Correlation between Galaxies and IGM at $z \approx 2$ Mapped by Subaru/HSC

A dissertation presented

by

Yongming LIANG

to

The Department of Astronomical Science

in partial fulfillment of the requirements

for the degree of

Ph.D. Thesis

in the subject of

Astronomy & Astrophysics

The Graduate University for Advanced Studies, SOKENDAI

NAOJ, Mitaka, Tokyo

June 2022

© 2022 — Yongming LIANG

All rights reserved.

Dissertation Advisors:

Prof. Nobunari Kashikawa (Univ. Tokyo), Prof. Masayuki Tanaka,

Prof. Yuichi Matsuda

Yongming LIANG

Correlation between Galaxies and IGM at $z \approx 2$ Mapped by Subaru/HSC

Abstract

The relation between galaxies and intergalactic medium (IGM) attracts great interest nowadays because it bridges galaxy formation and large-scale structure formation. Influenced by gravitational instability, the relatively dense fluctuation began to collapse from the earliest homogeneous HI density field. Therefore, galaxies preferentially form in high-density regions where more pristine HI in IGM sinks into massive potential wells. As large-scale filaments are the largest HI reservoir, the intersection of these filaments have provided an ideal platform to accrete matter along cold streams. Therefore, overdense regions in our universe are the most vital laboratories to inspect the galaxy and IGM HI correlation. Nevertheless, to unveil the complete picture of the galaxy formation concerning this topic, one must overcome cosmic variance to address the possible field variation of the interplay between the galaxy and IGM.

In this thesis, we first search for high-density regions using a novel technique that utilizes the grouping of strong IGM Lyα absorption systems imprinted on the SDSS/BOSS quasar spectra as the tracers. To expand the classes of different galaxy environments for discussing field variation, we have also collected field candidates that enclose grouping quasars with $\gtrsim 5$ within 40×40 cMpc^2 sky area. Then we perform the narrowband imaging observations with the filters NB387 and NB400 on Subaru/Hyper Suprime-Cam to map the $z = 2.18$ and $z = 2.29$ Lyα emitters (LAEs) in six fields, BOSS0210, BOSS0222, BOSS0924, BOSS1419, BOSS0240, and BOSS0755. Based on the narrowband images, we select LAEs using the narrowband excess technique, and finally, we have built up a catalog consisting of 3,687 LAEs over a survey area of ~ 8 deg^2. Moreover, there is a recent opportunity from the Subaru HSC-SSP survey, and a new $z = 2.18$ LAE catalog consisting of 3,720 LAEs based

on the NB387 survey in the Deep/Ultra-Deep fields, i.e., DEEP2-3, E-COSMOS, and the XMM-SDSS, which covers the sky area of $\sim$ 19 deg^2. We combine both catalogs to discuss the field variation of the Lyα emitters and the IGM HI correlation. This sample is the largest in terms of the LAEs and the covered sky area, giving us unprecedented power to address this topic.

Based on the LAE catalog, we have constructed the LAE overdensity maps for nine HSC fields, including the HSC-SSP Deep/Ultra-deep fields and those traced by IGM HI absorbers or grouping quasars at the redshifts $z = 2.18$ and $z = 2.29$. These maps have identified structures such as large filaments, clumpy overdensities, and voids. We perform the effective optical depth of the line-of-sight/LoS – the LAE overdensity ($\tau_{\rm LoS}$–$\delta_{\rm LAE}$) correlation analysis with the overdensity estimate. $\tau_{\rm LoS}$ is measured on 15 h^{-1}Mpc scales in the background SDSS/BOSS quasars. This correlation is further transformed into the Lyα forest transmission fluctuation – LAE overdensity ($\delta_{\langle F\rangle}$-$\delta_{\rm LAE}$) relation, whose slope can be compared among fields to inspect how the Lyα absorption is dependent on the local galaxy overdensity. As a result, we find a moderate and positive correlation between LAE overdensity and IGM HI absorption based on the Lyα absorber-trace fields, but the $\delta_{\langle F\rangle}$-$\delta_{\rm LAE}$ slope is much flatter in general fields from HSC-SSP and in the grouping quasar fields. This result implies the variation in the galaxy–IGM HI correlation and has revealed that the IGM HI absorption is stronger at the given $\delta_{\rm LAE}$ in the HI-rich regions compared to general fields. Meanwhile, grouping quasars may also play a role via the feedback to suppress the IGM HI, as the BOSS0210 that contains a significant quasar overdensity can essentially flatten the slope of the sample in the Lyα absorber-trace fields and make the slope more consistent with model predictions. However, this effect is weaker in the general fields because of the initially thin IGM HI in the mean regions at $z \approx 2$, and therefore, grouping quasar fields can have a similar $\delta_{\langle F\rangle}$-$\delta_{\rm LAE}$ slope to that of general fields.

We also perform correlation analysis using the two-point cross-correlation functions (CCFs) between the high-/low-$\tau_{\rm LoS}$ lines-of-sight (LoSs) and LAEs. From the CCFs, we find the positive correlation signal between the high-$\tau_{\rm LoS}$ LoSs and LAEs up to a scale of 4 ± 1 pMpc in the Lyα absorber-trace fields, and the measured values can be well modeled by a power law. This suggests that the LAEs reside in the HI-rich regions up to a scale of 4 pMpc. By comparing the original CCFs among different fields, we find that the data points of CCFs in all samples broadly span around a power law and the optimized power-law fitting gives consistent clustering lengths r_0 among different samples, which suggests that the galaxy–IGM HI correlation is dominated by underlying halos regardless of the various environments. However, the only negative signal of the CCF between high-$\tau_{\rm LoS}$ LoSs and LAEs at $d \approx 200$ kpc

showing up in the grouping quasar still emphasizes the effects of quasar feedback, at least on the circumgalactic scales, corroborating the suggestion from an analysis of the fluctuation of average $\tau_{\rm LoS}$ profile $\delta_{\langle\tau\rangle(d)}$ results.

Throughout this thesis, we confirm the correlation between galaxies and IGM in HI-rich environments at $z \approx 2$, the epoch of the *Cosmic Noon*, with the largest sample size and the widest survey area. These unprecedented samples enable us, for the first time, to witness the possible existence of the field variation of the galaxy–IGM HI interplay. The steeper slope of $\delta_{\langle F\rangle}$-$\delta_{\rm LAE}$ relation that we find in the Lyα absorber-trace fields, and that is flattened with the emergence of grouping quasars can be explained as a result of the exceeding HI supply from the large-scale filaments, which causes the stronger IGM Lyα absorption in high-density regions than in the field regions. Alternatively, the slope can also hint at the deficient LAE detection due to HI suppression, stressing the importance of studying different galaxy populations in surveys. However, the similarity of CCFs and their power-law fitting among different samples suggest the dominance of the gravitational potential of underlying halos in the galaxy–IGM correlation. Combining the two analyses, we conclude that despite the environments, dark matter halos still dominate the galaxy–IGM correlation at $z \approx 2$, although exceeding HI supply and AGN/quasar feedback can play secondary roles. The results from this thesis provide us with a unique key to open the door to understanding galaxy formation in the context of large-scale structure formation.

Contents

Acknowledgments

It is a memorable time to come to this step, which means that I am about to finish my five-year doctoral studies at NAOJ after all those precious journeys and charming sceneries of the universe.

I definitely should show my most tremendous appreciation to my supervisor, Prof. Nobunari Kashikawa, who I always called Kashikawa-san. But today, I sincerely prefer to call him, Kashikawa-sensei. He has always shown great enthusiasm and patience in offering research guidance to me in these years. His gentleness and kindness can easily warm up a child with depression about his failure in doing research or with excitement when finding something inspiring. Besides his personality, he is also professional and experienced as an astronomical researcher. I still remember the first day I met him before entering the doctoral course; he already encouraged me to apply for the time of Subaru telescope for doing my future research. As a student majoring in astronomy in my undergraduate study, I have always heard about large telescopes like Subaru, Keck, or VLT, but I have never gained experience in realistically working with such state-of-the-art instruments. 2.4-m telescope once astonished me when I was a third-grade undergraduate. However, when it looked like a dream to me that day, I came to know that Kashikawa-sensei was serious from the very beginning. Dreams become a reality. My research throughout the five-year doctoral study has always intensively relied on Subaru, and its most advanced instrument at that time, the Hyper Suprime-Cam. Since the first semester of my doctoral course, Kashikawa-sensei has led me to join an exciting project to spectroscopically confirm the most distant galaxy to date, the GN-z11, using the Keck/MOSFIRE in Hawaii, USA. Again, a name I have always heard in an astronomy class that makes me think it is a dreamland comes out, but I got to reach it and work there for something that keeps motivating me, the universe constructed by galaxies and cold gas. During that period, I began to consider my research and was encouraged by Kashikawa-sensei to write my proposals. Step by step, I gradually grow to be a guy who can think deeply about the pros and cons of every instrument and choose the most proper one to develop my interest topics. For five years, dreams become daily. I feel fortunate to have had a chance to work with him, which has made me progress both in academics and being a standout person.

I am also thankful to those friendly people at NAOJ. Since Kashikawa-sensei moved from NAOJ to the University of Tokyo in May 2019, Prof. Masayuki Tanaka has become my main supervisor and provided me with plenty of kind assistance without complaint. Although I have not worked with him on a dedicated research project, he is generous and always willing to give me an instructive direction when I come to him for suggestions in my studies. A person as busy as Tanaka-sensei,

who should be responsible for the ongoing HSC-SSP survey, who can still pay patient concerns to students, should win every respect of us who benefited a lot. I also appreciate the colleagues that we have closely worked together at NAOJ, including Drs. Hisakazu Uchiyama, Kei Ito, Masafusa Onoue, and Jun Toshikawa all graduated from the Kashikawa group. Since the first day I came to Japan and NAOJ, I have received countless help and support from these cool guys. They helped me settle down in a new country, let me know the tips for living a happy life, guided me through every tricky but essential research problem, and were always willing to share the most exciting things with me. Also, I have got lots of support from many other people, including Professors, and Drs. Matsuda, Komiyama, Shimakawa, Imanishi, Izumi; the seniors, Drs. Kikuta, Matsuno, Michiyama, Watanabe, Tsukui, Namiki; the juniors, Kasagi-san, Nagano-san, Masai-san, Takenmura-san, Ikeda-san, Naufal,; the staffs, both who have left NAOJ and who are still staying, Suzuki-san, Suehiro-san, Megumi-san, Shirato-san, Inoue-san, Mastuda-san, Omura-san, Shibayashi-san ... there are many many people that I cannot count completely, that have left great memories with me. It is you who makes my life at NAOJ never dull!

Outside NAOJ, I have encountered many people who share the same memories. The young people of the University of Tokyo in the Kashikawa group can always make something new and beyond expectations. Ishimoto-san is the first student I met in this group from the University of Tokyo, and then Yoshioka-san, Takeda-san, and Arita-san. All of them are talented and active in proceeding at the frontier. Working with these young and smart guys is really a pleasure. I have also benefited a lot from working with my collaborators, and they can be from anywhere in the world! I am thankful to Drs. Zheng Cai, Xiaohui Fan, Xavier Prochaska, Kazuhiro Shimasaku, Kentaro Nagamine, Ikko Shimizu, and Dongdong Shi. All of them are the top experts in dedicated fields in this world. A chance to co-work with them has improved me a lot, thanks to their visions, knowledge, experience, and intelligence. I cannot accomplish my thesis without their genius support.

Finally, I want to thank my family from my heart. Because of the pandemic situation from 2020, the world has been isolated for more than two years, and I have not been home for three years until my graduation. However, my families always show me the warmest part of the world and always keep concerning everything about me, which makes me not alone in this Ph.D. journey. I am sincerely grateful to have their understanding. Also, my beloved girl has never complained, even if I cannot accompany her. She has always warmed my heart, no matter what I went through. I am fortunate to have a precious angel!

Dedicated to the Lovely Universe.

Chapter 1

Introduction

1.1 Background

1.1.1 Structure Formation

In the context of the current standard model of cosmology (Springel et al. 2006), the present universe and its structures form from uniformly distributed matters in the early phase shortly after the Big Bang. At the redshift, $z \sim 1100$, the epoch of so-called cosmic recombination, electrons, and protons for the first time combine as the universe gets cooler. Then space is filled with homogeneous neutral hydrogen (HI). The Cosmic Microwave Background (CMB), measured as an isotropic and homogeneous temperature map of ~ 2.7 K for all-sky, is the remnant of the cosmic recombination. With a more precise measurement available from satellites WMAP (Hinshaw et al. 2007) and the later Planck (Planck Collaboration et al. 2018), we gradually learn that slight fluctuations are lying in the temperature map with an amplitude of an order of 10^{-5} K, hinting the earliest seed of the structure formation.

Such earliest fluctuations are just like ripples. As time goes by, gravitational effects gradually amplify these ripples into surges and collapse to the more massive, condensed, and fragmentary halos of dark matter and gas. With more complicated processes, e.g., gas cooling, the first stars and galaxies emerge at $10 < z < 30$ when gas collapses to trigger the star formation. The gravitational instability will also result in a hierarchical schematic for mass assembly (Baugh 2006), in which dark matter halos gradually assemble into more giant halos through mergers. More massive structures are then formed at large scales of or even over ten cMpc scale, which we call large-scale structures (LSSs). LSSs can be clumpy, filamentary, or sheet-like in the universe. The

later formation of galaxies is predicted to preferentially occurs along filamentary or sheet-like mass overdense regions at high-z in theory, making the intersections of such kinds of structures evolve into the present-day dense galaxy cluster (Bond et al. 1996; Cen & Ostriker 2000). An example is shown in Figure. 1.1 from Cen & Ostriker

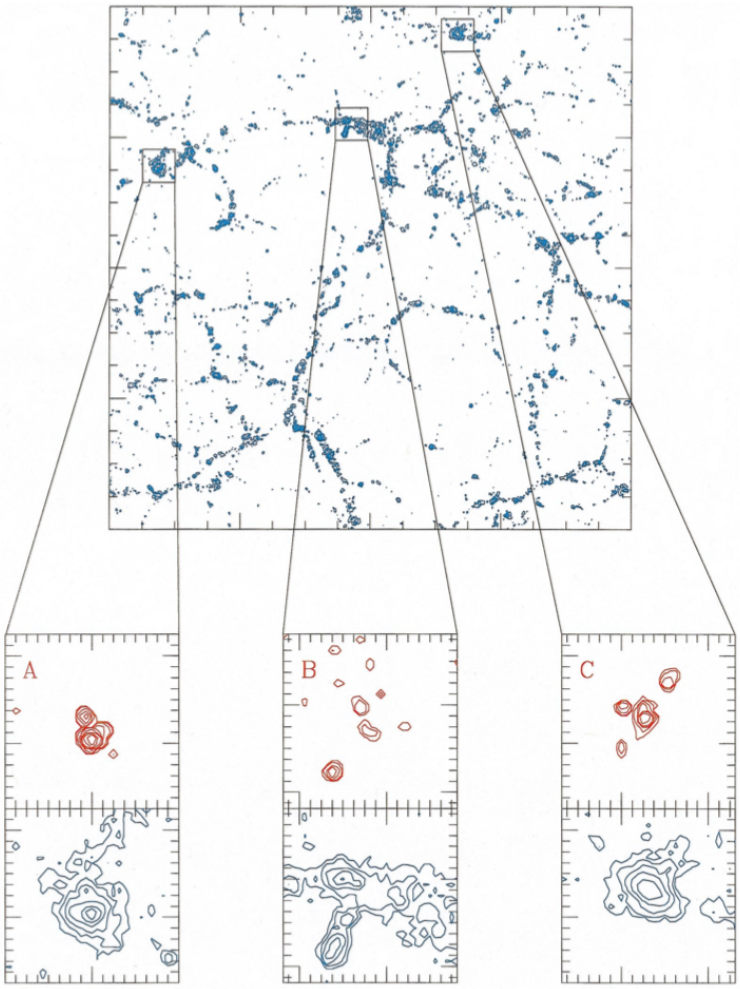

Figure 1.1: Dark matter density at redshift $z = 0$ from numerical simulation (Cen & Ostriker 2000). The top panel is a slice of a snapshot with size $100 \times 100\ h^{-2}$ cMpc^2 and a thickness of 586 h^{-1} kpc. The three zoom-in plots in the lower panel show groups of galaxies, the red contour shows galaxy density, and the blue contour shows the corresponding dark matter density.

(2000). The upper panel shows dark matter density in a $100 \times 100\ h^{-2}$ cMpc^2 box with a thickness of $586h^{-1}$ kpc from the simulation. Clear filamentary structures are seen among the whole box. And some clumpy nodes can also be easily found at the intersection regions. Zoom-in figures for three of such special regions are also plotted in the lower panel, with an upper row indicating the density contour of galaxy

particles with red curves and a lower row showing the density contour of dark matter halos with blue curves. The figures prove a large halo in the node structures can host several groups of galaxies, or sometimes a galaxy cluster, in a confined region, making it overdense compared to the cosmic mean.

Such kind of overdensities can be a very interesting laboratory for exploring or testing the current model of structure formation as well as the galaxies' formation history, especially at $z \sim 2-3$. The earliest mature galaxy cluster is found up to $z = 2.5$ (Wang et al. 2016) with X-ray emission from a hot intracluster medium (ICM), which means the progenitor of such kind of structures should appear at an earlier period. While we consider the galactic scale problems, like the star formation − density and morphology − density relations found at $z \sim 1$ (Elbaz et al. 2007), higher redshift at $z \sim 2$ should be also critical to unveil when they come into effect, as there is no strong evolution at $z < 1$ is reported (De Lucia et al. 2012). Especially if we look back at the cosmic star formation history, we will find that the SFR density peaks at $z \sim 2$, which means star formation activities at the period contribute a significant role in building up stellar masses in galaxy clusters we found today. Motivated by this, we focus on $2 < z < 3$, the *Cosmic Noon* epoch, in this thesis.

1.1.2 High-z Overdensity/Protocluster

Though galaxy cluster progenitors, protoclusters[1], at $z \sim 2-3$ are interesting targets to understand structure formation, it is not an easy mission to search for them. Chiang et al. (2017) has used Millennium cosmological N-body simulation (Springel 2005) with two semi-analytic models to estimate the (proto)cluster occupied fraction of cosmic volume from $z = 0$ to $z = 7$. A snapshot for a slice of the simulated box at $z = 2$ is shown in the left panel of Figure 1.2, where the box size is $200\times200\times50$ cMpc^3 with red points denoting the most massive core halos and blue surfaces indicating the Lagrangian boundaries of protoclusters. The right panel is a very vivid plot for us to imagine how rarely protoclusters distribute in the universe and how small the volume they occupy. Quantitatively, we can also confirm that the filling factor is only $\sim 2\%$ at $z \sim 2$ from the right panel.

That means it will be inefficient if we randomly point the telescope to the sky to look for protoclusters. Therefore, various kinds of tracers for hinting at the possible existence of protoclusters are used in surveys of protoclusters. Some works try to use radio galaxies/radio loud active galactic nucleus (AGN) (Cooke et al. 2014;

[1]Normally a protocluster is defined as a structure that will collapse into a galaxy cluster on the present day, i.e. a virialized object of 10^{14} $M_\odot$ at $z \sim 0$.

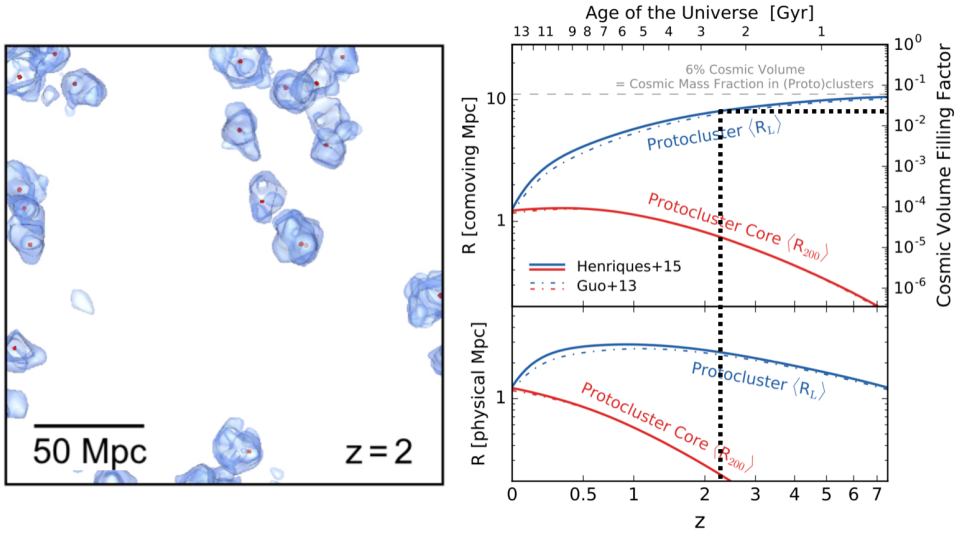

Figure 1.2: *Left panel*: Cosmic volumes occupied by protoclusters of $M_{z=0} > 10^{14}$ $M_\odot$ at $z = 2$ in a slice of $200 \times 200 \times 50$ cMpc. Blue surface shows the Lagrangian boundaries of protoclusters, while red dots are the most massive core; *Right panel*: Average size evolution of protoclusters, blue curves denote protoclusters and red curves shows protocluster cores in comoving and physical coordinates, and the filling factor is also shown in the right-hand y-axis in the upper plot. (Chiang et al. 2017)

Shimakawa et al. 2014; Husband et al. 2016; Noirot et al. 2018), dusty star-forming galaxies (DSFGs) (Casey et al. 2015), luminous quasi-stellar objects (QSOs) (Kikuta et al. 2019)/QSO pairs (Onoue et al. 2018), or sub-mm galaxies (Miller et al. 2019; Oteo et al. 2018) as tracers, because such kind of extreme objects are expected to reside in the center of massive halos. Besides the rare objects, gaseous systems are good tracer candidates, but such systems tend to have a physical connection with close host galaxies. Ogura et al. (2017) have found a possible $z \sim 2.3$ overdensity by using damped Lyα system (DLAs) as a tracer, which also works for a higher redshift search at $z \sim 3$ (Fumagalli et al. 2017). Not only does the absorption system, Bădescu et al. (2017) report a successful discovery of a protocluster associated with a pair of the extended nebular emission system around Lyα emitters, or LAEs, the Lyα Blob (LAB), at $z \sim 2.3$. Blind searches are also plausible when the survey field is large enough for tracers detected with broadband photometry, like the photo-z galaxies (Spitler et al. 2012) or Lyman-break galaxies (LBGs) (Toshikawa et al. 2016, 2018).

The aforementioned searches have managed to find many interesting structures and protoclusters that are important for answering astrophysical questions requiring an extreme laboratory. But if we want to study the large-scale structure formation itself, especially on the relation between the mass assembly history of gas and galaxies,

protoclusters found in these ways have many problems. The protoclusters that are traced by extreme systems, such as radio galaxies or luminous QSOs, are likely biased by the interaction or feedback from these dominant objects (Tumlinson et al. 2017). The number of protocluster cores that have been found tracing sub-mm galaxies has been increasing recently, though it is not clear how ubiquitous they are. Even if the systems are gaseous like DLAs, they are still not the HI reservoirs with high purity in the intergalactic scale; instead, they are probably hosted by a typical halo associated with nearby galaxies (Mackenzie et al. 2019). Meanwhile, photo-z galaxies or LBGs selected in blind surveys always have a considerable redshift uncertainty, which can equal a distance uncertainty larger than the galaxy-HI in intergalactic medium (IGM) correlation scale.

Therefore, another efficient and systematic method for discovering massive overdensities that are bias-free in the mass assembly history is highly desired. We will continue the discussion on this point in Sec 1.3 and look at LSSs from a different perspective in the next section.

1.2 Galaxy-IGM Correlation

1.2.1 Basic Idea on Studying IGM

The most abundant baryon in the universe is hydrogen, and the distribution of hydrogen can also tell us about the structure of the universe. For studying correlations between galaxies and HI, we consider two regimes. Gaseous environment far from any nearby galaxies over 300 pkpc is viewed as part of the IGM, which is more related to the LSSs, while a smaller scale is a circumgalactic medium (CGM), which is dominated by astrophysical processes on a galactic scale. Although the boundary is somewhat arbitrary, 300 pkpc is often adopted as a working criterion (Tumlinson et al. 2017). In this thesis work, we mainly focus on correlations between galaxies and IGM neutral hydrogen. We also have a plan to work on CGM in the future based on the sample built in this work.

The neutral part of IGM at high-z, mainly the HI, has been studied from the absorption imprinted in the spectra of background quasars (or quasi-stellar objects/QSOs). A more common case can be seen from the schematic in Figure 1.3 from Springel et al. (2006), which is the so-called Lyα forest. After the photons are emitted from a background quasar, cosmic expansion elongates their wavelength, i.e., cosmic redshift z. Once they pass through a clump of gas, they will leave a footprint as absorption in the corresponding redshifted wavelength in the spectrum. Therefore, we

can learn the gas properties from the absorption features, such as its column density, temperature, or metallicity if a metal line is available, in the intervening space along the line-of-sight (LoS) towards the background quasar.

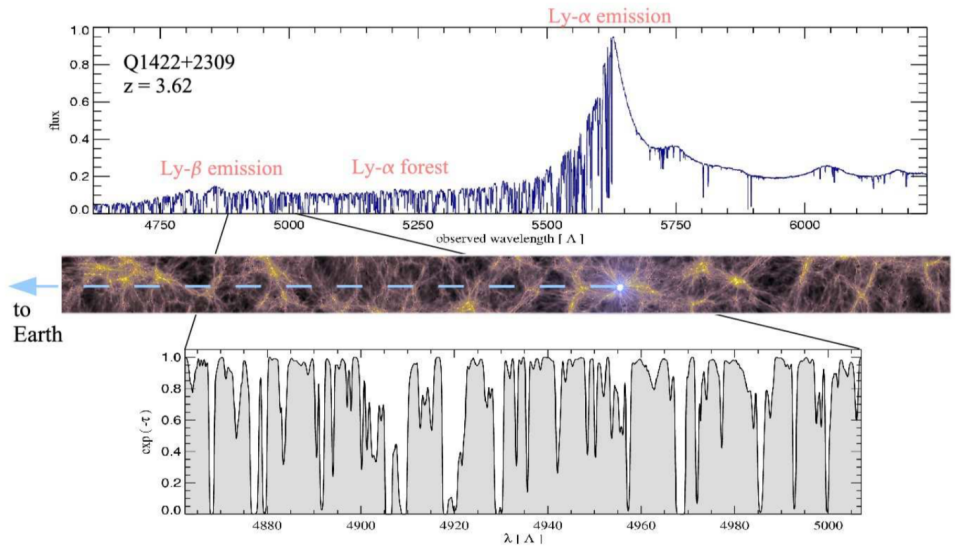

Figure 1.3: Lyα forest as a probe of large-scale structure. *Top panel*: A typical high-resolution spectrum of a $z = 3.62$ quasar. Shortward of the redshifted Lyα emission line at $1216(1 + z)$ Å, the spectrum shows a "forest" of absorption lines of different strengths produced by intervening neutral hydrogen gas along the Los from the quasar to the Earth. *Bottom panel*: Hydrodynamical simulations reproduce the observed absorption spectra with remarkable fidelity, as illustrated by the simulated spectrum, corresponding to intervening large-scale structure at z$\sim$3. *Middle panel*: An example of the gas distribution in a simulated ΛCDM model. (Springel et al. 2006)

If we want to study the correlation between halo mass/galaxy and IGM HI, we need to have appropriate tracers that allow us to infer the underlying mass around the absorption, both in not only the LoS direction, which can only be traced by quasar absorption systems but also the transverse direction (projected separation). Furthermore, the quality of the spectra and the number density of background quasars impose a practical limit on our ability to infer IGM properties. All of these difficulties make it a tough problem, especially at $z > 2$, where both the choices of tracer and the number of high-quality spectra become a real issue.

Despite the difficulties, however, many efforts have been made to push the frontier. High column density systems (HCDs) of HI reservoirs hinted in background quasar have been wildly used as an indicator of IGM/CGM gas. Prochaska et al. (2013) have searched for quasar pairs with projected separation $R_\perp < 1$ pMpc from

Sloan Digital Sky Survey/Baryon Oscillation Spectroscopic Survey (SDSS/BOSS) to use the spectra of background quasars to reveal the environment of the foreground quasars. As a result, they find the most strong and optically thick HI absorbers with column density $N_{HI} > 10^{17.3}$ cm^{-2} show a very high covering fraction within a $R_\perp < 200$ kpc. The fraction decreases by 20% as the $R_\perp$ becomes larger, although still significant up to $R_\perp \sim 1$ pMpc. They conclude that HI does trace LSSs instead of getting illuminated by foreground quasars, given that a smaller impact parameter yields a stronger absorption. Recent work also focuses on the extreme HI reservoirs, e.g., the Damped Lyα Absorption systems (DLAs), by using the VLT/MUSE integral field spectroscopy to reveal the vicinity of 6 DLAs selected purely by HI absorption (Mackenzie et al. 2019). They search for associated LAEs around the DLAs up to a large impact parameter of 280 kpc. Three LAEs they find around a metal-poor DLA may be too far to be the host, but the nearly one-plane locations suggest they are tracing a massive halo that hosts the DLA in comparison to simulations. Though these researches are not directly related to the IGM scale, they leave a potential message on the galaxy–IGM correlation study.

1.2.2 3D IGM Tomography and the Galaxy–IGM HI Relation

Explorations have been made for a long time to reconstruct the spatial distribution of IGM HI gas from the Lyα forest of background sources (Caucci et al. 2008). Background sources are often quasars, given their luminosity and well-modeled power-law continuum. However, the number density of high-z quasars is not sufficient to reconstruct a robust map due to a coarse spatial resolution. Luminous LBGs can be an alternative, but LBGs are typically fainter than $g > 23$, which takes unreasonable time to get high-resolution spectra even with 8-10 m telescopes. However, Lee et al. (2014a) demonstrate that a moderate spectral resolution will be sufficient for reconstructing the 3D IGM distribution, allowing a reasonable exposure time for large telescopes to take spectra for the more abundant though fainter high-z star-forming galaxies.

They have made a pilot survey to demonstrate this technique in a $5' \times 4'$ patch in the COSMOS field (Lee et al. 2014b). The reconstructed map for 2D slices and 3D volume are shown in the left panel of Figure 1.4. Following results over a larger area are available in Lee et al. (2016), one of which is shown in the right panel of Figure 1.4. Although their target is a blank field with a footprint of only $11.5' \times 13.5'$, they find a galaxy overdensity by chance after collecting the literature on galaxies, e.g., the MOSDEF survey (Kriek et al. 2015). It is likely due to the large survey length

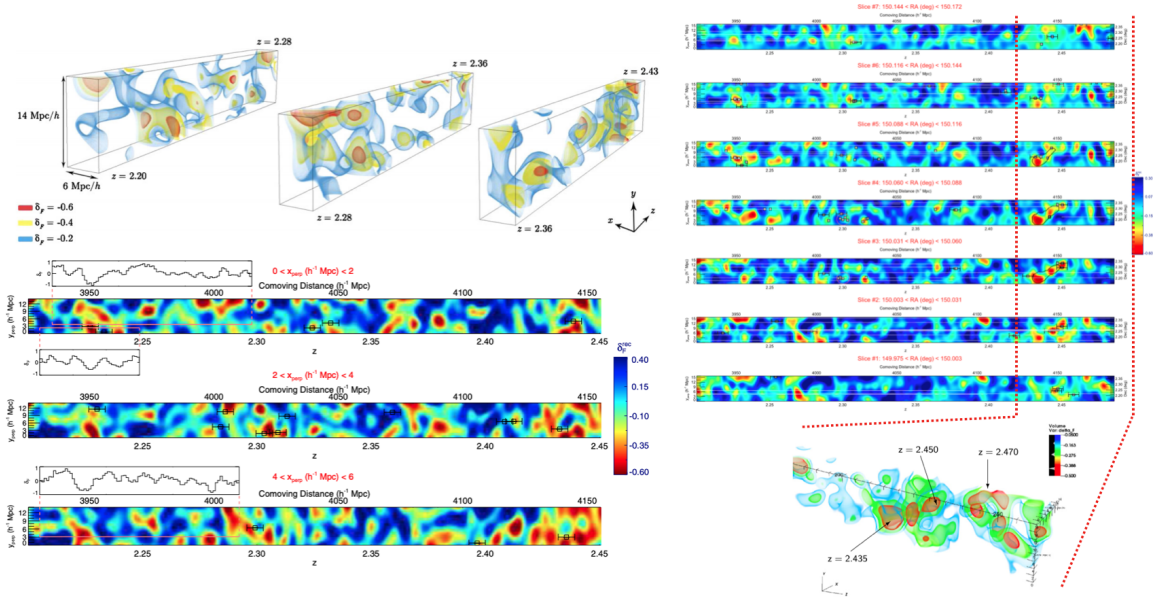

Figure 1.4: Projections of Lyα forest tomographic map and the corresponding reconstructed 3D map, where color is scaled by the Lyα forest transmission. *Left panel*: Data from 24 LBGs over a $5' \times 5'$ region (Lee et al. 2014b); *Right panel*: Data accepted from 58 LBGs over a $11.5' \times 13.5'$ region (Lee et al. 2016).

along the redshift direction.

Interestingly, while in most cases, galaxies distribute in consistency with the high-density IGM HI in the IGM tomography map, they do not find the corresponding strong Lyα absorption peak in a significant overdensity of coeval galaxies at $z = 2.300$ (the pink circle in the right panel of Figure 1.4). This seems to suggest a general correlation between IGM HI and the galaxies with some forms of field variation. They also put forward a larger survey called COSMOS Lyman-Alpha Mapping And Tomography Observations (CLAMATO) survey for mapping the $z \sim 3$ IGM in a ~ 1 deg^2 area of the COSMOS field and the first data release is now available (Lee et al. 2018).

Besides CLAMATO, another multi-object spectroscopy survey, the Lyα Tomography IMACS Survey (LATIS), is also carried out with the instrument Magellan/IMACS to map the IGM HI at the redshift interval of $z = 2.2 - 2.8$ in three of the Canada-France-Hawaii Telescope Legacy Survey (CFHTLS) Deep fields, including COSMOS (Newman et al. 2020). The LATIS has accomplished 1.5 deg^2 of the sky, corresponding to a survey volume of 3.7×10^6 h^{-3}cMpc that is ten times larger than that of the present CLAMATO, enabling the IGM HI map to find 149 absorption peaks with Lyα flux contrast $\delta_F < 2\sigma$ and 76 of them are expected to be the pro-

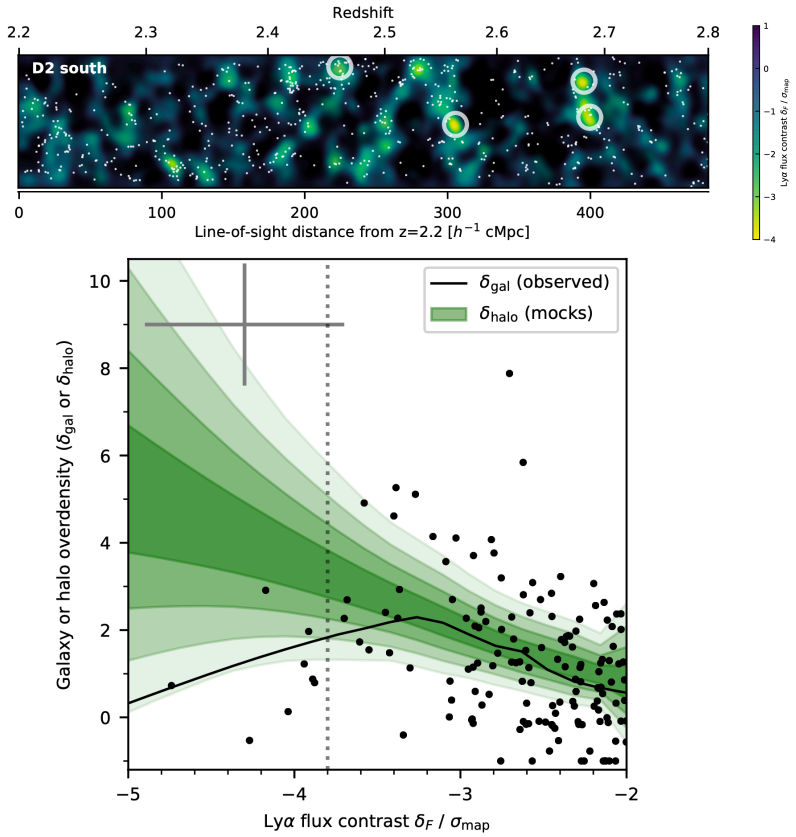

Figure 1.5: *Upper panel*: An example of the IGM map from LATIS. The background contour suggest the Lyα flux contrast δ_F smoothed by a Gaussian kernel with standard deviation 4 h^{-1}cMpc, and yellow means low δ_F, where the density of IGM is high. The points are the positions of LATIS galaxies with spectral confirmation in the survey, and the strongest absorption peaks are encircled. *Lower panel*: The galaxy overdensity $\delta_{\rm gal}$ at the location of Lyα absorption peaks. Black dots suggest each observed absorption peak, while the black curve represents the trendline. The green bands are the trendlines of 1, 2, 3, and 4 standard deviations of mock surveys from the IllustrisTNG simulation that mimic the LATIS survey. (Newman et al. 2022)

toclusters of galaxies (Newman et al. 2022). The survey volume also includes 2,241 LBGs and quasars with spectroscopic redshifts, which are used for the correlation analysis of IGM absorption – galaxy overdensity.

An example of the IGM map reconstructed in LATIS is shown in the upper panel of Figure 1.5, where the yellower background contour suggests the stronger Lyα flux contrast δ_F on a scale of 4 h^{-1}cMpc. Several strongest absorption peaks are shown by white circles, and the location of galaxies is denoted as white dots. In the moderate to relatively strong absorption regions, one can visually find that the galaxies roughly

have consistent distributions. At the same time, they seem to be absent at some of the most substantial absorption peaks. This result is consistent with the suggestion from CLAMATO survey (Lee et al. 2016). Thanks to the large survey volume, the IGM maps from LATIS allow statistical analysis between the Lyα flux contrast and the local galaxy overdensity $\delta_{\rm gal} = N/\left(\overline{N}V\right) - 1$, where N, $\overline{N}$, and V are the number of galaxies within a cylinder with a radius of $R = 8\ h^{-1}$cMpc and length of $2R$, the mean galaxy space density, and the search volume. The lower panel of Figure 1.5 shows the correlation results, with comparisons to a model prediction from the mock surveys based on IllustrisTNG simulation (Nelson et al. 2019). As expected from visual inspection, at the locations with Lyα flux contrast $-3.8 < \delta_F < -2$, the $\delta_{\rm gal}$ shows out anti-correlation with Lyα δ_F, which means the galaxies distribute more in the region with denser IGM, and this trend is also consistent with the model prediction. However, at the extremely dense region with $\delta_F < -3.8$, this relation seems to be broken and owns an offset to the model predictions. As an interpretation, Newman et al. (2022) suggests there are possibly missing galaxy populations in the UV-selected galaxies in these extremely high-density regions, such as the extremely dusty or quiescent galaxies, or rather the moderately UV-dim galaxies. Again, this result is likely another proof of the field variation of the galaxy-IGM HI correlation.

1.2.3 Other Methods on the Galaxy–IGM HI Relation

One can rely not only on the 3D IGM tomography but also use other methods to estimate the Lyα absorption at specific redshifts with known coeval galaxies to inspect the correlation. Mukae et al. (2017) probe larger scale statistics by using photo-z galaxies at $2 < z < 3$ and Lyα absorption in 16 SDSS/BOSS LoS over an area of 1.62 deg^2 in the COSMOS/Ultra VISTA. From such a global view, they discover a moderate correlation between the galaxy overdensity and spectral fluctuations in the quasar spectra at the corresponding redshift range, albeit with a significant scatter. They also perform a simulation for further investigation on the extreme cases among their LoSs; they find that the large scatter can be a natural result of the LoSs occasionally passing through a gas filament, a large void, or an orthogonal filament with low density. The simulation result from Mukae et al. (2017) can also be one reason for the non-detection of local overdensity dependence for ΔNB497 in Mawatari et al. (2017). The Keck Baryonic Structure Survey (KBSS) has taken high SNR spectra of 15 brightest quasars at $2.5 < z < 2.9$ as well as 886 foreground UV-color selected KBSS galaxies at $2.0 < z < 2.8$ with an impact parameter ranging from 50 pkpc ~ 3 pMpc. Rudie et al. (2012) present their results on the surrounding HI properties as a

function of separation to the nearby LoSs. They find N_{HI} is strongly correlated with galaxies within a CGM scale (< 300 pkpc). At the separation between 300 pkpc $-$ 2 pMpc, the correlation becomes weaker but still over a level of random IGM field. However, both papers summarized above are still limited in terms of the survey area and the dynamic range of environments, i.e., the overdensity range. Furthermore, Rudie et al. (2012) focus on relatively bright galaxies and photo-z galaxies in Mukae et al. (2017) have a large redshift uncertainty.

Besides the statistical study in general fields, a well-studied protocluster, the $z = 3.1$ SSA22 (Yamada et al. 2012), is also worth mentioning. Mawatari et al. (2017) have put forward a unique method by using the offset ΔNB497 between the magnitude of their narrowband (NB) $\mathrm{NB497_{obs}}$ as the indicator of gas richness, in which the Lyα absorptions at $z \sim 3.1$ are expected to come, and the template narrow band $\mathrm{NB497_{temp}}$ predicted from the best-fit spectral energy distribution (SED) fitting as the indicator of gas richness. With this technique, no expensive spectroscopy for bright objects is required, but just NB image and a 2-D absorption will then be available. They apply the technique to the SSA22 data and also to two control fields, GOODS-N and SXDS. They find that the whole SSA22 field shows a systematic ΔNB497 excess compared to the control fields on a ~ 50 cMpc scale, while there is no dependence on the local overdensity at several cMpc scales. This result indicates the extended HI gas in the overdensity cannot be due to individual CGM of each galaxy but instead may be in a pre-heated phase as warm–hot intergalactic medium (WHIM) or ionized by the galaxy feedback in such an extreme overdense environment. It partly agrees with another exception reported in Lee et al. (2016). The galaxy-IGM HI correlation in SSA22 is also investigated along the LoS dimension by a spectra-stacking study reported (Hayashino et al. 2019). Interestingly, they find two Lyα absorption dips indicating gas-rich regions appear at $z = 3.1$ and $z = 3.29$, where overdensities LBGs are found, while Lyα transparency peaks at $z = 2.98$, is also lacking LBGs. The results thus show a good correlation in the LoS direction, except for $z = 3.455$, where a dip is found, but there is no LBG overdensity, but is only an AGN pair.

