Environmental dependence of galactic properties traced by Lyα forest absorption (I): variation according to galaxy stellar mass and star-formation activity

RIEKO MOMOSE,1 IKKOH SHIMIZU,2 KENTARO NAGAMINE,3,4,5 KAZUHIRO SHIMASAKU,1,6 NOBUNARI KASHIKAWA,1 HARUKA KUSAKABE,7

1Department of Astronomy, School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
2National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan
3Theoretical Astrophysics, Department of Earth and Space Science, Graduate School of Science, Osaka University, 1-1 Machikaneyama, Toyonaka, Osaka 560-0043, Japan
4Kavli-IPMU (WPI), The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8583, Japan
5Department of Physics 8 Astronomy, University of Nevada, Las Vegas, 4505 S. Maryland Pkwy, Las Vegas, NV 89154-4002, USA
6Research Center for the Early Universe, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
7Observatoire de Genève, Université de Genève, 51 chemin de Pégase, 1290 Versoix, Switzerland

ABSTRACT

This is the first paper in a series to systematically investigate the environmental dependence of galactic properties traced by Lyα forest absorption in the intergalactic medium (IGM). Using our cosmological hydrodynamic 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 ($\delta F$); and (II) comparing the overdensity of neutral hydrogen (H\textsc{i}) and galaxies. Our simulations reproduce observed cross-correlation functions (CCF) between the Lyα forest and Lyman-break galaxies. We further investigate the variation of the CCF using subsamples divided by dark matter halo mass ($M_{\text{DH}}$), galaxy stellar mass ($M_*$), and star-formation rate (SFR), and find that the CCF signal becomes stronger with increasing $M_{\text{DH}}$, $M_*$, and SFR. The CCFs between galaxies and gas-density fluctuation are also found to have similar trends. Therefore, the variation of the $\delta F$–CCF depending on $M_{\text{DH}}$, $M_*$, and SFR is due to varying gas density around galaxies. We conclude that $M_{\text{DH}}$ is the most sensitive parameter to characterize the correlation between galaxies and the IGM distribution. Method (II) also finds correlations between galaxies and H\textsc{i} overdensities. Overall, our analyses confirm the spatial correlation between galaxies and IGM H\textsc{i}, with more massive galaxies being clustered in higher-density regions as expected from the ΛCDM paradigm.

Keywords: methods: numerical, galaxies: evolution – intergalactic medium, quasars: absorption lines, cosmology: large-scale structure of universe

1. INTRODUCTION

The standard picture of galaxy formation within the gravitational instability paradigm indicates that galaxy formation and evolution is closely linked to its 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 hole (SMBH). 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 Mpc-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 CGM/IGM–galaxy connection is the 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 HI overdensities have also been discussed in the literature (Mukae et al. 2017, 2019; Mawatari et al. 2017). 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 Mpc scales.

Physical properties of HI gas in the CGM or IGM have been studied in detail theoretically (e.g., Meiksin 2009; Meiksin et al. 2014, 2017; Fumagalli et al. 2011; van de Voort & Schaye 2012; van de Voort et al. 2012; Rahmati et al. 2015), aided by powerful cosmological hydrodynamic simulations such as EAGLE (Schaye et al. 2015), Illustris (Vogelsberger et al. 2014a,b; Genel et al. 2014; Sijacki et al. 2015), and IllustrisTNG (Villaescusa-Navarro et al. 2018; Nelson et al. 2019). Turner et al. (2017) presented the median HI optical depth vs. line-of-sight or transverse distance around galaxies in the EAGLE simulation, and found that it is sensitive to dark matter halo mass. Sorini et al. (2018) compared the radial profile of mean Lyα absorption around galaxies in both observations and simulations, and have shown a reasonable match between them beyond 2 proper Mpc (pMpc), but significant differences at 0.02 – 2 pMpc.

The correlation between galaxies and IGM morphology has also been examined in the literature. Martizzi et al. (2019) 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.

In order to unveil the gas distribution on Mpc-scales around galaxies, we systematically investigate it in a series of papers using both numerical simulations and observations. In this first theory paper, we aim to: 1) establish the methodology to evaluate the CGM/IGM–galaxy connection; 2) examine its dependency on galaxy properties; and 3) verify the cause of dependencies using GADGET3-Osaka cosmological hydrodynamic simulations (Shimizu et al. 2019). In this study, we particularly focus on statistical comparisons between simulations and observations, using the same parameters for a fair comparison. As we describe in the companion observational paper (Momose et al. 2020; hereafter Paper II), for observations, we use the CLAMATO (COSMOS Lyα Mapping And Tomography Observations) which is publicly-available Lyα forest 3D tomography data (Lee et al. 2014, 2016, 2018), and several other catalogs in the archives. Details of the observational data are given in Paper II.

This paper is organized as follows. We introduce our numerical simulations in Section 2 and the methodology to examine the IGM–galaxy connection in Section 3. Results and discussion are presented in Sections 4 and 5, respectively. Finally, we give our summary in Section 6. In the appendix, we discuss the dependencies 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 HI 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 the cosmological hydrodynamic simulations performed with GADGET3-Osaka (Shimizu et al. 2019), which is a modified version of the Tree-PM SPH code GADGET-3 (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 stellar mass function, SFR function, stellar-to-halo-mass ratio, and cosmic star formation history within observational uncertainties at $z \geq 2$ (Shimizu et al. 2020, in preparation).

Here, we briefly describe our simulations, which employed $N = 2 \times 512^3$ particles in a comoving volume of $(100\, h^{-1}\, \text{Mpc})^3$. The particle masses of dark matter and gas are $5.38 \times 10^8 \, M_\odot$ and $1.00 \times 10^8 \, h^{-1} \, M_\odot$, respectively. The gravitational softening length is set to be $8\, h^{-1} \, \text{kpc}$ in comoving units. 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 supernova and AGB stars).

