The H I Column Density Distribution of the Galactic Disk and Halo

David M. French1, Andrew J. Fox2, Bart P. Wakker3, Nicolas Lehner5, J. Christopher Howk5, Blair D. Savage3, Sanchayeeta Borthakur8, J. Christopher Howk5, Bart P. Wakker3, Timothy Heckman4, John O'Meara7, Sanchayeeta Borthakur8, and Colin Norman1,4

1 Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA; dfrench@stsci.edu
2 AURA for ESA, Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA
3 Department of Astronomy, University of Wisconsin—Madison, Madison, WI 53703, USA
4 Center for Astrophysical Sciences, Department of Physics & Astronomy, The Johns Hopkins University, Baltimore, MD 21218, USA
5 Department of Physics, University of Notre Dame, 225 Nieuwland Science Hall, Notre Dame, IN 46556, USA
6 Institut für Physik und Astronomie, Universität Potsdam, Haus 28, Karl-Liebknecht-Str. 24/25, D-14476 Golm (Potsdam), Germany
7 W.M. Keck Observatory, 65-1120 Mamalahoa Highway, Kamuela, HI 96743, USA
8 School of Earth and Space Exploration, Arizona State University, 781 Terrace Mall, Tempe, AZ 85287, USA

Received 2020 November 24; revised 2021 August 2; accepted 2021 August 5; published 2021 December 9

Abstract

We present a census of neutral gas in the Milky Way disk and halo down to limiting column densities of \(N(\text{HI}) \sim 10^{14} \text{cm}^{-2}\) using measurements of H I Lyman series absorption from the Far Ultraviolet Spectroscopic Explorer. Our results are drawn from an analysis of 25 AGN sight lines spread evenly across the sky with Galactic latitude \(|b| \gtrsim 20^\circ\). By simultaneously fitting multi-component Voigt profiles to 11 Lyman series absorption transitions covered by FUSE (Ly\(\alpha\)–Ly\(\mu\)) plus HST measurements of Ly\(\alpha\), we derive the kinematics and column densities of a sample of 152 H I absorption components. While saturation prevents accurate measurements of many components with column densities \(17 \lesssim \log N(\text{HI}) \lesssim 19\), we derive robust measurements at \(\log N(\text{HI}) \gtrsim 17\) and \(\log N(\text{HI}) \gtrsim 19\). We derive the first ultraviolet H I column density distribution function (CDDF) of the Milky Way, both globally and for low-velocity (ISM), intermediate-velocity clouds (IVCs), and high-velocity clouds (HVCs). We find that IVCs and HVCs show statistically indistinguishable CDDF slopes, with \(\beta_{\text{IVC}} = -1.01^{+0.15}_{-0.14}\) and \(\beta_{\text{HVC}} = -1.05^{+0.07}_{-0.06}\). Overall, the CDDF of the Galactic disk and halo appears shallower than that found by comparable extragalactic surveys, suggesting a relative abundance of high column density gas in the Galactic halo. We derive the sky-covering fractions as a function of H I column density, finding an enhancement of IVC gas in the northern hemisphere compared to the south. We also find evidence for an excess of inflowing H I over outflowing H I, with \(-0.88 \pm 0.40 \, M_\odot \, \text{yr}^{-1}\) of HVC inflow versus \(\approx 0.20 \pm 0.10 \, M_\odot \, \text{yr}^{-1}\) of HVC outflow, confirming an excess of inflowing HVCs seen in UV metal lines.

Unified Astronomy Thesaurus concepts: Interstellar medium (847); Ultraviolet astronomy (1732); Circumgalactic medium (1879); High-velocity clouds (735); Quasar absorption line spectroscopy (1317); Milky Way Galaxy (1054)

Supporting material: figure set

1. Introduction

The extended gaseous halos of galaxies (the circumgalactic medium; CGM) play essential roles in the life cycles of galaxies. These vast reservoirs of gas, extending out to hundreds of kiloparsecs around galaxy disks, contain at least as much baryonic matter as the galaxy disks themselves (e.g., Peeples et al. 2014; Werk et al. 2014; Keeney et al. 2017) and are fed by a number of sources, including infalling intergalactic matter and outflowing feedback material from their host galaxies. As a consequence, they are complex, turbulent, and multiphase. The cool phase\(^9\) in particular represents star formation fuel that has yet to be accreted. Understanding how this gas is distributed in galactic halos is a key step toward understanding how galaxies grow and evolve across cosmic time (see reviews by Putman et al. 2012; Tumlinson et al. 2017).

CGM studies have largely been separated into distinct Galactic and extragalactic camps due to their unique viewpoints and associated observational challenges. Sitting inside our own Galaxy, we can achieve full sky coverage and thus see the entire halo at once, using both absorption and emission. The cool gas in the Milky Way disk and halo has been extensively studied via H I 21 cm emission and UV metal-line absorption for decades. The resulting gas distribution has been historically separated by LSR velocity, with the assumption that low-velocity (\(|v_{\text{LSR}}| \lesssim 40 \text{ km s}^{-1}\)) gas traces the Galactic ISM (disk), and higher \(v_{\text{LSR}}\) gas traces the Galactic halo via intermediate-velocity (\(40 \lesssim |v_{\text{LSR}}| < 100 \text{ km s}^{-1}\)) and high-velocity (\(|v_{\text{LSR}}| \gtrsim 100 \text{ km s}^{-1}\)) clouds (IVCs/HVCs; see reviews by Wakker & van Woerden 1997; Putman et al. 2012; Richter 2017). This velocity-based division between disk and halo is generally supported by the available distance information. The H I disk has a scale height of \(\approx 200–300 \, \text{pc}\) (Dickey & Lockman 1990), whereas IVCs are found at distances of \(d \approx 0.5–2.0 \, \text{kpc}\) (Richter 2017) from the Sun, and HVCs are at \(d \approx 5–15 \, \text{kpc}\) (Wakker 2001; Thom et al. 2008; Wakker et al. 2008; Lehner & Howk 2011; Smoker et al. 2011; Richter et al. 2015). However, the HVCs with good distance constraints are 21 cm bright clouds, which only uncover the tip of the Galactic-halo iceberg because current 21 cm surveys are insensitive to H I column densities below \(\sim 10^{18} \text{cm}^{-2}\). We follow the velocity-based distinction between disk and halo in this paper, given the lack of a better alternative for a sample of absorbers with no direct distance information. However, this approach is inherently imprecise, as it relies on the assumption that halo gas exists exclusively at high velocity. Simulations show that a

\(9\) In this paper, we use the IGM/CGR definition of temperature where cool, warm, and hot denote \(T_{\text{gas}}/K = 4–5, 5–6, \) and 6, respectively.
large fraction of halo gas may be hidden at low velocity, where it blends with the disk (e.g., Zheng et al. 2020).

One key measure of the gas distribution in the CGM is the column density distribution function (CDDF), which was developed to characterize the distribution drawn from blind extragalactic surveys. The CDDF describes the relative number density of absorbers within a column density interval, $dN$, normalized by the redshift path, $d\tau$, over which an absorption survey has been completed. Tytler (1987) was the first to show that the distribution of extragalactic Ly$\alpha$ absorbers at $z = 0.2$–3.5 can be well fit by a single power-law model over seven orders of magnitude of H$\alpha$ column, from log $N$(H$\alpha$) = 14–21, with $f(N) = C N^{-\beta}$. The slope, $\beta$, of this power-law model encodes fundamental information about the mass distribution of the gas clouds and the ionizing radiation field incident on the gas. More recent studies have found evidence for at least one break in the power-law distribution for the extragalactic H$\alpha$ CDDF (e.g., Lehner et al. 2007; Ribaudo et al. 2011; Kim et al. 2013; O’Meara et al. 2013; Prochaska et al. 2014, and Rudie et al. 2013).

Our own Galaxy provides the unique point of view necessary to construct a complete CDDF from a single galaxy. However, observing Galactic H$\alpha$ presents its own set of difficulties. Absorption from the Galactic disk completely dominates Ly$\alpha$ at a $\sim$2000 km s$^{-1}$ window, masking the presence of any weaker components from the halo. H$\alpha$ 21 cm radio observations have the velocity resolution to separate individual components, but lack the sensitivity to probe gas below $\sim$10$^{18}$ cm$^{-2}$, and such gas is possibly lost near $v = 0$ km s$^{-1}$. The best solution to this issue is to observe the higher-order Lyman series lines, which become progressively less saturated as one moves up the series, and allow one to probe a much wider range of H$\alpha$ column densities.

For this reason, we are conducting a survey of the Galactic halo using archival Far Ultraviolet Spectroscopic Explorer (FUSE) sight lines toward background AGN. FUSE produced the largest and most sensitive sample of Galactic halo sight lines covering the Lyman series in absorption to date. The FUSE wavelength coverage allows for the simultaneous fitting of Galactic Lyman series absorption down to Ly$\mu$ 917.1805 (the 12th transition of the Lyman series), allowing us to reliably probe column densities down to log $N$(H$\alpha$) $\lesssim$ 14. Although these measurements are technically challenging, owing to saturation, blending with H$\beta$ and intergalactic absorption, and low signal-to-noise ratio (S/N), there is enough useful information in the Lyman series to reliably measure key H$\alpha$ properties including column densities and kinematics. In particular, even though the lower-order lines are saturated for typical H$\alpha$ column densities, the higher-order lines are often unsaturated and therefore on the linear part of the curve of growth, allowing more accurate column density measurements. This technique has been demonstrated via Lyman series absorption studies of a number of individual sight lines (Sembach et al. 2002; Collins et al. 2004; Fox et al. 2005; Ganguly et al. 2005; Zech et al. 2008; Richter et al. 2009) and as part of an O VI survey (Fox et al. 2006), but never in a blind H$\alpha$ survey.

This paper is organized as follows. In Section 2, we discuss the sample, data reduction, and Voigt-profile fitting procedures. In Section 3, we discuss the observational properties of the H$\alpha$ sample, including their basic properties, CDDF, sky-covering fraction, and breakdown into inflow and outflow. We present a summary in Section 4. Finally, in Appendix A, we present an

| Line       | $\lambda_0$ | $f$-value | log $N(\tau_0=1)^a$ |
|------------|-------------|-----------|---------------------|
| Ly$\alpha$ | 1215.6700   | 0.416400  | 13.30 13.60         |
| Ly$\beta$  | 1025.7223   | 0.079120  | 14.09 14.39         |
| Ly$\gamma$ | 972.5368    | 0.029000  | 14.55 14.85         |
| Ly$\delta$ | 949.7431    | 0.013940  | 14.88 15.18         |
| Ly$\epsilon$ | 937.8035  | 0.007799  | 15.14 15.44         |
| Ly$\zeta$  | 930.7483    | 0.004814  | 15.35 15.65         |
| Ly$\eta$   | 926.2257    | 0.003183  | 15.53 15.83         |
| Ly$\theta$ | 923.1504    | 0.002162  | 15.69 15.99         |
| Ly$\iota$  | 920.9631    | 0.001605  | 15.83 16.13         |
| Ly$\kappa$ | 919.3514    | 0.001200  | 15.96 16.26         |
| Ly$\lambda$| 918.1294    | 0.000921  | 16.07 16.37         |
| Ly$\mu$    | 917.1806    | 0.000723  | 16.18 16.48         |

Notes. Atomic data for the Lyman series transitions under study are taken from Jitrik & Bunge (2004). These values are close to but slightly different than those in Morton (2003).

