Connection between Galaxies and H I in Circumgalactic and Intergalactic Media: Variation according to Galaxy Stellar Mass and Star Formation Activity

Rieko Momose1, Ikkoh Shimizu2,3, Kentaro Nagamine4,5,6, Kazuhiro Shimasaku1,7, Nobunari Kashikawa1,7, Haruka Kusakabe8

1 Department of Astronomy, School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan; momose@astron.s.u-tokyo.ac.jp
2 Shikoku Gakuin University, 3-2-1 Bunkyocho, Zentsuji, Kagawa 765-0013, Japan
3 National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan
4 Theoretical Astrophysics, Department of Earth and Space Science, Graduate School of Science, Osaka University, 1-1 Machikaneyama, Toyonaka, Osaka 560-0043, Japan
5 Kavli-IPMU (WPI), The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8583, Japan
6 Department of Physics & Astronomy, University of Nevada, Las Vegas, 4505 South Maryland Parkway, Las Vegas, NV 89154-4002, USA
7 Research Center for the Early Universe, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
8 Observatoire de Genève, Université de Genève, 51 Chemin de Pégase, 1290 Versoix, Switzerland

Received 2020 February 17; revised 2020 December 21; accepted 2021 January 29; published 2021 April 20

Abstract

This paper systematically investigates the comoving megaparsec-scale intergalactic medium (IGM) environment around galaxies traced by the Lyα forest. Using our cosmological hydrodynamical simulations, we investigate the IGM–galaxy connection at z = 2 by two methods: (i) cross-correlation analysis between galaxies and the fluctuation of Lyα forest transmission (δf) and (ii) comparison of the overdensity of neutral hydrogen (H I) and galaxies. Our simulations reproduce observed cross-correlation functions (CCFs) between Lyα forest and Lyman-break galaxies. We further investigate the variation of the CCF using subsamples divided by dark matter halo mass (M_{DM}), galaxy stellar mass (M_∗), and star formation rate (SFR) and find that the CCF signal becomes stronger with increasing M_{DM}, M_*, and SFR. The CCFs between galaxies and gas density fluctuation are also found to have similar trends. Therefore, the variation of δf–CCF depending on M_{DM}, M_*, and SFR is due to varying gas densities around galaxies. We find that the correlation between galaxies and the IGM HI distribution strongly depends on M_{DH} as expected from linear theory. Our results support the ΛCDM paradigm, confirming a spatial correlation between galaxies and IGM H I, with more massive galaxies being clustered in higher-density regions.

Unified Astronomy Thesaurus concepts: Galaxy evolution (594); Galaxy formation (595); Intergalactic medium (813); Lyman alpha forest (980); Hydrodynamical simulations (767); Large-scale structure of the universe (902); Circumgalactic medium (1879)

1. Introduction

The standard picture of galaxy formation within the gravitational instability paradigm indicates that galaxy formation and evolution is closely linked to the surrounding gas called the circumgalactic medium (CGM) and intergalactic medium (IGM; e.g., Rauch 1998; Mo et al. 2010). The inflowing gas from the IGM provides the fuel for star formation in galaxies and promotes the growth of galaxies and their central supermassive black holes (SMBHs). As important as the inflow is the energetic feedback from massive stars and SMBHs, which blows the gas away into the CGM and IGM. Therefore, determining the megaparsec-scale distribution of gas as a function of time and space is quite important for understanding galaxy formation and evolution.

The connection between the CGM/IGM and galaxies has been studied using Lyα forest absorptions in quasar spectra. The most common method to clarify the IGM/CGM–galaxy connection is cross-correlation analysis between Lyα forest absorption and galaxies (e.g., Adelberger et al. 2003, 2005; Chen et al. 2005; Ryan-Weber 2006; Faucher-Giguère et al. 2008; Rakic et al. 2011, 2012; Rudie et al. 2012; Font-Ribera et al. 2013; Prochaska et al. 2013; Tejos et al. 2014; Bielby et al. 2017). Alternatively, comparisons of galaxy and neutral hydrogen (HI) overdensities have also been discussed in the literature (Mawatari et al. 2017; Mukae et al. 2017, 2020). Specific high-density regions with abundant HI gas and highly clustered galaxies have also been studied (e.g., Cai et al. 2016; Lee et al. 2016; Mawatari et al. 2017; Hayashino et al. 2019). All of the above studies have revealed that galaxy distribution correlates with the IGM up to tens of comoving megaparsec (Mpc) scales.

The physical properties of HI gas in the CGM or IGM have been studied in detail theoretically (e.g., Meiksin 2009; Fumagalli et al. 2011; van de Voort & Schaye 2012; van de Voort et al. 2012; Meiksin et al. 2014, 2017; Rahmati et al. 2015), aided by powerful cosmological hydrodynamical simulations such as EAGLE (Schaye et al. 2015), Illustris (Genel et al. 2014; Vogelsberger et al. 2014a, 2014b; Sijacki et al. 2015), and IllustrisTNG (Villaescusa-Navarro et al. 2018; Nelson et al. 2019). Turner et al. (2017) have presented the median H I optical depth versus the line of sight (LoS) or transverse distance around galaxies in an EAGLE simulation and found that the H I optical depths are sensitive to dark matter halo mass. Sorini et al. (2018) have compared the mean Lyα absorption profiles as a function of the transverse distance from galaxies in both observations and simulations and showed a reasonable match between them beyond 2 proper Mpc (pMpc) but significant differences at 0.02–2 pMpc. A subsequent study by Sorini et al. (2020) has investigated the impact of feedback...
around mock quasars and found that stellar feedback is the dominant driver (rather than active galactic nuclei (AGNs)) in determining the average properties of the CGM.

The correlation between galaxies and IGM morphology has also been examined in the literature. Martizzi et al. (2020) have demonstrated that galaxies with lower stellar masses than the median are in voids and sheets of the IGM, whereas galaxies with higher stellar masses are more likely to be in filaments and knots of the IGM with higher gas densities. In addition, the correlation of mock Lyα forest absorption spectra with galaxy overdensity has been studied and compared with observations (e.g., Stark et al. 2015; Miller et al. 2019). Although these theoretical studies provide further evidence of a strong link between the CGM/IGM and galaxies, there are many aspects that are still unclear, and our understanding is still insufficient.

Why do we need to study the distribution of H I and its bias beyond the well-established linear theory of the ΛCDM model? One of the main reasons is that the bias between baryons, galaxies, halos, and dark matter is nontrivial, and it is of substantial interest for future projects of IGM tomography, such as the Subaru PFS, Euclid, DESI, and 21 cm cosmology by the Square Kilometre Array. Understanding nonlinear bias at intermediate and small scales is becoming increasingly more important for baryon acoustic oscillation studies of precision cosmology. What we are trying to achieve in this paper is to dig deeper into this nonlinear bias between baryons (particularly H I), galaxies, and dark matter, beyond the simple halo bias, and evaluate the impact of limited observational data on the cross-correlation function (CCF). Feedback effects from galaxies can alter the H I distribution in the intergalactic space due to ionization, and it is crucial to use a cosmological hydrodynamic simulation that solves heating and cooling self-consistently with the impact of the UV background radiation field. As observations can only trace a certain phase of gas, it is unclear if the simple bias of linear theory is applicable to H I. In fact, it has been shown that H I has a complicated scale-dependent bias by many authors (e.g., Villaescusa-Navarro et al. 2018; Ando et al. 2019, 2020; Sinigaglia et al. 2020). It is thus crucial to examine the IGM H I–galaxy connection more deeply, including the dependence on galaxy mass and galaxy population. Current observational data are still very limited to study the CCF, in terms of both the survey volume of Lyα forest absorption spectra and the number of galaxies with spectroscopic redshifts (Momose et al. 2021, hereafter M21). In comparison, numerical simulations provide an alternative opportunity to study the IGM H I–galaxy connection, and this paper presents such systematic investigations for the first time. Our results certainly include the effect of halo bias, but they also reflect the intricate nonlinear bias between H I, galaxies, and dark matter, beyond the simple linear theory.

In order to unveil the gas distribution on Mpc scales around galaxies, we systematically investigate it using numerical simulations. We aim to (1) establish the methodology for evaluating the CGM/IGM–galaxy connection, (2) examine its dependence on galaxy properties, and (3) verify the cause of dependences using GADGET3-Osaka cosmological hydrodynamic simulations (Shimizu et al. 2019). What is new in this paper is a systematic investigation of the IGM–galaxy connection depending on several galactic properties. We make subsamples based on galaxies’ stellar mass $M_*$, halo mass $M_{\text{BH}}$, star formation rate (SFR), and specific SFR (sSFR = SFR/$M_*$) and compare the CCFs. Calculating the CCFs in three dimensions taking various galaxy properties into account is also a novel aspect of this work. Studying the IGM–galaxy correlation with the same resolution as the latest observations and comparing the simulation results with them is clearly a worthwhile exercise to check the validity of the ΛCDM model because the application of linear theory to the IGM H I–galaxy connection has not yet been fully investigated. In this study, we particularly focus on statistical comparisons between simulations and observations. For a fair comparison, we use the same parameters as in our companion observational paper (M21). As we describe in more detail in M21, for observations, we use the COSMOS Lyα Mapping And Tomography Observations (CLAMATO), which is publicly available Lyα forest 3D tomography data (Lee et al. 2014, 2016, 2018), and several other catalogs in the archives.

