fraction of $\Omega_{\rm H\,I}$, however. This fraction is 15–30 per cent at $z = 3$, depending on the assumed $N_{\rm H\,I}$

distribution (e.g. O'Meara et al. 2007; Noterdaeme et al. 2009; Prochaska, O'Meara & Worseck 2010; Zafar et al. 2013). To parametrize this uncertainty, we introduce a correction factor $\delta_{\rm H\,I} \equiv \Omega_{\rm H\,I}/\Omega_{\rm H\,I}^{\rm DLA}$ to convert between $\Omega_{\rm H\,I}^{\rm DLA}$, which we measure, and $\Omega_{\rm H\,I}$. We assume that the $N_{\rm H\,I}$ distribution at $z > 4$ is not dramatically different from that at $z \sim 3$ and take $\delta = 1.2$, which implies a 20 per cent contribution from lower column density systems. Zafar et al. find that the contribution of sub-damped systems to $\Omega_{\rm H\,I}$ increases with redshift, possibly due to a weakening of the UV background as the number density of QSOs drops at high redshift. Therefore, a goal of future surveys should be to measure the contribution of these sub-damped systems at $z > 4$.

4 METHOD

4.1 Procedure for identifying DLAs

We measure the frequency of DLAs, $f_{\rm meas}$, by identifying DLA candidates by eye in the GMOS spectra, and then correcting for any biases in identification using mock spectra. To identify candidates, we performed the following steps for each QSO spectrum.

(i) Estimate the continuum as a spline, placing the spline knot points by hand. We used the low-z composite QSO spectrum from Shull, Stevans & Danforth (2012) to indicate the position of likely QSO emission lines which fall inside the Lyα forest.

(ii) Look for a possible damped Lyα line in the Lyα forest between the QSO Lyα and Lyβ emission lines. Estimate its redshift and $N_{\rm H\,I}$ by plotting a single-component Voigt profile with $b = 30\,{\rm km\,s^{-1}}$ over the spectrum, and varying $N_{\rm H\,I}$ and z until it matches the data by eye.[2] If necessary the continuum was varied at the same time $N_{\rm H\,I}$ was estimated to obtain a plausible fit. At higher redshifts, blending with IGM absorption can make estimating $N_{\rm H\,I}$ challenging, as the damping wings can be very heavily blended with IGM absorption. In this case, the best constraint on $N_{\rm H\,I}$ is not from the shape of the damping wings, but instead from the extent of the Lyα trough consistent with zero flux and from any higher order Lyman transitions.

(iii) If a candidate DLA is found based on the Lyα profile, use its higher order Lyman series (if available in the spectrum) to refine its redshift and $N_{\rm H\,I}$.

(iv) Repeat steps (ii) and (iii) for all DLA candidates in the Lyα forest.

DLA absorption can also be detected bluewards of the QSO Lyβ emission line. However, we chose to search only between Lyα and Lyβ emission in our sample to maximize the chance of having useful Lyman series lines in addition to Lyα, and to avoid any additional systematic effects caused by further blending with the Lyβ forest. While most DLAs also have associated metal lines detected by the GMOS spectra, we did not use any metal line information when measuring the DLA candidate redshift or $N_{\rm H\,I}$. This was done to avoid any bias against finding low-metallicity systems, which may not have detectable metals in the GMOS spectra.

Two of the authors (NHMC and JXP) searched the spectra for DLAs independently. The above steps were followed either using custom-written PYTHON code or with X_FITDLA from XIDL, depending on which author performed the search. For each QSO, we also noted any properties of the spectrum which might complicate the identification of DLAs, such as the presence of broad absorption lines associated with the background QSO, or of possible problems with the sky background subtraction. Two example DLA candidates are shown in Fig. 2. In these two cases, higher resolution spectra confirm that both candidates are indeed DLAs. The $N_{\rm H\,I}$ and redshift estimated from the GMOS spectra differ slightly from the values inferred from the higher resolution spectra – we discuss this issue further in Section 4.3. Once we assembled a list of DLA candidates, we selected only those within a redshift path limit defined by

$$z_{\rm min} = (1 + z_{\rm qso})\frac{\lambda_{{\rm Ly}\beta}}{\lambda_{{\rm Ly}\alpha}} - 1$$
$$z_{\rm max} = (1 + z_{\rm qso})(1 - \delta v/c) - 1, \quad (8)$$

where $\lambda_{{\rm Ly}\alpha} = 1215.6701\,$Å, $\lambda_{{\rm Ly}\beta} = 1025.72\,$Å and $\delta v = 5000\,{\rm km\,s^{-1}}$. This δv was chosen to exclude 'proximate' DLAs, whose incidence rate is likely affected by a combination of ionizing radiation from the background QSO and the overdensity associated with the QSO host galaxy halo (e.g. Ellison et al. 2002, 2010; Russell, Ellison & Benn 2006; Prochaska, Hennawi & Herbert-Fort 2008). Table 3 lists the redshift path limits used for each QSO in the GGG sample. We then convert the redshift path for each QSO to an absorption distance path using equation (2).

With these DLA candidate lists, we can derive the measured incidence rate of DLAs, $f_{\rm meas}$. However, despite our attempt to take continuum uncertainties and IGM absorption into account when measuring $N_{\rm H\,I}$ for each DLA, large systematic uncertainties may remain. The following sections describe how we quantify these uncertainties using the correction factors $k_{\rm real}$ and $k_{\rm found}$ to $f_{\rm meas}$.

4.2 Estimation of $k_{\rm real}$ and $k_{\rm found}$

We expect $k_{\rm real}$ to be less than unity, meaning that there are some spurious DLA candidates. The rate of these spurious candidates is estimated in two ways. First, we use the sample of higher resolution spectra to identify DLAs, and compare these with the DLA candidates found in the low-resolution sample. Secondly, we create mock low-resolution spectra which closely match the GMOS spectra and contain DLAs generated from a distribution at $z = 3$, and then search these spectra for DLAs in the same way as the real spectra.

$k_{\rm found}$ is also expected to be less than unity, which means some true DLAs exist which we do not flag as DLA candidates in the low-resolution spectra. Again we estimate the fraction of true DLAs recovered in two independent ways, using higher resolution spectra and mocks. In the first case, DLAs identified in the higher resolution QSO spectra were used as a reference list of true DLAs, and compared to the candidate DLAs found in the lower resolution spectra of the same QSOs. In the second case, we used mock GMOS spectra, which allow us to directly compare known DLAs in the spectra to the DLA candidates.

Our motivation for using two different ways to estimate the correction factors (mocks and high-resolution spectra) is to test different systematic effects. The main advantage of the mocks is that the true DLA properties are known precisely. However, while we attempt to reproduce the real spectra as closely as possible, including Lyα forest clustering, QSO redshift and S/N distribution, it is still possible that the mocks may differ from the real GMOS spectra. Metal absorption (not included in the mocks) or clustering of strong absorbers that is different from the mocks may cause more spurious DLAs. Alternatively, non-Gaussian noise in the real spectra at low fluxes may mean that true DLAs are more likely to be missed

[2] This b value was chosen for convenience. The precise b used does not strongly affect the Lyα profile.

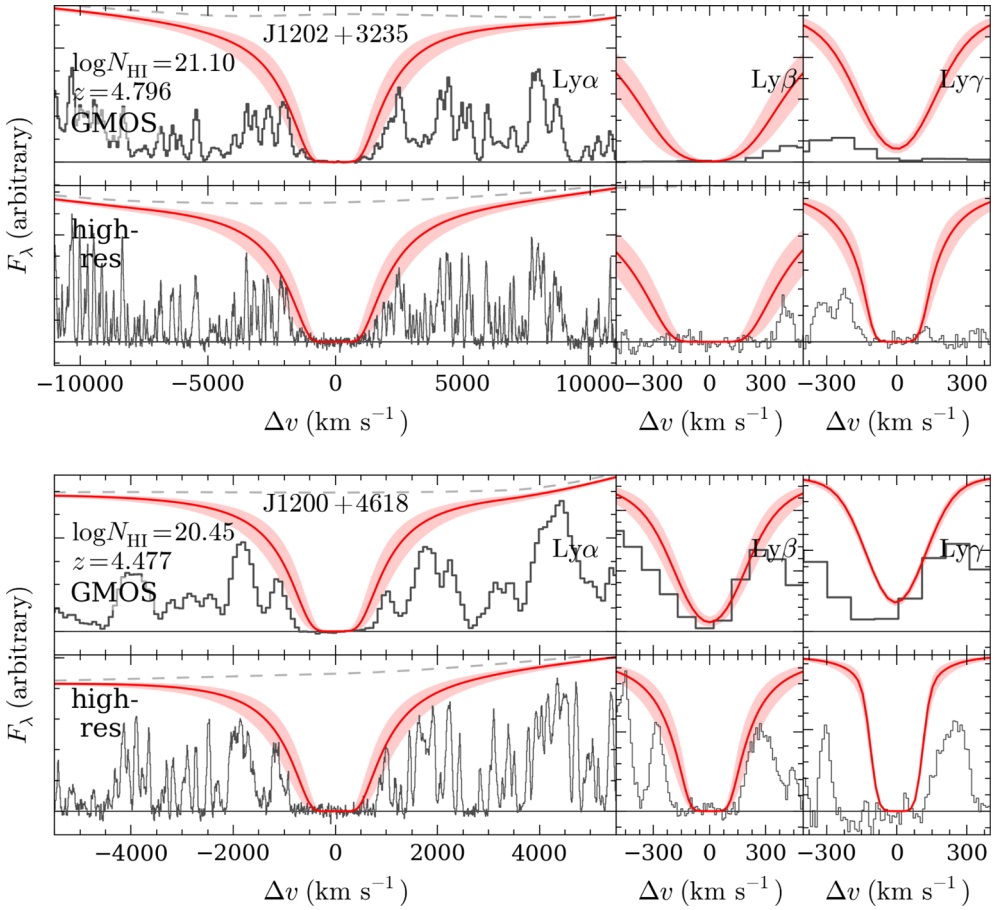

Figure 2. DLAs identified in the GMOS spectra (resolution FWHM ∼230 km s^{-1}) which are confirmed in higher resolution ESI spectra (resolution FWHM ∼30 km s^{-1}). In each case, the top panels show the GMOS spectrum and the bottom panels the ESI spectrum of the same QSO. The model shows the $N_{\rm H\,I}$ and redshift estimated from the ESI spectra with the redshift fixed by low-ion metal absorption. The shaded region shows an uncertainty in log $N_{\rm H\,I}$ of 0.2. The $N_{\rm H\,I}$ and redshift estimated from the GMOS spectra are given in Table 2.

in the real spectra. Conversely, for the high-resolution sample, the true DLA properties are not known with complete certainty, but the correct clustering, IGM blending, noise and metal absorption are all included. Therefore, these two approaches provide complementary estimates of $k_{\rm found}$ and $k_{\rm real}$. The following sections describe these approaches in more detail.

4.2.1 Corrections using high-resolution spectra

DLAs can be found more easily in our sample of high-resolution spectra, and their $N_{\rm H\,I}$ and redshift are more accurately measured, in comparison to the lower resolution GMOS spectra. Therefore, we independently identify DLAs in these spectra for the purpose of deriving the correction factors $k_{\rm real}$ and $k_{\rm found}$, and to test for any systematics in estimating $N_{\rm H\,I}$ and z for each DLA. When identifying the DLAs in the 59 high-resolution spectra, we follow the same process outlined for the lower resolution spectra in Section 4.1, using the Lyman series to estimate the redshift and $N_{\rm H\,I}$. However, we also refine the redshift and $N_{\rm H\,I}$ using the position of low-ionization metal lines (O I, Si II, C II and Al II) where possible. For the 20 QSOs with high-resolution spectra which are not in the GGG sample, we created low-resolution spectra by convolving the high-resolution spectra to the same FWHM resolution, and rebinning to the same pixel size as the GMOS spectra. The same noise array was used for these spectra as for the GGG QSO with a redshift closest to each QSO, normalizing such that the median S/N within rest-frame wavelengths 1260–1280 Å match. These low-resolution spectra were searched for DLAs in the same way as the GMOS spectra.

