The Astrophysical Journal, 948:114 (13pp), 2023 May 10 https://doi.org/10.3847/1538-4357/acc85b

CGM^2 + CASBaH: The Mass Dependence of H I Lyα–Galaxy Clustering and the Extent of the CGM

Matthew C. Wilde[1], Kirill Tchernyshyov[1], Jessica K. Werk[1], Todd M. Tripp[2], Joseph N. Burchett[3,4], J. Xavier Prochaska[3,5,6], Nicolas Tejos[7], Nicolas Lehner[8], Rongmon Bordoloi[9], John M. O'Meara[10], Jason Tumlinson[11], and J. Christopher Howk[8]

[1] University of Washington, Department of Astronomy, Seattle, WA 98195, USA; mwilde@uw.edu
[2] Department of Astronomy, University of Massachusetts, 710 North Pleasant St., Amherst, MA 01003-9305, USA
[3] University of California, Santa Cruz; 1156 High St., Santa Cruz, CA 95064, USA
[4] Department of Astronomy, New Mexico State University, PO Box 30001, MSC 4500, Las Cruces, NM 88001, USA
[5] Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU) The University of Tokyo; 5-1-5 Kashiwanoha, Kashiwa, 277-8583, Japan
[6] Division of Science, National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan
[7] Instituto de Física, Pontificia Universidad Católica de Valparaíso, Casilla 4059, Valparaíso, Chile
[8] Department of Physics and Astronomy, University of Notre Dame, Notre Dame, IN 46556, USA
[9] North Carolina State University, Department of Physics, Raleigh, NC 27695-8202, USA
[10] W.M. Keck Observatory, 65-1120 Mamalahoa Hwy., Kamuela, HI 96743, USA
[11] Space Telescope Science Institute, Baltimore, MD, USA

Received 2023 January 5; revised 2023 March 24; accepted 2023 March 27; published 2023 May 15

Abstract

We combine data sets from the CGM^2 and CASBaH surveys to model a transition point, $R_{\rm cross}$, between circumgalactic and intergalactic media (CGM and IGM, respectively). In total, our data consist of 7244 galaxies at $z < 0.5$ with precisely measured spectroscopic redshifts, all having impact parameters of 0.01–20 comoving Mpc from 28 QSO sightlines with high-resolution UV spectra that cover H I Lyα. Our best-fitting model is a two-component model that combines a 3D absorber–galaxy cross-correlation function with a simple Gaussian profile at inner radii to represent the CGM. By design, this model gives rise to a determination of $R_{\rm cross}$ as a function of galaxy stellar mass, which can be interpreted as the boundary between the CGM and IGM. For galaxies with $10^8 \leqslant M_\star/M_\odot \leqslant 10^{10.5}$, we find that $R_{\rm cross}(M_\star) \approx 2.0 \pm 0.6 R_{\rm vir}$. Additionally, we find excellent agreement between $R_{\rm cross}(M_\star)$ and the theoretically determined splashback radius for galaxies in this mass range. Overall, our results favor models of galaxy evolution at $z < 0.5$ that distribute $T \approx 10^4$K gas to distances beyond the virial radius.

Unified Astronomy Thesaurus concepts: Galaxy evolution (594); Circumgalactic medium (1879); Surveys (1671); Quasar absorption line spectroscopy (1317); Galactic and extragalactic astronomy (563); Galaxies (573); Galaxy photometry (611); Galaxy spectroscopy (2171); Astronomy data modeling (1859); Astronomical models (86); Bayesian statistics (1900)

1. Introduction

The formation and evolution of galaxies involves a complex interplay between gravitational collapse of gas from the intergalactic medium (IGM), galaxy mergers, and feedback due to stellar evolution and active galactic nuclei (AGN) that drive gaseous outflows and change the ionization state of the galaxies' gaseous halos. Together, these processes drive the "cosmic baryon cycle," which takes place largely in the region of a galaxy referred to as the circumgalactic medium (CGM). Indeed, understanding the CGM is critical for developing a complete theory of galaxy evolution, as highlighted by the recent decadal survey (National Academies of Sciences, Engineering & Medicine 2021). In particular, the extent of the gaseous CGM relative to the extent of the dark matter halo is a subject of great interest for models that aim to reproduce the properties of gaseous halos.

The existence of the CGM, first predicted by Bahcall & Spitzer (1969), was initially revealed by detection of Mg II and H I absorption at large projected distances ($R_\perp > 20$ kpc) from L^* galaxies (Bergeron 1986; Bergeron & Boissé 1991; Morris et al. 1993; Lanzetta et al. 1995; Chen et al. 2005), and it was subsequently traced via higher-energy metal-line transitions such as Si III, C IV, and O VI that are observed to correlate with galaxies and their global properties (e.g., Tripp & Savage 2000; Tripp et al. 2008; Prochaska et al. 2011; Tumlinson et al. 2011; Werk et al. 2013). Within 0.5 $R_{\rm vir}$ of $L \sim L^*$ galaxies, the metal-line incidence is found to be 60%–90% for a range of ionized metal species (Werk et al. 2013). Conversely, Berg et al. (2022) find an 80% chance of finding a massive galaxy nearby to any high-metallicity absorber. The CGM of $M_\star > 10^8\ M_\odot$ galaxies is now well established to be metal-enriched (Bordoloi et al. 2014; Liang & Chen 2014; Prochaska et al. 2017; Berg et al. 2022), and to extend to at least 1 $R_{\rm vir}$—and very likely beyond it (Wakker & Savage 2009; Burchett et al. 2015; Finn et al. 2016; Wilde et al. 2021; Borthakur 2022).

Generally, hydrodynamical simulations of galaxy evolution, which exhibit complex interactions between gravitational collapse from the cosmological large-scale structure and subsequent feedback from supernovae and AGN-driven winds that heat and enrich the CGM and IGM (EAGLE, Schaye et al. 2015; IllustrisTNG, Pillepich et al. 2018; SIMBA, Davé et al. 2019; and CAMELS, Villaescusa-Navarro et al. 2022), are consistent with the range of observations of the CGM in absorption. Yet these models still rely on simplistic implementations of the "subgrid" physics in order to model entire

galaxies (e.g., Ford et al. 2013; Hummels et al. 2013), and physical properties of the CGM are dependent on the simulation resolution (Hummels et al. 2019; Peeples et al. 2019). More sensitive observations of the CGM, including the ability to detect the diffuse gas in emission, are needed both to break degeneracies in these models, e.g., between heating and cooling mechanisms, and to develop a flexible parametric model of the CGM (Singh et al. 2021).

The two-point correlation function between H I absorption along QSO sightlines and galaxies has proven to be an essential tool to understand the connection of galaxies to the IGM (e.g., Morris et al. 1993; Chen et al. 2005; Ryan-Weber 2006; Prochaska et al. 2011, 2019; Tejos et al. 2014). The primary advantages of leveraging the clustering of these two entities over one-to-one association analyses is that it provides results for large scales (1–10 Mpc) as well as the relatively smaller scales where the baryonic processes associated with the CGM play out, and the correlation function statistically characterizes absorber–galaxy relationships when multiple galaxies are close to the sightline and a one-to-one assignment is ambiguous. Because H I traces both enriched material from galaxies as well as primordial accretion from the IGM, observations of the CGM, IGM, and galaxies in the same volume are fundamental to both testing the predictions of galaxy evolution models and providing a means to differentiate between them (e.g., Fumagalli et al. 2011; Oppenheimer et al. 2012; Stinson et al. 2012; Ford et al. 2013; Hummels et al. 2013; Butsky et al. 2020; Singh et al. 2021).

Understanding the physical profile and size of the CGM sheds light on the nonlinear processes of galaxy formation: on what spatial scale(s) do virialization, accretion, and feedback transform these galactic atmospheres? Astronomers have long used some version of the virial radius as an estimator for the size of galaxy halos, but this estimate is somewhat arbitrary and is based on the distribution of unobservable dark matter. By observing the radial gas profile around galaxies out to large scales, we can effectively map the gaseous halo, which in turn constrains the physics of galaxy-scale feedback processes. Observationally determining the galactic atmosphere's extent has additional implications for constraining galaxy evolution and assembly models. For example, the galaxy baryon and metal budgets require a scale to integrate the total mass (e.g., Peeples et al. 2014; Werk et al. 2014). Furthermore, the gaseous halo likely plays an important role in the quenching of dwarf satellite galaxies as they become stripped by ram pressure in a low-density CGM (Putman et al. 2021), and it is useful to constrain where this occurs, i.e., the extent of the CGM, and how this depends on central galaxy mass.

The presence of H I absorption beyond the virial radius is now widely accepted for a range of galaxy stellar masses (e.g., Prochaska et al. 2011; Tejos et al. 2012, 2014; Bouma et al. 2021; Wilde et al. 2021; Borthakur 2022). In Wilde et al. (2021) (henceforth Paper I), we found an empirical relation between galaxy stellar mass and the extent of the CGM as indicated by H I covering fractions. For galaxies with stellar masses $10^8 < M_*/M_\odot < 10^{11.5}$, we found that the CGM extends to two times the virial radius. In this paper, we focus on the functional forms of the mass dependence of the H I-traced CGM using a power-law model similar to the two-halo correlation function. We also investigate other two-component models that differentiate the CGM from the IGM. We combine the CGM^2 Survey, which focuses on sightlines at low galaxy impact parameters (<1 Mpc), with the COS Absorption Survey of Baryon Harbors (CASBaH), which probes larger spatial scales (<20 Mpc). In doing so, we greatly increase the absorber–galaxy sample from 543 spectroscopically confirmed absorber–galaxy pairs to 7244 pairs spanning $0.003 < z < 0.48$. Our goal is to provide the most reliable constraints to date on the spatial extent of the CGM as traced by H I absorption.

The paper is structured as follows: In Section 2, we briefly review each of the galaxy–absorber surveys and discuss their combined properties. In Section 3, we introduce two models of the H I–galaxy correlation functions, and we cover our main results in Section 4. We compare our results with simulations and previous results and discuss their implications for galaxy evolution models in Section 5. Finally, we summarize our results in Section 6.

2. Data—Combining CGM^2 and CASBaH

Both surveys feature far-ultraviolet spectroscopy of QSOs with HST, using both the Cosmic Origins Spectrograph (COS; Green et al. 2012) and the Space Telescope Imaging Spectrograph (STIS; Woodgate et al. 1998), as well as deep, ground-based optical spectroscopy of foreground galaxies in the QSO fields. CASBaH is well suited to the study of the interface between the CGM and the IGM at scales $\gtrsim$1 Mpc. CGM^2 provides a relatively more complete mapping of the inner CGM at scales $\lesssim$1 Mpc. By combining CGM^2 and CASBaH data, we leverage the strengths of each survey, as described below. Figure 1 shows the distributions of galaxy stellar masses and impact parameters versus redshift from both surveys out to $z = 0.5$. Together, the surveys allow us to probe the CGM as it transitions into the IGM for a large sample of galaxies.

