The Astrophysical Journal, 907:3 (21pp), 2021 January 20 https://doi.org/10.3847/1538-4357/abcd93
© 2021. The American Astronomical Society. All rights reserved.

Statistical Correlation between the Distribution of Lyα Emitters and Intergalactic Medium H I at $z \sim 2.2$ Mapped by the Subaru/Hyper Suprime-Cam

Yongming Liang[1,2,3], Nobunari Kashikawa[2,3,4], Zheng Cai[5], Xiaohui Fan[6], J. Xavier Prochaska[7], Kazuhiro Shimasaku[3,4], Masayuki Tanaka[1,2], Hisakazu Uchiyama[2], Kei Ito[1,2,3], Rhythm Shimakawa[2], Kentaro Nagamine[8,9,10], Ikkoh Shimizu[2,11], Masafusa Onoue[12], and Jun Toshikawa[13,14]
[1] Department of Astronomical Science, SOKENDAI (The Graduate University for Advanced Studies), Mitaka, Tokyo 181-8588, Japan; ym.liang@grad.nao.ac.jp
[2] National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan
[3] Department of Astronomy, School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
[4] Research Center for the Early Universe, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
[5] Department of Astronomy, Tsinghua University, Beijing 100084, People's Republic of China; zcai@mail.tsinghua.edu.cn
[6] Steward Observatory, University of Arizona, 933 N. Cherry Avenue, Tucson, AZ 85721, USA
[7] UCO/Lick Observatory, University of California, 1156 High Street, Santa Cruz, CA 95064, USA
[8] Department of Earth and Space Science, Osaka University, 1-1 Machikaneyama, Toyonaka, Osaka 560-0043, Japan
[9] Kavli IPMU (WPI), The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8583, Japan
[10] Department of Physics and Astronomy, University of Nevada, Las Vegas, 4505 S. Maryland Parkway, Las Vegas, NV 89154-4002, USA
[11] Shikoku Gakuin University, 3-2-1 Bunkyocho, Zentsuji, Kagawa, 765-8505, Japan
[12] Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany
[13] Institute for Cosmic Ray Research, The University of Tokyo, Kashiwa, Chiba 277-8582, Japan
[14] Department of Physics, University of Bath, Claverton Down, Bath, BA2 7AY, UK
Received 2020 April 20; revised 2020 October 29; accepted 2020 November 23; published 2021 January 19

Abstract

The correlation between neutral hydrogen (H I) in the intergalactic medium (IGM) and galaxies attracts great interest. We select four fields that include several coherently strong Lyα absorption systems at $z \sim 2.2$ detected using background quasars from the whole SDSS/(e) Baryon Oscillation Spectroscopic Survey (BOSS) database. Deep narrowband and g-band imaging are performed using the Hyper Suprime-Cam on the Subaru Telescope. We select 2642 Lyα emitter (LAE) candidates at $z = 2.177 \pm 0.023$ down to the Lyα luminosity of $L_{\mathrm{Ly}\alpha} \approx 2 \times 10^{42}$ erg s^{-1} to construct the galaxy overdensity maps, covering an effective area of 5.39 deg^2. Combining the sample with the Lyα absorption estimated from 64 (e) BOSS quasar spectra, we find a moderate to strong correlation between the LAE overdensity δ_{LAE} and the effective optical depth τ_{LoS} in lines of sight, with P value $= 0.09\%$ or $<0.01\%$ when the field containing a significant quasar overdensity is included or excluded. Cross-correlation analysis also clearly suggests that up to 4 ± 1 pMpc, LAEs tend to cluster in regions rich in H I gas, as indicated by the high τ_{LoS}, and avoid the low τ_{LoS} regions where the H I gas is deficient. By averaging the τ_{LoS} as a function of the projected distance (d) to LAEs, we find a 30% excess signal at 2σ level at $d < 200$ pkpc, indicating the dense H I in the circumgalactic medium, and a tentative excess at $400 < d < 600$ pkpc in the IGM regime, corroborating the cross-correlation signal detected at about 0.5 pMpc. These statistical analyses suggest that galaxy−IGM H I correlations exist on scales ranging from several hundred pkpc to several pMpc at $z \sim 2.2$.

Unified Astronomy Thesaurus concepts: Large-scale structure of the universe (902); Intergalactic medium (813); Two-point correlation function (1951); Galaxy formation (595); Lyman-alpha galaxies (978)

1. Introduction

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 overdense regions where the neutral hydrogen (H I) in the intergalactic medium (IGM) is more abundant (Baugh 2006; Springel et al. 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 galaxies and IGM H I.

However, it is not easy to find the overdense regions at $z > 2$, which occupy only a small fraction of the cosmic volume, e.g., <2% for protoclusters (Chiang et al. 2017). To perform efficient surveys for galaxies, some studies have used galaxies with radio-loud active galactic nuclei (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. Such rare objects are expected to reside in massive halos that are likely to host protoclusters. Damped Lyα systems (DLAs) (Fumagalli et al. 2017; Ogura et al. 2017) or systems with extended nebular emission around galaxies (Bădescu et al. 2017) are also good candidates as tracers. 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, in decades of simulations, LSSs in terms of IGM H I have also been demonstrated to be possibly revealed by absorption imprinted in the spectra of background quasars (Hernquist et al. 1996; Springel et al. 2006), and it has also been proven to be a nontrivial question in the high-z universe, as most baryons at $z > 2$ may reside in Lyα clouds (Miralda-Escudé et al. 1996). Strong H I absorbers have been studied around quasars (Prochaska et al. 2013) or by searching the associated galaxies (Mackenzie et al. 2019), from which hints of a galaxy–IGM H I correlation can be found.

Based on a specific field SSA 22 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 studied the same structure, found a similar correlation in redshift space.

Galaxy–IGM H I correlations have also been studied in a statistical manner 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. 2020a; Chen et al. 2020), and most of these studies found correlations on various scales. However, these studies were confined either by bright galaxy populations or a relatively small dynamic range of the IGM absorption due to the limited sample sizes and survey areas.

Recently, IGM tomography has also become a feasible method to construct 3D IGM H I maps from background star-forming galaxies (Lee et al. 2014a, 2014b, 2016, 2018; Newman et al. 2020). However, the tomography surveys performed to date have still been limited by the survey area $\lesssim$1 deg^2, and most were conducted on general fields. Larger survey areas covering various overdense regions are essential to take full advantage of the technique.

The MApping the Most Massive Overdensity Through Hydrogen project (MAMMOTH; Cai et al. 2016, 2017a, 2017b) was proposed to search for massive overdensities using coherently strong Lyα absorption systems (CoSLAs; see detailed definition in Section 2.5 in Cai et al. 2016), which originate from overlapping of the Lyα forest. N-body simulations imply that CoSLAs can effectively trace the most massive halos on scales over 15 h^{-1} cMpc. Although it is a matter of debate whether CoSLAs can trace the most massive overdensities (Miller et al. 2019), a pilot MAMMOTH program found BOSS1441, one of the most massive structures to date at $z > 2$, together with six associated BOSS quasars (Cai et al. 2017a).

While helping to pinpoint regions that tend to host overdensities, the grouping of rare lines of sight (LoSs) with high IGM absorption (e.g., CoSLAs) also significantly enhances the dynamic range of statistics for studying the galaxy–IGM H I correlation. Targeting fields centered on MAMMOTH candidates on 15 h^{-1} cMpc, the Subaru/Hyper Suprime-Cam (HSC; Miyazaki et al. 2018), equipped with a diameter $d = 1.^\circ5$ field of view (FoV), makes it possible to map the most diverse universe efficiently at $z \sim 2$ on a scale of over 100 cMpc. In addition, the narrowband technique for identifying Lyα emitters (LAEs), the redshifts of which can be well constrained in a narrow range ($\Delta z < 0.05$), also opens a window toward a fainter and less massive galaxy population for the correlation.

Here, we first summarize the SDSS/(e) BOSS data, the field selection, the Subaru/HSC observations, and the data processing in Section 2. LAE sample construction is then presented in Section 3. Section 4 shows our LAE overdensity maps for the four HSC fields. Analyses of the galaxy–IGM H I correlation are also provided in this section. Section 5 compares our results with those of other studies and explores the scale dependence of correlations. The underlying physics is also discussed in the last section. Finally, we end with a summary and discuss perspectives for future work in Section 6. The cosmological parameters used in this paper are based on Planck Collaboration et al. (2016): $H_0 = 67.7$ km Mpc^{-1} s^{-1} and $\Omega_0 = 0.307$. 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 (eBOSS) of SDSS-IV (Dawson et al. 2016) are used in this study both to select candidate fields and to evaluate the effective optical depths in correlation analysis. BOSS is a spectroscopic survey that was specifically designed to study intergalactic science through the Lyα forest. Using the 2.5 m Sloan telescope, it takes spectra of more than 150,000 background quasars at $z \gtrsim 2.15$, reaching a depth as faint as $g < 22$. The eBOSS observes 60,000 BOSS quasars for better spectra, complemented with 60,000 new targets. The combined surveys include more than 200,000 quasar spectra covering a survey area of over 10,000 deg^2, corresponding to a survey volume of >1 Gpc^3.

The (e) BOSS database provides abundant quasar spectra as LoSs, in which the IGM distribution can be traced by Lyα absorption. To evaluate Lyα absorption, we calculate the effective optical depth in the LoS, $\tau_{\rm LoS}$, within the Lyα redshift range traced by the narrowband filter NB387 ($\lambda_0 = 3862$ Å, FWHM $= 56$ Å).

We first smooth the flux along the wavelength dimension over a scale of 15 h^{-1} cMpc. Absorption features are searched by scanning through the spectra over a range of ±35 Å centered around 3862 Å. The effective optical depth is then calculated at the strongest absorption spike following Cai et al. (2016):

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

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

When evaluating the Lyα absorption by using the (e) BOSS spectra, the quasar continuum is constructed by applying the mean-flux-regulated principal component analysis (MF-PCA) technique to the fitting (Lee et al. 2013). The continuum is first fitted from the redward side of the Lyα emission line by using traditional PCA (Suzuki et al. 2005). Additional constraints on the slope and amplitude of the continuum are adjusted by using the mean optical depth of the Lyα forest (Lee et al. 2012; Becker et al. 2013).

The $\tau_{\rm LoS}$ are used throughout this paper for both field selection and galaxy–Lyα absorption correlation analysis.

2.2. Field Selection

This study is performed to examine galaxy–IGM H I correlations in a wide range of environments based on less massive galaxy populations. The principle for our field selection is to enclose a sufficient number of LoSs, especially those with strong Lyα absorptions, while we also target potential overdensities.

First, we briefly summarize the selection of CoSLAs, but please see Cai et al. (2016) for details. LoSs with $\tau_{\rm LoS} \gtrsim 3\langle\tau\rangle_{\rm cos}$ are chosen as the preliminary absorber candidates, where $\langle\tau\rangle_{\rm cos}$ is the cosmic mean optical depth, and we assume a value of 0.15 at $z = 2.2$ (Becker et al. 2013) with slight adjustments according to fields. To eliminate the non-IGM contaminants, we perform systematic inspection of the

Table 1
Summary of Field Information

Field	R.A. (J2000)	Decl. (J2000)	Obs Period	$\mathrm{FWHM_{PSF,NB}}$	$\mathrm{FWMH_{PSF,}}_{g}$	$m_{\mathrm{NB},5\sigma}$	$m_{g,5\sigma}$	*hscPipe*
BOSS J0210+0052	02:09:58.90	+00:53:43.0	Jan, 2018	1.″22	0.″90	24.36[a]	26.24[a]	v5.4
				1.″22	0.″90	24.25[a]	26.34[a]	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

Notes. 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 the observations; columns 5 and 6 are the FWHMs of star PSFs measured using the final stacked images of NB387 and the g band; columns 7 and 8 are the 5σ limiting magnitudes measured in an aperture with the radius of 1.″7 using the final stacked NB387 images and the PSF-matched g-band, respectively; and column 9 is the hscPipe version used for the data reduction.
[a] Measured in a 2.5″ aperture, and the g band is the PSF-matched image.

Table 2
Information on the LoSs and LAEs in Each Field

Field	$N_{\mathrm{LoS,All}}$	$N_{\mathrm{LoS,Ana}}$	N_{LAE}	Area [deg^2]	$E(B-V)$
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	\

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

criteria proposed in Cai et al. (2016) to reject the high-column-density systems, i.e., DLAs, sub-DLAs, or Lyman-limit systems (LLS). We also perform visual checks to remove broad absorption-line (BAL) quasars that can confuse the interpretation of IGM Lyα absorption in the NB387 wavelength range. Besides, for the high-τ_{LoS} LoSs, all of these processes for excluding non-IGM contaminants are also performed for the potential LoSs used in the following analysis, which we refer to here as clean LoSs.

Based on clean LoSs, the target fields of HSC FoV should (1) contain high-τ_{LoS} LoSs to expand the dynamic range; (2) enclose as many LoSs as possible to increase the sample size for inferring the galaxy–IGM H I relation; (3) contain a concentration of high-τ_{LoS} LoSs to find protoclusters, i.e., $\gtrsim$4 LoSs within a $(\sim 20\,h^{-1}\mathrm{cMpc})^3$ box, which is the typical scale of a Coma-type protocluster at $z \sim 2$ (Chiang et al. 2013); and (4) in special cases, contain associated quasars at $2.15 \leqslant z \leqslant 2.20$, i.e., proximity quasars at a similar redshift to our LAEs, to see any possible difference.

Our field selections are, however, further compromised by the field visibility and the distance to the Moon or the nearby bright stars in a specific observation run. Four fields—BOSS *J*0210+0052 (or *J*0210), BOSS *J*0222−0224 (or *J*0222), BOSS *J*0924+1503 (or *J*0924), and BOSS *J*1419+0500 (or *J*1419)—in our observations are selected, all of which satisfy (1) and most of which satisfy (2). *J*0222 and *J*0924 are selected based mainly on (3), the typical regions hinting at coherent IGM H I on a large scale. *J*1419 is initially selected for (3), but one of the two concentrating LoSs is found to be a possible BAL quasar[15] after observation and is excluded from the analysis. However, the field is still one of the best candidates considering (1) and (2), although coherent IGM Lyα absorption is not as significant as in other selected fields. *J*0210 is selected with consideration of (4), given that a group of 11 proximity quasars is associated within a region of 40×40 cMpc^2 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. (2017a). One of the proximity quasars also shows hints of strong IGM Lyα absorption at the wing of Lyα emission, but being conservative, it is not included in our correlation analysis. The coordinates of the field centers are listed in the Table 1.

We note that before applying the mask in the following sections, there are 26, 23, 19, and 22 clean LoSs in *J*0210, *J*0222, *J*0924, and *J*1419, respectively, which are summarized in Table 2. The τ_{LoS} distribution of these clean LoSs is shown in Figure 1, in which a dashed blue vertical line indicates the criterion for the clean LoSs with $\tau_{\mathrm{LoS}} \gtrsim 3\langle\tau\rangle_{\cos}$.

