PUBLICATIONS OF THE ASTRONOMICAL SOCIETY OF THE PACIFIC, **116**:622–633, 2004 July

The Sloan Digital Sky Survey Damped Lyα Survey: Data Release 1

JASON X. PROCHASKA AND STEPHANE HERBERT-FORT
UCO/Lick Observatory, University of California, 1156 High Street, Santa Cruz, CA 95064; xavier@ucolick.org, shf@ucolick.org
Received 2004 March 17; accepted 2004 April 20; published 2004 June 8

ABSTRACT. We present the results from an automated search for damped Lyα (DLA) systems in the quasar spectra of Data Release 1 from the Sloan Digital Sky Survey (SDSS-DR1). At $z \approx 2.5$, this homogeneous data set has greater statistical significance than the previous two decades of research. We derive a statistical sample of 71 DLA systems (>50 previously unpublished) at $z > 2.1$ and measure H I column densities directly from the SDSS spectra. The number of DLA systems per unit redshift is consistent with previous measurements, and we expect our survey has more than 95% completeness. We examine the cosmological baryonic mass density of neutral gas, Ω_g, inferred from the DLA systems from the SDSS-DR1 survey and a combined sample drawn from the literature. Contrary to previous results, the Ω_g values do not require a significant correction from Lyman limit systems at any redshift. We also find that the Ω_g values for the SDSS-DR1 sample do not decline at high redshift, and the combined sample shows a (statistically insignificant) decrease only at $z > 4$. Future data releases from SDSS will provide the definitive survey of DLA systems at $z \approx 2.5$ and will significantly reduce the uncertainty in Ω_g at higher redshift.

On-line material: extended table

1. INTRODUCTION

It has now been two decades since the inception of surveys for high-redshift galaxies through the signature of damped Lyα (DLA) absorption in the spectra of background quasars (Wolfe et al. 1986). Owing to large neutral hydrogen column densities N(H I), these absorption lines exhibit large rest equivalent widths ($W_\lambda > 10$ Å) and show the Lorentzian wings characteristic of quantum mechanic line damping. Through dedicated surveys of high- and low-redshift quasars with optical and ultraviolet telescopes, over 300 DLA systems have been identified (Curran et al. 2002). These galaxies span redshifts $z = 0$ (the Milky Way, LMC, SMC) to $z = 5.5$, where the opacity of the Lyα forest precludes detection (Songaila & Cowie 2002).

Statistics of the DLA systems impact a wide range of topics in modern cosmology, galaxy formation, and physics. These include studies on the chemical enrichment of the universe in neutral gas (Pettini et al. 1994; Prochaska et al. 2003b), nucleosynthetic processes (Lu et al. 1996; Prochaska et al. 2003a), galactic velocity fields (Prochaska & Wolfe 1997), the molecular and dust content of young galaxies (Vladilo 1998; Ledoux et al. 2003), star formation rates (Wolfe et al. 2003), and even constraints on temporal evolution of the fine-structure constant (Webb et al. 2001). Perhaps the most fundamental measurement from DLA surveys, however, is the evolution of the cosmological baryonic mass density in neutral gas, Ω_g (Storrie-Lombardi & Wolfe 2000; Rao & Turnshek 2000; Péroux et al. 2003, hereafter PMSI03). Because the DLA systems dominate the mass density of neutral gas from $z = 0$ to at least $z = 3.5$, a census of these absorption systems directly determines Ω_g. These measurements express global evolution in the gas, which feeds star formation (Pei & Fall 1995; Mathlin et al. 2001), and are an important constraint for models of hierarchical galaxy formation (e.g., Somerville et al. 2001; Nagamine et al. 2004a, 2004b).

The most recent compilation of DLA systems surveyed in a "blind," statistical manner combines the effects of observing programs using over 10 telescopes, 10 unique instruments, and the data reduction and analysis of ∼10 different observers (PMSI03). In short, the results are derived from a heterogeneous sample of quasar spectra derived from heterogeneous quasar surveys. While considerable care has been taken to collate these studies into an unbiased analysis, it is difficult to assess the completeness and potential selection biases of the current sample. These issues are particularly important when one aims to address the impact of effects like dust obscuration (Ostriker & Heisler 1984; Fall & Pei 1993; Ellison et al. 2001).

In this paper we present the first results in a large survey for DLA systems drawn from a homogeneous data set of high-z quasars with well-defined selection criteria. Specifically, we survey the quasar spectra from Data Release 1 of the Sloan Digital Sky Survey (SDSS-DR1), restricting our search to SDSS-DR1 quasars with Petrosian magnitudes of $r' < 19.5$ mag.

The DR1 sample alone (the first of five data releases from SDSS) offers a survey comparable to—although not strictly independent from—the efforts of 20 years of work. We introduce algorithms to automatically identify DLA candidates in the fluxed (i.e., non-normalized) quasar spectra and perform

Voigt profile analyses to confirm and analyze the DLA sample. This survey was motivated by a search for "metal strong" DLA systems like the $z = 2.626$ DLA system toward FJ0812+32 (Prochaska et al. 2003a). A discussion of the "metal strong" survey will be presented in a future paper (S. Herbert-Fort et al. 2004, in preparation).

This paper is organized as follows. In § 2, we present the quasar sample and discuss the automatic DLA candidate detection. In § 3, we present the Voigt profile fits to the full sample. We present a statistical analysis in § 4, and a summary and concluding remarks are given in § 5.

2. QUASAR SAMPLE AND DLA CANDIDATES

The quasar sample was drawn from Data Release 1 of the Sloan Digital Sky Survey to a limiting Petrosian magnitude of $r' = 19.5$ mag. This criterion was chosen primarily to facilitate follow-up observations with 10 m class telescopes, and it includes more than 60% of all SDSS-DR1 quasars at $z > 2$. With rare exception, the fiber-fed SDSS spectrograph provides FWHM ≈ 150 km s^{-1} spectra of each quasar for the wavelength range $\lambda \approx 3800$–9200 Å. All of the spectra were reduced using the SDSS spectrophotometric pipeline and were retrieved from the SDSS data archive[1] (Abazajian et al. 2003).

The first step of a DLA survey is to establish the redshift path length available to the discovery of DLA systems. The minimum starting wavelength of 3800 Å corresponds to $z = 2.12$ for the Lyα transition, and this sets the lowest redshift accessible to this survey. For each quasar, however, we define a unique starting redshift $z_{\rm start}$ by identifying the first pixel where the median signal-to-noise ratio (S/N) over 20 pixels exceeds 4. This criterion was chosen to (1) minimize the likelihood of identifying noise features as DLA systems, (2) achieve a high completeness limit, and (3) account for the presence of Lyman limit absorption. Consistent with previous studies, the ending redshift $z_{\rm end}$ corresponds to 3000 km s^{-1} blueward of Lyα emission. This criterion limits the probability of identifying DLA systems associated with the quasar, which may bias the analysis.

Special consideration is given to quasar spectra that show significant absorption lines at the quasar emission redshift (e.g., C IV, O VI). In previous studies, broad absorption line (BAL) quasars have been removed from the analysis primarily to prevent confusion with intrinsic O VI and/or N V absorption. We take a less conservative approach here. We visually inspected the 1252 quasars with $z_{\rm em} > 2.1$ and $r' < 19.5$ to identify quasars with associated absorption. In these cases, we limit the DLA search to 100 Å redward of O VI emission and 100 Å blueward of Lyα emission. However, if BAL contamination is determined to be too severe, the quasar is rejected from further analysis.