1.3 MAMMOTH Project

1.3.1 Methodology

Cai et al. (2016) has put forward a new tracer to meet the requirement discussed at the end of Section 1.1.2 in the Mapping the Most Massive Overdensities

through Hydrogen (MAMMOTH) project. They carry out a large-scale cosmological simulation with the N-body code GADGET (Springel 2005). A deterministic scheme and the LyMAS scheme (Peirani et al. 2014) are used for making mock Lyα forest at $z \sim 2.5$ for comparison. By varying the calculation scale for total mass overdensity δ_m

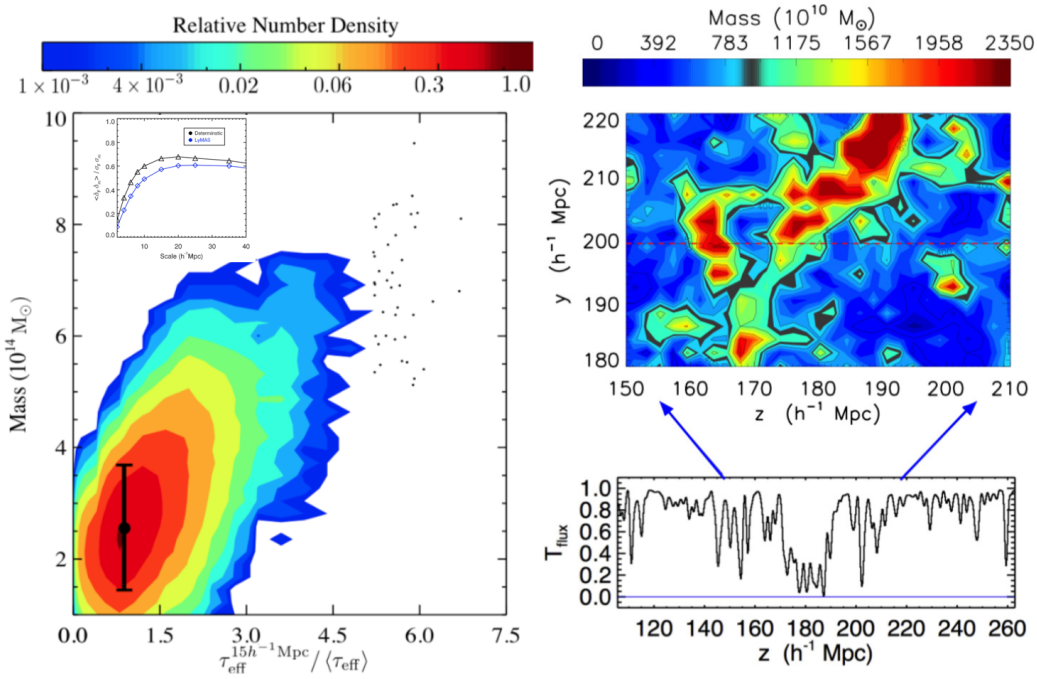

Figure 1.6: *Left panel*: The correlation between the distribution of mass and gas from MAMMOTH simulation. The inset figure shows the scale dependence of the correlation with two sets of simulation models, both of which reach a peak over 15 h^{-1} cMpc. The major figure shows when the scale is fixed at 15 h^{-1} cMpc, the mass within (15 h^{-1} cMpc)^3 is correlated with the $\tau_{\rm eff}^{15h^{-1}{\rm cMpc}}$. The most powerful systems with the highest $\tau_{\rm eff}^{15h^{-1}{\rm cMpc}}$ are plotted as black points, which are named as CoSLAs; *Right panel*: The mock spectrum and the slice corresponding to the LoSs in simulation box, the color is coded by mass. (Cai et al. 2016)

and the fluctuation of the transmitted flux δ_F, they find a good correlation at scales over 15 h^{-1} cMpc, as shown in the inset plot of the left panel in Figure 1.6. When they fix the scale at 15 h^{-1} cMpc to examine the relationship between an effective optical depth $\tau_{\rm eff}$ and the masses within the volume, they find that some extreme systems trace the most massive halos, as shown in the major plot of the left panel in Figure 1.6, where the color represent the contour of the number density of points and the black points at the upper right are the most extreme system with high $\tau_{\rm eff}$. They define the systems with $\tau_{\rm eff} > 4\sigma$ than the cosmic mean as Coherently Strong Lyα Absorption systems (CoSLAs). The CoSLAs can be a good tracer for overdense

regions, and it has the potential to pick up unbiased overdense regions from 1 Gpc^3 by using BOSS background quasars.

The name originated from their nature. When Cai et al. (2016) checks the corresponding spectra and the mass distribution in the simulation box, they find that most of them are a population of unique absorbers. An example is shown in the right panel of Figure 1.6. The upper plot shows the mass contour, and a large massive structure appears on the 15 h^{-1} cMpc scale. The lower panel is the mock spectra of the line-of-sight (LoS) crossing the center of the upper plot, indicated by a red dashed line. No strong damping wing, which are hints of DLAs with high HI column density, is found in the mock spectra. It means CoSLAs are different from DLAs and are not hosted by galaxies. Instead, the strong absorption is a result of the overlap of several absorptions in the groups of gas reservoirs with lower HI column density. This nature makes them trace the underlying halo mass because they are distributed in intergalactic space and free of galactic or AGN feedback.

In Cai et al. (2016), besides the excess in optical depth, they have also developed a set of well-defined criteria for picking out CoSLAs from the real quasar spectra in large redshift surveys, etc. SDSS/(e)BOSS survey (Dawson et al. 2013, 2016). The main contaminants are the high HI column density systems (HCDs), such as DLAs, sub-DLAs, and the Lyman limit systems (LLSs). Details can be found in Cai et al. (2016), but we have a brief summary for the key points of rejecting these contaminants at $z \sim 2.2/2.3$ that we are working on here: (a) The width at flux/continuum=0.8 $\omega_{0.8} < 70\mathring{A}$, for eliminating DLAs whose $\omega_{0.8}$ will be always larger; (b) The mean flux of absorption through $F_{\rm through} > 0.15$, again for rejecting DLAs whose $F_{\rm through} \sim 0$; (c) Non-detection of low-ionization metal lines, for rejecting all HCD tending to be associated. (d)Presents a group of absorbers at the surveying redshift because HCDs tend to trace field galaxies with a smaller cross-section.

1.3.2 Pilot Projects for MAMMOTH

As the pilot project for the implication of MAMMOTH technique, they have searched for CoSLAs from the SDSS-III DR9 quasar catalog containing $\sim$ 8000 spectra over 3000 deg^2 and selected out 11 candidate fields hinted by CoSLAs (Cai et al. 2017b). Follow-up narrowband and broadband imaging, multi-wavelength imaging, and also spectroscopy are done with the KPNO-4m/MOSAIC, LBT/LBC, and LBT/MODS, respectively, for one of the chosen fields, BOSS1441. They picked out LAEs within a field of view (FoV) of size $36' \times 36'$ and found a large and rarely massive overdensity at $z = 2.3$ (see the left panel of Figure 1.7). Spectroscopic confirmations are also done, and 12 out of 16 LAEs in the most overdense peak, indicated as the filled points

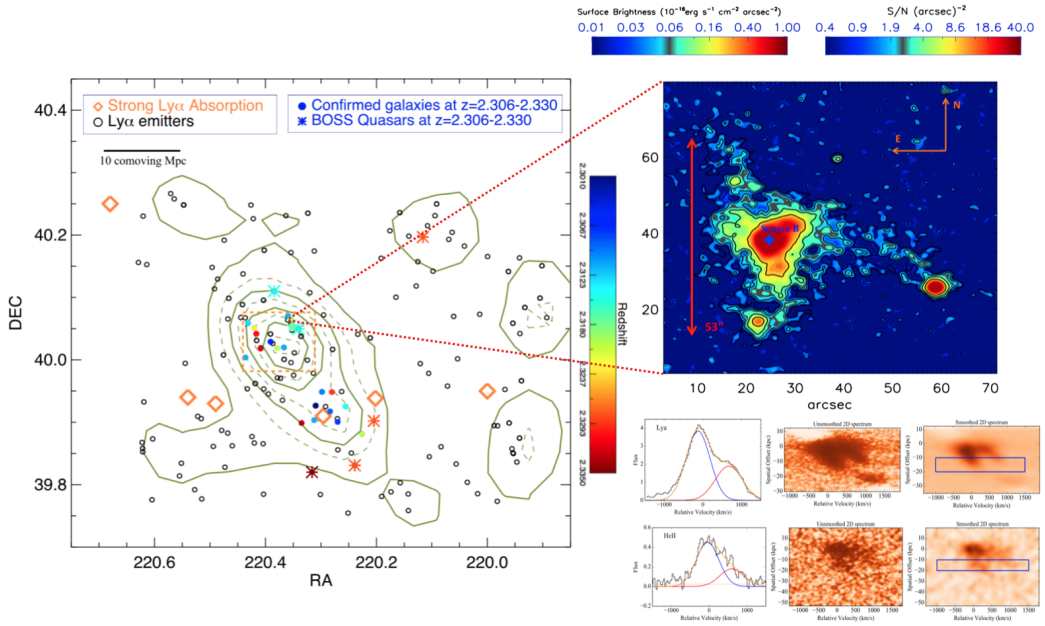

Figure 1.7: *Left panel*: Galaxy overdensity of BOSS1441 at $z = 2.32 \pm 0.02$ selected from the SDSS-III/BOSS data traced by CoSLAs (marked as orange diamonds) and quasars (marked as brown asterisks). The overdensity extends over 30 h^{-1} cMpc at $z = 2.3$. The cylinder at the position of the orange dash box with size of $5' \times 5'$ corresponds to $(15h^{-1}\text{cMpc})^3$ volume, inside which the δ_g =10.8 with 12/16 LAEs spectroscopically confirmed (Cai et al. 2017b); *Right panel*: The extended Enormous Lyα nebulae found at the BOSS1441 overdensity peak, with color code the Lyα emission. It extends over 400 kpc at $z = 2.3$ with a dominant core seeming like a Type-II AGN. Spectrum in the lower panel shows a two components in emission line profile, proving the existence of outflow (Cai et al. 2017a).

in the left panel, have been confirmed. They yield a refined galaxy overdensity of $\delta_g = 10.8 \pm 2.8$ within the yellow box regions, which will likely evolve to a present-day cluster of mass $M_{z=0} \gtrsim 10^{15}\ M_\odot$ (Chiang et al. 2013). Such a rare overdensity is detected with just one pointing with FoV of $\sim 30 \times 30$ arcmin^2, proving the strength of CoSLAs. More interestingly, an enormous Lyα nebulae (ELANe) extending over 400 pkpc is also found near the BOSS1441 overdensity peak, as shown in the right panel of Figure 1.7. The spectroscopic follow-up for one of the most Lyα luminous cores reveals a strongly obscured source, probably a type-II AGN, with a double peak structure in emission lines, indicating outflows. Such discovery demonstrates the structures traced by CoSLAs can not only be used for large-scale structures given the massive overdensity but also possibly offer us special objects for studying the interplay between galaxy and gas in a circumgalactic scale.

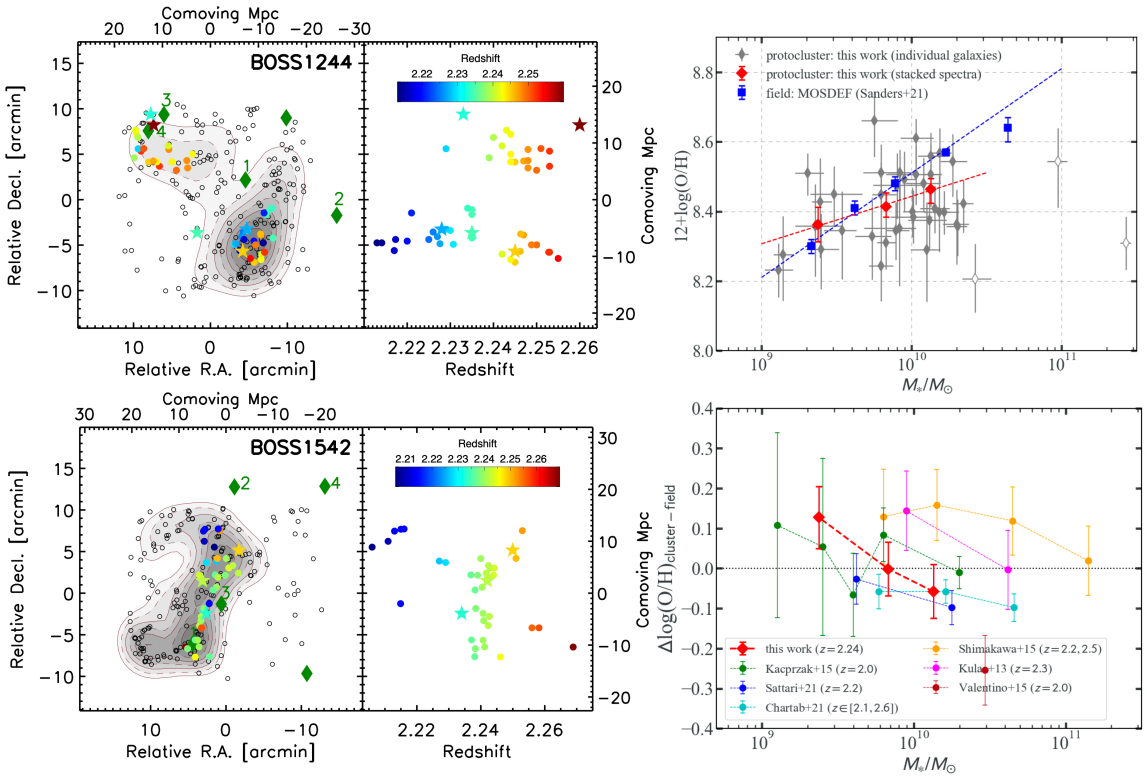

Figure 1.8: *Left panel*: The 3D distributions of the member galaxies in the BOSS1244 (upper panel) and BOSS1542 (lower panel) (Shi et al. 2021). The colored points and stars are the spectroscopically confirmed HAEs and quasars at $z = 2.213 - 2.260$ and $z = 2.206 - 2.269$ in the two fields, respectively. The black circles are the HAE candidates, and the green diamonds are the CoSLAs used for searching the fields. *Right panel*: The metallicity measurements for the star-forming member galaxies in a protocluster in BOSS1244 at the $z \approx 2.24$ (Wang et al. 2022). The upper right panel shows the member galaxies' mass-metallicity relation (MZR). The grey points are the individual galaxies measured in the protocluster, and the red points are the measurement after stacking the spectra with binning in the stellar mass M_*. The lower right panel shows the metallicity offset between protocluster and field galaxies in the MZR compared to the literature results.

Besides BOSS1441 studied by LAEs, two new massive structures at $z = 2.24$, the BOSS1244 and BOSS1542, are found by the CFHT/WIRCam (FoV: $26' \times 26'$) near-infrared NB imaging survey that aims to select Hα emitter ($\lambda_0 = 6,562$Å; HAE) in the regions traced by CoSLAs, which are extracted from the entire SDSS-III/BOSS quasar spectra dataset covering a sky area of 10,000 deg^2 (Zheng et al. 2021). With MMT/MMIRS and LBT/LUCI, a spectroscopic campaign with the high success rate of $\sim 80\%$ in finding Hα emission has confirmed the overdense structure at $z = 2.24$,

as shown with the colored points (galaxies) and stars (quasars) encoded with redshifts in the left panels of Figure 1.8. The grey contour indicate the galaxy overdensity $\delta_{\rm gal}$ and they are as high as $\delta_{\rm gal} = 5.5 \pm 0.7$ and 5.2 ± 0.6 in BOSS1244 and BOSS1542, respectively. Especially in the densest area, the overdensity can even be $\delta_{\rm gal} \sim 20$ (Shi et al. 2021). Based on the galaxy overdensity, the structures are predicted to evolve into the Coma-type galaxy clusters with halo masses of $> 8 \times 10^{15}$ $M_\odot$, making them the most massive structures to date at $z = 2 - 3$.

Follow-up grism spectroscopic observations with the WFC/G141 on the Hubble Space Telescope (HST) are carried out for BOSS1244 by Wang et al. (2022). In addition to further confirmation of the member galaxy redshifts at $z = 2.24$, galaxy properties like the nebular emission lines, stellar masses, instantaneous star formation rates (SFR), and metallicity are also measured. The upper right panel in Figure 1.8 shows the mass-metallicity relation (MZR) of the member galaxies in the two protoclusters, whose slope is shallower than that from the field galaxies at $z \sim 2.3$ from MOSDEF survey (Kriek et al. 2015). The lower right panel shows the comparisons among different $z \leq 2$ protocluster surveys of the metallicity offset in MZR between protocluster and field galaxies. The consistent trend that metallicity is enhanced in low-mass galaxies but slightly deficient in high-mass galaxies residing in overdense regions has suggested the existence of effects including efficient recycling of feedback-driven winds and cold-mode gas accretion in protoclusters (Wang et al. 2022). These findings highlight that the protoclusters, especially those traced by CoSLAs, can provide us with distinctive environments from general fields to test statistics.

1.4 Lyman-Alpha Emitters

Lyman-Alpha (Lyα) Emitters (LAEs), mentioned in the previous sections, are the primary objects representing the galaxies used in this thesis. This section introduces this galaxy population in more detail.

Lyα emission at the rest $\lambda = 1215.67$ Å is produced when the atomic electron transitions from an excited quantum state $n = 2$ to the ground state $n = 1$ (as shown in the upper left panel of Figure 1.9), i.e., a resonant line of the hydrogen (Dijkstra 2014). In astronomical objects, Lyα primarily originates from either the dense interstellar gas recombining when they are ionized by the ionizing radiation emitted by hot young O/B stars or the excited HI gas powered by collision (Dijkstra 2014). Therefore, LAEs at high-z are young galaxies with relatively faint continuum emission. In observations, they usually are very bright in the narrowband while they are very faint in the adjacent broadband(s), and a case of the NB387 ($\lambda_c = 3862$ Å)

and g-band observations carried out by Subaru/Hyper-Suprime Cam (HSC) is shown in the upper right panel of Figure 1.9.

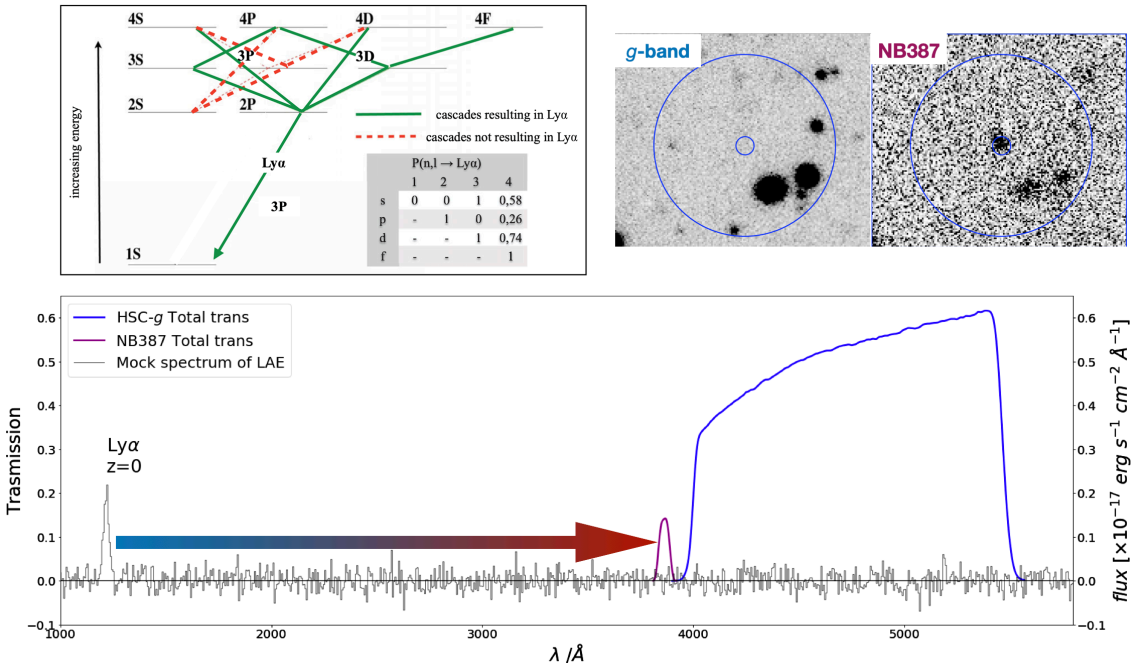

Figure 1.9: *Upper left*: A schematic diagram to show the energy levels of a hydrogen atom (Dijkstra 2014). The Lyα emission comes from the photon emitted when the electron transitions from the quantum state $n = 2$ to the ground state $n = 1$ in hydrogen; *Upper right*: An example of the LAEs detected in the imaging observations. It can be detected in the specific narrowband (NB387 with $\lambda_c = 3,862$ Å in the case for the $z = 2.18$ LAE), while the nearby broadband mapping its continuum finds nothing. *Lower*: A schematic diagram to show the idea of LAE searching. The rest-frame Lyα at $\lambda_0 = 1,215.67$ Å of high redshift objects shifts to a longer wavelength when observed, and therefore, a narrowband sensitive to a specific wavelength can detect the objects at the corresponding redshift with small uncertainty.

Nowadays, LAEs have become a popular choice of galaxies to be studied at the high-z. There are several merits, and one of them is that the young galaxies like LAEs can be more representative of the less massive galaxies with stellar mass $M_* = 10^9 M_\odot$ (Kusakabe et al. 2018) compared to other galaxies used in high-z galaxy surveys, e.g., LBGs and HAEs. It is essential to map the faint and less massive galaxy populations in the large-scale structure studies because they are the iceberg beneath the surface, which are the majority constructing our universe, according to the galaxy luminosity functions (Bouwens et al. 2022). Moreover, LAEs can be economically and systematically searched with relatively small redshift uncertainty by using the

narrowband imaging technique. As the schematic diagram in the lower panel of Figure 1.9 shows, the rest-frame Lyα emission at $\lambda_0 = 1215.67$ Å of high-z galaxies is getting its wavelength shifted to be longer when the universe expands. Then, a narrowband sensitive to a specific wavelength range can map the galaxies with strong Lyα emission only at the corresponding redshift range. At the same time, the adjacent broadbands can detect weaker or no signal with similar imaging depth (the case in the upper right panel). This advantage makes constructing a statistical galaxy sample with well-constrained redshifts feasible in many surveys, both for general fields or for protoclusters (e.g., (Venemans et al. 2005; Mawatari et al. 2012)).

1.5 This Thesis

As described in Section 1.1, it is vital to understand the complete picture of galaxy formation by inspecting the correlation between galaxies and IGM HI, especially in the overdense regions at $2 < z < 3$, and this is the primary objective of this thesis. However, previous efforts described in Section 1.2 at resolving the galaxy–IGM HI correlation either by using 3D IGM tomography with coeval galaxies or by estimating the Lyα absorption at specific redshifts of the known galaxies from the quasar spectra or NB offsets, are either limited in the relatively massive galaxy populations (e.g., LBGs) or small survey area (i.e., up to the order of 1 deg^2). Meanwhile, Sections 1.2.2 and 1.2.3 also show the possible existence of the field variation of the galaxy–IGM HI correlation, which may rely on the local environments of the IGM density or the influence of star-forming or AGN feedback. This stresses the importance of a large sample in different environments to fully consider all these factors.

In this thesis, we first search for the high-density region by the novel technique introduced in Section 1.3 that utilizes the grouping of strong IGM Lyα absorption systems imprinting in the SDSS/BOSS quasar spectra as the tracers. To expand the classes of different galaxy environments for discussing field variation, we have also collected field candidates that enclose grouping quasars, i.e., five quasars within a $40\times$ 40 cMpc^2 sky area from the SDSS/(e)BOSS database. We then conduct narrowband imaging observations with the filters NB387 and NB400 to map the $z = 2.18$ LAEs in four fields, BOSS0210, BOSS0222, BOSS0924, and BOSS1419, and $z = 2.29$ LAEs in two fields, BOSS0240 and BOSS0755, on the wide-field camera Subaru/HSC with a large field-of-view of 1.7 deg^2. The $z = 2.18$ fields are selected primarily by overdensity of strong IGM Lyα absorbers, except for the BOSS0210, which also encloses a group of quasars, and the $z = 2.29$ fields are traced by the grouping quasars. All these samples

expand the classes of different galaxy environments for discussing field variation. Also, these samples have covered 3,687 LAEs over a survey area of ~ 8 deg^2. Moreover, there is a recent opportunity from the Subaru HSC-SSP (Subaru Strategic Program) survey, and a new $z = 2.18$ LAE catalog consisting of 3,720 LAEs is built based on the NB387 survey in the Deep/Ultra-Deep fields, i.e., DEEP2-3, E-COSMOS, and the XMM-SDSS, which cover the sky area of ~ 19 deg^2 (Ono et al. 2021). We now combine all the catalogs in the three categories as summarized in Figure 1.10 together for discussing the field variation on the LAEs and the IGM HI correlation.

The final sample includes the largest number of LAEs at $z \sim 2$ to date, and the sizes of the LoS sample and the survey area are ~ 15 times the previous research. LAEs, rather than other UV-selected galaxies, are representative of the less massive galaxies, which construct the majority of the large-scale structures. Meanwhile, three distinct environments are included for decisive comparisons. These merits have provided us with unprecedented power to address the topic of field variation in the galaxy–IGM HI correlation.

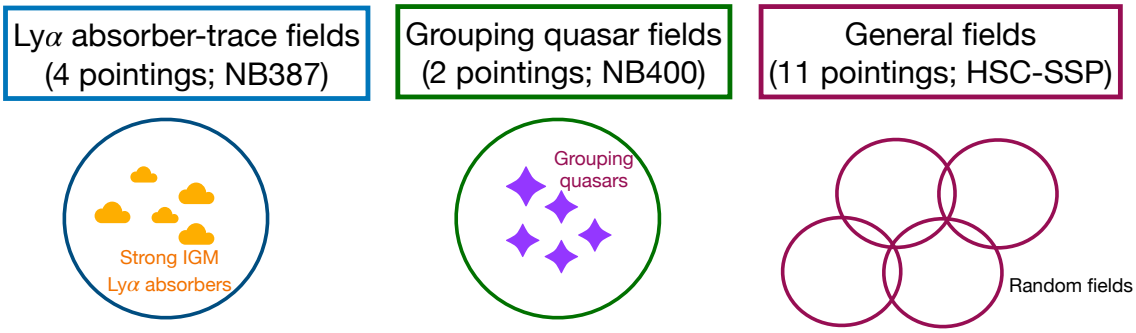

Figure 1.10: The sample categories in this thesis. They are the 4, 2, and 11 Subaru/HSC pointings (~ 1.7 deg^2 per pointing) of grouping Lyα absorbers-trace fields, grouping quasars fields and general fields. Note that the BOSS0210 in the first category also encloses a group of quasars; thus, discussions pay some special consideration to this field throughout the thesis.

In Chapter 2, we introduce the selection of the LoSs from the SDSSS-eBOSS and the effective optical depth estimated based on these quasar spectra. We will also review our narrowband imaging observations in the same chapter, including field selection, observation running, data reduction, and photometric calibration. Chapter 3 will start with our LAE selection with the Subaru/HSC narrowband imaging data and the detailed completeness estimate of the HSC-SSP $z = 2.18$ LAE catalog. After that, we show our overdensity map based on these LAE catalogs in the nine fields and summarize the information from the maps. Then we perform the correlation analysis between the LAE overdensity $\delta_{\rm LAE}$ and the effective optical depth on $15h^{-1}$cMpc

$\tau_{\rm LoS}$ in Chapter 4 and show our results and the comparisons among different fields in the same chapter. In this chapter, we also perform another statistical analysis, the two-point cross-correlation function between the low-/high-$\tau_{\rm LoS}$ LoSs and LAEs, to compare the results in different environments. Chapter 5 explores to have some more detailed discussions based on the NB387 Lyα absorber-trace fields, including the average $\tau_{\rm LoS}$ profile centered at LAEs, the scale dependence of the $\delta_{\rm LAE}$–$\tau_{\rm LoS}$ correlations, and the underlying physics behind the correlation. Finally, we summarize the entire thesis in Chapter 6. Througout the thesis, we use the cosmological parameters based on Planck Collaboration et al. (2016): $H_0 = 67.7$ km Mpc^{-1} s^{-1}, $\Omega_0 = 0.307$, $T_{\rm cmb,0} = 2.725$ K.

Parts of the results are already published in Liang et al. (2021), and more developments are in preparation for future publications.

Chapter 2

Data and Observations

2.1 SDSS/BOSS Spectral Data

We make use of the archival quasar spectra from the Baryon Oscillation Spectroscopic (BOSS) and the later extended-BOSS (eBOSS) Surveys of SDSS-III and SDSS-IV (Dawson et al. 2013, 2016) to search for the background sources. BOSS and eBOSS are spectroscopic surveys dedicated to studying intergalactic science by taking more than 200,000 quasar spectra covering the survey area of more than 10,000 deg^2. This enables us to find the rare high $\tau_{\rm LoS}$ LoSs, e.g., CoSLAs, in a searching area of over 1 Gpc^3.

We first obtain the spectra from the SDSS 16$^{\rm th}$ Data Release database with the search condition limited to $z \gtrsim 2.25$ and $z \lesssim 2.8$ quasars, because we mainly use the narrowband filters NB387 ($\lambda_0 = 3,862$ Å, FWHM = 56 Å) and NB400 ($\lambda_0 = 4,001$ Å, FWHM = 94 Å) that are installed on the Subaru HSC, which correspond to detect the Lyα at the redshift $z = 2.177 \pm 0.02$ and $z = 2.291 \pm 0.04$. For this purpose, the Lyα absorption imprinting in the quasar spectra from these redshifts avoids being too close to the wing of Lyα emission from the quasar, while we can also apply the principal component analysis (PCA) to predict the intrinsic continuum (Suzuki et al. 2005).

As a proxy of the Lyα absorption strength, the effective optical depth $\tau_{\rm LoS}$ can be calculated along the line-of-sight (LoS) towards the background quasars at the same the Lyα redshift ranges traced by these narrowbands. In practice, we first perform smoothing for the spectra with a flat window of 20 spectral pixels, equivalent to $\approx 15h^{-1}$cMpc at both redshifts. The normalization of the spectra is also made with respective to the 3.2 Å interval at 1280 Å. Then we search for an absorption valley

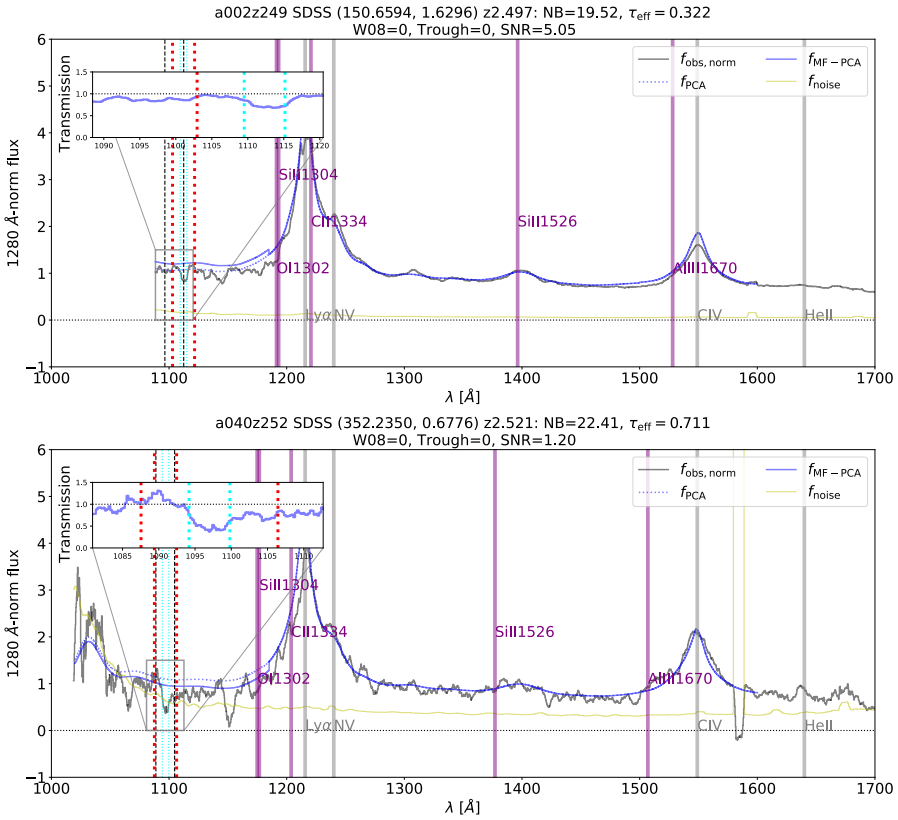

Figure 2.1: Examples of the rest-frame SDSS/BOSS quasar spectra that were categorized as good and kept in our analysis. The black solid, blue dotted, blue solid, and yellow solid curves are the 1280 Å-normalized and 15 h^{-1}cMpc-scale smoothed flux, PCA-predicted continuum, MF-PCA-regulated continuum, and the spectral noise. The gray vertical lines suggest the position of the possible emission lines of the host quasars. The purple vertical lines suggest the position of the possible low-ionization metal lines, including the O I λ1302, C II λ1334, Si II λ1304, Si II λ1526, and Al III λ1670. The black, cyan, and red dashed lines suggest the FWHM of the narrowband filters, the 15 h^{-1}cMpc for calculating $\tau_{\rm LoS}$ and 70 Å window at the observation-frame to indicate the $w_{0.8}$ checking width. The inset figure shows the continuum transmission around the narrowband filters. The lower panel deliberately shows a spectrum with relatively high $\tau_{\rm LoS}$= 0.71 but still well passes all the criteria.

over a range of ±35 Å centered at 3,862 Å and 4,001 Å, respectively. Once the absorption is identified, we calculate the effective optical depth $\tau_{\rm LoS}$ in the similar manner proposed in (Cai et al. 2016), where it expresses as:

$$\tau_{\rm LoS,eff} = -\ln \langle F \rangle_{15\ h^{-1}{\rm cMpc}}\,, \tag{2.1}$$

and the $\langle F \rangle_{15\ h^{-1}}$ cMpc is the continuum-normalized flux estimated on 15 h^{-1}cMpc

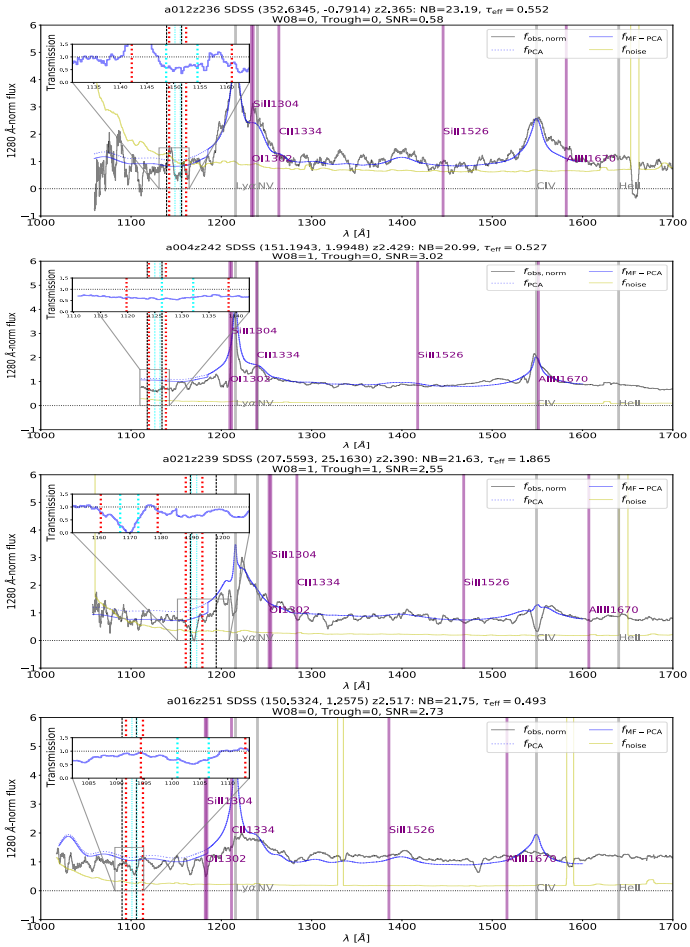

Figure 2.2: Examples of the rest-frame SDSS/BOSS quasar spectra that were categorized as bad and discarded in our analysis. The labels are the same as those in Figure 2.1. The upper to lower panels show the examples discarded because of S/N<1 at the narrowband coverage, $w_{0.8} > 70$ Å, $F_{\rm through} < 0.15$, and visual detection of the low-ionization metal lines.

scale. Therefore, the precise prediction of the intrinsic continuum is critical.

We apply the mean-flux-regulated PCA (MF-PCA) technique (Lee et al. 2016) to estimate the intrinsic continuum. Firstly, we fit the continuum at the blueward end of the quasar Lyα emission from its spectral features at the redward side. The principal components are constructed based on 334 high-resolution QSO spectra taken by the Hubble Space Telescope/Faint Object Spectrograph (FOS) (Suzuki et al. 2005). To avoid the uncertainty of BOSS spectra flux calibration, the Lyα blueward continuum should be further normalized by using the mean optical depth of the $z \sim 2$ Lyα forest (Lee et al. 2012; Becker et al. 2013).

The processes mentioned above are performed in an automated routine, but one should be careful in these $\tau_{\rm LoS}$ estimate because it may not be originated from intergalactic absorption. The high-column density systems (HCDs), e.g., Damped Lyα Asbroption systems (DLAs), sub-DLAs, or the Lyman-limit systems (LLSs) from nearby galaxies, can also cause strong absorption and result in over-estimated $\tau_{\rm LoS}$ for intergalactic absorption. To exclude these contaminants, we perform several tests and leave a flag if it cannot meet the conditions that Cai et al. (2016) have proposed:

- The observed flux has S/N $\gtrsim 1$ at the narrowband wavelength range.
- $\omega_{0.8} < 70$ Å to eliminate DLAs;
- $F_{\rm through} > 0.15$ to reject DLAs again;
- Non-detection of low-ionization metal lines to reject all HCDs;

In detail, if the signal-to-noise ratio S/N < 1, the Lyα estimate will be largely uncertain, and an example is shown in the upper panel of Figure 2.2. The $\omega_{0.8}$ is the width of the absorption at flux/continuum=0.8, since there is rarely a chance for the extreme IGM Lyα absorption cause the feature left in the spectra with $\omega_{0.8} > 70$ Å (e.g. the second panel of Figure 2.2). The $F_{\rm through}$ is the transmitted flux at the ± 5 Å trough centered at the absorption, and only DLAs are likely to cause such a strong trough (e.g., the third panel of Figure 2.2). As to the metal lines, we visually inspect the OI λ1302, CII λ1334, SI II λ1304, SI II λ1526, and Al III λ1670 according to the redshift of the absorption spike (e.g., the lowest panel of the Figure 2.2). A large fraction ($\sim 30\%$) of the LoSs are discarded mainly due to the first two criteria, especially for NB387, due to its bluer coverage and the less sensitivity of the SDSS spectrograph at the wavelength range. Here are two examples of the clean LoSs left in our final sample in Figure 2.1. The upper panel shows an example with a weak absorption and without any break to our constraints for rejecting HCDs. The lower panel one shows an LoS with relatively strong absorption, but there is still no concern from the possible contaminants.

Finally, we also checked the counterpart of our final sample with the published broad absorption line quasar (BAL) catalog and the DLA catalogs (Pâris et al. 2017; Noterdaeme et al. 2012), but there is none of the LoSs matched. The following analyses will include masks on specific sky areas due to the imaging data, and this effect also alters the number of the final LoSs for analysis. The $\tau_{\rm LoS}$ distribution of the inspected LoSs, the LoSs satisfying the above conditions, and the final LoSs are shown in Figure 2.3. The details of the final LoSs including their $\tau_{\rm LoS}$ are listed in the Table 2.1.

Table 2.1:: Summary of LoS information.
[1] The field that the LoS is in;
[2] ID of the LoS;
[3] RA in equinox with an epoch of J2000;
[4] DEC in equinox with an epoch of J2000;
[5] The effective optical depth $\tau_{\rm LoS}$ measured by the MF-PCA on a 15 h^{-1}cMpc scale.

Field [1]	ID [2]	RA (J2000) [3]	DEC (J2000) [4]	$\tau_{\rm LoS}$ [5]
BOSS0210	MAMMOTH-L001	32.11000	0.63630	$0.69^{+0.05}_{-0.05}$
BOSS0210	MAMMOTH-L002	32.20800	0.59630	$0.37^{+0.03}_{-0.03}$
BOSS0210	MAMMOTH-L003	32.44630	0.70730	$0.14^{+0.04}_{-0.04}$
BOSS0210	MAMMOTH-L004	32.47410	0.51860	$0.32^{+0.02}_{-0.02}$
BOSS0210	MAMMOTH-L005	32.50840	0.72190	$0.82^{+0.13}_{-0.12}$
BOSS0210	MAMMOTH-L006	32.51000	0.62790	$1.00^{+0.32}_{-0.24}$
BOSS0210	MAMMOTH-L007	32.52240	1.27380	$0.39^{+0.17}_{-0.14}$
BOSS0210	MAMMOTH-L008	32.54150	1.17070	$0.53^{+0.03}_{-0.03}$
BOSS0210	MAMMOTH-L009	32.61330	1.21040	$0.30^{+0.03}_{-0.02}$
BOSS0210	MAMMOTH-L010	32.62370	0.60610	$0.50^{+0.08}_{-0.08}$
BOSS0210	MAMMOTH-L011	32.75520	0.64090	$0.42^{+0.15}_{-0.13}$
BOSS0210	MAMMOTH-L012	32.84500	0.78210	$0.47^{+0.11}_{-0.10}$
BOSS0210	MAMMOTH-L013	32.85950	0.58580	$0.75^{+0.11}_{-0.10}$
BOSS0210	MAMMOTH-L014	32.86660	0.49800	$0.30^{+0.04}_{-0.04}$
BOSS0210	MAMMOTH-L015	32.90690	0.41600	$0.22^{+0.06}_{-0.06}$
BOSS0210	MAMMOTH-L016	32.92560	0.78510	$0.49^{+0.05}_{-0.05}$
BOSS0210	MAMMOTH-L017	32.94750	0.75050	$0.27^{+0.15}_{-0.13}$
BOSS0210	MAMMOTH-L018	32.22540	0.32020	$0.31^{+0.15}_{-0.13}$
BOSS0210	MAMMOTH-L019	32.15220	0.82480	$0.71^{+0.15}_{-0.13}$
BOSS0210	MAMMOTH-L020	33.17300	1.05830	$0.42^{+0.11}_{-0.10}$
BOSS0210	MAMMOTH-L021	33.17940	1.04050	$0.25^{+0.12}_{-0.11}$
BOSS0210	MAMMOTH-L022	33.18540	0.81380	$0.28^{+0.11}_{-0.10}$
BOSS0222	MAMMOTH-L023	35.79990	-2.27190	$0.59^{+0.09}_{-0.08}$
BOSS0222	MAMMOTH-L024	35.71920	-2.55390	$0.72^{+0.09}_{-0.08}$
BOSS0222	MAMMOTH-L025	35.46760	-2.52540	$0.78^{+0.11}_{-0.10}$
BOSS0222	MAMMOTH-L026	35.47340	-2.11100	$0.44^{+0.11}_{-0.10}$
BOSS0222	MAMMOTH-L027	34.98770	-2.09640	$0.21^{+0.03}_{-0.03}$
BOSS0222	MAMMOTH-L028	35.07430	-2.08700	$0.12^{+0.12}_{-0.11}$
BOSS0222	MAMMOTH-L029	35.20280	-2.94810	$0.23^{+0.26}_{-0.21}$
BOSS0222	MAMMOTH-L030	35.48780	-1.88850	$0.21^{+0.09}_{-0.08}$

Table 2.1:: Summary of LoS information.
[1] The field that the LoS is in;
[2] ID of the LoS;
[3] RA in equinox with an epoch of J2000;
[4] DEC in equinox with an epoch of J2000;
[5] The effective optical depth $\tau_{\rm LoS}$ measured by the MF-PCA on a 15 h^{-1}cMpc scale.