2.3. Imaging Observations

Observations to identify LAEs were carried out using the HSC installed at the prime focus of the 8.2 m Subaru telescope located at the summit of Maunakea, Hawaii. HSC is a high-performance camera with a wide FoV of 1.°5 in diameter. As a gigantic mosaic CCD camera, HSC consists of 104 Hamamatsu Photonics KK CCDs (2048×4096 pixels) for science, 4 for auto-guiders and 8 for focusing. The pixel scale of the CCD reaches 0.″168.

Here, we perform deep NB imaging using NB387, which enables detection of Lyα 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

[15] This is J141934.64+050327.1, which is categorized as a probable Pv BAL quasar in Capellupo et al. (2017).

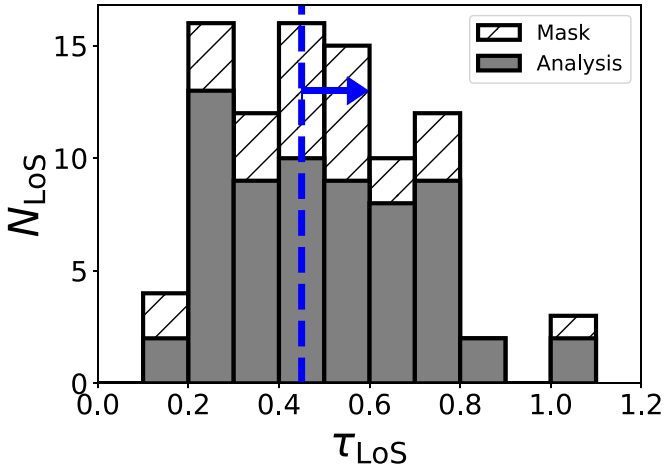

Figure 1. The $\tau_{\rm LoS}$ distribution of all inspected clean LoSs. The hatched area indicates the masked LoSs, while the filled area shows the LoSs used in the following correlation analysis. The dashed blue vertical line suggests the $\tau_{\rm LoS}$ criterion of the clean LoSs with $\tau_{\rm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

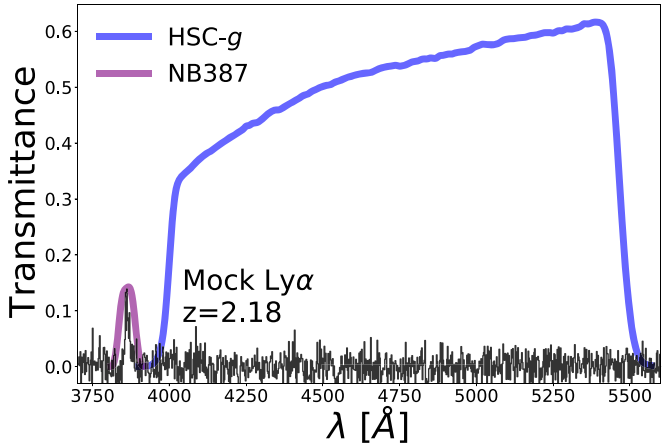

Figure 2. Transmission curve of HSC-g and the NB387 band. The solid purple and blue lines are the total transmittance of NB387 and HSC-g based on the CCD quantum efficiency, dewar window, the primary focus unit, and the reflectivity of the prime mirror. The black curve indicates a mock LAE spectrum at $z = 2.18$, the Lyα emission of which is precisely located at the sensitive wavelength range of NB387.

transmission curves of the filters, which take transmittance, CCD quantum efficiency, the dewar window, the primary focus unit, and the reflectivity of the prime mirror into account, are shown in Figure 2.

To ensure the depth needed to detect a sufficient number of LAEs, the observation is designed to have total exposures of 3 hours for NB387 and 40 minutes for the g band in each field. An S17B observation was carried out in queue mode in 2018 and 2019 January, and exposures were split into 900 and 600 s for the NB387 and g-band observations, respectively, except for the first two exposures of 1200 s for *J*0210. In the S19A observations carried out on site on 2019 March 8th, the exposures were split into 900 s and 300 s for the NB387 and g-band observations, respectively, to avoid saturation in the broad band, which can affect the selection completeness of LAEs around bright stars. However, saturation can still occur around the brightest stars even with such a shorter exposure time; therefore, we mask such regions in the photometric processing, as described in Section 2.5. From S17B, we obtain both the NB387 and g-band data for fields *J*0210 and *J*0222 and the g-band data for *J*0924. In the S19A run, NB387 data for both *J*0924 and *J*1419 are obtained, and the g-band imaging is taken for *J*1419 only. In summary, both the NB387 and g-band imaging data for all four fields are from two major runs.

With the exception of the NB imaging of *J*0210 under relatively poor conditions with seeing over 1.″2, all of the observations were performed under moderate to good conditions. Some exposures are discarded because of the occasionally poor seeing or low transparency. In the case of *J*0222, severe stray light from a nearby Mira southwest of the pointing with magnitude $g = 5.1$ contaminated some exposures because of exposure dithers, and these exposures 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 Section 2.4. Detailed information for each field is summarized in the Table 1.

2.4. Data Reduction

The NB387 and g-band imaging data are reduced using the HSC pipeline, *hscPipe* (Bosch et al. 2018; Aihara et al. 2019). *J*0210 and *J*0222 are reduced using *hscPipe* 5.4, and *J*0924 and *J*1419 are reduced using *hscPipe* 6.6. Given the relatively poor quality of the *J*0210 NB387 data, we also reduce both the g-band and the NB387 data for *J*0210 using *hscPipe* 6.6 and combine the catalog with that from *hscPipe* 5.4. The overlaid detections with separations smaller than 2″ are kept with only the latter version. Bosch et al. (2018) and Aihara et al. (2019) describe the data reduction process as well as the code updates in detail, but we present a brief summary here with emphasis on the processes that differ from the standard usage.

The *hscPipe* first creates calibration data, including the bias, dark, dome-flat, and global sky. Then, it applies them to each CCD in a single visit, and the local sky background on the 128 pixel scale is subtracted. Bright objects are then extracted for astrometric and photometric calibration. The point-spread function (PSF) models used within the pipeline are also created at this step. Astrometry and photometry are then calibrated against PS1 references. For each filter, the zero point is adjusted by fitting a multiband relation, which is derived from the template magnitudes predicted by spectroscopic Pickles star references (Pickles 1998) and filter transmissions, e.g., an NB $-$ g versus $g - r$ relation for the NB387:

$$\begin{aligned} {\rm NB387_{HSC}} - g_{\rm PS1} = & \ 0.541 \times (g_{\rm PS1} - r_{\rm PS1})^2 \\ & + 1.87 \times (g_{\rm PS1} - r_{\rm PS1}) \\ & + 0.428\,[+C_{\rm metal} + C_{\rm fit}], \end{aligned} \quad (2)$$

where $\rm NB387_{HSC}$ is the magnitude of NB387 based on HSC observations, and $g_{\rm PS1}$ and $r_{\rm PS1}$ are the g-band and r-band magnitudes, respectively, from the PS1 catalog. $C_{\rm metal}$ and $C_{\rm fit}$ are the correction factors for the original relation in *hscPipe*, as described below. A tract is defined as a large patch enclosing all observed sky, and then a sky map is made as the reference for the following coadding process. The global sky background is subtracted without masked regions. In the mosaicking of the CCD data, both the WCS and the flux scale are corrected by a spatially varying correction term. Finally, the coadding process warps the images to the sky map and coadds all visits of the image together, scaled with the WCS and flux correction.

For our data, it is necessary to optimize some configurations further in addition to the aforementioned process. The NB387 image in a single frame was generally too shallow to include enough bright stars in each CCD for calibration. Therefore, we lower the parameter set used to choose calibration stars by $\sim 0.5\times$ the default value. In addition, when fitting Equation (2) to determine the photometric zero point of NB387, we take into account additional corrections, including a systematic correction $C_{\rm metal}$ of -0.448 mag to correct stellar metallicity and/or stellar age biases, and a field-dependent term, $C_{\rm fit}$, ranging within 0.2 mag to calibrate the fitting uncertainties. Details are described in the Appendix A.

2.5. Photometric Processing

We use SExtractor 2.19.5 (Bertin & Arnouts 1996) for the photometry processing. First, we perform 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 source detection and measurement by setting the NB387 image as the reference. The detection threshold is set as 15 continuous pixels over the 1.2σ sky background. Due to the large HSC FoV and the mosaic CCD structure, there are slight fluctuations of 0.1–0.2 mag in the image depth within 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. Masking of regions with low signal-to-noise ratio (S/N) signals, saturation around bright stars, or severe stray light is also applied in background estimation, object detection, and photometric measurement.

Note that after masking, the final numbers of clean LoSs in *J*0210, *J*0222, *J*0924, and *J*1419 are 22, 11, 14, and 17, respectively, as summarized in Table 2. The masked clean LoSs are hatched in Figure 1, and the remaining 64 clean LoSs are used for all of the following correlation analyses, unless further removed if the nearby masked regions exceed a certain fraction, as described in Sections 4.2 and 5.2.

We use aperture magnitudes for the color selection; the aperture diameters are 15 pixels ($\sim 2\farcs5$) for *J*0210 and 10 pixels ($\sim 1\farcs7$) for *J*0222, *J*0924, and *J*1419. Auto-Mag, which applies automatically determined elliptical aperture for Kron photometry in SExtractor, is used for the estimation of total magnitudes. Galactic extinction is also taken into account in each band. Referring to the Galactic Dust Reddening and Extinction Service provided by IRSA, which is based on the results of Schlafly & Finkbeiner (2011), color reddening $E(B-V)$ is estimated as listed in Table 2. As $R_{\rm NB387} = A_{\rm NB387}/E(B-V)$ is estimated to be 4.009, taking into account the transmission curve, we apply the dust extinction correction to the detection catalogs. We replace the g-band magnitude with the corresponding 2σ limiting magnitude when the objects are fainter than the 2σ limit.

The measured PSF FWHM and the 5σ limiting magnitudes in the $1\farcs7$ aperture ($2\farcs5$ for *J*0210) of the final stacked images of NB387 and the g band are listed in the Table 1. The quality of the *J*0210 data is poorer than that of the other three fields in both the seeing and the final image depth.

3. Sample Selection

3.1. Lyα Emitter Selection

We use the color excess of the narrow band to the broad band as our LAE selection criterion, which has been widely used in previous studies (Guaita et al. 2010; Mawatari et al. 2012; Nakajima et al. 2012; Konno et al. 2016; Zheng et al. 2016). Although we have only the broadband data from the g band on HSC for estimating the continuum, we confirm here that the data are sufficient for the $z = 2.18$ LAE selection.

To define the selection criteria, we assume that the LAE spectrum model at $z = 2.0$–2.5 has a simple power law $f_\lambda = \lambda^\beta$ continuum and a correspondingly redshifted Lyα emission with a Gaussian profile with rest equivalent width ${\rm EW}_0 = 20$ Å. 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 determine the redshift evolution on the two-color diagrams. The tracks are shown in Figure 3. The black curves indicate the color tracks of $g-$NB387 versus $u-$NB387 in the left panel, and $g-$NB387 versus $r-$NB387 in the right panel. Three different UV slopes β, 0, -1.6, and -3.0, are shown in the both figures.

We also overplot the predicted tracks of possible contaminants, such as elliptical galaxies (ages of 2, 5, and 13 Gyr denoted as Ell2, Ell5, and Ell13, respectively), 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 faint stars used in Appendix A are also plotted, which are homogeneously archived from SDSS without the selection by prior knowledge of the colors.

From the color tracks, we find that the r band is of little help in LAE selection, while the u band can help to recover extremely red ($\beta \sim 0$) populations. However, as the typical UV slope of a $z \sim 2$ LAE has been found to be $\beta \sim -1.6$ (Kusakabe et al. 2019; Santos et al. 2020), we conclude that use of only the g band is sufficient for our $z \sim 2.18$ LAE selection, and a reasonable threshold for the color excess is $g-{\rm NB387} > 0.3$ to exclude most of the contaminants.

To increase confidence in selection by photometry, the color criteria are further defined as

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

The lower limit of 20.5 is set for the NB387 magnitude to avoid saturation, while the upper limit applies the 5σ limiting magnitude to ensure the reliability of the NB387 detection. This upper limit for the field *J*0924 with a moderate depth, 24.74, corresponds to $1.94\times10^{42}\,{\rm erg\,s^{-1}}$, which is $0.37\times L^*_{{\rm Ly}\alpha}$ and the characteristic luminosity $L^*_{{\rm Ly}\alpha} = 5.3\times10^{42}\,{\rm erg\,s^{-1}}$ (Konno et al. 2016). The definition of the color error follows Shibuya et al. (2018):

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

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

Selected objects that pass the criteria are naturally filtered by the spatial masks, as the original object detection is performed

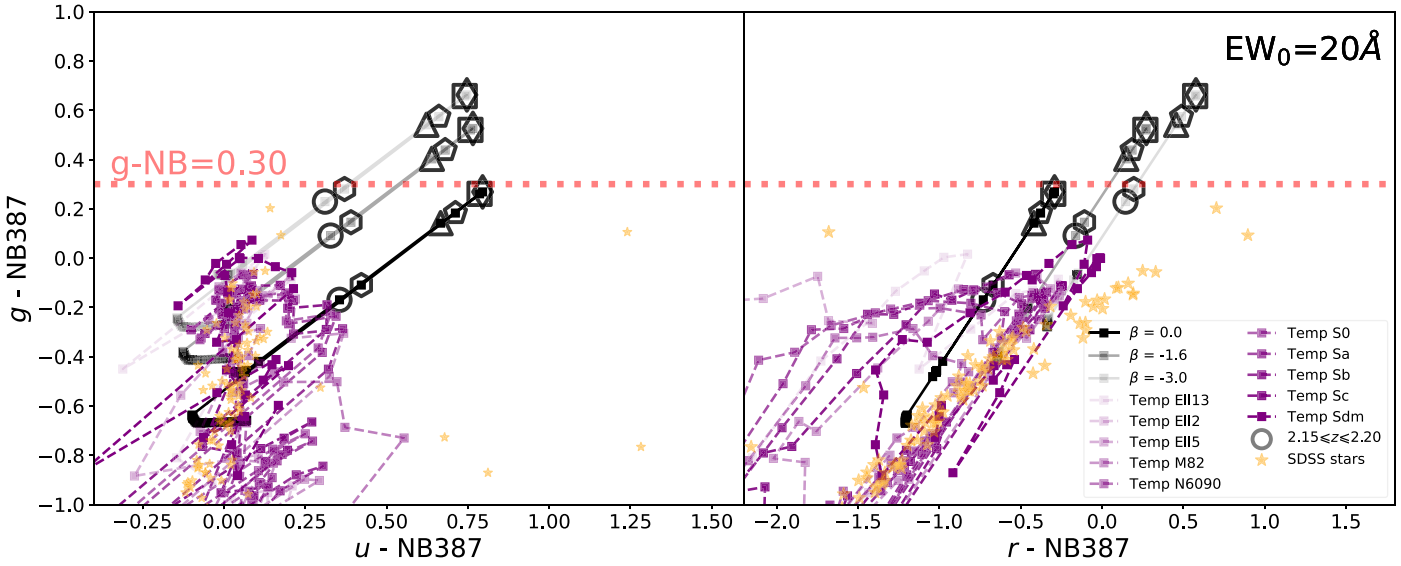

Figure 3. The color tracks with redshift evolution for the EW$_0$ = 20 Å LAE at the z = 2–2.5: g − NB387 vs. u − NB387 (left panel) and g − NB387 vs. u − NB387 (right panel). The black curves are the tracks for LAE models (with UV slope β = 0, −1.6, −3.0) with a redshift step Δz = 0.01 from z = 2 to z = 2.5, and the dashed purple curves are for galaxy templates, including elliptical galaxies (ages of 2, 5, and 13 Gyr denoted as Ell2, Ell5, and Ell13, respectively), starburst galaxies (M82 and N6090), and spiral galaxies (S0, Sa, Sb, Sc, Sd, and Sdm; Polletta et al. 2007) with a redshift step Δz = 0.1 from z = 0 to z = 3. Homogeneously archived SDSS stars with $g > 19$ (Yanny et al. 2009) are also plotted as yellow stars. The gray circle, triangle, square, diamond, pentagon and hexagon indicate the LAE models at z = 2.15, 2.16, 2.17, 2.18, 2.19, and 2.20, respectively. The narrowband excess g − NB387 > 0.30 works as a reasonable threshold to select the $z \sim 2.18$ LAEs.

with the masks applied. Finally, we perform a visual check of each candidate to reject false detections, such as hot pixels in the NB387 or saturated pixels in the g-band images. We also check for 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 *J*0210, six in *J*0222, and none in *J*0924 or *J*1419. These known low-z quasars are removed from our LAE sample. As a result, 465, 956, 585, and 636 LAE candidates are selected in fields *J*0210, *J*0222, *J*0924, and *J*1419, respectively, i.e., 2642 in total for all four fields, covering an effective area of 5.39 deg^2. There are 4, 3, 0, and 1 proximity quasars from the SDSS/(e) BOSS matched to these LAE candidates in each field. Specifically, in the *J*0210 central region where 11 quasars reside, three of the quasars in our sample are selected as LAEs, while most of the others are too bright in the NB387 images and do not fulfill the selection criterion NB387 > 20.5.