2.1. CGM^2

The CGM^2 survey, first presented in Wilde et al. (2021), includes precise spectroscopic redshifts and bulk galaxy properties (e.g., stellar masses, M_*, and star formation rates, henceforth SFR) from a combination of Gemini-GMOS spectra and deep, broadband photometry for ~1000 galaxies in the foreground of 22 QSOs, each with S/N ≈ 10 HST/COS G130M+G160M spectra. By matching galaxy and absorber redshifts in ±500 km s^{-1} windows, the CGM^2 survey is ultimately a large collection of measurements pertaining to the CGM of $z < 1$ galaxies over a wide range of stellar masses, $10^8 \lesssim M_*/M_\odot \lesssim 10^{11.5}$. The data acquisition and analysis are explained in detail in Wilde et al. (2021). Here, we present a brief overview of the survey data relevant to the present analysis.

The CGM^2 galaxy spectra were obtained using Gemini-GMOS spectrographs on the twin Gemini North and South telescopes (Hook et al. 2004; Gimeno et al. 2016). Galaxy redshifts were inferred from the template fitting code, Redrock[12] (v0.14) and manually inspected with VETRR.[13] The typical statistical uncertainly of our redshifts is $\sigma_z \sim 50$–100 km s^{-1} ($\Delta z \simeq 0.00016$–0.00030). Photometry of the CGM^2 galaxy catalog was obtained from the Gemini-GMOS pre-imaging in the *g* and *i* bands as well as all available bands from DESI Legacy Imaging Surveys Data Release 8

[12] https://github.com/desihub/redrock
[13] https://github.com/mattcwilde/vetrr

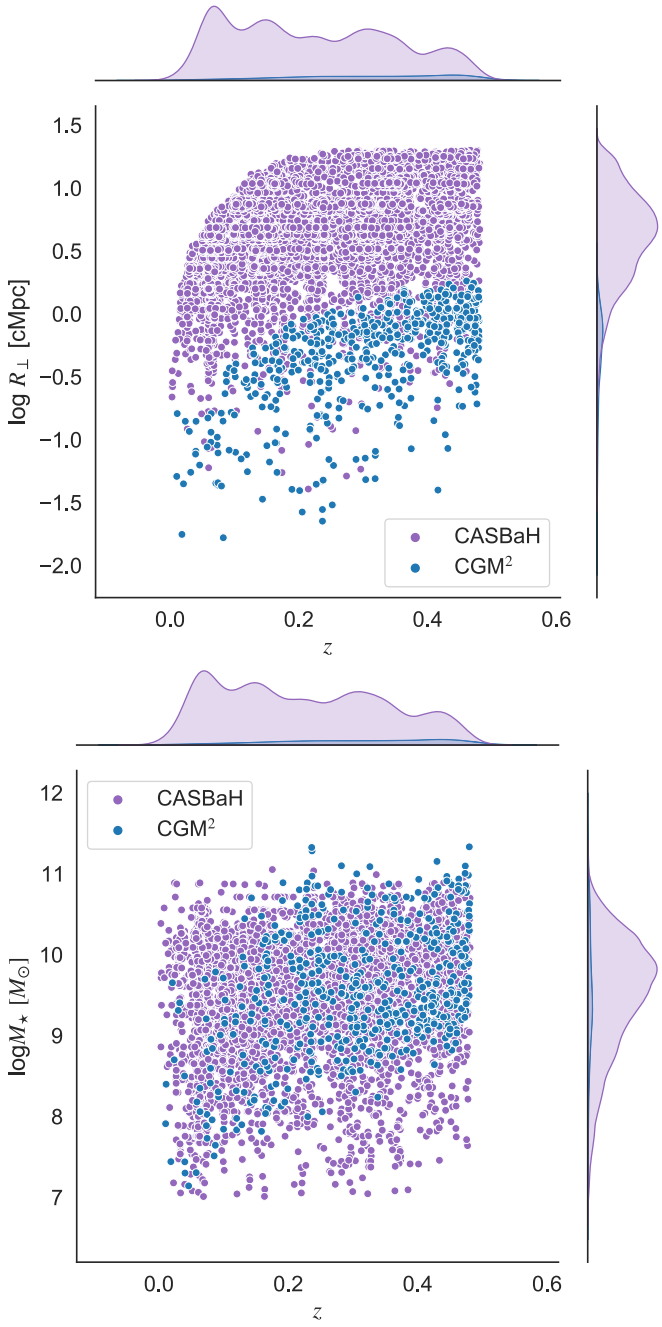

Figure 1. Top: distribution of the combined CGM^2 (blue dots) and CASBaH (purple dots) data sets in both logarithmic impact parameter and redshift. The data are roughly uniform in redshift space, but we can see the relative contributions of the data sets in impact parameter space; CGM^2 is highly concentrated at lower impact parameters, while CASBaH explores much greater impact parameters. Bottom: galaxy stellar mass distribution as a function of redshift for the two data sets.

(DR8; Dey et al. 2019), WISE (Cutri et al. 2013), Pan-STARRS Data Release 2 (Chambers et al. 2016), and SDSS DR14 (Abolfathi et al. 2018).

The 22 QSOs included in the CGM^2 survey have HST/COS spectra selected from the COS-Halos (GO11598, GO13033; Tumlinson et al. 2013) and COS-Dwarfs (GO12248; Bordoloi et al. 2014) surveys. In general, the CGM^2 QSO targets have $z_{\rm QSO} > 0.6$ and available HST imaging, which permits detailed analysis of absorption-hosting galaxies with $z < 0.5$. All COS spectra include both the G130M and G160M gratings and have an S/N $\simeq$ 8–12 per resolution element (FWHM $\simeq$ 16–18 km s^{-1}) or better over 1150–1800 Å. The COS data and their reduction are presented in detail by Tumlinson et al. (2013) and Bordoloi et al. (2014), following the same method used by Tripp et al. (2011), Meiring et al. (2011), Tumlinson et al. (2011), and Thom et al. (2012).

2.2. CASBaH

The CASBaH program was designed to take advantage of the multitude of resonance transitions at rest-frame wavelengths <912 Å to probe the physical conditions, metallicity, and physics of the multiphase CGM. A wide variety of elements and ionization stages have resonance lines only at $\lambda < 912$ Å (see, e.g., Verner et al. 1994), so observations of this wavelength range provide new diagnostics and precise constraints using banks of adjacent ions such as N I through N V, O I through O VI, and Ne II through Ne VIII (see Tripp et al. 2011, for examples of lines detected by CASBaH). The Ne VIII 770.4, 780.3 Å doublet has received particular attention as a probe of warm–hot gas at $\approx 10^5$–10^6 K (e.g., Savage et al. 2005; Burchett et al. 2019; Wijers et al. 2020). In many contexts such as the Milky Way interstellar medium, these lines are inaccessible because they are blocked by the H I Lyman limit. CASBaH overcomes this limitation by observing QSO absorbers with sufficient redshift to bring the lines into the observable band of HST.

The motivation and design of the CASBaH program is summarized in Section 1 of Haislmaier et al. (2021), and the CASBaH galaxy-redshift survey is presented in Prochaska et al. (2019). Briefly, CASBaH obtained both HST/COS and HST/STIS spectra of nine QSOs at $0.92 < z_{\rm QSO} < 1.48$, with two primary selection criteria. First, because some of the most important target lines (e.g., Ne VIII) are weak, the QSOs were required to be UV-bright so that good signal-to-noise ratios and sensitivity to weak lines would be attained. Second, the targets were required to have $z_{\rm QSO} > 0.9$ in order to provide a total redshift path that is sufficient to accumulate a statistically useful sample of absorbers of interest. No considerations were given to known foreground galaxies or absorbers, so the targets were not selected in a way that would favor particular types of foreground absorbers or galaxies, except that sightlines with known fully absorbed Lyman limits at $\lambda_{\rm ob} > 1150$ Å were excluded in order to avoid using HST time on sightlines that would not contribute useful path lengths to the samples (see Burchett et al. 2019). The CASBaH UV spectra were reduced in the same way as the CGM^2 data.

The CASBaH galaxy-redshift survey (Prochaska et al. 2019) measured thousands of redshifts in the fields of seven of the CASBaH QSOs using the Keck DEIMOS and MMT Hectospec spectrographs, with typical redshift uncertainties of ≈30 km s^{-1}. The survey used a wedding-cake strategy with the Hectospec covering galaxies in the ≈1° fields centered on the QSOs and the DEIMOS survey providing a deeper survey with a smaller field of view (81.5 arcmin^2) (see Prochaska et al. 2019). Using the CASBaH galaxy database, supplemented with data from public surveys such SDSS, we selected a sample of 6701 galaxies with spectroscopic redshifts $z < 0.481$ and comoving impact parameters less than 13 cMpc, appropriate for the H I analysis presented here.

2.3. Synergy of CGM^2 + CASBaH

The CASBaH and CGM^2 surveys have complementary designs. On the one hand, CGM^2 is built on COS-Halos and thus favors at least one L^* galaxy close to the sightline. CGM^2 also covers a smaller FOV. On the other hand, CASBaH is a blind survey that covers a larger FOV. Consequently, CASBaH provides more information about galaxies and large-scale structures at larger impact parameters, but as a blind survey, it is cross-section weighted in favor of galaxies at larger impact parameters. Also, because CASBaH avoided sightlines with fully absorbed Lyman limits in the HST band (i.e., at $\lambda_{ob} \geqslant 1150$ Å), it will not include galaxies at $z_{gal} > 0.26$ that harbor absorbers with $N(\mathrm{H\,I}) \gtrsim 10^{17}$ cm^{-2}. Thus, CGM^2 probes the inner CGM including higher $N(\mathrm{H\,I})$ absorbers, while CASBaH complements CGM^2 by adding very large samples of galaxies and structures at larger distances.

2.4. Galaxy Properties

To estimate the galaxy properties for both surveys, we used CIGALE (Noll et al. 2009; Boquien et al. 2019) to fit the spectral energy distribution (SED) and retrieve stellar mass and star formation rates (SFR). We used the Bruzual & Charlot (2003) stellar population models, assuming a Chabrier (2003) initial mass function (IMF). We chose a grid of metallicities ranging from 0.001 to $2.5Z_{\odot}$. A delayed star formation history (SFH) model was employed with an exponential burst. The e-folding time of the main stellar population models ranged from 0.1 to 8 Gyr. We varied the age of the oldest stars in the galaxy from 2 to 12 Gyr. We included an optional late burst with an e-folding time of 50 Myr and an age of 20 Myr. The burst mass fraction varied 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 to 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 an 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 to 1.98. Finally, we applied a reduction factor of 0.44 to the color excess for the old population compared to the young stars.