This paper is organized as follows. We introduce our numerical simulations in Section 2 and the methodology for examining the IGM–galaxy connection in Section 3. The results and discussions are presented in Sections 4 and 5, respectively. Finally, we give our summary in Section 6. In the appendix, we discuss the dependences of our results on redshift width, redshift uncertainty, sample size, and cosmic variance. We note that “cosmic web” and “IGM” are used specifically for those traced by H I gas unless otherwise specified in this paper. In addition, we mainly use $h^{-1}$ Mpc in comoving units in the following sections.

2. Simulations

In this paper, we use cosmological hydrodynamic simulations performed with GADGET3-Osaka (Shimizu et al. 2019), which is a modified version of the TreePM smoothed particle hydrodynamics (SPH) code GADGET3 (originally described in Springel 2005). Some physical processes important for galaxy formation such as star formation, supernova (SN) feedback, and chemical enrichment have been implemented and described in detail by Shimizu et al. (2019). Our simulations reproduce various observational results such as the stellar mass function, SFR function, stellar-to-halo-mass ratio, and cosmic star formation history within observational uncertainties at $z \geq 2$ (I. Shimizu et al., in preparation).

Here, we briefly describe our simulations, which employ $N = 2 \times 512^3$ particles in a comoving volume of $(100 h^{-1} \text{ Mpc})^3$. The particle masses of dark matter and gas are $5.38 \times 10^5 M_\odot$ and $1.00 \times 10^8 h^{-1} M_\odot$, respectively. The gravitational softening length is set to $8 h^{-1}$ kpc in comoving units. For SPH particles, the smoothing length is allowed to be 10% of the softening length ($800 h^{-1}$ pc in comoving units) in regions where baryons dominate over dark matter. This means that at $z = 3$ the physical maximum resolution for gas may reach 200 pc (proper) in our simulations. In this study, we pay particular attention to the physical properties of gas at $r > 100 h^{-1}$ kpc from galaxies, and our resolution is sufficient for this study.

Star particles are generated from gas particles when a set of criteria are satisfied. Note that the mass of gas particles changes over time due to star formation and stellar feedback (by SNe and AGB stars). As described in detail in Shimizu et al. (2019), we employ the CELib package (Saitoh 2017), which can treat the time- and metallicity-dependent metal yields and energies from SNeII, SNeIa, and AGB stars, allowing for a more realistic chemical evolution of simulated galaxies. In order to identify the simulated galaxies, we run a friends-of-friends (FoF) group finder with a comoving linking length of 0.2 in
units of the mean particle separation to identify groups of dark matter particles as dark matter halos. We then identify gravitationally bound groups of a minimum of 32 particles (dark matter + SPH + star) as substructures (subhalos) in each FoF group using the SUBFIND algorithm (Springel et al. 2001), which is a standard practice in many simulation works. We regard substructures that contain at least five star particles as our simulated galaxies. Moreover, we define the most massive galaxy in a halo as the central galaxy and the rest as satellite galaxies. We also calculate the virial halo mass \( M_{\text{vir}} \), which is defined by the total enclosed mass inside a sphere with 200 times the critical density of the universe. This means that the member galaxies (the central and satellite galaxies) in a dark matter halo have the same \( M_{\text{vir}} \), even though the substructures can have different subhalo masses as calculated by SUBFIND.

Note that each gas (star) particle has some associated physical properties such as mass, SFR, and metallicity. In our SPH simulation, these values are automatically computed following hydrodynamic and gravitational interactions, rather than given by hand, which is a fundamental difference from semianalytic models of galaxy formation. The properties (gas mass, stellar mass \( M_{*} \), and SFR) of a simulated galaxy are defined by the summation of these quantities in each subhalo.

In order to directly compare our simulations and observations, we create the light-cone output of gas particles and galaxies by connecting 10 simulation boxes of different redshifts following our previous work (Shimizu et al. 2012, 2014, 2016). The redshift range of our light-cone output is \( z \sim 1.8 \) to 3.1, which can cover the redshift range of recent \( \text{Ly}\alpha \) absorption line surveys (e.g., CLAMATO; Lee et al. 2014, 2016, 2018) and future PFS \( \text{Ly}\alpha \) absorption surveys (Takada et al. 2014). We then randomly shift and rotate each simulation box so that the same objects do not appear multiple times on a single LoS at different epochs.

With this light-cone output, we calculate the \( \text{Ly}\alpha \) optical depth \( \tau (\text{Ly}\alpha) \) along the LoS. First, we calculate the important physical quantities, \( A_\text{grid}(x) \), at each grid point \( x \) along the LoS, such as the H I density, LoS velocity, and temperature, as follows:

\[
A_\text{grid}(x) = \sum_j \frac{m_j}{\rho_j} A_j W(r, h_j) , 
\]

(1)

where \( A_j, m_j, \rho_j, \) and \( h_j \) are the physical quantity of concern, gas particle mass, gas density, and smoothing length of the \( j \)th particle, respectively. \( W \) is the SPH kernel function, and \( r \) is the distance between LoS grid points and gas particles. For simplicity, the grid size (\( d l \)) is set to a constant value of 100 \( h^{-1} \) kpc in comoving units, which corresponds to 10 km s\(^{-1} \) in the velocity space at \( z \sim 2 \). This is a higher resolution than that of any of the relevant \( \text{Ly}\alpha \) observations. Then, we calculate the \( \text{Ly}\alpha \) optical depth \( \tau (\text{Ly}\alpha) \) using these physical values at each grid point as follows:

\[
\tau (\text{Ly}\alpha) = \frac{\pi e^2}{m_e c} \sum_j f_{ij} \phi(x - x_j) n_{\text{HI}}(x_j) dl , 
\]

(2)

where \( e, m_e, c, f_{ij}, n_{\text{HI}}, \) and \( x_j \) are the electron charge, electron mass, speed of light, absorption oscillator strength, H I number density, and \( j \)th grid point location, respectively. \( \phi \) is the Voigt profile, and we use the fitting formula of Tasitsiomi (2006) without direct integration. In this study, after making the high-resolution LoS data, we reduce our resolution by coarse-graining the grid size to match the observations: 1024 (=32\(^2 \)) LoSs are drawn with regularly spaced intervals. The mean separation of each LoS is 3.3 \( h^{-1} \) Mpc, which is similar to that in the CLAMATO survey.

Finally, we note that AGN feedback is not considered in the simulation that we use for this paper. We note that the impact of AGN feedback on \( \text{H}\alpha \) distribution is still under debate. For example, Sorini et al. (2020) have found that AGN feedback has only little impact on the surrounding CGM in general, and SN feedback has more dominant effects. Besides, statistical results of the \( \text{H}\alpha \)-galaxy connection obtained from the same simulations in this study have shown good agreement with current observations (Nagamine et al. 2020). Some observations have found the proximity effect around quasars (e.g., Mukae et al. 2020; Momose et al. 2021), which requires full radiative transfer calculations to compare with simulations. Therefore, we defer the study of AGN feedback using radiation transfer to future work.

### 3. Methodology

One of the main purposes of this study is to compare results with CLAMATO, which is 3D tomography data of \( \text{Ly}\alpha \) forest transmission fluctuations \( \delta T \), see the definition below over \( 2.05 \leq z < 2.55 \) in 0.157 deg\(^2 \) of the COSMOS field (Scoville et al. 2007; Lee et al. 2016, 2018). CLAMATO consists of \( 60 \times 48 \times 876 \) pixels corresponding to a \( 30 \times 24 \times 436 h^{-1} \) Mpc cube with a pixel size of 0.5 \( h^{-1} \) Mpc. The average separation of background galaxies for measuring \( \text{Ly}\alpha \) absorption is \( [2.61, 3.18] h^{-1} \) Mpc in the [R.A., decl.] directions, and the separation in the LoS direction is \( 2.35 h^{-1} \) Mpc at \( z \sim 2.3 \). (See M21 for a more detailed comparison with CLAMATO.)