As shown in Figure 4, the selected LAE candidates in the final catalog are plotted as yellow points in the g − NB387 versus NB387 diagram, in which all detections are binned in the two-dimensional histogram coded by the blue color.[16]

The g-band and NB387 magnitude distributions in the surface density of the LAE candidates are shown 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 mag. The Auto-Mags are shown here for comparison of total magnitudes with the literature, and we overplot the $z \sim 2.2$ LAE sample in the Cosmic Evolution Survey (COSMOS) field from Nakajima et al. (2012) that was selected by the Subaru/Suprime-Cam NB387 (λ_0 = 3870 Å, FWHM = 94 Å). As their FWHM is almost twice that of the HSC NB387, corresponding to roughly double the survey volume, we also show the case scaled by a factor of 0.5. The distribution shapes are almost consistent, but all four of our fields show number excesses up around the limiting depth compared to the scaled numbers in COSMOS, although the excesses are less significant in *J*0924 and *J*1419 than in *J*0210 and *J*0222. These excesses are not surprising, as our fields are selected to contain potential overdensities.

In comparison 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 the Keck Baryonic Structure Survey (KBSS; Rudie et al. 2012; Chen et al. 2020) and *Ks*-selected photo-z galaxies (Mukae et al. 2017), our samples also reach deeper into 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

In addition to the LAEs at $z \sim 2.2$, some of the lower-z emitters may also pass our selection criteria. For the NB387 filter, the contaminants are considered to be mainly [O II] emitters at $z = 0.036 \pm 0.008$. However, the survey volume is much smaller in this redshift range than at $z \sim 2.2$, and the ratio reaches 0.2%. Given the low-z [O II] emitter luminosity function from Ciardullo et al. (2013) and our NB387 image depth, we estimate that the number detected is ∼0.05 in one HSC FoV. We conclude that the contamination rate of low-z [O II] emitters in our sample is negligible. In addition, C III] λ1909 at $z \sim 1$ and C IV λ1548 at $z \sim 1.5$ may also be interlopers. However, according to Konno et al. (2016), these emitters should be probable AGNs, as the objects passing our selection criteria yield EW$_0 \gtrsim$ 30 Å, which is much larger than that in typical star-forming galaxies.

In the literature on fields such as SXDS, COSMOS, HDFN, SSA 22, and E-CDFS (Guaita et al. 2010; Konno et al. 2016; Zheng et al. 2016), detection is performed in databases

[16] A sequence appearing around g − NB387 ∼ −2.5 is likely the stellar locus consisting of K- and M-type stars, as suggested by the stellar locus in the *ugr* diagram (Smolčić et al. 2004), and this is also supported by our random checks in matched SDSS spectra.

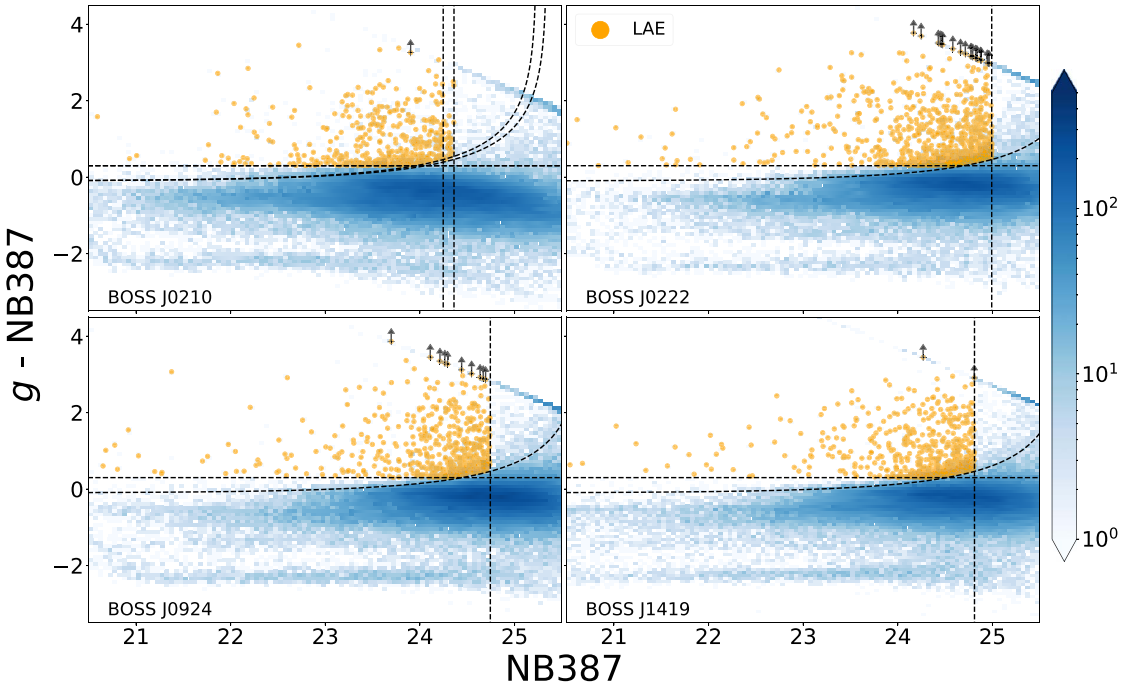

Figure 4. g − NB387 vs. NB387 diagram for LAE selection in each field. The two-dimensional histogram bins all detections from SExtractor, and the bluer bins contain more objects. The yellow points are the LAE candidates selected after visual inspection. The three selection criteria are shown as black dotted lines. Specifically, for the field *J*0210, the data are reduced in two versions of the *hscPipe*, and there is a slight difference in the final image depth, so the criteria are overplotted for clarification. The black arrows indicate LAE candidates with a g band fainter than the respective 2σ limiting magnitude of each image, and the g − NB387 values shown for these objects are the lower limits.

covering multiple wavelengths, e.g., the X-ray, UV, and radio wavelengths, to exclude low-z AGN contaminants. In our case, however, we search for overdense fields in the entire (e) BOSS survey; therefore, deep multiwavelength data are not available for testing the AGNs in this study. Instead, we refer to the literature mentioned above and find that the contamination rate of the LAE selections at $z \sim 2.2$ is commonly ∼10%–15%, and Sobral et al. (2017) 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 HydrOen Reionization Unveiled with Subaru (CHORUS; Inoue et al. 2020) survey, and the DEIMOS 10 K spectroscopic survey catalog (Hasinger et al. 2018). This yields a contamination rate of ∼15% in our LAE selection, and about two-thirds of the interlopers are likely to be the C IV emitters at $z \sim 1.5$, showing good consistency with previous studies. As this contamination level is secondary to Poisson noise and is therefore not crucial to our statistical study, we keep all of 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 Figure 6. We calculate the galaxy overdensity of each field to construct the overdensity maps. Overdensity is defined as

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

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

The aperture size is set to 10 cMpc (∼6′) in radius, which yields a mean number of LAEs counted in an aperture >10, giving a mean S/N > 3 signal assuming Poisson statistics for counting. The map is constructed using a 128 × 128 mesh grid for each field, corresponding to a resolution of ∼1 cMpc. When calculating the mean number density, we exclude the apertures covering the masked regions comprising more than 10%, a strict criterion to keep the mean number estimate robust. While drawing the overdensity map, we exclude apertures that are masked by more than 50%, a relaxed criterion to show more regions. The mean numbers $\langle N_{\rm LAE}\rangle$ (standard deviation $\sigma_{\rm LAE}$) within a radius $r = 10$ cMpc aperture for *J*0210, *J*0222, *J*0924, 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 *J*0210 and the larger number in *J*0222 are due to the difference in image depth.

In Figure 6, the blue contours in the background show the overdensity. Masked regions that are defined in Section 2.5 are shown as 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 positions of LoSs are marked as the red stars with the color coded by the effective optical depth $\tau_{\rm LoS}$. More quantitative discussions on the overdensity catalog will be presented in Z. Cai et al. (2020, in preparation), and here we present only a brief overview in Appendix B.

4.2. Correlation between Galaxy and IGM H I

With this LAE overdensity sample, we can first quantify the galaxy–IGM H I correlation based on the local overdensity estimate. We calculate the overdensity on the scale of 10 cMpc in radius, at the positions of the clean LoSs. Similar to Section 4.1, we discard LoSs whose vicinities are masked out by more than 50%, but as a result, no LoSs are removed in this

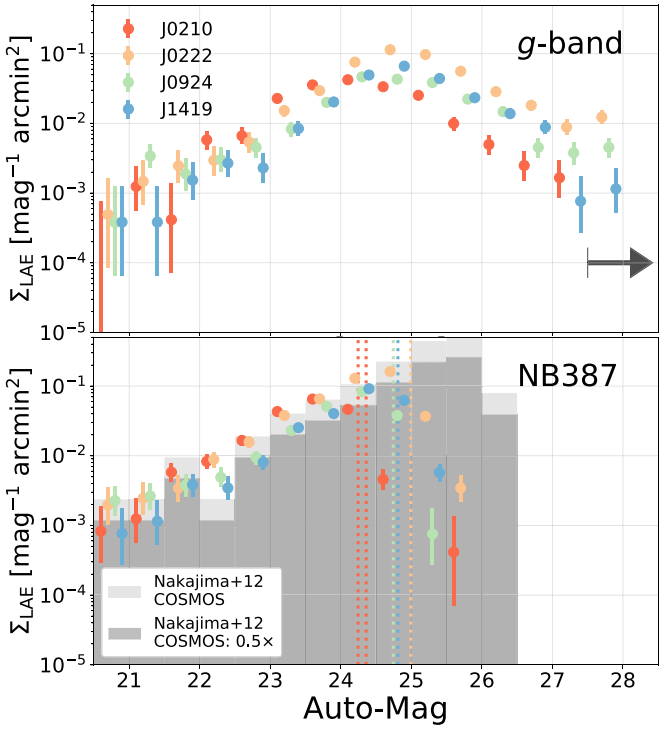

Figure 5. The Auto-Mag surface density distributions of the LAE candidates. Red, yellow, green, and blue histograms represent the LAEs in *J*0210, *J*0222, *J*0924, and *J*1419, respectively. Error bars indicate the Poisson errors. Upper: *g*-band magnitudes. The black arrow indicates that the rightmost bins include the faintest objects, the magnitudes of which are larger than the 2σ limiting magnitudes. Lower: NB387 magnitudes. The 5σ limiting aperture magnitudes are indicated by the vertical dotted lines in respective colors. The $z \sim 2.2$ LAEs in the Cosmic Evolution Survey (COSMOS) field selected by the Subaru/Suprime-Cam NB387 (Nakajima et al. 2012) are also plotted for comparison as the light shaded histogram. We also scale the numbers by a factor of 0.5 to roughly match the survey volume, and the results are shown as the dark shaded histogram.

process and the number of remaining LoSs is still 64. We assume the densities in the masked regions to be the mean values in each field. 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. 2017a).

Then, we can compare the LAE overdensity $\delta_{\rm LAE}$ and the effective optical depth $\tau_{\rm LoS}$ measured for the LoSs, the error of which is derived from the error of the mean flux in the pixel statistics, to investigate the correlation. The results are shown in Figure 7, from which it can be seen that the error of $\delta_{\rm LAE}$ suffers from the Poisson statistics with a small number of shots (LAEs) in each measured aperture, while the large error of $\tau_{\rm LoS}$ is mainly due to the relatively low S/N of the quasar spectra at the NB387-sensitive wavelength. Note that we discard LoSs with a continuum-to-noise ratio smaller than 2.

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

The large scatter in Figure 7 is largely contributed by the LoSs in *J*0210 (red triangles), which has a unique structure and has a shallower limiting magnitude. If we exclude *J*0210, the Spearman's rank correlation increases markedly to $\rho_S = 0.541^{+0.037}_{-0.051}$ with a P value of <0.01%. As the NB387 limiting magnitude of *J*0210 is shallower than that of the others, we test the effect of this by performing the same correlation analysis, limiting the LAE NB387 magnitude to 24.3 for all four fields; this leaves 451, 288, 264, and 248 LAEs in *J*0210, *J*0222, *J*0924, and *J*1419, respectively. The results with ($\rho_S = 0.388^{+0.026}_{-0.045}$) and without ($\rho_S = 0.502^{+0.031}_{-0.061}$) *J*0210 are consistent with those shown previously within the limits of uncertainty and cannot explain the significant difference. Therefore, the limit of the bright end is unlikely to be the dominant factor.

Alternatively, the difference in ρ_S can also be derived from the field variation in the correlation. The large filament and the existence of the grouping proximity quasars indicate that the structures in *J*0210 are probably different from those of other fields.