$a$ H$\alpha$ column density at which the optical depth at line center is unity, for two representative values of the line width ($b$ in km s$^{-1}$). This indicates the regime where each line is unsaturated, and therefore amenable to precise measurement.

## 2. Methods

### 2.1. The Lyman Series of Neutral Hydrogen

H$\alpha$ is the only directly observable ionization stage of the dominant element in the Universe. The most sensitive lines of H$\alpha$ available at any wavelength are the Lyman series, representing the resonance transitions out of the ground state ($n = 1$). Absorption in the Lyman series probes H$\alpha$ in the UV to very sensitive levels, provided that FUV-bright background sources are identified. Since FUSE covered the wavelength range 905–1195 Å, its bandpass encompasses the entire H$\alpha$ Lyman series except Ly$\alpha$. This makes FUSE a well-suited instrument to conduct a sensitive H$\alpha$ survey.

The basic atomic data, rest wavelengths ($\lambda_0$) and oscillator strengths ($f$), for the Lyman series transitions are given in Table 1. This table includes a calculation of the H$\alpha$ column density at which each line reaches an optical depth at line center of 1, $N(\tau_0 = 1) = m_e e^b /\sqrt{\pi} e^{\lambda_0 / 2\lambda}$, where $m_\text{e}$ is the electron mass, $e$ is the speed of light, $\lambda$ is the electronic charge, and $b$ is the Doppler line width. We consider two different values of $b$, 15 and 30 km s$^{-1}$, chosen to be representative of H$\alpha$ at $|b| > 20^\circ$. The purpose of calculating $N(\tau_0 = 1)$ is that it gives an indication of the column density where each line is unsaturated and therefore amenable to precise measurement. The 12 Lyman lines covered by our data cover 3 dex in log $N(\tau_0)$, from $\approx 13.3$ for Ly$\alpha$ to $\approx 16.2$ for Ly$\mu$, revealing the diagnostic power of the Lyman series for characterizing diffuse H$\alpha$ over a wide range of conditions.

### 2.2. Sample Selection

We selected our sample from a preliminary pool of the 67 FUSE AGN spectra with a signal-to-noise ratio per resolution element at 977 Å of (S/N)$_977 \geq 4$. Many of these proved to be unusable due to their low redshifts, which resulted in strong blending of intrinsic AGN emission or absorption. Several others were simply too low (S/N) to be useful. After careful inspection of each spectrum and removing the unusable cases,
we arrived at a final sample of 25 sight lines for which we could produce fits, all with \((S/N)_{\text{977}} \geq 6\) (see Table 4). The on-sky locations of both our final sample and the rejected sight lines are shown in Figure 1.

Of these 25 sight lines, 19 have HI measurements presented in Fox et al. (2006); however, that paper only measured one Lyman series line per sight line using apparent optical depth integrations; here we model the entire series using a full Voigt-profile fitting analysis. In addition, the majority of the sight lines were analyzed as part of the FUSE O VI survey (Savage et al. 2003; Sembach et al. 2003; Wakke et al. 2003). The majority also have Hubble Space Telescope/Cosmic Origins Spectrograph data covering UV metal lines (Richter et al. 2017; Fox et al. 2020), though (with the exception of the Ly\(\alpha\)) these are outside the scope of this analysis.

### 2.3. Data Reduction

For each sight line, we reduced the FUSE data with the reduction pipeline CALFUSE version 2.4 or higher (Dixon et al. 2007). This pipeline provides flux-calibrated, wavelength-calibrated spectra from the eight detector segments (LiF1A, LiF1B, LiF2A, LiF2B, SiC1A, SiC1B, SiC2A, SiC2B) with a pixel size of 2 km s\(^{-1}\) and velocity resolution of 20 km s\(^{-1}\) (FWHM). Details of the FUSE satellite and its on-orbit performance are given in Moos et al. (2000) and Sahnow et al. (2000).

All of our FUSE sight lines except for NGC 1068 and 1H0707-495 have also been observed by the Cosmic Origins Spectrograph (COS) and/or the Space Telescope Imaging Spectrograph (STIS) on board HST. For these targets, we obtained their G130M (COS) or E140M (STIS) spectra from the Mikulski Archive for Space Telescopes in order to include the Ly\(\alpha\) transition in our fitting procedure. From an analysis of four sight lines at varying signal-to-noise ratios, we found that excluding Ly\(\alpha\) from the fits (i.e., fitting the FUSE data alone) can lead to under-predicting the column density of the strongest components, in some cases by orders of magnitude. See Appendix A for a detailed discussion of this effect. Therefore we include Ly\(\alpha\) in the fits whenever possible, even though this means including data taken by a different instrument. These Ly\(\alpha\) data have been independently fit by Wakker et al. (2011), and we use their results as an independent check on the total HI column densities.

### 2.4. Voigt-Profile Fitting Procedure

We used the Voigt-profile fitting software VoigtFit (Krogager 2018) to perform all of our Lyman series fits, with atomic data sourced from Morton (2003). The code allows for simultaneous \(\chi^2\)-minimization fitting of multiple components across all 12 Lyman series transitions observed by both COS/STIS (Ly\(\alpha\)) and FUSE (Ly\(\beta\) – Ly\(\delta\)). The FUSE and STIS/E140M line-spread functions (LSFs) were approximated as Gaussian distributions, with FWHM = 20.0 km s\(^{-1}\) for FUSE and FWHM = 6.5 km s\(^{-1}\) for STIS/E140M. For COS spectra, we adopted the LSF tables made available by STScI for the appropriate central wavelength setting, wavelength, and detector lifetime position. These tabulated COS LSFs account for extended non-Gaussian wings in the line profiles. The LSFs are provided as inputs to VoigtFit for convolution with the model components when matching the data. Our detailed fitting procedure is described in the following six steps.

First, all spectra were binned by a factor of two, resulting in rebinned pixels of 4 km s\(^{-1}\) for FUSE, 5 km s\(^{-1}\) for COS, and 4 km s\(^{-1}\) for STIS.

Second, we searched for absorption components in the interval \(v_{\text{LSR}} \lesssim 500\) km s\(^{-1}\). This interval was chosen after inspection of all the spectra, since all components were found to lie within this window. It includes both strong absorption from the Milky Way ISM as well as strong HI airglow emission down to approximately Ly\(\alpha\) 937.803. These airglow regions are manually masked out before fitting. We endeavored to keep additional

---

**Figure 1.** All-sky map in Galactic coordinates showing the locations of the survey sight lines (red circles), with the four sight lines shown as examples in Figure 2 highlighted as yellow stars. Black circles indicate the locations of rejected sight-line candidates. The map is centered on the Galactic anti-center and includes the high-velocity 21 cm data from Westmeier (2018) color coded by HI column density. Note that our FUSE sample is split across the northern and southern hemispheres, and across the east and west.
masking to a minimum, although it is occasionally necessary due to intervening higher redshift absorption systems.

Third, we fitted continua around each line of interest using low-order Legendre polynomials, and applied an H$_2$ model from the FUSE data based on results from Wakker (2006). This H$_2$ correction sometimes results in “ghost-flux”, a spurious signal artificially pulled out from a saturated profile, or a simple over-correction resulting in too much flux being added back to a profile. We mask these regions only if they appear to affect the quality or convergence of a fit.

Fourth, we isolated and fitted the OI absorption seen at $\lambda$1302, 1039, 988, 971, 950, 948, 936, 929, 924, and SII absorption at $\lambda$1250, 1253, and 1259 using VoigtFit, to use as an initial-guess model for the H1 component structure. This required a combination of FUSE data and HST/STIS or COS data covering the OI $\lambda$1302 and SII $\lambda$1250, 1253, 1259 lines. The primary challenge for fitting the H1 profiles is the velocity structure, as it can be nearly impossible to unambiguously recover multiple absorbers overlapping in velocity space. The low-velocity, ISM region ($v_{\text{LSR}} < 40$ km s$^{-1}$) is the most saturated and suffers the most. However, OI and SII are abundant and not highly depleted onto dust grains, and thus are excellent tracers of H1. OI is a particularly strong tracer of H1 because the two are closely coupled by resonant charge exchange (Field & Steigman 1971). We thus use these OI and SII fits to inform the initial guesses for the H1 component structure. These metal lines complement each other effectively, as the more sensitive OI transitions can help reveal weak H1 lines, and the weaker and narrower SII lines provide clues to the velocity structure of closely blended components at low velocities. The combination of OI and SII can trace H1 down to roughly log $N$(H1) $\sim$ 17 at 20% solar metallicity. Additionally, all our targets have been observed in 21 cm emission, so where necessary we consulted these emission profiles for help untangling the H1 component structure.

Fifth, the actual fits were then performed simultaneously to Ly$\alpha$–Ly$\mu$. The OI line fits at $\lambda$971, 950, 948, 936, 929, and 924 are included during the H1 fitting but are held fixed (i.e., they are no longer allowed to vary while H1 is fit). We do this so that VoigtFit knows about and fits around the OI, instead of masking out all these regions.

Sixth, the final H1 fits are obtained by iteratively adding, removing, shifting, and fixing H1 component guesses until the fit both converges and visually matches the data. Generally we strive for the minimum number of components to adequately fit the data, while keeping in mind which components are most likely present based on OI, SII, and H1 21 cm data. When possible, we also reference other metal transitions such as SiII, SiIII, and CII for additional hints to component structure. At this point, we also check for intervening deuterium absorption associated with the strong, low-velocity components. Deuterium absorption can masquerade as weak IVC absorbers, but we find little evidence of such contamination. On occasion, an OI, SII, or H1 21 cm feature appears to be two blended components that are very close in velocity. These features are very difficult to reproduce in the Lyman series fits, and often are instead fit with a single component. This can have the effect of biasing our resulting H1 distribution toward higher columns, particularly in the log $N$(H1) = 18–19 column density range. However, as this affects less than 20% of our sight lines, we expect this bias to be minor.

Occasionally we found it necessary to fix a particular H1 component’s velocity, line width, and/or column density for a fit to converge. In these cases, we start by fixing only the component velocity and/or line width, which is informed by visually inspecting the spectra alongside OI, SII, and H1 21 cm results. In only one case (PG1211+143) did we find it necessary to fix a component column density, using the value from 21 cm observations, and we do not include this component in any subsequent analyses. The components with fixed parameters are overwhelmingly in the low-velocity ISM regime, where high-$N$(H1) components are preferentially found, and where the component velocity, line width, and column density are otherwise well constrained by OI, SII, and H1 21 cm values.