CIGALE then provides us with Bayesian estimates for the stellar mass and SFR for each galaxy in the combined catalog. In order to calculate the virial radius, we used the abundance-matching method of Moster et al. (2013) with the modifications used in Burchett et al. (2016). We adopt the convention of using $R_{vir} = R_{200m}$, 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.

2.5. Combining the CGM^2 and CASBaH Surveys

In order to combine the surveys, we modified both catalogs to ensure the same matching criteria between galaxies and absorbers. In the original CGM^2 survey, we measured the 2σ upper limit on absorption within $\delta v = \pm 30$ km s^{-1} of the galaxies redshift using the normalized error of the quasar flux when no absorption system was found within our $|\delta v| < 500$ km s^{-1} window. In order to match the CASBaH survey, we adjusted this to a 3σ upper limit. This did not change our results in a meaningful way. The original CASBaH survey used a velocity window of $|\delta v| < 400$ km s^{-1} to match the galaxies to absorption systems. We adjusted the window for this work to $|\delta v| < 500$ km s^{-1} to match the CGM^2 survey. As in Paper I, we restrict our H I measurements to those less than $z < 0.481$, because at this redshift, the Lyα line redshifts out of the G160 grating band, and thus we are only sensitive to higher-order transitions at higher redshifts.

Having made these two small changes to each survey, both could be combined to give us a total survey that includes 7244 galaxies spanning ∼0.01–8 comoving Mpc in impact parameter around 28 QSO sightlines. The distributions of impact parameter, redshift, and stellar mass are shown in Figure 1. In this paper, we will focus on galaxies with $8 < \log M_{\star}/M_{\odot} < 10.5$, a stellar mass range with good coverage in both surveys, which trims our galaxy sample to 6136 galaxies from CASBaH and 453 galaxies from CGM^2, for a total sample of 6589 absorber–galaxy pairs. The number of absorber–galaxy pairs is summarized in Table 1.

3. Modeling Absorber–Galaxy Clustering

We model the CGM using an absorber–galaxy cross-correlation analysis. This technique is based on modeling the covering fraction, f_c, as a binomial probability distribution of detections. To ensure high completeness in the absorber sample, based on the S/N of the data, we require a total column density $N_{\mathrm{H\,I}} \geqslant 10^{14}$ cm^{-2} to consider the sightline to have a "detection." Likewise, a nondetection is the case where we do not detect gas above this threshold. The models used here are based on the models employed in Paper I, which was inspired by the model developed by Hennawi & Prochaska (2007) and Prochaska et al. (2019). A more detailed explanation can be found in those three papers. In Paper I, we found a mass dependence of the extent of the CGM based on dividing the data into three mass bins. In this work, we wish to quantify the mass dependence of the clustering as well as determine the redshift dependence given our data.

3.1. Single Power-law Model

The single power-law model consists of two terms: the base rate of detection due to the random incidence of absorbers greater than this threshold and an excess above this base rate due to the clustering of galaxy–absorber pairs.

Much like Prochaska et al. (2019), we define the 3D absorber–galaxy cross-correlation function, $\xi_{ag}(r)$ as

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

To model the galaxy mass dependence of the clustering, we add a new mass dependence to the clustering scale, r_0:

$$r_{0,m}(m) = r_0 \left(\frac{M_{\star}}{M_0}\right)^{\beta}. \quad (2)$$

As before, we examine the projected 2D correlation function, which is obtained by integrating the 3D correlation function over the line of sight,

$$\chi_{\perp}(r) = \frac{1}{\Delta r_{\parallel}} \int_{r_{\parallel}} \xi_{ag}\left(\sqrt{r_{\parallel}^2 + r_{\perp}^2}\right) dr_{\parallel}, \quad (3)$$

Table 1
Number of Absorber–Galaxy Pairs

Survey (1)	$10^{7-11.3}M_*/M_\odot$ (2)	$10^{8-10.5}M_*/M_\odot$ (3)	$10^{8-9}M_*/M_\odot$ (4)	$10^{9-10}M_*/M_\odot$ (5)	$10^{10-10.5}M_*/M_\odot$ (6)
CGM^2	543	453	103	271	79
CASBaH	6701	6136	1265	3545	1326
Total	7244	6589	1368	3816	1405

Notes. Summary of absorber–galaxy pairs used in this manuscript. (1) The number of absorber–galaxy pairs in each survey and total of the combined surveys; (2) the number of absorber–galaxy pairs in the entire mass range; (3) the mass range used to perform the model fitting; (4, 5, 6) the number of absorber–galaxy pairs within each mass bin used for model verification.

where $r_\parallel$ is the line-of-sight distance, $r_\perp$ is the transverse distance, and $\Delta r_\parallel$ is the size of the redshift window.

For simplicity of notation, r is equivalent to $r_\perp$ in the following analysis.

In the following definitions, we label the single power-law clustering terms "two-halo," as the galaxy clustering method we adopt here describes the clustering of separate dark matter halos. This approach distinguishes the two-halo-only method from the two-component model we develop later in this manuscript.

In order to model f_c, we assume that the number of detected absorbers above the column-density threshold has a Poisson distribution. We consider two cases: (1) one or more absorbers are detected or (2) no absorbers are detected. In this framework, the probability of seeing no absorbers is

$$P^{\rm miss} = \frac{\lambda^0 \exp(-\lambda)}{0!}, \tag{4}$$

where we denote the rate of incidence (see below) as λ. The probability of finding one or more absorbers is just the complement of Equation (4),

$$f_c = 1 - P^{\rm miss}. \tag{5}$$

We model the rate of absorber incidence λ as a combination of the projected correlation function $\chi_\perp^{2\rm h}$ and a base detection rate due to the random incidence of absorbers $\langle d\mathcal{N}/dz\rangle\,\delta z$:

$$\lambda = (1 + \chi_\perp^{2\rm h})\,\langle d\mathcal{N}/dz\rangle\,\delta z. \tag{6}$$

δz is the width of the line-of-sight redshift window.

In addition to parameterizing the mass dependence as in Equation (2), we also parameterize the redshift dependence of $\langle dN/dz\rangle$ as follows:

$$\frac{d\mathcal{N}(N_{\rm HI} \geqslant N_{\rm HI}^{14}, z)}{dz} = C_0(1+z)^\alpha, \tag{7}$$

where $N_{\rm HI}^{14}$ denotes absorbers with column densities of 10^{14} cm^{-2}, C_0 is the random rate of incidence at $z = 0$, and δz is the redshift window. We adopt a redshift window to be $\pm$500 km s^{-1} in velocity units.

Thus, we have a rate of incidence of the form

$$\lambda = (1 + [\chi_\perp^{2\rm h}(r, m|r_0^{2\rm h}, \gamma^{2\rm h}, \beta^{2\rm h})])\,\langle d\mathcal{N}(z|C_0, \alpha)/dz\rangle\,\delta z. \tag{8}$$

Finally, we construct the likelihood function,

$$\mathcal{L} = \prod_i P^{\rm hit}(r_i, z_i, m_i|\theta) \prod_j P^{\rm miss}(r_j, z_j, m_j|\theta), \tag{9}$$

where $\theta = [r_0^{2\rm h}, \gamma^{2\rm h}, \beta^{2\rm h}, C_0, \alpha]$.

In constructing our Bayesian model, we must choose priors. For the single power-law parameters, we chose the priors based on the results of cross-correlation analysis by Tejos et al. (2014), except for our new mass-dependent term, $\beta^{2\rm h}$, which was motivated by physical arguments:

$$r_0^{2\rm h} \sim \mathcal{N}(\mu = 3.2, \sigma = 0.3),\ r_0^{2\rm h} > 0, \tag{10}$$

$$\gamma^{2\rm h} \sim \mathcal{N}(\mu = 1.7, \sigma = 0.1),\ \gamma^{2\rm h} > 0, \tag{11}$$

$$\beta^{2\rm h} > 0, \tag{12}$$

where $\mathcal{N}$ is the normal distribution with mean μ and variance σ^2.

The priors for the redshift dependence were chosen based on the findings in Kim et al. (2021):

$$C_0 \sim \text{Lognormal}(\mu = 1.25, \sigma = 0.11),\ C_0 > 0, \tag{13}$$

$$\alpha \sim \mathcal{N}(\mu = 0.97, \sigma = 0.87),\ -3 < \alpha < 3. \tag{14}$$

We note that we chose to use the more recent results of Kim et al. (2021) in modeling the redshift evolution instead of those from Danforth et al. (2016), as were used in Paper I.

As in Paper I, we apply the Bayesian Markov Chain Monte Carlo (MCMC) sampler `emcee` (Foreman-Mackey et al. 2013) to generate samples from the posterior probability distribution function in order to estimate the parameters of interest and their distributions, using Equation (9) and the priors described above.

In Figure 2, we show the posterior distributions of our single power-law model with $M_0 = 10^{9.5}M_\odot$. These were fit only to data with $8 < \log M_\star/M_\odot < 10.5$, as above this range there is a change in the virial radius due to the $M_\star$–$M_{\rm halo}$ relation from abundance matching (Moster et al. 2013). Below this mass range, we find a very flat covering fraction profile, which does not show a clustering signal.