In this way, we made two lists of DLAs, one from the high-resolution spectra and another from low-resolution spectra of the same QSOs. The DLAs identified in the higher resolution sample are listed in columns 5 and 6 of Table 2. We then estimated $k_{\rm real}$ as $N_{\rm cand,\,true}/N_{\rm cand}$, where $N_{\rm cand}$ is the number of DLA candidates from the low-resolution spectra and $N_{\rm cand,\,true}$ is the number of those candidates confirmed to be DLAs by the high-resolution spectra. $k_{\rm found}$ is estimated as $N_{\rm cand,\,true}/N_{\rm true}$, where $N_{\rm true}$ is the number of DLAs found in the high-resolution spectra and $N_{\rm cand,\,true}$ is the number of those also flagged as DLA candidates in the low-resolution spectra. We calculate the binomial confidence intervals on $k_{\rm real}$ and $k_{\rm found}$ using the method described by Cameron (2011).

With this procedure, we find $k_{\rm real} = 0.80^{+0.07}_{-0.08}$ and $k_{\rm found} = 0.84^{+0.06}_{-0.08}$ using DLAs identified by JXP (see Figs 5 and 6) with similar values found by NHMC. Both are below unity, and so there are both spurious DLA candidates and real DLAs missed. Spurious DLAs usually occur when flux spikes are smoothed away at GMOS resolution, making a lower $N_{\rm H\,I}$ system appear to have strong damping wings. An example spurious DLA is shown in Fig. 3. Real DLAs are generally missed due to flux fluctuations in the core of the Lyα line: an example is shown in Fig. 4.

Table 2. DLAs identified in the GGG spectra and other higher resolution spectra. The first four columns list the redshift and $N_{\rm H\,I}$ estimated in the GMOS-resolution spectra by two of us (JXP and NHMC). The fifth and sixth columns give measurements from a high-resolution spectrum of the QSO, if one exists. The seventh and eight columns give the 'best' estimate of the DLA redshift and $N_{\rm H\,I}$. This is the value from the high-resolution spectrum if one exists, otherwise it is the mean of the estimates from JXP and NHMC. In the first eight columns, an en-dash means no DLA was identified. The ninth column gives the QSO name, and the tenth column lists whether the QSO has a high-resolution spectrum. The last column lists whether the QSO is part of the GGG sample used to measure $\Omega_{\rm H\,I}$.

$z_{\rm JXP}$	$\log_{10}N_{\rm JXP}$	$z_{\rm NC}$	$\log_{10}N_{\rm NC}$	$z_{\rm hires}$	$\log_{10}N_{\rm hires}$	$z_{\rm best}$	$\log_{10}N_{\rm best}$	Label	Hi-res exists?	GGG?
4.740	20.25	4.739	20.40	4.7395	20.30	4.7395	20.30	J0040−0915	y	y
4.187	20.50	–	–	–	–	4.1888	20.60	J0125−1043	n	y
4.887	20.75	4.886	20.70	4.8836	20.50	4.8836	20.50	J0231−0728	y	y
4.658	20.95	4.657	21.00	4.6576	20.75	4.6576	20.75	J0759+1800	y	y
4.098	21.05	4.096	21.05	–	–	4.0985	21.05	J0800+3051	n	y
–	–	4.472	20.40	4.4720	20.30	4.4720	20.30	J0824+1302	y	y
4.830	20.85	4.830	20.90	4.8305	20.75	4.8305	20.75	J0824+1302	y	y
4.341	20.85	4.343	20.90	4.3441	20.60	4.3441	20.60	J0831+4046	y	n
3.713	20.75	3.712	20.95	3.7100	20.75	3.7100	20.75	J0834+2140	y	n
4.391	21.20	4.391	21.30	4.3920	21.15	4.3920	21.15	J0834+2140	y	n
4.424	21.05	4.425	21.02	–	–	4.4227	21.05	J0854+2056	n	y
4.795	20.45	4.794	20.45	–	–	4.7945	20.45	J0913+5919	n	y
3.979	20.35	–	–	–	–	3.9790	20.35	J0941+5947	n	y
–	–	4.862	20.40	–	–	–	–	J0957+0519	y	n
4.473	20.40	4.472	20.55	–	–	–	–	J1004+4347	y	y
–	–	–	–	4.4596	20.75	4.4596	20.75	J1004+4347	y	y
4.798	20.55	4.805	20.50	4.7979	20.60	4.7979	20.60	J1013+4240	y	n
4.257	20.70	4.259	20.30	–	–	4.2580	20.50	J1023+6335	n	y
4.087	20.70	4.086	20.90	4.0861	20.75	4.0861	20.75	J1042+3107	y	n
–	–	–	–	4.8165	20.70	4.8165	20.70	J1054+1633	y	y
–	–	–	–	4.8233	20.50	4.8233	20.50	J1054+1633	y	y
4.429	20.85	–	–	–	–	–	–	J1100+1122	y	y
4.397	21.60	4.395	21.55	4.3954	21.65	4.3954	21.65	J1100+1122	y	y
4.346	21.40	4.347	21.35	4.3441	21.35	4.3441	21.35	J1101+0531	y	y
4.380	21.20	–	–	4.3801	21.15	4.3801	21.15	J1132+1209	y	y
5.015	20.75	5.015	20.60	5.0165	20.70	5.0165	20.70	J1132+1209	y	y
4.476	20.60	4.476	20.65	4.4767	20.45	4.4767	20.45	J1200+4618	y	y
3.799	21.35	3.807	21.20	3.7961	21.25	3.7961	21.25	J1201+2117	y	y
4.156	20.60	–	–	4.1579	20.50	4.1579	20.50	J1201+2117	y	y
4.793	20.75	4.798	20.75	4.7956	21.10	4.7956	21.10	J1202+3235	y	y
4.811	20.75	–	–	4.8106	20.75	4.8106	20.75	J1221+4445	y	y
4.926	20.35	4.931	20.70	4.9311	20.55	4.9311	20.55	J1221+4445	y	y
4.711	20.50	–	–	–	–	–	–	J1233+0622	y	y
4.448	20.80	4.447	20.70	4.4467	20.45	4.4467	20.45	J1245+3822	y	y
4.213	20.50	4.213	20.40	–	–	4.2130	20.45	J1301+2210	n	y
3.937	21.10	3.937	21.10	–	–	3.9387	21.10	J1309+1657	n	y
4.303	20.55	4.303	20.50	–	–	4.3027	20.52	J1332+4651	n	y
–	–	–	–	4.7636	20.35	4.7636	20.35	J1334+1220	y	y
4.348	20.55	4.348	20.50	–	–	4.3480	20.52	J1337+4155	n	y
5.003	20.85	–	–	–	–	–	–	J1340+2813	y	y
–	–	5.096	20.30	–	–	–	–	J1340+2813	y	y
4.826	21.05	4.827	21.05	4.8258	21.20	4.8258	21.20	J1340+3926	y	y
4.109	20.35	–	–	4.1093	20.35	4.1093	20.35	J1412+0624	y	y
–	–	4.322	20.40	–	–	–	–	J1418+3142	y	n
3.958	20.55	–	–	–	–	3.9628	21.00	J1418+3142	y	n
4.453	20.35	4.453	20.45	–	–	–	–	J1418+3142	y	n
4.114	20.60	4.112	20.70	–	–	4.1140	20.65	J1420+6155	n	y
–	–	4.665	20.35	4.6644	20.30	4.6644	20.30	J1421+3433	y	n
4.093	20.30	–	–	–	–	4.0929	20.30	J1427+3308	n	y
4.526	20.60	4.527	20.60	–	–	4.5218	20.60	J1436+2132	n	y
4.800	21.10	4.801	21.10	4.8007	21.20	4.8007	21.20	J1437+2323	y	y
4.400	20.80	4.398	20.85	4.3989	20.80	4.3989	20.80	J1438+4314	y	y
–	–	4.355	20.35	–	–	–	–	J1443+0605	y	n
4.223	20.95	4.222	21.05	4.2237	20.95	4.2237	20.95	J1443+2724	y	y
4.088	21.45	4.089	21.57	–	–	4.0885	21.51	J1511+0408	n	y
4.304	21.05	4.305	21.10	–	–	4.3043	21.08	J1524+1344	n	y
3.818	20.45	3.817	20.30	–	–	3.8175	20.38	J1532+2237	n	y
–	–	4.466	20.30	4.4740	20.40	4.4740	20.40	J1607+1604	y	y

Table 2 – *continued*

$z_{\rm JXP}$	$\log_{10}N_{\rm JXP}$	$z_{\rm NC}$	$\log_{10}N_{\rm NC}$	$z_{\rm hires}$	$\log_{10}N_{\rm hires}$	$z_{\rm best}$	$\log_{10}N_{\rm best}$	Label	Hi-res exists?	GGG?
4.915	20.90	4.912	21.00	4.9091	21.00	4.9091	21.00	J1614+4640	y	y
4.462	20.70	4.462	20.85	–	–	–	–	J1626+2751	y	y
4.312	21.20	4.313	21.30	4.3105	21.30	4.3105	21.30	J1626+2751	y	y
4.498	20.95	4.498	21.00	4.4973	21.05	4.4973	21.05	J1626+2751	y	y
4.605	20.55	–	–	4.6067	20.55	4.6067	20.55	J1626+2858	y	y
4.083	20.60	4.082	20.50	–	–	4.0825	20.55	J1634+2153	n	y
–	–	4.101	20.60	–	–	–	–	J1654+2227	y	n
4.001	20.60	4.003	20.75	4.0023	20.55	4.0023	20.55	J1654+2227	y	n
4.742	20.70	4.740	20.60	4.7424	20.80	4.7424	20.80	J1737+5828	y	y
–	–	–	–	4.7475	20.55	4.7475	20.55	J2252+1425	y	y
4.257	21.10	4.256	20.80	–	–	–	–	J2312+0100	y	n

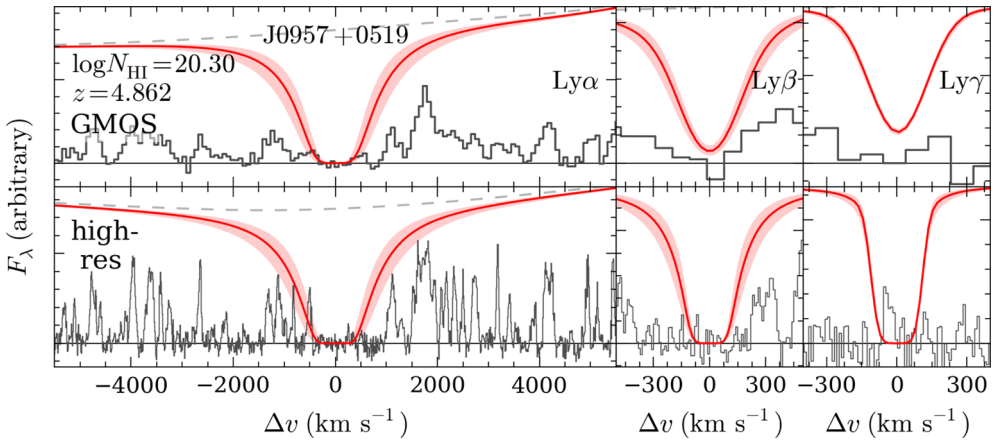

Figure 3. Example of a spurious DLA candidate. This was identified as a DLA with $N_{\rm H\,I} = 10^{20.4\pm0.2}$ cm^{-2} in the GMOS spectrum shown in the top panels. However, the residual flux spikes at Lyα and Lyγ in the higher resolution (FWHM ∼30 km s^{-1}) ESI spectrum in the bottom panels show that this system must have $N_{\rm H\,I} < 10^{20.3}$ cm^{-2}.

