A DEFINITIVE SURVEY FOR LYMAN LIMIT SYSTEMS AT \(z \approx 3.5\) WITH THE SLOAN DIGITAL SKY SURVEY

J. Xavier Prochaska\(^1\), John M. O’Meara\(^2\), and Gabor Worseck\(^1\)

\(^1\) Department of Astronomy and Astrophysics, UCO/Lick Observatory, University of California, 1156 High Street, Santa Cruz, CA 95064, USA
\(^2\) Department of Chemistry and Physics, Saint Michael’s College, One Winooski Park, Colchester, VT 05439, USA

Received 2009 November 25; accepted 2010 May 26; published 2010 June 30

ABSTRACT

We perform a semi-automated survey for \(\tau_{912} \geq 2\) Lyman limit systems (LLSs) in quasar spectra from the Sloan Digital Sky Survey, Data Release 7. From a starting sample of 2473 quasars with \(z_{\text{em}} = 3.6–5.0\), we analyze 429 spectra meeting strict selection criteria for a total redshift path \(\Delta z = 93.8\) and identify 190 intervening systems at \(z_{\text{LLS}} \geq 3.3\). The incidence of \(\tau_{912} \geq 2\) LLSs per unit redshift, \(\ell_{912}(\gamma)\), is well described by a single power law at these redshifts: \(\ell_{912}(\gamma) = C_{\text{LLS}}[(1 + z)/(1 + z_*)]^{3\gamma_{\text{LLS}}},\) with \(z_* = 3.7\), \(C_{\text{LLS}} = 1.9 \pm 0.2,\) and \(\gamma_{\text{LLS}} = 5.1 \pm 1.5\) (68% c.l.). These values are systematically lower than previous estimates (especially at \(z < 4\)) but are consistent with recent measurements of the mean free path to ionizing radiation. Extrapolations of this power law to \(z = 0\) are inconsistent with previous estimations of \(\ell(z)\) at \(z < 1\) and may indicate a break at \(\approx 2\), similar to that observed for the Ly\(\alpha\) forest. Our results also indicate that the systems giving rise to LLS absorption decrease by \(\approx 50\%\) in comoving number density and/or physical size from \(z = 4\) to 3.3, perhaps due to an enhanced extragalactic ultraviolet background. The observations place an integral constraint on the \(\text{H}\(\alpha\) frequency distribution \(f(N_{\text{HI}}, X)\) and indicate that the power-law slope \(\beta \equiv d \ln f(N_{\text{HI}}, X)/d \ln N_{\text{HI}}\) is likely shallower than \(\beta = -1\) at \(N_{\text{HI}} \approx 10^{18}\) cm\(^{-2}\). Including other constraints on \(f(N_{\text{HI}}, X)\) from the literature, we infer that \(\beta\) is steeper than \(\beta = -1.7\) at \(N_{\text{HI}} \approx 10^{15}\) cm\(^{-2}\), implying at least two inflections in \(f(N_{\text{HI}}, X)\). We also perform a survey for proximate LLSs (PLLs) and find that \(\ell_{\text{PLL}}(\gamma)\) is systematically lower (\(\approx 25\%\)) than intervening systems. Finally, we estimate that systematic effects impose an uncertainty of 10%–20% in the \(\ell(z)\) measurements; these effects may limit the precision of all future surveys.

Key words: diffuse radiation – intergalactic medium

Online-only material: color figures, machine-readable tables

1. INTRODUCTION

Studies of hydrogen absorption in the lines of sight toward distant quasars have served to both define, and in recent years bring precision to, our cosmological models. The low-density, highly ionized Ly\(\alpha\) forest lines (aka the intergalactic medium, IG\(M\)), with \(\text{H}\(\alpha\) column densities \(N_{\text{HI}} < 10^{17.2}\) cm\(^{-2}\), have through their aggregate statistical properties (e.g., their flux power spectrum, mean flux, and column density distributions) constrained cosmological parameters such as the primordial power spectrum and the baryonic mass density, and astrophysical parameters like the amplitude of the ionizing background (e.g., Rauch 1998; Croft et al. 2002; McDonald et al. 2005; Tytler et al. 2004; Faucher-Giguère et al. 2008b). The high-density, predominantly neutral damped Ly\(\alpha\) systems (DL\(\alpha\)s), with \(N_{\text{HI}} > 10^{20.3}\) cm\(^{-2}\), trace the gas that forms stars, and likely represent the progenitors of modern-day galaxies (e.g., Wolfe et al. 1995, 2005; Prochaska & Wolfe 2009).

The majority of Ly\(\alpha\) forest lines and the DL\(\alpha\)s have, through analysis of their Ly\(\alpha\) lines, precisely measured \(N_{\text{HI}}\) values that permit detailed study of their physical properties (e.g., metallicity). For systems with intermediate \(N_{\text{HI}}\) values (\(\approx 10^{18}\) cm\(^{-2}\)), however, Ly\(\alpha\) and most of the Lyman series lines lie on the flat portion of the curve of growth making the \(N_{\text{HI}}\) value difficult to constrain. On the other hand, these systems are optically thick to ionizing radiation and impose a readily identified signature in a quasar spectrum at the Lyman limit. These so-called Lyman limit systems (LL\(\alpha\)s), currently the least well studied \(\text{H}\(\alpha\) absorption systems at high redshift, are the focus of this paper.

Historically, the L\(\alpha\)s were among the first class of quasar absorption line (QAL) systems to be surveyed (Tytler 1982). This is because their spectral signature is obvious in low-resolution, low signal-to-noise ratio (S/N) spectra. The principal challenge is that the Lyman limit occurs redward of the atmospheric cutoff only for systems with redshifts \(z > 2.6\). For lower redshifts, one requires spectrometers on space-borne ultraviolet satellites. By the mid 1990s, samples of several tens of L\(\alpha\)s were generated, spanning redshifts \(0 < z < 4\) (Sargent et al. 1989; Lanzetta 1991; Storrie-Lombardi et al. 1994; Stegler-Larrea et al. 1995). These results were derived from heterogeneous sets of quasars discovered from a combination of color selection, radio detection, and slitless spectroscopic surveys. The spectra, too, were acquired with a diverse set of instrumentation and therefore varying S/N and spectral resolution mitigating differing sensitivity to the precise optical depth at the Lyman limit. Although the results were not fully consistent with one another, the general picture that resulted was a rapidly evolving population of absorption systems reasonably described by a \((1 + z)^{1.5}\) power law.

Cosmologically, the L\(\alpha\)s contribute much if not most of the universe’s opacity to ionizing radiation. And, until recently, the observed incidence of the L\(\alpha\) provided the only direct means of estimating the mean free path \(\lambda_{\text{mfp}}\) at any redshift (e.g., Meiksin & Madau 1993; Madau et al. 1999; Faucher-Giguère et al. 2008a). In a companion paper (Prochaska et al. 2009, hereafter PWO09), we have presented a new technique to measure \(\lambda_{\text{mfp}}\) that circumvents any knowledge of the L\(\alpha\)s. A more precise census of the L\(\alpha\)s will serve as a consistency check for this \(\lambda_{\text{mfp}}\) calculation, but is unlikely to ever again be
a competitive approach. Instead, the incidence of LLS can be used in combination with estimates of $\lambda_{912}$ to assess the $N_{\text{HI}}$ frequency distribution for gas with $N_{\text{HI}} \approx 10^{16} - 10^{18}$ cm$^{-2}$, a regime that is very difficult to explore by studying individual absorption systems. Surveys of the LLSs are also likely to place tight constraints on $z \approx 3$ cosmological simulations that include radiative transfer.

Physically, the nature of systems that give rise to an LLS remains an open question. The systems with the largest $N_{\text{HI}}$ values (i.e., $N_{\text{HI}} \approx 10^{19}$ cm$^{-2}$, the so-called super-LLS or SLLs and DLAs) are likely associated with the interstellar medium and outer regions of high-$z$ galaxies. These high $N_{\text{HI}}$ systems, however, are only a subset of the LLS population. Unfortunately, a proper modeling of the LLSs almost certainly requires careful modeling of radiative transfer in cosmological simulations, which has thus far been beyond the scope of modern computations in cosmological simulations. Indeed, the few studies to date have tended to severely underestimate the incidence of LLS (Katz et al. 1996; Gardner et al. 2001; but see Kohler & Gnedin 2007). This has led to the suggestion that “mini-halos” contribute significantly to the LLS phenomenon (Abel & Mo 1998), although it is not clear that such halos have sufficient cold gas (Maller et al. 2003). In recent simulations of high-$z$ galaxy formation, however, theorists have placed great attention on “streams” of cold gas that carry fresh material from the IGM to star-forming galaxies (Keres et al. 2005; Dekel et al. 2009). These cold streams have relatively large hydrogen surface densities ($N_{\text{HI}} \sim 10^{20}$ cm$^{-2}$) and could therefore produce Lyman limit absorption provided the material has a non-negligible neutral fraction. Consequently, an accurate census of the LLSs with redshift may directly constrain the nature and prevalence of cold streams in the young universe.

A final, yet perhaps most important, motivation for studying the LLSs is that these systems may dominate the census of metals at all epochs. The majority of LLSs are metal-bearing, showing metal-line transitions of common low and high ions (e.g., Prochaska 1999; Prochter et al. 2010). Because the estimated ionization corrections for LLSs with $N_{\text{HI}} \approx 10^{18}$ cm$^{-2}$ are large, observations of ions in an LLS likely track only a trace amount of the metals actually present in the gas. LLSs may show a wider spread in their ionization and metal content relative to the IGM or DLA, further emphasizing the need for a robust LLS survey.

In this paper, we survey the homogeneous data set of quasar spectra from the Sloan Digital Sky Survey (SDSS), using all seven public data releases. Our observational analysis aims to produce the most precise measurement of the LLS incidence paying careful attention to systematic biases. The wavelength coverage and data quality of the SDSS quasar spectra focus the survey at $z \approx 3.5$. Future work will depend on follow-up observations of well-defined quasar samples at other wavelengths.

The paper is organized as follows. In Section 2, we present a set of LLS definitions used throughout the paper. The selection criteria and data quality of quasars from the SDSS database are described in Section 3. The procedure to model the absorbed quasar continuum is presented in Section 4 and the search and characterization of LLSs is detailed in Section 5. The criteria used to measure the survey path are described in Section 6 and an assessment of systematic error and bias from analysis of mock spectra is provided in Section 7. Section 8 describes the principal results and the implications for the IGM and cosmology are discussed in Section 9. Finally, Section 10 presents a summary of the main findings. Throughout the paper, we adopt a ΛCDM cosmology with $H_0 = 72$ km s$^{-1}$ Mpc$^{-1}$, $\Omega_m = 0.3$, and $\Omega_{\Lambda} = 0.7$ and report proper lengths unless otherwise indicated.

2. LYMAN LIMIT SYSTEM DEFINITIONS

The photon cross section of a hydrogen atom at energies above the Lyman limit may be approximated by

$$\sigma_{\text{LL}}(\nu \geq \nu_{912}) \approx 6.35 \times 10^{-18} \left( \frac{\nu}{\nu_{912}} \right)^{-3} \text{cm}^2, \quad (1)$$

with

$$\nu_{912} = E_{912}/h = c/\lambda_{912}, \quad (2)$$

and $E_{912} = 1$ Ryd. Specifically, $\nu_{912} = 3.29 \times 10^{15}$ Hz and $\lambda_{912} = 911.7641$ Å. This implies an optical depth at wavelengths $\lambda \lesssim \lambda_{912}$,

$$\tau_{\text{LL}}(\lambda \lesssim \lambda_{912}) \approx \frac{N_{\text{HI}}}{10^{17.2}} \frac{\lambda}{\lambda_{912}}^{-3}, \quad (3)$$

where $N_{\text{HI}}$ is the H I column density. For a gas “cloud” intersecting a background source with intrinsic flux $F_{\text{int}}(\lambda)$, the observed flux $F_{\text{obs}}(\lambda)$ blueward of the Lyman limit is

$$F_{\text{obs}}(\lambda \lesssim \lambda_{912}) = F_{\text{int}}(\lambda) \exp \left[ -\tau_{\text{LL}}(\lambda) \right]. \quad (4)$$

In what follows, we define a “standard” LLS to be one where the optical depth at $\lambda_{912}$ is $\tau_{912} \geq 2$, i.e., $N_{\text{HI}} \geq 10^{17.5}$ cm$^{-2}$. We refer to these systems as the $\tau_{912} \geq 2$ LLS. This corresponds to greater than 85% attenuation of an incident ionizing radiation field at $\nu = \nu_{912}$. By this definition, the class of LLS includes systems with $10^{20.3}$ cm$^{-2} \geq N_{\text{HI}} \geq 10^{19}$ cm$^{-2}$ (the so-called super-LLS or sub-DLAs, hereafter referred to as SLLS; e.g., O’Meara et al. 2007) and systems with $N_{\text{HI}} \geq 10^{20.3}$ cm$^{-2}$ (the DLAs; e.g., Wolfe et al. 2005). In a few cases, we will distinguish between these “strong” LLSs from those with lower $N_{\text{HI}}$, referring to the latter as $\tau_{912} \lesssim 10$ LLS. We also note that our $\tau_{912} \geq 2$ definition for an LLS differs from other works which adopted $\tau_{912} \geq 1$ or $\tau_{912} \geq 1.5$. These are all observationally driven, not physically motivated definitions.

Observationally, the absorption of a background source by a $\tau_{912} \geq 2$ LLS is readily apparent, even in low S/N spectra. We define absorbers with $\tau_{912} < 2$ (i.e., $N_{\text{HI}} < 10^{17.5}$ cm$^{-2}$) as the partial Lyman limit systems (pLLS). To survey these systems, one requires higher quality spectra or an alternate approach to the analysis.

We define the redshift of an LLS as

$$z_{\text{LLS}} \equiv \frac{\lambda_{912}}{\lambda_{912}^{\text{LLS}}} - 1, \quad (5)$$

where $\lambda_{912}^{\text{LLS}}$ marks the observed onset of LL absorption. In practice, this is often estimated from strong Lyman series lines (e.g., Lyα and Lyβ) that accompany the Lyman limit opacity.

We define the subset of LLSs that occur within 3000 km s$^{-1}$ of the emission redshift of the background source as proximate LLSs (pLLSs). We separate the analysis of these systems from the rest to investigate changes in the incidence of optically thick gas near high-$z$ quasars due to, e.g., the quasar’s radiation field and local environment.

Finally, we define the observable $\ell_{\tau_{912}}(z)$ as the average number of $\tau_{912} \geq 2$ LLS detected per unit redshift at a given
redshift. In the previous literature, this quantity is also expressed as \( n(z) \), \( dN/dz \), and \( dn/dz \). For comparison with previous results in the literature, we also consider \( \ell_{\tau \geq 1}(z) \), the number of \( \tau_{12} \geq 1 \) LLSs detected per unit redshift. We also attempt to separate the contributions to \( \ell_{\tau \geq 2}(z) \) from SLLSs \( \ell_{\text{SLLS}}(z) \). DLA \( \ell_{\text{DLA}}(z) \), and attribute the remainder to the LLSs with \( 10^{17.5} < N_{\text{HI}} < 10^{19} \) cm\(^{-2} \), \( \ell_{\text{LLS}}(z) \).

3. SDSS QUASAR SAMPLE AND SPECTROSCOPY

One of the primary objectives of the SDSS was to discover \( \sim 100,000 \) new quasars across the northern sky (York et al. 2000). The strategy of the SDSS team to achieve this ambitious goal was a four-fold process: (1) obtain deep, multi-band images across a large area of the sky; (2) select quasar candidates by demanding a point-like, point-spread function and imposing color criteria that separate the candidates from the Galactic stellar locus; (3) obtain follow-up spectra for a magnitude-limited sample with a fiber-fed spectrograph. The details of target selection and quasar completeness with redshift are described at length in a series of SDSS papers (e.g., Richards et al. 2002), but see Worseck & Prochaska (2010) for a new and more accurate analysis; and (4) automatically identify quasars and estimate their redshifts (\( z_{\text{em}} \)) through template fitting to the optical spectroscopy.