3.2. Two-component Models

The single power-law model used in galaxy–galaxy clustering and adapted above to model the galaxy–absorber clustering makes no assumption of a CGM or overlapping (in projection) gaseous halos. However, the existence of the CGM is now well established (Tumlinson et al. 2017). In particular, the trends of ionized metal species with impact parameter around L^* and sub-L^* galaxies from $z = 0$–3.5 distinctly show that metal-enriched gaseous atmospheres are a fundamental component of galaxies (e.g., Werk et al. 2013; Bordoloi et al. 2014; Lehner et al. 2014; Borthakur et al. 2015; Rudie et al. 2019). In the following section, we therefore assume the existence of the CGM and use a simple Gaussian profile to model the excess clustering signal due to the presence of the CGM. In addition, we investigated several other functional forms of the CGM component, which we describe in Section 3.2.2. We find that

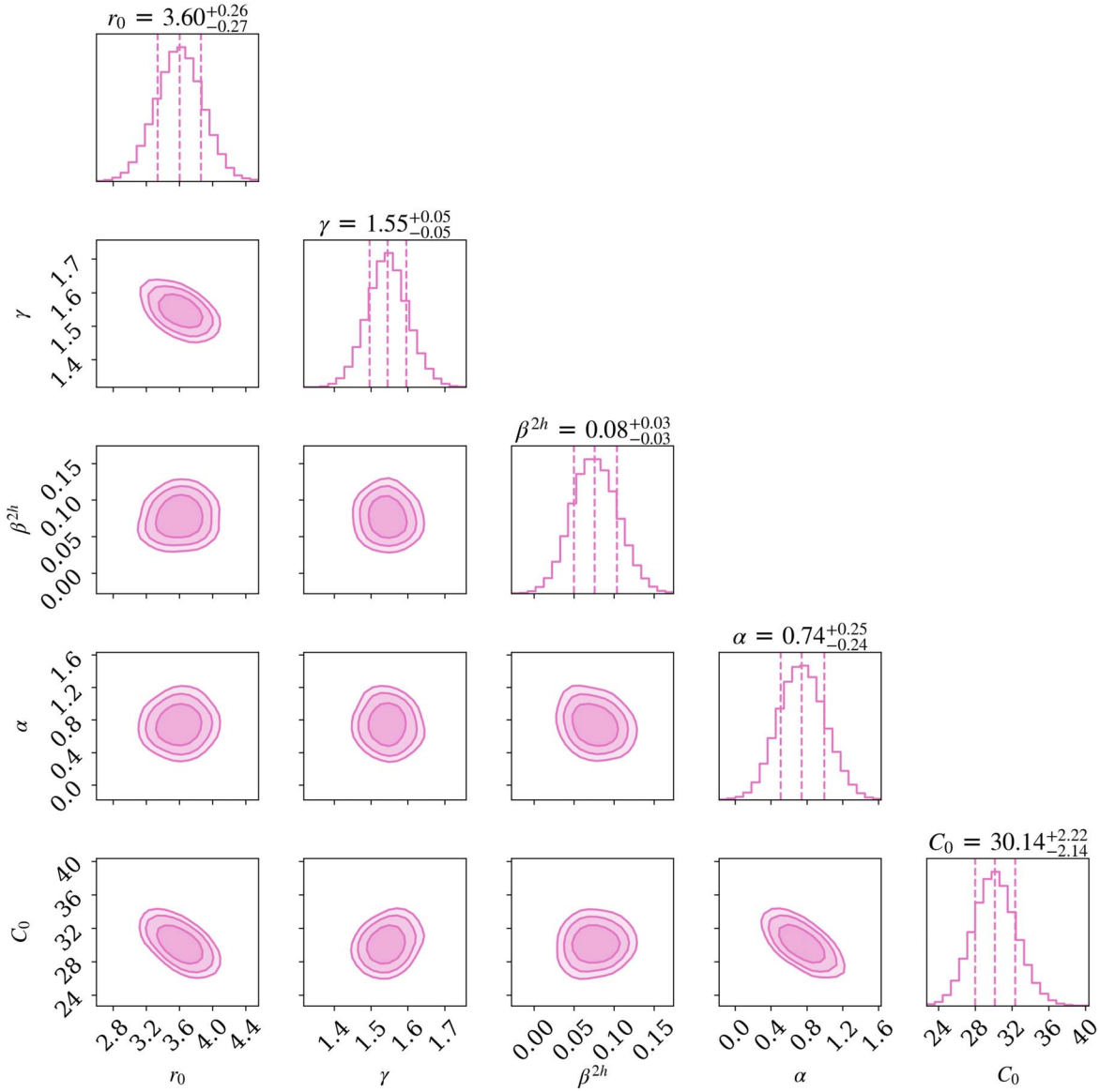

Figure 2. Corner plots showing the posterior parameter probabilities for the parameters in the single power-law clustering model. We find a nonzero, positive mass dependence term in the two-halo absorber–galaxy clustering, β^{2h}.

the particular functional form of this component has little impact on the results.

3.2.1. The Gaussian CGM Two-component Model

We now add a third term to the detection rate: a Gaussian one-halo component. The detection rate now consists of a baseline random incidence rate, an enhancement due to large-scale absorber–galaxy clustering, and an additional enhancement due to the CGM. We employ an exclusion model where the contribution from the two-halo term terminates at the distance it reaches the one-halo component. This scheme, shown in Figure 3, also allows us to determine a natural estimate of the extent of the CGM: the crossing point of the one- and two-halo components. More explicitly, within some radius, the galaxy has a CGM that we define as the gas of that galaxy and any other satellite galaxies within its halo. Our formalism then defines the $R_{\rm cross}$ where this CGM component exceeds the two-halo.

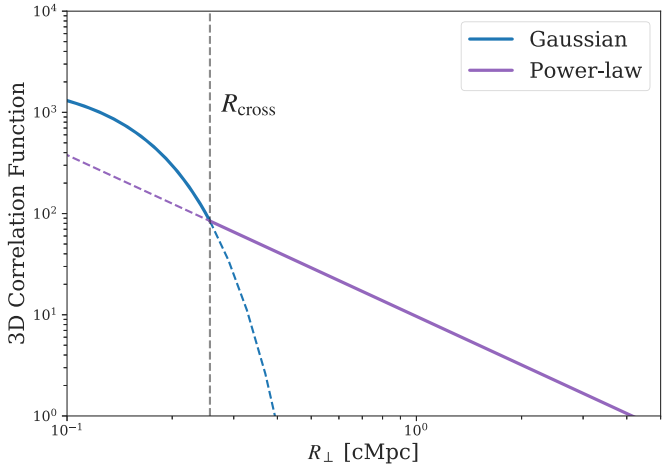

Figure 3. A schematic depiction of our two-component exclusion model and the determination of $R_{\rm cross}$. The two-halo component cuts off interior to $R_{\rm cross}$.

The model is similar to that single power law we introduced before, with a few key differences. We introduce a Gaussian one-halo term defined as

$$G(r)^{1h} = Ae^{-(r/\sigma)^2}. \quad (15)$$

Where the two models intersect, R_{cross}, we can solve for σ as

$$\sigma = \sqrt{\frac{1}{2}\frac{R_{cross}^2}{\ln(A) + \gamma \ln(R_{cross}/r_0)}}. \quad (16)$$

It should be noted that R_{cross} here is the 3D distance and not the projected distance. In order to characterize the mass dependence of R_{cross}, we define

$$R_{cross} = R_{cross,0}\left(\frac{M_\star}{M_0}\right)^{\beta^{1h}}, \quad (17)$$

where $R_{cross,0}$ is the one-halo term extent for a galaxy at the fixed pivot mass M_0. The galaxy mass dependence of σ includes contributions from the mass dependencies of R_{cross} and r_0.

This parameterization allows us to compare the mass dependence of the one-halo term, β^{1h}, with that of the two-halo term, β^{2h}.

In order to solve for the projected clustering signal, ξ, we first make some definitions to ease the notation. We use $s = r_\parallel$ in the remainder of the analysis. The integration is performed over different portions of the line-of-sight distance, s, corresponding to the one- and two-halo components. We define the line-of-sight crossing point s_{cross} as

$$s_{cross} = \sqrt{\max(R_{cross}^2 - r_\perp^2, 0)}, \quad (18)$$

and we can then integrate Equation (15) to $s_{eval} = \min(s_{cross}, s_{max})$, where s_{max} is the maximum interval we wish to integrate over, which in our case is $[-500, 500]$ km s^{-1}. Thus, we have

$$\chi(r_\perp) \propto 2\int_0^{s_{eval}} G(r_\perp, s)^{1h} ds + 2\int_{s_{eval}}^{s_{max}} \xi(r_\perp, s)^{2h} ds, \quad (19)$$

where the factor of 2 comes from the fact that both components are symmetric. Here, we integrate the one-halo component over the more nearby regime out to s_{eval} and only integrate the twp-halo term beyond s_{eval} out to the maximum line-of-sight distance, thus excluding the regimes in which the models do not apply. For the two-component model, we choose fairly weak priors on unknown parameters based on physical arguments while following the same priors as described above for the parameters in the single power-law model:

$$\beta^{1h} > -3, \quad (20)$$

$$A > 0, \quad (21)$$

$$R_{cross} > 0. \quad (22)$$

We can then follow the same MCMC fitting procedure described above to determine the posteriors for the parameters in this model as well as the crossing radius, R_{cross}. These are shown in Figure 4. As before, we only fit data with $8 < \log M_\star/M_\odot < 10.5$ and use $M_0 = 10^{9.5} M_\odot$.

3.2.2. Other Two-component Models

While the single power-law clustering model does an adequate job reproducing the data on large spatial scales, its contribution is insufficient at $R_\perp \lesssim 200$ kpc, as can be seen in Figure 5 (pink curve). Furthermore, the primary goal of our study is to find the boundary between the CGM and IGM, and thus including a CGM component is essential for this purpose. We explored several candidate functional forms for this CGM component.

We first investigated a two-component model where each component is represented by a power law, inspired by the one-halo and two-halo terms that are used to model the clustering of galaxies. The 3D and projected forms of the two absorber–galaxy correlation functions are given by Equations (1) and (3), respectively, and the two-component correlation function is the sum of these parts. We also considered a model where the two-component correlation function is, in 3D, the maximum of the two power laws. This is similar to our chosen model, but with an inner power law rather than an inner Gaussian profile.

To rise above the outer power-law component at small radii, the inner power law has to be steeper. In practice, the two power-law indices turned out to be similar, yielding essentially the same result as a single power-law fit. This outcome is not unexpected: the enhancement in the incidence rate or surface density of gas near galaxies often does not resemble a steepening power law at small radii (Zhu et al. 2014; Lan 2020).

In those studies, the enhancement is better described by a function that declines gradually (compared to a power law) at small radii and quickly at large radii. The top-hat function, which has amplitude A inside a boundary and amplitude 0 outside the boundary, is an extreme example of this class. Our adopted Gaussian profile allows a smoother transition between the CGM-like and outer components of the model. However, we note that a fit to the data combining a inner 3D top-hat with an outer power law yields an $R_{cross}(M_*)$ that is effectively indistinguishable from the one that emerges from the Gaussian component model.

3.3. Model Comparison

In addition to comparing the two models to each other, Figure 5 compares the models to the empirical covering fraction as a function of impact parameter and mass. The data are shown in black with 1σ error bars. The single power-law model is shown in pink, while the two-component model is shown in purple. Both models recreate the covering fractions in all mass bins at all values of $R_\perp$, except for one data point in the $\log M_*/M_\odot = 9$–10 bin at $R_\perp \approx 200$ kpc. Moreover, the two models make different predictions at low $R_\perp$, except for in the lowest-mass bin ($\log M_*/M_\odot < 9$) where there is no discernible excess above the clustering signal. This does not preclude the presence of a CGM around these galaxies, but rather it suggests that we require more data at lower $R_\perp$ for galaxies with $\log M_*/M_\odot < 9$ to be able to constrain R_{cross} at these masses.