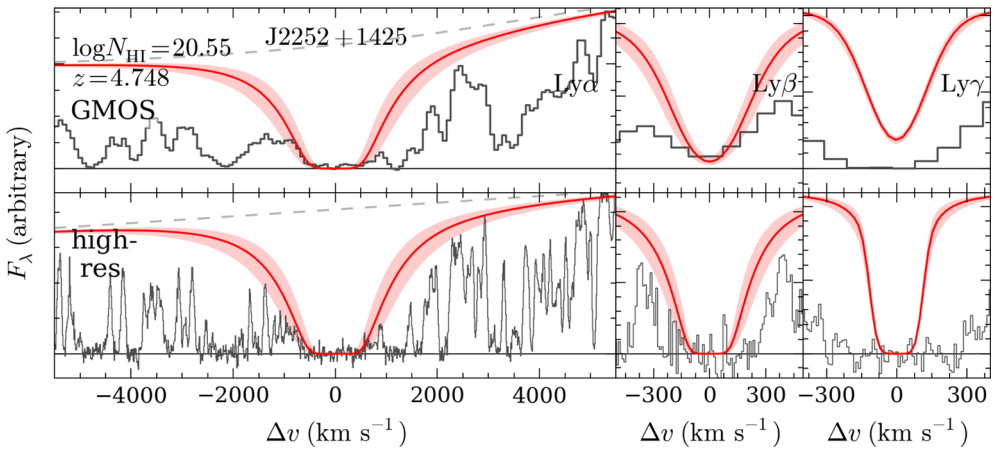

Figure 4. A DLA that was not correctly identified in the GMOS spectra. Lower panels show the DLA in the ESI spectrum, with $N_{\rm H\,I} = 10^{20.55\pm0.2}$ cm^{-2}. The residual flux in the core of the Lyα line in the GMOS spectrum, however (top-left panel), meant this system was missed. This residual flux around velocities $\Delta v \sim 0$ km s^{-1} may be caused by either statistical fluctuations or systematics associated with sky background level.

4.2.2 Corrections using mock spectra

Our method for generating mock spectra is described in Appendix A. In this case, the $N_{\rm H\,I}$ for each DLA is known, and so can be directly compared to the candidates identified in the low-resolution mocks. Again $k_{\rm real}$ is estimated as $N_{\rm cand,\,true}/N_{\rm cand}$, where $N_{\rm cand}$ is the number of DLA candidates from the low-resolution mock spectra and $N_{\rm cand,\,true}$ is the number of those candidates that are DLAs. $k_{\rm found}$ is estimated as $N_{\rm cand,\,true}/N_{\rm true}$, where $N_{\rm true}$ is the true number of DLAs in the mocks and $N_{\rm cand,\,true}$ is the number of those recovered as DLA candidates. Again we calculate the errors on $k_{\rm real}$ and $k_{\rm found}$ assuming a binomial confidence interval. For the mocks, we find $k_{\rm real} = 0.71 \pm 0.06$ and $k_{\rm found} = 0.92^{+0.04}_{-0.07}$ using DLAs identified by JXP (see Figs 5 and 6). Similar values are found by NHMC (see Figs A2 and A3).

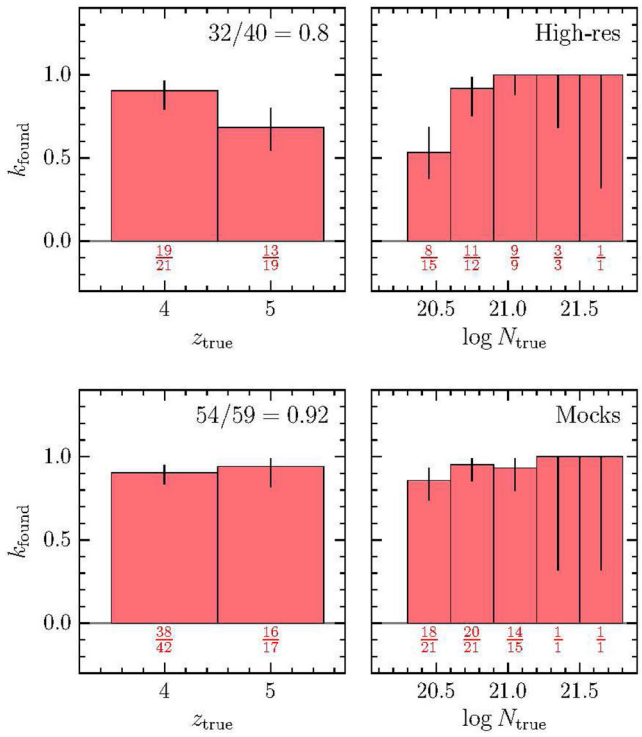

Figure 5. The fraction of true DLAs that were correctly identified by one of the authors (JXP), $k_{\rm found}$, as a function of the true redshift and $N_{\rm H\,I}$. Top panels are for the high-resolution sample, and bottom panels are for mocks. The upper row of numbers under each histogram gives the number of DLA candidates that are correct per bin, and the lower row the total number of candidates. The total numbers for all bins are given at the top in the left-hand panels. Vertical lines show the binomial 68 per cent uncertainties.

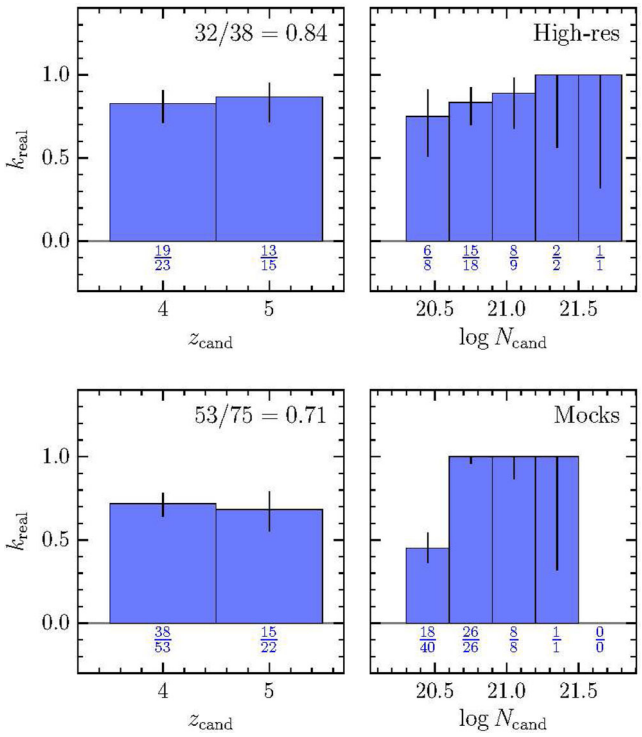

Figure 6. The fraction of non-spurious DLA candidates, $k_{\rm real}$, by one of the authors (JXP), as a function of the candidate redshift and $N_{\rm H\,I}$. Top panels are for the high-resolution sample, and bottom panels are for mocks. The upper row of numbers under each histogram gives the number of true DLAs that are recovered per bin, and the lower row the total number of true DLAs. The total numbers for all bins are given at the top of the left-hand panels. Vertical lines show the binomial 68 per cent uncertainties.

4.2.3 Comparison of correction factors and their dependence on redshift and column density

We expect $k_{\rm real}$ and $k_{\rm found}$ to be a function of a DLA's $N_{\rm H\,I}$ (high-$N_{\rm H\,I}$ candidates should be more reliable), spectral S/N (low-S/N spectra will produce more spurious candidates) and redshift (more spurious DLAs will be found at high redshift where there is more IGM absorption). The most important of these for our measurement of $\Omega_{\rm H\,I}$ is any redshift or $N_{\rm H\,I}$ dependence. Noterdaeme et al. (2009, 2012) show that at $z \sim 2.5$, systems with $N_{\rm H\,I} = 10^{20.6-21.5}$ cm^{-2} make the largest contribution to $\Omega_{\rm H\,I}$. Thus, we expect completeness corrections in this column density range to have the largest effect on the final derived $\Omega_{\rm H\,I}$.[3]

The top panels of Fig. 5 show the correction factor $k_{\rm found}$ from the high-resolution spectra binned by the true DLA redshift and $N_{\rm H\,I}$, and the bottom panels show the same correction factor estimated from the mocks. Fig. 6 shows the correction factor $k_{\rm real}$ binned by the candidate DLA redshift and $N_{\rm H\,I}$, again for the high-resolution spectra and mocks. These are derived from DLAs identified by one of the authors (JXP) who search the spectra for DLAs, but values for the other author (NHMC) are similar. There is no evidence for a strong dependence of $k_{\rm real}$ or $k_{\rm found}$ on redshift, using either the high-resolution spectra or the mocks. However, there is a weak dependence of $k_{\rm real}$ and $k_{\rm found}$ on $N_{\rm H\,I}$, with the lowest $N_{\rm H\,I}$ bin having a significantly lower $k_{\rm real}$ than for higher $N_{\rm H\,I}$ bins. This matches our expectations: weaker candidate DLAs are more likely to be spurious, and true DLAs that are weak are more likely to be missed. We take this $N_{\rm H\,I}$ dependence into account when applying the correction factors as described in Section 5. We find no strong dependence of the correction factors on S/N in either the mocks or the high-resolution sample for the range of S/N the GMOS spectra cover.

Figs 5 and 6 also show that corrections derived from the mocks and high-resolution spectra are in reasonable agreement. The main difference is in the number of spurious systems with $N_{\rm H\,I} \sim 10^{20.3-20.6}$ cm^{-2}. The right-hand panels of Fig. 6 show that there are more weak, spurious DLAs found in the mocks compared to the real GMOS spectra. However, we show in the following section that the correction factor in this $N_{\rm H\,I}$ range is not important for estimating $\Omega_{\rm H\,I}$, and for the remaining bins the mocks and high-resolution corrections match to within 20 per cent. As we discussed earlier, the high-resolution sample and mocks test different systematic uncertainties which may affect $\Omega_{\rm H\,I}$. Therefore, the consistency of the correction factors between these two methods suggests that the mocks reproduce the true GMOS spectra well, and that DLAs have been identified correctly in the higher resolution spectra.

4.3 Uncertainties in $N_{\rm H\,I}$ and redshift

If DLA column densities estimated from the GMOS spectra are systematically in error, our measurement of $\Omega_{\rm H\,I}$ may be biased.

[3] Due to our relatively small DLA sample, we may be missing some very high $N_{\rm H\,I}$ systems with $N_{\rm H\,I} \geq 10^{22}$ cm^{-2}. These contribute only 10 per cent of $\Omega_{\rm H\,I}$ at $z \sim 3$ (Noterdaeme et al. 2012) and thus we do not expect their absence from our sample to strongly bias our results.

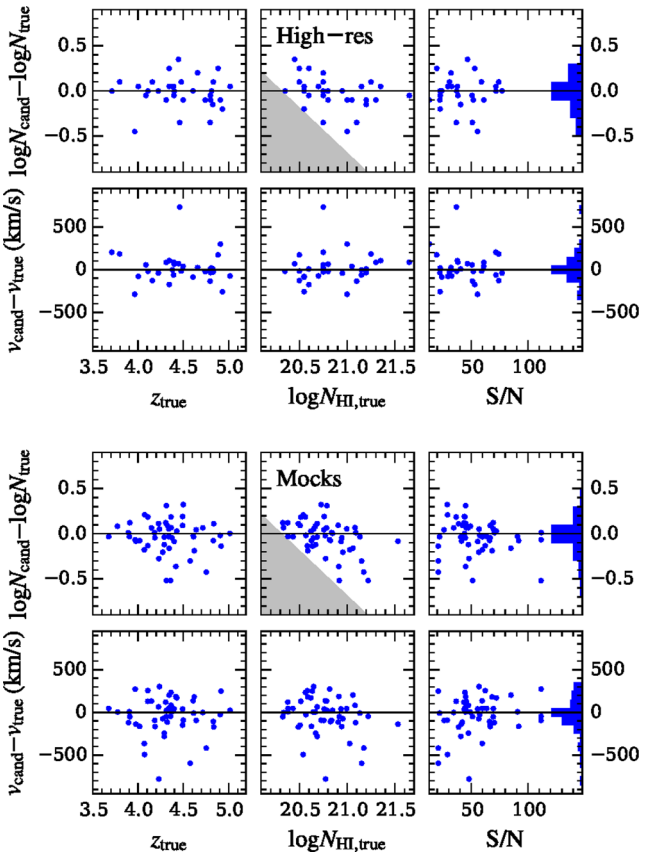

Figure 7. The difference between $N_{\rm H\,I}$ estimated for DLA candidates in low-resolution spectra ($N_{\rm H\,I,cand}$) and $N_{\rm H\,I,true}$ measured from high-resolution spectra (top) or known from mock line lists (bottom), and similarly for the velocity offset from the true DLA redshift. This is for one of the authors (JXP), but the results for NHMC are similar. These show that there is no strong systematic offset in the estimated $N_{\rm H\,I}$ as a function of redshift, S/N or $N_{\rm H\,I}$ which might systematically bias $\Omega_{\rm H\,I}$ significantly. Grey shading shows the regions that cannot be populated due to the requirement that both $N_{\rm H\,I,cand}$ and $N_{\rm H\,I,true}$ be $>10^{20.3}$ cm^{-2}.

