Title
Cosmic Noon at 3D

Author
Qezlou, Mahdi

Publication Date
2024

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Cosmic Noon in 3D

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Physics

by

Mahdi Qezlou

March 2024

Dissertation Committee:

Dr. Simeon Bird, Chairperson
Dr. Bahram Mobasher
Dr. Anson D'Aloisio

The Dissertation of Mahdi Qezlou is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I would like to express my deepest gratitude to the amazing people whom without I would have not been able to get to this milestome of my life. I list them in chrocological order of their help and support.

First and foremost, I am deeply indebted to my parents for instilling in me the values of hard work and friendships. My sister deserves special recognition for igniting my passion for science from an early age. I am grateful to my high school teachers, Mohsen Iraji and Kazem Kookaram, for introducing me to the captivating realms of Physics and Astronomy.During my undergraduate studies, I was fortunate to have had the mentorship of Dr. Shant Baghram and Dr. Hessamaddin Arfaei, whose dedication to teaching the fundamentals of Physics left a lasting impression on me. I am particularly grateful to my advisor, Dr. Simeon Bird, whose mentorship transcended academic boundaries, offering both guidance and friendship. His unwavering support has been invaluable throughout my journey. My advisors at Carnegie Observatories, Dr. Andrew Newman and Dr. Gwen Rudie, have patiently mentored me on the art of observational Astrophysics and have been a great support during my PhD program. Lastly, I extend my heartfelt appreciation to all my friends from college and graduate school. Their camaraderie has been a source of immense joy and fulfillment, contributing significantly to my personal growth during this period.

To my parents, teachers, mentors and friends for all the support.

ABSTRACT OF THE DISSERTATION

Cosmic Noon in 3D

by

Mahdi Qezlou

Doctor of Philosophy, Graduate Program in Physics
University of California, Riverside, March 2024
Dr. Simeon Bird, Chairperson

The Ly-α forest has emerged as a potent observational tool for probing the gas dynamics within the Intergalactic Medium (IGM) during cosmic noon. Surveys such as BOSS and DESI, which analyze the forest within the spectra of background quasars, have recently provided unique insights into cosmology and relevant astrophysical processes, including HeII reionization. Conversely, Lyman-alpha tomography surveys enhance the density of background sources by observing the more abundant Lyman Break Galaxies (LBGs), furnishing a 3D map of neutral hydrogen at redshifts around z 2-3 with exceptional spatial resolution of approximately 2 cMpc/h. Furthermore, Line Intensity Mapping (LIM) represents a novel technique aimed at detecting the collective emission from all galaxies across the sky. In my dissertation talk, I will delve into the indispensable role of state-of-the-art cosmological hydrodynamic simulations in modeling these observations, constraining cosmology and elucidating the intricate connection between galaxies and their gaseous environment on IGM scales.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Cosmic Time

During the early epochs of the evolution of the universe, the gravitational pull from the matter in the universe overcomes the expansion of the universe, and the matter collapses into the first structures. Later on, the first stars and galaxies are formed, the distribution and composition of which tell us more about the universe and its evolution to date. During the first gravitational collapse of the matter, another counteracting force, the radiation pressure, created an acoustic wave in the baryon-photon fluid. This sound wave propagated through the universe until the universe had expanded enough to cool down, and the gravitational force overcomes the radiation pressure. This sound wave, known as Baryonic Acoustic Oscillation (BAO), left an imprint on the distribution of the matter in the universe [147,]. The size of this imprint is a standard ruler and can be used to measure the expansion history of the universe across cosmic time. Beyond the BAO signature, the clustering of the matter in the universe from the early epochs to the present time can be used to measure the growth of the structures in the universe and hence the nature of the dark energy and the dark matter. This clustering has been observed in the spatial correlation function (or power spectrum in Fourier space) of the signal from the first light observed in the universe after the atoms recombined and it became transparent to photons, i.e., Cosmic Microwave Background (CMB) [153,]. The clustering of the matter in the universe has also been observed in the spatial correlation function of the galaxies and the Ly-α forest. The latter is a forest of the absorption lines of the neutral hydrogen in the spectra of the background quasars (QSOs). Both galaxies and the gas in the Intergalactic Medium (IGM) are tracers of the underlying matter distribution; therefore, expected to have similar sensitivity to the cosmological parameters [36,].