The two-halo-only model underpredicts the observed signal for galaxies at intermediate masses ($\log M_*/M_\odot = 9$–10). The two-component model does better for galaxies of $M_\star = 10^{9-10} M_\odot$ at the lowest impact parameters, where the single power-law model underestimates the covering fraction, although not significantly so. For $R_{cross} < 300$ kpc, one detects 52 H I systems where 46 systems are predicted. Assuming Poisson statistics, the two-halo-only model is consistent with the data at the 1σ level. Analogously to the one-halo term of galaxy–galaxy clustering, the data themselves do not require an

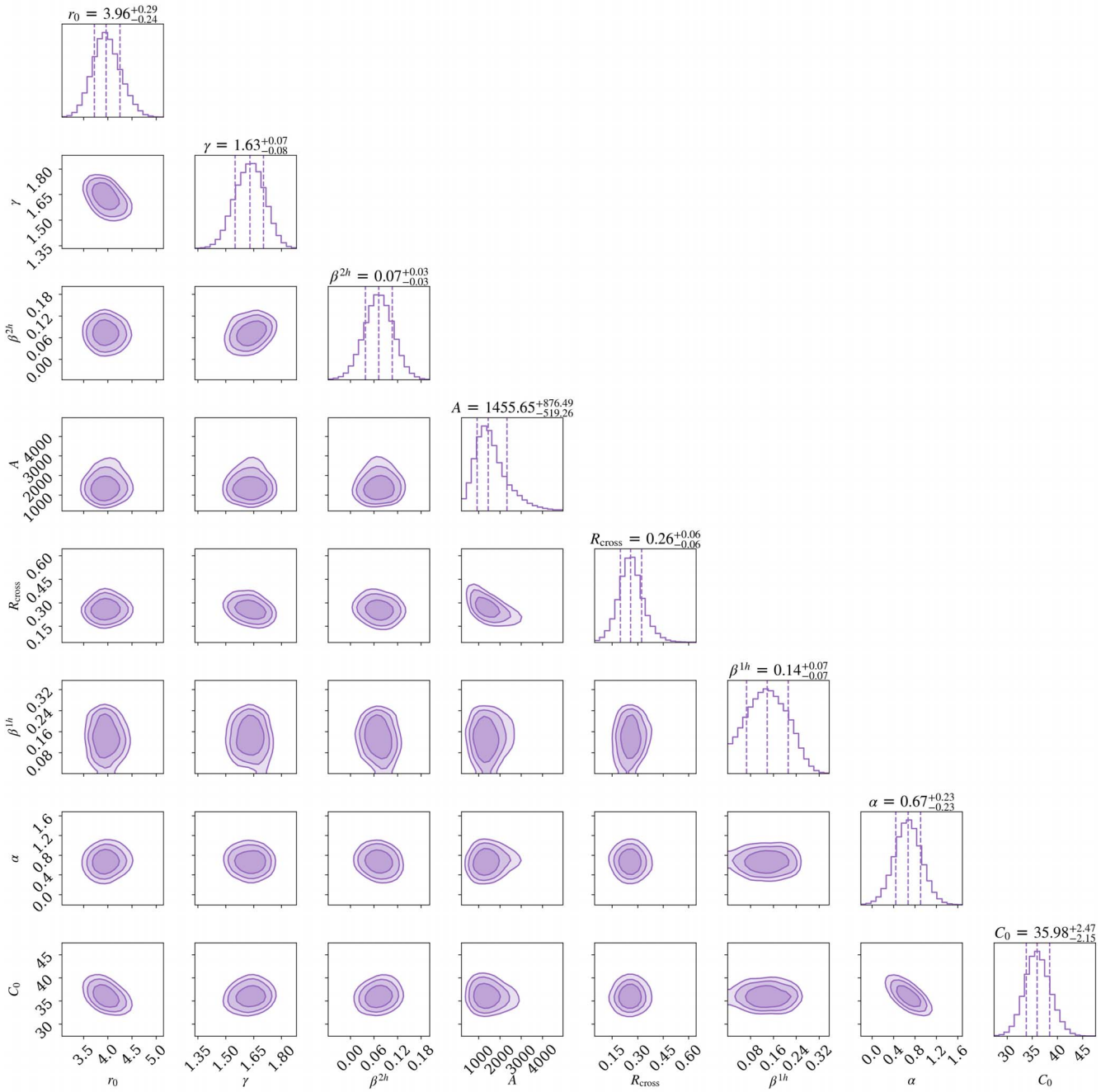

Figure 4. Posterior probabilities for the parameters in the two-component clustering model. We again recover a nonzero, positive mass dependence term in the two-halo absorber–galaxy clustering, β^{2h}, but we find an even stronger one-halo CGM clustering mass dependence $\beta^{1h} \simeq 0.14 \pm 0.07$.

enhanced covering fraction of H I absorption that we identify as the CGM.

We find the one-halo component has a stronger clustering mass dependence, $\beta^{1h} \simeq 0.14 \pm 0.07$, than the two-halo term, $\beta^{2h} \simeq 0.08 \pm 0.03$. We also find the two-halo clustering terms in each model to be internally consistent with each other, as seen in Figure 6.

4. Results

4.1. Clustering Mass Dependence

As seen in Figure 2, we find the clustering parameters to be $r_0 = 3.6 \pm 0.3$ cMpc and $\gamma = 1.6 \pm 0.5$. These values of r_0 and γ are consistent with those given by Tejos et al. (2014), who find $r_0 = 3.7 \pm 0.1$ cMpc and $\gamma = 1.7 \pm 0.3$. We also find a mass dependence of the absorber–galaxy clustering of $\beta^{2h} = 0.07^{+0.3}_{-0.2}$.

We find the two-component model fits the data better, as can be seen in Figure 5. Specifically, the two-component model better matches the covering fraction for galaxies of $M_\star > 10^{9-10} M_\odot$ at the lower impact parameters, where the single power-law model underestimates the covering fraction. In addition, we find the two-component model reproduces the mass dependence of the two-halo clustering term, $\beta^{2h} \simeq 0.07$, while also producing a stronger mass dependence of the one-halo clustering term, $\beta^{1h} \simeq 0.14$.

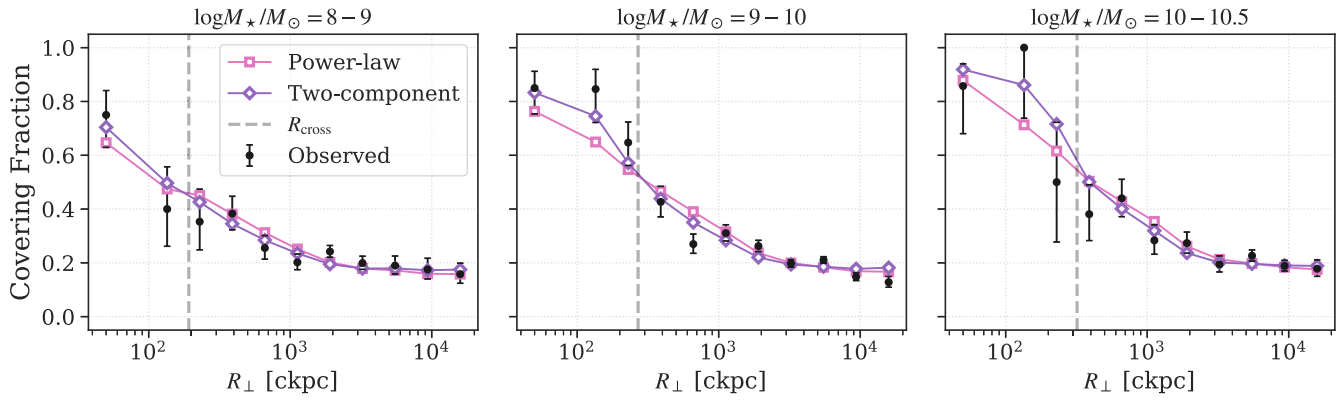

Figure 5. Comparison of our two models to the empirical covering fraction as a function of impact parameter in comoving kpc in mass bins of $10^{8-9}M_\odot$, $10^{9-10}M_\odot$, and $10^{10-10.5}M_\odot$. The data are shown in black with 1σ error bars. The single power-law model is shown in pink, while the two-component model is shown in purple. The vertical dotted line denotes $R_{\rm cross}$ in each mass bin. Both models recreate the covering fraction of the data in all mass bins except for the lowest-mass bin, where the clustering signal disappears. The two-component model provides a better match to the data for galaxies of $M_\star > 10^9 M_\odot$ at the lowest impact parameters, where the single power-law model underestimates the covering fraction.

4.2. Physically Motivated Extent of the CGM

As mentioned above, using the two-component model produces an estimate of $R_{\rm cross}$, a natural metric for the extent of the CGM. This 3D distance demarcates where the contribution to the clustering begins to be dominated by the CGM above the expected two-halo clustering, due to isolated galaxy halos traced by H I. $R_{\rm cross}$ can be viewed as the maximum radius to which an enhancement from the CGM could extend without overpredicting the data at large radii.

In Figure 7, we see $R_{\rm cross}$ (blue) compared with the spread in virial radii of the galaxy sample (gray filled region). The filled blue region represents the 1σ limits of the distribution in $R_{\rm cross}$, while the blue line denotes the median of this distribution. We find $R_{\rm cross}$ is $\sim 2.0 \pm 0.6 R_{\rm vir}$ for galaxies in the range $8 < \log(M_*/M_\odot) < 10.5$. The black crosses correspond to the values published in Paper I defined as the extent where there is a 50% chance to see H I absorption above 10^{14} cm^{-2}. The vertical dotted lines denote the mass range of $8 < \log(M_*/M_\odot) < 10.5$ that was used in our MCMC analysis. Above this range, we see a change in the relation of the virial radius with stellar mass, and below this mass range, we find little to no correlation between absorbers and galaxies. 5).

We also calculated the splashback radius, $R_{\rm sp}$, using the method from Diemer (2018) and encoded in the COLOSSUS[14] package. This radius denotes the location at which particles reach the apocenter of their first orbit. We find excellent agreement of $R_{\rm cross}$ with the results in Paper I, and $R_{\rm cross}$ neatly matches the splashback radius for galaxies in this mass range. We discuss these results in more detail below.

5. Discussion

Both of the models we investigate do an adequate job of recreating the cross-correlation signal at all impact parameters and masses $10^8 < M_\star < 10^{10.5} M_\odot$, as seen in Figure 5. It is not entirely clear that the single power-law model has any physically consistent meaning, however. Effectively, it would seem to signify that every time one measures H I absorption at the same redshift as a particular galaxy ($|\Delta v| < 500$ km s^{-1}), the absorption is always due to *another galaxy's CGM*. It is worth noting that we would conclude this for all galaxies, i.e., each has no CGM and only neighbors with a CGM. This is clearly impossible. The two-halo-only model for the CGM effectively breaks down when the galaxies lie within the halo under consideration, i.e., when they "mix." We cannot and do not try to distinguish between the two. However, our formalism does allow one to identify the outer extent of this "mixing."