Of these steps, the second has the greatest impact on a survey for high-\( z \) LLSs. The key issue for our survey is whether the presence of an intervening LLS biases the targeting of the background quasar for follow-up spectroscopy. In effect, a high-\( z \) LLS severely “reddens” the quasar at the bluest optical wavelengths of the SDSS imaging. With this effect in mind, the SDSS team imposed cuts on the \(( u - g )\) color which better separated the quasar locus in color pace from the stellar locus. The net effect, however, is to bias the spectroscopic follow-up against quasar sight lines \( \text{without} \) a foreground LLS (PWO09).

Our analysis indicates an important bias for quasars with \( z_{\text{em}} < 3.6 \). For this reason, we limit the statistical analysis to quasars with \( z_{\text{em}} \geq 3.6 \), but we also explore the bias by considering the incidence of LLLs toward quasars with \( z_{\text{em}} = 3.4-3.6 \).

The quasar spectra analyzed in this paper were taken from the SDSS Data Release 7 (Abazajian et al. 2009). We retrieved the “best” one-dimensional spectrum for every source flagged as a QSO or HIZ_QSO. This totaled 102,418 unique spectra. The SDSS survey employs a fiber-fed, dual-camera spectrometer that provides continuous wavelength coverage from \( \lambda \approx 3800-9200 \) Å at a spectral resolution of FWHM \( \approx 150 \) km s\(^{-1} \). The SDSS team employs a custom, data-reduction pipeline that performs sky subtraction using empirical measurements from fibers placed to avoid objects detected in the SDSS images. The majority of data suffer from excessive sky noise at long wavelengths (\( \lambda > 8000 \) Å) and the instrument throughput and atmospheric absorption limits the sensitivity at the shortest wavelengths (\( \lambda < 4200 \) Å).

A survey for \( \text{H}^\text{i} \) Lyman limit absorption in quasar spectra involves two principle steps. First, one must assess the flux at wavelengths near the Lyman limit in the quasar’s rest frame, \( \lambda \lesssim \lambda_{12}(1 + z_{\text{em}}) \). We discuss our procedure for this step in the following section. The second step is to estimate the flux

\( \text{Figure 1. Histograms of (top) the emission redshifts } z_{\text{em}} \text{ for the quasars comprising our survey and (bottom) the measurements of } S/\text{NA}^2 \text{ to noise of the absorbed continuum at the Lyman limit of each quasar. The sample is restricted to } S/\text{NA}^2 \geq 2. \)
a corresponding overestimate, but this has not been rigorously established. The SDSS fibers are sufficiently wide (diameter of 3′′) that they will occasionally include flux from a projected neighbor. These coincident objects may be much fainter than the quasar at redder wavelengths, but they could contribute all of the flux blueward of a strong LLS and lead to an underestimate of the LL opacity.4

4. THE ABSORBED QUASAR CONTINUUM

Absent any other sources of opacity, one can trivially estimate τ912 from the quasar spectrum through measurements of the flux both redward and blueward of the observed Lyman limit (Equation (4)). In practice, however, the quasar flux is also attenuated by line opacity from the so-called Lyα forest (aka, the IGM). For example, consider an LLS at zLLS = 3.5 intervening a zem = 4 quasar. The LLS attenuates the quasar flux blueward of λLLS = 4103 Å. At this wavelength, the quasar spectrum recorded on Earth will also include opacity from the Lyα forest at zLyα = (1 + zLLS)(λ912/λLyα) − 1, Lyβ absorption from the IGM at zLyβ = (1 + zLLS)(λ912/λLyβ) − 1, etc. It is necessary, therefore, to account for these additional sources of opacity when estimating τ912.

We can express the observed (rest-frame) quasar flux Fobs in terms of the intrinsic flux (just) redward of the Lyman limit Fint as

\[ F_{\text{obs}}(\lambda \gtrsim \lambda_{912}) = F_{\text{int}} \exp[-\tau_{\text{IGM}}(\lambda)] , \]

where τIGM is the effective opacity of the IGM from Lyman series line opacity.5 The LLS introduces an additional, continuous opacity blueward of the Lyman limit:

\[ F_{\text{obs}}(\lambda \leq \lambda_{912}) = F_{\text{int}} \exp[-\tau_{\text{IGM}}(\lambda) - \tau_{\text{LL}}(\lambda)] . \]

A precise estimate of τ912, therefore, requires an estimation of the absorbed quasar flux not its intrinsic flux. Conveniently, this quantity is the observed flux recorded in the spectrum at \( \lambda \gtrsim \lambda_{912} \). There are still significant challenges because the IGM opacity is stochastic on both small (individual Lyman lines) and modest scales (many 10 Å) and the intrinsic quasar spectrum (both shape and normalization) varies from source to source. We now describe an automated procedure to estimate the absorbed continuum from each quasar spectrum.

The traditional method of estimating the quasar continuum is to first identify regions of unabsorbed quasar flux and then to interpolate a continuum level between these regions. For quasars at high redshift, this method is particularly error prone, because at wavelengths below Lyα emission we expect few (and at very high redshift, none) of the pixels to be free of absorption from the IGM. Moreover, this traditional method frequently requires by-hand modification, which is time intensive and subjective to individual biases. Methods do exist to automatically generate a quasar continuum from emission-line characteristics (e.g., Suzuki 2006), but these are designed to infer the intrinsic quasar spectrum not the IGM-absorbed continuum.

Our approach is to match a template model of the average absorbed continuum to each spectrum, allowing for a large-scale tilt (i.e., a unique underlying power-law slope) and arbitrary normalization. We emphasize that this approach is not intended to recover the intrinsic spectral energy distribution (SED) of the quasars. Our scientific interest, in this paper, is to model the absorbed continuum of a quasar near its Lyman limit with the fewest number of parameters. Indeed, for some quasars we derive power-law slopes that are likely unrelated to the physical properties of high-\( z \) quasars.

Our first step is to derive the templates for the average absorbed quasar continuum. Because the line density of the IGM and therefore \( \tau_{\text{IGM}} \) increases with redshift, we perform this analysis in small redshift intervals (\( \delta z = 0.3\)–0.6; Figure 2). For every quasar within the redshift interval, we shift the spectrum to the quasar rest frame using the SDSS-reported \( z_{\text{em}} \) value. Next, we “detilt” the spectrum by removing a power-law shape. This power law is determined as follows.

We have constructed from the SDSS-DR3 data set an average template spectrum, archived in XIDL6 as “full_SDSS_LLS.fits.” For each individual quasar spectrum, we sample quasar pixels with wavelengths greater than Lyα emission, divide by the template spectrum, and measure the slope of the resulting spectrum. After aligning the spectra in the quasar rest frame (nearest pixel), we median-combine the data in each redshift interval.7 The resultant template represents the median intrinsic quasar continuum modulated by the median flux decrement of the IGM. Due to the presence of LLSs, the quasar templates will not be useful at wavelengths near and below the rest-frame Lyman limit; this portion of the spectrum, however, can be used to constrain the mean free path to ionizing radiation (PWO09). For wavelengths blueward of 920 Å, therefore, we set the template to have the value recorded at 920 Å. Conveniently, there are no strong emission features in the quasar SED at these

---

4 An amusing (and plausible) systematic effect related to this is contamination by the light reflected from terrestrial satellites crossing the night sky. Even a brief “exposure” through the 3′′ fiber could dominate the flux at the bluest wavelengths, although this should generally be mitigated by the fact that the SDSS team acquires three unique exposures per target.

5 In the following, we do not explicitly derive the opacity from metals in the IGM, but these may be considered included in \( \tau_{\text{IGM}} \).

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

7 This stack is not optimal for deriving Lyα forest statistics (Dall’Aglio et al. 2009).
The resultant template spectra are shown in Figure 2.

With the templates constructed, a model of the absorbed continuum for each quasar is determined as follows. First, we shift the observed quasar spectrum to the rest frame and divide by the appropriate template spectrum (i.e., according to \(z_{\text{em}}\)). Second, we sample the quasar in the wavelength range \(950 \text{ Å} < \lambda < 1800 \text{ Å}\) and fit a power law \(p(\lambda) = A + B \log(\lambda/\text{Å})\) to the observed flux, weighting by the inverse variance array. The emission lines in this spectral range may bias the fit, but we do review and modify these fits (see below). The product of this power law with the template, when shifted to the observed frame, provides our model for the absorbed quasar continuum. The power-law parameters derived in this fashion are listed in Table 1. Sample fits are shown in Figure 3. With these models, we can calculate the ratio of the observed continuum to the \(1\sigma\)-error array each quasar’s Lyman limit, which we denote as \(S/N_{\lambda_{912}}\). Figure 1(b) shows the distribution of \(S/N_{\lambda_{912}}\) values for the statistical survey. We have explored whether the \(S/N\) of the spectra vary as a function of quasar redshift or the redshift of the survey path (i.e., \(\lambda_{\text{obs}}/\lambda_{912}\)). We find only a mild trend of lower \(S/N\) at higher \(z\).

5. LYMAN LIMIT SYSTEM SEARCH AND CHARACTERIZATION

In the following, we parameterize an LLS by two quantities: (1) its absorption redshift \(z_{\text{LLS}}\) and (2) the total H\(^{\text{i}}\) column density \(N_{\text{H}^{\text{i}}}\). Although the H\(^{\text{i}}\) Lyman series lines are sensitive to the component structure and the Doppler parameters (also known as \(b\)-values) of the “clouds” comprising an LLS, the opacity blueward of \(\lambda_{912}\) is insensitive to these details. Furthermore, the SDSS spectra are generally of too poor quality to constrain such structure using the observed Lyman series lines. Therefore, our model of an LLS assumes a single cloud with a Doppler parameter of \(b = 30 \text{ km s}^{-1}\). An implication of this parameterization is that two systems with small redshift separation are modeled as a single system with the total of the \(N_{\text{H}^{\text{i}}}\) values. Our tests with mock spectra (Section 7) indicate that two absorbers with \(\delta z < 0.1\) are often indistinguishable from a single LLS. The survey presented here, therefore, refers to LL absorption smoothed...
over a redshift interval of $\delta z \approx 0.1$. We return to this point in our presentation and discussion of the survey results.

We have developed an algorithm (sdss_findlls) to automatically search for and characterize LLS absorption in quasar spectra. In brief, the code generates a set of model spectra for the line and continuum opacity of a single LLS with redshifts covering $z = z_0 = (\lambda_0/\lambda_{12} - 1)$ to $(z_{\text{em}} + 0.2)$ where $\lambda_0$ is the starting wavelength of the SDSS spectrum and $z_{\text{em}}$ is the emission redshift reported in the DR7. For the grid of models assumes $N_{\text{HI}}$ column densities $\log N_{\text{HI}} = 16.0, 16.2, 16.4, \ldots, 19.8$ and a Doppler parameter $b = 30 \text{ km s}^{-1}$. We implement a grid with 0.2 dex spacing in $N_{\text{HI}}$ because very few of the spectra have sufficient $S/N$ to provide a more precise estimate. Furthermore, we estimate systematic uncertainties (e.g., related to continuum placement and sky subtraction) to be of this order. These models are convolved with the SDSS instrumental resolution and then applied to the absorbed quasar continuum. Finally, the code constructs a $\chi^2$ grid in $z_{\text{LLS}}$ and $N_{\text{HI}}$, space, identifies the minimum $\chi^2$, and records the “best-fit” values. This approach differs from previous methods which focused solely on the Lyman limit (e.g., Storrie-Lombardi et al. 1994) or relied on “by-eye” analysis (Lanzetta 1991). These previous techniques were sufficient for the small data sets considered which generally had high $S/N$ data. In general, the spectra provide very little constraint on $N_{\text{HI}}$, for values exceeding $10^{17.5} \text{ cm}^{-2}$ until the Ly$\alpha$ profile becomes damped (e.g., Prochaska et al. 2005). Therefore, we report lower limits to $N_{\text{HI}}$ for any LLS with $\tau_{12} \geq 2$.

For sight lines with a single LLS having $N_{\text{HI}} > 10^{17.2} \text{ cm}^{-2}$ and good $S/N_{12}$ (i.e., greater than 5 pixel$^{-1}$ at $\lambda_{12} = (1 + z_{\text{em}})$), we find that the automated algorithm is highly successful on its own. In practice, however, there are several aspects of the data and analysis that require visual inspection of the spectra and interactive modification to the model: first, many spectra have such low $S/N$ that $z_{\text{LLS}}$ and $N_{\text{HI}}$ are poorly determined.

In these cases, a local minimum in $\chi^2$ can occur which gives a misestimate of these quantities. Second, the presence of multiple absorbers along the sight line (e.g., one or more pLLSs with a lower redshift LLS) gives a spectrum that cannot be well modeled by a single LLS. Third, a non-negligible number of the spectra retrieved from the SDSS database purported to be high-$z$ quasars are either at a lower redshift or are another class of astronomical object altogether. Fourth, we found that half of the absorbed continuum models required scaling to higher or lower value by greater than 10%. Finally, we prefer to avoid quasars with strong broad absorption line (BAL) or associated systems to focus the analysis on intervening LLS.

Given the above complications to an automated analysis, we built a graphical user interface (GUI) within the IDL software package (sdss_ehklss; bundled within XIDL) that inputs the data and best-fit LLS model for each object. Two of the authors (J.X.P. and J.M.O.) used this GUI to validate and/or modify all of the models. These authors flagged erroneous spectra (159 examples), strong BAL or associated absorption (quasars showing very strong C IV, N v, and O vi absorption; 290 quasars), or data with such low $S/N$ that any analysis was deemed impossible (114 spectra). For the remainder of sight lines, the authors could modify the continuum (via a multiplicative scalar; Table 1) and/or change the model of LLS absorption (i.e., $z_{\text{LLS}}, N_{\text{HI}}$). This includes absorption due to candidate pLLSs. In many cases, $z_{\text{LLS}}$ was modified to correspond to the strongest, local Ly$\alpha$ absorption line at $\lambda = (1 + z_{\text{LLS}}) \times 1215.67 \text{ Å}$, especially for those systems that also showed absorption at the expected wavelength for Ly$\beta$.

After every sight line was analyzed in this manner, the results from the two authors were compared to assess consistency. Roughly half of the spectra were reviewed because of conflicts in the models. The majority of these were associated with the absorbed continuum placement (typically offsets of 5%–10%) which implied differences in the search path of $|\Delta z| > 0.1$ (see Section 6). In the majority of these cases, we simply

---

### Table 1

| Plate | MID | FiberID | Object Name | $z_{\text{em}}$ | $f_{\text{g}}$ | $A^b$ | $B^b$ | Scale$^c$ | $S/N_{12}^d$ |
|-------|-----|---------|-------------|----------------|-------------|-------|-------|----------|------------|
| 650   | 52143 | 111     | J000238.41—101149.8 | 3.938 | 0 | 2.60 | −0.29 | 1.11 | 0.9 |
| 750   | 52235 | 608     | J000300.34+160027.7 | 3.698 | 0 | 16.35 | −3.54 | 1.00 | 1.5 |
| 650   | 52143 | 48      | J000303.34—105150.6 | 3.646 | 0 | 46.77 | −11.08 | 0.96 | 2.0 |
| 750   | 52235 | 36      | J000335.21+144743.6 | 3.484 | 0 | 13.96 | −3.23 | 1.12 | 0.8 |
| 751   | 52251 | 207     | J000536.38+135949.4 | 3.686 | 0 | 9.96 | −2.00 | 1.00 | 1.1 |
| 751   | 52251 | 562     | J000730.82+160732.5 | 3.501 | 0 | 7.35 | −1.26 | 1.15 | 1.2 |
| 651   | 52141 | 534     | J001001.02—090519.1 | 3.720 | 2 |       |       |       |         |
| 751   | 52251 | 39      | J001115.23+144600.8 | 4.967 | 0 | 80.36 | −18.34 | 1.00 | 3.0 |
| 752   | 52251 | 378     | J001134.52+155137.3 | 4.325 | 0 | 15.67 | −3.60 | 1.00 | 1.1 |
| 752   | 52251 | 204     | J001328.21+135828.0 | 3.575 | 0 | 16.28 | −2.75 | 1.07 | 2.5 |
| 752   | 52251 | 5       | J001747.90+141015.7 | 3.955 | 0 | 21.69 | −5.06 | 0.92 | 1.0 |
| 753   | 52233 | 501     | J001813.88+142455.6 | 4.235 | 0 | 22.74 | −5.04 | 1.00 | 1.7 |
| 753   | 52233 | 291     | J001820.71+141851.5 | 3.936 | 0 | 70.03 | −16.86 | 1.32 | 3.6 |
| 753   | 52233 | 391     | J001918.43+150611.3 | 4.153 | 0 | 1.91 | 0.13 | 1.13 | 1.2 |
| 390   | 51900 | 271     | J001950.05—004040.7 | 4.327 | 0 | 21.55 | −4.96 | 0.89 | 1.9 |

---

Notes. List of all objects tagged as QSO or HIZ_QSO in the SDSS Data Release 7 with estimated redshift $5 > z > 3.4$.