In order to identify 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 minimum 32 particles (dark matter + SPH + star) as substructures (subhalos) in each FoF group using the SUBFIND algorithm (Springel et al. 2001). 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 ($M_{\text{DH}}$), which is defined by the total enclosed mass inside a sphere of 200 times the critical density of the Universe. This means that the member galaxies (central and satellite galaxies) in a dark matter halo have the same $M_{\text{DH}}$, even though the substructures can have different subhalo masses calculated by SUBFIND.

Note that each gas (star) particle has some associated physical properties such as mass, star-formation rate (SFR) and metallicity. In this study, the properties (gas mass, stellar mass $M_*$, and SFR) of a simulated galaxy are defined by summation of these quantities in each subhalo.

In order to directly compare our simulations and observations, we created 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 from $z \sim 1.8$ to 3.1 which can cover the redshift range of recent Lyα absorption line surveys (e.g., CLAMATO; Lee et al. 2014, 2016, 2018) and future PFS Lyα absorption survey (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 line-of-sight (LoS) at different epochs.

With this light-cone output, we calculate the Lyα 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 LoS, such as HI density, LoS velocity and temperature as follows:

$$A_{\text{grid}}(x) = \sum_{j} \frac{m_j}{\rho_j} A_j W(r, h_j),$$

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 $j$-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 ($dl$) is set to a constant value of $0.1 \ h^{-1}$ Mpc in comoving units which is higher resolution than any of the relevant Lyα observations. Then, we calculate the Lyα optical depth $\tau_{\text{Ly} \alpha}(x)$ using these physical values at each grid point as follows:

$$\tau_{\text{Ly} \alpha}(x) = \frac{\pi e^2}{m_e c} \int_{j} \phi(x - x_j) n_{\text{HI}}(x_j) dl,$$

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, HI 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 ($= 32^3$) LoSs are drawn with regularly spaced intervals. The mean separation of each LoS is $3.3 \ h^{-1}$ Mpc which is similar to the CLAMATO survey.

Finally, we note that we do not consider the feedback by the central SMBHs (AGN feedback) in our current code. Thus, a significant amount of gas might remain in massive galaxies without being ejected into the CGM/IGM by the feedback, but this effect is considered to be not so strong at $z > 2$ with black holes not being super-massive yet.

3. METHODOLOGY

One of the main purposes of this series of study is to compare with the CLAMATO, which is a 3D tomography data of Lyα forest transmission fluctuation ($\delta_F$; see the following definition) over $2.05 < z < 2.55$ in 0.157 deg$^2$ of the COSMOS field (Scoville et al. 2007; Lee et al. 2016, 2018). The CLAMATO consists of $60 \times 48 \times 876$ pixels corresponding to $30 \times 24 \times 436 \ h^{-1}$ Mpc cubic with a pixel size of 0.5 $h^{-1}$ Mpc. The average separation of background galaxies for measuring Lyα absorption are [2.61, 3.18] $h^{-1}$ Mpc in [RA, DEC] directions, and the separation in LoS-direction is $2.35 \ h^{-1}$ Mpc at $z \sim 2.3$. (See Paper II for more detailed comparison with CLAMATO.)

To produce a similar data cube of $\delta_F$ from our LoS data, we first evaluate the Lyα transmission fluctuation $\delta_F$ in each LoS pixel by

$$\delta_F \equiv \frac{F(x)}{\langle F_2(x) \rangle} - 1,$$

where $F(x) = \exp[-\tau_{\text{Ly} \alpha}(x)]$ is the Lyα flux transmission, and $\langle F_2(x) \rangle$ is the cosmic mean transmission. We adopt the following value derived by Faucher-Giguère et al. (2008):

$$\langle F_2(x) \rangle = \exp[-0.00185 (1 + z)^{3.92}],$$

because it is used in CLAMATO. Additionally, we also use the same setup used in CLAMATO with Hubble constant $h = 0.7$ and 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
The cross-correlation analysis is often used in the literature to characterize the correlation between galaxies and 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 cross-correlation function (CCF):

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

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

To calculate CCFs, we prepare two LoS data with different LoS grid resolution. One is the LoS data with original resolution of comoving 0.1 \(h^{-1}\) Mpc. The other is a lower resolution data with 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 dataset as ‘LoS-4’. For comparison with observations, we use the LoS-4 dataset and calculate CCFs at \(1 - 100 h^{-1}\) Mpc scale around galaxies.

At the same time, we also use the original LoS dataset to derive CCFs at \(r = 0.16 - 1 h^{-1}\) Mpc, because the redshift resolution of LoS-4 data is larger 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 1–1(e).

The \(\xi_F\) value is evaluated in each shell with a thickness of \(\log(\Delta r h^{-1}\text{ Mpc}) = 0.2\) and \(\log(\Delta r h^{-1}\text{ Mpc}) = 0.1\) for the CGM and IGM regimes, respectively. We confirm that our CCFs by 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 \(\xi_F\) value, and adopt 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 3 – 5 subsamples according to \(M_*, M_{DH}\), SFR and specific SFR (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 the 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 as the “Overdensity analysis”. We first generate a 2D LoS map by binning the LoS data with \(\Delta z = 0.032\), which corresponds to \(27.9 h^{-1}\) Mpc. We then estimate the mean IGM fluctuation (\(\delta\)) within circles of radius 4.74 \(h^{-1}\) Mpc centered on local minima and maxima of the map. The sizes of both \(\Delta z\) and cylinder radius were chosen to be comparable to the actual observations (see Paper II). Note that we use a cylinder of \(\Delta z = 0.08\) \((\Delta z = 0.032)\) corresponding