To produce a similar data cube of \( \delta T \) from our LoS data, we first evaluate the \( \text{Ly}\alpha \) transmission fluctuation \( \delta T \) in each LoS pixel by

\[
\delta T = \frac{F(x)}{\langle F(x) \rangle} - 1 , 
\]

(3)

where \( F(x) = \exp[-\tau (\text{Ly}\alpha)(x)] \) is the \( \text{Ly}\alpha \) flux transmission and \( \langle F(x) \rangle \) is the cosmic mean transmission. We adopt the following value derived by Faucher-Giguère et al. (2008):

\[
\langle F(x) \rangle = \exp[-0.00185(1 + z)^{3.92}] , 
\]

(4)

because it is used in CLAMATO. Additionally, we use the same setup used in CLAMATO with a Hubble constant \( h = 0.7 \) and a redshift coverage of \( 2.05 \leq z \leq 2.55 \).

In the following two subsections, we present the methods for two analyses: (i) cross-correlation and (ii) overdensity analysis.

#### 3.1. Cross-correlation Analysis

Cross-correlation analysis is often used in the literature to characterize the correlation between galaxies and the CGM/IGM (e.g., Adelberger et al. 2005; Tejos et al. 2014; Croft et al. 2016, 2018; Bielby et al. 2017). In this study, we adopt the following definition of the CCF:

\[
\xi_{\text{CF}}(r) = \frac{1}{N(r)} \sum_{i=1}^{N(r)} \delta_{\text{CF},i} - \frac{1}{M(r)} \sum_{j=1}^{M(r)} \delta_{\text{ran},j} 
\]

(5)
where $\xi_{gf}$ is the CCF between $\delta F$ and galaxies, $\delta_{g,i}$ and $\delta_{r,j}$ are the values of $\delta F$ for pixel $i$ and $j$ at distance $r$ from galaxies and random points (Croft et al. 2016), and $N(r)$ and $M(r)$ are the numbers of pixel–galaxy and pixel–random pairs in the bin with distance $r$.

To calculate CCFs, we prepare two LoS data with different LoS grid resolutions. One is LoS data with the original resolution of comoving $0.1 h^{-1}$ Mpc. The other is lower-resolution data with a coarse-grained grid size of $0.4 h^{-1}$ Mpc (as described in Section 2) to match the CLAMATO resolution of $0.5 h^{-1}$ Mpc at $z = 2.35$. We call this latter, lower-resolution data set “LoS-4.” For comparison with observations, we use the LoS-4 data set and calculate CCFs at $1–50 h^{-1}$ Mpc scales around galaxies.

At the same time, we use the original LoS data set to derive CCFs at $r = 0.16–1 h^{-1}$ Mpc, because the redshift resolution of LoS-4 data is greater than the smallest radius for CCF calculation. Hereafter, we refer to the scales of $r < 1 h^{-1}$ Mpc and $r \geq 1 h^{-1}$ Mpc as the “CGM regime” and “IGM regime” as indicated in Figure 1(e1).

The $\xi_{gf}$ value is evaluated in each shell with a thickness of $\Delta \log (r / h^{-1} \text{Mpc}) = 0.2$ and 0.1 for the CGM and IGM regimes, respectively. We confirm that our CCFs by the LoS and LoS-4 data are smoothly connected at $r = 1 h^{-1}$ Mpc within the error. We perform jackknife resampling by leaving one object out and calculating the $\xi_{gf}$ value and adopt the jackknife standard error as the error in each shell.

To examine how the CCF varies according to the physical properties of galaxies, we divide the galaxy sample into three subsamples according to $M_*$, $M_{\text{DH}}$, SFR, and sSFR. The number of galaxies in each subsample and its name are summarized in Table 1.

### 3.2. Overdensity Analysis

Another analysis that we perform in this paper is a direct comparison of the IGM absorption and galaxy overdensity within cylinders along the LoS direction. This method was originally proposed by Mukae et al. (2017), and we call it “ overdensity analysis.”

We first generate a 2D LoS map by binning the LoS data with $\Delta \tau = 0.032$, which corresponds to $27.9 h^{-1}$ Mpc centered on the local minima and maxima of the IGM absorption signal. We determine the pixel positions of the local minima/maxima, we first mark the positions of the local minima/maxima of $\delta F$ within a few $h^{-1}$ Mpc scale and then repeat the same procedure to identify the minima/maxima on even smaller scales.

The galaxy overdensity is also computed together with $\langle \delta F \rangle$ in the same cylinders as follows:

$$\Sigma_{\text{gal}} = \frac{N_{\text{gal}}}{\langle N_{\text{gal}} \rangle} - 1, \quad (6)$$

where $N_{\text{gal}}$ and $\langle N_{\text{gal}} \rangle$ are the exact number of galaxies and the mean number of galaxies in the cylinder, respectively. Note that we calculate $N_{\text{gal}}$ and $\langle N_{\text{gal}} \rangle$ for each of the three stellar mass–divided subsamples given in Table 1. Since we first generate a 2D LoS map, the galaxy overdensity computed above can be regarded as the galaxy surface density, and thus we denote it as $\Sigma_{\text{gal}}$.

We also perform overdensity analysis for randomly selected positions to examine possible bias due to the positioning of the cylinders (see also Mukae et al. 2017).

### 4. Results

#### 4.1. Cross-correlation Analysis

#### 4.1.1. Ly$\alpha$ Absorption Fluctuation

The CCFs of all galaxies and subsamples are shown in Figure 1 (top). We also present the sign-flipped CCF plotted in log-scale in Figure 1 (bottom) for the discussion in Sections 4.1.2 and 5.1. We detect a CCF signal up to $r \sim 40 h^{-1}$ Mpc, which is in good agreement with observations by Adelberger et al. (2005) and Bielby et al. (2017).

Further investigations of the CCFs for four subsamples (divided by $M_*$, $M_{\text{DH}}$, SFR, and sSFR) are presented in Figures 1(a1)–(d1). Most of the CCFs show a monotonic increase from the center to $r = 20–50 h^{-1}$ Mpc, except for the $M_{\text{DH}}$–13 and $M_{\text{DH}}$–10 samples, which show irregular shapes. Considering our tests for CCF reproducibility by a small sample size in Appendix D, the swelling at $r \sim 0.4 h^{-1}$ Mpc in $M_{\text{DH}}$–10 can be attributed to the small sample size. For $M_{\text{DH}}$–13, we regard the slight bump at $r = 0.3–0.8 h^{-1}$ Mpc as a real feature.

We find that there is a clear tendency of the CCF signal depending on the subsample, except for the sSFR subsamples. The trend is that the CCF signal becomes stronger with increasing galaxy masses and SFRs, but the SFR–(v) subsamples do not follow this trend. A turnover radius where $\xi_{gf}$ reaches about zero also shows a trend for galaxies in the $M_*$ and $M_{\text{DH}}$ subsamples (see Figures 1(a2) and (b2)), which a sample with a stronger CCF signal drops rapidly to zero at a smaller radius—hereafter we call this the “turnover radius.” For the SFR samples, however, the CCF signal does not seem to correlate with the turnover radius. On the other hand, the sSFR samples do not show any obvious trend in their CCFs. Nonetheless, the turnover radius of the sSFR samples increases with an increasing sSFR. Likewise, Turner et al. (2017) have demonstrated the halo mass dependence of the median $\tau_{\text{HI}}$ as a function of distance from their modeled galaxies. Meiksin et al. (2017) have investigated $\delta F$ for galaxies in $M_{\text{DH}}$ subsamples against the projected impact parameter and found an increase in $\delta F$ with halo mass. Observationally, Chen et al. (2005) have measured the two-point cross-correlation $\xi_{ga}$ between Ly$\alpha$ absorbers and absorption line–dominated or emission line–dominated galaxies, which are presumably massive early-type and star-forming galaxies, and presented a different amplitude of $\xi_{ga}$ (see also Chen & Mulchaey 2009). Wilman et al. (2007) have also possibly found differences in the cross-correlation signal between absorption line–dominated and emission line–dominated galaxies (see their Figure 4), although they did not claim that they were statistically significant detections.
Figure 1. (Top) CCFs obtained from our simulation as a function of radius in comoving units. Dashed vertical and horizontal lines represent half of the pixel size in the transverse direction and the mean separation of the LoS. The definitions of the CGM and IGM regimes used in this paper are also shown in panel (e1). Panels (a1)–(d1): CCFs for each subsample divided by $M_{\star} - 11, M_{\star} - 12, M_{\star} - 10, M_{\star} - 9$. Panels (a2)–(e2): CCFs for each subsample divided by $M_{\star}$, $M_{\text{halo}}$, SFR, and sSFR. Panel (e1): the CCF calculated using all galaxies. Blue circles and squares indicate observational estimates of the CCFs between $\delta_F$ and Lyman-break galaxies (LBGs) from Adelberger et al. (2005) and Bielby et al. (2017), respectively. (Bottom) Sign-flipped CCFs of panels (a1)–(e1) plotted in log-scale. The vertical dashed lines in panels (a2)–(e2) are the same as those in panels (a1)–(e1).
To characterize our CCFs, we fit them with a power law:

$$\xi_{\text{CF}}(r) = \left( \frac{r}{r_0} \right)^{-\gamma},$$

where \(r_0\) and \(\gamma\) are the clustering length and slope. We apply the power-law fitting to the CCFs in Figure 1 (bottom) over 0.1–1 \(h^{-1}\) Mpc and 1–10 \(h^{-1}\) Mpc separately, corresponding to the CGM and IGM regimes. Note that \(r_0\) and \(\gamma\) reflect the amplitude and slope of the CCF, respectively. The best-fit parameters of all the CCFs are presented in Figure 2. Filled and open circles represent the best-fit parameters of the CGM and IGM regimes, respectively.