1. Flag indication—0: normal; 1: not at SDSS-reported redshift and/or not a quasar (excluded); 2: too low $S/N$ for evaluation (excluded); 3: strong BAL (excluded).

2. Absorbed continuum fitting parameters of the form: $C = A + B \log_{10}(\lambda/\text{Å})$. We caution the reader that these models are not meant to describe the intrinsic spectral energy distributions of the quasars (see the text).

3. Additional scaling factor imposed by the authors on the best-fit absorbed quasar continuum.

4. Estimate of the signal to noise for the absorbed quasar continuum at $\lambda = \lambda_{12}(1 + z_{\text{em}})$.

(This table is available in its entirety in a machine-readable form in the online journal. A portion is shown here for guidance regarding its form and content.)
averaged the two estimations of the continuum. The second most frequent conflict was on the definition of strong BAL absorption, primarily because we did not adopt uniform or strict criteria. In most cases, we conservatively excluded the sight line. There were also ≈100 cases where one author estimated log $N_{H_1} = 17.4$ when the other estimated log $N_{H_1} = 17.6$, i.e., straddling the $\tau_{12} = 2$ boundary that defines our LLS search. These were especially scrutinized for the presence of higher-order Lyman series lines. Where necessary, the final best estimate for $N_{H_1}$ was deferred to the third author (G.W.).

Tables 2 and 3 list the set of LLSs identified in the SDSS-DR7 for systems that (respectively) influence our statistical and non-statistical LLS surveys. For each system, we list our best estimate for $z_{em}$ and otherwise. For each system, we list our best $\tau_{12}$ for systems that (respectively) influence our statistical and non-statistical LLS surveys. In most cases, we conservatively excluded the sight line. There were also ≈100 cases where one author estimated log $N_{H_1} = 17.4$ when the other estimated log $N_{H_1} = 17.6$, i.e., straddling the $\tau_{12} = 2$ boundary that defines our LLS search. These were especially scrutinized for the presence of higher-order Lyman series lines. Where necessary, the final best estimate for $N_{H_1}$ was deferred to the third author (G.W.).

Tables 2 and 3 list the set of LLSs identified in the SDSS-DR7 for systems that (respectively) influence our statistical and non-statistical LLS surveys. For each system, we list our best estimate for $z_{em}$ and otherwise. For each system, we list our best $\tau_{12}$ for systems that (respectively) influence our statistical and non-statistical LLS surveys. In most cases, we conservatively excluded the sight line. There were also ≈100 cases where one author estimated log $N_{H_1} = 17.4$ when the other estimated log $N_{H_1} = 17.6$, i.e., straddling the $\tau_{12} = 2$ boundary that defines our LLS search. These were especially scrutinized for the presence of higher-order Lyman series lines. Where necessary, the final best estimate for $N_{H_1}$ was deferred to the third author (G.W.).

Figure 3 shows a representative sample of four quasar spectra, zoomed into the region blueward of Ly$\beta$, with the absorbed continuum and LLS models indicated. We provide snapshots of the LL region for all quasar spectra in the statistical sample online.9

---

### Table 2

| Quasar       | $z_{em}$ | $\tau_{12}$ | $S/N_{em}$ | $z_{em} < 3.6$ | $\tau_{12}$ |
|--------------|----------|-------------|------------|----------------|-------------|
| J001328.21+135828.0 | 3.575    | 3.443       | ...        | ...            | ...         |
| J015741.56−010629.6 | 3.564    | 3.387       | 3.387      | 3.387          | ...         |
| J073947.17+445236.7 | 3.575    | 3.300       | ...        | ...            | ...         |
| J074914.13+305605.8 | 3.436    | 3.300       | 3.300      | 3.300          | ...         |
| J080503.05+324138.7 | 3.425    | 3.300       | 3.300      | 3.300          | ...         |
| J001115.23+144601.8 | 4.967    | 4.567       | ...        | ...            | ...         |
| J011820.71+141851.5 | 3.936    | 3.536       | 3.596      | 3.596          | ...         |
| J004219.74−102009.4 | 3.880    | 3.633       | 3.633      | 3.633          | 3.633       |
| J010619.24+004823.3 | 4.449    | 4.049       | 4.049      | 4.049          | 4.049       |
| J011351.96−093551.0 | 3.668    | 3.615       | 3.615      | 3.615          | 3.615       |
| J012403.77+044332.7 | 3.834    | 3.434       | 3.434      | 3.434          | 3.434       |
| J015048.82+004126.2 | 3.702    | 3.302       | 3.302      | 3.302          | 3.302       |
| J015339.61−001104.8 | 4.194    | 3.879       | ...        | 3.879          | 3.879       |

Notes. The quasars with $z_{em} < 3.6$ are not included in the final analysis. The starting redshifts correspond to the wavelength at which the absorbed continuum model, starting from $z_{em}$, no longer exceeds the smoothed noise array by the specified $S/N_{em}$ threshold. This value is limited to a maximum offset from $z_{em}$ of $\delta z = 0.4$.

(This table is available in its entirety in a machine-readable form in the online journal. A portion is shown here for guidance regarding its form and content.)

### Table 3

| Quasar       | $z_{em}$ | $\tau_{12}$ | $S/N_{em}$ | $\log N_{H_1}$ | $S/N^a$ |
|--------------|----------|-------------|------------|----------------|---------|
| J000238.41−101149.8 | 3.938    | 3.809       | 17.2       | 0.7            |         |
| J000300.34+160027.7 | 3.698    | 3.570       | 17.2       | 1.3            |         |
| J000303.34−105150.6 | 3.646    | 3.476       | 17.6       | 1.5            |         |
| J000335.21+144743.6 | 3.484    | 3.498       | 16.6       | 0.9            |         |
| J000536.38+153949.4 | 3.686    | 3.580       | 19.0       | 1.1            |         |
| J000730.82+160732.5 | 3.501    | 3.511       | 19.8       | 1.3            |         |
| J001115.23+144601.8 | 4.967    | 3.995       | 17.8       | 2.6            |         |
| J001134.52+151537.3 | 4.325    | 4.348       | 19.8       | 1.1            |         |
| J001328.21+135828.0 | 3.575    | 3.282       | 19.8       | 1.7            |         |
| J001747.90+141015.7 | 3.955    | 3.925       | 17.4       | 1.0            |         |
| J001813.88+142455.6 | 4.235    | 4.151       | 17.2       | 1.6            |         |
| J001820.71+141851.5 | 3.936    | 3.456       | 17.2       | 2.4            |         |
| J001918.43+150611.3 | 4.153    | 4.053       | 17.8       | 1.2            |         |
| J001950.05−004040.7 | 4.327    | 4.047       | 18.4       | 1.8            |         |
| J002120.05+155125.7 | 3.698    | 3.671       | 17.8       | 2.0            |         |
| J002614.69+143105.2 | 3.973    | 3.895       | 19.8       | 1.3            |         |
| J002618.67+140946.7 | 4.572    | 4.414       | 17.4       | 0.9            |         |
| J002627.28+152446.3 | 3.590    | 3.461       | 17.8       | 1.2            |         |
| J002627.28+152446.3 | 3.590    | 3.574       | 16.8       | 1.3            |         |

Notes.

4 The $N_{H_1}$ values listed serve as a rough estimate. Typical uncertainties for systems with $N_{H_1} < 10^{17.5}$ cm$^{-2}$ are at least 0.2 dex.

6 Estimate of the $S/N$ of the absorbed continuum (ignoring the effects of pLLS) at $\lambda = \lambda_{\alpha}(1 + \tau_{12})$.

(This table is available in its entirety in a machine-readable form in the online journal. A portion is shown here for guidance regarding its form and content.)
6. SURVEY PATH

Analogous to galaxy surveys where one defines a search volume based on the depth of imaging and spectroscopic follow-up, measurements of the incidence of QAL systems requires an estimate of the total spectral path sensitive to a robust search. This is generally referred to as the redshift path covered (by translating observed wavelength into redshift). For the survey of LLSs, we have adopted the following criteria for including spectral regions in the search. These are based on our automated and interactive analysis of the SDSS spectra, our modeling of Keck/LRIS follow-up spectra, and our analysis of simulated spectra (Section 7, the Appendix).

1. The search path will begin at a minimum redshift of \( z_{\text{start}} \geq 3.3 \).
2. For the intervening LLS sample, the search path ends at the redshift \( z_{\text{end}} \) corresponding to 3000 km s\(^{-1}\) (relativistic) blueward of the quasar redshift \( z_{\text{em}} \).
3. The absorbed continuum flux must exceed twice the estimated error array, i.e., \( S/N_{\text{NA}}^A \).
4. The search path begins at a maximum offset of \( \delta z = z_{\text{em}} - z_{\text{start}} \leq 0.4 \).

The first criterion is motivated by the starting wavelength of the SDSS spectra (\( \lambda_0 \approx 3800 \) Å) and the poorer quality of the data at the bluest wavelengths. For redshifts less than 3.3, there is insufficient spectral coverage and/or data quality to confidently assess the presence of an LLS. The second criterion minimizes the influence of the quasar and its environment on the analysis. This criterion is relaxed in the study of PLLSs.

The third criterion is the most subjective, yet important, for setting the redshift survey path of each quasar. Algorithmically, we impose this constraint by identifying the first pixel blueward of \( \lambda_{\text{start}}^{\text{NA}} \equiv \lambda_{\text{start}}^{1+z_{\text{em}}} \) where our model of the absorbed quasar continuum falls below twice the median-smoothed (15 pixels) 1σ-error array. This pixel defines the starting redshift \( z_{\text{start}}^{\text{NA}} \) corresponding to an \( S/N_{\text{NA}}^A \geq 2 \) limit. If the first pixel blueward of \( \lambda_{\text{start}}^{\text{NA}} \) does not satisfy the \( S/N \) threshold, the quasar has zero redshift path, i.e., \( z_{\text{start}} = z_{\text{end}} \). One can, of course, define starting redshifts corresponding to higher (or lower) \( S/N_{\text{NA}}^A \) limits; indeed, our fiducial choice of \( S/N_{\text{NA}}^A = 2 \) should be considered arbitrary, although it is guided by our analysis of real and simulated spectra. And to avoid a systematic bias associated with PLLSs (see Section 7.3), one must choose the \( S/N_{\text{NA}}^A \) criterion to be sufficiently high to discover \( z_{\text{start}} \geq 2 \) LLSs even in the presence of a PLLS. We investigate the impact of this choice on our results later in the paper. Finally, our fourth criterion is imposed to mitigate the cumulative effects that PLLSs can have on our ability to detect LLSs with \( z_{\text{start}} \geq 2 \). That is, trials with mock spectra (Section 7) indicate that multiple PLLSs along a single sight line may prevent the detection of a \( z_{\text{start}} \geq 2 \) LLS and that this bias is minimized provided \( \delta z < 0.4 \). Furthermore, we find that the extrapolation of the absorbed continuum from the quasar’s Lyman limit often is a poor model for \( \delta z > 0.4 \).

The starting redshift is further modified by the presence of Lyman limit absorption. In the case of a \( z_{\text{start}} \geq 2 \) absorber, the quasar flux is severely depressed below \( \lambda = \lambda_{\text{start}} (1 + z_{\text{LLS}}) \) and we terminate the search path at this wavelength. Specifically, this implies \( z_{\text{start}} \geq z_{\text{LLS}} \) for all sight lines with \( z_{\text{start}} \geq 2 \) LLS. For sight lines where one or more PLLSs are identified, we again consider arbitrary, although it is guided by our analysis of real and simulated spectra. And to avoid a systematic bias associated with PLLSs (see Section 7.3), one must choose the \( S/N_{\text{NA}}^A \) criterion to be sufficiently high to discover \( z_{\text{start}} \geq 2 \) LLSs even along sight lines where one or more PLLSs are present.

Table 2 presents the list of quasars in SDSS-DR7 that (1) have \( 3.6 \leq z_{\text{em}} \leq 5 \), (2) were not identified to exhibit strong BAL signatures, and (3) have \( 3.3 \leq z_{\text{start}}^{\text{NA}} < z_{\text{em}} \). There are 429 quasars satisfying these criteria. For each sight line, we list the starting redshifts for \( S/N_{\text{NA}}^A = 2 \) and 3 limits. We report results for these two values of the \( S/N \) threshold to search for data-quality biases. Table 2 also lists all LLSs with \( z_{\text{start}} \geq z_{\text{LLS}} \).

We do list these quantities for quasars with \( z_{\text{em}} < 3.6 \), but these were not included in our final statistical analysis because of the bias related to SDSS targetting criteria previously mentioned by PWO09.

Using the values presented in Table 2, it is straightforward to calculate \( z_{\text{end}} \) and the redshift search path for each quasar: \( \Delta z_i = z_{\text{end}} - z_{\text{start}} \). For a given \( S/N_{\text{NA}}^A \) limit, the total search path for the full data set is

\[
\Delta z_{\text{TOT}} = \sum \Delta z_i . \tag{8}
\]

We calculate \( \Delta z_{\text{TOT}}^{S/N_{\text{NA}}^A=2} = 96 \). To maintain a homogeneous sample and set of search criteria, we do not include previous studies in our analysis. We compare against previous results in Section 9.

In Figure 4, we present the sensitivity function \( g(z) \) which expresses the number of SDSS quasars at redshift \( z \) where a robust search for LLSs is possible. The several solid curves represent differing \( S/N_{\text{NA}}^A \) limits for the survey. We also present the sensitivity function for PLLSs which corresponds to the sample of quasar spectra that satisfy the \( S/N_{\text{NA}}^A \geq 2 \) criterion at 3000 km s\(^{-1}\) blueward of \( z_{\text{em}} \).
With the definition of the search path and the identification of the LLSs along each sight line (Section 5), it is straightforward to calculate the incidence of intervening \( r_{912} \geq 2 \) LLS per redshift interval, \( \ell_{912}(z) \). The standard estimator is to compare the total number of LLSs against the total survey path in discrete redshift intervals as derived from survival statistics (Tytler 1982). We will return to evaluate \( \ell_{912}(z) \) and discuss the values after exploring several sources of systematic error.

### 7. Mock Spectra and Systematic Errors

With a survey the size of SDSS, one can quickly reduce the statistical noise in measurements to very small levels. In this regime, one must carefully assess all sources of systematic error as these may dominate the measurement uncertainty. To this end, we have conducted a range of tests with mock spectra as described in this section. The casual reader may wish to skip to the summary of this section (Section 7.7).

#### 7.1. Mock Spectra Construction

We generated two sets of mock SDSS quasar spectra and have analyzed them in the same way as the real data in order to assess bias and completeness in our LLS survey. In the first set, we examined 800 sight lines with a range of emission redshift and S/N characteristics to span the values observed in the SDSS data set. These spectra also assumed an IGM with a higher incidence of LLSs than we recover from the SDSS data set. In the second sample of mock spectra, we generated 429 spectra with \( z_{\text{em}} \) and S/N values chosen to very nearly match the SDSS survey analyzed in this paper, including the statistics of LLSs.

