Accepted to ApJ: August 15, 2005
Preprint typeset using LaTeX style emulateapj v. 20/04/00

THE SDSS DAMPED Lyα SURVEY: DATA RELEASE 3

Jason X. Prochaska & Stéphane Herbert-Fort
Department of Astronomy and Astrophysics, UCO/Lick Observatory; University of California, 1156 High Street, Santa Cruz, CA 95064; xavier@ucolick.org, shf@ucolick.org
AND
Arthur M. Wolfe
Department of Physics, and Center for Astrophysics and Space Sciences, University of California, San Diego, Gilman Dr., La Jolla; CA 92093-0424; awolfe@ucsd.edu
Accepted to ApJ: August 15, 2005

ABSTRACT

We present the results from a damped Lyα survey of the Sloan Digital Sky Survey, Data Release 3. We have discovered over 500 new damped Lyα systems at $z > 2.2$ and the complete statistical sample for $z > 1.6$ has more than 600 damped Lyα galaxies. We measure the H I column density distribution $f_{\rm HI}(N, X)$ and its zeroth and first moments (the incidence $\ell_{\rm DLA}$ and gas mass-density $\Omega_g^{\rm DLA}$ of damped Lyα systems, respectively) as a function of redshift. The key results include: (1) the full SDSS-DR3 $f_{\rm HI}(N, X)$ distribution ($z \sim 3.06$) is well fit by a Γ-function (or double power-law) with 'break' column density $N_\gamma = 10^{21.5\pm0.1}\,{\rm cm}^{-2}$ and 'faint-end' slope $\alpha = -1.8 \pm 0.1$; (2) the shape of the $f_{\rm HI}(N, X)$ distributions in a series of redshift bins does not show evolution; (3) the incidence and gas mass density of damped systems decrease by $35 \pm 9\%$ and $50 \pm 10\%$ during $\approx 1\,{\rm Gyr}$ between the redshift interval $z = [3., 3.5]$ to $z = [2.2, 2.5]$; and (4) the incidence and gas mass density of damped Lyα systems in the lowest SDSS redshift bin ($z = 2.2$) are consistent with the current values. We investigate a number of systematic errors in damped Lyα analysis and identify only one important effect: we measure $40 \pm 20\%$ higher $\Omega_g^{\rm DLA}$ values toward a subset of brighter quasars than toward a faint subset. This effect is contrary to the bias associated with dust obscuration and suggests that gravitational lensing may be important. Comparing the results against several models of galaxy formation in ΛCDM, we find all of the models significantly underpredict $\ell_{\rm DLA}$ at $z = 3$ and only SPH models with significant feedback (Nagamine et al.) may reproduce $\Omega_g^{\rm DLA}$ at high redshift. Based on our results for the damped Lyα systems, we argue that the Lyman limit systems contribute $\approx 33\%$ of the universe's H I atoms at all redshifts $z = 2$ to 5. Furthermore, we infer that the $f_{\rm HI}(N, X)$ distribution for $N_{\rm HI} < 10^{20}\,{\rm cm}^{-2}$ has an inflection with slope $d\log f/d\log N > -1$. We advocate a new mass density definition – the mass density of predominantly neutral gas $\Omega_g^{\rm Neut}$ – to be contrasted with the mass density of gas associated with H I atoms. We contend the damped Lyα systems contribute $> 80\%$ of $\Omega_g^{\rm Neut}$ at all redshifts and therefore are the main reservoirs for star formation.

Subject headings: galaxies: evolution — intergalactic medium — quasars: absorption lines

1. INTRODUCTION

The damped Lyα systems are the class of quasar absorption line systems with H I column density $N_{\rm HI} \geq 2 \times 10^{20}\,{\rm cm}^{-2}$ (see Wolfe, Gawiser, & Prochaska 2005, for a review). Unlike the Lyα forest, the damped Lyα systems are comprised of predominantly neutral gas and are proposed to be the progenitors of galaxies like the Milky Way (e.g. Kauffmann 1996). Wolfe et al. (1986) established the $N_{th} = 2 \times 10^{20}\,{\rm cm}^{-2}$ threshold primarily to correspond to the surface density limit of local 21 cm observations at that time. It is somewhat fortuitous that this threshold roughly corresponds to the transition from primarily ionized gas to predominantly neutral gas (e.g. Viegas 1995; Prochaska & Wolfe 1996; Prochaska 1999; Vladilo et al. 2001).

For the past two decades, several groups have surveyed high z quasars for the damped Lyα systems (Wolfe et al. 1986, 1995; Storrie-Lombardi et al. 1996a; Storrie-Lombardi & Wolfe 2000; Péroux et al. 2003). These surveys measured the H I frequency distribution function and its moments: the incidence of the damped Lyα systems $\ell_{\rm DLA}$ and the gas mass density of these galaxies $\Omega_g^{\rm DLA}$. The latter quantity has cosmological significance. Its evolution constrains the build-up of structure within hierarchical cosmology (e.g. Ma & Bertschinger 1994; Klypin et al. 1995), it serves as the important neutral gas reservoir for star formation at high redshift and describes the competition between gas accretion and star formation (e.g. Fall & Pei 1993), and it constrains models of galaxy formation in hierarchical cosmology (e.g. Somerville, Primack, & Faber 2001; Cen et al. 2003; Nagamine, Springel, & Hernquist 2004). Previous surveys have reported statistical error on $\Omega_g^{\rm DLA}$ of $\approx 30\%$ in redshift intervals $\Delta z \approx 0.5$ at high redshift. As we enter the so-called era of precision cosmology, we aspire to constrain $\Omega_g^{\rm DLA}$ to better than 10%. Although not formally a cosmological parameter, a precise determination of $\Omega_g^{\rm DLA}$ and its redshift evolution are fundamental constraints on any cosmological theory of galaxy formation.

In Prochaska & Herbert-Fort (2004), hereafter PH04, we initiated a survey for the damped Lyα systems in the quasar spectra of the Sloan Digital Sky Survey (SDSS) Data Release 1 (DR1). We demonstrated that the spectral resolution, signal-to-noise ratio (SNR), and wavelength coverage of the SDSS spectra are well suited to survey the damped Lyα systems at $z > 2.2$. We reported on the number of damped Lyα systems per unit redshift and the

TABLE 1
SDSS-DR3 QUASAR SAMPLE

Plate	MJD	FiberID	Name	z_{qso}	f^a_{BAL}	z_i	z^b_i	z_f	$z_{candidate}$
650	52143	178	J000050.60−102155.8	2.640	0	2.200	2.200	2.604	
750	52235	550	J000143.41+152021.4	2.638	0	2.200	2.200	2.602	
650	52143	519	J000159.12−094712.4	2.308	0	2.203	2.250	2.275	
387	51791	556	J000221.11+002149.4	3.057	0	2.200	2.209	3.016	1.958
650	52143	111	J000238.41−101149.8	3.941	0	3.203	...	3.891	3.271,3.523,3.605,3.655
750	52235	608	J000300.34+160027.7	3.675	0	3.285	3.480	3.629	
650	52143	48	J000303.34−105150.6	3.647	2	...	...	2.897,3.465	
750	52235	36	J000335.21+144743.6	3.484	0	...	...		
650	52143	604	J000413.63−085529.5	2.424	0	2.200	2.200	2.389	
650	52143	605	J000424.16−085047.9	2.433	0	2.200	2.200	2.399	

[a]0=No intrinsic absorption; 1=Mild intrinsic absorption, included in analysis with restriction; 2=Strong intrinsic absorption, excluded.
[b]Starting redshift for $\mathrm{SNR}_{lim} = 5$.

Note. — [The complete version of this table is in the electronic edition of the Journal. The printed edition contains only a sample.]

neutral gas mass density. In this paper, we extend the survey to include the full Data Release 3. In addition to substantially increasing the PH04 sample, this paper extends the analysis to include a determination of the N(HI) frequency distribution, $f_{\rm HI}(N, X)$. Furthermore, we perform a series of tests to examine systematic error related to the analysis. With the increased sample size, we believe the systematic uncertainty is as important as statistical uncertainty. Similarly, uncertainty related to selection bias (e.g. dust obscuration) must be given careful attention. Future progress will require significant strides on each of these three fronts.

This paper is organized as follows: § 2 describes the SDSS quasar sample and defines the sub-set used to survey the damped Lyα systems; we present the damped Lyα candidates and the $N_{\rm HI}$ measurements in § 3, we perform standard analysis of the H I distribution in § 4, and we discuss the results in § 6. Aside from comparisons against local observations, we mainly restrict the analysis to optical surveys for the damped Lyα systems (i.e. $z > 1.6$). Unless it is otherwise stated all log values and expressions correspond to log base 10. Here and throughout the paper we adopt cosmological parameters consistent with the latest WMAP results (Bennett et al. 2003): $\Omega_\Lambda = 0.7, \Omega_m = 0.3, H_0 = 70 \mathrm{km\ s^{-1}\ Mpc^{-1}}$.

2. SDSS SAMPLE

2.1. *Redshift Path*

The quasar sample considered in this paper includes every object identified spectroscopically as a quasar with $z > 2.2$ in the SDSS-DR3[1]. We also include the full compilation of quasars and damped Lyα systems (avoiding duplication) from the previous two decades of research as compiled by Péroux et al. (2003). In each of these SDSS quasars we have defined a redshift interval where we search for the presence of damped Lyα systems. In this paper, we define the redshift path using an algorithm similar to that introduced in PH04. We also apply more conservative criteria to investigate the effects of SNR (the estimated flux divided by estimated noise).

The starting redshift z_i is defined as follows. First, we identify the minimum wavelength λ_i where the median SNR in a 20 pixel interval exceeds SNR_{lim}. In PH04, we took $\mathrm{SNR}_{lim} = 4$ and in this paper we also consider larger values. An assumption of our prescription is that the median SNR does not decrease significantly at wavelengths greater than λ_i unless a damped Lyα candidate is present. This is a good assumption unless λ_i coincides with the Lyβ/O VI emission feature of the quasar. At these wavelengths, the SNR of the data is elevated and it is possible the median SNR may be above SNR_{lim} in this spectral region but below SNR_{lim} otherwise. Therefore, in the 10000km s^{-1} region centered at $\lambda = 1025.7(1 + z_{qso})$ we demand that the median SNR be greater than 3*SNR_{lim} if λ_i is to be set within that region. In practice, this removes a number of faint quasars from the sample whose spectra have median SNR greater than SNR_{lim} only in the Lyβ/O VI emission line. Finally, we define the starting redshift for the damped Lyα search:

$$z_i \equiv (\lambda_i/1215.67) - 1. + 0.005 \quad , \tag{1}$$

where the 0.005 increment offsets z_i by 1500 km s^{-1} . Finally, we demand $z_i \geq 2.2$.

Similarly, we define an ending redshift for the statistical pathlength of a given quasar,

$$z_f \equiv 0.99 z_{qso} - 0.01 \quad , \tag{2}$$

which corresponds to 3000 km s^{-1} blueward of the Lyα emission feature. The offset minimizes the likelihood that a damped Lyα system is physically associated with the quasar. For reasons unknown to us, a small fraction of the quasar spectra have extended regions (> 50 pixels) with zero flux and null values in their error array. In these cases, we set z_f to 1000 km s^{-1} blueward of the 'null region'.

Each quasar spectrum was visually inspected and characterized according to the presence or absence of features identified with intrinsic absorption (e.g. BAL quasars).

[1]Specifically, where SPECCLASS=3 or 4 in SPECPHOTO for SDSS-DR3.

This is a necessary step in damped Lyα surveys because quasars with strong intrinsic N V and O VI absorption can be confused with the Lyα transition of an intervening quasar absorption line system. Following PH04, we divided the quasars into three categories: (a) quasars without significant intrinsic absorption; (b) mild absorption-line quasars which show modest absorption at the C IV emission feature; and (c) strong absorption-line quasars whose C IV, O VI lines have large equivalent widths and could be confused with a damped Lyα system. The latter category is discarded from all subsequent analysis. For the mild absorption-line quasars, however, we search for damped Lyα systems in the redshift interval $max(z_i, z_{BAL}) < z < min(z_{qso}-0.08226, z_f)$ where $z_{BAL} \equiv (1+z_{qso})(1060/\lambda_{Ly\alpha})-1$ and the modification to the ending redshift minimizes the likelihood of misidentifying N V absorption as a damped Lyα system.

Table 1 presents the full list of SDSS-DR3 quasars analyzed here. Columns 1-9 designate the SDSS plate, MJD, and fiber numbers, the quasar name, the emission redshift, a flag describing intrinsic absorption, z_i and z_f for SNR$_{lim}$=4, and the absorption redshift of any damped Lyα candidates along the sightline. The latter may include candidates where z_{abs} is not in the $[z_i, z_f]$ interval. Also, many candidates are false positive detections, in particular BAL absorption lines.

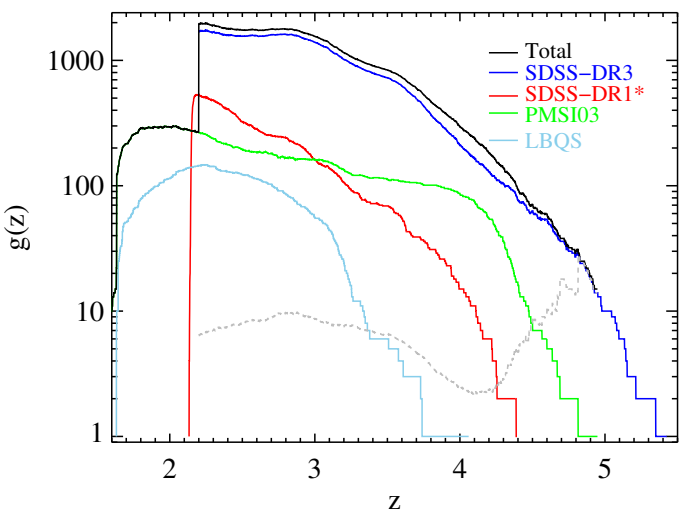

FIG. 1.— Redshift sensitivity function $g(z)$ as a function of redshift for the LBQS survey, the compilation of Péroux et al. (2003), the SDSS-DR1 sample from PH04, the SDSS-DR3 sample studied here, and the total sample. The gray dashed-line traces the ratio of the SDSS-DR3 sample to the previous 20 year compilation (Péroux et al. 2003).

2.2. $g(z)$

The survey size of a quasar absorption line study is characterized through the redshift sensitivity function $g(z)$. Figure 1 presents a series of $g(z)$ curves for the the Large Bright Quasar Survey (LBQS) search for damped Lyα systems (Wolfe et al. 1995), the compilation of Péroux et al. (2003), and the SDSS-DR1 sub-sample of PH04. These are compared against the full SDSS-DR3 sample with SNR$_{lim}$ =4. Note that the SDSS data releases are inclusive and also that the Péroux et al. (2003) compilation includes the LBQS sample. Comparing the curves, we find the current SDSS sample now exceeds the previous surveys by an order of magnitude at $z \sim 3$, several times at $z \sim 4$, and more than a factor of 10 at $z > 4.6$ (the gray dashed-line traces the ratio of SDSS-DR3 to the previous surveys). The dip at $z \sim 4$ in the ratio of SDSS-DR3 to the previous surveys is largely due to the impressive high z survey carried out by the Cambridge group. We also stress that the increased sensitivity of the SDSS-DR3 at $z > 4$ should be viewed conservatively; follow-up observations should be performed to confirm the high redshift results. In Figure 2 we show the $g(z)$ curves for several cuts on SNR$_{lim}$ and sample selection for the SDSS-DR3 quasars. These cuts are summarized in Table 2 and will be referred to throughout the text.

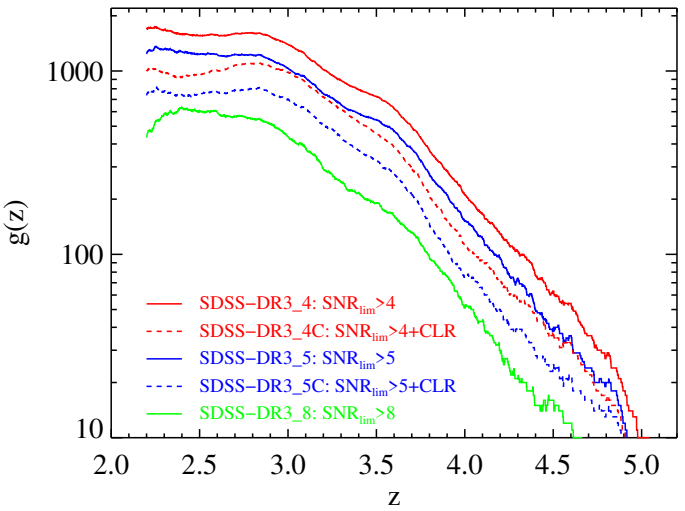

FIG. 2.— Redshift sensitivity function $g(z)$ as a function of redshift for several cuts of the SDSS-DR3 data (Table 2).

In contrast with the $g(z)$ curves for the Lyα forest or Mg II systems (e.g. Prochter, Prochaska, & Burles 2005), one notes no significant features due to strong sky lines. This is because the sky lines can only lead to false positive detections in our algorithm. These are easily identified and ignored.

TABLE 2
SDSS CUTS

Label	SNR$_{lim}$	CLR[a]	n$_{DLA}$
SDSS-DR3_4	4	no	525
SDSS-DR3_4C	4	yes	340
SDSS-DR3_5	5	no	395
SDSS-DR3_5C	5	yes	183
SDSS-DR3_8	8	no	155
SDSS-DR3_8C	8	yes	88

[a]This entry refers to whether the quasar sample adheres to the color criteria de
color-cut by only including quasars drawn from plates 761 and greater.

3. $N_{\rm HI}$ MEASUREMENTS

3.1. *Damped Lyα Candidates*

Damped Lyα candidates were identified in the quasar spectra using the same prescription introduced by PH04. In short, the algorithm compares the SNR of the pixels in a running window of width $6(1+z)$Å against x_{SNR} specifically definied to be the median SNR of the 151 pixels starting 200 pixels redward of the Lyα emission peak divided by 2.5, i.e. x_SNR gauges the SNR of the quasar continuum just redward of Lyα emission. Furthermore, x_{SNR} is restricted to have a minimum value of 1 and a maximum value of 2. If the fraction of pixels with SNR$\geq$ x_{SNR} is $\leq$ 60%, a damped Lyα candidate is recorded at the center of the window. This candidate list is supplemented by systems associated with relatively strong metal-line transitions outside the Lyα forest (see Herbert-Fort et al. 2005, in preparation). Finally, the list is further supplemented by our visual inspection of each quasar spectrum when characterizing its intrinsic absorption. The damped Lyα candidates are listed in Table 1.

In PH04, we reported that Monte Carlo tests of our automated algorithm on synthetic spectra implied an idealized damped Lyα completeness $\sim 95\%$ for $N_{\rm HI} \approx 2 \times 10^{20}\,{\rm cm}^{-2}$ and 100% for $\log N_{\rm HI} > 20.4$. Given the supplemental candidates from the metal-line search and our visual inspection, we believe the completeness definitely exceeds 95% for all absorbers with $N_{\rm HI} > 2 \times 10^{20}\,{\rm cm}^{-2}$. We discuss a new set of completeness tests below.