The majority of previous DLA surveys relied on low resolution "discovery" spectra to first identify DLA candidates. Follow-up observations were then made of these candidates to confirm DLA systems and measure their N(H I) values. A tremendous advantage of the SDSS spectra is that they have sufficient resolution to both readily identify DLA candidates and measure their N(H I) values. DLA candidates were identified using an algorithm tuned to the characteristics of the DLA profile, in particular its wide, saturated core. Our DLA-searching algorithm first determines a characteristic S/N (S/N$_{\rm qso}$) for each quasar spectrum. Ideally, we calculate this value blueward of Lyα emission specifically by taking the median S/N of 150 pixels lying 51–200 pixels blueward of Lyα emission. If the Lyα emission peak is at less than 200 pixels from the start of the spectrum, then we calculate S/N$_{\rm qso}$ from the median S/N of the 150 pixels, starting 50 pixels redward of Lyα emission. We then define a quantity $n_1 = {\rm S/N_{qso}}/2.5$ restricted to having a value between 1 and 2. At each pixel j in the spectrum, we then measure the fraction of pixels with S/N$_j < n_1$ in a window $6(1 + z_j)$ pixels wide where $z_j = \lambda_j/1215.67$ Å $- 1$. This window was chosen to match the width of the core of a DLA profile with SDSS spectral resolution and sampling. Importantly (for fiber data), the algorithm is relatively insensitive to the effects of poor sky subtraction. Furthermore, we stress that continuum fitting is unnecessary; the algorithm works directly on the fluxed data, because it focuses primarily on the core of the DLA profile.

This algorithm was developed through tests on both simulated spectra with resolution and S/N comparable to SDSS data, and also on a subset of SDSS spectra with known DLA systems. Our tests indicate that DLA candidates correspond to windows where the fraction of pixels with S/N $< n_1$ exceeds 60%. We recorded all regions satisfying this criterion and reduced them to individual candidates by grouping within 2000 km s^{-1} bins. In a sample of 1000 trials on simulated spectra with random N(H I) and redshift, we recover 100% of all DLA systems with log$N$(H I) > 20.4, and all but ~5% of the DLA systems with N(H I) $\approx 2 \times 10^{20}$ cm^{-2}. The algorithm is conservative in that it triggers many false-positive detections, the majority of which are BAL features or blended Lyα clouds. With custom software, it is easy to visually identify and account for these cases.

Table 1 lists the full sample of SDSS-DR1 quasars. The columns give the name, $z_{\rm em}$, $z_{\rm start}$, $z_{\rm end}$, a flag for BAL characteristics, and redshifts of DLA candidates, including the false-positive detections.

3. N(H I) ANALYSIS

The automated algorithm described in the previous section triggered 286 DLA candidates. We visually inspected the full set of candidates and identified ~100 as obvious false-positive detections. For the remainder of the systems, we fit a local continuum and a Voigt profile with FWHM = 2 pixels to the data. The Voigt-profile fits to the DLAs quoted in this paper are centered on the redshift determined by associated metal-line absorption. Because the metal lines are narrow, these redshifts are determined precisely. As emphasized by Prochaska

[1] See http://www.sdss.org.

TABLE 1
SDSS Quasar Sample

Name	z_{em}	z_{start}	z_{end}	f_{BAL}[a]	$z_{candidate}$
J094454.24−004330.3	2.292	2.150	2.259	0	
J095253.84+011422.1	3.024	2.154	2.984	0	2.204, 2.381
J100412.88+001257.5	2.239	2.156	2.207	0	
J100553.34+001927.1	2.501	2.155	2.466	0	
J101014.25−001015.2	2.190	2.143	2.158	0	
J101748.90−003124.5	2.283	2.156	2.250	0	
J101859.96−005420.2	2.183	2.147	2.151	0	
J102606.67+011459.0	2.266	2.157	2.233	0	
J102636.96+001530.2	2.178	...	...	0	
J102650.39+010518.3	2.274	2.177	2.192	1	

Note.—Table 1 is published in its entirety in the electronic edition of the *PASP*. A portion is shown here for guidance regarding its form and content.
[a] 0 = No BAL activity; 1 = modest BAL activity, included in analysis; 2 = strong BAL activity, excluded.

et al. (2003c), the N(H I) analysis is dominated by systematic error associated with continuum fitting and line blending of coincident Lyα clouds. The statistical error based on a χ^2 minimization routine would be unrealistically low and largely meaningless. Therefore, we perform a visual fit to the data and report a conservative systematic error that we believe encompasses an interval in N(H I) corresponding to a 95% confidence level. For a majority of the profiles, this corresponds to ± 0.15 dex, independent of N(H I) value.

The Lyα fits for all Lyα profiles satisfying $N(\mathrm{H\,I}) \geq 2 \times 10^{20}$ cm^{-2} criterion are plotted in Figure 1. Overplotted in each figure are the best fit and our assessment of the error corresponding to a 95% confidence level interval. Table 2 summarizes the absorption redshift, lists the N(H I) value and estimated uncertainty, and gives a brief comment for each profile (e.g., difficult continuum, severe line blending, poor S/N).

For ~10 of the DLA systems in the SDSS-DR1 sample, we have acquired higher resolution spectroscopy (FWHM ≈ 30 km s^{-1}) of the Lyα profile with the Echellette Spectrometer and Imager (ESI; Sheinis et al. 2002) on the Keck II Telescope. The ESI spectra suffer less from line blending and also allow for a more accurate determination of the quasar continuum. Furthermore, several of these systems were observed in previous studies. We find that our N(H I) values agree with all previous measurements to within 0.15 dex with no systematic offset. Therefore, we are confident in the N(H I) values reported here and the reported uncertainties.

4. ANALYSIS

4.1. $g(z)$ and $n(z)$

A simple yet meaningful description of the statistical significance of any quasar absorption-line survey is given by the redshift path density $g(z)$ (e.g., Lanzetta et al. 1991). This quantity corresponds to the number of quasars searched at a given redshift for the presence of a particular absorption feature (e.g., a DLA system). We have constructed $g(z)$ for the SDSS-DR1 sample by implementing the starting and ending redshifts listed in Table 1. Figure 2 presents $g(z)$ for (1) the SDSS-DR1 sample (*red dotted lines*), (2) the PMSI03 compilation (*dashed blue lines*), and (3) the combined surveys taking into account overlap between the two samples (*black solid line*). It is evident from Figure 2 that the SDSS-DR1 sample has greatest statistical impact at $z = 2$–3.2. With less than 20% of the projected SDSS database, the SDSS-DR1 exceeds the redshift path density of the previous two decades of research at $z = 2.5$. Although the SDSS-DR1 systems have only a modest contribution at $z > 3$, the projected $10\times$ increase in $g(z)$ for the full SDSS sample promises to have a major impact on DLA studies to at least $z = 4$.

Given a determination of $g(z)$, it is trivial to calculate the number density of DLA systems per unit redshift, $n(z)$. Integrating $n(z)$ over several redshift bins, we derive the results presented in Figure 3 for the SDSS-DR1 sample (*red*) and the combined surveys (*black*). Overplotted on the figure is the power-law fit to $n(z)$ from Storrie-Lombardi & Wolfe (2000): $n(z) = 0.055(1 + z)^{1.11}$. The SDSS-DR1 sample is in good agreement with previous analysis; this bolsters the assertion that our analysis has more than 95% completeness. The combined data sample has uncertainties in $n(z)$ of 10%–15% for $\Delta z = 0.5$ intervals. With future SDSS data releases, we will measure $n(z)$ in $\Delta z = 0.25$ intervals to better than 5% uncertainty. This measurement provides an important constraint on the H I cross section of high-redshift galaxies (e.g., Nagamine et al. 2004a) and thereby models of galaxy formation with cold dark matter cosmology (e.g., Kauffmann 1996; Ma et al. 1997). Table 3 lists the $n(z)$ values for the total sample for the redshift bins shown in Figure 3.