Such a systematic could occur because of incorrect placement of the continuum or blending of damping wings with the Lyα forest. This is an additional effect not accounted for by the correction factor, k, to $f_{\rm meas}$. Therefore, we search for any systematic offset in $N_{\rm H\,I}$ by matching DLA candidates from the low-resolution spectra to known DLAs in the high-resolution sample and mocks.

The results of this test are shown in Fig. 7. The log $N_{\rm H\,I}$ difference is plotted as a function of redshift, the true $N_{\rm H\,I}$ and S/N for the high-resolution sample (top panels) and mocks (bottom panels). For both the mocks and high-resolution samples, both $\Delta \log N_{\rm H\,I}$ and Δv are centred on 0. The standard deviation of the velocity and log $N_{\rm H\,I}$ offsets are 184/216 km s^{-1} and 0.165/0.196 for the high-resolution sample and mocks, respectively. We therefore adopt 0.2 dex as our uncertainty in $N_{\rm H\,I}$. There is no trend seen with redshift or S/N. There may be a trend with $N_{\rm H\,I}$, but above $N = 10^{20.3}$ cm^{-2}, it is too weak to significantly affect $\Omega_{\rm H\,I}$. We conclude that there is no systematic bias in $N_{\rm H\,I}$ which might adversely affect the $\Omega_{\rm H\,I}$ measurement.

Fig. 7 also shows the redshift difference between matched DLAs expressed as a velocity difference. DLAs identified in the higher resolution spectra use low-ionization metal lines to set a precise DLA redshift with an error of a few km s^{-1}. Both the mocks and high-resolution sample show that an uncertainty of $\sim$200 km s^{-1}

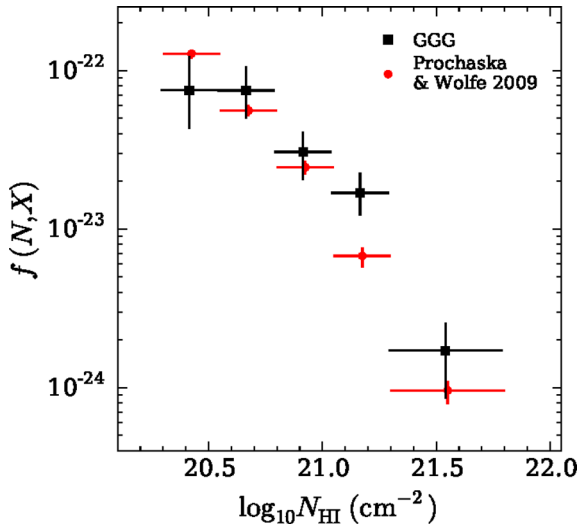

Figure 8. The column density distribution, $f(N_{\rm H\,I})$, for the GGG sample of DLAs. Black points show the measurements with correction $k(N_{\rm H\,I})$ applied. Both are slightly offset in $N_{\rm H\,I}$ for clarity. Red points show the SDSS DR5 measurements from Prochaska & Wolfe (2009) after applying the correction to the redshift search path recommended by Noterdaeme et al. (2009). The errors are 1σ, and include statistical and systematic errors (see Section 4.4 for more details).

results from estimating redshifts using Lyman series absorption alone (without reference to metal absorption) in the low-resolution spectra.

4.4 DLA incidence rate and differential $N_{\rm H\,I}$ distribution

Fig. 8 shows the differential $N_{\rm H\,I}$ distribution from the GGG sample compared to that from the SDSS sample from Prochaska & Wolfe (2009), which is consistent with the more recent estimate from Noterdaeme et al. (2012). We have four different measurements of the correction factor $k(N_{\rm H\,I})$, from two different authors using the mocks and high-resolution spectra, so there are four different estimates of $f(N_{\rm H\,I}, X)$. We find the final $f(N_{\rm H\,I}, X)$ by averaging these four estimates. The uncertainties on this value include a statistical and systematic component. The statistical uncertainty is found by bootstrap resampling, using 1000 samples from the observed DLA distribution, and averaging these uncertainties for the four different estimates. The systematic uncertainty is then assumed to be the standard deviation in the four estimates. These systematic and statistical components are added in quadrature to give the errors shown in Fig. 8. The two distributions are similar overall, although there is a clear discrepancy between the GGG and $z = 3$ $f(N_{\rm H\,I}, X)$ for the bin at log $N_{\rm H\,I} \sim 21.2$, which hints at evolution in the shape of $f(N_{\rm H\,I}, X)$ at high redshift. However, a simple change in the normalization is also consistent with the data.

The DLA incidence rate, $\ell(X)$, is shown in Fig. 9. This observable is more sensitive to the lowest $N_{\rm H\,I}$ DLAs than $\Omega_{\rm H\,I}$. Since the correction factors we derive are strongest for low-$N_{\rm H\,I}$ DLAs and these DLAs have a strong effect on $\ell(X)$, we expect $\ell(X)$ to be sensitive to the particular choices of correction factors. This is indeed the case – there are systematic differences at least as large as the statistical errors, and they depend on whether the mocks or the high-resolution spectra are used to estimate the correction factor. Similarly, large differences are found between $\ell(X)$ by each of the two authors who searched for DLAs. The $\ell(X)$ values we measure are consistent with a smooth increase from $z = 2$ to 5. However,

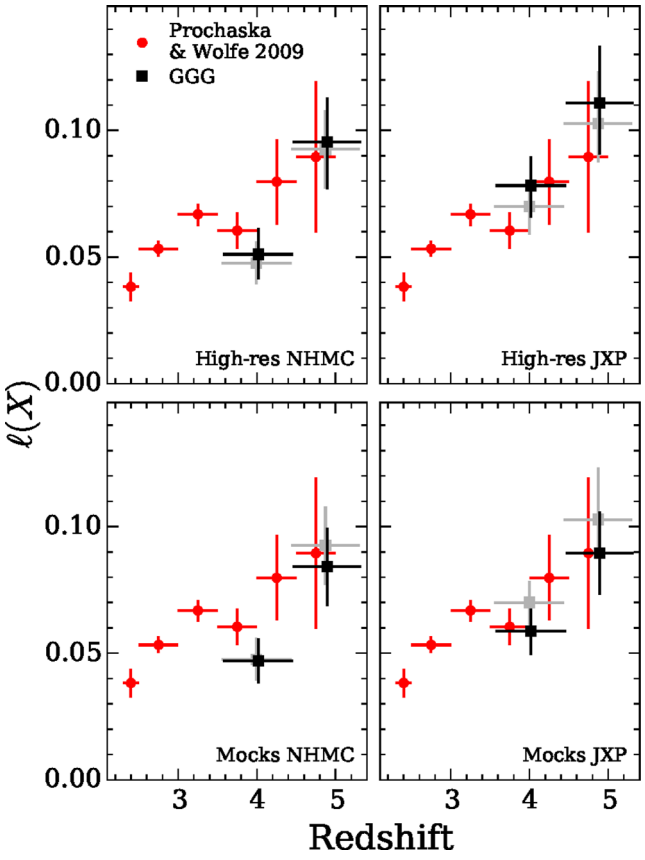

Figure 9. The DLA incidence rate $\ell(X)$. This observable is more sensitive to the lowest $N_{\rm H\,I}$ DLAs than $\Omega_{\rm H\,I}$. Grey points show the uncorrected GGG measurement, and black squares with the corrections applied. Each panel shows a different correction, using either mock or high-resolution spectra for two different authors. There are systematic differences comparable to the statistical errors, and they depend on whether the mocks or the high-resolution spectra are used to estimate correction factors. Similar differences are also found between the two different authors who searched for DLAs. These illustrate that significant systematic uncertainties affect the measurement of $\ell(X)$.

since we do not know which $k(N_{\rm H\,I})$ correction factors are best, we do not attempt to present a definitive $\ell(X)$ measurement here. A large sample of higher resolution spectra, where low column density DLAs can be identified with more certainty, will be necessary to robustly measure $\ell(X)$ at $z > 4$.

We can still make a more robust measurement of $\Omega_{\rm H\,I}$, however, regardless of the uncertainty in $\ell(X)$, as Fig. 10 illustrates. DLAs with the largest contribution to $\Omega_{\rm H\,I}$ have $N_{\rm H\,I}$ in the range $10^{20.8}$–$10^{21.6}$ cm^{-2}, and DLAs with lower $N_{\rm H\,I}$ make a substantially smaller contribution. Therefore, while systematic effects may give rise to a large uncertainty in the number of low column density systems (and thus $\ell(X)$), $\Omega_{\rm H\,I}$ can still be measured accurately. This point is discussed further in Section 5.

5 RESULTS AND DISCUSSION

5.1 $\Omega_{\rm H\,I}$ measurement

We can now use the $N_{\rm H\,I}$-dependent correction factor k estimated in the previous section to find $f_{\rm DLA}$ and thus $\Omega_{\rm H\,I}$. For the GGG sample, we count the number of DLAs in a given absorption path, giving each DLA a weight $k(N_{\rm H\,I})$, where $k(N_{\rm H\,I}) = k_{\rm found}(N_{\rm H\,I})/k_{\rm real}(N_{\rm H\,I})$.

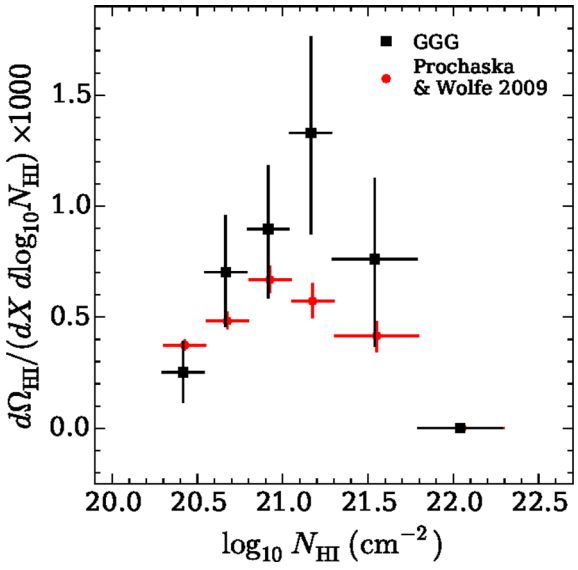

Figure 10. The differential $\Omega_{\rm H\,I}$ distribution, $\mathrm{d}\Omega_{\rm H\,I}/(\mathrm{d}X\,\mathrm{d}\log N_{\rm H\,I})$. DLAs with the largest contribution to $\Omega_{\rm H\,I}$ have $N_{\rm H\,I}$ in the range $10^{20.8}$–$10^{21.6}$ cm^{-2}; DLAs with lower $N_{\rm H\,I}$ are less important. Therefore, while there is uncertainty in the number of low column density systems (and thus $\ell(X)$), $\Omega_{\rm H\,I}$ can still be measured accurately. This is also illustrated by Fig. 11.

$k(N_{\rm H\,I})$ is then estimated as the ratio of the $\log N_{\rm H\,I}$ histograms shown in Figs 5 and 6, with the uncertainty on each bin given by the uncertainties in $k_{\rm found}$ and $k_{\rm real}$ added in quadrature.

There are two main contributions to the final error on $\Omega_{\rm H\,I}$. The dominant contribution is the statistical error due to the finite sampling of DLAs: there are 25–30 DLA candidates in each redshift bin, dependent on whether NHMC or JXP's results are used. We estimate this error using 1000 bootstrap samples from the DLA sample. The second is the systematic uncertainty in the correction factor, $k(N_{\rm H\,I})$. We estimate the effect of this uncertainty using a Monte Carlo technique. $\Omega_{\rm H\,I}$ is calculated 1000 times, each time drawing $k(N_{\rm H\,I})$ from a normal distribution with a mean given by the $k(N_{\rm H\,I})$ histogram bin value and σ determined by the uncertainty on that bin, assuming no correlation between uncertainties in adjacent bins. Then the final error in $\Omega_{\rm H\,I}$ is given by adding these two uncertainties in quadrature. We confirmed that $N_{\rm H\,I}$ error of each DLA (0.2 dex, see Section 4.3) has a negligible contribution compared to these statistical and systematic uncertainties. We also check that using $N_{\rm H\,I}$ measurements from the high-resolution spectra, where available, does not significantly change $\Omega_{\rm H\,I}$.