| Category          | Range         | Number | Sample Name |
|-------------------|---------------|--------|-------------|
| All galaxies      | \(10^{11} \leq M_*\) | 1662   | \(M_*-11\)  |
|                   | \(10^{10} \leq M_* < 10^{11}\) | 21975  | \(M_*-10\)  |
|                   | \(10^9 \leq M_* < 10^{10}\) | 65809  | \(M_*-9\)   |
| Stellar mass \(M_\odot\) | \(10^{13} \leq M_{DH}\) | 1874   | \(M_{DH}-13\) |
|                   | \(10^{12} \leq M_{DH} < 10^{13}\) | 20407  | \(M_{DH}-12\) |
|                   | \(10^{11} \leq M_{DH} < 10^{12}\) | 66803  | \(M_{DH}-11\) |
|                   | \(10^{10} \leq M_{DH} < 10^{11}\) | 362    | \(M_{DH}-10\) |
| log SFR \(M_\odot \text{ yr}^{-1}\) | \(2 \leq \log \text{ SFR}\) | 1152   | SFR-(i) |
|                   | \(1 \leq \log \text{ SFR} < 2\) | 24349  | SFR-(ii) |
|                   | \(0 \leq \log \text{ SFR} < 1\) | 46425  | SFR-(iii) |
|                   | \(-1 \leq \log \text{ SFR} < 0\) | 14654  | SFR-(iv) |
|                   | \(-1 > \log \text{ SFR}\) | 2866   | SFR-(v) |
| log sSFR \(\text{yr}^{-1}\) | \(-9 \leq \log \text{ sSFR}\) | 38078  | sSFR-(i) |
|                   | \(-10 \leq \log \text{ sSFR} < -9\) | 47419  | sSFR-(ii) |
|                   | \(-10 > \log \text{ sSFR}\) | 3949   | sSFR-(iii) |
to 69.7 (27.9) $h^{-1}$ Mpc in length, and with a radius of 3 (4.74) $h^{-1}$ Mpc in our observational analysis because the mean separation of our LoS data is 2.35 $h^{-1}$ Mpc. To determine the pixel positions of local min/max, we first mark the positions of local min/max of $\delta_F$ within a few $h^{-1}$ Mpc scale, and then further repeat the same procedure to identify the min/max on even smaller scales.

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

$$\Sigma_{\text{gal}} = \frac{N_{\text{gal}}}{\langle N_{\text{gal}} \rangle} - 1,$$

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. Since we first generate a 2D LoS map, the galaxy overdensity computed above can be regarded as galaxy surface density, and thus we denote it as $\Sigma_{\text{gal}}$.

To increase the number of data points, we randomly select four different redshift slices. The number of galaxies in each redshift slice is summarized in Table 2. We also perform the overdensity analysis for randomly selected positions to examine possible bias due to the positioning of cylinders (see also Mukae et al. 2017).

4. RESULTS

4.1. Cross-correlation analysis

4.1.1. $\text{Ly}_\alpha$ absorption fluctuation

The CCFs of all galaxies and subsamples are shown in Figure 1–1. We also present the sign-flipped CCF plotted in log-scale in Figure 1–2 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 CCF for four subsamples (divided by $M_*$, $M_{\text{DH}}$, SFR, and sSFR) are presented in Figure 1–1(a–d). Most of the CCFs show monotonic increase from the center to $r = 20 - 60$ $h^{-1}$ Mpc, except for the $M_{\text{DH}}-13$ and $M_{\text{DH}}-10$ sample which show irregular shapes. Considering our tests for CCF reproducibility by a small sample size in Appendix D, a swelling at $r \sim 0.4$ $h^{-1}$ Mpc in $M_{\text{DH}}-10$ can be attributed to the small sample size. While for $M_{\text{DH}}-13$, we regard a loosely bump at $r = 0.3 - 0.8$ $h^{-1}$ Mpc as a real feature.

We find that there is a clear tendency of CCF signal depending on the subsample, except for sSFR subsample. It is that the CCF signal becomes stronger with increasing galaxy masses and SFRs, but SFR–$(v)$ subsamples do not follow this trend. A turnover radius where $\xi_{\text{SF}}$ reaches about zero also shows a trend for galaxies in $M_*$ and $M_{\text{DH}}$ subsamples (see Figure 1–2(a) and (b)), that a sample with a higher CCF signal drops rapidly to zero at a smaller radius, and hereafter we call this as ‘turnover radius’. For SFR samples, however, the CCF signal does not seem to correlate with turnover radius. On the other hand, sSFR samples do not show any obvious trend in their CCFs. Nonetheless, the turnover radius of sSFR samples increases with increasing sSFR. Likewise, in previous studies, Turner et al. (2017) have demonstrated the halo mass dependency of the median $\tau_{\text{HI}}$ as a function of distance from their modeled galaxies. Melkins et al. (2017) have investigated $\delta_F$ for galaxies in $M_{\text{DH}}$ subsamples against projected impact parameter and found an increase in $\delta_F$ with halo mass. Observationally, Chen et al. (2005) have measured two-point cross-correlation $\xi_{\text{ga}}$ between $\text{Ly}_\alpha$ 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 $\xi_{\text{ga}}$ (see also Chen & Mulchaey 2009). However Wilman et al. (2007) did not find any significant differences in the cross-correlation signal between absorption-line-dominated and emission-line-dominated galaxies.

To characterize our CCFs, we fit them with a power-law of

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

where $r_0$ and $\gamma$ are a clustering length and slope. We apply the power-law fitting over $0.1 - 1$ $h^{-1}$ Mpc and $3 - 20$ $h^{-1}$ Mpc, corresponding to the CGM and IGM regimes. The fitting range for the IGM regime is set for the comparison with observations in Paper II. Best-fit parameters of the all 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 $(0.62 \pm 0.04, 1.37 \pm 0.04)$ for the CGM and IGM regimes (see also Figure 2(e)). Several observational studies have performed a power-law fitting to their CCFs between $\text{Ly}_\alpha$ absorption and galaxies. Tummuangpak et al. (2014) have calculated a CCF between $\text{Ly}_\alpha$ absorption and LBGs at $z \approx 3$, 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 the IGM–LBG clustering by Bielby et al. (2017) have been 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 a CCF of weak $\text{H}i$ ($N_{\text{HI}} < 10^{14}$ cm$^{-2}$) IGM and galaxies at $z < 1$ have been also 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
Figure 1. 1): CCFs obtained from our simulation as a function of radius in comoving units. Dashed vertical and horizontal lines represent a half of the pixel size in transverse direction and a mean separation of LoS. The definition of the CGM and IGM regimes used in this paper is also shown in panel (e). Panels (a)–(d): CCFs for each subsample divided by $M_\star$, $M_{\text{DH}}$, SFR, and sSFR. Panel (e): CCF calculated using all galaxies. Blue circles and squares indicate the observational estimates of CCF between $\delta_F$ and LBGs from Adelberger et al. (2005) and Bielby et al. (2017), respectively. 2): Sign-flipped CCF of Fig. 1–1 is plotted in log-scale. The vertical dashed line is the same as in Fig. 1–1.
Figure 2. Best-fit parameters of a power-law fitting. Filled and open circles represent the best-fit parameters obtained from the CGM and IGM regime, respectively. When error bars are not recognized, they are smaller than symbol sizes. Panels (a)–(e): the parameters for the CCFs in $M_\star$, $M_{\text{DH}}$, SFR, and sSFR categories, and for the CCF from all galaxies.