Moreover, studying the evolution of galaxies and their connection to the surrounding gaseous ecosystem is an exciting area of research that can provide us with a better understanding of the universe, its content, and the important physical processes. Measuring the clustering of matter or understanding the important processes in galaxy evolution requires observing a large sample of galaxies across a wide area of the sky. In Section 1.2, I introduce the observational probes overcoming these challenges by surveying the galaxies and gas around them at cosmic noon, $z \sim 2$. In Section 1.3, I discuss how this period of cosmic history can help us to resolve the tensions in the cosmological parameters measured from the CMB and the low-redshift universe. In Section 1.4, I discuss how the cosmic noon observations inform us about the important astrophysical processes such as the HeII reionization and the galaxy-IGM connection.

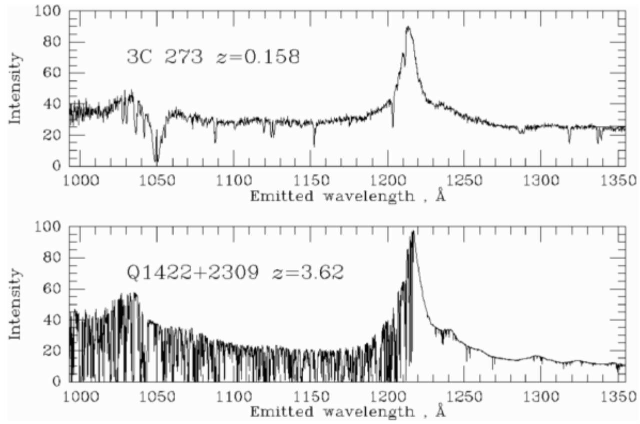


Figure 1.1: Credit: Edward Wright. Comparing the spectra of two QSOs, one in the nearby universe and the other at $z \sim 3.6$. The larger abundance of the HI at higher redshifts results in a denser Ly-α forest.

1.2 Probes at Cosmic Noon ($z \sim 2$)

1.2.1 Ly-α forest

The resonant absorption of light from background galaxies by the neutral hydrogen (HI) in the intergalactic medium (IGM) creates a forest of absorption lines in the spectrum. The most prominent of these HI absorption lines are the Ly-α forest [72,]. Figure 1.1 compares the spectrum of two QSOs, one in the nearby universe and the other at $z \sim 3.6$. The larger abundance of HI at higher redshifts results in a denser Ly-α forest. The Ly-α forest serves as a biased tracer of the matter distribution in the universe. This tracer has been identified as a powerful probe of the matter power spectrum at cosmic noon [36,]. It has been shown that the Ly-α forest power spectrum has strong constraining power on cosmological parameters such as the matter density Ω_m and the primordial power spectrum, $A_p(\sigma_8), n_p$ [120,]. Moreover, the importance of the ionizing background and the thermal history of the IGM in shaping the Ly-α forest has been demonstrated [200,]. The completed

SDSS QSO survey has already shown promising results in constraining such parameters [58,]. The ongoing DESI survey is expected to increase the number of observed QSOs and the spectral resolution, which will provide more robust statistics on the Ly-α forest power spectrum [28,].