Every damped Lyα candidate was subjected to the following analysis. First, the Lyα profile was visually inspected and obvious false-positive candidates were eliminated. This visual analysis included overplotting a Lyα profile with $N_{\rm HI} = 2 \times 10^{20}\,{\rm cm}^{-2}$. To minimize the labor of fitting Lyα profiles, we chose not to fit many systems where the overplotted profile clearly was a poor solution. It was our experience that nearly all of these candidates have best-fit values $N_{\rm HI} < 1.5 \times 10^{20}\,{\rm cm}^{-2}$. Second, we searched for metal-line absorption at redshifts near the estimated redshift centroid of the Lyα profile. The extensive wavelength coverage of SDSS spectra is a tremendous advantage in the analysis of damped Lyα candidates. In general, we focused on the strongest low-ion transitions observed in the damped Lyα systems (e.g. Prochaska et al. 2003a): Si II λ1260, 1304, 1526, O I λ1302, C II λ1334, Al II λ1670, and Fe II λ1608,2382,2600. We characterized the metal-line absorption for each candidate as: (a) no metals detected[2]; (b) weak or ambiguous metal absorption; and (c) secure metal-line absorption. Damped Lyα candidates in the latter category generally exhibit two or more metal-line features outside the Lyα forest. For damped Lyα systems with secure metal-line absorption, we constrain the subsequent Lyα profiles to coincide with the metal-line redshift.

3.2. *Lyα Fits*

Those damped Lyα candidates which were not rejected by visual inspection were fitted with Voigt profiles in a semi-quantitative fashion. As discussed by Prochaska et al. (2003a), systematic error associated with continuum placement and line-blending generally dominates the statistical error associated with Poisson noise in the quasar flux. We have fitted the Lyα profile with an IDL tool *x_fitdla* which allows the user to interactively modify the Voigt profiles and continuum placements. We then estimated a conservative uncertainty to each $N_{\rm HI}$ value based on the constraints of the quasar continuum near the Lyα profile, the degree of line blending, and the Poisson noise of the data. Bolton et al. (2004) have emphasized that the SDSS spectroscopic data has non-Gaussian fluctuations, e.g. 3σ departures from the 'true' value with higher than 0.27% frequency. Therefore, we have generally ignored single pixel outliers when performing the Lyα fits.

To be conservative, we adopt a minimum uncertainty of 0.15 dex in $\log N_{\rm HI}$ for the absorption systems with unambiguous metal-line detections and 0.20 dex otherwise. These uncertainties are larger than those reported in PH04. As noted above, the effects of line blending and continuum uncertainty dominate the measurement uncertainty of the $N_{\rm HI}$ values. Furthermore, it is very unlikely that the errors related to these systematic effects are Gaussian distributed. At present, it is very difficult to accurately assess the magnitude and distribution of the errors in the $N_{\rm HI}$ values. We discuss below an attempt based on the analysis of mock spectra, yet even this approach has limited value. Nevertheless, we contend that 95% of the true $N_{\rm HI}$ values will lie within 2σ of the reported $N_{\rm HI}$ values, but we cannot argue that the reported errors are Gaussian distributed. This remains a significant failing in the analysis of DLA surveys.

Figure 3 shows four example profile fits from the SDSS sample which illustrate several of the key issues. In Figure 3a we present a high signal-to-noise case with a well constrained quasar continuum and the redshift centroided

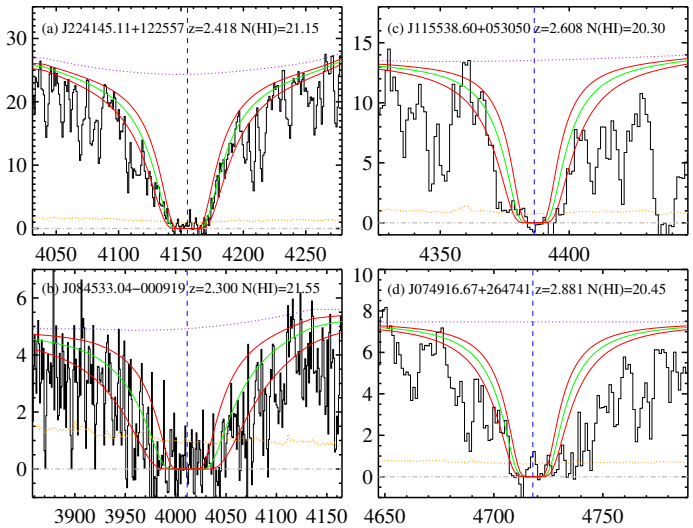

FIG. 3.— Example profile fits to four damped Lyα systems from our SDSS-DR3 survey. The examples represent (a) a high SNR case with a well constrained quasar continuum and minimal line blending ($\sigma(N_{\rm HI}) = 0.15$); (b) a low SNR case ($\sigma(N_{\rm HI}) = 0.30$); (c) an example with severe line blending ($\sigma(N_{\rm HI}) = 0.20$); and (d) an example where determining the redshift from associated metal-line absorption is very important ($\sigma(N_{\rm HI}) = 0.15$).

[2]We note that these damped Lyα candidates may very well exhibit metal-line absorption at higher spectral resolution and/or within the Lyα forest.

by metal-line detections. The fit is constrained by data both in the core and wings of the Lyα profile and the adopted 0.15 dex uncertainty is overly conservative. Figure 3b presents a low signal-to-noise case. In this case, it is very difficult to constrain the fit based on the core of the Lyα profile. Nevertheless, the continuum level is reasonably well constrained and the wings of the profile set the $N_{\rm HI}$ value. Figure 3c presents a case where line-blending is severe and the $N_{\rm HI}$ value is only constrained by the wings of the profile. In these cases the continuum placement is critical. Finally, Figure 3d emphasizes the importance of metal-line detection. In this example, the Lyα profile is centered on the redshift of the metal-line transitions. Without this constraint, we would have centered the Lyα profile a few Angstroms redward of the current position and derived a larger $N_{\rm HI}$ value. We suspect that some damped Lyα systems without metal-line detections have $N_{\rm HI}$ values that are biased high ($\approx$ 0.1 to 0.2 dex). Of course, these damped Lyα systems tend to have lower $N_{\rm HI}$ values and this systematic effect is primarily an issue only for candidates with $N_{\rm HI} \approx 2 \times 10^{20}\,{\rm cm}^{-2}$. Finally, we should stress that the $N_{\rm HI}$ values of damped Lyα candidates near the quasar O VI or Lyα emission features are difficult to constrain because of uncertain continuum placement.

All of the fits from the DR3 sample and the Lyα fits from are presented at http://www.ucolick.org/ ∼xavier/SDSSDLA/index.html. In addition to postscript plots, we provide the files (IDL format) which contain the Lyα fits and quasar continua, and also a description of the software used to perform the fits. Interested readers can reanalyze any of our Lyα fits.

TABLE 3
SDSS-DR3 DLA STATISTICAL SAMPLE

Quasar	z_{abs}	DR	SNR	CLR	f_m^b	log $N_{\rm HI}$
J000238.4−101149.8	3.2674	2	5	n	2	$21.20^{+0.15}_{-0.15}$
J000536.3+135949.4	3.4896	2	5	n	2	$20.30^{+0.15}_{-0.15}$
J001115.2+144601.8	3.6118	2	8	n	1	$20.60^{+0.25}_{-0.25}$
J001240.5+135236.7	3.0217	2	8	n	2	$20.55^{+0.15}_{-0.15}$
J001328.2+135827.9	3.2808	2	8	n	2	$21.55^{+0.15}_{-0.15}$
J001918.4+150611.3	3.9710	2	5	n	2	$20.40^{+0.20}_{-0.20}$
J002614.6+143105.2	3.3882	2	5	n	2	$20.70^{+0.15}_{-0.15}$
J003126.7+150739.5	3.3583	2	8	n	2	$20.95^{+0.20}_{-0.20}$
J003501.8−091817.6	2.3376	1	8	n	1	$20.55^{+0.15}_{-0.15}$
J003749.1+155208.3	3.4790	2	4	n	1	$20.35^{+0.20}_{-0.20}$

[a]0=No metals; 1=Weak metals; 2=Metals detected

Note. — [The complete version of this table is in the electronic edition of the Journal. The printed edition contains only a sample.]

Tables 3 to 5 present the $N_{\rm HI}$ values for the damped Lyα systems comprising the statistical sample, the damped Lyα systems discovered which are not in any of our statistical samples, and all of the damped Lyα candidates with central $N_{\rm HI}$ values less than $2 \times 10^{20}\,{\rm cm}^{-2}$ (termed super-LLS). Regarding Table 3, we list the damped Lyα absorption redshift, indicate its membership within the various statistical cuts, list the data release of the actual spectrum analyzed[3], present the metal-line characteristics, and list the $N_{\rm HI}$ value and error. The non-statistical damped Lyα sample is comprised of damped Lyα systems with $z_{abs} \approx z_{qso}$, $z_{abs} < z_i$, and/or systems toward quasars with strong intrinsic absorption. We present this Table for completeness. We emphasize that while the absorbers listed in Table 5 have a central value below $2 \times 10^{20}\,{\rm cm}^{-2}$, the true $N_{\rm HI}$ value of many of these absorbers will exceed the damped Lyα threshold.

TABLE 4
SDSS-DR3 DLA NON-STATISTICAL SAMPLE

Name	z_{abs}	DR	f_{mtl}^a	log $N_{\rm HI}$
J001115.2+144601.8	3.4523	2	2	$21.65^{+0.20}_{-0.20}$
J001134.5+155137.4	4.3175	2	1	$20.50^{+0.20}_{-0.20}$
J001134.5+155137.4	4.3592	2	2	$21.10^{+0.20}_{-0.20}$
J004950.9−093035.6	3.2858	2	2	$20.70^{+0.20}_{-0.20}$
J014214.7+002324.3	3.3482	1	2	$20.40^{+0.15}_{-0.15}$
J020651.4−094141.3	2.4702	2	2	$20.30^{+0.20}_{-0.20}$
J021232.1−100422.1	2.7140	1	0	$21.00^{+0.15}_{-0.15}$
J023148.8−073906.3	2.2982	3	2	$21.75^{+0.20}_{-0.20}$
J033344.4−060625.1	3.9349	3	2	$21.60^{+0.20}_{-0.20}$
J073149.5+285448.7	2.6859	2	1	$20.55^{+0.15}_{-0.15}$

[a]0=No metals; 1=Weak metals; 2=Metals detected

Note. — [The complete version of this table is in the electronic edition of the Journal. The printed edition contains only a sample.]

TABLE 5
SDSS-DR3 SUPER-LLS SAMPLE

Quasar	z_{abs}	DR	SNR	CLR	f_m^b	log $N_{\rm HI}$
J001328.2+135828.0	2.6123	3	5	n	1	$20.10^{+0.15}_{-0.15}$
J002614.6+143105.2	3.9039	2	5	n	2	$20.10^{+0.15}_{-0.15}$
J004732.7+002111.3	2.4687	2	5	n	2	$20.00^{+0.15}_{-0.15}$
J013317.7+144300.3	2.4754	2	4	n	2	$20.00^{+0.20}_{-0.20}$
J013317.7+144300.3	2.9766	2	8	n	2	$20.15^{+0.15}_{-0.15}$
J014609.3−092918.2	3.6804	2	8	n	1	$20.25^{+0.20}_{-0.20}$
J021143.3−084723.8	2.2684	2	4	y	2	$20.10^{+0.15}_{-0.15}$
J021232.1−100422.1	2.2738	1	8	n	2	$20.25^{+0.15}_{-0.15}$
J025039.1−065405.1	4.3894	3	4	n	2	$20.00^{+0.15}_{-0.15}$
J031036.8+005521.7	3.1142	2	5	n	2	$20.20^{+0.15}_{-0.15}$

[a]0=No metals; 1=Weak metals; 2=Metals detected

Note. — [The complete version of this table is in the electronic edition of the Journal. The printed edition contains only a sample.]

3.3. *Completeness Tests and $N_{\rm HI}$ Accuracy*

There are several tests one can perform to assess the reliability and completeness of the $N_{\rm HI}$ analysis. A valuable test of the $N_{\rm HI}$ accuracy is to compare the SDSS measurements against values derived from observations at higher spectral resolution and/or SNR. Over the past several years, we have observed a subset of the SDSS damped Lyα sample for other scientific studies (e.g. chemical evolution and C II* analysis; Prochaska et al. 2003b; Wolfe

[3]In general this corresponds to the first data release when the quasar was observed by SDSS although errors in our bookkeeping may have delayed its analysis.

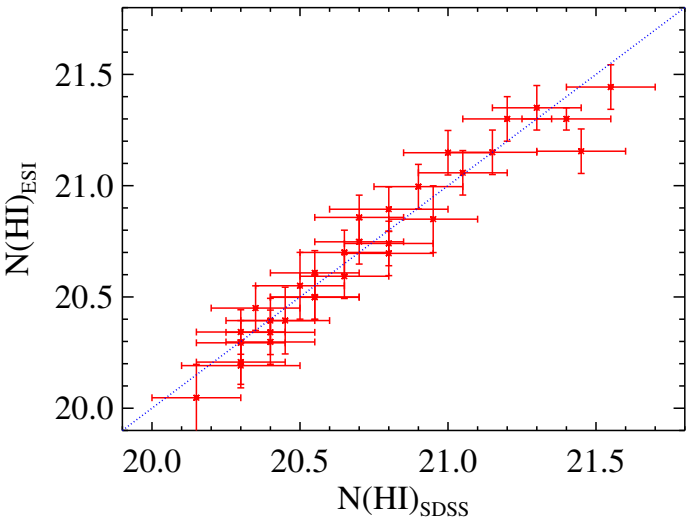

FIG. 4.— Comparison of the $N_{\rm HI}$ values for a sub-set of the SDSS-DR3 sample as measured from the SDSS quasar spectra and, independently, data acquired with the Echellette Spectrometer and Imager.

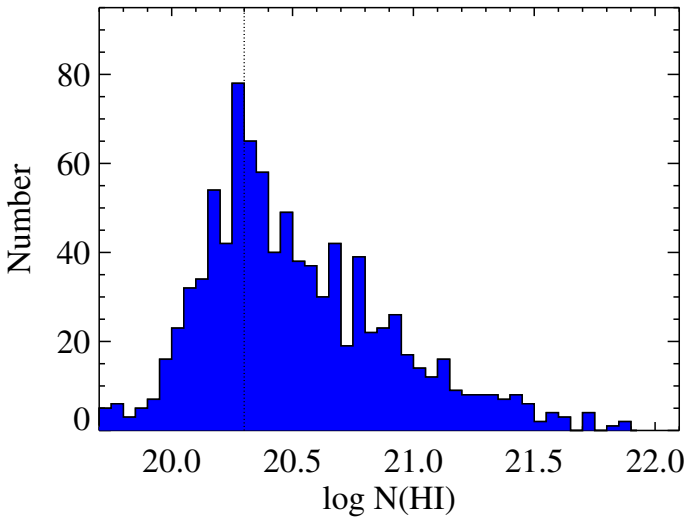

FIG. 5.— Histogram of the $N_{\rm HI}$ values for every damped Lyα candidate fit in the SDSS-DR3 sample. Note that the distribution peaks at $N_{\rm HI} < 2 \times 10^{20}\,{\rm cm}^{-2}$ suggesting a high level completeness at the damped Lyα threshold (dotted line). We currently estimate the completeness level to be $\approx 99\%$.

et al. 2004) with the Echellette Spectrometer and Imager (ESI; Sheinis et al. 2002). Figure 4 compares the values of the SDSS fits with those derived from the ESI spectra. The agreement is excellent. It is important to note, however, that the fits were performed primarily by one of us (JXP) and that the systematic error related to continuum placement is likely correlated. Similarly, the effects of line-blending are qualitiatively similar for the two data sets. Indeed, the reduced $\chi^2_\nu = 0.4$ indicates the estimated uncertainties significantly exceed the statistical uncertainty. There is a notable case from PH04 (at $z = 2.77$ toward J084407.29+515311) where the SDSS value is 0.3 dex higher than the ESI value due to improper continuum placement in the SDSS analysis. We believe this one case was the result of our initial inexperience with the SDSS spectra and that such errors are very rare in the current analysis. Nevertheless, it does underscore the fact that the error in the $N_{\rm HI}$ values are predominantly systematic and in some cases large.

One qualitative assessment of sample completeness is an inspection of the $N_{\rm HI}$ distribution. Figure 5 presents a histogram of $N_{\rm HI}$ values for all of the fitted damped Lyα candidates. The distribution shows a steady increase to lower $N_{\rm HI}$ value which extends just below $2 \times 10^{20}\,{\rm cm}^{-2}$. We note that the relatively steep decline in the distribution at $N_{\rm HI} < 10^{20.2}\,{\rm cm}^{-2}$ is partly due to incompleteness in our search algorithm. It is also due to the fact that we eliminated many damped Lyα candidates after visual inspection.

Since the publication of PH04, we have performed a damped Lyα survey on a new set of mock spectra kindly provided by P. McDonald. These mock spectra were carefully constructed to match the SNR and redshifts of the full SDSS-DR3 quasar sample. A synthetic Lyα forest was added to the quasar continuum according to the prescriptions detailed in McDonald et al. (2005a). Finally, a random set of damped Lyα systems (with a column density frequency distribution similar to the results below) were added to the spectra and also a set of Lyman limit systems with $N_{\rm HI} = 1 \times 10^{17}$ to $2 \times 10^{20}\,{\rm cm}^{-2}$ following the prescriptions of McDonald et al. (2005b). We analyzed these mock spectra using the same algorithms as the scientific search and recovered 99% of the damped Lyα systems. The few that we did not 'discover' have $N_{\rm HI} < 10^{20.4}\,{\rm cm}^{-2}$ and just missed satisfying the damped Lyα candidate criteria. Furthermore, we recovered 90% of the absorbers with $N_{\rm HI} \approx 1 \times 10^{20}\,{\rm cm}^{-2}$. Therefore, we are confident that the sample is nearly complete although it is likely we are missing $\approx 5\%$ of the absorbers with $N_{\rm HI} \approx 2 \times 10^{20}\,{\rm cm}^{-2}$.

To further investigate uncertainty in our $N_{\rm HI}$ measurements, we fitted Voigt profiles to 50 of the damped Lyα systems from the mock spectra. With only one exception, our fitted values are within 0.3 dex of the true value. In this one exception we fitted one damped Lyα profile to a case that was actually two damped systems separated by $\approx 500\,{\rm km\,s}^{-1}$. In practice, this situation can be accounted for if both damped Lyα systems exhibit metal-line absorption. In any case, we expect these systems to be rare (Lopez & Ellison 2003), but a non-zero systematic error to the damped Lyα analysis. In future papers, we will expand this type of mock analysis to probe other aspects of fitting errors and completeness limits for the SDSS samples.

4. RESULTS

In this section, we report the main results of the damped Lyα survey. These include the H I frequency distribution function and its zeroth and first moments. It is amusing to note that many of the techniques are directly analogous to the derivation and analysis of galaxy luminosity functions. Indeed, we rely on many of the standard techniques and face similar challenges and systematic uncertainties.