4.2. Ω_g

We now turn our attention to the cosmological baryonic mass density in neutral gas, Ω_g, as determined by DLA surveys. As first described by Wolfe (1985), one can calculate Ω_g for a given redshift interval by summing the N(H I) values of all DLA systems within that interval and comparing them against the total cosmological distance ΔX surveyed

$$\Omega_g = \frac{\mu m_{\mathrm{H}} H_0}{c\rho_c} \frac{\Sigma N(\mathrm{H\,I})}{\Delta X}, \quad (1)$$

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. We have calculated ΔX and Ω_g for the SDSS-DR1 sample and the PMSI03 compilation for a $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $H_0 = 70$ km s^{-1} Mpc^{-1} cosmology consistent with the current "concordance" cosmology (e.g., Spergel et al. 2003).

Implicit to equation (1) is the presumption that the DLA systems dominate Ω_g at all redshift. A principal result of PMSI03 was that at $z > 3.5$ there are fewer DLA systems with $N(\mathrm{H\,I}) > 10^{21}$ cm^{-2}, and, therefore, that absorption systems with

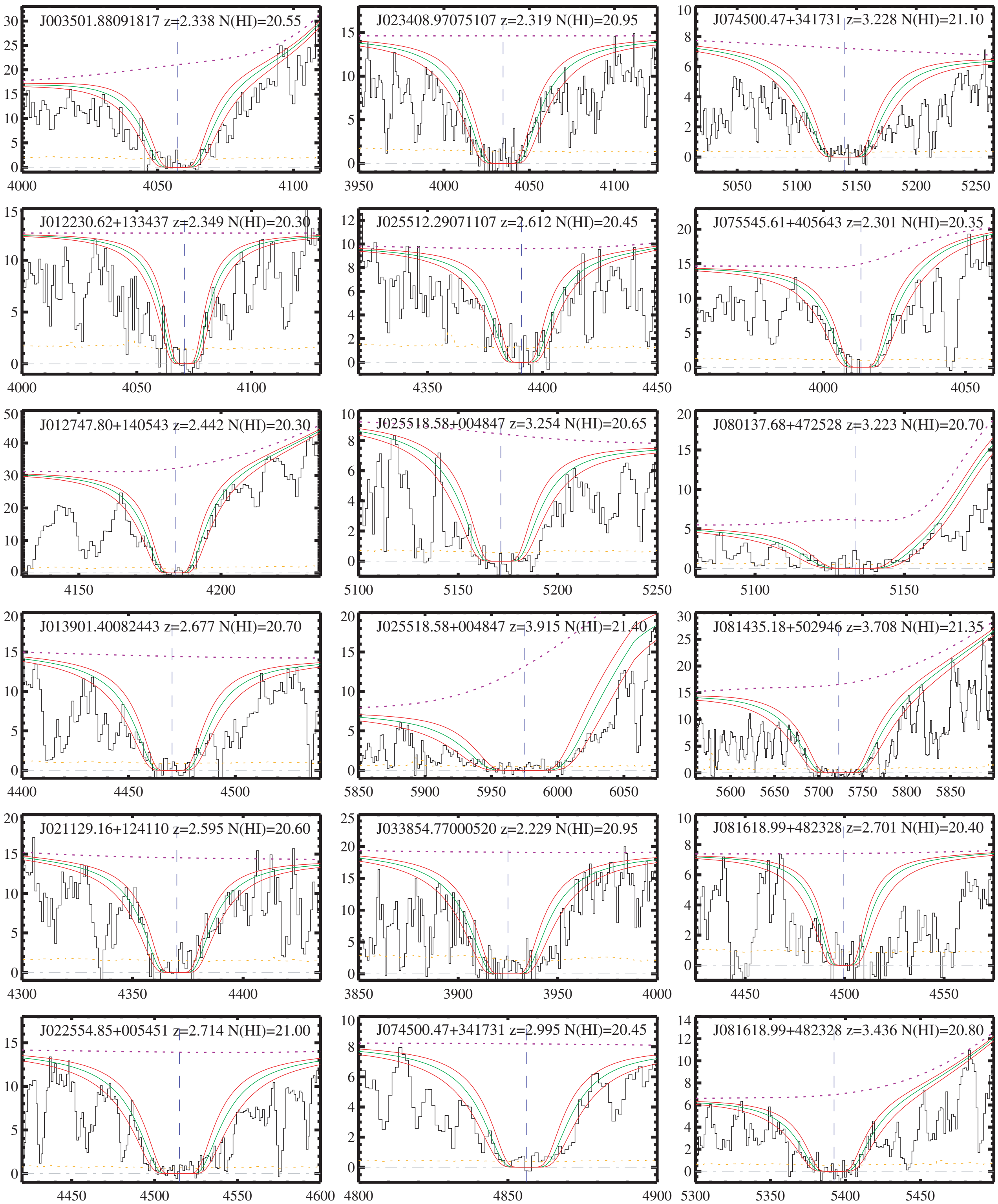

FIG. 1.—Lyα profiles of the 71 DLA systems comprising the full statistical sample from the SDSS Data Release 1. The dotted line traces the assumed continuum of the quasar, and the green solid line is a Voigt profile corresponding to the N(H I) values given in Table 2. All plots have angstroms along the x-axis and flux ($f_\lambda \times 10^{17}$ cgs) along the y-axis.