The two-component model asserts that galaxies with $M_\star > 10^8 M_\odot$ have a CGM, an assumption that is motivated by previous survey results (e.g., Werk et al. 2013). Additionally, this model is able to better recreate the data—from the combined data sets of CGM^2 + CASBaH, which together represent the largest sample of galaxies with confirmed spectroscopic redshifts in the foregrounds of UV-bright QSOs with high-resolution absorption spectroscopy—both at smaller impact parameters and at $M_* > 10^9 M_\odot$.

The much larger number of galaxies at larger impact parameters drives the fit of the models to the data. There is, however, a $>1\sigma$ inconsistency between the two-halo-only model and the data at $R_\perp \sim 200$ kpc and for both models at $R_\perp \sim 600$ kpc in the $\log M_\star = 9$–$10\,M_\odot$ mass range. The latter inconsistency may be due to cosmic variance or the assumption that the absorber–galaxy measurements are independent and are not correlated, which would increase the scale of the error bars at $R_\perp \sim 600$ kpc.

5.1. Comparing the Mass Dependence of the Single- and Two-component Models

Our galaxy sample includes a large number of galaxies at low (<500 kpc) impact parameters, which allows us to better model the regime in which the two-halo galaxy clustering becomes dominated by the signal of galaxies that inhabit the same dark matter halo, the one-halo term. By separating these two terms in the manner presented here, we can disentangle the large-scale clustering as well as the contribution of the CGM to the 3D correlation of absorbers and galaxies.

Our analysis finds nearly identical terms for the mass dependence of the clustering at large scales, $\beta^{\rm 2h}$, as well as the contribution of absorbers at random, C_0 and α. We do find a stronger mass dependence in the one-halo term, $\beta^{\rm 1h}$, than at larger scales. This can be seen in Figure 5, where the correlation steepens in higher-mass bins.

[14] https://bdiemer.bitbucket.io/colossus/

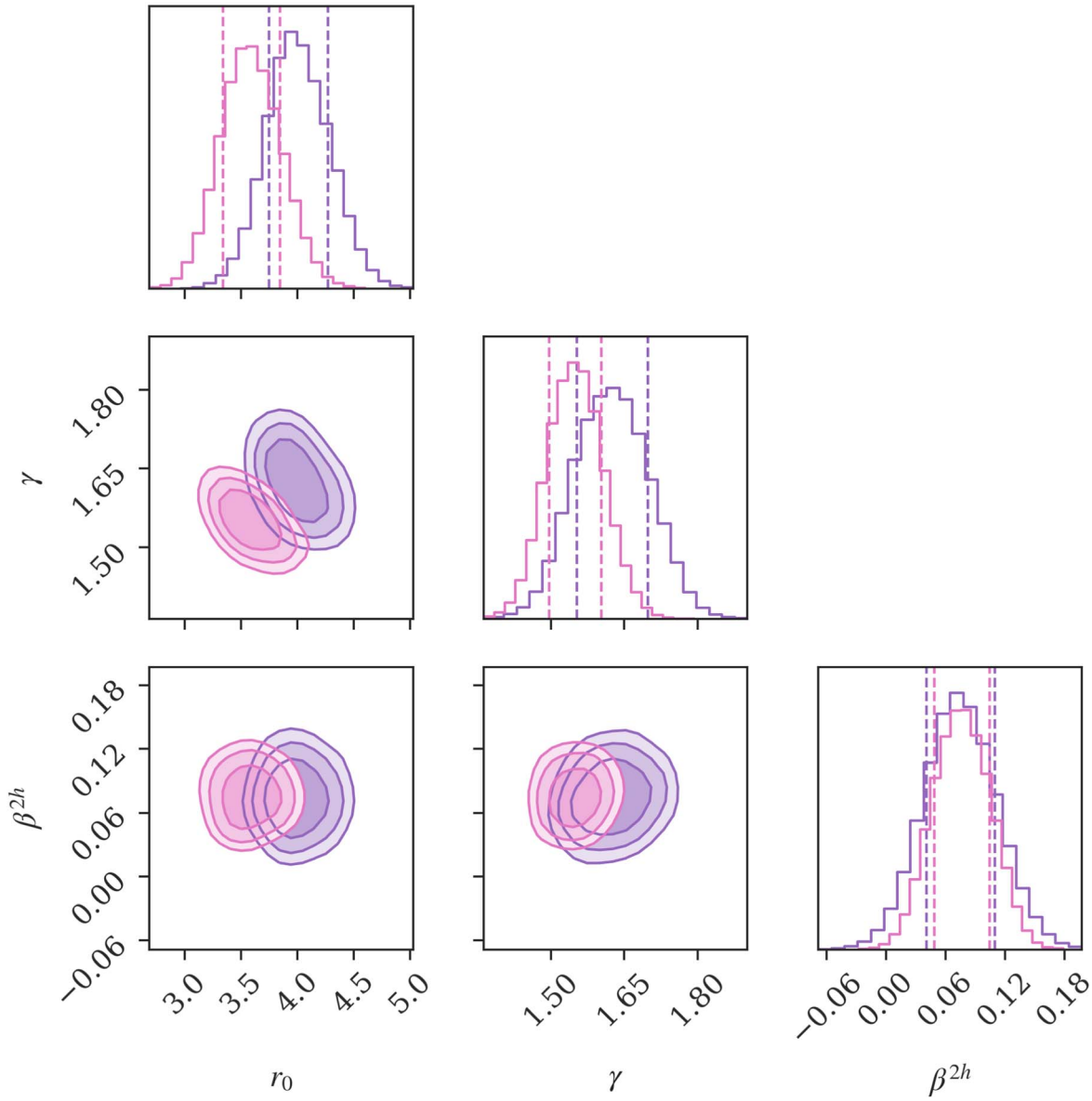

Figure 6. Comparison of the two-halo 3D cross-correlation posteriors between the two-component model ($r_0 = 3.99^{+0.28}_{-0.24}$ cMpc, $\gamma = 1.62 \pm 0.07$) and the single power-law model ($r_0 = 3.58^{+0.28}_{-0.24}$ cMpc, $\gamma = 1.55 \pm 0.05$). The two models are consistent with each other within the 1σ limits and have a power-law slope consistent with the absorber–galaxy 3D cross-correlation found in the literature (e.g., Tejos et al. 2014) of $\gamma = 1.7 \pm 0.1$.

5.2. Absorber–Galaxy Bias

Our covering fraction analyses provide an estimate of the galaxy–absorber correlation function, $\xi_{\rm ag}$ (Equation (1)). Here, we test if the mass dependence of $\xi_{\rm ag}$ outside the CGM is consistent with absorption systems and galaxies simply being two independent tracers of the same underlying dark matter distribution. Assuming both tracers have linear bias, $\xi_{\rm ag}$ should be equal to $b_a b_g \xi_{\rm DM}$, where b_a and b_g are the absorber and galaxy bias, respectively, and $\xi_{\rm DM}$ is the dark matter 3D correlation function. Following Tinker et al. (2010) (hereafter T10), we assume the dark matter correlation function can be described by a power-law function of radius with index $\gamma = 1.62$. We fix the power-law index in the $\xi_{\rm ag}$ determined by fitting a single power law to the data to this same value, with which it is consistent. With the above assumptions, $\xi_{\rm ag} = (r/r_0(M))^{-\gamma} = b_a b_g \xi_{\rm DM}(r)$. The radial dependence cancels, leaving the proportionality $r_0(M)^\gamma \propto b_a b_g$.

We show a scaled $r_0(M)^\gamma$ in Figure 8, along with the galaxy bias as a function of stellar mass from T10 and implemented in the `COLOSSUS` package (Diemer 2018). If b_a is constant and the assumptions stated above hold, $r_0(M)^\gamma$ should have the same mass dependence as galaxy bias. While there is a visually apparent difference between the galaxy bias and the best-fit $r_0(M)^\gamma$, this difference is not significant at a 2σ level, and so it is merely suggestive. If the difference is real, it could be a consequence of the H I mass per dark matter mass being a function of overdensity. Up to the overdensities at which $M_* = 10^{10.5}\,M_\odot$ galaxies tend to be found, this function would be increasing: H I would be less common in low-density regions than in higher-density filaments. This behavior would be consistent with theoretical expectations (e.g., Hui &

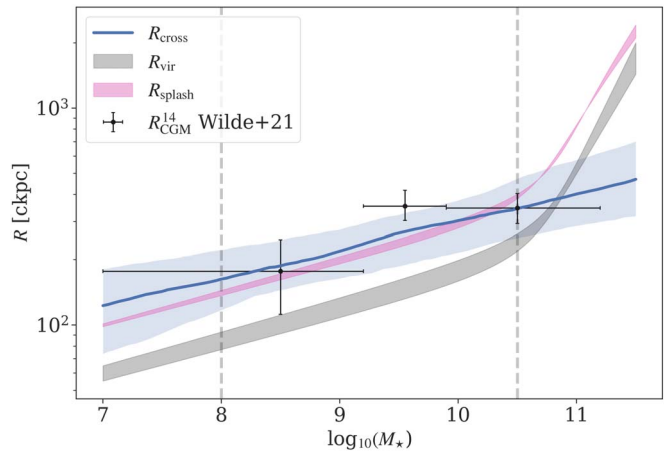

Figure 7. A comparison of $R_{\rm cross}$ with the virial radius ($R_{\rm vir}$, gray filled region) as well as the splashback radius ($R_{\rm splash}$, pink shaded region) of the galaxy sample. The filled regions in $R_{\rm vir}$ and $R_{\rm splash}$ denote the redshift range for the galaxies in our sample ($0.1 \lesssim z \lesssim 0.48$). The filled blue region represents the 1σ limits of the distribution in $R_{\rm cross}$, while the blue line denotes the median of this distribution. The black crosses correspond to the values published in PaperI. The vertical dotted lines denote the mass range of $8 < \log(M_*/M_\odot) < 10.5$ to which we limited the fitting in our MCMC analysis in Figure 5.