For all galaxies, we obtain the best-fit parameters of \((r_0, \gamma) = (0.07 \pm 0.004, 0.50 \pm 0.01)\) and \((0.34 \pm 0.03, 1.07 \pm 0.04)\) for the CGM and IGM regimes (see also Figure 2(e)). Several observational studies have performed a power-law fitting to their CCFs between Ly\(\alpha\) absorption and galaxies. Tummuangpak et al. (2014) have calculated the CCF between Ly\(\alpha\) absorption and LBGs at \(z \sim 3\) and fit it by a double power law, showing \((r_0, \gamma) = (0.08 \pm 0.04, 0.47 \pm 0.10)\) and \((0.49 \pm 0.32, 1.47 \pm 0.91)\) for the CGM and IGM regimes used at \(r = 1.6 h^{-1}\) Mpc as a border. A subsequent study of IGM–LBG clustering by Bielby et al. (2017) described their CCF by a single power law with \((r_0, \gamma) = (0.27 \pm 0.14, 1.1 \pm 0.2)\). A power-law fitting to the CCF of a weak H\(\text{i}\) (\(N_{\text{HI}} < 10^{14} \text{ cm}^{-2}\)) IGM and galaxies at \(z = 2\) has also been attempted and resulted in \((r_0, \gamma) = (0.2 \pm 0.4, 1.1 \pm 0.3)\) (Tejos et al. 2014). Although the fitting range for power-law fitting is different among studies, our best-fit parameters for both the CGM and IGM regimes are comparable to those of previous observations within the error.

We next show the best-fit parameters obtained from all categories. For \(M_*\) and \(M_{\text{DM}}\), they show similar trends in both \(r_0\) and \(\gamma\) of both the CGM and IGM regimes. The clustering length \(r_0\) becomes longer with increasing mass independently of the regime. We remind the reader that \(r_0\) reflects the amplitude of the CCF. Thus, we can confirm particularly in Figure 1 (bottom) that the higher the mass becomes, the stronger the CCF amplitude is. Meanwhile, the slope \(\gamma\) becomes smaller with increasing mass in the CGM regime but is consistent with being constant within the error in the IGM regime. In the literature, Tummuangpak et al. (2014) have calculated the mass-dependent CCFs by dividing their simulated galaxies into two categories of \(M_* > 10^8\) and \(10^8 h^{-1} M_*\). Their double power-law fit for these \(M_* > 10^8\) and \(10^8 h^{-1} M_*\) samples have given \((r_0, \gamma) = (0.10 \pm 0.07, 0.46 \pm 0.22)\) and \((0.16 \pm 0.09, 0.46 \pm 0.19)\) in the CGM regime and \((0.51 \pm 0.39, 1.25 \pm 0.61)\) and \((0.61 \pm 0.34, 1.18 \pm 0.43)\) in the IGM regime. Although the differences of \(r_0\) and \(\gamma\) between the \(M_* > 10^8\) and \(10^8 h^{-1} M_*\) samples are within the error, a similar trend is confirmed in \(r_0\). For the SFR sample except SFR–(v), we confirm that \(r_0\) becomes greater with an increasing SFR. This is naturally explained by the fact that our galaxy sample are mostly on the star formation main sequence between SFR and stellar mass. On the other hand, \(\gamma\) has no clear trend in either the CGM or the IGM regime. Although the CCFs of all three sSFR categories are comparable as presented in Figure 1(d1), the best-fit parameters of both \(r_0\) and \(\gamma\) are different from each other. Interestingly, both \(r_0\) and \(\gamma\) show similar trends in both the CGM and IGM regimes, where sSFR–(i) shows the smallest value. We briefly discuss the reason in Sections 5.1 and 5.2.

4.1.2. Gas Density Fluctuation

The \(\delta_p\) value of the LoS data should correlate with the H\(\text{I}\) gas density at the position. Thus, the variation of CCF amplitudes can be attributed to the variation of the local gas density around galaxies. To verify this hypothesis, we evaluate the CCFs of gas density fluctuations around galaxies defined by

$$\delta_p \equiv \frac{\rho}{\langle \rho \rangle},$$

$$\xi_{\delta \rho} = \frac{1}{N(r)} \sum_{i=1}^{N(r)} \frac{\delta_{\rho_{ij}}}{M(r)} - \frac{1}{M(r)} \sum_{j=1}^{M(r)} \delta_{\rho_{ij}},$$

where \(\delta_p\) is the gas density fluctuation defined by the ratio of the gas density at one LoS pixel \(\rho\) to the mean gas density at each redshift of \(\Delta z = 0.01\). \(\langle \rho \rangle\) is the CCF; and \(\delta_{\rho_{ij}}\) and \(\delta_{\rho_{ij}}\) are the gas density fluctuations for pixel \(i\) and \(j\) at distance \(r\) from galaxies and random points. As in Equation (5), \(N(r)\) and \(M(r)\) are the numbers of pixel–galaxy and pixel–random pairs in the bin with distance \(r\). The CCFs of gas density fluctuations are measured for both total gas and H\(\text{I}\) (\(\delta_{\rho_{\text{gas}}}\) and \(\delta_{\rho_{\text{H}}}\)) and are shown in Figures 3 (top) and 4 (bottom). For comparison to the CCFs of \(\delta_p\), the log-scale inversion CCFs of \(\delta_p\) (i.e., \(-\log \delta_p\)) are also shown in Figure 1 (bottom). Due to several negative values in \(\delta_{\rho_{\text{gas}}}\), we also present the mean gas density fluctuations around galaxies in Figures 3 (bottom) and 4 (bottom).

First, we start from the CCFs of the total gas density fluctuation \(\delta_{\rho_{\text{gas}}}\) in Figure 3. The overall trends for each category (i.e., \(M_*\), \(M_{\text{DM}}\), SFR, and sSFR) are almost the same as those of \(\delta_{\rho_{\text{gas}}}^p\)’s CCFs. That is, the overall CCFs monotonically decrease with increasing galaxy mass (either \(M_*\) or \(M_{\text{DM}}\)) or SFR. However, several samples show notable behavior in their CCFs’ strength.
or shape at a certain radius. For $M_{\star}–11$, $M_{\text{BH}}–13$, and SFR–(i), we find a slight decline of $\xi_{\delta_{\gamma_i},1}$ at the center. This indicates the decline of the relative total gas density in the vicinity of those galaxies. For SFR–(v) and sSFR–(iii), their CCFs show a convex feature at $r=0.3–2\, h^{-1}$ Mpc and even have the strongest signal among each category over that radius range. It is noted for $M_{\text{BH}}–10$ that the swallowing at $r=0.5–1\, h^{-1}$ Mpc may be due to its small sample size.

The CCFs of H I gas density fluctuation $\delta_{\gamma,i}$, are shown in Figure 4. Compared to the H I optical depth distribution on the LoS (Figure 1, top, $\delta_{\gamma}$), the raw H I gas particles have a slightly discrete distribution (Figure 4, bottom, $\delta_{\gamma,i}$). This is because we consider line broadening based on the Voigt profile (see also Equation (2)). As a result, the CCFs in $\delta_{\gamma}$ have a smoother shape than those in $\delta_{\gamma,i}$. Similar to the overall trends seen in the CCFs of $\delta_{\gamma,i}$, we find that the CCF signal becomes stronger with increasing $M_{\star}$, $M_{\text{BH}}$, or SFR in general. The trend is also about the same as that found in the CCFs of $\delta_{\gamma}$. The consistency of the CCFs' trends is naturally explained by considering Equations (2) and (3), where $\delta_{\gamma}$ is proportional to $H$ number density. We also find that the irregularity identified in the CCFs of $\delta_{\gamma,i}$ is seen in several samples: a decline of $\xi_{\delta_{\gamma,i}}$, at the center in $M_{\star}–11$, $M_{\text{BH}}–13$, and SFR–(i) and a convex profile and the strongest signal at $r=0.3–2\, h^{-1}$ Mpc in SFR–(v) and sSFR–(iii). In addition to the above irregular CCF shapes, SFR–(v) and sSFR–(iii) show a significant decline of $\xi_{\delta_{\gamma,i}}$ values at the center. This suggests that H I densities around galaxies in $M_{\star}–11$, $M_{\text{BH}}–13$, SFR–(i), SFR–(v), and sSFR–(iii) are also relatively low in general. We discuss this in Section 5.2.