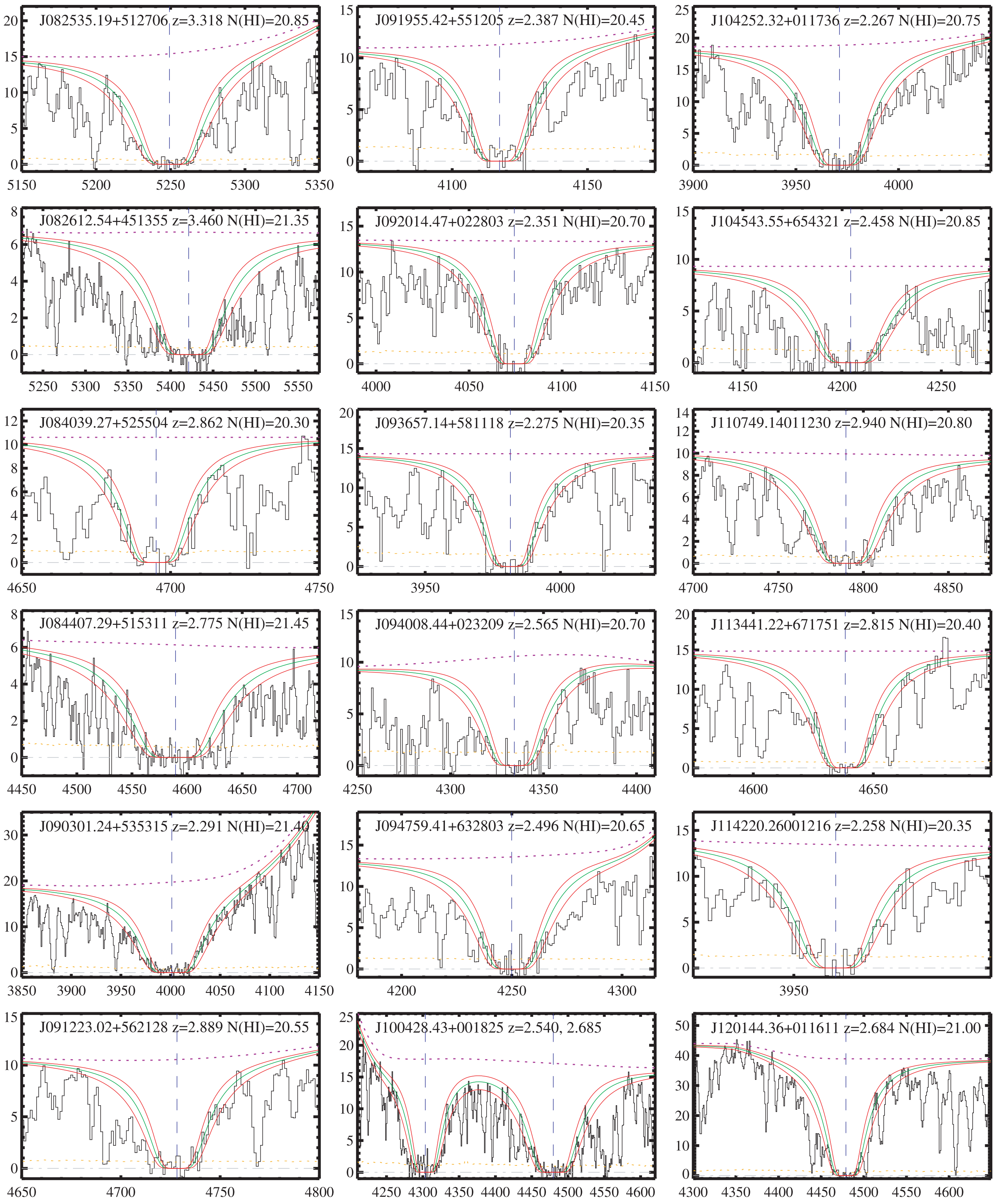

FIG. 1.—*Continued*

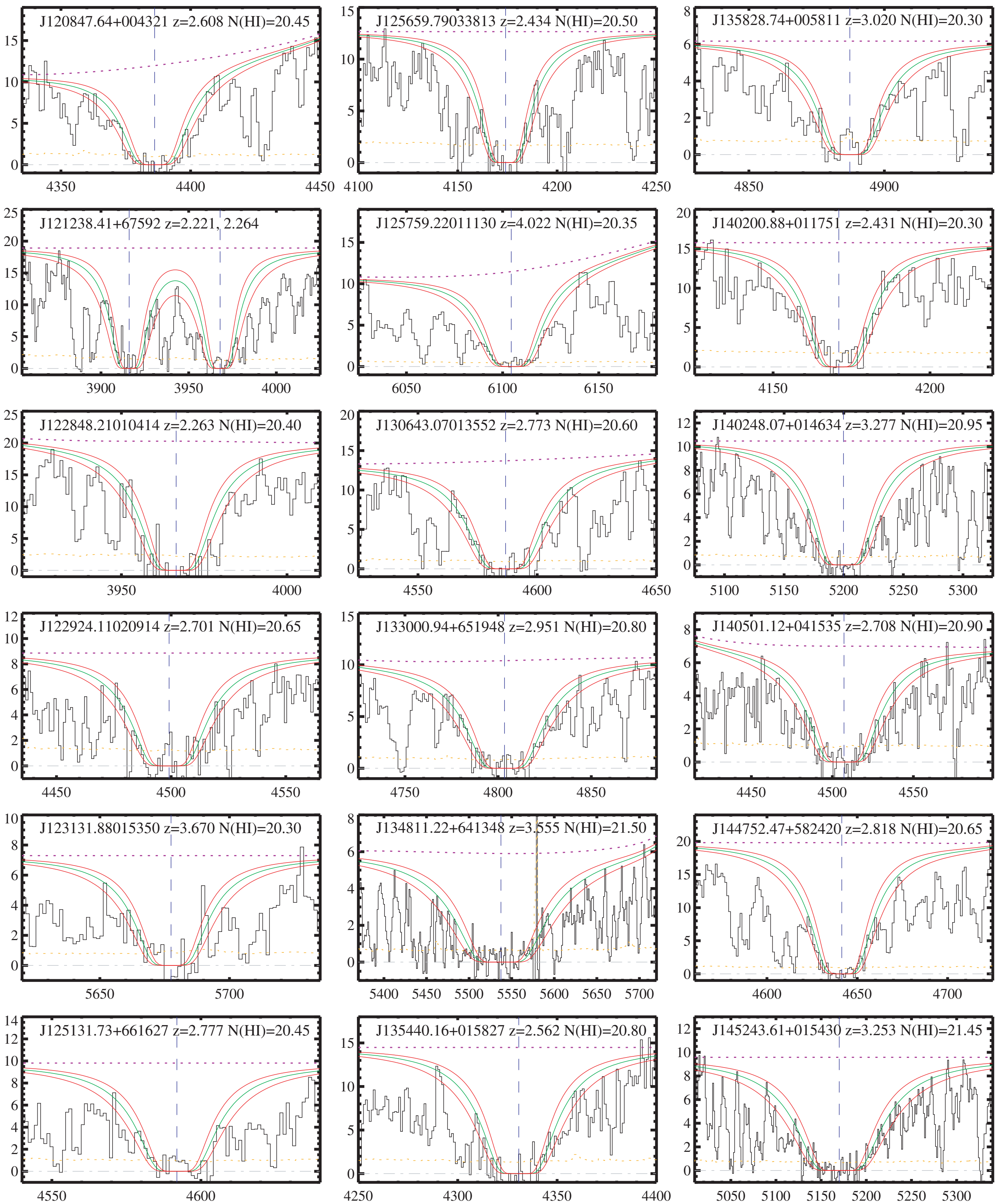

Fig. 1.—*Continued*

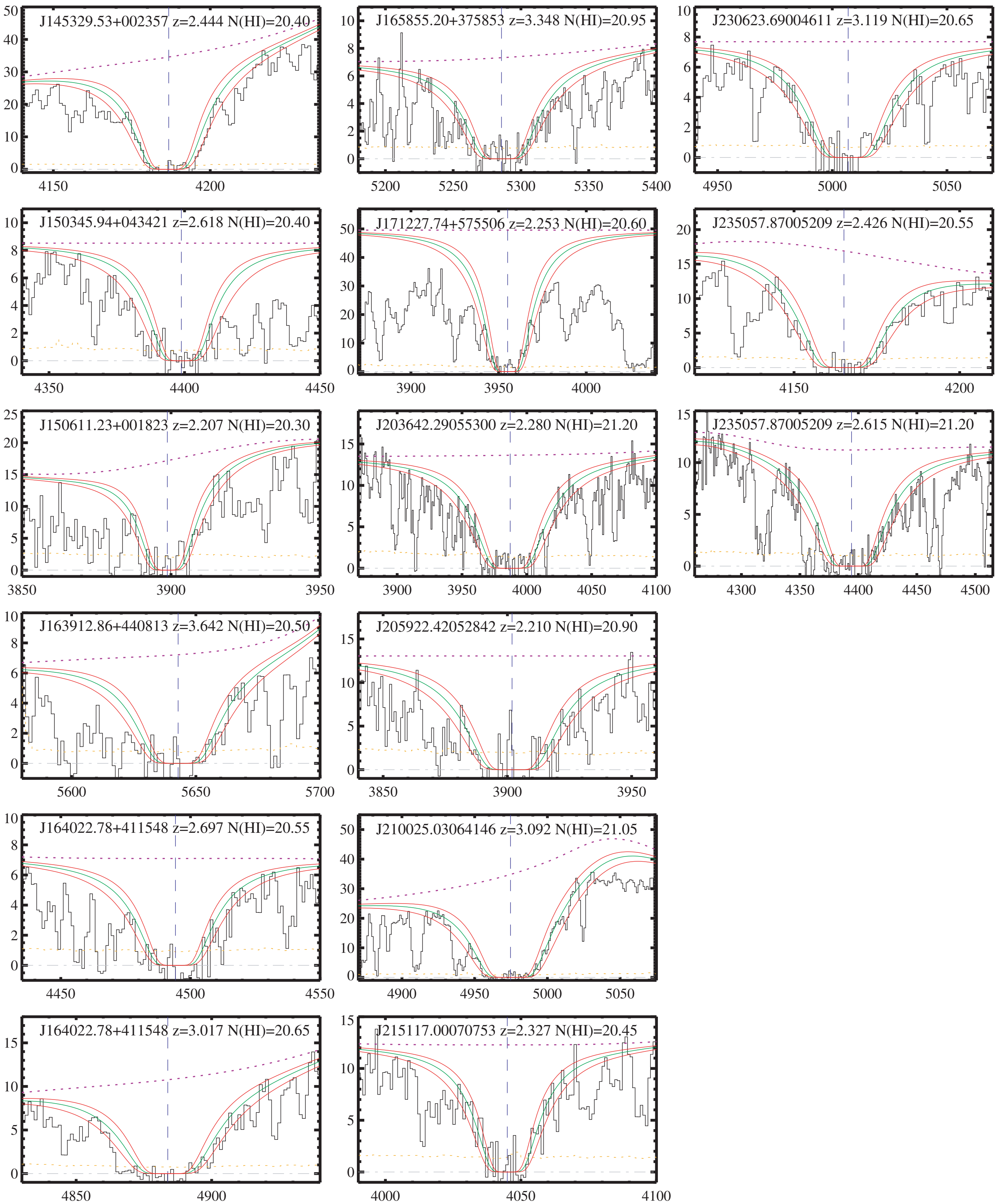

FIG. 1.—*Continued*

TABLE 2
SDSS DLA Sample

Name	r'	z_{em}	z_{abs}	log N(H I)	Comment
J003501.88−091817	19.10	2.420	2.338	$20.55^{+0.15}_{-0.15}$	Continuum
J012230.62+133437	19.32	3.010	2.349	$20.30^{+0.15}_{-0.15}$	
J012747.80+140543	18.73	2.490	2.442	$20.30^{+0.15}_{-0.15}$	Continuum, blending
J013901.40−082443	18.68	3.020	2.677	$20.70^{+0.15}_{-0.15}$	
J021129.16+124110	18.87	2.950	2.595	$20.60^{+0.15}_{-0.15}$	
J022554.85+005451	18.97	2.970	2.714	$21.00^{+0.15}_{-0.15}$	Blending
J023408.97−075107	18.97	2.540	2.319	$20.95^{+0.15}_{-0.15}$	
J025512.29−071107	19.43	2.820	2.612	$20.45^{+0.15}_{-0.20}$	
J025518.58+004847	19.27	3.990	3.254	$20.65^{+0.15}_{-0.15}$	Continuum, blending
			3.915	$21.40^{+0.15}_{-0.15}$	Continuum, blending
J033854.77−000520	18.78	3.050	2.229	$20.90^{+0.15}_{-0.15}$	Continuum, blending, poor S/N
J074500.47+341731	19.25	3.710	2.995	$20.45^{+0.15}_{-0.15}$	
			3.228	$21.10^{+0.15}_{-0.15}$	
J075545.61+405643	19.23	2.350	2.301	$20.35^{+0.20}_{-0.15}$	Blending
J080137.68+472528	19.42	3.280	3.223	$20.70^{+0.15}_{-0.15}$	Continuum, blending
J081435.18+502946	18.34	3.880	3.708	$21.35^{+0.15}_{-0.15}$	
J081618.99+482328	19.17	3.570	2.701	$20.40^{+0.20}_{-0.15}$	Continuum
			3.436	$20.80^{+0.15}_{-0.15}$	Continuum
J082535.19+512706	18.36	3.510	3.318	$20.85^{+0.15}_{-0.15}$	
J082612.54+451355	19.23	3.820	3.460	$21.35^{+0.15}_{-0.15}$	Blending, poor S/N
J084039.27+525504	19.34	3.090	2.862	$20.30^{+0.15}_{-0.15}$	Continuum
J084407.29+515311	19.44	3.210	2.775	$21.45^{+0.15}_{-0.15}$	Continuum
J090301.24+535315	18.56	2.440	2.291	$21.40^{+0.15}_{-0.15}$	
J091223.02+562128	19.09	3.000	2.889	$20.55^{+0.15}_{-0.15}$	Continuum