The H I forest absorption spectra were generated via a Monte Carlo routine similar to the one described in Dall’Aglio et al. (2008), assuming that the Ly\( \alpha \) forest is well characterized by three independent distributions: (1) the Ly\( \alpha \) line-incidence \( \ell_{\alpha}(z) \propto (1 + z)^{\gamma} \), (2) the H I column density distribution \( f(N_{\text{HI}}) \propto N_{\text{HI}}^\beta \), and (3) the Doppler parameter distribution parameterized as \( f(b) \propto b^{-5} \exp[-b^4/b^4] \) (Hui & Rutledge 1999). Each simulated line of sight was filled with H I Ly\( \alpha \) absorption lines at \( 2 < z < 4.6 \) until the H I Ly\( \alpha \) effective optical depth was consistent with Faucher-Giguère et al. (2008b), both in normalization and redshift evolution. We did not model the \( z \sim 3.2 \) dip in the effective optical depth measured by Faucher-Giguère et al., and instead adopted a simple power-law \( \tau_{\text{eff}}(z) = 0.0011(1 + z)^{4.23} \). If the number of lines in a given patch of the forest is Poisson distributed, a power-law line-density evolution \( \ell_{\alpha}(z) \propto (1 + z)^{\gamma} \) yields a power-law effective optical depth evolution \( \tau_{\text{eff}}(z) \propto (1 + z)^{\gamma + 1} \) (Zuo 1993). The column density distribution was modeled with a single power-law index \( \beta = -1.5 \) for \( 12 < \log(N_{\text{HI}}) < 19 \), but with a 0.5 dex break at \( \log(N_{\text{HI}}) = 14.5 \) in order to account for the dearth of high column density lines, consistent with observations (e.g., Hu et al. 1995; Kim et al. 2002; and our own inferences, Section 8.3). For the Doppler parameter distribution, we set \( v_0 = 24 \text{ km s}^{-1} \) (Kim et al. 2001).

Because SLLSs (19 \( \leq \log(N_{\text{HI}}) < 20.3 \)) and DLAs (log(N\( _{\text{HI}} \)) \( \geq 20.3 \)) have different column density distributions and are usually excluded in measurements of \( \tau_{\text{eff}} \), these were added after the log(N\( _{\text{HI}} \)) < 19 line forest converged to the chosen \( \tau_{\text{eff}}(z) \). To constrain the redshift evolution of SLLSs, we combined the sample by O’Meara et al. (2007) and the lower limit given in Rao et al. (2006), yielding \( \ell_{\text{SLLS}}(z) \approx 0.066(1 + z)^{1.70} \). For the SLLS column density distribution, we adopted \( \beta = -1.4 \) (O’Meara et al. 2007). The DLAs were modeled via \( \tau_{\text{DLA}}(z) = 0.044(1 + z)^{1.27} \) (Rao et al. 2006) and \( \beta = -2 \) (Prochaska et al. 2005), ignoring deviations in \( f(N_{\text{HI}}, X) \) from a single power law. The Doppler parameter distribution was left unchanged.

With the overall opacity of the modeled Ly\( \alpha \) forest consistent with observations, and the high column density systems incorporated, we used the generated line lists to compute H I Lyman series (up to Ly30) and Lyman continuum absorption spectra. In total, we computed 800 different lines of sight for quasars in the redshift range of our sample, 160 each at \( z = 3.4, z = 3.6, z = 3.8, z = 4.0 \), and \( z = 4.2 \), respectively. From these, we then generated mock SDSS spectra. First, the resolved H I forest spectra were multiplied onto synthetic quasar SEDs generated from principal component spectra (Suzuki 2006). We then degraded the resolution of the mock spectra to \( R = 2000 \) by convolving them with a Gaussian, and rebinned them to \( \delta v = 69 \text{ km s}^{-1} \), matching the approximate resolution and pixel size of the SDSS spectra. Finally, we added Gaussian noise to the mock SDSS spectra. In each mock spectrum, the S/N was normalized in the quasar continuum at 1450 Å and varied as a function of flux and wavelength according to the throughput of the SDSS spectrograph. The sky level was approximated as a constant and readout noise was also incorporated. Finally, for a subset of the mocks we imposed a sky-subtraction error implemented by subtracting/adding a constant to the spectrum. These were generated to mimic such systematic errors that occasionally occur in the SDSS spectra.

Four representative examples are shown in Figure 5.

#### 7.2. LLS Recovery and Sky-subtraction Bias

In this subsection, we report on results from analysis of the first set of mock spectra: those with a wide range of \( z_{\text{em}} \) and S/N values. Two of the authors (J.X.P. and J.M.O.) analyzed these mock spectra using the identical tools and procedures applied to the real SDSS spectra. These steps were done without knowledge of the mock line distribution and column densities (constructed by author G.W.). The integrated results for the two authors were nearly identical; the following discussion and figures refer to the results from the analysis of J.X.P.

Figure 6 summarizes the completeness and several biases uncovered by our analysis. In the top panel, all cases where a mock \( r_{912} \geq 2 \) LLS exists with \( z_{\text{abs}} > 3.2 \) and an LLS was “observed” are presented; these correspond to \( >80\% \) of the cases. Specifically, we plot the offset \( \delta z \) between the true and observed LLS absorption redshifts as a function of the S/N\( _{912} \) of the spectrum. We find excellent agreement (small \( \delta z \), nearly independent of the S/N\( _{912} \) of the data). There are, however, a number of cases with \( \delta z > 0.01 \), primarily related to the blending of absorption lines (see Section 7.4).

The lower panel in Figure 6 presents those cases that are false negatives (an LLS in the mock spectrum that was not detected by our analysis; diamonds) and false positives (an LLS identified by our analysis that was not present in the mock spectrum; triangles). The latter were very rare because these mock spectra had a high inputted incidence of LLS such that at least one LLS was typically present in each spectrum. The former, however, did occur quite often in our analysis and can be divided into two classes: (1) cases with \( |\delta z| > 0.05 \), which are proper false negatives (i.e., true \( r_{912} \geq 2 \) LLS that were modeled as a pLLS); and (2) cases with \( \delta z > 0.1 \), which are sight lines where a higher pLLS precluded the detection of a lower LLS. The majority of the latter cases are due to bona-fide pLLS at higher \( z \) which
greatly diminish the S/N of the spectra at shorter wavelengths and “obscure” the presence of a τ_{912} ≥ 2 LLS. Almost none of these cases, however, satisfy the selection criteria established in Section 6; either the data have too low S/N_{912} or the absorption redshift of the LLS gives z_{em} − z_{abs} ≥ 0.4.

The first class of false negatives, meanwhile, are almost exclusively associated with spectra that had systematically low estimates for the sky background. In these cases, a τ_{912} ≥ 2 LLS has an apparent flux at λ < 912 Å and therefore was modeled as a pLLS. This is the dominant effect of a sky-subtraction bias. Although our mock spectra had even numbers of over and undersubtracted sky backgrounds, only the former are relatively easy to identify (large regions of spectra are significantly negative) and ignore. The net effect of a random sky-subtraction error is a systematic underestimate in the incidence of LLS. We stress, however, that the magnitude and frequency of poor sky subtraction in the mock spectra was intentionally elevated so that we could explore these effects. The incidence of such effects within the SDSS spectra is much lower, an assertion supported by our follow-up spectra with Keck/LRIS (the Appendix) and a careful search of the SDSS spectra for cases where strong Lyα lines do not show zero flux levels. Therefore, we are confident that the sky-subtraction bias gives rise to a less than 10% systematic error for τ_{912} ≥ 2(z).

Figure 7 shows the same analysis as in Figure 6 but as a function of redshift of the absorption system. We identify no trend with redshift except for a decreasing δz for higher z in the false negatives which is trivially due to the definition of the false negatives (cases where the redshift of the true LLS is lower will by default give higher δz values).

If we limit the analysis to the spectra with small sky-subtraction errors (less than 10%), then we recover to within 5% the inputted incidence of τ_{912} ≥ 2 LLS. There is no trend with redshift. We caution, however, that the median S/N of this first set of mock spectra was skewed to higher values than the S/N_{912} = 2 cut from the SDSS. It was used, in large part, to inform the procedures and selection criteria to be applied to the SDSS spectra. After completing the SDSS analysis, we generated a new set of mock spectra designed to closely mimic the properties of the SDSS spectra. The analysis of these mock spectra are discussed in a following subsection.

7.3. The pLLS Bias

Originally, we intended to perform a search for LLSs in spectra with S/N_{912} = 1 to maximize the pathlength of the survey (a nearly 4× increase over S/N_{912} = 2). Our tests with the mock spectra and follow-up observations with Keck/LRIS (the Appendix) indicated that we could robustly identify τ_{912} ≥ 2 LLS in such data. We also noted, however, that many of the spectra showed pLLS candidates, which reduced the S/N_{912} to below 1 and made the search for τ_{912} ≥ 2 LLS much more challenging. Our response was to redefine the search path by attenuating the absorbed continuum due to any identified pLLS candidates and then reapply the S/N_{912} criterion for the remaining z < z_{pLLS} spectral range. With this approach, the search path was frequently terminated by the presence of a pLLS candidate.

In principle, this modification should provide an unbiased search for τ_{912} ≥ 2 LLS. Our trials with mock spectra, however, revealed an insidious bias associated with this redefinition of the
mock spectra with significant underestimates in the sky background. For the all of the survey criteria (e.g., $z > \ell \tau$) of this systematic bias for it depends sensitively on find that it is very difficult to precisely estimate the magnitude with $\delta z > (A color version of this figure is available in the online journal.)

Figure 6. Offset in redshift $\delta z$ for the “observed” LLS and pLLS from the $z_{\text{abs}}$ value of the nearest “true” $r_{12} > 2$ LLS in our mock spectra. The analysis is restricted to the highest $z_{\text{abs}}$ LLS along each sight line. In the upper panel, we show the $\delta z$ value for sight lines where an LLS was “observed” in our analysis of the spectrum and is also truly present in the mock data. Lighter/darker (green/black) points correspond to LLSs discovered outside/within the statistical survey path (i.e., $S/N_{12} = 2$). These are the majority of cases ($>80\%$) and we find small $\delta z$ values with a small, but important bias to $\delta z = 0.01$. In the lower panel, we show false negatives (diamonds) and false positive (triangles) detections. The false negatives correspond to cases where a $r_{12} > 2$ LLS is present in the mock spectrum but it was analyzed as a pLLS (those with $\delta z < 0.1$) or a higher redshift pLLS masked its presence altogether (those with $\delta z > 0.1$). The misidentifications with $\delta z < 0.1$ are dominated by the mock spectra with significant underestimates in the sky background. For the misidentifications with $\delta z < 0.1$, the dark points indicate systems that satisfy all of the survey criteria (e.g., $z > 3.3, S/N_{12} \geq 2$).

(A color version of this figure is available in the online journal.)

Figure 7. Same as in Figure 6 but plotted as a function of the LLS redshift instead of the spectral $S/N$. In the upper panel, the redshift plotted is that of the LLS in the mock spectrum. In the lower panel, it is the same for cases where an LLS really exists in the mock spectrum. Otherwise, it is for the LLS identified in our analysis of the mock spectrum.

(A color version of this figure is available in the online journal.)

7.4. The Blending Bias

As noted in the previous subsection, LLSs and pLLSs that lie close to one another in redshift are very difficult to distinguish as individual systems. This is even true in the limit where one has spectra with exquisite $S/N$ and resolution when $\delta z < 0.1$ (or less in the case of high-resolution echelle observations). With SDSS spectra, the limited information provided by the Lyman limit and the strongest Lyman series lines is insufficient to robustly distinguish multiple LLSs from a single system. This leads to a “blending bias” that manifests itself in several ways.

First, the blending bias increases the number of LLSs observed because pairs of pLLSs blend together to give a single system with $r_{12} \geq 2$. Second, the absorption redshifts of the LLSs are shifted to higher redshifts because one generally adopts $z_{\text{LLS}}$ from the higher of the pair of systems. This leads to a smaller survey path and possibly a higher inferred incidence of LLSs. More importantly (see below), many LLSs are shifted into the proximate region of the quasar. This causes an underestimate of $\ell_{r \geq 2}$ for intervening LLSs and an overestimate of pLLSs.

We explored the quantitative effects of the blending bias with the following analysis. We constructed a set of mock absorption lines for each quasar in the statistical survey (Table 2) with an incidence set to match our measurements (Section 8.1). Specifically, we adopted an $N_{\text{H}}$ frequency distribution

$$f(N_{\text{H}}; z) = CN_{\text{H}}^{\beta} \left( \frac{1 + z}{1 + z_s} \right)^{\gamma}$$

with $\beta = -1.3$ for $10^{16.5} \leq N_{\text{H}} \leq 10^{19.5}$ cm$^{-2}$ and $\beta = -2$ for $N_{\text{H}} \geq 10^{19.5}$, $z_s = 3.7, \gamma = 5.1$ and...
C = 1.244 × 10^5. From the mock absorber list, we identified all LLSs with \( N_{\text{H}1} \geq 10^{17.5} \) cm\(^{-2}\). This formed the control sample. Then, we blended together all systems with \(|\delta z_j| < \delta z_j\) where \( \delta z_j = [0.01, 0.05, 0.1, 0.2]\) and reidentified systems satisfying \( N_{\text{H}1} \geq 10^{17.5} \) cm\(^{-2}\). When blending two or more systems together, we set \( z_{\text{abs}} \) to the maximum of all the lines. Finally, we calculated the incidence of LLSs using the survey path and LLSs for each \( \delta z_j \).

Figure 8(a) presents the results of the blending bias in terms of the enhancement/decrement of \( \ell_{\tau \geq 2}(z) \) relative to no bias) for mock absorption line systems for (upper) intervening \( t_{912} \geq 2 \) LLSs and (lower) proximate \( t_{912} \geq 2 \) LLSs. The absorption-line statistics were set to roughly match the observed incidence. Even for blending with redshifts \( \delta z_j = 0.2 \), the bias for intervening systems is relatively small. In contrast, the blending bias systematically elevates \( \ell_{\tau \geq 2}(z) \) for PLLSs, especially at \( z > 4 \).

(A color version of this figure is available in the online journal.)

and we will report our results on \( \ell_{\tau \geq 2}(z) \) for these absorbers as upper limits, especially for \( z_{\text{abs}} > 4 \).

Before closing this section, we stress that the blending bias affects all previous and future LLS surveys. In particular, we caution that the incidence of LLSs cannot trivially be used to constrain the H\( \alpha \) frequency distribution \( f(N_{\text{H}1}, \lambda) \), because the latter assumes that every absorption system is a unique, identifiable line. For the LLSs in particular (but this is also true for the Ly\( \alpha \) forest), blending smears these lines over non-negligible redshift intervals (\( \delta z \approx 0.1 \)) and this effect must be considered when comparing against theoretical line densities.

7.5. Continuum Uncertainty

An important systematic uncertainty in our analysis is the placement of the absorbed quasar continuum. As described in Section 4, the continuum for each quasar spectrum was determined from an automated fit of a template model to the data. Each continuum was then reviewed by two authors (J.X.P. and J.M.O.) and frequently scaled up/down by 5%–10%. For our LLS survey, modifications to the continuum primarily modify the survey path; the estimates of \( t_{912} \) are only affected if \( t_{912} \approx 2 \).

To test the sensitivity of our results to continuum placement, we reanalyzed the data after scaling each continuum up/down by 5% and 10%. We also considered a scenario where the scaling was random between ±10%. We find that the estimates on the incidence of LLSs vary by 5%–10% as the continua are modified. The effect, while due to systematic changes in the continuum, is not systematic. That is, a systematic increase in the continuum does not systematically increase/decrease the incidence of LLSs at all redshifts. Therefore, we conclude that continuum placement errors yield a random, non-negligible (≈5%–10%) uncertainty in the final results.

7.6. A Mock SDSS Survey

As described in Section 7.1, we generated a second set of mock quasar spectra with the same emission redshifts and S/N as the SDSS spectra analyzed in our statistical sample. In this set, we restricted the sky-subtraction errors to typically be less than 5% but we did allow for a small set of spectra with significantly larger error. These mock spectra were analyzed by the lead author with the same algorithms and procedures used on the real SDSS data set.

