

CGM² I: The Extent of the Circumgalactic Medium Traced by Neutral HydrogenMATTHEW C. WILDE,¹ JESSICA K. WERK,¹ JOSEPH N. BURCHETT,^{2,3} J. XAVIER PROCHASKA,^{2,4} KIRILL TCHERNYSHYOV,¹
TODD M. TRIPP,⁵ NICOLAS TEJOS,⁶ NICOLAS LEHNER,⁷ RONGMON BORDOLOI,⁸ JOHN M. O'MEARA,⁹ AND
JASON TUMLINSON¹⁰¹University of Washington, Department of Astronomy, Seattle, WA 98195, USA²University of California, Santa Cruz; 1156 High St., Santa Cruz, CA 95064, USA³Department of Astronomy, New Mexico State University, PO Box 30001, MSC 4500, Las Cruces, NM 88001⁴Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU) The University of Tokyo; 5-1-5 Kashiwanoha, Kashiwa, 277-8583, Japan⁵Department of Astronomy, University of Massachusetts, 710 North Pleasant Street, Amherst, MA 01003-9305, USA⁶Instituto de Física, Pontificia Universidad Católica de Valparaíso, Casilla 4059, Valparaíso, Chile⁷Department of Physics, University of Notre Dame, Notre Dame, IN 46556⁸North Carolina State University, Department of Physics, Raleigh, NC 27695-8202⁹W. M. Keck Observatory, 65-1120 Mamalahoa Hwy., Kamuela, HI 96743, USA¹⁰Space Telescope Science Institute, Baltimore, MD, USA

ABSTRACT

We present initial results from the *COS* and *Gemini Mapping the Circumgalactic Medium* (CGMCGM $\equiv$ CGM²) survey. The CGM² survey consists of 1689 galaxies, all with high-quality Gemini GMOS spectra, within 1 Mpc of twenty-two $z \lesssim 1$ quasars, all with S/N $\sim$ 10 *HST/COS* G130M+G160M spectra. For 572 of these galaxies having stellar masses $10^7 M_\odot < M_\star < 10^{11} M_\odot$ and $z \lesssim 0.5$, we show that the H I covering fraction above a threshold of $N_{\text{HI}} > 10^{14} \text{ cm}^{-2}$ is $\gtrsim 0.5$ within 1.5 virial radii ($R_{\text{vir}} \sim R_{200m}$). We examine the H I kinematics and find that the majority of absorption lies within $\pm 250 \text{ km s}^{-1}$ of the galaxy systemic velocity. We examine H I covering fractions over a range of impact parameters to infer a characteristic size of the CGM, R_{CGM}^{14} , as a function of galaxy mass. R_{CGM}^{14} is the impact parameter at which the probability of observing an absorber with $N_{\text{HI}} > 10^{14} \text{ cm}^{-2}$ is $> 50\%$. In this framework, the radial extent of the CGM of $M_\star > 10^{9.9} M_\odot$ galaxies is $R_{\text{CGM}}^{14} = 346_{-53}^{+57} \text{ kpc}$ or $R_{\text{CGM}}^{14} \simeq 1.2 R_{\text{vir}}$. Intermediate-mass galaxies with $10^{9.2} < M_\star/M_\odot < 10^{9.9}$ have an extent of $R_{\text{CGM}}^{14} = 353_{-50}^{+64} \text{ kpc}$ or $R_{\text{CGM}}^{14} \simeq 2.4 R_{\text{vir}}$. Low-mass galaxies, $M_\star < 10^{9.2} M_\odot$, show a smaller physical scale $R_{\text{CGM}}^{14} = 177_{-65}^{+70} \text{ kpc}$ and extend to $R_{\text{CGM}}^{14} \simeq 1.6 R_{\text{vir}}$. Our analysis suggests that using R_{vir} as a proxy for the characteristic radius of the CGM likely underestimates its extent.

1. INTRODUCTION

The Circumgalactic Medium (CGM) represents the complex interface between stellar evolution, feedback from super-massive massive black holes, and the cosmic web, dictated by the large scale cosmology of the universe (e.g., Λ CDM) Tumlinson et al. (2017). These phenomena span many orders of magnitude in relevant spatial scales and involve several different sub-disciplines of astrophysics. In the canonical cosmological picture of galaxy evolution of White & Rees (1978), gas accretes onto galactic halos from the intergalactic medium (IGM), cools and condenses to form the interstellar medium, and then eventually stars. At the same time,

mass and energy is returned to the CGM via supernovae and AGN feedback, with some fraction potentially deposited back into the IGM.

In spite of its role as the harbor for a significant fraction of galactic baryons (e.g. Davé et al. 2010; Shull et al. 2012; Werk et al. 2014), the CGM is predicted to have low gas densities ($n \ll 0.01 \text{ cm}^{-3}$) due to its vast volume. Such low-density gas is most efficiently observed using absorption lines found in the spectra of bright background objects, usually quasars (QSOs). Generally, QSO absorption spectroscopy provides very high sensitivity to extremely low column densities in various ions, down to $N \sim 10^{12} \text{ cm}^{-2}$, and access to a broad range of ionized metal transitions that trace gas at both high and low densities in the CGM.

Bahcall & Spitzer (1969) first posited that galaxies are responsible for the high-column density absorption lines

in QSO spectra. In the 40 years since the prediction of Bahcall & Spitzer (1969), there has been much effort to link QSO absorption lines to both the IGM and CGM. In particular, the IGM traces the large scale structure of the universe, and is observed mainly through the Ly α forest (see Rauch 1998 and McQuinn 2016 for reviews). Previous works have found that the H I Ly α forest is associated with two distinct populations based on the strength of absorption systems; a low-density population that traces the cosmic-web, and a high-column density population that lies within dark matter halos (e.g., Morris et al. 1993; Tripp et al. 1998; Chen et al. 2005; Tejos et al. 2014; Burchett et al. 2020a). Linking sets of absorption lines arising at the same redshift (absorber systems) with redshifts of foreground galaxies in close proximity to the QSO sightline has essentially defined the last 20 years of CGM work (e.g. Chen & Tinker 2008; Prochaska et al. 2011; Tumlinson et al. 2013; Rudie et al. 2019). Boksenberg & Sargent (1978) and Bergeron & Stasińska (1986) reported some of the first identifications of intervening galaxies responsible for intermediate redshift absorption line systems in quasar spectra. Subsequently, Lanzetta et al. (1995) found that all the luminous galaxies in their sample exhibit extended gaseous envelopes (not yet called CGM), out to at least ~ 160 kpc.

Most of the early studies either searched for galaxies associated with known absorbers or they conducted blind surveys that started with no previously known information about either the absorbers or the galaxies at the outset of the survey. However, Bowen et al. (1991, 1995, 2002) used an inverted approach in which they searched for specific absorption features (e.g., Ca II Mg II, or H I) affiliated with galaxies with redshifts and properties that were known prior to the absorption survey. The advantage offered by this approach is that the CGM of galaxies with specific properties of interest could be accumulated more efficiently than in blind or absorption-selected surveys. Unfortunately, while many QSOs can be found behind low- z galaxies at interesting impact parameters, the first generation of HST spectrographs could not go deep enough to access most of those QSOs with a practical allocation of telescope time.

In 2009, the installation of the *Cosmic Origins Spectrograph* (COS) on HST, 20–30 times more sensitive than its UV spectrograph predecessors, allowed for targeted CGM studies of statistical samples of galaxies nearby in projection to background QSOs. The studies carried out with HST/COS over the last decade have revealed complex connections between the bulk properties of galaxies and their CGM. For example, the presence of O VI in the CGM, 10–150 kpc from the star-forming

host galaxy’s disk, is somehow linked to the present-day star-forming properties of the disk (Tumlinson et al. 2011). The H I content of the CGM scales with that of the host galaxies’ H I disks, perhaps implying that the disk is being fed by accretion from the IGM and CGM (Borthakur et al. 2015). In addition, large quiescent galaxies that have long since ceased star formation were observed to contain rich reservoirs of cool ($\sim 10^4$ K) gas (e.g. Thom et al. 2012; Chen et al. 2018) which should provide fuel for new star formation.

Thus far, the empirical picture of the CGM remains somewhat piecemeal, assembled from a variety of targeted observations of small-moderate sized samples carried out by different teams. Generally, various surveys of the CGM at low- z have revealed a multi-phase medium extending as far as $0.5 - 1 R_{\text{vir}} (R_{200m})$, which exhibits absorption from H I and a range of metal ions with ionization potential energies ranging from 10 - 239 eV (e.g. Lanzetta et al. 1995; Stocke et al. 2013; Tripp et al. 2011; Muzahid et al. 2011; Tumlinson et al. 2013; Werk et al. 2013; Burchett et al. 2019). Absorption signatures of low-ionization state metals like Mg II and Si II are rarely detected greater than ~ 100 kpc away from their host galaxies (Bordoloi et al. 2011; Nielsen et al. 2013; Werk et al. 2013; Bordoloi et al. 2014). By comparison, highly ionized metal lines such as O VI are detected out to ~ 200 kpc (Tumlinson et al. 2011; Johnson et al. 2015, 2017; Prochaska et al. 2019) and in M31, out to ~ 500 kpc (Lehner et al. 2020). This progress at low- z has been mirrored at $z \sim 2-3$ with the Keck Baryonic Structure Survey (KBSS; Rudie et al. (2012)), which found typical scales of the CGM to be ~ 300 kpc for L^* galaxies.

Observational progress enabled by HST/COS has presented several challenges to our current models of galaxy evolution. Photoionization modeling of the 10^4 K cool CGM gas around L^* galaxies has revealed it to be an order of magnitude under-pressurized with respect to the envisioned 10^6 K virialized, ambient hot halo (e.g. Werk et al. 2014; Prochaska et al. 2017; Haislmaier et al. 2020), and thus likely not in thermal or pressure equilibrium as expected from simulations or simple theoretical arguments. Complex, non-equilibrium physical processes may include precipitation (e.g. Voit 2019), shocks (e.g. McQuinn & Werk 2018), cosmic rays (e.g. Ji et al. 2019), kpc-scale fountain cycles (e.g. Fraternali & Binney 2008; Kim & Ostriker 2018), multi-filament gas flows fueling star formation (Martin et al. 2019), galactic winds (Bordoloi et al. 2011; Burchett et al. 2020b; Huang et al. 2020) and turbulent mixing layers (Fielding et al. 2020). Furthermore, simulations have generally under-predicted observed column densities of various ions, of both high-and-low ionization potential, com-

pared to observations of galactic halos (e.g. [Stinson et al. 2012](#); [Hummels et al. 2013](#); [Oppenheimer et al. 2016](#); [Liang et al. 2016](#); [Fielding et al. 2017](#)). Newer simulations (e.g. [Oppenheimer et al. 2018](#); [Lehner et al. 2020](#); [Nelson et al. 2020](#)) have more faithfully reproduced observations.

Recent simulations have shown that the CGM content, thermal structure, and kinematic properties are highly dependent on the mass and spatial resolution of simulations (e.g. [van de Voort et al. 2019](#)). High-resolution simulations that focus explicitly on the CGM, with resolved spatial scales of ~ 100 pc and mass scales of $\sim 10^3 M_\odot$ out to 250 kpc, have helped to ease the discrepancies with the observed column densities ([Peeples et al. 2019](#); [Hummels et al. 2019](#)), more naturally producing cool, denser clumps. However, it remains important to characterize the full extent of the CGM, as simulations still struggle to recreate all key aspects of the extended CGM.

For the above reasons, it has been challenging to put quantitative, physically-motivated bounds on the radial extent of the CGM. Generally, statistical QSO-galaxy samples are limited to single sightlines per galaxy due to the rarity of QSOs (but see [Chen et al. 2014](#); [Bowen et al. 2016](#)) and thus assembling large enough samples to place constraints on the extent of the CGM is challenging work. A comprehensive $z < 1$ CGM survey requires both space and ground-based spectroscopy. The former are required to measure key far-ultraviolet transitions such as $\text{Ly}\alpha$, $\text{Ly}\beta$ and metal species (e.g., C II, Si II, C IV, O VI). At low redshift, only Ca II, Na I, Fe II, and Mg I and Mg II can be observed from the ground. However, these species are confused by ionization and dust depletion and are hard to interpret by themselves. On the other hand, large samples of galaxies are accessible to ground based spectrographs to measure precise ($z_{\text{err}} < 10^{-4}$) redshifts of potentially associated galaxies.

To fill this need for a comprehensive, uniform CGM study, we have completed a deep spectroscopic survey of > 1000 galaxies around $z > 0.5$ QSOs, all within a few arcminutes of the background QSO sightline. The galaxies and absorption systems are blindly selected such that no preference to galaxy type or absorber optical depth is explicitly imposed. Our survey, *COS and Gemini Mapping the Circumgalactic Medium* (CGMCGM $\equiv$ CGM²), was designed to address many puzzles related to the CGM, such as: (i) What is the physical state of the gaseous halo when galaxies quench their star formation? (ii) What effects do environment and merger history have on the physical state and content gaseous halo? (iii) How do galaxy properties relate to the metal content and kinematic structure of their gaseous halos and vice

versa? (iv) What is the physical extent of the CGM? In this first presentation of the CGM² survey, we address the last question and in particular, how its spatial extent may depend upon halo mass and galaxy type. Ultimately, this study will inform how galaxy evolution and the large scale structure of the universe are connected.

CGM² includes a significant investment of effort from a team of University of Washington Undergraduate Astronomy majors led by CGM² PI Jessica Werk, the Student Quasar Absorption Diagnosticians (aka the *Werk SQuAD*) whose work is described in detail in the relevant sections. The outline of this paper is as follows. Section 2 describes the observations and data reduction for the CGM² spectroscopic galaxy survey. Section 3 describes an analysis of the galaxies' spectroscopic and photometric properties. In Section 4, we discuss our initial empirical results from associating galaxies with absorbers. We then derive a characteristic physical scale for the CGM in Section 5 by examining the H I-galaxy clustering. Section 6 compares our results to previous surveys of similar design and hydrodynamical cosmological simulations. We summarize and conclude this work in Section 7. Throughout this analysis we adopt the Planck15 ([Planck Collaboration et al. 2016](#)) cosmology as encoded in the *ASTROPY* package ([Astropy Collaboration et al. 2013](#); [Price-Whelan et al. 2018](#)). All distances are in physical space (not co-moving) unless otherwise noted.

2. OBSERVATIONS AND DATA REDUCTION

2.1. QSO Sample Selection

The CGM² survey is built upon the COS-Halos (GO11598, GO13033; [Tumlinson et al. 2013](#)) and COS-Dwarfs (GO12248; [Bordoloi et al. 2014](#)) surveys, which used 263 orbits of COS time to observe ~ 80 QSOs with $z = 0.2 - 1.0$. These surveys were designed to probe the halo(s) of one or two foreground galaxies well inside R_{vir} using suitable QSO-galaxy pairs. The COS-Halos QSO catalog was selected from the SDSS DR5 quasar catalog for QSOs that are UV-bright (GALEX FUV $\lesssim 18$) and lie at $z \lesssim 1$. The target QSOs were further selected to avoid Mg II absorbers at $z > 0.4$ to avoid losing a large range of the FUV QSO spectrum to LLSs. While this criterion did not affect the original surveys, which typically targeted specific galaxies at $z < 0.4$, it does affect the CGM² sample because it selects against high column density systems at $z \gtrsim 0.4$.

COS-Halos galaxies were selected via photometric redshifts to target stellar masses of $M_\star \simeq 10^{10-11} M_\odot$ at $0.1 < z < 0.3$ with projected separations from a nearby QSO $\rho < 150$ kpc. The precise redshifts (and other galaxy properties) were subsequently constrained with

follow-up spectroscopy (Werk et al. 2012). The COS-Dwarfs survey was designed to probe the CGM of galaxies with $\log M_*/M_\odot \lesssim 10^{10}$ at $z \simeq 0.01 - 0.05$ with masses and redshifts derived via SDSS photometry and spectroscopy. Both surveys yield galaxies that are the closest spectroscopically-identified galaxy to each QSO sightline and were designed to avoid biases with respect to galaxy neighbors, large-scale environment, or status as a satellite of a larger halo (in cases where neighbors were known).

For the CGM² multislit spectroscopic follow-up, we targeted a subset of “high-value” QSO fields with our Gemini program, either those with $z_{\text{QSO}} > 0.6$ or those with HST imaging available, which allows us to obtain morphology of any absorption-hosting galaxies with $z < 0.5$. To-date, we have obtained multislit galaxy spectra in 22 QSO fields as part of CGM². The fields we have surveyed, and some basic properties of the background QSOs, are tabulated in Table 1.

2.1.1. COS Data Reduction

The COS spectra were taken using both the G130M and G160M gratings. The balance between the time allocated to G130M and G160M was designed to achieve $S/N \simeq 10 - 12$ per resolution element ($\text{FWHM} \simeq 16 - 18 \text{ km s}^{-1}$) or better over 1150-1800 Å. The reduction of the COS spectra is explained in detail in Tumlinson et al. (2013) and Bordoloi et al. (2014) and follows the same method used by Tripp et al. (2011), Meiring et al. (2011), Tumlinson et al. (2011) and Thom et al. (2012) but we provide a brief description here. The COS data were obtained from MAST¹ and processed via CALCOS (v2.12) (Holland 2012) with standard parameters and reference files. First, exposures taken at the same grating and CENWAVE were combined. These coadded spectra were then coadded with exposures in the same grating at different CENWAVEs. This was followed by a combination of the two grating spectra to produce a single one-dimensional (1D) trace from 1150-1800 Å. At each coadd, exposures were combined by aligning common Milky Way interstellar absorption lines. Features related to the design of COS show up in the raw data and must be removed. The photocathode grid wires cast shadows on the detector as well as other fixed-pattern noise features and were removed with a flatfield and moreover are mitigated by the use of FP SPLITS. Flat-field reference files, prepared and communicated to us by D. Massa at STScI and filtered for high-frequency noise by E. Jenkins, were used to correct these fixed pattern features. These flats do not however, correct for gain-sag

depressions in the spectra created by prolonged exposure to bright geocoronal emission lines (Sahnou et al. 2011). The affected regions are flagged by the CALCOS pipeline and are rejected in our coaddition process. The 1D spectra are binned to Nyquist sampling resulting in a 1D, flat-corrected spectra with two bins per COS resolution element ($\text{FWHM} \simeq 18 \text{ km s}^{-1}$) and a $S/N \sim 8-10$. Errors arising from counting statistics (Poisson) are propagated through each step in the calibration.

2.2. Spectroscopic Galaxy Survey

2.2.1. Survey Design

All galaxy spectra were obtained at the Gemini North and South Observatories using the GMOS instrument (Hook et al. 2004; Gimeno et al. 2016) in multi-object spectrograph mode. The GMOS observations enabled us to obtain accurate, precise redshifts for low-redshift ($0.1 \lesssim z \lesssim 1$) galaxies as faint as $g \sim 24.5$ for star-forming galaxies and $g \sim 24$ for early-type galaxies. Over 3 observing semesters from 2014–2015 at both GMOS-N and GMOS-S, we surveyed galaxies in the high-quality QSO fields listed in Table 1.