J091955.42+551205	19.02	2.510	2.387	$20.45^{+0.15}_{-0.15}$	Continuum
J092014.47+022803	19.21	2.940	2.351	$20.70^{+0.15}_{-0.15}$	Continuum
J093657.14+581118	19.03	2.540	2.275	$20.35^{+0.15}_{-0.15}$	Blending
J094008.44+023209	19.41	3.220	2.565	$20.70^{+0.15}_{-0.15}$	
J094759.41+632803	19.17	2.620	2.496	$20.65^{+0.15}_{-0.15}$	
J100428.43+001825	18.50	3.050	2.540	$21.00^{+0.15}_{-0.15}$	
			2.685	$21.35^{+0.15}_{-0.15}$	
J104252.32+011736	18.69	2.440	2.267	$20.75^{+0.15}_{-0.15}$	Continuum, poor S/N
J104543.55+654321	19.10	2.970	2.458	$20.85^{+0.15}_{-0.15}$	Continuum
J110749.14−011230	19.22	3.400	2.940	$20.80^{+0.15}_{-0.15}$	Blending
J113441.22+671751	18.59	2.960	2.815	$20.40^{+0.15}_{-0.15}$	Blending
J114220.26−001216	18.91	2.490	2.258	$20.35^{+0.15}_{-0.15}$	
J120144.36+011611	17.53	3.230	2.684	$21.00^{+0.15}_{-0.15}$	Blending
J120847.64+004321	19.19	2.720	2.608	$20.45^{+0.15}_{-0.15}$	
J121238.41+675920	18.68	2.570	2.221	$20.40^{+0.15}_{-0.15}$	
			2.264	$20.35^{+0.20}_{-0.20}$	
J122848.21−010414	18.23	2.660	2.263	$20.40^{+0.15}_{-0.15}$	
J122924.11−020914	19.27	3.620	2.701	$20.65^{+0.15}_{-0.15}$	Blending
J123131.88−015350	19.30	3.900	3.670	$20.30^{+0.15}_{-0.15}$	
J125131.73+661627	19.34	3.020	2.777	$20.45^{+0.15}_{-0.15}$	Blending
J125659.79−033813	19.08	2.970	2.434	$20.50^{+0.15}_{-0.15}$	
J125759.22−011130	18.87	4.110	4.022	$20.35^{+0.15}_{-0.15}$	
J130643.07−013552	18.82	2.940	2.773	$20.60^{+0.15}_{-0.15}$	
J133000.94+651948	18.89	3.270	2.951	$20.80^{+0.15}_{-0.15}$	Blending, poor S/N
J134811.22+641348	19.12	3.840	3.555	$21.50^{+0.15}_{-0.15}$	
J135440.16+015827	19.07	3.290	2.562	$20.80^{+0.15}_{-0.15}$	
J135828.74+005811	19.40	3.910	3.020	$20.30^{+0.15}_{-0.15}$	Blending
J140200.88+011751	18.81	2.950	2.431	$20.30^{+0.15}_{-0.15}$	
J140248.07+014634	18.84	4.160	3.277	$20.95^{+0.15}_{-0.15}$	

TABLE 2
(CONTINUED)