Precise modeling of the Ly-α forest requires high-resolution hydrodynamical simulations to resolve the Jeans length of the IGM gas and incorporate important physical processes such as gas cooling and heating, and the ionization by local sources which cause major phase changes in the IGM, i.e., hydrogen and helium reionization. The current constraints show that hydrogen reionization has happened at redshifts $z > 5$ [209,]; however, residual large-scale temperature fluctuations affect the Ly-α forest observed at lower redshifts too [129,]. The HeII reionization happening at lower redshifts also has a significant impact on the Ly-α forest observations at cosmic noon [207,]. Incorporating all the important cosmological and astrophysical model parameters requires a suite of hydrodynamic simulations sampling the parameter space along with reliable interpolation schemes to avoid the need for full coverage of the parameter space [17, 59,].

The QSO spectra observed with SDSS have been very important in constraining cosmological parameters [58,]. However, QSOs are rare bright sources, and QSO surveys like SDSS [3,] and DESI [28,] sample the IGM with coarse spatial resolution along the transverse direction. The current generation of telescopes is capable of observing the Ly-α forest within the spectra of fainter star-forming galaxies, i.e., Lyman Break Galaxies(LBGs), and it has been shown that the signal-to-noise of these spectra is sufficient to make a 3D map of the

Ly-α forest [106,] at cosmic noon with megaparsec resolution. Such tomographic maps of the IGM are unique tools to provide high-resolution maps of the cosmic web and to study the relation between galaxies and their gaseous ecosystem at extended scales. The pioneering of such Ly-α tomography surveys is the CLAMATO project, mapping the intergalactic medium at $2.05 < z < 2.55$ across $\sim 0.2, \mathrm{deg}^2$ [77,]. The LATIS survey is another example of the Ly-α tomography surveys, which provides a 3D map of the IGM in a volume 13 times larger ($2.2 < z < 2.8$ and $1.7, \mathrm{deg}^2$). The upcoming tomography survey with PFS instrument is planned to increase the observed volume by another order of magnitude [71,]. A larger volume provides robust statistics on the overdense and underdense regions within the cosmic web. Of particular interest are the large matter overdensities, which are the progenitors of the galaxy clusters and groups at the present time. The high spatial resolution of the Ly-α tomography surveys is essential for reliable identification of these galaxy protoclusters. However, accurate theoretical modeling using hydrodynamical simulations is required for proper characterization of the protoclusters, i.e., their spatial extent and mass. In section 1.4, I will discuss the applications of these tomography maps in understanding the galaxy evolution at cosmic noon.

1.2.2 Line Intensity Mapping

The emerging technique of Line Intensity Mapping (LIM) is another promising probe to map the large-scale structure of the universe and to study the gaseous interstellar medium (ISM) and intergalactic medium (IGM) across cosmic time [201, 99, 10,]. LIM relies on measuring the integrated emission from the atomic and molecular lines of the gas

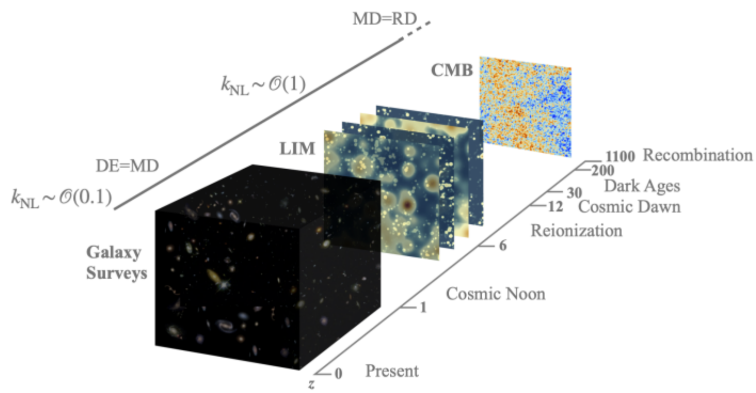


Figure 1.2: Credit: [10], Qualitatively comparing the available observation surveys across cosmic time. Line Intensity Mapping (LIM) surveys explore large volumes at redshifts inaccessible by the majority of galaxy surveys and early universe probes, e.g., the Cosmic Microwave Background (CMB).