Table 4 presents the complete list of fit components with errors and notes indicating which components were fixed. We make a final quality cut for all subsequent analysis by including only components with total column density fit errors less than 3 dex, which removes three particularly uncertain components.

Figure 2 showcases example fits for four sight lines ranging in signal-to-noise along with an all-sky plot with the positions of all of our targets, which we discuss in more detail in Section 2.5. Detailed figures showing our fits for each sight line are presented in Appendix B.

It is important to consider whether any bias is introduced by beginning with OI and SII component structure and applying this to the H1. Since hydrogen is the dominant baryon in the Universe, there must be H1 coexisting with the observed OI and SII, and our fitting procedure will, by design, reveal the spatial portion of the H1, but in principle there could be low-metallicity H1 clouds without corresponding OI and SII, and any such clouds would be missed by our metal-based fitting procedure. However, we have addressed this issue by the iterative process of adding and removing H1 components until the data are matched with the minimum number of components as described above. The finding that there are no strong H1 components without OI counterparts suggests that we are not missing a population of extremely metal-poor components, and therefore that this bias is small. In a future paper, we will present the associated OI and SII absorber population along with an analysis of the resulting metallicities.

2.5. Determining Component Structure: Examples

In this section, we examine the component structure and fitting choices in detail for the four example sight lines shown in Figure 2. The purpose of this exercise is to illustrate the effectiveness of our fitting methodology and to show how information can be extracted on the H1 properties even from data sets that suffer from saturation and low S/N. The four sight lines cover a range of S/N and sky location. In each direction, Figure 2 shows six Lyman series transitions and two metal-line transitions, and the sight lines are ordered left to right from high to low S/N. The colored tick marks show the locations of individual components and correspond to the matching-color Voigt-profile fits. The colors of the components from left to right in velocity additionally match between metal and H1 plots, further indicating how components match between metal and H1 data. The final combined fit profile is shown in red for the H1 data. The details of the component structure are now discussed for each of these four sight lines in turn.

Many of the individual FUSE spectra presented here have been published in earlier work on the Galactic halo.
2.5.1. MRK 817

The FUSE spectrum of MRK 817, a high-quality data set with (S/N)$_{977}$ = 18.5, shows strong H I absorption at both positive and negative velocities. This sight line passes through HVC Complex C (Collins et al. 2003). All five of the H I components we fit in the FUSE data are also seen in O I and S II, albeit somewhat shifted in velocity. In Figure 2, we show O I $\lambda$1302 and S II $\lambda$1253 as metal-line examples. The weaker S II $\lambda$1253 transition clearly shows two components near zero velocity (fit in green and pink). The stronger O I transition clearly shows a higher negative velocity component near $-100$ km s$^{-1}$ (orange), and suggests additional weaker components on the highest negative and positive velocity edges of the absorption profile (blue and lime green).

2.5.2. HE 0226-4110

The FUSE spectrum of HE 0226-4110, with (S/N)$_{977}$ = 11.2, illustrates an example of low-metallicity gas detected in the Lyman series data (see Fox et al. 2005). The S II $\lambda$1253 transition clearly shows one strong, near zero-velocity component, and the O I $\lambda$1302 transition suggests two very weak additional components around $+80$ and $+150$ km s$^{-1}$ (shown in orange and green). All three of these components are clearly visible in the H I. However, two additional higher velocity components (pink and lime green) are clearly seen in the H I spectra. As these components both have column densities of log $N$(H I)$\leq17$, their H I properties would be unknown if not for

Figure 2. Top: Four example FUSE spectra from our survey spanning a range in sky location and signal-to-noise. Each panel shows the normalized flux vs. LSR velocity for a different Lyman series absorption line, with the best-fit Voigt model shown in red, and each individual component plotted in a different color. Bottom: Metal-line absorption fits toward the same four sight lines using COS, STIS, and FUSE data. The matching colored tick marks and Voigt profiles illustrate the component structure of these unsaturated H I tracers, which help guide the placement of H I components in the FUSE Lyman series fits.
the FUSE data. We note that the orange component appears to be poorly fit in Lyα (H I, 12), but this is due to edge effects and model H₂ over-correction artificially pulling up flux. This component is well fit throughout the rest of the Lyman series.

2.5.3. PG 1011-040

The FUSE spectrum of PG 1011-040, with a moderate (S/N)_{Hβ} = 8.9, is an example of a highly saturated profile with a number of absorbers at positive velocities. Nevertheless, the constraints imposed by the combination of H I, O I, and S II data illustrate how a saturated, moderate S/N FUSE spectrum is still useful and can yield reliable H I fits. Here the O I λ1302 and S II λ1253 suggest two overlapping ~0 km s⁻¹ components, one broad (shown in blue) and one narrow (orange). The H I is also best fit with both components as a single, higher column density component would lead to overly strong wings in Lyα and Lyβ. Three additional components are seen in O I and S II around +100 km s⁻¹, which also clearly appear in the H I. From the H I spectra, we see that there are additional, low-metallicity components needed. We start by adding a component at ~325 km s⁻¹ (brown). However, there must be an additional component near ~200 km s⁻¹ (yellow) as this region of the spectra remains saturated. Finally, the edges of the spectra between these yellow and brown components suggest the presence of an additional weak component (gray).

Hence, while the FUSE data alone looks too saturated to achieve a unique solution, the constraints imposed by the metal lines leave little room for degeneracy even when there is metal-poor gas present. We also note that there is a flux-calibration issue with this COS sight line, such that the saturated core of the O I λ1302 still shows substantial flux. The orange fit line, which diverges below the data here, is well constrained by other O I transitions.

2.5.4. 1H0707-495

The FUSE spectrum of 1H0707-495 is our lowest S/N example (with (S/N)_{Hβ} = 6.4), and does not have complementary COS or STIS data. However, O I absorption seen in the FUSE data alone is still useful for determining the component structure. O I λ1039 shows a strong ~0 km s⁻¹ component, and good evidence for three higher velocity components. Other O I transitions, such as the O I λ308 shown, provide additional constraints on the strength and velocity of each component. These four components appear to nicely match the Lyman series profiles, even though the low S/N makes fitting difficult.

3. Results

The intention of this paper is to provide an empirical survey of the H I absorption in the Galactic halo. We therefore focus our results on the following observational quantities: the basic properties of the H I sample (sample size, kinematics, and column densities; Section 3.1), the H I errors (Section 3.2), the dependence of H I properties on sky location (Section 3.3), the H I sky-covering fraction (Section 3.4), the H I CDDF (Section 3.5), and the relative prevalence of inflow and outflow (Section 3.6).

3.1. Basic H I Properties

From our sample of 25 sight lines, we arrive at a final sample of 152 individual H I components, or an average of six components per target. The sample is composed of 39 ISM (|v_{LSR}| < 40 km s⁻¹), 42 IVC (40 ≤ |v_{LSR}| < 100 km s⁻¹), and 71 HVC (|v_{LSR}| ≥ 100 km s⁻¹) components. All sight lines have at least one H VC, and 24/25 have at least one IVC. Taken together, there are 113 HVCs and IVCs in the sample. 64 of these 113 (58%) have log N(H I) < 17.5 and are not detectable in 21 cm emission, showing that the majority of the HVCs and IVCs in our sample are invisible to 21 cm observations.

Figure 3 shows histograms of the component column densities for the full sample and for three bins of velocity: ISM, IVC, and HVC. These histograms show the basic distributions from which all subsequent H I quantities are derived. The ISM components tend to show high column densities of log N(H I) ≥ 19 (as expected, since the strong H I components in the Galactic disk are already well characterized by 21 cm observations), whereas IVCs and HVCs show much lower values, each peaking near log N(H I) = 16.5. The HVC distribution also shows a low column density tail extending down to log N(H I) ≈ 14, which is not seen in the IVC distribution. We also visualize the combined, normalized data set in Figure 4. This figure includes a Kernel Density Estimation (KDE) of the full sample of column densities (shown by the gray line). The KDE is an estimate of the probability density function (PDF) of the sample, and in practice produces a smoothed, nonparametric, functional form for the column density distribution of the Galactic halo.

Figure 5 shows histograms of the component b values (Doppler line widths), with the same figure format as Figure 3. The observed H I b values range from 0 to 50 km s⁻¹ for each of the ISM, IVC, and HVC populations. In each case, the distribution peaks at b ≈ 10 km s⁻¹ and then shows a long tail.
extending out to ≈50 km s$^{-1}$. Note that the FUSE instrumental velocity resolution (FWHM = 20 km s$^{-1}$) corresponds to a limiting $b$ value of 12 km s$^{-1}$, so a large fraction of the components are unresolved. We do not show components with $b > 50$ km s$^{-1}$ (two such cases were found, both ISM components), since such broad lines may be unresolved blends of narrower components, especially given the low S/N of the data.

3.2. The Saturation Problem

The difficulty of accurately measuring saturated lines from the Lyman series results in large column density errors for many H$_I$ components. To illustrate this, Figure 6 shows the errors on log $N$ (H$_I$) as a function of log $N$(H$_I$) for all components in the full sample. This error is a quadrature sum of the statistical and systematic errors listed in Table 4. The errors are largest in the log $N$(H$_I$) ∼ 17–19 region where the lines are saturated but the wings are not yet strong enough to fit. The errors reach ∼1–3 dex in this region. In Figure 6, we also include a red-shaded horizontal bar illustrating the mean error as a function of column density. We include this bar in subsequent figures as a visual indicator of the regions most affected by saturation.

For some of these components, the statistical fitting errors may be artificially inflated and/or unreliable due to numerical errors in the fitting routine. This can happen because the code cannot resolve degeneracy between components, even though we can be confident of component location and approximate strength from the combination of other constraints (e.g., H$_I$ 21 cm and UV metal-line absorption). We highlight these highly uncertain components in Table 4 and plot them in gray in Figure 6.

Below log $N$(H$_I$) ∼ 17, the errors become considerably smaller as the higher-order Lyman lines become unsaturated. While a few uncertain components remain in this region, the vast majority have errors $\lesssim$ 0.5 dex. Above log $N$(H$_I$) ∼ 19, the errors become smaller as the damping wings of Ly$\alpha$ allow precise measurements. We find that 45 (out of 152) components have an error >0.5 dex and 24 have an error >1 dex, illustrating the challenging (but not insurmountable) nature of the measurements. These large errors provide confidence that the VoigtFit code is doing a reasonable job of identifying uncertain components.