### 4.2. Overdensity Analysis

We present the results of the overdensity analysis in Figures 5(a) and (b), which are derived from local minima/maxima and random positions in the gas density field. The analysis is performed for all galaxies and $M_{\star}$-dependent subsamples of $M_{\star}–11$, $M_{\star}–10$, and $M_{\star}–9$. We evaluate the $\Sigma_{\text{gal}}$ and $\langle \delta_{\gamma} \rangle$ values in each of the four redshift slices indicated by open or filled circles colored red or blue. The exact galaxy count used in each redshift slice is shown in Table 2.

First, we find possible anticorrelations between $\Sigma_{\text{gal}}$ and $\langle \delta_{\gamma} \rangle$ in Figure 5(a). To statistically assess those correlations, we perform Spearman’s rank correlation tests and obtain correlation coefficients ranging from $R_s = -0.33$ to $R_s = -0.42$, indicating a mild anticorrelation. Similarly, Mukae et al. (2017) have identified a mild anticorrelation in their $\langle \delta_{\gamma} \rangle–\Sigma_{\text{gal}}$ distribution with $R_s = -0.39$.

We should comment on the effect of outlier data points in $M_{\star}–10$, $M_{\star}–9$, and ALL of Figure 5(a). We repeat Spearman’s rank correlation test for all data points but without the outliers and obtain $R_s = (-0.37, -0.28, -0.32)$ for the ($M_{\star}–10$, $M_{\star}–9$, ALL) samples. Therefore, weak anticorrelations are still confirmed even without the outliers.

To characterize the $\langle \delta_{\gamma} \rangle–\Sigma_{\text{gal}}$ distribution, we follow Mukae et al. (2017) and apply chi-square fitting in Figure 5 with the linear model

$$\langle \delta_{\gamma} \rangle = \alpha + \beta \Sigma_{\text{gal}}. \quad (10)$$

The best-fit parameters of $\alpha$ and $\beta$ are summarized in Table 2. We find that $\alpha \sim -0.13$, which is about the same for all of the four samples within the error, while the $\beta$ become slightly larger with increasing $M_{\star}$, although they are still similar within the error: $\beta = -0.007 \pm 0.003, -0.014 \pm 0.005$, and $-0.020 \pm 0.007$ for $M_{\star}–11$, $M_{\star}–10$, and $M_{\star}–9$, respectively.

Note that the best-fit parameters of anticorrelations for $M_{\star}–10$, $M_{\star}–9$, and ALL without outliers appear to be comparable within the error. We compare the best-fit parameters of the ALL sample to those in Mukae et al. (2017), which are $(\alpha, \beta) = (-0.17 \pm 0.06, -0.14 \pm 0.06)$. We find a similarity in $\alpha$ but a large difference in $\beta$, showing a much shallower slope for our sample. A shallower slope of the simulated galaxy sample has also been found in Nagamine et al. (2020). It may be attributed to photo-z errors in the observational data. If the photo-z errors are large, then some galaxies would contaminate the sample and the value of $\Sigma_{\text{gal}}$ would be smeread out. As a result, the observed $\Sigma_{\text{gal}}$ only has a narrow dynamic range, which could make the apparent correlation steeper than the real one.

We also examine the result of overdensity analysis based on randomly selected points in order to verify the effect of position bias (Figure 5(b)). Spearman’s rank correlation tests for randomly selected positions yield mild anticorrelations with $R_s = -0.51$ to $R_s = -0.57$. The best-fit parameters of the linear model $(\alpha, \beta)$ are also comparable to those from Figure 5(a) within the error. Moreover, the trend found in the best-fit $\alpha$ and $\beta$ as a function of stellar mass is confirmed. Therefore, we

### Table 2

| Sample Name | $N^{(a)}$ | $N^{(b)}$ | $N^{(c)}$ | $N^{(d)}$ | $R_s^{(2)}$ | $p^{(2)}$ | $\alpha^{(5)}$ | $\beta^{(5)}$ |
|-------------|-----------|-----------|-----------|-----------|------------|----------|-------------|-------------|
| $M_{\star}–11$ | 134       | 139       | 100       | 85        | -0.41      | 0.12     | -0.129 ± 0.011 | -0.007 ± 0.003 |
| $M_{\star}–10$ | 1709      | 1555      | 1226      | 1182      | -0.42      | 6.27e–3 | -0.126 ± 0.006 | -0.014 ± 0.005 |
| $M_{\star}–9$   | 4658      | 4356      | 3984      | 3921      | -0.33      | 0.03     | -0.126 ± 0.006 | -0.020 ± 0.007 |
| ALL            | 6501      | 6050      | 5310      | 5188      | -0.37      | 0.02     | -0.126 ± 0.006 | -0.018 ± 0.006 |

Notes. (1) Number of galaxies in (1a) $2.07 < z < 2.102$, (1b) $2.215 < z < 2.247$, (1c) $2.3 < z < 2.332$, and (1d) $2.45 < z < 2.482$. (2) Spearman’s coefficient and $p$-value. The $\langle \delta_{\gamma} \rangle–\Sigma_{\text{gal}}$ relation is examined around the local minima and maxima. (3) The best-fit parameters of the chi-square fitting of the $\langle \delta_{\gamma} \rangle–\Sigma_{\text{gal}}$ relation examined around the local minima and maxima. (4) Spearman’s coefficient and $p$-value. The $\langle \delta_{\gamma} \rangle–\Sigma_{\text{gal}}$ Relation is examined around random points. (5) The best-fit parameters of the chi-square fitting of the $\langle \delta_{\gamma} \rangle–\Sigma_{\text{gal}}$ relation examined around random points.
conclude that the position bias of overdensity analysis does not affect the $\langle \delta V \rangle - \Sigma_{\text{gal}}$ correlation seriously.

5. Discussion

5.1. Origin of CCF Variations

We present in Section 4.1.1 that the CCF of $\delta V$ varies depending on the galaxy mass and SFR. To find the origin of its variation, we also calculate the CCFs of $\delta_{\text{HI}}$ and $\delta_{\text{H}1}$ in Section 4.1.2 and find that their signal strengths also depend on $M_*$, $M_{\text{DM}}$, SFR, and sSFR. This suggests that the different relative gas densities around galaxies cause the variation of the CCFs of $\delta V$. Considering the relation between $\delta V$ and HI number density in Equations (2)–(4) a similar trend of CCFs in $\delta V$ and $\delta_{\text{HI}}$ is reasonable. The same trend even for $\delta_{\text{H}1}$ probably means that the total gas density correlates with H I gas density in general. Therefore, we argue that the variation of $\delta V$ CCFs is caused by different gas distributions around galaxies.

We find that not only the $\delta V$ CCFs but also the best-fit parameters of their power-law fitting vary depending on the galaxies’ mass and SFR (see also Figure 2). In particular, our best-fit clustering lengths $r_0$ for the IGM regime show an increase with an increasing mass or SFR of galaxies. A dependence of CCF signal strength and its best-fit parameters on H I column density ($N_{\text{HI}}$) of the CGM and IGM has also been reported in observational studies. For example, Ryan-Weber (2006) has calculated CCFs by splitting their absorber sample into two based on the absorber’s $N_{\text{HI}}$ and found that the high-$N_{\text{HI}}$ subsample shows stronger correlation than the low-$N_{\text{HI}}$ subsample. Tejos et al. (2014) have calculated CCFs depending on $N_{\text{HI}}$ for galaxies at $z < 1$ and found a positive correlation between best-fit parameters (both $r_0$ and $\gamma$) and $N_{\text{HI}}$. They have also demonstrated a dramatic change of CCF signals depending on $N_{\text{HI}}$, showing more than a factor of 10 higher CCF signals in the $N_{\text{HI}} > 10^{14} \text{ cm}^{-2}$ sample compared to those of the $N_{\text{HI}} < 10^{14} \text{ cm}^{-2}$ sample. Similarly, Bielby et al. (2017) have analyzed the cross-correlation between Ly$\alpha$ absorption with different $N_{\text{HI}}$ measurements and LBGs at $z = 3$ and presented a positive correlation between the best-fit parameters of their CCFs and $N_{\text{HI}}$. These studies have also argued for a relation between the best-fit parameters ($r_0$ and $\gamma$) and gas, particularly H I. Larger $r_0$ implies a stronger clustering of H I systems around galaxies. On the other hand, for $\gamma$, previous studies have suggested the necessity for additional baryonic physics to explain its changes with $N_{\text{HI}}$.