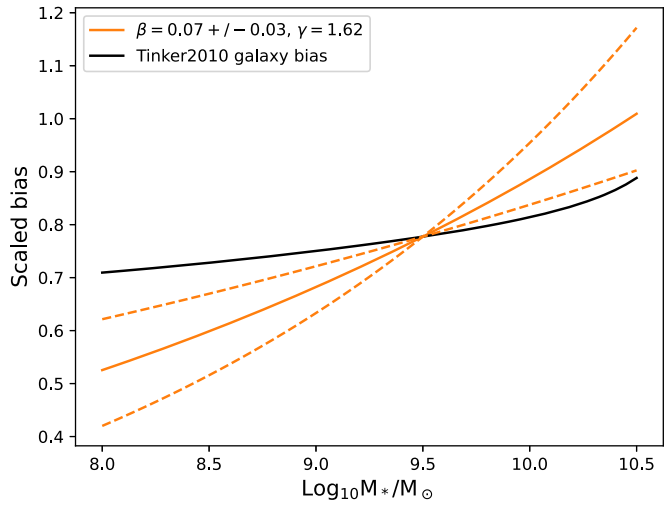

Figure 8. A comparison of the slopes of the relative bias as a function of mass derived from our analysis (orange) compared to Tinker et al. (2010) (T10, black). The dashed lines correspond to the ranges spanned by the 1σ limits in in $\beta^{\rm 2h}$. The relative bias, $r_0(M) \propto (M_\star/M_0)^{\gamma\beta}$, is normalized to the value of T10 at $\log M_\star/M_\odot = 9.5$. We find a steeper mass dependence than T10, but the significance of the difference is less than 2σ.

Gnedin 1997; Schaye 2001; Davé et al. 2010) and observations (e.g., Rudie et al. 2012; Burchett et al. 2020).

5.3. Comparison to Previous Work

One of the key aspects of this analysis is determining the mass dependence of the extent of CGM gas with $N_{\rm H\,I} > 10^{14}$ cm^{-2}, for which our model provides a direct metric, $R_{\rm cross}(M_\star)$. We compare our resulting $R_{\rm cross}(M_\star)$ to the method and results from Paper I in Figure 7. The results of Paper I, $R_{\rm CGM}^{14}$, which are based only on the CGM^2 survey, are shown as black crosses in the mass bins they span in that paper. We also compare the method used in that paper to determine $R_{\rm CGM}^{14}$, the radius at which the probability of detecting $N_{\rm H\,I} > 10^{14}$ cm^{-2} is >50%, calculated with the two-component model using the combined CGM^2 + CASBaH surveys, and we find it to be consistent within 1σ with our newer model for $R_{\rm cross}(M_\star)$. We find that our mass-dependent estimate of the extent of the CGM, $R_{\rm cross}(M_\star)$, corroborates the findings of Paper I that the $N_{\rm H\,I} > 10^{14}$ cm^{-2} extends to approximately twice the virial radius ($\sim 2 \pm 0.6 R_{\rm vir}$).

One of the main strengths of the CGM^2 + CASBaH sample is the large number of galaxies at small projected separations (<1 Mpc). This allows us to investigate the smaller-scale regime in more detail within the context of similar studies such as Tejos et al. (2014) (hereafter, T14), who use a single power-law model to measure the two-point correlation between H I and galaxies above $N_{\rm H\,I} > 10^{14}$ cm^{-2}. In that work, they break up their measurements into SF versus non-SF samples, whereas we do not. Our sample, however, is dominated by the more common SF galaxies, and we will compare our results to their SF sample. Comparing our cross-correlation results with T14, we find good agreement between the results in T14 ($r_0^{\rm T14} = 3.8 \pm 0.2$ Mpc and $\gamma = 1.7 \pm 0.1$) and the results from both models presented here ($r_0 = 3.99^{+0.28}_{-0.24}$ Mpc and $\gamma = 1.62 \pm 0.07$) as well as the single power-law model ($r_0 = 3.58^{+0.28}_{-0.24}$ Mpc and $\gamma = 1.55 \pm 0.05$). We find a mass dependence of this cross-correlation, however, as parameterized by $\beta^{\rm 2h}$.

Our results are slightly in tension with those of Momose et al. (2021), who find galaxies in the $10^{9-10} M_\odot$ range dominate their H I–galaxy cross-correlation signal. We find the largest-mass bin sample to have the most elevated covering fractions at low impact parameters.

5.4. Physical Extent of Galaxy Halos

Astronomers often use the virial radius as a means to describe the characteristic size of galaxy halos, and it is convenient to compare this to the extent of the gaseous galactic atmosphere as we have done here and in Paper I. The virial radius is typically defined in terms of the spherical overdensity mass definition, which is based on the radius that encloses an overdensity of 200 times the critical or mean density, i.e., R_{200c} and R_{200m}. Because the mean and critical densities are decreasing over cosmic time, this can lead to a pseudo-evolution, as pointed out in Diemer et al. (2013). In addition, subhalos show evidence of being stripped outside the virial radius of clusters (Behroozi et al. 2014).

An alternative physically motivated halo scale is the splashback radius, $R_{\rm sp}$ (Adhikari et al. 2014; Diemer & Kravtsov 2014; More et al. 2015). This radius effectively distinguishes infalling material from matter orbiting in the halo. We compare our results to the splashback radius in Figure 7 and find that our estimate of the extent of the H I CGM, $R_{\rm cross}$, neatly aligns with $R_{\rm sp}$ over the mass range $10^8 < M_\star/M_\odot < 10^{10.5}$. This result implies that $R_{\rm sp}$ is a better approximation of the CGM extent than the more commonly used virial radius.

O'Neil et al. (2021) compared $R_{\rm sp}$ as estimated from dark matter and gas profiles in the IllustrisTNG simulations and found that the gas $R_{\rm sp}$ is consistently smaller than the dark matter $R_{\rm sp}$. However, they were looking at much more massive halos, $M_{\rm halo} > 10^{13}$ $M_\odot$, in which shocks dominate the gas distribution. Nonetheless, the fact that $R_{\rm cross} \approx R_{\rm sp}$ at the mass ranges considered here ($M_{\rm halo} \sim 10^{10-12} M_\odot$) is intriguing. The halo mass-accretion rate generally sets whether $R_{\rm sp}$ exceeds

$R_{\rm vir}$; a rapid accretion rate will impact the growth of the gravitational potential well, leading to $R_{\rm sp} < R_{\rm vir}$. If the location of $R_{\rm cross}$ reflects the extent of orbiting gas in a halo, then our observational results imply a halo mass-accretion rate that is slow enough to keep the apocenters of orbiting structures at large radii.

Another way of defining the extent of the CGM is to use the boundary of the pressure-supported CGM. For galaxies with halo masses $\gtrsim 10^{11.5} M_{\odot}$ ($M_{\star} \approx 10^{9.8} M_{\odot}$), this pressure support comes from fact that the gas that has fallen into the gravitational potential well is virially shocked and cannot cool within a Hubble time (Binney 1977; Rees & Ostriker 1977; Silk 1977). For the galaxies in our survey, which are predominately below this halo mass, however, the gas would rapidly cool—and thus this pressure support might come from galactic winds. Fielding et al. (2017) and Lochhaas et al. (2018) show that supernovae winds with reasonable mass-loading efficiencies could shock the gas to distances past the virial radius and account for the survival of cool gas at these large radii. Using a more comprehensive model of the multiphase CGM, Fielding & Bryan (2022) show that SF in the galactic disk can slow cooling and accretion as part of a global preventive self-regulation mechanism. In addition, the winds can transport cold clouds to large radii, consistent with these constraints from our combined survey data.

6. Summary

Herein, we have examined the associations of galaxies with Lyα absorption $z < 0.48$, to explore the spatial profile of this gas and the mass dependence of the profile. Specifically, we have combined the CGM^2 and CASBaH H I measurements and constructed a catalog of 7244 absorber–galaxy pairs around 28 QSO sightlines (6589 absorber–galaxy pairs when we restrict our galaxy sample to galaxies with $8 < \log M_{\star}/M_{\odot} < 10.5$). The CGM^2 survey has better sampling of galaxies at low impact parameter, while CASBaH samples galaxies out to 20 cMpc. This allows us to characterize the H I profile via the covering fraction as a tracer of the gas.

1. By modeling the covering fraction as a power law with a mass-dependent length scale, we find good agreement with previous studies, such as T14, of our clustering amplitude and power-law slope parameters.
2. In Section 3.1, we find the clustering scale has a mass dependence with a power-law slope of $\beta^{\rm 2h} = 0.08 \pm 0.03$.
3. We compare the slope of our absorber–galaxy bias to the galaxy–dark matter bias of Tinker et al. (2010). The absorber–galaxy bias is a steeper function of galaxy mass than the galaxy–dark matter bias. However, this difference is only significant at a sub-2σ level.
4. We model the data with an exclusionary two-component model where we adopt an inner-CGM Gaussian profile to describe the data at smaller impact parameters and the customary two-halo single power-law model at larger impact parameters. This model faithfully reproduces the data for galaxies $M_{\star} > 10^{8} M_{\odot}$.
5. The two-component model allows us to calculate the crossover radius, $R_{\rm cross}(M_{\star})$, where the models are equal. $R_{\rm cross}(M_{\star})$ represents a soft upper estimate of the furthest impact parameter needed to optimally fit the inner-CGM component. We then use $R_{\rm cross}$ as an estimate of the extent of the CGM and find $R_{\rm cross}(M_{\star}) \approx 2.0 \pm 0.6 R_{\rm vir}$ for galaxies $10^{8} \leqslant M_{\star}/M_{\odot} \leqslant 10^{10.5}$. This metric is consistent with $R_{\rm CGM}^{14}$ from Paper I but provides a mass dependence. Additionally, we find excellent agreement between $R_{\rm cross}(M_{\star})$ and the splashback radius, $R_{\rm sp}$, for galaxies in this mass range.

We thank the referee for a constructive report that improved the clarity of this work. M.C.W., K.T., and J.K.W. acknowledge support for this work from NSF-AST 1812521, NSF-CAREER 2044303, and the Research Corporation for Science Advancement, grant ID No. 26842. Support for the CASBaH HST programs HST-GO-11741 and HST-GO-13846 was provided through grants from the Space Telescope Science Institute under NASA contract NAS5-26555.

Support for the CASBaH HST programs HST-GO-11741 and HST-GO-13846 was provided through grants from the Space Telescope Science Institute under NASA contract NAS5-26555.

The CGM^2 Survey would not have been possible without the substantial contributions from a dedicated group of UW undergraduate Student Quasar Absorption Diagnosticians, the Werk SQuAD, with over 50 individual undergraduate research assistants since 2016. The SQuAD confirmed all auto-fitted galaxy spectroscopic redshifts by eye, identified absorption systems along every quasar line of sight, and measured various spectroscopic properties (e.g., SFRs) of the nearly 1000 galaxies included in the survey. We are deeply grateful to work with such motivated and enthusiastic students.