Field [1]	ID [2]	RA (J2000) [3]	DEC (J2000) [4]	$\tau_{\rm LoS}$ [5]
BOSS0222	MAMMOTH-L031	35.73610	-2.23460	$0.82^{+0.09}_{-0.08}$
BOSS0222	MAMMOTH-L032	35.57320	-2.38750	$0.74^{+0.13}_{-0.12}$
BOSS0222	MAMMOTH-L033	35.44700	-2.47470	$0.60^{+0.06}_{-0.06}$
BOSS0924	MAMMOTH-L034	141.08069	14.95697	$0.57^{+0.04}_{-0.03}$
BOSS0924	MAMMOTH-L035	140.84493	14.81363	$0.30^{+0.06}_{-0.05}$
BOSS0924	MAMMOTH-L036	140.93623	14.99946	$0.66^{+0.04}_{-0.04}$
BOSS0924	MAMMOTH-L037	140.95420	15.69943	$0.47^{+0.19}_{-0.16}$
BOSS0924	MAMMOTH-L038	140.71501	15.55039	$0.64^{+0.12}_{-0.10}$
BOSS0924	MAMMOTH-L039	141.10702	15.07088	$0.65^{+0.12}_{-0.10}$
BOSS0924	MAMMOTH-L040	141.45351	14.69011	$0.77^{+0.12}_{-0.10}$
BOSS0924	MAMMOTH-L041	141.27276	14.77130	$0.35^{+0.14}_{-0.12}$
BOSS0924	MAMMOTH-L042	140.73459	15.34292	$0.46^{+0.10}_{-0.09}$
BOSS0924	MAMMOTH-L043	140.90029	15.47429	$0.27^{+0.08}_{-0.08}$
BOSS0924	MAMMOTH-L044	141.09274	15.26301	$0.79^{+0.07}_{-0.06}$
BOSS0924	MAMMOTH-L045	141.61389	14.97427	$0.63^{+0.10}_{-0.09}$
BOSS0924	MAMMOTH-L046	140.54426	14.78670	$1.04^{+0.34}_{-0.25}$
BOSS0924	MAMMOTH-L047	140.68599	14.91317	$0.69^{+0.10}_{-0.10}$
BOSS1419	MAMMOTH-L048	214.99153	4.44898	$0.30^{+0.06}_{-0.05}$
BOSS1419	MAMMOTH-L049	214.43355	4.83006	$0.39^{+0.05}_{-0.04}$
BOSS1419	MAMMOTH-L050	214.38686	4.81351	$0.47^{+0.44}_{-0.30}$
BOSS1419	MAMMOTH-L051	214.55276	5.23412	$0.25^{+0.08}_{-0.07}$
BOSS1419	MAMMOTH-L052	214.58907	4.67769	$0.65^{+0.12}_{-0.11}$
BOSS1419	MAMMOTH-L053	214.60573	5.01409	$0.46^{+0.21}_{-0.17}$
BOSS1419	MAMMOTH-L054	214.56614	5.04207	$0.48^{+0.10}_{-0.09}$
BOSS1419	MAMMOTH-L055	214.58009	4.69308	$0.28^{+0.03}_{-0.03}$
BOSS1419	MAMMOTH-L056	214.84842	5.42960	$0.35^{+0.12}_{-0.11}$
BOSS1419	MAMMOTH-L057	214.75582	4.96796	$0.78^{+0.46}_{-0.32}$
BOSS1419	MAMMOTH-L058	214.76557	4.58137	$0.58^{+0.09}_{-0.09}$
BOSS1419	MAMMOTH-L059	214.92877	5.45505	$0.51^{+0.05}_{-0.05}$
BOSS1419	MAMMOTH-L060	214.96227	5.50922	$0.73^{+0.47}_{-0.32}$

Table 2.1:: Summary of LoS information.
[1] The field that the LoS is in;
[2] ID of the LoS;
[3] RA in equinox with an epoch of J2000;
[4] DEC in equinox with an epoch of J2000;
[5] The effective optical depth $\tau_{\rm LoS}$ measured by the MF-PCA on a 15 h^{-1}cMpc scale.

Field [1]	ID [2]	RA (J2000) [3]	DEC (J2000) [4]	$\tau_{\rm LoS}$ [5]
BOSS1419	MAMMOTH-L061	214.90256	4.90854	$0.52^{+0.05}_{-0.05}$
BOSS1419	MAMMOTH-L062	215.07084	4.58767	$0.36^{+0.15}_{-0.13}$
BOSS1419	MAMMOTH-L063	215.19108	5.24957	$0.53^{+0.27}_{-0.21}$
BOSS1419	MAMMOTH-L064	215.22404	5.00992	$0.61^{+0.16}_{-0.14}$
BOSS0240	QSO-L001	39.77097	-4.93288	$0.39^{+0.07}_{-0.06}$
BOSS0240	QSO-L002	39.30519	-5.31255	$0.37^{+0.04}_{-0.04}$
BOSS0240	QSO-L003	39.61937	-5.45144	$0.19^{+0.02}_{-0.01}$
BOSS0240	QSO-L004	39.79528	-5.83616	$0.52^{+0.10}_{-0.09}$
BOSS0240	QSO-L005	39.88710	-6.02066	$0.40^{+0.09}_{-0.08}$
BOSS0240	QSO-L006	40.05549	-5.67282	$0.34^{+0.09}_{-0.08}$
BOSS0240	QSO-L007	40.21837	-5.90067	$0.45^{+0.05}_{-0.05}$
BOSS0240	QSO-L008	40.12843	-5.62813	$0.91^{+0.32}_{-0.25}$
BOSS0240	QSO-L009	40.36814	-5.59818	$0.49^{+0.07}_{-0.07}$
BOSS0755	QSO-L010	118.93147	30.70434	$0.50^{+0.16}_{-0.14}$
BOSS0755	QSO-L011	119.42920	30.77425	$0.21^{+0.07}_{-0.06}$
BOSS0755	QSO-L012	118.73824	31.23223	$0.37^{+0.15}_{-0.12}$
BOSS0755	QSO-L013	118.81332	31.07877	$0.26^{+0.02}_{-0.02}$
BOSS0755	QSO-L014	119.53365	30.85766	$0.17^{+0.05}_{-0.04}$
BOSS0755	QSO-L015	118.75277	31.13443	$0.31^{+0.09}_{-0.08}$
BOSS0755	QSO-L016	118.75877	31.04206	$0.25^{+0.09}_{-0.08}$
BOSS0755	QSO-L017	119.34363	30.95714	$0.09^{+0.21}_{-0.15}$
BOSS0755	QSO-L018	119.67846	30.97011	$0.51^{+0.10}_{-0.09}$
BOSS0755	QSO-L019	118.84348	31.57476	$0.33^{+0.07}_{-0.06}$
BOSS0755	QSO-L020	118.54199	31.15767	$0.61^{+0.19}_{-0.16}$
BOSS0755	QSO-L021	118.18931	31.06606	$0.23^{+0.02}_{-0.02}$
BOSS0755	QSO-L022	118.81979	31.89191	$0.13^{+0.06}_{-0.05}$
DEEP2-3	Field-L001	350.63659	-0.61113	$0.21^{+0.06}_{-0.06}$
DEEP2-3	Field-L002	350.32496	-1.19861	$0.16^{+0.06}_{-0.05}$
DEEP2-3	Field-L003	350.28315	-0.46573	$0.32^{+0.07}_{-0.07}$

Table 2.1:: Summary of LoS information.
[1] The field that the LoS is in;
[2] ID of the LoS;
[3] RA in equinox with an epoch of J2000;
[4] DEC in equinox with an epoch of J2000;
[5] The effective optical depth $\tau_{\rm LoS}$ measured by the MF-PCA on a 15 h^{-1}cMpc scale.

Field [1]	ID [2]	RA (J2000) [3]	DEC (J2000) [4]	$\tau_{\rm LoS}$ [5]
DEEP2-3	Field-L004	351.15582	0.20713	$0.22^{+0.03}_{-0.03}$
DEEP2-3	Field-L005	352.34672	-0.70453	$0.41^{+0.13}_{-0.11}$
DEEP2-3	Field-L006	352.09897	-0.73078	$0.30^{+0.08}_{-0.07}$
DEEP2-3	Field-L007	352.10998	-0.88481	$0.23^{+0.11}_{-0.10}$
DEEP2-3	Field-L008	351.67114	-0.42002	$0.59^{+0.23}_{-0.18}$
DEEP2-3	Field-L009	351.66980	-0.37682	$0.28^{+0.10}_{-0.09}$
DEEP2-3	Field-L010	351.54264	-1.07441	$0.54^{+0.27}_{-0.20}$
DEEP2-3	Field-L011	351.40576	-1.08866	$0.44^{+0.14}_{-0.12}$
DEEP2-3	Field-L012	351.45870	-0.84636	$0.57^{+0.15}_{-0.13}$
DEEP2-3	Field-L013	351.36826	-0.99831	$0.25^{+0.10}_{-0.09}$
DEEP2-3	Field-L014	351.37701	0.89939	$0.28^{+0.04}_{-0.04}$
DEEP2-3	Field-L015	351.58319	0.67443	$0.22^{+0.06}_{-0.06}$
DEEP2-3	Field-L016	352.03992	0.73245	$0.21^{+0.03}_{-0.03}$
DEEP2-3	Field-L017	352.21420	0.72986	$0.36^{+0.14}_{-0.12}$
DEEP2-3	Field-L018	352.08445	0.76485	$0.12^{+0.09}_{-0.08}$
DEEP2-3	Field-L019	352.23502	0.67759	$0.71^{+0.33}_{-0.25}$
DEEP2-3	Field-L020	352.14619	0.41636	$0.06^{+0.11}_{-0.09}$
DEEP2-3	Field-L021	352.26605	0.27708	$0.19^{+0.14}_{-0.11}$
DEEP2-3	Field-L022	352.62007	0.46593	$0.24^{+0.06}_{-0.05}$
DEEP2-3	Field-L023	352.80954	0.85776	$0.26^{+0.08}_{-0.07}$
DEEP2-3	Field-L024	353.28629	-0.24294	$0.26^{+0.03}_{-0.03}$
DEEP2-3	Field-L025	352.78834	-1.02229	$0.25^{+0.12}_{-0.10}$
DEEP2-3	Field-L026	352.99424	-0.90290	$0.51^{+0.04}_{-0.04}$
DEEP2-3	Field-L027	352.60275	-0.52732	$0.05^{+0.04}_{-0.04}$
DEEP2-3	Field-L028	352.95674	0.82740	$0.09^{+0.07}_{-0.06}$
DEEP2-3	Field-L029	353.10299	0.86461	$0.44^{+0.18}_{-0.15}$
DEEP2-3	Field-L030	353.35574	0.62739	$0.32^{+0.14}_{-0.12}$
DEEP2-3	Field-L031	353.69387	0.45583	$0.21^{+0.05}_{-0.04}$
DEEP2-3	Field-L032	353.75530	0.66033	$0.16^{+0.17}_{-0.13}$
DEEP2-3	Field-L033	352.03310	-1.40950	$0.28^{+0.16}_{-0.13}$

Table 2.1:: Summary of LoS information.
[1] The field that the LoS is in;
[2] ID of the LoS;
[3] RA in equinox with an epoch of J2000;
[4] DEC in equinox with an epoch of J2000;
[5] The effective optical depth $\tau_{\rm LoS}$ measured by the MF-PCA on a 15 h^{-1}cMpc scale.

Field [1]	ID [2]	RA (J2000) [3]	DEC (J2000) [4]	$\tau_{\rm LoS}$ [5]
DEEP2-3	Field-L034	352.22480	-1.56021	$0.26^{+0.22}_{-0.17}$
DEEP2-3	Field-L035	350.59431	-1.14904	$0.37^{+0.17}_{-0.13}$
DEEP2-3	Field-L036	351.02139	-1.52297	$0.55^{+0.16}_{-0.14}$
DEEP2-3	Field-L037	350.91668	-0.19960	$0.48^{+0.28}_{-0.21}$
DEEP2-3	Field-L038	351.54814	-0.14523	$0.34^{+0.19}_{-0.15}$
DEEP2-3	Field-L039	351.77254	-0.21087	$0.92^{+0.29}_{-0.24}$
DEEP2-3	Field-L040	353.18435	-0.91931	$0.19^{+0.06}_{-0.05}$
DEEP2-3	Field-L041	351.80762	-0.12759	$0.33^{+0.17}_{-0.14}$
DEEP2-3	Field-L042	352.37300	0.26353	$0.21^{+0.22}_{-0.16}$
DEEP2-3	Field-L043	352.41786	-0.29769	$0.25^{+0.17}_{-0.13}$
DEEP2-3	Field-L044	352.47680	-0.47115	$0.21^{+0.08}_{-0.07}$
DEEP2-3	Field-L045	352.44391	-0.71385	$0.13^{+0.18}_{-0.14}$
DEEP2-3	Field-L046	352.71501	-0.76115	$0.31^{+0.15}_{-0.13}$
DEEP2-3	Field-L047	353.24743	-0.06225	$0.17^{+0.07}_{-0.06}$
DEEP2-3	Field-L048	353.44551	-0.23535	$0.64^{+0.28}_{-0.22}$
DEEP2-3	Field-L049	353.39158	-0.36158	$0.04^{+0.06}_{-0.05}$
DEEP2-3	Field-L050	353.73714	-0.06347	$0.76^{+0.52}_{-0.35}$
DEEP2-3	Field-L051	353.73262	0.02517	$0.31^{+0.14}_{-0.11}$
DEEP2-3	Field-L052	351.07554	0.73349	$0.23^{+0.06}_{-0.05}$
DEEP2-3	Field-L053	353.66055	0.50845	$0.18^{+0.14}_{-0.11}$
DEEP2-3	Field-L054	353.72449	0.72929	$0.71^{+0.26}_{-0.21}$
DEEP2-3	Field-L055	353.65881	0.74354	$0.22^{+0.11}_{-0.09}$
DEEP2-3	Field-L056	353.54375	0.53255	$0.20^{+0.11}_{-0.09}$
DEEP2-3	Field-L057	352.88378	0.12813	$0.20^{+0.14}_{-0.11}$
DEEP2-3	Field-L058	352.14063	0.59999	$0.29^{+0.13}_{-0.11}$
DEEP2-3	Field-L059	352.04681	0.79413	$0.11^{+0.16}_{-0.13}$
DEEP2-3	Field-L060	350.98094	0.59727	$0.50^{+0.02}_{-0.02}$
E-COSMOS	Field-L061	150.65939	1.62959	$0.32^{+0.04}_{-0.04}$
E-COSMOS	Field-L062	150.69650	0.86987	$0.29^{+0.02}_{-0.02}$
E-COSMOS	Field-L063	149.05408	1.32013	$0.29^{+0.07}_{-0.07}$

Table 2.1:: Summary of LoS information.
[1] The field that the LoS is in;
[2] ID of the LoS;
[3] RA in equinox with an epoch of J2000;
[4] DEC in equinox with an epoch of J2000;
[5] The effective optical depth $\tau_{\rm LoS}$ measured by the MF-PCA on a 15 h^{-1}cMpc scale.

Field [1]	ID [2]	RA (J2000) [3]	DEC (J2000) [4]	$\tau_{\rm LoS}$ [5]
E-COSMOS	Field-L064	149.09109	1.08982	$0.32^{+0.09}_{-0.08}$
E-COSMOS	Field-L065	149.21490	0.75237	$0.36^{+0.02}_{-0.02}$
E-COSMOS	Field-L066	149.24015	1.21947	$0.31^{+0.04}_{-0.04}$
E-COSMOS	Field-L067	151.30130	1.26487	$0.17^{+0.05}_{-0.05}$
E-COSMOS	Field-L068	150.05182	1.68273	$0.24^{+0.07}_{-0.07}$
E-COSMOS	Field-L069	149.74450	2.02750	$0.51^{+0.02}_{-0.02}$
E-COSMOS	Field-L070	148.99584	1.60231	$0.27^{+0.06}_{-0.06}$
E-COSMOS	Field-L071	148.97821	1.76501	$0.44^{+0.33}_{-0.23}$
E-COSMOS	Field-L072	148.85477	3.24193	$0.46^{+0.14}_{-0.12}$
E-COSMOS	Field-L073	148.93647	3.16080	$0.67^{+0.30}_{-0.23}$
E-COSMOS	Field-L074	148.99779	3.50208	$0.46^{+0.04}_{-0.03}$
E-COSMOS	Field-L075	149.49884	2.78266	$0.38^{+0.07}_{-0.07}$
E-COSMOS	Field-L076	149.81144	3.25589	$0.27^{+0.21}_{-0.16}$
E-COSMOS	Field-L077	149.77175	2.96417	$0.47^{+0.28}_{-0.21}$
E-COSMOS	Field-L078	149.82509	3.29686	$0.25^{+0.14}_{-0.11}$
E-COSMOS	Field-L079	150.24369	3.04813	$0.30^{+0.04}_{-0.04}$
E-COSMOS	Field-L080	150.28073	3.15921	$0.40^{+0.09}_{-0.08}$
E-COSMOS	Field-L081	150.89389	3.46884	$0.44^{+0.09}_{-0.08}$
E-COSMOS	Field-L082	150.76650	3.53129	$0.44^{+0.20}_{-0.16}$
E-COSMOS	Field-L083	150.83346	2.98650	$0.35^{+0.28}_{-0.20}$
E-COSMOS	Field-L084	150.90873	3.15851	$0.34^{+0.07}_{-0.07}$
E-COSMOS	Field-L085	151.15042	3.49425	$0.23^{+0.12}_{-0.10}$
E-COSMOS	Field-L086	151.33928	3.46245	$0.15^{+0.28}_{-0.19}$
E-COSMOS	Field-L087	151.36440	3.44887	$0.40^{+0.05}_{-0.05}$
E-COSMOS	Field-L088	148.84850	1.81234	$0.27^{+0.09}_{-0.08}$
E-COSMOS	Field-L089	148.90947	1.96490	$0.49^{+0.24}_{-0.19}$
E-COSMOS	Field-L090	149.06607	2.41097	$0.50^{+0.32}_{-0.23}$
E-COSMOS	Field-L091	148.74890	1.90552	$0.31^{+0.18}_{-0.14}$
E-COSMOS	Field-L092	148.61897	1.53245	$0.13^{+0.20}_{-0.15}$
E-COSMOS	Field-L093	148.84324	2.74978	$0.53^{+0.17}_{-0.14}$

Table 2.1:: Summary of LoS information.
[1] The field that the LoS is in;
[2] ID of the LoS;
[3] RA in equinox with an epoch of J2000;
[4] DEC in equinox with an epoch of J2000;
[5] The effective optical depth $\tau_{\rm LoS}$ measured by the MF-PCA on a 15 h^{-1}cMpc scale.

Field	ID	RA (J2000)	DEC (J2000)	$\tau_{\rm LoS}$
[1]	[2]	[3]	[4]	[5]
E-COSMOS	Field-L094	151.15466	3.31580	$0.20^{+0.25}_{-0.18}$
E-COSMOS	Field-L095	150.49000	3.34548	$0.26^{+0.20}_{-0.16}$
XMM-SXDS	Field-L096	37.53605	-4.52658	$0.34^{+0.02}_{-0.02}$
XMM-SXDS	Field-L097	36.67074	-4.50114	$0.21^{+0.08}_{-0.07}$
XMM-SXDS	Field-L098	36.11203	-5.65574	$0.29^{+0.10}_{-0.09}$
XMM-SXDS	Field-L099	35.69926	-4.45778	$0.25^{+0.16}_{-0.13}$
XMM-SXDS	Field-L100	34.94567	-4.29388	$0.56^{+0.16}_{-0.14}$
XMM-SXDS	Field-L101	35.43735	-3.96262	$0.34^{+0.16}_{-0.13}$
XMM-SXDS	Field-L102	36.15873	-4.09184	$0.26^{+0.19}_{-0.15}$
XMM-SXDS	Field-L103	36.78030	-4.00468	$0.24^{+0.08}_{-0.07}$
XMM-SXDS	Field-L104	36.93597	-3.70152	$0.31^{+0.07}_{-0.06}$
XMM-SXDS	Field-L105	35.83829	-5.10745	$0.15^{+0.03}_{-0.03}$
XMM-SXDS	Field-L106	35.94669	-5.09803	$0.17^{+0.06}_{-0.06}$
XMM-SXDS	Field-L107	36.04567	-5.11501	$0.33^{+0.09}_{-0.08}$
XMM-SXDS	Field-L108	36.95357	-4.47252	$0.35^{+0.08}_{-0.08}$
XMM-SXDS	Field-L109	36.89897	-4.30776	$0.34^{+0.06}_{-0.06}$
XMM-SXDS	Field-L110	36.49009	-4.83485	$0.39^{+0.05}_{-0.04}$
XMM-SXDS	Field-L111	36.14308	-4.53347	$0.15^{+0.04}_{-0.03}$
XMM-SXDS	Field-L112	36.05571	-5.59600	$0.26^{+0.06}_{-0.05}$
XMM-SXDS	Field-L113	35.97837	-4.80419	$0.58^{+0.02}_{-0.02}$
XMM-SXDS	Field-L114	34.89977	-4.42949	$0.39^{+0.10}_{-0.09}$
XMM-SXDS	Field-L115	36.21664	-4.20306	$0.20^{+0.09}_{-0.08}$
XMM-SXDS	Field-L116	36.05614	-4.86957	$0.81^{+0.55}_{-0.37}$
XMM-SXDS	Field-L117	36.11531	-5.78167	$0.07^{+0.08}_{-0.07}$
XMM-SXDS	Field-L118	36.12201	-5.65272	$0.19^{+0.24}_{-0.17}$
XMM-SXDS	Field-L119	36.23531	-4.92142	$0.21^{+0.10}_{-0.09}$
XMM-SXDS	Field-L120	35.75123	-5.76315	$0.42^{+0.12}_{-0.10}$
XMM-SXDS	Field-L121	35.64018	-5.81742	$0.37^{+0.31}_{-0.22}$
XMM-SXDS	Field-L122	35.50003	-5.80216	$0.49^{+0.13}_{-0.11}$
XMM-SXDS	Field-L123	35.29179	-5.84498	$0.45^{+0.16}_{-0.13}$

Table 2.1:: Summary of LoS information.
[1] The field that the LoS is in;
[2] ID of the LoS;
[3] RA in equinox with an epoch of J2000;
[4] DEC in equinox with an epoch of J2000;
[5] The effective optical depth $\tau_{\rm LoS}$ measured by the MF-PCA on a 15 h^{-1}cMpc scale.

Field [1]	ID [2]	RA (J2000) [3]	DEC (J2000) [4]	$\tau_{\rm LoS}$ [5]
XMM-SXDS	Field-L124	35.18930	-5.78328	$0.52^{+0.21}_{-0.17}$
XMM-SXDS	Field-L125	37.24015	-4.12591	$0.40^{+0.07}_{-0.06}$
XMM-SXDS	Field-L126	37.24852	-4.00885	$0.39^{+0.07}_{-0.06}$
XMM-SXDS	Field-L127	36.51721	-4.83903	$0.34^{+0.11}_{-0.09}$
XMM-SXDS	Field-L128	36.04090	-4.49687	$0.38^{+0.10}_{-0.09}$
XMM-SXDS	Field-L129	36.01729	-3.67431	$0.41^{+0.18}_{-0.15}$
XMM-SXDS	Field-L130	36.28740	-5.38547	$0.38^{+0.30}_{-0.21}$
XMM-SXDS	Field-L131	35.97944	-5.26097	$0.10^{+0.23}_{-0.16}$
XMM-SXDS	Field-L132	35.59754	-5.64821	$0.26^{+0.15}_{-0.12}$
XMM-SXDS	Field-L133	36.50680	-3.88422	$0.17^{+0.13}_{-0.11}$
XMM-SXDS	Field-L134	36.49530	-4.30025	$0.21^{+0.17}_{-0.13}$
XMM-SXDS	Field-L135	35.47590	-4.24962	$0.27^{+0.02}_{-0.02}$
XMM-SXDS	Field-L136	35.05974	-4.48811	$0.47^{+0.02}_{-0.02}$
XMM-SXDS	Field-L137	37.10288	-4.26277	$0.29^{+0.02}_{-0.02}$

2.2 Narrowband Imaging Observations with Subaru/HSC

2.2.1 Field Selection

We aim to examine galaxy-IGM HI correlations in a wide range of environments based on less-massive and young galaxy populations. The principle for our field selection is to enclose a sufficient number of LoSs, especially those with strong Lyα absorptions. At the same time, we also target potential overdensities traced by CoSLAs or the grouping quasars.

We mainly choose the LoSs with $\tau_{\rm LoS} \gtrsim 3\,\langle\tau\rangle_{cos}$ as the preliminary absorber

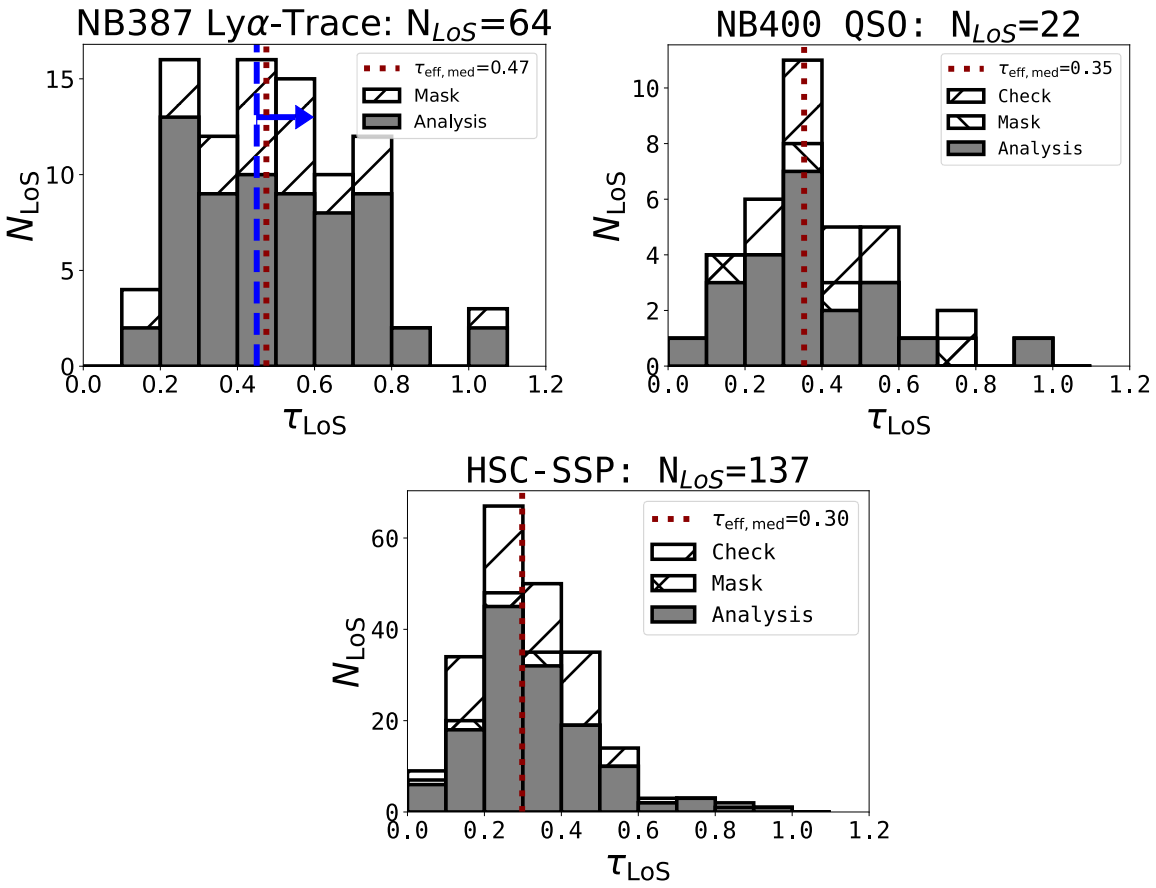

Figure 2.3: The distributions of the effective optical depth meansured on 15 h^{-1}cMpc scales $\tau_{\rm LoS}$ in each samples. The gray histogram shows the LoSs used in the analysis throughout this thesis. The hatch histograms denote the LoSs that were checked but are discarded because they are flagged as bad or they lie in the masked sky. The brown pointed vertical lines indicate the median of the $\tau_{\rm LoS}$ in each sample, and the blue dashed vertical line in the upper left panel denotes the criterion to be viewed as $\tau_{\rm LoS} \gtrsim 3\,\langle\tau\rangle$ when we make the field selection. Following the order of upper left, upper right, and lower panels, they are for the NB387 Lyα absorber-trace fields (BOSS0210, BOSS0222, BOSS0924, BOSS1419), the NB400 grouping quasar fields (BOSS0240, BOSS0755), and the HSC-SSP Deep/Ultra-deep fields (DEEP2-3, E-COSMOS, XMM-SXDS), respectively. The $\rm N_{LoS}$ in the title of each figure shows the number of all the accepted LoSs for analysis.

candidates, where $\langle\tau\rangle_{cos}$ is the cosmic mean optical depth that is given by Becker et al. (2013) as 0.15 at $z \sim 2$. Note that these LoSs are already carefully inspected to exclude the contaminants, including HCDs and BALs, when estimating the $\tau_{\rm LoS}$. Based on clean LoSs, the target fields of HSC-FoV should: (1) contain high $\tau_{\rm LoS}$ LoSs

to expand the dynamic range; (2) enclose as many LoSs as possible to increase the sample size for inferring the galaxy-IGM HI relation; (3) contain a concentration of high $\tau_{\rm LoS}$ LoSs to find protoclusters, i.e., $\gtrsim$ 4 LoSs within a ($\sim$ 20 h^{-1}cMpc)^3 box, which is the typical scale of a Coma-type protocluster at $z \sim 2$ (Chiang et al. 2013); (4) in exceptional cases, contain associated quasars at $2.15 \leq z \leq 2.20$ or $2.25 \leq z \leq 2.33$, i.e., proximity quasars at the 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 J0210+0052 (or BOSS0210), BOSS J0222-0224 (or BOSS0222), BOSS J0924+1503 (or BOSS0924), and BOSS J1419+0500 (or BOSS1419) – in our NB387 observations are selected, all of which satisfy (1) and most of which satisfy (2). BOSS0222 and BOSS0924 are selected based mainly on (3), the typical regions hinting at coherent IGM HI on a large scale. BOSS J1419 is initially selected for (3), but one of the two concentrating LoSs is found to be a possible BAL quasar [1] 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. BOSS0210 is selected with consideration of (4), given that a group of 11 proximity quasars is associated within a region of 40 $\times$ 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). Besides, we also have two other fields BOSS J0240-0521 (or BOSS0240) and BOSS J0755+3109 (BOSS0755), from the NB400 observation proposed by Tsinghua University's team (PI: Z. Cai) that also encloses grouping quasars ($\sim$ 5 in 40 $\times$ 40 cMpc^2 at $2.25 < z < 2.33$) to enlarge the sample size of this special class.

The $\tau_{\rm LoS}$ distribution of these clean LoSs is shown in Figure 2.3, in which a dashed blue vertical line in the left panel indicates the criterion for the clean LoSs with $\tau_{\rm LoS} \gtrsim 3 \langle\tau\rangle_{cos}$ for the Lyα absorption–trace field and the dotted red line shows the median of the $\tau_{\rm LoS}$in each narrowband sample. Note that although the NB387 and NB400 map LAEs at different redshifts $z = 2.18 \pm 0.02$ and $z = 2.29 \pm 0.04$, the cosmic mean optical depth $\langle\tau\rangle_{cos} = 0.17$ and 0.20 at the respective redshifts vary by about 0.03 (Faucher-Giguère et al. 2008), which is much smaller than the typical error of the $\tau_{\rm LoS}$ estimate ($\tau_{\rm LoS} > 0.1$). We thus do not consider the redshift evolution effects on our following analysis.

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

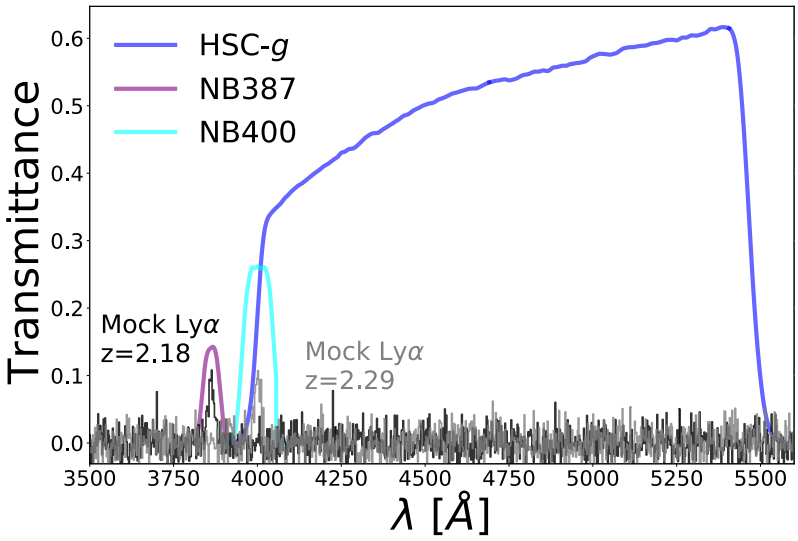

Figure 2.4: The total transmittance of each filter. The blue, purple, and cyan curve represent the HSC-g, NB387, and NB400 bands. The black and gray curves indicate mock LAE spectra at $z = 2.18$ and $z = 2.29$, respectively, the Lyα emission of which is precisely located at the sensitive wavelength range of NB387 and NB400.

2.2.2 Imaging Observations

Observations to identify LAEs were carried out using the HSC installed at the prime focus of the 8.2-m Subaru telescope at the summit of Mauna Kea, Hawaii. HSC is a high-performance camera with a wide FoV of 1.5 deg in diameter. As a gigantic mosaic CCD camera, HSC consists of 104 Hamamatsu Photonics KK CCDs (2048 $\times$ 4096 pixels) for science, 4 for auto-guiders, and 8 for focusing. The pixel scale of the CCD reaches 0.168″.

Here, we perform deep NB imaging using NB387 and NB400, enabling the detection of Lyα emission at the corresponding redshift of $z = 2.177 \pm 0.023$ and $z = 2.291 \pm 0.041$. The g-band is also used to evaluate the detected objects' continuum level. The 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 2.4.

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/NB400 and 20-40 minutes for the g-band in each field. An S17B observation was carried out in queue mode in January 2018, and January 2019, and exposures were split into 900 s and 600 s for the NB387 and g-band observations, respectively, except for the first two exposures

of 1,200 s for BOSS0210. In the S19A observations carried out on-site on March 8$^{\text{th}}$, 2019, the exposures were split into 900 s and 300 s for the NB387 and g-band observations, respectively, to avoid saturation in the broadband, which can affect the selection completeness of LAEs around bright stars. However, saturation can still occur around the brightest stars even with a shorter exposure time; therefore, we mask such regions in the photometric processing, as described in Section 2.2.5. From S17B, we obtain the NB387 and g-band data for fields BOSS0210 and BOSS0222 and the g-band data for BOSS0924. In the S19A run, NB387 data for both BOSS0924 and BOSS1419 are obtained, and the g-band imaging is taken for BOSS1419 only. In summary, the NB387 and g-band imaging data for all four fields are from two major runs.

As to the NB400 campaign, BOSS0240 and BOSS0755 were observed in the S20B semester, with the exposure also split into 900 s for the narrowband and 300 s for g-band. Based on the experience of past NB387 observations, we choose to keep rotating the telescope when dithering because there are some bad HSC CCD channels, and without rotation, we are likely to lose a certain part of the sky.

With the exception of the NB imaging of BOSS0210 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 occasionally poor seeing or low transparency. In the case of BOSS0222, 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. The condition is generally better for NB400 than for NB387 since the CCD response performs better at 4,000 Å than 3,860 Å, and there is a lower chance of encountering accidents like losing tracking stars. Detailed information of the image quality for each field is summarized in the Table 2.2.

Standard stars are not used, considering the large FoV and 104 CCDs. Instead, we use Pan-STARS DR1 (PS1) photometric data (Chambers et al. 2016) for calibration as described in the Sec. 2.2.3.

Field [1]	RA (J2000) [2]	DEC (J2000) [3]	Obs Period [4]	$\mathrm{FWHM_{PSF,NB}}$ [5]	$\mathrm{FWMH_{PSF,}}_{g}$ [6]	$m_{\mathrm{NB},5\sigma}$ [7]	$m_{g,5\sigma}$ [8]	*hscPipe* [9]
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
BOSS J0240-0521	02:40:05.11	-05:21:06.7	Apr. 2021	1.01″	0.82″	25.61	26.80	v6.6
BOSS J0755+3109	07:55:35.89	+31:09:56.9	Apr. 2021	0.82″	1.12″	25.83	26.50	v6.6

Table 2.2:: Summary of field information.
[1] The full name of fields;
[2] RA in equinox with an epoch of J2000;
[3] DEC in equinox with an epoch of J2000;
[4] The period of the observations;
[5] The FWHMs of star PSFs measured using the final stacked images of NB387 or NB400;
[6] The FWHMs of star PSFs measured using the final stacked images of g-band;
[7] The 5σ limiting magnitudes measured in an aperture with the radius of the 1.7″(BOSS0222, BOSS0924, BOSS1419), 2.0″(BOSS0240, BOSS0755) or 2.5″(BOSS0210) using the stacked and the PSF-matched NB387/NB400;
[8] The 5σ limiting magnitudes measured g-band with the same treatment as [7];
[9] The *hscPipe* version used for the data reduction.

2.2.3 Data Reduction

The NB387, NB400, and g-band imaging data are reduced using HSC pipeline, *hscPipe* (Bosch et al. 2018; Aihara et al. 2019). BOSS0210 and BOSS0222 are reduced using *hscPipe* 5.4, and BOSS0924, BOSS1419 are reduced using *hscPipe* 6.6. When we finished the NB400 observations, the latest version of *hscPipe* was not yet available for this new filter. Therefore, we modified the *hscPipe* 6.6 with the software team helpdesk assistance. Given the relatively poor quality of the BOSS0210 NB387 data, we also reduce both the g-band and the NB387 data for BOSS0210 using *hscPipe* 6.6, and combine the catalog with that from *hscPipe* 5.4. The overlaid detections with separations more minor than $2''$ are kept with only the latter version. Bosch et al. (2018) and Aihara et al. (2019) describe the data reduction process and the code updates in detail, but we present a 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 multi-band relation, which is derived from the template magnitudes predicted by spectroscopic Pickles star references (Pickles 1998) and filter transmissions, e.g., a NB $-$ g vs. $g-r$ relation for the NB387:

$$\begin{aligned} \mathrm{NB387_{HSC}} - g_{\mathrm{PS1}} = & \, 0.541 \times (g_{\mathrm{PS1}} - r_{\mathrm{PS1}})^2 \\ & + 1.87 \times (g_{\mathrm{PS1}} - r_{\mathrm{PS1}}) \\ & + 0.428 \, [+C_{\mathrm{metal}} + C_{\mathrm{fit}}], \end{aligned} \tag{2.2}$$

where $\mathrm{NB387_{HSC}}$ is the magnitude of NB387 based on HSC observations, and g_{PS1} and r_{PS1} are the g-band and r-band magnitudes, respectively, from the PS1 catalog. C_{metal} and C_{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 to reference the following coadding process. The global sky background is subtracted without masked regions. In the mosaicking of the CCD data, both the World Coordinate System (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 co-adds all visits of the image together, scaled with the WCS and flux correction.

For our data, it is necessary to optimize some configurations for NB387 because its 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 the Equation 2.2 to determine the photometric zero-point of NB387, we take into account additional corrections, including a systematic correction $C_{\rm metal}$ of -0.448 magnitude for correcting stellar metallicity and/or stellar age biases, and a field-dependent term, $C_{\rm fit}$, ranging within 0.2 mag for calibrating the fitting uncertainties. Details are described in the next Section 2.2.4.

2.2.4 Narrowband Photometric Zero-point Calibration

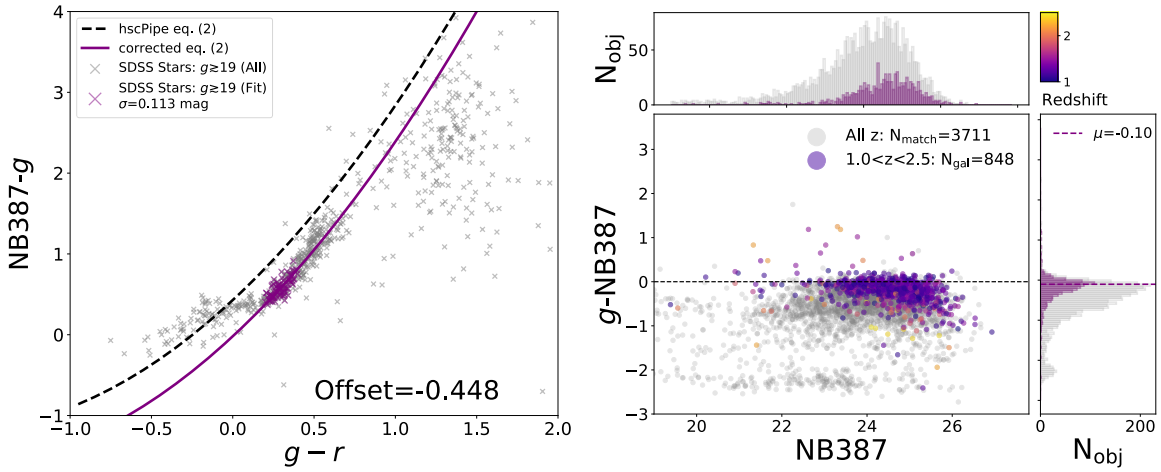

Figure 2.5: (*Left*) The predicted NB387 − g *vs.* $g-r$ diagram for homogeneously selected SDSS stars with $g \gtrsim 19$. The gray crosses are the selected stars, and the purple crosses are those with $0.2 < g-r < 0.4$ after visual inspection, which is used for fitting the correction factor $C_{\rm metal}$. The dashed black curve is Equation 2.2 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 10K catalog (Hasinger et al. 2018). The gray dots are all 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.

There is a systematic offset in the Equation 2.2 for NB387, and a constant should be introduced for correction as the colors between NB387, g, and r are influenced by the 4,000 Å *break*, which is sensitive to metallicity and stellar age (Kauffmann et al. 2003). The Pickles templates are constructed mainly from the 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 degenerate here, we denote the correction factor as $C_{\rm metal}$.

To estimate the $C_{\rm metal}$, 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 considering their total transmission curves. These stars are plotted as the gray crosses in Figure 2.5. To maintain consistency with the fitting in *hscPipe* and reduce the fitting uncertainty, we use only stars with $0.2 < g - r < 0.4$, showing the smallest scatter in the relation. Most selected stars are flagged as the SEGUE targets in the SDSS (Yanny et al. 2009). For robust estimation, we perform a visual inspection of each spectrum of these stars to discard those with peculiar features in the NB387 wavelength range. After this check, the stars used for the zero-point correction are plotted as purple crosses in Figure 2.5. We use these real stars instead of the Pickles templates to fit the relation shown in Equation 2.2, and the $C_{\rm metal}$ 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. Similar calibration is applied to the NB400 data, and the offset is measured as -0.140, much smaller than that of NB387 because its coverage is partly overlapped with the $4,000$ Å *break* in local star spectra.

When fitting the Equation 2.2, the scatter of references is prominent in the case of NB387, resulting in fitting uncertainty as large as 0.2 mags, thus causing field-to-field variation. We perform a 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 $4,000$ Å *break* in 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 $1 < z < 3$ galaxies given their typical UV slope (Kurczynski et al. 2014).

The $g -$ NB387 $= -0.10$ can also be verified by utilizing the HSC/NB387 data from Cosmic HydrOgen Reionization Unveiled with Subaru survey (CHORUS Inoue et al. 2020) and the spectral redshift $z_{\rm spec}$ from the DEIMOS 10K spectroscopic survey catalog (Hasinger et al. 2018) in the COSMOS field. We first cross-match the CHORUS objects with the spectroscopic catalog within a $2''$ aperture; there are 3,711 matches with flag $q > 1$, suggesting good spectroscopic redshift measurement. The $g-$NB387 *vs.* NB387 with $C_{\rm metal}$ correction of these objects is plotted as the gray dots

in the right panel of Figure 2.5. We pick out all 848 high-z galaxies with $1.0 < z < 2.5$ from the matched catalog, which is coded by the heat map in the figure, to measure the mean of the $g - \mathrm{NB387}$ in a dual-Gaussian distribution, as faint objects are likely in a flattened distribution due to photometric errors. The 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 the C_{metal} as a confident correction. 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 the C_{fit} in each of our fields in this paper. The resulting C_{fit} fluctuates in $-0.002 - 0.191$ mag among the four fields, consistent with the fitting uncertainty of 0.2 mag. We perform a similar measure for NB400 data, and this effect is also more minor as of < 0.1 mag, given that its coverage is partly covered by g-band, and the color term is less significant.