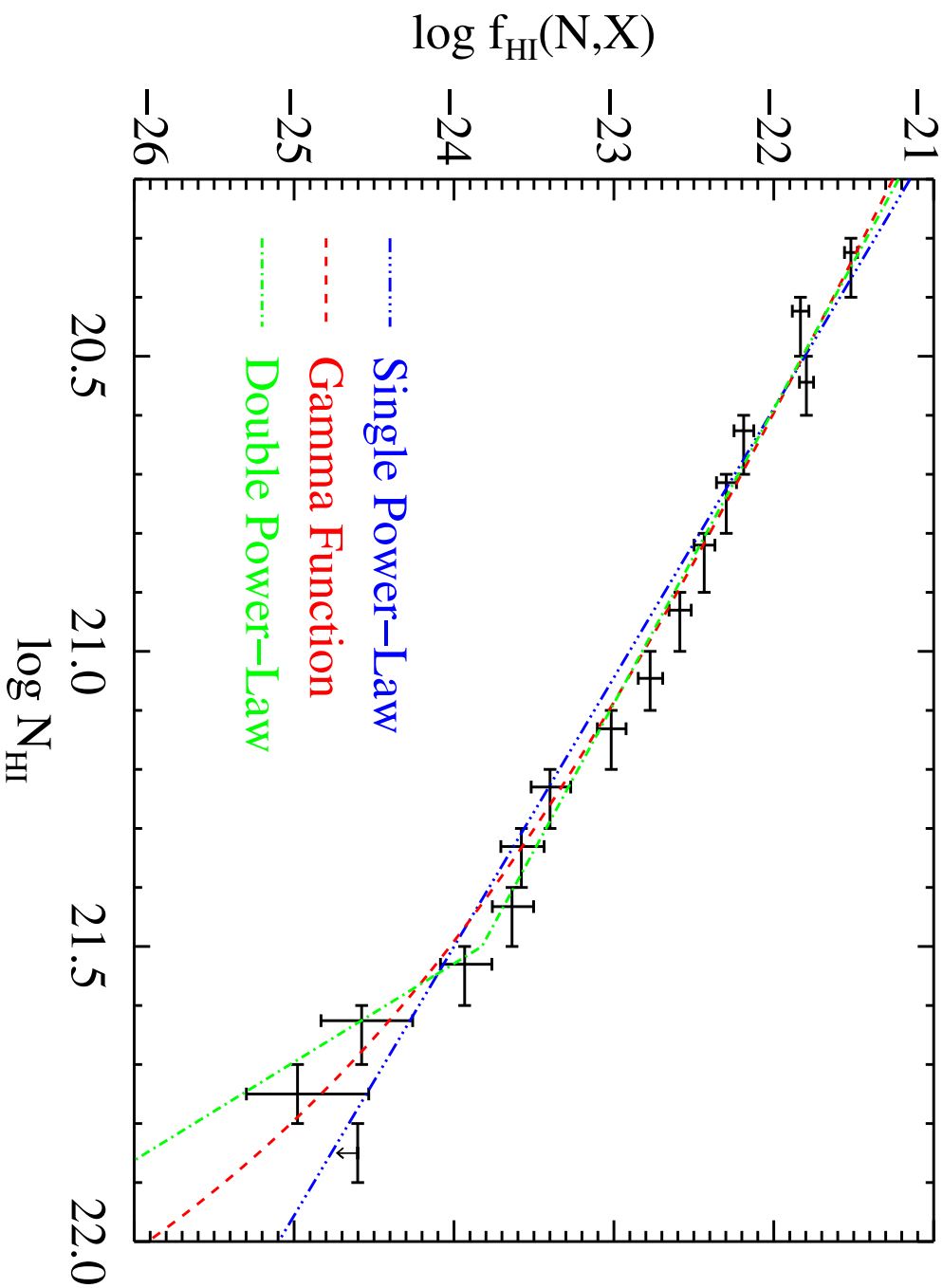

FIG. 6.— The H I frequency distribution $f_{\rm HI}(N,X)$ for all of the damped Lyα systems identified in the SDSS-DR3_4 sample (mean redshift $z = 3.06$). Overplotted on the discrete evaluation of $f_{\rm HI}(N,X)$ are the fits of a single power-law, a Γ-function, and a double power-law. Only the latter two are acceptable fits to the observations.

4.1. $f_{\rm HI}(N,X)$: *The H I Column Density Frequency Distribution*

Following previous work (e.g. Lanzetta et al. 1991), we define the number of damped Lyα systems in the intervals $(N, N+dN)$ and $(X, X+dX)$:

$$f_{\rm HI}(N,X)dNdX \quad , \tag{3}$$

where $f_{\rm HI}(N,X)$ is the frequency distribution and the 'absorption distance'

$$dX \equiv \frac{H_0}{H(z)}(1+z)^2 dz \quad . \tag{4}$$

With this definition, an unevolving population of objects will have constant $f_{\rm HI}(N,X)$ in time provided one adopts the correct cosmology.

In Figure 6, we present $f_{\rm HI}(N,X)$ for the full SDSS-DR3_4 survey. This sample spans the redshift interval $z = [2.2, 5.5]$ with an integrated absorption pathlength $\Delta X = 7417.5$ and a $N_{\rm HI}$-weighted mean redshift of 3.2. The solid points with error bars describe $f_{\rm HI}(N,X)$ for $N_{\rm HI}$ bins of $\Delta N = 0.1$ dex:

$$f_{\rm HI}(N,X) = \frac{m_{DLA}(N, N+\Delta N)}{\Delta X} \quad , \tag{5}$$

with m_{DLA} the number of damped Lyα systems within $(N, N+\Delta N)$ in the ΔX interval. The error bars reflect Poisson uncertainty (68.3% c.l.) according to the value of m_{DLA} and the upper limits correspond to 95% c.l. The derivation of $f_{\rm HI}(N,X)$ in this discrete manner is analogous to the derivation of galaxy luminosity functions using the V_{max} method.

Lacking a physical model, we have considered three functional forms to describe $f_{\rm HI}(N,X)$: (i) a single power-law

$$f_{\rm HI}(N,X) = k_1 N^{\alpha_1} \quad , \tag{6}$$

(ii) a Γ-function (e.g. Fall & Pei 1993)

$$f_{\rm HI}(N,X) = k_2 \left(\frac{N}{N_\gamma}\right)^{\alpha_2} \exp\left(\frac{-N}{N_\gamma}\right) \quad , \tag{7}$$

and (iii) a double power-law

$$f_{\rm HI}(N,X) = k_3 \left(\frac{N}{N_d}\right)^{\beta} \text{ where } \beta = \{\, \alpha_3 : N < N_d; \alpha_4 : N \geq N_d \} \tag{8}$$

In each case, we perform a maximum likelihood analysis to constrain the functional parameters and set the normalization constants (k_1, k_2, k_3) by imposing the integral constraint

$$\int_{N_{th}}^{\infty} f(N,X)dN = \frac{m_{DLA}}{\Delta X} \quad . \tag{9}$$

To estimate the parameter uncertainties and construct a correlation matrix, we performed a 'jack-knife' analysis (e.g. Lupton 1993; Blanton et al. 2003). Specifically, we derived the best fit parameters for 21 subsamples ignoring 222 random quasars in each case. The parameter uncertainty is then given by

$$\sigma^2 = \frac{N-1}{N}\sum_i (\, x_i - \bar{x}\,)^2 \tag{10}$$

with $N = 21$. In some cases whese values are close to the one-parameter confidence limits derived from the maximum likelihood function although there are notable exceptions. We calculate the correlation matrix in standard fashion (see Blanton et al. 2003),

$$r_{ij} = \frac{< \Delta x_i \Delta x_j >}{\left(< \Delta x_i^2 >< \Delta x_j^2 > \right)^{\frac{1}{2}}} \tag{11}$$

with

$$< \Delta x_i \Delta x_j >= \frac{N-1}{N}\sum_i (x_i - \bar{x}_i)(x_j - \bar{x}_j) \quad . \tag{12}$$

As is typically the case for fits to galaxy luminosity functions, we find significant correlation between the parameters. It is important to keep this in mind when comparing the fits to various redshift intervals. Table 6 presents the best-fit values for the parameters for the full SDSS-DR3 sample and also the fits to the damped Lyα systems in several redshift intervals. In this table, the error bars refer to 'one-parameter uncertainties' that correspond to the 68% c.l. of the maximum likelihood function when keeping the other parameters fixed at their best-fit values. The exceptions are the normalization values where we have only reported the Poissonian error based on the best fit. Table 7 provides the parameter uncertainties and the correlation matrix for the full SDSS-DR3_4 sample. When the absolute value of the off-diagonal terms is much less than 1, then there is little correlation between the parameters. Unfortunately, there are too few damped Lyα systems in the redshift bins to perform a meaningful jack-knife analysis for these subsets. We suspect, however, that the parameters are correlated in a similar way as to the results presented in Table 7.

The best-fit solutions are overplotted on the binned evaluation of $f_{\rm HI}(N,X)$ in Figure 6. First, consider the single power-law solution (dotted line) with a best fit slope of $\alpha_1 = -2.19 \pm 0.05$. This functional form is a poor description of the data. A one-sided Kolmogorov-Smirnov (KS) test indicates there is a less than 0.1% probability that the cumulative distributions of the observations and the power-law are drawn from the same parent population. In short, the power-law is too steep at low $N_{\rm HI}$ and too shallow at large $N_{\rm HI}$. Although previous surveys suggested a single power-law was a poor description (e.g. Wolfe et al. 1995), their sample size was too small to rule out this solution. Note that this result contrasts the damped Lyα systems with the Lyα forest (absorbers with $N_{\rm HI} < 10^{15}\,{\rm cm}^{-2}$) where a single power-law with exponent $\alpha_1 \approx -1.5$ is a good description of the observations (e.g. Kirkman & Tytler 1997).

Although a single power-law is a poor description of the observations, the fit does highlight an important new result: the $f_{\rm HI}(N,X)$ distribution is steeper than a $N_{\rm HI}^{-2}$ power-law at large column density. We will further develop this point in § 4.3. The other two curves on Figure 6 show the Γ-function (dashed line) and double power-law fit (dash-dot line). Both of these functional forms are a fair

TABLE 6
FITS TO $f_{\rm HI}(N,X)$

Form	Parameters	$z\epsilon[2.2,5.5)^{a}$	$z\epsilon[1.7,2.2)$	$z\epsilon[2.2,2.5)$	$z\epsilon[2.5,3.0)$	$z\epsilon[3.0,3.5)$	$z\epsilon[3.5,5.5)$
Single	$\log k_1$	$23.16^{+0.02}_{-0.02}$	$21.81^{+0.08}_{-0.08}$	$27.15^{+0.04}_{-0.04}$	$22.12^{+0.03}_{-0.03}$	$20.84^{+0.03}_{-0.03}$	$23.33^{+0.04}_{-0.04}$
	α_1	$-2.19^{+0.05}_{-0.05}$	$-2.13^{+0.18}_{-0.23}$	$-2.39^{+0.13}_{-0.15}$	$-2.14^{+0.08}_{-0.08}$	$-2.08^{+0.08}_{-0.09}$	$-2.20^{+0.10}_{-0.11}$
Gamma	$\log k_2$	$-23.52^{+0.02}_{-0.02}$	$-23.03^{+0.08}_{-0.08}$	$-23.40^{+0.04}_{-0.04}$	$-23.70^{+0.03}_{-0.03}$	$-23.03^{+0.03}_{-0.03}$	$-23.99^{+0.04}_{-0.04}$
	$\log N_\gamma$	$21.48^{+0.09}_{-0.06}$	$21.31^{+0.35}_{-0.15}$	$21.31^{+0.29}_{-0.12}$	$21.56^{+0.16}_{-0.10}$	$21.36^{+0.12}_{-0.08}$	$21.69^{+0.14}_{-0.14}$
	α_2	$-1.80^{+0.06}_{-0.06}$	$-1.56^{+0.13}_{-0.27}$	$-1.94^{+0.16}_{-0.16}$	$-1.78^{+0.09}_{-0.10}$	$-1.52^{+0.09}_{-0.11}$	$-1.93^{+0.12}_{-0.12}$
Double	$\log k_3$	$-23.83^{+0.02}_{-0.02}$	$-24.09^{+0.08}_{-0.08}$	$-24.16^{+0.04}_{-0.04}$	$-23.69^{+0.03}_{-0.03}$	$-23.51^{+0.03}_{-0.03}$	$-23.76^{+0.04}_{-0.04}$
	$\log N_d$	$21.50^{+0.05}_{-0.04}$	$21.65^{+0.15}_{-0.08}$	$21.50^{+0.18}_{-0.07}$	$21.45^{+0.12}_{-0.05}$	$21.45^{+0.09}_{-0.06}$	$21.50^{+0.20}_{-0.07}$
	α_3	$-2.00^{+0.06}_{-0.06}$	$-1.96^{+0.23}_{-0.26}$	$-2.25^{+0.16}_{-0.17}$	$-1.94^{+0.10}_{-0.10}$	$-1.81^{+0.10}_{-0.11}$	$-2.03^{+0.12}_{-0.13}$
	α_4	$-6.00^{+4.06}_{-3.93}$	$-10.00^{+8.27}_{-7.78}$	$-10.00^{+7.91}_{-7.59}$	$-4.26^{+2.41}_{-2.22}$	$-5.72^{+4.02}_{-3.81}$	$-4.78^{+2.87}_{-2.63}$

[a]Restricted to the SDSS-DR3_4 sample. The remaining columns include the full sample.

Note. — The errors reported are one-parameter errors which do not account for correlations among the parameters. See Table 7 for the correlation matrix.

TABLE 7
CORRELATION MATRIX FOR $f_{\rm HI}(N,X)$

Param	σ	$\delta\log k_1$	$\delta\alpha_1$	$\delta\log k_2$	$\delta\log N_\gamma$	$\delta\alpha_2$	$\delta\log k_3$	$\delta\log N_d$	$\delta\alpha_3$	$\delta\alpha_4$
$\delta\log k_1$	1.04	1.00	−1.00	−0.48	0.28	−0.80	−0.33	−0.01	−0.93	−0.06
$\delta\alpha_1$	0.05	−1.00	1.00	0.48	−0.29	0.80	0.32	0.02	0.93	0.05
$\delta\log k_2$	0.29	−0.48	0.48	1.00	−0.98	0.91	0.47	−0.25	0.66	−0.28
$\delta\log N_\gamma$	0.10	0.28	−0.29	−0.98	1.00	−0.80	−0.45	0.29	−0.50	0.29
$\delta\alpha_2$	0.11	−0.80	0.80	0.91	−0.80	1.00	0.48	−0.17	0.89	−0.15
$\delta\log k_3$	0.18	−0.33	0.32	0.47	−0.45	0.48	1.00	−0.93	0.61	0.53
$\delta\log N_d$	0.08	−0.01	0.02	−0.25	0.29	−0.17	−0.93	1.00	−0.29	−0.64
$\delta\alpha_3$	0.07	−0.93	0.93	0.66	−0.50	0.89	0.61	−0.29	1.00	0.09
$\delta\alpha_4$	2.13	−0.06	0.05	−0.28	0.29	−0.15	0.53	−0.64	0.09	1.00

Note. — Restricted to the SDSS-DR3_4 sample.

fit to the observations; a one-sided KS test gives values $> 10\%$ for the full 1σ range of the parameters. Furthermore, there is good agreement between the 'break' column densities (N_γ and N_d) and the power-law indices at low column densities are consistent. Both solutions indicate that $f_{\rm HI}(N,X)$ drops very steeply ($\alpha \ll -2$) at $N_{\rm HI} \approx 10^{21.5}\,{\rm cm}^{-2}$ and that the distribution has a 'faint-end' slope of $\alpha \approx -2$.

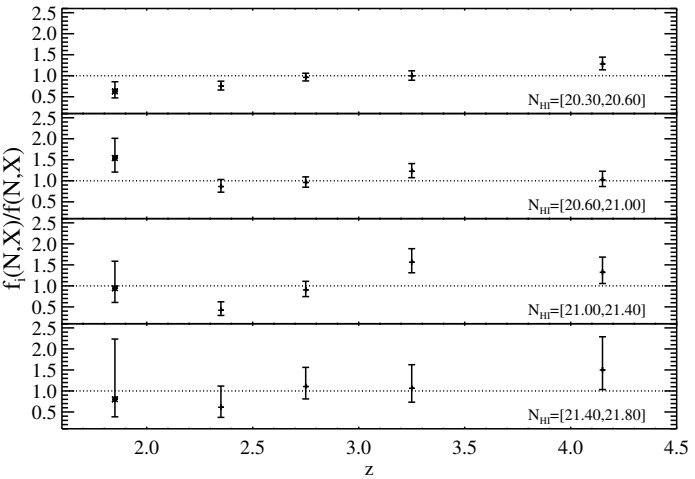

FIG. 7.— Plot of the $N_{\rm HI}$ frequency distribution of damped Lyα systems in 5 redshift intervals against the full distribution. The four $N_{\rm HI}$ intervals are: (1) $20.3 \leq \log N_{\rm HI} < 20.6$; (2) $20.6 \leq \log N_{\rm HI} < 21.0$; (3) $21.0 \leq \log N_{\rm HI} < 21.4$; and (4) $21.4 \leq \log N_{\rm HI} < 21.8$. One observes significant evolution in $f_{\rm HI}(N,X)$ at low $N_{\rm HI}$ value but that the shape of $f_{\rm HI}(N,X)$ is nearly invariant with redshift. Finally, note that the data points for $z < 2.2$ (marked with a cross) do not include measurements from the SDSS survey.

In the following sub-sections, we will consider evolution in the zeroth and first moments of $f_{\rm HI}(N,X)$. Figure 7 qualitatively describes the redshift evolution of the full $f_{\rm HI}(N,X)$ distribution. The figure plots $f_{\rm HI}(N,X)$ as a function of redshift against the $f_{\rm HI}(N,X)$ distribution of the full SDSS-DR3 sample in four $N_{\rm HI}$ intervals: (1) $20.3 \leq \log N_{\rm HI} < 20.6$; (2) $20.6 \leq \log N_{\rm HI} < 21.0$; (3) $21.0 \leq \log N_{\rm HI} < 21.4$; and (4) $21.4 \leq \log N_{\rm HI} < 21.8$. At low $N_{\rm HI}$, $f_{\rm HI}(N,X)$ increases monotonically with redshift. The peak-to-peak range is modest, however, only evolving by a factor of ≈ 2. One identifies similar evolution in the high $N_{\rm HI}$ bins with the exception of the highest redshift interval.

Perhaps the most important result is that the shape of $f_{\rm HI}(N,X)$ is independent of redshift. We have compared the $f_{\rm HI}(N,X)$ distributions from each redshift bin using a two-sided Komolgorov-Smirnov (KS) test. This test compares the shape of the distributions but is insensitive to the normalization. Our analysis found KS probabilities greater than 10% for every pair of redshift intervals. Contrary to previous claims, therefore, there is no evidence for a significant evolution in $f_{\rm HI}(N,X)$ for the DLAs with redshift (e.g. a steeping beyond $z = 3$). With the current sample, there may only be significant evolution in the normalization of $f_{\rm HI}(N,X)$.

4.2. $\ell_{\rm DLA}(X)dX$: *The Incidence of Damped Lyα Systems*

The zeroth moment of $f_{\rm HI}(N,X)$ gives the number of damped Lyα systems encountered per unit absorption pathlength dX, i.e. the line-density of damped Lyα systems[4]:

$$\ell_{\rm DLA}(X)dX = \int_{N_{th}}^{\infty} f_{\rm HI}(N,X)dNdX \quad . \qquad (13)$$

The line density is related to the comoving number density of damped Lyα systems $n_{DLA}(X)$ and the cross-section $A(X)$ as follows:

$$\ell_{\rm DLA}(X) = (c/H_0)n_{DLA}(X)A(X) \quad . \qquad (14)$$

In this manner, $\ell_{\rm DLA}$ is related to the covering fraction of damped Lyα systems on the sky.

Figure 8 shows $\ell_{\rm DLA}(X)dX$ for the damped Lyα systems at $z > 1.6$ and for the local universe (Zwaan, Briggs, & Sprayberry 2001; Rosenberg & Schneider 2003; Ryan-Weber, Webster, & Staveley-Smith 2003, 2005). The values for the damped Lyα systems were calculated in the discrete limit, e.g.

$$\ell_{\rm DLA}(X) = \frac{m_{DLA}}{\Delta X} \qquad (15)$$

and the error bars reflect Poisson uncertainty in m_{DLA} (68.3% c.l.). The solid line traces the value of $\ell_{\rm DLA}(X)$ derived in a series of 0.5 Gyr time intervals to reveal any differences in binning.

It is evident from the figure that the line density of damped Lyα systems increases from $z = 2$ to 4; the increase between the $z = [2.2, 2.5]$ interval and the $z =$

[4]Previous work has represented this quantity with various terminology that is often confused with the column density or number density (e.g. $dN/dX, dn/dz$).