GMOS uses slit masks that must be cut for each field based on highly-accurate astrometry from either a ‘pre-image’ taken in GMOS imaging mode or derived from an accurate catalog. For most of our fields, we used Gemini GMOS pre-imaging to create the masks. However, for fields observed in 2014A, specifically J0809+4619, J0943+0531, J1133+0327, J1134+2555, J1241+5721, and J1555+3628, we used a combination of HST-WFC3 imaging for angular separations less than $1'$ from the QSO along with SDSS imaging for galaxies further from the sightline to construct a slit mask target catalog. We refer to the observational program that observed these fields as the “Rollup” program and are differentiated in Table 2 with a ‘Q’ in the Maskname column. Subsequent observations used Gemini-GMOS g and i band pre-imaging to create the slit mask target lists and are referred to as “Large Program” (LP) observations. From the pre-imaging, we derived astrometry as well as magnitudes from SExtractor (Bertin & Arnouts 1996) using Gemini-calibrated photometric zero points and color corrections. We then optimized to get as many $g < 24$ galaxies into each slit mask as possible (see below).

In our GMOS programs, we used the R400 grating with $1''$ slits to balance wavelength coverage with spectral resolution. The fields from the Rollup programs listed above were dithered across two exposures per mask at central wavelengths of $\lambda\lambda 6000, 6900\text{Å}$ in order to avoid losing information to the chip gaps. The LP fields have spectra consisting of three exposures per mask at three separate central wavelengths,

¹ <http://archive.stsci.edu>

Table 1. QSO Fields in CGM²

QSO	QSO Long	RA	Dec	z_{QSO}	m_g
(1)	(2)	(3)	(4)	(5)	(6)
J0226+0015	J022614.46+001529.7	36.56028	0.25827	0.615	17.15
J0809+4619	J080908.13+461925.6	122.2839	46.3238	0.657	16.54
J0843+4117	J084349.49+411741.6	130.9562	41.2949	0.99	17.31
J0914+2823	J091440.38+282330.6	138.66829	28.39184	0.735	17.79
J0935+0204	J093518.19+020415.5	143.82581	2.07098	0.649	16.99
J0943+0531	J094331.61+053131.4	145.88173	5.52541	0.564	17.16
J1001+5944	J100102.55+594414.3	150.2606	59.7373	0.746	16.08
J1016+4706	J101622.60+470643.3	154.09418	47.11204	0.822	17.12
J1022+0132	J102218.99+013218.8	155.57913	1.53856	0.789	16.75
J1059+1441	J105945.23+144142.9	164.9385	14.6953	0.631	16.93
J1059+2517	J105958.82+251708.8	164.9951	25.2858	0.662	17.39
J1112+3539	J111239.11+353928.2	168.16296	35.65784	0.636	17.73
J1133+0327	J113327.78+032719.1	173.36578	3.45533	0.525	17.54
J1134+2555	J113457.62+255527.9	173.7401	25.9244	0.71	16.8
J1233-0031	J123304.05-003134.1	188.26688	-0.52616	0.471	17.76
J1241+5721	J124154.02+572107.3	190.4751	57.35205	0.583	17.58
J1342-0053	J134251.60-005345.3	205.71503	-0.89592	0.326	16.92
J1419+4207	J141910.20+420746.9	214.79251	42.1297	0.873	17.04
J1437+5045	J143726.14+504555.8	219.35892	50.76551	0.783	17.57
J1553+3548	J155304.92+354828.6	238.27052	35.80795	0.722	16.46
J1555+3628	J155504.39+362848.0	238.76833	36.48001	0.714	17.76
J2345-0059	J234500.43-005936.0	356.2518	-0.99335	0.789	16.8

NOTE—The full sample of QSOs included in the CGM² Survey: (1) QSO Short Name; (2) QSO Long Name, RA in hms and Dec in dms, all J2000; (3 & 4) RA, Dec in decimal degrees; (5) QSO redshift; (6) SDSS g -band magnitude of QSO.

$\lambda\lambda 6900, 7000, 7100\text{\AA}$. Despite these differences, both programs yielded similar depths with total exposure times of 1 hour per mask. Our final spectra achieved a S/N of at least a few per pixel at $\lambda_{obs} = 4800\text{\AA}$, the approximate wavelength of the 4000 $\text{\AA}$ break at $z \sim 0.2$.

Because precise redshift determination from our spectra was the primary goal, we chose a grating to cover 4400 to 9000 $\text{\AA}$ in order to detect emission from $z \lesssim 0.5$ [OII], $H\beta$ (for both redshifts and star formation metrics) and $H\alpha$ as well as absorption from Ca H+K and Na D in passively-evolving galaxies. The R400 grating provided a spectral resolving power of $R = \lambda/\Delta\lambda \simeq 950$, which corresponds to a velocity resolution of 300 km s^{-1} per resolution element. Ultimately, this setup allowed for determining the velocity centroid to $\sim 50 \text{ km s}^{-1}$ for precise redshift determination. We note that every spectrum does not exhibit uniform wavelength coverage given the design of GMOS. Galaxies that are not placed

near the center of the field have redder or bluer coverage than the range quoted above.

The slit masks were designed such that slits were placed on $\sim 85\%$ of objects within 1 arcminute of the QSO, with additional slits placed to fill the 5.5×5.5 arcminute area of the detector. Slit placement constraints meant that $\sim 40 - 50$ slits could be placed on one mask. We aimed to obtain spectra of $\sim 80 - 120$ unique galaxies per QSO field. Some fields in the earlier Rollup programs (those observed in 2014A) had as few as two and as many as four masks per field. In the later observing campaigns, all fields (except J1059+1441, which was only observed with one mask due to problems in the mask making) have three masks per field. This is shown in Table 2. An example of the targeting strategy can be seen in Figure 1 for the field J0843+4117. The red circles highlight the galaxies with $z < z_{\text{QSO}}$ for which we obtained reliable redshifts while the white circles were

Table 2. CGM² Multislit Mask Observations

QSO	Maskname	Date	Instrument	N(slits)	N(z)
(1)	(2)	(3)	(4)	(5)	(6)
J0809+4619	GN2014AQ001-01	2014-11-19	GMOS-N	31	24
J0809+4619	GN2014AQ001-02	2014-11-20	GMOS-N	27	21
J0809+4619	GN2014AQ001-03	2014-11-21	GMOS-N	29	21
J0809+4619	GN2014AQ001-04	2014-11-23	GMOS-N	17	15
J1134+2555	GN2014AQ001-05	2014-06-21	GMOS-N	29	26
J1134+2555	GN2014AQ001-06	2014-06-22	GMOS-N	29	26
J1134+2555	GN2014AQ001-07	2014-06-25	GMOS-N	24	20
J1134+2555	GN2014AQ001-08	2014-12-21	GMOS-N	16	13
J1241+5721	GN2014AQ001-09	2014-06-25	GMOS-N	27	18
J1241+5721	GN2014AQ001-10	2014-06-30	GMOS-N	24	23
J1555+3628	GN2014AQ001-11	2014-06-24	GMOS-N	33	29
J1555+3628	GN2014AQ001-12	2014-06-24	GMOS-N	26	24
J0914+2823	GN2014BLP003-01	2015-01-17	GMOS-N	45	29
J0914+2823	GN2014BLP003-02	2015-01-17	GMOS-N	44	27
J0914+2823	GN2014BLP003-03	2015-01-17	GMOS-N	45	26
J0843+4117	GN2014BLP003-04	2015-01-16	GMOS-N	39	28
J0843+4117	GN2014BLP003-05	2015-01-16	GMOS-N	39	31
J0843+4117	GN2014BLP003-06	2015-01-16	GMOS-N	39	31
J1059+1441	GN2014BLP003-08	2015-01-16	GMOS-N	45	38
J1001+5944	GN2014BLP003-10	2015-01-18	GMOS-N	45	36
J1001+5944	GN2014BLP003-11	2015-01-18	GMOS-N	45	29
J1001+5944	GN2014BLP003-12	2015-01-18	GMOS-N	41	33
J1016+4706	GN2014BLP003-13	2015-01-19	GMOS-N	42	34
J1016+4706	GN2014BLP003-14	2015-01-19	GMOS-N	40	30
J1016+4706	GN2014BLP003-15	2015-01-19	GMOS-N	42	33
J1112+3539	GN2014BLP003-19	2015-01-17	GMOS-N	49	33
J1112+3539	GN2014BLP003-20	2015-01-17	GMOS-N	46	31
J1112+3539	GN2014BLP003-21	2015-01-17	GMOS-N	43	37
J1059+2517	GN2015ALP003-01	2015-05-18	GMOS-N	47	31
J1059+2517	GN2015ALP003-02	2015-05-20	GMOS-N	43	29
J1059+2517	GN2015ALP003-03	2015-06-08	GMOS-N	42	25
J1419+4207	GN2015ALP003-04	2015-05-17	GMOS-N	45	27
J1419+4207	GN2015ALP003-05	2015-05-17	GMOS-N	46	33
J1419+4207	GN2015ALP003-06	2015-05-17	GMOS-N	43	32
J1419+4207	GN2015ALP003-07	2015-05-17	GMOS-N	41	25
J1437+5045	GN2015ALP003-08	2015-05-24	GMOS-N	40	33
J1437+5045	GN2015ALP003-09	2015-05-22	GMOS-N	43	31
J1437+5045	GN2015ALP003-10	2015-06-19	GMOS-N	42	30
J1553+3548	GN2015ALP003-11	2015-05-17	GMOS-N	45	32
J1553+3548	GN2015ALP003-12	2015-05-18	GMOS-N	45	35
J1553+3548	GN2015ALP003-13	2015-05-18	GMOS-N	41	28

Table 2. Continued

QSO	Maskname	Date	Instrument	N(slits)	N(z)
(1)	(2)	(3)	(4)	(5)	(6)
J0943+0531	GS2014AQ002-01	2015-04-19	GMOS-S	19	13
J0943+0531	GS2014AQ002-02	2015-04-18	GMOS-S	7	3
J0943+0531	GS2014AQ002-03	2015-02-18	GMOS-S	28	22
J0943+0531	GS2014AQ002-04	2015-02-19	GMOS-S	18	9
J1133+0327	GS2014AQ002-05	2015-03-15	GMOS-S	32	27
J1133+0327	GS2014AQ002-06	2015-04-25	GMOS-S	24	18
J1133+0327	GS2014AQ002-07	2015-04-25	GMOS-S	26	21
J1133+0327	GS2014AQ002-08	2015-05-14	GMOS-S	22	17
J1233-0031	GS2014AQ002-09	2015-04-19	GMOS-S	30	19
J1233-0031	GS2014AQ002-10	2015-05-14	GMOS-S	29	24
J1342-0053	GS2014AQ002-11	2015-01-04	GMOS-S	31	18
J1342-0053	GS2014AQ002-12	2015-06-06	GMOS-S	29	22
J0226+0015	GS2014BLP004-01	2014-10-26	GMOS-S	47	40
J0226+0015	GS2014BLP004-02	2014-10-26	GMOS-S	44	32
J0226+0015	GS2014BLP004-03	2014-10-26	GMOS-S	43	29
J2345-0059	GS2014BLP004-04	2015-06-19	GMOS-S	41	25
J0935+0204	GS2014BLP004-07	2015-02-13	GMOS-S	45	26
J0935+0204	GS2014BLP004-08	2015-02-13	GMOS-S	32	20
J0935+0204	GS2014BLP004-09	2015-02-17	GMOS-S	38	25
J1022+0132	GS2015ALP004-01	2015-02-17	GMOS-S	40	26
J1022+0132	GS2015ALP004-02	2015-02-17	GMOS-S	38	30
J1022+0132	GS2015ALP004-03	2015-02-17	GMOS-S	36	27
J1022+0132	GS2015ALP004-04	2015-02-17	GMOS-S	34	23

NOTE—Multislit Observation with GMOS taken as part of the CGM² Survey:
 (1) QSO short name; (2) Unique program mask name with project ID; (3) Date
 of mask observation; (4) Instrument, either GMOS-N or GMOS-S; (5) Number
 of slits placed on each mask; (6) Number of slits that yielded a reliable redshift.

targeted for spectra but for which we did not recover a reliable redshift.

All masks were observed between June 2014 and June 2015 on either GMOS-N or GMOS-S. Table 2 provides details of each multislit mask observed. Columns 5 and 6 list the number of slits placed and the number of slits that yielded reliable redshifts, respectively. A summary of the completeness of our galaxy spectroscopic survey is shown Table 3 for each field as a function of angular separation from the central QSO and limiting magnitude. The completeness is given as the fraction of reliable, redshift-yielding spectra, with ‘ZQ’ > 2 of galaxies (CLASSSTAR < 0.5) by SExtractor within some angular separation to the QSO. The fields J2345-0059 and J1059+1441 have only one mask per field and thus have a lower completeness relative to the rest of the sample. In general, we are nearly 100% complete out to 2 arcminutes from the QSO for galaxies with an *i*-band magnitude of < 22. The completeness drops to ~50% for a limiting magnitude of *i* < 23. 68% of the galaxies used in this analysis lie within 2 arcminutes of the central quasar. Details of the survey design, completeness and their impact on scientific results will be discussed in the full presentation of the CGM² survey, currently in preparation (Werk et al. 2021).

2.2.2. Gemini Data Reduction

The spectra were reduced using a combination of Gemini’s PyRAF package and PypeIt² (Prochaska et al. 2020). The initial reduction closely follows the GMOS Data Reduction Cookbook³. After obtaining the raw spectra from Gemini Observatory Archive, the data were organized according to field and mask. Biases were created by downloading all bias exposures of the same 2 × 2 binning taken within ~ 1 month from the observations if they were the same detector. Flat field and NeAr wavelength calibration exposures were taken along with each mask and were prepared using the standard methods in the cookbook. This procedure performs the bias subtraction, performs automatic slit edge finding, cuts the slits out of the image and isolates them, flat-fields each slit with the flats taken at the same central wavelength, and then performs a wavelength transformation to each slit. This results in multiple 2-D spectra. In general, the typical RMS of the wavelength solution is 0.1 pixel, which corresponds to 0.04Å given the dispersion of GMOS of 0.4Å per pixel. We then turned to PypeIt

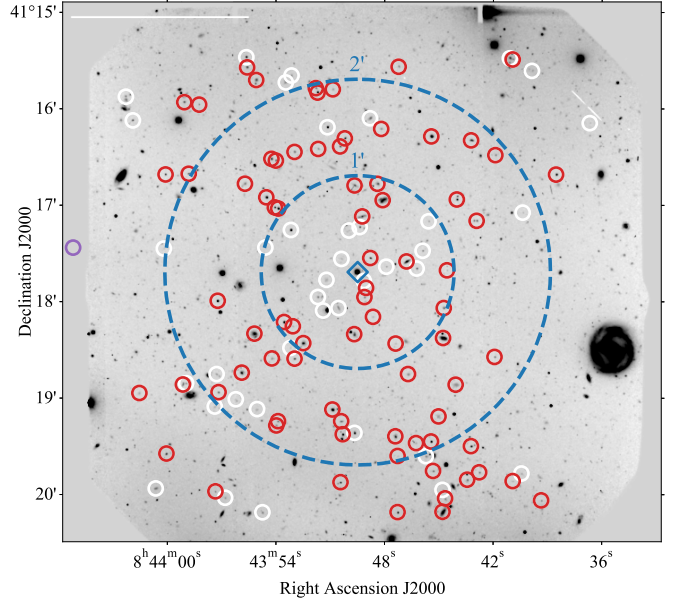


Figure 1. An example of the survey design and targeting strategy of CGM² showing the slit placements centered around QSO J0843+4117 (blue diamond) overlayed on the *g*-band pre-image from the Gemini-GMOS detector. The circles denote where slits were placed on the slitmasks. White circles indicate galaxies with slits whose final spectra did not yield a reliable redshift, while the red circles indicate galaxies that produced reliable redshifts. Large blue dashed circles show the one and two arcminute radii from the QSO. The purple circle just off the left of the detector is the COS-Dwarfs (Bordoloi et al. 2014) galaxy target for this field whose spectra was obtained by SDSS with an impact parameter of $\rho = 113$ kpc at $z = 0.0300$.

to perform sky subtraction, spectral continuum tracing, and combination of the wavelength-dithered. The spectra were then flux calibrated using a sensitivity function based on a selection of spectro-photometric flux calibration standards, BD28+4211, EG21, EG131, and Wolf1346, choosing whichever was closest on the sky. The final 1D spectra were co-added in vacuum wavelength space weighted by the inverse variance of the individual exposures.

3. ANALYSIS

3.1. Galaxy Redshift Determination

The process of determining galaxy redshifts was done in two stages. First, each 1D extracted spectrum was passed through an automated redshift fitting code,

² <https://github.com/pypeit/PypeIt>

³ Shaw, Richard A. 2016, GMOS Data Reduction Cookbook (Version 1.2; Tucson: National Optical Astronomy Observatory), available online at: http://ast.noao.edu/sites/default/files/GMOS_Cookbook/

Table 3. Completeness

QSO	$C_{22,1'}$	$C_{22,2'}$	$C_{22,FOV}$	$C_{23,1'}$	$C_{23,2'}$	$C_{23,FOV}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)
J0226+0015	0.57	0.79	0.47	0.34	0.41	0.23
J0809+4619	1.00	1.00	1.00	0.77	0.70	0.37
J0843+4117	1.00	1.00	0.63	0.70	0.51	0.31
J0914+2823	1.00	1.00	1.00	0.80	0.63	0.40
J0935+0204	0.93	0.86	0.42	0.33	0.40	0.20
J0943+0531	1.00	1.00	0.54	0.61	0.42	0.20
J1001+5944	1.00	1.00	0.65	0.87	0.53	0.32
J1016+4706	1.00	1.00	0.55	0.81	0.62	0.31
J1022+0132	1.00	0.97	0.79	0.59	0.51	0.35
J1059+1441	0.60	0.47	0.32	0.24	0.24	0.15
J1059+2517	1.00	1.00	0.68	0.72	0.69	0.36
J1112+3539	1.00	1.00	0.49	0.54	0.43	0.22
J1133+0327	1.00	0.95	0.48	0.68	0.46	0.23
J1134+2555	1.00	0.99	0.66	0.70	0.57	0.33
J1233-0031	1.00	0.44	0.25	0.43	0.19	0.11
J1241+5721	1.00	0.79	0.35	0.90	0.36	0.14
J1342-0053	1.00	0.58	0.29	0.54	0.30	0.14
J1419+4207	1.00	1.00	1.00	1.00	1.00	0.76
J1437+5045	1.00	1.00	0.66	1.00	0.63	0.33
J1553+3548	1.00	1.00	0.82	1.00	0.74	0.40
J1555+3628	1.00	1.00	0.62	0.67	0.53	0.24
J2345-0059	0.28	0.40	0.16	0.16	0.19	0.09
Median	1.00	1.00	0.57	0.69	0.51	0.27

NOTE—Completeness of the CGM² galaxy catalog along with the median in each radial and magnitude limited bin. The completeness is defined as the fraction of reliable spectra, with ‘ZQ’ > 2, to objects designated as a galaxy (CLASSSTAR < 0.5) by SExtractor (1) QSO Field; (2) Completeness within 1’ of the QSO at a limiting *i*-band magnitude of 22; (3) Completeness within 2’ of the QSO at a limiting *i*-band magnitude of 22; (4) Completeness within the GMOS field of view of the QSO at a limiting *i*-band magnitude of 22; (5) Completeness within 1’ of the QSO at a limiting *i*-band magnitude of 23; (6) Completeness within 2’ of the QSO at a limiting *i*-band magnitude of 23; (7) Completeness within the GMOS field of view of the QSO at a limiting *i*-band magnitude of 23. The fields J2345-0059 and J1059+1441 have only one mask per field and thus have a low completeness relative to the rest of the sample.