3.3. Spatial Distribution

Next we consider the dependence of H$_I$ component properties on Galactic longitude $l$ and latitude $b$, to look for spatial variation in the neutral gas. Figure 7 shows the H$_I$ column densities against $l$ and $b$, with ISM, IVC, and HVC components colored separately. The H$_I$ columns are found to be fairly evenly distributed in longitude; this is true both in general and for each velocity category (ISM, IVC, and HVC), with the range of observed columns staying fairly constant across longitude. The range in columns observed is also fairly constant in latitude; however, we observe more IVCs and HVCs in the north than the south. This may simply reflect that there are more northern sight lines in the survey than southern (our sample includes 14 N and 11 S), and many of those northern sight lines pierce large well-known H$_I$ HVC complexes like Complex C and Complex A. To statistically check for differences, we performed Anderson-Darling two-sample distribution tests for north versus south latitudes and longitudes above and below 180°, which found no significant deviations between these split samples. The gap at $|b| < 20^\circ$ is due to a lack of background AGN in this region, which is a consequence of dust extinction in the Galactic plane.

Figure 8 shows the component velocity centroids in the Galactic standard of rest (GSR) reference frame against Galactic longitude (top) and latitude (bottom). Since the local standard of rest (LSR) frame in which the data are fit is a rotating frame, the transformation to GSR attempts to remove the effect of Galactic rotation and thus reveal a more valid distribution of cloud velocities in the Galactic rest frame. In this figure, we have normalized the size and color of each point to reflect the H$_I$ column density of each component, allowing a visual assessment of where the components are located in velocity-N(H$_I$) space.
The height of each column of points indicates the velocity interval over which components are detected, and therefore provides a measure of the kinematic complexity of each sight line.

In the top panel, we also plot dashed red lines to indicate where ISM gas at LSR velocities of $|v_{\text{LSR}}| \leq 40$ km s$^{-1}$ resides in GSR space. The highest column density components, as indicated by larger point size and yellow color, mostly fall within this velocity range. We again performed Anderson-Darling two-sample distribution tests to check for differences between components at longitudes of $0 \leq l < 180^\circ$ and $180 \leq l < 360^\circ$. The resulting $p$ value ($p = 0.002$) suggests that there is a difference, with preferentially positive velocity components at $0 \leq l < 180^\circ$ and negative at $180 \leq l < 360^\circ$. 

Figure 6. The total HI column density error (summed in quadrature) as a function of HI column density for all components included in the final sample. The three components with error $> 3$ dex are not shown, but have log $N$(HI) = 18.8, 18.4, and 18.4. The red-shaded horizontal bar illustrates the mean error as a function of column density, where darker red implies larger errors. On the right, we include a marginal histogram showing the distribution of errors for the ISM (orange), IVC (purple), HVC (green), and combined sample (black outline). The quartiles of the distribution are indicated by the black solid and dashed lines and matching labels.

Figure 7. HI column densities as a function of Galactic longitude (top) and latitude (bottom) for HVCs (green pluses), IVCs (purple crosses), and ISM components (orange rhombuses). Each column of points shows the components from a given sight line. No components are seen at $|b| < 20^\circ$ due to a lack of observable background AGN in this region.
However, this is exactly what we would expect if the conversion of velocities from LSR-to-GSR was imperfect (i.e., notice how the points appear to follow the dashed red lines). As we know the LSR-to-GSR conversion is indeed imperfect, we do not assign much significance to this finding.

In the bottom panel, we also observe some structure with latitude. In particular, we notice an absence of high positive velocity (v > 100 km s\(^{-1}\)) absorbers at high negative latitude (the top-left corner of the plot), whereas high positive velocity components are found at high positive latitudes (the top-right corner). Both distributions show more negative velocity components than positive, and this is particularly true in the southern sky, or a relative excess of inflow gas in the southern sky, or a relative excess of inflow. An Anderson-Darling test produces a \(p\) value of \(p = 0.01\), suggesting a mildly significant difference between the high- and low-latitude distributions.

### 3.4. Sky-covering Fraction

One fundamental HI quantity we can calculate is the sky-covering fraction \(f_{\text{sky}}(N)\), which is the fraction of sight lines containing at least one component of a given column density. We can calculate this in two primary ways. First, we derive \(f_{\text{sky}}\) for discrete column density intervals \(\log N + \Delta \log N\) in Figure 9, where \(\Delta \log N = 0.5\), chosen because this gives bin sizes of 1 dex. This plot is similar to the raw distribution of components in Figure 3, but with the key difference that a sight line may contain multiple components in any given \(\log N\) interval, whereas the sky-covering fractions only count one component per bin even if multiple are present. In the full sample, we observe a high value \(f_{\text{sky}} > 0.5\) down to \(10^{16}\) cm\(^{-2}\), with a peak in the number of components in the \(\log N = 16.5 - 17.5\) column density bin. This peak is observed in both the IVC and HVC subsamples, although we note that the measurement errors are generally high around these column densities.

Second, we calculate the cumulative sky-covering fraction as a function of column density, \(f_{\text{sky}}(> N)\), in Figure 10. This shows the percentage of sight lines in which we observe an absorber of column density \(N\) or higher. We can repeat the exercise separately for the northern and southern absorbers, as shown in the lower two panels in the figure. This analysis shows that the sky is 100% covered with ISM HI above \(\log N = 16.5\). IVC sky coverage only increases to 80% sky coverage at \(\log N = 16.5\) in both the northern and southern hemispheres, with higher sky-covering fractions for IVCs in the north than in the south. To quantify this N-S asymmetry, we determine the HI column at which a certain \(f_{\text{sky}}\) is reached. For IVCs, we find that \(f_{\text{sky}} = 0.8\) (80% sky coverage) is reached at \(\log N = 17.0\) in the north and 16.5 in the south, showing that one must go 0.5 dex lower in \(\log N\) in the south to reach this sky-covering fraction. This excess of northern IVCs has already been reported in HI and O VI studies (Savage et al. 2003). For HVCs, we find that \(f_{\text{sky}} = 0.8\) is reached at \(\log N = 16.5\) in both the north and in the south, so there is less asymmetry for HVCs than
for IVCs in this low-column regime; however, there is a tail of high-$N$(H I) HVC gas extending to log $N$(H I) = 20 in the north that is not present in the south. This is not surprising, as there are several large, well-known HVC complexes in the north, such as Complex A and Complex C.

The fact that the cumulative $f_{\text{sky}}$ of our HVC sample reaches unity indicates that we have probed the HVC population in far more depth than past studies using 21 cm emission data. However, we can make a more fair comparison by placing a lower column density limit to replicate the sensitivity of relevant 21 cm results. Wakker (1991) find an HVC covering fraction of 18% down to a detection limit of log $N$(H I) = 18.3, compared to our 40% result at this column density. The more sensitive surveys of Murphy et al. (1995) and Lockman et al. (2002) probe down to log $N$(H I) $\sim$ 17.8 and both find a 37% covering fraction, compared to our 48% result at this column density limit. Our slightly higher covering fractions may be due to the locations of our sight lines not being entirely random, but with a slight bias toward known structures (e.g., five of our sight lines probe Complex C). Incompleteness in the 21 cm results may also play a role (Wakker & van Woerden 1997).

### 3.5. The Column Density Distribution Function

#### 3.5.1. Definition of CDDF

The H I column density distribution function (CDDF) is defined (Tytler 1987) as

$$\text{CDDF} = f(N) = \frac{d^2 n}{dN(H I) dX},$$

where $n$ is the number of H I absorbers per column density interval $dN(H I)$ per unit path length $dX$, where

$$dX = dv/c,$$

with $dv$ equal to the velocity path length over which we searched for absorption. This is derived from the cosmological description of co-moving path length used by higher redshift surveys and defined as $dX = \frac{H_0}{H(z)}(1 + z)^2 dz$, which at $z = 0$ reduces to Equation (2).

The CDDF is a fundamental description of the distribution of a sample of QSO absorbers, just as the luminosity function is a fundamental description of a sample of galaxies. It encodes information on which absorbers host most of the H I mass.

The first two moments of the CDDF contain important physical information. The zeroth moment represents the incidence of absorbers per absorption distance $dn/dX$ [sometimes called $l(X)$ in the literature] between two column density limits $N_1$ and $N_2$:

$$dn/dX = \int_{N_1}^{N_2} f(N)dN.$$  

The first moment of the CDDF integrated over path length gives the H I mass density between the two limits $N_1$ and $N_2$:

$$\rho(H I) = m_{H I}(H_0/c) \int_{N_1}^{N_2} N f(N) dN.$$  

For a power-law slope $\beta$ where $f(N) = C N^{-\beta}$, a critical slope of $-2$ exists, above which the H I mass budget will be...
dominated by high column density systems, so for slopes shallower than $-2$ the mass will be dominated by the choice of $N_2$ in the integral. This is because $\rho(H I) \propto \int N N^\beta dN$ so if $\beta = -2$, then $\rho(H I) \propto \int N^{-1}dN = \ln N$, which diverges (Prochaska & Wolfe 2009; Danforth et al. 2016).

3.5.2. Construction of the CDDF

We construct the Galactic CDDF by automatically binning the H I column density distribution using the maximum of the FreedmanDiaconis Estimator or Sturges methods, which takes into account both data variability and size. The number of absorbers in each bin, $n$, is then divided by the column density size of each bin, $\Delta N(H I)$, multiplied by the path length searched for absorption, $dX$.

The path length $dX$ effectively sets the vertical normalization of the CDDF, but also encodes key information concerning the incompleteness of our sample. Each absorption component obscures a section of the spectrum and thus reduces the path length over which additional components can be detected. To account for this incompleteness, we have simulated the observability of additional components in each sight line.

Our observable-path-length simulation works by first recreating the absorption profile for each target, then adding an additional component to the profile from a grid of column densities and velocities spanning the entire observable range, i.e., column densities from $13 \leq \log N(H I) \leq 21$, and velocities from $-500 \leq v_{LSR} \leq 500$ km s$^{-1}$. This modified profile is then subtracted from the original to yield a residual $R$, which represents the impact of the additional component on the line profile. If the maximum value of the residual is not significant at $3\sigma$, i.e., if $R_{\max} < 3/(S/N)$, then we would not be able to detect that component if it were present, because it would be lost in the noise, so we declare such a component unobservable. For example, in a spectrum with $S/N = 20$, we can detect components that create residuals at the 15% level, whereas if $S/N = 10$, we can only detect residuals at the 30% level, and if $S/N = 5$, we are only sensitive to 60% residuals. The output of this simulation is a function for each Lyman series transition that gives the observable path length $dX$ as a function of column density, taking into account the noise properties of each individual spectrum. Figure 11 shows the resulting path length $dX$ summed for the entire sample. The path length used in our CDDF calculations is the upper envelope drawn from this curve, i.e., the Lyman line that gives the maximum path length at a given $N(H I)$. This method ensures that each CDDF bin is individually normalized vertically to account for the incompleteness due to velocity shielding within each column density interval.