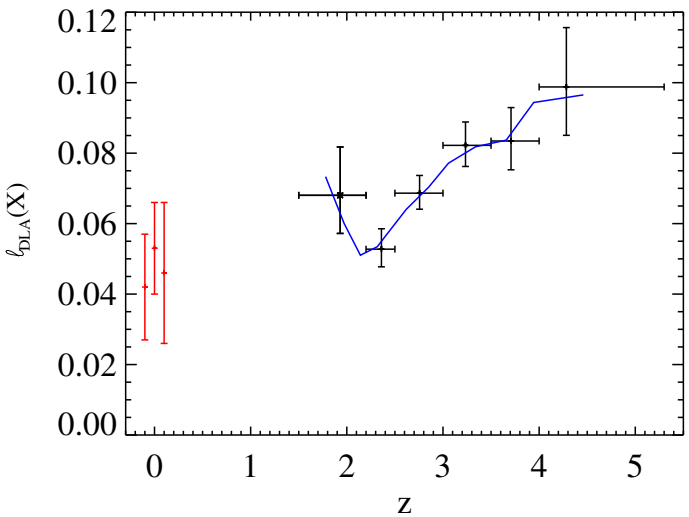

FIG. 8.— Plot of the line density of damped Lyα systems $\ell_{\rm DLA}(X)$ versus redshift. The points at $z = 0$ are from three analyses of 21cm observations by Zwaan, Briggs, & Sprayberry (2001); Rosenberg & Schneider (2003); Ryan-Weber, Webster, & Staveley-Smith (2003, 2005). The curve overplotted on the data traces the evolution of $\ell_{\rm DLA}(X)$ in a series of 0.5 Gyr intervals. Contrary to previous studies (which focused on the line density in redshift space $\ell_{\rm DLA}(z)$), we find statistically significant evolution in the line density per unit absorption distance.

[3.5, 5, 5] bin is a factor of 1.7 ± 0.2. The largest change in $\ell_{\rm DLA}(X)$ occurs during a $\Delta z = 1$ interval from $z \approx 2.3$ to 3.3 corresponding to $\Delta t = 800\,{\rm Myr}$. We discuss possible explanations for this evolution in § 6. Another interesting result is that $\ell_{\rm DLA}(z \approx 2.3)$ is less than 2σ larger than $\ell_{\rm DLA}(z = 0)$, i.e. the data suggest little evolution in the line density of damped systems over the past ≈ 10Gyr. Within the context of hierarchical cosmology, this is a surprising result. It implies that when smaller galaxies merge and accrete other systems, the product of the comoving density and the total gas cross-section satisfying the damped Lyα criterion is conserved. Perhaps this is related to the fact that the damped Lyα threshold is near the surface density threshold for star formation.

4.3. Ω_g: *The Cosmological Neutral Gas Mass Density*

4.3.1. *Definition of* Ω_g

The first moment of the $f_{\rm HI}(N, X)$ distribution gives a cosmologically meaningful quantity, the gas mass density of H I atoms:

$$\Omega_g^{\rm HI}(X) dX \equiv \frac{\mu m_H H_0}{c\rho_c} \int_{N_{min}}^{N_{max}} N_{\rm HI} f_{\rm HI}(N, X)\, dX \qquad (16)$$

where μ is the mean molecular mass of the gas (taken to be 1.3), H_0 is Hubble's constant, and ρ_c is the critical mass density. Traditional treatments of $\Omega_g^{\rm HI}$ and the damped Lyα systems have integrated Equation 16 from $N_{min} = 2 \times 10^{20}\,{\rm cm}^{-2}$ to $N_{max} = \infty$ yielding $\Omega_g^{\rm DLA}$. As emphasized by Péroux et al. (2003), $\Omega_g^{\rm DLA}$ may be significantly less than $\Omega_g^{\rm HI}$ if absorbers below the damped Lyα threshold contribute to the H I mass density. In this case, $\Omega_g^{\rm HI}$ would include a large contribution from gas which is predominantly ionized because the H I atoms in the Lyman limit systems are a mere tracer of the gas. It is difficult, however, to assign any physical significance to $\Omega_g^{\rm HI}$ aside from a mere census of the H I atoms in the universe. In contrast, $\Omega_g^{\rm DLA}$ offers a good estimate of the mass density of gas which is predominantly neutral (see below).

In the following, we will consider both quantities with primary emphasis on $\Omega_g^{\rm DLA}$ because we contend it best defines the gas reservoir available for star formation at high redshift. In practice, one generally evaluates $\Omega_g^{\rm DLA}$ in the discrete limit

$$\Omega_g^{\rm DLA} = \frac{\mu m_H H_0}{c\rho_c} \frac{\Sigma N_{\rm HI}}{\Delta X} \quad , \qquad (17)$$

where the sum is performed over the $N_{\rm HI}$ measurements of the damped Lyα systems in a given redshift interval with total pathlength ΔX.

In the following, we will consider several definitions for Ω_g based on the values of N_{min} and N_{max}. Table 8 summarizes the various definitions.

4.3.2. *Convergence of* Ω_g

Consider first the upper limit, N_{max}. To verify Ω_g converges, it is necessary to integrate $f_{\rm HI}(N, X)$ until

$$\frac{d \log f_{\rm HI}(N, X)}{d \log N_{\rm HI}} \ll -2 \quad , \qquad (18)$$

TABLE 8
DEFINITIONS OF Ω_g

Def.	N_{min} ($\rm cm^{-2}$)	N_{max} ($\rm cm^{-2}$)	Description
$\Omega_g^{\rm HI}$	0	∞	Density of gas associated with H I atoms
Ω_g^{Neut}	??	∞	Density of predominantly neutral gas
$\Omega_g^{\rm DLA}$	2×10^{20}	∞	Density of gas associated with the DLA
$\Omega_g^{\rm LLS}$	1.6×10^{17}	2×10^{20}	Density of the gas associated with LLS
Ω_g^{21cm}	—	—	Density of H I gas in collapsed objects

i.e. until one establishes that $f_{\rm HI}(N, X)$ is significantly steeper than a $N_{\rm HI}^{-2}$ power-law. As noted in § 4.1 and Figure 6, the Γ-function and double power-law fits to the $f_{\rm HI}(N, X)$ distribution indicate $f_{\rm HI}(N, X)$ steepens at $N_{\rm HI} \approx 10^{21.5}\,{\rm cm}^{-2}$. Furthermore, even a single power-law fit to the data almost satisfies Equation 18. This point is emphasized in Figure 9 which presents the exponent of the best fitting single power-law to the SDSS-DR3 $f_{\rm HI}(N, X)$ distribution function as a function of the lower $N_{\rm HI}$ bound to the distribution function, e.g., the point at $\log N_{\rm HI}^* = 21$ shows the exponent for the single power-law fit to $f_{\rm HI}(N_{\rm HI} > 10^{21}\,{\rm cm}^{-2}; X)$. Not surprisingly, the curve starts at $N_{\rm HI} = 2 \times 10^{20}\,{\rm cm}^{-2}$ with $\alpha_1 = -2.2$ and decreases with increasing $N_{\rm HI}^*$. In addition, it is important that the value of α_1 decreases below -3 in Figure 9 because a power-law with $\alpha_1 = -2.2$ converges very slowly with N_{max}.

Figure 10 shows the cumulative $\Omega_g^{\rm DLA}$ value against $\log N_{max}$. Both the double power-law and Γ-function converge to the value indicated by the discrete evaluation of $\Omega_g^{\rm DLA}$ (Equation 17). This adds additional confidence that these func-

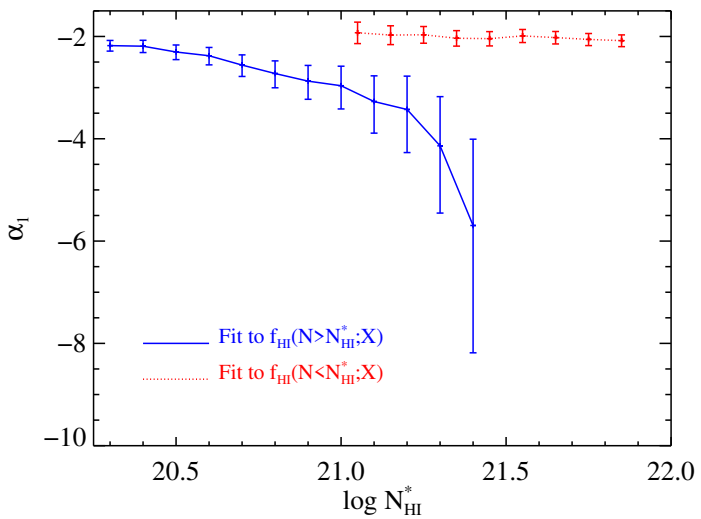

FIG. 9.— The solid line plots the maximum likelihood value of the exponent for a single power-law fit to the $f_{\rm HI}(N, X)$ distribution of the SDSS-DR3_4 sample as a function of the minimum $N_{\rm HI}$ value of the distribution. The value is less than -2 for all values of $N_{\rm HI}^*$ and the curve indicates the power-law steepens as one restricts the analysis to the high $N_{\rm HI}$ end of the distribution function. The dotted curve, meanwhile, plots the exponent value as a function of the maximum $N_{\rm HI}$ value of the $f_{\rm HI}(N, X)$ distribution. There is little evolution in this value although the slope is more shallow at the low $N_{\rm HI}$ region of the distribution.

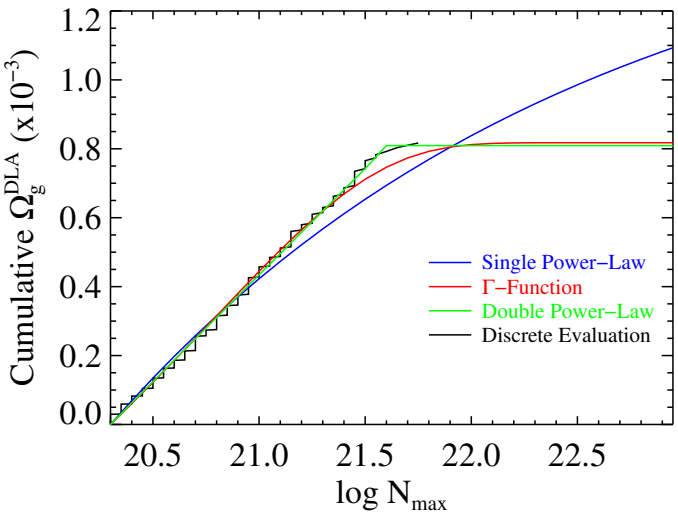

FIG. 10.— Cumulative evaluation of $\Omega_g^{\rm DLA}$ for the SDSS-DR3_4 sample as a function of the maximum $N_{\rm HI}$ value. The black curves trace the evaluation of $\Omega_g^{\rm DLA}$ in the discrete limit (Equation 17) while the smooth curves show the evaluation by integrating the functional fits to the $f_{\rm HI}(N,X)$ distribution. Both the double power-law and Γ-function converge to the discrete value by $N_{\rm HI} = 10^{22}\,{\rm cm}^{-2}$. An extrapolation of the single power-law, however, has not converged even at $N_{\rm HI} = 10^{23}\,{\rm cm}^{-2}$.

tional forms are good representations of $f_{\rm HI}(N,X)$. The single power-law, however, has not converged at the highest $N_{\rm HI}$ value observed for a damped Lyα system; it implies significantly larger $\Omega_g^{\rm DLA}$ values than currently supported by the data. Nevertheless, even this functional form converges formally as $N_{max} \rightarrow \infty$: $\alpha_1 < -2$ at 3σ significance. *The results presented here are the first definitive demonstration that* $\Omega_g^{\rm DLA}$ *converges and most likely by* $N_{\rm HI} \approx 10^{22}\,{\rm cm}^{-2}$. We emphasize that smaller redshift intervals (i.e. subsets of the SDSS-DR3 sample) have insufficient sample size to convincingly argue for convergence. Therefore, one may cautiously view the $\Omega_g^{\rm DLA}$ values presented as a function of redshift.

Now consider the lower bound to the integrand, N_{min}. One evaluates $\Omega_g^{\rm HI}$ with $N_{min} = 0$, i.e. including the contribution from all classes of quasar absorption line systems. We have argued that this definition for Ω_g lacks special physical significance, yet let us first consider its value. In particular, we will investigate the contribution of $\Omega_g^{\rm DLA}$ to $\Omega_g^{\rm HI}$. Tytler (1987) first demonstrated that the H I frequency distribution of the Lyα forest (absorbers with $N_{\rm HI} < 10^{17}\,{\rm cm}^{-2}$) is well described by a single power-law with exponent $\alpha_{{\rm Ly}\alpha} \approx -1.5$. More recent studies give similar results over the $N_{\rm HI}$ interval $10^{12} - 10^{15}\,{\rm cm}^{-2}$ (e.g. Hu et al. 1995; Kirkman & Tytler 1997). This functional form is sufficiently shallow that the contribution of the Lyα forest to $\Omega_g^{\rm HI}$ is negligible in comparison to absorbers with $N_{\rm HI} > 10^{15}\,{\rm cm}^{-2}$.

4.3.3. *Impact of the Lyman Limit Systems*

Of greater interest to this discussion is the contribution of the Lyman limit systems[5] to $\Omega_g^{\rm HI}$. Indeed, Péroux et al. (2003) claimed that Lyman limit systems with $N_{\rm HI} > 10^{19}\,{\rm cm}^{-2}$ contribute ≈ 50% of $\Omega_g^{\rm HI}$ at $z > 3.5$. In PH04, we argued that Péroux et al. (2003) overstated the contribution of these 'super-LLS' to $\Omega_g^{\rm HI}$ and, in particular, underestimated the uncertainty of their contribution. Let us return to this issue with the increased sample of SDSS-DR3.

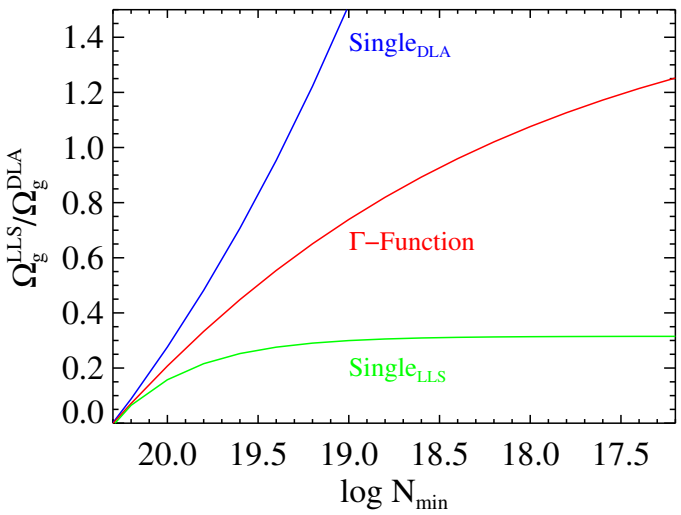

FIG. 11.— Contribution to $\Omega_g^{\rm HI}$ of the Lyman limit systems relative to the damped Lyα systems as a function of the minimum $N_{\rm HI}$ value of the Lyman limit systems. The top two curves refer to extrapolations of the single power-law and Γ-function fits to $f_{\rm HI}(N,X)$ for the damped Lyα systems. The single-power law is divergent and even the Γ-function has not converged at $10^{17.2}\,{\rm cm}^{-2}$. These extrapolations, however, significantly overpredict the line density of Lyman limit systems and should be considered upper limits to $\Omega_g^{\rm LLS}/\Omega_g^{\rm DLA}$. The lower curve is the result for a single power-law fit to $f_{\rm HI}(N,X)$ in the $N_{\rm HI}$ interval for Lyman limit systems. Because the first derivative of this single power law is not continuous at the damped Lyα system threshold or the transition to the Lyα forest, it should be considered a lower limit to $\Omega_g^{\rm LLS}/\Omega_g^{\rm DLA}$.

The first point to stress is that $d\log f_{\rm HI}(N,X)/d\log N_{\rm HI} \approx -2$ at $N_{\rm HI} \approx 10^{20.3}\,{\rm cm}^{-2}$. This is sufficiently steep that absorbers with $N_{\rm HI} \approx 10^{20}\,{\rm cm}^{-2}$ must contribute at least a few percent of $\Omega_g^{\rm HI}$. Indeed, if we extrapolate our fits to $f_{\rm HI}(N,X)$ for the damped Lyα systems from $N_{\rm HI} = 2\times10^{20}\,{\rm cm}^{-2}$ to $10^{17}\,{\rm cm}^{-2}$, we find the Lyman limit systems would give a larger neutral gas mass density than the value of the damped Lyα systems. Figure 11 presents the contribution of the Lyman limit systems to Ω_g relative to the damped Lyα systems $\Omega_g^{\rm LLS}/\Omega_g^{\rm DLA}$. The upper curves show the results for extrapolations of the the single power-law and the Γ-function fits to $f_{\rm HI}(N,X)$ for the damped Lyα systems. For the single power-law, the gas mass density diverges as $N_{min} \rightarrow 0$ because $\alpha_1 < -2$. Even with the Γ-function (with exponent $\alpha_2 = -1.8$) the Lyman limit systems dominate the contribution to $\Omega_g^{\rm HI}$.

There is, however, an additional observational constraint which limits $\Omega_g^{\rm LLS}/\Omega_g^{\rm DLA}$. Although the functional form of $f_{\rm HI}(N,X)$ is essentially unconstrained for the Lyman limit systems, their line density,

[5]Lyman limit systems are defined to be all absorbers with optical depth exceeding unity at 1 Ryd. In this discussion, we generally restrict the definition to be absorbers with $17.2 \le \log(N_{\rm HI}/\,{\rm cm}^{-2}) < 20.3$. The one exception is the line density, as defined in Equation 19.

$$\ell_{LLS}(X) \equiv \ell(N > 10^{17.2}\,\mathrm{cm}^{-2}; X) \quad , \qquad (19)$$

has been measured at $z > 2$ (Péroux et al. 2001). The extrapolations of the single power-law and Γ-function fits to the damped Lyα frequency distribution overpredict the observed value of the incidence of Lyman limit systems by more than an order of magnitude. Therefore, the estimate of $\Omega_g^{\rm LLS}$ for these curves should be considered unrealistic upper limits.

We can set a lower bound to $\Omega_g^{\rm LLS}$ by assuming $f_{\rm HI}(N, X)$ is a single power-law over the $10^{17.2}\,\mathrm{cm}^{-2} < N_{\rm HI} < 10^{20.3}\,\mathrm{cm}^{-2}$ interval:

$$f_{\rm HI}^{\rm LLS}(N, X) = k_{\rm LLS} \left(\frac{N}{10^{20.3}\,\mathrm{cm}^{-2}} \right)^{\alpha_{\rm LLS}} \quad . \qquad (20)$$

The two parameters of this function are determined by demanding

$$f_{\rm HI}^{\rm LLS}(10^{20.3}\,\mathrm{cm}^{-2}; X) = f_{\rm HI}^{\rm DLA}(10^{20.3}\,\mathrm{cm}^{-2}; X) \qquad (21)$$

and

$$\int_{10^{17.2}}^{10^{20.3}} f_{\rm HI}^{\rm LLS} dN\, dX = \ell_{\rm LLS} dX - \ell_{\rm DLA} dX \quad . \qquad (22)$$

To estimate $\ell_{\rm LLS}(X)$, we adopt the functional form provided by Péroux et al. (2003) as a function of redshift $\ell_{\rm LLS}(z) = 0.07(1+z)^{2.45}$, and adopt the mean value for the entire SDSS-DR3 survey, i.e.

$$\ell_{\rm LLS}(X) = \frac{\int \ell_{\rm LLS}(z) g(z) dz}{\Delta X} \quad . \qquad (23)$$

The uncertainty in $\ell_{\rm LLS}(z)$ implies an approximately 20% uncertainty in $\ell_{\rm LLS}(X)$. To estimate $f_{\rm HI}^{\rm DLA}(N = 10^{20.3}\,\mathrm{cm}^{-2})$ and determine $k_{\rm LLS}$, we adopt the central value of our Γ-function fit. The relative contribution of $\Omega_g^{\rm LLS}/\Omega_g^{\rm DLA}$, of course, is relatively insensitive to this parameter. For the full SDSS-DR3_4 sample, we find $\log k_{\rm LLS} = -21.43$ and $\alpha_{\rm LLS} = -1.01^{+0.09}_{-0.02}$. It is not coincidental that this slope matches the best-fit values of the 'faint-end' slope for the Γ-function fits to $f_{\rm HI}(N, X)$ performed by Storrie-Lombardi & Wolfe (2000) and Péroux et al. (2003). Their maximum likelihood analyses included the integral constraint imposed by the Lyman limit systems and the greater than three orders of magnitude lever arm in $N_{\rm HI}$ forces this result.