REDROCK⁴ (v0.14). REDROCK was developed by the DESI team and uses a template fitting algorithm to generate a set of ranked best-fitting models, identifying the object’s type (QSO, Star, Galaxy) and corresponding redshift.

However, much of our galaxy sample has moderate to poor S/N, and the automated REDROCK redshift guesses

can thus fail catastrophically by fixating on spurious features. We constructed a method of manually vetting the REDROCK redshifts by eye using a custom GUI, VETRR⁵. Each redshift was visually assessed and assigned a quality flag, Z_Q , by one of the authors, and by at least two members of the Werk SQuAD. A Z_Q of either 0,

⁴ <https://github.com/desihub/redrock>

⁵ <https://github.com/mattcwilde/vetrr>

1, 3, 4 were assigned to each spectrum, where a $Z_Q = 0$ indicates the spectrum has a S/N that is too low to be useful. $Z_Q = 1$ denotes a good spectrum but we are not confident in identifying a redshift (i.e. has no clear absorption or emission lines). $Z_Q = 3$ are spectra with one absorption or emission line that was confidently identified. These are usually strong [OII], H α emission or weak CaII absorption. A solitary H α emission only falls into this category if H β is off the detector and the strong emission line is too narrow to be [OII], which is a marginally unresolved doublet in these data. $Z_Q = 4$ represents a spectrum for which we are most confident in the redshift, with at least two absorption or emission lines identified. $Z_Q = 2$ was not used. The fully vetted galaxy survey database that we use for our analysis contains only spectra with $Z_Q > 2$. The statistical uncertainty of our redshift identification was determined by computing the standard deviation of redshift identifications from at least three humans for a sample of 50 galaxies and is typically in the range of $\sigma_z \sim 50\text{--}100 \text{ km s}^{-1}$ ($z \sim 0.00016\text{--}0.00030$). The redshift distribution of the vetted galaxy database containing 971 galaxies with redshifts less than that of the field quasar is shown in Figure 2.

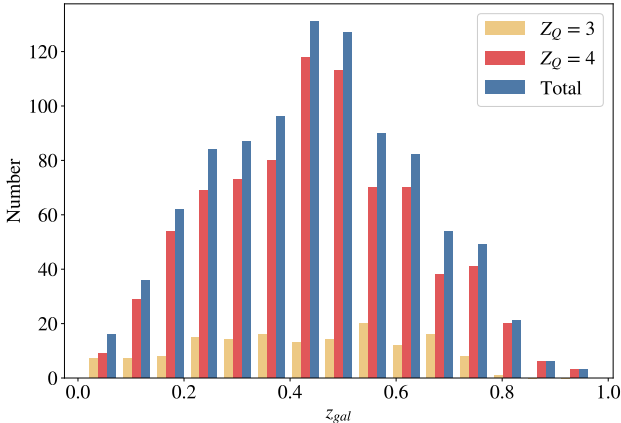


Figure 2. The redshift distribution of the CGM² galaxy catalog for galaxies with $z < z_{\text{QSO}}$. The redshift reliability is encoded in yellow and red. Red represents our most reliable redshift quality flag of ‘4’ with spectra displaying more than one strong absorption or emission line. A quality flag of ‘3’ was reserved for spectra with only one strong emission line and thus a less reliable redshift designation. Approximately 85% (820 out of 971) of our spectra were given the highest reliability flag. The typical statistical uncertainty of our redshifts is $\sigma_z \sim 50\text{--}100 \text{ km s}^{-1}$ ($z \simeq 0.00016\text{--}0.00030$).

In addition to manually vetting the galaxy redshifts, we also visually identified a galaxy spectral type during

the process of examining each 1D spectrum. If strong emission lines and weak continuum were present, we classified the galaxy as “star forming” or “SF”. If only absorption lines along a detected continuum were found, we classified the galaxy as “quiescent” or “E”. If both emission and absorption lines were identified, we classified the galaxy as a combination of the two, “SF+E.” In several cases, stars were mistakenly targeted in our slit masks, and they have the identification as star. Stars are never included as part of our vetted galaxy database. Examples of galaxy spectra of each type are shown in the three panels of Figure 3. These spectra contain some poorly-subtracted sky lines, and telluric absorption lines at $\sim 7600\text{\AA}$, which are the sorts of spurious spectral features that can cause REDROCK to fail. The top panel displays an emission-line ‘SF’ galaxy that shows emission from several highlighted strong emission lines, the middle panel shows a combination-type, ‘SF+E’ spectrum with both [OII] emission and strong CaII absorption against a bright stellar continuum, and the bottom panel shows an example of an absorption-only spectrum, type ‘E.’

3.2. Galaxy Photometry and Spectral Energy Distribution Fitting

In addition to the galaxy spectroscopic catalog, we constructed a photometric galaxy catalog to derive stellar masses and galaxy star formation rates (SFRs). Our spectra are generally insufficient in signal and flux calibration to analyze them via spectral fitting codes (e.g. Cappellari 2017). To estimate stellar masses, we used CIGALE (Noll et al. 2009; Boquien et al. 2019) to fit the spectral energy distribution and retrieve stellar mass and SFRs. We note that while there are myriad other SED fitting codes available (as well as direct color-mass relations) we chose CIGALE for its options for stellar population models, dust models, etc. across large swaths of parameter space. We also chose it to compare directly to complementary surveys such as CASBaH (Burchett et al. 2019; Prochaska et al. 2019).

3.2.1. Galaxy Photometry

One major challenge in constructing a photometric catalog for the CGM² survey was that the spectroscopic target catalog is generally fainter than available public all-sky surveys and thus the photometric coverage is not uniform in all fields. We chose to gather photometric data for every spectroscopic target in the galaxy catalog, totalling 2310 unique targets.

We created the photometric catalog by cross-matching our target to the DESI Legacy Imaging Surveys Data Release 8 (DR8) (Dey et al. 2019). The imaging survey

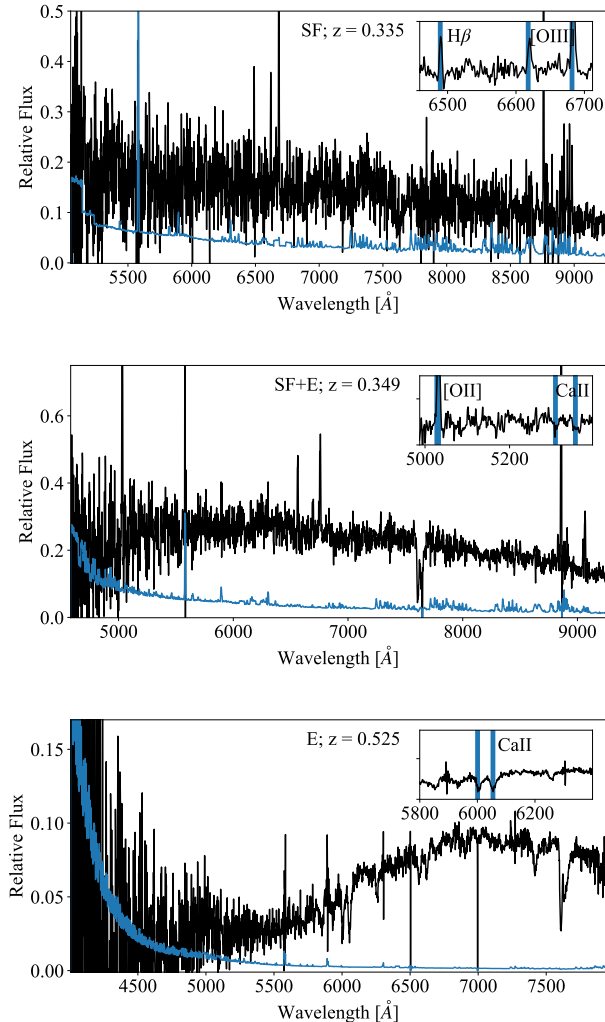


Figure 3. Examples of typical Gemini-GMOS spectra with a quality flag of $Z_Q = 4$, along with the error shown in blue. These spectra highlight our method of visual galaxy spectral typing. The cut out insert in each panel shows an example of the key spectral features used in redshift determination. The top spectrum is classified as star-forming galaxy, the middle panel shows a galaxy with both emission lines and absorption lines, and the bottom spectrum is an example of galaxy with an older stellar population with strong Ca H+K absorption.

for DESI is composed of data from three telescopes covering $\sim 14000 \text{ deg}^2$ over $-18^\circ < \delta < +84^\circ$ ($|b| > 18^\circ$). These three Programs include The Beijing-Arizona Sky Survey (BASS), The DECam Legacy Survey (DECaLS), and the Mayall z-band Legacy Survey (MzLS) which provide g , r , and z band photometry to ~ 23.3 mag.

The photometry is corrected for Galactic extinction. This is the deepest publicly available optical survey and provided the bulk of the photometry. In cases of overlap

between the North and South catalogs, we chose the Southern DECaLS observations. We limited matches to objects with $S/N > 2$ and chose the closest match within 1.3 arcseconds of our targets in order to limit mismatches between our faint sources and the Legacy Survey catalogs. This gives us 1985 targets with at least one band of photometry. In addition to the g , r , and z bands, DR8 provides cross-matched WISE (Cutri et al. 2013) observations in 3.4, 4.6, 12, and $22 \mu\text{m}$.

In order to cover a larger wavelength range to better estimate the SED, we also cross-matched our catalog with the Pan-STARRS Data Release 2 (Chambers et al. 2016) with coverage of the *grizy* bands, utilizing the MAST cross-match service, using a 1.3 arcsecond threshold. We limited photometry to those marked as extended objects with good stack photometry and *grizy* $< 23.3, 23.2, 23.1, 22.3, 21.3$, giving 393 objects with photometry in at least one band.

In addition we also queried the SDSS DR14 (Abolfathi et al. 2018) survey with *ugriz* coverage where we restricted matches to less than *ugriz* $< 22.15, 23.13, 22.70, 22.20, 20.71$, totalling to 331 targets with photometry in at least one band.

To make sure we included photometry for every object in the CGM² survey that may be too faint or in crowded areas for the public surveys, we included photometry from the GMOS imaging. The target selection and slit mask design was based on Gemini GMOS *i*-band and g -band imaging except in the case of the Rollup programs, GN-2014A-Q1 and GS-2014A-Q2, which involved a combination of HST WFC-ACS and Gemini imaging. This allowed us to use the Gemini images to get photometry for each field, the details of which are presented in Table 4. Two exposures in each band were combined and processed by Gemini. We obtained g and i band magnitudes of the Gemini sources using the SExtractor software. We used pre-calibrated photometric zero-points and color corrections for GMOS-N and GMOS-S for an initial pass on the Gemini photometry. In order to further calibrate the Gemini photometry, we cross-matched our $m < 21$ sources to the g and i band SDSS photometric sources, and then bootstrapped the Gemini photometry below the SDSS magnitude limit within each field. This consisted of applying a constant magnitude offset to the Gemini sources to match the SDSS photometry in the g and i bands.

All photometry from Gemini, SDSS, and Pan-STARRS was corrected for Galactic reddening based on the values in Schlafly & Finkbeiner (2011) provided

Table 4. CGM² Gemini GMOS Imaging

Project ID	g - t_{exp} [s]	i - t_{exp} [s]
(1)	(2)	(3)
GN-2014A-Q1	X	150
GS-2014A-Q2	X	150
GN-2014B-LP-4	450	200
GS-2014B-LP-3	450	200
GN-2015A-LP-3	450	200
GS-2015A-LP-4	450	200

NOTE—Imaging observations with GMOS-N and GMOS-S taken as part of the CGM² Survey: (1) Project ID; (2) g -band exposure time in seconds; (3) i -band exposure time in seconds

by the NASA Extragalactic Database⁶ by querying our targets coordinates via **ASTROQUERY**⁷. We employed the SVO Filter Profile Service (Rodrigo et al. 2012) to obtain our filter transmission curves for the telescopes used in these surveys as required input to CIGALE.

CIGALE includes many models as options to include in fitting. For stellar populations, we used the Bruzual & Charlot (2003) models, assuming a Chabrier (2003) initial mass function (IMF). We chose a grid of metallicities ranging from 0.001 – $2.5Z_{\odot}$. We used a delayed star formation history (SFH) model with an exponential burst. The e-folding time of the main stellar population models ranged from 0.1–8 Gyr. We varied the age of the oldest stars in the galaxy from 2–12 Gyr. We included an optional late burst with an e-folding time of 50 Myr and an age of 20 Myr. We varied the burst mass fraction from 0.0 or 0.1 to turn this feature on or off. Nebular emission and reprocessed dust models (Dale et al. 2014) were also included with the default values. The dust models have slopes ranging from 1–2.5 and the nebular models include no active galactic nuclei.

We employed the Calzetti et al. (1994) dust attenuation law, but we also included a “bump” in the UV (see discussion in Prochaska et al. (2019)) at 217.5 nm with a FWHM of 35.6 nm. The bump amplitude is set at 1.3 and the power law slope is -0.13 (Lo Faro et al. 2017). We varied the color excess of the stellar continuum from

the young population, $E(B-V)$, from 0.12–1.98. Finally, we used a reduction factor of 0.44 to the color excess for the old population compared to the young stars.

3.3. Galaxy Stellar Masses and Derived Properties

After fitting each galaxy’s SED, CIGALE then outputs several useful parameters including stellar mass, SFR, and the rest-frame absolute luminosity in each band. Figures 4 and 5 show the resultant mass distributions with masses spanning $M_{\star} \approx 10^6$ – $10^{11} M_{\odot}$ and $\bar{M}_{\star} = 10^{9.3} M_{\odot}$ at $\bar{z} = 0.44$. In order to calculate the virial radius of the galaxies, we first calculate the halo mass using the abundance matching method of Moster et al. (2013) with the modifications used in Burchett et al. (2016). We adopt R_{200m} , the radius within which the average mass density is 200 times the mean matter density of the universe, as the virial radius (R_{vir}) of a galaxy halo.

Figure 4 shows the CIGALE-derived SFRs vs stellar masses, which exhibit, as expected, a positive correlation known as the “star-forming main-sequence” (SFMS). We compare our models to the redshift dependent fit from Schreiber et al. (2015) over a range of galaxies spanning $z = [0.23, 0.63]$. This range represents the 16th and 84th percentiles of the CGM² galaxy catalogs redshift distribution. Typical uncertainties on photometry-derived stellar masses range from a factor of 3–5 (e.g. Blanton & Roweis 2007; Werk et al. 2012) corresponding to ~ 0.5 dex, and errors on CIGALE-derived SFRs are similar.

Figure 5 shows galaxy stellar masses vs. galaxy systemic redshift, and differentiates among our three galaxy spectral types visually derived from the GMOS spectra. Above $z \sim 0.5$, we are no longer detecting galaxies with $M_{\star} \lesssim 10^8 M_{\odot}$. Both SF and SF+E galaxy spectral types are preferentially distributed among lower stellar masses as seen in the marginal distributions. We find that, as expected, there are more ‘E’ type galaxies at higher galaxy stellar masses than a random distribution would predict. As shown in Figure 6, the CIGALE-derived color-mass diagram of our galaxy sample shows the bimodality of the star-forming and non-star-forming galaxies found in large galaxy surveys such as SDSS (e.g. Chang et al. 2015).

3.4. Absorber Catalog

To identify absorption features in the QSO spectra, the HST/COS UV spectra were visually inspected by members of the *Werk SQuAD* in a multi-step process which closely follows the procedure described in Tejos et al. (2014) and is designed to leave no QSO absorption feature unidentified. This process includes identifying all of the most ubiquitous metal ions present

⁶ The NASA/IPAC Extragalactic Database (NED) is funded by the National Aeronautics and Space Administration and operated by the California Institute of Technology.

⁷ <http://dx.doi.org/10.6084/m9.figshare.805208>

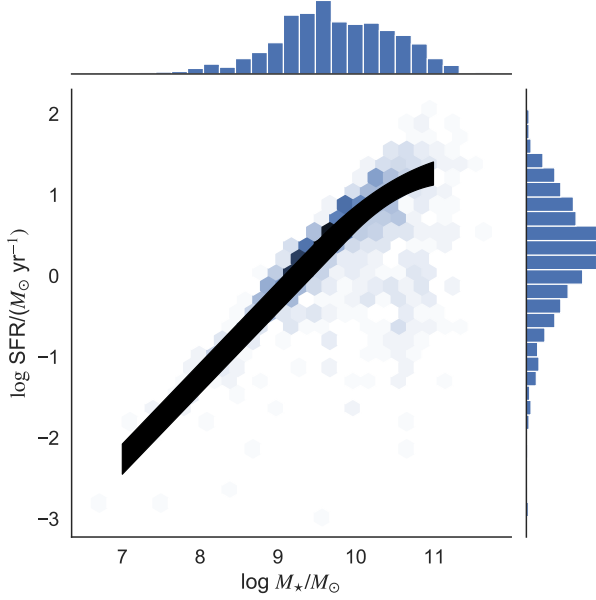


Figure 4. Star formation rate (SFR) vs. stellar mass for the CGM² galaxy sample as estimated by CIGALE. The locus of galaxies tracks a monotonic increase of SFR with stellar mass, known informally as the “star-forming main-sequence” (SFMS). The density of galaxies in this space is indicated via shading of the hexagonal bins. The grey shaded region corresponds to the redshift dependent fit of the SFMS from Schreiber et al. (2015) of galaxies spanning $z = [0.23, 0.63]$. This range represents the 16th and 84th percentiles of the galaxy catalogs redshift distribution.

in QSO spectra at the wavelengths covered by the HST/COS spectra, not just the H I Lyman series. These other metal ions will be explored in future CGM² papers. To identify the absorption features, the Werk SQuAD used a module from the PYIGM⁸ software package, IGMUESSES. The software allows for a straightforward comparison of multiple transitions from different elements, as multiple lines are displayed with their expected relative intensities given by their atomic parameters simultaneously for a given redshift.