without individually resolving the sources. This approach enables LIM to probe the large volumes accessible at higher redshifts as opposed to traditional galaxy surveys at lower redshifts, as shown in Figure 1.2. Starting from the 21-cm emission of neutral hydrogen, LIM has been extended to other lines such as CO, [CII], and Ly-α lines, particularly at cosmic noon. The ongoing pathfinder program with COMAP [33,] measures the CO(1-0) rotational line at $z = 2.4 - 3.4$ and other transitions at higher redshifts across $12.3, \text{deg}^2$. The EXCLAIM project [193,] will observe the [CII] emission lines at similar redshifts at cosmic noon over $305, \text{deg}^2$. The observing footprint of EXCLAIM is designed to overlap with the rich extragalactic data archives of SDSS, BOSS [104,], and HSC photometric surveys [4,]. Many other examples of LIM surveys observing across cosmic time are listed in this recent review [10,].

The interpretation of the signals observed in LIM surveys is challenging due to the low sensitivity of the detectors and the difficulty in disentangling the signal from interlopers. These interlopers consist of foreground contamination by extragalactic sources or emission from the Milky Way. To overcome this issue, one solution suggested for many LIM surveys is to cross-correlate the LIM signal with other tracers of the large-scale structure at the same redshifts. The utility of this approach has been shown in the context of cross-correlation with galaxy redshift surveys or other line emission surveys, e.g., 21-cm $\times$ galaxies [63,], Ly-α emission $\times$ galaxies [168,], [CII] $\times$ QSOs [160,]. However, it has been shown that the cross-correlation with galaxy surveys is limited by the redshift uncertainties and the mass completeness of the galaxy surveys [32,]. In Chapter 3, I will introduce a novel approach to cross-correlate the LIM signal with the Ly-α forest observed in Ly-α tomography surveys. We show that the Ly-α forest surveys, even the sparse ones which rely on observing only QSOs, significantly boost the signal-to-noise of the LIM surveys such as COMAP and EXCLAIM. Moreover, foreground contaminations have the least impact on the LIM $\times$ Ly-α forest compared to LIM $\times$ photometric galaxy surveys.

1.3 Cosmology at Cosmic Noon

The majority of the probes constraining cosmological parameters are carried out either at the early epochs of the universe, i.e., CMB, or at the present time, i.e., galaxy surveys. The tension between the cosmological parameters measured from the CMB and the low-redshift universe has been a long-standing problem in cosmology. The Hubble tension at 5.0σ, i.e., the discrepancy between the Hubble constant measured from the CMB and the local universe, has been a major problem in the standard cosmological model [95,]. Moreover, there is another inconsistency between the early and late universe measurements known as the "S_8 tension," although with smaller significance. The strength of matter clustering at $z = 0$, i.e., $S_8 = \sigma_8\sqrt{\Omega_m/0.3}$, can be inferred from the CMB observations, assuming the standard ΛCDM model to forward model the structure growth to the present time. On the other hand, the matter clustering at $z = 0$ can be directly measured from the galaxy surveys and the weak lensing measurements. Measuring the amplitude of the galaxy-galaxy weak lensing leads to a $10 - 15\%$ smaller value of S_8 compared to the CMB measurements. This $2-4\sigma$ tension is slightly less significant for the redshift-space clustering measurements [1,] and Figure 1.3.