Finally, we calculate errors in each CDDF bin via a bootstrap resampling method. This method randomly resamples and rebins the entire distribution 10,000 times, where each component is itself drawn from the normal distribution defined by its VoigtFit measurement errors (a quadrature sum of the statistical and systematic fitting errors shown in Column 7 of Table 4). The final error for each bin is then the 16th and 84th percentile ($\pm 1\sigma$) of the resampled distribution. This method thus takes into account both measurement and sampling uncertainties.

3.5.3. Fitting the CDDF

We perform power-law fits to the CDDF using the affine-invariant Markov Chain Monte Carlo (MCMC) Ensemble sampler Emscee (Foreman-Mackey et al. 2013). Emscee implements an MCMC sampling of likelihoods across the parameter space using an ensemble of walkers, each of which evolves for a number of steps to sample the posterior distribution. Our best fits are then reported as the median, 16th, and 84th percentiles ($1\sigma$) of the posterior distributions. Because the $17 \leq \log N(H I) \leq 19$ region is affected by saturation, we sum all components in this region into a single, larger bin prior to fitting the CDDF to help minimize the effect of highly uncertain components. All of the final fits are performed with this larger bin for the saturated absorbers, although we also display the original, finer-binned data points in Figures 12 and 13 for completeness.

We also investigate the possibility of breaks in the full CDDF as a function of column density only. Figure 13 presents the CDDF fit with a three-component, piecewise, power-law model. The fitting process is again carried out using Emscee, but with six free parameters representing the $y$-intercept, slopes for the three segments, and the locations of the two break points. We perform this fit both for our complete sample of absorbers (“Galactic Halo + Disk”) as well as a “Galactic Halo-only” version, which excludes the ISM components (i.e., $|v_{LSR}| < 40$ km s$^{-1}$). We find a best fit with breaks at $\log N(H I) = 16.6$ (16.6) and 17.8 (17.7) for the Galactic halo and disk (Galactic halo-only). The results of these MCMC fits can be found in Appendix B (Figure 17) displayed as corner plots. The advantage of this piecewise method is that the break points are identified automatically in a statistically rigorous manner, rather than being inserted manually.

3.5.4. Displaying the CDDF

We display the H I CDDF in several formats to highlight different features of the distribution. First, in the left panel of Figure 12, we show the global H I CDDF (over all velocities) fit with a single power-law model. Second, in the right panel of Figure 12, we show a version split by velocity into HVC, IVC, and ISM bins, each fit with a separate power-law model. Third, in Figure 13, we show the H I CDDF together with the three-component piecewise power-law model described in Section 3.5.3, split into Galactic halo and disk (all velocities) and Galactic halo-only (v$_{LSR} \geq 40$ km s$^{-1}$) versions. The key
difference of Figure 13 is that it breaks the sample down by HI column density, rather than by velocity. In addition, Figure 13 includes a comparison of the Galactic CDDF to a selection of extragalactic CDDFs from the literature, including results from O’Meara et al. (2013), Rudie et al. (2013), Danforth et al. (2016), and Shull et al. (2017), and simulation results from Rahmati et al. (2015).

We emphasize that the extragalactic CDDFs are derived from blind surveys, whereas the Galactic CDDF is derived from our vantage point inside the Milky Way, which is not a random position. That is why the Galactic CDDF has a much higher normalization than the extragalactic CDDF; we are not in an unbiased location. Of more physical interest is the difference in slope. The slopes we measure from the various forms of the HI CDDF are summarized in Table 2. A discussion of these slopes is given in Section 3.5.5. Each figure also includes a red-shaded bar across the top edge, which indicates the mean component fit error as a function of column density (derived from Figure 6). This gives a visual indication of which regions are most affected by saturation and blending issues.

### 3.5.5. Slope of the CDDF

In Figure 12 (left), we find that a single power-law model with a slope of $\beta = -0.94^{\pm}0.06$ provides a reasonable fit to the full CDDF over six orders of magnitude from $\log N = 14-20$. The fit is not perfect, however, as several data points lie at or beyond $1\sigma$ from the line of best fit, particularly near $\log N(\text{HI}) \approx 16.5$ where there is a marked excess of IVCs and HVCs.

Next, in Figure 12 (right), we individually fit the ISM, IVC, and HVC subsamples each with a single power-law model. We find that IVCs and HVCs have nearly indistinguishable CDDF slopes, with $\beta_{\text{IVC}} = -1.01^{+0.15}_{-0.14}$ and $\beta_{\text{HVC}} = -1.05^{+0.07}_{-0.06}$, in line with the basic similarities we noted from their number counts seen in Figure 3. Conversely, the ISM components (which are heavily biased toward higher HI column densities) show a shallower slope, $\beta_{\text{ISM}} = -0.68^{+0.17}_{-0.20}$. This shallower slope is largely due to a lack of ISM components in the $17 \lesssim \log N(\text{HI}) \lesssim 19$ region, where saturation and velocity blending make it difficult to detect components.

These differing CDDF slopes may indicate that the ISM HI shows a different behavior than ISV and HVC HI, which perhaps reflects the different physical conditions in the disk and halo. However, the ISM fit in particular does not appear to capture the detailed structure of the CDDF data points. While the high-column end of the ISM is quite shallow, there appears to be a steepening below $\log N(\text{HI}) \approx 18$, evidenced by the last ISM point more closely matching the IVCs and HVCs. Hence, while a difference likely exists between the high- and low-velocity regimes, the overall shape of the CDDF may be mostly a function of column density.

### Table 2

Summary of CDDF Slopes

| Sample | $\beta$ |
|--------|---------|
| All HI | $-0.94^{+0.06}_{-0.06}$ |
| ISM ($|v_{LSR}| < 40$ km s$^{-1}$) | $-0.68^{+0.20}_{-0.17}$ |
| IVC ($40 \leq |v_{LSR}| < 100$ km s$^{-1}$) | $-1.01^{+0.15}_{-0.14}$ |
| HVC ($|v_{LSR}| \geq 100$ km s$^{-1}$) | $-1.05^{+0.07}_{-0.06}$ |
| Galactic Halo | | |
| $\log N(\text{HI}) \gtrsim 17.7$ | $-0.99^{+0.11}_{-0.11}$ |
| $16.5 \lesssim \log N(\text{HI}) \lesssim 17.7$ | $-1.71^{+0.43}_{-0.33}$ |
| $\log N(\text{HI}) \lesssim 16.6$ | $-0.64^{+0.11}_{-0.11}$ |
| Galactic Halo + Disk | | |
| $\log N(\text{HI}) \gtrsim 17.8$ | $-0.72^{+0.10}_{-0.07}$ |
| $16.5 \lesssim \log N(\text{HI}) \lesssim 17.8$ | $-1.78^{+0.42}_{-0.29}$ |
| $\log N(\text{HI}) \lesssim 16.6$ | $-0.62^{+0.11}_{-0.10}$ |

Note. This table reports the best-fit power-law slopes of HI CDDF as shown in Figure 12 and 13, where $f(N) = CN^{-\beta}$.
An alternative model for the full CDFD is a three-component piecewise power-law model as shown in Figure 13, in which the CDDF is fit over three separate column density regimes. We include two versions of this fit. First, for the complete survey including absorption at all velocities (Figure 13 left; “Galactic Halo + Disk”), and second for just $v_{LSR} \geq 40$ km s$^{-1}$ velocity absorption (Figure 13 right; “Galactic Halo”). As mentioned earlier, our unique location within the Galaxy means that our sight lines include gas from the Galactic disk as well as the halo. Extragalactic CDDF results, however, mostly consist of halo or IGM material far from galaxies, so our “Galactic halo” CDFD may provide a more fair comparison to these studies.

It is worth noting that the slope of the Galactic halo and disk CDFD is similar at the low end [$\beta_{\text{low}} = -0.62^{+0.11}_{-0.10}$ for log $N$(H I) $\leq 16.6$] and high end [$\beta_{\text{high}} = -0.72^{+0.10}_{-0.07}$ for log $N$(H I) $\geq 17.8$], on either side of the saturation zone. So if we were to interpolate between the low and high ends, we would recover a slope similar to what is actually observed in the intermediate region where saturation is important. This indicates that, despite the large uncertainties on the individual saturated components, their effect on the CDFD is minor, because we would have derived a similar global CDFD slope even if we had excluded them from the sample entirely.

To further test this point, we performed all of the CDFD fits again after excluding all components with errors larger than 1 dex. Doing so results in almost the same CDDF slopes as before: $\beta_{\text{low}} \leq 1$ dex $= -0.96^{+0.09}_{-0.07}$, $\beta_{\text{mid}} \leq 1$ dex $= -0.66^{+0.11}_{-0.07}$, $\beta_{\text{high}} \leq 1$ dex $= -0.66^{+0.08}_{-0.07}$. This occurs because the highly uncertain components are much less common than the well-constrained components and are spread fairly evenly across a wide range of H I column density. The finding that the highly uncertain data points in our sample do not substantially change the derived slopes of the CDFD indicates that our results are robust against the presence of a minority of uncertain components. When excluding the low-velocity ISM components for our Galactic halo-only version, the high-column end rises to $\beta_{\text{high}} \sim -1$, as expected from the individual IVC and HVC power-law fit results.

3.5.6. Discussion of the CDFD

Our survey has extended our knowledge of the Galactic H I distribution by four orders of magnitude, pushing down from the limits of 21 cm-based H I (which reach log $N$(H I) $\approx 18$, Wakker 1991; Murphy et al. 1995; Lockman et al. 2002; Moss et al. 2013; Westmeier 2018) down to our UV limit of log $N$(H I) $\approx 14$.

Our measured near-unity slope for the full-sample CDFD ($\beta = -0.94^{+0.06}_{-0.08}$) is much shallower than the critical slope of $-2$, telling us that the H I mass is dominated by high $N$(H I) systems. This measured slope is also significantly shallower than extragalactic results from blind surveys, which tend to find CDFD slopes with $\beta = 1.4 - 1.7$ (Kim et al. 2002; Lehner et al. 2007; Prochaska et al. 2010, 2014; Ribaudo et al. 2011; Tilton et al. 2012; Kim et al. 2013; O’Meara et al. 2013; Rudie et al. 2013; Danforth et al. 2016; Shull et al. 2017). Hence, the distribution of H I in the Galactic halo appears to be tipped toward higher H I columns. This result may not be unique to our own Galaxy; although most extragalactic CDFD measurements are composed of absorbers in the IGM and extended CGM, far from galaxies, Rudie et al. (2013) constructed CDFD from H I both “near” and “far” from galaxies, and found shallower power-law slopes for absorbers within 300 proper kpc and 300 km s$^{-1}$ of galaxies. This supports the idea that galaxy environments lessen the slope of the CDFD and host different absorber populations than the IGM, a result also supported by metal-line analyses (Richter et al. 2009).