The binning data[17] show a clearer trend, which is overlaid as the gray crosses in Figure 7. $\tau_{\rm LoS}$ increases with $\delta_{\rm LAE}$, although interestingly, the trend becomes flatter when $\delta_{\rm LAE} \gtrsim 0.2$. The trend at the overdense end is likely dominated by the *J*0210 LoSs in the $\delta_{\rm LAE} > 0.5$ bins. Some of these LoSs are spatially close, and these special LoSs, hereinafter referred to as outliers, are highlighted by circling them out in Figure 7, and their sky distributions are also shown using the same symbol in Figure 6. The outliers cluster in two regions of the *J*0210 field, which are close to the node of the filament. This may indicate that different physical processes have taken place in the *J*0210 filament compared to typical structures at the same redshift.

4.3. Cross-correlation Analysis

Along with the analysis based on the local overdensity of LoSs, a more general analysis can be performed for the galaxy–IGM H I correlation. The correlation of the spatial distribution can be translated as the clustering properties between the two populations of objects, and the two-point cross-correlation function (CCF) is an ideal tool to quantify the clustering strength.

We divide the LoSs into two subsamples based on $\tau_{\rm LoS}$. To ensure there were comparable numbers of LoSs in the two subsamples, we set the criterion as $\tau_{\rm LoS} = 0.5$. LoSs with $\tau_{\rm LoS} > 0.5$ and $\tau_{\rm LoS} < 0.5$ are referred to as high-$\tau_{\rm LoS}$ LoSs and low-$\tau_{\rm LoS}$ LoSs, respectively. In this case, if we use the full sample, the numbers of LoSs for the high- and low-$\tau_{\rm LoS}$ subsamples are 30 and 34, respectively, and if we exclude field *J*0210, these numbers change to 23 and 19, respectively. We mainly discuss the case with inclusion of the *J*0210 LoSs for CCFs. Although we find outlying LoSs in Section 4.2 in *J*0210, only six outliers are pinpointed, while there are a total of 64 LoSs. The statistics like CCF is unlikely to be biased.

[17] The bins are made by sorting data points according to their $\delta_{\rm LAE}$ and splitting the nearest eight LoSs into one bin. Eight bins are set for the 64 LoSs, and the error is the 1σ standard deviation at each bin.

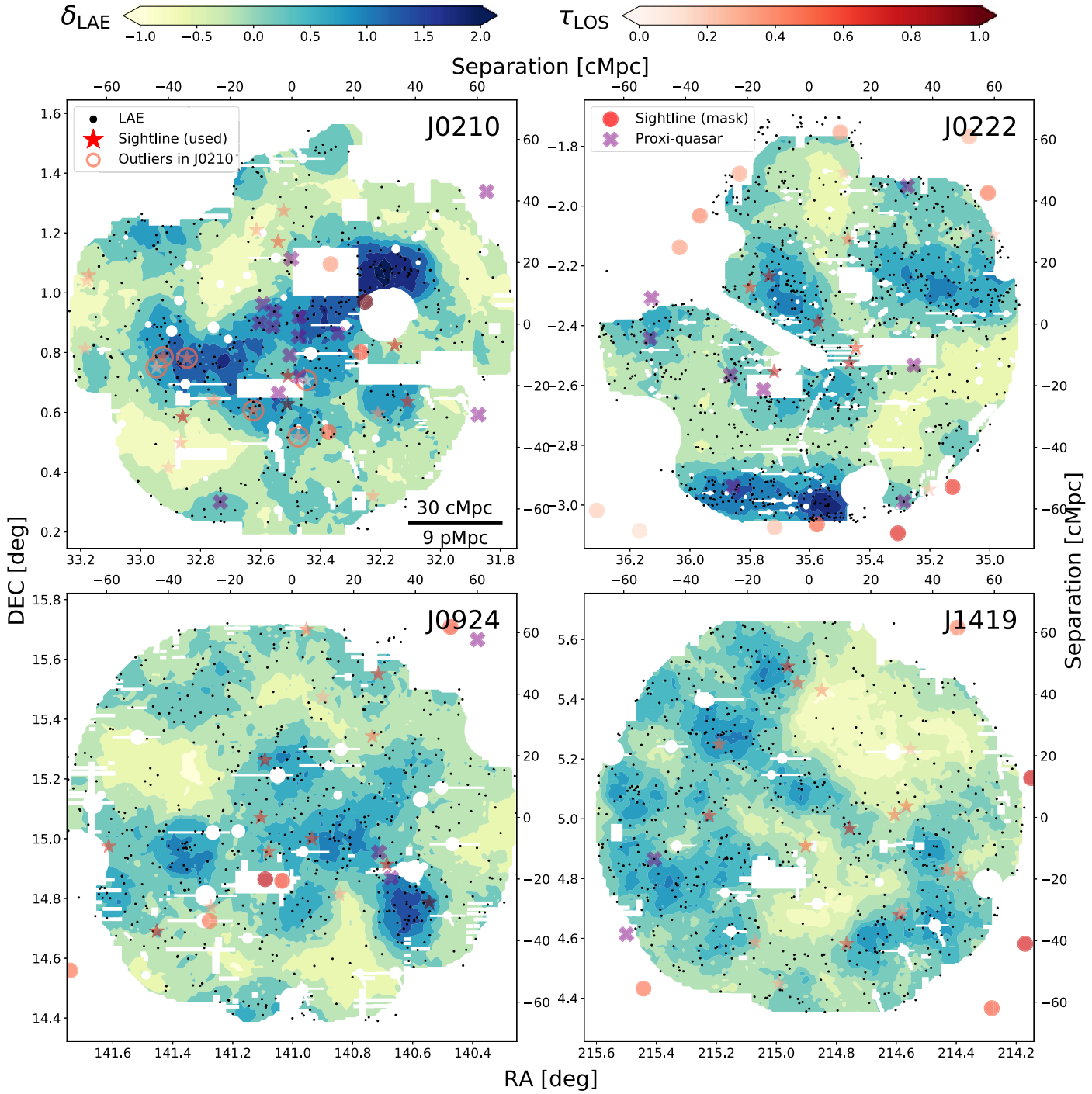

Figure 6. Overdensity maps for the four selected fields *J*0210, *J*0222, *J*0924, and *J*1419. Black points represent the LAE candidates. The blue contour in the background is scaled by the LAE overdensity $\delta_{\rm LAE}$ on a scale of $r = 10$ cMpc. The red stars and dots are the positions of the LoSs used 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 $2.15 < z < 2.20$. The 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 *J*0210, and details can be checked in the text in Section 4.2.

We use the angular CCF $\omega(\theta)$, or the projected CCF $\omega(r)$ if the angular separation is translated into projected physical distance, in the 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:

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

where $D_{\rm LAE}D_{\rm LoS}(\theta)$, $D_{\rm LAE}R(\theta)$, $D_{\rm LoS}R(\theta)$, and $RR(\theta)$ are the normalized LAE–LoS, LAE–Random, LoS–Random, and Random–Random pairs counted at the separation of angle θ within the 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 to be $0\overset{\circ}{.}013$, i.e., ∼0.4 proper-Mpc (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

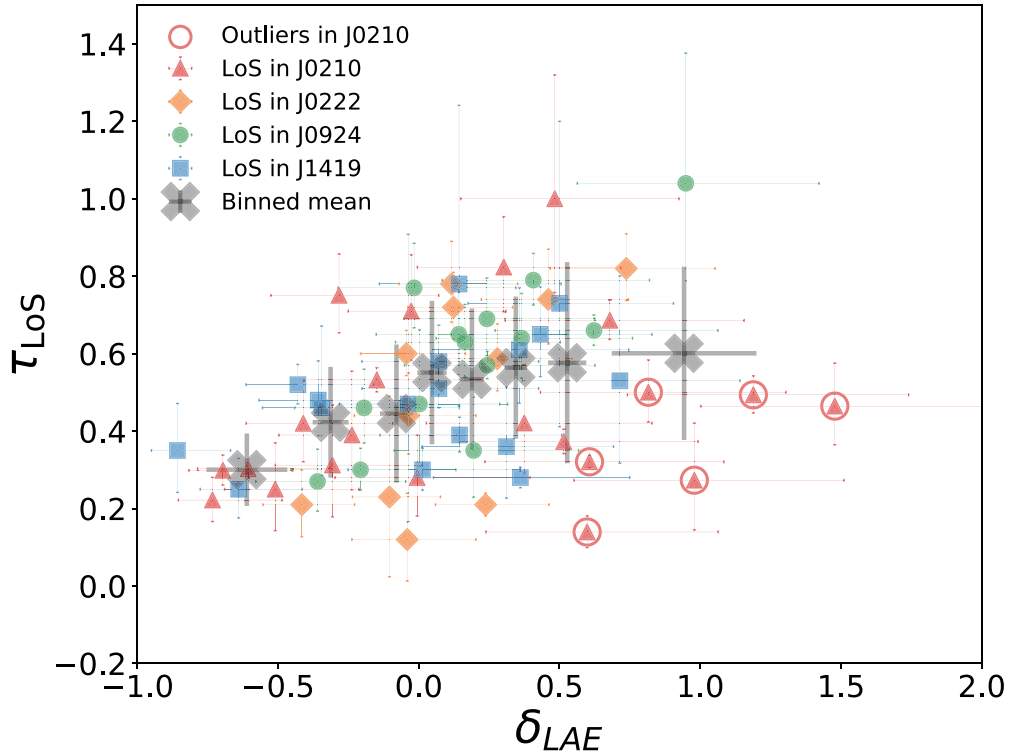

Figure 7. Correlation between LAE overdensity $\delta_{\rm LAE}$ and effective optical depth $\tau_{\rm LoS}$ at the positions of LoSs. Red triangles, yellow diamonds, green circles, and blue squares represent the original data points in fields *J*0210, *J*0222, *J*0924, and *J*1419, respectively. The number of the LoSs is 64 for four fields on a scale of 10 cMpc with less than 50% of the vicinity masked. The gray crosses represent the binned data points with 1σ standard deviation. With the exception of the most overdense bin, which is dominated by the data points in *J*0210, a clear increasing trend is observed. The outliers in *J*0210 with close spatial distribution are highlighted by red circles.

bins are set for the calculation extending up to 0°.6.[18] It should be noted here that $D_{\rm LAE}D_{\rm LoS}(\theta)$ represents the LAE–LoS pairs instead of the LAE–absorber pairs. We do not use the information regarding location along the LoS direction of the absorbers, because accurate LAE redshifts are unknown within $2.15 < z < 2.20$. Therefore, we use the projected CCF, but not the 3D CCF.

The CCF error is estimated by jackknife resampling, which can also take the field fluctuation into account. For resampling, we split each HSC field into 5×5 patches, and the one overlapping with masked regions by more than 50% is excluded to ensure sufficient numbers of pairs in each patch. Following Norberg et al. (2009), we denote the calculating bin by i and the resampling run by k. In the kth run, we skip the kth patch and perform an identical CCF calculation as for the full field sample. Then, the variance of the statistics of interest, i.e., $\omega(r)$, is derived for the ith bin:

$$\sigma_i = \frac{N_{\rm sub}-1}{N_{\rm sub}} \sum_{k=1}^{N_{\rm sub}} (\omega_{i,k} - \overline{\omega_i})^2, \quad (7)$$

where $\overline{\omega_i}$ is the mean over all resampling runs given by $\overline{\omega_i} = \sum_{k=1}^{N_{\rm sub}} \frac{\omega_{i,k}}{N_{\rm sub}}$ at the ith bin.

As described above, the projected CCF does not rely on the information regarding LoS direction. The LoSs can be viewed as being selected homogeneously from the sky and as not dependent on the foreground IGM at $z \sim 2.2$. Therefore, if the LoS number is infinite, a full sample that is not split by the $\tau_{\rm LoS}$ is expected to have a null CCF signal.

However, our sample size is limited, and this may involve an artificial signal into the CCF. We first test the CCF for the full sample that combines the high $\tau_{\rm LoS}$ and low $\tau_{\rm LoS}$ LoSs, and the results are shown as the green points in the left panel of Figure 8. Although the full sample has a much weaker signal than any subsample, they do not equal exactly to zero. For clearer comparison, we subtract the amplitude of the full sample CCF $\omega(r)_{\rm full}$ from that of each subsample CCF $\omega(r)$ and call the reduced signal the relative CCF, i.e., $\omega(r) - \omega(r)_{\rm full}$, which is shown in the right panel of Figure 8. For clarity, the data of the subsamples at each bin are slightly shifted along the r-axis in the figures.

From both panels in Figure 8, the high-$\tau_{\rm LoS}$ subsample shows a continuous positive signal from the innermost bins up to separation $r \approx 4$ pMpc. In contrast, the low-$\tau_{\rm LoS}$ CCF remains negative in the same distance range. As there can be 15% randomly distributed low-z interlopers in our LAE sample, the intrinsic signal may be even stronger. By varying the bin size, this characteristic distance changes by less than 1 pMpc. This result suggests that up to a scale of 4 ± 1 pMpc ($\sim 13 \pm 3$ cMpc at $z = 2.2$), LAEs tend to cluster in regions rich in gas and avoid the low-$\tau_{\rm LoS}$ regions where gas is less abundant.[19] The two bins at ~0.8–1.0 pMpc also tend to be consistent with zero, suggesting a weak signal at this distance.

[18] We test the following results by varying the bin size, and we confirm that our major results are not sensitive to the bin determination.

[19] As the accurate LAE redshifts are unknown, the $\tau_{\rm LoS}$ estimated at the absorption spike acts as an upper limit constraining the intrinsic H I associated with the LAEs around an LoS. This is why the $\tau_{\rm LoS} < 0.5$ subsample can show a negative CCF, even though many of them still have $\tau_{\rm LoS}$ higher than the cosmic mean value.