Observations at an intermediate redshift range, the cosmic noon, play a crucial role in resolving these tensions. Currently, one of the cosmological constraints at these redshifts are from the Ly-α forest. The the 3D correlation function of the Ly-α forest and the cross-correlation with QSOs place precise constraints on the BAO features [43, 18,]. The BAO constrains the expansion rate of the universe at present, H_0. The H_0 infered from CMB measurments is in a 5σ tension with the direct measurements obtained through

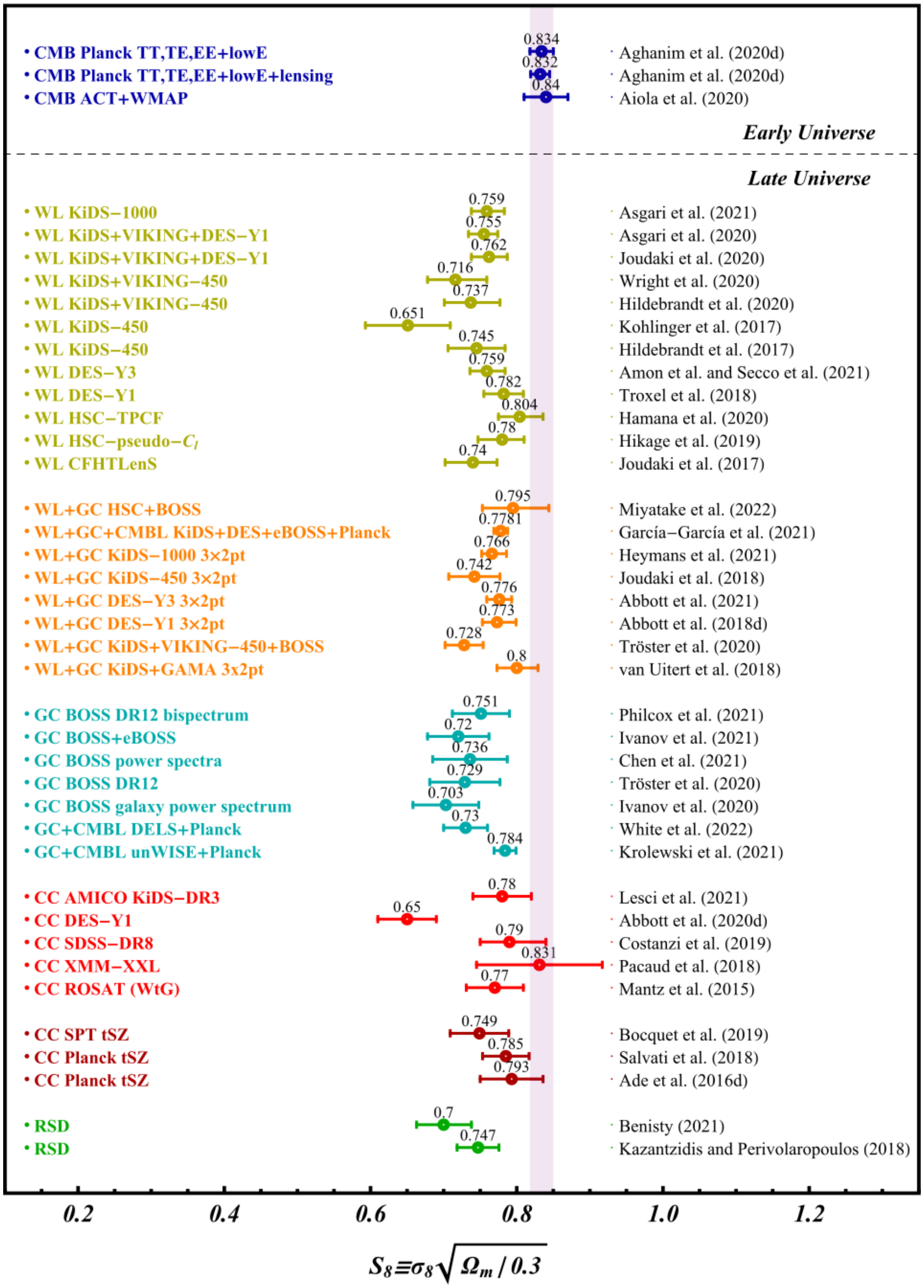


Figure 1.3: Credit: [1]. Listing the S_8 values inferred from the early universe experiments and the late-time estimates from weak lensing (WL), galaxy clustering (GC and RSD), cluster counts (CC).