Moving to our alternative piecewise power-law model, the high-column density end (log $N$(H I) $\geq 17.8$) has a slope $\beta = -0.72^{+0.07}_{-0.06}$ (or $\beta = -0.99^{+0.11}_{-0.11}$ for the Galactic halo only), which aligns well with extragalactic results from Ribaudo et al. (2011), who report $\beta = -0.8_{-0.3}^{+0.3}$ for their SLLS regime of...
19.1 \leq \log N(HI) \leq 20.2. However, this value is significantly shallower than results from Galactic 21 cm studies, such as those found by Putman et al. (2002), and Moss et al. (2013), who find $\beta \sim -2$. This could be the result of several competing effects, including beam-smearing, velocity resolution, and differences in definition between absorption lines and emission peaks. In a future paper, we will compare in detail our HI results to those from a variety of 21 cm observatories in an attempt to untangle these effects.

Our intermediate column density region, $16.6 \leq \log N(HI) \leq 17.8$, shows a significantly steepened slope of $\beta = -1.78^{+0.29}_{-0.42}$ ($\beta = -1.71^{+0.33}_{-0.43}$ for the Galactic halo only). A steepening below $\log N(HI) \leq 17.5$ has also been reported in blind surveys at $z > 1$ by Prochaska et al. (2010), Ribaudo et al. (2011), O'Meara et al. (2013), and Rudie et al. (2013). At lower redshift, Shull et al. (2017) found a slightly shallower slope of $\beta = -1.48^{+0.05}_{-0.05}$ for partial Lyman Limit Systems (pLLS), although Danforth et al. (2016) also found evidence of $\beta$ steepening with redshift between $z = 0.01$ and $0.47$.

The steepening in the Galactic CDFD may be related to ionization: the optical depth to ionization reaching $\sim 1$ at $\log N(HI) \approx 17.2$, so absorbers below this threshold are expected to be significantly ionized. However, this effect would be expected to lead to a flattening of the HI slope, not a steepening, so this does not explain the observed excess of clouds in this range. This excess is visible in the raw component distributions shown in Figure 3 as well, so it is not related to the way the CDFD is constructed. Recent theoretical work on the size scale of fragmenting, pressure-confined cool gas favors characteristic columns of $\log N(HI) = 17$ (McCourt et al. 2018) to $\log N(HI) = 18$ (Gronke & Oh 2018), so this is one potential explanation for the observed abundance of these absorbers and resulting upick in the CDFD slope. However, we caution that this intermediate-column density region also corresponds to the regime where saturation effects and fitting errors are the highest (as illustrated by the red shading in the mean $\log N(HI)$ error bar along the top edge of each CDFD figure). The transition to unsaturated absorption at the low end of this region could also cause some fitting artifacts, which could be contributing to the distribution peak. Thus, it is possible the break at $\log N(HI) \sim 18$ is exaggerated by fitting errors. That being said, the fact that a majority of other CDDF studies also see a slope change near this region (as seen in Figure 13) does provide reason to believe that a slope change is real. Indeed, we note that the overall shape of the CDFD from $16.6 \leq \log N(HI) \leq 17.7$ for the Galactic halo-only model matches well with the extragalactic CDFDs, as seen in Figure 13.

In the low-column density regime ($\log N(HI) \lesssim 17$), we find a significantly shallower slope ($\beta = -0.62^{+0.11}_{-0.10}$) than the rest of the Galactic CDFD and extragalactic CDFD results (e.g., Kim et al. 2002; Lehner et al. 2007; Tilton et al. 2012; Danforth et al. 2016). This flattened slope is likely due to a combination of incompleteness and physical differences between the Galactic halo and the Ly$\alpha$ forest environments probed by other UV surveys.

In this lower column density range, low-$z$ extragalactic absorbers are generally far from galaxies and reside in dark-matter filaments. They have lower densities and higher ionization parameters than Galactic halo gas (e.g., Collins et al. 2005; Shull et al. 2011). Conversely, the low column density absorbers in our sample are all IVCs and HVCs, which are likely relatively close to the Galaxy; Richter (2017) find $d \leq 2$ kpc for IVCs, and Lehner & Howk (2011) find $d \approx 12$ kpc for HVCs, although the lowest column density clouds are less well constrained. This difference in environment and location also results in an ionization effect, in which our Galactic H1 sample is exposed to a significantly enhanced radiation field compared to the extragalactic sample, because the Galactic ionizing radiation field dominates the extragalactic UV background within $\approx 50$ kpc (Giroux & Shull 1997; Bland-Hawthorn & Maloney 1999; Fox et al. 2005). This ionization effect means that hydrogen exists preferentially in the form of HII toward lower $N(HI)$, which would lead to a depression in the relative amount of low column density HI. Hence, it is not altogether unexpected for the low column density end of the Galactic CDDF to differ from that found in extragalactic surveys.

The other potential effect is incompleteness: the inability to detect some components because they are hidden in the data beneath our ability to detect them. This effect increases toward lower $N(HI)$, since weaker components are easier to hide. We have carefully accounted for the effect of reduced path length $dX$ for lower $N(HI)$ absorbers using our observable-path-length simulation (Section 3.5.1), so we have dealt with incompleteness as much as is possible, but we are still likely still incomplete in the low $N(HI)$ regime because the velocity density of absorbers (number per unit velocity) is not uniform across $dX$. We most readily detect weak ($\log N(HI) \lesssim 16.6$) absorbers at high velocities ($|v| \gtrsim 100$ km s$^{-1}$), where the contamination from ISM gas is lower, but conversely find that the density of absorption across all $N(HI)$ is highest at low velocities ($|v| < 100$ km s$^{-1}$).

If there was a high velocity-density of weak absorbers at low velocities, we would be insensitive to them, and thus they would be missing from our statistics. This bias has been estimated in recent simulations of the Galactic CGM (Zheng et al. 2020), but observational results from the QuaStar project on HST have found surprisingly little missed low-velocity gas (Bish et al. 2020). This discrepancy could be explained by invoking a flared disk or flattened halo model, in which the majority of gas is located at low latitudes along an extended disk (e.g., Qu et al. 2020). In this case, high-latitude sight lines, such as in our survey, would miss the most significant reservoir of Galactic halo gas. There is indeed some observational evidence supporting such flattened halos in extragalactic H1 studies (French & Wakker 2017) and several studies of rotating cold gas in the CGM (Ho et al. 2017; Martin et al. 2019; French & Wakker 2020).

An alternative approach to quantifying the incompleteness is to use upper limits on the CDFD slope to back out an absorber distribution. If we extend the intermediate $N_{HI}$ region of our fit out to $\log N(HI) = 14$, closely approximating the slopes found by O'Meara et al. (2013), Prochaska et al. (2014), Rahmati et al. (2015), and others, we calculate we would have to have missed over 800 low column density components, or $\approx 33$ absorbers per sight line, with a distribution peaking around $\log N(HI) = 14.5$. While we cannot completely rule out such a large missed distribution, we emphasize that they would be forced to live entirely in the $|v_{LSR}| \lesssim 100$ km s$^{-1}$ region. Beyond this velocity in the HVC range, only a handful of additional components could be hidden. Thus, while the overall low-column slope could be steeper than our best fit, the HVC-only CDFD results are relatively robust. Furthermore, the similarity between our HVC and IVC CDFDs indicates that the $40 \lesssim |v_{LSR}| \lesssim 100$ km s$^{-1}$ velocity range is not strongly affected by incompleteness either.
outmetallicity correction, the HI survey results presented here rely on advantage, because whereas that metal-based survey relied on a subsets, for all absorbers and out...n Haud 2006; Peake et al. 2008 the aggregate distribution in position, but we expect these imperfections to average out across distant clouds, so is more appropriate for HVCs than IVCs.

Figure 14. The H I sky-covering fraction split into inflowing and outflowing subsets, for all absorbers (black) and for each velocity bin (colored lines). The shaded regions around each line show the 1σ error regions derived from a bootstrap resampling routine.

3.6. Inflow versus Outflow

Our data set allows us to assess the relative prevalence of inflowing and outflowing gas. Following Fox et al. (2019), we approximate clouds with $v_{\text{GSR}} < 0$ as inflowing and those with $v_{\text{GSR}} > 0$ as outflowing, thereby splitting the observed cloud population into two. This may artificially identify some clouds as inflowing or outflowing depending on their distance and Galactic position, but we expect these imperfections to average out across the aggregate distribution (see, e.g., Wakker 1991; Kalberla & Haud 2006; Peek et al. 2008). This assumption is only valid for distant clouds, so is more appropriate for HVCs than IVCs.

Compared with Fox et al. (2019) we have a significant advantage, because whereas that metal-based survey relied on a metallicity correction, the H I survey results presented here rely on no such correction, since we observe the neutral baryons directly. We calculate the sky-covering fraction separately for inflowing and outflowing samples, as shown in Figure 14, finding that the outflowing gas covers less of the sky than inflowing gas, with a large jump in covering fraction below log $N$(H I) = 17. This excess of inflowing gas is seen independently in the IVC and HVC subsamples, as well as in the full sample.

We can then compare the mean H I column density of the inflowing and outflowing clouds, and look for an excess of one versus the other. We perform this exercise in Figure 15, which shows the H I column density of each component as a function of $v_{\text{GSR}}$, with histograms overplotted showing the mean log $N$(H I) in 50 km s$^{-1}$ bins for each of the ISM, IVC, and HVC categories.

In order to calculate the H I mass and mass flow rate represented by our absorbers, we adopt a partially covered spherical shell model following Fox et al. (2019). This calculation should be viewed as a back-of-an-envelope estimation rather than a precision calculation; nonetheless, the results still offer useful physical insight into the relative strength of inflow and outflow. In this model, the halo gas clouds are located in a shell at distance $d$ with a sky-covering fraction $f_{\text{sky}}$ and mean H I column $\langle N$(H I) $\rangle$. Then the total H I mass in inflowing and outflowing clouds can be written as

$$M_{\text{in}} = 1.4m_{\text{H}} f_{\text{sky, in}} \langle N$(H I) $\rangle 4\pi d^2$$ and

$$M_{\text{out}} = 1.4m_{\text{H}} f_{\text{sky, out}} \langle N$(H I) $\rangle 4\pi d^2,$$

where $m_{\text{H}}$ is the mass of the hydrogen atom, the factor 1.4 accounts for the mass in helium and metals, and the covering fractions are assumed to be independent of distance. We then combine $M_{\text{in}}$ and $M_{\text{out}}$ with the mean flow velocities ($v_{\text{in}}$) and ($v_{\text{out}}$) to derive the mass flow rates $dM/dt$ for the inflowing and outflowing cloud populations:

$$dM_{\text{in}}/dt = M_{\text{in}} \langle v_{\text{in}} \rangle /d$$ and

$$dM_{\text{out}}/dt = M_{\text{out}} \langle v_{\text{out}} \rangle /d.$$  

Each of the terms on the right-hand side of Equations (5) to (8) can be directly constrained by our observations, except the distance, which we take from literature values: We adopt a mean distance to HVCs of 12 ± 4 kpc from Lehner & Howk (2011) for both the inflowing and outflowing populations. We caution that this distance is highly uncertain, particularly for clouds with low H I column densities.