During the line identification process, the absorption lines are assigned a reliability score. ‘a’ signifies a *certain* feature. For example, the always-present Milky Way (MW) ISM lines at $z = 0$ fall into this reliability category. Other examples include absorption lines observed in two or more transitions of H I, or multiple metal lines that align with hydrogen lines within $\pm 30 \text{ km s}^{-1}$, and metal doublets or multiplets that show the expected relative strengths as derived from their oscillator strengths

⁸ <https://github.com/pyigm/pyigm>

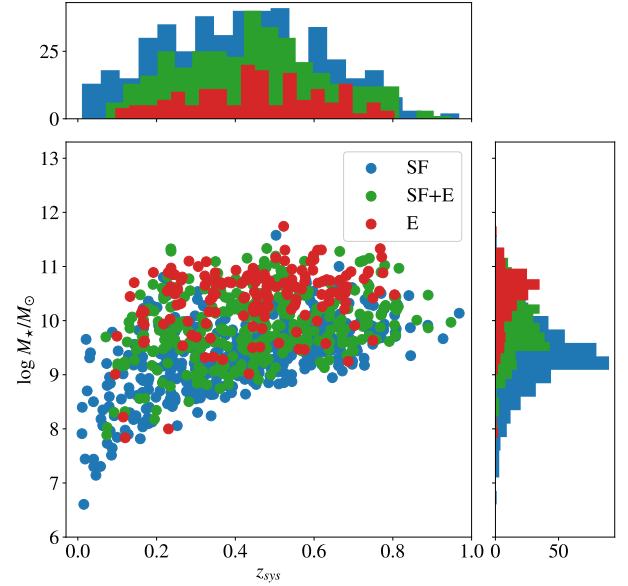


Figure 5. The distribution of galaxy stellar mass as a function of galaxy systemic redshift with marginal distributions on the right and top. The red, green, blue colors correspond to the galaxy spectral type determined from visual inspection of GMOS spectra. Red circles show absorption-line only, or elliptical (E) type galaxies, green circles show galaxies displaying a combination of absorption and emission lines associated with star formation (SF+E), and blue circles show emission-line only, or star forming (SF) galaxies.

and wavelengths (Morton 2003) and similar velocity profiles. A reliability score of ‘b’, or *possible*, includes single H I lines with no associated metal lines, and metal ions having only one transition within the observed wavelength range. Other cases of assigning ‘b’ values involve messy blends from absorption lines at different redshifts, weak or uncommon metal lines in an otherwise strong absorption system, and velocity offsets $> 30 \text{ km s}^{-1}$ from other ‘a’ lines. If a line did not fall into either of the previous categories we gave it a reliability score of ‘c’, or *unreliable*. In our analysis, we did not include any absorbers in the *unreliable* ‘c’ category. The full catalog of identified absorbers will be presented in detail in future work.

The end result of this line identification process for all 22 QSOs is a catalog of 2914 distinct absorption components, 2071 of which have a reliability rating of *certain* or *possible*. An absorption component is defined by an absorption line or lines with a distinct central velocity (or redshift). In practice, individual components offset by $< 20 \text{ km s}^{-1}$ may not be separable in the HST/COS

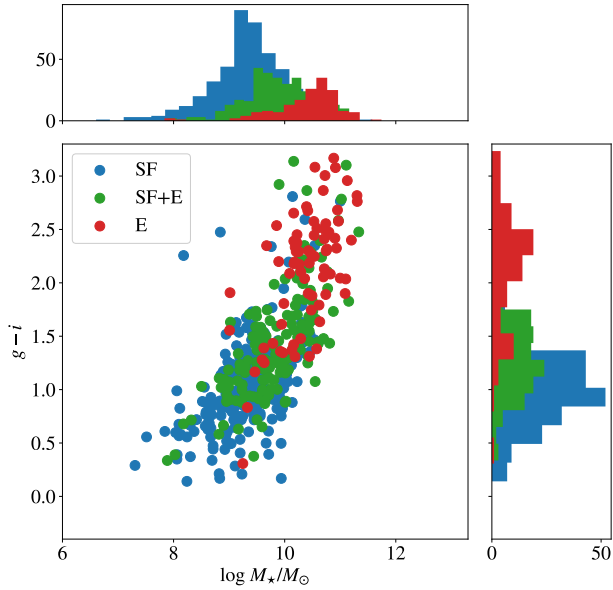


Figure 6. CGM² sample in a color-mass diagram using the $g - i$ color and the galaxy stellar mass, M_* . Multi-band

photometry was not available for all of the galaxy targets, only objects with both bands are shown here. The bi-modal populations of the star-forming and passive galaxies are evident. Due to the nature of our survey, we are slightly biased against faint, passive galaxies since retrieving a redshift in the case of absorption lines requires large continuum flux.

spectra. An absorber, or absorption system, is a set of absorption line components within $|\delta v| \approx 1000 \text{ km s}^{-1}$. For example, Ly- α and Ly- β are distinct lines but would be part of the same H I component if aligned within the COS resolution velocity (redshift). This component may be grouped with other H I or metal ion components to form an absorption system.

Different absorbers lie at distinct redshifts and may physically correspond to clouds or layers of gas in the CGM of galaxies at their respective redshifts. Absorbers may also be clouds or filaments of gas in the IGM referred to as the Ly α forest, not directly associated with a nearby galaxy. The term absorber is often used interchangeably with the term “system.” However, absorber distinctly does not imply an association with a galaxy and is a more empirical term. Absorption systems may have multiple components at different velocities within their assigned redshift ranges. The vast majority of our H I absorption systems cover > 3 Lyman series lines, many with multiple absorption components.

In order to retrieve more physical quantities such as column density, N_{HI} , we used the apparent optical depth

method (AODM) from [Savage & Sembach \(1991\)](#) as encoded in the `linetools`⁹ package. Because we have column density measurements from several Lyman Series lines in most cases, the mean 1σ uncertainties on column density is 0.17 dex for unsaturated H I lines to column densities $\simeq 10^{17.5} \text{ cm}^{-2}$.

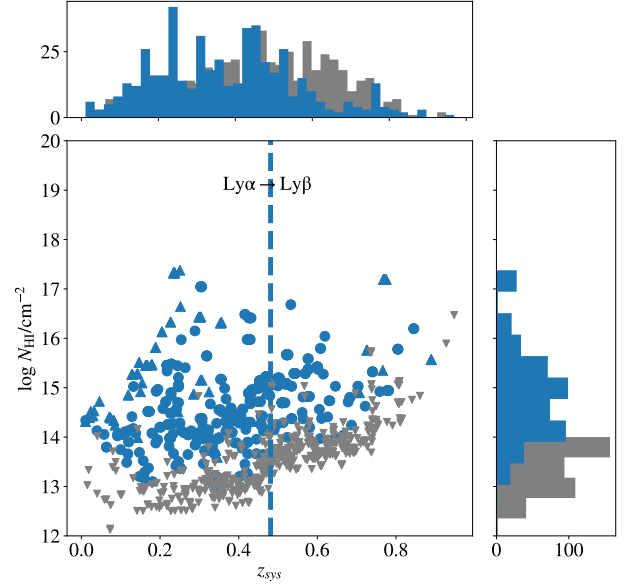


Figure 7. Scatter plot and marginal distributions of column densities vs. redshift for the H I systems detected in the CGM² survey. The mean 1σ uncertainties on column density is 0.17 dex for unsaturated H I lines to column densities $\simeq 10^{17.5} \text{ cm}^{-2}$, which is of order the size of the symbol (see Figure 9 for the size of our uncertainties). The measurements designated with upward triangles are saturated absorption lines and are thus lower limits while circles represent detections. Smaller gray downward facing triangles are 2σ upper limits for galaxies where no corresponding absorption was measured. The visible break in the minimum N_{HI} at $\gtrsim 0.48$ is shown by the vertical dashed line, and marks the redshift at which Ly α shifts out of the COS G160M bandpass and thus becomes inaccessible. There can be multiple systems at the same column density and redshift which appear as darker points in this plot. This occurs when multiple galaxies lie within $|\delta v| < 500 \text{ km s}^{-1}$ of the absorption systems.

4. CONNECTING GALAXIES AND ABSORBERS: THE CGM² H I SURVEY

4.1. Defining CGM H I Absorption Systems

⁹ <https://github.com/linetools/linetools>

With our separately completed galaxy and absorber databases, we can now begin to connect the two as a study of the CGM. In order to construct CGM systems, we first group the individual absorption component identifications into absorption systems, or absorbers. The grouping of absorption components was done using a clustering algorithm from SKLEARN (Pedregosa et al. 2011), MEANSHIFT. This algorithm groups individual absorption components together within a window function of 1000 km s^{-1} . The resultant absorber catalog consists of groups of components we call absorption systems.

We then cross-matched the galaxies and absorption systems if the relative velocity difference of the galaxy and the velocity centroid of at least one of the components of an absorption system exhibits $|\delta v| < 500 \text{ km s}^{-1}$. We chose this velocity threshold to include absorption systems that could be at or above the escape velocity of the most massive galaxies in our sample. If no absorption system is found at the redshift of a galaxy, we measure the 2σ upper limit within $\delta v = \pm 30 \text{ km s}^{-1}$ of the galaxies redshift using the normalized error of the quasar flux. If there was an interloping line at this redshift, we measure the AODM column density as a conservative upper limit. Our results are not sensitive to the choice in the velocity window. We find 181 systems consisting of 416 distinct components that exhibit HI column densities above our 2σ detection threshold, giving an average of 2.3 detected components per galaxy (absorption system) within our stated velocity window. We find 2.4 average components per galaxy for a smaller window of $|\delta v| < 250 \text{ km s}^{-1}$, while for a larger velocity window, we find a small decrease to 2.2 detected components per galaxy for a window of $|\delta v| < 1000 \text{ km s}^{-1}$. The average column density of detected, but not saturated, components remains $10^{14.9} \text{ cm}^{-2}$ in each case.

Thus, we are left with a galaxy-centric CGM survey that consists of absorption line column density measurements (or limits) around 971 galaxies with reliable redshifts that lie $< 1000 \text{ km s}^{-2}$ from the quasar. We denote these galaxy-absorber pairs as CGM systems. Figure 7 shows the AODM HI column densities for all 971 CGM systems as a function of galaxy systemic redshift. Saturated lines provide only lower limits to the HI column density and are shown as upward facing triangles. Non-detections are shown as 2σ upper limits, and as downward facing triangles. There is an obvious “knee” of amplitude 0.5 dex in the lowest HI column densities at a redshift of $z \sim 0.5$. This decreased sensitivity to weak HI absorption features is driven by the redshifting of Ly α out of the wavelength range of the COS G160M grating. The column density measurements at $z > 0.481$ are derived from measurements of the Ly β ab-

sorption line and/or weaker Lyman series transitions, leading to a decrease in sensitivity for a fixed S/N. This shift in sensitivity motivates us to limit our HI analysis to $z < 0.481$, corresponding to $\lambda_{\text{Ly}\alpha}(1+z) = 1800 \text{ \AA}$, or the reddest end of the COS-G160M grating. By imposing this limit, we ensure nearly uniform sensitivity to HI column density for our CGM sample.

After limiting our CGM system survey to the aforementioned redshift, we construct the CGM² HI survey, consisting of 572 total galaxy-absorber pairs. The survey includes all galaxies with reliable redshifts and HI absorption systems (including the non-detections within $|\delta v| < 500 \text{ km s}^{-1}$), collectively referred to as CGM systems. In the HI covering fraction analysis that follows, it is possible to have multiple galaxies at similar redshifts but at differing impact parameters that match with individual absorption systems (74 systems total). For understanding the HI extent of the CGM, we want to understand the correlation of galaxies and absorbers and thus do not limit our matches to the closest or most massive galaxy. However, in our complementary HI velocity analysis, we limit the survey to the galaxy with the smallest impact parameter, thus leaving 522 CGM systems. Future studies, depending on their specific aims, will make independent choices about how to include galaxy-absorber pairs.

In the following analyses, we will examine trends with the impact parameter, ρ , which quantifies the projected distance between the QSO and the galaxy in the rest frame of the galaxy. Figure 8 shows the impact parameter–redshift distribution of CGM systems. The grey curve approximates the distance to the edge of the detector in the $5.5'$ GMOS FOV, assuming the QSO is in the center, in order to highlight the survey coverage as a function of redshift. The QSO was slightly offset from the center in certain fields, either to better place guide stars or to avoid bright foreground stars; thus a few galaxies fall on or near this approximate field-size limit.

4.2. HI Covering Fraction: Definition and Threshold

To quantify the radial profile and extent of the CGM, we use the covering fraction f_c as a measure of the probability of the presence of HI. The covering fraction is the comparison of “hits” (H) and “misses” (M), with a hit being defined as a galaxy with a corresponding absorber at or above the detection threshold for the full ensemble (see below), while a miss occurs when the 2σ upper limit on a detection is below the threshold at the redshift of the galaxy. A system with a 2σ upper limit above the threshold is ignored altogether because this indicates that the S/N of the spectrum is not adequate

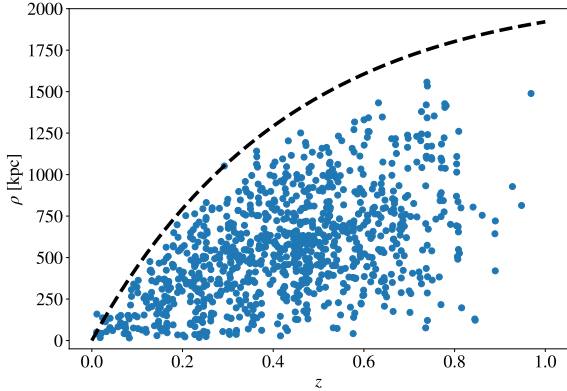


Figure 8. The distribution of impact parameters as a function of redshift. The grey curve approximates the distance to the edge of the detector in the 5.5' GMOS FOV assuming the QSO is in the center to highlight the survey coverage as a function of redshift. Although most of the QSOs are centered in the FOV, in a few cases they had to be offset to avoid bright stars.

for detection of the H I lines. The covering fraction is the fraction of hits versus the total number of CGM systems above the threshold in a given bin, $f_c = H/(H + M)$.

We choose a threshold of $N_{\text{HI}} \geq 10^{14} \text{ cm}^{-2}$ for the covering fraction calculations which is supported by previous survey work. In particular, [Chen et al. \(2005\)](#) find that H I column densities below $10^{13.6} \text{ cm}^{-2}$ are consistent with being randomly distributed with respect to known galaxies. They also show the correlation of galaxies and absorbers does not depend sensitively on N_{HI} for strong absorbers $N_{\text{HI}} > 10^{13.6} \text{ cm}^{-2}$. Furthermore, a value of $N_{\text{HI}} \geq 10^{14} \text{ cm}^{-2}$ was shown in [Tejos et al. \(2014\)](#) to be more highly correlated with galaxies than gas at lower column densities. Adopting a threshold of $N_{\text{HI}} \geq 10^{14} \text{ cm}^{-2}$ additionally provides us with a sample against which we can compare to the cumulative column density distributions of H I systems found in [Danforth et al. \(2016\)](#), who also use this value.

4.3. The Empirical H I-Galaxy connection

This section presents an empirical analysis of the H I-galaxy connection. In the following analysis we generally avoid differentiating the systems via their spectroscopic galaxy classifications due in part to the fact that we do not expect to see differences in covering fractions of H I in star forming and quiescent galaxies ([Thom et al. 2012](#); [Keeney et al. 2017](#)) but include it in Figure 9 to illustrate this. Additionally, Figures 6 and 5 show that the galaxy spectral classification is correlated with the galaxy stellar mass, and thus we have relatively few galaxies with comparable masses but with differing SF classification.

Future analyses that examine the metal ions present in the CGM will focus on galaxy spectroscopic type.

In Figure 9, we present the H I column density (left axis) and covering fraction (right axis) as a function of stellar mass for galaxies within 300 kpc (top panel) and $1.5R_{\text{vir}}$ (bottom panel). The covering fraction for each bin is shown as a dotted line with the grey boxes corresponding to the 68% binomial confidence intervals. At $R < 1.5R_{\text{vir}}$, f_c remains consistent with $f_c \gtrsim 0.5$ at all ranges in mass. We also notice a trend of increasing covering fraction and column density as a function of galaxy stellar mass.

The H I column densities and impact parameters are shown in Figure 10 as an anti-correlation of column density with increasing separation between the galaxy and QSO sightline. Figure 10 separates our galaxy sample into three mass bins, each containing approximately equal numbers of Galaxy-Absorber systems. The top panel contains 191 systems with $M_* > 10^{9.9} M_\odot$. The middle panel also contains 191 systems with $10^{9.2} < 10^{10} M_*/M_\odot < 10^{9.9}$ while the bottom panel corresponds to a sample of 190 systems with $M_* < 10^{9.2} M_\odot$.

We find that for the highest-and-intermediate mass galaxy samples, f_c drops off monotonically but remains $\geq 50\%$ out to an impact parameter of $\rho \simeq 400$ kpc, corresponding to $R \simeq 2R_{\text{vir}}$ for these masses, before flattening out in the case of the intermediate mass sample. In the lowest-mass regime, we find lower ($f_c \sim 60\%$) covering fractions at the smallest separations. We also observe a shallower profile in covering fraction out to $R \simeq 2R_{\text{vir}}$, beyond which the covering fraction remains elevated and consistent with $\sim 50\%$ out to 600 kpc, or $5R_{\text{vir}}$.

The extended distribution seen in the lower two panels, where the covering fraction remains elevated, could imply that the galaxies are located inside (or close to) the halos of other (massive) galaxies (e.g. [Burchett et al. 2016](#)). In an upcoming CGM² paper, we plan to explore the impact of galaxies' environments on the properties of their CGM. Briefly, we find that 56% (243/435) of galaxies with $M_* < 10^{10} M_\odot$ have $M_* > 10^{10} M_\odot$ neighbors within 300 kpc and 1000 km/s of the sightline. We observe 75% (86/121) of galaxies with $M_* > 10^{10} M_\odot$ have neighbors above this mass threshold within the same physical window, consistent with the fact that larger galaxies are more highly clustered. Furthermore, we generally see that low mass galaxies with nearby massive neighbors tend to have elevated HI covering fractions out to $\sim 2-3 R_{\text{vir}}$ compared to galaxies with no detected massive neighbors. In tandem, these effects support our conclusion that the elevated covering fractions out to

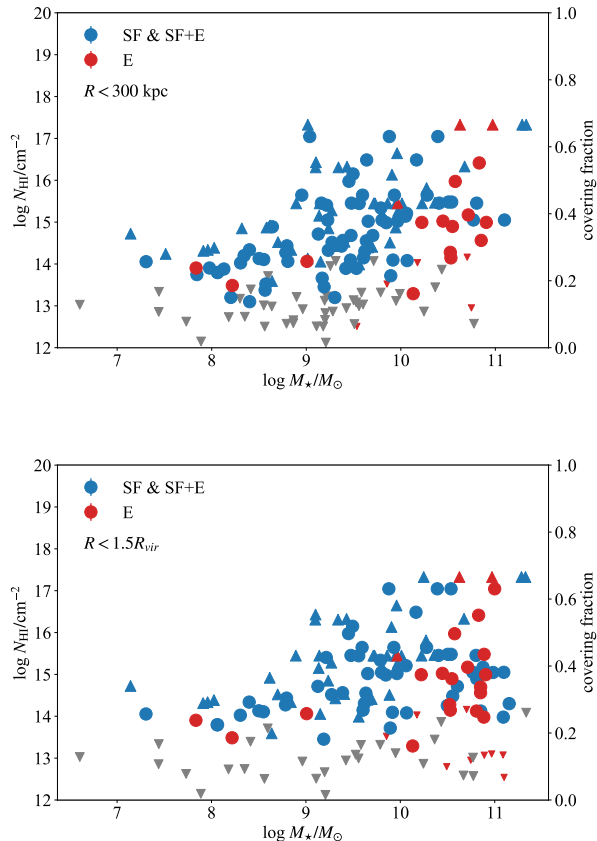


Figure 9. Column density as a function of mass for $\rho < 300$ kpc (top) and $R < 1.5R_{\text{vir}}$ (bottom). CGM systems with saturated absorption are marked with upward facing triangles, while non-detections are displayed as lighter, downward facing triangles at their corresponding 2σ upper limits. Circles represent CGM systems with measured N_{HI} . Measured 1σ uncertainties in the column density of the detected CGM systems are shown as lines inside the markers. Red markers indicate a spectroscopically-determined quiescent galaxy classification ‘E’, while blue corresponds to those galaxies with emission lines present in their spectra, classified as ‘SF’ and ‘SF+E’. Our sample of quiescent galaxies predominately reside in the highest-mass bin. Covering fractions f_c are plotted with respect to the right axes and are calculated without differentiating spectroscopic galaxy categories. The grey boxes correspond to the binomial confidence interval of the covering fraction ($N_{\text{HI}} > 10^{14} \text{ cm}^{-2}$) with the mean f_c in each bin denoted with a dotted line. The column density increases as a function of mass while the covering fraction remains greater than $f_c > 0.5$ for galaxies with masses of $M_* = 10^{8-11} M_\odot$.

large impact parameters for our $M_* < 10^{10} M_\odot$ sample are consistent with environmental effects.