strong gravitational lensing and the distance ladders at low redshifts, $z < 1$ [1,]. However, the constraints from the BAO measurements of the Ly-α forest at $z \sim 2$ and galaxies at $z < 1$, are closer to the Planck CMB measurements hinting to potential deviations from the ΛCDM model assuemd in all these measurements [38,]

The 1D Ly-α power spectrum of the BOSS QSOs at the same redshift place tight constraints on the amplitude, A_p, and the slope, n_p, of the primordial power spectrum. The clustering amplitude measured by the forest stays consistent with the lower redshift measurements, confirming the S_8 tension with the inferred values from CMB [58, 145,]. The comparison between the slope of the primordial power inferred from the Ly-α forest, sensitive to small-scale signals, and the CMB measurements which probe the power on larger scales will shed light on the scale-dependency of the matter power spectrum. A similar comparison on the amplitude of the power, A_p, can place constraints on the sum of the neutrino masses [146,]. In Chapter 4, I will discuss the promise of the 3D Ly-α forest power spectrum on nonlinear scales in improving the cosmological constraints. Our theoretical predictions indicate that the 3D Ly-α power will be a powerful probe of such parameters. Line Intensity Mapping (LIM) is another promising cosmological probe at cosmic noon. The large volume explored by the LIM surveys is ideal for constraining dark matter models [34,] and the infliationary models [132,].

1.4 Galaxy Evolution at Cosmic Noon

1.4.1 HeII reionization

The thermal state of the IGM at $2 \lesssim z \lesssim 5$ is shaped by the reionization of the second electron of helium, HeII [174,]. The HeII reionization was likely caused by the UV photons emitted from QSOs forming large-scale fluctuations in the temperature of the IGM [198,]. This change in the thermal state of the gas impacts the Ly-α forest statistics at cosmic noon, altering the 1D Ly-α power spectrum by a few percent and a larger predicted change on the 3D Ly-α forest power spectrum [122,]. Therefore, robust modeling of the patchy HeII reionization reduces any potential biases in the cosmological constraints from the Ly-α forest and provides a unique tool to characterize the ionizing sources.

The PRIYA simulation suite [13,] incorporates a flexible HeII reionization model developed by [196]. Using this suite to model the 1D Ly-α power spectrum observed with the BOSS QSO survey [25,] and the thermal history of the IGM from high-resolution QSO spectra [65,], [58] constrains the timing and the strength of the HeII reionization. Inspired by the success of this work, in Chapter 4, I will introduce an accurate emulator for the 3D Ly-α forest power spectrum and study its sensitivity to the HeII reionization model parameters.

1.4.2 Galaxy-IGM Relation with Ly-α Tomography

The link between galaxies and their gaseous ecosystem at extended distances, the intergalactic medium (IGM), can be precisely determined through the emerging Ly-

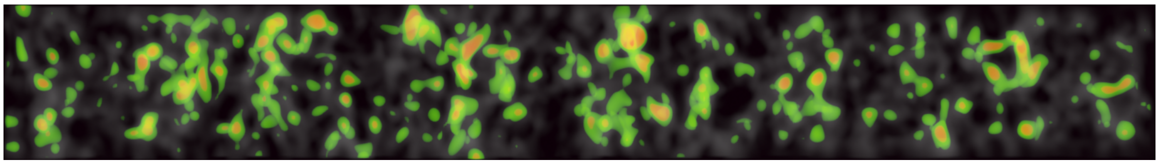


Figure 1.4: A subset of the Ly-α tomography maps observed in the LATIS survey [137,]. The color map represents the strength of the absorption, with red regions indicating the highest absorption. The dimensions of this map are $93 \times 51 \times 483\ h^{-1}$cMpc with the line of sight directions pointing to the right.