2.2.5 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 field. To minimize this influence, we apply the sky background root-mean-square (RMS) map as the weighting map in *SExtractor*. In addition, we use a local background with a thickness of 128 pixels. Masking of regions with low S/N signals, saturation around bright stars, or severe stray light is also applied in background estimation, object detection, and photometric measurement.

We use aperture magnitudes for the color selection; the aperture diameters are 15 pixels ($\sim 2.5''$) for BOSS0210 and 10 pixels ($\sim 1.7''$) for BOSS0210, BOSS0210 and BOSS0210. While for NB400 fields, we apply 12 pixels ($\sim 2.0''$) as the diameter for the aperture photometry as the seeing is $\sim 1''$ when doing the observation Auto-Mag is used to estimate total magnitudes, automatically determining elliptical aperture for Kron photometry in *SExtractor*. 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. 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 of the final stacked

images of NB387/NB400 and the g-band are listed in the Table 2.2. The quality of the BOSS0210 data is poorer than that of the other three fields in both the seeing and the final image depth.

2.3 HSC-SSP Data

The Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP; Aihara et al. 2019) is an imaging survey performed with the Subaru/HSC. In the second public data release (PDR2), the wide-field survey targets about 300 deg^2 sky area in all five $grizY$ broadband filters, while the Deep (26 deg^2) and UltraDeep (4 deg^2) fields include NB387, NB816, NB921, and NB1010 narrowbands. This thesis utilizes the LAE catalogs built based on the NB387 image data in the Deep/UltraDeep layers covering the DEEP2-3, Extended-COSMOS (E-COSMSO), and XMM-SXDS fields (Ono et al. 2021), which will be further introduced in Section 3.2.

Up to the PDR2 that is used in Ono et al. (2021), the HSC-SSP survey has taken 49^{+24}_{-17}, 45^{+24}_{-17}, 65^{+46}_{-37}, and 130^{+46}_{-51} minutes for the g, r, i, and z bands respectively, which reaching the depth of $27.3^{+0.4}_{-0.3}$, $26.9^{+0.2}_{-0.3}$, $26.7^{+0.3}_{-0.5}$, and $26.3^{+0.2}_{-0.4}$ (Aihara et al. 2019). NB387 images are taken for $\sim$ 1.3 hours per pointing, and it aims to reach a 5σ limiting magnitude of $\sim$ 24.5 magnitude. To be precise, we measure the 5σ limiting magnitude of the NB387 by each $12' \times 12'$ patch by randomly distributing 10,000 random apertures with the diameter size$d = 2''$ in the sky. The result of spatial fluctuation is shown in Figure 2.6 for the DEEP2-3, E-COSMOS, and XMM-SXDS fields, respectively; it is clear that the limiting magnitude fluctuates from 24.0 to 24.6 among fields and also within a field. The depth enables us to search for LAEs with Lyα luminosity comparable to those in BOSS0210 in our previous observations.

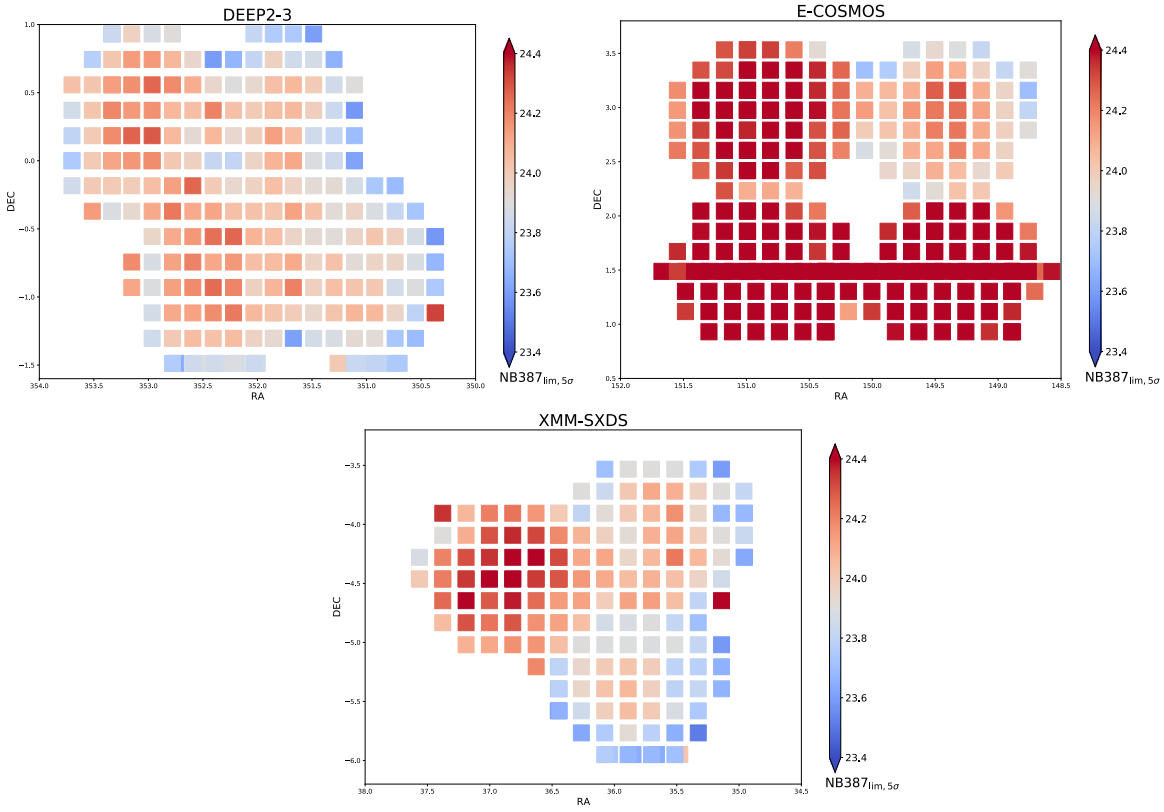

Figure 2.6: The sky distributions of 5σ limiting magnitude measured in each patch of the NB387 images from the HSC-SSP DR2. The measurement is performed by distributing 10,000 random apertures in the sky of each $12' \times 12'$ patchy image to measure the standard deviation of the Gaussian distribution. The redder color suggests the deeper depth of the image, while the bluer corresponds to the shallower depth. There are overlapped regions because sky areas can be covered in different patches from different tracts defined in the survey. The upper left, the upper right, and the lower panels are for the DEEP2-3, E-COSMOS, and XMM-SXDS fields.

Chapter 3

Lyα Emitters and their Spatial Distributions

3.1 Lyα Emitter Catalogs from Subaru/HSC Observations targeting MAMMOTH/QSO fields

3.1.1 Lyα Emitter Selection

We use the color excess of the narrowband to the broadband 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 $z = 2.18$ and $z = 2.29$ 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 $EW_0 = 20$ Å. The IGM absorption is considered 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 3.1 and 3.2. The black curves indicate the color tracks of $g-\mathrm{NB}$ vs. $u-\mathrm{NB}$ in the left panel, and $g-\mathrm{NB}$ vs. $r-\mathrm{NB}$ in the right panel. Three different UV slopes $\beta = 0$, -1.6, and -3.0, are shown in the both figures.

We also overplot the predicted tracks of possible contaminants, such as elliptical

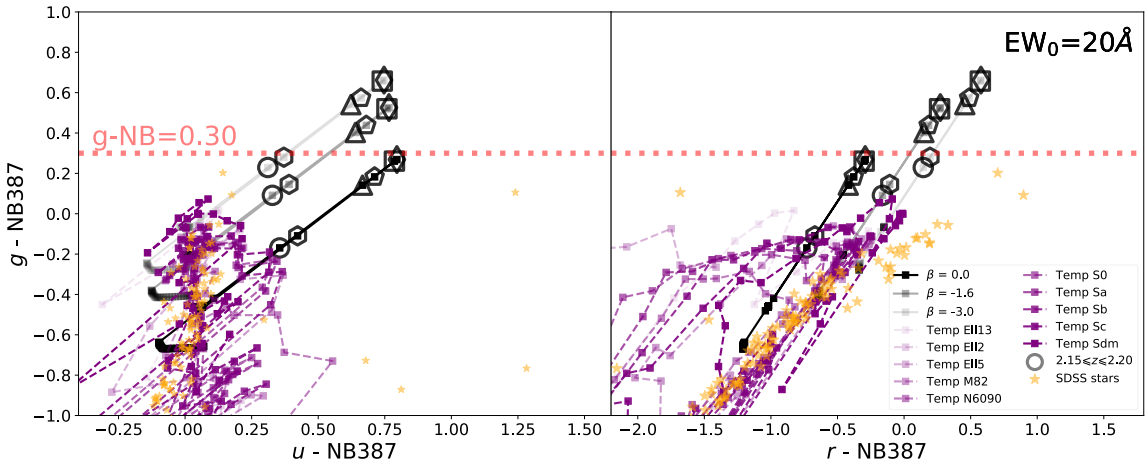

Figure 3.1: The color tracks with redshift evolution for the $EW_0 = 20$ Å LAE at the $z = 2.0 - 2.5$: $g - \mathrm{NB387}$ vs. $u - \mathrm{NB387}$ (*left panel*) and $g - \mathrm{NB387}$ vs. $u - \mathrm{NB387}$ (*right panel*). The black curves are the tracks for LAE models (with UV slope $\beta = 0, -1.6, -3.0$) with a redshift step $\Delta z = 0.01$ from $z = 2.0$ 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 $\Delta z = 0.1$ 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 $g - \mathrm{NB387} > 0.30$ works as a reasonable threshold to select the $z \sim 2.18$ LAEs.

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 are homogeneously archived from SDSS without the selection by prior knowledge of the colors.

The color tracks show that the r-band is of little help in LAE selection, while the u-band can help 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 $g - \mathrm{NB} > 0.3$ for NB387 and $g - \mathrm{NB} > 0.4$ for NB400 to exclude most of the contaminants.

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

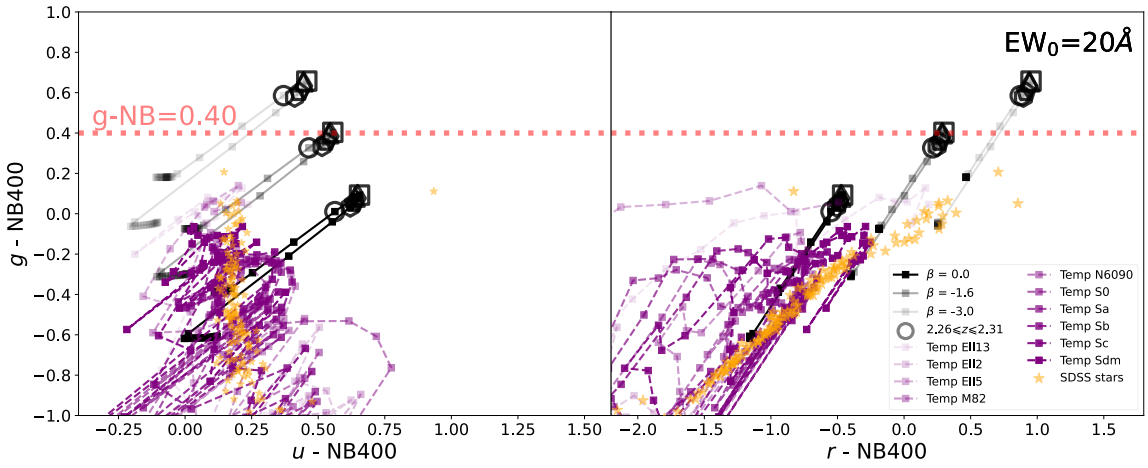

Figure 3.2: The same as Figure 3.1, but for NB400. The gray circle, triangle, square, diamond, pentagon, and hexagon indicate the LAE models at $z = 2.26$, 2.27, 2.28, 2.2.29, 2.30, and 2.31, respectively.

defined as

$$\begin{aligned} 20.5 \;<\; \mathrm{NB} \;&\lesssim \mathrm{NB}_{\mathrm{lim},5\sigma}, \\ g - \mathrm{NB} \;&> 0.3 \text{ or } 04, \\ g - \mathrm{NB} \;&> 2\sigma(\mathrm{NB387}) - 0.1 \\ &\quad \text{or } 3\sigma(\mathrm{NB400}) \;- 0.1. \end{aligned} \tag{3.1}$$

The lower limit of 20.5 is set for the NB387 and NB400 magnitude to avoid saturation, while the upper limit applies the 5σ limiting magnitude to ensure the reliability of narrowband detection. This upper limit for the field BOSS0924 with a moderate depth, 24.74, corresponds to 1.94×10^{42} erg s^{-1}, which is $0.37 \times L^*_{\mathrm{Ly}\alpha}$ and the characteristic luminosity $L^*_{\mathrm{Ly}\alpha} = 5.3 \times 10^{42}$ erg s^{-1} (Konno et al. 2016). The definition of the color error follows Shibuya et al. (2018):

$$n\sigma(\mathrm{NB}) = -2.5 \log_{10}\left(1 - n \times \frac{\sqrt{f^2_{1\sigma,\ \mathrm{NB}} + f^2_{1\sigma,\ g}}}{f_{\mathrm{NB}}}\right). \tag{3.2}$$

where the $n = 2$ or 3 follows the proper choice used in Nakajima et al. (2012), and the $f_{1\sigma,\ \mathrm{NB}}$ and $f_{1\sigma,\ g}$ are the 1σ flux limit of the NB and g-band images respectively. This aims to reject the false selection of faint objects that pass the criteria due to statistical fluctuation around the $g - \mathrm{NB} = -0.1$, where the high-z galaxy sequence lies.

Selected objects that pass the criteria are naturally filtered by the spatial masks, as the original object detection is performed with the masks applied. Finally, we per-

form 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 the cross-matches between our selected objects and the SDSS/(e)BOSS quasars with spectral redshift to discard the low-z contaminants. Eight are found in BOSS0210, six in BOSS0222, none in BOSS0924 or BOSS1419, eight in BOSS0240, and nine in BOSS0755. These known low-z quasars are removed from our LAE sample. As a result, 465, 956, 585, 636, 509, and 536 LAE candidates are selected in fields BOSS0210, BOSS0222, BOSS0924, BOSS1419, BOSS0240, and BOSS0755, respectively, i.e., 2,642 in total for the four NB387 fields and 1,045 LAEs for the two NB400 fields, covering an effective area of ~ 8 deg^2. These numbers are summarized in Table 3.1.

There are 4, 3, 0, 1, 2, and 2 proximity quasars from the SDSS/(e)BOSS that matched these LAE candidates in each field. Specifically, in the BOSS0210 central region, where 11 quasars reside, three of the quasars in our sample are selected as LAEs. At the same time, most others are too bright in the NB387 images and do not fulfill the selection criterion NB387 > 20.5. Figure 3.3 and 3.4 show the selected LAE candidates in the final catalog are plotted as yellow points in the $g-$NB387 vs. NB387 diagram, in which all detections are binned in the two-dimensional histogram coded by the blue color[1].

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 regarding the UV continuum given the depth limit of $R \sim 25.5$ in KBSS and $g \sim 26.4$ in Mukae et al. (2017).

3.1.2 Potential Contaminants

In addition to the LAEs at $z \sim 2.2$ and $z \sim 2.3$, some of the lower-z emitters may also pass our selection criteria. The contaminants are considered to be mainly [OII] emitters at $z = 0.036 \pm 0.008$ and $z = 0.074 \pm 0.013$ for NB387 and NB400, respectively. However, the survey volume is much smaller in this redshift range than at $z \sim 2.2/2.3$, and the ratio reaches 0.2%. Given the low-z [OII] 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

[1] 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 ugr diagram (Smolčić et al. 2004), and this is also supported by our random checks in matched SDSS spectra.

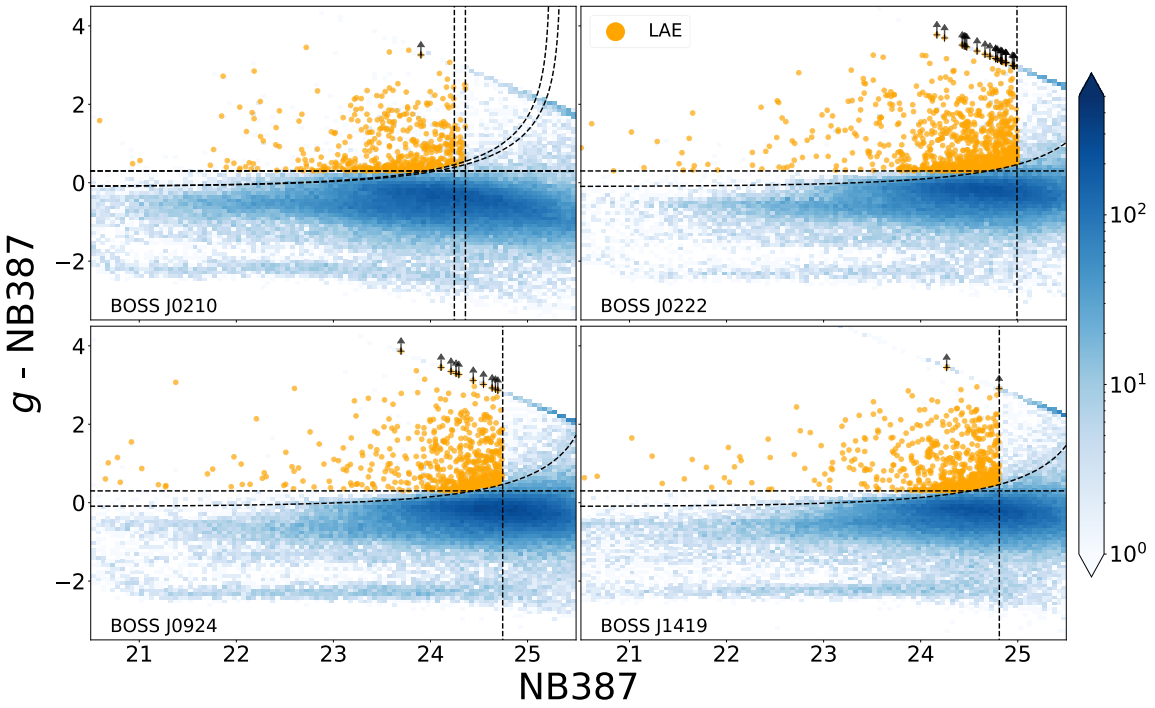

Figure 3.3: $g - \mathrm{NB387}$ vs. NB387 diagram for LAE selection in each field. The two-dimensional histogram bins all detections from *SExtractor*, and the bluer bins contain more objects. The yellow points are the LAE candidates selected after visual inspection. The three selection criteria are shown as black dotted lines. Specifically, for the field BOSS0210, 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 $g - \mathrm{NB387}$ values shown for these objects are the lower limits.

rate of low-z [OII] emitters in our sample is negligible. In addition, CIII] λ1909 at $z \sim 1$ and CIV 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 SXDS, COSMOS, HDFN, SSA 22, and E-CDFS (Guaita et al. 2010; Zheng et al. 2016; Konno et al. 2016), detection is performed in databases covering multiple wavelengths, e.g., the X-ray, UV, and radio wavelengths, to exclude low-z AGN contaminants. However, we search for overdense fields in the entire (e)BOSS survey; therefore, deep multi-wavelength data are unavailable for testing this study's AGNs. Instead, we refer to the literature mentioned above and find that

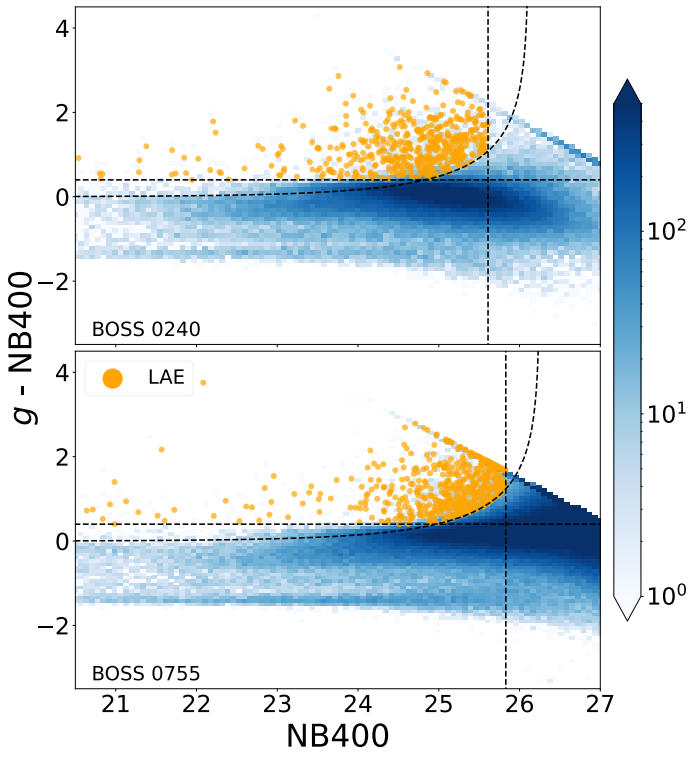

Figure 3.4: $g - \mathrm{NB400}$ vs. NB400 diagram for LAE selection in each field. The two-dimensional histogram bins all detections from *SExtractor*, and the bluer bins contain more objects. The yellow points are the LAE candidates selected after visual inspection. The three selection criteria are shown as black dotted lines.

the contamination rate of the LAE selections at $z \sim 2.2$ is normal $\sim 10 - 15\%$, and Sobral et al. (2017) confirms this number spectroscopically.

We test this contamination estimate for the case of HSC/NB387 by utilizing the COSMOS data, the NB387 data from CHORUS (Inoue et al. 2020), and the DEIMOS 10K spectroscopic survey catalog (Hasinger et al. 2018). This yields a contamination rate of $\sim 15\%$ in our LAE selection. About 2/3 of the interlopers are likely to be the CIV 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.

3.2 Lyα Emitter Catalogs from HSC-SSP Survey

3.2.1 Lyα Emitter Selection

In addition to our observations, the Hyper Suprime-Cam Subaru Strategic Survey (HSC-SSP) has also finished the Deep/Ultra-deep fields available with the NB387 data. Based on the second Data Release of the HSC-SSP, Ono et al. (2021) have constructed a $z = 2.18$ LAE sample including 3,720 candidates spanning the sky areas of ≈ 19 deg^2 in DEEP2-3, E-COSMOS, XMM-SXDS fields by applying the color excess selection:

$$\begin{aligned} \mathrm{NB387} &\lesssim NB387_{\mathrm{lim},5\sigma}, \\ g - \mathrm{NB387} &> 0.2, \\ r - \mathrm{NB387} &> 0.3, \\ i - \mathrm{NB387} &> 0.4, \\ z - \mathrm{NB387} &> 0.5, \end{aligned} \tag{3.3}$$

and the colors are also measured using the $2''$ aperture magnitudes from *SExtractor*, similar to our treatment in the previous data. Instead of performing a visual inspection to exclude contaminants, they utilize the CNN neural network to train the machine to remove the artifacts systematically. Therefore, this LAE sample can be of relatively high purity. For a more robust usage, we cross-check the LAE catalog with archival objects with spectral information to exclude the known low-z interlopers. There are 14 LAEs rejected with spectral redshift suggesting a different redshift at $z \sim 2.18$, 13 of them are based on the SDSS spectra, while 1 of them is based on the VIMOS Public Extragalactic Redshift Survey (Vergani et al. 2018, VIPERS). Finally, the sample size slightly drops from 3,720 to 3,706.

3.2.2 Completeness Estimate

Due to the fluctuation of survey conditions, the fields are not always homogeneous, which may affect the LAE completeness. As shown in Figure 2.6, the limiting magnitudes appear to fluctuate among patches. Especially, the (R.A., Decl.) = (149.3 deg, 3.0 deg) pointing in E-COSMOS and (R.A., Decl.) = (37.0 deg, -4.5 deg) pointing in XMM-SXDS are the most extreme cases. The former is systematically shallower than its adjacent pointings by ~ 0.5 magnitude, while the latter can be deeper by the others by another ~ 0.5 magnitude. This difference can also be checked in the histogram of the limiting magnitude in each field (i.e., Figure 3.5),

and E-COSMOS can be deeper than the other two regions. A similar inhomogeneous distribution also appears in the seeing PSF measure based on the survey PSF picker.

Therefore, one should perform careful calibration to account for incompleteness before using the LAE catalog. We use the *Balrog*, which is a Python script based on the *GalSim*, to simulate the galaxy images with the consideration of PSF convolution and then to insert the mock galaxies into the science image (Suchyta et al. 2021; Rowe et al. 2015). To better account for field variation, this process is made on each patchy image with the size of $12' \times 12'$. Before running the code, we must perform PSF matching among the images of different filters. Following Ono et al. (2021), we select the interpolated PSF at the center of each patchy image based on the PSF picker estimate from the HSC-SSP database. We then determine the largest PSF size of the filters in each patch and make the images with better PSF convolved with a proper Gaussian kernel to degrade to match the worst images. To run the *Balrog*, we also need the PSF exported from *PSFex* code (Bertin 2011), which helps to check further whether the PSF matching is appropriate. We make the mock galaxies share a flat probability of the magnitude from 20 to 26, and set the half-light radius=0.15 and the Sersic index as 1.5, mimicking the LAEs at $z \sim 2$. After the simulation, we do the source extraction and photometric measurement in the same manner of Ono et al. (2021), i.e., setting the DETECT_MINAREA=5, DETECT_THRESH=2.0, and using the $2''$ aperture as the photometric measurement. We then select the LAEs following the Criterion 3.3, and compare the detected objects and selected objected to the truth sample of mock galaxies to determine a total completeness $f_{\rm tot,comp} = f_{\rm detect,comp} \times f_{\rm select,comp}$, where $f_{\rm detect,comp} = {\rm N_{detect}}/{\rm N_{truth}}$ is the detection rate by *SExtractor* of the mock galaxies, and the $f_{\rm select,comp} = {\rm N_{select}}/{\rm N_{detect}}$ is the selection rate by the Criterion 3.3 of the detected objects. The $\rm N_{truth}$, $\rm N_{detect}$, and $\rm N_{select}$ are the total numbers of truth mock galaxies, detected objects, and selected objects with the magnitude up to the 5 limiting magnitude in each patch. The result of this procedure is shown in Figure 3.6. It shows that in most regions, the completeness can be as high as 95% even though the limiting magnitude is vastly fluctuating. The only exception is the (R.A., Decl.) = (37.0 deg, -4.5 deg) pointing in XMM-SXDS, where the completeness is as low as 75%. It seems this is mainly caused by its relatively poor seeing ($\sim 1''$ compared to $\sim 0.6''$ in the other fields).

The completeness will be used in the following sections to calibrate the overdensity estimate and work as the criterion to discard specific regions for our homogeneous inspection. For example, in Figure 3.6, the black boxes denote the patches with completeness over 80%, and if we apply this constraint in the two-point cross-correlation analysis, the abnormal pointing in XMM-SXDS will be masked.

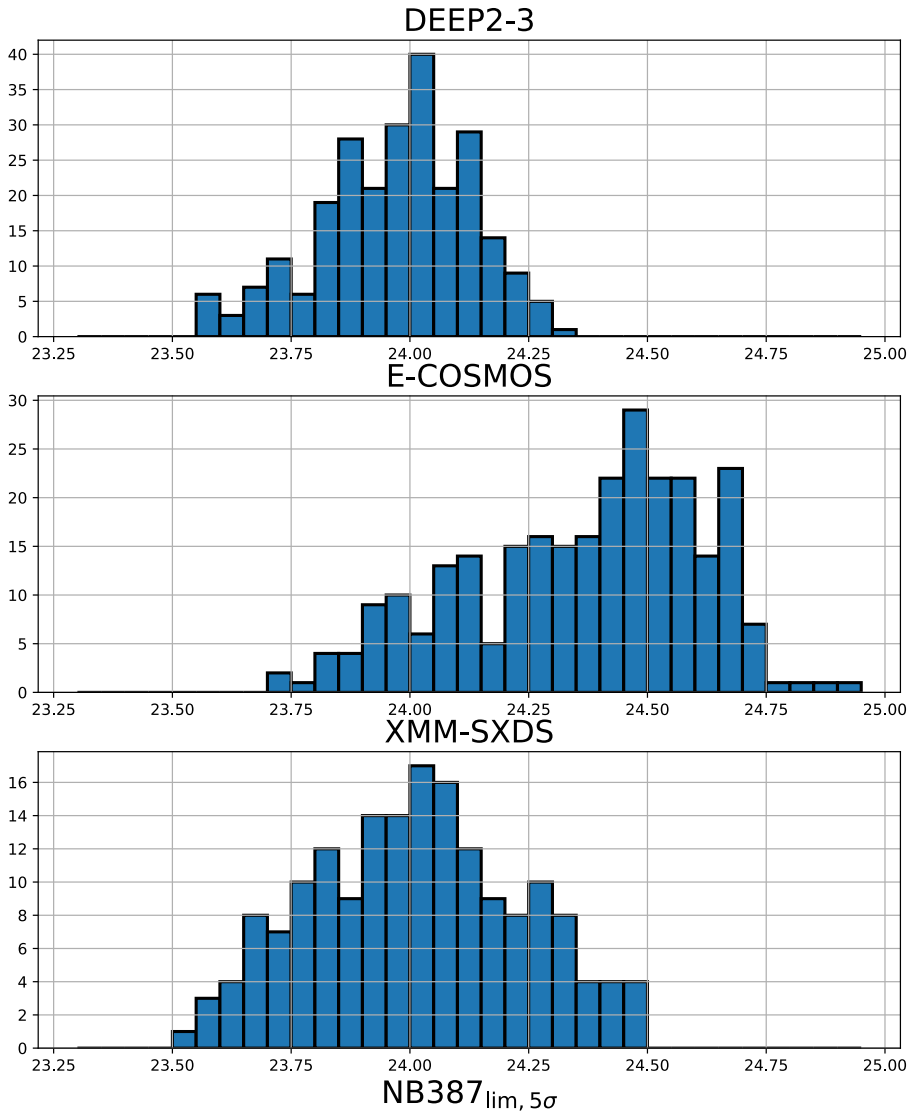

Figure 3.5: The distribution of the 5σ limiting magnitude measured in each patch of the NB387 images from the HSC-SSP DR2, whose data is the sample as Figure 2.6. DEEP2-3, E-COSMOS, and XMM-SXDS fields are from top to bottom, respectively. From these histograms, it is more apparent to see the inhomogeneous distributions of the limiting magnitude among fields.

3.3 Lyα Overdensity Map

3.3.1 NB387 LAE Overdensity – Traced by Lyα Absorbers

The sky distribution of the selected LAE candidates is shown in Figure 3.7.

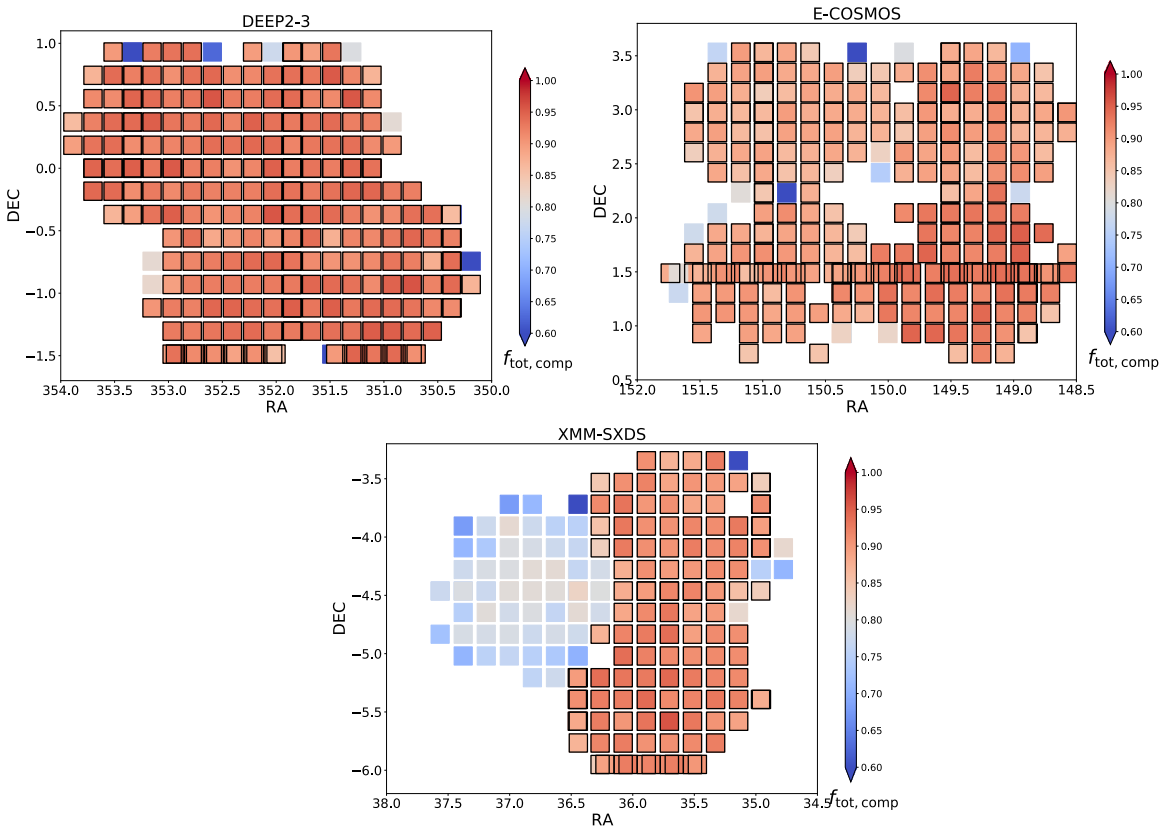

Figure 3.6: The sky distributions of total completeness $f_{\rm tot,\ comp}$ estimated in each patch for the HSC-SSP $z = 2.18$ LAE selection (Ono et al. 2021). The estimate is performed by simulating mock LAE images and inserting them into the NB387 images of each $12' \times 12'$ patch with *Balrog* and *GalSim* (Rowe et al. 2015; Suchyta et al. 2021). We then follow the selection presented in Ono et al. (2021) and perform object detection with *SExtractor* and color excess selection to calculate the completeness of the mock galaxies. The redder color suggests the deeper depth of the image, while the bluer corresponds to the shallower depth. The boxes with black boundary suggest the patch is with $f_{\rm tot,\ comp} > 0.8$. There are overlapped regions because there can be sky areas covered in different patches from different tracts that are defined in the survey. The upper left, the upper right, and the lower panels are for the DEEP2-3, E-COSMOS, and XMM-SXDS fields. We note a special pointing at the west of XMM-SXDS field with systematical lower completeness of $f_{\rm tot,\ comp} < 0.8$.

We calculate the galaxy overdensity of each field to construct the overdensity maps. Overdensity is defined as

$$\delta_{\rm LAE} = \frac{{\rm N}_{i,\ {\rm LAE}} - \langle {\rm N}_{\rm LAE} \rangle}{\langle {\rm N}_{\rm LAE} \rangle}, \tag{3.4}$$

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

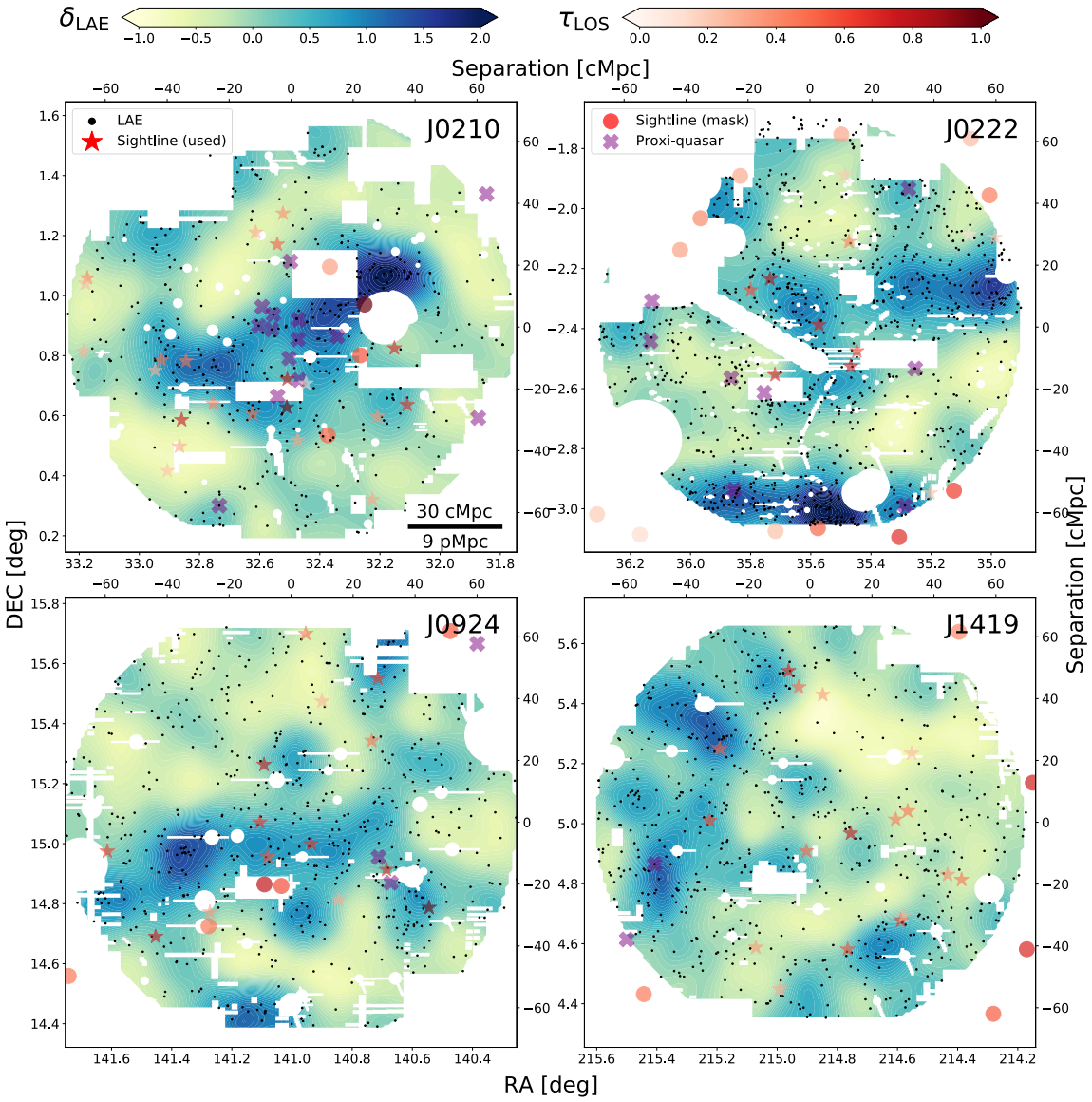

Figure 3.7: Overdensity maps for the four selected fields BOSS0210, BOSS0222, BOSS0924, and BOSS1419. Black points represent the LAE candidates. The green contour in the background is scaled by the LAE overdensity δ_{LAE} on a scale of $r = 10$ cMpc. The red stars and dots are the positions of the LoSs 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 BOSS quasars with $2.15 < z < 2.20$. The white regions are masks that exclude regions with low S/N signals, saturation around bright stars, or serious stray light.

The aperture size is set at ten cMpc ($\sim 6'$) 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 $\sim$ one 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 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 BOSS0210, BOSS0222, BOSS0924, and BOSS1419 are 10.7 (6.0), 23.5 (10.1), 12.6 (4.9), and 14.0 (5.4), respectively. The smaller mean number in BOSS0210 and the more significant number in BOSS0222 are due to image depth differences.

In Figure 3.7, the blue contours in the background show the overdensity. The masked regions defined in Section 2.2.5 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 adequate optical depth $\tau_{\rm LoS}$. The more quantitative discussion on the large-scale structures found in these fields is under preparation by our collaborators (Z. Cai et al. in prep.), but one can find a brief overview of the features of these fields below:

⊙ BOSS0210

There is a sizeable filamentary structure at the center of the field, and the structures with weaker significance are likely to be sheet-like 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 the combined use of both tracers can effectively trace the unique LSSs, as 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. 2020b), the emergence of the grouping of quasars implies that the filament is different from the typical environments at $z \sim 2$, and BOSS0210 should be considered individually in the correlation analysis.

⊙ BOSS0222

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 is 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.

⊙ BOSS0924

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 BOSS0924, several peaks have moderate overdensities $\delta_{\rm LAE} > 0.8$ comparable to or surpassing the central structure. The most overdense structure is found southwest of the field, close to two LoSs with strong Lyα absorption. The structure's peak has an overdensity measured as over 1.2, extending for about 30 cMpc.

⊙ BOSS1419

There are more structures in clumpy shapes in BOSS1419. Although there are four LoSs with $\tau_{\rm LoS} > 0.6$, they are more scattered than those in other fields, with distances of $\sim 40-100$ 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 $\delta_{\rm LAE} > 0.6$ 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 northwest of the FoV.

3.3.2 NB400 LAE Overdensity – Traced by Grouping Quasars

Following the identical manners performed for NB387 data as introduced in Section 3.3.1, we calculate the overdensity based on the fixed aperture method with the aperture size of 10 cMpc. We note that although NB400 images of these two fields are relatively deeper than the NB387 images, the stronger constraint on $g-$NB color has suppressed the selection of the LAEs, leading to comparable LAE numbers to that of the NB387 fields, except for BOSS0210. Therefore the aperture size of 10 cMpc is an appropriate setting.

There are some interesting features one can find from the overdensity map, and one of the common points for these fields is that they both enclose grouping quasars

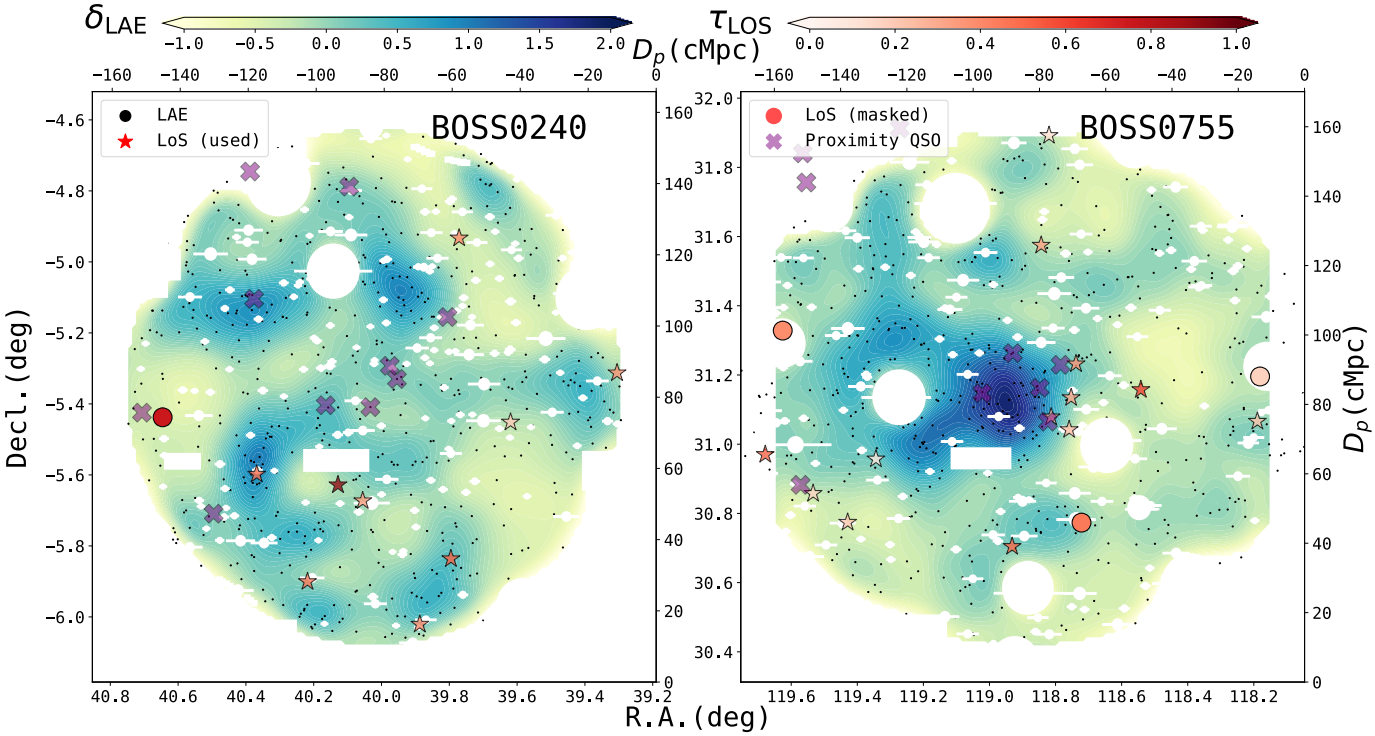

Figure 3.8: Overdensity maps for the two selected grouping quasar fields BOSS0240 and BOSS0755. Labels have the same meaning as those in the Figure 3.7, but the purple crosses represent the proximity BOSS quasars at the redshift interval $2.25 < z < 2.34$ for the NB400.

at the central region, i.e., ~ 5 BOSS quasars within 40×40 cMpc^2.