Our results are consistent (within 2σ) with those presented by [Wakker & Savage \(2009\)](#) who find that Ly α absorbers at $z < 0.017$ with equivalent width > 50

mÅ ($N_{\text{HI}} \simeq 10^{13} \text{ cm}^{-2}$) have covering fractions of 100% at $\rho < 400$ kpc. Similarly, [Prochaska et al. \(2011\)](#) find high covering fractions out to $\rho = 300h_{72}^{-1}$ kpc for absorbers with equivalent widths > 30 mÅ. This observed enhancement of covering fraction in the low-mass sample could also be due in part to the fact that our CGM systems were defined to have separations $|\delta v| < 500 \text{ km s}^{-1}$, which is around the expected escape velocities of high-mass galaxies but is more likely to encompass unassociated absorbers in the low mass sample that trace the cosmic web. However, low-mass galaxies have smaller escape velocities and shallower potential wells, and thus would likely exhibit gas being ejected at larger velocities. There are clearly many competing effects in this mass range.

Figure 11 shows the covering fraction and confidence intervals for the total sample at $N_{\text{HI}} \in 10^{13-15} \text{ cm}^{-2}$ (blue, green, yellow). For the lowest column density threshold in the low-mass sample, the H I covering fractions remain elevated at $\sim 80\%$ out to at least $4R_{\text{vir}}$. This signal must be dominated by galaxy-galaxy clustering as the f_c of random incidence in a velocity window at the mean redshift is $f_c = 0.14$ for $N_{\text{HI}} > 10^{13} \text{ cm}^{-2}$. (see Section 5). In contrast, stronger absorbers preferentially occur at smaller impact parameters. For example, it is at $\log N_{\text{HI}} \gtrsim 14.5$ that the covering fraction drops below 50% by $R = R_{\text{vir}}$.

4.4. H I velocity offsets

Here, we examine the kinematics of the H I absorption in our CGM² sample of $z < 0.481$ absorbers that are associated with galaxies. Velocity distributions of CGM absorbers quantify the amount of material gravitationally associated with the assumed host halo. For simplicity and clarity, we trim the sample to those galaxies closest to the absorber impact parameter (ρ), thus each absorber is associated with only one galaxy. We are left with a sample of 171 unique galaxy-absorber pairs at $z < 0.48$ that were detected with a signal $> 2\sigma$.

In Figure 12, we show the distribution of velocity centroids of the 416 detected absorption components associated with these 181 CGM systems in the rest-frame of the galaxy systemic redshift as a function of column density. The boxes show the quartiles of distribution in velocity while the whiskers show the extent of the distribution. Outliers beyond 1.5 times the innerquartile range are displayed as points. We see that systems with $N_{\text{HI}} < 10^{14.0} \text{ cm}^{-2}$ have a higher median and larger spread in velocity. The component velocity centroids do not exhibit any clear trends with impact parameter. Relevant to this discussion, we recall that absorption components are required to lie within $\pm 1000 \text{ km s}^{-1}$ of

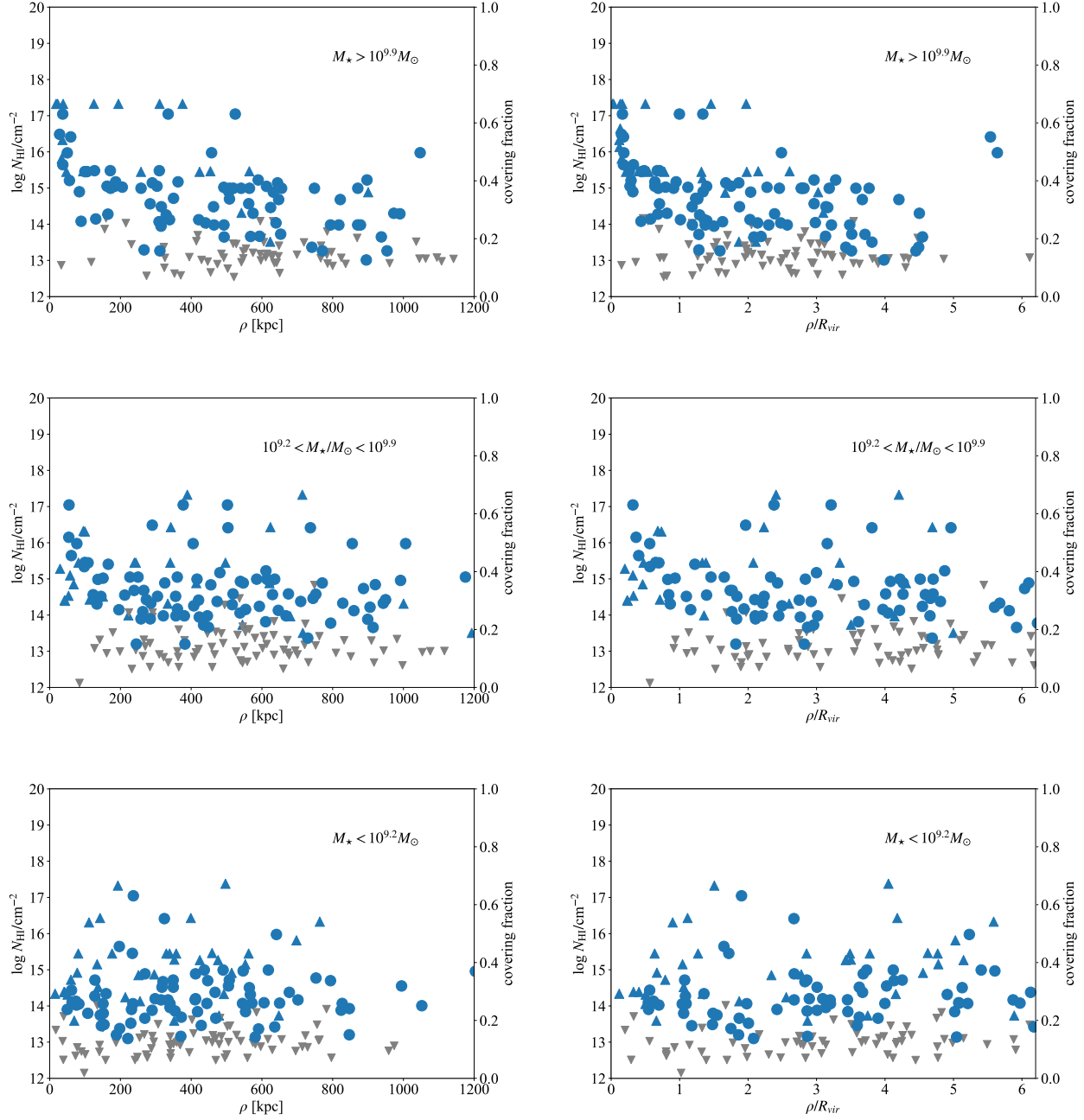


Figure 10. Column density N_{HI} as a function of impact parameter, ρ (left) and R_{vir} (right). The panels are arranged vertically in order of decreasing mass. Galaxy-absorber systems with dark circles represent galaxies with N_{HI} detections in the corresponding QSO spectrum. Systems with dark, up-arrow symbols show our inferred lower limits due to saturation in N_{HI} . Similarly, systems with gray, down-arrow symbols denote the 2σ upper limit on N_{HI} (non-detections). Measured 1σ uncertainties in the column density of the detected CGM systems are shown as grey lines but are smaller than the marker size in every case. N_{HI} covering fraction, f_c , corresponds to the right axes in bins of 100 kpc (left) and 1 R_{vir} (right). The dotted lines represent f_c assuming a detection threshold of $N_{\text{HI}} = 10^{14} \text{ cm}^{-2}$ with the 68% binomial confidence interval shown as shaded gray regions about the mean f_c .

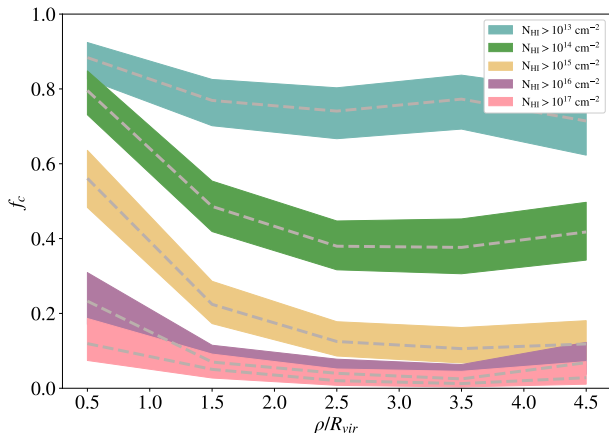


Figure 11. Covering fraction of H I as a function of ρ/R_{vir} for column density thresholds of $N_{\text{HI}} > 10^{13} \text{ cm}^{-2}$ (blue), $N_{\text{HI}} > 10^{14} \text{ cm}^{-2}$ (green), $N_{\text{HI}} > 10^{15} \text{ cm}^{-2}$ (yellow), $N_{\text{HI}} > 10^{16} \text{ cm}^{-2}$ (purple), and $N_{\text{HI}} > 10^{17} \text{ cm}^{-2}$ (pink). Shaded regions represent the $1\text{-}\sigma$ (68%) binomial confidence intervals. Here we connect the center of the radial bins to highlight the difference in the distributions. We see that for column densities less than 10^{14} cm^{-2} show little correlation with galaxies. The covering fraction at $R < R_{\text{vir}}$ in for the highest column densities ($N_{\text{HI}} > 10^{15} \text{ cm}^{-2}$) never gets higher than 0.7.

each other. The only velocity constraint imposed with respect to the galaxy is that at least one of these absorption components must lie within $\pm 500 \text{ km s}^{-1}$ of a CGM² galaxy systemic redshift in order to be classified as a CGM absorption system.

We find that $> 53\%$ of the detected H I components are located within $\pm 250 \text{ km s}^{-1}$ of the galaxy systemic redshift, independent of H I column density. Furthermore, only 27% of all absorption components lie at $|\delta v| > 500 \text{ km s}^{-1}$. Finally, approximately 83% of the total H I column density in the CGM² survey lies within $\pm 250 \text{ km s}^{-1}$ of the associated host galaxy, consistent with the results of COS-Halos (Tumlinson et al. 2013). Some absorption systems show a total extent $> 500 \text{ km s}^{-1}$, which likely captures both bulk motions of galaxies and possible “missassociations” due to incompleteness of our galaxy survey, especially at magnitudes > 23 .

Were absorption components distributed uniformly in velocity space relative to the galaxies, limited only by the 1000 km s^{-1} window selection function, we would not expect to find such a high concentration of absorbers at low δv . To first order, Figure 12 shows that our CGM survey is capturing gas that is mostly not exceeding the escape velocity of its host halo, where the escape velocity of an $M_{\star} \approx 10^{9.5} M_{\odot}$ galaxy at R_{vir} is $\sim 250 \text{ km s}^{-1}$.

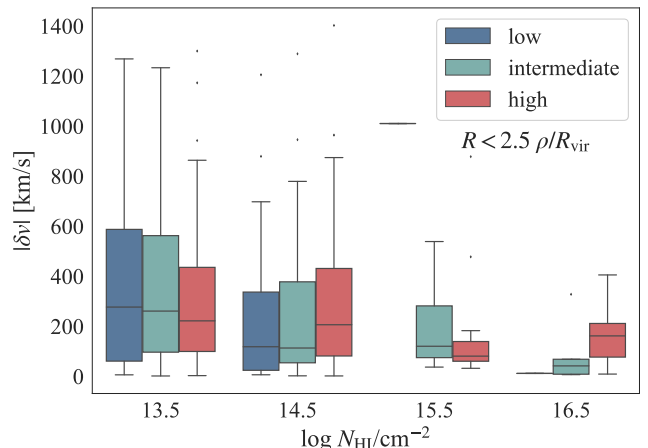


Figure 12. The distribution of absolute velocities as a function of N_{HI} displayed in a box and whisker plot for each mass sample. The boxes display the quartiles of the distribution centered at each bin spanning 1 decade in column density while the whiskers extend to show the rest of the distribution of the bins. Outliers are defined as points that lie outside 1.5 times the innerquartile range and are displayed as small diamonds. Each bin is split into a high (blue), intermediate (green) and low-mass (red) sample. We see a strong anticorrelation between velocity spread and column density. The single line seen at higher column densities in the low-mass sample indicate there are only one point in each column density bin. This highlights how rare high-column density systems are around low-mass galaxies.

There is very clearly a gravitational association between the galaxies and the absorption at higher column densities. The component velocity centroids do not exhibit a clear trend with impact parameter but we find that the median velocity for components at $\rho > R_{\text{vir}}$ to be 233 km s^{-1} with a standard deviation of 453 km s^{-1} while components at $\rho < R_{\text{vir}}$ have a median velocity of 185 km s^{-1} with a standard deviation of 354 km s^{-1} . In this unique galaxy-absorber pairing framework, 80% of the absorption components with $N_{\text{HI}} > 10^{15} \text{ cm}^{-2}$ lie within $1 R_{\text{vir}}$. We caution that for this kinematic analysis we paired specific galaxies with absorbers on the criterion that they were the closest in physical separation from the absorption sightline. Without this bias, the trend remains, however.

Lastly, Figure 12 also shows the column-density dependence of the absorption component velocity distribution. Here, we see that absorption systems with $N_{\text{HI}} > 10^{14} \text{ cm}^{-2}$ concentrate more strongly at low $|\delta v|$ than weaker systems, and there is a clear column density dependence to the overall velocity concentration at $|\delta v| < 250 \text{ km s}^{-1}$. Counting by column density we find that 53% of the total fitted column density lies within

$\pm 250 \text{ km s}^{-1}$. In contrast, we find that only 44% of the $N_{\text{HI}} < 10^{14} \text{ cm}^{-2}$ absorption components lie at $|\delta v| < 250 \text{ km s}^{-1}$, while 67% of the $N_{\text{HI}} > 10^{14.0} \text{ cm}^{-2}$ absorption components lie at $|\delta v| < 250 \text{ km s}^{-1}$. This concentration increases with increasing column density. For a limit of $N_{\text{HI}} = 10^{15} \text{ cm}^{-2}$, the percentages of components within $|\delta v| < 250 \text{ km s}^{-1}$ shift to 50% and 86% for the low and high column density thresholds, respectively. Similarly, the high velocity components tend to have low column density. 27% of all absorption components lie at $|\delta v| > 500 \text{ km s}^{-1}$, but less than 5% of those components have $N_{\text{HI}} > 10^{14} \text{ cm}^{-2}$. Generally, there is no systematic effect preventing high column density components from appearing at higher velocities, so this trend captures an important characteristic of the CGM.

5. HI-GALAXY CLUSTERING: R_{CGM}^{14}

5.1. Setup

In order to quantitatively describe the radial dependence of the CGM and estimate its extent, we perform an absorber-galaxy cross-correlation analysis. We aim to measure the excess probability of detecting an absorber given the proximity of a galaxy over the proximity-agnostic average rate. Our analysis of the HI-galaxy clustering is similar to the one developed by Hennawi & Prochaska (2007) and follows more closely the analysis by Prochaska et al. (2019). We define the 3D cross-correlation function, $\xi_{ag}(r)$ as

$$\xi_{ag}(r) = \left(\frac{r}{r_0}\right)^{-\gamma}. \quad (1)$$

In order to determine the best fitting parameters, r_0 and γ , we define a likelihood function as

$$\mathcal{L} = \prod_i P_i^{\text{hit}}(r, z) \prod_j P_j^{\text{miss}}(r, z), \quad (2)$$

where P^{hit} is defined to be the probability of detecting one or more H I systems and P^{miss} is the probability of detecting none. This probability has both a radial and redshift dependence. An absorber is considered a “hit” if it falls within our window of $\delta v = \pm 500 \text{ km s}^{-1}$ of a galaxy and we measure a column density above a threshold $N_{\text{HI}}^{\text{thresh}}$. We define P^{miss} to be the probability of observing zero events from a Poisson distribution where the rate is the number of events expected from the average density of absorbers and our clustering term:

$$P^{\text{miss}} = \exp(-[1 + \chi_{\perp}(r)] \langle \frac{dN}{dz} \rangle \delta z). \quad (3)$$

Here $\langle dN/dz \rangle \delta z$ is the mean number of absorbers in a window of redshift δz . $[1 + \chi_{\perp}(r)]$ represents an excess in

the number of absorbers due to clustering. This boost due to clustering can be expressed in terms of the 3-dimensional correlation function as

$$\begin{aligned} \chi_{\perp}(r) &= \frac{1}{V} \int_V \xi_{ag}(r) dV \\ &\approx \frac{aH(z)}{2\delta v} \int_{-\delta v/[aH(z)]}^{-\delta v/[aH(z)]} \left(\frac{\sqrt{R_{\perp}^2 + R_{\parallel}^2}}{r_0} \right)^{-\gamma} dR_{\parallel} \end{aligned} \quad (4)$$

where we are integrating Equation (1) along the length of a cylinder of length $2\delta v/aH(z)$. Here $a = 1/(1+z)$ and $H(z)$ is the Hubble parameter. The probability of a “hit,” P^{hit} , is the complement of the probability of a “miss”: $P^{\text{hit}} = 1 - P^{\text{miss}}$. P^{hit} is equivalent to the covering fraction f_c . The covering fraction of a random sightline is then:

$$f_c = 1 - \exp\left(-\left\langle \frac{dN}{dz} \right\rangle \delta z\right). \quad (5)$$

We take $\langle dN/dz \rangle$ from Danforth et al. (2016), who measured the occurrence of 5138 individual extragalactic absorption lines of H I in 82 QSO/AGN spectra at redshifts $z_{\text{AGN}} < 0.85$. The occurrence rate of H I absorbers dN/dz is expressed with a functional form as follows:

$$\frac{dN(N_{\text{HI}} \geq N_{\text{HI}}^{\text{thresh}}, z)}{dz} = C_0(1+z)^{\gamma}, \quad (6)$$

with $C_0 = 16$ and $\gamma = 2.3$. This measurement is valid only for absorbers with $N_{\text{HI}}^{\text{thresh}} \geq 10^{14} \text{ cm}^{-2}$, which coincides with the threshold in our definition of “hits” and “misses.”