Because the LLS power-law is more shallow than the 'faint-end' slope of the fits to $f_{\rm HI}^{\rm DLA}$, the estimate of $\Omega_g^{\rm LLS}$ based on $f_{\rm HI}^{\rm LLS}$ must be considered a lower limit. The $\Omega_g^{\rm LLS}/\Omega_g^{\rm DLA}$ curve for the single-power law fit to $f_{\rm HI}^{\rm LLS}$ is also shown in Figure 11. We find $\Omega_g^{\rm LLS} > 0.3\,\Omega_g^{\rm DLA}$ and that the majority of the contribution comes from absorbers with $N_{\rm HI} > 10^{19}\,\mathrm{cm}^{-2}$. As noted above, Péroux et al. (2003) first cautioned that the super-LLS could contribute significantly to $\Omega_g^{\rm HI}$, especially at very high redshift. In PH04, we argued that Péroux et al. (2003) had overstated the effect at $z > 3.5$ and underestimated the uncertainty. Our new results indicate the super-LLS contribute from 20 to 50% to $\Omega_g^{\rm HI}$ for the complete SDSS-DR3 survey. Also, recall that the shape of $f_{\rm HI}(N, X)$ is statistically invariant with redshift (Figure 7). Therefore, it is likely that the Lyman limit systems offer a non-negligible contribution to $\Omega_g^{\rm HI}$ at all redshifts. To determine the value of $\Omega_g^{\rm LLS}$ as a function of redshift, one compares the relative evolution of $\ell_{\rm LLS}$ and $\ell_{\rm DLA}$. Current observations suggest $\ell_{\rm LLS}$ has a stronger redshift dependence than $\ell_{\rm DLA}$. In fact, we do find $\Omega_g^{\rm LLS}$ (as determined from the single power-law fit to $f_{\rm HI}^{\rm LLS}$) to increase with redshift (Table 9). There is a comparable increase in $\Omega_g^{\rm DLA}$, however, and the ratio $\Omega_g^{\rm LLS}/\Omega_g^{\rm DLA}$ is roughly constant in time.

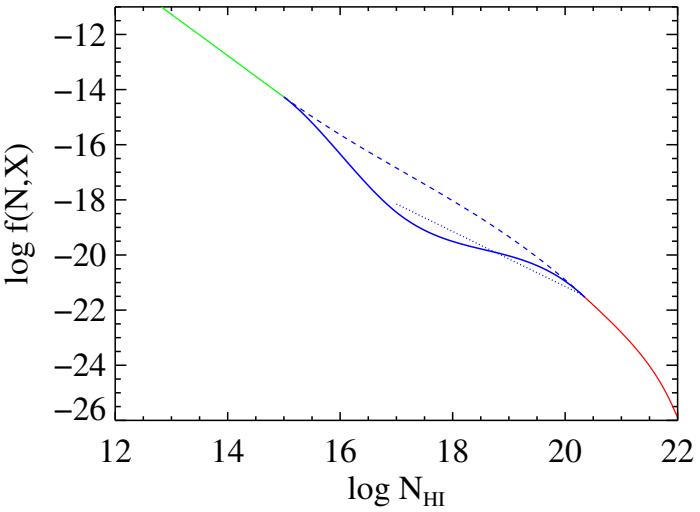

FIG. 12.— Figure depicting the $f_{\rm HI}(N, X)$ distribution of the quasar absorption line systems at $z = 2.7$. The distribution for the damped Lyα systems corresponds to the Γ-function fit to the SDSS-DR3_4 sample. The distribution for the Lyα forest ($N_{\rm HI} < 10^{15}\,\mathrm{cm}^{-2}$) is taken from Kirkman & Tytler (1997) scaled to the ΛCDM cosmology. We show three curves for the Lyman limit system regime which should only be considered as a sketch: (1) the dotted curve is the single power-law fit to $f_{\rm HI}^{\rm LLS}$. It is constrained to match the incidence of Lyman limit systems at $z \sim 2.7$; (2) the dashed curve is a spline function constrained to match the Lyα forest and damped Lyα systems. It clearly overpredicts the incidence of Lyman limit systems; and (3) the solid curve is a spline function fit to the Lyα forest and damped Lyα systems and also constrained by a single point at $N_{\rm HI} = 10^{17}\,\mathrm{cm}^{-2}$ defined such that the integral constraint for the Lyman limit systems is satisfied.

We reemphasize that the value of $\Omega_g^{\rm LLS}$ derived from the single-power law fit to $f_{\rm HI}^{\rm LLS}$ is an underestimate of the true value. In Figure 12, we show a spline solution which matches the $f_{\rm HI}(N, X)$ functional forms of the Lyα forest at $N_{\rm HI} = 10^{15}\,\mathrm{cm}^{-2}$ and the damped Lyα systems at $N_{\rm HI} = 2 \times 10^{20}\,\mathrm{cm}^{-2}$ and also gives the correct line density of Lyman limit systems. We can give an estimate of the LLS contribution based on this spline function: $\Omega_g^{\rm LLS}/\Omega_g^{\rm DLA} = 0.57$. Therefore, the likely contribution of the Lyman limit systems to $\Omega_g^{\rm HI}$ is $\approx 50\%\,\Omega_g^{\rm DLA}$ and they comprise $\approx 1/3$ of $\Omega_g^{\rm HI}$. In § 6, we consider an alternate definition of the gas mass density: the mass density of gas that is predominantly neutral, $\Omega_g^{\rm Neut}$. Under this definition, one may disregard the majority of the Lyman limit systems on the grounds that they are highly ionized (Prochaska & Wolfe 1996; Prochaska 1999; Prochaska & Burles 1999; Dessauges-Zavadsky et al. 2003), but we

suspect that at least a subset with $N_{\rm HI} \approx 10^{20}\,{\rm cm}^{-2}$ is significantly neutral and could feed star formation. This neutral subset, however, may contribute only a few percent relative to $\Omega_g^{\rm Neut}$ and we contend that the damped Lyα systems contain $> 80\%$ of $\Omega_g^{\rm Neut}$ at high redshift. Until a robust determination of $f_{\rm HI}(N,X)$ and the ionization state of the gas in the Lyman limit systems is made, the only conservative practice is to restrict the discussion of $\Omega_g^{\rm HI}$ to include only those regions of the universe with surface density $N_{\rm HI} \geq 2 \times 10^{20}\,{\rm cm}^{-2}$ (i.e. $\Omega_g^{\rm DLA}$). For the near future, this is the only quantity that will be precisely measured.

As an aside, consider the implications of the $\alpha_{\rm LLS}$ value. First, this value is significantly more shallow than the power-law dependence of the Lyα forest. Second, it is shallower than the 'faint-end' slope of the damped Lyα $f_{\rm HI}(N,X)$ distribution. If we demand that $f_{\rm HI}(N,X)$ and its first derivative are continuous across the Lyα forest/LLS and LLS/DLA boundaries, then our results indicate that $f_{\rm HI}^{\rm LLS}$ has an inflection with $d\log f_{\rm HI}^{\rm LLS}/d\log N > -1$ (Figure 12). It is not surprising that $f_{\rm HI}^{\rm LLS}$ would have an unusual functional form given that it spans the regime from highly ionized gas to primarily neutral gas. In fact, the inflection indicated by our analysis resembles the functional forms suggested by Zheng & Miralda-Escudé (2002) and Maller et al. (2003). In passing, we advertise that a principal goal of the MIKE/HIRES Lyman Limit Survey is to study $f_{\rm HI}^{\rm LLS}$ and the ionization state of the Lyman limit systems (Prochaska, Burles, Prochter, O'Meara, Bernstein, & Schectman, in prep.).

4.3.4. $\Omega_g^{\rm DLA}$ *Evolution*

Let us consider[6] the redshift evolution of $\Omega_g^{\rm DLA}$ by evaluating Equation 17 in a series of redshift intervals. Figure 13 presents the results. Following PH04, we estimate the uncertainty in $\Omega_g^{\rm DLA}$ through a modified bootstrap technique. Specifically, we perform 5000 trials where we randomly draw p damped Lyα systems from the observed sample where p is a normally distributed random number with mean equal to the number of damped systems in the interval (m_{DLA}) and with variance also equal to m_{DLA}. We calculate $\Omega_g^{\rm DLA}$ for each trial and calculate the upper and lower values corresponding to 68.3% of the distribution.

The results presented in Figure 13 present the first statistically significant evidence that $\Omega_g^{\rm DLA}$ evolves with redshift. Comparing the measurements in the $z = [2.2, 2.5]$ and $z = [3., 3.5]$ intervals, we find a $50 \pm 10\%$ decrease in the lowest SDSS redshift bin. This evolution occurs during the same time interval when the decrease in $\ell_{\rm DLA}(X)$ is observed (Figure 8). Regarding the line density, we have attributed the decline to a decrease in the typical cross-section of the damped Lyα systems. Given the decrease in $\Omega_g^{\rm DLA}$ it is evident that the gas is not simply contracting into smaller structures (e.g. via dissipative cooling), but that the gas is being consumed and/or expelled from the system.

[6]The reader should refer to § 5.2 for a discussion of an important systematic effect related to $\Omega_g^{\rm DLA}$. For the remainder of this paper, we present results which ignore the effect. Our present concern is that we may be overestimating $\Omega_g^{\rm DLA}$.

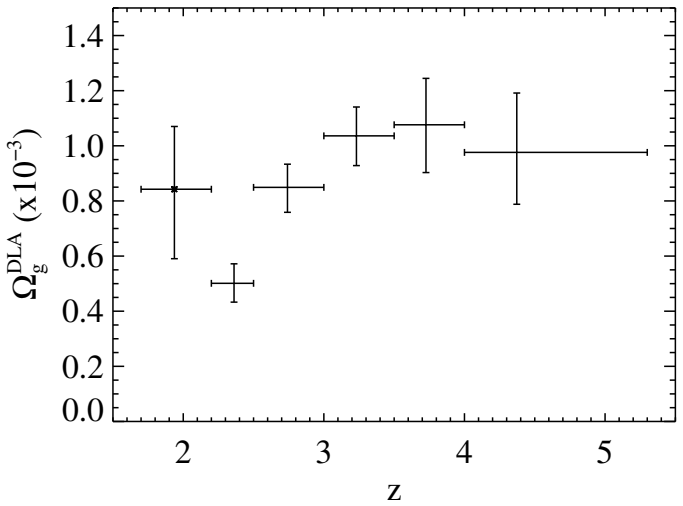

FIG. 13.— Neutral gas mass density of the damped Lyα systems alone as a function of redshift. There is an increase of approximately a factor of two in $\Omega_g^{\rm DLA}$ from $z = 2$ to 3 with the majority of rise occurring in only 500 Myr. Note the data point at $z < 2.2$ (marked with a cross) does not include measurements from the SDSS survey.

Contrary to previous claims, we find no evidence for evolution in $\Omega_g^{\rm DLA}$ at $z > 3.5$ although a modest increase (or decrease) is permitted by the observations. Similarly, the results at $z \approx 2$ from previous work are consistent (albeit uncomfortably high[7]) with the $z = 2.3$ data point. We also caution the reader that the results at $z > 4$ should be confirmed by higher resolution observations. At these redshifts, blending in the Lyα forest is severe and we worry that the SDSS spectroscopic resolution is insufficient. In passing, we also note that a significant contribution to $\Omega_g^{\rm DLA}$ at $z > 4$ comes along the single sightline to J162626.5+275132 which shows two $N_{\rm HI} \approx 10^{21}\,{\rm cm}^{-2}$ damped Lyα systems at $z \approx 5$ and even two additional systems with comparable $N_{\rm HI}$ value at $z \sim 4.5$ which is not in the statistical sample.

4.4. *Summary*

Table 9 presents a summary of $f_{\rm HI}(N,X)$ and its zeroth and first moments for a series of redshift intervals. A qualitative summary of the new results is that they are in broad agreement with the previous analyses based on substantially smaller datasets (Wolfe et al. 1995; Storrie-Lombardi & Wolfe 2000; Péroux et al. 2003). With the increased sample size, however, we have measured evolution in the incidence and H I content of these galaxies at $z > 2$ to unprecedented precision.

[7]Note that the value for $\Omega_g^{\rm DLA}$ is reduced by $\approx 50\%$ if one removes the single damped system at $z = 2.04$ toward Q0458−02. This stresses the sensitivity of these low redshift results to small number statistics (as stressed by Chen et al. 2005) and the bootstrap error estimate may be overly optimistic in this case. Coincidentally, Q0458−02 is an optically variable quasar which in its low state has a magnitude below the original survey detection limit. Also note that this redshift interval does not include any data from the SDSS survey.

TABLE 9
SUMMARY

z	dX	$m_{\rm DLA}$	$\bar{z}^a$	$\log f_{\rm HI}(N,X)$				$\ell_{\rm DLA}(X)$	$\Omega_g^{\rm DLA}$	$\alpha_{\rm LLS}$	$\Omega_g^{\rm LLS}$
				$N_{\rm HI}\epsilon[20.3,20.6)$	$N_{\rm HI}\epsilon[20.6,21.0)$	$N_{\rm HI}\epsilon[21.0,21.4)$	$N_{\rm HI}\epsilon[21.4,21.8)$		($\times 10^{-3}$)		($\times 10^{-3}$)
[2.2,5.5][b]	7333.2	525	3.06	$-21.71^{+0.03}_{-0.03}$	$-22.45^{+0.04}_{-0.03}$	$-23.21^{+0.05}_{-0.05}$	$-24.14^{+0.10}_{-0.09}$	$0.072^{+0.003}_{-0.003}$	$0.817^{+0.050}_{-0.052}$	-1.0	0.26
[1.7,2.2]	420.7	30	1.94	$-21.88^{+0.14}_{-0.12}$	$-22.26^{+0.12}_{-0.11}$	$-23.20^{+0.23}_{-0.19}$	$-24.20^{+0.45}_{-0.32}$	$0.071^{+0.015}_{-0.011}$	$0.842^{+0.228}_{-0.252}$		
[2.2,2.5]	1796.4	95	2.36	$-21.84^{+0.06}_{-0.06}$	$-22.52^{+0.08}_{-0.07}$	$-23.59^{+0.17}_{-0.15}$	$-24.36^{+0.26}_{-0.22}$	$0.053^{+0.006}_{-0.005}$	$0.501^{+0.071}_{-0.068}$	-0.9	0.21
[2.5,3.0]	2984.7	205	2.75	$-21.73^{+0.04}_{-0.04}$	$-22.47^{+0.06}_{-0.05}$	$-23.26^{+0.09}_{-0.08}$	$-24.10^{+0.15}_{-0.13}$	$0.069^{+0.005}_{-0.005}$	$0.849^{+0.084}_{-0.090}$	-1.0	0.24
[3.0,3.5]	2068.7	169	3.22	$-21.72^{+0.05}_{-0.05}$	$-22.36^{+0.06}_{-0.06}$	$-23.02^{+0.08}_{-0.08}$	$-24.12^{+0.18}_{-0.16}$	$0.082^{+0.007}_{-0.006}$	$1.036^{+0.105}_{-0.108}$	-1.0	0.27
[3.5,4.0]	1066.5	89	3.72	$-21.65^{+0.06}_{-0.06}$	$-22.47^{+0.10}_{-0.09}$	$-23.03^{+0.12}_{-0.11}$	$-23.91^{+0.20}_{-0.18}$	$0.083^{+0.009}_{-0.008}$	$1.076^{+0.168}_{-0.173}$	-1.1	0.28
[4.0,5.5]	426.1	42	4.35	$-21.51^{+0.09}_{-0.08}$	$-22.37^{+0.14}_{-0.12}$	$-23.21^{+0.23}_{-0.19}$	$-24.21^{+0.45}_{-0.32}$	$0.099^{+0.017}_{-0.014}$	$0.974^{+0.215}_{-0.188}$	-1.0	0.45
[3.5,5.5]	1492.6	131	3.92	$-21.61^{+0.05}_{-0.05}$	$-22.43^{+0.08}_{-0.07}$	$-23.07^{+0.10}_{-0.10}$	$-23.98^{+0.18}_{-0.16}$	$0.088^{+0.008}_{-0.007}$	$1.047^{+0.127}_{-0.137}$	-1.1	0.32

5. SYSTEMATIC EFFECTS

With the large sample size of the SDSS-DR3, we now have sufficient signal to examine a number of systematic effects on the results of our analysis. This section describes an investigation into a few of these effects. It is advantageous that the entire SDSS sample was observed with the same instrument, reduced with the same software, and analyzed by one group. In contrast, the compilation of Péroux et al. (2003) is based on many surveys carried out on over 10 telescopes and reduced and analyzed by at least five different individuals. This undoubtedly leads to at least minor systematic differences in the surveys, e.g. differences in the continuum placement, systematic effects related to instrumental resolution, and/or color selection of the quasar populations. Of course, if multiple groups report the same value, this lends confidence to the sets of procedures related to damped Lyα surveys.

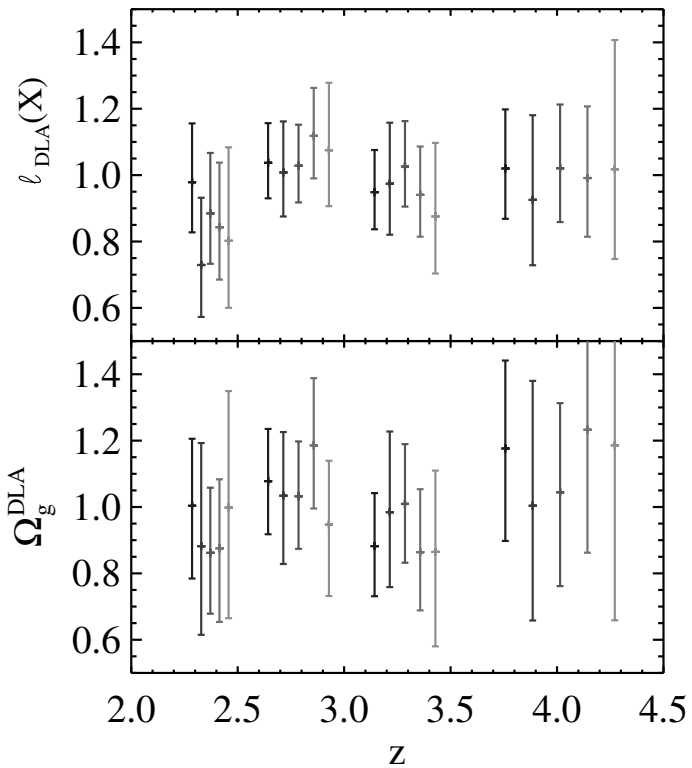

FIG. 14.— The top panel shows the line density as a function of redshift for the various sample cuts on SNR_{lim} and color selection criteria listed in Table 2. The data are plotted relative to the SDSS-DR3_4 sample and are offset in redshift for presentation purposes only. The values are calculated in the redshift intervals: $z = [2.2, 2.5], z = [2.5, 3.], z = [3., 3.5], z = [3.5, 4.4]$. The lower panel shows the results for $\Omega_g^{\rm DLA}$ for the same cuts and also relative to SDSS-DR3_4. One notes very little difference between the various samples.