⊙ BOSS0240

Interestingly, the overdensity in BOSS0240 seems to avoid the region where quasars intensively emerge, while two weak nodes are located at the northeast and southwest sides, respectively. The quasars likely reside in the filament that connects the two nodes, which is very similar to the case in BOSS0210. There are few LoSs located in this field, especially not in the high-density area, but a high-$\tau_{\rm LoS}$ LoS at around (R.A., Decl.) = (40.1 deg, -5.75 deg) is located in an underdense region.

⊙ BOSS0755

A significant overdensity with $\delta_{\rm LAE} > 2$ is found in this field, which extends for ~ 20 cMpc. Different from BOSS0240, there is an overdensity closely located at the position of the grouping quasars. However, interestingly, the quasars do not lie in the densest peak but rather distribute around the outskirt of the overdensity. Also, there are four LoSs penetrating the outskirt, but none shows strong Lyα absorption. Just

comparing the two fields, the variation of large-scale structures is already distinctive.

3.3.3 HSC-SSP LAE Overdensity

As explained in Section 3.2.2, we have performed the completeness estimate for each $12'$ patchy image in the three Deep/Ultra-deep fields. We only apply the completeness calibration to the HSC-SSP LAEs because the observations carried out by the survey are in multiple periods for several pointings in each field, which results in a large field fluctuation of the image quality. Meanwhile, the Subaru/HSC observations targeting Lyα absorber-trace fields and grouping quasar fields are performed in a single night with one pointing per field, where the completeness is nearly homogeneous within a pointing, as also demonstrated in Figure 3.6, so the completeness calibration is not applied to the LAEs from these runs.

We match the LAEs with their originated patches and apply this completeness $f_{\mathrm{i,\ tot,comp}}$ in the i^{th} patch when estimating the δ_{LAE}. Similar to previous work in Section 3.3, we use the fixed aperture method with the aperture of $r = 10$ cMpc size to estimate the mean number density $\langle N \rangle_{\mathrm{LAE}}$ over all the regions in one field, and then calculate the overdensity follow Equation 3.5, Except for a minor change that whenever an LAE is counted in the aperture, it is weighted by the invert completeness $1/f_{\mathrm{i,\ tot,comp}}$, and thus, the revised LAE overdensity will be:

$$\delta_{\mathrm{LAE,cor}} = \frac{\mathrm{N}_{i,\ \mathrm{LAE}} \times 1/f_{i,\ \mathrm{tot,comp}} - \sum_i (\mathrm{N}_{i,\ \mathrm{LAE}} \times 1/f_{i,\ \mathrm{tot,comp}})}{\sum_i (\mathrm{N}_{i,\ \mathrm{LAE}} \times 1/f_{i,\ \mathrm{tot,comp}})}, \tag{3.5}$$

In addition to the completeness, the limiting magnitude fluctuation within one field is also not negligible since each consists of several pointings, and their condition can be vastly different. Referring to Figure 3.5, we decide to set the faintest magnitude cut at NB387$<$ 24.2 to ensure the sample is mostly homogeneous and guarantee that we still have sufficient numbers of LAEs for performing statistical analysis. After applying this magnitude cut, the sample of LAEs dropped from 3,706 to 1,361. This is a drastic drop in the sample size in LAEs, but we note that the survey area is the more critical factor when inspecting field variation. The final number of LAEs from HSC-SSP is summarized in Table 3.1.

In addition, we apply the masks based on the bright stars with Gaia g-band brighter than 11^{th} magnitude, which is bright enough to cause ghosts in any of the broadband images and, thus, has the possibility of affecting the LAE selection. This mask will also be applied in the following analysis for the HSC-SSP fields.

The final result of the overdensity maps for DEEP2-3, E-COSMOS, and XMM-SXDS fields are shown in Figure 3.9, 3.10, and 3.11, repectively. The background

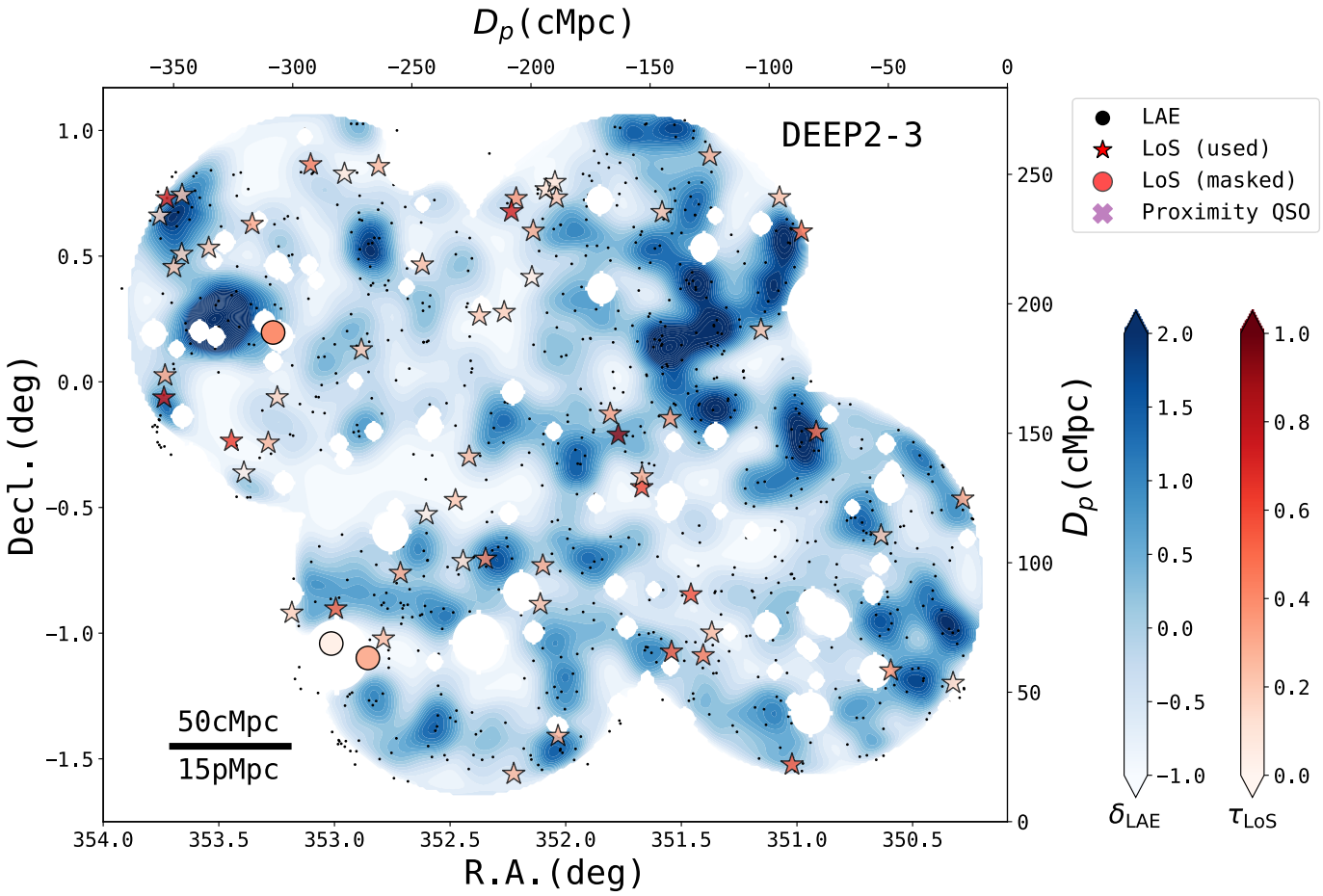

Figure 3.9: Overdensity maps for the $z = 2.18$ LAEs in the DEEP2-3 field from HSC-SSP Deep/Ultral-Deep layer. Black points represent the LAE candidates. The blue contour in the background is scaled by the LAE overdensity $\delta_{\rm LAE}$ on a scale of $r = 10$ cMpc. The red stars and circles are the 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 white regions are masks that exclude regions with low S/N signals, saturation around bright stars, or serious stray light.

contour suggests the $\delta_{\rm LAE}$ is smoothed with a 3σ Gaussian kernel for showing.

One can easily find that there are some structures, such as the significant ($\sim 6\sigma$) and extended (~ 50 cMpc) overdensity at (R.A., Decl.) = (149.5 deg, 1.4 deg) in the E-COSMOS field. However, unfortunately, there is no appropriate LoS penetrating this region, and it has little contribution to our overdensity-based correlation analysis.

As to DEEP2-3 field, we can also visually notice that a region of LAE under density at (R.A., Decl.) = (352.1 deg, 0.2 deg) happens to enclose 3-4 LoSs, and

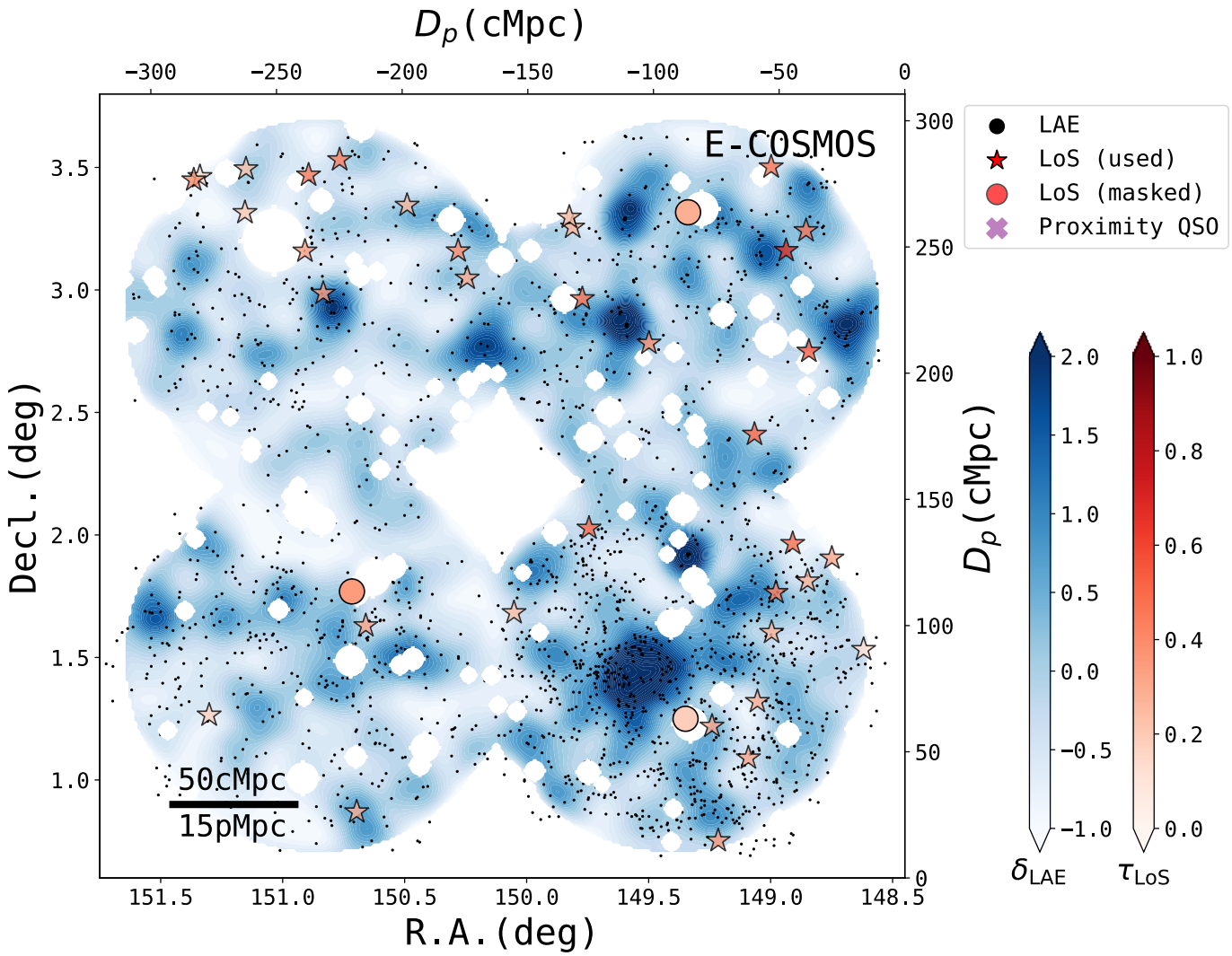

Figure 3.10: Overdensity maps for the $z = 2.18$ LAEs in the E-COSMOS field from HSC-SSP Deep/Ultral-Deep layer. Labels have the same meaning to those in Figure 3.9

none of them shows strong absorption. This suggests a consistent picture of the density distribution in terms of galaxies and IGM HI in such low-density areas.

The east pointing in XMM-SXDS field shows distinctively low completeness, but after the calibration, the overdensity map shows no robust features. Although the overdensity peaks are rare in this pointing compared to the other two, the LoSs in this pointing show moderate values, agreeing to the rarity of LAE overdensities. Therefore, this abnormal pointing is not likely to affect the statistics with significant weight, but we will exclude it when completeness calibration does not work for analysis, e.g., the

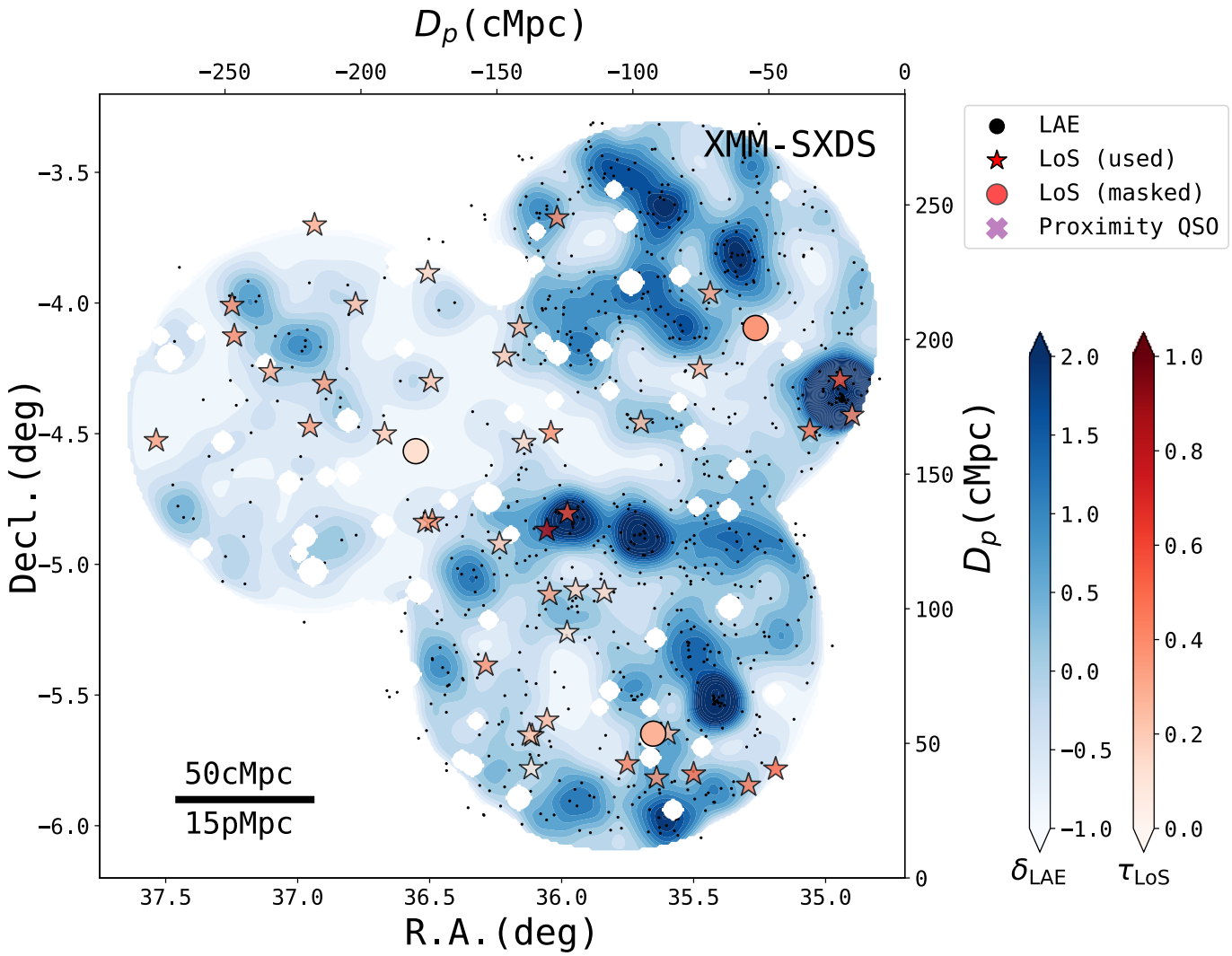

Figure 3.11: Overdensity maps for the $z = 2.18$ LAEs in the XMM-SXDS field from HSC-SSP Deep/Ultral-Deep layer. Labels have the same meaning to those in Figure 3.9

two-point cross-correlation analysis.

3.4 Summary of the Final Sample

Combining all three categories of LAEs, we finally have 5,048 LAEs in total. Among them, 2,642 LAEs at $z = 2.18$ are from our PI observation with NB387 on Subaru/HSC, 1,045 LAE at $z = 2.18$ are from our PI observation with NB400 on Subaru/HSC, and 1,361 LAEs at $z = 2.18$ are from the NB387 data in HSC-

SSP Deep/UltraDeep fields as summarized in Table 3.1. We further look into our LAE samples by plotting the g-band and NB387/NB400 magnitude distributions in the surface density of the LAE candidates in Figure 3.12, 3.13 and 3.14 with the Poisson errors estimated by the statistics proposed in Gehrels (1986). In the NB387 Lyα absorber-trace fields, the BOSS0210 and BOSS0222 are found with the excess at around 23–24 mags. In the NB400 grouping quasar fields, LAEs in BOSS0240 are likely more abundant than those in BOSS0755, despite the fact that an LAE density peak is found in BOSS0755; this finding suggests that on a scale of $\sim$ 160 cMpc, galaxy numbers may not fluctuate significantly due to the existence of individual protoclusters. In the NB387 field LAEs, no clear difference is found in the NB387 Auto-Mag except for an excess in E-COSMOS at the faintest end of NB$>$ 24.0, which is likely due to the deeper depth of E-COSMOS field, as shown in Figure 3.5. Interestingly, excess at the bright end can also be found in the g-band of LAEs in E-COSMOS, which suggests an excess of relatively massive galaxies with a bright continuum. This excess possibly originates from the $\sim 6\sigma$ extended overdensity at (R.A., Decl.) = (149.5 deg, 1.4 deg), and the bright galaxies, which are relatively rare, can stand out due to such an enormous structure.

The Auto-Mags are shown here for a comparison of total magnitudes with the literature, and we overplot the $z \sim 2.2$ LAE sample in the COSMOS field from Nakajima et al. (2012) that was selected by the Subaru/Suprime-Cam NB387 ($\lambda_0 =$ 3870 Å, FWHM = 94 Å). As their FWHM is similar to the HSC NB400 but 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 there is a relative excess in BOSS0210 and BOSS0222, which are not surprising, as our fields are intentionally selected to contain potential overdensities. On the other hand, the general fields from HSC-SSP are consistent in both shapes and amplitudes.

After masking, the final numbers of clean LoSs in BOSS0210, BOSS0222, BOSS0924, BOSS1419, BOSS0240, and BOSS0755 are 22, 11, 14, 17, 9, and 13, respectively, as summarized in Table 3.1. The masked clean LoSs are hatched in Figure 2.3, and the remaining 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 Section 4.1.1 and 5.2.

Field	N_{LAE}	N_{LoS}	Area [deg^2]	NB Filter	Sample Category
BOSS J0210+0052	465	22	1.34	NB387	Lyα absorbers & Grouping quasars
BOSS J0222-0224	956	11	1.13	NB387	Lyα absorbers
BOSS J0924+1503	585	14	1.47	NB387	Lyα absorbers
BOSS J1419+0500	636	17	1.45	NB387	Lyα absorbers
Total in PI NB387	2642	64	5.39	...	...
BOSS J0240-0521	509	9	1.53	NB400	Grouping quasars
BOSS J0755+3109	536	13	1.54	NB400	Grouping quasars
Total in PI NB400	1045	22	3.07	...	...
HSC-SSP	1361	137	19.03	NB387	General fields
Total	5048	223	27.49	...	...

Table 3.1:: Information on the LoSs and LAEs in each field.
[1] The field name;
[2] The number of LAE candidates;
[3] The number of clean LoSs after being masked, which are used in the correlation analysis in this thesis;
[4] The effective survey area for selecting LAEs after being masked;
[5] The narrowband filter used for searching LAEs;
[6] The sample category of each field, BOSS0210 was selected primarily by the grouping strong Lyα absorbers, but it also enclose a group of BOSS quasars.

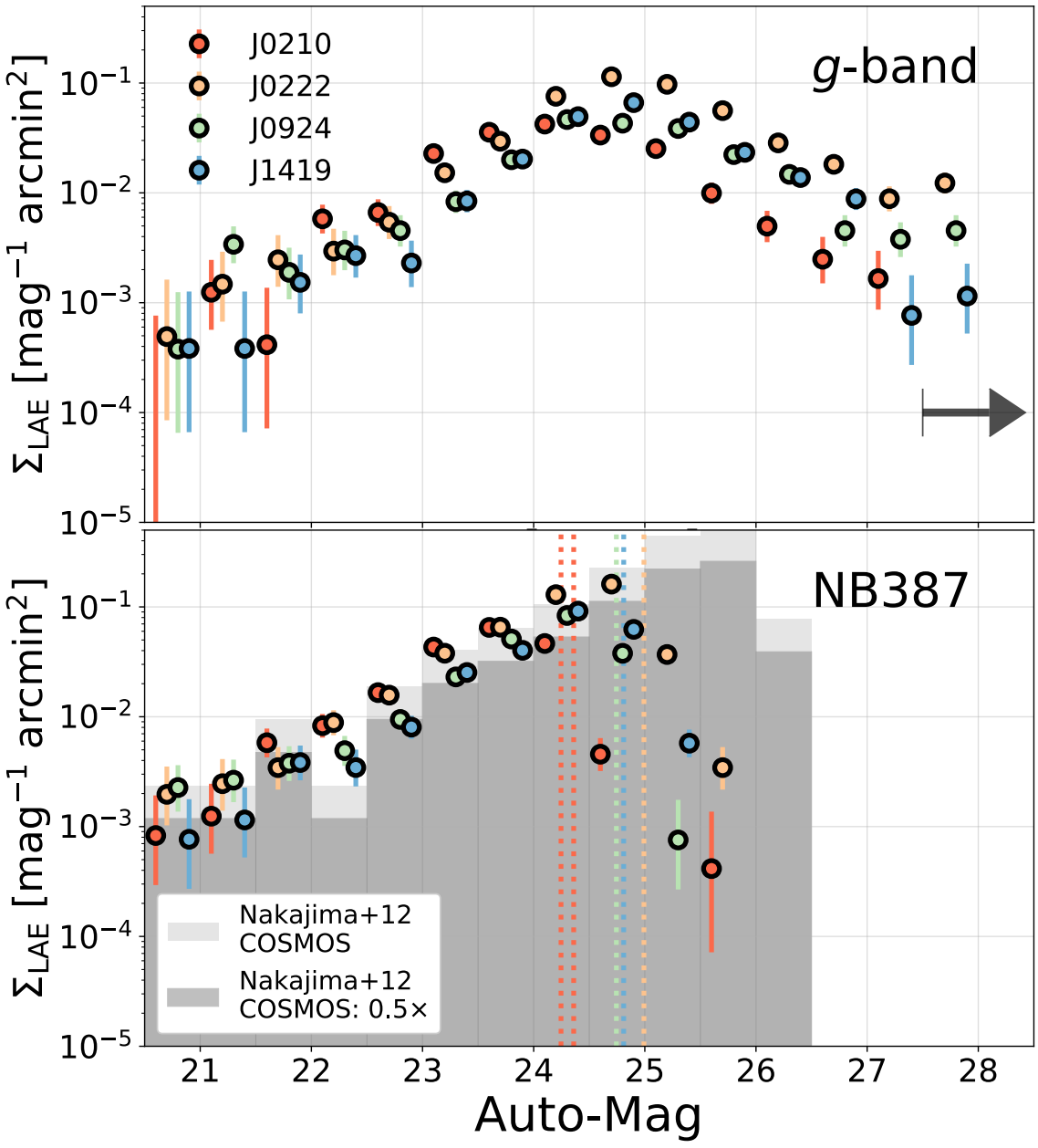

Figure 3.12: The Auto-Mag surface density distributions of the LAE candidates. Red, yellow, green, and blue histograms represent the LAEs in BOSS0210, BOSS0222, BOSS0924, and BOSS1419, respectively. Error bars indicate the Poisson errors. *Upper:* g-band magnitudes. The black arrow indicates that the 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.

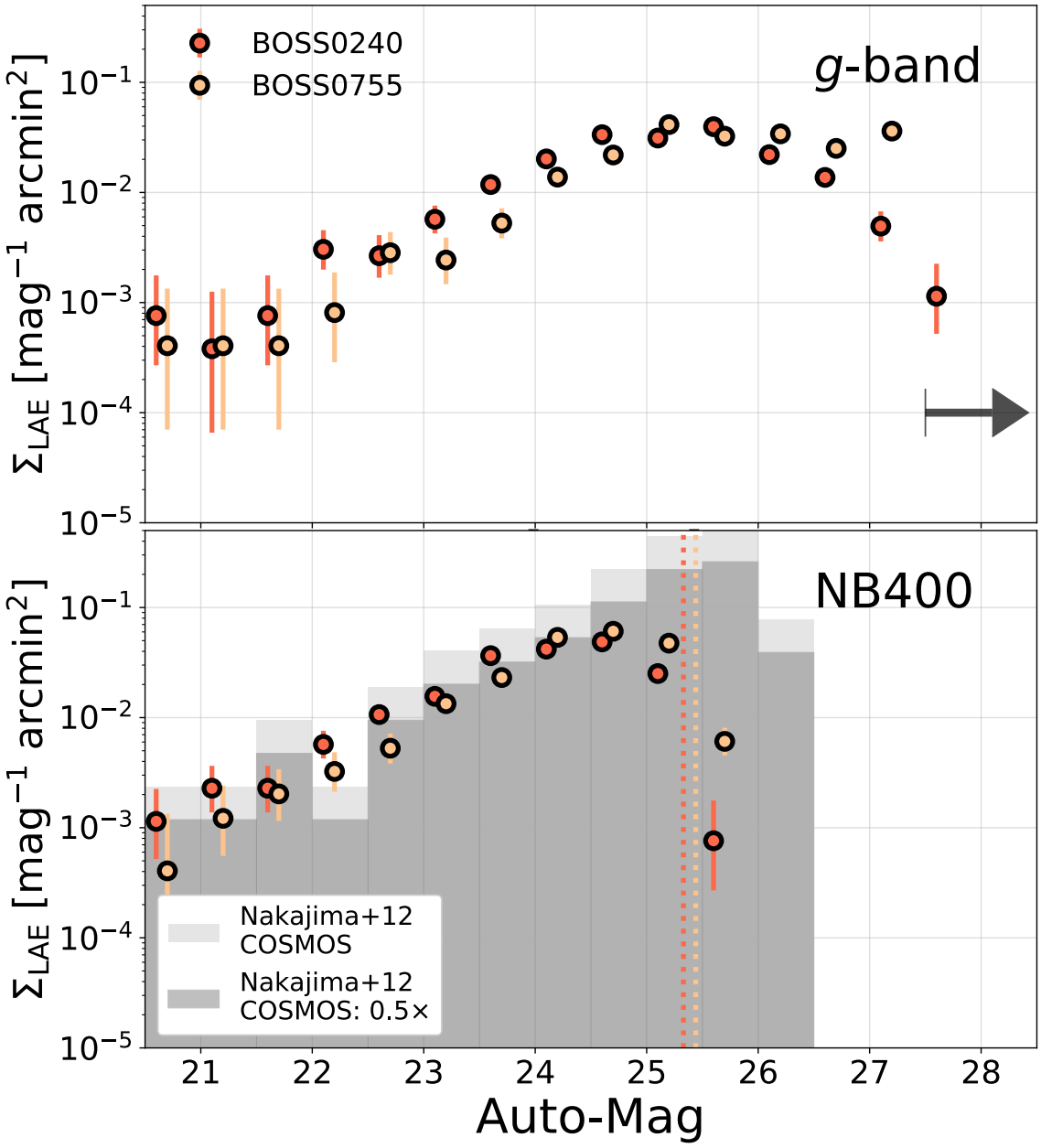

Figure 3.13: Same as Figure 3.12, but for the LAEs selected from the NB400 images of the grouping quasar fields. Red and yellow histograms represent the LAEs in BOSS0240 and BOSS0755, respectively.

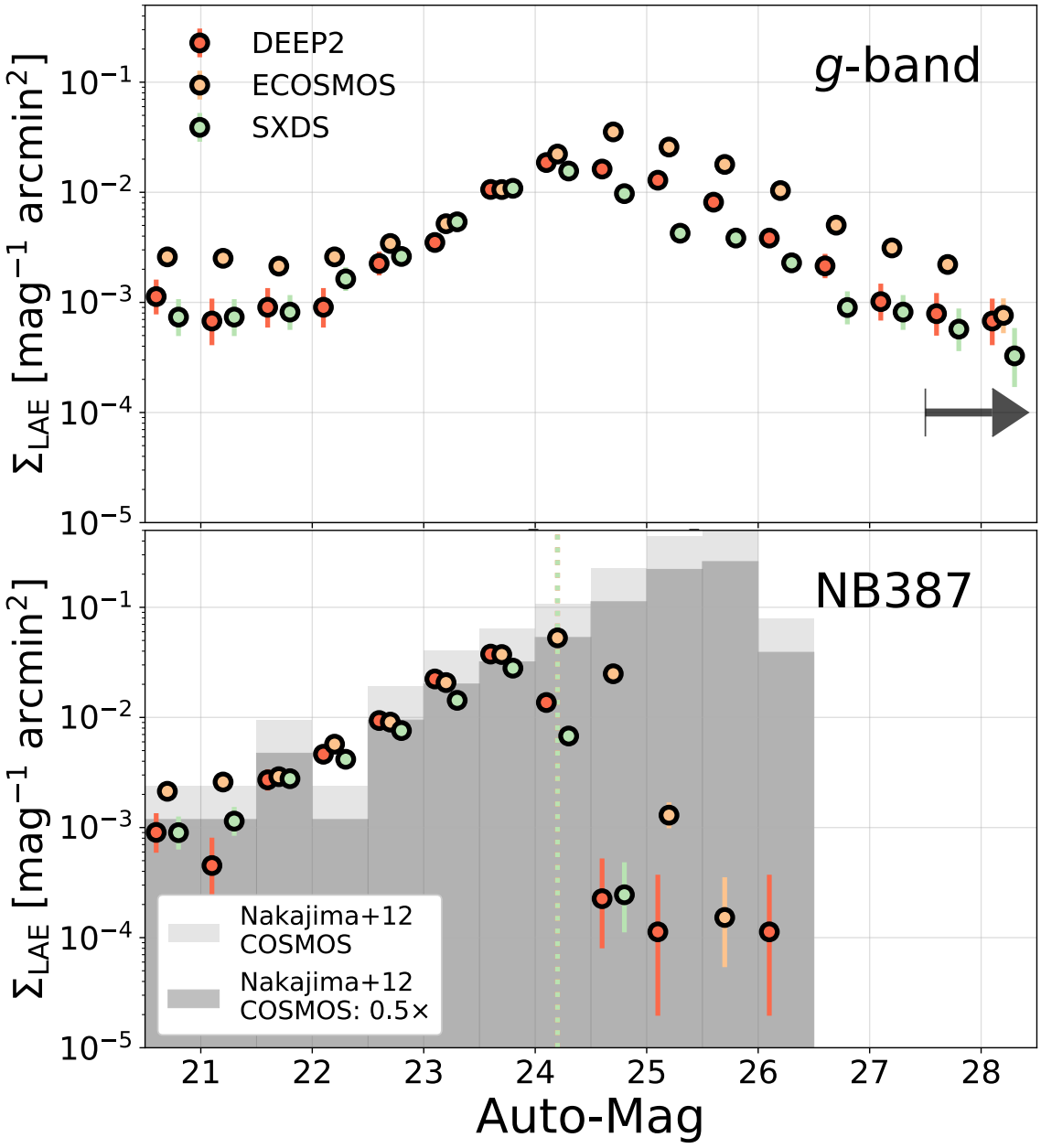

Figure 3.14: Same as Figure 3.12, but for the LAEs from HSC-SSP survey (Ono et al. 2021). Red, yellow, and green histograms represent the LAEs in DEEP2-3, E-COSMOS, and XMM-SXDS field, respectively. The vertical line shows the NB = 24.2 used for cutting the sample.

Chapter 4

The LAE-IGM HI Correlation

4.1 Correlation between $\delta_{\rm LAE}$ and $\tau_{\rm LoS}$ & $\delta_{\langle F \rangle}$

4.1.1 Fields Traced by Lyα Absorbers

We first quantify the galaxy–IGM HI correlation based on the local overdensity estimate for the field traced by grouping strong IGM Lyα absorbers. We calculate the overdensity on the scale of 10 cMpc in radius at the positions of the clean LoSs. Similar to Section 3.3, we discard LoSs whose vicinities are masked out by more than 50%, but as a result, no LoSs are removed in this 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 of the LAE overdensity are estimated as the Poisson noise using the statistics proposed in Gehrels (1986), which is the dominant uncertainty due to the small number of 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 4.1, from which it can be seen that the typical error of $\delta_{\rm LAE}$ shown at the upper left corner 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.

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 entire 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 solid 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 $z > 2$.

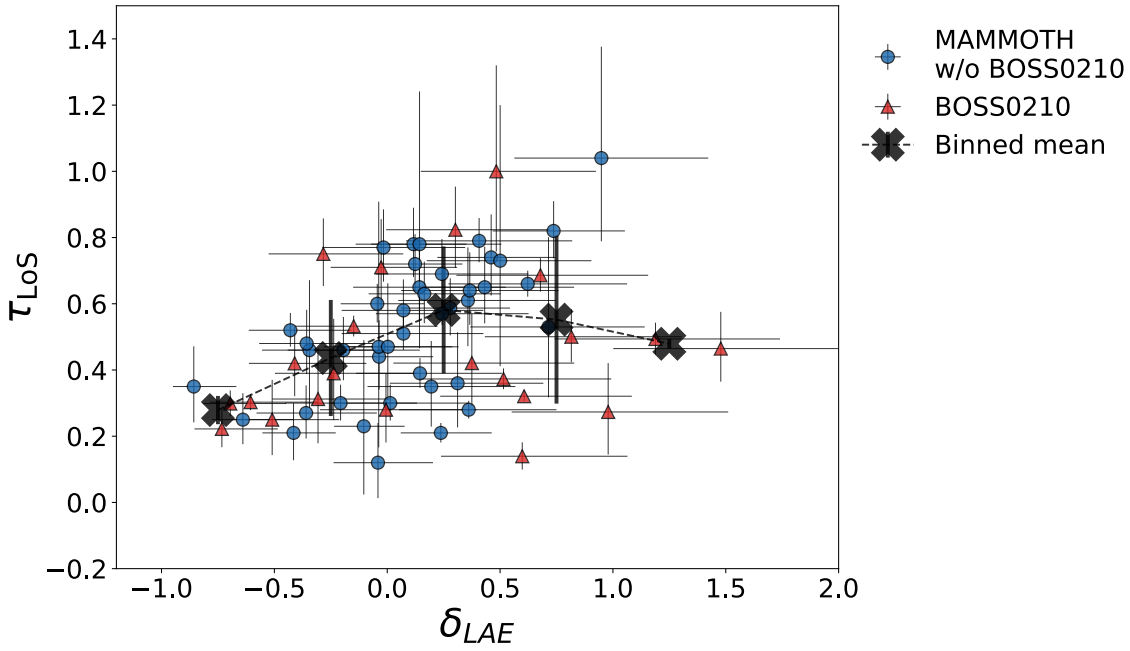

Figure 4.1: Correlation between LAE overdensity $\delta_{\rm LAE}$ and effective optical depth $\tau_{\rm LoS}$ at the positions of LoSs. Red triangles, yellow diamonds, green circles and blue squares represent the original data points in fields BOSS0210, BOSS0222, BOSS0924, and BOSS1419, respectively. 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 BOSS0210, a clearly increasing trend is observed.

The large scatter in Figure 4.1 is primarily contributed by the LoSs in BOSS0210 (red triangles), which has a unique structure and a shallower limiting magnitude. If we exclude BOSS0210, 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 BOSS0210 is shallower than that of the others, we test this effect 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 BOSS0210, BOSS0222, BOSS0924,

and BOSS1419, respectively. The results with ($\rho_S = 0.388^{+0.026}_{-0.045}$) and without ($\rho_S = 0.502^{+0.031}_{-0.061}$) BOSS0210 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 driven origin.

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 BOSS0210 are probably different from those of other fields. The binning data, show a clearer trend, which is overlaid as the gray crosses in Figure 4.1. The $\tau_{\rm LoS}$ increases with the $\delta_{\rm LAE}$, although interestingly, the trend becomes flattered when $\delta_{\rm LAE} \gtrsim 0.2$. The trend at the overdense end is likely dominated by the BOSS0210 LoSs in the $\delta_{\rm LAE} > 0.5$ bins. This may indicate different physical processes in the BOSS0210 filament compared to typical structures at the same redshift.

One can also convert the optical depth derived in Section 2.1 into the Lyα forest transmission fluctuation $\delta_{\langle F\rangle}$ according to:

$$\delta_{\langle F\rangle} = \frac{\langle F\rangle_{dz}}{F_{\rm cos}(z)} - 1, \tag{4.1}$$

where $\langle F\rangle_{dz}$ is the transmission calculated within the redshift uncertainty $dz = 0.025(1+z)$ from the spectra and the $F_{\rm cos}(z)$ is the cosmic Lyα forest mean transmission that is estimated from $F_{\rm cos}(z) = e^{-0.001845(1+z)^{3.924}}$ (Faucher-Giguère et al. 2008). 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 Lyα absorber-trace field sample is shown in Figure 4.2.

The symbols used for the data points are the same as in Figure 4.1, but to clarify the effects of including or excluding BOSS0210 in fitting, the LoSs in BOSS0210 are shown in red, and the LoSs in the other fields are shown in blue. We also present the linear fits, i.e., the dotted-dashed lines in Figure 4.2, with and without the BOSS0210 corresponding to the light and deep blue curves, respectively. The uncertainty of parameters is given by the $16\% - 84\%$ ranks from the 10,000 Monte Carlo simulations with perturbation, shown as the colored confidence band. The fitted relation for all four fields is:

$$\delta_{\langle F\rangle} = -0.116^{+0.018}_{-0.022}\ \delta_{\rm LAE} - 0.248^{+0.082}_{-0.093}, \tag{4.2}$$

If we exclude the LoSs in BOSS0210, the relation becomes:

$$\delta_{\langle F\rangle} = -0.227^{+0.026}_{-0.023}\ \delta_{\rm LAE} - 0.258^{+0.096}_{-0.114}, \tag{4.3}$$

which shows a steeper slope, meaning that the $\tau_{\rm LoS}$ is more sensitive to the $\delta_{\rm LAE}$. The results are summarized in Table 4.1.

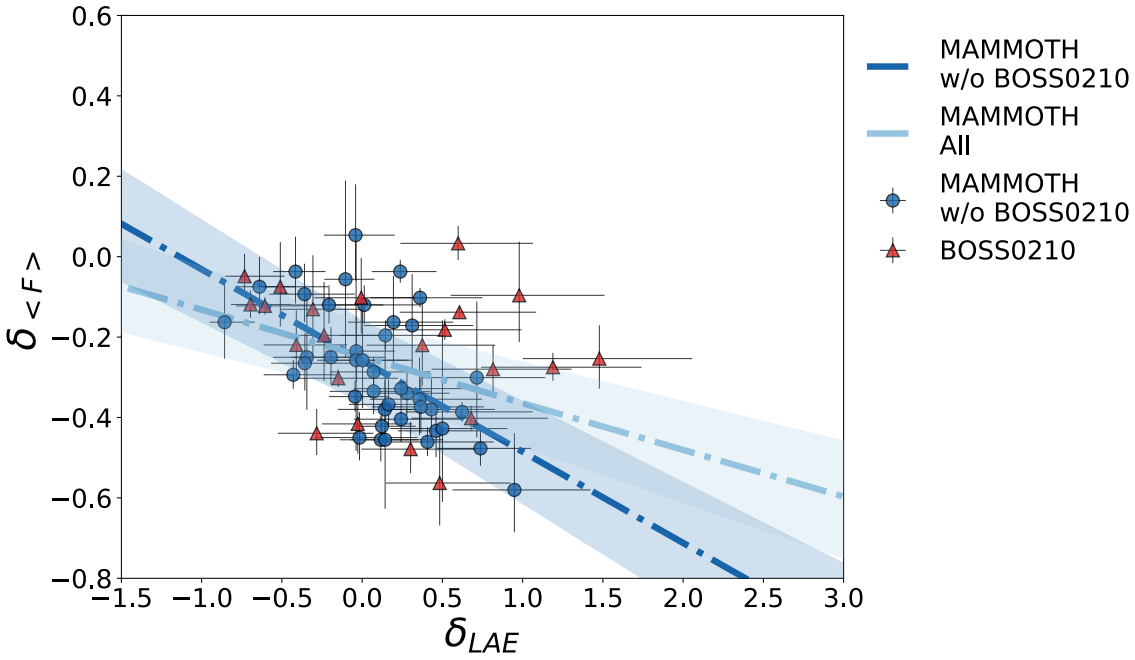

Figure 4.2: Transmission fluctuation $\delta_{\langle \mathrm{F} \rangle}$ against the LAE overdensity δ_{LAE} in NB387 fields traced by Lyα absorbers. LoSs in BOSS0210 are shown in red, while LoSs in the other three fields are colored blue. The error bars suggest the uncertainty of the measurements: The errors of δ_{LAE} originate from Poisson error and the errors of τ_{LoS} originate from pixel noise. The light and deep blue dotted dashed curves are the best-fit model for the data points, including and excluding BOSS0210. The shaded regions are the 16%−84% rank from the perturbation simulations for the case excluding BOSS0210.