In studies, our best-fit parameters for both CGM and IGM regimes are comparable to those previous observations within the error.

We next show best-fit parameters obtained from all categories. For $M_\star$ and $M_{\text{DH}}$, they show similar trend in both $r_0$ and $\gamma$ of the both CGM and IGM regimes. A clustering length $r_0$ becomes longer with increasing mass independent of the regimes. Meanwhile, a slope $\gamma$ becomes smaller with increasing mass in the CGM regime, but becomes larger 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 $M_\star > 10^8$ and $M_\star > 10^9\, h^{-1} M_\odot$. Their double power-law fit for these $M_\star > 10^8$ and $M_\star > 10^9\, h^{-1} M_\odot$ samples have been given ($r_0, \gamma$) = (0.10 ± 0.07, 0.46 ± 0.22) and (0.16 ± 0.09, 0.46 ± 0.19) for the CGM regime, and (0.51 ± 0.39, 1.25 ± 0.61) and (0.61 ± 0.34, 1.18 ± 0.43) in the IGM regime. Although differences of $r_0$ and $\gamma$ between $M_\star > 10^8$ and $M_\star > 10^9\, h^{-1} M_\odot$ samples are within the error, the similar trend is confirmed in $r_0$ but is absent in $\gamma$ estimates. Because $M_\star > 10^8\, h^{-1} M_\odot$ sample of Tummuangpak et al. (2014) also includes objects with $M_\star > 10^9\, h^{-1} M_\odot$, their best-fit parameters might not characterize the CCF from galaxies with $10^8 < M_\star \leq 10^9\, h^{-1} M_\odot$. For the SFR sample except SFR–(v), we identify that $r_0$ becomes greater with increasing SFR. It 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 IGM regime. Although the CCFs of all three sSFR category are comparable as presented in Figure 1–1(d), best-fit parameters of both $r_0$ and $\gamma$ are different each other. Interestingly, both $r_0$ and $\gamma$ show identical trends in the both CGM and IGM regimes, that sSFR value becomes smaller with increasing the parameters’ estimates. We briefly discuss its reason in Section 5.1.

4.1.2. Gas density fluctuation
Figure 3. 1) CCFs of total gas density fluctuation ($\delta_{\rho_{\text{gas}}} = \rho_{\text{gas}}/\langle \rho_{\text{z}} \rangle$) as a function of comoving distance from galaxies. 2) Mean total gas density fluctuation around galaxies.
Figure 4. 1) CCFs of H\textsubscript{i} gas density fluctuation ($\delta_{\rho_{\text{H}\text{i}}} = \rho_{\text{H}\text{i}}/\langle \rho_{z} \rangle$) as a function of comoving distance from galaxies. 2) Mean H\textsubscript{i} density fluctuation around galaxies.
A $\delta_F$ value of LoS data shall correlate with $\text{Hi}$ gas density at the position. Thus, a variety of the CCFs amplitudes can be attributed to a variety of the local gas density around galaxies. To verify the hypothesis, we evaluate CCFs of gas density fluctuations around galaxies defined by

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

where $\delta_p$ is the gas density fluctuation defined by the ratio of a gas density at one LoS pixel $\rho$ to the mean gas density at each redshift of $\Delta z = 0.01$, $\langle \rho \rangle$: $\xi_{\delta p}$ is the CCF; $\delta_{\rho_{p,i}}$ and $\delta_{\rho_{ran,j}}$ are the gas density fluctuation for the pixel $i$ and $j$ at the distance $r$ from galaxies and random points. Similarly, in Equation (5), $N(r)$ and $M(r)$ are the numbers of pixel-galaxy and pixel-random pairs in the bin with the distance $r$. The CCF of gas density fluctuations are measured for both total gas and $\text{Hi}$ in Figure 4. Compared to the corresponding CCFs' strength or shape at a certain radius. For $\delta_{\rho_{\text{Hi}}}$ (i.e., $-\log \delta_F$) are also shown in Figure 1–2. Due to several negative values in $\delta_{\rho_{\text{Hi}}}$, we also present the mean gas density fluctuations around galaxies in Figures 3–2 and 4–2.

First, we start from the CCFs of total gas density fluctuation $\delta_{\rho_{\text{gas}}}$, in Figure 3. Overall trends for each category (i.e., $M_{\star}$, $M_{\text{DH}}$, SFR, and sSFR) are almost the same as that of $\delta_F$'s CCFs. It is that overall CCFs are monotonically decreasing with radius, and a CCF signal becomes higher with increasing galaxies mass (either $M_{\star}$ or $M_{\text{DH}}$) or SFR. However several samples show a notable behavior in their CCFs in the context of a CCF's strength or shape at a certain radius. For $M_{\star}$, $M_{\text{DH}}$ and SFR–(i), we find a slight decline of $\xi_{\delta_{\rho_{\text{gas}}}}$ at the center. It indicates the decline of relative total gas density in the proximity of those galaxies. For SFR–(v) and sSFR–(iii), their CCFs show a convex feature at $r = 0.3 - 2 \, h^{-1} \, \text{Mpc}$ and even have a highest signal among each category over that radius. It is note for $M_{\text{DH}}$–10 that a swelling at $r \sim 0.5 \, h^{-1} \, \text{Mpc}$ can be due to its small sample size.