Since we have separate estimates of $k(N_{\rm H\,I})$ from the mocks and high-resolution sample, and two authors performed these estimates, we can make four different measurements of $\Omega_{\rm H\,I}$. We use these to gauge the effect on $\Omega_{\rm H\,I}$ of estimating corrections from the mocks versus the high-resolution sample, or of any differences in the way the two authors identified DLAs. The results are shown in Fig. 11. The differences between the mocks compared to the high-resolution sample, and between the two authors, are significantly smaller than the uncertainty on any individual $\Omega_{\rm H\,I}$ measurement. Therefore, we conclude that neither the methods we use to estimate $k(N_{\rm H\,I})$ nor any differences in DLA detection between methods contribute a significant uncertainty to the final $\Omega_{\rm H\,I}$. We caution that this conclusion only holds for the sample of spectra we analyse. New tests of systematic effects may be required for measurements of $\Omega_{\rm H\,I}$ using larger samples of DLAs, or using different resolution or S/N QSO spectra.

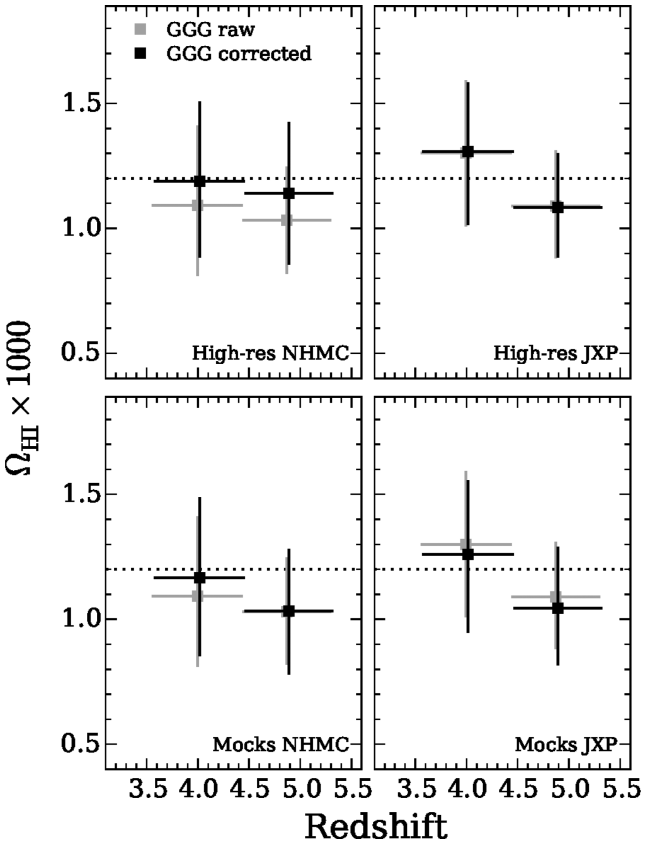

Figure 11. $\Omega_{\rm H\,I}$ measured by the two authors using the high-resolution sample (top) and mocks (bottom). The uncertainties on $\Omega_{\rm H\,I}$ introduced by any differences in selecting DLA candidates between the authors, or between using mocks or the high-resolution sample, are much smaller than the errors shown, which are a combination of the statistical error and uncertainty in the correction factor (see Section 5).

For the remainder of the paper, we use the measurement of $\Omega_{\rm H\,I}$ derived using k from the higher resolution sample and measured by author JXP, which is shown in the top-right panel of Fig. 11. This measurement and the 68 per cent confidence interval are given in Table 4. We assume a 20 per cent contribution to $\Omega_{\rm H\,I}$ from systems below the DLA threshold, as described in Section3.2.

5.1.1 Is there a bias from gravitational lensing?

There is a 30 ± 20 per cent increase in $\Omega_{\rm H\,I}$ for sightlines towards the brighter half of our QSO sample ($z \leq 19.2$ mag) relative to $\Omega_{\rm H\,I}$ towards the fainter QSOs ($z > 19.2$ mag). If this effect is caused by gravitational lensing of a background QSO by a galaxy associated with a foreground DLA, then our measured $\Omega_{\rm H\,I}$ will be artificially enhanced. A detailed lensing analysis is beyond the scope of this work. However, if we follow Ménard & Fukugita (2012) and assume that the lensing DLA galaxies are isothermal spheres, we can estimate their Einstein radius as

$$\zeta_0 = 4\pi \left(\frac{\sigma_{\rm v}}{c}\right)^2 \frac{D_{\rm l} D_{\rm ls}}{D_{\rm s}} \tag{9}$$

where $\sigma_{\rm v}$ is the velocity dispersion, c is the speed of light and $D_{\rm l,\,s,\,ls}$ are the angular diameter distances from the observer to the lens and to the source, and from the lens to the source. Assuming a typical dispersion of 100 km s^{-1}, we find that the effective radius

Table 3. The start and end redshifts for each QSO used to calculate the redshift search path for DLAs. This is a stub; the full table is available online.

Name	$z_{\rm min}$	$z_{\rm max}$	$z_{\rm qso}$
SDSS J001115.23+144601.8	4.037	4.870	4.970
SDSS J004054.65−091526.8	4.046	4.880	4.980
SDSS J010619.24+004823.3	3.598	4.358	4.449
SDSS J012509.42−104300.8	3.639	4.406	4.498
SDSS J021043.16−001818.4	3.868	4.674	4.770
SDSS J023137.65−072854.4	4.417	5.313	5.420
SDSS J033119.66−074143.1	3.838	4.638	4.734
SDSS J033829.30+002156.2	4.096	4.939	5.040
SDSS J073103.12+445949.4	4.061	4.898	4.998
SDSS J075907.57+180054.7	3.911	4.723	4.820
SDSS J080023.01+305101.1	3.789	4.581	4.676
SDSS J080715.11+132805.1	3.961	4.782	4.880
SDSS J081806.87+071920.2	3.746	4.531	4.625
SDSS J082212.34+160436.9	3.649	4.418	4.510
SDSS J082454.02+130217.0	4.237	5.103	5.207

Table 4. $\Omega_{\rm H\,I}$ for the GGG sample, assuming a flat cosmology with $H_0 = 70$ km s^{-1} Mpc^{-1} and $\Omega_{\rm m,\,0} = 0.3$. The redshift bins were chosen to cover roughly equal redshift widths, and to yield approximately equal numbers of DLAs in each bin. To convert between $\Omega_{\rm H\,I}$ and $\Omega_{\rm g}^{\rm DLA}$, which is often quoted by other DLA studies, we use $\Omega_{\rm H\,I} = \delta_{\rm H\,I}\Omega_{\rm g}^{\rm DLA}/\mu$, where $\mu = 1.3$ accounts for the mass of helium and $\delta_{\rm H\,I} = 1.2$ estimates the contribution from systems below the DLA threshold of $10^{20.3}$ cm^{-2}.

z	$10^3\,\Omega_{\rm H\,I}$	$10^3\,\Omega_{\rm H\,I}\,(1\sigma)$	ΔX
3.56–4.45	1.18	0.92–1.44	356.9
4.45–5.31	0.98	0.80–1.18	194.6

Table 5. $\Omega_{\rm H\,I}$ measurements from the literature shown in Fig. 12. Each has been converted to a flat cosmology with $H_0 = 70$ km s^{-1} Mpc^{-1} and $\Omega_{\rm m,\,0} = 0.3$, and represent the mass density from H I gas alone, without any contribution from helium or molecules. For previous analyses which quote the gas mass in DLAs, $\Omega_{\rm g}^{\rm DLA}$, we have converted to $\Omega_{\rm H\,I}$ using $\Omega_{\rm H\,I} = \delta_{\rm H\,I}\Omega_{\rm g}^{\rm DLA}/\mu$, where $\mu = 1.3$ accounts for the mass of helium and $\delta_{\rm H\,I} = 1.2$ estimates the contribution from systems below the DLA threshold of $10^{20.3}$ cm^{-2} (see Section 3.2).

z	$10^3\,\Omega_{\rm H\,I}$	Reference
0	0.375 ± 0.061	Zwaan et al. (2005)
0	0.548 ± 0.091	Braun (2012)
0.026	0.430 ± 0.030	Martin et al. (2010)
0.028	$0.403^{+0.043}_{-0.084}$	Delhaize et al. (2013)
0.096	$0.456^{+0.061}_{-0.084}$	Delhaize et al. (2013)
0.1	0.33 ± 0.05	Rhee et al. (2013)
0.2	0.34 ± 0.09	Rhee et al. (2013)
0.24	0.70 ± 0.31	Lah et al. (2007)
0.15–0.90	$0.88^{+0.36}_{-0.33}$	Rao et al. (2006)
0.9–1.6	$0.86^{+0.30}_{-0.27}$	Rao et al. (2006)
2.0–2.3	0.872 ± 0.044	Noterdaeme et al. (2012)
2.3–2.6	0.765 ± 0.035	Noterdaeme et al. (2012)
2.6–2.9	0.914 ± 0.044	Noterdaeme et al. (2012)
2.9–3.2	0.966 ± 0.070	Noterdaeme et al. (2012)
3.2–3.5	1.11 ± 0.11	Noterdaeme et al. (2012)
3.5–4.3	$0.82^{+0.30}_{-0.27}$	Songaila & Cowie (2010)
4.3–5.1	$0.77^{+0.30}_{-0.27}$	Songaila & Cowie (2010)

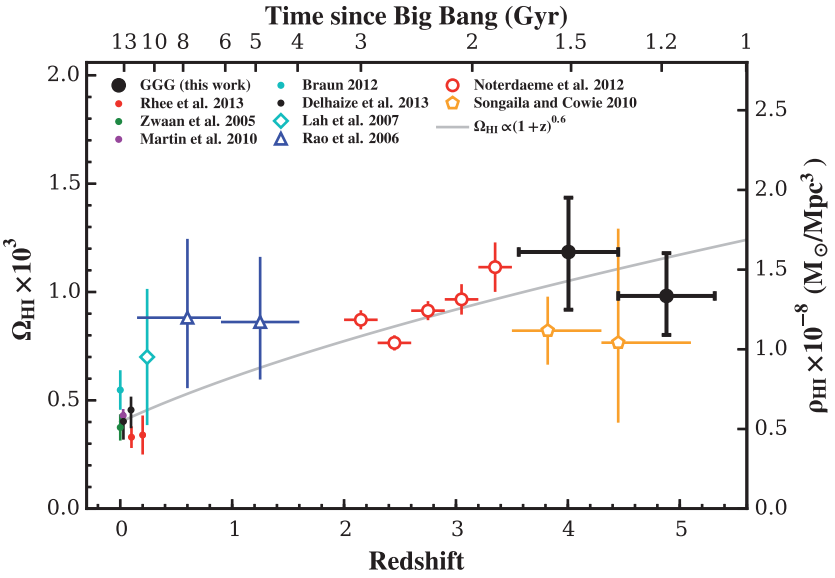

Figure 12. Measurements of $\Omega_{\rm H\,I}$ at different redshift, from Zwaan et al. (2005), Rao, Turnshek & Nestor (2006), Lah et al. (2007), Braun (2012), Martin et al. (2010), Noterdaeme et al. (2012), Rhee et al. (2013), Delhaize et al. (2013) and S10 (see Table 5). We do not show the measurement using SDSS QSOs by Prochaska & Wolfe (2009); it is consistent with the measurement by Noterdaeme et al. (2012), who use a superset of SDSS QSOs. We also do not show the Péroux et al. (2003) and Guimarães et al. (2009) results, which have a large overlap with the QSO sample used by S10 and are consistent with that measurement. Finally, for clarity, we do not show the measurements at lower redshift from Freudling et al. (2011) and Meiring et al. (2011); they are consistent with the plotted values. All measurements have been converted to the same cosmology ($h = 0.7$, $\Omega_{\rm m} = 0.3$, $\Omega_\Lambda = 0.7$) and include H I mass only, with no contribution from helium or molecular hydrogen.

for lensing is very small, 0.1 kpc for a $z = 4.5$ DLA towards a $z = 5$ QSO. This is half the radius for a DLA at $z = 2.5$ towards a QSO at $z = 3.5$. Since the magnitude of the increase in $\Omega_{\rm H\,I}$ due to the putative lensing at $z \sim 3$ is relatively small ($\sim$20 $\pm$ 10 per cent; Prochaska et al. 2005), we do not expect it to have a large effect at higher redshifts. We conclude that it is more likely that the difference in $\Omega_{\rm H\,I}$ between the bright and faint QSO samples is caused by a statistical fluctuation, rather than a lensing bias.

