

*Citation for published version:*

Liang, Y, Kashikawa, N, Cai, Z, Fan, X, Prochaska, JX, Shimasaku, K, Tanaka, M, Uchiyama, H, Ito, K, Shimakawa, R, Nagamine, K, Shimizu, I, Onoue, M & Toshikawa, J 2021, 'Statistical Correlation between the Distribution of Ly Emitters and Intergalactic Medium Hi at z2.2 Mapped by the Subaru/Hyper Suprime-Cam', *The Astrophysical Journal*, vol. 907, no. 1, pp. 3. <https://doi.org/10.3847/1538-4357/abcd93>

*DOI:*

[10.3847/1538-4357/abcd93](https://doi.org/10.3847/1538-4357/abcd93)

*Publication date:*

2021

*Document Version*

Peer reviewed version

[Link to publication](#)

Liang, Y, Kashikawa, N, Cai, Z, Fan, X, Prochaska, JX, Shimasaku, K, Tanaka, M, Uchiyama, H, Ito, K, Shimakawa, R, Nagamine, K, Shimizu, I, Onoue, M & Toshikawa, J 2021, 'Statistical Correlation between the Distribution of Ly Emitters and Intergalactic Medium Hi at z2.2 Mapped by the Subaru/Hyper Suprime-Cam', *The Astrophysical Journal*, vol. 907, no. 1, pp. 3. <https://doi.org/10.3847/1538-4357/abcd93>

**University of Bath**

## **Alternative formats**

If you require this document in an alternative format, please contact:  
[openaccess@bath.ac.uk](mailto:openaccess@bath.ac.uk)

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Statistical correlation between the distribution of Ly $\alpha$  emitters and IGM HI at  $z \sim 2.2$   
mapped by Subaru/Hyper Suprime-Cam

YONGMING LIANG,<sup>1,2,3</sup> NOBUNARI KASHIKAWA,<sup>3,2</sup> ZHENG CAI,<sup>4,5</sup> XIAOHUI FAN,<sup>6</sup> J. XAVIER PROCHASKA,<sup>5</sup>  
KAZUHIRO SHIMASAKU,<sup>3</sup> MASAYUKI TANAKA,<sup>1,2</sup> HISAKAZU UCHIYAMA,<sup>2</sup> KEI ITO,<sup>1,2,3</sup> RHYTHM SHIMAKAWA,<sup>2</sup>  
KENTARO NAGAMINE,<sup>7,8,9</sup> IKKOH SHIMIZU,<sup>2,10</sup> MASAFUSA ONOUE,<sup>11</sup> AND JUN TOSHIKAWA<sup>12,13</sup>

<sup>1</sup>*Department of Astronomical Science, SOKENDAI (The Graduate University for Advanced Studies), Mitaka, Tokyo 181-8588, Japan*

<sup>2</sup>*National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan*

<sup>3</sup>*Department of Astronomy, School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan*

<sup>4</sup>*Department of Astronomy, Tsinghua University, Beijing 100084, China*

<sup>5</sup>*UCO/Lick Observatory, University of California, 1156 High Street, Santa Cruz, CA 95064, USA*

<sup>6</sup>*Steward Observatory, University of Arizona, 933 N. Cherry Avenue, Tucson, AZ 85721, USA*

<sup>7</sup>*Department of Earth and Space Science, Osaka University, 1-1 Machikaneyama, Toyonaka, Osaka 560-0043, Japan*

<sup>8</sup>*Kawli IPMU (WPI), The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8583, Japan*

<sup>9</sup>*Department of Physics and Astronomy, University of Nevada, Las Vegas, 4505 S. Maryland Pkwy, Las Vegas, NV 89154-4002, USA*

<sup>10</sup>*Shikoku Gakuin University, 3-2-1 Bunkyocho, Zentsuji, Kagawa, 765-8505, Japan*

<sup>11</sup>*Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany*

<sup>12</sup>*Institute for Cosmic Ray Research, The University of Tokyo, Kashiwa, Chiba 277-8582, Japan*

<sup>13</sup>*Department of Physics, University of Bath, Claverton Down, Bath, BA2 7AY, UK*

(Received April 20, 2020; Revised August 6, 2020; Accepted ZZZ)

## ABSTRACT

The correlation between neutral Hydrogen (HI) in the intergalactic medium (IGM) and galaxies now attracts great interests. We select four fields which include several coherently strong Ly $\alpha$  absorption systems at  $z \sim 2.2$  detected by using background quasars from the whole SDSS/(e)BOSS database. Deep narrow-band and  $g$ -band imaging are performed using the Hyper Suprime-Cam on the Subaru Telescope. We select out 2,642 Ly $\alpha$  emitter (LAE) candidates at  $z = 2.177 \pm 0.023$  down to the Ly $\alpha$  luminosity of  $L_{\text{Ly}\alpha} \approx 2 \times 10^{42} \text{ erg s}^{-1}$  to construct the galaxy overdensity maps, covering an effective area of  $5.39 \text{ deg}^2$ . Combining the sample with the Ly $\alpha$  absorption estimated from 64 (e)BOSS quasar spectra, we find a moderate to strong correlation between the LAE overdensity  $\delta_{\text{LAE}}$  and the effective optical depth  $\tau_{\text{LoS}}$  in line-of-sights, with  $P$ -value = 0.09% ( $< 0.01\%$ ) when the field that contains a significant quasar overdensity is in(ex)cluded. The cross-correlation analysis also clearly suggests that up to  $4 \pm 1 \text{ pMpc}$ , LAEs tend to cluster in the regions rich in HI gas, indicated by the high  $\tau_{\text{LoS}}$ , and avoid the low  $\tau_{\text{LoS}}$  region where the HI gas is deficient. By averaging the  $\tau_{\text{LoS}}$  as a function of the projected distance ( $d$ ) to LAEs, we find a 30% excess signal at  $2\sigma$  level at  $d < 200 \text{ pkpc}$ , indicating the dense HI in circumgalactic medium, and a tentative excess at  $400 < d < 600 \text{ pkpc}$  in IGM regime, corroborating the cross-correlation signal detected at about  $0.5 \text{ pMpc}$ . These statistical analyses indicate that galaxy–IGM HI correlations exist on scales ranging from several hundred pkpc to several pMpc at  $z \sim 2.2$ .

*Keywords:* Galaxy formation — Large-scale structure of the universe — Intergalactic medium — Lyman-alpha galaxies — Two-point correlation function

## 1. INTRODUCTION

The gravitational instability leads mass to assemble in a hierarchical manner from a uniform phase in the early universe, and galaxy formation occurs preferentially along large-scale filamentary and sheet-like over-

dense regions where the neutral Hydrogen (HI) in the intergalactic medium (IGM) is more abundant (Springel et al. 2006; Baugh 2006; Hinshaw et al. 2007). The intersections of such filaments or sheets then evolve into dense clusters of galaxies at a later epoch (Bond et al. 1996; Cen & Ostriker 2000). Therefore, the overdensities at high- $z$  are the crucial laboratories to study the large-scale structure (LSS) formation and evolution, especially the correlation between galaxy and IGM HI.

However, it is not easy to find the overdense regions at  $z > 2$ , which only occupies a small fraction of the cosmic volume, e.g.,  $< 2\%$  for protoclusters (Chiang et al. 2017). To make efficient surveys for galaxies, some studies use galaxies with radio loud active galactic nucleus (AGNs) (Cooke et al. 2014; Shimakawa et al. 2014; Noirot et al. 2018), dusty star forming galaxies (Casey et al. 2015), luminous quasars (Kikuta et al. 2019) or quasar pairs (Onoue et al. 2018) as overdensity tracers. Because such rare objects are expected to reside in massive halos, which are likely to host the protoclusters. In addition, damped Ly $\alpha$  systems (DLAs) (Ogura et al. 2017; Fumagalli et al. 2017) or systems with extended nebular emission around galaxies (Bădescu et al. 2017) are the good candidates as tracers as well. Wide-field surveys also enable blind searches of protoclusters via photo- $z$  galaxies (Spitler et al. 2012) and Lyman-break galaxies (LBGs) (Toshikawa et al. 2016, 2018).

In addition to galaxy surveys, for decades in simulations, the LSSs in terms of IGM HI have also been demonstrated to be possibly revealed by the absorptions imprinting in the spectra of background quasars (Hernquist et al. 1996; Springel et al. 2006), and it is also proved to be a non-trivial question at high- $z$  universe, as most baryons at  $z > 2$  may reside in Ly $\alpha$  clouds (Miralda-Escudé et al. 1996). Strong HI absorbers are studied around quasars (Prochaska et al. 2013) or with searching the associated galaxies (Mackenzie et al. 2019), from which a hint of the galaxy-IGM HI correlation is found. Based on a specific field SSA22 with the protocluster found at  $z = 3.1$ , Mawatari et al. (2017) have found a global correlation on a scale of tens of comoving-Mpc (cMpc) via the narrowband absorption technique. Hayashino et al. (2019), who study the same structure, find the similar correlation in the redshift space.

The galaxy-IGM HI correlation can also be studied in a statistical way with large galaxy surveys for foreground LBG/photo- $z$  galaxy and the background quasar/galaxy pairs (Adelberger et al. 2003, 2005; Steidel et al. 2010; Rudie et al. 2012; Turner et al. 2014; Mukae et al. 2017; Momose et al. 2020b; Chen et al. 2020), and most of these researches find the correlation on various

scales. However, these studies are confined by either the bright galaxy populations, or the relatively small dynamic range of the IGM absorption due to the limited sample size and survey area.

Recently, the IGM tomography also becomes feasible to construct 3D IGM HI maps from the background star-forming galaxies (Lee et al. 2014a,b, 2016, 2018; Newman et al. 2020). But, the tomography surveys to date are still limited by the survey area  $\lesssim 1 \text{ deg}^2$ , and it is mainly conducted on blank fields. A larger survey area covering various overdense regions is essential to take full advantage of the technique.

In the MAPPING the Most Massive Overdensity Through Hydrogen (MAMMOTH) project (Cai et al. 2016, 2017b,a), N-body simulations suggest that coherently strong Ly $\alpha$  absorption system (CoSLAs), originated from the overlapping of the Ly $\alpha$  forest, can effectively trace the most massive halos on the scale over  $15 h^{-1} \text{ cMpc}$ . Although whether CoSLAs traces well the most massive overdensity is now under debate (Miller et al. 2019), a pilot MAMMOTH program has found the BOSS1441, one of the most massive structures to date at  $z > 2$ , with also six BOSS quasars associated (Cai et al. 2017b).

While helping to pinpoint the regions that tend to host overdensities, the grouping rare lines-of-sight (LoSs) with high IGM absorption (e.g., CoSLAs) also enable us to significantly enhance the dynamic range in statistics for studying the galaxy-IGM HI correlation. Targeting the fields centered at MAMMOTH candidates on  $15 h^{-1} \text{ cMpc}$ , the Subaru/Hyper Suprime-Cam (HSC; Miyazaki et al. 2018) equipped with a diameter  $d = 1.5 \text{ deg}$  field-of-view (FoV) makes it possible to efficiently map the most diverse universe at  $z \sim 2$  on a scale over  $100 \text{ cMpc}$ . Additionally, the narrowband technique for identifying Ly $\alpha$  emitters (LAEs) whose redshifts can be well constrained in a narrow range ( $\Delta z < 0.05$ ), also opens a window towards a fainter and less massive galaxy population for the correlation.

In this paper, we first summarize the SDSS/(e)BOSS data, the field selection, the Subaru/HSC observations and the data processing in Section 2. The LAE sample construction are then presented in Section 3. Section 4 shows our LAE overdensity maps for the four HSC fields. The analyses of the galaxy-IGM HI correlation are also shown in this section. Section 5 compares our results with other works, and explores the scale dependence of the correlation. The underlying physics is also discussed in the last part. Finally, we end with a summary and give an outlook of the future work in Section 6. The cosmological parameters used in this paper are based on Planck Collaboration et al. (2016):

$H_0 = 67.7 \text{ km Mpc}^{-1} \text{ s}^{-1}$ ,  $\Omega_0 = 0.307$ .  $AB$  magnitudes are used throughout the paper.

## 2. DATA

### 2.1. SDSS/BOSS Spectral Data

The background quasar spectra from the Baryon Oscillation Spectroscopic Survey (BOSS) of SDSS-III (Dawson et al. 2013) and the later upgraded extended-BOSS, or eBOSS, of SDSS-IV (Dawson et al. 2016) are used in this work for both selecting candidate fields and evaluating the effective optical depth in the correlation analysis. BOSS is a spectroscopic survey specially designed to study the intergalactic science through Ly $\alpha$  forest. It takes spectra with the 2.5-m Sloan telescope for over 150,000 background quasar at  $z \gtrsim 2.15$  reaching a depth as faint as  $g < 22$ . The eBOSS observes 60,000 BOSS quasars for spectra in better quality and 60,000 new targets in complement. The surveys combined offer more than 200,000 quasar spectra covering a survey area of over 10,000 deg $^2$ , corresponding to a survey volume of  $> 1 \text{ Gpc}^3$ .

The (e)BOSS database offers us abundant quasar spectra working as LoSs, in which the IGM distribution can be traced by the Ly $\alpha$  absorption. To evaluate the Ly $\alpha$  absorption, we calculate the effective optical depth in the LoS,  $\tau_{\text{LoS}}$ , within the Ly $\alpha$  redshift range traced by the narrowband filter NB387 ( $\lambda_0 = 3,862 \text{ \AA}$ , FWHM =  $56 \text{ \AA}$ ).

We first smooth the flux along the wavelength dimension over a scale of  $15 h^{-1} \text{ cMpc}$ . Absorption features are searched by scanning through the spectra over a range of  $\pm 35 \text{ \AA}$  centered around  $3,862 \text{ \AA}$ . The effective optical depth is then calculated at the strongest absorption spike following Cai et al. (2016):

$$\tau_{\text{LoS}} = -\ln \langle F \rangle_{15 h^{-1} \text{ cMpc}}, \quad (1)$$

where the  $\langle F \rangle_{15 h^{-1} \text{ cMpc}}$  is the continuum normalized flux estimated on the  $15 h^{-1} \text{ cMpc}$  scale. Note that the  $\tau_{\text{LoS}}$  estimated here can be systematically larger than the cosmic mean, as we are targeting at IGM Ly $\alpha$  absorbers as the gas tracers, instead of the random forest.

When evaluating the Ly $\alpha$  absorption by using the (e)BOSS spectra, quasar continuum is constructed using the mean-flux-regulated principal component analysis (MF-PCA) technique to the fitting (Lee et al. 2013). The extra constraints on the slope and amplitude of the continuum are adjusted by using the mean optical depth of the Ly $\alpha$  forest (Lee et al. 2012; Becker et al. 2013).

The  $\tau_{\text{LoS}}$  will be used throughout the paper for both the field selection and the galaxy-Ly $\alpha$  absorption correlation analysis.

### 2.2. Field Selection

Our goal in this paper is to study the galaxy-IGM HI correlation on a wide range of the environments based on the less massive galaxy populations. The principle for our field selection is to enclose a sufficient number of LoSs, especially those with strong Ly $\alpha$  absorptions, while we also target the possible overdensities.

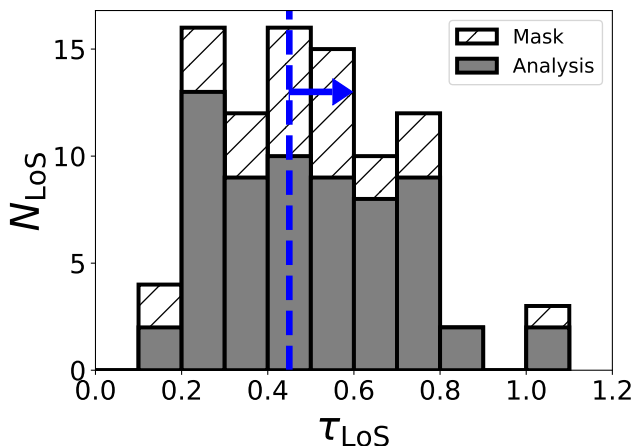
First, we briefly summarize the selection of COSLAs, but please see the details in Cai et al. (2016). The LoSs with  $\tau_{\text{LoS}} \gtrsim 3 \langle \tau \rangle_{\text{cos}}$  is chosen as the preliminary absorber candidates, where the  $\langle \tau \rangle_{\text{cos}}$  is the cosmic mean optical depth, and we assume it as 0.15 at  $z = 2.2$  (Becker et al. 2013) with slight adjustments according to fields. To eliminate the non-IGM contaminants, we make the systematic inspections of the criteria proposed in Cai et al. (2016) to reject the high column density systems (HCDs), i.e., DLAs, sub-DLAs or Lyman-limit systems (LLS), and we also do the visual checks to remove the broad absorption line (BAL) quasars that may confuse the interpretation of IGM Ly $\alpha$  absorption in the NB387 wavelength range. Besides for the high  $\tau_{\text{LoS}}$  LoSs, all of these processes for excluding the non-IGM contaminants are also done for the potential LoSs used in our following analysis, which we call as the clean LoSs here.

Therefore, based on the clean LoSs, we own several preferences when selecting the target fields: the target fields of HSC-FoV should: (1) contain the high  $\tau_{\text{LoS}}$  LoSs to expand the dynamic range; (2) enclose as many LoSs as possible to increase sample size for drawing the galaxy-IGM HI relation; (3) contain a concentration of high  $\tau_{\text{LoS}}$  LoSs to find a protocluster, i.e.,  $\gtrsim 4$  LoSs within a ( $\sim 20 h^{-1} \text{ cMpc}$ ) $^3$  box, which is the typical scale of a Coma-type protocluster at the  $z \sim 2$  (Chiang et al. 2013); (4) in special case, contain the associated quasars at  $2.15 \leq z \leq 2.20$ , i.e., proximity quasars at the similar redshift of our LAEs, to see any possible difference.

Our field selection were, however, further compromised by the field visibility, the distance to the moon or the nearby bright stars in a specific observation run. As a result, four fields BOSSJ0210 + 0052 (or J0210), BOSSJ0222 - 0224 (or J0222), BOSSJ0924 + 1503 (or J0924) and BOSSJ1419+0500 (or J1419) are selected in our observations, and all of them satisfy (1) and mostly (2). J0222 and J0924 are selected mainly based on (3), the typical regions hinting the coherent IGM HI on large scale. J1419 was once selected also for (3), but one of the two concentrating LoSs, which is found to be a possible

BAL quasar<sup>1</sup> after the observation, is excluded from our analysis. However, the field is still one of the best candidates considering (1) and (2), although the coherent IGM Ly $\alpha$  absorption is not as significant as other selected fields here. Specially, *J0210* is selected with the consideration of (4), given that a group of 11 proximity quasars is associated within a region of  $40 \times 40$  cMpc<sup>2</sup> at  $2.15 < z < 2.20$ , a length of 62 cMpc along the LoS direction, which is more extreme than the BOSS1441 found in Cai et al. (2017b). One of the proximity quasars also shows the hint of strong IGM Ly $\alpha$  absorption at the wing of Ly $\alpha$  emission, but being conservative, we do not include it in our correlation analysis. The coordinates of the field centers are listed in the Table 1.