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

4.2. Overdensity analysis

We present the results of overdensity analysis in Figures 5–(a) and 5–(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 $\Sigma_{\text{gal}}$ and $\langle \delta_P \rangle$ values in each of four redshift slices indicated by the open or filled circles colored in red or blue. Exact galaxy counts used in each redshift slice is shown in Table 2.

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

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

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

$$\langle \delta_F \rangle = \alpha + \beta \Sigma_{\text{gal}}.$$  

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 four samples within the error, while $\beta$ becomes slightly larger with increasing $M_{\star}$, although they are still similar within the error: $\beta = -0.007 \pm 0.003$, $-0.014 \pm 0.005$, $-0.020 \pm 0.007$ for $M_{\star}$–11, $M_{\star}$–10, $M_{\star}$–9, respectively.
Figure 5. Overdensity analysis obtained from (a) local minima and maxima of gas density field, and (b) random points, respectively. Results for $M_{-11}$, $M_{-10}$, $M_{-9}$ and ALL are shown from left to right. Open and filled circles colored in red and blue indicate data points from each of four 2D LoS map. The best-fit linear regression is shown in black line with its errors shown by grey shade.

Table 2. Measurements results of over-density examinations

| Sample Name | $N^{(1a)}$ | $N^{(1b)}$ | $N^{(1c)}$ | $N^{(1d)}$ | $R_s^{(2)}$ | $p^{(2)}$ | $\alpha^{(3)}$ | $\beta^{(3)}$ |
|-------------|-------------|-------------|-------------|-------------|-------------|------------|-------------|-------------|
| $M_{-11}$   | 134         | 139         | 100         | 85          | -0.41       | 0.12       | -0.129 ± 0.011 | -0.007 ± 0.003 |
| $M_{-10}$   | 1709        | 1555        | 1226        | 1182        | -0.42       | 6.27e-3    | -0.126 ± 0.006 | -0.014 ± 0.005 |
| $M_{-9}$    | 4658        | 4356        | 3984        | 3921        | -0.33       | 0.03       | -0.126 ± 0.006 | -0.020 ± 0.007 |
| ALL         | 6501        | 6650        | 5310        | 5188        | -0.37       | 0.02       | -0.126 ± 0.006 | -0.018 ± 0.006 |

Note— (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 $\langle \delta F \rangle - \Sigma_{gal}$ relation is examined around local minima and maxima. (3) The best-fit parameters of chi-square fitting of the $\langle \delta F \rangle - \Sigma_{gal}$ relation examined around local minima and maxima. (4) Spearman’s coefficient and $p$–value. The $\langle \delta F \rangle - \Sigma_{gal}$ relation is examined around random points. (5) The best-fit parameters of chi-square fitting of the $\langle \delta F \rangle - \Sigma_{gal}$ relation examined around random points.
Note that the best-fit parameters of anti-correlations for $M_*-10$, $M_*-9$ 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 $(\alpha, \beta) = (-0.17 \pm 0.06, -0.14^{+0.06}_{-0.16})$. We find a similarly in $\alpha$ but a larger difference in $\beta$, showing a much shallower slope for our sample. The shallower slope of simulated galaxy sample has also been found Nagamine et al. (2020, in preparation). It may be attributed to photo-z errors in the observational data. If photo-z errors are large, then some galaxies would contaminate the sample, and the value of $\Sigma_{\text{gal}}$ would be smeared 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)). The Spearman’s rank correlation tests for randomly-selected positions yield mild anti-correlations with $R_s = -0.51$ to $R_s = -0.57$. The best-fit parameters of the linear model ($\alpha$ and $\beta$) are also comparable to those from Figure 5–(a) within the error. Moreover, the trend found in best-fit $\alpha$ and $\beta$ as a function of stellar-mass is also confirmed. Therefore, we conclude that the position bias of overdensity analysis is not affecting the $(\delta_F) - \Sigma_{\text{gal}}$ correlation seriously.

5. DISCUSSIONS

5.1. Origin of CCF variation

We presented in Section 4.1.1 that the CCF of $\delta_F$ varies depending on galaxy mass and SFR. To find the origin of its variation, we also calculated the CCFs of $\delta_{\rho_{\text{gas}}}$ and $\delta_{\rho_{\text{H}_2}}$ in Section 4.1.2, and found that their signal strengths also depend on $M_*$, $M_{\text{DHI}}$ and SFR. It suggests that different relative gas density around galaxies is causing the variation of the CCFs of $\delta_F$. Considering the relation between $\delta_F$ and $\text{H}_i$ number density in Equations (2)–(4), a similar trend of CCFs in $\delta_F$ and $\delta_{\rho_{\text{gas}}}$ is reasonable. The same trend even for $\delta_{\rho_{\text{H}_2}}$ probably means that the total gas density correlates with $\text{H}_i$ gas density in general. Therefore, we argue that the variation of $\delta_F$ CCF is caused by different gas distribution around galaxies.

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

Considering the above discussion, the signal strength and best-fit parameters ($r_0$ and $\gamma$) of CCF depends on relative gas densities on Mpc-scale 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 higher signal and give larger best-fit parameters for the IGM-regime. The variation of CCFs is hence attributed to different gas density around each galaxy and the strength of galaxy–IGM connection.

5.2. Which type of galaxies strongly correlate with the IGM?

Under the ΛCDM 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 confirmed that the CCF signal becomes stronger with increasing mass (both $M_*$ and $M_{\text{DHI}}$) of a galaxy. In addition, we also find that the turnover radius of CCFs becomes smaller with mass in $M_* - 11$ and $M_{\text{DHI}} - 13$, which is likely to be the result of stronger connection between massive galaxies and higher density regions.