In order to estimate r_0 and γ in our 3D correlation function, Equation (1), we follow a Bayesian approach. The posterior probability function can be defined as

$$p(r_0, \gamma | \{k_i, r_i, z_i\}_{i=1}^N) \propto p(r_0, \gamma) p(\mathbf{k} | \mathbf{r}, \mathbf{z}, r_0, \gamma). \quad (7)$$

where $k_i \in \{0, 1\}$ specifies whether system i is a “hit” or a “miss” and $p(\mathbf{k} | \mathbf{r}, \mathbf{z}, r_0, \gamma)$ is the likelihood function $\mathcal{L}$ defined in Equation (2). We define the priors as follows:

$$p(r_0) = \begin{cases} 1/10, & \text{if } 0 < r_0/\text{Mpc} < 10 \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

and

$$p(\gamma) = \begin{cases} \mathcal{N}(\mu = 1.6, \sigma = 1), & \text{if } \gamma > 0 \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

where r_0 is measured in h_{68}^{-1} comoving Mpc and $\mathcal{N}(\mu, \sigma)$ is the normal distribution. These priors were chosen

based on physical arguments and previous results on absorber-galaxy clustering (e.g. [Tejos et al. 2014](#)). We used the Markov Chain Monte Carlo (MCMC) sampler **emcee** ([Foreman-Mackey et al. 2013](#)) to generate samples from the posterior probability distribution function over r_0, γ .

The data were cut as in the previous analysis with a redshift cut of $z < 0.481$ and subdivided into the three equal sized mass samples as before with identical priors used for each sample.

5.2. HI-Galaxy Clustering Results

Figure 13 illustrates the results of the MCMC parameter estimation. The plots on the left show the covering fraction as a function of the perpendicular separation in comoving Mpc.

The grey boxes on the bottom of the plots on the left indicate the covering fraction for sightlines taken at random, $f_c^{rand} = 1 - \exp[-(dN/dz)\delta z]$ at the mean redshift of the sample, $\bar{z} = 0.34$, $\bar{z} = 0.34$ and $\bar{z} = 0.26$ for the high, intermediate and low-mass samples, respectively. The plots on the right further illustrate the marginal distributions of the posteriors of our parameters as well as indicate the median values for each. The covering fraction due a random sightline is $f_{c,rand} = 0.14$.

5.3. Estimating R_{CGM}^{14}

In order to estimate a characteristic size of the $N_{HI} > 10^{14} \text{ cm}^{-2}$ CGM, R_{CGM}^{14} , we devise a method in which we use the parameters to calculate the impact parameter at which the covering fraction ($f_c = P^{hit}$) exceeds 0.5. Within this impact parameter, a sightline has a greater than 50% chance of exhibiting a H I column with $N_{HI} > 10^{14} \text{ cm}^{-2}$. We can then estimate the posterior distribution of R_{CGM} by calculating it for each sample taken from the posterior distributions of γ and r_0 . Using the samples from the posterior distributions in r_0 and γ we calculate $R_{CGM,p}^{14} = 346_{-53}^{+57}$ kpc ($R_{CGM,c}^{14} = 463_{-71}^{+76}$ comoving kpc), $R_{CGM,p}^{14} = 353_{-50}^{+64}$ kpc ($R_{CGM,c}^{14} = 469_{-66}^{+85}$ comoving kpc) and $R_{CGM,p}^{14} = 177_{-65}^{+70}$ kpc ($R_{CGM,c}^{14} = 222_{-81}^{+88}$ comoving kpc) in order of decreasing mass samples. The extent of the CGM remains relatively similar in size (~ 350 physical kpc) except for the lowest mass sample. These correspond to $R_{CGM}^{14} = 1.2_{-0.2}^{+0.2} R_{vir}$, $R_{CGM}^{14} = 2.4_{-0.3}^{+0.4} R_{vir}$ and $R_{CGM}^{14} = 1.6_{-0.6}^{+0.6} R_{vir}$ for the mass samples in order of decreasing mass, respectively, where R_{vir} was calculated using the mean redshift and mass of each sample. These estimates are in agreement with our qualitative empirical estimates from the previous analysis. These results are summarized in Table 5.

6. COMPARISON WITH OTHER WORK

Table 5. Results

M_*	$R_{CGM,p}^{14}$	$R_{CGM,c}^{14}$	R_{CGM}^{14}	$\bar{z}$
[$M_\odot$]	[kpc]	[kpc]	[R_{vir}]	
(1)	(2)	(3)	(4)	(5)
$M_* > 10^{9.9}$	346_{-53}^{+57}	463_{-71}^{+76}	$1.3_{-0.2}^{+0.2}$	0.34
$10^{9.2} < M_* < 10^{9.9}$	353_{-50}^{+64}	469_{-66}^{+85}	$2.4_{-0.4}^{+0.4}$	0.33
$M_* < 10^{9.2}$	177_{-65}^{+70}	222_{-81}^{+88}	$1.6_{-0.6}^{+0.6}$	0.26

NOTE—Results of H I-galaxy clustering analysis and R_{CGM}^{14} : (1) Stellar Mass limits of each sample; (2) R_{CGM}^{14} in physical kpc; (3) R_{CGM}^{14} in physical kpc; (4) R_{CGM}^{14} normalized to the average virial radius of the sample; (5) Average redshift of each sample

6.1. Previous Surveys

CGM² contains the largest sample by a factor of 5 of CGM absorption systems within ~ 600 kpc of putative host galaxies. Our increased coverage is partly due to the fact that we probe a larger volume than previous surveys by confirming sub-L* galaxies out to higher redshifts. For this reason, CGM² presents an excellent opportunity to study the HI-traced CGM at galaxy-absorber separations that exceed R_{vir} for a wide range of stellar masses. Here, we compare our results to those from other observational surveys of the CGM.

Figures 13 and 14 and Table 5 provide a summary of our covering fraction results for our mass segregated galaxy samples. The red, green, and blue-shaded curves in Figure 14 correspond to the samples cut at masses of $M_{*,cut} = 10^{9.204}, 10^{9.888} M_\odot$ as described above. This figure allows us to better compare the covering fraction profiles of each mass sample to each other. As we saw in the last section, we find a mass dependence on the extent of the CGM; the covering fraction remains higher for larger galaxy mass when looking at the physical impact parameter. When we consider the impact parameter normalized to the average virial radii of each mass sample, we find the covering fraction remains elevated to larger radii ($\gtrsim 2R_{vir}$) for the intermediate mass sample, while for the other mass samples the extent of the CGM is similar ($\gtrsim 1R_{vir}$). We also find that the covering fraction of the lowest mass sample never exceeds 80% while for both higher mass samples, the covering fraction is consistent with $\sim 100\%$.

Many previous studies have examined the covering fraction of H I as a function of galaxy projected separation, although most of these surveys are limited to $\rho \lesssim 300$ kpc. These surveys are all generally constructed in the same manner, with HST UV spectroscopy with

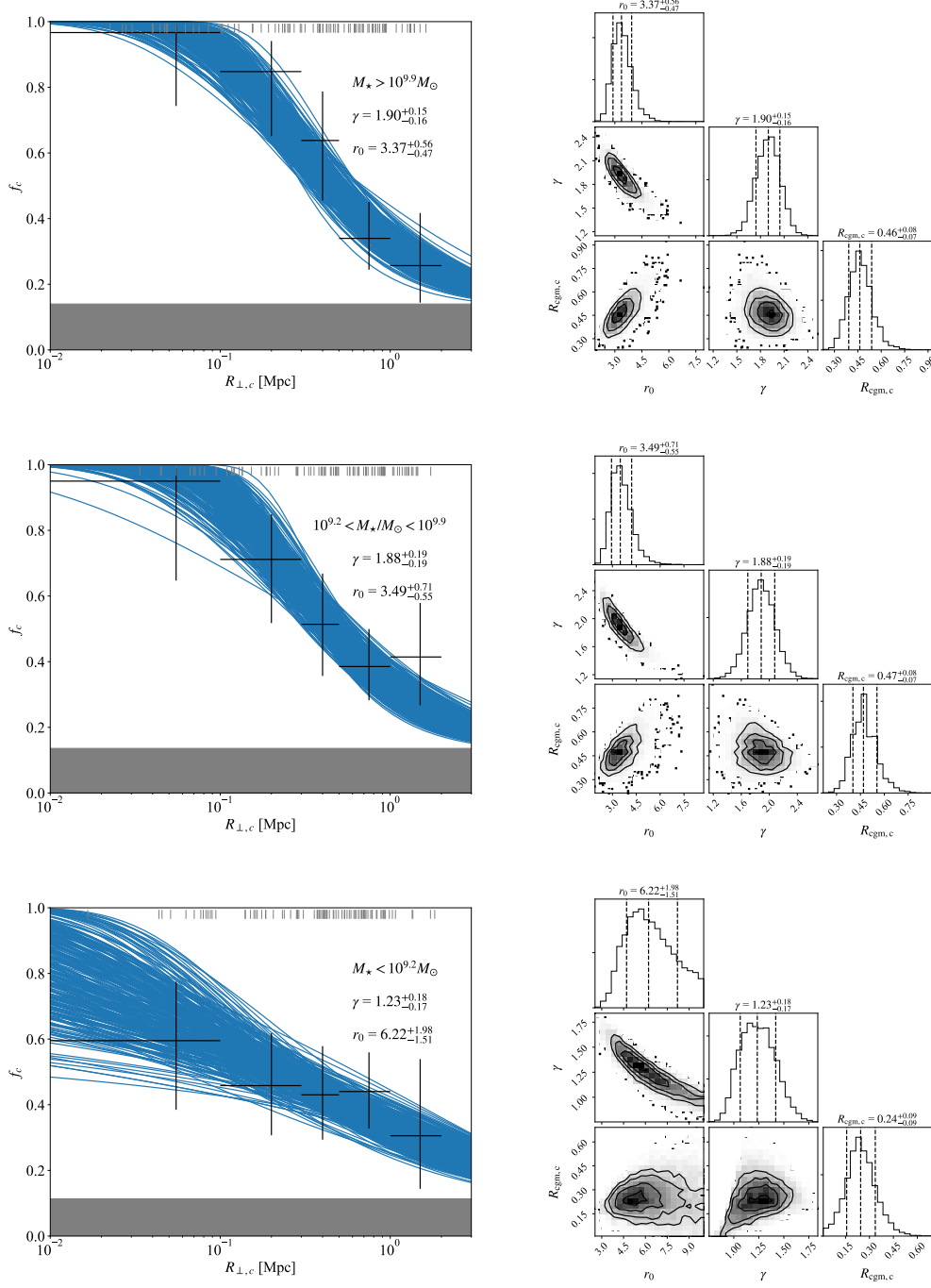


Figure 13. Left: covering fraction of H I gas with $N_{\text{HI}} > 10^{14} \text{ cm}^{-2}$ as a function of the physical impact parameter in comoving Mpc for galaxies in order of decreasing mass. The black crosses are binned evaluations of the covering fraction in each bin with 95% binomial confidence limits on covering fraction. The small grey ticks at the top and bottom of the figures indicate impact parameter for those systems that were *hits* (top) and *misses* (bottom). Beneath the black crosses are samples drawn from the posterior distribution showing the range in the γ - r_0 parameter space. The grey boxes on the bottom of the plots on the left indicate the covering fraction for sightlines taken at random, $\langle dN/dz \rangle \delta z$. The plots on the right further illustrate the one and two dimensional projections of the posterior probability distributions of our parameters as well as indicate the median values for each. The highest mass sample contains 191 CGM systems at mean redshift of $\bar{z} = 0.34$ and a random sightline covering fraction of $f_{c,\text{rand}} = 0.14$. The intermediate-mass sample contains 191 CGM systems at a mean redshift of $\bar{z} = 0.33$ and a random covering fraction of $f_{c,\text{rand}} = 0.14$. The lowest mass sample contains 190 galaxies at a mean redshift of $\bar{z} = 0.26$ and a random covering fraction of $f_{c,\text{rand}} = 0.14$. Using the samples from the posterior distributions in r_0 and γ we calculate $R_{\text{CGM},p}^{14} = 346^{+57}_{-53}$ kpc, $R_{\text{CGM},p}^{14} = 353^{+64}_{-50}$ kpc and $R_{\text{CGM},p}^{14} = 177^{+70}_{-65}$ kpc in order of decreasing mass samples. The extent of the CGM drops off in the lowest mass sample.

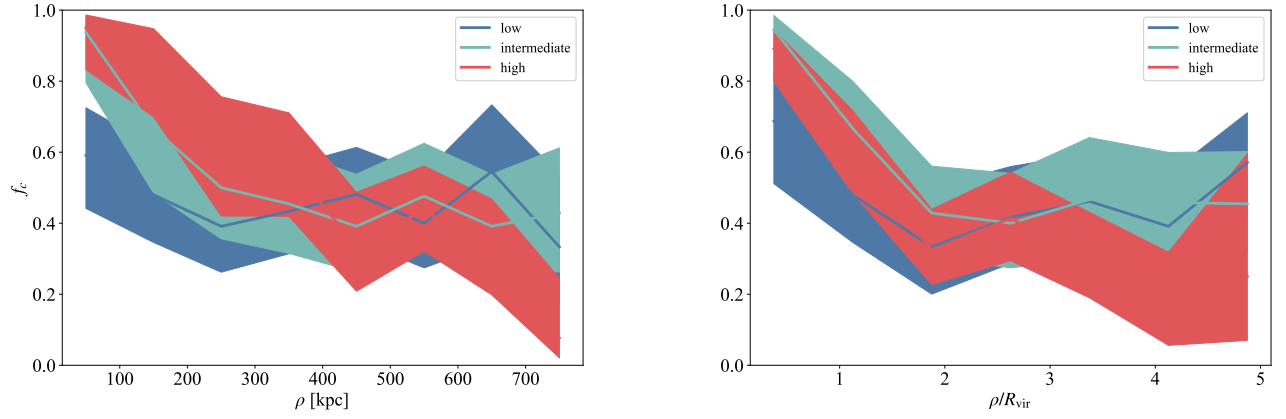


Figure 14. Covering fraction of H I as a function of impact parameter ρ (left) and ρ/R_{vir} (right) highlighting $N_{\text{HI}} > 10^{14} \text{ cm}^{-2}$ broken down into the three mass samples. The solid lines correspond to the covering fraction in bins of 100 physical kpc (left) and $1 R_{\text{vir}}$ (right). The shaded region encodes the 1σ error in the covering fraction measurement. The covering fraction decreases monotonically in either measure to $\sim 300 \text{ kpc}$ or $2R_{\text{vir}}$. We find that the covering fraction in the lowest radial bins is consistent with $\sim 100\%$ in each case except for the lowest mass sample. We note that the intermediate mass sample extends to further radius when looking at the normalized impact parameter (right) as compared to either of the other mass samples.

COS and/or STIS and a spectroscopic galaxy catalog to connect galaxies and absorbers.

One of the largest such surveys was carried out by [Tejos et al. \(2014\)](#) who attempted to explore the connection between the IGM and galaxies by measuring the HI-galaxy cross-correlation at $z < 1$ at distances between ~ 1 and 10 Mpc. They used multiple ground based instruments to build a new spectroscopic survey of 2143 galaxies in 8 QSO fields with HST spectroscopy. They combined their catalog with existing catalogs to build a survey of ~ 17500 galaxies. They did not limit their galaxy survey to only the nearest galaxies but used a statistical approach to examine the cross-correlation of galaxies and absorption systems. They found that the H I Ly α forest to be divided into two main categories: a population of low column density absorbers tracing the cosmic web and a higher-column density population that traces the dark matter halos in which galaxies reside. CGM² has the benefit of being more sensitive to fainter galaxies closer to the QSO itself allowing us to better constrain the extent of the CGM. However, their larger galaxy sample at larger separations (which explores different scales) presents a great opportunity to compare our cross-correlation analysis. Using the same 3D cross-correlation power law, Equation 1, they found $r_0 = 3.8 \pm 0.2 h_{70}^{-1}$ Mpc and $\gamma = 1.7 \pm 0.1$ in their sample of SF-galaxies. This sample is similar to our high-mass sample since we are dominated by SF galaxies and indeed, we find agreement consistent within 1σ .

A mass dependent CGM was examined in [Bordoloi et al. \(2018\)](#) who used the 85 galaxies in the COS-Halos and COS-Dwarfs surveys at $z \sim 0$ with M_* ranging from 8 to $11.6 \log M_*/M_\odot$. This sample was limited to impact parameters ($\rho < 160$ kpc). They found a mass and radius dependence of the strength of H I absorption where the equivalent width of H I increases with M_* and with decreasing impact parameter. Here we are primarily focused on the mass and radius dependence of the covering fraction, but find trends generally consistent with those observed by [Bordoloi et al. \(2018\)](#).

[Chen et al. \(2001\)](#) found Ly α with column densities $N_{\text{HI}} \gtrsim 10^{14} \text{ cm}^{-2}$ in 34/47 galaxies ($f_c \approx 0.7$) out to $\rho \simeq 330$ kpc. Their sample consists of $\sim L^*$ HI-Galaxy pairs with $|\delta v| < 500 \text{ km s}^{-1}$ spanning $0.1 < z < 0.9$ ($\bar{z} = 0.36$). They also found a sharp decline thereafter. We find close agreement $f_c = 0.78^{+0.06}_{-0.89}$ (162/209) when applying the same criterion to our sample also seeing a sharp decline around 400 kpc. (Figure 14).

[Prochaska et al. \(2011\)](#) used 14 QSO sightlines with previously published equivalent width (W_0) measurements of Ly α to carry out a galaxy survey that targeted 37 $L > 0.01L_*$ galaxies at $\bar{z} = 0.18$. They con-

nected absorbers and galaxies with $|\delta v| < 400 \text{ km s}^{-1}$, although they show it makes no qualitative difference in their results using a larger velocity window of $|\delta v| < 600 \text{ km s}^{-1}$. They found covering fractions of order unity ($\approx 90\%$) for $N_{\text{HI}} > 10^{13} \text{ cm}^{-2}$ gas out to $\rho = 300$ kpc. Comparing our covering fractions with $N_{\text{HI}} > 10^{13} \text{ cm}^{-2}$ and for galaxies with $M_* > 10^{8.55} M_\odot$ to approximate their $L > 0.01L_*$ sample, we find $f_c = 0.85^{+0.04}_{-0.05}$ (128/150), which is roughly consistent with their value.

[Wakker & Savage \(2009\)](#) conducted a large survey of the HI-galaxy connection at $z \lesssim 0.017$, consisting of 76 QSO sightlines and ~ 20000 local galaxies. They found covering fractions of 77% for Ly α absorbers $> 50 \text{ mÅ}$ within $\rho < 400$ kpc and $|\delta v| < 400 \text{ km s}^{-1}$ of $L > 0.1L_*$ galaxies. If we limit our sample to $L > 0.1L_*$ galaxies and use a Ly α absorber threshold of $\sim 50 \text{ mÅ}$ ($N_{\text{HI}} \sim 10^{13} \text{ cm}^{-2}$), we find covering fractions, $f_c = 0.89^{+0.04}_{-0.05}$ (99/111) for galaxies within 400 kpc. This discrepancy could imply that covering fractions increase with redshift.