We note that before applying mask in the following sections, there are 26, 23, 19 and 22 clean LoSs in *J0210*, *J0222*, *J0924* and *J1419* respectively, and they are summarized in Table 2. The  $\tau_{\text{LoS}}$  distribution of these clean LoSs is shown in Figure 1, in which a vertical blue dash line indicates the criterion for the clean LoSs with  $\tau_{\text{LoS}} \gtrsim 3 \langle \tau \rangle_{\text{cos}}$ .



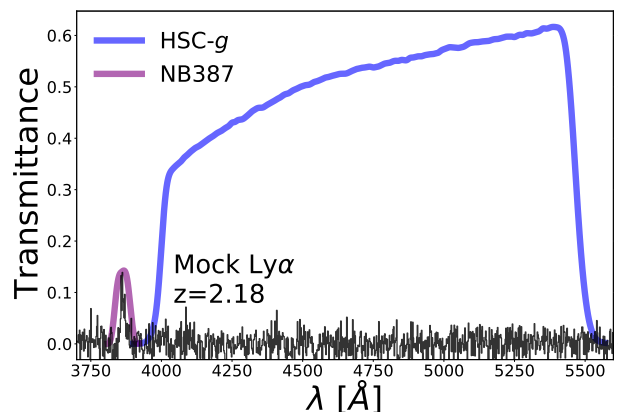
**Figure 1.** The  $\tau_{\text{LoS}}$  distribution of all inspected clean LoSs. The hatched area indicates the masked LoSs, while the filled area shows the LoSs used in the following correlation analysis. The blue vertical dash line suggests the  $\tau_{\text{LoS}}$  criterion of the clean LoSs with  $\tau_{\text{LoS}} \gtrsim 3 \langle \tau \rangle_{\text{cos}}$ .

### 2.3. Imaging Observations

Observations to identify LAEs were carried out with the HSC installed at the prime focus of the 8.2-m Subaru telescope located at the summit of the Mauna Kea, Hawaii. HSC is a high performance camera with a wide FoV of 1.5 deg in diameter. As a gigantic mosaic

CCD camera, HSC consists of 104 Hamamatsu Photonics KK CCDs ( $2048 \times 4096$  pixels) for science, 4 for auto-guiders and 8 for focusing. The pixel scale of the CCD reaches  $0.168''$ .

In this paper, we perform the deep NB imaging using the NB387, which enables us to detect Ly $\alpha$  emission at the corresponding redshift of  $z = 2.177 \pm 0.023$ . The *g*-band is also used for the evaluation of the continuum level of the detected objects. The transmission curves of the filters, which has taken the transmittance accounting in CCD quantum efficiency, dewar window, the Primary Focus Unit and the reflectivity of the Prime Mirror into account, are shown in Figure 2. To ensure the depth for



**Figure 2.** Transmission curve of HSC-*g* and NB387 band. The purple and blue solid lines are the total transmittance of NB387 and HSC-*g* counting the CCD quantum efficiency, dewar window, the Primary Focus Unit and the reflectivity of the Prime Mirror. The black curve indicates a mock LAE spectrum at  $z = 2.18$ , whose Ly $\alpha$  emission is exactly located at the sensitive wavelength range of the NB387.

detecting a sufficient number of LAEs, the observation is designed to have total exposures of 3 hours for NB387 and 40 min for *g*-band in each field. An S17B observation was carried out in a queue mode in Jan. 2018 and Jan. 2019, and exposures are split into 900 s and 600 s for the NB387 and *g*-band respectively, except for the first 2 exposures of 1,200 s for *J0210*. In the S19A observation carried out on-site on Mar. 8<sup>th</sup>, 2019, the exposures are split into 900s and 300s for NB387 and *g*-band respectively to avoid saturations in the broadband. From S17B, we have obtained both the NB387 and *g*-band data for fields *J0210* and *J0222* and the *g*-band data for *J0924*. In the S19A run, NB387 data for both *J0924* and *J1419* are achieved, and the *g*-band observation is taken for *J1419* only. In summary, both NB387 and *g*-band imaging data for all four fields are from two major runs.

<sup>1</sup> This is J141934.64+050327.1, which is categorized as a *probable* Pv BAL quasar in Capellupo et al. (2017)

Field	RA (J2000)	DEC (J2000)	Obs Period	FWHM <sub>PSF,NB</sub>	FWHM <sub>PSF,g</sub>	$m_{\text{NB},5\sigma}$	$m_{g,5\sigma}$	<i>hscPipe</i>
BOSS J0210+0052	02:09:58.90	+00:53:43.0	Jan., 2018	1.22''	0.90''	24.36 <sup>a</sup>	26.24 <sup>a</sup>	v5.4
				1.22''	0.90''	24.25 <sup>a</sup>	26.34 <sup>a</sup>	v6.6
BOSS J0222-0224	02:22:24.66	-02:23:41.2	Jan., 2018	0.90''	0.90''	24.99	27.01	v5.4
BOSS J0924+1503	09:24:00.70	+15:04:16.7	Jan. & Mar., 2019	0.84''	0.79''	24.74	26.63	v6.6
BOSS J1419+0500	14:19:33.80	+05:00:17.2	Mar., 2019	0.86''	0.70''	24.81	26.80	v6.6

<sup>a</sup>Measured in 2.5'' aperture, and  $g$ -band is the PSF-matched image.

**Table 1.** Summary of field information. Column 1 is the full name of fields; Columns 2 and 3 are the coordinates RA and DEC in equinox with an epoch of J2000; Column 4 is the period of executing the observations; Columns 5 and 6 are the FWHMs of star PSFs measured for the final stacked images of NB387 and  $g$ -band; Columns 7 and 8 are the  $5\sigma$  limiting magnitudes measured in an aperture with the radius of 1.7'' for the final stacked NB387 image and PSF-matched  $g$ -band, respectively; Column 9 is the *hscPipe* version used for the data reduction.

Field	N <sub>LoS,All</sub>	N <sub>LoS,Ana</sub>	N <sub>LAE</sub>	Area [deg <sup>2</sup> ]	E(B-V)
BOSS J0210+0052	26	22	465	1.34	0.0246
BOSS J0222-0224	23	11	956	1.13	0.0222
BOSS J0924+1503	19	14	585	1.47	0.0217
BOSS J1419+0500	22	17	636	1.45	0.0264
Total	90	64	2642	5.39	\

**Table 2.** Information of LoSs and LAEs in each field. Column 1 is the respective field; Column 2 is the number of all the clean LoSs inspected in/around the four fields; Column 3 is the number of the clean LoSs after being masked, which are used in the correlation analysis in this work; Column 4 is the number of LAE candidates; Column 5 is the effective survey area for selecting LAEs after being masked; Column 6 is the galactic reddening accounting the Milky Way based on the measurement and calibration from Schlafly & Finkbeiner (2011).

Except for the NB imaging of *J0210* in relatively poor conditions with seeing over 1.2'', all of the observations were executed under moderate to good conditions. Some exposures are discarded because of the occasionally poor seeing or low transparency. In *J0222*, the severe stray light from a nearby Mira contaminates some exposures and they are also discarded. Standard stars are not used considering the large FoV and 104 CCDs. Instead, we use Pan-STARS DR1 (PS1) photometric data (Chambers et al. 2016) for calibration as described in the Sec. 2.4. Detailed information of each field is summarized in the Table 1.

#### 2.4. Data Reduction

The NB387 and  $g$ -band imaging data are reduced with the HSC pipeline, *hscPipe* (Bosch et al. 2018; Aihara et al. 2019). *J0210* and *J0222* are reduced with the *hscPipe* 5.4, and *J0924* and *J1419* are reduced with the *hscPipe* 6.6. Given the relatively poor quality of the *J0210* NB387 data, we also reduce both the  $g$ -band and NB387 data for *J0210* with *hscPipe* 6.6 and combine the catalog with the one produced from *hscPipe* 5.4. The overlaid detections with separations smaller than 2'' are only kept for the latter version. Bosch et al. (2018) and Aihara et al. (2019) describe the data reduction process as well as the code updates in details, but we give a brief

summary here with an emphasis on the processes that are different from the standard usage.

The *hscPipe* first makes calibration data, including the bias, dark, dome-flat and global sky. Then it applies them to each CCD in single visit, and a local sky background on 128 pixel scale is subtracted. Bright objects are then extracted for astrometric and photometric calibration. Point-spread function (PSF) models used inside the pipeline are also made at this step. Astrometry and photometry are then calibrated against PS1 references. For each filter, the zero-point is adjusted by fitting a multi-band relation, e.g., a NB -  $g$  vs.  $g - r$  relation for the NB387:

$$\begin{aligned} \text{NB387}_{\text{HSC}} - g_{\text{PS1}} = & 0.541 \times (g_{\text{PS1}} - r_{\text{PS1}})^2 \\ & + 1.87 \times (g_{\text{PS1}} - r_{\text{PS1}}) \\ & + 0.428 [+C_{\text{metal}} + C_{\text{fit}}], \end{aligned} \quad (2)$$

which is derived from the template magnitudes predicted by spectroscopic Pickles star references (Pickles 1998) and the filter transmissions. A tract ID is defined to enclose all observed sky, and then a sky map is made as the reference for the following coadding process. A global sky background was subtracted without masked regions. In the mosaicking of the CCDs data, both the WCS and the flux scale were corrected by a spatially-

varying correction term. Finally, the coadding process warp the images to the sky map and co-add all visits of the image together scaled with the WCS and flux correction from the mosaicking process.

For our data, some configurations need to be further optimized in addition to the aforementioned process. The NB387 image is always too shallow in a single frame image to have enough bright stars in each CCD for calibration. Therefore, we set the parameter set for choosing calibration stars lower by  $\sim 0.5 \times$  default value. In addition, when fitting the Equation 2 to determine the photometric zero-point of NB387, we take into accounts additional corrections, including a systematic correction  $C_{\text{metal}}$  of 0.448 mag for correcting the stellar metallicity bias, and a field dependent term  $C_{\text{fit}}$  ranging within 0.2 mag for calibrating the fitting uncertainties. Details are described in Appendix A.

### 2.5. Photometric Processing

We use the *SExtractor* 2.19.5 (Bertin & Arnouts 1996) for the photometry processing. First, we do the PSF matching for the  $g$ -band and NB387 images by convolving a proper Gaussian kernel in each field. Then we run the dual-image mode for the source detection and measurement by setting the NB387 image as the reference. The detection threshold is set as 15 continuous pixels over the  $1.2\sigma$  sky background. Because of the large HSC FoV and the mosaic CCD structure, there are slight fluctuations of 0.1 - 0.2 mag of the image depth among the whole field. We apply the sky background root-mean-square (RMS) map as the weighting map in *SExtractor* to minimize this influence. In addition, we use a local background with the thickness of 128 pixels. Masks are also applied when doing the background estimation, object detection and measurement. The masks are defined as regions with low S/N signals, saturation around bright stars or severe stray lights.

Note that after applying the masks, the final numbers of clean LoSs in *J0210*, *J0222*, *J0924* and *J1419* are 22, 11, 14 and 17 respectively, as summarized in the Table 2. The masked clean LoSs are hatched in Figure 1, and the remaining 64 clean LoSs will be used for all of the following correlation analysis, unless some of them are further removed with the nearby masked regions over a certain fraction, as described in Section 4.3 and 5.2.

We use aperture magnitudes for the color selection, and the aperture diameters are 15 pixels ( $\sim 2.5''$ ) for *J0210* and 10 pixels ( $\sim 1.7''$ ) for *J0222*, *J0924* and *J1419*. The Auto-Mag is used for the estimate of total magnitude, which applies automatically determined elliptical aperture for Kron photometry in *SExtractor*. Galactic extinction is also accounted in each band. Re-

ferring to the Galactic Dust Reddening and Extinction Service provided by IRSA, which is based on the results of Schlafly & Finkbeiner (2011), color reddening  $E(B - V)$  can be estimated and is listed in Table 2. As the  $R_{\text{NB387}} = A_{\text{NB387}}/E(B - V)$  is estimated as 4.009, taking into account of the transmission curve, we manage to apply the dust extinction correction for the detection catalogs. We replace the  $g$ -band magnitude with the corresponding  $2\sigma$  limiting magnitude, when the objects are fainter than the  $2\sigma$  limit.

The measured PSF FWHM and the  $5\sigma$  limiting magnitudes in the  $1.7''$  aperture ( $2.5''$  for *J0210*) of the final stacked images of NB387 and  $g$ -band are listed in the Table 1. The quality of *J0210* data is relatively poorer compared to the other three fields, in both of the seeing and the final image depth.

## 3. SAMPLE SELECTION

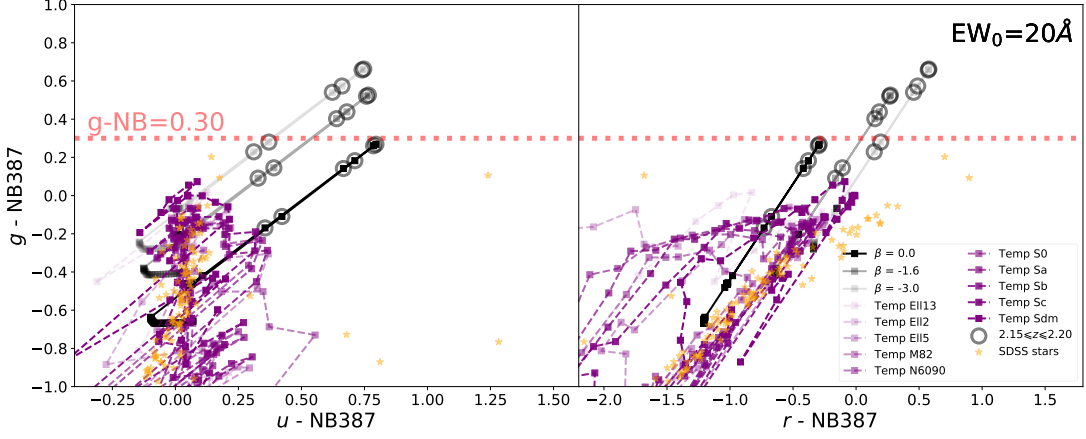
### 3.1. Ly $\alpha$ Emitters Selection

We use the color excess of the narrowband to the broadband as our LAE selection criteria, which is widely used in previous works (Guaita et al. 2010; Mawatari et al. 2012; Nakajima et al. 2012; Konno et al. 2016; Zheng et al. 2016). Though we have only the broadband data from the  $g$ -band on HSC for estimating the continuum, we prove here it is sufficient enough for the  $z = 2.18$  LAE selection.

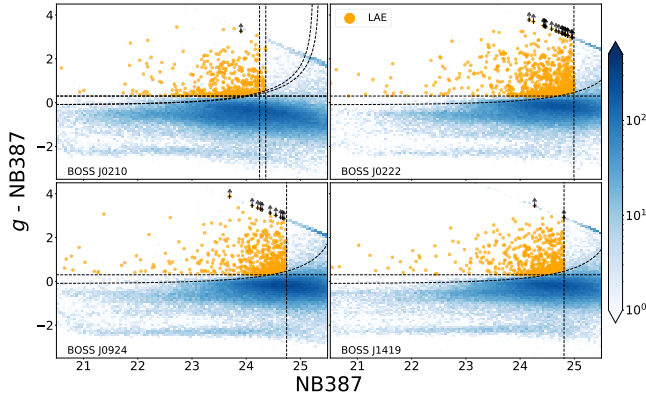
In order to define the selection criteria, we assume the LAE spectrum model at  $z = 2.0 - 2.5$  has a simple power law  $f_{\lambda} = \lambda^{\beta}$  continuum and a correspondingly redshifted Ly $\alpha$  emission with a Gaussian-like profile, whose rest equivalent width  $EW_0 = 20 \text{ \AA}$ . The IGM absorption is taken into account when we calculate the observed magnitude in each filter (Inoue et al. 2014). In addition to the  $g$ -band on Subaru/HSC, we include the adjacent broadband filters, the  $u$ -band on CFHT/MegaCam and the  $r$ -band on Subaru/HSC, to see the redshift evolution on the two color diagrams. The tracks are shown in the Figure 3. The black curves indicate the color tracks of  $g$ -NB387 vs.  $u$ -NB387 in the left panel, and  $g$ -NB387 vs.  $r$ -NB387 in the right one. Three different UV slopes  $\beta$ s, 0, -1.6 and -3.0, are shown in the both figures.

Meanwhile, we also overplot the predicted tracks of the possible contaminants, such as elliptical galaxies (age of 2, 5 and 13 Gyr denoted as Ell2, Ell5 and Ell13), starburst galaxies (M82 and N6090) and spiral galaxies (S0, Sa, Sb, Sc, Sd and Sdm) from the SWIRE library (Polletta et al. 2007) from redshift 0 to 3.0. The homogeneously archived faint stars from SDSS used in Section 2.4 are also plotted.

From the color tracks, we find that the  $r$ -band is hardly helpful for the LAE selection, while the  $u$ -band



**Figure 3.** The color tracks with redshift evolution for the  $EW_0 = 20 \text{ \AA}$  LAE at the  $z = 2 - 2.5$ :  $g - \text{NB387}$  vs.  $u - \text{NB387}$  (left panel) and  $g - \text{NB387}$  vs.  $r - \text{NB387}$  (right panel). The black curves are the tracks for LAE models (with UV slope  $\beta = 0, -1.6, -3.0$ ) with a redshift step  $\Delta z = 0.01$  from  $z = 2$  to  $z = 2.5$ , and the purple dash curves are for galaxy templates, including elliptical galaxies (age of 2, 5 and 13 Gyr denoted as Ell2, Ell5 and Ell13), starburst galaxies (M82 and N6090) and spiral galaxies (S0, Sa, Sb, Sc, Sd and Sdm) (Polletta et al. 2007) with a redshift step  $\Delta z = 0.1$  from  $z = 0$  to  $z = 3$ . The homogeneously picked SDSS stars with  $g > 19$  (Yanny et al. 2009) are also plotted as yellow stars. Circles indicate the LAE models at  $2.15 \leq z \leq 2.20$ . The narrowband excess  $g - \text{NB387} > 0.30$  works as a reasonable threshold to select out the  $z \sim 2.18$  LAEs.