The calculated values for the masses and mass flow rates are presented in Table 3. They are calculated separately for inflowing and outflowing HVCs. Furthermore, for each we give the results for two bins of $N$(H I), [18–19] and [19–20], revealing how the mass and mass flow rate is distributed over $N$(H I). These two bins span the column density range containing the vast majority of the mass and mass flow rate, because they are both proportional to the mean $N$(H I) in the bin. Additionally, the [18–19] bin contains many poorly constrained saturated absorbers, while the [19–20] bin contains mostly well-measured components (as shown in Figure 6). By summing over these bins, we can calculate the total H I mass and total H I mass flow rate in each of the four subsamples.

We find that inflowing HVC components carry about two times as much H I mass as outflowing components, which results in a more than four times larger mass flow rate, with $\approx −0.88 ± 0.40 (d/12 \text{ kpc}) M_{\odot}\text{yr}^{-1}$ of HVC inflow versus $≈ 0.20 ± 0.10$.
The HI masses and mass flow rates are calculated using Equations (5) to (8) for HVCs in two $N(\text{H})$ intervals. Inflowing and outflowing clouds are defined as those with $V_{\text{LSR}} < 0$ and $V_{\text{LSR}} > 0$, respectively. A mean distance of 12 kpc is assumed for HVCs. Negative mass flow rates denote inflow.

The H I components cover approximately 7 dex of column density, though the range depends on velocity: ISM components cover a range of $\log N(\text{H})$ from $\approx18$–$21$ whereas HVCs cover $\approx14$–$19$ and IVCs cover $\approx16$–$20$. The differences in column density ranges appear to be mostly due to incompleteness from saturation and velocity crowding of components. These ranges of column density are observed at longitudes and latitudes across the sky with no strong spatial dependence.

2. Taken together, there are 113 HVCs and IVCs in the sample. 64 of these 113 (58%) have log $N(\text{H}) < 17.5$ and are not detectable in 21 cm emission, showing that the majority of the HVCs and IVCs in our sample are invisible to 21 cm observations.

3. Saturation prevents accurate measurements of many components with column densities $17 \lesssim \log N(\text{H}) \lesssim 19$, but we derive robust measurements of components with $\log N(\text{H}) < 17$ and $\log N(\text{H}) \gtrsim 19$. Saturated components are kept in the sample (with their associated large uncertainties) since excluding them does not significantly change our results on the H I distribution function.

4. The H I components have $b$ values $< 50$ km s$^{-1}$. For all three velocity categories (ISM, IVC, and HVC), the distribution of $b$ values is peaked near 10 km s$^{-1}$ (close to the FUSE instrumental resolution) and shows an extended tail out to 50 km s$^{-1}$.

5. We have computed the sky-covering fractions for all, ISM, IVC, and HVC H I components, both in the discrete $f_{\text{sky}}(N + \Delta N)$ and cumulative $f_{\text{sky}}(> N)$ forms. We find an excess of IVCs in the northern hemisphere compared to the south. To quantify this excess, we show that $f_{\text{sky}} = 0.7$ is reached at log $N(\text{H}) = 18.7$ in the north but 16.7 in the south. For HVCs, the sky-covering fractions are similar in the north and the south, though there is more high column density H I in the north than the south.

6. We have computed the first UV CDDF for Galactic H I. This extends the range of the Galactic CDDF by over four orders of magnitude deeper than existing 21 cm surveys (Wakker 1991; Murphy et al. 1995; Lockman et al. 2002; Moss et al. 2013; Westmeier 2018). Our CDDF includes an incompleteness correction based on simulating the observability of mock components inserted into the data as a function of velocity and $N(\text{H})$, and we assess the impact of saturation as a function of $N(\text{H})$.

7. We find that IVCs and HVCs show statistically indistinguishable CDDF slopes, with $\beta_{\text{HVC}} = -1.01^{+0.15}_{-0.14}$ and $\beta_{\text{IVC}} = -1.05^{+0.07}_{-0.06}$. In contrast, ISM clouds (at low velocity) have a shallower slope $\beta_{\text{ISM}} = -0.68^{+0.17}_{-0.20}$. These ISM clouds tend to have larger $N(\text{H})$ column densities. This slope difference from the HVC and IVC distributions suggests a significantly different behavior and reflecting the different physical conditions in the disk and halo. Although incompleteness affects the ISM and IVC samples at low column densities, and saturation affects all components at intermediate column densities,
the high column density results are relatively robust. Furthermore, the HVC slope is not strongly affected by incompleteness effects since that preferentially affects the low-velocity gas.

8. We fit the global H I CDDF as a function of column density with a piecewise three-component model, in which the breakpoints are automatically identified. We have presented separate “Galactic halo” and combined “Galactic halo + disk” versions of our CDDF to aid in the comparison of Galactic and extragalactic studies. We found evidence that three components are needed for an adequate fit, with a marked steepening of the slope $(\beta = -1.78^{+0.29}_{-0.42})$ in the intermediate-column density range $16.6 < \log(N(H)) < 17.8$, driven by an excess of absorbers observed near $\log N = 17$. This excess is present for both IVCs and HVCs, though we note it appears at the column densities where saturation becomes significant and so the values of $\log N$ are most uncertain. The slopes at the low end $(\log(N(H)) < 16.6; \beta = -0.62^{+0.11}_{-0.10})$ and high end $(\log(N(H)) > 17.8; \beta = -0.72^{+0.16}_{-0.07})$ are shallower.

9. By comparing the Galactic CDDF to a number of extragalactic CDDFs drawn from the literature, we have considered the relationship between the Galactic halo and extragalactic quasar absorption line systems. We find that the slope of the Galactic CDDF is most similar to the extragalactic CDDFs (Ribaudo et al. 2011; O’Meara et al. 2013; Rudie et al. 2013; Shull et al. 2017) in the intermediate-column density range (Lyman Limit Systems), but is flatter than that below $\log N(H) \approx 16.6$. While this flattening could be related to incompleteness, it is more likely due to the difference in physical environment between Galactic halo and Lyα forest absorbers. The shallower Galactic CDDF slope may reflect the enhanced ionizing radiation field in the Galactic halo compared to the IGM, as well as stripping and thermal instabilities resulting from infall. Overall, the shape of the high- and intermediate-column density range of the Galactic halo CDDF appears similar to extragalactic findings.

10. We find an excess of inflowing H I over outflowing H I in HVCs, as revealed by higher gas masses and mass flow rates for components with negative $v_{\text{GSR}}$ than for those with positive $v_{\text{GSR}}$. Using a partially covered spherical shell model, we show that inflowing HVCs carry $-0.88 \pm 0.40 (d/12 \text{ kpc}) M_\odot \text{yr}^{-1}$ whereas outflowing HVCs represent only $0.20 \pm 0.10 (d/12 \text{ kpc}) M_\odot \text{yr}^{-1}$. The excess of inflow agrees with results from UV metal-line HVCs (Shull et al. 2009, Fox et al. 2019), but our new measurements are based on H I so they have the advantage of directly measuring the (neutral) baryons, without the need for metallicity or dust corrections. These mass flow rates show that the halo overall experiences a net accretion of cool gas, which can contribute to future star formation.

By using the diagnostic power of the Lyman series absorption lines combined with the information content of the FUSE archives, this survey has provided a new, metal-independent view of the Galactic halo down to very sensitive limits. The measurements presented here constitute a rich database of the cool gas properties of the Milky Way that will be valuable for constraining future simulations of the formation and evolution of the Galactic halo.

We thank the dedicated team that operated the FUSE satellite. More than a decade after the mission ended, the data are still leading to new scientific insights. We gratefully acknowledge support from the NASA Astrophysics Data Analysis Program (ADAP) under grant 80NSSC18K0421, Surveying the H I Content of the Galactic Halo via Lyman Series Absorption. We thank Vanessa Moss, J. X. Prochaska, and Max Gronke for useful discussion on the CDDF, and we are grateful to the referee for a thorough report.

Facilities: FUSE, HST/STIS, HST/COS, GBT.

Software: Astropy (The Astropy Collaboration et al. 2018), CalFUSE (Dixon et al. 2007), Corner (Foreman-Mackey 2016), Emcee (Foreman-Mackey et al. 2013), Matplotlib (Hunter 2007), NumPy (Harris et al. 2020), pandas (Wes McKinney 2010), SciPy (Virtanen et al. 2020), VoigtFit (Krogager 2018).

Appendix A

Lyα Inclusion

We have tested the effect of excluding the Lyα transition from our fitting analysis. Lyα 1215.6700 lies outside the wavelength coverage of FUSE, thus including it requires obtaining spectra from HST/COS or STIS. This adds significant overhead to the analysis, as well as complications due to combining spectra with different S/N, resolution, and line-spread functions. To test if this is necessary, we chose a sample of four sight lines of varying signal-to-noise ratios, and fit them while both including and excluding Lyα. The result is shown in Figure 16.

We find that below column densities of $\sim 10^{19} \text{ cm}^{-2}$, there is little difference between the fit results. Above this point, the no-Lyα fits often under-predict the column densities by substantial amounts (up to 3 dex). All four sight lines show at least one component where the no-Lyα fit produced a column density $1 \sigma$ or more away from the Lyα-included fits. For this reason, we have included Lyα in the fitting procedure whenever data is available.
Appendix B
Spectra and Fits

Figure 17 presents corner plots from our MCMC piecewise CDDF fits. The histograms along the outer diagonal show the probability density functions (PDFs) for the normalization (i.e., $y$-intercept, $b_0$), low-, mid-, and high-column density slopes ($\beta_1$, $\beta_2$, $\beta_3$), and low- and high-column density break locations. The contour plots below each diagonal PDF show the joint posterior PDFs for the row and column.

Figures 18 present our best Voigt fits for each target and Lyman series transition. Ly$\alpha$ fits are included for all sight lines with data available (all but 1H0707-495 and NGC1068).