From the overdensity analysis in Section 4.2, we find that the slope of the anti-correlation between $\langle \delta_F \rangle$ and $\Sigma_{\text{gal}}$ becomes shallower with increasing $M_*$, although its difference is still within the error. A shallower slope in $M_* - 11$ sample indicates stronger clustering of massive galaxies around dense $\text{H}_i$ IGM.
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 an opposite trend.

The same trend should be true for SFR subsamples by considering the star-formation main sequence (e.g., Brinchmann et al. 2004; Noeske et al. 2007; Elbaz et al. 2007; Daddi et al. 2007; Speagle et al. 2014; Schreiber et al. 2015; Tomczak et al. 2016). The overall trend for the SFR samples is that the galaxies with higher (lower) SFRs correlate with higher (lower) gas density of the IGM. However, 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 in SFR–(v), which has two peaks at \( M_{\text{DH}} \sim 10^{11.3} - 10^{11.5} M_\odot \). It implies that the host dark matter halos of SFR–(v) subsample can be roughly divided into two: one is less massive and the other is massive. Given the mass-dependency of IGM–galaxy connection, the strong CCF (i.e. high-density gas) seen at \( r = 0.3 - 2 h^{-1} \text{Mpc} \) of SFR–(v) subsample reflects high gas density around massive halos.

Our sSFR samples do not show any obvious trends in the CCF strength. We find that all three CCFs in sSFR subsample are comparable, but their best-fit parameters become greater with decreasing sSFR. Similar trends were observed for the SFR–(v) sample, which resulted from different halo mass distribution in each sSFR sample (see Figure 7): sSFR–(i) and sSFR–(ii) subsamples show a single peak at \( M_{\text{DH}} \sim 10^{11.3} - 10^{11.5} M_\odot \). On the other hand, the sSFR–(iii) subsample has a mild bimodal distribution which is similar to the one for SFR–(v). As a result, sSFR–(iii) subsample shows stronger correlation to higher density region than the other two subsamples.

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

Finally, we should briefly discuss about declining CCF signal at \( r < 0.3 - 0.4 h^{-1} \text{Mpc} \) around galaxies in \( M_{\star} - 11, M_{\text{DH}} - 13 \), SFR–(i), SFR–(v), sSFR–(iii) subsamples (see Figs. 3 and 4), which are hosted by the most massive halos among each category. There are several possible reasons for the lack of Hi 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 the 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 our future work.
5.3. *Photo-z vs. Spec-z data and the IGM–galaxy connection*

Observationally, it is usually difficult to completely identify galaxies, especially the faint galaxies which are probably low mass and/or massive but with little star formation. Moreover, even if we successfully find all galaxies from photometric data, we cannot measure the spectroscopic redshift for all of them. In this study, we use two methods to examine IGM–galaxy connection. One is the cross-correlation, and the other is the overdensity analysis. In this subsection, we briefly discuss the reliability of using either photo-z or spec-z data to investigate IGM–galaxy connection.

We have demonstrated in this paper that the cross-correlation method succeeds in identifying the variations of 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 would be attenuated if 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 only with photometric redshift cannot be used for the cross-correlation analysis.

On the other hand, galaxies only with photo-z are usable for the overdensity analysis, but still have following possible problems. First is a contamination of galaxies whose real redshifts are out of 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 the difficulty to statistically confirm a correlation, when the cylinder volume is large (see also Appendix A). Both this study and those in the literature show the presence of 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 the large-scale structure in $\delta_F$. If either the cylinder length or radius is larger than the above scale, the large-scale structure traced by cosmic web and galaxies will be attenuated, and thus both $\langle \delta_F \rangle$ and $\Sigma_{\text{gal}}$ would become close to zero. In that case, the slopes for $\langle \delta_F \rangle - \Sigma_{\text{gal}}$ relation would become indistinguishable.

Based on all these arguments, we propose that the cross-correlation method for galaxies with spec-z measurements is the most reliable way to investigate the IGM–galaxy connection over $1 h^{-1}$ Mpc scale using the actual observational data. It will also be useful to examine the CCF variation according to galactic properties. Alternatively, the overdensity analysis can be useful to confirm the presence of spatial correlation between IGM and galaxies, when only photo-z data is available.

6. SUMMARY

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

1. We calculate CCFs between Ly$\alpha$ forest transmission fluctuation ($\delta_F$) and galaxies as shown in Figure 1. The CCF obtained from all galaxies reproduce the one from LBGs in the literature.
2. We measure CCFs between gas density fluctuations \( \delta_{\text{gas}} \) and \( \delta_{\text{hi}} \) and galaxies in Figures 3 and 4. Overall trends of CCFs are similar to those of CCFs in \( \delta_F \) except for SFR–\( \langle v \rangle \) and sSFR–iii. It indicates that the variation in \( \delta_F \) reflects different relative gas densities around galaxies; i.e., galaxies with higher mass and SFR generally reside in higher density gas, and vice versa. For the SFR–\( \langle v \rangle \) and sSFR–iii subsamples, we find the highest CCF signal at \( r = 0.3 - 2 \, h^{-1} \, \text{Mpc} \) among all SFR and sSFR samples. Because the two subsamples have a mild bimodal halo mass distribution with two peaks at \( M_{\text{DH}} \sim 10^{11.3} \, M_\odot \) and \( M_{\text{DH}} \sim 10^{12.5} \, M_\odot \), their highest 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 \( \text{H} I \)) around galaxies.

3. Our overdensity analysis between galaxy overdensity \( \Sigma_{\text{gal}} \) and mean IGM fluctuation \( \delta_F \) is presented in Figure 5. We statistically identify anticorrelations from all subsamples of \( M_* - 11 \), \( M_* - 10 \), \( M_* - 9 \) and ALL. In addition, we also find that their slopes are decreasing with increasing \( M_* \), although within the error. It suggests that galaxies in the \( M_* - 11 \) subsample are more strongly correlated with higher density gas than those in \( M_* - 9 \) 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 determine 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 anti-correlation between \( \langle \delta_F \rangle \) and \( \Sigma_{\text{gal}} \) at \( r \geq 1 \, h^{-1} \, \text{Mpc} \).