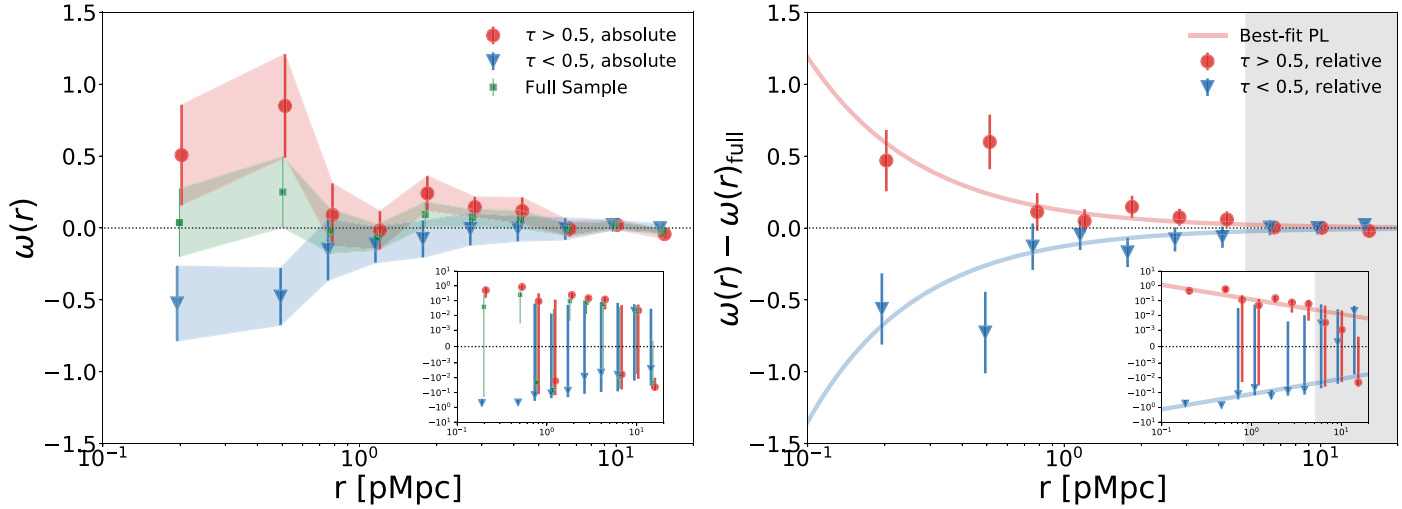

Figure 8. Cross-correlation function (CCF) between LAEs and LoSs for the high-$\tau_{\rm LoS}$/low-$\tau_{\rm LoS}$ subsamples. The red points and curves are the $\tau_{\rm LoS} > 0.5$ subsample and the corresponding fit power-law model, while the blue points and curves are for the $\tau_{\rm LoS} < 0.5$ subsample. Data points for different subsamples in each bin are shifted slightly along the r-axis for clarification. In both panels, the major figures are shown on a linear scale, while the inset figures are shown on a log scale. The axis labels of the inset figures are the same as those of the respective major figures. Left panel: original CCFs are shown for the two subsamples in red/blue, and the full sample in green, the LoSs of which 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 $\omega(r)_{\rm full}$ 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 gray-shaded region indicates the separation limit where the signal becomes noisy.

Table 3
The Parameters of CCF Power-law Fitting for Different Subsamples

Fields [1]	$\tau_{\rm LoS}>$ [2]	$N_{\rm LoS}$ [3]	γ [4]	r_0 [5]	$\tau_{\rm LoS}<$ [6]	$N_{\rm LoS}$ [7]	γ [8]	r_0 [9]
Incl. *J*0210	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. *J*0210	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. *J*0210	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. *J*0210	0.7	13	$0.79^{+1.97}_{-0.11}$	$0.06^{+0.07}_{-0.06}$	0.3	13	$0.84^{+2.07}_{-0.15}$	$0.09^{+0.14}_{-0.09}$

Note. [1]: cases regarding field *J*0210; [2]: high-$\tau_{\rm LoS}$ criterion; [3]: number of LoSs in the high-$\tau_{\rm LoS}$ subsample; [4]: γ fit for the high-$\tau_{\rm LoS}$ subsample; [5]: r_0 fit for the high-$\tau_{\rm LoS}$ subsample; [6]: low-$\tau_{\rm LoS}$ criterion; [7]: number of LoSs in the low-$\tau_{\rm LoS}$ subsample; [8]: γ fit for low-$\tau_{\rm LoS}$ subsample; [9]: r_0 fit for the low-$\tau_{\rm LoS}$ subsample.

It should be noted that the CCF shown in the right panel can be well fitted by a power law:

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

where r_0 is called the clustering length, which makes $\omega(r_0) = 1$ and can be taken as an indicator of clustering strength. We fit the binned data points with the power law, and the fitting curves are shown in corresponding colors in Figure 8. The best-fit parameters (γ, r_0) with the errors estimated from the 10,000 Monte Carlo perturbed simulations, similar to Section 4.2, are ($0.99^{+0.54}_{-0.17}$, $0.12^{+0.05}_{-0.03}$ pMpc) and ($1.03^{+0.83}_{-0.21}$, $0.13^{+0.06}_{-0.02}$ pMpc) for the high- and low -$\tau_{\rm LoS}$ subsamples, respectively, and they are also summarized in Table 3.

The r_0 for both subsamples are of the order of 0.1 pMpc, which is much smaller than the typical galaxy–galaxy clustering strength of the order of several pMpc. This indicates that the strength of the LAE−IGM H I correlation is much weaker than galaxy clustering, although it is still sufficiently significant to be detected based on our samples in the projected CCF. We note that Momose et al. (2020a) obtained somewhat stronger 3D CCF signals between LAEs and CLAMATO H I absorption data with $r_0 = 0.78\,h^{-1}$ cMpc, which corresponds to ∼0.35 pMpc for $z \sim 2.2$ in our cosmology.

We examine whether the results would be changed, if we excluded the field *J*0210 or if we changed the $\tau_{\rm LoS} = 0.5$ criterion to separate the LoSs into subsamples. We do not find that such factors have significant impacts on our results (the details are presented in Appendix C).

4.4. 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 as the overdensity-based analysis requires a sufficiently large aperture to overcome the small number statistics when counting galaxies. In the CCF analysis, it is necessary to divide LoSs into high- and low-$\tau_{\rm LoS}$ subsamples, which causes a drop in the sample size by at least a factor of 2o. This can be extremely problematic for the smallest separation bin.

To inspect the correlation on smaller scales down to sub-cMpc or ∼200 kpc in physical length, where the CGM is thought to surround the host galaxies, we perform another analysis that is similar in concept to the stacking technique. We derive the radial distribution of the IGM optical depth that is averaged over all LoSs within annuli centered on a specific

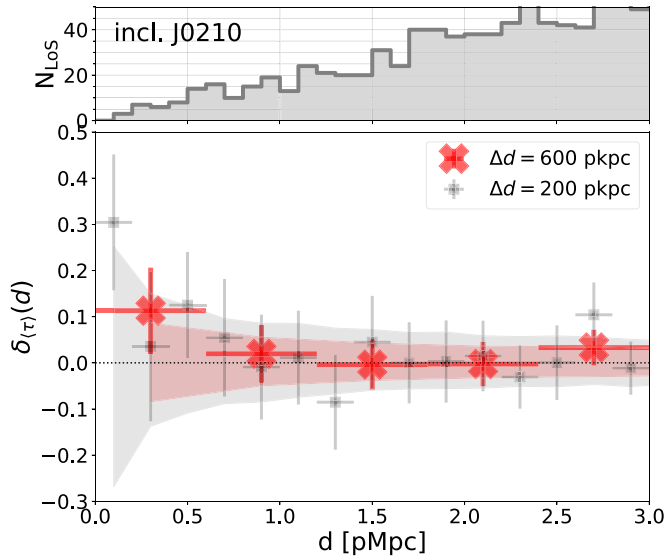

Figure 9. Fluctuations of the average $\tau_{\rm LoS}$ as a function of distance to LAEs, $\delta_{\langle\tau\rangle}(d)$, for the case including *J*0210. Errors are indicated by the 1σ standard deviation based on 1000 times bootstrap resampling. The gray squares represent the fine bins with resolution of 200 kpc, and the red crosses show the 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 2σ appears at $d < 200$ pkpc, indicating the detection of a CGM signal around LAEs, while a tentative 13% excess at $400 < d < 600$ pkpc shows a weaker signal in the IGM regime.

LAE. We then further calculate the mean over all LAEs, designated as the average $\tau_{\rm LoS}$ profile $\langle\tau\rangle$, where d is the distance from the stacked LAE. To emphasize the excess level, we define the fluctuation of $\langle\tau\rangle$ as

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

where $\langle\tau\rangle_{\rm tot}$ is the mean over the radial direction. We first calculate the $\langle\tau\rangle_{\rm tot}$ over a large distance $0 < d < 0.^{\circ}3$, or $0 < 0 < 9.2$ pMpc at $z = 2.2$, in two cases: a coarse bin with a spatial resolution of $\Delta d = 600$ pkpc and a finer bin with a higher resolution of $\Delta d = 200$ pkpc.[20]

The results derived from all LAEs and LoSs in four fields are shown in the lower panel of Figure 9. We mainly consider the case including *J*0210 here for the same reason as given for CCFs, i.e., global statistics are unlikely to be biased by six outliers in 64 LoSs. We test the case excluding *J*0210 in Appendix D, which shows consistent results, except for a larger scatter due to the smaller LoS number. The gray squares are for the fine bin, while the red crosses represent the coarse bin. The error shown in the *y*-axis is the 1σ standard deviation from the 1000 times bootstrap resampling with both the LAEs and LoSs, and the error in the *x*-axis indicates the bin size. The shaded regions are the 16%–84% ranks in 1000 simulations by assessing random positions of the 64 LoSs with corresponding resolutions, indicating that bins outside the shaded regions are confident for inferences. The number of LoSs counted can be checked with the gray step function above.

[20] The two cases are chosen because (1) they are significant 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; and (2) to draw meaningful results after testing various bin sizes, signals exceed the 84% rank in random LoSs only with these two choices.

As shown in Figure 9, a 10% excess beyond the error is found in the innermost region on the scale of <600 pkpc. Although the number of LoSs in the innermost fine bin is only three, a more significant 30% excess is found at the 2σ level at $d < 200$ kpc, which is the expected region of distribution with the CGM. Rudie et al. (2012) and Momose et al. (2020a) also find CGM signals at $d \lesssim 300$ pkpc to star-forming galaxies that were originally identified as LBGs. Here, we statistically find a similar result for LAEs at the smaller distance.

In addition, the fine bins also indicate a 13% excess at a distance of $400 < d < 600$ pkpc, and given the sufficient number of LoSs, this excess is likely to be the dominant signal accounting for the 10% excess in the coarse bin. However, this scale is larger than the region typically thought to be the CGM reservoir, especially with regard to LAEs, which are generally less massive than LBGs. Instead, the clustering of IGM H I is 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 ~0.5 pMpc, confirming the correlation between galaxies and IGM H I down to around 400 pkpc.

We also note that there is a sudden excess of 10% at $2.6 < d < 2.8$, although the coarse bin largely flattens the signal. We do not fully understand the origin of this signal, but a noncontinuous signal on such a large scale is unlikely to be physically meaningful.

5. Discussion

Here, we discuss the implication of the results shown in Section 4. We first compare our galaxy–IGM H I results with those of the previous studies. As hinted at in the CCF, the correlation is dependent on the scale. Therefore, we further explore the scale dependence of the overdensity-based correlation. Finally, we discuss the underlying physics related to our results.

5.1. Comparison with Previous δ_{LAE}–τ_{LoS} Studies

One directly related study is Mukae et al. (2017), in which the galaxy–IGM H I correlation is studied by using the K_s-selected photo-z galaxies at redshift $2 < z < 3$ and the Lyα forest sample in the background quasar spectra from the SDSS-III/BOSS survey.

The correlation shown in their Figure 2 is physically similar to our $\delta_{\rm LAE}$–$\tau_{\rm LoS}$ correlation, but the Lyα absorption is estimated in the Lyα forest fluctuation, which is defined as

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

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

For comparison with their results, we convert the optical depth derived in Section 2.1 into the transmission fluctuation $\delta_{\langle F\rangle}$ according to 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 $\delta_{\langle F\rangle} - \delta_{\rm LAE}$ relation based on our samples of LAEs and LoSs is shown in Figure 10.

The symbols used for the data points are the same as in Figure 7, but to clarify the effects of including or excluding *J*0210 in fitting, the LoSs in *J*0210 are shown in red and the LoSs in the other fields are shown in blue. We also present the

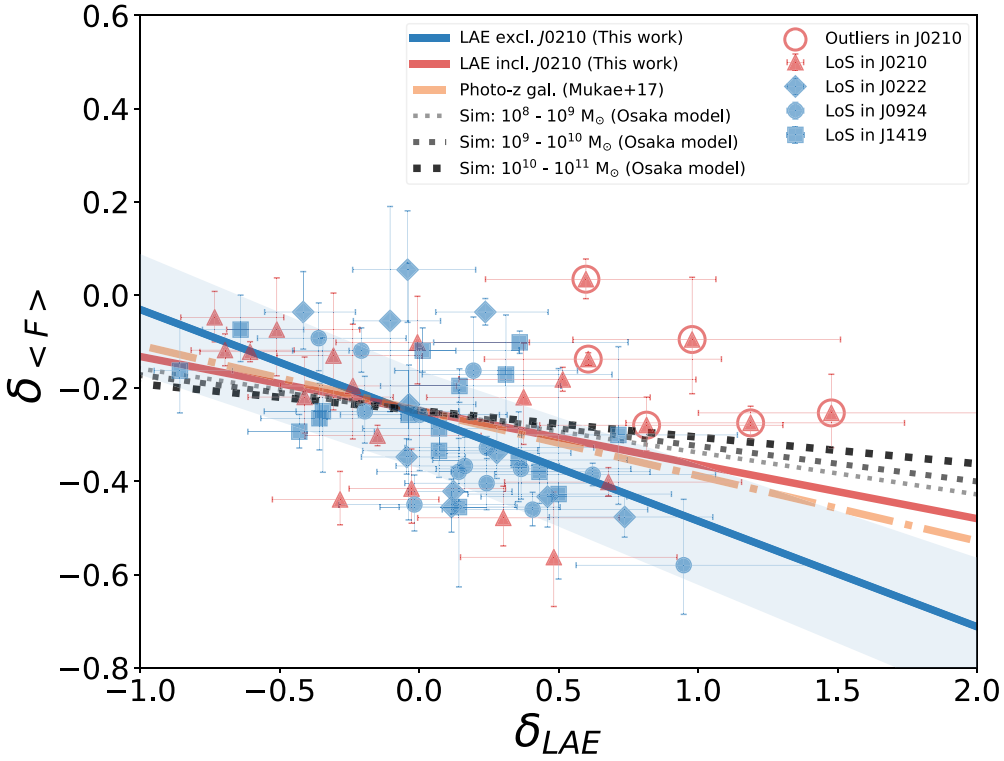

Figure 10. Transmission fluctuation $\delta_{\langle F\rangle}$ against the LAE overdensity $\delta_{\rm LAE}$, similar to Figure 2 of Mukae et al. (2017). Symbols are the same as in Figure 7. LoSs in *J*0210 are shown in red, while LoSs in the other three fields are colored blue. Outliers in *J*0210 are highlighted by red circles. The red and blue solid curves are the best-fit model for the data points including and excluding *J*0210, respectively. The orange dotted–dashed line is the result from Mukae et al. (2017) using the photo-z galaxies and 16 BOSS LoSs. The dotted lines are the predictions from the 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 *J*0210.

linear fits, i.e., the solid lines in Figure 10, with and without the *J*0210 corresponding to the red and blue curves, respectively. The uncertainty of parameters is given by the 16%–84% ranks from the 10,000 Monte Carlo simulations with perturbation. The fitted relation for all four fields is

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

Similar to Figure 7, the outliers in *J*0210 are highlighted in circles in the upper right of Figure 10. If we exclude the LoSs in *J*0210, the relation becomes

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

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

Both cases in our study give consistent slopes with the photo-z galaxies within their uncertainty, although the case excluding *J*0210 has a larger discrepancy and is steeper. One possible reason for the large discrepancy is the difference in galaxy masses, given that photo-z galaxies are generally more massive than LAEs. Massive galaxies are likely to form in deeper positions in the gravitational potential well, where H I is abundant for building up stellar masses M_*. In this case, the overdensity of less massive galaxies, such as LAEs, will be systematically lower than that of heavier populations, e.g., photo-z galaxies, resulting in a $\delta_{\langle F\rangle} - \delta_{\rm LAE}$. This interpretation is also supported by Momose et al. (2020a).