Most recently, a large survey of H I was carried out by [Keeney et al. \(2018\)](#) (K18). Their survey consisted of 47 COS sightlines (COS GTO) with higher signal-to-noise ($S/N \sim 15\text{--}50$ compared to our ~ 10) QSO spectra. Using ground based telescopes, they constructed a spectroscopic galaxy database of $\sim 9,000$ galaxies with the aim of $> 90\%$ completeness to 1 Mpc down to $0.1L_*$ at $z \lesssim 0.1$. Due to their higher S/N , they could consistently measure weaker absorption lines, down to $N_{\text{HI}} \geq 10^{12.8} \text{ cm}^{-2}$. Leveraging the high completeness of K18 at low redshifts, they were able to measure the H I column densities out to $4R_{\text{vir}}$ with enough galaxies in this range (243) to make precise statements about the radial profile of HI. They find the covering fraction for $L < L^*$ galaxies (corresponding to our low-mass sample) to feature a shallower decline than that of their $L > L^*$ sample. Qualitatively, we show similar results. This difference in covering fraction behavior between high and low mass samples can readily be seen in Figures 11 and 14. The elevated covering fractions at large radii in the $L < 0.1L_*$ sample imply that contributions to $N_{\text{HI}} > 10^{14} \text{ cm}^{-2}$ CGM gas are dominated by low-mass galaxies. This result is consistent with [Prochaska et al. \(2011\)](#) in relation to H I and with high-ionization metals such as O VI ([Tumlinson & Fang 2005](#); [Pratt et al. 2018](#); [Prochaska et al. 2019](#)).

We note that the high mass sample drops to a very low covering fraction $f_c \simeq 0.20$ at $4R_{\text{vir}}$ while in the lower mass sample, f_c remains elevated. At radii greater than $3R_{\text{vir}}$, we may be limited to only higher- z high-mass galaxies due to the detector size but this detector size bias should not affect the lower mass samples.

In addition, previous low-redshift studies have found that low column density gas ($N_{\text{HI}} = 10^{13-14} \text{ cm}^{-2}$) is likely uncorrelated with galaxy halos (Chen et al. 2005; Prochaska et al. 2011; Danforth et al. 2016) (but, see Tejos et al. (2014) who find that 50% of weak lines can still be correlated with galaxies on 1-10 Mpc and Tripp et al. (1998) who show that the weakest absorbers are not randomly distributed). First, low column density material exhibits high covering fractions out to 1 Mpc which we can see in Figure 11. Second, low column density material exhibits less velocity correspondence with the systemic velocities of galaxies nearby in projection (Tumlinson et al. 2013), as we find a median velocity difference of 233 km s^{-1} with a standard deviation 453 km s^{-1} for $\rho > 1R_{\text{vir}}$ while we find a median velocity of 185 km s^{-1} with a standard deviation of 354 km s^{-1} for $\rho < 1R_{\text{vir}}$. Traditionally, such low column density material is attributed to the Ly α forest, or to gas in a filament like structure (Tejos et al. 2014), physically distinct from the CGM. This was examined in greater depth by Burchett et al. (2020a) who conclusively tie the diffuse IGM to the cosmic web. They find the H I absorption signature decreases past $\rho > R_{\text{vir}}$ and settles to the cosmic mean matter density.

Turning to higher redshifts, Rudie et al. (2012) use the ground based *Keck Baryonic Structure Survey; KBSS* to investigate the $z \sim 2 - 3$ CGM surrounding 886 galaxies. Their sample contains 48 galaxies at $\bar{z} = 2.3$ within $\rho \lesssim 300 \text{ kpc}$ for which they measure a covering fraction of $f_c = 0.81 \pm 0.06$ for absorbers with $N_{\text{HI}} > 10^{14} \text{ cm}^{-2}$. By comparison, if we choose a mass range of $10^{10.4} < M_*/M_\odot < 10^{11}$ to approximate their mass distribution (Erb et al. 2006) we find $f_c = 0.86_{-0.14}^{+0.08}$ (18/22) which is good agreement with their results. However, looking at the extended CGM $\rho < 1\text{Mpc}$, we find a discrepancy: we measure $f_c = 0.49_{-0.08}^{+0.08}$ (40/83) vs their 0.70 ± 0.03 . These numbers are in agreement at the 2σ level, however. Interestingly, the extended CGM may show a decrease in covering fraction as the universe evolves. This phenomenon may be due to the development of virial shocks that ionize the gas, as suggested by Burchett et al. (2018). At $z = 2.3$, $300\text{kpc} \sim 2R_{\text{vir}}$ for a $7 \times 10^{10} M_\odot$ galaxy while at $z = 0.3$ (CGM²) $300 \text{ kpc} \sim 0.8R_{\text{vir}}$. Alternatively, this could simply be due to the fact that a column density of $N_{\text{HI}} > 10^{14} \text{ cm}^{-2}$ traces lower density peaks at high- z .

6.2. Comparison with Hydrodynamical Simulations

We now turn to a brief comparison with hydrodynamical simulations. van de Voort et al. (2019) simulate a roughly $z \sim 0$ L^* galaxy using a new refinement technique to better resolve the CGM. They find an increase

in the H I column density and resultant covering fraction when the resolution is increased to resolve 1 kpc scales. We find their model to be in good agreement with our $N_{\text{HI}} \geq 10^{14} \text{ cm}^{-2}$ covering fraction measurements when we limit our sample to $M_* \sim 10^{10.5} M_\odot$ (see their Figure 3). Comparing our column density measurements to van de Voort et al. (2019) and Hummels et al. (2019) we also see good agreement out to their limiting distance of 200 kpc and 100 kpc, respectively (compare Figure 10 to van de Voort et al. (2019) Figure 2). Our high H I covering fractions and column densities at $R < R_{\text{vir}}$ are in conflict with earlier simulations that consistently underpredict the column density of low-ions in the CGM (e.g., Hummels et al. 2013; Liang et al. 2016; Stinson et al. 2012; Shen et al. 2012; Ford et al. 2013), validating the work that has gone into creating these new high resolution techniques. To understand the extent of the CGM around a diverse sample of galaxies, we encourage future simulations extending to at least $4R_{\text{vir}}$, and covering a larger range of galaxy masses, down to $0.01L^*$.

7. SUMMARY AND CONCLUSIONS

We have reported the first results from the CGM² survey, a comprehensive survey of the $z < 1$ CGM at least $5\times$ larger than previous surveys such as COS-Halos at comparable redshifts. This paper has presented the detailed properties of the survey design and the procedures followed in the collection and processing of the data. We present an H I study that combines high-resolution HST/COS UV spectra of 22 background QSOs with Gemini/GMOS spectra of 572 foreground galaxies having stellar masses $10^6 M_\odot < M_* < 10^{11.5} M_\odot$ and $z < 0.481$. The S/N ~ 10 -12 of these COS spectra and access to the H I Lyman limit enables us to constrain the H I column densities and kinematics of associated CGM absorption, and to ultimately examine the extent of the CGM as a function of galaxy mass and physical separation from the QSO sightline.

We find that high-column density circumgalactic material is associated with galaxies at high statistical significance out to $2R_{\text{vir}}$, whereas H I absorption with $N_{\text{HI}} < 10^{14} \text{ cm}^{-2}$ is more broadly distributed in both impact parameter and velocity space and may not be associated directly with massive galaxies. Our kinematic analysis reveals that the detected strong H I is most likely gravitationally associated with the most nearby galaxy, while weaker H I components seen at absolute velocity offsets $\gtrsim 500 \text{ km s}^{-1}$ may be instead associated with extended large scale structures. We find generally good agreement between our sample and the prior studies that have examined the CGM of low-redshift galaxies out to similar and larger separations.

We define the cool CGM as the region surrounding a galaxy in which the probability of observing an absorber with $N_{\text{HI}} > 10^{14.0} \text{ cm}^{-2}$ is $> 50\%$. Our column density threshold, $N_{\text{HI}} > 10^{14.0} \text{ cm}^{-2}$, is motivated by previous observational work that examines the statistical, large-scale (Mpc) correlations between galaxies and QSO absorption lines. In essence, our definition of R_{CGM}^{14} demands that around a given galaxy, one is more likely than not to find material that has been empirically associated with galaxies. The picture that emerges is of a diffuse, CGM extending to ~ 350 kpc around galaxies with stellar mass $\gtrsim 10^9 M_{\odot}$ at $z \lesssim 0.5$. For galaxies of lower stellar mass, the extent of the CGM is < 200 kpc. At all stellar masses, the extent of this CGM exceeds a virial radius, especially for galaxies with intermediate masses where the R_{CGM}^{14} exceeds $2R_{\text{vir}}$. Therefore, these results imply that using R_{vir} as a proxy for the characteristic edge of the CGM may significantly underestimate its true extent. The detailed nature of the mass dependence of the CGM will be examined in a forthcoming paper. Additional follow-up studies using CGM² data will consider transitions from a wide range of ionized metals and absorption-line profile analyses to characterize the ionization state, metallicity, kinematics, and mass of the CGM at low redshift.

8. ACKNOWLEDGEMENTS

MCW, JKW, and KT acknowledge the people of the Dkhw'Duw'Absh, the Duwamish Tribe, and other tribes on whose traditional lands we study and work at the University of Washington. MCW, JKW, and KT acknowledge support for this work from NSF-AST-1812521, and the contributions from the following University of Washington current and former undergraduate Astronomy majors who participated in "Werk SquAD" from 2017-2020: Steven Bet, Douglas Branton, Joseph Breneman, Dustin Burnham, Olivia Petry Caplow-Munro, Justin Collins, Apurva Goel, Samuel Johnson, Courtney Klein, Dalton LaCoste, Christina Lindberg, Camellia Magness, Travis Mandeville, Karalyn Ostler, Magdalyn Paige, Locke Patton, Brittany Platt, David Staker, Matthew Strasbourg, Sophia Taylor, Mercedes Thompson, and Leonardo Zhu. Additionally, the authors thank the anonymous referee for a productive report that improved the analysis.

MCW would like to thank Daniel J. D'Orazio and David Fleming for useful discussion.

This work was based on observations obtained at the Gemini Observatory, which is operated by the Association of Universities for Research in Astronomy, Inc., under a cooperative agreement with the NSF on behalf of the Gemini partnership: the National Science

Foundation (United States), National Research Council (Canada), CONICYT (Chile), Ministerio de Ciencia, Tecnología e Innovación Productiva (Argentina), Ministério da Ciência, Tecnologia e Inovação (Brazil), and Korea Astronomy and Space Science Institute (Republic of Korea). The Gemini North telescope is located within the Maunakea Science Reserve and adjacent to the summit of Maunakea. We are grateful for the privilege of observing the Universe from a place that is unique in both its astronomical quality and its cultural significance.

This research has made use of the SVO Filter Profile Service (<http://svo2.cab.inta-csic.es/theory/fps/>) supported from the Spanish MINECO through grant AYA2017-84089.

This work is partly based on tools and data products produced by GAZPAR operated by CeSAM-LAM and IAP.

The Pan-STARRS1 Surveys (PS1) and the PS1 public science archive have been made possible through contributions by the Institute for Astronomy, the University of Hawaii, the Pan-STARRS Project Office, the Max-Planck Society and its participating institutes, the Max Planck Institute for Astronomy, Heidelberg and the Max Planck Institute for Extraterrestrial Physics, Garching, The Johns Hopkins University, Durham University, the University of Edinburgh, the Queen's University Belfast, the Harvard-Smithsonian Center for Astrophysics, the Las Cumbres Observatory Global Telescope Network Incorporated, the National Central University of Taiwan, the Space Telescope Science Institute, the National Aeronautics and Space Administration under Grant No. NNX08AR22G issued through the Planetary Science Division of the NASA Science Mission Directorate, the National Science Foundation Grant No. AST-1238877, the University of Maryland, Eotvos Lorand University (ELTE), the Los Alamos National Laboratory, and the Gordon and Betty Moore Foundation.

The Legacy Surveys consist of three individual and complementary projects: the Dark Energy Camera Legacy Survey (DECaLS; NOAO Proposal ID 2014B-0404; PIs: David Schlegel and Arjun Dey), the Beijing-Arizona Sky Survey (BASS; NOAO Proposal ID 2015A-0801; PIs: Zhou Xu and Xiaohui Fan), and the Mayall z-band Legacy Survey (MzLS; NOAO Proposal ID 2016A-0453; PI: Arjun Dey). DECaLS, BASS and MzLS together include data obtained, respectively, at the Blanco telescope, Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory (NOAO); the Bok telescope, Steward Observatory, University of Arizona; and the Mayall telescope, Kitt Peak National Observatory, NOAO. The Legacy Surveys project is

honored to be permitted to conduct astronomical research on Iolkam Du’ag (Kitt Peak), a mountain with particular significance to the Tohono O’odham Nation.

NOAO is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

This project used data obtained with the Dark Energy Camera (DECam), which was constructed by the Dark Energy Survey (DES) collaboration. Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundacao Carlos Chagas Filho de Amparo, Financiadora de Estudos e Projetos, Fundacao Carlos Chagas Filho de Amparo a Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Cientifico e Tecnologico and the Ministerio da Ciencia, Tecnologia e Inovacao, the Deutsche Forschungsgemeinschaft and the Collaborating Institutions in the Dark Energy Survey. The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energeticas, Medioambientales y Tecnologicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenossische Technische Hochschule (ETH) Zurich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciencies de l’Espai (IEEC/CSIC), the Institut de Fisica d’Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universitat Munchen and the associated Excellence Cluster Universe, the University of Michigan, the National Optical Astronomy Ob-

servatory, the University of Nottingham, the Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, and Texas A&M University.

BASS is a key project of the Telescope Access Program (TAP), which has been funded by the National Astronomical Observatories of China, the Chinese Academy of Sciences (the Strategic Priority Research Program “The Emergence of Cosmological Structures” Grant XDB09000000), and the Special Fund for Astronomy from the Ministry of Finance. The BASS is also supported by the External Cooperation Program of Chinese Academy of Sciences (Grant 114A11KYSB20160057), and Chinese National Natural Science Foundation (Grant 11433005).

The Legacy Survey team makes use of data products from the Near-Earth Object Wide-field Infrared Survey Explorer (NEOWISE), which is a project of the Jet Propulsion Laboratory/California Institute of Technology. NEOWISE is funded by the National Aeronautics and Space Administration.

The Legacy Surveys imaging of the DESI footprint is supported by the Director, Office of Science, Office of High Energy Physics of the U.S. Department of Energy under Contract No. DE-AC02-05CH1123, by the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility under the same contract; and by the U.S. National Science Foundation, Division of Astronomical Sciences under Contract No. AST-0950945 to NOAO.

Software: SEXTRACTOR (Bertin & Arnouts 1996), PYIGM (Prochaska et al. 2017), PyeIt (Prochaska et al. 2020), ASTROPY (Astropy Collaboration et al. 2013; Price-Whelan et al. 2018), PyRAF (Science Software Branch at STScI 2012), linetools (Prochaska et al. 2016), REDROCK (<https://github.com/desihub/redrock>), CIGALE (Noll et al. 2009; Boquien et al. 2019), seaborn (Waskom & the seaborn development team 2020), pandas (pandas development team 2020; Wes McKinney 2010), corner (Foreman-Mackey 2016), emcee (Foreman-Mackey et al. 2013)

REFERENCES

- Abolfathi, B., Aguado, D. S., Aguilar, G., et al. 2018, ApJS, 235, 42, doi: [10.3847/1538-4365/aa9e8a](https://doi.org/10.3847/1538-4365/aa9e8a)
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33, doi: [10.1051/0004-6361/201322068](https://doi.org/10.1051/0004-6361/201322068)
- Bahcall, J. N., & Spitzer, Jr., L. 1969, ApJL, 156, L63, doi: [10.1086/180350](https://doi.org/10.1086/180350)
- Bergeron, J., & Stasińska, G. 1986, A&A, 169, 1
- Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393, doi: [10.1051/aas:1996164](https://doi.org/10.1051/aas:1996164)