Name	r'	z_{em}	z_{abs}	log N(H I)	Comment
J140501.12+041535	19.31	3.220	2.708	$20.90^{+0.15}_{-0.15}$	Poor S/N
J144752.47+582420	18.37	2.980	2.818	$20.65^{+0.15}_{-0.15}$	Blending
J145243.61+015430	18.87	3.910	3.253	$21.45^{+0.15}_{-0.15}$	
J145329.53+002357	18.58	2.540	2.444	$20.40^{+0.15}_{-0.15}$	
J150345.94+043421	19.49	3.060	2.618	$20.40^{+0.20}_{-0.15}$	Blending
J150611.23+001823	18.89	2.830	2.207	$20.30^{+0.15}_{-0.15}$	Continuum
J163912.86+440813	19.22	3.770	3.642	$20.50^{+0.15}_{-0.15}$	
J164022.78+411548	19.41	3.080	2.697	$20.55^{+0.15}_{-0.15}$	
			3.017	$20.65^{+0.15}_{-0.15}$	
J165855.20+375853	19.13	3.640	3.348	$20.95^{+0.15}_{-0.15}$	Continuum
J171227.74+575506	17.46	3.010	2.253	$20.60^{+0.15}_{-0.15}$	Blending
J203642.29−055300	18.80	2.580	2.280	$21.20^{+0.15}_{-0.15}$	Continuum, blending
J205922.42−052842	19.01	2.540	2.210	$20.90^{+0.15}_{-0.15}$	Continuum, blending, poor S/N
J210025.03−064146	18.12	3.140	3.092	$21.05^{+0.15}_{-0.15}$	Blending
J215117.00−070753	19.26	2.520	2.327	$20.45^{+0.15}_{-0.15}$	Continuum
J230623.69−004611	19.23	3.580	3.119	$20.65^{+0.15}_{-0.15}$	Continuum
J235057.87−005209	18.79	3.020	2.426	$20.55^{+0.15}_{-0.15}$	Continuum, blending
			2.615	$21.20^{+0.15}_{-0.15}$	Continuum, blending

$N(\text{H I}) \lesssim 10^{20}$ cm^{-2} (the so-called sub-DLA) will contribute ~50% of Ω_g. This point is partially described by Figure 4, which presents the cumulative cosmological number density of DLA systems as a function of H I column density. The red curves correspond to the compilation analyzed by PMSI03; as emphasized by these authors, there is a significant drop in the fraction of DLA systems with large N(H I) at $z > 3.5$ in their compilation. The authors then argued that the sub-DLA make an important contribution to Ω_g at high redshift. The black lines in Figure 4 correspond to the combined sample. There is only a modest difference between the PMSI03 and combined samples for the $z = [2.4, 3.5)$ interval, but at $z > 3.5$ (*dotted lines*) the SDSS-DR1 results have greatly changed the picture.[2] Al-

[2] We also note that more accurate N(H I) measurements from Prochaska et al. (2003c) indicate that PMSI03 systematically underestimated several DLA systems with large N(H I) values. These new results are not included in Fig. 4, but are included in the results presented below.

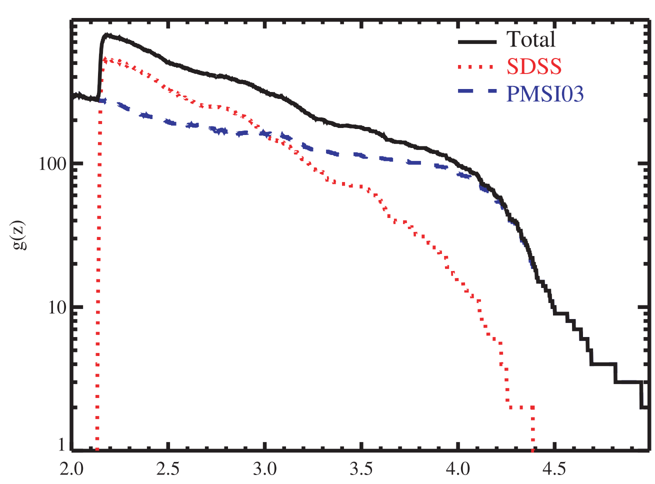

FIG. 2.—Redshift path density $g(z)$ as a function of redshift for (1) the SDSS-DR1 survey (*dotted red line*), (2) the PMSI03 compilation (*dashed blue line*), and (3) the combined surveys.

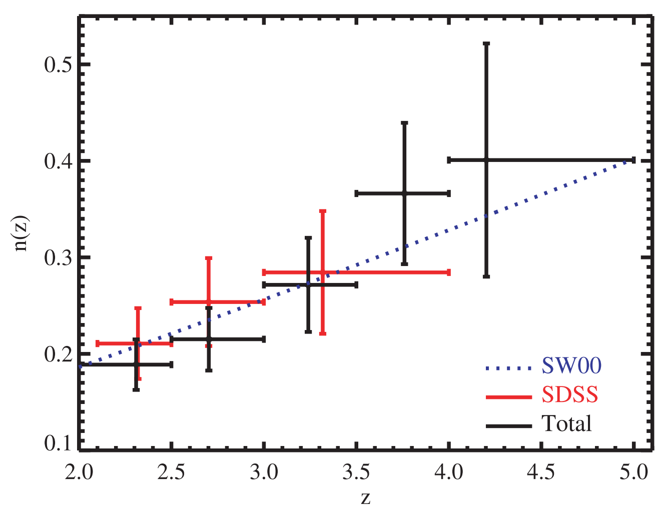

FIG. 3.—Incidence of DLA systems per unit redshift, $n(z)$, as a function of redshift for the SDSS-DR1 (*red points*) and total samples (*black points*). The vertical error bars reflect 1 σ uncertainty assuming Poissonian statistics, and the horizontal bars indicate the redshift interval. The dotted blue line is the fit to $n(z)$ from Storrie-Lombardi & Wolfe (2000): $n(z) = 0.055(1+z)^{1.11}$.

TABLE 3
Results

Sample	z	N	$n(z)$	ΔX[a]	$\Omega_g(10^{-3})$
SDSS	2.1–2.5	26	0.211 ± 0.037	505.0	$0.47^{+0.12}_{-0.12}$
	2.5–3.0	26	0.254 ± 0.046	420.9	$0.76^{+0.20}_{-0.18}$
	3.0–4.1	19	0.296 ± 0.065	266.3	$1.43^{+0.44}_{-0.45}$
Total	2.0–2.5	52	0.189 ± 0.026	880.8	$0.67^{+0.16}_{-0.14}$
	2.5–3.0	44	0.215 ± 0.032	704.5	$1.03^{+0.24}_{-0.26}$
	3.0–3.5	31	0.271 ± 0.049	421.8	$1.22^{+0.27}_{-0.25}$
	3.5–4.0	25	0.366 ± 0.073	268.0	$1.21^{+0.36}_{-0.36}$
	4.0–5.0	11	0.401 ± 0.121	113.5	$0.76^{+0.29}_{-0.26}$

[a] Assumes a $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $H_0 = 70$ km s^{-1} Mpc^{-1} cosmology.

though the SDSS-DR1 systems contribute only six new DLA systems at $z > 3.5$, half of these have $N(\text{H I}) > 10^{21}$ cm^{-2}. The resulting cumulative number density at $z > 3.5$ is now in rough agreement with the lower redshift interval (and the predictions of Nagamine et al. 2004a). Of course, we suspect the SDSS-DR1 sample shows an abnormally high fraction of DLA systems at $z > 3.5$ with $N(\text{H I}) > 10^{21}$ cm^{-2}. Similarly, we suspect the PMSI03 compilation had disproportionately too few systems with large N(H I). This speculation can only be tested through a significantly larger sample.