In Figure 9, we present the cumulative number of LLS recovered by our analysis against the true number under the following conditions: (1) we combined the \( N_{\text{H}1} \) values of all the absorbers that lie within \( |\delta z| \leq 0.05 \) of each other, (2) we restrict the analysis to spectral regions with \( S/N_{t_{912}} \geq 2 \), and (3) we avoid the proximity region of each quasar. We find excellent agreement between the “true” number and recovered number at all redshifts. Our final result is an overestimate of ≈5% for the recovered LLS, dominated by sky-subtraction errors and well within the Poisson error. Lastly, we analyzed these data to recover an estimate of the incidence of LLS with redshift \( \ell_{\tau \geq 2}(z) \) and find that it matches the inputted incidence to within 1σ uncertainty. We conclude from this analysis that there is no major bias in our procedures.

7.7. Summary

We have conducted an assessment of the systematic uncertainty related to surveying LLSs using mock spectra with idealized Ly\( \alpha \) forest absorption yet realistic spectral characteristics
In this section, we present the principal results of our survey. We defer extended discussion of previous work and the implications of our analysis to the following section. Systematic biases and uncertainties in these results were discussed in the previous section and are summarized in Section 7.7.

8. RESULTS

In this section, we present the principal results of our survey. We defer extended discussion of previous work and the implications of our analysis to the following section. Systematic biases and uncertainties in these results were discussed in the previous section and are summarized in Section 7.7.

8.1. $\ell_{z>2}(z)$: The Incidence of Intervening $\tau_{912} \geq 2$ LLSs per Redshift Interval

An LLS survey, by its nature, provides only a single observable quantity: the incidence of LLSs per redshift interval $\ell_{z>2}(z)$. This quantity is independent of any assumed cosmology and consequently has limited physical meaning. Nevertheless, it is the proper starting point for describing our results.

Following standard practice, we estimate $\ell_{z>2}(z)$ from the ratio of $N_{LLS}$ to the total number of LLSs detected in a redshift interval ($\Delta z_{\text{TOT}}$) the total search path for that redshift interval:

$$\ell_{z>2}(z) = \frac{N_{LLS}}{\Delta z_{\text{TOT}}}. \quad (10)$$

The statistical error in $\ell_{z>2}(z)$ from this estimator is assumed to be dominated by the Poisson uncertainty in $N_{LLS}$. We have discussed a range of possible systematic uncertainties with this estimator in the previous section. Figure 10 presents the values of $\ell_{z>2}(z)$ for the $S/N_{912} = 2$ criterion in a set of arbitrary redshift intervals chosen to give $m_{LLS} \geq 30$ systems per bin. Table 4 lists these values for $S/N_{912}$ thresholds of 2 and 3; the results are in excellent agreement.

Figure 10 reveals that the incidence of $\tau_{912} \geq 2$ LLSs increases monotonically for $z > 3.5$. Following previous work, we have modeled the redshift evolution in $\ell_{z>2}(z)$ as a power law with the functional form

$$\ell_{z>2}(z) = C_{LLS} \left[ \frac{1 + z}{1 + z_{\text{em}}} \right]^{\gamma_{LLS}}. \quad (11)$$

setting $z_{\text{em}} \equiv 3.7$. Using standard maximum likelihood techniques (e.g., Storrie-Lombardi et al., 1994), we find best-fit values to the data at $z \geq 3.5$ of $C_{LLS} = 1.9 \pm 0.2$ and $\gamma_{LLS} = 5.1 \pm 1.5$ (68% c.l.). The best-fit model is overplotted on the data in Figure 10. The relatively large uncertainty in $\gamma_{LLS}$ is due to the small redshift interval covered by our survey. Nevertheless, we conclude at high confidence ($>95\%$) that $\ell_{z>2}(z)$ is increasing at least as steeply as $\gamma_{LLS} = 2$ at $z > 3.5$. If we restrict the analysis to spectra with $S/N_{912} \geq 3$, we find consistent results ($C_{LLS} = 1.9 \pm 0.2$ and $\gamma_{LLS} = 5.6 \pm 2.1$).

The subpanel of Figure 10 compares the cumulative number of LLSs detected in the survey against redshift both as observed (solid) and as predicted (dotted) by the best-fit power-law model. For the latter, we adopt the $g(z)$ curves for $S/N_{912} = 2$ from Figure 4. A one-sided Kolmogorov–Smirnov (K–S) test yields a probability $P_{\text{K-S}} = 0.95$ that the observed distribution is drawn from the adopted power-law expression; the power-law model is a good description of the observations. We comment, however, that the best-fit slope ($\gamma_{LLS} = 5.2$) is considerably

---

**Figure 9.** Comparison of the cumulative number of LLS recovered in our analysis of SDSS mock spectra (red curve) against the true number of LLS within the mock spectra (black curve). For this comparison, we imposed the following conditions: (1) we combined the $N_{LLS}$ values of all the absorbers that lie within $|\Delta z| \leq 0.05$ of each other, (2) we restrict the analysis to spectral regions with $S/N_{912} \geq 2$, and (3) we avoid the proximity region of each quasar. The good agreement between the two curves indicates that there is no major bias in our procedures.

**Figure 10.** Incidence of intervening LLSs with $\tau_{912} \geq 2$ as a function of redshift (solid points). Only the darker points were included in a maximum likelihood analysis to determine best-fit power law (blue curve): $\ell_{z>2}(z) = C_{LLS}[(1 + z)/(1 + z_{\text{em}})]^{\gamma_{LLS}}$, with $z_{\text{em}} \equiv 3.7$, $C_{LLS} = 1.9 \pm 0.2$, and $\gamma_{LLS} = 5.1 \pm 1.5$ (68% c.l.). The dotted points, meanwhile, show $\ell_{z>2}(z)$ when one includes quasars with $3.4 \leq z_{\text{em}} \leq 3.6$. These measurements are significantly biased to higher values by the SDSS quasar-targeting criteria (PWO09). In the subpanel, the black (solid) curve shows the cumulative number of $\tau_{912} \geq 2$ LLSs detected in our survey of the SDSS-DR7 database adopting the $S/N_{912} = 2$ threshold. The blue (dotted) curve shows the predicted number of LLSs assuming the best-fit power law from the figure and adopting the $g(z)$ function from Figure 4. A one-sided K-S test does not rule out the null hypothesis that the model distribution is statistically different from the observations.

(A color version of this figure is available in the online journal.)

(noise and resolution). Our analysis revealed an insidious bias associated with pLLSs that is best minimized by restricting the analysis to data with $S/N_{912} \geq 2$. We identified an unavoidable bias related to the blending of LLS and pLLS that implies an $\approx 10\%$ uncertainty in the measured incidence of LLSs. This bias becomes even more significant at $z > 4$ when the incidence of LLSs exceeds 3 per unit $\Delta z$. Finally, we found that continuum placement errors yield a random, non-negligible ($\approx 5\%$–$10\%$) uncertainty. Although higher S/N and spectral resolution will reduce some of these effects, we conclude that it will be difficult to avoid a systematic error of $10\%$–$20\%$ using the standard approaches to surveying LLSs. We believe that future programs will require new techniques if higher precision measurements are desirable.
steep than most previous estimates for the LLSs at this redshift (see Section 9.1) and also steeper than the redshift evolution measured for the Lyα forest and DLAs (Kim et al. 2001; Prochaska et al. 2008). It is our expectation that γz,LLS is likely lower than the central value of our analysis. This assertion will be tested with future observations at z < 3 and z > 4.5.

For z < 3.5, Figure 10 shows two evaluations of ℓz≥2(z). The light, solid points show the values of ℓz≥2(z) which values derived from our statistical quasar sample with the restriction that zem ≥ 3.6. These values are consistent with an extrapolation of the best-fit power law. The dotted points in the figure, meanwhile, show the values of ℓz≥2(z) when one also surveys quasars with 3.4 ≤ zem ≤ 3.6. In this case, we find systematically higher ℓz≥2(z) values which would indicate a non-physical, non-monotonic evolution in ℓz≥2(z). These results confirm the findings of PW09 that the SDSS targeting criteria for quasar spectroscopy biases the sample against sight lines without an LLS. The values of ℓz≥2(z) reported in Table 4, therefore, are restricted to quasars with zem ≥ 3.6.

8.2. The Incidence of LLSs in ΛCDM

If one introduces a cosmological model, the observed incidence of τ912 ≥ 2 LLSs with redshift ℓz≥2(z) may be translated into physical quantities. Consider first, ΔrLLS, the average distance that a photon travels before encountering an LLS with τ912 ≥ 2. Specifically, we define

\[ Δr_{LLS} ≡ \ell_{z≥2}(z)^{-1} \frac{dz}{d\ell} , \]  

where

\[ \frac{dz}{d\ell} = \frac{c}{(1 + z)H(z)} \]  

and

\[ H(z) = \frac{H_0[ΩΛ + (1 + z)^3Ω_m]^{1/2}}{ \left[ 1 + \frac{(1 + z)^3Ω_m}{1 + \frac{ΩΛ}{Ω_m}} \right] } . \]

With our adopted cosmology, we estimate that ΔrLLS ranges from ≈100 to 40 h⁻¹ Mpc proper distance from z = 3.5 to z = 4.4 (Table 4). This is an order of magnitude or more larger than the separation of high-z quasars (several Mpc for L_B ≳ 10^{40} erg s⁻¹ Hz⁻¹; Faucher-Giguere et al. 2009).

An especially informative quantity for associating LLSs to structures in the universe (e.g., galaxies and filaments) is ℓ(λ) the number of systems per absorption length (Bahcall & Peebles 1969), where ℓ(λ)dℓ = ℓ(λ) dz and

\[ dX = \frac{H_0}{H(z)} (1 + z)^2 \frac{dz}{d\ell} . \]

The quantity ℓz≥2(z) is defined to remain constant if nLLS, the comoving number density of structures giving rise to a τ912 ≥ 2, times αLLS, the average physical size of the structure remains constant, i.e., ℓz≥2(z) ∝ nLLSαLLS. Figure 11 presents the evolution of ℓz≥2(z) for our cosmology as a function of redshift (see also Table 4). We observe a rise in ℓz≥2(z) with redshift of roughly two times over the ≈1 Gyr from z = 3.3 to 4.4. At 99% confidence, we infer an increase in ℓz≥2(z) over this redshift interval. This follows, of course, from the very steep redshift evolution observed for ℓz≥2(z) (Section 8.1); in a flat cosmology with Ωm on the order of Ωm, an ℓz≥2(z) evolution steeper than (1 + z)¹/² implies ℓz≥2(z) is also increasing. We conclude that nLLS and/or αLLS are increasing with redshift at z ≳ 3.5. We discuss the implications of this result in Section 9.4.

8.3. N_HI Frequency Distribution (f(NHI, ξ)) at z ≲ 3.7

In this subsection, we combine our results with previous work on the IGM to place constraints on the HI frequency distribution, f(NHI, ξ). We focus this analysis at a single redshift (z = 3.7) where our observations have greatest statistical power.

8.3.1. fLLS(NHI, ξ)

Although our observations and LLS analysis are insensitive to the HI column densities of the LLSs, they do provide an integral constraint on the frequency distribution of NHI, f(NHI, ξ). We constrain the NHI frequency distribution of τ912 ≥ 10 LLS per absorption length, fLLS(NHI, ξ), at column densities NHI = 10^{17.5}–10^{19} cm⁻² as follows. Previous surveys at z > 3 have measured f(NHI, ξ) for column densities NHI ≥ 10^{19} cm⁻² (Prochaska et al. 2005; O’Meara et al. 2007; Prochaska & Wolfe 2009; Noterdaeme et al. 2009; Guimaraes et al. 2009). These authors have parameterized the distribution functions as single
(SLLs) and double (DLAs) power laws of the following form:

\[ f_{\text{SLLS}}(10^{19} \text{ cm}^{-2}) \leq N_{\text{HI}} < 10^{20.3} \text{ cm}^{-2}, X) = k_{\text{SLLS}} N_{\text{HI}}^{\beta_{\text{SLLS}}} \]

and

\[ f_{\text{DLA}}(N_{\text{HI}} \geq 10^{20.3} \text{ cm}^{-2}, X) = k_{\text{DLA}} \left( \frac{N_{\text{HI}}}{N_d} \right)^{\beta_{\text{DLA}}} \]

where \( \beta_{\text{DLA}} = \{\beta_3 : N_{\text{HI}} < N_d, \beta_4 : N_{\text{HI}} \geq N_d\}. \]

Figure 12 presents these frequency distributions. For the SLLs at \( z = 3.7 \), we have taken \( \beta_{\text{SLLS}} = -1.2 \pm 0.2 \) and normalized the power law by taking

\[ \ell_{\text{SLLS}}(X) = \int_{10^{19}}^{10^{20.3}} f_{\text{SLLS}}(N_{\text{HI}}, X) dN_{\text{HI}} = 0.2. \]

These values are consistent with the range of published measurements at this redshift (Péroux et al. 2005; O’Meara et al. 2007; Guimaraes et al. 2009). For the DLAs, we have evaluated \( f_{\text{DLA}}(N_{\text{HI}}, X) \) from the SDSS-DR5 (Prochaska & Wolfe 2009) over the redshift interval \( z = [3.4, 4.0] \), giving \( N_d = 10^{21.75}, \beta_3 = -1.8, \beta_4 < -3, \) and \( k_{\text{DLA}} = 7 \times 10^{-25} \text{ cm}^2. \)

We estimate \( f(N_{\text{HI}}, X) \) for the interval \( N_{\text{HI}} = [10^{17.5}, 10^{19}] \) \text{ cm}^{-2}, \) which we refer to as \( f_{\text{SLLS}}(N_{\text{HI}}, X) \), under the following assumptions/constraints: (1) \( f_{\text{SLLS}}(N_{\text{HI}}, X) \) has a power law form

\[ f_{\text{SLLS}}(N_{\text{HI}}, X) = k_{\text{SLLS}} N_{\text{HI}}^{\beta_{\text{SLLS}}}, \]

and (2) \( f_{\text{SLLS}}(N_{\text{HI}}, X) \) at \( N_{\text{HI}} = 10^{19} \) \text{ cm}^{-2} is consistent with the range of values given by the SLLSs. Specifically, we demand

\[ \log f(N_{\text{HI}} = 10^{19} \text{ cm}^{-2}, X) = -20.05 \pm 0.2; \]

and

\[ \ell_{\tau \geq 2}(X) = \int_{10^{17.5}}^{10^{19}} f(N_{\text{HI}}, X) dN_{\text{HI}}. \]

At \( z \approx 3.7 \), we estimate \( \ell_{\tau \geq 2}(X) = 0.5 \pm 0.1 \) (Figure 11).

Overplotted on Figure 12 are the power-law frequency distributions (shown as dashed, dotted, and a solid line) that satisfy the extrema of those constraints. The shaded region shows the intersection of the curves and roughly represents the allowed region of \( f(N_{\text{HI}}, X) \) values. We find \( \beta_{\text{SLLS}} = -0.8 \pm 0.3, \) and derive \( k_{\text{SLLS}} = 10^{-4.5} \text{ cm}^2 \) for the central value. Table 5 further summarizes these results.