5.2 Comparison with previous measurements

Several groups have made measurements of $\Omega_{\rm H\,I}$ at $z > 4.5$ using DLA surveys (Péroux et al. 2003; Guimarães et al. 2009; S10). These are cumulative results – $\Omega_{\rm H\,I}$ measurements from each new QSO sample are combined with older $\Omega_{\rm H\,I}$ measurements which used a different DLA survey. While combining results in this way maximizes the statistical S/N of the final result, it results in a heterogeneous sample of quasar spectra with different data quality and different DLA identification methods. As shown in Sections 4 and 5.1, at $z > 4.4$ different identification methods can produce a systematic uncertainty in $\Omega_{\rm H\,I}$ which, although smaller than the statistical uncertainties for our current DLA sample, may still be considerable. Since these analyses did not use mock spectra to explore systematic effects, it is difficult to estimate the true uncertainty in $\Omega_{\rm H\,I}$ when combining heterogeneous quasar samples with different selection criteria. In contrast, our sample has homogeneous data quality, QSO selection method and DLA identification procedure, and we use mock spectra to test any systematic effects.[4]

[4] We note that eight of the QSOs used by S10 are also included in our sample, but the 155 remaining GGG QSOs are independent of previous samples.

Fig. 12 shows our new results together with previous measurements of $\Omega_{\rm H\,I}$, converted to our adopted cosmology. When multiple measurements of $\Omega_{\rm H\,I}$ have been made using overlapping QSO samples and the most recent measurement uses a superset of previous QSO samples, only the most recent measurement is shown. For example, the results of S10 include most of the quasars used by Péroux et al. (2003) and Guimarães et al. (2009), so we show only the S10 result. In all such cases, the most recent measurement is consistent with earlier results. Where previous DLA surveys have quoted $\Omega_{\rm H\,I}^{\rm DLA}$, we convert to $\Omega_{\rm H\,I}$ using the relationship $\Omega_{\rm H\,I} = 1.2\Omega_{\rm H\,I}^{\rm DLA}/1.3$. Our measurement at $\langle z \rangle = 4$ is higher than, but consistent with, earlier measurements by S10. As such and because we find a possible systematic increase in $\Omega_{\rm H\,I}$ towards bright QSOs, we checked whether the magnitude distribution of the S10 QSOs was lower than the GGG sample. z-band data were not available for the whole S10 sample, but the eight QSOs which overlap between their sample and ours have a similar fraction of QSOs with $z \leq 19.2$ and $z > 19.2$ mag. Therefore, a difference in QSO magnitudes is unlikely to cause a difference between our result and the S10 result, and it seems more likely that the difference is caused by a statistical fluctuation.

Our results at $\langle z \rangle = 4.9$ give the most robust indication to date that there is no strong evolution in $\Omega_{\rm H\,I}$ over the $\sim$1 Gyr period from $z = 5$ to 3. We see a slight drop in $\Omega_{\rm H\,I}$ between our $z \sim 4$ and $z \sim 4.9$ $\Omega_{\rm H\,I}$ measurements, but this difference is not statistically significant. If the metal content of DLAs does change suddenly at $z = 4.7$, as suggested by Rafelski et al. (2014), there is no evidence that it is accompanied by a concomitant change in $\Omega_{\rm H\,I}$. However, the uncertainties remain large and future observations should continue to test this possibility.

Fig. 12 also shows a power law with the form $\Omega_{\rm H\,I} = A(1 + z)^\gamma$ fitted to the binned data. This simple function provides a reasonable fit (χ^2 per degree of freedom =1.44) across the full redshift

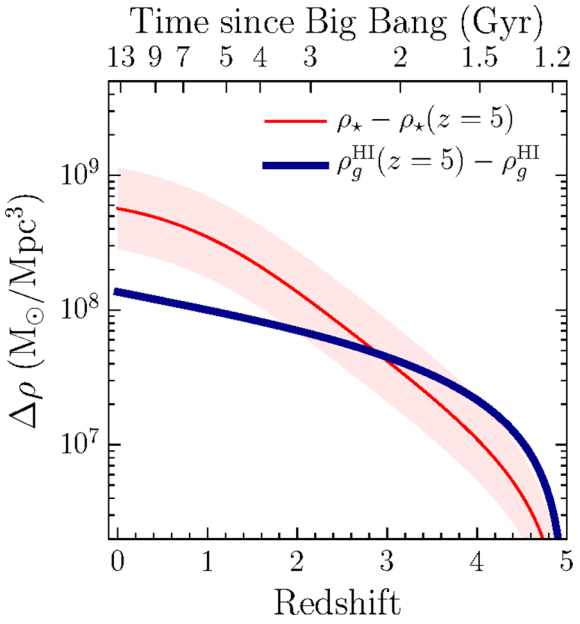

Figure 13. The increase in comoving stellar mass density from $z = 5$ to 0 (from Madau & Dickinson 2014, thin line and shading) and the corresponding decrease in H I gas mass density over the same period (thick line) using the fitting formula from Section 5.2. Before $z \sim 3$, the H I gas phase contains ample mass density to fuel all the observed star formation. However, from $z \sim 3$ to the present, it contributes less than ∼20 per cent of the mass necessary to form stars, and so must be continually replenished by more highly ionized gas.

range, with best-fitting parameters $A = (4.00 \pm 0.24) \times 10^{-4}$ and $\gamma = 0.60 \pm 0.05$. There is no obvious physical motivation for this relation nor any expectation that it should apply at redshifts >5. Nevertheless, it may provide a useful fiducial model to compare to simulations and future observations.

We also compare our new high-redshift value to lower redshift $\Omega_{\rm H\,I}$ measurements. As previous authors have noted (e.g. Prochaska et al. 2005; Noterdaeme et al. 2009; Prochaska & Wolfe 2009), $\Omega_{\rm H\,I}$ evolves from $z = 3$ to 0 by factor of $\lesssim 2$, at odds with the very strong evolution in the star formation rate over the same period. Moreover, the drop in $\Omega_{\rm H\,I}$ is much smaller than the increase in stellar mass over this period. Fig. 13 demonstrates this point by showing the increase in comoving mass density in stars from $z = 5$, $\rho_\star - \rho_\star(z = 5)$ and the contemporaneous decrease in H I comoving gas mass density,[5] $\rho_g^{\rm H\,I}(z = 5) - \rho_g^{\rm H\,I}$, using the power-law fit from Fig. 12. The mass in stars is calculated using the expression from Madau & Dickinson (2014), and the range shows an uncertainty of 50 per cent, indicative of the scatter in observations around this curve. While the evolution of $\Omega_{\rm H\,I}$ from $z = 5$ to 3 remains uncertain, the H I phase at $z = 5$ contains ample mass density to form all the stars observed at $z \sim 3$, and the evolution predicted by the simple power-law function is consistent with this scenario. From $z \sim 3$ to $z \sim 0$, however, there is a factor of 5–6 shortfall in H I mass density compared to amount needed to produce stars over the same period. This underscores that at $z \lesssim 3$, the H I phase must be continually replenished by more highly ionized gas, presumably through a combination of cold-mode accretion (e.g. Dekel et al. 2009) and recycled winds (e.g. Oppenheimer et al. 2010). The more highly ionized Lyman limit systems and sub-DLAs should then be important tracers of the interface between this H I phase and more highly ionized gas (e.g. Fumagalli et al. 2011).

There are several reasons to expect the neutral fraction of the Universe to evolve at $z > 3$. As we approach the epoch of reionization, the filling factor of neutral hydrogen in the Universe should increase, as large pockets of the Universe are no longer ionized. This is reflected in the decrease in the mean free path for H-ionizing photons (Fumagalli et al. 2013; Worseck et al. 2014) towards higher redshifts. While the bulk of reionization is thought to occur at $z > 6$, large neutral regions may persist to lower redshifts (e.g. Becker et al. 2015). Our results suggest that while regions of this kind may exist, they do not change the total neutral gas mass density appreciably from that observed at $z \sim 3$. This is consistent with the conclusions of Becker et al., who find that by $z = 5$ the bulk of IGM absorption is due to density fluctuations instead of large, neutral regions yet to be reionized.

This is perhaps not surprising. The distribution of these neutral pockets depends on the nature of reionization, which may progress from low-density regions to high-density regions ('outside-in') or the reverse ('inside-out'), or some combination of the two (e.g. Finlator et al. 2012). However, favoured scenarios see the highest density regions with $\Delta \equiv \rho/\langle\rho\rangle \gg 100$ reionized first, as they are populated by galaxies, believed to be the dominant source of ionizing photons. In this case, neutral pockets will persist only in underdense regions such as filaments or voids, with $\Delta < 100$. At $z \sim 2.5$ clustering measurements suggest that most DLAs are found inside haloes with masses 10^{10}–10^{12} $M_\odot$ (Cooke et al. 2006; Font-Ribera et al. 2012), which have a mean $\Delta > 100$. Therefore, even if large neutral regions do persist to $z = 5$, they may not occur at cosmic densities high enough to produce strong DLA absorption. The remnants of such regions may be observable as Lyman limit systems however, and so one might expect an increase in their incidence rate towards $z \sim 5$, which observations already hint may be the case (Prochaska et al. 2010; Fumagalli et al. 2013). The GGG sample can also be used to measure the LLS incidence rate at $z > 4$, which we will present in a future work.

5.3 Comparison with theory

In Fig. 14 we show $\Omega_{\rm H\,I}$ in comparison to some recent theoretical predictions for its evolution. These are by Lagos et al. (2014) using the semi-analytic GALFORM model (see also Popping, Somerville & Trager 2014; Berry et al. 2014), by Bird et al. (2014) from a simulation using the moving-mesh code AREPO, and by Tescari et al. (2009, see also Duffy et al. 2012; Rahmati et al. 2015) and Davé et al. (2013) using smoothed particle hydrodynamics (SPH) simulations. While these models broadly match the slow evolution of $\Omega_{\rm H\,I}$ since $z \sim 4$, most struggle to reproduce the trend of decreasing $\Omega_{\rm H\,I}$ with time (with Tescari et al. being a notable exception). Lagos et al. suggest that their model's underestimation of $\Omega_{\rm H\,I}$ at high redshift may be due to more neutral gas being found outside galaxy discs in the early Universe. If this interpretation is correct, then our observations suggest that more than half the neutral gas mass (and more than half of DLAs) are found outside galaxies at $z \sim 5$. Alternatively, Davé et al. (2013) show that agreement between their simulations and observations can be improved by assuming that a population of low-mass galaxies, unresolved by current SPH simulations, make a

[5] In Fig. 13, the H I gas mass density $\rho_g^{\rm H\,I}$ is used, which is related to the H I mass density by $\rho_g^{\rm H\,I} \equiv \mu\rho_{\rm H\,I}$ with $\mu = 1.3$ and $\rho_{\rm H\,I} = \rho_{\rm crit,0}\Omega_{\rm H\,I}$. We do not apply any correction for dust extinction by foreground DLAs. If this is present, it could increase $\rho_g^{\rm H\,I}$ by 20 per cent (Pontzen & Pettini 2009), which would not affect our discussion.

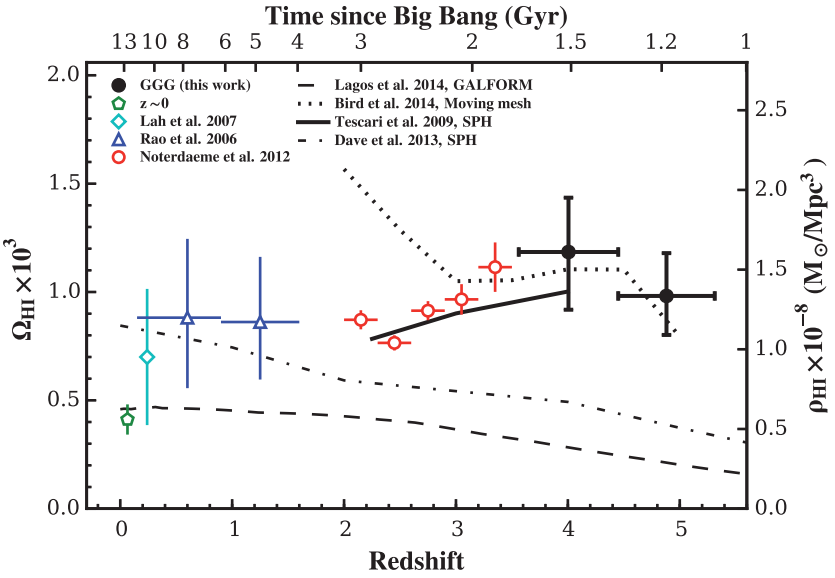

Figure 14. Measurements of $\Omega_{\rm H\,I}$ compared to recent theoretical predictions. For clarity, the mean of measurements at $z < 0.2$ (the error bar shows the standard deviation) is shown. Lines show predictions from a recent semi-analytic model (Lagos et al. 2014), along with SPH (Tescari et al. 2009; Davé et al. 2013) and moving-mesh (Bird et al. 2014) simulations. All the models have been converted to our adopted cosmology. While they do reproduce the roughly flat evolution of $\Omega_{\rm H\,I}$ from $z = 5$ to 0 (in comparison to the cosmic star formation rate density), they do not match the data across the full redshift range.

significant contribution to the DLA absorption cross-section at high redshift.