- Blanton, M. R., & Roweis, S. 2007, *AJ*, 133, 734, doi: [10.1086/510127](https://doi.org/10.1086/510127)
- Boksenberg, A., & Sargent, W. L. W. 1978, *ApJ*, 220, 42, doi: [10.1086/155880](https://doi.org/10.1086/155880)
- Boquien, M., Burgarella, D., Roehlly, Y., et al. 2019, *A&A*, 622, A103, doi: [10.1051/0004-6361/201834156](https://doi.org/10.1051/0004-6361/201834156)
- Bordoloi, R., Prochaska, J. X., Tumlinson, J., et al. 2018, *ApJ*, 864, 132, doi: [10.3847/1538-4357/aad8ac](https://doi.org/10.3847/1538-4357/aad8ac)
- Bordoloi, R., Lilly, S. J., Knobel, C., et al. 2011, *ApJ*, 743, 10, doi: [10.1088/0004-637X/743/1/10](https://doi.org/10.1088/0004-637X/743/1/10)
- Bordoloi, R., Tumlinson, J., Werk, J. K., et al. 2014, *The Astrophysical Journal*, 796, 136, doi: [10.1088/0004-637x/796/2/136](https://doi.org/10.1088/0004-637x/796/2/136)
- Borthakur, S., Heckman, T., Tumlinson, J., et al. 2015, *ApJ*, 813, 46, doi: [10.1088/0004-637X/813/1/46](https://doi.org/10.1088/0004-637X/813/1/46)
- Bowen, D. V., Blades, J. C., & Pettini, M. 1995, *ApJ*, 448, 662, doi: [10.1086/175994](https://doi.org/10.1086/175994)
- Bowen, D. V., Chelouche, D., Jenkins, E. B., et al. 2016, *ApJ*, 826, 50, doi: [10.3847/0004-637X/826/1/50](https://doi.org/10.3847/0004-637X/826/1/50)
- Bowen, D. V., Pettini, M., & Blades, J. C. 2002, *ApJ*, 580, 169, doi: [10.1086/343106](https://doi.org/10.1086/343106)
- Bowen, D. V., Pettini, M., Penston, M. V., & Blades, C. 1991, *MNRAS*, 249, 145, doi: [10.1093/mnras/249.1.145](https://doi.org/10.1093/mnras/249.1.145)
- Bruzual, G., & Charlot, S. 2003, *MNRAS*, 344, 1000, doi: [10.1046/j.1365-8711.2003.06897.x](https://doi.org/10.1046/j.1365-8711.2003.06897.x)
- Burchett, J. N., Elek, O., Tejos, N., et al. 2020a, *ApJL*, 891, L35, doi: [10.3847/2041-8213/ab700c](https://doi.org/10.3847/2041-8213/ab700c)
- Burchett, J. N., Rubin, K. H. R., Prochaska, J. X., et al. 2020b, *arXiv e-prints*, arXiv:2005.03017, <https://arxiv.org/abs/2005.03017>
- Burchett, J. N., Tripp, T. M., Wang, Q. D., et al. 2018, *MNRAS*, 475, 2067, doi: [10.1093/mnras/stx3170](https://doi.org/10.1093/mnras/stx3170)
- Burchett, J. N., Tripp, T. M., Bordoloi, R., et al. 2016, *ApJ*, 832, 124, doi: [10.3847/0004-637X/832/2/124](https://doi.org/10.3847/0004-637X/832/2/124)
- Burchett, J. N., Tripp, T. M., Prochaska, J. X., et al. 2019, *ApJ*, 877, L20, doi: [10.3847/2041-8213/ab1f7f](https://doi.org/10.3847/2041-8213/ab1f7f)
- Calzetti, D., Kinney, A. L., & Storchi-Bergmann, T. 1994, *ApJ*, 429, 582, doi: [10.1086/174346](https://doi.org/10.1086/174346)
- Cappellari, M. 2017, *MNRAS*, 466, 798, doi: [10.1093/mnras/stw3020](https://doi.org/10.1093/mnras/stw3020)
- Chabrier, G. 2003, *PASP*, 115, 763, doi: [10.1086/376392](https://doi.org/10.1086/376392)
- Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, *arXiv e-prints*, arXiv:1612.05560, <https://arxiv.org/abs/1612.05560>
- Chang, Y.-Y., van der Wel, A., da Cunha, E., & Rix, H.-W. 2015, *ApJS*, 219, 8, doi: [10.1088/0067-0049/219/1/8](https://doi.org/10.1088/0067-0049/219/1/8)
- Chen, H.-W., Gauthier, J.-R., Sharon, K., et al. 2014, *MNRAS*, 438, 1435, doi: [10.1093/mnras/stt2288](https://doi.org/10.1093/mnras/stt2288)
- Chen, H.-W., Lanzetta, K. M., Webb, J. K., & Barcons, X. 2001, *ApJ*, 559, 654, doi: [10.1086/322414](https://doi.org/10.1086/322414)
- Chen, H.-W., Prochaska, J. X., Weiner, B. J., Mulchaey, J. S., & Williger, G. M. 2005, *ApJL*, 629, L25, doi: [10.1086/444377](https://doi.org/10.1086/444377)
- Chen, H.-W., & Tinker, J. L. 2008, *ApJ*, 687, 745, doi: [10.1086/591927](https://doi.org/10.1086/591927)
- Chen, H.-W., Zahedy, F. S., Johnson, S. D., et al. 2018, *MNRAS*, 479, 2547, doi: [10.1093/mnras/sty1541](https://doi.org/10.1093/mnras/sty1541)
- Cutri, R. M., Wright, E. L., Conrow, T., et al. 2013, *Explanatory Supplement to the AllWISE Data Release Products*, Tech. rep.
- Dale, D. A., Helou, G., Magdis, G. E., et al. 2014, *ApJ*, 784, 83, doi: [10.1088/0004-637X/784/1/83](https://doi.org/10.1088/0004-637X/784/1/83)
- Danforth, C. W., Keeney, B. A., Tilton, E. M., et al. 2016, *ApJ*, 817, 111, doi: [10.3847/0004-637X/817/2/111](https://doi.org/10.3847/0004-637X/817/2/111)
- Davé, R., Oppenheimer, B. D., Katz, N., Kollmeier, J. A., & Weinberg, D. H. 2010, *MNRAS*, 408, 2051, doi: [10.1111/j.1365-2966.2010.17279.x](https://doi.org/10.1111/j.1365-2966.2010.17279.x)
- Dey, A., Schlegel, D. J., Lang, D., et al. 2019, *AJ*, 157, 168, doi: [10.3847/1538-3881/ab089d](https://doi.org/10.3847/1538-3881/ab089d)
- Erb, D. K., Steidel, C. C., Shapley, A. E., et al. 2006, *ApJ*, 646, 107, doi: [10.1086/504891](https://doi.org/10.1086/504891)
- Fielding, D., Quataert, E., McCourt, M., & Thompson, T. A. 2017, *MNRAS*, 466, 3810, doi: [10.1093/mnras/stw3326](https://doi.org/10.1093/mnras/stw3326)
- Fielding, D. B., Ostriker, E. C., Bryan, G. L., & Jermyn, A. S. 2020, *ApJL*, 894, L24, doi: [10.3847/2041-8213/ab8d2c](https://doi.org/10.3847/2041-8213/ab8d2c)
- Ford, A. B., Oppenheimer, B. D., Davé, R., et al. 2013, *MNRAS*, 432, 89, doi: [10.1093/mnras/stt393](https://doi.org/10.1093/mnras/stt393)
- Foreman-Mackey, D. 2016, *The Journal of Open Source Software*, 24, doi: [10.21105/joss.00024](https://doi.org/10.21105/joss.00024)
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, 125, 306, doi: [10.1086/670067](https://doi.org/10.1086/670067)
- Fraternali, F., & Binney, J. J. 2008, *MNRAS*, 386, 935, doi: [10.1111/j.1365-2966.2008.13071.x](https://doi.org/10.1111/j.1365-2966.2008.13071.x)
- Gimeno, G., Roth, K., Chiboucas, K., et al. 2016, in *Ground-based and Airborne Instrumentation for Astronomy VI*, ed. C. J. Evans, L. Simard, & H. Takami, Vol. 9908, International Society for Optics and Photonics (SPIE), 872 – 885, doi: [10.1117/12.2233883](https://doi.org/10.1117/12.2233883)
- Haislmaier, K. J., Tripp, T. M., Katz, N., et al. 2020, *MNRAS*, doi: [10.1093/mnras/staa3544](https://doi.org/10.1093/mnras/staa3544)
- Hennawi, J. F., & Prochaska, J. X. 2007, *ApJ*, 655, 735, doi: [10.1086/509770](https://doi.org/10.1086/509770)
- Holland, S. T. e. 2012, *Cosmic Origins Spectrograph Instrument Handbook for Cycle 21 v. 5.0*
- Hook, I. M., Jørgensen, I., Allington-Smith, J. R., et al. 2004, *PASP*, 116, 425, doi: [10.1086/383624](https://doi.org/10.1086/383624)
- Huang, S., Katz, N., Scannapieco, E., et al. 2020, *MNRAS*, 497, 2586, doi: [10.1093/mnras/staa1978](https://doi.org/10.1093/mnras/staa1978)

- Hummels, C. B., Bryan, G. L., Smith, B. D., & Turk, M. J. 2013, *MNRAS*, 430, 1548, doi: [10.1093/mnras/sts702](https://doi.org/10.1093/mnras/sts702)
- Hummels, C. B., Smith, B. D., Hopkins, P. F., et al. 2019, *ApJ*, 882, 156, doi: [10.3847/1538-4357/ab378f](https://doi.org/10.3847/1538-4357/ab378f)
- Ji, S., Chan, T. K., Hummels, C. B., et al. 2019, arXiv e-prints, arXiv:1909.00003.
<https://arxiv.org/abs/1909.00003>
- Johnson, S. D., Chen, H.-W., & Mulchaey, J. S. 2015, *MNRAS*, 449, 3263, doi: [10.1093/mnras/stv553](https://doi.org/10.1093/mnras/stv553)
- Johnson, S. D., Chen, H.-W., Mulchaey, J. S., Schaye, J., & Straka, L. A. 2017, *ApJL*, 850, L10, doi: [10.3847/2041-8213/aa9370](https://doi.org/10.3847/2041-8213/aa9370)
- Keeney, B. A., Stocke, J. T., Danforth, C. W., et al. 2017, *ApJS*, 230, 6, doi: [10.3847/1538-4365/aa6b59](https://doi.org/10.3847/1538-4365/aa6b59)
- Keeney, B. A., Stocke, J. T., Pratt, C. T., et al. 2018, *ApJS*, 237, 11, doi: [10.3847/1538-4365/aac727](https://doi.org/10.3847/1538-4365/aac727)
- Kim, C.-G., & Ostriker, E. C. 2018, *ApJ*, 853, 173, doi: [10.3847/1538-4357/aaa5ff](https://doi.org/10.3847/1538-4357/aaa5ff)
- Lanzetta, K. M., Bowen, D. V., Tytler, D., & Webb, J. K. 1995, *ApJ*, 442, 538, doi: [10.1086/175459](https://doi.org/10.1086/175459)
- Lehner, N., Berek, S. C., Howk, J. C., et al. 2020, arXiv e-prints, arXiv:2002.07818.
<https://arxiv.org/abs/2002.07818>
- Liang, L., Durier, F., Babul, A., et al. 2016, *MNRAS*, 456, 4266, doi: [10.1093/mnras/stv2840](https://doi.org/10.1093/mnras/stv2840)
- Lo Faro, B., Buat, V., Roehly, Y., et al. 2017, *MNRAS*, 472, 1372, doi: [10.1093/mnras/stx1901](https://doi.org/10.1093/mnras/stx1901)
- Martin, D. C., O'Sullivan, D., Matuszewski, M., et al. 2019, *Nature Astronomy*, 356, doi: [10.1038/s41550-019-0791-2](https://doi.org/10.1038/s41550-019-0791-2)
- McQuinn, M. 2016, *Annual Review of Astronomy and Astrophysics*, 54, 313, doi: [10.1146/annurev-astro-082214-122355](https://doi.org/10.1146/annurev-astro-082214-122355)
- McQuinn, M., & Werk, J. K. 2018, *ApJ*, 852, 33, doi: [10.3847/1538-4357/aa9d3f](https://doi.org/10.3847/1538-4357/aa9d3f)
- Meiring, J. D., Tripp, T. M., Prochaska, J. X., et al. 2011, *ApJ*, 732, 35, doi: [10.1088/0004-637X/732/1/35](https://doi.org/10.1088/0004-637X/732/1/35)
- Morris, S. L., Weymann, R. J., Dressler, A., et al. 1993, *ApJ*, 419, 524, doi: [10.1086/173505](https://doi.org/10.1086/173505)
- Morton, D. C. 2003, *ApJS*, 149, 205, doi: [10.1086/377639](https://doi.org/10.1086/377639)
- Moster, B. P., Naab, T., & White, S. D. M. 2013, *MNRAS*, 428, 3121, doi: [10.1093/mnras/sts261](https://doi.org/10.1093/mnras/sts261)
- Muzahid, S., Srianand, R., & Petitjean, P. 2011, *MNRAS*, 410, 2193, doi: [10.1111/j.1365-2966.2010.17592.x](https://doi.org/10.1111/j.1365-2966.2010.17592.x)
- Nelson, D., Sharma, P., Pillepich, A., et al. 2020, arXiv e-prints, arXiv:2005.09654.
<https://arxiv.org/abs/2005.09654>
- Nielsen, N. M., Churchill, C. W., & Kacprzak, G. G. 2013, *ApJ*, 776, 115, doi: [10.1088/0004-637X/776/2/115](https://doi.org/10.1088/0004-637X/776/2/115)
- Noll, S., Burgarella, D., Giovannoli, E., et al. 2009, *A&A*, 507, 1793, doi: [10.1051/0004-6361/200912497](https://doi.org/10.1051/0004-6361/200912497)
- Oppenheimer, B. D., Schaye, J., Crain, R. A., Werk, J. K., & Richings, A. J. 2018, *MNRAS*, 481, 835, doi: [10.1093/mnras/sty2281](https://doi.org/10.1093/mnras/sty2281)
- Oppenheimer, B. D., Crain, R. A., Schaye, J., et al. 2016, *MNRAS*, 460, 2157, doi: [10.1093/mnras/stw1066](https://doi.org/10.1093/mnras/stw1066)
- pandas development team, T. 2020, pandas-dev/pandas: Pandas, latest, Zenodo, doi: [10.5281/zenodo.3509134](https://doi.org/10.5281/zenodo.3509134)
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *J. Mach. Learn. Res.*, 12, 2825.
<http://dl.acm.org/citation.cfm?id=1953048.2078195>
- Peeples, M. S., Corlies, L., Tumlinson, J., et al. 2019, *ApJ*, 873, 129, doi: [10.3847/1538-4357/ab0654](https://doi.org/10.3847/1538-4357/ab0654)
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, *A&A*, 594, A13, doi: [10.1051/0004-6361/201525830](https://doi.org/10.1051/0004-6361/201525830)
- Pratt, C. T., Stocke, J. T., Keeney, B. A., & Danforth, C. W. 2018, *ApJ*, 855, 18, doi: [10.3847/1538-4357/aaaac](https://doi.org/10.3847/1538-4357/aaaac)
- Price-Whelan, A. M., Sipőcz, B. M., Günther, H. M., et al. 2018, *AJ*, 156, 123, doi: [10.3847/1538-3881/aabc4f](https://doi.org/10.3847/1538-3881/aabc4f)
- Prochaska, J. X., Hennawi, J. F., Westfall, K. B., et al. 2020, arXiv e-prints, arXiv:2005.06505.
<https://arxiv.org/abs/2005.06505>
- Prochaska, J. X., Weiner, B., Chen, H.-W., Mulchaey, J., & Cooksey, K. 2011, *ApJ*, 740, 91, doi: [10.1088/0004-637X/740/2/91](https://doi.org/10.1088/0004-637X/740/2/91)
- Prochaska, J. X., Tejos, N., Crighton, N., et al. 2016, *linetools/linetools: Second major release, v0.2*, Zenodo, doi: [10.5281/zenodo.168270](https://doi.org/10.5281/zenodo.168270)
- Prochaska, J. X., Werk, J. K., Worseck, G., et al. 2017, *ApJ*, 837, 169, doi: [10.3847/1538-4357/aa6007](https://doi.org/10.3847/1538-4357/aa6007)
- Prochaska, J. X., Tejos, N., cwotta, et al. 2017, *pygim/pygim: Initial release for publications, v1.0*, Zenodo, doi: [10.5281/zenodo.1045480](https://doi.org/10.5281/zenodo.1045480)
- Prochaska, J. X., Burchett, J. N., Tripp, T. M., et al. 2019, *The Astrophysical Journal Supplement Series*, 243, 24, doi: [10.3847/1538-4365/ab2b9a](https://doi.org/10.3847/1538-4365/ab2b9a)
- Rauch, M. 1998, *Annual Review of Astronomy and Astrophysics*, 36, 267, doi: [10.1146/annurev.astro.36.1.267](https://doi.org/10.1146/annurev.astro.36.1.267)
- Rodrigo, C., Solano, E., & Bayo, A. 2012, *SVO Filter Profile Service Version 1.0*, Tech. rep., doi: [10.5479/ADS/bib/2012ivoa.rept.1015R](https://doi.org/10.5479/ADS/bib/2012ivoa.rept.1015R)
- Rudie, G. C., Steidel, C. C., Pettini, M., et al. 2019, arXiv e-prints, arXiv:1903.00004.
<https://arxiv.org/abs/1903.00004>
- Rudie, G. C., Steidel, C. C., Trainor, R. F., et al. 2012, *The Astrophysical Journal*, 750, 67, doi: [10.1088/0004-637x/750/1/67](https://doi.org/10.1088/0004-637x/750/1/67)
- Sahnow, D. J., Oliveira, C., Aloisi, A., et al. 2011, *Gain sag in the FUV detector of the Cosmic Origins Spectrograph*, Instrument Science Report COS 2011-05

- Savage, B. D., & Sembach, K. R. 1991, *ApJ*, 379, 245, doi: [10.1086/170498](https://doi.org/10.1086/170498)
- Schlaflly, E. F., & Finkbeiner, D. P. 2011, *The Astrophysical Journal*, 737, 103, doi: [10.1088/0004-637x/737/2/103](https://doi.org/10.1088/0004-637x/737/2/103)
- Schreiber, C., Pannella, M., Elbaz, D., et al. 2015, *A&A*, 575, A74, doi: [10.1051/0004-6361/201425017](https://doi.org/10.1051/0004-6361/201425017)
- Science Software Branch at STScI. 2012, PyRAF: Python alternative for IRAF. <http://ascl.net/1207.011>
- Shen, S., Madau, P., Aguirre, A., et al. 2012, *ApJ*, 760, 50, doi: [10.1088/0004-637X/760/1/50](https://doi.org/10.1088/0004-637X/760/1/50)
- Shull, J. M., Smith, B. D., & Danforth, C. W. 2012, *ApJ*, 759, 23, doi: [10.1088/0004-637X/759/1/23](https://doi.org/10.1088/0004-637X/759/1/23)
- Stinson, G. S., Brook, C., Prochaska, J. X., et al. 2012, *MNRAS*, 425, 1270, doi: [10.1111/j.1365-2966.2012.21522.x](https://doi.org/10.1111/j.1365-2966.2012.21522.x)
- Stoeck, J. T., Keeney, B. A., Danforth, C. W., et al. 2013, *ApJ*, 763, 148, doi: [10.1088/0004-637X/763/2/148](https://doi.org/10.1088/0004-637X/763/2/148)
- Tejos, N., Morris, S. L., Finn, C. W., et al. 2014, *MNRAS*, 437, 2017, doi: [10.1093/mnras/stt1844](https://doi.org/10.1093/mnras/stt1844)
- Thom, C., Tumlinson, J., Werk, J. K., et al. 2012, *The Astrophysical Journal*, 758, L41, doi: [10.1088/2041-8205/758/2/L41](https://doi.org/10.1088/2041-8205/758/2/L41)
- Tripp, T. M., Lu, L., & Savage, B. D. 1998, *ApJ*, 508, 200, doi: [10.1086/306397](https://doi.org/10.1086/306397)
- Tripp, T. M., Meiring, J. D., Prochaska, J. X., et al. 2011, *Science*, 334, 952, doi: [10.1126/science.1209850](https://doi.org/10.1126/science.1209850)
- Tumlinson, J., & Fang, T. 2005, *ApJL*, 623, L97, doi: [10.1086/430142](https://doi.org/10.1086/430142)
- Tumlinson, J., Peebles, M. S., & Werk, J. K. 2017, *Annual Review of Astronomy and Astrophysics*, 55, 389, doi: [10.1146/annurev-astro-091916-055240](https://doi.org/10.1146/annurev-astro-091916-055240)
- Tumlinson, J., Werk, J. K., Thom, C., et al. 2011, *ApJ*, 733, 111, doi: [10.1088/0004-637X/733/2/111](https://doi.org/10.1088/0004-637X/733/2/111)
- Tumlinson, J., Thom, C., Werk, J. K., et al. 2013, *The Astrophysical Journal*, 777, 59, doi: [10.1088/0004-637x/777/1/59](https://doi.org/10.1088/0004-637x/777/1/59)
- van de Voort, F., Springel, V., Mandelker, N., van den Bosch, F. C., & Pakmor, R. 2019, *MNRAS*, 482, L85, doi: [10.1093/mnrasl/sly190](https://doi.org/10.1093/mnrasl/sly190)
- Voit, G. M. 2019, *ApJ*, 880, 139, doi: [10.3847/1538-4357/ab2bfd](https://doi.org/10.3847/1538-4357/ab2bfd)
- Wakker, B. P., & Savage, B. D. 2009, *ApJS*, 182, 378, doi: [10.1088/0067-0049/182/1/378](https://doi.org/10.1088/0067-0049/182/1/378)
- Waskom, M., & the seaborn development team. 2020, *mwaskom/seaborn*, latest, Zenodo, doi: [10.5281/zenodo.592845](https://doi.org/10.5281/zenodo.592845)
- Werk, J. K., Prochaska, J. X., Thom, C., et al. 2012, *ApJS*, 198, 3, doi: [10.1088/0067-0049/198/1/3](https://doi.org/10.1088/0067-0049/198/1/3)
- . 2013, *ApJS*, 204, 17, doi: [10.1088/0067-0049/204/2/17](https://doi.org/10.1088/0067-0049/204/2/17)
- Werk, J. K., Prochaska, J. X., Tumlinson, J., et al. 2014, *ApJ*, 792, 8, doi: [10.1088/0004-637X/792/1/8](https://doi.org/10.1088/0004-637X/792/1/8)
- Wes McKinney. 2010, in *Proceedings of the 9th Python in Science Conference*, ed. Stéfan van der Walt & Jarrod Millman, 56 – 61, doi: [10.25080/Majora-92bf1922-00a](https://doi.org/10.25080/Majora-92bf1922-00a)
- White, S. D. M., & Rees, M. J. 1978, *MNRAS*, 183, 341, doi: [10.1093/mnras/183.3.341](https://doi.org/10.1093/mnras/183.3.341)