To further examine this possibility, we refer to the results from the GADGET3-Osaka cosmological hydrodynamic simulation, which is based on the smoothed particle hydrodynamics (SPH) simulation code GADGET3 (Springel 2005) and takes full account of star formation and supernova feedback (Shimizu et al. 2019). More details on the simulation data processing are presented in Momose et al. (2020b) as well as in Nagamine et al. (2020), and we call the simulation as the Osaka model. The model curves for galaxies with M_* ranging in the range 10^8–10^9, 10^9–10^{10}, and 10^{10}–10^{11} $M_\odot$ have slopes of -0.090 ± 0.011, -0.076 ± 0.009, and -0.057 ± 0.006, as plotted in Figure 10, and the intercepts are again normalized at $\delta_{\rm LAE} = 0$, given that the absorption in the model is estimated at a fixed position, i.e., the central redshift $z = 2.175$, which is different from our estimate at the absorption spike.

There is an M_* dependence of the relation slopes in the Osaka model, and the less massive galaxy population shows a steeper trend. However, such dependence is not as sensitive as expected to fully account for the discrepancy, and more interestingly, the fit of the case with *J*0210 shows 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 *J*0210, 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 that the reason of the discrepancy originated from different galaxy

stellar masses, and the case excluding *J*0210 seems rather to be the biased case.

Another possible cause for the discrepancy may be H I suppression in the Lyα emission. As our observations targeted fields with clustering of strong IGM Lyα absorption, the Lyα emission from galaxies may have been suppressed in such H I-rich environments. In this case, the LAEs in *J*0210, which is likely to contain unique structures lacking the IGM H I as suggested by the outlying LoSs, should be less affected. Meanwhile, in the other fields, the completeness of detection of LAEs may have been lower and the $\delta_{\rm LAE}$ is underestimated. This interpretation is favored by the Osaka model prediction. The plateau that appears in the CCF at $r \lesssim 0.6$ pMpc is found in Momose et al. (2020a), which also suggests the deficit of LAEs on small scales.

However, we note that some uncertainties may remain in the simulation models (e.g., contributions from AGNs), and our sample size is still limited in terms of the field variance. In future studies, we hope to determine the true reason for the slope discrepancy, with more HSC fields targeting at various environments. Follow-up studies to search for Hα emitters (HAEs) residing in the same structures, which are less biased by the radiative transfer process, will also help to calibrate the $\delta_{\langle F\rangle} - \delta_{\rm LAE}$ slope robustly.

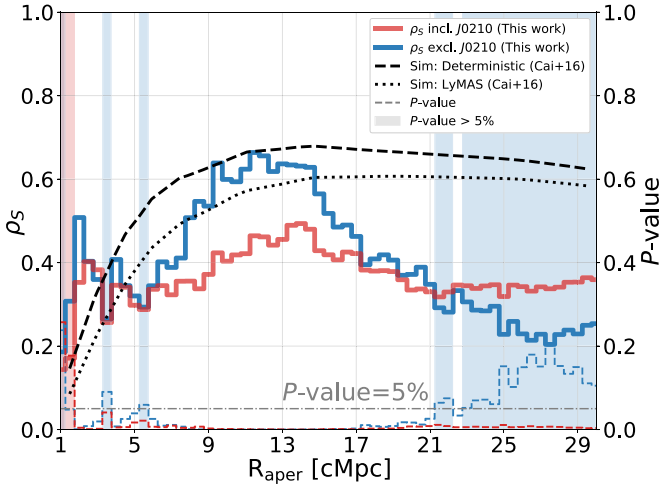

Figure 11. Scale dependence of the $\delta_{\rm LAE}$–$\tau_{\rm LoS}$ correlation. The bin size is 0.5 cMpc and the red and blue solid curves are the Spearman's rank correlation coefficient including and excluding *J*0210, respectively. The dashed colored curves indicate the corresponding *P* values at each bin. The scale range with the *P* value >5% is masked with shaded regions, indicating that the result is not confident. The two models from Cai et al. (2016), Deterministic and LyMAS models, are overplotted by scaling the box length in simulations by a factor of 1/2 to match the $R_{\rm aper}$. Both models reach the maximum at a comparable scale $R_{\rm aper} \sim 13$ cMpc.

5.2. Scale Dependence of the δ_{LAE}–τ_{LoS} Correlation

Possible scale dependence is already hinted at in the CCF in Section 4.3. To investigate the scale dependence in overdensity-based analysis, we perform the Spearman's rank correlation test for ($\delta_{\rm LAE}$, $\tau_{\rm LoS}$) with the $\delta_{\rm LAE}$ calculated for different aperture sizes set from 1 to 30 cMpc with a bin step of 0.5 cMpc for the radius. We note that LoSs are kept for analysis only when <50% of their vicinity is masked, which maintains a stable LoS number when the scale increased. The results are shown in Figure 11. The red and blue curve shows the Spearman's rank correlation coefficients on various scales for the cases including and excluding *J*0210, respectively. The corresponding *P* values are shown as the dashed line; the $R_{\rm aper}$ with *P* values >5% indicating results lacking confidence are shaded.

It is clear that when *J*0210 is included, the correlation remains moderate at a level of $\rho_S \sim 0.3$ for almost all scales, while in the case excluding *J*0210, the correlation becomes strong at $r \approx 9$–15 with $\rho_S \gtrsim 0.6$. In both cases, there is scale dependence of the correlation between $\delta_{\rm LAE}$ and $\tau_{\rm LoS}$, although the trend is more significant without *J*0210. The trend is tested by making 1000 realizations with randomly distributed LAEs and LoSs, and the results fluctuate around $\rho_S = 0$ with scatter $\Delta\rho_S = \pm 0.2$ in the 16%–84% ranks for both cases, which does not show scale dependence larger than the scatter magnitude. This supports that our results are not dominated by statistical fluctuation.

On a relatively small scale, the correlation becomes stronger as the scale increases and reaches a peak at $r = 13 \pm 2$ cMpc. With *J*0210, the correlation shows a flatter shape when $r > 13$ cMpc, while it tends to decrease on the same scales without *J*0210. The difference again indicates that *J*0210 may have a unique structure, the existence of which can alter the correlation significantly in overdensity-based analysis. Therefore, a large sample size covering various types of environments is essential in galaxy–IGM H I correlation studies regarding field variance. Here, however, the discussion focuses on the case with exclusion of *J*0210.

The CCF in Section 4.3 shows that the amplitude for high $\tau_{\rm LoS}$ subsample remains positive up to 4 pMpc (∼12.7 cMpc). As the $\delta_{\rm LAE}$–$\tau_{\rm LoS}$ correlation shown in Figure 11 can be viewed as the cumulative signal within $R_{\rm aper}$, the scale of the correlation peak agrees well with the CCF results.

The similar $\delta_{\rm LAE}$–$\tau_{\rm LoS}$ curves from the deterministic and the Lyα Mass Association Scheme (LyMAS) models in Cai et al. (2016) are also overlaid as black curves, with the original box lengths scaled by 1/2 on the projected plane to maintain consistency with $R_{\rm aper}$. Compared with the models in Cai et al. (2016), our results on the scale of maximum correlation are consistent with both the deterministic case and LyMAS on the projected plane, although the amplitude may be different due to the different tracers used. We note that the two models construct Lyα forest in different ways, and both are shown here to validate that the consistency is unlikely to be dependent on how the Lyα forest is modeled. This suggests the effectiveness of the current simulated cosmological models in terms of the IGM H I gas.

The decrease on large scales is interesting, as it is not predicted in the simulation. Note that the HSC FoV covers a region of over 100 cMpc, so it is unlikely to account for the weaker correlation on scales of 15–30 cMpc. We also test the mask-region criteria by changing <50% to <10% when estimating the ρ_S. With this change, the analysis uses only the clean LoS sample far from uncovered regions. This test shows a similar decreasing trend at $r > 15$ cMpc, supporting the suggestion that the decrease is unlikely to be caused by the FoV limit.

An alternative reason for the discrepancy on large scales is that the models use not only galaxies but rather the total matter in a defined box, which is less clustered, as indicated by the simulations. At $z \sim 2$, the galaxy bias continues to decrease toward the scale over $\sim 10\,h^{-1}$ cMpc (Cen & Ostriker 2000; Springel et al. 2018), while the H I bias is almost flat on such

large scales (Ando et al. 2019). Hence, the correlation between the total matter and H I in the models can remain strong on larger scales, while the LAE–IGM H I correlation simultaneously weakens. Another possible reason suggested by the simulations in Momose et al. (2020b) is that on large scales the signal in the projected correlation is diluted, as the uncertainty in the three-dimensional separation becomes larger when $R_{\rm aper}$ increases. In addition, the inclusion of more contaminants 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 $z \approx 2.2$, there is a moderate to strong positive correlation between $\delta_{\rm LAE}$ and $\tau_{\rm LoS}$ on the scale of $r = 10$ cMpc. Such a correlation implies that galaxies are clustered in regions associated with large amounts of H I gas. This correlation is dependent on scale, and the peak is located at $R_{\rm aper} \sim 13$ cMpc. The correlation seems to be natural in that IGM H I gas tends to accumulate in deeper potential wells, which are associated with more massive halos. The condensed H I gas then triggers star formation, and stars and galaxies will emerge in the same region. According to Madau & Dickinson (2014), such activity is especially intense at $z \sim 2$.

However, the situation becomes complicated when detailed processes are taken into account. Hot massive stars can emit ionizing photons with energy >13.6 eV, and once they escape from the host galaxies, the H I gas in the surrounding IGM will be ionized. Such processes are likely to be more active for the case of the LAEs used in this study, which are thought to be a population of young star-forming galaxies. Our results indirectly imply that the escape fraction of ionizing photons from LAEs at $z \sim 2$ or their SFR may still not sufficiently high as to fully ionize the IGM H I gas on the scale of several cMpc. Feedback from supernovae or AGNs potentially inhabiting them may also blow off the surrounding gas to more distant regions, although the power of such processes is still under debate and the scale at which their effects are evident is not clear.

Some studies have explored the two-point CCFs between Lyα 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 $z \lesssim 1$. However, at $z \gtrsim 2$, such correlations can only be constrained with limitations in either a bright galaxy population, small survey area, or small LoS sample size in a limited number of studies (Adelberger et al. 2003; Rudie et al. 2012; Mukae et al. 2017). Our results confirm the correlation between IGM H I and galaxies with rest-frame UV magnitudes down to $M_{\rm UV} \approx -18$ estimated from the g band, even at redshift $z \approx 2.2$, where star formation and feedback processes can be very active. A positive correlation is found up to 4 pMpc (or 13 cMpc) from CCF analysis (or with varying the aperture size for $\delta_{\rm LAE}$ in the $\delta_{\rm LAE}$–$\delta_{\langle F \rangle}$ correlation), and down to at least 400 pkpc (or 1.3 cMpc) from the average $\tau_{\rm LoS}$ profile centered on LAEs.

This implies that the ionization or feedback from galaxies (LAEs) and/or the inhabiting AGNs is not sufficient to cancel out gravitational effects on large scales. It also indicates that IGM H I still traces LSS well at $z \sim 2$ on the scale of 1.3 to ∼13 cMpc, although with large scatter. Alternatively, the correlation may also be the result of additional inflow providing pristine H I gas (Dekel & Birnboim 2006; Tumlinson et al. 2017). Turner et al. (2017) suggest that the observed redshift-space distortions in the KBSS survey (Rakic et al. 2012) are predominantly caused by infall, confirming that gas inflow can alter observables up to a scale of 5 pMpc.

Either or both the two possible scenarios can reproduce our results, and the two cannot be distinguished at present. However, it will be possible to answer this question by comparing our results with future numerical simulations that take into account photoionization effects.

On the other hand, little is known regarding how well the LAEs trace the underlying structures, especially in our fields, which are thought to be associated with neutral IGM gas. The physical similarity between LAEs and non-LAEs at $z \sim 2.2$ is hinted at in Hathi et al. (2016). Hough et al. (2020) also report that overdense regions traced by LAEs and HAEs show good consistency in their semianalytic models, and Shimakawa et al. (2017) observationally support this claim on the scale of >1 cMpc. All these indicate that LAEs can be good structure tracers on large scales.

Meanwhile, as reported in Shi et al. (2019), LAEs and LBGs may trace different structures formed in different periods or in different dynamic statuses. Especially on the small scale of $\lesssim$300 pkpc, or <1 cMpc, tentative deficits are always found for LAEs, both in this study, as suggested by the inner plateau in the CCF, and in the literature, e.g., LAE number deficits in a protocluster core (Shimakawa et al. 2017) or at the center of the massive overdensity (Cai et al. 2017a), and possible Lyα suppression in galaxy overdense regions (Toshikawa et al. 2016). Thus, LAEs may not be good for tracing the highest overdensity regions.

However, in statistical studies on large scales, LAEs with well-constrained redshift $\Delta z \approx 0.04$ are still the best tool and Subaru/HSC can map these objects with high efficiency. In the future, NIR surveys for HAEs, which are identified by the nonresonant Hα emission line, will help determine the performance of the LAE tracer in unique environments.

5.3.2. Hints from the Correlation Scatter

The scatter can also be an important indicator of underlying structures. As shown in Figures 7 and 10, data points are distributed with large scatter. This may be due to the uncertainties in our measurements. We summarize the possible factors here.

First, regarding the overdensity measured in our study, we can map the LAEs only on a projected plane, leaving an uncertainty of ∼60 cMpc along the redshift direction, and the aforementioned scales are all in transverse separation instead of in comoving volume.

Additionally, as discussed in Section 5.3.1, it is unclear how much bias is introduced by LAEs. As *J*0210 markedly changes the statistical results, there may be structures with field-to-field variation. With regard to the LoSs sample, although the CoSLAs have been checked carefully to exclude DLAs or LLSs, these systems may still contaminate the low-$\tau_{\rm LoS}$ LoSs. However, even if we focus only on high-$\tau_{\rm LoS}$ LoSs (see LoSs with $\tau_{\rm LoS} \gtrsim 0.6$ in Figure 7), there is still large scatter, just similar to what is reported by Miller et al. (2019) in their simulations with both high spatial and mass resolutions. This suggests that there may be some intrinsic origins.

The scatter may be a coincidence that occurs when LoSs pass through a gas filament, a large void, or an orthogonal

filament with low density, according to the simulation presented in Mukae et al. (2017), which also shows a large scatter in their correlation. In this sense, the outliers in *J*0210 may penetrate galaxy overdensity associated with a gas filament lying in a direction transverse to the LoS by chance.