Considering the above discussion, the signal strength and best-fit parameters ($r_0$ and $\gamma$) of the CCF depend on relative gas densities on Mpc scales near the galaxy. If the gas has a high density and clusters around a galaxy, the resultant CCF between Ly$\alpha$ absorbers and the galaxy must have a stronger signal and give larger best-fit parameters for the IGM regime. The variation of CCFs is hence attributed to the different gas density around each galaxy and the strength of the galaxy–IGM connection.

5.2. What Types of Galaxies Strongly Correlate with the IGM?

Under the $\Lambda$CDM paradigm, massive galaxies are expected to strongly correlate with the underlying dark matter (e.g.,
Mo & White 2002; Zehavi et al. 2005) and hence with IGM as well. In that sense, more massive galaxies should strongly cluster in higher-density regions compared to less massive galaxies. In Section 4.1.1, we confirm that the CCF signal becomes stronger with the increasing mass (both $M_*$ and $M_{DH}$) of a galaxy. In addition, we find that the turnover radius of

---

Figure 3. (Top) CCFs of total gas density fluctuation ($\delta_{\rho_{gas}} = \rho_{gas}/\langle \rho_z \rangle$) as a function of comoving distance from galaxies. (Bottom) Mean total gas density fluctuation around galaxies.
CCFs becomes smaller with mass in $M_*-11$ and $M_{\text{DM}}-13$, which is likely to be the result of a stronger connection between massive galaxies and higher-density regions.

From the overdensity analysis in Section 4.2, we find that the slope of the anticorrelation between $\langle \delta F \rangle$ and $\Sigma_{\text{gal}}$ becomes shallower with increasing $M_*$, although its difference is still

Figure 4. (Top) CCFs of H I density fluctuation ($\delta_{\text{HI}} = \rho_{\text{HI}}/\langle \rho_{\text{HI}} \rangle$) as a function of comoving distance from galaxies. (Bottom) Mean H I density fluctuation around galaxies.
within the error. A shallower slope in the $M_{\ast} - 11$ sample indicates stronger clustering of massive galaxies around a dense HI IGM.

The above results from both methods (CCF and overdensity analysis) imply that massive galaxies are strongly clustered in high-density regions in the cosmic web, while less massive galaxies have the opposite trend.

The same trend should be true for SFR subsamples by considering the star formation main sequence (e.g., Brinchmann et al. 2004; Daddi et al. 2007; Elbaz et al. 2007; Noeske et al. 2007; Speagle et al. 2014; Schreiber et al. 2015; Tomczak et al. 2016). The overall trend for the SFR samples is that galaxies with higher (lower) SFRs correlate with a higher (lower) gas density of the IGM. However, the SFR–(v) subsample does not follow the trend and even seems to reside in the highest density among all SFR samples at $r = 0.3 – 2 \, h^{-1} \, \text{Mpc}$ (see also Section 4.1.2). Such CCF behavior can be attributed to the halo mass distribution of galaxies in SFR–(v). Figure 6 represents the halo mass distribution of galaxies in all SFR samples. They generally have a single peak, but a mild bimodal distribution is seen in SFR–(v), which has two peaks at $M_{\Delta M} \sim 10^{11}$ and $10^{12.5} - 10^{13} \, M_{\odot}$. This implies that the host dark matter halos of the SFR–(v) subsample can be roughly divided into two: one is less massive and the other is massive. Given the mass dependence of the IGM–galaxy connection, the strong CCF (i.e., high-density gas) seen at $r = 0.3 – 2 \, h^{-1} \, \text{Mpc}$ of the SFR–(v) subsample reflects high gas density around massive halos.

Our sSFR samples do not show any obvious trends in CCF strength. We find that all three samples have comparable CCFs, although sSFR–(i) gives smaller $r_0$ and $\gamma$ in both the CGM and IGM regimes. This is due to the same reason as what we see in the SFR samples. Similar to the SFR–(v) sample, the sSFR–(iii) sample has a mild bimodal halo mass distribution (see Figure 7). The halo mass distribution of sSFR–(ii) may have a tail extending to higher mass. As a result, these two samples possibly reside in higher-density regions in the IGM compared to sSFR–(i).

Summarizing the above discussion, we conclude that the dark matter halo mass is the most sensitive parameter to determining the baryonic environment around galaxies. Galaxies that are hosted by massive halos are generally located in high-density gaseous environments, resulting in a stronger signal of the CCF.

Finally, we should briefly discuss the declining CCF signal at $r < 0.3 – 0.4 \, h^{-1} \, \text{Mpc}$ around galaxies in the $M_{\ast} - 11, M_{\Delta M} - 13, \text{SFR–(i), SFR–(v), and sSFR–(iii)}$ subsamples (see Figures 3 and 4), which are hosted by the most massive halos among each category. There are several possible reasons for the lack of H I gas in the central region of massive galaxies. The first possibility is that the gas particles in the central region are blown out to the CGM/IGM by SN feedback. The second possibility is that our feedback prescription without AGN contribution is still inadequate in pushing the gas away into the CGM/IGM for massive galaxies due to their deep gravitational potential. As a result, most gas particles in the central region are consumed by star formation. We need more detailed analysis of our simulation to confirm the reason, but this is beyond the scope of this paper. We will try to address this issue in future work.
5.3. Photo-z versus Spec-z Data and the IGM–Galaxy Connection

In this subsection, we briefly discuss the reliability of using photo-z and spec-z data to investigate the IGM–galaxy connection. We also argue that some precautions must be taken when one applies the two analyses to observational data. We have demonstrated that the cross-correlation method succeeds in identifying variations of the CCF according to galaxy properties. The difficulty for the CCF method is the necessity of relatively accurate redshift measurements for galaxies. As we demonstrate in Appendix B, the CCF signal will be attenuated if the redshift uncertainty is large. According to the general photo-z uncertainties at $z \sim 2$ in the literature, $\sigma_z = 0.05$–0.1 (e.g., Muzzin et al. 2013; Laigle et al. 2016; Straatman et al. 2016), galaxies with only photometric redshift cannot be used for cross-correlation analysis. In general, however, the majority in a given galaxy sample do not have spectroscopic redshift. In that sense, the results of our CCF analysis applied to observational data are biased toward the CGM and IGM around galaxies with spec-z measurements.

On the other hand, galaxies with only photo-z can be used for overdensity analysis that examines the IGM–galaxy connection on large scales (beyond tens of $h^{-1}$ Mpc). Indeed, photo-z galaxies are used as tracers of large-scale structures (e.g., Nakata et al. 2005; Blake et al. 2007; Trevese et al. 2007; Tanaka et al. 2008; Hayashi et al. 2019). However, there are still the following potential problems in using photo-z galaxies for analysis. First is the contamination of galaxies whose real redshifts are out of the redshift range for 2D LoS data. It must always happen, even if the redshift range is wider than the mean photo-z error. Second is a difficulty in statistically confirming a correlation when the cylinder volume is large (see also Appendix A). Both this study and those in the literature show the presence of an IGM–galaxy connection up to $\sim 10 \, h^{-1}$ Mpc (e.g., Adelberger et al. 2003, 2005; Chen et al. 2005; Ryan-Weber 2006; Faucher-Giguère et al. 2008; Rakic et al. 2011, 2012; Rudie et al. 2012; Font-Ribera et al. 2013; Prochaska et al. 2013; Tejos et al. 2014; Bielby et al. 2017). In addition, according to our tests in Appendix A, we suggest that a cylinder length less than $\Delta z = 0.01$ (corresponding to $9 \, h^{-1}$ Mpc) might be able to capture large-scale structures in $\delta_F$. If either the cylinder length or radius is larger than the above scale, the large-scale structures traced by the cosmic web and galaxies will be attenuated, and thus both $\langle \delta_F \rangle$ and $\Sigma_{\text{gal}}$ will become close to zero. In that case, the slope of the $\langle \delta_F \rangle$–$\Sigma_{\text{gal}}$ relation will become indistinguishable.

Figure 6. Normalized number histogram of galaxies as a function of $M_{\text{DH}}$ for each SFR subsample. The solid black line indicates the histogram of each subsample, and the SFR–(i) histogram is overlaid in the bottom four panels as a gray-shaded histogram for comparison.

Figure 7. Same as Figure 6 but for the sSFR subsamples. The solid black line indicates the histogram of each subsample, and the sSFR–(i) histogram is overlaid in the bottom two panels as a gray-shaded histogram for comparison.
Based on all these arguments, the cross-correlation method is a promising method to study the variation of the IGM–galaxy connection over a $1\, h^{-1}\text{Mpc}$ scale using actual observational data. On the other hand, overdensity analysis applied to photo-$z$ samples can probe the IGM–galaxy connection on scales beyond several tens of $h^{-1}\text{Mpc}$.