**Figure 4.**  $g - \text{NB387}$  vs.  $\text{NB387}$  diagram for LAE selection in each field. The two dimensional histogram bins the all detections from *SExtractor* and the bluer bins contain more objects. The yellow points are the selected LAE candidates after visual inspection. The three selection criteria are shown as the black dotted lines. Specially, for the field *J0210*, the data is reduced in two versions of the *hscPipe* and there is a slight difference of the final image depth, so the criteria are overplotted for clarification. The black arrows indicate the LAE candidates with the  $g$ -band fainter than the respective  $2\sigma$  limiting magnitude of each image, and the  $g - \text{NB387}$  shown for these objects are the lower limits.

may help to recover the extremely red ( $\beta \sim 0$ ) populations. However, given that the typical UV slope of the  $z \sim 2$  LAE is found to be  $\beta \sim -1.6$  (Kusakabe et al. 2019; Santos et al. 2020), we conclude that only the  $g$ -band is sufficient enough for our  $z \sim 2.18$  LAE selection and a reasonable threshold of the color excess is  $g - \text{NB387} > 0.3$  to exclude most of the contaminants.

To make the selection more confident in photometry, the color criteria are further defined as

$$\begin{aligned} 20.5 < \text{NB387} &\lesssim \text{NB}_{\text{lim}, 5\sigma}, \\ g - \text{NB387} &> 0.3, \\ g - \text{NB387} &> 2\sigma(\text{NB387}) - 0.1. \end{aligned} \quad (3)$$

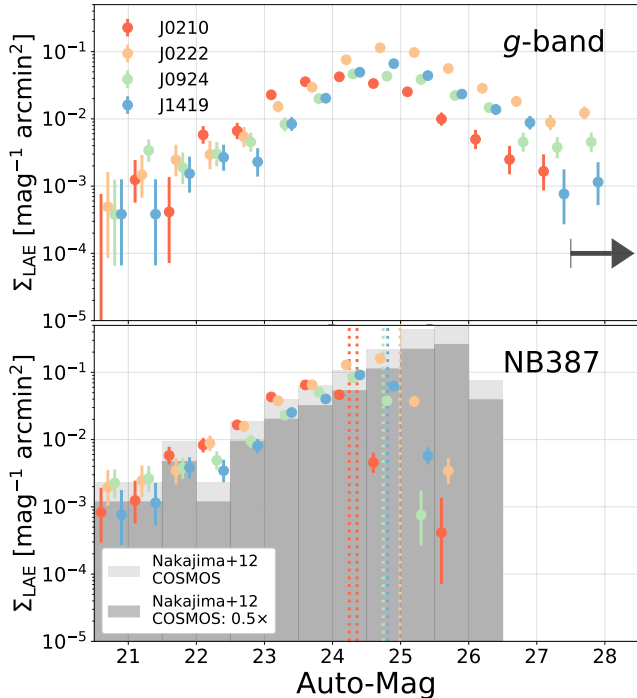
The lower limit of the NB387 mag, 20.5, is set to avoid saturations, while the upper limit applies the  $5\sigma$  limiting magnitude to promise the reliability of NB387 detection. This upper limit for the field *J0924* with a moderate depth, 24.74, corresponds to  $1.94 \times 10^{42} \text{ erg s}^{-1}$ , which is  $0.37 \times L_{\text{Ly}\alpha}^*$  and the characteristic luminosity  $L_{\text{Ly}\alpha}^* = 5.3 \times 10^{42} \text{ erg s}^{-1}$  (Konno et al. 2016). The definition of the color error follows Shibuya et al. (2018):

$$2\sigma(\text{NB387}) = -2.5 \log_{10} \left( 1 - 2 \times \frac{\sqrt{f_{1\sigma, \text{NB387}}^2 + f_{1\sigma, g}^2}}{f_{\text{NB387}}} \right). \quad (4)$$

where the  $2\sigma$  follows a proper choice used in Nakajima et al. (2012). It aims to reject the false selection of the faint objects that pass the criteria due to statistical fluctuation around the  $g - \text{NB387} = -0.1$ , where the high- $z$  galaxy sequence lies on as described in Appendix A.

The selected object passing the criteria are naturally filtered by the spatial masks, as the original object detection is performed with the masks applied. Finally, we perform the visual check for each candidate to reject fake detections, like the hot pixels in the NB387 or the saturated pixels in the  $g$ -band image. We also check





**Figure 5.** The Auto-Mag surface density distributions of the LAE candidates. Red, yellow, green and blue histograms represent for the LAEs in *J0210*, *J0222*, *J0924* and *J1419* respectively. Errorbars indicate the Poisson errors. *Upper*: *g*-band magnitudes. The black arrow indicates that the right-most bins include the faintest objects whose magnitudes are larger than the  $2\sigma$  limiting mags. *Lower*: NB387 magnitudes. The  $5\sigma$  limiting aperture mags are indicated by the vertical dotted lines in respective colors. The  $z \sim 2.2$  LAEs in COSMOS selected by the Subaru/Suprime-Cam NB387 (Nakajima et al. 2012), are also plotted for comparison as the light shade histogram. We also scale the number by a factor of 0.5 to roughly match the survey volume, and show it as the deep shade histogram.

the cross-matches between our selected objects and the SDSS/(e)BOSS quasars at  $z < 2.15$  to discard the low- $z$  contaminants. Eight are found in *J0210*, while six in *J0222* and none in *J0924* or *J1419*. These known low- $z$  quasars are removed from our LAE sample. As a result, there are 465, 956, 585 and 636 LAE candidates selected out in the field *J0210*, *J0222*, *J0924* and *J1419* respectively, i.e., 2642 in total for the all four fields covering an effective area of  $5.39 \text{ deg}^2$ . There are 4, 3, 0 and 1 proximity quasars from the SDSS/(e)BOSS matched to these LAE candidates in each field. Specially, in the *J0210* central region where 11 quasars reside in, three of the quasars are selected out as LAEs in our sample, while most others are too bright in the NB387 images and break the selection criterion  $\text{NB387} > 20.5$ .

As shown in Figure 4, the selected LAE candidates in the final catalog are plotted as the yellow points in

the  $g - \text{NB387}$  vs. NB387 diagram, in which the all detections are binned in the two dimensional histogram coded by the blue color<sup>2</sup>.

We show the  $g$ -band and NB387 magnitude distributions in surface density of the LAE candidates in Figure 5, with the Poisson errors estimated by the statistics proposed in Gehrels (1986). In both filters, the *J0210* and *J0222* are found with the excess at around 23–24 mags. The Auto-Mags are shown here for the fair comparison of total magnitudes with literature, and we overplot the  $z \sim 2.2$  LAE sample in Cosmic Evolution Survey (COSMOS) field from Nakajima et al. (2012) that are selected by the Subaru/Suprime-Cam NB387 ( $\lambda_0 = 3870 \text{ \AA}$ ,  $\text{FWHM} = 94 \text{ \AA}$ ). As their FWHM is almost twice to the HSC NB387, corresponding to a roughly double survey volume, we also show the case scaled with a factor of 0.5. The distribution shapes are almost consistent, but all of our four fields show out number excesses up around the limiting depth compared to the scaled numbers in COSMOS, although the excesses in *J0924* and *J1419* are not as significant as *J0210* and *J0222*. The excess is not surprising as our fields are selected to contain potential overdensities.

Comparing with other galaxy–IGM correlation studies, we note that while LAEs are expected to be younger and less massive than the more mature LBGs in Keck Baryonic Structure Survey (KBSS, Rudie et al. 2012; Chen et al. 2020) and  $K_s$ -selected photo- $z$  galaxies (Mukae et al. 2017), our samples also reach deeper regarding the UV continuum given the depth limit of  $R \sim 25.5$  in KBSS and  $g \sim 26.4$  in Mukae et al. (2017).

### 3.2. Potential Contaminants

Besides the LAEs at  $z \sim 2.2$ , some of the lower- $z$  emitters may also pass our selection criteria. For the filter NB387, the contaminants are mainly considered as [OII] emitters at  $z = 0.036 \pm 0.008$ . But, the survey volume at such redshift range is much smaller than that at the  $z \sim 2.2$ , and the ratio reaches 0.2%. Given the low- $z$  [OII] emitters luminosity function from Ciardullo et al. (2013) and our NB387 image depth, we can estimate that the detected number is  $\sim 0.05$  in one HSC FoV. We conclude that the contamination rate of low- $z$  [OII] emitters is negligible in our sample. In addition, CIII]  $\lambda 1909$  at  $z \sim 1$  and CIV  $\lambda 1548$  at  $z \sim 1.5$  can be also the interlopers. However, according to Konno et al. (2016), these emitters should be probable AGNs, as the

<sup>2</sup> A sequence appearing around  $g - \text{NB387} \sim -2.5$  is likely the stellar locus consisting of K and M-type stars, as suggested by the stellar locus in  $ugr$  diagram (Smolčić et al. 2004), and this is also supported by our random checks in matched SDSS spectra.

objects passing our selection criteria yield the  $EW_0 \gtrsim 30 \text{ \AA}$ , which is much larger than that in the typical star-forming galaxies.

In the literature working on fields like SXDS, COSMOS, HDFN, SSA22 and E-CDFS fields (Guaita et al. 2010; Zheng et al. 2016; Konno et al. 2016), they use the detection in the database covering multi-wavelengths, e.g., the X-ray, UV and radio, to exclude the low- $z$  AGN contaminants. In our case, however, we search for overdense fields from the whole (e)BOSS survey, and therefore, the deep multi-wavelength data is not available for testing the AGNs in this work. Instead, we refer to the literature aforementioned and find that the contamination rate of the LAE selections at  $z \sim 2.2$  is commonly  $\sim 10 - 15\%$ , and Sobral et al. (2017) also confirm this number spectroscopically.

We test this contamination estimate for the case of HSC/NB387 by utilizing the COSMOS data, the NB387 data from the Cosmic Hydrogen Reionization Unveiled with Subaru (CHORUS; Inoue et al. 2020, submitted) survey and the DEIMOS 10K spectroscopic survey catalog (Hasinger et al. 2018). It yields a contamination rate of  $\sim 15\%$  in our LAE selection, and about 2/3 of the interlopers are likely to be the CIV emitters at  $z \sim 1.5$ , which shows a good consistency to what have been stated in the previous studies. As this contamination level is not crucial to our statistical study, we keep all the selected LAE candidates in our overdensity maps as well as the correlation analysis performed in the following sections.

## 4. RESULTS

### 4.1. LAE Overdensity Map

The sky distribution of the selected LAE candidates is shown in the Figure 6. We calculate the galaxy overdensity over each field to construct the overdensity maps. The overdensity is defined as

$$\delta_{\text{LAE}} = \frac{N_{i, \text{LAE}} - \langle N_{\text{LAE}} \rangle}{\langle N_{\text{LAE}} \rangle}, \quad (5)$$

where the  $N_{i, \text{LAE}}$  is the number of galaxies counted within an aperture with the fixed radius, and the  $\langle N_{\text{LAE}} \rangle$  is the mean number of galaxies in an aperture averaged over each field respectively.

The aperture size is set as 10 cMpc ( $\sim 6'$ ) in radius, which yields a mean number of LAEs counted in an aperture  $> 10$ , and so we can have a mean S/N  $> 3$  signal assuming the Poisson statistics for counting. The map is constructed through a  $128 \times 128$  meshgrid for each field, which corresponds to a resolution of  $\sim 1$  cMpc. When calculating the mean number density, we exclude the apertures covering the masked regions for more than

10%, a strict criterion to keep the mean number estimate robust. While drawing the overdensity map, we exclude the apertures that are masked out over 50%, a relaxed criterion to show more accepted regions. The mean number  $\langle N_{\text{LAE}} \rangle$  (standard deviation  $\sigma_{\text{LAE}}$ ) within a radius  $r = 10$  cMpc aperture for  $J0210$ ,  $J0222$ ,  $J0924$  and 1419 are 10.7 (6.0), 23.5 (10.1), 12.6 (4.9) and 14.0 (5.4) respectively. The smaller mean number in  $J0210$  and larger number in  $J0222$  are mainly originated from the image depth difference.

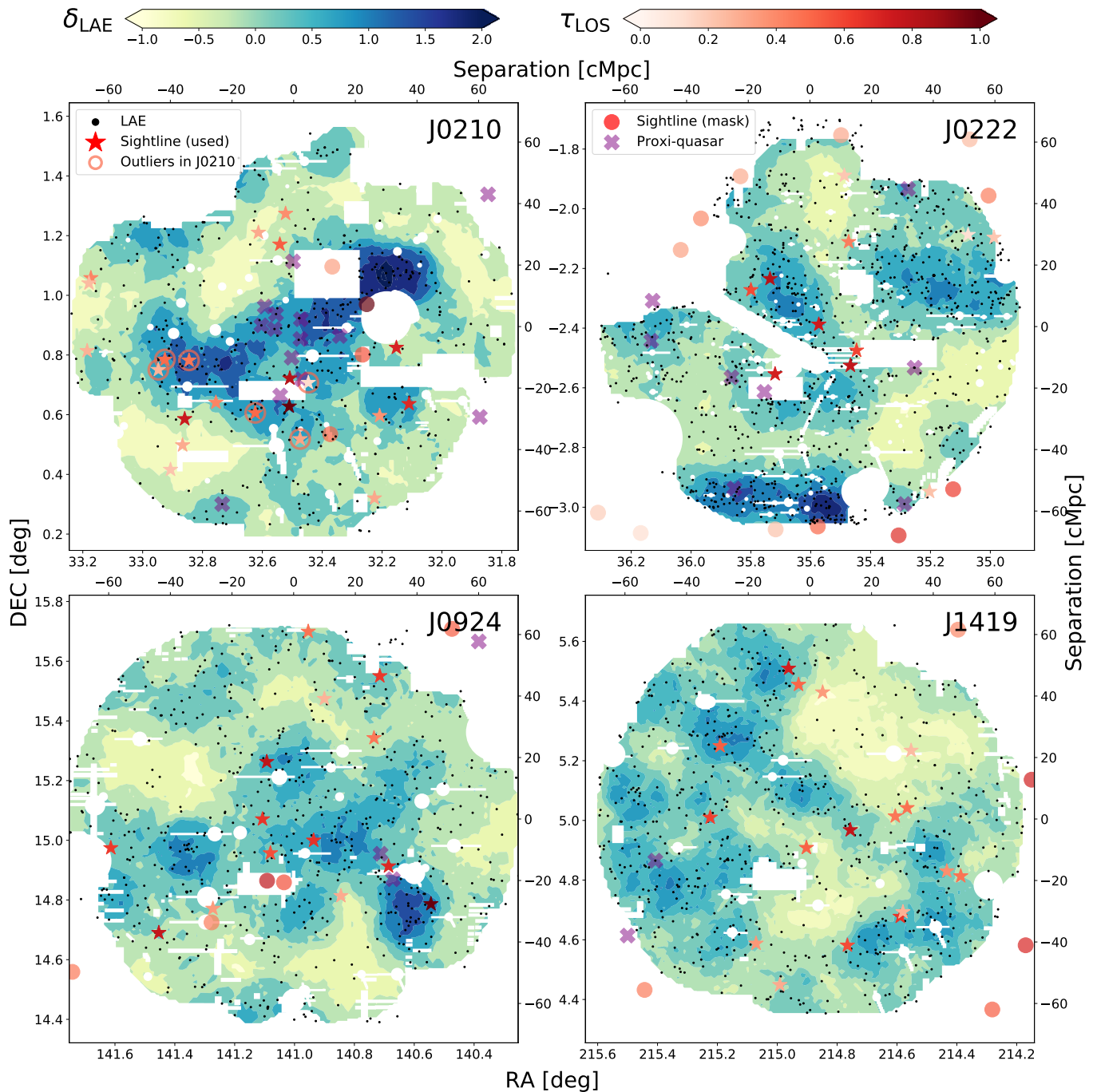
In Figure 6, the blue contours in the background show the overdensity. Masked regions that are defined in Section 2.5 are also shown as the white areas. The LAEs and the proximity quasars with  $2.15 < z < 2.20$  checked when selecting candidate fields are both shown for each field. The position of LoSs are also marked as the red stars with the color coded by the effective optical depth  $\tau_{\text{LoS}}$ .

### 4.2. Notes on Individual Fields

More quantitative discussions on the overdensity catalog will be presented in Cai et al. (in prep.), and we just have a brief overlook here.

In the Figure 6, we find a large filamentary structure at the center of the field  $J0210$  as well as the structures with weaker significance, which are likely to be the sheet-like structures, around the nodes at the ends. The field is traced by both the central grouping LoSs with strong Ly $\alpha$  absorption and a group of quasars clustering within an area of  $\sim (40 \text{ cMpc})^2$  at  $2.15 < z < 2.20$ . The filamentary structure they are associated with extends for about 100 cMpc, and the peak density of one node with  $\delta_{\text{LAE}} \sim 3$  reaches the significance of over  $6\sigma$ . This result supports that the combination of using both tracers seems to effectively hint the unique LSS, as also suggested in Cai et al. (2017b). Given what have reported in the previous studies on the correlation between multiple quasar environments and the galaxy overdensity (Hennawi et al. 2015; Cai et al. 2017a; Onoue et al. 2018; Mukae et al. 2019), the emergence of the grouping quasars suggests the filament is much different from the typical environments at  $z \sim 2$ , and  $J0210$  will be considered individually in the following parts. The uniqueness of the structure in  $J0210$  is out of the scope of this paper, and further discussion will be made in our future paper.

As to the  $J0222$ , one can find that this field is seriously affected by bright stars in and around the FoV, which results in large masked areas with strange patterns. A weak clump with an overdensity  $\delta_{\text{LAE}} \sim 1.0$  over a 20 cMpc length scale is found close to the central region, likely associated with the central group of



**Figure 6.** Overdensity maps for the four selected fields  $J0210$ ,  $J0222$ ,  $J0924$  and  $J1419$ . Black points represent the LAE candidates. The blue contour in the background is scaled by the LAE overdensity  $\delta_{\text{LAE}}$  on a scale of  $r = 10$  cMpc. The red stars and dots are the positions of used LoSs and masked LoSs respectively, with the color coded by effective optical depth on a scale of  $15 h^{-1}$  cMpc. The purple crosses represent the proximity (e)BOSS quasars with  $2.15 < z < 2.20$ . White regions are the masks used to exclude regions with low S/N signals, saturation around bright stars or serious stray light. The circles highlight the outliers in  $J0210$ , and details can be checked in the text in Section 7.

high  $\tau_{\text{LoS}}$  LoSs. Another clump with comparable significance appears at the west side, but it seems to be independent from the central structure. Interestingly, a large filamentary structure with an overdense peak  $\delta_{\text{LAE}} > 1.6$  appears at the southern boundary of the

FoV. There are nearby LoSs in the vicinity with relatively high  $\tau_{\text{LoS}}$ , but they are out of our pointing FoV. So, this structure is not found intendedly by the strong IGM Ly $\alpha$  absorption, but just by chance.