In addition to the morphological origin, the radiation from galaxies preheating the diffuse IGM H I in the most overdense region may also cause the observed scatter. This scheme is suggested by Mawatari et al. (2017), where Lyα absorption is shown to be associated with a $z \approx 3.1$ overdensity SSA 22 on a scale of $\sim$50 cMpc overall, but not dependent on local overdensity. A special system found in IGM tomography also shows similar characteristics (Lee et al. 2016). Actually, three outliers in *J*0210 located at regions with $\delta_{\rm LAE} \gtrsim 1.0$ and $\tau_{\rm LoS} \lesssim 0.4$ are likely to reside in regions that are abundant with galaxies but deficient in cold H I gas, similar to the environments mentioned above.

By further studying such cases in the scatter, it may be possible to identify more ideal laboratories for testing the theories of galaxy evolution and their interplay with IGM H I in extreme environments at $z \approx 2$.

5.3.3. Redshift Evolution of Correlation Scales

There have been several studies on the CCF between Lyα absorbers and galaxies at $z < 1$ (Ryan-Weber 2006; Chen & Mulchaey 2009; Tejos et al. 2014), and the CCFs provide information for both correlations and their effective scales. As the galaxy populations and the Lyα absorption systems used in our study are not identical, it is difficult to directly compare the CCF amplitude and the resulting clustering length r_0. Nevertheless, $r_{\rm up}$, defined here as the upper limit of the scale to identify the positive signal, can still be instructive.

From the CCFs, we find an underlying redshift evolution of the correlation scale by combining Ryan-Weber (2006) and Tejos et al. (2014), the CCFs of whom also extend over 10 cMpc, with our results. We find that (1) at $z \lesssim 0.04$, the CCF between Lyα absorbers with H I column density $10^{12.5} \lesssim N_{\rm H\,I} \lesssim 10^{15}\,{\rm cm}^{-2}$ and HIPSS galaxies shows a strong positive signal up to 10 h^{-1} cMpc (Ryan-Weber 2006), i.e., $r_{\rm up} \sim 15$ cMpc, slightly larger than our upper limit $r_{\rm up} = 13 \pm 3$ cMpc; (2) while at $0 \lesssim z \lesssim 1$, a CCF signal between Lyα absorption systems with H I column densities in the range $10^{14} \lesssim N_{\rm H\,I} \lesssim 10^{17}\,{\rm cm}^{-2}$ and galaxies can be found only up to $r_{\rm up} \sim 7$ cMpc (Tejos et al. 2014), which is significantly smaller than ours.

$r_{\rm up}$ decreases from $z > 2$ to $0 < z < 1$, and then increases again toward $z = 0$, which shows consistency with the evolution of the correlation length of galaxy clustering (Baugh et al. 1999; Springel et al. 2018). This supports a physical picture in which the redshift evolution of galaxy–IGM H I correlation may follow a similar pattern to that of the galaxy clustering. However, we note that when measuring $r_{\rm up}$, there may be uncertainty based on the different galaxy populations and absorbers.

5.3.4. A Sudden Dip in the $\delta_{\langle\tau\rangle}(d)$ Profile

As shown in Figure 9, $d = 200$–400 kpc interestingly shows a sudden dip between the two excess bins. The significance is low due to the small number of LoSs, but a similar trend is reported in the latest results by Chen et al. (2020), which show a sudden dip at 70–150 pkpc based on a sample of 2862 background galaxies (originally identified as LBGs). It suggests that the feature is related to the transition phase between the inflow in the inner region and outflow in the outer region, which may originate from star formation activities. In this picture, it is possible that LAEs, the young and less massive galaxies located in the shallower potential well that are active in star formation, host the stronger outflows and cause the transition that appears at a larger projected distance. The current weak signal in our data still precludes us from drawing any firm conclusions, but a larger sample size in future studies will help to resolve this question.

6. Summary

In this study, we perform deep NB387 and *g*-band imaging using the 8.2 m Subaru/HSC on the fields that are selected by a technique similar to that used in the MAMMOTH project, preferentially tracing a group of strong Lyα absorbers selected from the full (e) BOSS database. Using narrowband images, we select LAE candidates at $z = 2.18$ and constructed $\delta_{\rm LAE}$ maps. To estimate the IGM H I overdensity, we use the (e) BOSS LoS data to calculate the $\tau_{\rm LoS}$ at the same redshift. Based on the $\delta_{\rm LAE}$ and $\tau_{\rm LoS}$ data, we perform correlation analyses to study the galaxy−IGM H I correlation up to a scale of $\sim$100 cMpc. In addition, we also examine the correlation at CGM scales down to 200 pkpc based on the statistical sample.

The results are summarized as follows:

1. We construct LAE overdensity maps for four HSC fields traced by IGM H I at $z = 2.18$, with a total of 2642 LAE candidates detected down to $L_{{\rm Ly}\alpha} \approx 2 \times 10^{42}\,{\rm erg\,s^{-1}}$ over a survey area of $5.39\,{\rm deg}^2$. The selected LAE candidates reside in a variety of environments, including filaments, sheets, and clumps.
2. There is a moderate to strong correlation between the $\delta_{\rm LAE}$ and $\tau_{\rm LoS}$ based on 64 LoSs from SDSS/(e) BOSS, which shows rough consistency with the results of Mukae et al. (2017), although the $\delta_{\langle F\rangle} - \delta_{\rm LAE}$ slope is steeper when we exclude the field *J*0210. Based on comparison with the GADGET3-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 Lyα emission in regions of high H I density. We further find that the correlation is dependent on the scale of the $\delta_{\rm LAE}$ estimate. The peak of the correlation is located around $R_{\rm aper} = 13 \pm 2$ cMpc.
3. By dividing the LoSs into high and low $\tau_{\rm LoS}$ subsamples with a cutoff criterion of $\tau_{\rm LoS} = 0.5$, the cross-correlation analysis shows a significant correlation signal up to 4 ± 1 pMpc ($\sim 12.7 \pm 3.2$ cMpc). The results clearly imply that LAEs tend to reside in gas-rich regions, as indicated by the high $\tau_{\rm LoS}$ in the background LoS, and avoid low-$\tau_{\rm LoS}$ areas where the H I is deficient. The plateau shape at $r \lesssim 600$ pkpc implies the offset of LAEs and IGM H I at small scales.
4. Analysis of the average $\tau_{\rm LoS}$ profile centered on LAEs can trace the absorption signal down to a scale of 200 pkpc. We find a 30% excess at $d < 200$ kpc with 2σ significance, although only with three LoSs counted, indicating statistical detection of the CGM signal around LAEs. We also detect a $>1\sigma$ signal of 13% excess at

400–600 pkpc that is thought to be in the IGM regime, supporting IGM signal detection down to ∼400 pkpc.
5. The positive correlation indicates that, at $z \sim 2$, neither ionization nor supernova/inhabiting AGN feedback from LAEs is sufficient to erase the gravitational effects on galaxy−IGM H I correlation, or alternatively, inflows continue to supply the H I gas in large-scale filaments to the surrounding environment of galaxies.
6. We also find a large scatter in the $\delta_{\rm LAE}$–$\tau_{\rm LoS}$ correlation. According to the simulation in Mukae et al. (2017), outliers may be the cases where LoSs penetrate regions with specific morphological arrangements. At the high-overdensity end, exceeding ionization and preheating processes may explain the deficit in cold IGM H I, similar to the $z = 3.1$ protocluster in the SSA 22 field (Mawatari et al. 2017).
7. Comparison of our correlation scale with CCFs between Lyα absorbers and galaxies at $z < 1$ (Ryan-Weber 2006; Tejos et al. 2014) implies that the redshift evolution of galaxy−IGM H I correlation may follow the evolution of galaxy clustering.
8. A tentative sudden dip at $d = 200$–400 pkpc in the $\delta_{\langle\tau\rangle}(d)$ profile may imply a transition between the inflow and outflow of LAEs compared to LBGs.

This project is still ongoing to obtain more LAEs in different fields and more LoSs in overdense regions to strengthen statistical robustness and allow us to compare observables with simulations to identify the best models of structure formation and evolution in terms of IGM H I in the future. The upcoming Subaru/Prime Focus Spectroscopy will be highly efficient for spectroscopic confirmation of our LAE candidates and will also provide us with a good opportunity to perform 3D IGM tomography in various environments, especially those with coherently distributed IGM H I and overdensities.

We thank the anonymous referees 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 JSPS KAKENHI grant No. JP17H01111, 19H05810 (K.N.). K.N. 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

There is a systematic offset in Equation (2) for NB387, and a constant should be introduced for correction as the colors between NB387, g, and r are influenced by the 4000 Å break, which is sensitive to metallicity and stellar age (Kauffmann et al. 2003). The Pickles templates are constructed mainly from stars with solar metallicity (Pickles 1998), while the number of star references used in the *hscPipe* has a peak at around $19 < g < 21$ and so tends to include metal-poor and old halo stars that are more distant from us at high Galactic latitude. This difference may be a source of systematic bias. As the effects of metallicity and stellar age are degenerated here, we denote the correction factor as $C_{\rm metal}$.

To estimate $C_{\rm metal}$, we homogeneously select the faint stars ($g \gtrsim 19$) with S/N > 3 spectra in the NB387 wavelength range from the SDSS database around the COSMOS field, the Galactic latitude of which 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 account. These stars are plotted as the gray crosses in Figure 12. To maintain consistency with the fitting in *hscPipe* and also to reduce the fitting uncertainty, we use only stars with $0.2 < g - r < 0.4$, which show 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 estimation, we perform visual inspection of each spectrum of all of these stars to discard those with peculiar features in the NB387 wavelength range. After this check, stars used for the zero-point correction are plotted as purple crosses in Figure 12. We use these realistic stars instead of the Pickles templates to fit the relation shown in Equation (2), and $C_{\rm metal}$ is estimated to be −0.448. The original relation fit from *hscPipe* is shown as the dashed black curve, and the corrected relation fit is shown as the solid purple curve.

When fitting Equation (2), the scatter of references is large in the case of NB387, resulting in a fitting uncertainty as large as 0.2 mag, thus causing field-to-field variation. We perform more subtle calibration for this by introducing another constant, $C_{\rm fit}$. We first select the extended sources with $23.5 <$ NB387 < 24.5, which are most likely the high-z galaxies that are free from the 4000 Å break in the g band, in each field. Then, the field-dependent $C_{\rm fit}$ is estimated by adjusting the $g -$ NB387 of these sources to −0.10, the expected mean color of the $1 < z < 3$ galaxies given their typical UV slope (Kurczynski et al. 2014).

$g -$ NB387 $= -0.10$ can also be verified by utilizing the HSC/NB387 data from the CHORUS survey (Inoue et al. 2020) and the spectral redshift $z_{\rm spec}$ from the DEIMOS 10 K 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; there are 3711 matches with flag $q > 1$, suggesting good spectroscopic redshift measurement. The $g -$ NB387 versus NB387 with $C_{\rm metal}$ correction of these objects is plotted as the gray dots in the right panel of Figure 12. We pick out all 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 $g -$ NB387 in a dual-Gaussian distribution, as faint objects are likely in a flattened distribution due to photometric errors. The

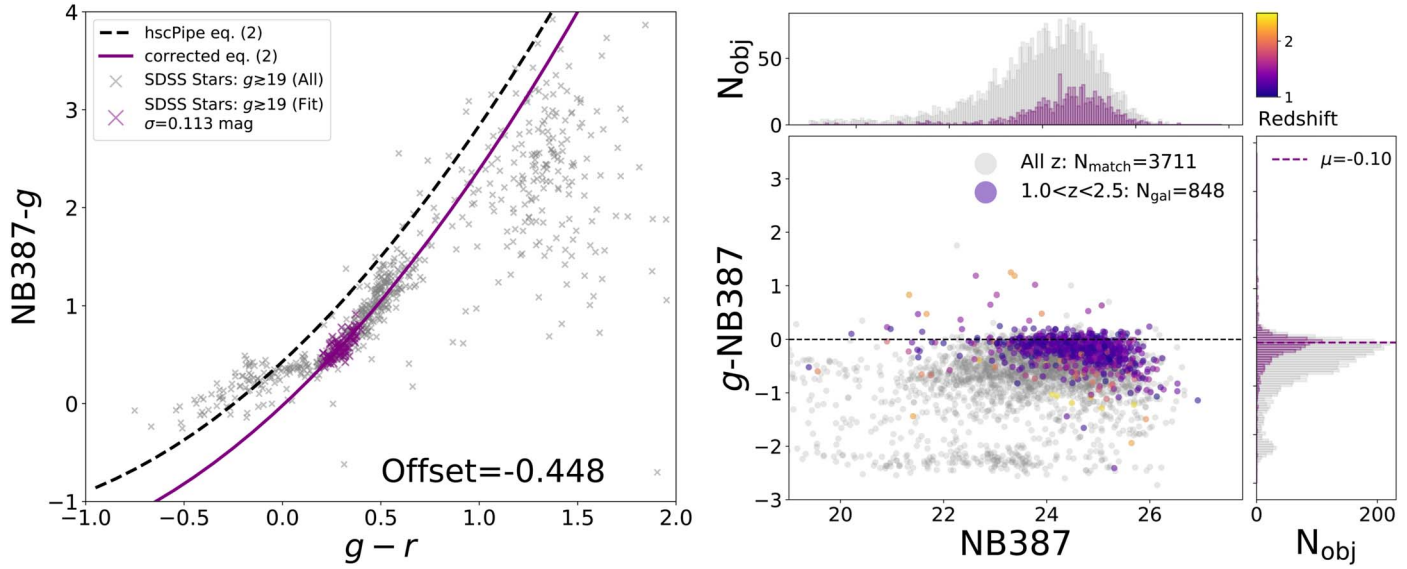

Figure 12. (Left) The predicted NB387 − g vs. $g - r$ diagram for homogeneously selected SDSS stars with $g \gtrsim 19$. The gray crosses are the selected stars, and the purple crosses are those with $0.2 < g - r < 0.4$ after visual inspection, which are used for fitting the correction factor $C_{\rm metal}$. The dashed black curve is Equation (2) from *hscPipe*, and the solid purple curve is the corrected relation with $C_{\rm metal} = -0.448$. (Right) The g − NB387 vs. NB387 diagram for the 2″ cross-matches between the CHORUS objects (Inoue et al. 2020) and the DEIMOS 10 K catalog (Hasinger et al. 2018). The gray dots are all of the 3711 matches with flag $q > 1$, suggesting a robust spectral redshift $z_{\rm spec}$ measurement, and the dots coded with the hot map are the 848 high-z matches with $1.0 < z_{\rm spec} < 2.5$, where hotter indicates greater redshift.

result for the main sequence peak is $\mu = -0.10$, which is consistent with the expected color of high-z galaxies. This consistency also validates $C_{\rm metal}$ as a confident correction, and because the CHORUS NB387 data were observed under excellent conditions and have a plausible depth, it is reasonable to use the suggested value of −0.10 for calibrating $C_{\rm fit}$ in each of our fields in this paper. The resulting $C_{\rm fit}$ fluctuates in −0.002–0.191 mag among the four fields, which is consistent within the fitting uncertainty of 0.2 mag.

Appendix B
Notes on Individual Fields

More quantitative discussions on the overdensity catalog will be presented in Z. Cai et al. (2020, in preparation), and we present a brief overview from Figure 6 in this appendix.