6. Summary

We systematically investigate the connection between galaxies and the CGM/IGM, particularly as traced by Ly$\alpha$ forest absorption. In this study, we use cosmological hydrodynamic simulation (Shimizu et al. 2019; Nagamine et al. 2020) and demonstrate the CGM/IGM–galaxy connection using two methods: one is cross-correlation analysis, and the other is the overdensity analysis proposed by Mukae et al. (2017). Using our simulation, we also calculate the CCFs of relative gas density (both total and H I) around galaxies. All parameters for our analyses are chosen to match the observations presented in M21. The main results of this paper are summarized below.

1. We calculate the CCFs between Ly$\alpha$ forest transmission fluctuation ($\delta_F$) and galaxies as shown in Figure 1. The CCF obtained from all galaxies reproduces the one from LBGs in the literature (Adelberger et al. 2005; Bielby et al. 2017). Further investigations based on subsamples divided by the $M_s$, $M_{\text{DM}}$, SFR, and sSFR of simulated galaxies show the following trends and variations in the CCF. For the $M_s$ and $M_{\text{DM}}$ subsamples, we find a clear trend that the CCF signal becomes stronger with increasing galaxy mass. We also confirm that the turnover radius of CCFs becomes smaller with increasing mass (see Figure 1, bottom), indicating stronger clustering of massive galaxies around density peaks. Additionally, from the best-fit parameters of power-law fitting for the CCFs, the clustering length $r_0$ and slope $\gamma$ are found to become longer and steeper with increasing galaxy mass in the IGM regime ($r \gtrsim 1\, h^{-1}\text{Mpc}$; see also Figure 2). For the SFR samples, we find that they tend to have stronger signals with an increasing SFR. Such a trend for SFR samples is also linked to the mass dependence of CCFs, because $M_s$ and SFR are almost linearly related to each other through the star formation main sequence of galaxies. However, we do not identify clear trends in the CCFs of sSFR samples.

2. We measure CCFs between gas density fluctuation ($\delta_{\text{HI}}$) and galaxies as shown in Figures 3 and 4. The overall trends of the CCFs are similar to those of CCFs in $\delta_F$ except for SFR–(v) and sSFR–(iii). This indicates that the variation in $\delta_F$ CCFs reflects different relative gas densities around galaxies; i.e., galaxies with higher mass and SFRs generally reside in higher-density gas and vice versa. For the SFR–(v) and sSFR–(iii) subsamples, we find the strongest CCF signal at $r = 0.3–2\, h^{-1}\text{Mpc}$ among all the SFR and sSFR samples. Because the two subsamples have a mild bimodal halo mass distribution with two peaks at $M_{\text{DM}} \sim 10^{11.3}$ and $M_{\text{DM}} \sim 10^{12.5}$, their strongest CCF signals are probably due to high-density regions where massive host halos reside. We suggest that the observed variation in the CCF is caused by the dependence of gas density (both total and H I) around galaxies.

3. Our overdensity analysis between galaxy overdensity $\Sigma_{\text{gal}}$ and mean IGM fluctuation $\langle \delta_F \rangle$ is presented in Figure 5. We statistically identify anticorrelations from all subsamples of $M_s$, $M_{SFR}$, $M_{\text{HI}}$, and ALL. In addition, we find that their slopes decrease with increasing $M_s$, albeit still within the error. This suggests that galaxies in the $M_s$ subsample are more strongly correlated with higher-density gas than those in $M_{SFR}$ in terms of their spatial distribution.

4. Considering all of our results together, we conclude that the mass, particularly the dark matter halo mass, is the most sensitive parameter to determining the Mpc-scale gas density environment around galaxies. Galaxies in massive halos tend to be clustered in higher-density regions of the cosmic web, resulting in a CCF with a higher amplitude, greater $r_0$, steeper $\gamma$, and shallower anticorrelation between $\langle \delta_F \rangle$ and $\Sigma_{\text{gal}}$ at $r \gtrsim 1\, h^{-1}\text{Mpc}$.

Overall, our analyses confirm the strong connection between galaxies, dark matter halos, and the IGM, providing further support for the gravitational instability paradigm of galaxy formation within the concordance $\Lambda$CDM model. Future observations of CCF studies between galaxies, HI, and metals will provide useful information on the interaction between them and the details of feedback mechanisms that are important for the theory of galaxy formation and evolution such as galactic wind and the associated ejection of metals into the IGM.

By comparing the results of Figures 1 and 2 against predictions of linear perturbation theory, we can infer the mean bias parameters of the Ly$\alpha$ forest and galaxies relative to the underlying dark matter density field (e.g., Croft et al. 2016; Bielby et al. 2017; Kakihara et al. 2018; Meyer et al. 2019). In addition, we can perform cross-checks by computing such bias parameters directly from our simulation output and compare them with those inferred from the linear theory framework. Such bias parameters will further provide additional checks against the gravitational instability paradigm, and we plan to carry out such analyses in further work.

We are grateful to Drs. M. Rauch, A. Meiksin, H. Yajima, D. Sorini, T. Suarez Noguez, K. Kakiichi, and R.A. Meyer for helpful discussions. We also thank the referee and the editor for providing constructive suggestions and comments to improve our manuscript. R.M. acknowledges a Japan Society for the Promotion of Science (JSPS) fellowship in Japan. K.N. is grateful to Volker Springel for providing the original version of GADGET3, on which the GADGET3-Osaka code is based. Our numerical simulations and analyses were carried out on the XC50 systems at the Center for Computational Astrophysics of the National Astronomical Observatory of Japan, on the XC40 system at the Yukawa Institute for Theoretical Physics in Kyoto University, and on the OCTOPUS at the Cybermedia Center, Osaka University, as part of an HPCI System research project (hp180063, hp190050). This work is supported by JSPS KAKENHI grant Nos. JP18J40088 (R.M.), JP17H01111, 19H05810 (K.N.), and JP19K03924 (K.S.). K.N. acknowledges travel support from Kavli IPMU, World Premier Research Center Initiative, where part of this work was conducted. We acknowledge the Python programming language and its packages of numpy, matplotlib, scpy, and astropy (Astropy Collaboration et al. 2013).
addition, the dynamic range of increasing redshift width for the binning of the LoS data. In the intensity of IGM overdensity analysis in Section 4.2. Figure 8 clearly shows that each column indicates a different redshift used for the Mpc, which correspond to each column. A black circle indicates the cylinder size to estimate Figure 8.

The cross-correlation analysis used in this study requires a sample of galaxies (see Section 5.3). However, spectroscopic redshift measurements are not always available for galaxies in photometric images. Hence, here we test the usability of photo-z galaxies in photometric images. Therefore, here we test the CCF of each routine and the mean of 10 tests are shown by thin and thick black lines in Figure 9. We also carry out the same tests with $\sigma_z \leq 0.05$, $\sigma_z \leq 0.02$, and $\sigma_z \leq 0.01$. The actual distances corresponding to the $\sigma_z$ values are $(\sigma_z=0.1, 0.05, 0.02, 0.01) = (90, 45, 18, 9) \text{ h}^{-1}\text{ Mpc at } \langle z \rangle = 2.3$. In the bottom panel of Figure 9, the ratio of the CCF from all galaxies in red ($\xi_{org})$ to the mean of the CCFs from galaxies in consideration of redshift uncertainties ($\xi_e$) is also presented ($\Xi = \xi_{org}/\xi_e$).

We find that all CCF signals become weaker with respect to the original CCF in red, though the scatter of CCF signals among 10 tests is quite small. In particular, the CCF signal becomes insignificant for the data with $\sigma_z \leq 0.1$ or 0.05. This suggests that galaxy data sets with such large photo-z errors are useless for cross-correlation analysis. However, the other two cases with $\sigma_z \leq 0.02$ or 0.01 still show some signals at the center, although they are weak. Due to the signal detection, galaxies with $\sigma_z \leq 0.02$ may be usable for calculating CCFs.

Another interesting result from this test is the radius where $\Xi$ becomes approximately unity. Within $r < 40 \text{ h}^{-1}\text{ Mpc}$, each sample reaches $\Xi = 0$ at a radius that is equivalent to the actual distance of $\sigma_z$ (see also vertical lines in the bottom panel of Figure 9). These results indicate that a data set with redshift uncertainties less than $1 \text{ h}^{-1}\text{ Mpc}$ (or $\sigma_z \sim 0.001$) would be necessary to obtain a true CCF.

Many high-$z$ galaxies often have photo-z estimates. In the literature, such photo-z uncertainties have been evaluated as $\sigma_z = (0.007-0.021) \times (1+z)$, which corresponds to $\sigma_z = 0.023-0.070$ at $z = 2.35$ (e.g., Laigle et al. 2016; Straatman et al. 2016). Considering our tests shown in Figure 9, galaxy data sets with current photo-z measurements only are not useful for cross-correlation analysis. In order to derive an accurate CCF, galaxy samples with good spectroscopic redshifts are needed.