5.1. *SNR_{lim} and Color Selection*

All of the results presented in this paper are based on searches for damped Lyα systems in the spectral regions of quasars where the median SNR exceeds 4 in a 20 pixel bin (§ 2). Our experience via visual inspection and Monte Carlo simulations is that this limit is just satisfactory. Lowering the limit to a $\mathrm{SNR}_{lim} = 3$ gives many more false positives and a greater number of missed damped Lyα systems. Of course, it is important to test the sensitivity of our results to SNR_{lim}. Figure 14 presents the (a) line density $\ell_{\rm DLA}(X)$ and (b) neutral gas mass density Ω_g relative to the full sample as a function of redshift for several values of SNR_{lim} or the samples restricted to the strict color criteria (Richards et al. 2002). See Table 2 for a description of each cut of the SDSS-DR3.

Examining the figure, we find little difference between the various samples. We find, therefore, no systematic bias related to SNR_{lim} or the color-selection criteria. In fact, the scatter in the data is smaller than that expected for a Poissonian distribution. This is because the samples are not independent. Nevertheless, this analysis demonstrates that our results are relatively insensitive to SNR_{lim} and the color selection criteria.

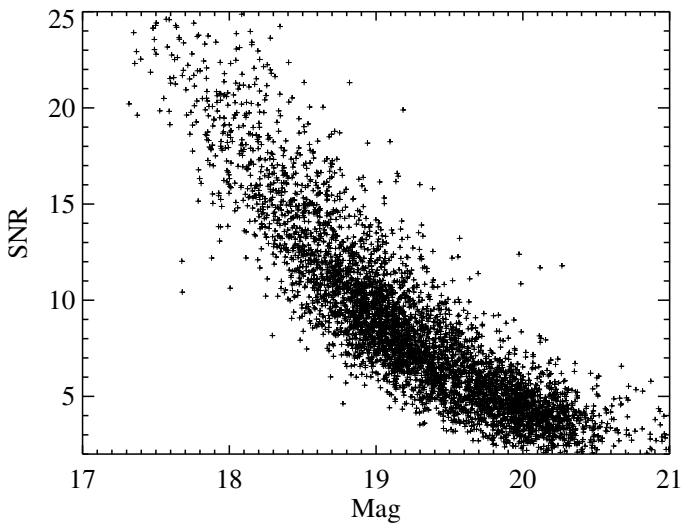

FIG. 15.— Plot of the quasar magnitude at $\lambda_{rest} = 1300$ to 1350Å against the median spectral SNR at $\lambda \approx 1400$Å. There is a good correlation between the two quantities although there are many outliers due primarily to BAL quasars.

5.2. *Quasar magnitude and SNR*

There are a number of reasons to examine the results as a function of the quasar magnitude and/or the SNR of its spectrum. These include considerations of dust obscuration (Fall & Pei 1993), the precision of the $N_{\rm HI}$ values as a function of SNR, and gravitational lensing. We consider the quasar magnitude and spectral SNR together because we have found a similar dependence on the results for these two characteristics. Although there is not a strict one-to-one relationship between quasar magnitude and SNR for the SDSS survey, the two characteristics are closely associated.

We have calculated the magnitude of the quasar as follows: First, we identify the Sloan filter closest to $\lambda = (1 + z_{qso})1325$Å. Second, we adopt the PSF magnitude for the quasar as reported in SDSS-DR3. Third, we flux calibrate the 1D quasar spectrum assuming its relative flux is

accurate. Finally, we calculate the median flux in the interval $\lambda_{rest} = 1300$ to 1350Å (which avoids strong emission line features) and convert to AB magnitude. To estimate the SNR of the spectrum, we calculate the median SNR for $\lambda_{rest} = 1440$ to 1490Å where $z_{qso} < 5$ and $\lambda_{rest} = 1287$ to 1366Å for $z_{qso} > 5$. Figure 15 plots the quasar magnitude versus the spectral SNR. There is a reasonably tight correlation with a few outliers which are primarily exotic BAL quasars.

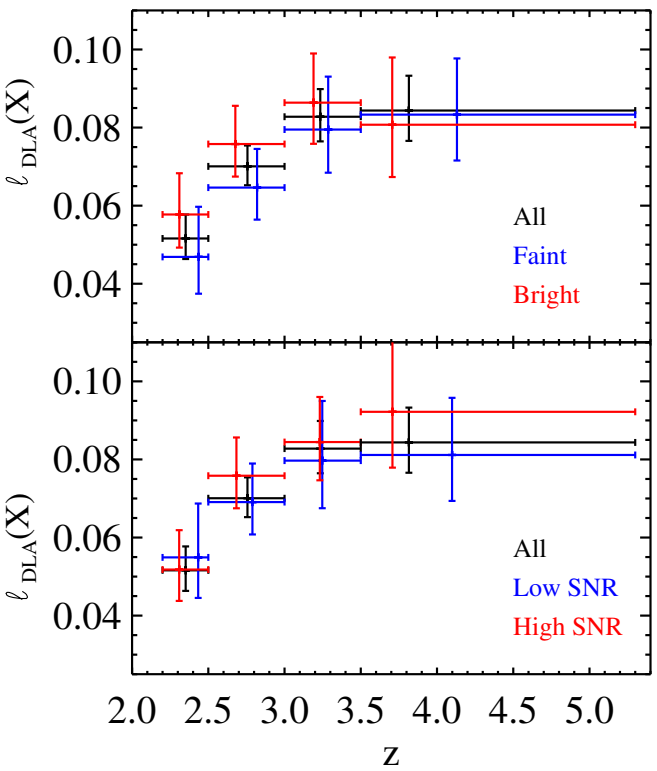

FIG. 16.— Line density as a function of redshift for the full sample, the brightest 33% of the quasars and the faintest 33% of the quasars in each redshift interval. There is a mild trend toward higher line density toward the brighter quasars, but the effect is not statistically significant. Similar results are seen for the quasars cut on spectral SNR.

The upper panel in Figure 16 presents the line density of damped Lyα systems as a function of redshift for three cuts of the SDSS-DR3_4 sample: (i) the complete sample; (ii) the brightest 33% of the quasars; and (iii) the faintest 33% of the quasars. One notes no significant dependence of $\ell_{\rm DLA}$ on quasar magnitude. The lower panel shows the same quantity for cuts on the SNR. Note that the median magnitude of the bright/faint sample for the lowest redshift bin is 18.4 and 19.5 mag respectively. This difference in magnitudes is significant and it nearly brackets the 'break magnitude' of the quasar luminosity function at $z = 2$ (Boyle et al. 2000). The results indicate that if dust obscuration is relevant to damped Lyα surveys, then the line density is insensitive to its effects. While the line density may not be very sensitive to dust obscuration, it is possible that our results violate the predictions of Fall & Pei (1993). A quantitative statement awaits a full treatment which includes the SDSS quasar color selection criteria (Murphy et al., in prep).

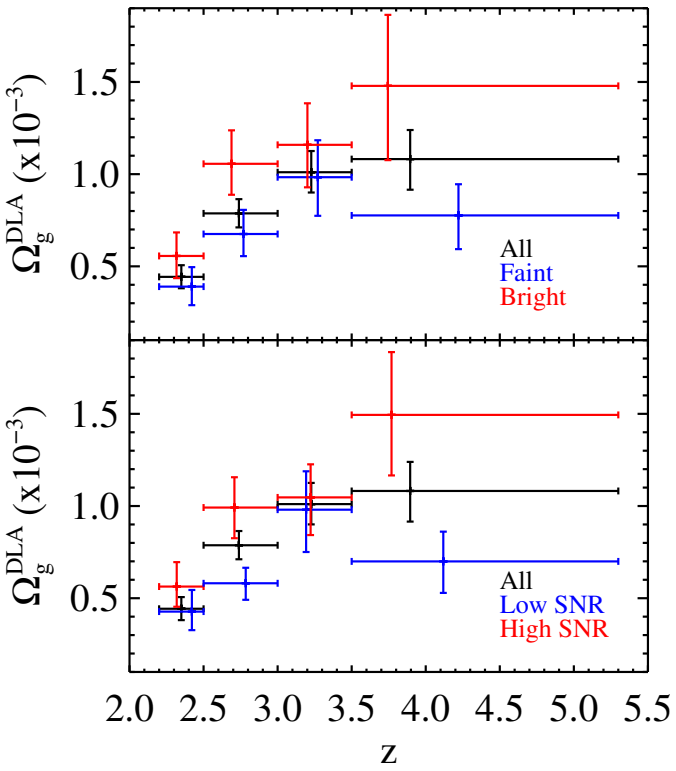

FIG. 17.— Gas mass density as a function of redshift for the full sample, the brightest 33% of the quasars and the faintest 33% of the quasars in each redshift interval. There is a systematic trend of higher $\Omega_g^{\rm DLA}$ values toward the brighter quasars. Although the effect is not significant at the 2σ level in any given redshift bin, the effect is significant at $> 95\%$ c.l. when considering all bins together. In the text we discuss a number of possible explanations, the most likely being gravitational lensing.

In contrast, we do find a systematic relationship between quasar magnitude (and SNR) with $\Omega_g^{\rm DLA}$. Figure 17 presents the results for the same cuts of SDSS-DR3_4 as in Figure 16. It is evident that the bright sub-sample shows systematically higher $\Omega_g^{\rm DLA}$ values than the faint sub-sample. Of course, in order to explain the independence of $\ell_{\rm DLA}$ to quasar magnitude, all of the difference in $\Omega_g^{\rm DLA}$ must arise from the frequency of large $N_{\rm HI}$ damped systems. Indeed, Figure 18 shows that the $f_{\rm HI}(N, X)$ distributions are comparable for the bright and faint sub-samples for $N_{\rm HI} < 10^{21}\,{\rm cm}^{-2}$ but there is a factor of 5 difference in the incidence of systems with $N_{\rm HI} > 10^{21.4}\,{\rm cm}^{-2}$. This has the greatest impact on the $\Omega_g^{\rm DLA}$ results which we now discuss.

Let us consider possible explanations for the results in Figures 17 and 18:

1. Chance (small number statistics): In any given redshift interval in Figure 17 the results are significant at only the $\approx 1\sigma$ level. Put together, however, the effect is significant at greater than the 95% c.l.; the weighted mean of the ratio of the bright to faint sub-samples is $\Omega_g^{\rm DLA}({\rm bright})/\Omega_g^{\rm DLA}({\rm faint}) = 1.4 \pm 0.2$. Therefore, we con-

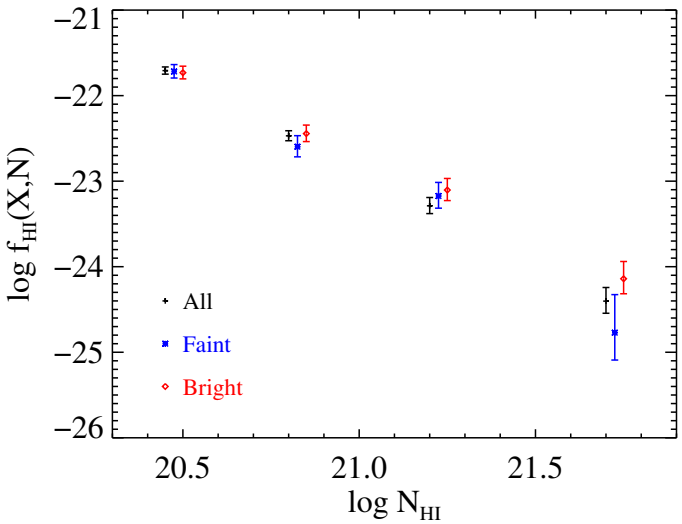

FIG. 18.— The H I column density distribution at $z = 2.5$ to 3 for the full sample, the brightest 33% of the quasars and the faintest 33%. One observes no statistically significant difference in $f_{\rm HI}(N,X)$ at low $N_{\rm HI}$ value but that the bright sample has 5× higher value at large $N_{\rm HI}$.

sider it unlikely that the effect is related only to statistical fluctuations.

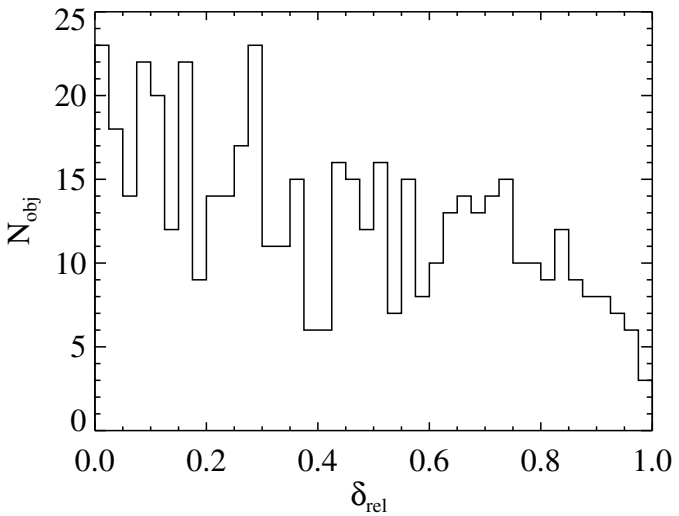

FIG. 19.— Histogram of the relative separation of the damped Lyα absorption redshift from the redshifts defining the survey path of the background quasar (z_f and z_i) as defined by Equation 24. One notes fewer damped Lyα systems with $z \approx z_i$ as expected because the incidence of damped Lyα systems increases with redshift.

2. DLA are intrinsic to the quasar: This is a difficult assertion to disprove. Figure 19 plots a histogram of the relative statistic of the damped Lyα systems from the quasar

$$\delta_{rel} \equiv \frac{z_f - z_{DLA}}{z_f - z_i} \qquad (24)$$

restricted to sightlines where $z_f - z_i > 0.2$. While there are fewer systems with $z_{DLA} \approx z_i$, this is expected because the incidence of the damped Lyα systems increases with redshift. At present, we do not consider association with the quasar to be an important issue.

3. The survey has missed damped Lyα systems with large $N_{\rm HI}$ toward faint quasars: Given that our algorithm is most sensitive to spectral regions of very low SNR, it is very unlikely that we would miss a damped Lyα system whose core is much larger than the search window $6(1+z)$Å. Furthermore, the profiles with very large $N_{\rm HI}$ value are the easiest to identify and we would have recovered most in our visual inspection of the spectra. Finally, these systems will probably all show significant metal-line absorption and we would have identified them independently of the Lyα profile (e.g. § 5.4).

There is, however, one bias related to this point. If a damped Lyα profile coincides with the wavelength λ_i in the quasar spectrum where the median SNR would have otherwise exceeded SNR$_{lim}$ (see § 2), then the presence of a damped Lyα system would preclude its inclusion in the statistical sample (and possibly its discovery altogether). That is, the presence of a damped Lyα profile at wavelength λ_{DLA}, where the median SNR of the unabsorbed flux exceeds 4, lowers the SNR at λ_{DLA} such that λ_i occurs redward of the damped Lyα system. This is true for damped Lyα systems of all $N_{\rm HI}$ values, although the effect will be greatest for larger $N_{\rm HI}$. We have accounted for this bias, in part, by defining z_1 to be 1000 km s^{-1} redward of the redshift corresponding to λ_i. Nevertheless, this issue is a concern, albeit one we consider minor.

4. Overestimating $N_{\rm HI}$ in the bright sample: We have reexamined the damped Lyα profiles for every system with $N_{\rm HI} > 10^{21.4}$ cm^{-2}. In no case would we modify the best $N_{\rm HI}$ value beyond the reported uncertainty. Furthermore, the bright sample has the highest SNR data and we consider the $N_{\rm HI}$ values to be well constrained. All have associated metal-line absorption and reasonably well determined quasar continua. Therefore, we consider this explanation to be unimportant.

5. Underestimating $N_{\rm HI}$ in the faint sample: Because these quasars are faint, the spectra have lower SNR and more uncertain $N_{\rm HI}$ values. While this could lead to a significant underestimate of $N_{\rm HI}$, several effects work against it. First, $f_{\rm HI}(N,X)$ is sufficiently steep (particularly at high $N_{\rm HI}$) that a Malmquist bias will actually lead to a flattening of $f_{\rm HI}(N,X)$ at the high $N_{\rm HI}$ end. This leads to an overestimate of $\Omega_g^{\rm DLA}$ in the faint subset because the bias will be more significant for $N_{\rm HI}$ values with larger uncertainty. Second, the effects of low SNR are most significant at low $N_{\rm HI}$ values where only a handful of pixels may constrain the result. But these systems do not contribute significantly to $\Omega_g^{\rm DLA}$. Third, at large $N_{\rm HI}$ value, the profile extends over ≈ 100Å and is generally constrained by tens of pixels. It is true that uncertainty in the continuum level is greater, but we do not believe this to be the problem.

To further pursue the possibility that we have underestimated $N_{\rm HI}$ in the faint sub-sample, we reexamined every Lyα profile with $N_{\rm HI} > 10^{21}$ cm^{-2} in the $z \sim 2.7$ redshift bin. While we could allow for 0.1 to 0.2 dex larger values in a number of cases, this would not account for the differences in $\Omega_g^{\rm DLA}$. Furthermore, we note that the $f_{\rm HI}(N,X)$ value in the $N_{\rm HI}$ interval $10^{21} - 10^{21.4}$ cm^{-2} is lower in the

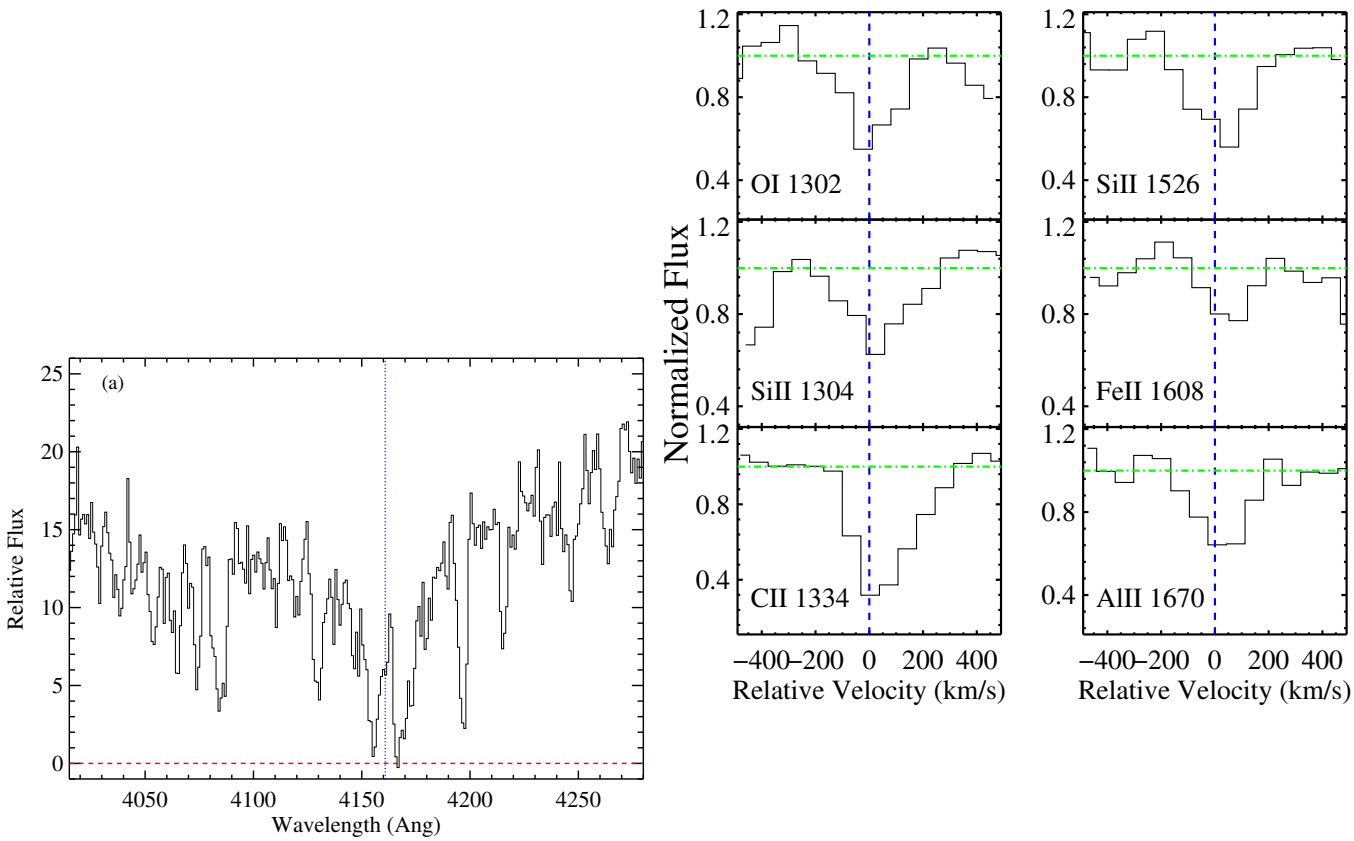

FIG. 20.— The (a) Lyα profile and (b) metal-line profiles of the super-LLS at $z = 2.42$ toward J130634.6+523250. The emission-line feature at the center of the Lyα profile is due to [OII] emission from a galaxy at $z = 0.112$ in the SDSS fiber. If this absorber had no significant metal-line absorption, it would have been dropped from our sample altogether.

faint sub-sample. Therefore, it would require a systematic underestimate in $N_{\rm HI}$ at nearly all values to explain the observations. At present, we consider an underestimate of $N_{\rm HI}$ values in lower SNR data to be a possible contribution to the differences in $\Omega_g^{\rm DLA}$ but unlikely to be the main explanation.