Overall, our analyses confirm strong connection between galaxies, dark matter halos, and 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, \( \text{H} I \), and metals will provide useful information on the interaction between them and the details of feedback mechanisms which is important for the theory of galaxy formation and evolution such as galactic wind (kinetic feedback) and associated ejection of metals into IGM.

By comparing the results of Figures 1 & 2 against predictions of linear perturbation theory, we can infer the mean bias parameters of Ly\( \alpha \) forest and galaxies relative to underlying dark matter density field (e.g., Croft et al. 2016; Bielby et al. 2017; Kakiichi 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 with those inferred from 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 our further work.
ACKNOWLEDGMENTS

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. RM acknowledges a Japan Society for the Promotion of Science (JSPS) Fellowship at Japan. KN is grateful to Volker Springel for providing the original version of GADGET-3, on which the GADGET3-Osaka code is based on. Our numerical simulations and analyses were carried out on the XC50 systems at the Center for Computational Astrophysics (CiCA) of the National Astronomical Observatory of Japan (NAOJ), the XC40 system at the Yukawa Institute for Theoretical Physics (YITP) in Kyoto University, and the OCTOPUS at the Cybermedia Center, Osaka University as part of the HPCI system Research Project (hp180063, hp190050). This work is supported by the JSPS KAKENHI Grant Numbers JP18J40088 (RM), JP17H01111 (KN), and JP19K03924 (KS). KN acknowledges the travel support from the Kavli IPMU, World Premier Research Center Initiative (WPI), where part of this work was conducted.
APPENDIX

We conduct several tests for generating 2D LoS maps and the cross-correlation analysis based on the supposition of an actual observational data. In this appendix, we briefly show our results. Note that we only calculate the CCFs beyond $r \geq 1 \ h^{-1} \text{Mpc}$ for those tests in order to directly compare with the observational results presented in Paper II.

A. 2D LOS MAPS

We show 2D LoS $\delta_F$ maps binned by 9, 18, 45, and $90 \ h^{-1} \text{Mpc}$ which correspond to $\Delta z = 0.01$, 0.02, 0.05, and 0.1 in Figure 8. Each column indicates 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 binning the LoS data. In addition, the dynamic range of $\langle \delta_F \rangle$ of 2D LoS maps binned by more than 45 $h^{-1} \text{Mpc}$ seems to be too narrow to differentiate the environment based on galaxy properties.

![Figure 8](image-url)

**Figure 8.** 2D LoS maps of four arbitrary redshift slices are shown in four rows. Different binning widths to generate 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 the overdensity analysis ($r = 4.74 \ h^{-1} \text{Mpc}$).

B. THE IMPACT OF REDSHIFT MEASUREMENT UNCERTAINTIES ON THE CCF

The cross-correlation analysis used in this study requires spec-z sample of galaxies (see Section 5.3). Nonetheless, 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 $z_{\text{use}} = z_{\text{real}} \pm \sigma_z$, where $z_{\text{use}}$ and $z_{\text{real}}$ 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 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 $\sigma_z$ values are $(\sigma_z = 0.1, 0.05, 0.02, 0.01) = (90, 45,$
Figure 9. (Top) The CCFs from galaxies with redshift uncertainties of $\sigma_z \leq 0.1$, 0.05, 0.02, and 0.01. Thin lines indicate the CCFs of 10 tests. The dotted, dash-dotted, dashed and solid lines represent the mean of 10 tests with $\sigma_z \leq 0.1$, 0.05, 0.02, and 0.01. The original CCF derived from all galaxies is colored in red. (Bottom) The CCF ratio of original to the mean of 10 tests ($\Xi = \xi_{org}/\xi_{\sigma z}$). The gray shade shows the error of $\Xi$. Vertical four lines represent the effective radius of $\sigma_z \leq 0.1$, 0.05, 0.02, and 0.01 corresponding to 90, 45, 18, and 9 $h^{-1}$ Mpc.

18, 9) $h^{-1}$ Mpc at $\langle z \rangle = 2.3$. In the bottom panel of Figure 9, the ratio of CCF from all galaxies colored in red ($\xi_{org}$) to the mean of the CCFs from galaxies in consideration of redshift uncertainties ($\xi_{\sigma z}$) is also presented ($\Xi = \xi_{org}/\xi_{\sigma z}$).

We find that all CCF signals become weaker with respect to the original CCF colored in red, though the scatter of CCF signal 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. It suggests that galaxy data set with such a large photo-z errors is 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, albeit 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 one. Within $r < 40$ $h^{-1}$ Mpc, each sample reaches $\Xi = 0$ at a radius which 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 $h^{-1}$ 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 set 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.

C. COSMIC VARIANCE

If a survey volume is not large enough, cosmic variance must affect the CCFs (both amplitude and shape). To evaluate this effect, we measure CCFs by limiting the volume to following two sizes: \( \Delta z = 0.1 \) & 0.05, corresponding to 91 \( h^{-1} \) Mpc and 45 \( h^{-1} \) Mpc in redshift direction, respectively. Note that we miss large-scale fluctuations due to the limited simulation volume, and thus we inevitably underestimate the effect of cosmic variance. The CCFs derived from limited volumes are presented in Figure 10, where we see the scatter due to cosmic variance around the original CCF colored in red.

To quantify the CCF variation, we fit them with a power-law of Equation (8) over \( 3 - 20 h^{-1} \) Mpc in radius. Best-fit parameters are shown in Figure 11. Both \( r_0 \) and \( \gamma \) show a large variation. It suggests that cosmic variance also influences both the slope and clustering-length of CCF. Therefore, when we compare the CCFs of two different galaxy populations, it would be desirable to match their redshift coverage for CCF calculation. We should note that we do not find any redshift dependence for both \( r_0 \) and \( \gamma \) measurements from our simulations.
D. CCFS OBTAINED FROM SMALL SAMPLE SIZE

In Appendix B, we discussed 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 are limited, and therefore the derived CCF from spec-z data may suffer from small sample size and differ from true signal. Thus, we carry out following two tests in order to verify the effect of sample size on resultant CCF.