Our analysis reveals that \( f(N_{\text{HI}}, X) \) becomes increasingly shallow with decreasing \( N_{\text{HI}} \). Only for the most extreme values of our analysis, low \( f(N_{\text{HI}}, X) \) at \( N_{\text{HI}} = 10^{19} \text{ cm}^{-2} \) and a large \( \ell_{\tau \geq 2}(z) \) value, do we recover \( \beta_{\text{SLLS}} < -1. \) This flattening of \( f(N_{\text{HI}}, X) \) was suggested by previous authors based on a similar analysis but with much poorer observational constraints on \( \ell_{\tau \geq 2}(X) \) (Péroux et al. 2003; Prochaska et al. 2005; O’Meara et al. 2007). Remarkably, our results indicate \( \beta_{\text{SLLS}} > -1 \) which means that the IGM has a higher total covering fraction per unit pathlength for sight lines with \( N_{\text{HI}} = 10^{19} \text{ cm}^{-2} \) than those with \( N_{\text{HI}} = 10^{18} \text{ cm}^{-2}. \)

8.3.2. Constraints from Measurements of the Mean Free Path

Traditionally, the mean free path to ionizing radiation in the IGM (\( \lambda_{\text{mfp}} \)) has been estimated from the observed incidence of LLSs (e.g., Meiksin & Madau 1993). Recently, PW009 introduced a new approach to measure \( \lambda_{\text{mfp}} \) from stacked quasar spectra, without any consideration of LLSs. We can reverse the problem, therefore, and use the \( \lambda_{\text{mfp}} \) results to constrain properties of \( f(N_{\text{HI}}, X) \). One expects to have the greatest
Table 5
Summary of f(N(H\textsubscript{i}), X) Results

| ε\textsubscript{LLS}(z)\textsuperscript{a} | log f\textsubscript{10}\textsuperscript{b} | β\textsubscript{LLS} | log N\textsubscript{pLLS}\textsuperscript{c} | β\textsubscript{pLLS} |
|----------------|----------------|-------------|----------------|-------------|
| Preferred values |
| 0.23 | −20.05 | −0.8 | −4.5 | 17.3 | −1.9 |
| Conservative range of allowed values |
| 0.23 | −20.05 | −0.8 | −4.5 | 17.1 | −2.0 |
| | | | | 17.3 | −1.9 |
| | | | | 17.5 | −1.9 |
| 0.15 | −20.25 | −0.9 | −4.1 | 17.5 | −1.9 |
| 0.35 | −20.25 | −1.3 | 4.7 | 15.0 | −3.5 |
| | | | | 15.2 | −3.0 |
| | | | | 15.4 | −2.6 |
| 0.15 | −19.85 | −0.1 | −18.0 | ... | ... |
| 0.35 | −19.85 | −0.8 | −4.6 | 15.2 | −5.2 |
| | | | | 15.4 | −4.3 |
| | | | | 15.5 | −3.7 |
| | | | | 15.7 | −3.3 |
| | | | | 15.9 | −3.0 |
| | | | | 16.1 | −2.7 |
| | | | | 16.2 | −2.5 |
| | | | | 16.4 | −2.4 |
| | | | | 16.6 | −2.2 |
| | | | | 16.8 | −2.1 |
| | | | | 17.0 | −2.0 |
| | | | | 17.1 | −1.9 |
| | | | | 17.3 | −1.9 |

Notes. The analysis throughout assumes that the DLAs contribute $\ell(X) = 0.09$ to $\ell_{\tau > 2}(X)$.

* The incidence of $\tau < 10^{-2}$, attributed to LLSs with $N(\text{HI}) < 10^{19}$ cm$^{-2}$.

The adopted value of $f(N(\text{HI}), X)$ at $N(\text{HI}) = 10^{20}$ cm$^{-2}$.

The break column density within the Ly\textalpha forest as defined in the text. Entries without values have power-law descriptions for the LLS that cannot satisfy the mean free path and Ly\textalpha forest constraints.

sensitivity to absorption systems with $\tau < 10^{-2}$, i.e., the LLSs and pLLSs.

At $z = 3.7$, PWO09 estimate $\lambda_{\text{mfp}} = 47 h^{-1}$ Mpc proper distance. This means that in the absence of an expanding universe, a packet of 1 Ryd photons at $z = 3.7$ would be attenuated by $\exp(-1)$ after traveling $\lambda_{\text{mfp}}$. One can also express the mean free path as an opacity, $\kappa_{\text{mfp}} = 1/\lambda_{\text{mfp}}$, which can be related to the optical depth of a 1 Ryd photon as

$$\kappa_{\text{mfp}} = \frac{d\tau_{\text{mfp}}}{d\tau} = \frac{d\tau_{\text{mfp}}}{d\tau} \frac{dz}{dr}. \quad (21)$$

Finally, we can relate the differential optical depth to the H$\text{\textsubscript{i}}$ frequency distribution of absorbers:

$$\frac{d\tau_{\text{mfp}}}{dz} = \int_{N_{\text{HI}}^{\text{min}}}^{N_{\text{HI}}} f(N(\text{HI}), X) [1 - \exp(-N(\text{HI}, \sigma_{\text{ph}}))] dN(\text{HI}) \quad (22)$$

with $\sigma_{\text{ph}}$ the photoionization cross section evaluated at 1 Ryd and $dr/dz$ given by Equation (13). Although the integral should be evaluated with $N_{\text{HI}} = 0$, in practice $d\tau_{\text{mfp}}/dz$ is insensitive to the minimum $N_{\text{HI}}$ column density for any value $N_{\text{HI}}^{\text{min}} < 10^{12}$ cm$^{-2}$.

Figure 13 shows the $\kappa_{\text{mfp}}$ value at $z = 3.7$ from PWO09 as a horizontal band that illustrates the 1σ error interval. The solid curve, meanwhile, corresponds to the evaluation of Equation (21) using our best estimation of $f(N(\text{HI}, X)$ (Figure 12) as a cumulative function of $N(\text{HI})$. At the limiting $N(\text{HI})$ value of our LLS survey ($10^{17.5}$ cm$^{-2}$), we estimate that $\approx 55\%$ of the opacity to ionizing radiation is contributed by $\tau > 10^{-2}$. The uncertainty in the results is roughly proportional to the uncertainty in $f(\tau_{\geq 2}(X)$, i.e., $\approx 20\%$ as indicated by the error bars on the figure. It is notable that $\approx 1/3$ of the contribution to $\kappa_{\text{mfp}}$ is from very optically thick absorbers ($\tau > 10$), i.e., the SLLSs and DLAs.

It is also evident from Figure 13 that systems with $\tau_{\geq 2} < 10^{-2}$ must contribute to $\kappa_{\text{mfp}}$. For $N_{\text{HI}}^{\text{min}} < 10^{17.5}$ cm$^{-2}$, we continue the calculation by assuming that $f(N(\text{HI}, X)$ follows a power law

$$f_{\text{pLLS}}(N(\text{HI}) = 10^{17.5} \text{ cm}^{-2}, X) = k_{\text{pLLS}}N_{\text{HI}}^{\beta_{\text{pLLS}}} \quad (23)$$

constrained to match $f_{\text{LLS}}(N(\text{HI}, X)$ at $N(\text{HI}) = 10^{17.5}$ cm$^{-2}$. We find that models with $\beta_{\text{pLLS}} > 1.8$ cannot reproduce the $\lambda_{\text{mfp}}$ results. In fact, the data favor $\beta_{\text{pLLS}} < 1.8$, i.e., a much steeper power law than inferred for the LLS and also that that commonly observed for the Ly\textalpha forest. These conclusions depend rather insensitively on our estimate of $\tau_{\geq 2}(X)$; slopes only as shallow as $\approx 1.7$ are allowed if we adopt our highest estimates for $\tau_{\geq 2}(X)$.

Thus far, these inferences on $f(N(\text{HI}, X)$ for absorption systems with $\tau_{\geq 2} < 2$ have ignored observations of the Ly\textalpha forest. By including these data, we provide further constraints on $f(N(\text{HI}, X)$ for $N_{\text{HI}} = 10^{15} - 10^{19}$ cm$^{-2}$. To derive these constraints, however, we must adopt a functional form for $f(N(\text{HI}, X)$. Absent a physical model, we take an empirical approach. We express $f(N_{\text{HI}}, X)$ as a series of six power laws that intersect at $N_{\text{HI}} = [10^{14.5}, N_{\text{pLLS}}, 10^{19.0}, 10^{20.3}, 10^{21.75}] \text{ cm}^{-2}$, where $N_{\text{pLLS}}$ is constrained to lie between $N_{\text{HI}} = 10^{15} - 10^{17.5}$ cm$^{-2}$. Other than a small “kink” at $N_{\text{HI}} = 10^{20.3}$ cm$^{-2}$, the power laws are required to match at each intersection point.

Figure 13. Opacity at the Lyman limit $\kappa_{\text{mfp}}$ contributed by absorbers with $N_{\text{HI}} > N_{\text{HI}}^{\text{min}}$. For $N_{\text{HI}}^{\text{min}} > 10^{17.5}$ cm$^{-2}$ (solid curve), which corresponds to our LLS survey, we have adopted the estimate of $f(N(\text{HI}, X)$ from Figure 12 in the calculation. We estimate a 20% uncertainty in the contribution of LLSs to $\kappa_{\text{mfp}}$, as shown in the figure. For $N_{\text{HI}}^{\text{min}} < 10^{17.5}$ cm$^{-2}$, we assume $f(N(\text{HI}, X)$ follows a simple power law with exponent $\beta_{\text{pLLS}}$, and show a series of extrapolations (dashed and dotted curves). The solid (red) horizontal band centered at $\kappa_{\text{mfp}} = 0.0225 2172 \text{ Mpc}^{-2}$ indicates the measurement at $z \approx 3.7$ by PWO09. These results imply that LLSs contribute $\approx 55\%$ of the mean free path ($\approx 33\%$ for systems with $\tau_{\geq 2} > 1$) and that $N_{\text{pLLS}}$ must be steeper than $\approx 1.5$ to explain all of these observations.

(A color version of this figure is available in the online journal.)
The power laws are forced to satisfy the following observational constraints.

1. The power laws for \( N_{\text{HI}} \geq 10^{17.5} \) are constrained as described at the start of this subsection.

2. For the Ly\( \alpha \) forest, we assume \( f(N_{\text{HI}}, X) \propto N_{\text{HI}}^{-1.5} \) and normalize at \( z = 3.7 \) by the effective Ly\( \alpha \) optical depth \( \tau_{\text{eff}} \) measured by Faucher-Giguère et al. (2008b) assuming a \( b \)-value distribution \( f(b) \propto b^{-5} \exp[-b^4/b^4] \) (Hui & Rutledge 1999).

3. The integrated opacity of the IGM at the Lyman limit is constrained by the measurement of PWO09, i.e., \( N_{\text{HI}} = 47 h^{-1} Mpc \).

For the range of power laws derived from our analysis of the LLS results (Figure 12), we show in Figure 14 the range of \( f(N_{\text{HI}}, X) \) distributions that also satisfy all of the constraints. We find viable models with \( N_{\text{pLLS}} \) values that range across the allowed interval. These are correlated with \( \beta_{\text{pLLS}} \) values ranging from \( \beta_{\text{pLLS}} = -1.9 \) to \( -5 \) (Table 5).

The principal results of this analysis are threefold. First, the single power law connecting the Ly\( \alpha \) forest to the SLLS satisfies neither the LLS nor mean free path constraints. There is at least one break between \( N_{\text{HI}} = 10^{14.5} \) cm\(^{-2} \) and \( N_{\text{HI}} = 10^{19} \) cm\(^{-2} \) where \( f(N_{\text{HI}}, X) \) steepens to \( \beta < -1.8 \) and then flattens to \( \beta \approx -1 \). This is consistent with conclusions drawn from line-counting statistics of Ly\( \alpha \) forest lines (e.g., Petitjean et al. 1993; Kim et al. 2002). Second, the added \( \chi_{\text{pLLS}}^2 \) and Ly\( \alpha \) forest constraints rule out the lowest values of \( f_{\text{SLLS}}(N_{\text{HI}}, X) \) at \( N_{\text{HI}} \approx 10^{17.5} \) cm\(^{-2} \) that were otherwise allowed by our LLS results. Specifically, the data require \( \log f(N_{\text{HI}} = 10^{17} \) cm\(^{-2} \), \( X \) \( \geq -19 \) and the (blue) shaded region in Figure 14 shows the proper allowed range for \( f_{\text{SLLS}}(N_{\text{HI}}, X) \). Finally, we find that the slope of \( f(N_{\text{HI}}, X) \) must steepen to \( \beta_{\text{SLLS}} \leq -1.8 \) at columns \( N_{\text{HI}} \geq 10^{14.5} \) cm\(^{-2} \). We note that these results are largely independent of our LLS survey because the other observations have higher statistical significance.

### 8.4. The Incidence of Proximate LLSs (PLLs)

The results presented thus far all refer to intervening LLSs, i.e., systems restricted to have \( z_{\text{LLS}} \) blueward of 3000 km s\(^{-1} \) from the quasar emission redshift. This restriction was imposed to isolate the “ambient” IGM and avoid biases related to having performed the search for bright, background quasars. In the space surrounding a bright quasar, one predicts at least two such biases: (1) bright, high-\( z \) quasars are known to cluster strongly (\( r_b > 15 h^{-1} \) Mpc; Shen et al. 2007) suggesting these objects trace massive structures in the young universe. The local environment of bright quasars, therefore, has an uncommonly high density (at least in dark matter) which may give a higher incidence of LLSs; (2) the radiation field of the quasar will ionize gas to large distances, reducing the incidence of LLSs. For DLAs, the first effect dominates as one observes an enhanced rate of proximate DLAs (PDLAs) relative to the intervening systems (Russell et al. 2006; Prochaska et al. 2008).

Following the formalism presented in Prochaska et al. (2008), we have estimated the incidence of PLLs in a series of redshift intervals. First, we re-measured the quasar emission redshifts for all systems with \( S/N_{912} \geq 2 \) at the Lyman limit and with a \( t_{912} \geq 2 \) LLS within 5000 km s\(^{-1} \) of \( z_{\text{em}} \). The SDSS quasar redshifts reported in the standard DR7 data release are known to have significant systematic errors. Following the prescriptions described in Shen et al. (2007), J. Hennawi has kindly re-measured the redshifts for all of these quasars. These have typical uncertainties of \( \approx 200 \) km s\(^{-1} \) and believed to have no large systematic bias. Second, we analyzed the quasars whose absorbed continuum at the wavelength of the Lyman limit corresponding to 3000 km s\(^{-1} \) blueward of \( z_{\text{em}} \) is twice the median-smoothed, 1σ error array. This establishes the survey path. All PLLs identified redward of this 3000 km s\(^{-1} \) offset form the statistical sample (Table 6). The incidence, \( \epsilon_{\text{PLL}}(z) \), is then estimated in arbitrary redshift intervals assuming the same estimator for intervening LLSs (Equation (10)). These results are presented in Figure 15 and compared against the incidence of intervening LLSs.

Ignoring the data at \( z < 3.6 \) (which we suspect to be biased high by the SDSS targeting criteria; PWO09), the incidence of PLLs roughly tracks that of intervening LLSs but is \( \approx 25\% \) lower than the intervening systems. The inset figure shows the observed cumulative number of PLLs (dark curve) versus the predicted number (light curve) assuming the best-fit power law for \( \epsilon_{\text{PLL}}(z) \) of intervening LLSs. A one-sided K-S test yields only a 1% probability that the two distributions are drawn from the same parent population. We also remind the reader that corrections for the blending bias described in Section 7.4 will likely reduce \( \epsilon_{\text{PLL}}(z) \) further, especially at \( z \sim 4 \). The principal implication is that \( t_{912} \geq 2 \) LLSs toward \( i < 20 \) mag quasars at \( z > 3.5 \) suffer from a proximity effect, presumably due to the ionizing radiation field of the quasar itself. Given the observed enhancement of strong LLSs at \( z \approx z_{\text{em}} \) along sight lines transverse to such quasars, our results lend further evidence that quasar emission at \( z \approx 1 \) Ryd is usually anisotropic (Hennawi & Prochaska 2007).

Before concluding this section, we comment that the PLLs analysis is subject to another systematic error. In performing

---

11 Unless the emission redshifts are systematically offset by a large amount (greater than several hundred km s\(^{-1} \)), this will not affect the results.
follows that of intervening systems but is systematically lower by万千瓦." Ignoring that last point, we find that the incidence of PLLSs roughly curve) against the predicted number using the power-law model for $\ell_{\tau}$.

The inset figure shows the cumulative number of PLLSs observed (black $z<\theta$ with very strong associated absorption, e.g., BALs. The intent

our LLS survey, we have identified and removed all quasars with very strong associated absorption, e.g., BALs. The intent

of this procedure was to remove the signatures of absorption from gas very local ($<1$ kpc) to the quasar from the analysis. It is possible, however, that the associated absorption in some of these removed quasars is due to a PLLS at distances $\gg1$ kpc and not very local gas. This would lead to an underestimate of $\ell_{\text{PLLs}}(z)$. Alternatively, we may not have identified all of the local absorbers and therefore might have overestimated $\ell_{\text{PLLs}}(z)$. In either case, we caution that a systematic error of the order of 10%–20% should be attributed to this effect.