6. Dust Obscuration: To date, every damped Lyα system observed at high resolution has a measured metallicity exceeding 1/1000 solar (Prochaska et al. 2003b). With the presence of metals, there is the prospect that the gas has a non-negligible dust content (e.g. Pettini et al. 1994). Indeed, damped Lyα systems with metallicity $> 1/10$ solar have enhanced Si/Fe ratios which are best explained by differential depletion (Prochaska 2003). As Ostriker & Heisler (1984) first noted, the damped Lyα systems with the highest optical depths of dust will obscure the background quasar and possibly remove it from a magnitude limited quasar survey.

The prediction, as formalized by Fall & Pei (1993), is that optical quasar samples will underestimate $\Omega_g^{\rm DLA}$ owing to this bias. As Ellison et al. (2004) have stressed, the effect will be less significant if the quasar survey extends beyond the peak in the luminosity function. In any case, this bias implies that the $\Omega_g^{\rm DLA}$ values from brighter quasars should be lower than the value inferred from faint quasars. We observe the opposite trend in Figure 17. Therefore, the results are unlikely to be explained by dust obscuration.

7. Gravitational Lensing: If the damped Lyα systems arise in massive halos or disks, then it is possible that they would gravitationally magnify the background quasar (Bartelmann & Loeb 1996; Smette, Claeskens & Surdej 1997; Maller, Flores, & Primack 1997). In their analysis of the SDSS-DR2 quasar database, Murphy & Liske (2004) reported an approximately 2σ result that the luminosity function of quasars with foreground damped Lyα systems is brighter than those without. Unlike dust obscuration, this systematic effect could explain the results in Figure 17. In fact, treatments based on describing damped Lyα systems as exponential disks predict that the effect will be greatest for damped Lyα systems with large $N_{\rm HI}$ value (Bartelmann & Loeb 1996; Maller, Flores, & Primack 1997).

At present, we believe gravitational lensing to be the most viable explanation for the results in Figures 17 and 18. The key implications for $\Omega_g^{\rm DLA}$ are that (i) our reported values may be too high and (ii) the evolution of $\Omega_g^{\rm DLA}$ revealed by the entire sample is qualitatively correct. Therefore, while this systematic effect is important, our conclusions are relatively invariant to it. We note, however, that the effect should be even more pronounced at $z < 2$ and may significantly affect damped Lyα statistics at these redshifts (Rao & Turnshek 2000).

If gravitational lensing does explain the effect, it may allow for a statistical mass measurement of at least the high $N_{\rm HI}$ sightlines. It also motivates a high spatial resolution

survey of all damped Lyα systems with $N_{\rm HI} > 10^{21}\,{\rm cm}^{-2}$. In a future paper (Murphy et al. 2005, in prep.) we will present a full analysis of gravitational lensing and dust obscuration for the SDSS-DR3 sample.

5.3. *Human Error*

There is a non-negligible likelihood that we have overlooked a few systems or have made significant errors in a few select systems. Regarding PH04, for example, we note the incorrect $N_{\rm HI}$ value for the damped system at $z = 2.77$ toward J084407.29+515311 (§ 3.3) and also a bug in our calculation of $\ell(z)$ for the SDSS-DR1 sample. To quickly disseminate corrected and updated results to the community, we have established a public web site where all of the fits and analysis will be presented (http://www.ucolick.org/∼xavier/SDSSDLA/index.html). We encourage the community to report any mistakes with our analysis to the lead author via the email address sdss-dla@ucolick.org.

5.4. *Unusual Systematic Effects: "Things that go bump in the night"*

With a sample of damped Lyα systems approaching 1000, it is not surprising that unexpected systematic errors will arise. Figure 20 presents the Lyα profile and metal-line profiles for the damped Lyα candidate at $z = 2.42$ toward J130634.6+523250. This damped Lyα candidate was not identified by our automated algorithm because of significant flux at the center of the Lyα profile. Instead, the system was identified because of its metal-line absorption and we immediately hypothesized that the flux in the Lyα profile was due to Lyα emission from the host galaxy. The emission line would be amazingly strong, however, and we considered alternate explanations. In due time, we realized that the feature is an emission line: [OII] emission from a $z = 0.116$ galaxy which lies within the 3″ SDSS fiber. Emission lines of Hα, Hβ, and [OIII] are also apparent in the quasar spectrum. Ignoring the [OII] emission, we have fit the Lyα profile and its central value places it beneath the statistical threshold for damped Lyα systems. Nevertheless, this is a systematic effect which leads to an underestimate of $f_{\rm HI}(N, X)$ at all $N_{\rm HI}$ value. It is difficult to quantify the overall effect here, but it is presumably less than 1%.

6. DISCUSSION AND SPECULATIONS

The emphasis of this paper is to describe the results of the damped Lyα survey of the SDSS-DR3 quasar database. These results and a discussion of the systematic errors were presented in the previous sections. We now consider a few of the implications with emphasis on the new results. We also compare the observations against theoretical treatments of the damped Lyα systems within ΛCDM models of galaxy formation. We consider the results from the smooth particle hydrodynamic (SPH) simulations of Nagamine, Springel, & Hernquist (2004), the Eulerian simulations of Cen et al. (2003), and the semi-analytic model (SAM) of the Santa Cruz group (Somerville, Primack, & Faber 2001; Maller et al. 2001, 2003). It is important to stress that each model includes its own set of star formation and feedback recipes which do bear on the results for the damped Lyα systems.

Consider first the $N_{\rm HI}$ frequency distribution, $f_{\rm HI}(N, X)$. Perhaps the most remarkable result from the SDSS-DR3 sample is that there is no statistical evidence for any evolution in the shape of $f_{\rm HI}(N, X)$ with redshift (Figure 7). There is, however, evidence for evolution in the normalization of $f_{\rm HI}(N, X)$ as traced by the trends in the zeroth and first moments of the distribution function. These results suggest that the gas distribution within galaxies is similar at all redshifts and that only the number and/or sizes of these galaxies evolve significantly. Another interesting result is that the faint-end slope of the $f_{\rm HI}(N, X)$ distribution is $\alpha_3 \approx -1.8$. This slope matches the faint-end slope of the dark matter halo mass function for CDM (e.g. Sheth, Mo, & Tormen 2001). If this is not a coincidence, it indicates that low mass halos dominate the incidence of damped Lyα systems at low $N_{\rm HI}$ values. Furthermore, it suggests that the cross-section $A(X)$ of low mass galaxies is nearly independent of mass. At present, however, we consider the correspondence to be a coincidence.

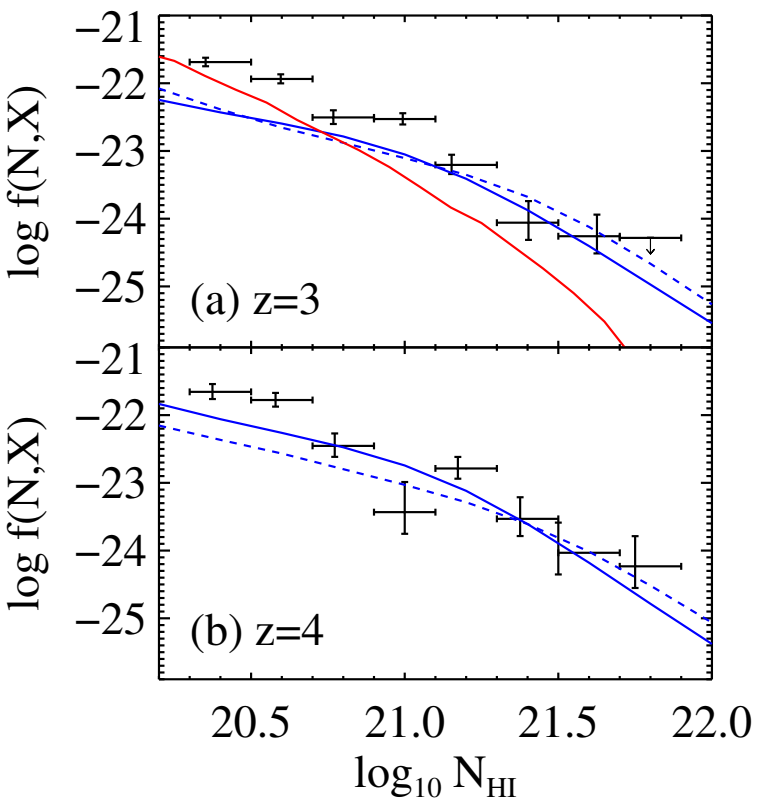

FIG. 21.— The H I frequency distribution of the damped Lyα systems at $z = 3$ (top; specifically $z = 2.8$ to 3.2) and $z = 4$ (bottom; specifically $z = 3.7$ to 4.3) compared against the theoretical curves of Maller et al. (2001) (SAMS; lighter curve) and Nagamine, Springel, & Hernquist (2004) (SPH; darker curves; dashed is the D5 model and solid is the Q5 model).

A comparison of the results against ΛCDM models of galaxy formation is presented in Figure 21 at $z = 3$ and $z = 4$. The $f_{\rm HI}(N, X)$ curves are for the SPH simulations of Nagamine, Springel, & Hernquist (2004) and the SAM model of Maller et al. (2001). The SAM model shows a reasonable match to the shape of $f_{\rm HI}(N, X)$ at $z = 3$, yet systematically underpredicts the observations.

Maller et al. (2001) were primarily interested in modeling the kinematics of the damped Lyα systems and we suspect their results are insensitive to the normalization of $f_{\rm HI}(N,X)$. Nevertheless, the discrepancy suggests that the SAM model has too few damped Lyα systems at $z=3$. Similarly, the SPH theoretical curves offer a reasonable match to the observations at large $N_{\rm HI}$ value yet significantly underpredict $f_{\rm HI}(N,X)$ at low $N_{\rm HI}$ values.

The discrepancy has several important implications. First, the SPH simulations will predict a significantly smaller contribution to $\Omega_g^{\rm HI}$ for the Lyman limit systems at these redshifts. In part, this is a restatement of a current problem in numerical cosmology: the simulations underpredict the incidence of Lyman limit systems by an order of magnitude (Gardner et al. 2001). Second, we reemphasize the critical constraint imposed by two sets of observations of the damped Lyα systems: (a) $f_{\rm HI}(N,X)$ and (b) the velocity width distribution of the low-ion profiles (Prochaska & Wolfe 1997). Jedamzik & Prochaska (1998) first emphasized that there is a tension between $f_{\rm HI}(N,X)$ and the velocity width distribution. Specifically, if one introduces enough low mass halos to match the low $N_{\rm HI}$ end of $f_{\rm HI}(N,X)$ this implies far too many damped Lyα systems with small velocity width. Therefore, if the SPH simulations were modified to match the $f_{\rm HI}(N,X)$ observations (e.g. via different treatments of feedback or radiative transfer), we predict that (i) the dependence of the cross-section $A(X)$ on the halo mass as reported by Nagamine, Springel, & Hernquist (2004) would change significantly and (ii) the simulations would not reproduce the damped Lyα kinematics. We also note that the results of Nagamine, Springel, & Hernquist (2004) show that $f_{\rm HI}(N,X)$ is relatively insensitive to treatments of feedback. It is possible that $f_{\rm HI}(N,X)$ and the damped Lyα kinematics present a fundamental challenge to scenarios of galaxy formation within the ΛCDM cosmology.

Now consider the evolution in the line density of damped Lyα systems. Because $\ell_{\rm DLA}(X)$ is the product of the comoving number density and cross-section of damped Lyα systems (Equation 14), the rise in $\ell_{\rm DLA}(X)$ reflects an increase in one or both of these quantities. We can estimate the evolution in $n_{\rm DLA}$ by considering the Press-Schechter formalism for the mass function of dark matter halos (e.g. Peacock 1999). Within this formalism one can define a mass scale M_* as a function of redshift which identifies the typical mass of assembly. At $z=3$ in the ΛCDM cosmology with $\sigma_8 = 1$, $M_* \approx 10^{10}\,{\rm M}_\odot$. Halos with masses $M \gg M_*$ will have a number density which increases significantly from $z=3$ to 2. Therefore, damped Lyα systems are unlikely to arise in halos with $M \gg 10^{10}\,{\rm M}_\odot$ at these redshifts. Furthermore, although halos with $M < M_*$ do have a decreasing comoving number density, the decrease is small (i.e. $< 15\%$ for $M = 10^9\,{\rm M}_\odot$). Within the context of the hierarchical cosmology, therefore, the most likely explanation for the decrease in $\ell_{\rm DLA}(X)$ is a corresponding decline in $A(X)$. There are several physical effects which could reduce $A(X)$, e.g., star formation, AGN feedback, photoionization, and galactic winds. Current estimates of the star formation rates are relatively uncertain over this redshift range (e.g. Wolfe, Gawiser, & Prochaska 2003) but probably cannot account for the bulk of evolution. Similarly, the intensity of the extragalactic background radiation field is believed to be roughly constant (e.g. Haardt & Madau 1996) and photoionization effects should be minor. Therefore, we contend that one or more feedback mechanisms have significantly reduced the typical cross-section of galaxies to damped Lyα absorption from $z=3$ to 2.

Turning our attention to the neutral gas mass density, we present the damped Lyα observations in Figure 22 including recent results at $z < 1.6$ from Rao, Turnshek, & Nestor (in prep.) compared against several ΛCDM models, and also current estimates of the mass density of stars (star), neutral gas (diamond), and Irr galaxies (+ sign) at $z \sim 0$. It is important to note that the theoretical models of Somerville et al. and Nagamine et al. include contributions to Ω_g from all quasar absorption line systems (i.e. they calculate $\Omega_g^{\rm HI}$), whereas the observational measurements are restricted to the damped Lyα systems. Therefore, if the Lyman limit systems do contribute significantly to $\Omega_g^{\rm HI}$, we must increment the observations accordingly. Alternatively, we recommend that future theoretical analysis be restricted to sightlines with $N_{\rm HI} \geq 2\times10^{20}\,{\rm cm}^{-2}$ (e.g. Cen et al. 2003).

Examining the damped Lyα observations alone, we note a relatively confusing picture. While the results based primarily on the SDSS-DR3 observations ($z > 2.2$) show a well behaved trend with redshift, the estimates of $\Omega_g^{\rm DLA}$ at $z < 2$ are all consistent with one another with a central value higher than the $z=2.3$ measurement. While each individual measurement at $z < 2$ is consistent with the SDSS data point at $z=2.3$, taken together the difference is significant at $> 95\%$ c.l. In fact, if one were to ignore the redshift interval at $z=2.3$, the observations are consistent with no evolution in $\Omega_g^{\rm DLA}$ from $z=0.1$ to $z=4.5$. Before reaching such a conclusion, however, we wish to emphasize several points: (1) the value in the $z=[2.2,2.5]$ interval is very well determined because it is based on ~ 100 damped Lyα systems; (2) the $z\sim2$ data point is derived from a heterogeneous sample of observations and is dominated by a single damped Lyα system; (3) the low redshift values are based on the novel yet non-standard technique of Rao & Turnshek (2000). Their approach has its own set of systematic errors which are uniquely different from the damped Lyα survey described in this paper; and (4) we argued in § 5.2 that the $\Omega_g^{\rm DLA}$ results may be biased by gravitational lensing. If this is confirmed, the effect should be largest at $z < 2$. These points aside, it is clear that achieving better than 10% precision on $\Omega_g^{\rm DLA}$ at $z < 2$ is a critical goal of future damped Lyα surveys. At $z\sim2$, this will require a large observing campaign with a spectrometer efficient down to 3200Å. At lower redshift, one will require a new UV space observatory.

Comparing the models, we note a wide range of predictions. The most successful models at $z > 2$ are from Nagamine, Springel, & Hernquist (2004), in particular their D5 run. This model reproduces both the shape and normalization of the observed data. In contrast, the Eulerian and SAM models overpredict $\Omega_g^{\rm DLA}$ at all redshifts and at $z < 3$ respectively even if one adopts a 1.5 multiplicative correction due to the Lyman limit systems. Because the cooling processes and time-scales are comparable in all of the models (e.g. Pearce et al. 2001), the differences must be due to processes which consume or ionize the

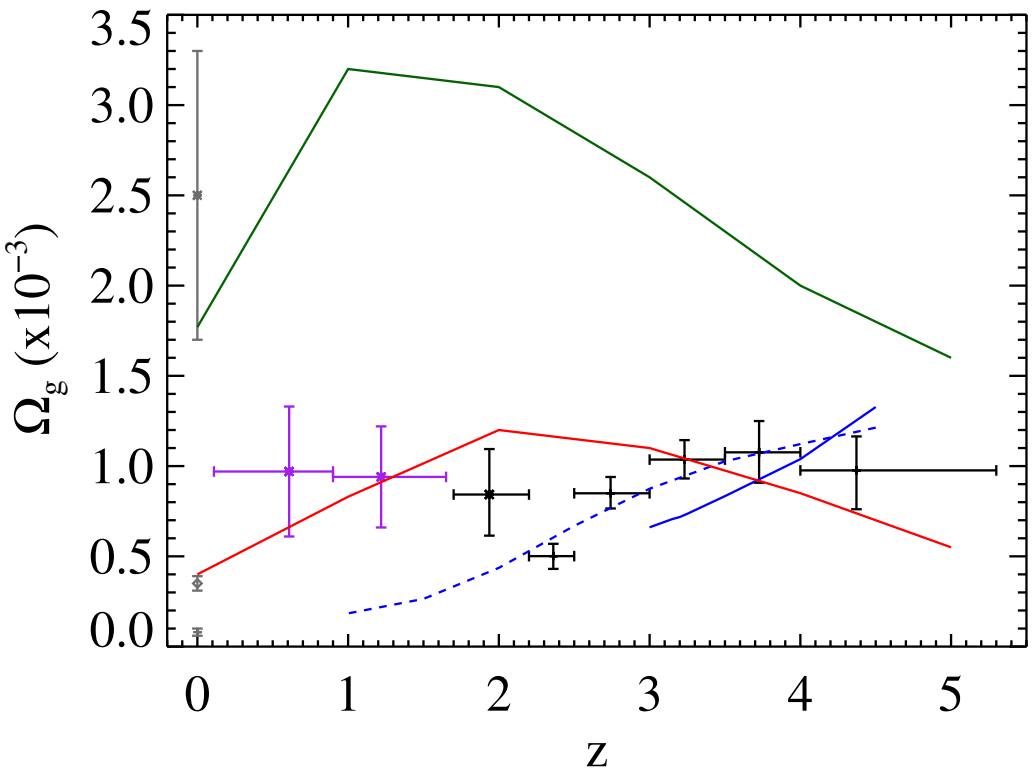

FIG. 22.— The gas mass density of neutral gas for the damped Lyα systems from our analysis ($z > 1.5$ with SDSS restricted to $z > 2.2$) and recent results (purple) from Rao, Turnshek, & Nestor (in prep.). These observations are compared against the theoretical curves of Cen et al. (2003) (EULER; green), Somerville, Primack, & Faber (2001) (SAMS; red), and Nagamine, Springel, & Hernquist (2004) (SPH; blue – dotted is the D5 model and solid is the Q5 model). The data points at $z = 0$ correspond to the stellar mass density from Cole et al. (2001) (star), the neutral gas mass density (diamond), and the mass density of Irr galaxies (+-sign; Fukugita et al. 1998).

neutral gas (e.g. star formation, galactic winds, AGN feedback). The indication from our observations is that the Eulerian and SAM models underpredict these processes at $z \gtrsim 2$ and therefore overestimate Ω_g.