Figure 16. A comparison of component column densities resulting from including or excluding Ly$\alpha$ data when fitting four sight lines. At high values of $N$(HI), excluding Ly$\alpha$ often results in a substantial underestimate of the column density.
Figure 17. Corner plots from our MCMC piecewise CDDF fits. Top: Full data set fit including ISM, IVC, and HVC components (Galactic halo and disk). Bottom: Galactic halo only fit ($v_{LSR} \geq 40$ km s$^{-1}$). Both: The histograms along the outer diagonals show the probability density functions (PDFs) for the normalization (i.e., $y$-intercept; $b_0$), low-, mid-, and high-column density slopes ($\beta_1$, $\beta_2$, $\beta_3$), and low- and high-column density break locations. The contour plots below each diagonal PDF show the joint posterior PDFs for the row and column.
Figure 18. Fits for PG1116+215. The data are shown in black, errors and/or masked regions in light gray, and the composite fit in red. Each contributing component is plotted with a unique color, and the matching tick marks in the top of each panel show the centroid velocity. (The complete figure set (25 images) is available.)
| Sight line | $l$ (deg) | $b$ (deg) | ($S/N)_{977}$ | $v_\pm \sigma_v \pm \sigma_{v_{sys}}$ (km s$^{-1}$) | $b_\pm \sigma_b \pm \sigma_{b_{sys}}$ (km s$^{-1}$) | $\log_{10} N(HI) \pm \sigma_e \pm \sigma_{sys}$ | Note |
|------------|----------|----------|--------------|---------------------------------|---------------------------------|---------------------------------|------|
| HR707-495  | 15.004   | -17.67   | 6.41         | 0 \pm 3 \pm 1                   | 36 \pm 1 \pm 1                 | 20.24 \pm 0.038 \pm 0.029     |      |
| 3C723      | 289.95   | 64.36    | 19.59        | -117 \pm 2 \pm 4               | 3 \pm 1 \pm 2                  | 14.933 \pm 0.155 \pm 0.131    |      |
| ESO141-G55 | 338.18   | -26.71   | 8.26         | -67 \pm 2 \pm 1                | 9 \pm 1 \pm 1                  | 19.52 \pm 0.299 \pm 0.123     |      |
| HR821+643  | 94.00    | 27.42    | 7.64         | -214 \pm 13 \pm 1              | 20 \pm 20 \pm 20              | 14.073 \pm 0.280 \pm 0.177    |      |
| HE0226-411 | 253.94   | -65.77   | 11.2         | -4 \pm 1 \pm 0                 | 15 \pm 0 \pm 3                 | 20.098 \pm 0.006 \pm 0.004    |      |
| MRK279     | 115.04   | 46.86    | 20.62        | 75 \pm 1 \pm 2                 | 2 \pm 0 \pm 0                 | 18.868 \pm 0.360 \pm 0.263    |      |
| MRK335     | 108.76   | -41.42   | 13.22        | -415 \pm 2 \pm 1               | 15 \pm 1 \pm 2                | 18.179 \pm 0.439 \pm 0.371    |      |
| MRK421     | 179.83   | 65.03    | 14.62        | -170 \pm 1 \pm 1               | 3 \pm 0 \pm 1                 | 17.558 \pm 0.243 \pm 0.233    |      |
| MRK509     | 35.97    | -29.86   | 7.02         | -322 \pm 4 \pm 2               | 10 \pm 1 \pm 1                | 19.03 \pm 0.113 \pm 0.065     |      |
| MRK817     | 100.30   | 53.48    | 18.54        | -147 \pm 13 \pm 1              | 18 \pm 4 \pm 1                | 19.222 \pm 0.394 \pm 0.172    |      |

Note: $f_1$, $f_2$, $f_3$, $f_4$, $f_5$.
| Sight line  | $f$  | $b$   | $(S/N)_{rot}$ | $r \pm \sigma_r \pm \sigma_{r_{sys}}$ (km s$^{-1}$) | $b \pm \sigma_b \pm \sigma_{b_{sys}}$ (km s$^{-1}$) | $\log_2 N$(H$^+$) $\pm \sigma_{N} \pm \sigma_{N_{sys}}$ | Note |
|------------|------|-------|---------------|--------------------------------------------|---------------------------------|---------------------------------|------|
| MRK876     | 98.27| 40.38 | 10.9         | $-205 \pm 4 \pm 1$ | $22 \pm 2 \pm 1$ | $16.628 \pm 0.110 \pm 0.005$ | ... |
| MRK1383    | 349.22| 55.13 | 12.11        | $-106 \pm 1 \pm 4$ | $4 \pm 1 \pm 1$ | $16.932 \pm 0.773 \pm 0.638$ | ... |
| NGC1068    | 172.10| $-51.93$ | 9.8    | $-267 \pm 1 \pm 9$ | $10 \pm 1 \pm 1$ | $18.432 \pm 0.192 \pm 0.022$ | ... |
| PG0804+761 | 138.28| 31.03 | 13.97        | $-103 \pm 7 \pm 1$ | $30 \pm 4 \pm 0$ | $16.664 \pm 0.125 \pm 0.028$ | ... |
| PG0844+349 | 188.56| 37.97 | 6.44         | $-154 \pm 2 \pm 0$ | $3 \pm 2 \pm 1$ | $17.166 \pm 0.181 \pm 0.022$ | $f_6$ |
| PG0953+414 | 179.79| 51.71 | 10.52        | $-135 \pm 6 \pm 1$ | $22 \pm 4 \pm 0$ | $16.307 \pm 0.170 \pm 0.024$ | $f_6$ |
| PG1011-040 | 246.50| 40.75 | 8.89         | $-46\ldots$ | $15\ldots$ | $16.060 \pm 0.760 \pm 0.001$ | $f_6$ |
| PG1116+215 | 223.36| 68.21 | 12.19        | $-120 \pm 1 \pm 1$ | $3 \pm 1 \pm 1$ | $19.632 \pm 0.465 \pm 0.006$ | $f_6$ |
| PG1211+143 | 267.55| 74.31 | 11.58        | $-114 \pm 1 \pm 4$ | $11 \pm 1 \pm 2$ | $15.633 \pm 0.073 \pm 0.018$ | $f_6$ |

*Note: $f_6$*
| Sight line       | f (deg) | b (deg) | S/N_{tot} | v ± \sigma_v ± \sigma_{sys} (km s\(^{-1}\)) | b ± \sigma_b ± \sigma_{sys} (km s\(^{-1}\)) | log_{10}^{}N(H I) ± \sigma_n ± \sigma_{sys} | Note |
|-----------------|---------|---------|-----------|----------------------------------------|----------------------------------------|----------------------------------------|------|
| PG1259+593      | 120.56  | 58.05   | 13.7      | -200 ± 3 ± 16                          | 24 ± 4 ± 13                            | 15.559 ± 0.066                          |      |
| PKS0405-12      | 204.93  | -41.76  | 6.88      | -75 ± 4 ± 2                            | 5 ± 1 ± 1                              | 17.33 ± 0.389                           |      |
| PKS0558-504     | 257.96  | -28.87  | 8.6       | -114 ± 4 ± 2                           | 5 ± 1 ± 1                              | 20.376 ± 0.068                          |      |
| PKS2005-489     | 350.37  | -32.60  | 6.62      | -3 ± 6 ± 1                             | 19 ± 3 ± 1                             | 18.511 ± 0.035                          |      |
| PKS2155-304     | 17.73   | -52.25  | 12.5      | -315 ± 15 ± 2                          | 33 ± 21 ± 1                           | 14.638 ± 0.348                          |      |
| TON_S210        | 224.97  | -83.16  | 7.69      | -165 ± 7 ± 1                           | 12 ± 3 ± 1                            | 19.034 ± 0.125                          |      |

Note. The Note column indicates if the velocity (f\(_v\)), b-parameter (f\(_b\)), or log N (H I) (f\(_L\)) was manually fixed for this component. Fixed-component parameters consequently do not have error estimates. A dagger (†) and approximate symbol (~) indicate column density measurements that are very uncertain or possibly suffer from unreliable error estimates.

**ORCID iDs**

David M. French 🐣 https://orcid.org/0000-0003-3681-0016
Andrew J. Fox 🐣 https://orcid.org/0000-0003-0724-4115
Bart P. Wakker 🐣 https://orcid.org/0000-0002-5057-7096
Nicolas Lehner 🐣 https://orcid.org/0000-0001-9158-0829
J. Christopher Howk 🐣 https://orcid.org/0000-0002-2591-3792
Blair D. Savage 🐣 https://orcid.org/0000-0001-8016-6980
Philipp Richter 🐣 https://orcid.org/0000-0002-1188-1435
John O'Meara 🐣 https://orcid.org/0000-0002-7893-1054
Sanjaya Borthakur 🐣 https://orcid.org/0000-0002-2724-8289
Timothy Heckman 🐣 https://orcid.org/0000-0001-6670-6370
Colin Norman 🐣 https://orcid.org/0000-0002-5222-5717

**References**

Bish, H. V., Werk, J. K., Peek, J., Zheng, Y., & Putman, M. 2020, ApJ, 912, 1
Bland-Hawthorn, J., & Maloney, P. R. 1999, ApJL, 510, L33
Collins, J. A., Shull, J. M., & Giroux, M. L. 2003, ApJ, 585, 336
Collins, J. A., Shull, J. M., & Giroux, M. L. 2004, ApJ, 605, 216
Collins, J. A., Shull, J. M., & Giroux, M. L. 2005, ApJ, 623, 196
Collins, J. A., Shull, J. M., & Giroux, M. L. 2007, ApJ, 657, 271
Danforth, C. W., Keeney, B. A., Tilton, E. M., et al. 2016, ApJ, 817, 111
Dickey, J. M., & Lockman, F. J. 1990, ARA&A, 28, 215
Dixon, W. V., Sahnow, D. J., Barrett, P. E., et al. 2007, PASP, 119, 527
Field, G. B., & Steigman, G. 1971, ApJ, 166, 59
Foreman-Mackey, D. 2016, JOS, 1, 24
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Fox, A. J., Frazer, E. M., Bland-Hawthorn, J., et al. 2020, ApJ, 897, 23
Fox, A. J., Richter, P., Ashley, T., et al. 2019, ApJ, 884, 53
Fox, A. J., Savage, B. D., Wakker, B. P., et al. 2004, ApJ, 602, 738
Fox, A. J., Savage, B. D., & Wakker, B. P. 2006, ApJS, 165, 229
Fox, A. J., Wakker, B. P., Savage, B. D., et al. 2005, ApJL, 630, 332
French, D. M., & Wakker, B. P. 2017, ApJ, 837, 138
French, D. M., & Wakker, B. P. 2020, ApJ, 897, 151
Ganguly, R., Sembach, K. R., Tripp, T. M., & Savage, B. D. 2005, ApJS, 157, 251
Giroux, M. L., & Shull, J. M. 1997, AJ, 113, 1505
Gronke, M., & Oh, S. P. 2018, MNRAS, 480, L111
Harris, C. R., Jarrod Millman, K., van der Walt, S. J., et al. 2020, Nature, 585, 357
Ho, S. H., Martin, C. L., Kacprzak, G. G., & Churchill, C. W. 2017, ApJ, 835, 267
Hunter, J. D. 2007, CSE, 9, 90
Jitrik, O., & Bunge, C. F. 2004, IPCRD, 33, 1059
Kalberla, P. M. W., & Haud, U. 2006, A&A, 455, 481
Keeney, B. A., Stocke, J. T., Danforth, C. W., et al. 2017, ApJS, 230, 6
Kim, T. S., Carswell, R. F., Cristiani, S., D’Odorico, S., & Giallongo, E. 2002, MNRAS, 335, 555
Kim, T. S., Partl, A. M., Carswell, R. F., & Müller, V. 2013, A&A, 552, A77