The first test is to change the completeness of galaxy sample. We randomly select 0.1, 1, and 10% of galaxies from the entire sample, and recalculate CCF. This routine is repeated 10 times, which is shown by the black thin lines in Figure 12). The mean of all trials are shown by the dotted line in each panel. As shown in Figure 12(a), the CCFs from 0.1%–sample shows some scatter especially at \( r < 4 \, h^{-1} \) Mpc. However, the scatter around the mean becomes smaller with increasing fraction of galaxy sample from panel (a) to (c). Surprisingly, all 10 CCFs in the 1%–sample are very close to the original CCF, and those in 10%–sample show almost no scatter. Therefore, we argue that at least 1% of the total sample is required for spec-z measurements in order 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 extremely small sample of randomly selected 5 and 10 galaxies, and repeat it 100 times. All 100 CCFs and their mean are shown in Figure 13–1, together with the original CCF colored in red. We demonstrate that all 100 CCFs in 5 and 10 samples show a large dispersion around the original CCF. However the dispersion becomes smaller as the sample size increases from five to ten. In fact, the CCFs derived from 0.1% of all galaxies in Figure 12 have much smaller dispersion than those in Figure 13–1.

We also perform a second test assuming a situation that all galaxies in a small sample preferentially reside in a similar IGM density. To examine such a case, we impose a condition when we select galaxies, that galaxies whose \( \langle \delta F \rangle \) within 1.7 \( h^{-1} \) Mpc in radius is less than \(-0.2\). The results of such 100 tests are shown in Figure 13–2. Although the amplitudes of CCFs and their mean are larger than the original CCF until \( r \sim 20 \, h^{-1} \) Mpc, the scatter among 100 CCFs becomes smaller than those of CCFs in Figure 13–1. Additionally, most of the CCFs show strongest signals at the center. This additional second test suggest that the true CCF cannot be obtained from a randomly-selected, extremely small sample size. Whereas the CCF derived from a few galaxies which are located in similar IGM densities, is perhaps still able to reflect their IGM environments.

REFERENCES

Adelberger, K. L., Shapley, A. E., Steidel, C. C., et al. 2005, ApJ, 629, 636, doi: 10.1086/431753
Adelberger, K. L., Steidel, C. C., Shapley, A. E., & Pettini, M. 2003, ApJ, 584, 45, doi: 10.1086/345660
Bielby, R. M., Shanks, T., Crighton, N. H. M., et al. 2017, MNRAS, 471, 2174, doi: 10.1093/mnras/stx1772
Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, MNRAS, 351, 1151, doi: 10.1111/j.1365-2966.2004.07881.x
Cai, Z., Fan, X., Peirani, S., et al. 2016, ApJ, 833, 135, doi: 10.3847/1538-4357/833/2/135
Figure 13. The CCFs obtained from extremely small sample sizes of 1) randomly selected galaxies (N = 5, 10), and 2) with an additional condition of \( \langle \delta_F \rangle \leq -0.2 \). The red curve indicate the CCF of all galaxies. Black thin and dotted lines represent each CCF of 100 tests and their mean.

Chen, H.-W., & Mulchaey, J. S. 2009, ApJ, 701, 1219, doi: 10.1088/0004-637X/701/2/1219
Chen, H.-W., Prochaska, J. X., Weiner, B. J., Mulchaey, J. S., & Williger, G. M. 2005, ApJL, 629, L25, doi: 10.1086/444377
Croft, R. A. C., Miralda-Escudé, J., Zheng, Z., Blomqvist, M., & Pieri, M. 2018, MNRAS, 481, 1320, doi: 10.1093/mnras/sty2302
Croft, R. A. C., Miralda-Escudé, J., Zheng, Z., et al. 2016, MNRAS, 457, 3541, doi: 10.1093/mnras/stw204
Daddi, E., Dickinson, M., Morrison, G., et al. 2007, ApJ, 670, 156, doi: 10.1086/521818
Elbaz, D., Daddi, E., Le Borgne, D., et al. 2007, A&A, 468, 33, doi: 10.1051/0004-6361:20077525
Faucher-Giguère, C.-A., Prochaska, J. X., Lidz, A., Hernquist, L., & Zaldarriaga, M. 2008, ApJ, 681, 831, doi: 10.1086/588648
Font-Ribera, A., Arnau, E., Miralda-Escudé, J., et al. 2013, JCAP, 2013, 018, doi: 10.1088/1475-7516/2013/05/018
Fumagalli, M., Prochaska, J. X., Kasen, D., et al. 2011, MNRAS, 418, 1796, doi: 10.1111/j.1365-2966.2011.19599.x
Genel, S., Vogelsberger, M., Springel, V., et al. 2014, MNRAS, 445, 175, doi: 10.1093/mnras/stu1654
Hayashino, T., Inoue, A. K., Kousai, K., et al. 2019, MNRAS, 484, 5868, doi: 10.1093/mnras/stz388
Kakiichi, K., Ellis, R. S., Laporte, N., et al. 2018, MNRAS, 479, 43, doi: 10.1093/mnras/sty1318
Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, ApJS, 224, 24, doi: 10.3847/0067-0049/224/2/24
Lee, K.-G., Hennawi, J. F., Stark, C., et al. 2014, ApJL, 795, L12, doi: 10.1088/2041-8205/795/1/L12
Lee, K.-G., Hennawi, J. F., White, M., et al. 2016, ApJ, 817, 160, doi: 10.3847/0004-637X/817/2/160
Lee, K.-G., Krolewski, A., White, M., et al. 2018, ApJS, 237, 31, doi: 10.3847/1538-4365/aace58
Martizzi, D., Vogelsberger, M., Torrey, P., et al. 2019, arXiv e-prints, arXiv:1907.04333.
https://arxiv.org/abs/1907.04333