We can perform an analysis similar to PMSI03 to estimate the contribution of Lyman limit systems (LLSs) with $N(\text{H I}) = 10^{17.2}$–$10^{20.3}$ cm^{-2} to Ω_g in the combined sample. Adopting their power-law fit to the incidence of LLS $n(z)_{\rm LLS} = 0.07(1+z)^{2.45}$, one predicts 318 LLS with $z > 3.5$ for the combined sample in which 36 DLA systems are observed. Assuming the LLS column density distribution follows a power law $f(N)_{\rm LLS} = f_0 N(\text{H I})^\gamma$, we derive $\gamma = -1.31$ and $f_0 = 10^{4.66}$. We estimate the contribution of LLS to Ω_g by integrating

$$\Omega_{\rm LLS} = \frac{\mu m_{\rm H} H_0}{c\rho_c}\int_{10^{17.2}}^{10^{20.3}} N f(N)_{\rm LLS}\, dN = 0.00015. \qquad (2)$$

This value corresponds to less than 15% of Ω_g derived from $z > 3.5$ DLA systems (see below). The fractional contribution is 3 times lower (and $>4\,\sigma$ lower) than the results from PMSI03. It is important to note that this result has large statistical and systematic uncertainty. This includes the parameterization of $n(z)_{\rm LLS}$, the assumed functional form of $f(N)_{\rm LLS}$, and the statistical uncertainties in all quantities including Ω_g. Nevertheless, we conclude that there is no longer compelling evidence that LLSs with $N(\text{H I}) < 2 \times 10^{20}$ cm^{-2} contribute significantly to Ω_g at any redshift. Given the current uncertainties, however, the exact contribution of the Lyman limit and DLA systems to Ω_g will await future studies.

Restricting our analysis of Ω_g to the DLA systems, we derive Ω_g for the SDSS-DR1 sample and the combined data sets (Fig. 5, Table 3). The points plotted in Figure 5 are centered at the N(H I)-weighted redshift in each interval, and the horizontal errors correspond to the redshift bins analyzed. It is

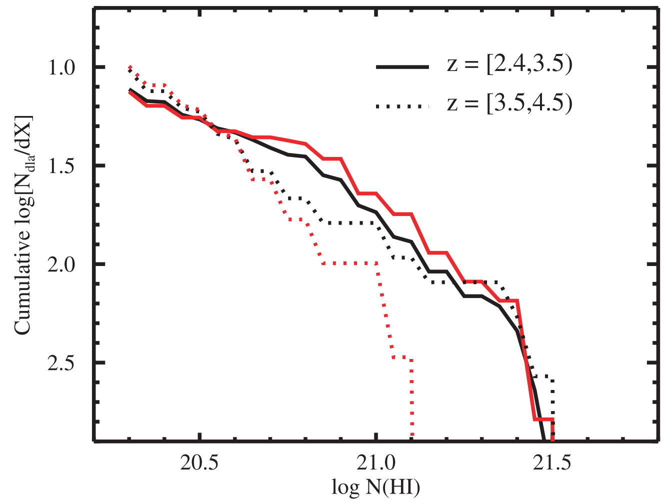

Fig. 4.—Cumulative logarithmic incidence of DLA systems per unit absorption distance interval dX as a function of log$N$(H I). The red curves correspond to the DLA compilation of PMSI03, and the black curves refer to the combined sample. Note that the high-redshift results have changed significantly by including the SDSS-DR1 sample.

difficult to estimate the error in Ω_g, because the uncertainty is dominated by sample size, especially the column density frequency distribution at $N(\text{H I}) > 10^{21}$ cm^{-2}. In the current analysis, we estimate 1 σ uncertainties through a modified bootstrap error analysis. Specifically, we examine the distribution of Ω_g values for 1000 trials in which we randomly select $m \pm p$ DLA

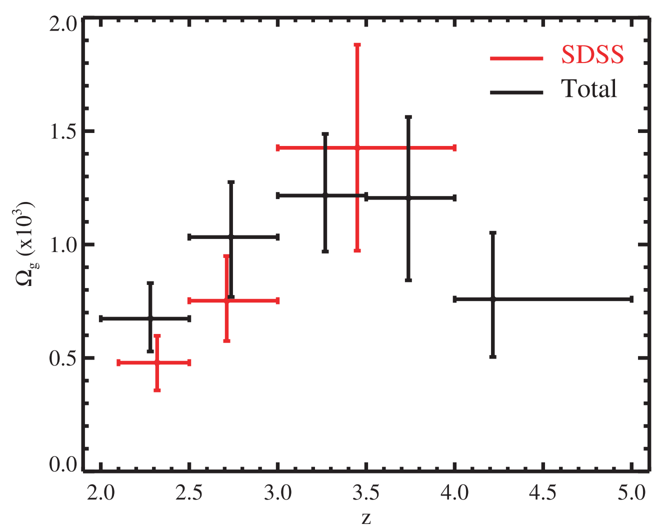

Fig. 5.—Cosmological baryonic-mass density in neutral gas, Ω_g, as derived from the DLA systems for the SDSS-DR1 sample (*red points*) and the combined sample (*black points*). The 1 σ vertical error bars were derived from a modified bootstrap analysis described in the text. We find that Ω_g is rising or unchanged to $z = 4$, and there is only a statistically insignificant decline at $z > 4$.

systems for each redshift interval containing m DLA systems and where p is a normally distributed random integer with standard deviation $\sqrt{m}$. The bootstrap technique provides a meaningful assessment of the uncertainty related to sample size, provided the observed data set samples a significant fraction of the intrinsic distribution. At present, we are not confident that this is the case at any redshift interval, particularly at $z > 3$. The results for the $z > 4$ redshift interval are an extreme example of this concern. The addition of one or two new DLAs with $N(\mathrm{H\,I}) > 10^{21}$ cm^{-2} would significantly increase Ω_g and its 1 σ uncertainty. Therefore, we caution the reader that the 1 σ errors reported in Table 3 likely underestimate the true uncertainty.

The SDSS-DR1 sample shows no evidence for a decline in Ω_g at high redshift; the results are even suggestive of an increasing baryonic-mass density at $z > 3$. We caution, however, that the uncertainties are large. Combining the SDSS-DR1 sample with the previous studies,[3] we reach a similar conclusion, except at $z > 4$, where the current results indicate a drop in Ω_g. As noted above, the results in the highest redshift interval are very uncertain, owing to the small sample size. At present, we consider it an open question as to whether Ω_g declines at high redshift.

One means of assessing the robustness of the Ω_g values to sample size is to cumulatively examine the total $N(\mathrm{H\,I})$ in the various redshift intervals. This quantity is presented in Figure 6 as a function of $N(\mathrm{H\,I})$ for the combined DLA sample. On the positive side, the total $N(\mathrm{H\,I})$ for the $z < 4$ samples all approach $10^{22.5}$ cm^{-2}, which is ~10 × larger than the highest $N(\mathrm{H\,I})$ values observed to date. Therefore, the results in these intervals are reasonably robust to the inclusion of an "outlier" with $N(\mathrm{H\,I}) \approx 10^{22}$ cm^{-2}. On the other hand, the curves in Figure 6 demonstrate that DLA systems with $N(\mathrm{H\,I}) > 10^{21}$ cm^{-2} do contribute ~50% of the total $N(\mathrm{H\,I})$ in each interval. This point stresses the sensitivity of Ω_g to sample size; there are relatively few DLA systems with $N(\mathrm{H\,I}) > 10^{21}$ cm^{-2} in each interval. Sample variance will be important in any given interval for Ω_g until it includes many systems with $N(\mathrm{H\,I}) > 10^{21}$ cm^{-2}.

5. SUMMARY AND CONCLUDING REMARKS

In this paper, we have introduced an automated approach for identifying DLA systems in the SDSS quasar database. We have applied our method to the Data Release 1 quasar sample and have identified a statistical sample of 71 DLA systems, including more than 50 previously unpublished cases. Remarkably, the SDSS Data Release 1 exceeds the statistical significance of the previous two decades of DLA research at $z \approx 2.5$. More importantly, this sample was drawn from a well-defined, homogeneous data set of quasar spectroscopy. We present measurements of the number per unit redshift, $n(z)$, of the DLA population and the contribution of these systems to the cosmological baryonic-mass density in neutral gas, Ω_g. Although the SDSS-DR1 sample does not offer a definitive assessment of either of these quantities, future SDSS data releases will provide a major advancement over all previous work.