It is evident that further improvements are needed to theoretical models to reproduce the evolution of $\Omega_{\rm H\,I}$ across the full redshift range. If much of the neutral gas is found in galactic outflows or recycled winds, the sub-grid prescription for outflows in SPH simulations may have a strong influence on the predicted $\Omega_{\rm H\,I}$ (e.g. Bird et al. 2014). Furthermore, given the small sizes of DLAs (∼5 kpc; Cooke et al. 2010), it may also be important to correct for any smoothing over small-scale density peaks where DLAs are produced, and account for hydrodynamic instabilities which are not resolved by current cosmological simulations (e.g. Crighton et al. 2015).

6 SUMMARY

We have measured $\Omega_{\rm H\,I}$ at $3.5 < z < 5.3$ using the GGG survey, a homogeneous sample of 163 QSO spectra with emission redshifts >4.4. All the QSOs were colour-selected from the SDSS and so have a well-understood selection function which is independent of any strong absorption in the QSO spectra. Using a combination of higher resolution spectra of DLA candidates and mock spectra, we explore systematic uncertainties in identifying DLAs due to strong IGM absorption at high redshift and the low spectral resolution of the GMOS spectra. The main conclusions from our analysis are as follows.

(i) We derive the most precise measurement of $\Omega_{\rm H\,I}$ at $\langle z \rangle = 4.9$ to date, with a redshift path length at $z > 4.5$ a factor of 8 larger than previous analyses. $\Omega_{\rm H\,I}$ at $z = 4.5$ is consistent with the value measured at $z = 3$–3.5, and there is no evidence that $\Omega_{\rm H\,I}$ evolves strongly over the Gyr period from redshift 5 to 3. There is also no evidence for an abrupt change in $\Omega_{\rm H\,I}$ between $z = 4$ and 5, which may be associated with a sudden change in metallicity reported at a similar redshift (Rafelski et al. 2014). However, such a change is not strictly ruled out by the data.

(ii) We quantify and correct for the fraction of spurious DLA candidates, and for any DLAs missed in the low-resolution spectra, using higher resolution and mock spectra. We also estimate the uncertainty in the DLA column densities. For this DLA sample, the uncertainty introduced by these systematic effects on the $\Omega_{\rm H\,I}$ measurement is smaller than the statistical uncertainties.

(iii) Using the higher resolution spectra and mocks, we show that the typical uncertainty on the DLA $N_{\rm H\,I}$ and redshift is 0.2 dex and 200 km s^{-1}, respectively. Despite the increased IGM absorption at higher redshifts and the low spectral resolution, we find no strong systematic offset in the estimated $N_{\rm H\,I}$ for DLAs either as a function of redshift or $N_{\rm H\,I}$.

(iv) We find an excess in $\Omega_{\rm H\,I}$ (30 ± 20 per cent) from the brighter half of our QSO sample compared to the fainter half. This is consistent with similar effects found in previous analyses at $z \sim 2.5$, which posited gravitational lensing as a possible explanation. Given the smaller Einstein radius at $z = 4.5$ compared to $z = 2.5$, for our sample this effect seems more likely to be caused by a statistical fluctuation. As such it should not significantly bias our result.

(v) Recent theoretical models do not match the data across their full redshift range ($z = 5$ to 0). A simple power-law model of the form $\Omega_{\rm H\,I} = A(1+z)^\gamma$ with $A = (4.00 \pm 0.24) \times 10^{-4}$ and $\gamma = 0.60 \pm 0.05$, while not physically motivated, does describe the observations over the entire redshift range.

ACKNOWLEDGEMENTS

We thank Marcel Neeleman for comments on an earlier version of this paper, Regina Jorgenson for providing a MagE spectrum of one of the GGG DLAs and the referee for their suggestions. Simeon Bird, Claudia Lagos, Romeel Davé and Edoardo Tescari kindly provided tables of their model predictions. We also thank the late Arthur M. Wolfe for providing unpublished ESI spectra and for early contributions to this work.

Our analysis made use of ASTROPY (Astropy Collaboration 2013), XIDL[6] and MATPLOTLIB (Hunter 2007). NC and MM thank the Australian Research Council for *Discovery Project* grant DP130100568 which supported this work. MF acknowledges support by the Science and Technology Facilities Council (grant number ST/L0075X/1). SL has been supported by FONDECYT grant number 1140838 and received partial support from Center of Excellence in Astrophysics and Associated Technologies (PFB 06).

Based on observations obtained at the Gemini and W. M. Keck Observatories. We wish to acknowledge the very significant cultural role and reverence that the summit of Mauna Kea has always had within the indigenous Hawaiian community. We are most fortunate to have the opportunity to conduct observations from this mountain.

REFERENCES

Astropy Collaboration, 2013, A&A, 558, A33
Barkana R., Loeb A., 2007, Rep. Prog. Phys., 70, 627
Becker G. D., Hewett P. C., Worseck G., Prochaska J. X., 2013, MNRAS, 430, 2067
Becker G. D., Bolton J. S., Madau P., Pettini M., Ryan-Weber E. V., Venemans B. P., 2015, MNRAS, 447, 3402
Berry M., Somerville R., Haas M., Gawiser E., Maller A., Popping G., Trager S., 2014, MNRAS, 441, 939
Bird S., Vogelsberger M., Haehnelt M., Sijacki D., Genel S., Torrey P., Springel V., Hernquist L., 2014, MNRAS, 445, 2313
Bouwens R. J. et al., 2012, ApJ, 752, L5
Braun R., 2012, ApJ, 749, 87
Cameron E., 2011, Publ. Astron. Soc. Aust., 28, 128
Chang T. C., Pen U. L., Peterson J. B., McDonald P., 2008, Phys. Rev. Lett., 100, 091303
Cooke J., Wolfe A. M., Gawiser E., Prochaska J. X., 2006, ApJ, 652, 994
Cooke R., Pettini M., Steidel C. C., King L. J., Rudie G. C., Rakic O., 2010, MNRAS, 409, 679
Crighton N. H. M., Hennawi J. F., Simcoe R. A., Cooksey K. L., Murphy M. T., Fumagalli M., Prochaska J. X., Shanks T., 2015, MNRAS, 446, 18
Davé R., Katz N., Oppenheimer B. D., Kollmeier J. A., Weinberg D. H., 2013, MNRAS, 434, 2645
Dekel A. et al., 2009, Nature, 457, 451
Delhaize J., Meyer M. J., Staveley-Smith L., Boyle B. J., 2013, MNRAS, 433, 1398
Duffy A. R., Kay S. T., Battye R. A., Booth C. M., Dalla Vecchia C., Schaye J., 2012, MNRAS, 420, 2799
Ellison S. L., Yan L., Hook I. M., Pettini M., Wall J. V., Shaver P., 2001, A&A, 379, 393
Ellison S. L., Yan L., Hook I. M., Pettini M., Wall J. V., Shaver P., 2002, A&A, 383, 91
Ellison S. L., Prochaska J. X., Hennawi J., Lopez S., Usher C., Wolfe A. M., Russell D. M., Benn C. R., 2010, MNRAS, 406, 1435
Finlator K., Oh S. P., Özel F., Davé R., 2012, MNRAS, 427, 2464
Font-Ribera A. et al., 2012, J. Cosmol. Astropart. Phys., 11, 059
Freudling W. et al., 2011, ApJ, 727, 40
Fumagalli M., Prochaska J. X., Kasen D., Dekel A., Ceverino D., Primack J. R., 2011, MNRAS, 418, 1796
Fumagalli M., O'Meara J. M., Prochaska J. X., Worseck G., 2013, ApJ, 775, 78
Guimarães R., Petitjean P., de Carvalho R. R., Djorgovski S. G., Noterdaeme P., Castro S., Poppe P. C. D. R., Aghaee A., 2009, A&A, 508, 133
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Jorgenson R. A., Wolfe A. M., Prochaska J. X., Lu L., Howk J. C., Cooke J., Gawiser E., Gelino D. M., 2006, ApJ, 646, 730
Jorgenson R. A., Murphy M. T., Thompson R., 2013, MNRAS, 435, 482
Kim T. S., Partl A. M., Carswell R. F., Müller V., 2013, A&A, 552, A77
Lagos C. D. P., Baugh C. M., Zwaan M. A., Lacey C. G., Gonzalez-Perez V., Power C., Swinbank A. M., van Kampen E., 2014, MNRAS, 440, 920
Lah P. et al., 2007, MNRAS, 376, 1357
Lanzetta K. M., Wolfe A. M., Turnshek D. A., Lu L., McMahon R. G., Hazard C., 1991, ApJS, 77, 1
Liske J. et al., 2008, MNRAS, 386, 1192
Madau P., Dickinson M., 2014, ARA&A, 52, 415
Martin A. M., Papastergis E., Giovanelli R., Haynes M. P., Springob C. M., Stierwalt S., 2010, ApJ, 723, 1359
Meiring J. D. et al., 2011, ApJ, 732, 35
Ménard B., Fukugita M., 2012, ApJ, 754, 116
Murphy M. T., Liske J., 2004, MNRAS, 354, L31
Noterdaeme P., Petitjean P., Ledoux C., Srianand R., 2009, A&A, 505, 1087
Noterdaeme P. et al., 2012, A&A, 547, L1
O'Meara J. M., Prochaska J. X., Burles S., Prochter G., Bernstein R. A., Burgess K. M., 2007, ApJ, 656, 666
O'Meara J. M., Prochaska J. X., Worseck G., Chen H. W., Madau P., 2013, ApJ, 765, 137
Oppenheimer B. D., Davé R., Kereš D., Fardal M., Katz N., Kollmeier J. A., Weinberg D. H., 2010, MNRAS, 406, 2325
Padmanabhan H., Choudhury T. R., Refregier A., 2015, MNRAS, 447, 3745
Péroux C., McMahon R. G., Storrie-Lombardi L. J., Irwin M. J., 2003, MNRAS, 346, 1103
Péroux C., Dessauges-Zavadsky M., D'Odorico S., Sun Kim T., McMahon R. G., 2005, MNRAS, 363, 479
Popping G., Somerville R., Trager S., 2014, MNRAS, 442, 2398
Pontzen A., Pettini M., 2009, MNRAS, 393, 557
Prochaska J. X., Herbert-Fort S., 2004, PASP, 116, 622
Prochaska J. X., Wolfe A. M., 2009, ApJ, 696, 1543
Prochaska J. X., Herbert-Fort S., Wolfe A. M., 2005, ApJ, 635, 123
Prochaska J. X., Hennawi J. F., Herbert-Fort S., 2008, ApJ, 675, 1002
Prochaska J. X., Worseck G., O'Meara J. M., 2009, ApJ, 705, L113
Prochaska J. X., O'Meara J. M., Worseck G., 2010, ApJ, 718, 392
Prochaska J. X., Madau P., O'Meara J. M., Fumagalli M., 2014, MNRAS, 438, 476
Rafelski M., Wolfe A. M., Prochaska J. X., Neeleman M., Mendez A. J., 2012, ApJ, 755, 89
Rafelski M., Neeleman M., Fumagalli M., Wolfe A. M., Prochaska J. X., 2014, ApJ, 782, L29
Rahmati A., Schaye J., Bower R., Crain R., Furlong M., Schaller M., Theuns T., 2015, preprint (arXiv:1503.05553v1)
Rao S. M., Turnshek D. A., Nestor D. B., 2006, ApJ, 636, 610
Rhee J., Zwaan M. A., Briggs F. H., Chengalur J. N., Lah P., Oosterloo T., van der Hulst T., 2013, MNRAS, 435, 2693
Rudie G. C., Steidel C. C., Shapley A. E., Pettini M., 2013, ApJ, 769, 146
Russell D. M., Ellison S. L., Benn C. R., 2006, MNRAS, 367, 412
Saslaw W. C., 1989, ApJ, 341, 588
Shull J. M., Stevans M., Danforth C. W., 2012, ApJ, 752, 162
Songaila A., Cowie L. L., 2010, ApJ, 721, 1448 (S10)
Storrie-Lombardi L. J., Wolfe A. M., 2000, ApJ, 543, 552
Storrie-Lombardi L. J., McMahon R. G., Irwin M. J., 1996, MNRAS, 283, L79
Suzuki N., Tytler D., Kirkman D., O'Meara J. M., Lubin D., 2005, ApJ, 618, 592
Tescari E., Viel M., Tornatore L., Borgani S., 2009, MNRAS, 397, 411
Vladilo G., Prochaska J. X., Wolfe A. M., 2008, A&A, 478, 701
Wolfe A. M., Lanzetta K. M., Foltz C. B., Chaffee F. H., 1995, ApJ, 454, 698
Wolfe A. M., Gawiser E., Prochaska J. X., 2005, ARA&A, 43, 861
Worseck G., Prochaska J. X., 2011, ApJ, 728, 23
Worseck G. et al., 2014, MNRAS, 445, 1745
Wyithe J. S. B., Loeb A., 2008, MNRAS, 383, 606
Zafar T., Péroux C., Popping A., Milliard B., Deharveng J. M., Frank S., 2013, A&A, 556, A141
Zwaan M. A., Meyer M. J., Staveley-Smith L., Webster R. L., 2005, MNRAS, 359, L30