In the field  $J0924$ , we mainly use the central four grouping LoSs with high  $\tau_{\text{LoS}}$  as tracers. But in the central area, we do not find a structure with significant overdensity based on this LAE sample. Within the  $J0924$ , several peaks have moderate overdensities  $\delta_{\text{LAE}} > 0.8$  that are comparable with or surpass the central structure. The most overdense structure is found at the southwest of the field, which is close to two LoSs with strong Ly $\alpha$  absorption. The peak of the structure has an overdensity measured over 1.2, and it extends for about 30 cMpc.

The field  $J1419$  shows more structures in the clumpy shapes. Although there are four LoSs with  $\tau_{\text{LoS}} > 0.6$ , they are more scattered with distances of  $\sim 40 - 100$  cMpc to each others compared to those in other fields. Hence the coherently strong absorption is expected to be less significant. But instead, the number of LoSs in this field is appreciable for the correlation analysis. Five peaks with the moderate  $\delta_{\text{LAE}} > 0.6$  can be found in various regions, but no extreme overdense or extended structure is in this field. On the contrary, a large void with a size of  $\sim 50 \times 60$  cMpc<sup>2</sup> emerges at the northwest of the FoV.

#### 4.3. Correlation between Galaxy and IGM HI

The past observational studies on a large scale correlation are still restricted by the limits in both the FoV and the depth. We have described some relations between the LoSs and the overdensities qualitatively in Section 4.1, and from this section, we will have more quantitative analysis on such correlation in statistics.

To quantify the correlation, we calculate the overdensity on the scale of 10 cMpc in radius, at the positions of the clean LoSs. Similar to the Section 4.1, we discard the LoSs whose vicinity are masked out by more than 50%, but as the result, no LoS is removed in this process and the number of remaining LoSs is still 64. We assume the density in the masked regions to be the mean value in each field respectively. Errors are estimated as the Poisson noise using the statistics proposed in Gehrels (1986), which is the dominant uncertainty due to the small number statistics (Cai et al. 2017b). Then we can compare the LAE overdensity  $\delta_{\text{LAE}}$  and the effective optical depth  $\tau_{\text{LoS}}$  measured for the LoSs, whose error is derived from the error of mean flux in pixel statistics, to investigate the correlation. Figure 7 shows the result. As we can see from the figure, the error for  $\delta_{\text{LAE}}$  suffers from the Poisson statistics with a small number of shots (LAEs) in each measured aperture. While for  $\tau_{\text{LoS}}$ , the large error is mainly due to the relatively low S/N of the quasar spectra at the NB387 sensitive wavelength, which is close to the blue-end of the response range of

the SDSS spectrograph. Note that we have discarded LoSs with continuum-to-noise ratio smaller than 2.

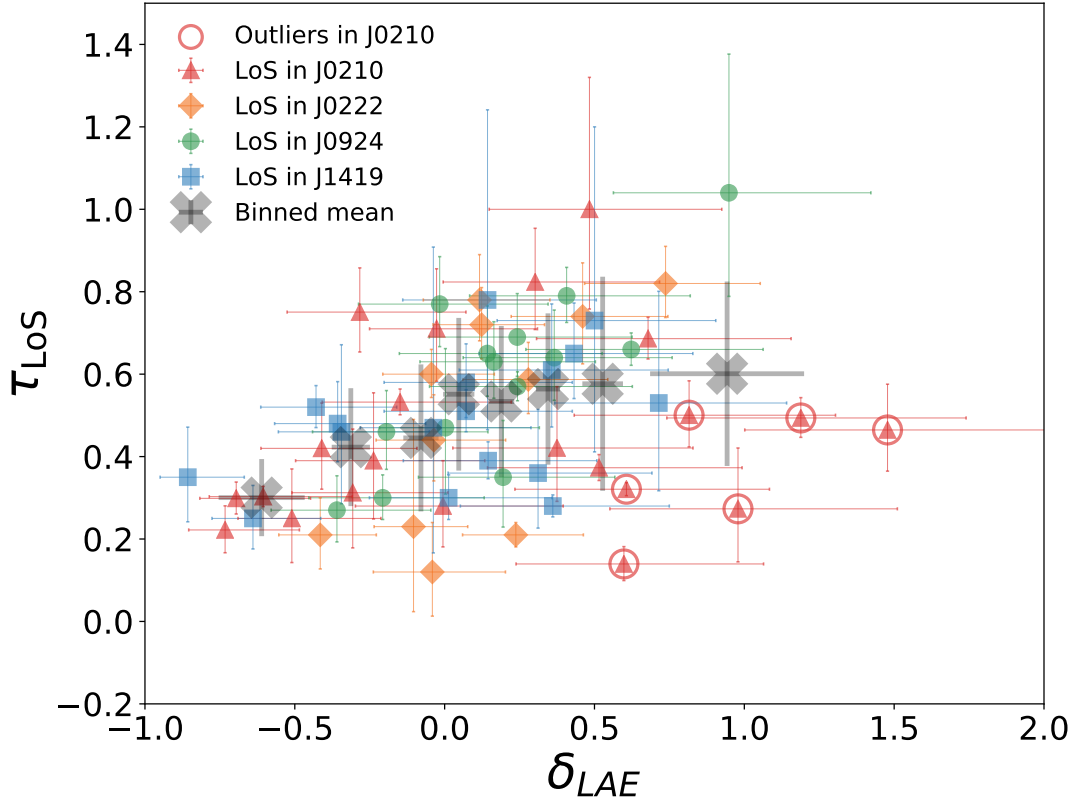
A tentative positive correlation can be found intuitively in the figure albeit, though with a large scatter. We perform the Spearman's rank correlation test for the full data sample, and the result shows the Spearman's rank correlation coefficient as  $\rho_S = 0.384^{+0.015}_{-0.037}$  with a  $P$ -value = 0.09%. The uncertainty of  $\rho_S$  is estimated by performing a Monte Carlo simulation by fluctuating the data points within their errors. We make 10,000 runs to pull sets of pseudo data from the Gaussian distributions, whose mean  $\mu$  and standard deviation  $\sigma$  are the observed data and the corresponding error. The shown values are the 16%, 50% and 84% rank of the simulated  $\rho_S$  results. It proves a moderately positive correlation with strong confidence between the LAE overdensity and IGM effective optical depth, based on the LoSs that are randomly distributed on the areas extended to scales over 100 cMpc at  $z > 2$ .

We find that the large scatter in Figure 7 might be largely contributed by the LoSs in  $J0210$  (red triangles), which contains a unique structure and has a shallower limiting magnitude. If we exclude  $J0210$ , the Spearman's rank correlation increases largely to an  $\rho_S = 0.541^{+0.037}_{-0.051}$  with a  $P$ -value < 0.01%. The NB387 limiting magnitude of  $J0210$  is shallower than others and the selected LAEs distribute at the relatively bright-end. The bright-end LAEs can result in the overestimated overdensities compared to other fields (Lee et al. 2014b; Casey et al. 2015). In this case, the bias from the potentially different spatial distributions of bright and faint galaxies can enlarge the scatter of overdensity. We perform the same correlation analysis by limiting the LAE NB387 magnitude to 24.3 for all four fields, and there are 451, 288, 264 and 248 LAEs left in  $J0210$ ,  $J0222$ ,  $J0924$  and  $J1419$  respectively, but the results with ( $\rho_S = 0.388^{+0.026}_{-0.045}$ ) and without ( $\rho_S = 0.502^{+0.031}_{-0.061}$ )  $J0210$  are consistent with those shown previously within the uncertainty, and cannot explain the significant difference. Therefore, the limit of the bright-end is unlikely to be the driven origin.

Alternatively, the difference of  $\rho_S$  can also be originated from the field variation in the correlation. The found large filament and the existence of the grouping proximity quasars indicate that the structures in  $J0210$  are probably different from other fields. More fields will be required for the more robust statistics in the future.

The binning data<sup>3</sup> shows a clearer trend intuitively, which is overlaid as the grey crosses in Figure 7. The

<sup>3</sup> The bins are made by sorting data points according to their  $\delta_{\text{LAE}}$  and splitting the nearest eight LoSs into one bin. There



**Figure 7.** Correlation between LAE overdensity  $\delta_{\text{LAE}}$  and effective optical depth  $\tau_{\text{LoS}}$  at the positions of LoS. Red triangle, yellow diamond, green circle and blue square represent the original data points in fields *J0210*, *J0222*, *J0924* and *J1419* respectively. The number of the LoSs are 63 for 4 fields on a scale of 10 cMpc with less than 50% masked vicinity. The grey crosses are the binned data points with the  $1\sigma$  standard deviation. Except for the most overdense bin which is dominated by the data points in *J0210*, a clear increasing trend shows out. The outliers in *J0210* with close spatial distribution are highlighted by red circles.

$\tau_{\text{LoS}}$  increases with the  $\delta_{\text{LAE}}$  at all range, though interestingly, the pace of increasing seems to be slower and the trend becomes flatter when  $\delta_{\text{LAE}} \gtrsim 0.2$ . We notice the trend at the overdense end is likely dominated by the *J0210* LoSs contributing in the  $\delta_{\text{LAE}} > 0.5$  bins. Especially, some of these LoSs are spatially close, and we highlight these special LoSs, hereinafter referred to as *outliers*, by circling them out in Figure 7, and their sky distributions are also shown in Figure 6 with the same symbol. We can find that the outliers cluster at two regions in *J0210*, which are close to the node of the filament. Considering that the *J0210* LoSs do not show a large scatter at the smaller  $\delta_{\text{LAE}}$  bins, the field variation, instead of the bright-end limit, is again favored to be the reason for the  $\rho_S$  difference between the cases with and without *J0210*.

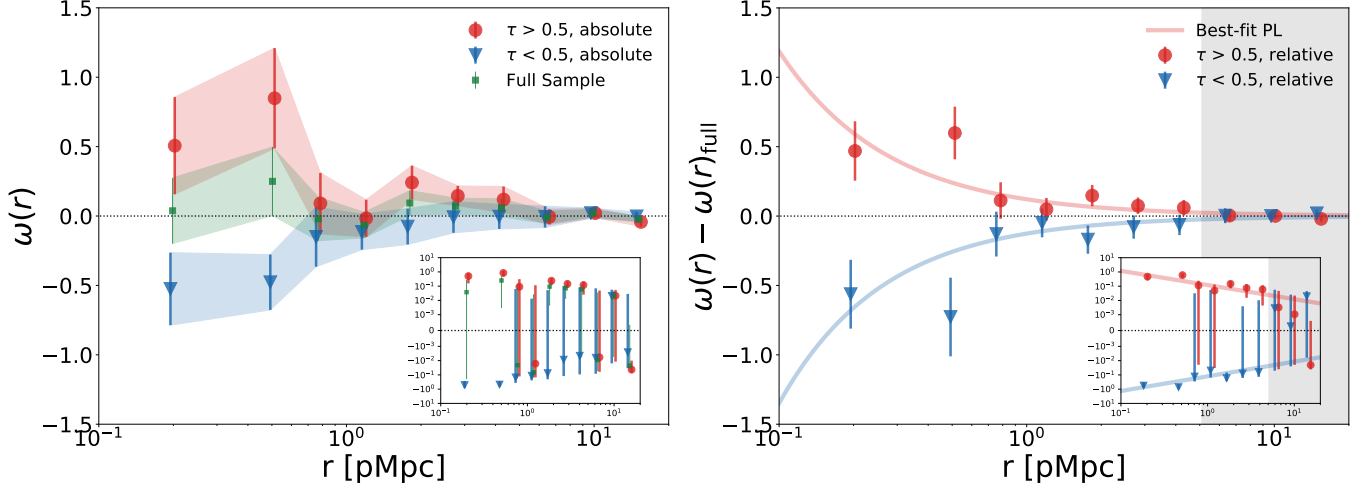
Therefore, it might indicate that different physical processes may have taken place in the *J0210* filament

compared to the typical structures at the same redshift. The lower  $\tau_{\text{LoS}}$  of the outliers than those of other LoSs can suggest either the lack of IGM HI in *J0210* or the LAE deficit in other fields, or both. As mentioned above, when limiting the LAE sample with NB387 magnitude up to 24.3, there are 451 LAEs in *J0210*. The number is a factor of  $> 1.5$  larger than the cases in the other fields, suggesting the number excess in *J0210*. Meanwhile in passing, we note that Momose et al. (2020b) have found that the LAEs might be residing slightly off-centered from the most highest density regions identified by the HI tomography. This is suggestive and is consistent with the lower  $\tau_{\text{LoS}}$  values in *J0210*, although a larger and deeper sample of LAEs and many more higher-resolution LoSs would be needed to say something more definitive.

#### 4.4. Cross-correlation Analysis

Along with the analysis based on the local overdensity of LoSs, a more general analysis can be made with the galaxy–IGM HI correlation. Correlation of the spatial distribution can be translated as the clustering prop-

are eight bins for the 63 LoSs in total, with 7 LoSs in the largest  $\delta_{\text{LAE}}$  bin, and the error is the  $1\sigma$  standard deviation at each bin.



**Figure 8.** Cross-correlation function (CCF) between LAEs and LoSs for the high  $\tau_{\text{LoS}}$ /low  $\tau_{\text{LoS}}$  subsamples. Red points and curves are the  $\tau_{\text{LoS}} > 0.5$  subsample and the corresponding fit power law model, while blue points and curves are for the  $\tau_{\text{LoS}} < 0.5$  subsample. Data points for different subsamples at each bin are slight shifted along the  $r$ -axis for clarification. In both panels, the major figures are shown in linear scale while the inset figures are shown in log scale. *Left panel:* original CCFs are shown for the two subsamples in red/blue, and the full sample in green, whose LoSs consist of the two subsamples. Shaded regions are the uncertainties from Jackknife resampling. *Right panel:* relative CCFs are calculated by subtracting the full sample signal  $\omega(r)_{\text{full}}$  from the original CCF of each subsample  $\omega(r)$ . The solid curves are the best-fit power law models for all of the data points. The grey shaded region indicates the separation limit where the signal becomes noisy.

erties between the two populations of objects, and to quantify the clustering strength, the two-point cross-correlation function (CCF) can work as an ideal tool.

We divide the LoSs into two subsamples according to the measured  $\tau_{\text{LoS}}$ . For the purpose of having a comparable number of LoSs in the two subsamples, we set the criterion as  $\tau_{\text{LoS}} = 0.5$ . LoSs with  $\tau_{\text{LoS}} > 0.5$  are called as the high  $\tau_{\text{LoS}}$  LoSs, while LoSs with  $\tau_{\text{LoS}} < 0.5$  are correspondingly called as the low  $\tau_{\text{LoS}}$  LoSs. In this case, if we use the full sample, then the number of LoSs for high/low  $\tau_{\text{LoS}}$  subsample is 30/34 respectively, and if we exclude field *J0210*, then the number of LoSs for high/low  $\tau_{\text{LoS}}$  subsample changes to 23/19. We mainly discuss the case including the *J0210* LoSs for CCFs. Although there are outlying LoSs found in Section 4.3 in *J0210*, we note that only 6 outliers are pinpointed while there are 64 LoSs in total. The statistics like CCF is unlikely to be biased.

We use the angular CCF  $\omega(\theta)$ , or the so-called projected CCF  $\omega(r)$  if the angular separation is translated into projected physical distance, for our analysis. To estimate the  $\omega(\theta)$ , we apply the estimator proposed by Landy & Szalay (1993), which can be better constrained in errors, to compare the data pairs against the randomly distributed points:

$$\omega(\theta) = \frac{D_{\text{LAE}}D_{\text{LoS}}(\theta) - D_{\text{LAE}}R(\theta) - D_{\text{LoS}}R(\theta) + RR(\theta)}{RR(\theta)}, \quad (6)$$

where  $D_{\text{LAE}}D_{\text{LoS}}(\theta)$ ,  $D_{\text{LAE}}R(\theta)$ ,  $D_{\text{LoS}}R(\theta)$  and  $RR(\theta)$  are the normalized LAE–LoS, LAE–Random, LoS–Random and Random–Random pairs counted at the separation of an angle  $\theta$  within an interval of  $\delta\theta$ . The normalization factor is the total pair number of each term.

To keep the statistics significant, we choose the right boundary of the innermost bin as 0.013 deg ( $\sim 0.4$  pMpc at  $z = 2.2$ ) so that there are  $> 10$  pairs at the bin in one subsample, reaching  $S/N > 3$  in Poisson statistics. Ten bins are set for the calculation extending up to 0.6 deg<sup>4</sup>. Note here that  $D_{\text{LAE}}D_{\text{LoS}}(\theta)$  represents the LAE–LoS pairs, not the LAE–absorber pairs. We do not use the information of location along the LoS of the absorbers, because the exact LAE redshifts are unknown within  $2.15 < z < 2.20$ , and the LoS-direction distance is meaningless even if we know where the absorbers are. This is why we use the projected CCF, but not the 3D one.

The error for the CCF is estimated by the Jackknife resampling, which can also take the field fluctuation into account. To do the resampling, we split each HSC field into  $5 \times 5$  square sub-fields, and the sub-fields that are overlapped by over 50% mask regions are excluded to ensure a sufficient number of pairs in each sub-field. Fol-

<sup>4</sup> We test the following results by varying the bin size, and we confirm that our major results are not sensitive to the bin determination.

lowing [Norberg et al. \(2009\)](#), we denote  $i$  as the calculating log scale bin, and  $k$  as the resampling run. In the  $k^{\text{th}}$  run, we skip the  $k^{\text{th}}$  sub-field and perform an identical CCF calculation as for the full field sample. Then the variance of the statistics of our interest, i.e.,  $\omega(r)$ , can be derived for the  $i^{\text{th}}$  bin:

$$\sigma_i = \frac{N_{\text{sub}} - 1}{N_{\text{sub}}} \sum_{k=1}^{N_{\text{sub}}} (\omega_{i,k} - \bar{\omega}_i)^2, \quad (7)$$

where the  $\bar{\omega}_i$  is the mean over all resampling runs given by  $\bar{\omega}_i = \sum_{k=1}^{N_{\text{sub}}} \frac{\omega_{i,k}}{N_{\text{sub}}}$  at the  $i^{\text{th}}$  bin.

As we described above, the projected CCF does not rely on the information of LoS-direction location. Without considering the Ly $\alpha$  absorption, the LoSs should be viewed as being selected homogeneously from the sky and they are not dependent on the foreground IGM at  $z \sim 2.2$ . Therefore, if the LoS number is infinite, a full sample without being split by the  $\tau_{\text{LoS}}$  is expected to have a null CCF signal.

However, our sample size is limited in fact, and this may involve an artificial signal into the CCF. We firstly check the CCF for the full sample combining the high  $\tau_{\text{LoS}}$  and low  $\tau_{\text{LoS}}$  LoSs, and the result is shown as the green points in the left panel of Figure 8. Although the full sample has much weaker signal than any subsample, which is clearer in linear scale by comparing the green points with blue/red points, they do not exactly equal to zero. This effect is due to a limited sample size. For the purpose of the clearer comparison, we subtract the amplitude of the full sample CCF  $\omega(r)_{\text{full}}$  from that of each subsample CCF  $\omega(r)$ , and we call the reduced signal as the relative CCF, i.e.,  $\omega(r) - \omega(r)_{\text{full}}$ , which is shown in the right panel of Figure 8. Data of subsamples at each bin is slightly shifted along the  $r$ -axis in the figures for clarification.