Figure 8. Two-dimensional LoS maps of four arbitrary redshift slices are shown in four rows. Different binning widths to generate the 2D LoS maps are arranged in each column. A black circle indicates the cylinder size to estimate $\langle \delta_F \rangle$ and $\Sigma_{gal}$ by overdensity analysis ($r = 4.74 \text{ h}^{-1}\text{Mpc}$).

Appendix A
2D LoS Maps

To understand the impact on actual observational data, we conduct several tests for generating 2D LoS maps and cross-correlation analysis with the actual observational data in M21 in mind. In these appendices we briefly show our results. Note that we only calculate the CCFs at $r \geq 1 \text{ h}^{-1}\text{ Mpc}$ for those tests in order to directly compare them with the observational results presented in M21.

We show 2D LoS $\delta_F$ maps binned by 9, 18, 45, and 90 $\text{ h}^{-1}\text{ Mpc}$, which correspond to $\Delta z = 0.01, 0.02, 0.05$, and 0.1, in Figure 8. Each column indicates a different redshift used for the overdensity analysis in Section 4.2. Figure 8 clearly shows that the intensity of IGM fluctuation becomes attenuated with increasing redshift width for the binning of the LoS data. In addition, the dynamic range of $\langle \delta_F \rangle$ of the 2D LoS maps binned by more than 45 $\text{ h}^{-1}\text{ Mpc}$ seems to be too narrow to differentiate the environment based on galaxy properties.

Appendix B
The Impact of Redshift Measurement Uncertainties on the CCF

The cross-correlation analysis used in this study requires a spec-z sample of galaxies (see Section 5.3). However, spectroscopic redshift measurements are not always available for galaxies in photometric images. Hence, here we test the usability of photo-z galaxies by adding photo-z errors.

We randomly add redshift uncertainties with $\sigma_z \leq 0.1$ to all galaxies. Then, we recalculate the CCF using the reassigned galaxy redshift $z_{reassign} = z_{real} \pm \sigma_z$, where $z_{reassign}$ and $z_{real}$ are the redshift used to calculate the CCF and the original one in our simulation, respectively. This process is carried out 10 times. The CCF of each routine and the mean of 10 tests are shown by
therefore the derived CCF from spec-z data may suffer from a small sample size and differ from the true signal. Thus, we carry out the following two tests in order to verify the effect of sample size on the resultant CCF.

The first test is to change the completeness of the galaxy sample. We calculate CCFs by randomly selecting 0.1%, 1%, and 10% of galaxies from the entire sample and repeat this procedure 10 times. We find that the resultant CCFs scatter around the original one. Additionally, this scatter becomes smaller with an increasing fraction and negligible in the 10% samples. Therefore, we argue that at least 1% of the total sample must have spec-z measurements to reproduce the true CCF.

Unfortunately, we do not always know the true total number of galaxies in real observations. Therefore as a second test, we examine the effect of using an extremely small sample of randomly selected 5 and 10 galaxies and repeat this step 100 times. We conduct this test twice by changing the galaxy selection method: one is completely random, while the other is to select only galaxies whose \( \langle \delta v \rangle \) within 1.7 \( h^{-1} \) Mpc of the radius are less than −0.2. The resultant CCFs of both methods show a large dispersion around the original CCF, and the dispersion becomes smaller as the sample size increases from 5 to 10. However, intriguingly, the scatter among 100 CCFs in the latter method becomes smaller than that in the former method. This suggests that the true CCF cannot be obtained from a randomly selected, extremely small sample. On the other hand, a CCF derived from a few galaxies that are located in similar IGM densities (i.e., in a limiting \( \langle \delta v \rangle \)) could still reflect their IGM environments.

**Appendix C**

**Cosmic Variance**

If a survey volume is not large enough, the cosmic variance must affect the CCFs (in terms of both amplitude and shape). We evaluate the effect on the CCF by limiting the volume to \( \Delta z = 0.1 \) and 0.05, corresponding to 91 \( h^{-1} \) Mpc and 45 \( h^{-1} \) Mpc in redshift direction. Note that because large-scale fluctuations are missed due to a limited simulation volume, we inevitably underestimate the effect of cosmic variance on large scales. We find that the resultant CCFs scatter around the original one obtained from all galaxies in the entire volume with a wide variation of the power-law \( \gamma \) and \( r_0 \) parameters of Equation (9). This suggests that cosmic variance also influences both the slope and the clustering length of the CCF. Therefore, when we compare the CCFs of two different galaxy populations, it would be desirable to match their redshift coverage.

**Appendix D**

**CCFs Obtained from Small Sample Size**

In Appendix B, we discuss the uncertainty introduced by using photo-z data and the need for more accurate spec-z measurements for cross-correlation analysis. However, the number of galaxies with spec-z measurements is limited, and therefore the derived CCF from spec-z data may suffer from a small sample size and differ from the true signal. Thus, we carry out the following two tests in order to verify the effect of sample size on the resultant CCF.

The first test is to change the completeness of the galaxy sample. We calculate CCFs by randomly selecting 0.1%, 1%, and 10% of galaxies from the entire sample and repeat this procedure 10 times. We find that the resultant CCFs scatter around the original one. Additionally, this scatter becomes smaller with an increasing fraction and negligible in the 10% samples. Therefore, we argue that at least 1% of the total sample must have spec-z measurements to reproduce the true CCF.

Unfortunately, we do not always know the true total number of galaxies in real observations. Therefore as a second test, we examine the effect of using an extremely small sample of randomly selected 5 and 10 galaxies and repeat this step 100 times. We conduct this test twice by changing the galaxy selection method: one is completely random, while the other is to select only galaxies whose \( \langle \delta v \rangle \) within 1.7 \( h^{-1} \) Mpc of the radius are less than −0.2. The resultant CCFs of both methods show a large dispersion around the original CCF, and the dispersion becomes smaller as the sample size increases from 5 to 10. However, intriguingly, the scatter among 100 CCFs in the latter method becomes smaller than that in the former method. This suggests that the true CCF cannot be obtained from a randomly selected, extremely small sample. On the other hand, a CCF derived from a few galaxies that are located in similar IGM densities (i.e., in a limiting \( \langle \delta v \rangle \)) could still reflect their IGM environments.

**ORCID iDs**

Rieko Momose @ https://orcid.org/0000-0002-8857-2905
Kentaro Nagamine @ https://orcid.org/0000-0001-7457-8487
Kazuhiro Shimakatu @ https://orcid.org/0000-0002-2597-2231
Nobunari Kashikawa @ https://orcid.org/0000-0001-5493-6259
Haruka Kusakabe @ https://orcid.org/0000-0002-3801-434X

**References**

Adelberger, K. L., Shapley, A. E., Steidel, C. C., et al. 2005, ApJ, 629, 636
Adelberger, K. L., Steidel, C. C., Shapley, A. E., & Pettini, M. 2003, ApJ, 584, 45
Ando, R., Kohno, K., Umemata, H., et al. 2019, ApJ, 871, 256
Ando, R., Nishizawa, A. J., Ikôh, S., & Nagamine, K. 2020, arXiv:2011.13165
Astropy Collaboration, Robitaille, T. P., & Tollerud, E. J. 2013, A&A, 558, A33
Bielby, R. M., Shanks, T., Crighton, N. H. M., et al. 2017, MNRAS, 471, 2174
Blake, C., Collister, A., Bridle, S., & Lahav, O. 2007, MNRAS, 374, 1527
Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, MNRAS, 351, 1151
Cai, Z., Fan, X., Peirani, S., et al. 2016, ApJ, 833, 135
Chen, H.-W., & Mulchaey, J. S. 2009, ApJ, 701, 1219
Chen, H.-W., Prochaska, J. X., Weiner, B. J., Mulchaey, J. S., & Williger, G. M. 2005, ApJL, 629, L25
Croft, R. A. C., Miralda-Escudé, J., Zheng, Z., et al. 2016, MNRAS, 457, 3541
Croft, R. A. C., Miralda-Escudé, J., Zheng, Z., Blomqvist, M., & Pieri, M. 2018, MNRAS, 481, 1320
Daddi, E., Dickinson, M., Morrison, G., et al. 2007, ApJ, 670, 156
Elbaz, D., Daddi, E., Le Borgne, D., et al. 2007, A&A, 468, 33
Faucher-Giguère, C.-A., Prochaska, J. X., Lidz, A., Hernquist, L., & Zaldarriaga, M. 2008, ApJ, 681, 831
Font-Ribera, A., Arnau, E., Miralda-Escudé, J., et al. 2013, JCAP, 2013, 018
Fumagalli, M., Prochaska, J. X., Kasen, D., et al. 2011, MNRAS, 418, 1796
Genel, S., Vogelsberger, M., Springel, V., et al. 2014, MNRAS, 445, 175