4.1.2 Fields Traced by Grouping Quasars

From Section 3.3.1, we already learn that BOSS0210 is not only traced by the grouping strong Lyα absorption systems but also inhabited with a very significant quasar overdensity. Therefore, the clear difference of the slopes between the fitting curves 4.2 and 4.3 may be originated from these quasars. We try to include the grouping quasar fields from the NB400 sample to inspect the τ_{LoS}–δ_{LAE}correlation. Results are shown in Figure 4.3. Although the number of data points is much smaller (22 LoSs counted), there is a distinctly different trend from that in NB387 Lyα absorber-trace fields, which is almost flat over the entire δ_{LAE} range. We also inspect the Spearman's rank correlation for this sample, and the result is $\rho_S = 0.091^{+0.076}_{-0.116}$ with P-value = 65%, leaving no confident conclusion on the correlation. Furthermore, we perform a similar calculation for the δ_{LAE} and $\delta_{\langle F \rangle}$ linear relation for each LoSs in BOSS0240 and BOSS0755.

We perform the same fitting and error estimate for these two fields, and the fitting result is:

$$\delta_{\langle F \rangle} = -0.015^{+0.006}_{-0.006}\ \delta_{\rm LAE} - 0.141^{+0.072}_{-0.072}, \tag{4.4}$$

and the curve is shown in Figure 4.4. The results are summarized in Table 4.1. One can find that the relation in the pure quasar fields has much flatter slopes than those in NB387 samples.

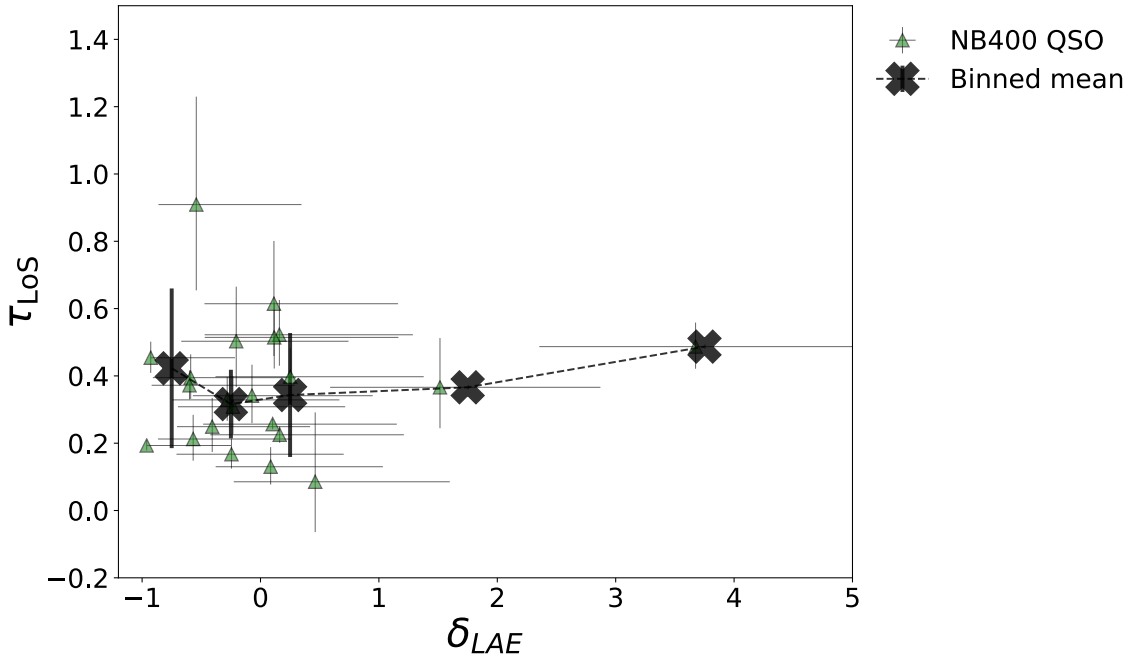

Figure 4.3: Same as the Figure 4.1, but for the samples in NB400 grouping quasar field. The green triangle are the individual LoSs in the NB400 grouping quasar field.

4.1.3 General Fields in HSC-SSP

We also include the LAEs from HSC-SSP to perform a similar analysis. Because this sample covers the total sky area of over 19 deg^2, it can be viewed as the most general one representing the field case. Firstly, we inspect again the $\tau_{\rm LoS}$–$\delta_{\rm LAE}$ correlation, and the results are plotted in Figure 4.5. The three individual fields seem to have similar distributions. Performing the Spearman's rank correlation test, results shows $\rho_S = 0.225^{+0.011}_{-0.037}$ with P-value = 8.4%, $\rho_S = 0.180^{+0.050}_{-0.134}$ with P-value = 21%, and $\rho_S = 0.278^{+0.081}_{-0.125}$ with P-value = 7.8% for DEEP2-3, E-COSMOS, and XMM-SXDS, respectively. Individual field of the general fields seems hard to give any statistically significant conclusion on the correlation, but $\rho_S = 0.241^{+0.034}_{-0.084}$ with P-value = 0.37% when all the three fields are combined, which indicates a positive

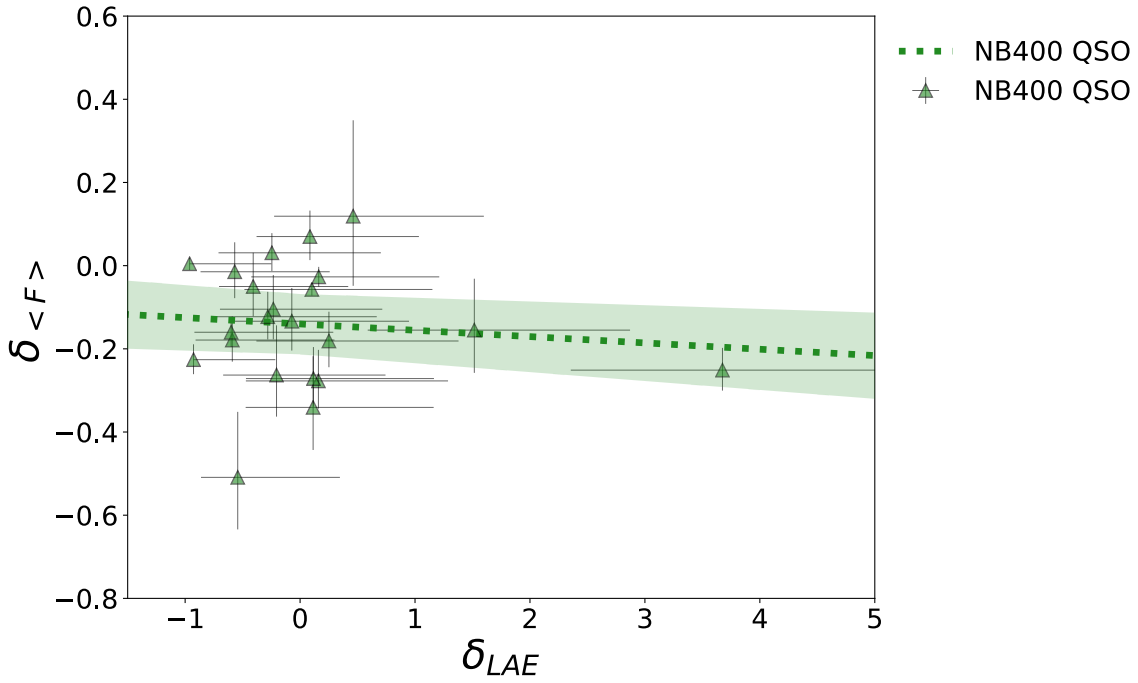

Figure 4.4: Same as the Figure 4.2, but for the samples in NB400 grouping quasar field. The green triangles, dotted curve, and transparent band are the individual LoSs, the fitting curve, and the confident band of the fitting in the NB400 grouping quasar field.

correlation between LAE and the IGM HI distributions, and the possibility that the correlation is produced by the random sample is smaller than 1%. Note, though, that the correlation coefficient ρ_S is much smaller than that of the NB387 Lyα absorber-trace field, suggesting the correlation in general fields is relatively weaker than that in HI-rich regions. The binning data show a gradual trend that the $\tau_{\rm LoS}$ enhances as the $\delta_{\rm LAE}$ increases, though at a mild pace, except for the highest $\delta_{\rm LAE}$ bin that is possibly affected by the limited LoS number.

We then perform the linear fitting of the $\delta_{\langle F\rangle}$–$\delta_{\rm LAE}$ relation for each field, i.e., DEEP2-3, E-COSMOS, and XMM-SXDS. The results of the fittings are shown in Figure 4.6, and they are:

$$\begin{aligned} {\rm DEEP2-3} \quad & \delta_{\langle F\rangle} = -0.037^{+0.011}_{-0.019}\ \delta_{\rm LAE} - 0.125^{+0.104}_{-0.115}, \\ {\rm E-COSMOS} \quad & \delta_{\langle F\rangle} = -0.037^{+0.023}_{-0.033}\ \delta_{\rm LAE} - 0.157^{+0.112}_{-0.123}, \\ {\rm XMM-SXDS} \quad & \delta_{\langle F\rangle} = -0.022^{+0.004}_{-0.005}\ \delta_{\rm LAE} - 0.129^{+0.093}_{-0.093}. \end{aligned} \tag{4.5}$$

The large scatter of the of the relation, reflected in the large uncertainty of the intercept, may partially originated from the shallow depth of the NB387 images in HSC-SSP or the more fluctuating $\tau_{\rm LoS}$ in the large general fields. Nevertheless, we

find that different fields in the HSC-SSP survey do not vary much in the slope, which demonstrates that the large survey area makes the relation less biased in cosmic variance.

Finally, we combine all the three fields in HSC-SSP and obtain the fitting:

$$\delta_{\langle F \rangle} = -0.025^{+0.006}_{-0.006}\ \delta_{\rm LAE} - 0.134^{+0.100}_{-0.101}, \tag{4.6}$$

and the slope is still flatter then that of the NB387 Lyα absorber-trace fields. The results are summarized in Table 4.1.

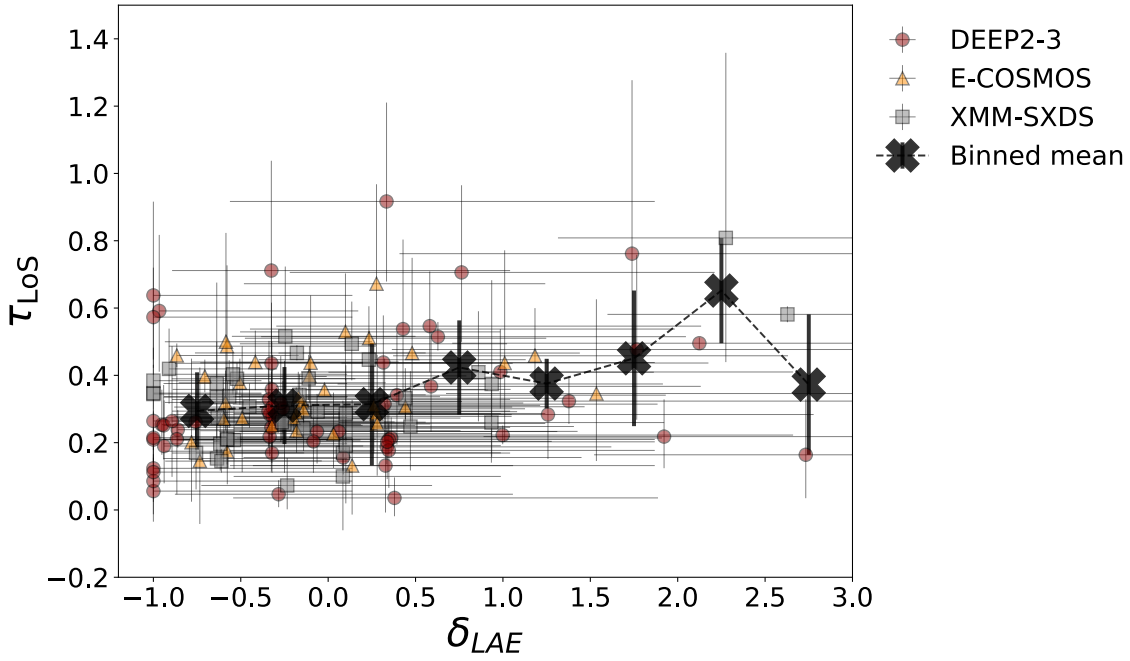

Figure 4.5: Same as the Figure 4.1, but for the samples in HSC-SSP general fields. The transparent red circles, orange triangles, and gray squares are the individual LoSs in the DEEP2-3, E-COSMOS, and XMM-SXDS field, respectively.

4.1.4 Comparisons to Models & Literature

We overplot the fitting result $\delta_{\langle F \rangle}$–$\delta_{\rm LAE}$ relation from the NB387 Lyα absorber-trace fields, NB400 grouping quasar fields, and the HSC-SSP fields together for clearer comparisons in Figure 4.7. The slope of the general fields is close to that of grouping quasar fields, but the slope of the Lyα absorber-trace fields is significantly steeper than that of either the general fields or the grouping quasar fields, regardless of whether

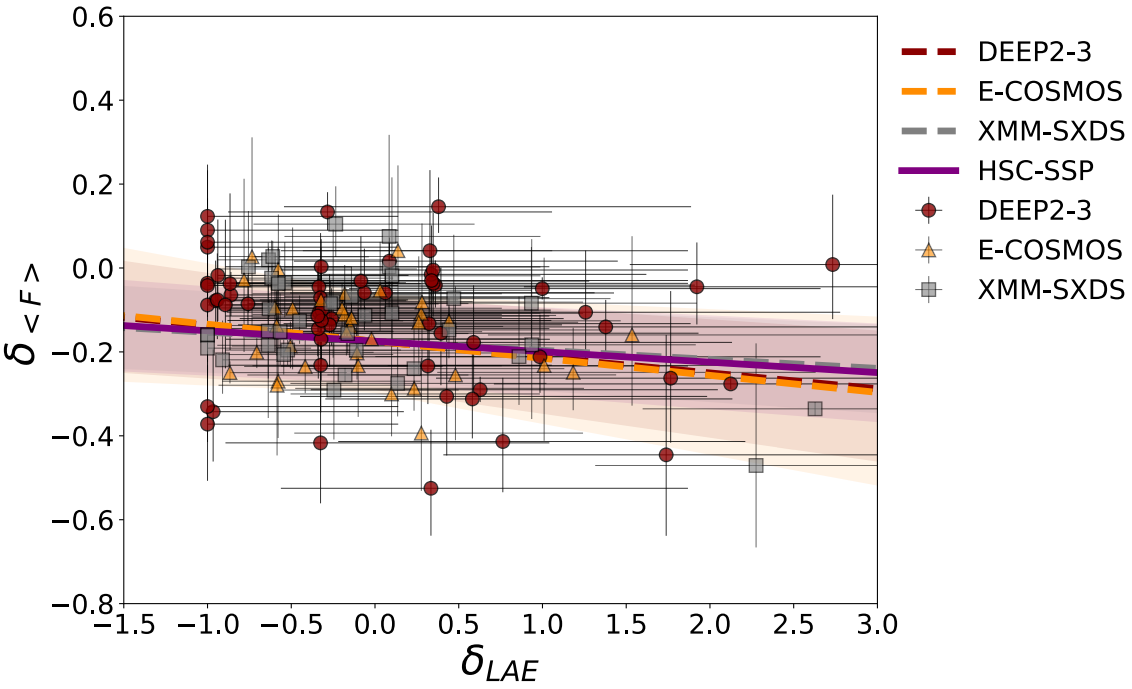

Figure 4.6: Same as the Figure 4.2, but for the samples in HSC-SSP general fields. The red, orange, and gray labels represent the DEEP2-3, E-COSMOS, and XMM-SXDS fields. The data points, dashed curves, and transparent bands are the individual LoSs, the respective fitting curves, and the confident band of the respective fittings. The solid purple curve is the fitting curve by combining the three individual fields in HSC-SSP.

BOSS0210 is included. However, the BOSS0210 field, which is not only traced by strong Lyα absorbers but also found with significant BOSS quasar overdensity, flattens the slope in the NB387 Lyα absorber-trace sample.

The results seem to suggest that in the HI rich regions, the $\delta_{\langle F\rangle}$ is much more sensitive to galaxy overdensity than in other environments. Comparing the HSC-SSP fields and the grouping quasar samples, one finds that the quasar residing environments still own a slightly flatter slope, which means the HI is even thinner in these regions. Since BOSS0210 can also relax the $\delta_{\langle F\rangle}$–$\delta_{\rm LAE}$ relation from Lyα absorber-trace sample, a potential suggestion is that the emergence of the grouping quasars can have an influential effect on regulating the correlation between galaxies and the intergalactic HI, by means of suppressing the HI density around galaxy overdensity.

We further inspect the possible origin of variations by comparing the fitting of $\delta_{\langle F\rangle}$–$\delta_{\rm LAE}$ relation with models and literature.

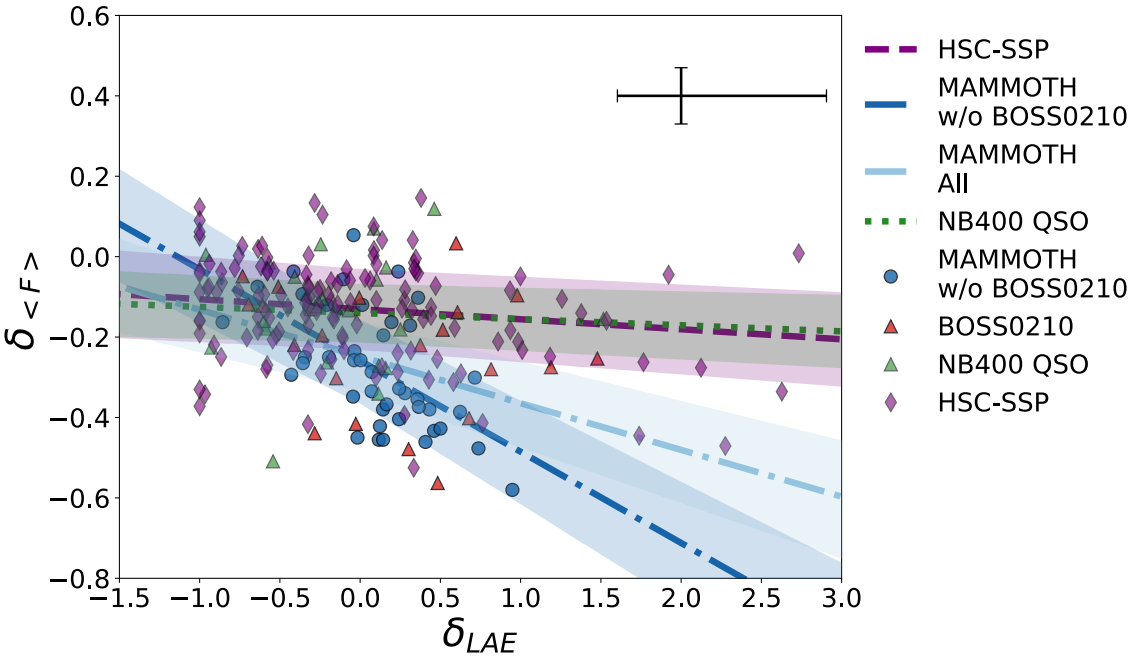

Figure 4.7: Comparison of the $\delta_{\langle \mathrm{F} \rangle}$–$\delta_{\mathrm{LAE}}$ among the three samples: NB387 Lyα absorber-trace fields, NB400 grouping quasar fields, and the HSC-SSP general fields. All labels have the same meaning as in Figure 4.2, 4.4, and 4.6, except for the hidden individual errorbars. Instead, the typical error bar is shown in the upper right corner.

Mukae et al. (2017) have measured the similar $\delta_{\langle F \rangle}$ but for the more massive Lyman-break galaxies spanning in a larger redshift interval at $2 < z < 3$, and we overplot their result with ours in the Figure 4.8:

$$\delta_{\langle F \rangle} = -0.14^{+0.06}_{-0.16}\ \delta_{\mathrm{LAE}} - 0.17^{+0.06}_{-0.06}, \tag{4.7}$$

The intercepts are normalized to match that in general fields for clearer comparisons of the slopes. They estimate $\delta_{\langle F \rangle}$ at the position of the highest $\mathrm{S/N}_{\langle F \rangle}$, defined as the ratio between Lyα absorption and its error, on a $\sim$ 100cMpc scale within the redshift $2 < z < 3$, while we target the absorption spike based on the τ_{LoS} on $\sim$ 20 cMpc scale within $2.15 < z < 2.20$ or $2.25 < z < 2.33$. Our NB387 Lyα absorber-trace fields give consistent slopes with the photo-z galaxies within their uncertainty, although the case for BOSS0210 has a flatter slope. Both their and Lyα absorber-trace samples' fittings are inconsistent with those of HSC-SSP fields and show steeper slopes compared to the general fields, but the case for the NB387 Lyα absorber-trace fields excluding BOSS0210 shows a more significant discrepancy.

We also derive the results of $\delta_{\langle F \rangle}$–$\delta_{\mathrm{LAE}}$ relation with similar procedures of mea-

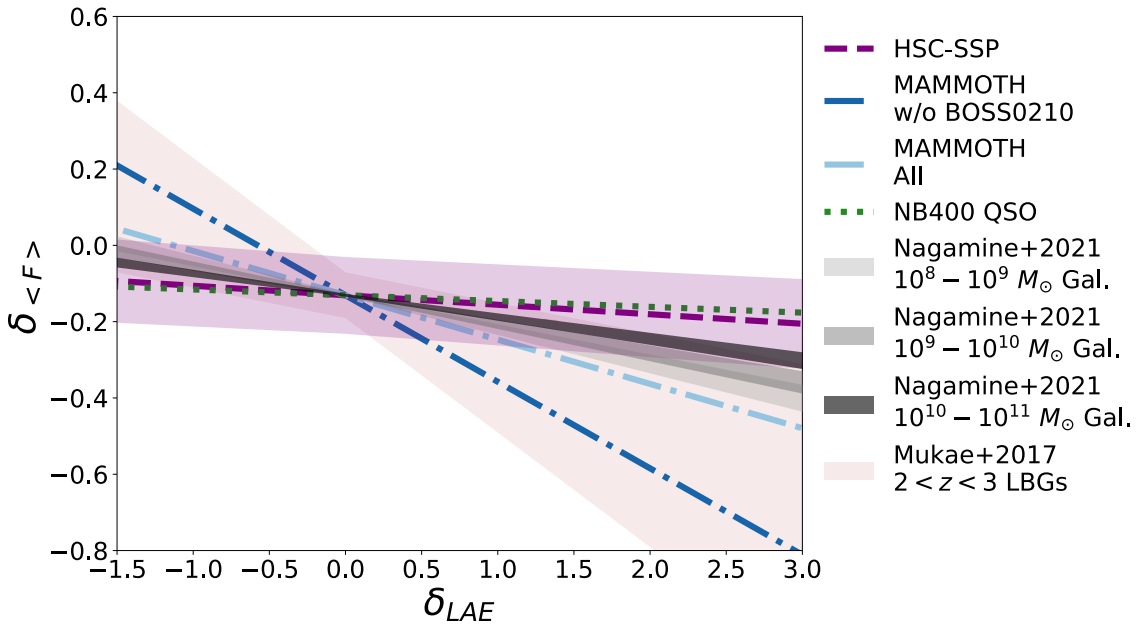

Figure 4.8: Comparison of the $\delta_{\langle F\rangle}$–$\delta_{\rm LAE}$ between the LAE samples and the Photo-z at $2 < z < 3$ (Mukae et al. 2017) and the Osaka models (Shimizu et al. 2019; Nagamine et al. 2021). All intercepts are normalized to match HSC-SSP general fields for a better comparison of slopes. The light, medium, and deep gray shaded regions indicate the fitting of the predictions within uncertainties from the GADGET3-Osaka model (Shimizu et al. 2019; Nagamine et al. 2021) for galaxies with 10^8–10^9 $M_\odot$, 10^9–10^{10} $M_\odot$ and 10^{10}–10^{11} $M_\odot$. The brown shaded region indicates the fitting of Mukae et al. (2017) results within their uncertainty. Other labels are the same as Figure 4.7.

surements in the GADGET3-Osaka cosmological hydrodynamic simulation, which is based on the smoothed particle hydrodynamics (SPH) simulation code GADGET-3 (Springel 2005) and takes full account of star formation and supernova feedback (Shimizu et al. 2019). More details on the simulation data processing are presented in Momose et al. (2021b) as well as Nagamine et al. (2021), and we denote the simulation as the *Osaka model*. The curves fit from the models for galaxies with different M_* are:

$$\begin{aligned}
10^8 - 10^9\ M_\odot \quad & \delta_{\langle F\rangle} = -0.0900^{+0.0105}_{-0.0105}\ \delta_{\rm LAE} - 0.0984^{-0.0036}_{-0.0036}, \\
10^9 - 10^{10}\ M_\odot \quad & \delta_{\langle F\rangle} = -0.0761^{+0.0086}_{-0.0086}\ \delta_{\rm LAE} - 0.0984^{+0.0036}_{-0.0036}, \\
10^{10} - 10^{11}\ M_\odot \quad & \delta_{\langle F\rangle} = -0.0571^{+0.0061}_{-0.0061}\ \delta_{\rm LAE} - 0.0984^{+0.0036}_{-0.0036},
\end{aligned} \tag{4.8}$$

as plotted with the intercept normalized in Figure 4.8.

In summary, we find the general fields in HSC-SSP and the grouping quasar fields

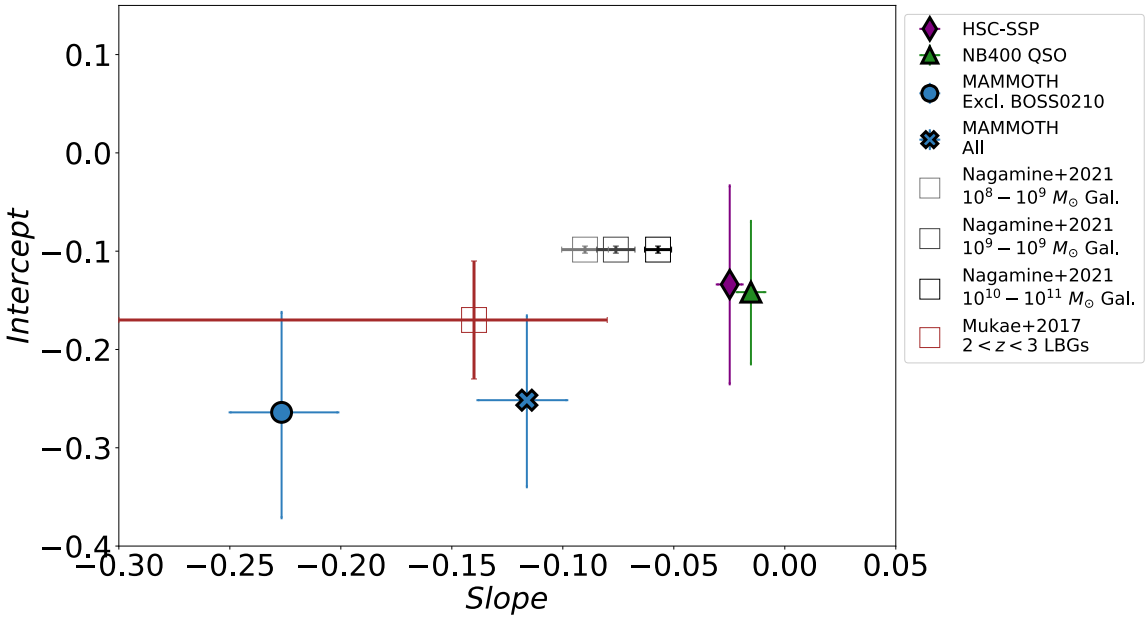

Figure 4.9: Comparison of the fitting parameters in the $\delta_{\langle F \rangle}$–$\delta_{\rm LAE}$ relation among the three samples: NB387 Lyα absorber-trace fields, NB400 grouping quasar and fields, and the HSC-SSP general fields. The blue cross and circle are the samples in NB387 Lyα absorber-trace fields with and without the BOSS0210 field. The green triangle and purple diamond represent samples in NB400 grouping quasar fields and the HCS-SSP entire sample, respectively. The result from Mukae et al. (2017) using the photo-z galaxies & 16 BOSS LoSs is overplotted as the brown open square. The predictions from the GADGET3-Osaka model (Shimizu et al. 2019; Nagamine et al. 2021) for galaxies with 10^8–10^9 $M_\odot$, 10^9–10^{10} $M_\odot$ and 10^{10}–10^{11} $M_\odot$ are shown in the light, medium, and deep gray open squares, respectively.

agree more with the Osaka model's prediction. Meanwhile, it shows a distinct discrepancy between the simulation and the IGM HI-rich fields traced by Lyα absorbers, but the offset will be reduced when the BOSS0210 field containing a significant grouping of quasars is included. These comparisons can also be confirmed in the parameter space, as shown in Figure 4.9.

The LBG result from Mukae et al. (2017) also shows an offset compared to the model, although not quite significant due to their considerable uncertainty. We note that there is an M_*-dependence in the Osaka model, and the less massive galaxy population shows a steeper trend. However, such dependence is not sufficiently sensitive to be discriminated beyond the uncertainties of observational measurements. This likely disproves that the reason for the offsets among samples originated from different galaxy stellar masses.

Another possible cause for the discrepancy between Lyα absorber-trace fields and other samples may be HI suppressing 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 HI-rich environments, and thus the completeness of detection of LAEs may have been lower, resulting in the underestimated $\delta_{\rm LAE}$ at the given $\delta_{\langle F \rangle}$. In this case, the LAEs in BOSS0210, which contains a group of quasars that possibly ionize the surrounding IGM HI as suggested by the abnormal LoSs with relatively weak Lyα absorption at the high $\delta_{\rm LAE}$ regions (see Figure 4.2), compensates the suppression and becomes a more typical environment similar to general fields without high density HI. This hypothesis gets support from the NB400 grouping quasar fields that have a slope closer to general fields – because no dense HI can be ionized to make a big difference. Moreover, the Osaka model prediction favors this interpretation, as the fit with BOSS0210 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).

The similar argument is also motivated in literature. The plateau in the CCF at $r \lesssim 0.6$ pMpc is found in Momose et al. (2021b), which also suggests the deficit of LAEs on a small scale. A recent δ_F–$\delta_{\rm gal}$ based on the LATIS 3D IGM HI tomography also shows the deficient number of UV-selected galaxies at the extremely strong IGM absorption peaks (Newman et al. 2022, see also in Section 1.2.2).

Also, we can easily find the intercept difference among different samples and models. Naturally, the general fields and quasar fields have a relatively higher intercept, which means a statistically weaker IGM Lyα absorber signal compared to the samples in NB387 Lyα absorber-trace fields because these fields are not intentionally searched with IGM priority (also see the $\tau_{\rm LoS}$ distribution in each sample in Figure 2.3). Their intercepts are consistent with that of the LBGs in COSMOS and the Osaka model. However, we note that the intercept comparison is meaningless, as the measured amplitudes can depend on the sample bias, as explained above, and the $\delta_{\langle F \rangle}$ among different studies. Instead, we will primarily focus on the slope of the relation.

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 field variance. We hope to determine the valid reason for the slope discrepancy in future studies, with more HSC fields targeting various environments. Follow-up studies to search for 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.

Sample [1]	ρ_S [2]	P-value [3]	Slope [4]	Intercept [5]
MAMMOTH w/o BOSS0210	$0.541^{+0.037}_{-0.051}$	$< 0.01\%$	$-0.227^{+0.026}_{-0.023}$	$-0.258^{+0.096}_{-0.114}$
MAMMOTH All	$0.384^{+0.015}_{-0.038}$	0.09%	$-0.116^{+0.018}_{-0.022}$	$-0.248^{+0.082}_{-0.093}$
NB400 QSO	$0.091^{+0.076}_{-0.116}$	65%	$-0.015^{+0.006}_{-0.006}$	$-0.141^{+0.072}_{-0.072}$
HSC-SSP	$0.241^{+0.034}_{-0.084}$	0.37%	$-0.025^{+0.006}_{-0.006}$	$-0.134^{+0.100}_{-0.101}$
DEEP2-3	$0.225^{+0.011}_{-0.037}$	8.4%	$-0.037^{+0.011}_{-0.019}$	$-0.125^{+0.104}_{-0.115}$
E-COSMOS	$0.180^{+0.050}_{-0.134}$	21%	$-0.037^{+0.023}_{-0.033}$	$-0.157^{+0.112}_{-0.123}$
XMM-SXDS	$0.278^{+0.081}_{-0.125}$	7.8%	$-0.022^{+0.004}_{-0.005}$	$-0.129^{+0.093}_{-0.093}$

Table 4.1:: The parameters of the Spearman's rank correlation test and the fitting of the linear δ_F–$\delta_{\rm LAE}$ relation.
[1] The sample category;
[2] The correlation coefficient in Spearman's rank correlation test;
[3] The P-value in Spearman's rank correlation test;
[4] The slope of the linear δ_F–$\delta_{\rm LAE}$ relation fitting;
[5] The intercept of the linear δ_F–$\delta_{\rm LAE}$ relation fitting.

4.2 Two-point Cross-correlation Analysis

4.2.1 Cross-correlation function

Along with the analysis based on the local overdensity of LoSs, a more general analysis can be performed for the galaxy–IGM HI 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 use the angular CCF $\omega(\theta)$ or projected CCF $\omega(r)$ if the angular separation is translated into the 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:

$$\omega(\theta) = \frac{D_{\rm LAE}D_{\rm LoS}(\theta) - D_{\rm LAE}R(\theta) - D_{\rm LoS}R(\theta) + RR(\theta)}{RR(\theta)}, \tag{4.9}$$

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 0.013 deg, i.e., ~ 0.4 proper-Mpc (pMpc) at $z = 2.2$ and $z = 2.3$, so that there are > 10 pairs at the bin in one subsample, reaching S/N > 3 in Poisson statistics. Ten bins are set for the calculation extending up to 0.6 deg 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 the accurate LAE redshifts are unknown within $2.15 < z < 2.20$ or $2.25 < z < 2.33$. Therefore, we use the projected CCF but not the 3D CCF.

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

$$\sigma_i = \frac{N_{\rm sub} - 1}{N_{\rm sub}} \sum_{k=1}^{N_{\rm sub}} \left(\omega_{i,k} - \overline{\omega_i}\right)^2 , \tag{4.10}$$

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 $i^{\rm th}$ bin.

4.2.2 First Inspection with NB387 Fields Traced by Lyα Absorbers

We divide the LoSs into two subsamples based on $\tau_{\rm LoS}$. To ensure comparable numbers of LoSs in the two subsamples, we set the criterion as $\tau_{\rm LoS} = 0.5$, which is approximate to the median of $\tau_{\rm LoS} = 0.47$ in the sample. 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 entire sample, the numbers of LoSs for high and low $\tau_{\rm LoS}$ subsample are 30 and 34, respectively, and if we exclude field BOSS0210, these numbers change to 23 and 19, respectively. In the first inspection, we mainly work with the inclusion of the BOSS0210 LoSs for CCFs.

As described above, the projected CCF does not rely on 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 $z \sim 2.2$ or $z \sim 2.3$. 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, which may involve an artificial signal in 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 4.10. Although the full sample has a much weaker signal than any subsamples, they do not equal precisely zero. For clearer comparison, we subtract the amplitude of the full sample CCF $\omega(r)_{\rm full}$ (or all sample $\omega(r)_{\rm all}$) 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 4.10. For clarity, the data of subsamples at each bin are slightly shifted along the r-axis in the figures.

From both panels in Figure 4.10, 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. The intrinsic signal may be even more vital as there can be 15% randomly distributed low-z interlopers in our LAE sample. 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 $\tau_{\rm LoS}$ higher than the cosmic mean value. The two bins at $\sim 0.8 - 1.0$ pMpc also tend to be consistent with zero, suggesting a weak signal at this distance.

It should be noted that the CCF shown in the right panel of Figure 4.10 can be well fitted by a power law:

$$\omega(r) = \pm \left(\frac{r}{r_0} \right)^{-\gamma}, \tag{4.11}$$

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 4.10. Noted that the data points whose CCF of the total sample suggests do not equivalent to zero means that the parent is biased to the low- or high $\tau_{\rm LoS}$, which will cause artificial signals. Thus we only use the data points with CCF of the total sample consistent with zero within their errors to perform the model fitting. The best-fit parameters (γ, r_0) with the errors estimated from the 10,000 Monte Carlo perturbed simulations, similar to Section 4.1.1, 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.

The r_0 for both subsamples are of the order of 0.1 pMpc, much smaller than the typical galaxy–galaxy clustering strength of the order of several pMpc. This indicates that the strength of the LAE−IGM HI correlation is much weaker than galaxy clustering. However, it is still sufficiently significant to be detected based on our samples in the projected CCF. We note that Momose et al. (2021b) obtained

somewhat stronger 3D CCF signals between LAEs and CLAMATO HI 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 BOSS0210 or 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.

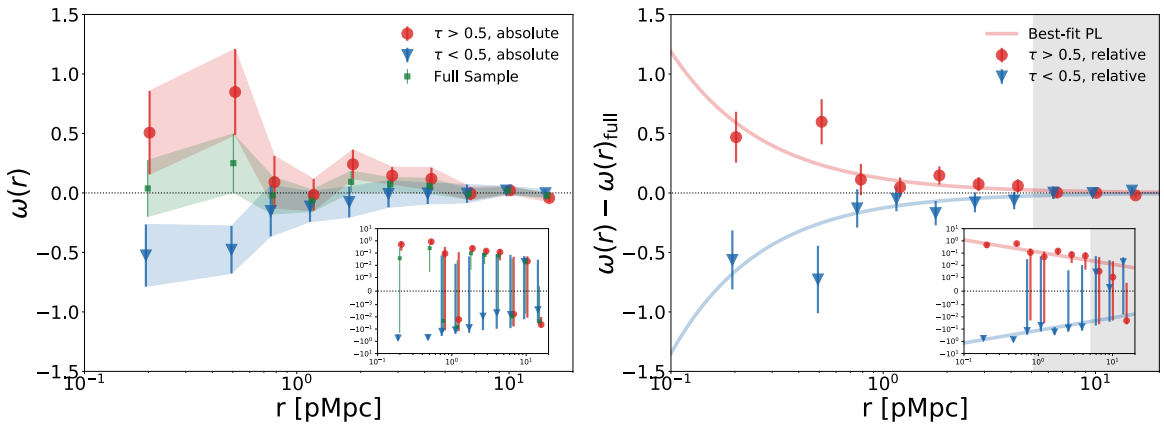

Figure 4.10: Cross-correlation function (CCF) between LAEs and LoSs for the high $\tau_{\rm LoS}$/low $\tau_{\rm LoS}$ subsamples in NB387 Lyα absorber-trace field. The red points and curves are the $\tau_{\rm LoS} > 0.5$ subsample and the corresponding fit power-law model, while the blue points and curves are for the $\tau_{\rm LoS} < 0.5$ subsample. For clarification, data points for different subsamples in each bin are shifted slightly along the r-axis. 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 data points. The gray-shaded region indicates the separation limit where the signal becomes noisy.

4.2.3 Comparisons Among Different Fields

From the Section 4.2.2, we learn that the CCFs of high-$\tau_{\rm LoS}$ LoSs–LAEs and low-$\tau_{\rm LoS}$ LoSs–LAEs likely follow the power laws, which show excellent symmetry between

the two subsamples, but one should also be concerned with the bias from the limited sample size that can be tested in the all LoS sample. Meanwhile, it is not intuitive to interpret the physical meaning of the parameters in the relative CCF $\omega(r) - \omega(r)_{\rm all}$. We thus try to optimize the power-law fitting of the absolute CCF $\omega(r)$ in this section by excluding the data points with biased $\omega(r)_{\rm all}$, i.e., $\omega(r)_{\rm all} > err\,[\omega(r)_{\rm all}]$, where $err\,[\omega(r)_{\rm all}]$ is the error estimated for $\omega(r)_{\rm all}$. Also, to overcome the difficulty in fitting the noisy data points, we exclude the negative (positive) values in the CCFs of high- (low-) $\tau_{\rm LoS}$ and fix the indifferent parameter $\gamma = 1.0$, as suggested by the first inspection in Section 4.2.2. Furthermore, we expand the CCF analysis from NB387 Lyα absorber-trace sample to the NB400 grouping quasar sample and the HSC-SSP general fields as the control sample to investigate if the galaxy–IGM HI correlation can be dependent on fields, just as what the slopes in the $\delta_{\rm LAE}$–$\delta_{\langle F \rangle}$ relation shows.

As to the sample in NB400 grouping quasar fields, the calculation procedure is identical to that for the sample in NB387 Lyα absorver-trace field because both regions focus on single-pointing in each surveyed field, and the photometric fluctuation is typical < 0.1 magnitude in NB400 and < 0.2 magnitude in NB387 images, as stated in Section 2.2.4. However, the situation for HSC-SSP data is different, with ineligible fluctuations among different pointings. Conservatively, we do not only limit the sample by the NB387 magnitude cut at NB < 24.2 but also limit the completeness estimated in Section 3.2.2 as $f_{\rm tot,comp} > 0.8$. As Figure 3.6 suggests, this treatment gives us the most homogeneous sample distribution. At the same time, there is a significant fraction of the sky area sufficient for CCF calculations. We then separate the samples into high-$\tau_{\rm LoS}$ and low-$\tau_{\rm LoS}$ subsamples according to the median of the $\tau_{\rm LoS}$ in each field, i.e., $\tau_{\rm LoS,\ median} = 0.47$, 0.35, and 0.30 in NB387, NB400, and HSC-SSP fields respectively (see Figure 2.3).