B.1. J0210 Field

There is a large filamentary structure at the center of the field as well as the structures with weaker significance, which are likely to be sheet-like structures, around the nodes at the ends. The field is traced by both the central grouping LoSs with strong Lyα absorption and a group of quasars clustering within an area of $\sim 40 \times 40$ cMpc^2 at $2.15 < z < 2.20$. The filamentary structure with which they are associated extends for about 100 cMpc, and the peak density of one node with $\delta_{\rm LAE} \sim 3$ reaches a significance of over 6σ. This result implies that combined use of both tracers can effectively trace the unique LSS, as also suggested in Cai et al. (2017a). Given what has been reported in the previous studies regarding the correlation between multiple quasar environments and galaxy overdensity (Hennawi et al. 2015; Cai et al. 2017b; Onoue et al. 2018; Mukae et al. 2019), the emergence of the grouping of quasars implies that the filament is different from the typical environments at $z \sim 2$, and *J*0210 should be considered individually in the correlation analysis. The uniqueness of the structure in *J*0210 is outside the scope of this paper, and further discussion will be presented in a future paper.

B.2. J0222 Field

This field is seriously affected by bright stars in and around the FoV, resulting in large masked areas with strange patterns. A weak clump with overdensity $\delta_{\rm LAE} \sim 1.0$ over a 20 cMpc length scale is found close to the central region, likely associated with the central group of high-$\tau_{\rm LoS}$ LoSs. Another clump with comparable significance appears on the west side, but it is not connected with the central structure on our projected overdensity map. It should be noted that a large filamentary structure with an overdense peak $\delta_{\rm LAE} > 1.6$ appears at the southern boundary of the FoV. There are nearby LoSs in the vicinity with relatively high $\tau_{\rm LoS}$, but they are outside our pointing FoV. Therefore, this structure is not found intendedly by strong IGM Lyα absorption, but just by chance.

B.3. J0924 Field

We mainly use the central four grouping LoSs with high $\tau_{\rm LoS}$ as tracers. However, in the central area, we do not find a structure with significant overdensity based on this LAE sample. Within the *J*0924, several peaks have moderate overdensities $\delta_{\rm LAE} > 0.8$ that are comparable to or surpass the central structure. The most overdense structure is found in the southwest of the field, which is close to two LoSs with strong Lyα absorption. The peak of the structure has an overdensity measured as over 1.2, and it extends for about 30 cMpc.

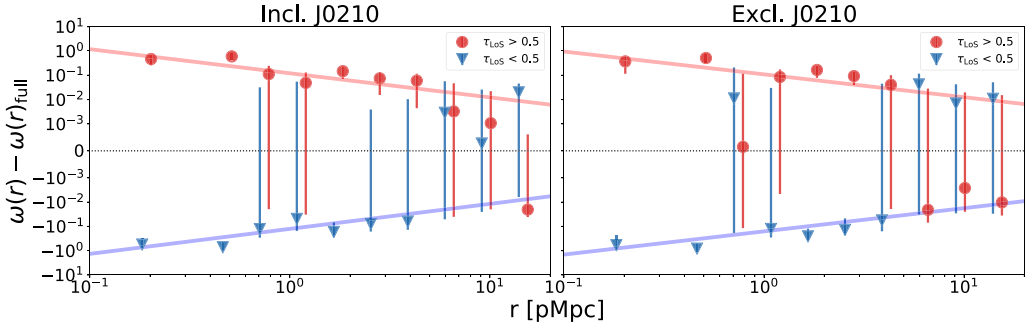

Figure 13. Relative CCFs between LAEs and LoSs for the high-τ_{LoS}/low-τ_{LoS} subsamples on a log scale to check the results for the cases including (left panel) and excluding (right panel) *J*0210, similar to the inset figures in the right panel of Figure 8. Bins are set on a log scale with the right boundary from 0.4 pMpc to 18.3 pMpc. Red points and curves are the $\tau_{LoS} > 0.5$ subsample and corresponding fitting power-law model, while blue points and curves are the $\tau_{LoS} < 0.5$ subsample. The fitting parameters can be checked in Table 3; they do not differ significantly between the two cases.

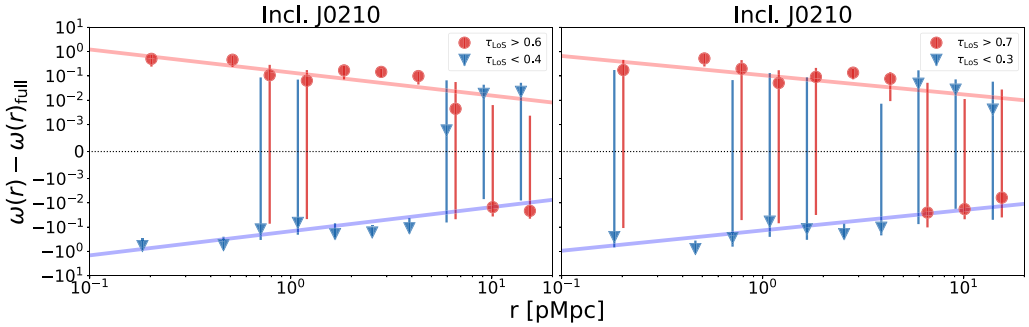

Figure 14. Relative CCFs between LAEs and LoSs for the high-τ_{LoS}/low-τ_{LoS} subsamples on a log scale to check the results with varying the subsample criteria. Symbols are similar to Figure 13. Both figures show the results for the case including *J*0210. Left panel: for the subsamples $\tau_{LoS} > 0.6/\tau_{LoS} < 0.4$; right panel: for the subsamples $\tau_{LoS} > 0.7/\tau_{LoS} < 0.3$. The fitting parameters can be checked in Table 3; they do not differ significantly.

B.4. J1419 Field

There are more structures in clumpy shapes in *J*1419. Although there are four LoSs with $\tau_{LoS} > 0.6$, they are more scattered than those in other fields, with distances of ∼40–100 cMpc from each other. Hence, coherently strong absorption is expected to be less significant. However, the number of LoSs in this field is appreciable for correlation analysis. Five peaks with moderate $\delta_{LAE} > 0.6$ are found in various regions, but no extreme overdense or extended structure is seen in this field. In contrast, a large void with a size of $\sim 50 \times 60$ cMpc^2 emerges at the northwest of the FoV.

Appendix C
Test of the CCF Results

The CCF presented in Section 4.3 may show some variation by changing the sample size. Here, we first test the difference between cases including and excluding field *J*0210. The results on a log scale are shown in Figure 13, where the left and right panels show results including and excluding *J*0210, respectively. The comparison indicates no significant changes in the results with the exclusion of *J*0210, except for one bin around 0.8 pMpc, and generally larger errors, probably due to the smaller sample size. There is also not much variation in the clustering strength indicated by r_0, as summarized in Table 3.

As the definition of subsamples is based on the essentially arbitrary criterion, i.e., τ_{LoS} over/lower than 0.5, we also test whether varying the criterion changes the results. We divide LoSs into other subsamples with $\tau_{LoS} > 0.6/\tau_{LoS} < 0.4$ and $\tau_{LoS} > 0.7/\tau_{LoS} < 0.3$ to ensure a comparable sample size for each subsample as of 21/24 and 13/13. The results are shown in Figure 14. Comparison of the results with those shown in the left panel of Figure 13 indicates no significant changes in the trend of CCFs, except for greater uncertainty because of the smaller sample size. Table 3 summarizes the fitted r_0; they are still of the same order as the ∼0.1 pMpc scale.

Appendix D
Average Optical Depth Profile Excluding *J*0210

In Section 4.4, given the importance of the LoS number for the statistics when inspecting on small scales, we discuss mainly the case with inclusion of *J*0210, which contains a large filament associated with a group of quasars. Here, we show the results for the case excluding *J*0210 in Figure 15. There is no significant change in the general $\langle\tau\rangle$ varying trend along the distance to LAEs in the inner region as discussed in Section 4.4, although the scatter is larger due to the smaller number of LoSs. This supports our assumption that the six outliers of the 64 LoSs are unlikely to affect the statistics, such as $\langle\tau\rangle$. We note that two fine bins at ∼2.7 pMpc show the tentative excess, more significant than the case including *J*0210, although the coarse bin still shows a weak signal.

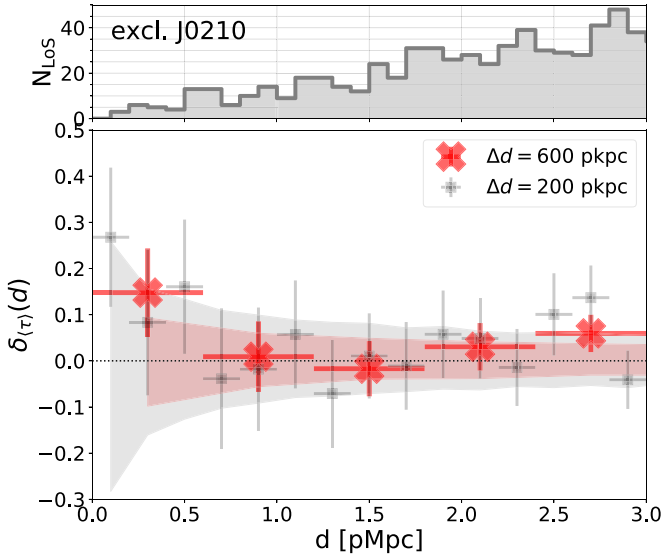

Figure 15. The fluctuation of the average $\tau_{\rm LoS}$ as a function of distance to LAEs, $\delta_{\langle\tau\rangle}(d)$, for the case excluding *J*0210. The symbols are the same as in Figure 9. *J*0210 does not alter the general trend.

ORCID iDs

Yongming Liang https://orcid.org/0000-0002-2725-302X
Nobunari Kashikawa https://orcid.org/0000-0003-3954-4219
Zheng Cai https://orcid.org/0000-0001-8467-6478
Xiaohui Fan https://orcid.org/0000-0003-3310-0131
J. Xavier Prochaska https://orcid.org/0000-0002-7738-6875
Kazuhiro Shimasaku https://orcid.org/0000-0002-2597-2231
Masayuki Tanaka https://orcid.org/0000-0002-5011-5178
Kei Ito https://orcid.org/0000-0002-9453-0381
Rhythm Shimakawa https://orcid.org/0000-0003-4442-2750
Kentaro Nagamine https://orcid.org/0000-0001-7457-8487
Masafusa Onoue https://orcid.org/0000-0003-2984-6803
Jun Toshikawa https://orcid.org/0000-0001-5394-242X

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, PASJ, 71, 114
Ando, R., Nishizawa, A. J., Hasegawa, K., Shimizu, I., & Nagamine, K. 2019, MNRAS, 484, 5389
Bădescu, T., Yang, Y., Bertoldi, F., et al. 2017, ApJ, 845, 172
Baugh, C. M. 2006, RPPh, 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, Natur, 380, 603
Bosch, J., Armstrong, R., Bickerton, S., et al. 2018, PASJ, 70, S5
Cai, Z., Fan, X., Bian, F., et al. 2017a, ApJ, 839, 131
Cai, Z., Fan, X., Peirani, S., et al. 2016, ApJ, 833, 135
Cai, Z., Fan, X., Yang, Y., et al. 2017b, ApJ, 837, 71
Capellupo, D. M., Hamann, F., Herbst, H., et al. 2017, MNRAS, 469, 323
Casey, C. M., Cooray, A., Capak, P., et al. 2015, ApJL, 808, L33
Cen, R., & Ostriker, J. P. 2000, ApJ, 538, 83
Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, arXiv:1612.05560
Chen, H.-W., & Mulchaey, J. S. 2009, ApJ, 701, 1219
Chen, Y., Steidel, C. C., Hummels, C. B., et al. 2020, MNRAS, 499, 1721
Chiang, Y.-K., Overzier, R., & Gebhardt, K. 2013, ApJ, 779, 127
Chiang, Y.-K., Overzier, R. A., Gebhardt, K., & Henriques, B. 2017, ApJL, 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., Kneib, J.-P., Percival, W. J., et al. 2016, AJ, 151, 44
Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, AJ, 145, 10
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., & Arrigoni-Battaia, F. 2015, Sci, 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
Hough, T., Gurung-López, S., Orsi, Á, et al. 2020, MNRAS, 499, 2104
Inoue, A. K., Shimizu, I., Iwata, I., & Tanaka, M. 2014, MNRAS, 442, 1805
Inoue, A. K., Yamanaka, S., Ouchi, M., et al. 2020, PASJ, 72, 101
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., Momose, R., et al. 2019, PASJ, 71, 55
Kusakabe, H., Shimasaku, K., Ouchi, M., et al. 2018, PASJ, 70, 4
Landy, S. D., & Szalay, A. S. 1993, ApJ, 412, 64
Lee, K.-G., Bailey, S., Bartsch, L. E., et al. 2013, AJ, 145, 69
Lee, K.-G., Hennawi, J. F., Stark, C., et al. 2014a, ApJL, 795, L12
Lee, K.-G., Hennawi, J. F., White, M., et al. 2016, ApJ, 817, 160
Lee, K.-G., Hennawi, J. F., White, M., Croft, R. A. C., & Ozbek, M. 2014b, ApJ, 788, 49
Lee, K.-G., Krolewski, A., White, M., et al. 2018, ApJS, 237, 31
Lee, K.-G., Suzuki, N., & Spergel, D. N. 2012, AJ, 143, 51
Mackenzie, R., Fumagalli, M., Theuns, T., et al. 2019, MNRAS, 487, 5070
Madau, P., & Dickinson, M. 2014, ARA&A, 52, 415
Mawatari, K., Inoue, A. K., Yamada, T., et al. 2017, MNRAS, 467, 3951
Mawatari, K., Yamada, T., Nakamura, Y., Hayashino, T., & Matsuda, Y. 2012, ApJ, 759, 133
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., Shimasaku, K., Kashikawa, N., et al. 2020a, Environmental Dependence of Galactic Properties Traced by Lyα Forest Absorption (II): Diversity Among Galaxy Populations, arXiv:2002.07335
Momose, R., Shimizu, I., Nagamine, K., et al. 2020b, Environmental Dependence of Galactic Properties Traced by Lyα Forest Absorption (I): Variation According to Galaxy Stellar Mass and Star-formation Activity, arXiv:2002.07334
Mukae, S., Ouchi, M., Cai, Z., et al. 2019, ApJ, 896, 45
Mukae, S., Ouchi, M., Kakiichi, K., et al. 2017, ApJ, 835, 281
Nagamine, K., Shimizu, I., Fujita, K., et al. 2020, 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
Noirot, 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., Hayashi, M., et al. 2017, MNRAS, 468, L21
Shimakawa, R., Kodama, T., Tadaki, K. I., et al. 2014, MNRAS, 441, L1
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, ApJL, 748, L21
Springel, V. 2005, MNRAS, 364, 1105
Springel, V., Frenk, C. S., & White, S. D. M. 2006, Natur, 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
Suzuki, N., Tytler, D., Kirkman, D., O'Meara, J. M., & Lubin, D. 2005, ApJ, 618, 592
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., Peeples, 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