9. DISCUSSION

9.1. Comparisons with Previous Work

Surveys for LLSs have been carried out for several decades now (Tytler 1982; Sargent et al. 1989; Lanzetta 1991; Storrie-Lombardi et al. 1994; Stengler-Larrea et al. 1995; Péroux et al. 2003). These have been performed primarily at optical wavelengths on heterogeneous quasar samples drawn from a diverse set of survey approaches: color-selection, radio detection, slitless spectroscopy, etc. The authors adopted differing completeness limits for $z_{912}$ (ranging from 1 to 3) and used different approaches to establishing the pathlength that establishes $\ell_{\text{PLLs}}(z)$. Little attention was given to assessing systematic error, and several of the effects described in Section 7 assuredly apply to the previous works. On the positive side, the previous data were generally slit spectra of bright quasars and several of the data sets have higher S/N than typical of the SDSS spectra analyzed in this paper. The higher S/N helps to mitigate several of the potential biases we have discussed in Section 7.

In Figure 16, we present estimates of $\ell(z)$ from several previous studies, parameterized as power laws with the parameters listed in Table 7. The dotted lines show estimates from Sargent et al. (1989) and Lanzetta (1991). The dashed curve is the estimate from Stengler-Larrea et al. (1995) and the solid blue and green curves are from Storrie-Lombardi et al. (1994) and Péroux et al. (2003), respectively. For the values of our survey, we have increased the $\ell_{\text{PLLs}}(z)$ results by 10% to match the $z_{912}=1$ threshold of the previous work. Our results indicate a significantly lower incidence of LLSs at $z<4$ than suggested by the previous estimates. This lower incidence is consistent with recent estimates of the mean free path to ionizing radiation (POW09).

(A color version of this figure is available in the online journal.)

Table 6

SDSS-DR7 Proximate $z_{912} \geq 2$ LLS Survey

| Quasar                  | $z_{\text{em}}$ | $z_{\text{LLS}}$ |
|-------------------------|-----------------|-----------------|
| J001115.23+144601.8     | 4.967           | ...             |
| J001820.71+141851.5     | 3.936           | ...             |
| J004219.74+102009.4      | 3.880           | ...             |
| J004240.65+141529.6      | 3.687           | 3.684           |
| J010619.24+044823.3      | 4.449           | ...             |
| J011351.96-035511.0      | 3.668           | ...             |
| J012403.77+044327.2      | 3.834           | ...             |
| J014049.18-083942.5      | 3.704           | 3.693           |
| J015048.82+044126.2      | 3.702           | ...             |
| J015339.61-01104.8       | 4.194           | ...             |
| J021318.98-090458.3      | 3.794           | 3.797           |
| J022518.35-011332.2      | 3.628           | ...             |
| J024447.78-081606.1      | 4.068           | ...             |
| J025518.58+044847.6      | 3.966           | ...             |
| J031213.97-062658.8      | 4.031           | ...             |
| J034402.85-065300.6      | 3.957           | ...             |
| J073149.50+285448.6      | 3.676           | ...             |
| J074154.59+342521.2      | 3.905           | ...             |
| J074500.47+341731.1      | 3.713           | ...             |
| J074640.16+344624.7      | 4.010           | ...             |
| J074711.15+273903.3      | 4.154           | ...             |
| J075006.62+491834.1      | 3.603           | ...             |
| J075017.49+405825.3      | 3.864           | 3.849           |
| J075103.95+424211.6      | 4.163           | ...             |
| J075347.41+218058.2      | 4.031           | ...             |
| J075552.41+134551.1      | 3.673           | ...             |
| J075732.89+441424.6      | 4.170           | ...             |

(This table is available in its entirety in a machine-readable form in the online journal. A portion is shown here for guidance regarding its form and content.)
Stengler-Larrea et al. (1995), who integrated previous work with a new measurement at $z < 1$ and a still unpublished survey by C. C. Steidel & W. L. W. Sargent. Finally, the solid curves show the results from Storrie-Lombardi et al. (1994) and Péroux et al. (2003) who surveyed LLSs at $z \sim 4$ using color-selected quasars. All of these analyses were claimed to correspond to the incidence of LLSs with $\tau_{912} \geq 1$, although a careful review of the literature raises doubts regarding this assertion. Nevertheless, to make comparisons with their reported $\tau_{912} \geq 1$ results we have boosted each of our $\ell_{\tau \geq 1}(z)$ estimations. Formally, we estimate a correction of $6\%$ from our derived $f(N_{\text{H}_1}, X)$ distribution but, in practice, we adopt a more conservative $10\%$ correction.

Our results on $\ell_{\tau \geq 1}(z)$ at $z > 4$ are lower than the values derived from the Automated Plate Measuring Machine surveys, but within $\pm 1\sigma$ of concordance. The more important differences are between the estimations at $z < 4$ which include the bulk of our SDSS statistical sample. All of the previous work was essentially derived from the surveys of Sargent et al. (1989) and Lanzetta (1991) and, therefore, the curves all intersect at $z \approx 3$ with $\ell_{\tau \geq 1}(z) \approx 2$. If our results are correct, they suggest that some of the previous work may have overestimated the incidence of LLSs at $z \sim 3$. The original survey by Sargent et al. (1989) is in fair agreement with an extrapolation of our power-law form for $\ell_{\tau \geq 1}(z)$ to $z = 3$, but the reanalysis by Stengler-Larrea et al. (1995) of unpublished spectra taken by Steidel & Sargent apparently yields a higher estimate at this redshift. We suggest that this later work gave $\ell_{\tau \geq 1}(z)$ values that were too high either because of sample variance (i.e., small number statistics), selection bias in the quasar sample (we note, however, that these quasars are not color selected), and/or the systematic effects described in Section 7.

We have also considered whether our SDSS survey has been biased low by an unidentified systematic error. To this end, we performed an a posteri reanalysis of all spectra that contributed to the measurement of $\ell_{\text{LLS}}(z)$ in the $z = 3.5$ to $z = 3.65$ interval. Specifically, we searched for any missed LLS or obvious cases of poor sky subtraction. We found no cases that warranted us to overturn our initial assessment. We also note that the sample of spectra with $S/N_{\lambda_{912}} > 3$ shows an even lower incidence of LLS at these redshifts (Table 4).

Finally, and perhaps most important, our independent measurement\footnote{Based on the analysis of stacked SDSS spectra including thousands of quasars not included in the LLS survey.} of $\lambda_{912}^{\text{SDSS}}$ (Prochaska et al. 2009) argues for the low incidence of LLSs recovered in this paper. In Section 8.3.2 (Figure 13), we showed that if one were to adopt a $50\%$ higher incidence of $\tau_{912} \geq 2$ LLSs, e.g., $\ell_{\tau \geq 2}(X) = 0.75$ at $z = 3.7$, then we would infer a much steeper slope for the LLSs ($\beta_{\text{LLS}} \approx -1.5$) and also a power law shallower than $\beta = -0.5$ for $N_{\text{H}_1} = 10^{15}-10^{17}$ cm$^{-2}$. This would force $f(N_{\text{H}_1}, X)$ to steepen to $\beta < -3$ at $N_{\text{H}_1} \approx 10^{14.5}$ cm$^{-2}$. Such an extreme $f(N_{\text{H}_1}, X)$ distribution is non-physical and, more importantly, ruled out by observed line statistics of the Ly$\alpha$ forest (Kim et al. 2002; Misawa et al. 2007). We conclude that the incidence of LLSs at $z \approx 3.7$ cannot be more than $30\%$ higher than our central value, and, at present, we cannot identify a bias that would lead to such a large systematic underestimate. We also note that even though the $\lambda_{\text{mpf}}$ analysis does not extend below $z = 3.6$, the observed evolution in the $\lambda_{\text{mpf}}$ is sufficiently steep that it is likely to apply at lower redshifts (see also Worseck & Prochaska 2010).

9.2. Evolution in $\ell_{\tau \geq 1}(z)$ from $z = 0$ to $4$

Previous work has debated whether $\ell_{\tau \geq 1}(z)$ evolves as a single-power-law $(1 + z)^{\gamma_{\text{LLS}}}$ from $z \approx 0$ to $4$ (e.g., Stengler-Larrea et al. 1995). In Section 8.1, we modeled our observations with a single power law having $\gamma_{\text{LLS}} = 5.1 \pm 1.5$. The observed evolution in the mean free path (PW009) also suggests a steep evolution ($\gamma_{\text{LLS}} > 2$) for the LLSs. We now consider whether a single power-law extrapolation is a good description of $\ell_{\tau \geq 1}(z)$ for $z < 3$.

Survey of other H$\text{I}$ absorption systems have demonstrated that a single $(1 + z)^{\gamma}$ power law is a poor description of the Ly$\alpha$ forest (Weymann et al. 1998) and the DLAs (Prochaska et al. 2008). In the former case, one observes a flattening in the Ly$\alpha$ forest density at $z \sim 1$ which has been interpreted to result from a corresponding decline in the intensity of the extragalactic UV background (EUVB; Weymann et al. 1998; Davé et al. 1999). It is plausible that a similar effect would influence the LLS. In Figure 17, we present estimates of $\ell_{\tau \geq 1}(z)$ at $z \sim 3.5$, estimated by increasing the measured $\ell_{\tau \geq 2}(z)$ values by $10\%$. These are compared against the $z \sim 1$ measurement from Stengler-Larrea et al. (1995; see J. Ribaudulo et al. 2010, in preparation for a new estimate). Overplotted on the data is a solid curve that shows the best-fit power law to the SDSS results. The dashed curves show $2\sigma$ departures from this model, extrapolated to $z = 0$. It is evident that none of these curves intersect the low redshift observations. We conclude, with high confidence, that a strict
(1+z)LLS power law does not describe the evolution of $\tau_{Ly\alpha}^{\geq 1}$ (z) from z = 0 to 4. Instead, the data suggest a "break" in the high-z power law at z ~ 2, similar to that observed for the Lyα forest although at somewhat higher redshift.

The dotted curve shows an attempt to model this break. For z < 2.3, we adopt the power-law form that matches the Lyα forest low redshift ($\gamma_{LLS} = 0.26$) and demand that it intersect the central value of the Stengler-Larrea et al. (1995) measurement. For z ≥ 2.3, the model breaks to a $\gamma_{LLS} = 2.78$ power law, again consistent with the high-z evolution of the Lyα forest (e.g., Kim et al. 2002). This is a reasonably good description of the data and we conclude that a break in the power-law description of $\ell_{Ly\alpha}(z)$ likely occurs at z ~ 2. We will test this prediction with an (ongoing) survey for LLSs at z ~ 2 in Hubble Space Telescope (HST)/Advanced Camera for Surveys (ACS) and WFC3 slitless spectra of z ~ 2.3 quasars (PI: O'Meara).

### 9.3. The Average LLS Spectrum

To gain additional insight into the absorption properties of the LLSs, we have constructed an average (stacked) spectrum by (1) shifting each quasar spectrum containing a $\tau_{912} \geq 2$ LLSs to its rest frame (190 systems total) and (2) averaging the fluxed data. This stacked spectrum is primarily illustrative; it is shown in Figure 18. The peak in emission at $\lambda \approx 1280$ Å is from the Lyα emission peak of the background quasars. The peak is offset from 1215 Å because the stack only includes intervening LLSs, i.e., those that are offset by at least 3000 km s$^{-1}$ from the quasar emission redshift. The peak’s proximity to 1215 Å and relatively narrow width, however, reflect that most LLSs in our survey are located within $\delta z = 0.2$ of the quasar emission redshift.

The second strongest feature in the spectrum is the Lyman limit at the expected wavelength of 912 Å. Shortward of the Lyman limit, one observes a non-zero flux that extends down to ≈770 Å. We estimate the average optical depth of the Lyman limit absorption in this stacked spectrum by assuming the flux at 980 Å provides a rough estimate of the absorbed continuum at the Lyman limit and measure

$$\bar{\tau}_{912} = -\ln \left[ \frac{f(990 \, \text{Å})}{f(980 \, \text{Å})} \right] = 3.5 \, ,$$

(24)

We estimate a 20% error in this value due to effects related to sky subtraction, uncertainty in the absorbed continuum $f(980 \, \text{Å})$, and the flux-weighted average of our stack.

The value of $\bar{\tau}_{912}$ may be compared to the average optical depth derived from our $f(N_{H_i}, X)$ distribution:

$$\bar{\tau}_{912} = -\ln \left( \frac{\int \exp[-N_{H_i} \tau_{LLS} f(N_{H_i}, X)] dN_{H_i}}{\int f(N_{H_i}, X) dN_{H_i}} \right) \, ,$$

(25)

where the integrals are evaluated over the interval $N_{H_i} = [10^{17.5}, 10^{22}]$ cm$^{-2}$. Evaluating at z = 3.7 using the $f(N_{H_i}, X)$ distribution function shown in Figure 12, we derive $\tau_{912} \approx 5.4$. The frequency distribution in the LLS regime is sufficiently flat that the higher $N_{H_i}$ systems (SLLS and DLAs) contribute significantly to the average.

The offset between these two evaluations is most likely due to the inclusion of a non-negligible number of systems having $\tau_{912} < 2$. Our analysis of mock spectra and our internal comparison of the $\tau_{912}$ estimates for the LLS indicate that this occurs frequently. As described in Section 5, we estimate a 0.2 dex uncertainty in the $\bar{\tau}_{912}$ values of LLSs with $\tau_{912} \approx 2$. Although our tests also suggest this does not significantly affect the estimate of $\ell_{Ly\alpha}(z)$, it can have a significant effect on the $\tau_{912}$ value in the stacked spectrum. We have repeated our calculation of $\tau_{912}$ extending the lower limit of the $N_{H_i}$ distribution to $N_{H_i} = 10^{17.2}$ cm$^{-2}$ instead of $N_{H_i} = 10^{17.5}$ cm$^{-2}$. For our favored $f(N_{H_i}, X)$ distribution, we calculate $\tau_{912} = 4.1$ and when allowing for sample variance with a bootstrap analysis we find consistency in a non-negligible fraction of the trials (>5%).

Returning to the stacked spectrum (Figure 18), we note a series of absorption lines corresponding to strong metal-line transitions of low and high ionization states. This suggests a highly ionized, and possibly multi-phase gas, similar to the stack constructed of strong Lyα absorbers by Pieri et al. (2010). The strong absorption of O\text{VI}, in particular, suggests a multi-phase medium consisting of at least one “cool” (T ~ 10^4 K), photoionized phase and another, more highly ionized phase which is presumably “warmer” (T ≥ 10^5 K) and possibly collisionally ionized. This highly ionized phase has been detected in the DLAs (Wolfe & Prochaska 2000; Fox et al. 2007a) and SLLSs (Fox et al. 2007b) and its presence in our average spectrum suggests it likely exists in lower $N_{H_i}$ systems too (e.g., Simcoe et al. 2004). A more quantitative analysis of the metal-line absorption in LLSs, however, awaits high-resolution spectroscopy (e.g., Prochter et al. 2010).

### 9.4. The Physical Nature of the LLSs

As described in Section 1, observations and numerical simulations associate the DLAs with high-z galaxies residing in virialized dark matter halos (e.g., Møller et al. 2002; Pontzen et al. 2008). The majority of absorption lines comprising the Lyα forest, meanwhile, are believed to trace Mpc-scale overdensities in the medium between such galaxies (e.g., Miralda-Escudé et al. 1996). By inference, one may associate LLSs with lower $N_{H_i}$ (<10^{20} cm$^{-2}$) with the interface between the IGM and galaxies.
This inference, however, has not yet been extensively tested by cosmological simulations or empirical observation. Early works on the topic generally yielded too few LLSs in cosmological volumes (Katz et al. 1996; Gardner et al. 2001; Maller et al. 2003). More recently, Kohler & Gnedin (2007) examined LLSs in a suite of simulations tuned to match the observed incidence of LLSs at $z \sim 4$. Their simulations suggest LLSs are highly ionized gas occupying volumes of space with dimension 1–100 kpc and physically associated with galaxies of a wide range in mass. The simulations, however, were not rigorously tested against observations nor did they have sufficient spatial resolution to “establish the physical nature of these systems.” The question remains: what is the physical nature of the LLSs?