Now consider a comparison of the high-z $\Omega_g^{\rm DLA}$ values with the mass density at $z = 0$ of stars Ω_*, neutral gas $\Omega_g^{\rm 21cm}$, and dwarf galaxies Ω_d. The stellar mass density was estimated by Cole et al. (2001) from an analysis of the 2dF survey. The uncertainty in this estimate is dominated by systematic error related to the assumed initial mass function. Wolfe et al. (1995) first stressed that the gas mass density of the damped Lyα systems is comparable to Ω_*. Adopting the ΛCDM cosmology and the current estimate of Ω_*, we now find that Ω_* exceeds $\Omega_g^{\rm DLA}$ by a factor of two to three at $z = 3$. Because star formation is ongoing at all redshifts probed by the damped Lyα systems (e.g. Chen & Lanzetta 2003; Wolfe, Prochaska, & Gawiser 2003; Møller et al. 2002), however, it would be wrong to interpret the maximum $\Omega_g^{\rm DLA}$ value as the total gas mass density contributed by the damped Lyα systems. It is more accurate to regard damped Lyα systems as neutral gas reservoirs in which gas consumed by star formation is replaced by neutral gas accreted from the IGM. In this manner, the mass density of damped Lyα systems would be less than Ω_* at any given epoch. Therefore, it is reasonable to assume that all of the stars observed today arose from gas originating in the damped Lyα systems. Indeed, this is the generic conclusion of current cosmological simulations. These points aside, it is evident that the damped Lyα systems contain sufficient gas mass to account for all of the stars observed in disks today provided current estimates ($\Omega_{\rm disk} \approx \Omega_*/3$; Fukugita et al. 1998).

Examining Figure 22, we note that the difference between $\Omega_g^{\rm 21cm}$ at $z = 0$ and $\Omega_g^{\rm DLA}$ at $z = 2.3$ is 0.15 ± 0.08, i.e. consistent with very little evolution if one ignores the results at $z \sim 1$ we present from Rao et al. (in prep). Is this a remarkable coincidence or is there a physical explanation (e.g. the gas accretion rate equaled the star formation rate over the past 10 Gyr)? Before addressing this question, consider the determination of $\Omega_g^{\rm 21cm}$. The value is derived from large area surveys of 21cm emission-line observations for H I 'clouds'. The most recent results are from the HIPASS survey as analyzed by Zwaan et al. (2005). The analysis proceeds by fitting a functional form (a Γ-function) to the H I mass distribution of all galaxies detected. The $\Omega_g^{\rm 21cm}$ value is simply proportional to the first moment of this distribution function. The key point to emphasize is that the analysis includes all H I gas within the beam of HIPASS, i.e., the values are independent of H I surface density and are not restricted to the damped Lyα threshold. In the extreme case that the H I gas is predominantly distributed in Mestel disks with $\Sigma(R) \propto R^{-1}$, the contribution to $\Omega_g^{\rm 21cm}$ from damped Lyα systems is $\Omega_g^{\rm DLA}/\Omega_g^{\rm 21cm} = \Sigma_{\rm trunc}/\Sigma_{\rm DLA} = N_{\rm trunc}/10^{20.3}\,{\rm cm}^{-2}$ where $\Sigma_{\rm trunc}$ and $N_{\rm trunc}$ are the surface density and H I column

density at the truncation radius of the Mestel disk and $\Sigma_{\rm DLA}$ is the surface density at the damped Lyα threshold. One may adopt a truncation radius for the Mestel disk according to the photoionization edge set by the intensity of the extragalactic UV background (EUVB) or observational constraints (e.g. Maloney 1993). Adopting a relatively conservative value of $N_{\rm trunc} = 10^{19.5}\,{\rm cm}^{-2}$, we find $\Omega_g^{\rm DLA} = 0.15\Omega_g^{\rm 21cm}$. Of course, this is a somewhat extreme model because many gas disks have exponential profiles. For an exponential disk, the value of $\Omega_g^{\rm DLA}/\Omega_g^{\rm 21cm}$ is sensitive to the ratio of the radii corresponding to the damped Lyα threshold and the exponential scale-length. Taking $R_{DLA}/R_{exp} = 2$, we find $\Omega_g^{\rm DLA} = 0.5\Omega_g$. These simple estimates imply that a significant fraction of $\Omega_g^{\rm 21cm}$ at $z = 0$ could come from Lyman limit systems.

One can also place an empirical constraint on $\Omega_g^{\rm DLA}/\Omega_g$ from the H I distribution function at $z = 0$ as measured by Ryan-Weber, Webster, & Staveley-Smith (2003). If one (a) adopts their double power-law fit to $f_{\rm HI}(N, X)$ with exponents $\alpha_3 = -1.4$ and $\alpha_4 = -2.1$ and $N_d = 10^{20.9}\,{\rm cm}^{-2}$, (b) extrapolates this frequency distribution to $N_{min} = 10^{19}\,{\rm cm}^{-2}$, and (c) integrates to $N_{max} = 10^{21.6}\,{\rm cm}^{-2}$ (the largest $N_{\rm HI}$ value measured), we find $\Omega_g^{\rm DLA} = 0.6\Omega_g$. There is significant uncertainty and degeneracy in the fitted parameters of $f_{\rm HI}(N, X)$, however, and $\Omega_g^{\rm DLA}/\Omega_g$ could be significantly higher or lower. Nevertheless, it is our present view[8] that $\Omega_g^{\rm DLA}$ at $z = 0$ is $\approx 0.5\Omega_g$. Therefore, we argue that a strict comparison of $\Omega_g^{\rm DLA}$ at $z = 0$ and $z = 2$ could imply a decrease in $\Omega_g^{\rm DLA}$ over the past 10 Gyr.

Given the uncertainties inherent to $\Omega_g^{\rm HI}$ and the fact that it has an ambiguous physical meaning, we now advocate an alternate definition for Ω_g motivated by star formation processes. Define $\Omega_g^{\rm Neut}$ to be the *mass density of gas that is predominantly neutral.* Under this definition, $\Omega_g^{\rm Neut}$ evolves as follows. The decrease of the EUVB intensity with decreasing redshift implies that lower surface densities will be neutral at lower redshift. Therefore, it is very possible that $\Omega_g^{\rm Neut}$ equals $\Omega_g^{\rm DLA}$ at $z = 2$ and the HIPASS value at $z = 0$. In this case, the gas reservoir available to star formation may have remained nearly constant over the past 10 Gyr (subject, of course, to the value of $\Omega_g^{\rm Neut}$ from $z = 0$ to 2). We emphasize that this is consistent with the presence of star formation provided gas replenishment from the IGM is available. Future progress on these issues will require an accurate determination of $f_{\rm HI}^{\rm LLS}$ at all redshifts and also an accurate assessment of the ionization state of the gas as a function of H I column density.

Finally, compare $\Omega_g^{\rm DLA}$ at $z = 3$ against the mass density in dwarf galaxies Ω_d as derived by Fukugita et al. (1998). The value presented here is their estimate for the stellar mass density in Irr galaxies boosted by a factor of 2 to account for the gas mass and by 30% for $H_0 = 75{\rm km\ s}^{-1}\ {\rm Mpc}^{-1}$. We emphasize that the estimate of Ω_d is difficult to make and may be more uncertain than depicted here. Nevertheless, it is evident that the mass density of dwarf galaxies is roughly $10\times$ smaller than $\Omega_g^{\rm DLA}$ at $z = 3$. A number of authors have argued that the majority of damped Lyα systems will evolve into dwarf galaxies today. It is obvious, however, that the *majority of damped Lyα systems (by mass) cannot evolve into dwarf galaxies.* And because damped Lyα systems with all values of $N_{\rm HI}$ contribute to $\Omega_g^{\rm DLA}$ (Figure 10), we argue that the damped Lyα systems by number are not the progenitors of dwarf galaxies. As we have argued previously, it is premature to interpret the observed abundance patterns as significant evidence for a link between damped Lyα systems and dwarf galaxies (Prochaska 2003). Furthermore, the kinematic characteristics of the damped Lyα systems preclude such an interpretation (Prochaska & Wolfe 1997; Prochaska, Ryan-Weber, & Staveley-Smith 2002). Revealing the true nature of the damped Lyα systems, of course, will require detailed follow-up studies (high resolution spectroscopy, deep imaging, etc.) of the galaxies identified in this survey.

[8]Since submission of our paper, we have been informed that a better estimate is 15% (Zwaan, priv. comm.).

We acknowledge the tremendous effort put forth by the SDSS team to produce and release the SDSS survey. JXP would like to acknowledge the visitor's program at Cambridge where the work on this paper was initiated. We would like to thank P. McDonald for providing the mock SDSS spectra for investigating systematic uncertainty in the analysis. We thank B. Metcalf and A. Maller for helpful comments and suggestions regarding gravitational lensing and the damped Lyα systems. We thank J. Primack and P. Madau for helpful discussions. Additional thanks to M. Murphy for helping to identify a bug in our search algorithm. Finally, we wish to thank S. Rao, D. Turnshek, and D. Nestor for sharing their results in advance of publication. JXP and SHF (through an REU fellowship) acknowledge support through NSF grant AST 03-07408. JXP and AMW are also partially supported by NSF grant AST 03-07824.

REFERENCES

Abazajian, K. et al. 2003, AJ, 126, 2081
Bartelmann, M. & Loeb, A. 1996, ApJ, 457, 529
Bennett, C.L., et al. 2003, ApJS, 148, 1
Blanton, M.R., et al. 2003, ApJ, 592, 819
Bolton, A. S., Burles, S., Schlegel, D. J., Eisenstein, D. J., & Brinkmann, J. 2004, AJ, 127, 1860
Boyle, B.J., Shanks, T., Croom, S.M., Smith, R.J., Miller, L., Loaring, N., & Heymans, C. 2000, MNRAS, 317, 1014
Burles, S.M., & Schlegel, D. 2004, in preparation
Cen, R., Ostriker, J.P., Prochaska, J.X., & Wolfe, A.M. 2003, ApJ, 598, 741
Chen, H.-W. & Lanzetta, K. M. 2003, ApJ, 597, 706
Chen, H.-W., Kenicutt, R.C., Jr., & Rauch, M. 2005, ApJ, 620, 703
Cole, S. et al. 2001, MNRAS, 326, 255
Dessauges-Zavadsky, M., Péroux, C., Kim, T.-S., D'Odorico, S., McMahon, R.G. 2003, MNRAS, 345, 447
Ellison, S.L., Yan, L., Hook, I.M., Pettini, M., Wall, J.V., & Shaver, P. 2001, A&A, 379, 393
Ellison, S.L., Churchill, C.W., Rix, S.A., & Pettini, M. 2004, ApJ, 615, 118
Fall, S.M. & Pei, Y.C. 1993, ApJ, 402, 479
Fukugita, M., Hogan, C.J, Peebles, P.J.E. 1998, Nature, 503, 518
Gardner, J.P., Katz, N., Hernquist, L., & Weinberg, D.H. 2001, ApJ, 559, 131
Haardt, F. & Madau, P. 1996, ApJ, 461, 20
Hu, E., Kim, T.-S., Cowie, L.L., Songaila, A., & Rauch, M. 1995, AJ, 110, 1526
Jedamzik, K. & Prochaska, J. X. 1998, MNRAS, 296, 430
Kauffmann, G. 1996, MNRAS, 281, 475
Kirkman, D. & Tytler, D. 1997, ApJ, 484, 672
Klypin, A., Borgani, S., Holtzman, J., & Primack, J. 1995, ApJ444, 1

Lanzetta, K. M., Wolfe, A. M., Turnshek, D. A., Lu, L., McMahon, R. G., & Hazard, C. 1991, ApJS, 77, 1
Lanzetta, K. M., Wolfe, A. M.,& Turnshek 1995, ApJ, 440, 435
Ledoux, C., Petitjean, P., & Srianand, R. 2003, MNRAS, 396, 429
Lopez, S. & Ellison, S.L. 2003, A&A, 403, 573
Lu, L., Sargent, W.L.W., Barlow, T.A., Churchill, C.W., & Vogt, S. 1996, ApJS, 107, 475
Lupton, R. 1993, Statistics in Theory and Practise (Princeton: Princeton Univ. Press)
Ma, C.-P., Bertschinger, E., Hernquist, L., Weinberg, D.H., & Katz, N. 1997, 484, 1L
Ma, C.-P., & Bertschinger, E. 1994, ApJ, 434, L5
Maller, A.H., Flores, R.A., & Primack, J.R. 1997, ApJ, 486, 681
Maller, A.H., Prochaska, J.X., Somerville, R.S., & Primack, J.R. 2001, MNRAS, 326, 1475
Maller, A.H., Prochaska, J.X., Somerville, R.S., & Primack, J.R. 2003, MNRAS, 343, 268
Maloney, P. 1993, ApJ, 414, 41
Mathlin, G.P., Baker, A.C., Churches, D.K., & Edmunds, M.G. 2001, MNRAS, 321, 743
McDonald, P., Seljak, U., Burles, S., Schlegel, D.J., Weinberg, D.H., Shih, D., Schaye, J., Schneider, D.P., Brinkmann, J., Brunner, R.J., & Fukugita, M. 2005a, ApJ, submitted (astro-ph/0405013)
McDonald, P., Seljak, U., Cen, R., Bode, P., & Ostriker, J.P. 2005b, MNRAS, in press (astro-ph/0407378)
Møller, P., Warren, S.J., Fall, S.M., Fynbo, J.U., & Jakobsen, P. 2002, ApJ, 574, 51

Murphy, M.T., & Liske, J. 2004, MNRAS, 354, L31
Nagamine, K., Springel, V., & Hernquist, L. 2004, MNRAS, 348, 421
Ostriker, J.P. & Heisler, J. 1984, ApJ, 278, 1
Peacock, J.A. 1999, *Cosmological Physics*, (Cambridge: Cambridge University Press)
Pearce, F. R., Jenkins, A., Frenk, C. S., White, S. D. M., Thomas, P. A., Couchman, H. M. P., Peacock, J. A., & Efstathiou, G. 2001, MNRAS, 326, 649
Pei, Y.C. & Fall, S.M. 1995, ApJ, 454, 69
Péroux, C., Storrie-Lombardi, L.J., McMahon, R.G., Irwin, M., & Hook, I.M. 2001, AJ, 121, 1799
Péroux, C., McMahon, R., Storrie-Lombardi, L., & Irwin, M.J. 2003, MNRAS, 346, 1103 (PMSI03)
Pettini, M., Smith, L. J., Hunstead, R. W., and King, D. L. 1994, ApJ, 426, 79
Prochaska, J.X. 1999, ApJ, 511, L71
Prochaska, J.X. 2003, IAU,
Prochaska, J.X. & Burles, S.M. 1999, AJ, 117, 1957
Prochaska, J.X. & Herbert-Fort, S. 2004, PASP, 116, 622 (PH04)
Prochaska, J. X., Ryan-Weber, E., & Staveley-Smith L. 2002, PASP, 114, 1197
Prochaska, J. X. & Wolfe, A. M. 1996, ApJ, 470, 403
Prochaska, J. X. & Wolfe, A. M. 1997, ApJ, 486, 73
Prochaska, J.X., Howk, J.C., & Wolfe, A.M. 2003, Nature, 423, 57
Prochaska, J.X., Gawiser, E., Wolfe, A.M., Cooke, J., & Gelino, D. 2003, ApJS, 147, 227
Prochaska, J.X., Gawiser, E., Wolfe, A.M., Castro, S., & Djorgovski, S.G. 2003, ApJ, 595, L9
Prochter, G.E., Prochaska, J.X., & Burles, S.M. 2005, ApJ, in press
Rao, S.M. & Turnshek, D.A. 2000, ApJS, 130, 1
Richards, G., et al. 2002, AJ, 123, 2945
Rosenberg, J.L. & Schneider, S.E. 2003, ApJ, 585, 256
Ryan-Weber, E., Webster, R., Staveley-Smith, L. 2003, MNRAS, 343, 1195
Ryan-Weber, E., Webster, R., Staveley-Smith, L. 2005, MNRAS, 356, 1600
Schaye, J. 2001, ApJ, 526, 95L
Sheinis, A.I., Miller, J., Bigelow, B., Bolte, M., Epps, H., Kibrick, R., Radovan, M., & Sutin, B. 2002, PASP, 114, 851
Sheth, R.K., Mo, H.J., & Tormen, G. 2001, MNRAS, 323, 1
Smette, A., Claeskens, J.-F., & Surdej, J. 1997, New Astr., 2, 53
Somerville, R. S., Primack, J. R., & Faber, S.M. 2001, MNRAS, 320, 504
Songaila, A. & Cowie, L.L. 2002, AJ, 123, 2183
Storrie-Lombardi, L.J., Irwin, M.J. 1996, & McMahon, R.G. MNRAS, 282, 1330
Storrie-Lombardi, L.J. & Wolfe, A.M. 2000, ApJ, 543, 552
Spergel, D. et al. 2003, ApJS, 148, 175
Tytler, D. 1987, ApJ, 321, 69
Viegas, S.M. 1995, MNRAS, 276, 268
Vladilo, G. 1998, ApJ, 493, 583
Vladilo, G., Centurión, M., Bonifacio, P., & Howk, J.C. 2001, ApJ, 557, 1007
Webb, J.K., Murphy, M.T., Flambaum, V.V., Dzuba, V.A., Barrow, J.D., Churchill, C.W., Prochaska, J.X., & Wolfe, A.M. 2001 Phys. Rev. Lett. 87, 1301
Wolfe, A.M. 1985, Phil. Trans. Roy. Soc. Lon., 321, 503
Wolfe, A.M., Turnshek, D.A., Smith, H.E., & Cohen, R.D. 1986, ApJS, 61, 249
Wolfe, A. M., Lanzetta, K. M., Foltz, C. B., and Chaffee, F. H. 1995, ApJ, 454, 698
Wolfe, A. M., Prochaska, J.X., & Gawiser, E. 2003, ApJ, 593, 215
Wolfe, A. M., Gawiser, E., & Prochaska, J.X. 2003, ApJ, 593, 235
Wolfe, A. M., Howk, J. C., Gawiser, E., Prochaska, J. X., & Lopez, S. 2004, ApJ, 615, 625
Wolfe, A. M., Gawiser, E., & Prochaska, J.X. 2005, ARA&A, in press
Zheng, Z. & Miralda-Escudé, J. 2002, ApJ, 568, L71
Zwaan, M.A., Briggs, F.H., & Sprayberry, D. 2001, MNRAS, 327, 1249
Zwaan, M.A., Meyer, M.J., Staveley-Smith, L., & Webster, R.L. 2005, MNRAS, in press