Results of the individual samples are shown in Figure 4.11 for the high-$\tau_{\rm LoS}$ sample and in Figure 4.12. The data points in the lower panel of each grid represent the CCF from all LoSs $\omega_{\rm all}(r)$ at the bin, and only those $\omega_{\rm all}(r) \approx 0$ are meaningful as the unbiased data. First, after checking the upper right panel of both figures, the data points which are the same as the left panel in Figure 4.10, one can find that the updated fitting methods does not change the fitting result significantly. Although the updated $r_0 = 0.107^{+0.103}_{-0.097}$ pMpc ($0.106^{+0.079}_{-0.082}$ pMpc) with relatively large error range is for $\omega(r)$ of high- (low-) $\tau_{\rm LoS}$ sample, while the $r_0 = 0.12^{+0.05}_{-0.03}$ pMpc ($0.13^{+0.06}_{-0.02}$ pMpc) fit in 4.2.2 is fit for the $\omega(r) - \omega(r)_{all}$ of high- (low-) $\tau_{\rm LoS}$ sample, they precisely agree to each other within the uncertainty. The upper left panel also shows the Lyα absorber-trace fields, but BOSS0210 is excluded for a sample only in the pure Lyα absorber-trace regions, and the CCF does not alter significantly except for the two noisy bins at around $r \approx 1$ pMpc. The $r_0 = 0.136^{+0.138}_{-0.136}$ pMpc ($0.091^{+0.100}_{-0.091}$

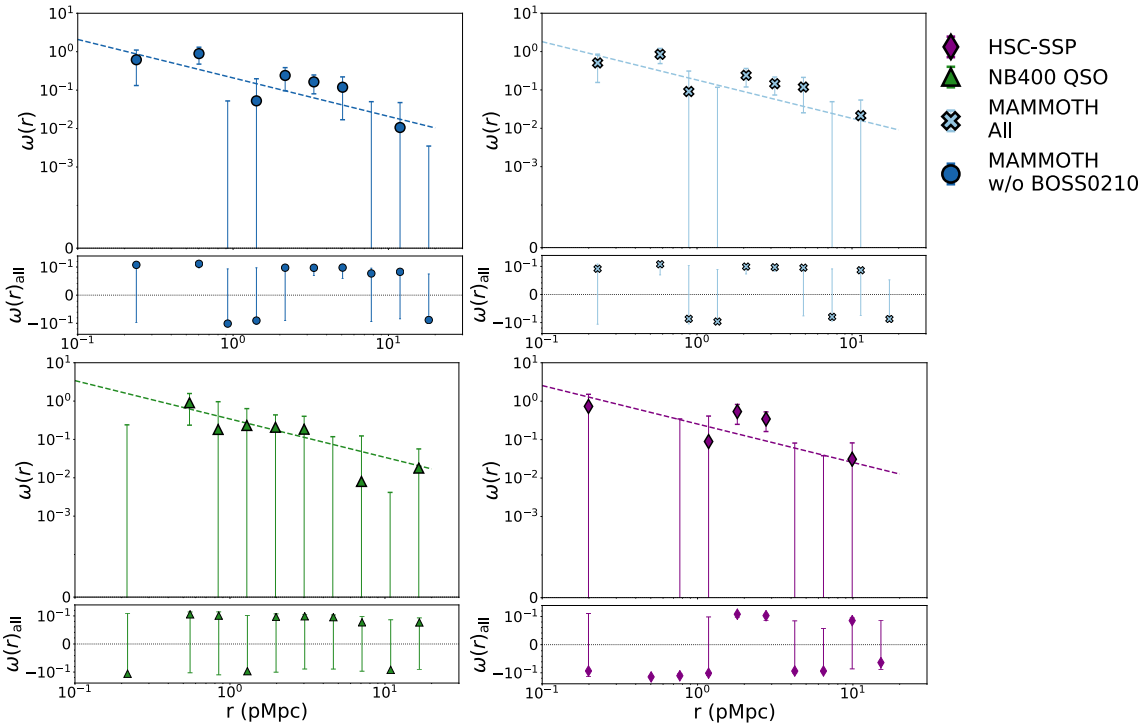

Figure 4.11: CCFs between LAEs and LoSs for the high-$\tau_{\rm LoS}$ subsample with $\tau_{\rm LoS} > 0.47, 0.38$ and 0.3 in the NB387 Lyα absorber-trace field, NB400 grouping quasar field, and HSC-SSP general fields. The upper panel of each grid shows the absolute CCF, $\omega(r)_{\rm h}$, of the high-$\tau_{\rm LoS}$ LoSs, while the lower panel is for the absolute CCF, $\omega(r)_{\rm all}$, of the full LoS samples to demonstrate the unbiased data points. The deep blue circles, light blue crosses, green triangles, and purple diamonds are the data points in the NB387 Lyα absorber-trace field, NB400 grouping quasar field, and HSC-SSP general fields, respectively. The dashed curves are the best-fit power-law model, and the colors correspond to the meaning of labels.

pMpc) is slightly larger (smaller) than the case with BOSS0210 in the high- (low-) $\tau_{\rm LoS}$ sample, but they are still comparable under the uncertainty, which approves that data points from the individual field may not affect the CCF statistics critically. The lower left panel is the sample in grouping quasar fields, and data points are generally noisy except for the second innermost bin at $r \approx 0.6$ pMpc in the high-$\tau_{\rm LoS}$ sample, which may be due to the small number (22) of LoSs in this category. Regardless of the large error bar, the $r_0 = 0.332^{+0.457}_{-0.332}$ pMpc is the largest among the high-$\tau_{\rm LoS}$ samples, although the difference is not beyond the uncertainty. We also note that this is the only high-$\tau_{\rm LoS}$ sample that shows the negative signal in the innermost bin at $r \approx 0.2$ pMpc, which seems to be a hint of deficient IGM HI around LAEs in the grouping quasar fields, somehow in agreement with the suggestion from $\delta_{\langle F \rangle}$–$\delta_{\rm LAE}$ relation in

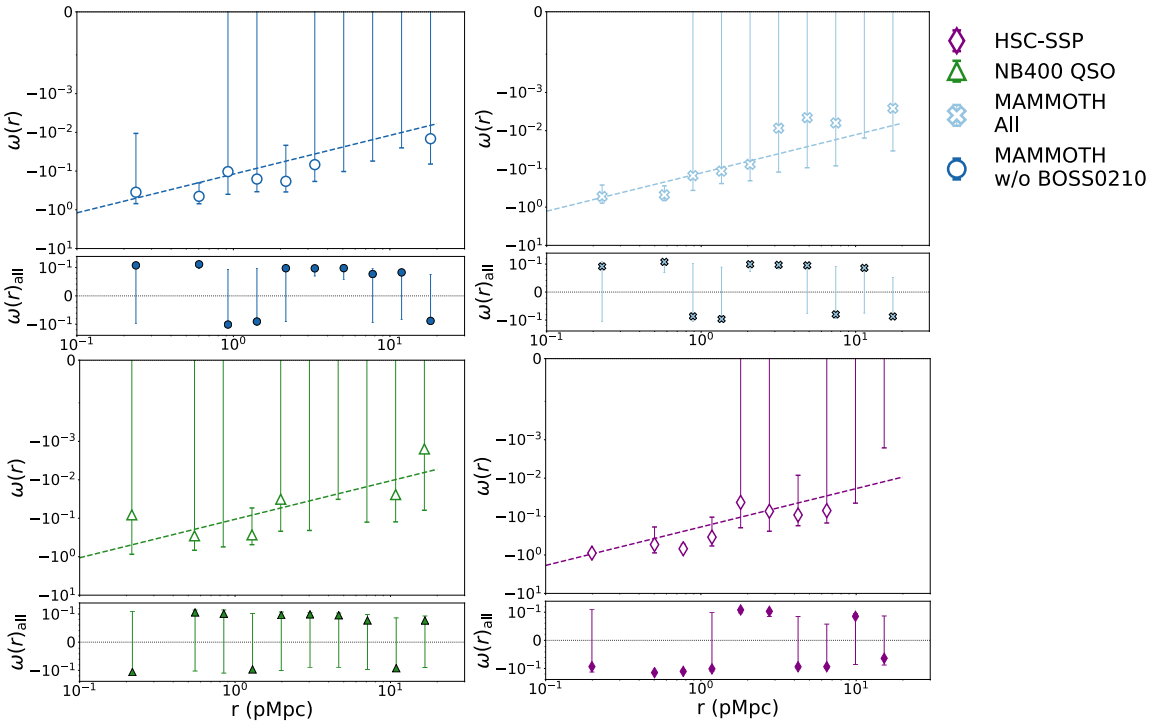

Figure 4.12: CCFs between LAEs and LoSs for the low-$\tau_{\rm LoS}$ subsample with $\tau_{\rm LoS} < 0.47, 0.38$ and 0.3 in the NB387 Lyα absorber-trace field, NB400 grouping quasar field, and HSC-SSP general fields. The upper panel of each grid shows the absolute CCF, $\omega(r)_{\rm h}$, of the low-$\tau_{\rm LoS}$ LoSs, while the lower panel is for the absolute CCF, $\omega(r)_{\rm all}$, of the full LoS samples to demonstrate the unbiased data points. The open deep blue circles, open light blue crosses, open green triangle, and open purple diamonds are the data points in the NB387 Lyα absorber-trace field, NB400 grouping quasar field, and HSC-SSP general fields, respectively. The dashed curves are the best-fit power-law model, and the colors correspond to the meaning of labels.

Section 4.1. The lower right panel is the sample in HSC-SSP general fields, and the fit $r_0 = 0.151^{+0.202}_{-0.151}$ pMpc ($0.179^{+0.020}_{-0.020}$ pMpc) is also comparable to the other samples. All the fitting results are summarized in Table 4.2.

Furthermore, we can look into the different fields of HSC-SSP to see whether these *general fields* vary in the CCF statistics. Figure 4.13 shows the results of the DEEP2-3, E-COSMOS, and XMM-SXDS. The measured values of the data points in each individual field, especially those distinct from zero beyond the errors, seem to give very consistent CCF signals that roughly span around a power law. In the innermost bin at $r = 200$ pkpc, the three regions are roughly consistent around $\omega(r) \approx 1$ for the high-$\tau_{\rm LoS}$ samples, although the DEEP2-3 and E-COSMOS/XMM-SXDS can be different on the order of a magnitude, and the DEEP2-3 has a stronger

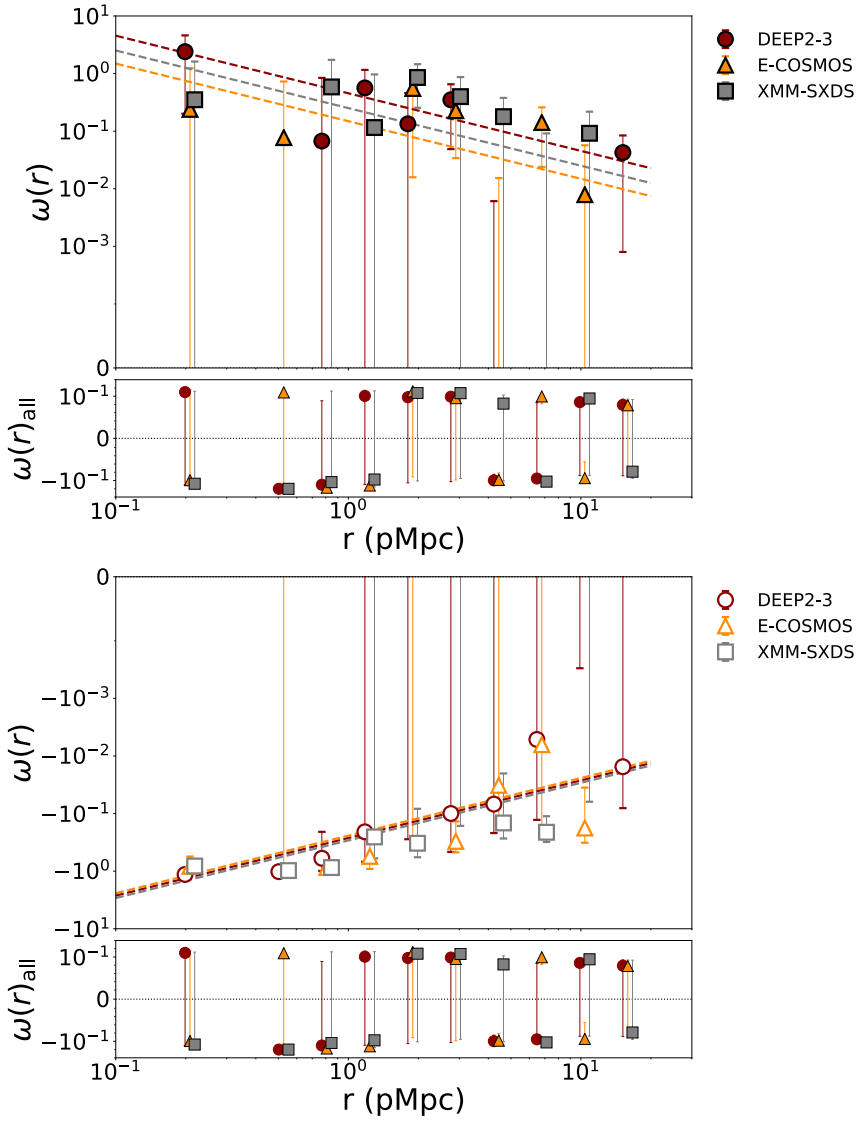

Figure 4.13: CCFs between LAEs and LoSs for the high-$\tau_{\rm LoS}$ and low-$\tau_{\rm LoS}$ subsample with respective to the median $\tau_{\rm eff}^{15h^{-1}{\rm cMpc}}$ of each sample in the HSC-SSP general fields. The brown circles, orange triangles, and gray squares are the points in DEEP2-3, E-COSMOS, and XMM-SXDS fields, respectively. The open labels in the lower panel share the same colors and markers for each sample in low-$\tau_{\rm LoS}$ subsamples. The large figure shows the CCF, $\omega(r)_{\rm h}$ in either high-$\tau_{\rm LoS}$ or low-$\tau_{\rm LoS}$ subsample, while the small figure is for the absolute CCF, $\omega(r)_{\rm all}$, of the full LoS samples to demonstrate the unbiased data points. The curves with corresponding colors are the best fit power-law model for the four categories of the samples.

signal. Such a small-scale signal reflects the HI information in the circumgalactic medium. Thus it is not surprising to have a significant signal. Interestingly, the

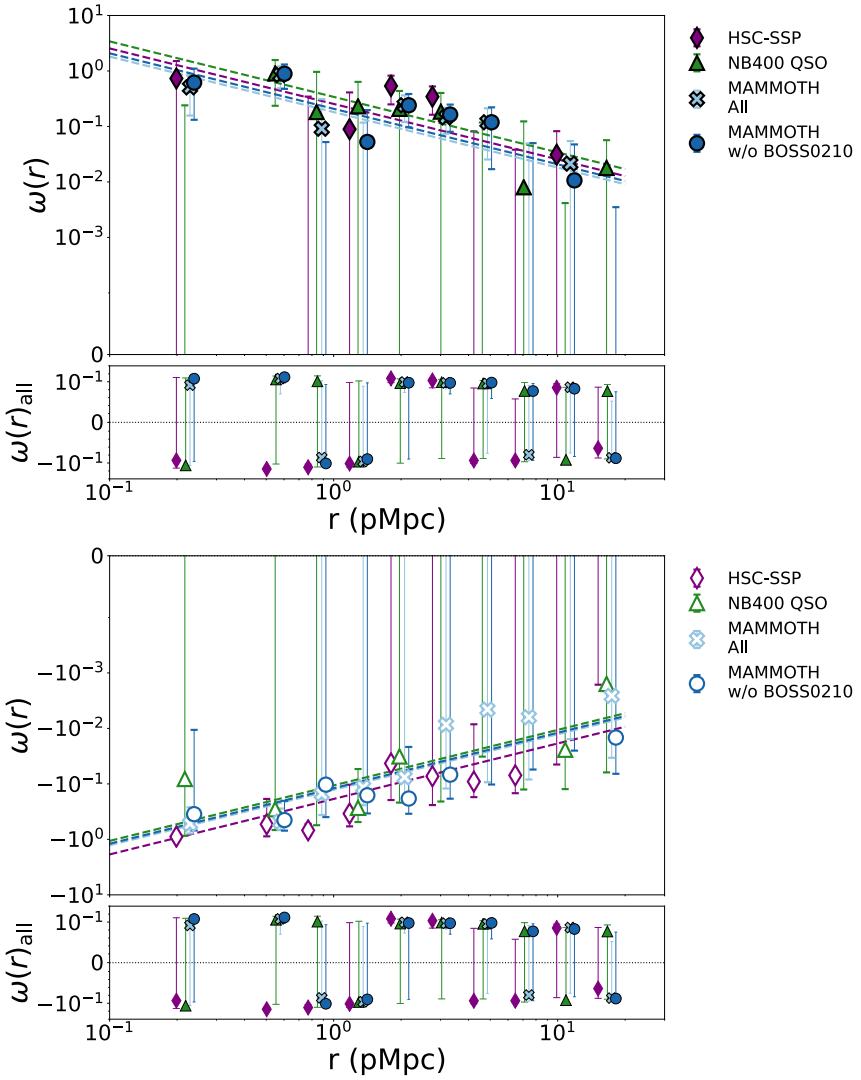

Figure 4.14: CCFs between LAEs and LoSs for the high-$\tau_{\rm LoS}$ and low-$\tau_{\rm LoS}$ subsample with respective to the median $\tau_{\rm eff}^{15h^{-1}{\rm cMpc}}$ of each sample in the NB387 Lyα absorber-trace field, the NB400 grouping quasar fields, and the HSC-SSP general fields. The light blue crosses and deep blue circles are the samples in Lyα absorber-trace fields with and without LoSs in BOSS0210. The green triangle is the sample in NB400 grouping quasar fields, while the purple diamonds represent the entire HSC-SSP sample. The open labels in the lower panel share the same colors and markers for each sample in low-$\tau_{\rm LoS}$ subsamples. The large figure shows the CCF, $\omega(r)_{\rm h}$ in either high-$\tau_{\rm LoS}$ or low-$\tau_{\rm LoS}$ subsample, while the small figure is for the absolute CCF, $\omega(r)_{\rm all}$, of the full LoS samples to demonstrate the unbiased data points. The curves with corresponding colors are the best fit power-law model for the four categories of the samples.

low-$\tau_{\rm LoS}$ samples have better consistency among fields, suggesting regions deficient in IGM have smaller variations. When checking the power-law fitting among the fields, we also find that they roughly share comparable r_0 within uncertainty in the high-$\tau_{\rm LoS}$ samples, while the differences are more negligible in the low-$\tau_{\rm LoS}$ samples, as shown in the Table 4.2.

Finally, we can compare the CCFs among all the samples shown in Figure 4.11 and 4.12: the NB387 Lyα absorber-trace fields without or with BOSS0210, the NB400 grouping quasar fields, and the HSC-SSP general fields. The results are shown in the figure 4.14. All the data span closely in a band, demonstrating that the high-$\tau_{\rm LoS}$ LoS–LAE relative CCF signal is consistent with a power law (or an inverse power law for the low-$\tau_{\rm LoS}$ samples), regardless of the environment. This may suggest that even though the high-density HI may reflect the exceeding accretion of the HI, or the grouping quasar may cause vital feedback, the underlying halo is still dominant in regulating the galaxy–IGM HI relation. Moreover, there is only a slight variation in the CCF when looking into the fitted power law.

We can directly compare the fit parameters r_0 of the power-law models in different samples in Figure 4.15. The violin plot shows the 10,000 Monte Carlo simulations of the fitting performed for estimating errors, i.e., by fluctuating the measured data points with their 1σ uncertainties and with the assumption of Gaussian distribution in each measurement. The hatched violins at the left-hand side of each panel show the distributions of r_0 simulations in the samples of the Lyα absorber-trace fields without or with BOSS0210, the quasar grouping fields, and the HSC-SSP general fields, and the black vertical bars suggest the percentile range in 16%–84% with the open circles indicating the median values. In the high-$\tau_{\rm LoS}$ samples, one easily finds that the r_0 in all four categories are roughly consistent with each other within the errors, especially those in the Lyα absorber-trace fields and the general fields, regardless of whether the BOSS0210 is accounted or not. Although there is a significant offset of the median value in the grouping quasar fields, this offset seems insignificant compared to the fluctuation in individual fields in HSC-SSP, as shown on the right-hand side of the violin plot. Note that the r_0 fluctuation among different samples is even smaller in the low-$\tau_{\rm LoS}$ samples. Especially, the $r_0 = 0.179^{+0.020}_{-0.020}$ can be well constrained in the general fields with an uncertainty of only 11%. By comparing the r_0 variation between high-$\tau_{\rm LoS}$ and low-$\tau_{\rm LoS}$ samples, we may witness how the high-density environments affect the galaxy-IGM HI correlation, even though the r_0 consistency persists.

In conclusion, the fact that CCF signals span around the band of a power-law in all these samples and that the fit of r_0 does not critically depend on the environments likely suggests the gravitational potential of underlying halos dominates the correlation between LAEs and IGM HI. Combining the results from the $\delta_{\rm LAE}$-

$\delta_{\langle F \rangle}$ and the CCFs, one may have a suggestion that when the gravitational potential of underlying halos still plays the dominant role in regulating the LAE–IGM HI correlation, quasar (or AGN) feedback and/or the excess of surrounding IGM HI may have secondary effects.

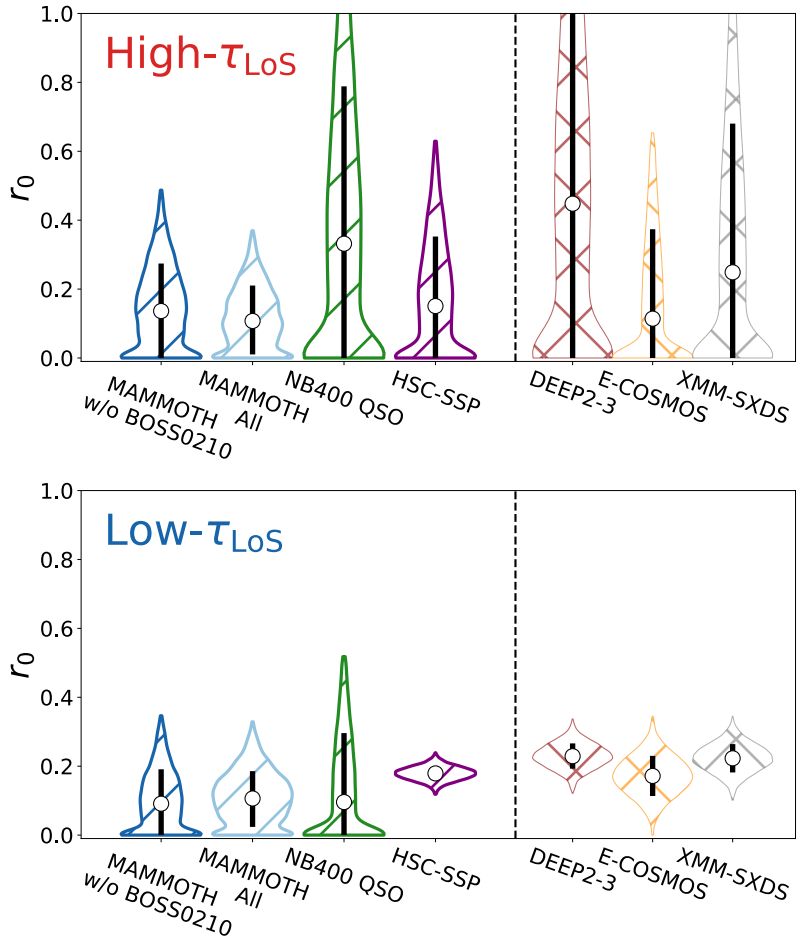

Figure 4.15: The violin plots show the 10,000 simulations of r_0 fitting of the CCF power-law in each sample. The simulations are performed by fluctuating the measured data points with their 1σ uncertainties and with the assumption of Gaussian distribution in each measurement. The dark blue, light blue, green, and purple shaded violins at the left-hand side are the r_0 distributions of the simulations in Lyα absorber-trace fields without BOSS0210, all the Lyα absorber-trace fields, grouping quasar fields, and the HSC-SSP general fields/. The brown, orange, and gray violins on the right-hand side are the r_0 distributions of the simulations in individual HSC-SSP fields: DEEP2-3, E-COSMOS, and XMM-SXDS fields, respectively. The black vertical bars suggest the range of 16%–84% percentile of r_0 in the simulations, and the open circles indicate the median value of r_0 in each sample.

Fields [1]	$\tau_{\rm LoS}$ [2]	$r_0^{{\rm high}-\tau_{\rm LoS}}$ [3]	$r_0^{{\rm low}-\tau_{\rm LoS}}$
MAMMOTH w/o BOSS0210	0.47	$0.136^{+0.138}_{-0.136}$	$0.091^{+0.100}_{-0.091}$
MAMMOTH All	0.47	$0.107^{+0.103}_{-0.097}$	$0.106^{+0.079}_{-0.082}$
NB400 QSO	0.35	$0.332^{+0.457}_{-0.332}$	$0.096^{+0.200}_{-0.096}$
HSC-SSP	0.30	$0.151^{+0.202}_{-0.151}$	$0.179^{+0.020}_{-0.020}$
DEEP2-3	0.30	$0.448^{+0.566}_{-0.448}$	$0.229^{+0.037}_{-0.036}$
E-COSMOS	0.30	$0.114^{+0.260}_{-0.114}$	$0.172^{+0.058}_{-0.058}$
XMM-SXDS	0.30	$0.249^{+0.432}_{-0.249}$	$0.223^{+0.041}_{-0.041}$

Table 4.2:: The parameters of CCF power law fitting for different samples.
[1] The sample category;
[2] The $\tau_{\rm LoS}$ criterion for separating high-$\tau_{\rm LoS}$ and low-$\tau_{\rm LoS}$ LoS subsamples;
[3] The r_0 fit for high-$\tau_{\rm LoS}$ subsample;
[4] The r_0 fit for low-$\tau_{\rm LoS}$ subsample.

Chapter 5

Discussions

Finally, we would like to make more discussions on IGM HI–LAE correlation on various aspects, including the average $\tau_{\rm LoS}$ as a function of distance to LAEs, the scale dependence on the overdensity-base analysis, and some discussions on the underlying physics. As the HSC-SSP fields are relatively shallow in the narrowband ($\rm NB387_{\it cut} < 24.2$) and the number of LoSs in NB400 quasar fields is small ($\rm N_{LoS} = 22$), the discussion in this section is primarily based on the NB387 Lyα absorber-trace fields with additional comparisons to ther other samples if possible. Most of the discussions in this chapter are published in Liang et al. (2021). However, we may have a chance to extend this discussion into field variation in the future if the Dark Energy Spectroscopic Instrument (DESI) data release and the Subaru Prime Focus Spectrograph (PFS) are available.

5.1 Average $\tau_{\rm LoS}$ Profile Centered at LAEs

To inspect the correlation on smaller scales down to sub-cMpc or $\sim$ 200 kpc in physical length, where circumgalactic medium (CGM) is thought to surround the host galaxies, we perform another analysis that in concept follows the stacking idea. We derive the radial distribution of the IGM optical depth averaged over all LoSs within annuluses centered at a specific LAE. We then calculate the mean of the optical depth

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 the $\langle\tau\rangle$ as

$$\delta_{\langle\tau\rangle}(d) = \frac{\langle\tau\rangle - \langle\tau\rangle_{\rm tot}}{\langle\tau\rangle_{\rm tot}}, \tag{5.1}$$

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 $0 < d < 0.3$ deg, 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. We mainly consider the case, including BOSS0210, 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 BOSS0210, which shows consistent results, except for a larger scatter due to the smaller LoS number. The gray squares represent the fine bin, while the red crosses represent the coarse bin in Figure 5.1. The error shown in the y-axis is the 1σ standard deviation from the $1,000$ 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 1,000 simulations by assessing random positions of the 64 LoSs with corresponding resolutions, indicating that bins outside the shaded areas are confident for inferences. The number of LoSs counted can be checked with the gray step function above.

As shown in Figure 5.1, 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 distributed with CGM. Rudie et al. (2012) and Momose et al. (2021b) also find CGM signals at $d \lesssim 300$ pkpc to star-forming galaxies that were originally identified as LBGs. Here, we find a similar result for LAEs at a smaller distance.

Besides, 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 concerning LAEs, which are generally less massive than LBGs. Instead, the clustering of IGM HI 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 HI down to around 400 pkpc.

We try to apply the similar estimate of the fluctuation of $\langle\tau\rangle$, $\delta_{\langle\tau\rangle}(d)$, to the other samples in the grouping quasar fields and general fields, and results are shown in Figure 5.2. Note that the grouping quasar fields only consist of the BOSS0240 and BOSS0755, and BOSS0210 is not included because of the significant difference in the

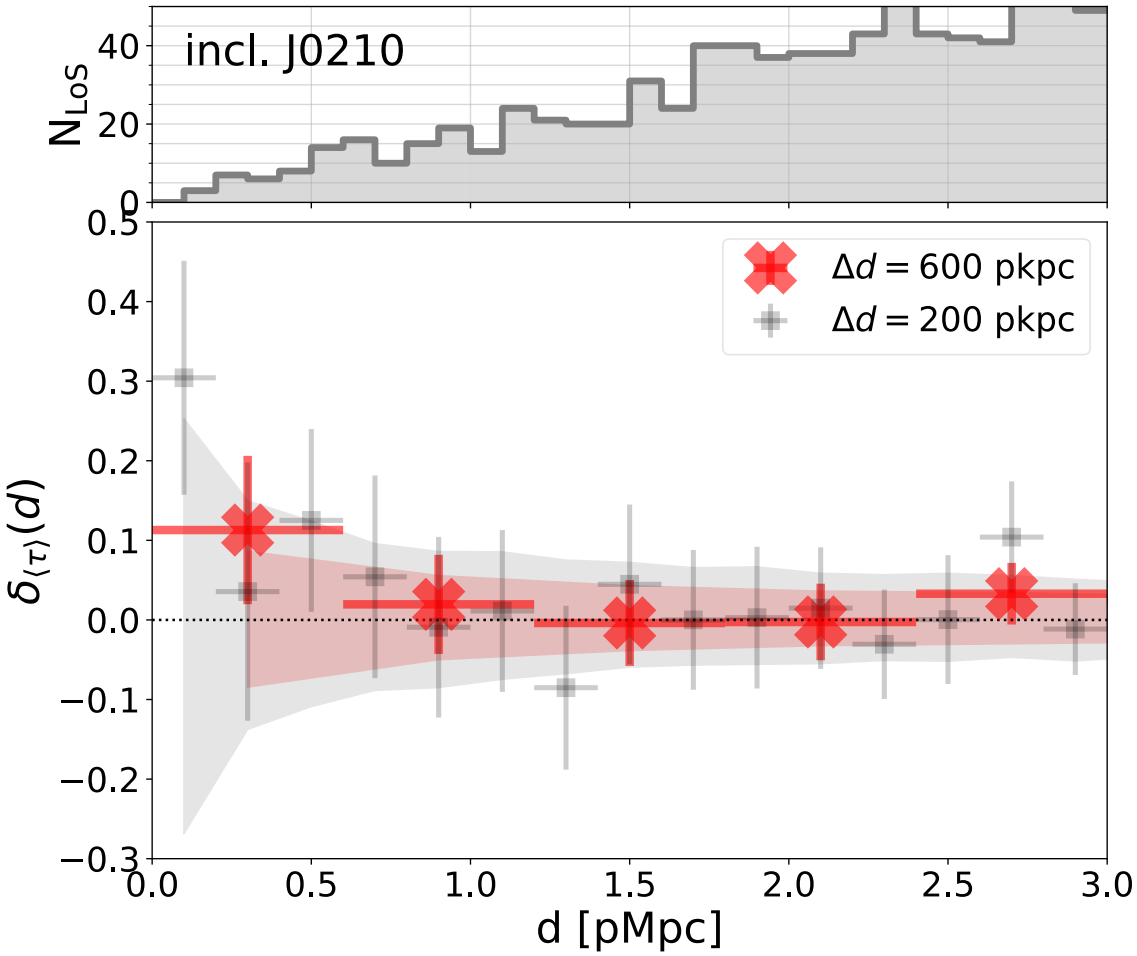

Figure 5.1: Fluctuations of the average $\tau_{\rm LoS}$ profile centered at LAEs, $\delta_{\langle\tau\rangle}(d)$, for the NB387 Ly$\alpha$ absorber-trace fields including BOSS0210. Errors are indicated by the 1σ standard deviation based on 1,000 times bootstrap resampling. The gray squares represent the fine bins with a resolution of 200 kpc, and the red crosses show the coarse bins with a resolution of 600 kpc. The shaded regions indicate the uncertainty with corresponding colors. 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.

faint-end limit of the samples, while the $\delta_{\langle\tau\rangle}(d)$ should be calculated with fair weight of each LoS. The blue squares are the same as the fine bins in Figure 5.1, and if we compare it with the estimates from grouping quasar fields and general fields, we find that at the innermost bin at $d = 200$ pkpc, the CGM signal is the most distinct in

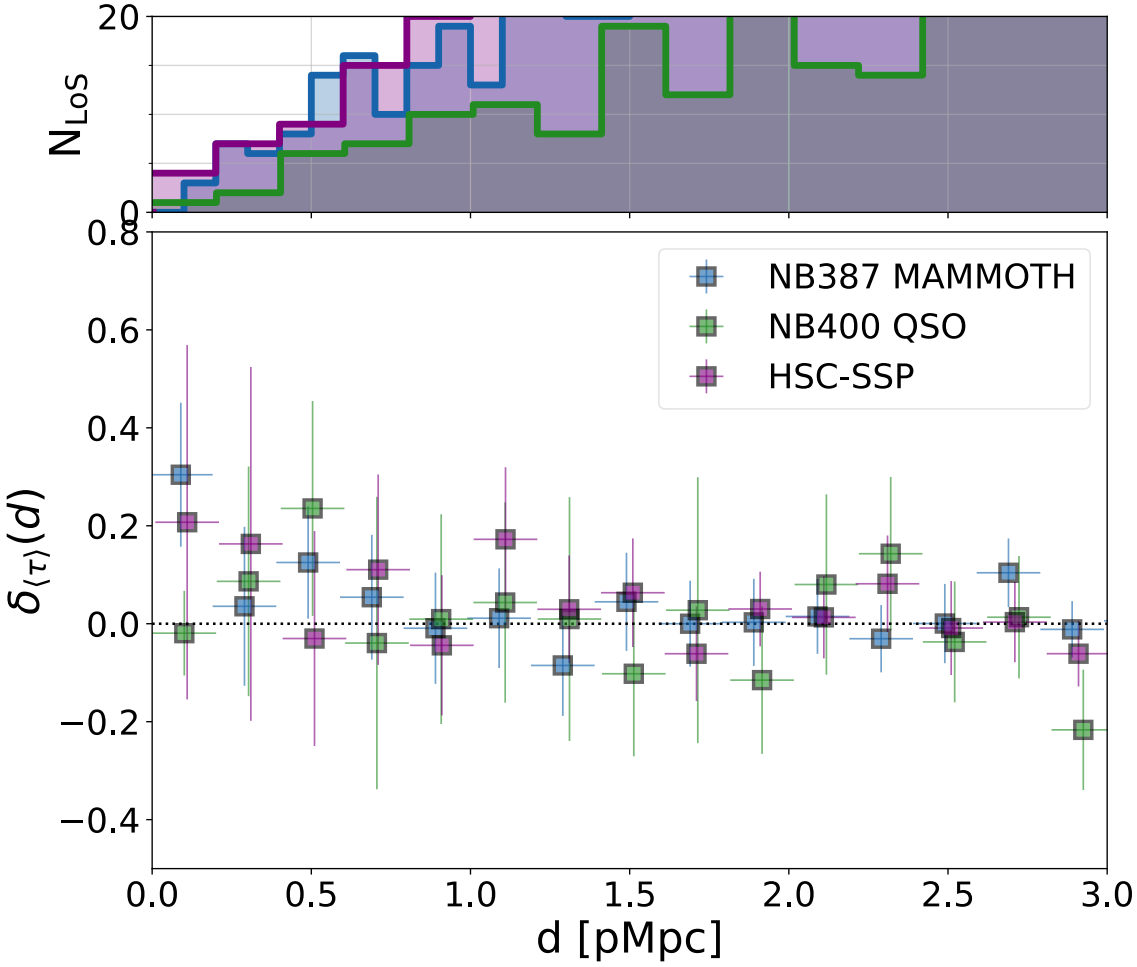

Figure 5.2: Fluctuations of the average $\tau_{\rm LoS}$ profile centered at LAEs, $\delta_{\langle\tau\rangle}(d)$, for different samples in the case of fine bin, i.e., $\Delta d = 200$ pkpc. Errors are indicated by the 1σ standard deviation based on 1,000 times bootstrap resampling. The blue, green, and purple squares are the results of the samples in all Lyα absorber-trace fields, the grouping quasar fields, and HSC-SSP general fields, respectively. The numbers of LoSs counted in the annulus at each step of 100 kpc are shown in the upper panel with corresponding colors.

the Lyα absorber-trace fields, and the sample in general fields shows some tentative signal, though with large uncertainty. Interestingly, the sample in grouping quasar fields gives the weakest signal that is different from the Lyα absorber-trace field at a $\sim 2\sigma$ level, which is almost comparable to the cosmic mean. This result suggests that, at least at the small scale concerned with CGM, the existence of quasars may play a significant role in changing the HI phase surrounding galaxies.

5.2 Scale Dependence of the $\delta_{\rm LAE}$–$\tau_{\rm LoS}$ Correlation

Possible scale dependence is already hinted at in the CCF in Section 4.2. 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. This test is based on the sample in Lyα absorber-trace fields, i.e., BOSS0210, BOSS0222, BOSS0924, and BOSS1419, because when varying the aperture size, the number density of LAEs is critical for statistics and the depth of Lyα absorber-trace fields is relatively homogeneous; on the other hand, we have also learned that the IGM HI–LAE correlation is most significant in this sample from Section 4.1, which can test more sensitively on the scale dependence.

As for the test procedure, we note that LoSs is kept for analysis only when $< 50\%$ of their vicinity is masked, which maintains a stable LoS number when the scale increases. The results are shown in Figure 5.3. The red and blue curve shows Spearman's rank correlation coefficients on various scales for the cases, including and excluding BOSS0210, respectively. The corresponding P-values are shown as the dashed line; the $\rm R_{aper}$ with P-values $> 5\%$ indicating results lacking confidence are shaded. It is clear that when BOSS0210 is included, the correlation remains moderate at a level of $\rho_S \sim 0.3$ for almost all scales, while in the case excluding BOSS0210, 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 BOSS0210. The trend is tested by making 1,000 realizations with randomly distributed LAEs and LoSs. The results fluctuate around $\rho_S = 0$ with scatter $\Delta\rho_S = \pm 0.2$ 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.

The correlation becomes stronger on a relatively small scale as the scale increases and peaks at $r = 13 \pm 2$ cMpc. With BOSS0210, the correlation shows a flatter shape when $r > 13$ cMpc, while it tends to decrease on the same scales without BOSS0210. The difference again indicates that BOSS0210 may have a unique structure, which can significantly alter the correlation in overdensity-based analysis. Therefore, a large sample size covering various environments is essential in the galaxy–IGM HI correlation studies regarding field variation. Here, however, the discussion focuses on the case with the exclusion of BOSS0210.

The CCF in Section 4.2 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

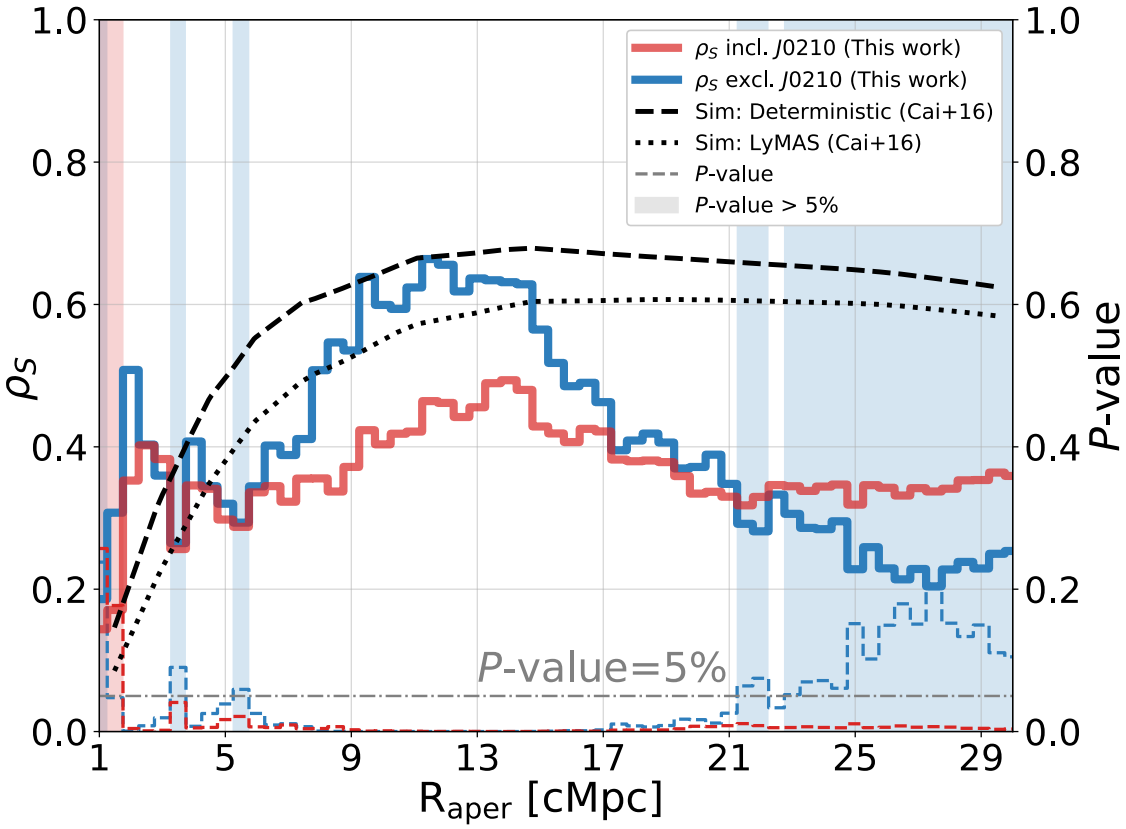

Figure 5.3: Scale dependence of the $\delta_{\rm LAE}$–$\tau_{\rm LoS}$ correlation based on NB387 Lyα absorber-trace fields. The bin size is 0.5 cMpc, and the red and solid blue curves are Spearman's rank correlation coefficient, including and excluding BOSS0210. The dashed colored curves indicate the 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 the Cai et al. (2016), Deterministic & 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.

5.3 can be viewed as the cumulative signal within the $\rm R_{aper}$, the scale of the correlation peak well agrees 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 $\rm R_{aper}$. Compared with the models in Cai et al. (2016), our results on the maximum correlation scale are consistent with the deterministic case and LyMAS on the projected plane. However, the amplitude may be different due to the other

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 depend on how Lyα forest is modeled. This suggests the effectiveness of the current simulated cosmological models in terms of the IGM HI gas.

The decrease on a large-scale 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 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 HI bias is almost flat on such large scales (Ando et al. 2019). Hence, the correlation between the total matter and HI in the models can remain strong on larger scales, while the LAE–IGM HI correlation simultaneously weakens. Another possible reason suggested by the simulations in Momose et al. (2021a) 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 of the Correlation

5.3.1 The Positive Correlation

We showed that at the redshift $z \approx 2.2 - 2.3$, there is a moderate to a 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 HI gas. This correlation is dependent on scale, and the peak is located at $R_{\rm aper} \sim 13$ cMpc. The correlation seems natural in that IGM HI gas tends to accumulate in deeper potential wells, which are associated with more massive halos. The condensed HI gas 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 HI gas in the surrounding IGM will be ionized. Such processes are likely to be more active in the case of the LAEs used in this study, which is thought to be a population of young star-forming galaxies. Our results finding the positive correlation indirectly imply that the escape fraction of ionizing photons from LAEs at $z \sim 2$ or their SFR may still not be sufficiently high as to fully ionize the IGM HI 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. However, although the power of such processes is still under debate as the scale at which their effects are evident is not precise, our finding of the possible field variation, especially in the region with grouping quasars, likely supports that the AGN feedback can have an influential effect on regulating the IGM–LAE correlations(Mukae et al. 2020a,b), even though this feedback effect may be secondary to the potential of dark matter halos.

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 HI and galaxies with rest-frame UV magnitudes down to $M_{UV} \approx -18$ estimated from 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 δ_{LAE} in the δ_{LAE}–$\delta_{\langle F \rangle}$ correlation), and down to at least 400 pkpc (or 1.3 cMpc) in the regions rich in IGM HI from the average τ_{LoS} profile centered on LAEs.

The results solely from Lyα absorber-trace fields imply that the ionization or feedback from galaxies (LAEs) and/or the inhabiting AGNs is insufficient to cancel large-scale gravitational effects. It also indicates that IGM HI still traces LSSs well at $z \sim 2$ on the scale of $1.3 \sim 13$ cMpc, although with a large scatter. Alternatively, as shown by the schematic diagram in Figure 5.4, the correlation in case (2) may result from additional inflow providing pristine HI 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.