[6] http://www.ucolick.org/~xavier/IDL

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this paper:

Table 3. The start and end redshifts for each QSO used to calculate the redshift search path for DLAs (http://mnras.oxfordjournals.org/lookup/suppl/doi:10.1093/mnras/stv1182/-/DC1).

Please note: Oxford University Press are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.

APPENDIX A: MOCK SPECTRA

We generated a set of mock spectra to quantify the reliability and completeness of our DLA candidates in the low-resolution GMOS spectra. Here we describe how these mocks were produced.

One mock spectrum was generated for each real GMOS spectrum, assuming the same noise properties and the same QSO redshift. Therefore, the sample of mocks has the same redshift and S/N distribution as the real GGG spectra. We model the forest absorption by a distribution of Voigt profiles. Due to the difficulty of profile-fitting the strongly absorbed Lyα forest at high redshifts, the $N_{\rm H\,I}$, b and z distribution of Lyα forest lines at $z > 4$ is not well known. However, the distribution at $z \sim 2.5$ has been measured (e.g. Kim et al. 2013; Rudie et al. 2013). Therefore, we assume that the shape of $f(N_{\rm H\,I})$ at $z \sim 4$–5 is the same as that used by Prochaska et al. (2014) at $z \sim 2.5$, and increase its normalization until the mean flux of the mock spectra at $z = 4.5$ matches the value from Becker et al. (2013). DLAs were generated using $f(N_{\rm H\,I}, X)$ from O'Meara et al. (2013), and we assume that $f(N_{\rm H\,I})$ is redshift independent, whereas $f(N_{\rm H\,I}, X)$ evolves as $(1 + z)^{1.5}$.

We initially did not include any line clustering in the Lyα forest, but found that this produced spectra which were markedly different from the real spectra: there were too few regions with very strong absorption and also too few regions with low absorption. To address this, we introduced line clustering, similar to that used by Liske et al. (2008) to model the Lyα forest at $z \sim 3$. This involves generating absorption at 'clump' positions rather than individual lines. For each clump, 0, 1 or more lines are produced, with the number taken from a Borel distribution (Saslaw 1989) with $\beta = 0.6$. Each line in a clump is offset from the clump redshift by a velocity drawn from a Gaussian distribution with $\sigma = 250$ km s^{-1}. These values of β and σ were chosen by a parameter grid search, varying each until values were found which produce in mock spectra with a Lyα forest which match the flux distribution of the real GMOS spectra. The number of clumps was set such that the mean transmission in the Lyα forest matches the effective optical depth at $z = 4.5$ derived by Becker et al. (2013).

We then generated a QSO continuum from the Principle Component Analysis presented by Suzuki et al. (2005), derived using a sample of low-redshift QSOs observed with the UV Faint Object Spectrograph. We set the QSO redshift to that of the matching GMOS QSO, and added noise to the mock using the same noise array as the GMOS spectrum, normalized so that the median S/N of the mock and the real spectra in the range 7600–7800 Å matches. Using the noise array from the real spectra for the mocks is an approximation, as the noise properties vary with the QSO spectrum (strong absorbers and strong emission lines affect the noise level). However, the variations in noise due to these effects are small in the Lyα forest, so we believe this is a good approximation.

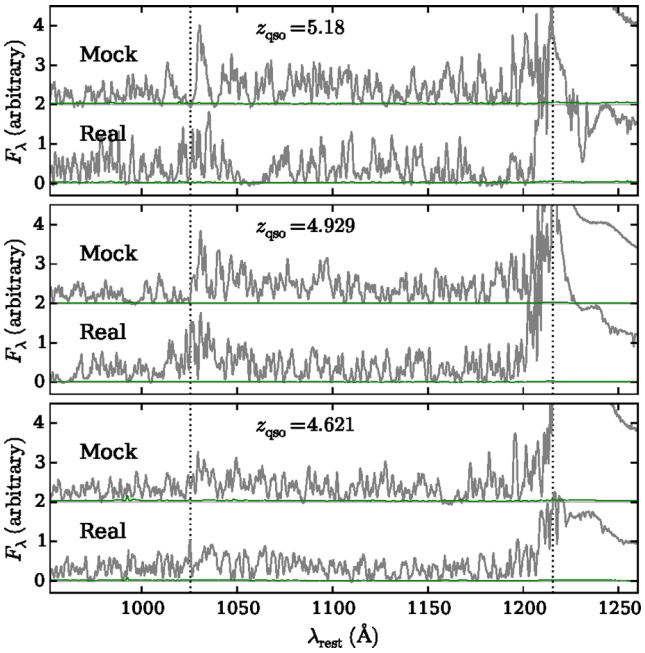

Figure A1. Three mock GMOS spectra, selected at random, with their corresponding real spectra. The real and mock spectra are normalized in the rest-frame wavelength region 940–1200 Å and offset for clarity. The flux distribution in the Lyα forest (between the two dotted vertical lines), where we search for DLAs, is very similar. The thin green lines show the 1σ error array.

Fig. A1 shows three example mock spectra and their corresponding real spectra, selected at random from our sample. The Lyα forest distribution in the mocks matches closely the distribution seen in the real spectra. We do not expect these mocks to correctly reproduce the mean optical depth at the Lyman limit or the power spectrum of Lyα flux absorption. However, our aim is not to reproduce all properties of the real spectra. Instead, we aim to create mock spectra which match by eye the Lyα forest at GMOS resolution, the most important characteristic for DLA identification.

We did not include metal absorption in the mocks. The similarity between the mocks and the real spectra, and the agreement between the correction factors $k_{\rm real}$ and $k_{\rm found}$ derived from the mocks and high-resolution spectra suggest that their inclusion is unnecessary.

A1 High-$N_{\rm H\,I}$ DLAs

DLAs in the column density range $N_{\rm H\,I} = 10^{21-21.8}$ cm^{-2} make the dominant contribution to $\Omega_{\rm H\,I}$, and it is thus important to correctly measure the uncertainty in $k_{\rm real}$ and $k_{\rm found}$ for this $N_{\rm H\,I}$ range. There are only ~10 DLAs in this column density range in both the mocks and the high-resolution sample, so the uncertainties in this correction are large. Therefore, we generated further mocks with an enhanced incidence rate of high-$N_{\rm H\,I}$ systems. We did this by generating 10 times more mocks than were used above, using the same line distribution. Due to time constraints, we were unable to search by eye every one of these mocks. Instead, we selected just 100 spectra: the 50 containing the highest $N_{\rm H\,I}$ DLAs, and a further 50 selected at random from the remainder. This formed a sample of 100 new mock spectra which we searched for high-$N_{\rm H\,I}$ systems. 50 were included without requiring a DLA to present so that when scanning the spectra by eye, the searcher would not be certain that

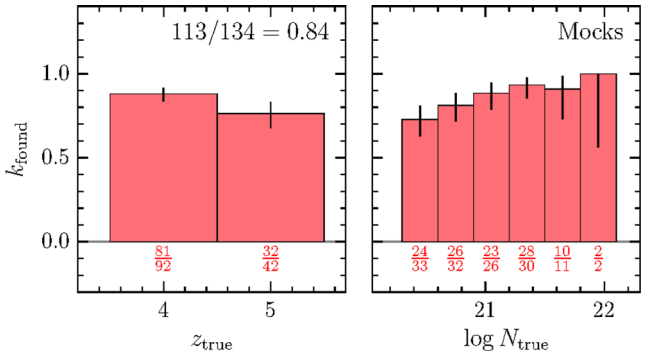

Figure A2. The fraction of true DLAs that were correctly identified by one of the authors (NHMC), $k_{\rm found}$, as a function of the true redshift and $N_{\rm H\,I}$ found using the mock spectra. This includes the mock sightlines with additional strong $N_{\rm H\,I}$ DLAs.

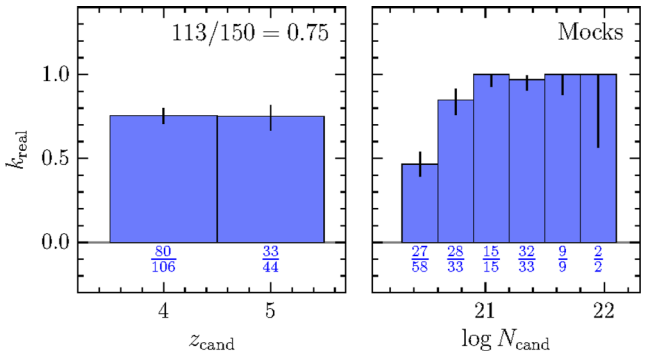

Figure A3. The fraction of non-spurious DLA candidates, $k_{\rm real}$, by one of the authors (NHMC), as a function of the candidate redshift and $N_{\rm H\,I}$ for the mock spectra. This includes the mock sightlines with additional strong $N_{\rm H\,I}$ DLAs.

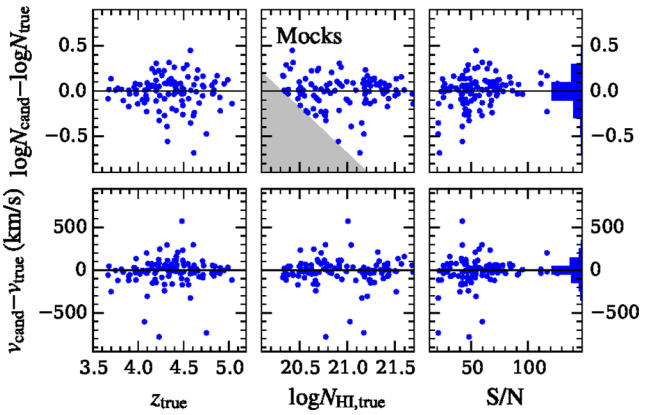

Figure A4. The difference between $N_{\rm H\,I}$ estimated for DLA candidates in low-resolution spectra ($N_{\rm H\,I,cand}$) and $N_{\rm H\,I,true}$ measured from mock line lists by one of the authors (NHMC), including the extra sightlines with additional strong $N_{\rm H\,I}$ DLAs. This shows that there is no strong systematic offset in the estimated $N_{\rm H\,I}$ as a function of redshift, even for $N_{\rm H\,I} \sim 10^{21.5}$ cm^{-2} DLAs.

every spectrum contains a DLA. The $k_{\rm found}$, $k_{\rm real}$ values found by including these extra sightlines into our mock sample are shown in Figs A2 and A3.

The log $N_{\rm H\,I}$ and velocity differences between the candidate and true values are shown in Fig. A4. This shows that even at high $N_{\rm H\,I}$, there is no strong systematic offset from the true value.

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