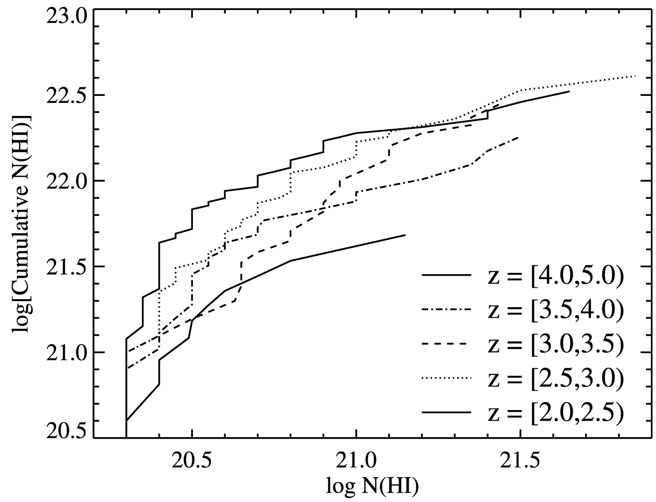

FIG. 6.—Cumulative total $N(\mathrm{H\,I})$ as a function of $\log N(\mathrm{H\,I})$ for the redshift intervals displayed in Fig. 5. These curves provide a qualitative assessment of the robustness of the Ω_g values to the addition of new DLA systems, especially "outliers" with large $N(\mathrm{H\,I})$.

Our measurements of $n(z)$ are consistent with previous results suggesting a high completeness level for our DLA survey of the SDSS-DR1. We find Ω_g increases with redshift to at least $z = 3$ and is consistent with increasing to $z = 4$ and beyond. This latter claim, however, is subject to significant uncertainty relating to sample size. Perhaps the most important result of our analysis is that the full DLA sample no longer shows significantly fewer DLA systems with large $N(\mathrm{H\,I})$ at $z > 3.5$. This contradicts the principal result of PMSI03 from their analysis of the pre-SDSS DLA compilation. Apparently, their maximum likelihood approach failed to adequately assess uncertainty related to sample size. With the inclusion of only six new DLAs, we no longer find that Lyman limit systems with $N(\mathrm{H\,I}) < 2 \times 10^{20}$ cm^{-2} are required in an analysis of Ω_g.

Before concluding, we offer several additional criticisms of the PMSI03 analysis and the role of sub-DLA systems. First, these authors assumed a three-parameter Γ-function for the column density frequency distribution of absorption systems with $N(\mathrm{H\,I}) > 10^{17.2}$ cm^{-2}: $f(N) = (f_*/N_*)(N/N_*)^{-\beta} e^{-N/N_*}$. Although this function gives a reasonable fit to the column density frequency distribution of the DLA systems, it is not physically motivated,[4] and more importantly, it places much greater emphasis on sub-DLA than other functions (e.g., a broken power law). Future assessments must include other functional forms

[3] We have updated the measurements presented in PMSI03 to match the ones presented in Prochaska et al. (2003b).

[4] In fact, this curve does not smoothly connect to the power law derived for quasar absorption lines with $N(\mathrm{H\,I}) < 10^{17.2}$ cm^{-2}.

to examine this systematic uncertainty. Second, the authors did not fit for the normalization of the distribution function f_*. The uncertainty in this parameter could easily contribute an additional 50% or more to the error budget. Third, their treatment did not account for sample variance; the uncertainties these authors reported were severe underestimates. Finally (and perhaps most importantly), a recent analysis of a sub-DLA sample by Dessauges-Zavadsky et al. (2003) has shown that these absorption systems have very high ionization fractions (see also J. Howk & A. Wolfe 2004, in preparation). Although these absorption systems may ultimately make an important contribution to the total H I mass density of the universe, they are intrinsically different from the DLA systems. Indeed, a more appropriate title for this subset of Lyman limit systems is the "super-LLS." This gas—in its present form—cannot contribute to star formation and is unlikely to be directly associated with galactic disks or the inner regions of protogalactic "clumps." Any interpretation of results related to the super-LLS must carefully consider these points (e.g., Maller et al. 2003; Péroux et al. 2003).

We acknowledge the tremendous effort put forth by the SDSS team to produce and release the SDSS survey. We thank Art Wolfe, Gabe Prochter, John O'Meara, J. Chris Howk, and Ben Weiner for helpful comments and suggestions. J. X. P. and S. H. F. are partially supported by NSF grant AST-0307408 and its REU subcontract.

REFERENCES

Abazajian, K., et al. 2003, AJ, 126, 2081
Curran, S. J., Webb, J. K., Murphy, M. T., Bandiera, R., Corbelli, E., & Flambaum, V. V. 2002, Publ. Astron. Soc. Australia, 19, 455
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
Fall, S. M., & Pei, Y. C. 1993, ApJ, 402, 479
Kauffmann, G. 1996, MNRAS, 281, 475
Lanzetta, K. M., Wolfe, A. M., Turnshek, D. A., Lu, L., McMahon, R. G., & Hazard, C. 1991, ApJS, 77, 1
Ledoux, C., Petitjean, P., & Srianand, R. 2003, MNRAS, 396, 429
Lu, L., Sargent, W. L. W., Barlow, T. A., Churchill, C. W., & Vogt, S. 1996, ApJS, 107, 475
Ma, C.-P., Bertschinger, E., Hernquist, L., Weinberg, D. H., & Katz, N. 1997, ApJ, 484, L1
Maller, A. H., Prochaska, J. X., Somerville, R. S., & Primack, J. R. 2003, MNRAS, 343, 268
Mathlin, G. P., Baker, A. C., Churches, D. K., & Edmunds, M. G. 2001, MNRAS, 321, 743
Nagamine, K., Springel, V., & Hernquist, L. 2004a, MNRAS, 348, 421
———. 2004b, MNRAS, 348, 435
Ostriker, J. P., & Heisler, J. 1984, ApJ, 278, 1
Pei, Y. C., & Fall, S. M. 1995, ApJ, 454, 69
Péroux, C., McMahon, R., Storrie-Lombardi, L., & Irwin, M. J. 2003, MNRAS, 346, 1103 (PMSI03)
Pettini, M., Smith, L. J., Hunstead, R. W., & King, D. L. 1994, ApJ, 426, 79
Prochaska, J. X., Howk, J. C., & Wolfe, A. M. 2003a, Nature, 423, 57
Prochaska, J. X., Gawiser, E., Wolfe, A. M., Castro, S., & Djorgovski, S. G. 2003b, ApJ, 595, L9
Prochaska, J. X., Gawiser, E., Wolfe, A. M., Cooke, J., & Gelino, D. 2003c, ApJS, 147, 227
Prochaska, J. X., & Wolfe, A. M. 1997, ApJ, 486, 73
Rao, S. M., & Turnshek, D. A. 2000, ApJS, 130, 1
Sheinis, A. I., Miller, J., Bigelow, B., Bolte, M., Epps, H., Kibrick, R., Radovan, M., & Sutin, B. 2002, PASP, 114, 851
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., & Wolfe, A. M. 2000, ApJ, 543, 552
Spergel, D., et al. 2003, ApJS, 148, 175
Vladilo, G. 1998, ApJ, 493, 583
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., Prochaska, J. X., & Gawiser, E. 2003, ApJ, 593, 215
Wolfe, A. M., Turnshek, D. A., Smith, H. E., & Cohen, R. D. 1986, ApJS, 61, 249