Furthermore, after comparing the IGM HI-rich fields with the general fields and the grouping quasar fields, especially the BOSS0210 with both IGM absorbers and

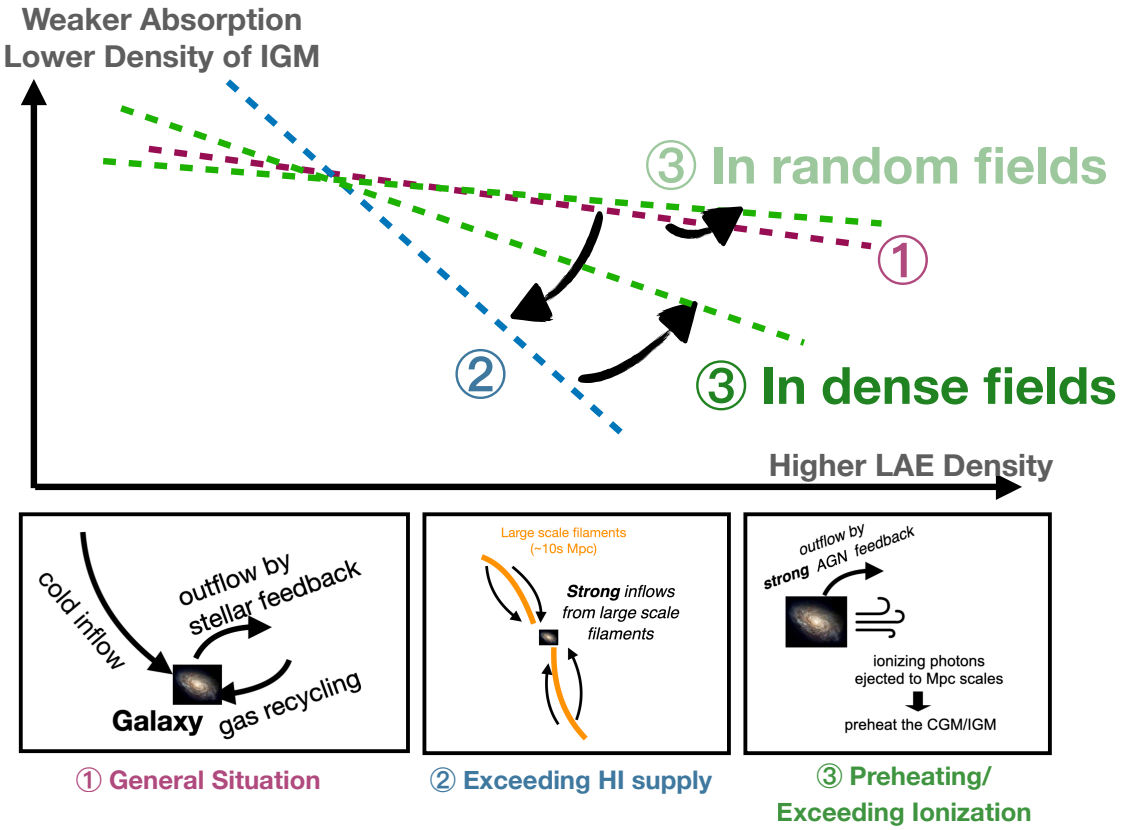

Figure 5.4: A schematic diagram to show the hypothesis interpreting the origins of different slopes in a variety of environments. The upper panel indicates the slopes we have found in different samples, and based on the characteristics of each field, we may get a suggestion that: Case (1) represents the general situation that inflows, outflows and gas recycling are well balanced; case (2) represents the case with exceeding HI supply from large-scale filaments in the high-density regions, and the IGM absorption is naturally enhanced, which may be the case of our Lyα absorber-trace fields.; case (3) represents the vital feedback from AGNs/quasars, and in the HI-rich regions, the preheating or ionizing effects can significantly alter the slopes; in general fields, HI is initially thin at $z \sim 2$, so the quasar feedback makes no essential difference.

grouping quasars, the latter hypothesis about the generous HI supply is also more favored. The grouping quasars possibly flatten the IGM–LAE correlation from the case (2) to case (3) in Figure 5.4 in the high-density HI regions by ejecting exceeding ionizing photons to ionize the surrounding IGM or preheat the cold HI gas. Nevertheless, in the general field, HI is almost fully ionized at $z = 2$, and therefore, the feedback of quasars can hardly change the relation (1) significantly in this case, and our NB400 grouping quasar fields may be just the examples of this situation.

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 semi-analytic 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 dynamic states. 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). Also, the recent galaxy-IGM correlation results reported by the LATIS team (Newman et al. 2022, also see Section 1.2.2) also show that there can be deficient UV-selected galaxies in the extremely strong Lyα absorption peaks compared to simulations. Thus, LAEs, or star-forming galaxies alone, may not be suitable for tracing the highest overdensity regions.

However, in large-scale statistical studies, 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 future, NIR surveys for HAEs, which are identified by the non-resonant Hα emission line, will help to determine the performance of the LAE tracer in unique environments

5.3.2 Hints from the Correlation Scatter

The scatter can also be an essential indicator of underlying structures. As shown in Figure 4.1 and 4.2, data points are distributed with large scatter. This may be due to the uncertainties in our measurements. However, we note that a similar large scatter can also be found in the results from high spatial resolution IGM tomography (e.g., LATIS introduced in Section 1.2.2) and the Osaka model. 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 $\sim 60 - 120$ cMpc along the redshift direction, and the scales mentioned above 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 BOSS0210 and other grouping quasar fields markedly change the statistical results, the field-to-field variation is a non-negligible factor. Concerning 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 4.1), there is still a 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.

Intrinsically, the scatter can also be a coincidence 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 BOSS0210 may penetrate galaxy overdensity associated with a gas filament lying in a direction transverse to the LoS by chance. Besides the morphological origin, the radiation from galaxies preheating the diffuse IGM HI 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 ~ 50 cMpc overall but not dependent on local overdensity. A unique system found in IGM tomography also shows similar characteristics (Lee et al. 2016). Actually, three outliers in BOSS0210 located at regions with $\delta_{\rm LAE} \gtrsim 1.0$ and $\tau_{\rm LoS} \lesssim 0.4$ are likely to reside in areas that are abundant with galaxies but deficient in cold HI gas, similar to the environments mentioned above.

Considering the possible physical origin of the scatter, it may be meaningful to quantify the statistics, which may offer a guide for future simulations to be compared with. Therefore, we first perform data binning along the $\delta_{\rm LAE}$ dimension in the $\delta_{\langle F\rangle}$–$\delta_{\rm LAE}$ relation with the width of $\Delta\delta_{\rm LAE}$= 0.5, and then derive the $\delta_{\langle F\rangle}$ at the give percentiles, i.e., 16%, 32%, 68%, and 84% etc. The left panel in Figure 5.5 shows the trends of the $32\% - -68\%$ and $16\% - -84\%$ in the three samples, the Lyα absorber-trace fields w/o BOSS0210, the grouping quasar fields combined with BOSS0210, and the general fields, in the dark and light black dashed curves, respectively. Thus we can further estimate the scatter by defining the width

$$\Delta\delta_{\langle F\rangle,\ 1}(\delta_{\rm LAE}) = \delta_{\langle F\rangle,\ 68\%} - \delta_{\langle F\rangle,\ 32\%}, \tag{5.2}$$

or

$$\Delta\delta_{\langle F\rangle,\ 2}(\delta_{\rm LAE}) = \delta_{\langle F\rangle,\ 84\%} - \delta_{\langle F\rangle,\ 16\%}. \tag{5.3}$$

For comparison, we can also define it according to the standard deviation in each bin:

$$\Delta\delta_{\langle F\rangle,\ 2}(\delta_{\rm LAE}) = \sqrt{\sum_{i}^{n} \left(\delta_{\langle F\rangle,\ i} - \left\langle\delta_{\langle F\rangle}\right\rangle\right)/(n-1)}. \tag{5.4}$$

The $\Delta\delta_{\langle F\rangle}(\delta_{\rm LAE})$ is shown in the right panel of Figure 5.5. Interestingly, the sample in Lyα absorber-trace fields is almost consistent with the general fields in

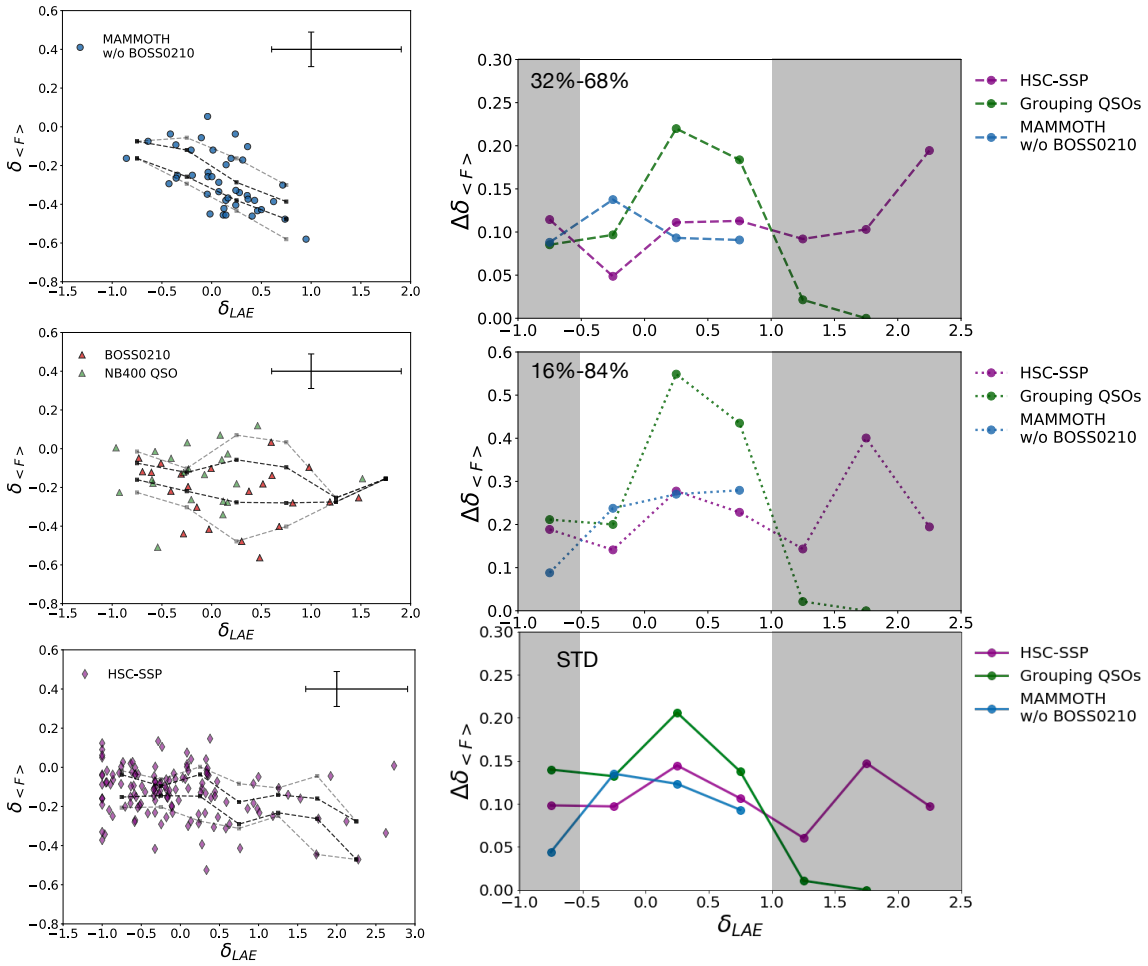

Figure 5.5: The Scatter of $\delta_{\langle F\rangle}$–$\delta_{\rm LAE}$ relation. *Left panel*: the $\delta_{\langle F\rangle}$–$\delta_{\rm LAE}$ correlation with trendlines of percentile ranges in the binning data. From upper to bottom figures are the pure Lyα absorber-trace field excluding BOSS0210, the grouping quasar fields including BOSS0210, and the general fields from HSCSSP. The dark and light gray dashed curves suggest the 32%–68% and 16%–84% ranks respectively. *Right panel*: the trendlines of the $\delta_{\langle F\rangle}$–$\delta_{\rm LAE}$ correlation scatter. From upper to bottom figures are the scatter with different definitions: the width between 32% and 68% ranks, the width between 16% and 84% ranks, and the standard deviation of the binning data. Purple, green, and blue curves represent samples in general fields, grouping quasar fields, and Lyα absorber-trace fields. The shaded region indicate the bins with few data points counted.

trend and amplitude. In contrast, the sample in grouping quasar fields has apparent excess at the mean to moderately overdense regions with $0 < \delta_{\rm LAE} < 0.5$, regardless of the $\Delta\delta_{\langle F\rangle}$ definition. This result likely supports that the radiation around extreme

objects can significantly alter the $\delta_{\langle F \rangle}$–$\delta_{\rm LAE}$. More specific physical quantities, e.g., the volume fraction of large-scale filaments and nodes at $z \sim 2$ and the HI fraction in large-scale filaments, may be possible to be derived by comparing the observational $\Delta\delta_{\langle F \rangle}(\delta_{\rm LAE})$ with the predictions from future simulations that carefully model the radiative transfer of the Lyα.

Chapter 6

Summary

Throughout this thesis, I, as the principal investigator, have led the project working on understanding the correlation between Lyα emitters and the IGM HI at the redshift $z \sim 2$ by using Subaru/HSC and the SDSS/BOSS spectra. We have conducted several narrowband imaging observations with the filters NB387 and NB400 to map the $z = 2.18$ and $z = 2.29$ LAE in six fields, BOSS0210, BOSS0222, BOSS0924, BOSS1419, BOSS0240, and BOSS0755 on Subaru/HSC. These fields are selected with unique tracers like the grouping strong IGM Lyα absorbers or the grouping quasars, which expand different galaxy environments' classes to discuss field variation. Also, these samples have covered 3,687 LAEs over a survey area of ~ 8 deg^2. Moreover, there is a recent opportunity from the Subaru HSC-SSP survey, and a new $z = 2.18$ LAE catalog consisting of 3,720 LAEs is built based on the NB387 survey in the Deep/Ultra-Deep fields, i.e., DEEP2-3, E-COSMOS, and the XMM-SDSS, which cover the sky area of ~ 19 deg^2 (Ono et al. 2021). We can now combine both catalogs for discussing the field variation on the Lyα emitters and the IGM HI correlation, and this sample is the largest one in terms of both the LAEs and the covered sky area, which has provided us with the unprecedented power to address this topic. Parts of the results are already published in Liang et al. (2021), and more developments are in preparation for future publications.

In specific, in this thesis, we have worked on items:

1. We construct six LAE overdensity maps for six HSC fields, four of which are traced by IGN HI and three of which are fully or partially traced by grouping quasars at the redshifts $z = 2.18$ and $z = 2.29$. These maps have identified many interesting structures, including large filamentary structures, clumpy

overdensity, and voids.

2. We carefully evaluate the data quality of the HSC-SSP $z = 2.18$ LAE catalog and have performed a completeness estimate by using mock galaxies simulated by *Balrog* and GalSim (Rowe et al. 2015; Suchyta et al. 2021). This completeness is used in our overdensity reconstruction of the HSC-SSP fields and the criterion to determine patches for CCF analysis. After this, we constructed three overdensity maps for the DEEP2-3, E-COSMOS, and XMM-SXDS fields based on the $z = 2.18$ LAE catalog. We can also find unique structures from this map, which will be ideal targets for future studies.

3. We perform the $\delta_{\rm LAE}$–$\tau_{\rm LoS}$ correlation analysis based on the estimated LAE overdensity and the measured effective optical depth on 15 h^{-1}Mpc scales in the background SDSS/BOSS quasars. This correlation is further transformed into the $\delta_{\langle F\rangle}$–$\delta_{\rm LAE}$ relation, whose slope can be compared among fields to inspect how the Lyα absorption is dependent on the local galaxy overdensity. As a result, we have found a moderate and positive correlation between LAE overdensity and IGM HI absorption based on the Lyα absorber-trace fields. Still, the $\delta_{\langle F\rangle}$–$\delta_{\rm LAE}$ slope is much flatter in general fields from HSC-SSP and in the grouping quasar fields. This result suggests the existence in the field variation of the galaxy–IGM HI correlation and has revealed that the IGM Lyα absorption is stronger at the given $\delta_{\rm LAE}$ in the HI rich regions than in the field regions. Meanwhile, the existence of grouping quasars may have vital feedback to ionize/preheat the dense IGM HI, although it makes no significant difference in the general fields due to the originally diffuse HI.

4. We perform correlation analysis using the two-point cross-correlation functions (CCFs) between the high-$\tau_{\rm LoS}$ and low-$\tau_{\rm LoS}$ LoSs and LAEs. The subsamples are separated regarding the median $\tau_{\rm LoS}$ of the full LoS sample in each field. As the first inspection on the possible artificial bias from the limited LoS number, we use the relative CCFs in contrast to the CCFs of the full sample instead of the original CCFs. As a result, we have found the positive correlation signal between the high-$\tau_{\rm LoS}$ LoSs and LAEs up to a scale of 4 ± 1 pMpc in the Lyα absorber-trace fields, and a power-law can well model the measured values. This suggests that the LAEs tend to reside in the HI rich regions up to a scale of 4 pMpc. Based on the suggestion from the first inspection, we optimize the power-law fitting process for the original CCFs, and also expand the analysis by comparing the CCFs among the three categories of samples. We find that the signals of CCFs are roughly consistent with being well modeled by a power-law, and the clustering strength r_0 also agrees with each other within the errors among different samples. Both findings suggest that the galaxy–underlying

halos can still dominate IGM HI correlation, although the negative signal in the innermost bin of the sample in the grouping quasar fields infers the possible influence of quasar feedback at small scales, which is in agreement with the fluctuation of average $\tau_{\rm LoS}$ profile $\delta_{\langle\tau\rangle}(d)$ results.

5. We perform further inspections primarily based on the data in Lyα absorber trace fields, such as looking into the fluctuation of average $\tau_{\rm LoS}$ profile $\delta_{\langle\tau\rangle}(d)$, the scale dependence of the $\tau_{\rm LoS}$–$\delta_{\rm LAE}$ correlation, and additional discussions on the underlying physics from the correlation we have identified. The average $\tau_{\rm LoS}$ profile shows a 30% excess at $d < 200$ pkpc with 2σ significance in the Lyα absorber-trace fields, broadly consistent with the general fields, which hints at the CGM signals around LAEs. Meanwhile, the $\delta_{\langle\tau\rangle}$ of grouping quasar fields shows only the cosmic mean level at $d < 200$ pkpc, implying the effect of quasar feedback on CGM scales. We also find that the aperture size can partly alter the correlation for measuring the LAE overdensity. It reaches a peak at $R_{\rm aper} = 13 \pm 2$ cMpc, agreeing to the model prediction from Cai et al. (2016).

6. We interpret the finding of the positive correlation between galaxies and IGM HI as a suggestion that the escape fraction of ionizing photons from LAEs or their SFR may not be sufficient to ionize the surrounding IGM HI in the Lyα absorber-trace fields. The steeper slope of $\delta_{\langle F\rangle}$-$\delta_{\rm LAE}$ relation in the Lyα absorber-trace fields can be a result of either the exceeding HI supply from the large-scale filaments or the deficient LAE detection due to HI suppression, as these hypotheses can be supported by that the slope is flattened with the emergence of grouping quasars. Combining the $\delta_{\langle F\rangle}$–$\delta_{\rm LAE}$ correlation and CCF results, we find that dark matter halos still dominate the galaxy–IGM correlation at $z \approx 2$, although exceeding HI supply and AGN/quasar feedback can play secondary roles.

7. We find the scatter of the $\delta_{\langle F\rangle}$–$\delta_{\rm LAE}$ correlation is larger in the grouping quasar fields with mean to moderate LAE density $0 < \delta_{\rm LAE} < 0.5$, compared to either the general fields or the Lyα absorber-trace fields, again emphasizing the effects of AGN feedback on regulating IGM HI–LAE correlation. The statistics from the $\delta_{\langle F\rangle}$–$\delta_{\rm LAE}$ scatter can be compared with future simulations that can well model the radiative transfer of Lyα emission to investigate meaningful quantities, e.g., the volume fraction of large-scale filaments and nodes at $z \sim 2$ and the HI fraction in large-scale filaments.

Conclusion

In this thesis, we have confirmed the correlation between galaxies and IGM with the largest sample size and the widest survey area at the epoch of the *Cosmic Noon*, which is stronger in HI-rich environments. This observational result supports the big picture predicted by the cosmological simulations that galaxies preferentially form in the high-density regions, particularly of the gas reservoirs, i.e., the intersects of the large-scale cosmic webs.

This homogeneous sample also enables us to inspect the variation of the galaxy–IGM HI interplay in regions that can be influenced by various factors. The steeper slope of $\delta_{\langle F \rangle}$-$\delta_{\rm LAE}$ relation shown in the Lyα absorber-trace fields suggests that generous IGM supply from large-scale filaments HI can strengthen the galaxy–IGM correlation by providing sufficient materials for building galaxies. This result is consistent with the suggestion from LATIS survey that UV-selected galaxies are deficient in dense HI regions (Newman et al. 2022, also see Introduction 1.2.2). The slope in HI-rich fields flattened with the emergence of grouping quasars, corroborated by the CCF signals and average $\tau_{\rm LoS}$, has emphasized the effective influence of AGN feedback, at least on relatively small scales. This result can be the case comparable to SSA22 protocluster (Mawatari et al. 2017, also see Introduction 1.2.3); the exceptional region with a group of galaxies but no strong absorption spike in CLAMATO is also similar to this case, although we have no information of the coeval AGN activities (Lee et al. 2016, also see Introduction 1.2.2).

Unlike previous studies encountering individual cases, we have systematically searched for the unique regions and can recognize the environmental or feedback effects with consistent suggestions. The statistics, including the slope of the $\delta_{\langle F \rangle}$-$\delta_{\rm LAE}$ slopes and the clustering lengths of CCFs, will be practical indicators for future studies to quantitatively understand how various physical processes can affect the galaxy formation and evolution, both observationally and theoretically. However, we note that despite these secondary effects from astrophysical processes, the gravitational potential of underlying halos is still the dominator of the galaxy–IGM correlation at $z \sim 2$, as proved by the similarity of CCFs and clustering lengths among different samples. This highlights the validity of the current cosmological simulations based on ΛCDM theory (Nagamine et al. 2021; Yajima et al. 2022)

Future Plan

The works presented in this thesis are not the end for studying the galaxy–IGM HI interplays at high-redshift but is instead a starting point.

Based on our constructed overdensity maps, we have identified many exciting structures for follow-up observation to perform a detailed case inspection. For example, in the E-COSMOS overdensity map shown in Figure 3.10 , we have identified a $\sim 6\sigma$ significant LAE overdensity around (R.A., Decl.) = (149.5 deg, 1.4 deg). Since it is located in the Deep/Ultra-deep field of the Subaru HSC-SSP survey, there are abundant synergy data in multi-wavelengths. We will utilize this rich dataset to thoroughly inspect its environments, including the surrounding IGM HI through the absorption imprinting in the narrowbands towards the background sources. We have also obtained the spectral data to confirm the galaxies at the densest peak from the Gemini/GMOS, which is under analysis.

In addition to the fields that are already presented in this thesis, there are also more MAMMOTH candidates traced by grouping strong Lyα absorbers now available. We have performed NIR imaging observations with Subaru/SWIMS targeting a MAMMOTH candidate that is already mapped by Subaru/HSC for LAEs, and this target is suggested to be filled with dense IGM HI but lacks a sufficient number of LAEs compared to the $\delta_{\langle F \rangle}$–$\delta_{\rm LAE}$ correlation. The observation has provided us with the data of K_s, $J1$, $J2$, Paβ-off and $NB2167$ bands, which will be used to select HAEs and distant red galaxies (DRGs). HAEs are the relatively massive star-forming galaxies that can be searched by the non-resonant line, Hα emission, while DRGs are even more massive with higher metallicity compared to HAEs. On the other hand, there are also the MAMMOTH candidates BOSS1244 and BOSS1524 that are already spectroscopically confirmed as the most massive protoclusters from HAE surveys (Zheng et al. 2021; Shi et al. 2021, see also in Section 1.3). We have a new accepted Subaru/HSC NB395 observation that will target these two structures to map the coeval LAEs. The comparisons of these different galaxy populations in the environments with high-density IGM HI will answer one of the key questions remaining in this thesis: whether LAEs are effective structure tracers in the present surveys, especially in the extremely overdense regions?

In the near future, the upcoming Subaru Prime Focus Spectrograph (PFS) will offer us an unprecedentedly powerful tool to inspect the IGM HI through Lyα absorption with a much higher spatial resolution – the average separation of LoSs in the PFS period can be as high as $r \sim 2.5$ cMpc, while it is $r \sim 20$ cMpc in the study of this thesis. Meanwhile, the VLT Multi-Object Optical and Near-infrared Spectrograph

(MOOS) will benefit us from the infrared side to better understand the properties of different galaxies. This drastic improvement in future spectroscopic ability will help us to reveal the IGM and its relation with galaxies in much more detail in the future. Moreover, the high-efficiency survey ability also enables us to go to higher redshift, and how the galaxy–IGM relation evolves across cosmic time will also be under the scope.

From the long-term perspective, next-generation instruments, including James Webb Space Telescope (JWST), Rubin Telescope (LSST), Roman Telescope (WFIRST) and ELTs (TMT, E-ELT, GMT and etc.), will definitely push forward the study on IGM across the cosmic time. For example, in this thesis, we have stressed the importance of the survey area for understanding the variation of the galaxy–IGM correlation in different environments, but in the epoch with the Rubin telescope, one can essentially set up robust statistics to determine the dependent factors of the variation with $100\times$ larger survey area of galaxies with relatively confident photometric redshifts down to $r \approx 27.8$ once the LSST survey is finished. Moreover, as suggested by Newman et al. (2022), UV-selected galaxies can be missed in the area with extremely strong IGM absorption, which is also supported in our study; surveys of galaxies in longer wavelengths will be critical to better understand the interplays between galaxies and IGM. While Roman Telescope can offer us to statistically survey the galaxies in the infrared wavelength due to its wide FoV, JWST, on the other hand, will offer the best sensitivity and resolution to target specific unique structures, e.g., the BOSS0210 that encloses a $\sim 30\sigma$ significance of SDSS/BOSS quasar overdensity in this thesis, for detailed case inspections.

There is no doubt that this is the true timing to start exploring the hidden majority of the baryonic matters that construct our universe – the cold, diffuse, but ubiquitous IGM.

Chapter 7

Lists

List of Acronyms

- **pkpc:** Proper kpc
- **cMpc:** Comoving Mpc
- **pMpc:** Proper Mpc
- **AGN:** Active Galactic Nucleus
- **BAL:** Broad Absorption Line Quasar
- **(e)BOSS:** (Extended) Baryon Oscillation Spectroscopic Survey
- **CCF:** Cross-Correlation Function
- **CFHTLS:** Canada-France-Hawaii Telescope Legacy Survey
- **CGM:** Circumgalactic Medium
- **CHORUS:** Cosmic HydrOgen Reionization Unveiled with Subaru Survey
- **CLAMATO:** COSMOS Lyman-Alpha Mapping And Tomography Observation
- **CMB:** Cosmic Microwave Background
- **CoSLAs:** Coherently Strong Lyα Absorption Systems
- **DESI:** Dark Energy Spectroscopic Instrumen

- **DLA:** Damped Lyα Systems
- **DRG:** Distant Red Galaxy
- **DSFG:** Dusty Star-Forming Galaxy
- **ELAN:** Enormous Lyα Nebula
- **FOS:** Faint Object Spectrograph
- **FoV:** Field of View
- **FWHM:** Full Width at Half Maximum
- **HAE:** Hα Emitter
- **HCD:** High Column Density Systems
- **HSC:** Hyper-Suprime Cam
- **HSC-SSP:** Hyper-Suprime Cam Subaru Strategic Program
- **HST:** Hubble Space Telescope
- **ICM:** Intracluster Medium
- **IGM:** Intergalactic Medium
- **KBSS:** Keck Baryonic Structure Survey
- **LAB:** Lyα Blob
- **LAE:** Lyα Emitter
- **LATIS:** Lyα Tomography IMACS Survey
- **LBG:** Lyman Break Galaxy
- **LoS:** Line-of-Sight
- **LSS:** Large-Scale Structure
- **LyMAS:** Lyα Mass Association Scheme
- **MAMMOTH:** Mapping the Most Massive Overdensities through Hydrogen
- **MF-PCA:** Mean-Flux-regulated Principal Component Analysis
- **MZR:** Mass-Metallicity Relation

- **NB:** Narrowband
- **PCA:** Principal Component Analysis
- **PDR2:** The Second Data Release
- **PFS:** Prime Focus Spectrograph
- **PSF:** Point-Spread Function
- **PS1:** Pan-STARS DR1
- **QSO:** Quasi-Stellar Object/Quasar
- **RMS:** Root-Mean-Square
- **SDSS:** Sloan Digital Sky Survey
- **SED:** Spectral Energy Distribution
- **SFR:** Star Formation Rate
- **S/N:** Signal-to-Noise Ratio
- **HI:** Neutral Hydrogen
- **WCS:** World Coordinate System
- **WHIM:** Warm–Hot Intergalactic Medium

List of Figures

List of Tables

References

Adelberger, K. L., Steidel, C. C., Shapley, A. E., & Pettini, M. 2003, ApJ, 584, 45 5.3.1

Aihara, H., AlSayyad, Y., Ando, M., et al. 2019, PASJ, 71, 114 2.2.3, 2.3

Ando, R., Nishizawa, A. J., Hasegawa, K., Shimizu, I., & Nagamine, K. 2019, MNRAS, 484, 5389 5.2

Baugh, C. M. 2006, Reports on Progress in Physics, 69, 3101 1.1.1

Becker, G. D., Hewett, P. C., Worseck, G., & Prochaska, J. X. 2013, MNRAS, 430, 2067 2.1, 2.2.1

Bertin, E. 2011, in Astronomical Society of the Pacific Conference Series, Vol. 442, Astronomical Data Analysis Software and Systems XX, ed. I. N. Evans, A. Accomazzi, D. J. Mink, & A. H. Rots, 435 3.2.2

Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393 2.2.5

Bond, J. R., Kofman, L., & Pogosyan, D. 1996, Nature, 380, 603 1.1.1

Bosch, J., Armstrong, R., Bickerton, S., et al. 2018, PASJ, 70, S5 2.2.3

Bouwens, R. J., Illingworth, G. D., Ellis, R. S., Oesch, P. A., & Stefanon, M. 2022, arXiv e-prints, arXiv:2205.11526 1.4

Bădescu, T., Yang, Y., Bertoldi, F., et al. 2017, ApJ, 845, 172 1.1.2

Cai, Z., Fan, X., Peirani, S., et al. 2016, ApJ, 833, 135 1.3.1, 1.6, 1.3.1, 2.1, 2.1, 5.3, 5.2, 5, 7

Cai, Z., Fan, X., Yang, Y., et al. 2017a, ApJ, 837, 71 1.7, 2.2.1, 3.3.1, 4.1.1, 5.3.1, 7

Cai, Z., Fan, X., Bian, F., et al. 2017b, ApJ, 839, 131 1.3.2, 1.7, 3.3.1, 7

Capellupo, D. M., Hamann, F., Herbst, H., et al. 2017, MNRAS, 469, 323 1

Casey, C. M., Cooray, A., Capak, P., et al. 2015, ApJ, 808, L33 1.1.2

Caucci, S., Colombi, S., Pichon, C., et al. 2008, MNRAS, 386, 211 1.2.2

Cen, R., & Ostriker, J. P. 2000, ApJ, 538, 83 1.1.1, 1.1.1, 1.1, 5.2, 7

Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, arXiv e-prints, arXiv:1612.05560 2.2.2

Chen, H.-W., & Mulchaey, J. S. 2009, ApJ, 701, 1219 5.3.1

Chen, Y., Steidel, C. C., Hummels, C. B., et al. 2020, MNRAS, 499, 1721 3.1.1

Chiang, Y.-K., Overzier, R., & Gebhardt, K. 2013, ApJ, 779, 127 1.3.2, 2.2.1

Chiang, Y.-K., Overzier, R. A., Gebhardt, K., & Henriques, B. 2017, ApJ, 844, L23 1.1.2, 1.2, 7

Ciardullo, R., Gronwall, C., Adams, J. J., et al. 2013, ApJ, 769, 83 3.1.2

Cooke, E. A., Hatch, N. A., Muldrew, S. I., Rigby, E. E., & Kurk, J. D. 2014, MNRAS, 440, 3262 1.1.2

Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, AJ, 145, 10 1.3.1, 2.1

Dawson, K. S., Kneib, J.-P., Percival, W. J., et al. 2016, AJ, 151, 44 1.3.1, 2.1

De Lucia, G., Weinmann, S., Poggianti, B. M., Aragón-Salamanca, A., & Zaritsky, D. 2012, MNRAS, 423, 1277 1.1.1

Dekel, A., & Birnboim, Y. 2006, MNRAS, 368, 2 5.3.1

Dijkstra, M. 2014, , 31, e040 1.4, 1.9, 7

Elbaz, D., Daddi, E., Le Borgne, D., et al. 2007, A&A, 468, 33 1.1.1

Faucher-Giguère, C.-A., Prochaska, J. X., Lidz, A., Hernquist, L., & Zaldarriaga, M. 2008, ApJ, 681, 831 2.2.1, 4.1.1

Fumagalli, M., Mackenzie, R., Trayford, J., et al. 2017, MNRAS, 471, 3686 1.1.2

Gehrels, N. 1986, ApJ, 303, 336 3.4, 4.1.1

Guaita, L., Gawiser, E., Padilla, N., et al. 2010, ApJ, 714, 255 3.1.1, 3.1.2

Hasinger, G., Capak, P., Salvato, M., et al. 2018, ApJ, 858, 77 2.5, 2.2.4, 3.1.2, 7

Hathi, N. P., Le Fèvre, O., Ilbert, O., et al. 2016, A&A, 588, A26 5.3.1

Hayashino, T., Inoue, A. K., Kousai, K., et al. 2019, MNRAS, 484, 5868 1.2.3

Hennawi, J. F., Prochaska, J. X., Cantalupo, S., & Arrigoni-Battaia, F. 2015, Science, 348, 779 3.3.1

Hinshaw, G., Nolta, M. R., Bennett, C. L., et al. 2007, ApJS, 170, 288 1.1.1

Hough, T., Gurung-López, S., Orsi, Á., et al. 2020, MNRAS, 499, 2104 5.3.1

Husband, K., Bremer, M. N., Stott, J. P., & Murphy, D. N. A. 2016, MNRAS, 462, 421 1.1.2

Inoue, A. K., Shimizu, I., Iwata, I., & Tanaka, M. 2014, MNRAS, 442, 1805 3.1.1

Inoue, A. K., Yamanaka, S., Ouchi, M., et al. 2020, PASJ, 72, 101 2.5, 2.2.4, 3.1.2, 7

Kauffmann, G., Heckman, T. M., White, S. D. M., et al. 2003, MNRAS, 341, 33 2.2.4

Kikuta, S., Matsuda, Y., Cen, R., et al. 2019, PASJ, 71, L2 1.1.2

Konno, A., Ouchi, M., Nakajima, K., et al. 2016, ApJ, 823, 20 3.1.1, 3.1.1, 3.1.2

Kriek, M., Shapley, A. E., Reddy, N. A., et al. 2015, ApJS, 218, 15 1.2.2, 1.3.2

Kurczynski, P., Gawiser, E., Rafelski, M., et al. 2014, ApJ, 793, L5 2.2.4

Kusakabe, H., Shimasaku, K., Ouchi, M., et al. 2018, PASJ, 70, 4 1.4, 4.1.4

Kusakabe, H., Shimasaku, K., Momose, R., et al. 2019, PASJ, 71, 55 3.1.1

Landy, S. D., & Szalay, A. S. 1993, ApJ, 412, 64 4.2.1

Lee, K.-G., Hennawi, J. F., White, M., Croft, R. A. C., & Ozbek, M. 2014a, ApJ, 788, 49 1.2.2

Lee, K.-G., Suzuki, N., & Spergel, D. N. 2012, AJ, 143, 51 2.1

Lee, K.-G., Hennawi, J. F., Stark, C., et al. 2014b, ApJ, 795, L12 1.2.2, 1.4, 7

Lee, K.-G., Hennawi, J. F., White, M., et al. 2016, ApJ, 817, 160 1.2.2, 1.4, 1.2.3, 2.1, 5.3.2, 6, 7

Lee, K.-G., Krolewski, A., White, M., et al. 2018, The Astrophysical Journal Supplement Series, 237, 31 1.2.2

Liang, Y., Kashikawa, N., Cai, Z., et al. 2021, ApJ, 907, 3 1.5, 5, 6

Mackenzie, R., Fumagalli, M., Theuns, T., et al. 2019, Monthly Notices of the Royal Astronomical Society 1.1.2, 1.2.1

Madau, P., & Dickinson, M. 2014, ARA&A, 52, 415 5.3.1

Mawatari, K., Yamada, T., Nakamura, Y., Hayashino, T., & Matsuda, Y. 2012, ApJ, 759, 133 1.4, 3.1.1

Mawatari, K., Inoue, A. K., Yamada, T., et al. 2017, MNRAS, 467, 3951 1.2.3, 5.3.2, 6

Miller, J. S. A., Bolton, J. S., & Hatch, N. 2019, MNRAS, 489, 5381 1.1.2, 5.3.2

Momose, R., Shimizu, I., Nagamine, K., et al. 2021a, ApJ, 911, 98 5.2

Momose, R., Shimasaku, K., Kashikawa, N., et al. 2021b, ApJ, 909, 117 4.1.4, 4.1.4, 4.2.2, 5.1

Mukae, S., Ouchi, M., Kakiichi, K., et al. 2017, ApJ, 835, 281 1.2.3, 3.1.1, 4.1.4, 4.8, 4.9, 4.1.4, 5.3.1, 5.3.2, 7

Mukae, S., Ouchi, M., Hill, G. J., et al. 2020a, ApJ, 903, 24 5.3.1

Mukae, S., Ouchi, M., Cai, Z., et al. 2020b, ApJ, 896, 45 3.3.1, 5.3.1

Nagamine, K., Shimizu, I., Fujita, K., et al. 2021, ApJ, 914, 66 4.8, 4.1.4, 4.9, 6, 7

Nakajima, K., Ouchi, M., Shimasaku, K., et al. 2012, ApJ, 745, 12 3.1.1, 3.1.1, 3.4, 3.12, 7

Nelson, D., Springel, V., Pillepich, A., et al. 2019, Computational Astrophysics and Cosmology, 6, 2 1.2.2

Newman, A. B., Rudie, G. C., Blanc, G. A., et al. 2020, ApJ, 891, 147 1.2.2

—. 2022, Nature, 606, 475 1.5, 1.2.2, 4.1.4, 5.3.1, 6, 6, 7

Noirot, G., Stern, D., Mei, S., et al. 2018, ApJ, 859, 38 1.1.2

Norberg, P., Baugh, C. M., Gaztañaga, E., & Croton, D. J. 2009, MNRAS, 396, 19 4.2.1

Noterdaeme, P., Petitjean, P., Carithers, W. C., et al. 2012, A&A, 547, L1 2.1

Ogura, K., Nagao, T., Imanishi, M., et al. 2017, PASJ, 69, 51 1.1.2

Ono, Y., Itoh, R., Shibuya, T., et al. 2021, ApJ, 911, 78 1.5, 2.3, 3.2.1, 3.2.2, 3.6, 3.14, 6, 7

Onoue, M., Kashikawa, N., Uchiyama, H., et al. 2018, PASJ, 70, S31 1.1.2, 3.3.1

Oteo, I., Ivison, R. J., Dunne, L., et al. 2018, ApJ, 856, 72 1.1.2

Pâris, I., Petitjean, P., Ross, N. P., et al. 2017, A&A, 597, A79 2.1

Peirani, S., Weinberg, D. H., Colombi, S., et al. 2014, ApJ, 784, 11 1.3.1

Pickles, A. J. 1998, PASP, 110, 863 2.2.3, 2.2.4

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, A&A, 594, A13 1.5

Planck Collaboration, Akrami, Y., Arroja, F., et al. 2018, arXiv e-prints, arXiv:1807.06205 1.1.1

Polletta, M., Tajer, M., Maraschi, L., et al. 2007, ApJ, 663, 81 3.1, 3.1.1, 7

Prochaska, J. X., Hennawi, J. F., Lee, K.-G., et al. 2013, ApJ, 776, 136 1.2.1

Rakic, O., Schaye, J., Steidel, C. C., & Rudie, G. C. 2012, ApJ, 751, 94 5.3.1

Rowe, B. T. P., Jarvis, M., Mandelbaum, R., et al. 2015, Astronomy and Computing, 10, 121 3.2.2, 3.6, 2, 7

Rudie, G. C., Steidel, C. C., Trainor, R. F., et al. 2012, ApJ, 750, 67 1.2.3, 3.1.1, 5.1, 5.3.1

Ryan-Weber, E. V. 2006, MNRAS, 367, 1251 5.3.1

Santos, S., Sobral, D., Matthee, J., et al. 2020, MNRAS, 493, 141 3.1.1

Schlafly, E. F., & Finkbeiner, D. P. 2011, ApJ, 737, 103 2.2.5

Shi, D. D., Cai, Z., Fan, X., et al. 2021, ApJ, 915, 32 1.8, 1.3.2, 6, 7

Shi, K., Huang, Y., Lee, K.-S., et al. 2019, ApJ, 879, 9 5.3.1

Shibuya, T., Ouchi, M., Konno, A., et al. 2018, PASJ, 70, S14 3.1.1

Shimakawa, R., Kodama, T., Tadaki, K. I., et al. 2014, MNRAS, 441, L1 1.1.2

Shimakawa, R., Kodama, T., Shibuya, T., et al. 2017, MNRAS, 468, 1123 5.3.1

Shimizu, I., Todoroki, K., Yajima, H., & Nagamine, K. 2019, MNRAS, 484, 2632 4.8, 4.1.4, 4.9, 7

Smolčić, V., Ivezić, Ž., Knapp, G. R., et al. 2004, ApJ, 615, L141 1

Sobral, D., Matthee, J., Best, P., et al. 2017, MNRAS, 466, 1242 3.1.2

Spitler, L. R., Labbé, I., Glazebrook, K., et al. 2012, The Astrophysical Journal Letters, 748, L21 1.1.2

Springel, V. 2005, MNRAS, 364, 1105 1.1.2, 1.3.1, 4.1.4

Springel, V., Frenk, C. S., & White, S. D. M. 2006, Nature, 440, 1137 1.1.1, 1.2.1, 1.3, 7

Springel, V., Pakmor, R., Pillepich, A., et al. 2018, MNRAS, 475, 676 5.2

Suchyta, E., AleksicÌ, J., Huff, E., & MacCrann, N. 2021, Balrog: Astronomical image simulation, Astrophysics Source Code Library, record ascl:2107.009, ascl:2107.009 3.2.2, 3.6, 2, 7

Suzuki, N., Tytler, D., Kirkman, D., O'Meara, J. M., & Lubin, D. 2005, ApJ, 618, 592 2.1, 2.1

Tejos, N., Morris, S. L., Finn, C. W., et al. 2014, MNRAS, 437, 2017 5.3.1

Toshikawa, J., Kashikawa, N., Overzier, R., et al. 2016, ApJ, 826, 114 1.1.2, 5.3.1

Toshikawa, J., Uchiyama, H., Kashikawa, N., et al. 2018, PASJ, 70, S12 1.1.2

Tumlinson, J., Peeples, M. S., & Werk, J. K. 2017, ARA&A, 55, 389 1.1.2, 1.2.1, 5.3.1

Turner, M. L., Schaye, J., Crain, R. A., et al. 2017, MNRAS, 471, 690 5.3.1

Venemans, B. P., Röttgering, H. J. A., Miley, G. K., et al. 2005, A&A, 431, 793 1.4

Vergani, D., Garilli, B., Polletta, M., et al. 2018, A&A, 620, A193 3.2.1

Wang, T., Elbaz, D., Daddi, E., et al. 2016, ApJ, 828, 56 1.1.1

Wang, X., Li, Z., Cai, Z., et al. 2022, ApJ, 926, 70 1.8, 1.3.2, 7

Yajima, H., Abe, M., Khochfar, S., et al. 2022, MNRAS, 509, 4037 6

Yamada, T., Nakamura, Y., Matsuda, Y., et al. 2012, AJ, 143, 79 1.2.3

Yanny, B., Rockosi, C., Newberg, H. J., et al. 2009, AJ, 137, 4377 2.2.4, 3.1, 7

Zheng, X. Z., Cai, Z., An, F. X., Fan, X., & Shi, D. D. 2021, MNRAS, 500, 4354 1.3.2, 6

Zheng, Z.-Y., Malhotra, S., Rhoads, J. E., et al. 2016, ApJS, 226, 23 3.1.1, 3.1.2