From the both panels in Figure 8, we find the high  $\tau_{\text{LoS}}$  subsample shows a continuous positive signal from the innermost bins up to a separation  $r \approx 4$  proper-Mpc (pMpc). On the contrary, the low  $\tau_{\text{LoS}}$  CCF stays negative in the same distance range. By varying the bin size, this characteristic distance changes by smaller than 1 pMpc. This result suggests that up to a scale of  $4 \pm 1$  pMpc ( $\sim 13 \pm 3$  cMpc at  $z = 2.2$ ), LAEs tend to cluster in the regions rich in gas, indicated by the high  $\tau_{\text{LoS}}$  LoS, and avoid the low  $\tau_{\text{LoS}}$  region where the gas is less abundant<sup>5</sup>. We also notice the two bins at  $\sim 0.8$  –

<sup>5</sup> As the accurate LAE redshifts are unknown, one can view the  $\tau_{\text{LoS}}$  estimated at the absorption spike works as the upper limit constraining the intrinsic HI associated with the LAEs around a LoS. This is why the  $\tau_{\text{LoS}} < 0.5$  subsample can show a negative

1.0 pMpc tends to be consistent with zero, suggesting a weak signal at the distance.

Interestingly, the CCF shown in the right panel can be well fitted by a power law:

$$\omega(r) = \pm \left( \frac{r}{r_0} \right)^{-\gamma}, \quad (8)$$

where the  $r_0$  is called as clustering length that makes  $\omega(r_0) = 1$ , and it can be an indicator of the clustering strength. We fit the binned data points with the power law by using least-square method with Levenberg-Marquardt algorithm, and the fitting curves are shown in corresponding colors in the Figure 8. The best fit parameters ( $\gamma$ ,  $r_0$ ) with the errors estimated from the 10,000 Monte Carlo perturbed simulations, similar to Section 4.3, are  $(0.99_{-0.17}^{+0.54}, 0.12_{-0.03}^{+0.05}$  pMpc) and  $(1.03_{-0.21}^{+0.83}, 0.13_{-0.02}^{+0.06}$  pMpc) for the high and low  $\tau_{\text{LoS}}$  subsamples respectively, and they are also summarized in Table 3 in Appendix B. The  $r_0$  for both subsamples are of an order of 0.1 pMpc, which is much smaller than the typical clustering strength in the case of galaxy-galaxy clustering derived from 3D CCF, i.e., several pMpc. This indicates that the strength of the LAE-IGM HI clustering is not very strong, though it is still significant enough for being detected based on our samples for the projected CCF. We note that [Momose et al. \(2020b\)](#) obtained somewhat stronger 3D CCF signal between LAEs and CLAMATO HI absorption data with  $r_0 = 0.78 h^{-1}$  cMpc, which corresponds to  $\sim 0.35$  pMpc for  $z \sim 2.2$  in our cosmology.

We test whether the results will be changed, if we exclude the field J0210, or if we change the  $\tau_{\text{LoS}} = 0.5$  criterion to separate the LoSs into subsamples. We do not find that such factors have significant impacts on our results, and details can be found in Appendix B.

#### 4.5. Average Optical Depth Profile to LAEs

We can further trace down to the circumgalactic medium (CGM) scale using our LAE and LoS samples. The aforementioned analyses mainly focus on  $r \gtrsim 1$  cMpc. Because the overdensity-based analysis requires a large enough aperture to overcome the small number statistics when counting galaxies. In the CCF analysis, we need to divide LoSs into high/low  $\tau_{\text{LoS}}$  subsamples, which makes a drop in the sample size by at least a factor of two. This can be extremely problematic for the smallest separation bin, which pushes us to set the innermost bin as large as 0.013 deg.

CCF, even though many of them still own the  $\tau_{\text{LoS}}$  higher than the cosmic mean value.

For the purpose of studying a smaller scale down to sub-cMpc, or  $\sim 200$  kpc in physical length, where CGM is supposed to surround the host galaxies, we perform another analysis that is similar to the stacking technique in concept. We derive the average radial distribution of the IGM optical depth that is averaged over all LoSs within a ring-like bin centered at a specific LAE. We then further calculate a mean over all LAEs, named as the average  $\tau_{\text{LoS}}$  profile  $\langle \tau \rangle$ , where  $d$  is the distance from the stacked LAE. To emphasize the excess level, we define the fluctuation of the  $\langle \tau \rangle$  as

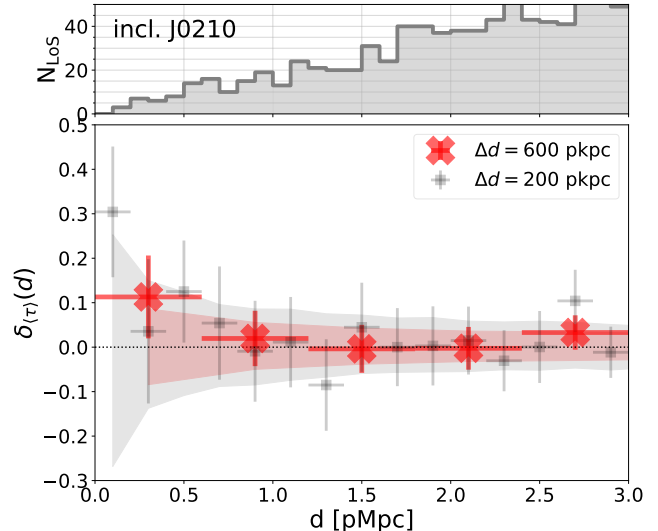
$$\delta_{\langle \tau \rangle}(d) = \frac{\langle \tau \rangle - \langle \tau \rangle_{\text{tot}}}{\langle \tau \rangle_{\text{tot}}}, \quad (9)$$

where  $\langle \tau \rangle_{\text{tot}}$  is the mean over the radial direction. We first calculate the  $\langle \tau \rangle_{\text{tot}}$  over a large distance range  $0 < d < 0.3$  deg, or  $0 < d < 9.2$  pMpc at  $z = 2.2$ , in the two cases, i.e., the coarse bin with a spatial resolution of  $\Delta d = 600$  pkpc and the finer bin with the a higher resolution  $\Delta d = 200$  pkpc<sup>6</sup>. Then the  $\langle \tau \rangle$  can be derived based on the bins, and the signal of CGM is expected in the inner regions. Momose et al. (2020b) also found CCF signal at CGM scales between LAEs and HI absorption with an interesting plateau in the central few hundreds pkpc (see their Fig. 9).

Results derived from all the LAEs and LoSs in four fields are shown in the lower panel of Figure 9. We mainly consider the case including *J0210* here given the same reason for CCFs, i.e., statistics is unlikely to be biased by 6 outliers out of 64 LoSs. We check the case excluding *J0210* in the Appendix C, and it shows the consistent results except for a larger scatter due to the LoS number decrease. The grey squares are for the finer bin, while the red crosses represent the coarse bin. The error shown in y-axis is the  $1\sigma$  standard deviation from the 1,000 times Bootstrap resampling with both the LAEs and LoSs, and the one in x-axis indicates the bin size. The shaded regions are the 16%–84% ranks in the 1,000 simulations assessing random positions to the 64 LoSs with corresponding resolutions, indicating that bins outside the shaded regions are confident for inferences. The number of LoSs counted at each finer bin can be checked with the grey step function above.

From the figure, we can learn that a 10% excess beyond the error is found in the innermost region, corresponding to a scale of  $< 600$  pkpc. Though the

<sup>6</sup> The two cases are chosen because: (1) they are concerned in physics as the 200 pkpc is a typical scale of CGM and 600 pkpc is persuasively far enough to be in the IGM regime; (2) signals only exceed the 84% ranks in random LoSs with these two choices to draw a meaningful result after testing various bin sizes.



**Figure 9.** The fluctuation of the average  $\tau_{\text{LoS}}$  as a function of distance to LAEs,  $\delta_{\langle \tau \rangle}(d)$ , for the case including *J0210*. Errors are indicated by the  $1\sigma$  standard deviation from the 1,000 times Bootstrap resampling. The grey squares represent the finer bins with resolution of 200 kpc, and the red crosses show coarse bins with resolution of 600 kpc. The shaded region indicates the uncertainty for coarse bins. The numbers of LoSs counted in the annulus at each step of 100 kpc are shown in the upper panel. A 30% excess at a level of  $2\sigma$  appears at  $d < 200$  pkpc indicating the detection of CGM signal around LAEs, while a tentative 13% excess at  $400 < d < 600$  pkpc shows a weaker signal in the IGM regime.

counted LoSs number in the innermost finer bin is only three, but we see a more significant 30% excess at a  $2\sigma$  level at  $d < 200$  kpc, which is the expected region distributed with CGM. Rudie et al. (2012) and Momose et al. (2020b) also found the CGM signals at  $d \lesssim 300$  pkpc to the star-forming galaxies that are firstly identified as LBGs. This time, we find the indication may be also true for LAEs at a smaller distance from the statistical point of view.

In addition, the finer bins seem to also indicate a 13% excess at the distance of  $400 < d < 600$  pkpc, and given its sufficient number of LoSs, this excess is likely to be the dominant signal accounting for the 10% excess in coarse bin. However, such scale is larger than the typical region thought to be the CGM reservoir, especially with regard to LAEs which are generally less massive than LBGs. Instead, the clustering of IGM HI works as a better interpretation for the excess found in  $\langle \tau \rangle$  at this distance. This excess in  $\langle \tau \rangle$  corroborates the signal detected in CCF at  $\sim 0.5$  pMpc, proving the correlation between galaxy and IGM HI down to around 400 pkpc.

Between the two excess bins,  $d = 200 - 400$  kpc interestingly shows a relative valley. A turnover seems to



appear at  $d \approx 400$  kpc. The significance is very low due to the small number of LoSs counted at the bin, but we note that a similar turnover is once also reported for the  $z \sim 3$  LBGs by [Adelberger et al. \(2003\)](#), though theirs appears at  $0.5 h^{-1}\text{Mpc}$ , or 190 kpc in our cosmology. The latest result based on a sample of 2,862 background galaxies has also revealed a more similar sudden dip at 70–150 pkpc ([Chen et al. 2020](#)). The former work has estimated that the supernova-driven outflow with the speed of  $600 \text{ km s}^{-1}$  may cause the turnover, while the latter suggest the feature can be related to the transition phase between the inflow outside and the outflow inside, which may be related with star formation activities. In this picture, it is possible that LAEs, the young and less massive galaxies which can be located at the shallower potential well and is active in forming stars, host the stronger outflows and cause the turnover appearing at a larger projected distance. The current weak signal in our data still prevents us drawing any firm conclusion, but a larger sample size in the future may help to resolve this question.

We also notice there is a sudden excess of 10% at  $2.6 < d < 2.8$ , though the coarse bin largely flatten the signal. We do not fully understand the origin of this signal, but a non-continuous signal at such a large scale is not likely to be physically meaningful.

## 5. DISCUSSION

Based on the results shown in Section 4, we discuss their implications. We first make a comparison between our galaxy–IGM HI result with the previous literature. As hinted in the CCF, we find the correlation is possibly dependent on the scale. Therefore, we further explore the scale dependence of the observed correlation. Finally, we discuss the possible underlying physics that are related to our results on the positive correlation, correlation scale and the visible scatter in the  $\delta_{\text{LAE}}-\tau_{\text{LoS}}$  diagram at  $z \sim 2$ .

### 5.1. Comparison with Previous Work

There are already a few studies in the literature working on the correlation between galaxies and IGM HI on the over tens of cMpc scales at  $z > 2$ . The directly related work is [Mukae et al. \(2017\)](#), in which the galaxy–IGM HI correlation is studied by using the  $K_s$ -selected photo- $z$  galaxies at the redshift  $2 < z < 3$  and the Ly $\alpha$  forest sample in the background quasar spectra from SDSS-III/BOSS survey.

The correlation shown in their Figure 2 is physically similar to our  $\delta_{\text{LAE}}-\tau_{\text{LoS}}$  correlation, but the Ly $\alpha$  absorption is estimated in Ly $\alpha$  forest fluctuation which is

defined as:

$$\delta_{\langle F \rangle} = \frac{\langle F \rangle_{dz}}{F_{\text{cos}}(z)} - 1, \quad (10)$$

where  $\langle F \rangle_{dz}$  is the transmission calculated within the redshift uncertainty  $dz = 0.025(1+z)$  from the spectra and the  $F_{\text{cos}}(z)$  is the cosmic Ly $\alpha$  forest mean transmission that is estimated from  $F_{\text{cos}}(z) = e^{-0.001845(1+z)^{3.924}}$  ([Faucher-Giguère et al. 2008](#)).

To compare with their results, we convert the optical depth derived in Section 2.1 into the transmission fluctuation  $\delta_{\langle F \rangle}$  according to the Equation 10. The cosmic mean is also assumed to be given by the relation in [Faucher-Giguère et al. \(2008\)](#) as 0.84 at  $z = 2.18$ . The translated  $\delta_{\langle F \rangle} - \delta_{\text{LAE}}$  relation from our LAEs and LoSs sample is shown in Figure 10.

The symbols of the data points are the same as Figure 7, but for clarifying the different cases including or excluding  $J0210$  for the fitting, we paint the LoSs in  $J0210$  red and the LoSs in other fields blue. We also make a linear fit using the Levenberg-Marquardt least-square fitting, shown as the solid lines in Figure 10, in(ex)cluding the  $J0210$  corresponds to the red(blue) curve. The uncertainty of parameters is again given by the 16%–84% ranks from the 10,000 Monte Carlo simulations with perturbation. The fitted relation for all four fields is:

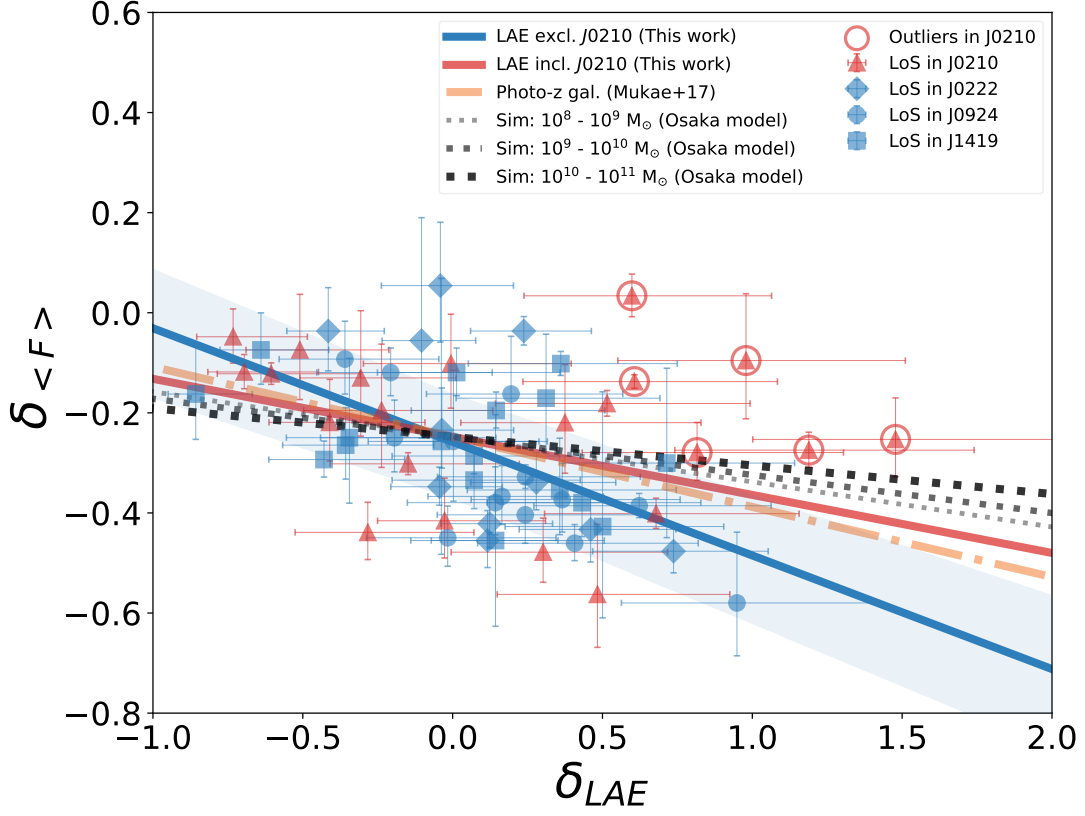
$$\delta_{\langle F \rangle} = -0.116_{-0.022}^{+0.018} \delta_{\text{LAE}} - 0.248_{-0.093}^{+0.082}, \quad (11)$$

Similar to Figure 7, we can find the outliers in  $J0210$  at the upper right in Figure 10, which is highlighted with circles. If we exclude the LoSs in  $J0210$ , the relation becomes:

$$\delta_{\langle F \rangle} = -0.227_{-0.023}^{+0.026} \delta_{\text{LAE}} - 0.258_{-0.114}^{+0.096}, \quad (12)$$

which shows a steeper slope, meaning the  $\tau_{\text{LoS}}$  is more sensitive to the  $\delta_{\text{LAE}}$ . We overplot the curve whose slope is  $-0.14_{-0.16}^{+0.06}$  from the [Mukae et al. \(2017\)](#), with the intercept normalized at  $\delta_{\text{LAE}} = 0$ . The normalization is necessary as our tracers of absorption are not defined in the same way, which causes systematic offset reflected on the intercept. They estimate the  $\delta_{\langle F \rangle}$  at the position of the highest  $\text{S/N}_{\langle F \rangle}$ , defined as the ratio between Ly $\alpha$  absorption and its error, on  $\sim 100\text{cMpc}$  scale within the redshift  $2 < z < 3$ , while we are targeting at the absorption spike based on the  $\tau_{\text{LoS}}$  on  $\sim 20 \text{ cMpc}$  scale within  $2.15 < z < 2.20$ .