Our observations place new constraints on the structures that give rise to LLSs absorption. The most informative measurements are $\ell_{\geq 2}(X)$ and the shape of $f(N_{H_1}, X)$ in the LLSs regime. Consider first $\ell_{\geq 2}(X)$, which is proportional to the comoving number density $n_{LLS}$ of the structures times their average physical size $\sigma_{LLS}$. In Figure 11, we present the $\ell(X)$ values (in cumulative form) for DLAs (Prochaska & Wolfe 2009) and SLLSs (O’Meara et al. 2007). For the latter, we assume $f_{SLLS}(X) = 0.20$ at all redshifts. We have adopted a 20% lower ($1\sigma$) value than reported by O’Meara et al. (2007) to crudely correct for the SDSS quasar-targeting bias that will affect their measurement (PW09). The figure demonstrates that the DLAs (especially) and the SLLSs have modest contributions to $\ell_{\geq 2}(X)$. At $z = 3.4$, they contribute roughly half of the observed incidence of $\tau_{\geq 12}$ LLS decreasing to $\approx 33\%$ by $z = 4$. This latter conclusion hinges on our assumptions for $\ell_{SLL}(X)$, in particular at $z \approx 4$ where the value is not well constrained, but we expect the SLLSs to behave similarly to the DLAs whose incidence is not increasing significantly at these redshifts. We conclude that the incidence of $\tau_{\geq 12} \lesssim 10$ LLSs $\ell_{SLL}(z)$ is comparable to that of the DLAs and SLLSs. As a result, it is reasonable to associate all LLSs with the same structures, i.e., gas within virialized halos as suggested by Kohler & Gnedin (2007). The principle challenge to this association is whether diffuse halo gas has sufficiently high cross section to LLS absorption. In particular, one should consider whether the “cold flows” identified in numerical simulations of galaxy formation (Kereš et al. 2005; Dekel et al. 2009) have sufficient density and size to explain the majority of LLSs. We are currently pursuing such analysis.

Now consider the evolution in $\ell_{\geq 2}(X)$ with redshift. As noted in Section 8.2, $\ell_{\geq 2}(X)$ is observed to decrease with decreasing redshift. Examining Figure 11 it is evident that this decrease is driven by LLS with lower $\tau_{\geq 12}$, i.e., by a significant decrease
Figure 19. (Continued)
2. A survey of LLSs in the SDSS spectra \( z_{\text{em}} < 3.6 \) quasars confirms a previously identified bias (PWO09) in the SDSS quasar-targeting criteria that biases the sample toward sight lines with foreground LLS absorption.

3. The number of \( r_{\text{012}} \geq 2 \) LLS per unit absorption length \( \ell_{\text{r} \geq 2}(X) \) is observed to decrease by \( \approx 50\% \) from \( z = 4 \) to 3.4. This indicates a decrease in the number of systems per comoving Mpc\(^3\) and/or a decrease in the average physical cross-section per system. We suggest it is the latter effect, possibly related to an increase in the EUVB with decreasing redshift or a rising radiation field local to LLSs.

4. The measured \( \ell_{\text{r} \geq 2}(X) \) values place an integral constraint on the \( \text{H}1 \) frequency distribution \( f(N_{\text{H}1}, X) \) at \( z \approx 3.7 \). Adopting previous estimates of \( f(N_{\text{H}1}, X) \) for \( N_{\text{H}1} \geq 10^{19} \) cm\(^{-2}\) (O’Meara et al. 2007; Prochaska & Wolfe 2009), we constrain \( f(N_{\text{H}1}, X) \) for \( N_{\text{H}1} = 10^{17.5} \)–\( 10^{19} \) cm\(^{-2}\) assuming a power-law form \( f_{\text{LLS}}(N_{\text{H}1}, X) = k_{\text{LLS}}N_{\text{H}1}^{-\beta_{\text{LLS}}} \) to have \( k_{\text{LLS}} \approx 10^{-4.5} \) and \( \beta_{\text{LLS}} = -0.8 \pm 0.3 \). This indicates a further shallowing of the slope as one decreases \( N_{\text{H}1} \) below \( 10^{19} \) cm\(^{-2}\).

5. Adopting constraints from the mean free path (PWO09) and \( \text{Ly} \alpha \) forest, we derived new constraints on \( f(N_{\text{H}1}, X) \) at \( z \approx 3.7 \) for \( N_{\text{H}1} \approx 10^{15.5} \)–\( 10^{18} \) cm\(^{-2}\). We find that \( \beta \equiv d \ln f(N_{\text{H}1}, X)/d \ln N_{\text{H}1} \) must be steeper than \( \beta = -1.5 \) at \( N_{\text{H}1} \approx 10^{15} \) cm\(^{-2}\).

6. We surveyed the spectra for PLLSs, those with redshifts that are within \( 3000 \) km s\(^{-1}\) of the quasar. We measure an \( \approx 25\% \) lower incidence of PLLSs than intervening systems at \( z > 3.5 \). This lends further support to the assertion that quasars have anisotropic emission (Hennawi & Prochaska 2007).

Compared with previous work, our estimates of \( \ell(z) \) show systematically lower values. Our results are supported by measurements of the mean free path (PWO09) which do not allow for a significantly higher incidence of LLSs at \( z > 3.6 \). We also find that the range of power laws that describe our results at \( z \approx 3.5–4 \) do not extrapolate to the results from \( z < 1 \) observations (Stengler-Larrea et al. 1995). We infer that the incidence of LLSs exhibits a break at \( z \approx 2 \), qualitatively similar to that observed for the Ly\( \alpha \) forest (e.g., Weymann et al. 1998). The declining incidence of LLSs per absorption length and the very shallow slope of \( f(N_{\text{H}1}, X) \) at \( N_{\text{H}1} < 10^{19} \) cm\(^{-2}\) suggest that \( r_{\text{012}} \geq 10 \) LLSs arise in flattened (e.g., filamentary) structures that have relatively sharp edges. We associate these structures to the virialized halos that presumably give rise to SLLS and DLA absorption. Finally, we encourage future work on whether such structures are consistent with the “cold flows” identified in numerical simulations.

Through detailed analysis of biases and careful sample selection from a large and homogeneous data set, this paper provides the first robust estimate of the incidence of LLSs at high redshift. We note that the systematic errors described in Section 7 likely limit the precision of any future \( \ell(z) \) estimates to the order of 20%–30%. Nevertheless, this is sufficient to further explore the true evolution in \( \ell(z) \) with redshift. Programs with the HST for absorption at \( z < 2 \) and with ground-based observatories for \( z > 4 \) are currently ongoing. Altogether, these projects will describe the evolution of the UV background, the growth of structure on galactic (and larger) scales, and the chemical enrichment history of the universe.

Some of the data presented herein were obtained at the W. M. Keck Observatory, which is operated as a scientific
partnership among the California Institute of Technology, the University of California and the National Aeronautics and Space Administration. The Observatory was made possible by the generous financial support of the W. M. Keck Foundation. The authors wish to recognize and acknowledge the very significant cultural role and reverence that the summit of Mauna Kea has always had within the indigenous Hawaiian community.

The authors wish to recognize and acknowledge the very significant cultural role and reverence that the summit of Mauna Kea has always had within the indigenous Hawaiian community.

The authors wish to recognize and acknowledge the very significant cultural role and reverence that the summit of Mauna Kea has always had within the indigenous Hawaiian community.

Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The Observatory was made possible by an NSF CAREER grant (AST–0548180) and by NSF grant AST-0908910.

To assess uncertainties (statistical and systematic) of surveying LLSs in the SDSS quasar spectra, we obtained independent, higher quality spectra using the LRIS spectrometer (Oke et al. 1995) on the Keck I telescope. LRIS employs a dichroic to split the data into two spectral channels, each with its own camera. For our observations, we employed the d560 dichroic which splits the light at \( \approx 5600 \) Å. For the blue channel, we used the 640/4000 grism which provides a dispersion of 0.63 Å per unbinned pixel and has a nominal wavelength coverage of 3100 Å < \( \lambda \) < 5600 Å. The blue channel data were binned by 2

### APPENDIX

### COMPARISONS WITH KECK+LRIS SPECTRA

Table 8

| Object Name | \( z_{\text{em}} \) | S/N\(_{Q12}^a \) | SDSS\(_{9g}^a \) | SDSS\(_{LLS} \) | SDSS\(_{pLLS} \) | LRIS\(_{LLS} \) | LRIS\(_{pLLS} \) |
|-------------|----------------|--------------|---------------|----------------|----------------|----------------|----------------|
| J000300.34+160027.7 | 3.70 | 1.5 | 3 | ... | 3.570 | ... | 3.611 |
| J002946.37-093540.9 | 3.61 | 1.3 | 2 | 3.481 | ... | 3.334 | 3.563 |
| J004142.52-085704.6 | 3.61 | 1.2 | 3 | ... | 3.616 | ... | 3.606 |
| J023923.47-081005.1 | 4.02 | 1.9 | 2 | 3.749 | ... | 3.826 | ... |
| J024447.78-081601.6 | 4.07 | 6.8 | 3 | ... | 3.957 | ... | 3.949 |
| J025105.13-001732.0 | 3.46 | 0.5 | 3 | ... | 3.438 | 3.262 | 3.434 |
| J171421.72+314807.4 | 3.42 | 0.7 | 3 | ... | 3.442 | ... | 3.275 |
| J171046.65+303931.2 | 3.35 | 3.5 | 3 | 3.840 | ... | 3.476 | ... |
| J171800.20+621325.6 | 3.67 | 1.6 | 3 | ... | 3.614 | 3.620 | ... |
| J173115.35+563641.0 | 3.71 | 1.5 | 3 | ... | 3.562 | 3.397 | ... |
| J205141.73-071906.2 | 3.84 | 0.9 | 3 | ... | 3.780 | 3.792 | ... |
| J205059.49-071748.6 | 4.01 | 1.3 | 2 | 3.859 | ... | 3.553 | ... |
| J210054.58-004842.9 | 3.51 | 1.0 | 2 | 3.331 | ... | 3.331 | ... |
| J212204.46-001012.5 | 3.48 | 0.7 | 2 | 3.267 | ... | 3.234 | 3.405 |
| J212357.56-005350.1 | 3.58 | 4.0 | 2 | 3.626 | ... | 3.626 | ... |
| J211433.95-005532.7 | 3.42 | 4.1 | 2 | 3.442 | ... | 3.443 | ... |
| J214049.84+103831.7 | 3.92 | 2.2 | 3 | ... | 3.737 | ... | 3.705 |
| J214227.48+055651.8 | 3.64 | 1.0 | 3 | ... | 3.598 | 3.289 | 3.601 |
| J220212.82-085221.7 | 3.53 | 2.2 | 1 | ... | ... | ... | ... |
| J221014.32+114551.9 | 3.49 | 0.8 | 1 | ... | ... | ... | ... |
| J221458.45+153544.8 | 3.67 | 1.3 | 3 | ... | 3.480 | ... | 3.447 |
| J222420.37-085338.6 | 3.60 | 1.0 | 3 | ... | 3.577 | ... | ... |
| J222824.19+134154.9 | 3.99 | 1.8 | 1 | ... | ... | ... | 3.442 |
| J224243.03-091543.9 | 4.21 | 1.9 | 3 | ... | 4.162 | 3.867 | 4.108 |
| J225052.66-084600.2 | 3.87 | 2.1 | 3 | ... | 3.716 | 3.299 | ... |
| J225109.09-083137.5 | 3.88 | 2.0 | 2 | 3.867 | ... | 3.887 | ... |
| J225151.55+125704.1 | 3.42 | 0.7 | 3 | ... | 3.386 | 3.360 | ... |
| J230022.18+125354.1 | 3.68 | 1.5 | 3 | ... | 3.509 | ... | 3.545 |
| J230301.45-093930.7 | 3.45 | 10.0 | 2 | 3.316 | ... | 3.311 | ... |
| J231137.05-084409.5 | 3.75 | 1.3 | 2 | 3.689 | ... | 3.710 | ... |
| J233534.53-085393.9 | 3.68 | 1.7 | 3 | ... | 3.621 | 3.389 | ... |
| J234349.41+104742.0 | 3.62 | 1.1 | 2 | ... | 3.403 | ... | 3.366 |

Notes. List of all objects which have LRIS comparison spectra. Discussion of LLS is limited to systems with \( z > 3.20 \).

\(^a\) Flag indication—1: no LLS; 2: \( \tau_{912} \geq 2 \) LLS; 3: \( \tau_{912} < 2 \) LLS.
in both the spatial and spectral dimensions. For the red channel, we used the 600/7500 grating which provides a dispersion of 1.28 Å per unbinned pixel, and which was tilted to provide a wavelength coverage of 5600 Å < λ < 8200 Å. The red channel data were unbinned. All observations were obtained using a 1 arcsec slit which provides an ≈4 pixel FWHM corresponding to ≈290 km s⁻¹ and ≈220 km s⁻¹ for the blue and red data, respectively. The data were obtained in good sky conditions during a four night run in 2008 October and had exposure times ranging from 300 to 500 s. The data were reduced using the LowRedux pipeline13 which bias subtracts, flat fields, optimally extracts, wavelength and flux calibrates the data to produce a final one-dimensional spectrum.

The SDSS targets for LRIS observations were chosen to sample a range of LLSs, e.g., LLSs with t₉₁₂ ≳ 2, pLLS candidates, PLLSs and spectra without apparent LLSs. Furthermore, an emphasis was placed on quasars with lower S/N SDSS spectra to assess the completeness of recovering LLSs. In all cases, the LRIS spectra have sufficient S/N to unambiguously detect the presence of absorbers with t₉₁₂ > 1 over the full SDSS wavelength range (i.e., λ > 3800 Å) for the intervening LLS survey.

Two of the authors (J.X.P. and J.M.O.) independently modeled LLS absorption in the LRIS and SDSS spectra and then reconciled any substantial differences to produce a final model for each data set. The LRIS and SDSS data were treated independently (months apart from one another). In Figure 19, we present all of the spectra (side by side) and include the model continuum and LLS absorption for each SDSS spectrum. In Table 8, we list the quasars, the S/N₉₁₂ value for the SDSS spectrum, and the results of the LLS modeling in each data set. For the majority of cases (≈80%), there is excellent agreement between the two analyses, including many examples where the SDSS spectra have S/N₉₁₂ ≲ 2. The agreement is particularly impressive given that we intentionally focused on cases with low S/N and/or examples where we considered the SDSS analysis to be poorly constrained.

There are five cases (e.g., J0041–0857) where the LRIS data reveal a t₉₁₂ ≳ 2 LLS but we modeled the SDSS spectra with a pLLS. All of these have S/N₉₁₂ < 2 and nearly all have S/N < 1 at the observed Lyman limit. This results, especially, motivated our criterion of S/N₉₁₂ ≥ 2 for the SDSS survey. There are several examples of the “blending bias” (Section 7.4) where a true pair of LLSs have been modeled as a single LLS at a higher z than is correct (e.g., J0029–0935 and J2035–0717). In fact, in one case (J2325+1432) we suspect the blending bias has even affected our analysis of the higher quality LRIS spectra; higher spectral resolution observations would be required to resolve the two LLS in this case.

REFERENCES

Abazajian, K. N., et al. 2009, ApJS, 182, 543
Abel, T., & Mo, H. J. 1998, ApJ, 494, L151
Bahcall, J. N., & Peebles, P. J. E. 1969, ApJ, 156, L7
Croft, R. A. C., Weinberg, D. H., Bolte, M., Burles, S., Hernquist, L., Katz, N., Kirkman, D., & Thyler, D. 2002, ApJ, 581, 20
Dall’Aglio, A., Wisotzki, L., & Worseck, G. 2008, A&A, 491, 465
Dall’Aglio, A., Wisotzki, L., & Worseck, G. 2009, arXiv:0906.1484
Davé, R., Hernquist, L., Katz, N., & Weinberg, D. H. 1999, ApJ, 511, 521
Dekel, A., et al. 2009, Nature, 457, 451
Faucher-Giguère, C.-A., Lidz, A., Hernquist, L., & Zaldarriaga, M. 2008a, ApJ, 688, 85

13 http://www.ucolick.org/~xavier/LowRedux/