ORCID iDs

Matthew C. Wilde https://orcid.org/0000-0003-1980-364X
Kirill Tchernyshyov https://orcid.org/0000-0003-0789-9939
Jessica K. Werk https://orcid.org/0000-0002-0355-0134
Todd M. Tripp https://orcid.org/0000-0002-1218-640X
Joseph N. Burchett https://orcid.org/0000-0002-1979-2197
J. Xavier Prochaska https://orcid.org/0000-0002-7738-6875
Nicolas Tejos https://orcid.org/0000-0002-1883-4252
Nicolas Lehner https://orcid.org/0000-0001-9158-0829
Rongmon Bordoloi https://orcid.org/0000-0002-3120-7173
John M. O'Meara https://orcid.org/0000-0002-7893-1054
Jason Tumlinson https://orcid.org/0000-0002-7982-412X
J. Christopher Howk https://orcid.org/0000-0002-2591-3792

References

Abolfathi, B., Aguado, D. S., Aguilar, G., et al. 2018, ApJS, 235, 42
Adhikari, S., Dalal, N., & Chamberlain, R. T. 2014, JCAP, 2014, 019
Bahcall, J. N., & Spitzer, L., Jr. 1969, ApJL, 156, L63
Behroozi, P. S., Wechsler, R. H., Lu, Y., et al. 2014, ApJ, 787, 156
Berg, M. A., Lehner, N., Howk, J. C., et al. 2023, ApJ, 944, 101
Bergeron, J. 1986, A&A, 155, L8
Bergeron, J., & Boissé, P. 1991, A&A, 243, 344
Binney, J. 1977, ApJ, 215, 483
Boquien, M., Burgarella, D., Roehlly, Y., et al. 2019, A&A, 622, A103
Bordoloi, R., Tumlinson, J., Werk, J. K., et al. 2014, ApJ, 796, 136
Borthakur, S. 2022, ApJ, 924, 123
Borthakur, S., Heckman, T., Tumlinson, J., et al. 2015, ApJ, 813, 46
Bouma, S. J. D., Richter, P., & Wendt, M. 2021, A&A, 647, A166
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Burchett, J. N., Elek, O., Tejos, N., et al. 2020, ApJL, 891, L35
Burchett, J. N., Tripp, T. M., Bordoloi, R., et al. 2016, ApJ, 832, 124
Burchett, J. N., Tripp, T. M., Prochaska, J. X., et al. 2015, ApJ, 815, 91
Burchett, J. N., Tripp, T. M., Prochaska, J. X., et al. 2019, ApJL, 877, L20

Butsky, I. S., Fielding, D. B., Hayward, C. C., et al. 2020, ApJ, 903, 77
Calzetti, D., Kinney, A. L., & Storchi-Bergmann, T. 1994, ApJ, 429, 582
Chabrier, G. 2003, PASP, 115, 763
Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, arXiv:1612.05560
Chen, H.-W., Prochaska, J. X., Weiner, B. J., Mulchaey, J. S., & Williger, G. M. 2005, ApJL, 629, L25
Cutri, R. M., Wright, E. L., Conrow, T., et al. 2013, Explanatory Supplement to the AllWISE Data Release Products, Technical Report
Dale, D. A., Helou, G., Magdis, G. E., et al. 2014, ApJ, 784, 83
Danforth, C. W., Keeney, B. A., Tilton, E. M., et al. 2016, ApJ, 817, 111
Davé, R., Anglés-Alcázar, D., Narayanan, D., et al. 2019, MNRAS, 486, 2827
Davé, R., Oppenheimer, B. D., Katz, N., Kollmeier, J. A., & Weinberg, D. H. 2010, MNRAS, 408, 2051
Dey, A., Schlegel, D. J., Lang, D., et al. 2019, AJ, 157, 168
Diemer, B. 2018, ApJS, 239, 35
Diemer, B., & Kravtsov, A. V. 2014, ApJ, 789, 1
Diemer, B., More, S., & Kravtsov, A. V. 2013, ApJ, 766, 25
Fielding, D., Quataert, E., McCourt, M., & Thompson, T. A. 2017, MNRAS, 466, 3810
Fielding, D. B., & Bryan, G. L. 2022, ApJ, 924, 82
Finn, C. W., Morris, S. L., Tejos, N., et al. 2016, MNRAS, 460, 590
Ford, A. B., Oppenheimer, B. D., Davé, R., et al. 2013, MNRAS, 432, 89
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Fumagalli, M., Prochaska, J. X., Kasen, D., et al. 2011, MNRAS, 418, 1796
Gimeno, G., Roth, K., Chiboucas, K., et al. 2016, Proc. SPIE, 9908, 872
Green, J. C., Froning, C. S., Osterman, S., et al. 2012, ApJ, 744, 60
Haislmaier, K. J., Tripp, T. M., Katz, N., et al. 2021, MNRAS, 502, 4993
Hennawi, J. F., & Prochaska, J. X. 2007, ApJ, 655, 735
Hook, I. M., Jørgensen, I., Allington-Smith, J. R., et al. 2004, PASP, 116, 425
Hui, L., & Gnedin, N. Y. 1997, MNRAS, 292, 27
Hummels, C. B., Bryan, G. L., Smith, B. D., & Turk, M. J. 2013, MNRAS, 430, 1548
Hummels, C. B., Smith, B. D., Hopkins, P. F., et al. 2019, ApJ, 882, 156
Kim, T. S., Wakker, B. P., Nasir, F., et al. 2021, MNRAS, 501, 5811
Lan, T.-W. 2020, ApJ, 897, 97
Lanzetta, K. M., Bowen, D. V., Tytler, D., & Webb, J. K. 1995, ApJ, 442, 538
Lehner, N., O'Meara, J. M., Fox, A. J., et al. 2014, ApJ, 788, 119
Liang, C. J., & Chen, H.-W. 2014, MNRAS, 445, 2061
Lo Faro, B., Buat, V., Roehlly, Y., et al. 2017, MNRAS, 472, 1372
Lochhaas, C., Thompson, T. A., Quataert, E., & Weinberg, D. H. 2018, MNRAS, 481, 1873
Meiring, J. D., Tripp, T. M., Prochaska, J. X., et al. 2011, ApJ, 732, 35
Momose, R., Shimasaku, K., Kashikawa, N., et al. 2021, ApJ, 909, 117
More, S., Diemer, B., & Kravtsov, A. V. 2015, ApJ, 810, 36
Morris, S. L., Weymann, R. J., Dressler, A., et al. 1993, ApJ, 419, 524
Moster, B. P., Naab, T., & White, S. D. M. 2013, MNRAS, 428, 3121
National Academies of Sciences, Engineering, and Medicine 2021, Pathways to Discovery in Astronomy and Astrophysics for the 2020s (Washington, DC: The National Academies Press)
Noll, S., Burgarella, D., Giovannoli, E., et al. 2009, A&A, 507, 1793
O'Neil, S., Barnes, D. J., Vogelsberger, M., & Diemer, B. 2021, MNRAS, 504, 4649
Oppenheimer, B. D., Davé, R., Katz, N., Kollmeier, J. A., & Weinberg, D. H. 2012, MNRAS, 420, 829
Peeples, M. S., Corlies, L., Tumlinson, J., et al. 2019, ApJ, 873, 129
Peeples, M. S., Werk, J. K., Tumlinson, J., et al. 2014, ApJ, 786, 54
Pillepich, A., Springel, V., Nelson, D., et al. 2018, MNRAS, 473, 4077
Prochaska, J. X., Burchett, J. N., Tripp, T. M., et al. 2019, ApJS, 243, 24
Prochaska, J. X., Weiner, B., Chen, H.-W., Mulchaey, J., & Cooksey, K. 2011, ApJ, 740, 91
Prochaska, J. X., Werk, J. K., Worseck, G., et al. 2017, ApJ, 837, 169
Putman, M. E., Zheng, Y., Price-Whelan, A. M., et al. 2021, ApJ, 913, 53
Rees, M. J., & Ostriker, J. P. 1977, MNRAS, 179, 541
Rudie, G. C., Steidel, C. C., & Pettini, M. 2012, ApJL, 757, L30
Rudie, G. C., Steidel, C. C., Pettini, M., et al. 2019, ApJ, 885, 61
Ryan-Weber, E. V. 2006, MNRAS, 367, 1251
Savage, B. D., Lehner, N., Wakker, B. P., Sembach, K. R., & Tripp, T. M. 2005, ApJ, 626, 776
Schaye, J. 2001, ApJ, 559, 507
Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, MNRAS, 446, 521
Silk, J. 1977, ApJ, 211, 638
Singh, P., Voit, G. M., & Nath, B. B. 2021, MNRAS, 501, 2467
Stinson, G. S., Brook, C., Prochaska, J. X., et al. 2012, MNRAS, 425, 1270
Tejos, N., Morris, S. L., Crighton, N. H. M., et al. 2012, MNRAS, 425, 245
Tejos, N., Morris, S. L., Finn, C. W., et al. 2014, MNRAS, 437, 2017
Thom, C., Tumlinson, J., Werk, J. K., et al. 2012, ApJL, 758, L41
Tinker, J. L., Robertson, B. E., Kravtsov, A. V., et al. 2010, ApJ, 724, 878
Tripp, T. M., Meiring, J. D., Prochaska, J. X., et al. 2011, Sci, 334, 952
Tripp, T. M., & Savage, B. D. 2000, ApJ, 542, 42
Tripp, T. M., Sembach, K. R., Bowen, D. V., et al. 2008, ApJS, 177, 39
Tumlinson, J., Peeples, M. S., & Werk, J. K. 2017, ARA&A, 55, 389
Tumlinson, J., Thom, C., Werk, J. K., et al. 2013, ApJ, 777, 59
Tumlinson, J., Werk, J. K., Thom, C., et al. 2011, ApJ, 733, 111
Verner, D. A., Barthel, P. D., & Tytler, D. 1994, A&AS, 108, 287
Villaescusa-Navarro, F., Genel, S., Anglés-Alcázar, D., et al. 2022, ApJS, 259, 61
Wakker, B. P., & Savage, B. D. 2009, ApJS, 182, 378
Werk, J. K., Prochaska, J. X., Thom, C., et al. 2013, ApJS, 204, 17
Werk, J. K., Prochaska, J. X., Tumlinson, J., et al. 2014, ApJ, 792, 8
Wijers, N. A., Schaye, J., & Oppenheimer, B. D. 2020, MNRAS, 498, 574
Wilde, M. C., Werk, J. K., Burchett, J. N., et al. 2021, ApJ, 912, 9
Woodgate, B. E., Kimble, R. A., Bowers, C. W., et al. 1998, PASP, 110, 1183
Zhu, G., Ménard, B., Bizyaev, D., et al. 2014, MNRAS, 439, 3139