Both cases in our work give the consistent slopes with the photo- $z$  galaxies within their uncertainty, though the case excluding  $J0210$  owns a larger discrepancy and is steeper. One possible reason for the large discrepancy can be the different galaxy masses, given that photo- $z$  galaxies are generally more massive than LAEs. The



**Figure 10.** Transmission fluctuation  $\delta_{\langle F \rangle}$  against the LAE overdensity  $\delta_{\text{LAE}}$ , similar to the Figure 2 of Mukae et al. (2017). Symbols are same as Figure 7. LoSs in *J0210* are in red, while LoSs in the other three fields are painted blue. Outliers in *J0210* are highlighted by red circles. The red (blue) solid curves are the best-fit model for the data points in(ex)cluding *J0210*. The orange dotted dash line is the result from Mukae et al. (2017) using the photo- $z$  galaxies & 16 BOSS LoSs. The dotted lines are the prediction from GADGET3-Osaka model (Shimizu et al. 2019; Nagamine et al. 2020) for galaxies with  $10^8$ – $10^9$   $M_{\odot}$ ,  $10^9$ – $10^{10}$   $M_{\odot}$  and  $10^{10}$ – $10^{11}$   $M_{\odot}$ . The shaded regions are the 16%–84% rank from the perturbation simulations for the case excluding *J0210*.

massive galaxies are likely to form in the deeper position of the gravitational potential well, where the HI is abundant for building up stellar masses  $M_*$ . In this case, the overdensity of less massive galaxies like LAEs will be systematically lower than that of the heavier populations, e.g., photo- $z$  galaxies, and thus make the  $\delta_{\langle F \rangle} - \delta_{\text{LAE}}$  steeper. A similar trend is also reported in a study based on the IGM tomography (Momose et al. 2020b). In this case, the shallower slope with *J0210* can be explained by the possible LAE number excess in the large filament in *J0210*, because the filament is associated with a group of quasars and this can be an indicator of the potential massive halos around the structure. It may boost the  $\delta_{\text{LAE}}$  given the same  $\delta_{\langle F \rangle}$ , especially at the regions where the LoS outliers reside in, making the slope shallower when *J0210* is included.

To further inspect the possibility, we refer to the results from the GADGET3-Osaka cosmological hydrodynamic simulation, which is based on the smoothed particle hydrodynamics (SPH) simulation code GADGET-3

(Springel 2005) and takes full account of the star formation and supernova feedbacks (Shimizu et al. 2019). More details on the simulation data processing is explained in Momose et al. (2020a) as well as in Nagamine et al. (2020), and we denote it as the *Osaka model* hereinafter. The model curves for galaxies with  $M_*$  ranging in  $10^8 - 10^9$ ,  $10^9 - 10^{10}$  and  $10^{10} - 10^{11}$   $M_{\odot}$  with respective slopes of  $-0.090 \pm 0.011$ ,  $-0.076 \pm 0.009$  and  $-0.057 \pm 0.006$  are also plotted in Figure 10, and the intercepts are again normalized at  $\delta_{\text{LAE}} = 0$ , given that the absorption in the model is estimated at the fixed position, i.e., the central redshift  $z = 2.175$ , which is different from our estimate at the absorption spike.

We do find there is an  $M_*$ -dependence of the relation slopes in the Osaka model, and the less massive galaxy population owns a steeper trend. However, such dependence is not as sensitive as we expected and more interestingly, our fitting for the case including *J0210* shows a good consistency with the Osaka model prediction for the galaxies with  $M_* \sim 10^9$   $M_{\odot}$ , the typical stellar

masses for  $z \approx 2.2$  LAEs (Kusakabe et al. 2018). Meanwhile, the slope for the case excluding  $J0210$ , which was expected to be more representative of the general fields at  $z \sim 2$ , is significantly steeper than the Osaka model. These comparisons are likely to disprove the reason originated from the galaxy stellar masses, and the case excluding  $J0210$  seems rather to be the biased case.

Another possibility can be the HI suppression on the Ly $\alpha$  emission. Given that our observations target at the fields with the clustering of strong IGM Ly $\alpha$  absorption, the Ly $\alpha$  emission from galaxies may get suppressed in such HI-rich environments before we can observe. In this case, LAEs in the  $J0210$ , which is likely to contain a special structure lacking the IGM HI as suggested by the outlying LoSs, should be less influenced. Meanwhile, in the other fields, the detection completeness of LAEs could be lower and the  $\delta_{\text{LAE}}$  might be underestimated. This interpretation seems to be more favored by the Osaka model prediction. Actually, the plateau appearing in the CCF at  $r \lesssim 0.6$  pMpc also supports such a possibility at least on the small scales.

However, we note that there can be some uncertainties left in the simulation models (e.g., contribution from AGNs), and our sample size is still limited for the discussions on field variation. In the future, we are hopefully to find out the true reason for the slope discrepancy with more HSC fields targeting at various environments. Follow-ups to search for H $\alpha$  emitters (HAEs) residing in the same structures, which are less biased by the radiative transfer process, can also help to robustly calibrate the  $\delta_{(F)} - \delta_{\text{LAE}}$  slope.

### 5.2. Scale Dependence of the Correlation

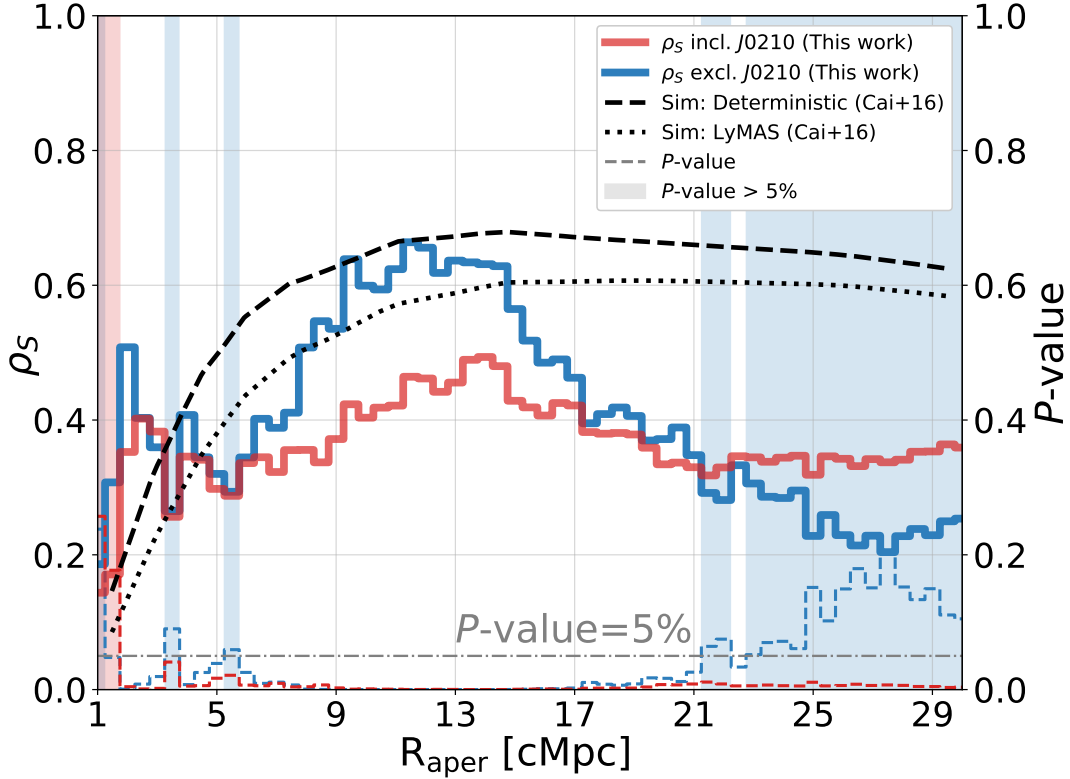
A possible scale dependence is already hinted in the CCF in Section 4.4 for the LAE and IGM HI correlation. To investigate the scale dependence, we perform the Spearman’s rank correlation test for  $(\delta_{\text{LAE}}, \tau_{\text{LoS}})$  with the  $\delta_{\text{LAE}}$  calculated in different aperture sizes. We again consider the two cases, including and excluding the field  $J0210$ , as we already find that it may significantly alter the overdensity-based analysis in Section 4.3 and 5.1. The aperture size is set from 1 to 30 cMpc with a bin step as 0.5 cMpc for the radius. We note that LoSs are kept for analysis only when  $< 50\%$  vicinity is masked, and this keeps a stable LoS number when the scale increases. The result is shown in Figure 11. The red (blue) curve shows the Spearman’s rank correlation coefficients on various scales for the case including (excluding)  $J0210$ . The corresponding  $P$ -value is shown as the dash line, and the  $R_{\text{aper}}$  with  $P$ -value  $> 5\%$  indicating an unconfident result is shaded. The similar results from the deterministic as well as Ly $\alpha$  Mass Association

Scheme (LyMAS) models in Cai et al. (2016) are also overlaid as the black curves. The LyMAS considers a stochastic relation described by a conditional probability distribution of the flux on the mass overdensity  $\delta_{\text{m}}$ , based on the hydrodynamic simulations. Note that the original box lengths in the simulations are scaled with a 1/2 to keep consistent with  $R_{\text{aper}}$ , and this scale-match for estimating galaxy/total matter overdensity is performed on the projected plane.

It is clear that when the  $J0210$  is included, the correlation keeps moderate at a level of  $\rho_S \sim 0.3$  for almost all scales, while in the case excluding  $J0210$ , the correlation becomes strong at  $r \approx 9 - 15$  with the  $\rho_S \gtrsim 0.6$ . In both cases, we can find the scale dependence of the correlation between  $\delta_{\text{LAE}}$  and  $\tau_{\text{LoS}}$ , though the trend is much more significant when the  $J0210$  is not included. At the relatively small scale, the correlation becomes stronger with the scale increases, and it reaches a peak at  $r = 13 \pm 2$  cMpc. With  $J0210$ , the correlation shows a flatter shape when  $r > 13$  cMpc, while it tends to decrease at such scales if  $J0210$  is rejected. The difference again, indicates that  $J0210$  may own a special structure and the existence of such structure can alter the correlation significantly in the overdensity-based analysis. So, when doing the galaxy-IGM HI correlation study, a large sample size covering various types of environments should be essential. But here, we will keep the discussions focused with  $J0210$  excluded.

The CCF in Section 4.4 shows that the amplitude for high  $\tau_{\text{LoS}}$  subsample keeps positive up to 4 pMpc ( $\sim 12.7$  cMpc). As the correlation shown in Figure 11 can be viewed as the cumulative signal within the  $R_{\text{aper}}$ , the scale of the correlation peak agrees well with the CCF result. Compared with the models in Cai et al. (2016), our result on the scale of maximum correlation is also well consistent to both of the deterministic one and the LyMAS on the projected plane, though the amplitude may be different due to the different tracers we use. This consistency suggests the effectiveness of the current simulated cosmological models in terms of the IGM HI gas.

The decrease at large scale is interesting, as it is not predicted in the simulation. Note the HSC FoV covers a region on the scale over 100 cMpc, so it is not likely to be the reason accounting for the weaker correlation on scales just over 15 cMpc. We also assess the mask region criterion by changing  $< 50\%$  to  $< 10\%$  when estimating the  $\rho_S$ . With this change, the analysis only uses the clean LoS sample free from the uncovered regions. This test also shows a similar decreasing trend at  $r > 15$  cMpc, supporting that the decrease is unlikely to be caused by the FoV limit.



**Figure 11.** Scale dependence of the  $\delta_{\text{LAE}}-\tau_{\text{LoS}}$  correlation. Bin size is 0.5 cMpc and the red (blue) solid curves are the Spearman’s rank correlation coefficient in(ex)cluding  $J0210$ . The dash colored curves indicate the corresponding  $P$ -value at each bin. Scale range with the  $P$ -value  $> 5\%$  is masked with shaded regions, indicating that the result is not confident. The two models, Deterministic & LyMAS models, from the Cai et al. (2016) are overplotted, by scaling the boxlength in simulations with a factor of 1/2 to match the  $R_{\text{aper}}$ . Both models reach the maximum at a comparable scale  $R_{\text{aper}} \sim 13$  cMpc.

An alternative reason of the discrepancy on large scale is that the models do not only use galaxies but rather use the total matter in a defined box, which is less clustering indicated by the simulations that at  $z \sim 2$  the galaxy bias keeps decreasing towards the scale over  $\sim 10 h^{-1}\text{cMpc}$  (Cen & Ostriker 2000; Springel et al. 2018), while the HI bias is almost flat at such large scale (Ando et al. 2019). Hence, the correlation in the models can stay strong on a larger scale, while the LAE–IGM HI correlation becomes weaker simultaneously. Another possible reason suggested by the simulations in Momose et al. (2020a) is that the signal on large scales is diluted in the projected correlation, as the uncertainty on the three dimensional separation becomes larger when  $R_{\text{aper}}$  increases. In addition, more contaminants included in a larger aperture can also weaken the signal.

### 5.3. Underlying Physics in the Correlation

#### 5.3.1. The Positive Correlation

We showed that, at the redshift  $z \approx 2.2$ , a moderate to strong positive correlation can be found between  $\delta_{\text{LAE}}$  and  $\tau_{\text{LoS}}$  on a scale of  $r = 10$  cMpc. Such a correlation suggests that galaxies are clustering in a region

associated with large amount of HI gas. This correlation is found to be scale dependent, and the peak locates at  $R_{\text{aper}} \sim 13$  cMpc. The correlation seems to be natural in a simple picture that IGM HI gas tends to be accumulated in the deeper potential wells which are inhabited by the more massive halos. The condensed HI gas then triggers star formation, and stars and galaxies will emerge at the same region. Especially at  $z \sim 2$ , such activity is extremely intensive according to Madau & Dickinson (2014).

However, when detailed processes are taken into account, the situation becomes complicated. Hot massive stars can emit ionizing photons with energy  $> 13.6$  eV, and once they succeed to escape from the host galaxies, the surrounding HI gas in the surrounding IGM will be ionized. Such process can be more active for the case of LAEs used in our work, which are thought to be a population of young star-forming galaxies. Our result indirectly suggests that the escape fraction of ionizing photons from LAEs at  $z \sim 2$  or their SFR is still not high enough to fully ionize the IGM HI gas on the scale of several cMpc. Feedback from supernovae or the potentially inhabiting AGNs can also be possible to blow

off the surrounding gas to more distant regions, though how powerful such processes can be is still under debate and it is not clear up to which scale they can affect.

Some literature has explored the two point CCFs between Ly $\alpha$  absorbers and galaxies in the lower- $z$  universe (Ryan-Weber 2006; Chen & Mulchaey 2009; Tejos et al. 2014), in which the correlations are confirmed under the redshift  $z \lesssim 1$ . But at  $z \gtrsim 2$ , such correlation can be only constrained with limitations in either bright galaxy population, small survey area or small LoS sample size in a limited number of works (Adelberger et al. 2003; Rudie et al. 2012; Mukae et al. 2017). Our result confirms the correlation between IGM HI and galaxies with rest-frame UV magnitude down to  $M_{UV} \approx -18$  estimated from observed  $g$ -band, even at the redshift  $z \approx 2.2$  where the star formation and feedback processes can be very active. The result shows a rough consistency with Mukae et al. (2017) based on photo- $z$  galaxies, but a factor of  $\sim 4$  larger sample size in both of the LoS number and survey area makes the statistics more robust with various overdense environments.

The identified positive correlation is found up to 4 pMpc (or 13 cMpc) from the CCF analysis (or varying the aperture size for  $\delta_{LAE}$  in the  $\delta_{LAE}-\delta_{(F)}$  correlation), and down to at least 400 pkpc (or 1.3 cMpc) from the average  $\tau_{LoS}$  profile centered at LAEs. This suggests the ionization or feedback from galaxies (LAEs) is not sufficient enough to cancel out the gravitational effects on large scale. This indicates that IGM HI still traces well LSS at  $z \sim 2$  on the scale 1.3  $\sim$  13 cMpc, though with large scatter. Alternatively, the correlation can also be a result of additional inflow providing exceeding pristine HI gas (Dekel & Birnboim 2006; Tumlinson et al. 2017). Turner et al. (2017) suggests the observed redshiftspace distortions in the KBSS survey (Rakic et al. 2012) are predominantly caused by infall, which proves gas inflow can alter observables up to a scale of 5 pMpc. The two possible scenarios can either or both reproduce our results and cannot be distinguished at this point. But it will be possible to answer this question by comparing our results with numerical simulations in the future.

Also, we still have little knowledge on how well the LAEs trace the underlying structures, especially in our fields which are expected to be associated with neutral IGM gas. Physical similarity between LAEs and non-LAEs at  $z \sim 2.2$  is hinted in Hathi et al. (2016), and Shimakawa et al. (2017) also find the overdense regions traced by LAEs and HAEs show good consistency on the scale of  $> 1$  cMpc, indicating that LAEs can be a good structure tracer on large scale. However, as reported in Shi et al. (2019), LAEs and LBGs do possibly trace different structures formed in different period or

in different dynamic status. Especially on small scale of  $\lesssim 300$  pkpc, or  $< 1$  cMpc, tentative deficit is always found for LAEs, both in this work hinted by the plateau shape in the CCF and in the literature, e.g., LAE number deficit in a protocluster core (Shimakawa et al. 2017) or at the center of the massive overdensity (Cai et al. 2017b), and the possible Ly $\alpha$  suppression in galaxy overdense regions (Toshikawa et al. 2016). This indicates that the LAE may be not a good tracer of the highest overdensity regions.

But for a statistical study on large scale, LAEs still work as the best tool with a well constrained redshift  $\Delta z \approx 0.04$  and Subaru/HSC can map the objects with high efficiency. In the future, we will perform the NB imaging with NIR instruments like Subaru/MOIRCS, on which the appropriate NB2083 ( $\lambda_0 = 2.083 \mu\text{m}$ ) filter is installed, to select the resonance-free HAEs to figure out the performance of the LAE tracers.

### 5.3.2. The Scale Dependence of Correlation

Our results on the scale dependence of the correlation was discussed in Section 5.3.1. The Subaru/HSC allows us to map extended structures as well as their environments up to a scale over 100 cMpc, down to a depth  $L_{Ly\alpha} \approx 2 \times 10^{42} \text{ erg s}^{-1}$  at  $z \approx 2.2$ . It makes our study unique for robustly confirming correlation on a large scale of several tens of Mpc in comoving at  $z > 2$ .

As mentioned previously, there are already some studies working on the CCF between Ly $\alpha$  absorbers and galaxies at  $z < 1$  (Ryan-Weber 2006; Chen & Mulchaey 2009; Tejos et al. 2014), and the CCFs provide us the information for both correlations and their effective scales. Given that the galaxy populations and the Ly $\alpha$  absorption systems used among our works are not identical, it is hard to directly compare the CCF amplitude and the resulting clustering length  $r_0$ . Nevertheless, the  $r_{up}$ , defined here as the upper limit of the scale to identify the positive signal, can be still instructive. From the CCFs, we find an underlying redshift evolution of the correlation scale by combining Ryan-Weber (2006), Tejos et al. (2014), whose CCFs also extend over 10 cMpc, with our result. We find that: (1) at  $z \lesssim 0.04$ , the CCF between Ly $\alpha$  absorbers with HI column density ranging in  $10^{12.5} \lesssim N_{HI} \lesssim 10^{15} \text{ cm}^{-2}$  and HIPSS galaxies shows a strong positive signal up to 10  $h^{-1}$  cMpc (Ryan-Weber 2006), i.e.,  $r_{up} \sim 15$  cMpc, slightly larger than our upper limit  $r_{up} = 13 \pm 3$  cMpc; (2) while at  $0 \lesssim z \lesssim 1$ , the signal of CCF between Ly $\alpha$  absorption systems with HI column density ranging in  $10^{14} \lesssim N_{HI} \lesssim 10^{17} \text{ cm}^{-2}$  and galaxies, can be only found up to  $r_{up} \sim 7$  cMpc (Tejos et al. 2014), significantly smaller than ours.

The  $r_{\text{up}}$  decreases from  $z > 2$  to  $0 \lesssim z \lesssim 1$ , and then increases again towards  $z = 0$ . This interestingly shows a consistency with the varying trend of the correlation length of galaxy clustering (Baugh et al. 1999; Springel et al. 2018). It supports a physical picture that the redshift evolution of galaxy-IGM HI correlation may follow a similar pattern of the galaxy clustering.

### 5.3.3. The Scatter of Correlation

The scatter can also be an important indicator of the underlying structures. As the Figure 7 and 10 show, data points are distributed with a large scatter. It may originate from the uncertainties in our measurements. We summarize the possible factors here. First, regarding the overdensity measured in our work, we can only map the LAEs on projected plane while an uncertainty of  $\sim 60$  cMpc is left along the redshift direction, and the aforementioned scales are all in transverse separation instead of in comoving volume. Additionally, we are not sure how much bias LAEs are introducing, as we already discussed in Section 5.3.1. As  $J0210$  changes the statistical results very much, structures with field-to-field variation may exist. Regarding the LoSs sample, though the CoSLAs have been carefully checked to exclude DLAs or LLSs, low  $\tau_{\text{LoS}}$  LoSs can be still possibly contaminated by these systems. But even if we only focus on CoSLAs (see LoSs with  $\tau_{\text{LoS}} \gtrsim 0.6$  in Figure 7), we can still find a large scatter, just like what Miller et al. (2019) report in their simulations with both the high spatial and mass resolutions. This indicates there should be some intrinsic origins.

The scatter can be coincidences that happen when LoSs pass through a gas filament, a large void or an orthogonal filament with low density. According to the simulation in Mukae et al. (2017), which also find a large scatter on their correlation, it may indicate the outliers in  $J0210$  penetrate a galaxy overdensity associated with a gas filament lying on a transverse direction to the LoS by chance. In addition to the morphological origin, the radiation from galaxies may preheat the diffuse IGM HI in the most overdense region, causing the scatter. This scheme is suggested by Mawatari et al. (2017), where Ly $\alpha$  absorption is found to be associated with a  $z \approx 3.1$  overdensity SSA22 on a scale  $\sim 50$  cMpc overall, but not dependent on local overdensity.

Actually, three outliers in  $J0210$  located at regions with  $\delta_{\text{LAE}} \gtrsim 1.0$  in the  $r_{\text{aper}} = 10$  cMpc aperture and  $\tau_{\text{LoS}} \lesssim 0.4$  are just likely to reside in the regions that are abundant with galaxies but in deficit of cold HI gas, similarly to the environments mentioned above. A special system found in the IGM tomography also shows the similar characteristics (Lee et al. 2016). By further

studying such cases in the scatter, we may be able to find more ideal laboratories for testing the theories of galaxy evolution and their interplay with IGM HI in the extreme environments at  $z \approx 2$ .

## 6. SUMMARY

In this paper, we perform deep NB387 and  $g$ -band imaging with the 8.2-m Subaru/HSC on the fields following the similar technique used in MAMMOTH project, which are preferentially traced by the group of strong Ly $\alpha$  absorbers selected from the full (e)BOSS database. Using the narrow-band images, we select out LAE candidates at  $z = 2.18$  and construct the  $\delta_{\text{LAE}}$  maps. To estimate the IGM HI overdensity, we use the (e)BOSS LoS data to calculate the  $\tau_{\text{LoS}}$  at the same redshift. Based on the  $\delta_{\text{LAE}}$  and  $\tau_{\text{LoS}}$  data, we perform correlation analyses to study the galaxy-IGM HI correlation up to a scale of  $\sim 100$  cMpc. In addition, we also examine the correlation on CGM scales down to 200 pkpc based on the statistical sample.

The results achieved are summarized as follows:

1. We construct the LAE overdensity maps for four HSC fields traced by IGM HI at  $z = 2.18$ , with a total of 2,642 LAE candidates detected down to  $L_{\text{Ly}\alpha} \approx 2 \times 10^{42}$  erg s $^{-1}$  over a survey area of 5.39 deg $^2$ . The selected LAE candidates reside in a variety of environments, including the filaments, sheets and clumps. The  $J0210$  field, which is associated with 11 quasars within  $\sim 40 \times 40$  cMpc $^2$  and  $\Delta z \approx 0.05$ , is found to be associated with a large LAE filament extending for about 100 cMpc, one of whose nodes reaches the overdensity significance of  $> 6\sigma$ .
2. We find a moderate to strong correlation between the  $\delta_{\text{LAE}}$  and  $\tau_{\text{LoS}}$  based on 64 LoSs from SDSS/(e)BOSS, which shows a rough consistency with the results in Mukae et al. (2017), though the  $\delta_{(F)} - \delta_{\text{LAE}}$  slope is steeper when we exclude the field  $J0210$ . Based on the comparison with the Osaka simulation model (Shimizu et al. 2019; Nagamine et al. 2020), the discrepancy is unlikely to be caused by different stellar masses, but rather due to the suppression of Ly $\alpha$  emission in high HI density regions. We further find that the correlation depends on the scale of  $\delta_{\text{LAE}}$  estimate. The peak of the correlation is located around  $R_{\text{aper}} = 13 \pm 2$  cMpc.
3. By dividing the LoSs into high and low  $\tau_{\text{LoS}}$  subsamples with a criterion of  $\tau_{\text{LoS}} = 0.5$ , the cross-correlation analysis shows a significant correlation signal up to  $4 \pm 1$  pMpc ( $\sim 12.7 \pm 3.2$  cMpc). The result clearly suggest that LAEs tend to reside in the gas-rich regions, which is indicated by the high

$\tau_{\text{LoS}}$  in the background LoS, and avoid the low  $\tau_{\text{LoS}}$  area where the HI is deficient. The plateau shape at  $r \lesssim 600$  pkpc suggests the offset of LAEs and IGM HI on the small scale.

4. The analysis of the average  $\tau_{\text{LoS}}$  profile centered at LAEs can trace the absorption signal down to a scale of 200 pkpc. We find a 30% excess at  $d < 200$  kpc, though only with three LoSs counted, indicating the statistical detection of the CGM signal around LAEs. We also detect a signal of 13% excess at 400 – 600 pkpc that is supposed to be in the IGM regime, supporting the IGM signal detection down to  $\sim 400$  pkpc.
5. The positive correlation indicates that, at  $z \sim 2$ , neither ionization nor supernova/inhabiting AGN feedback from LAEs are sufficient to erase the gravitational effects on galaxy–IGM HI correlation, or alternatively, the exceeding inflows keep supplying HI gas from a very large scale to the surrounding environment of galaxies.
6. By comparing our correlation scale with CCFs between Ly $\alpha$  absorbers and galaxies at  $z < 1$  (Ryan-Weber 2006; Tejos et al. 2014), we find that the redshift evolution of galaxy–IGM HI correlation may follow the evolution of galaxy clustering.
7. We also find a large scatter in the  $\delta_{\text{LAE}} - \tau_{\text{LoS}}$  correlation. Referring to the simulation in Mukae et al. (2017), outliers may be the cases that LoSs penetrate regions with specific morphological arrangement. In the high overdensity end, exceeding ionization and pre-heating process may be the reasons for the deficit of cold IGM HI, just like the  $z = 3.1$  protocluster in SSA22 field (Mawatari et al. 2017).

The project is still on-going for obtaining more LAEs in different fields and more LoSs in the overdense regions to strengthen the statistical robustness, so that we can compare the observables with simulations to tell the models of structure formation and evolution in terms of

IGM HI in the future. The upcoming Subaru/Prime Focus Spectroscopy (PFS) will be of high efficiency to make the spectroscopic confirmation for our LAE candidates, and also will provide us a good chance to perform IGM tomography in various environments, especially those with coherently distributed IGM HI and overdensities.

## ACKNOWLEDGMENTS

We thank the anonymous referee for the kind comments and suggestions, which have helped to improve this paper a lot. We also thank Drs. Ishikawa, S., Iwata, I., Kusakabe, H., Kakuma, R., Momose R., Nakanishi K., Onodera, M. and Ouchi, M. for their discussions and instructive advice on our works.

This work was supported in part by the Graduate University for Advanced Studies, SOKENDAI. This work is based on data collected at Subaru Telescope, which is operated by the National Astronomical Observatory of Japan. This paper makes use of software developed for the Large Synoptic Survey Telescope. We thank the LSST Project for making their code available as free software at <http://dm.lsst.org>.

Numerical simulations were carried out on the Cray XC50 at the Center for Computational Astrophysics, National Astronomical Observatory of Japan, and the OCTOPUS at the Cybermedia Center, Osaka University as part of the HPCI system Research Project (hp180063, hp190050). This work is supported in part by the JSPS KAKENHI Grant Number JP17H01111, 19H05810 (K.N.). KN acknowledges the travel support from the Kavli IPMU, World Premier Research Center Initiative (WPI), where part of this work was conducted.

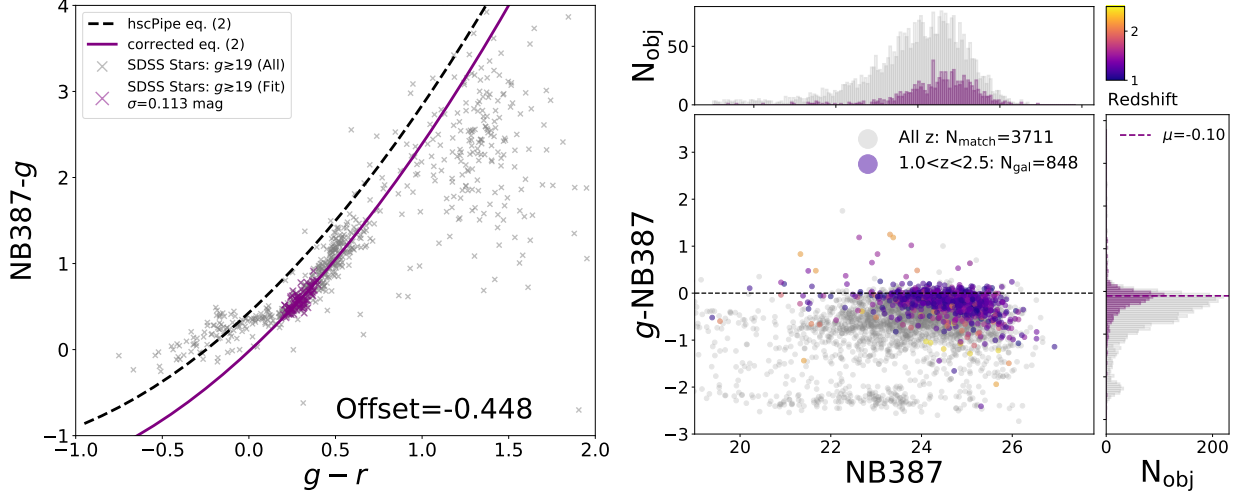
The authors wish to recognize and acknowledge the very significant cultural role and reverence that the summit of Maunakea has always had within the indigenous Hawaiian community. We are most fortunate to have the opportunity to conduct observations from this mountain.

*Facility:* Subaru (HSC)

## APPENDIX

### A. CORRECTION OF PHOTOMETRIC ZERO-POINT

We notice that there is a systematic offset in the Equation 2 for NB387, and we should introduce a constant  $C_{\text{metal}}$  for correction. Because the colors between NB387,  $g$  and  $r$  are influenced by the 4,000 Å *break*, which is sensitive to the metallicity (Kauffmann et al. 2003). The Pickles templates are mainly constructed from the stars with solar metallicity (Pickles 1998), while the number of star references used in the *hscPipe* has the peak around  $19 < g < 21$ , and so, tend to be the metal-poor halo stars that are more distant to us at high Galactic latitude. This difference may cause a systematic bias.



**Figure 12.** (Left) The predicted  $\text{NB387} - g$  vs.  $g - r$  diagram for homogeneously selected SDSS stars with  $g \gtrsim 19$ . The grey crosses are all the selected stars, and the purple ones are those with  $0.2 < g - r < 0.4$  after visual inspection, which are used for fitting the correction factor  $C_{\text{metal}}$ . The black dash curve is the Equation 2 from *hscPipe*, and the purple solid curve is the corrected relation with  $C_{\text{metal}} = -0.448$ . (Right) The  $g - \text{NB387}$  vs.  $\text{NB387}$  diagram for the  $2''$  cross-matches between the CHORUS objects (Inoue et al. 2020, submitted) and the DEIMOS 10K catalog (Hasinger et al. 2018). The grey dots are all the 3,711 matches with flag  $q > 1$ , suggesting a robust spectral redshift  $z_{\text{spec}}$  measurement, and the dots coded with the hot map are the 848 high- $z$  matches with  $1.0 < z_{\text{spec}} < 2.5$ , and the hotter means the higher redshift.

To estimate the  $C_{\text{metal}}$ , we homogeneously select the faint stars ( $g \gtrsim 19$ ) with  $\text{S/N} > 3$  spectra at the NB387 wavelength range from the SDSS database around the COSMOS field, whose Galactic latitude is comparable to our case. Then we calculate the predicted HSC/NB387, PS1/ $g$ -band and PS1/ $r$ -band magnitudes for these stars by taking their total transmission curves into accounts. These stars are plotted as the grey crosses in Figure 12. To keep consistency with the fitting in *hscPipe* and also to reduce the fitting uncertainty, we only use stars with  $0.2 < g - r < 0.4$ , which shows the smallest scatter in the relation. Most of the selected stars are flagged as the SEGUE targets in the SDSS (Yanny et al. 2009). For robust estimate, we perform visual inspection on each spectra of all these stars to discard those with weird features at the NB387 wavelength range. After this check, stars used for the zero-point correction is plotted as the purple crosses in Figure 12. We use these realistic stars, instead of the Pickles templates, to fit the relation shown in Equation 2 and the  $C_{\text{metal}}$  is estimated as  $-0.448$ . The original relation fit from *hscPipe* is shown as the black dash curve, and the corrected one is shown as the purple solid curve.

When fitting the Equation 2, the scatter of references is large in the case of NB387, making the fitting uncertainty as large as 0.2 mag and thus causing a field-to-field variation. We do the more subtle calibration for it by introducing another constant  $C_{\text{fit}}$ . We first select out the extended sources with  $23.5 < \text{NB387} < 24.5$ , which are most likely the high- $z$  galaxies that are free from the  $4,000 \text{ \AA}$  break in  $g$ -band, in each field. Then the field dependent  $C_{\text{fit}}$  is estimated by adjusting the  $g - \text{NB387}$  of these sources to  $-0.10$ , the expected mean color of  $1 < z < 3$  galaxies given their typical UV slope (Kurczynski et al. 2014).

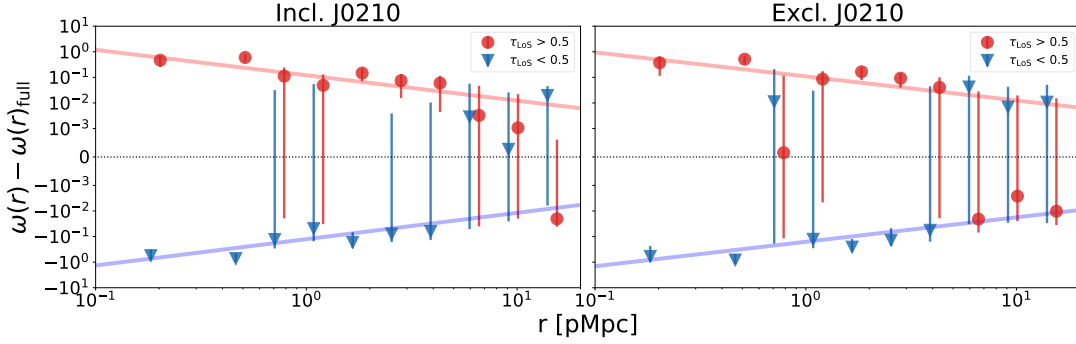
The  $g - \text{NB387} = -0.10$  can also be verified by utilizing the HSC/NB387 data from CHORUS survey (Inoue et al. 2020, submitted) and the spectral redshift  $z_{\text{spec}}$  from DEIMOS 10K spectroscopic survey catalog (Hasinger et al. 2018) in the COSMOS field. We first cross-match the CHORUS objects with the spectroscopic catalog within a  $2''$  aperture, and there are 3,711 matches with flag  $q > 1$  suggesting the good spectroscopic redshift measurement. The  $g - \text{NB387}$  vs.  $\text{NB387}$  with  $C_{\text{metal}}$  correction of these objects are plotted as the grey dots in the right panel of Figure 12. We pick out all the 848 high- $z$  galaxies with  $1.0 < z < 2.5$  from the matched catalog, which are coded by the heat map in the figure, to measure the mean of the  $g - \text{NB387}$  in a dual-Gaussian distribution, as the faint objects are likely in a flatten distribution due to photometric errors. The result for the main sequence peak is  $\mu = -0.10$ , being consistent with the expected color of high- $z$  galaxies. This consistency also validates the  $C_{\text{metal}}$  as the confident correction, and because the CHORUS NB387 data is observed in excellent conditions and has a plausible depth, it is reasonable to use the suggested value  $-0.10$  for calibrating the  $C_{\text{fit}}$  in each of our fields in this paper. The resulting



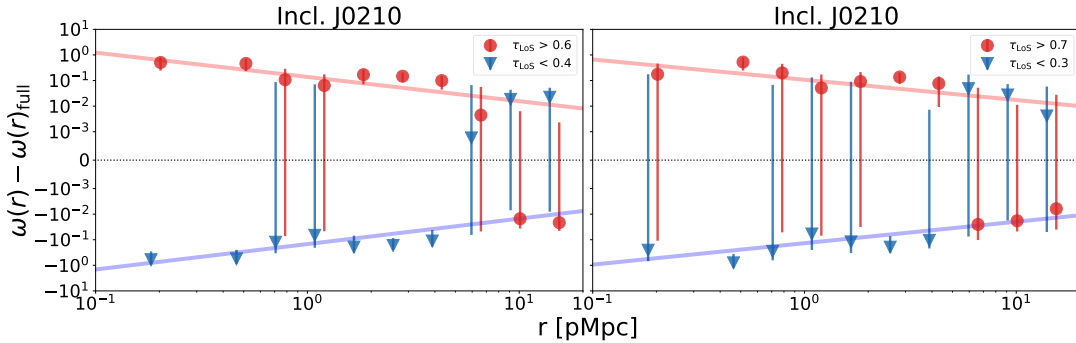
$C_{\text{fit}}$  fluctuates in  $-0.002 - 0.191$  mag among the four fields, which is consistent within the fitting uncertainty of 0.2 mag.

### B. TEST OF THE CCF RESULTS

The cross-correlation function (CCF) presented in Section 4.4 may show some variation by changing the sample size. Here, we first test the difference between cases including and excluding the field  $J0210$ . The results in log scale are shown in Figure 13, where the left panel shows results including  $J0210$  while the right one excludes it. From the comparison, no significant change in the results is found when we exclude  $J0210$ , except for one bin around 0.8 pMpc and generally larger errors, probably due to a smaller sample size. There is also not much variation in the clustering strength indicated by the  $r_0$ , which is summarize in Table 3.



**Figure 13.** Relative CCFs between LAEs and LoSs for the high  $\tau_{\text{LoS}}$ /low  $\tau_{\text{LoS}}$  subsamples in log scale for checking results for the cases including (*left panel*) and excluding (*right panel*)  $J0210$ , similar to the inset figures in the right panel of Figure 8. Bins are set in log scale with right boundary from 0.4 pMpc to 18.3 pMpc. Red points and curves are the  $\tau_{\text{LoS}} > 0.5$  subsample and corresponding fit power law model, while blue points and curves are the  $\tau_{\text{LoS}} < 0.5$  subsample. The fit parameters can be checked in Table 3, and they are not significantly changed between the two cases.



**Figure 14.** Relative CCFs between LAEs and LoSs for the high  $\tau_{\text{LoS}}$ /low  $\tau_{\text{LoS}}$  subsamples in log scale for checking results varying the subsample criteria. Symbols are similar to the Figure 13. Both figures are the results for the case including  $J0210$ . *Left panel:* for the subsamples  $\tau_{\text{LoS}} > 0.6/\tau_{\text{LoS}} < 0.4$ ; *right panel:* for the subsamples;  $\tau_{\text{LoS}} > 0.7/\tau_{\text{LoS}} < 0.3$ . The fit parameters can be checked in Table 3, and they do not significantly changed, either.

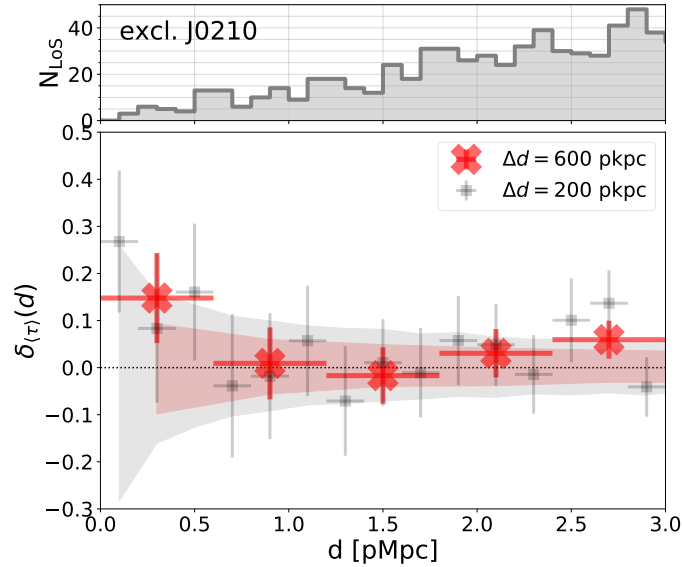
Because the definition of subsamples is based on a criterion, i.e.,  $\tau_{\text{LoS}}$  over/lower than 0.5, which is kind of arbitrary, we also test whether varying the criteria will change the result or not. We divide LoSs into other subsamples with  $\tau_{\text{LoS}} > 0.6/\tau_{\text{LoS}} < 0.4$ ,  $\tau_{\text{LoS}} > 0.7/\tau_{\text{LoS}} < 0.3$  respectively, to ensure sample size for each subsample is comparable as of 21/24 and 13/13. The results are shown in the Figure 14. When we compare the results with the one shown in the left panel of Figure 13, no significant changes can be found in the trend of CCFs, except for a larger uncertainty because of smaller sample size. The fitted  $r_0$  is summarized in Table 3, and they are still of the same order of  $\sim 0.1$  pMpc scale. These consistency prevents the galaxy–IGM HI correlation up to a scale of  $\sim 4$  pMpc at  $z = 2.2$  hinted in the CCF analysis from a coincidentally defined criterion.

Fields	$\tau_{\text{LoS}} >$	$N_{\text{LoS}}$	$\gamma$	$r_0$	$\tau_{\text{LoS}} <$	$N_{\text{LoS}}$	$\gamma$	$r_0$
[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Incl. <i>J0210</i>	0.5	30	$0.99^{+0.54}_{-0.17}$	$0.12^{+0.05}_{-0.03}$	0.5	34	$1.03^{+0.83}_{-0.21}$	$0.13^{+0.06}_{-0.02}$
Excl. <i>J0210</i>	0.5	23	$0.94^{+0.66}_{-0.16}$	$0.09^{+0.05}_{-0.03}$	0.5	19	$0.96^{+0.95}_{-0.19}$	$0.15^{+0.10}_{-0.05}$
Incl. <i>J0210</i>	0.6	21	$0.95^{+0.58}_{-0.13}$	$0.12^{+0.07}_{-0.03}$	0.4	24	$0.99^{+0.92}_{-0.16}$	$0.15^{+0.09}_{-0.04}$
Incl. <i>J0210</i>	0.7	13	$0.79^{+1.97}_{-0.11}$	$0.06^{+0.07}_{-0.06}$	0.3	13	$0.84^{+2.07}_{-0.15}$	$0.09^{+0.14}_{-0.09}$

**Table 3.** The parameters of CCF power law fitting for different subsamples. [1]: cases regarding field *J0210*; [2]: high  $\tau_{\text{LoS}}$  criterion; [3]: number of LoSs in the high  $\tau_{\text{LoS}}$  subsample; [4]:  $\gamma$  fit for high  $\tau_{\text{LoS}}$  subsample; [5]:  $r_0$  fit for high  $\tau_{\text{LoS}}$  subsample; [6]: low  $\tau_{\text{LoS}}$  criterion; [7]: number of LoSs in the low  $\tau_{\text{LoS}}$  subsample; [8]:  $\gamma$  fit for low  $\tau_{\text{LoS}}$  subsample; [9]:  $r_0$  fit for low  $\tau_{\text{LoS}}$  subsample;

### C. AVERAGE OPTICAL DEPTH PROFILE EXCLUDING *J0210*

In Section 4.5, given the importance of the LoS number for the statistics when inspecting small scales, we mainly discuss the case with *J0210*, which contains a large filament with a group of quasars associated. Here, we show the result for the case excluding *J0210*, and we do not find a significant change on the general  $\langle \tau \rangle$  varying trend along the distance to LAEs at the inner region that is discussed in Section 4.5, although the scatter is larger due to a smaller number of LoSs. This supports our assumption that the statistics, such as the  $\langle \tau \rangle$ , is unlikely to be affected by the six outliers out of 64 LoSs. We note that two finer bins at  $\sim 2.7$  pMpc show the tentative excess, more significant than the case including *J0210*, although the coarse bin still shows a weak signal.



**Figure 15.** The fluctuation of the average  $\tau_{\text{LoS}}$  as a function of distance to LAEs,  $\delta_{\langle \tau \rangle}(d)$ , for the case excluding *J0210*. The symbols are the same with Figure 9. The *J0210* does not alter the general trend.

### 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
- Aihara, H., AlSayyad, Y., Ando, M., et al. 2019, arXiv e-prints, arXiv:1905.12221
- Ando, R., Nishizawa, A. J., Hasegawa, K., Shimizu, I., & Nagamine, K. 2019, *MNRAS*, 484, 5389
- Baugh, C. M. 2006, *Reports on Progress in Physics*, 69, 3101
- Baugh, C. M., Benson, A. J., Cole, S., Frenk, C. S., & Lacey, C. G. 1999, *MNRAS*, 305, L21

- Becker, G. D., Hewett, P. C., Worseck, G., & Prochaska, J. X. 2013, *MNRAS*, **430**, 2067
- Bertin, E., & Arnouts, S. 1996, *A&AS*, **117**, 393
- Bond, J. R., Kofman, L., & Pogosyan, D. 1996, *Nature*, **380**, 603
- Bosch, J., Armstrong, R., Bickerton, S., et al. 2018, *PASJ*, **70**, S5
- Bădescu, T., Yang, Y., Bertoldi, F., et al. 2017, *ApJ*, **845**, 172
- Cai, Z., Fan, X., Peirani, S., et al. 2016, *ApJ*, **833**, 135
- Cai, Z., Fan, X., Yang, Y., et al. 2017a, *ApJ*, **837**, 71
- Cai, Z., Fan, X., Bian, F., et al. 2017b, *ApJ*, **839**, 131
- Capellupo, D. M., Hamann, F., Herbst, H., et al. 2017, *MNRAS*, **469**, 323
- Casey, C. M., Cooray, A., Capak, P., et al. 2015, *ApJ*, **808**, L33
- Cen, R., & Ostriker, J. P. 2000, *ApJ*, **538**, 83
- Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, arXiv e-prints, arXiv:1612.05560
- Chen, H.-W., & Mulchaey, J. S. 2009, *ApJ*, **701**, 1219
- Chen, Y., Steidel, C. C., Hummels, C. B., et al. 2020, arXiv e-prints, arXiv:2006.13236
- Chiang, Y.-K., Overzier, R., & Gebhardt, K. 2013, *ApJ*, **779**, 127
- Chiang, Y.-K., Overzier, R. A., Gebhardt, K., & Henriques, B. 2017, *ApJ*, **844**, L23
- Ciardullo, R., Gronwall, C., Adams, J. J., et al. 2013, *ApJ*, **769**, 83
- Cooke, E. A., Hatch, N. A., Muldrew, S. I., Rigby, E. E., & Kurk, J. D. 2014, *MNRAS*, **440**, 3262
- Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, *AJ*, **145**, 10
- Dawson, K. S., Kneib, J.-P., Percival, W. J., et al. 2016, *AJ*, **151**, 44
- Dekel, A., & Birnboim, Y. 2006, *MNRAS*, **368**, 2
- Faucher-Giguère, C.-A., Prochaska, J. X., Lidz, A., Hernquist, L., & Zaldarriaga, M. 2008, *ApJ*, **681**, 831
- Fumagalli, M., Mackenzie, R., Trayford, J., et al. 2017, *MNRAS*, **471**, 3686
- Gehrels, N. 1986, *ApJ*, **303**, 336
- Guaita, L., Gawiser, E., Padilla, N., et al. 2010, *ApJ*, **714**, 255
- Hasinger, G., Capak, P., Salvato, M., et al. 2018, *ApJ*, **858**, 77
- Hathi, N. P., Le Fèvre, O., Ilbert, O., et al. 2016, *A&A*, **588**, A26
- Hayashino, T., Inoue, A. K., Kousai, K., et al. 2019, *MNRAS*, **484**, 5868
- Hennawi, J. F., Prochaska, J. X., Cantalupo, S., & Arrighi-Battaia, F. 2015, *Science*, **348**, 779
- Hernquist, L., Katz, N., Weinberg, D. H., & Miralda-Escudé, J. 1996, *ApJL*, **457**, L51
- Hinshaw, G., Nolta, M. R., Bennett, C. L., et al. 2007, *ApJS*, **170**, 288
- Inoue, A. K., Shimizu, I., Iwata, I., & Tanaka, M. 2014, *MNRAS*, **442**, 1805
- Kauffmann, G., Heckman, T. M., White, S. D. M., et al. 2003, *MNRAS*, **341**, 33
- Kikuta, S., Matsuda, Y., Cen, R., et al. 2019, *PASJ*, **71**, L2
- Konno, A., Ouchi, M., Nakajima, K., et al. 2016, *ApJ*, **823**, 20
- Kurczynski, P., Gawiser, E., Rafelski, M., et al. 2014, *ApJL*, **793**, L5
- Kusakabe, H., Shimasaku, K., Ouchi, M., et al. 2018, *PASJ*, **70**, 4
- Kusakabe, H., Shimasaku, K., Momose, R., et al. 2019, *PASJ*, **71**, 55
- Landy, S. D., & Szalay, A. S. 1993, *ApJ*, **412**, 64
- Lee, K.-G., Hennawi, J. F., White, M., Croft, R. A. C., & Ozbek, M. 2014a, *ApJ*, **788**, 49
- Lee, K.-G., Suzuki, N., & Spergel, D. N. 2012, *AJ*, **143**, 51
- Lee, K.-G., Bailey, S., Bartsch, L. E., et al. 2013, *AJ*, **145**, 69
- Lee, K.-G., Hennawi, J. F., Stark, C., et al. 2014b, *ApJ*, **795**, L12
- Lee, K.-G., Hennawi, J. F., White, M., et al. 2016, *ApJ*, **817**, 160
- Lee, K.-G., Krolewski, A., White, M., et al. 2018, *The Astrophysical Journal Supplement Series*, **237**, 31
- Mackenzie, R., Fumagalli, M., Theuns, T., et al. 2019, *Monthly Notices of the Royal Astronomical Society*
- Madau, P., & Dickinson, M. 2014, *ARA&A*, **52**, 415
- Mawatari, K., Yamada, T., Nakamura, Y., Hayashino, T., & Matsuda, Y. 2012, *ApJ*, **759**, 133
- Mawatari, K., Inoue, A. K., Yamada, T., et al. 2017, *MNRAS*, **467**, 3951
- Miller, J. S. A., Bolton, J. S., & Hatch, N. 2019, *MNRAS*, **489**, 5381
- Miralda-Escudé, J., Cen, R., Ostriker, J. P., & Rauch, M. 1996, *ApJ*, **471**, 582
- Miyazaki, S., Komiyama, Y., Kawanomoto, S., et al. 2018, *PASJ*, **70**, S1
- Momose, R., Shimizu, I., Nagamine, K., et al. 2020a, Environmental dependence of galactic properties traced by Ly forest absorption (I): variation according to galaxy stellar mass and star-formation activity, arXiv:2002.07334 [astro-ph.GA]

- Momose, R., Shimasaku, K., Kashikawa, N., et al. 2020b, Environmental dependence of galactic properties traced by Ly forest absorption (II): diversity among galaxy populations, [arXiv:2002.07335 \[astro-ph.GA\]](#)
- Mukae, S., Ouchi, M., Kakiichi, K., et al. 2017, *ApJ*, **835**, 281
- Mukae, S., Ouchi, M., Cai, Z., et al. 2019, arXiv e-prints, [arXiv:1910.02962](#)
- Nagamine, K., Shimizu, I., Fujita, K., et al. 2020, arXiv e-prints, [arXiv:2007.14253](#)
- Nakajima, K., Ouchi, M., Shimasaku, K., et al. 2012, *ApJ*, **745**, 12
- Newman, A. B., Rudie, G. C., Blanc, G. A., et al. 2020, *ApJ*, **891**, 147
- Noirrot, G., Stern, D., Mei, S., et al. 2018, *ApJ*, **859**, 38
- Norberg, P., Baugh, C. M., Gaztañaga, E., & Croton, D. J. 2009, *MNRAS*, **396**, 19
- Ogura, K., Nagao, T., Imanishi, M., et al. 2017, *PASJ*, **69**, 51
- Onoue, M., Kashikawa, N., Uchiyama, H., et al. 2018, *PASJ*, **70**, S31
- Pickles, A. J. 1998, *PASP*, **110**, 863
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, *A&A*, **594**, A13
- Polletta, M., Tajer, M., Maraschi, L., et al. 2007, *ApJ*, **663**, 81
- Prochaska, J. X., Hennawi, J. F., Lee, K.-G., et al. 2013, *ApJ*, **776**, 136
- Rakic, O., Schaye, J., Steidel, C. C., & Rudie, G. C. 2012, *ApJ*, **751**, 94
- Rudie, G. C., Steidel, C. C., Trainor, R. F., et al. 2012, *ApJ*, **750**, 67
- Ryan-Weber, E. V. 2006, *MNRAS*, **367**, 1251
- Santos, S., Sobral, D., Matthee, J., et al. 2020, *MNRAS*, **493**, 141
- Schlafly, E. F., & Finkbeiner, D. P. 2011, *ApJ*, **737**, 103
- Shi, K., Huang, Y., Lee, K.-S., et al. 2019, *ApJ*, **879**, 9
- Shibuya, T., Ouchi, M., Konno, A., et al. 2018, *PASJ*, **70**, S14
- Shimakawa, R., Kodama, T., Tadaki, K. I., et al. 2014, *MNRAS*, **441**, L1
- Shimakawa, R., Kodama, T., Hayashi, M., et al. 2017, *MNRAS*, **468**, L21
- Shimizu, I., Todoroki, K., Yajima, H., & Nagamine, K. 2019, *MNRAS*, **484**, 2632
- Smolčić, V., Ivezić, Ž., Knapp, G. R., et al. 2004, *ApJL*, **615**, L141
- Sobral, D., Matthee, J., Best, P., et al. 2017, *MNRAS*, **466**, 1242
- Spitler, L. R., Labbé, I., Glazebrook, K., et al. 2012, *The Astrophysical Journal Letters*, **748**, L21
- Springel, V. 2005, *MNRAS*, **364**, 1105
- Springel, V., Frenk, C. S., & White, S. D. M. 2006, *Nature*, **440**, 1137
- Springel, V., Pakmor, R., Pillepich, A., et al. 2018, *MNRAS*, **475**, 676
- Steidel, C. C., Erb, D. K., Shapley, A. E., et al. 2010, *ApJ*, **717**, 289
- Tejos, N., Morris, S. L., Finn, C. W., et al. 2014, *MNRAS*, **437**, 2017
- Toshikawa, J., Kashikawa, N., Overzier, R., et al. 2016, *ApJ*, **826**, 114
- Toshikawa, J., Uchiyama, H., Kashikawa, N., et al. 2018, *PASJ*, **70**, S12
- Tumlinson, J., Peebles, M. S., & Werk, J. K. 2017, *ARA&A*, **55**, 389
- Turner, M. L., Schaye, J., Crain, R. A., et al. 2017, *MNRAS*, **471**, 690
- Turner, M. L., Schaye, J., Steidel, C. C., Rudie, G. C., & Strom, A. L. 2014, *MNRAS*, **445**, 794
- Yanny, B., Rockosi, C., Newberg, H. J., et al. 2009, *AJ*, **137**, 4377
- Zheng, Z.-Y., Malhotra, S., Rhoads, J. E., et al. 2016, *ApJS*, **226**, 23