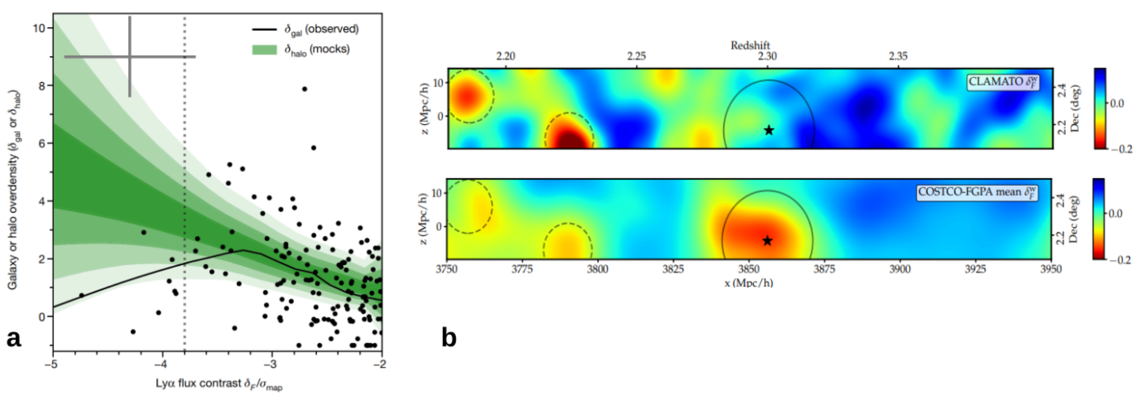


Figure 1.5: Galaxy Evolution with Ly-α Tomography: Anomalies in the Galaxy-Ly-α Relation as Tracers of Large-Scale Structures. **Panel a:** *Missing Galaxies in Opaque IGM Gas*. The displayed data points on the left side of the dashed curve show that most significant protoclusters detected with Ly-α absorption do not correspond to overdensities of observed galaxies, contrary to predictions based on halo overdensities in the mock-observed maps, the green contours (Figure from [139]). **Panel b:** *Transparent IGM around Galaxy Overdensities*. The star in the lower panel illustrates an overdensity of galaxies not visible in the upper panel's IGM absorption map (Figure from [47])

α tomography surveys. These surveys create a 3D map of the IGM gas with Mpc resolution by measuring the HI absorption (the Ly-α forest) in the spectra of densely packed background galaxies (Figure 1.4). In these tomography surveys, the Ly-α absorption in the intergalactic medium typically reflects the spatial distribution of galaxies on similar scales. Nevertheless, observational evidence suggests that this correlation occasionally deviates; instances of galaxy overdensities not evident in the Ly-α tomography signal [47,], or extensive absorbed regions with fewer observed galaxies than expected from the simulations [139,] have been found, refer to Figure 1.5. These anomalies in the galaxy-Ly-α forest correlation provide a unique opportunity to investigate galaxy evolution processes.

In Chapter 2, I discuss the accurate mock observations of the Ly-α tomography surveys through careful postprocessing of the cosmological simulations (TNG [135,] and ASTRID [16,]) while precisely incorporating the modeled observational uncertainties. Leveraging these precise mock observations, we have developed innovative tools for the identification of galaxy protoclusters during cosmic noon ($z \sim 2.5$) within Ly-α tomography maps [162,]. These evolving structures transition into massive galaxy clusters by $z = 0$, and our model accurately predicts their virialized mass. The comparison between the modeled galaxy protoclusters in our mocks and the observed population in the Ly-α tomography IMACS Survey (LATIS) has yielded a remarkable discovery [139,]. Notably, the largest galaxy protoclusters found in absorption, expected to correspond to large matter overdensities, host only a few observed galaxies, refer to the left panel of Figure 1.5. This observation implies that they have remained unnoticed as galaxy overdensities in traditional galaxy surveys. This discovery serves as a catalyst for further investigations into the intricate relation-

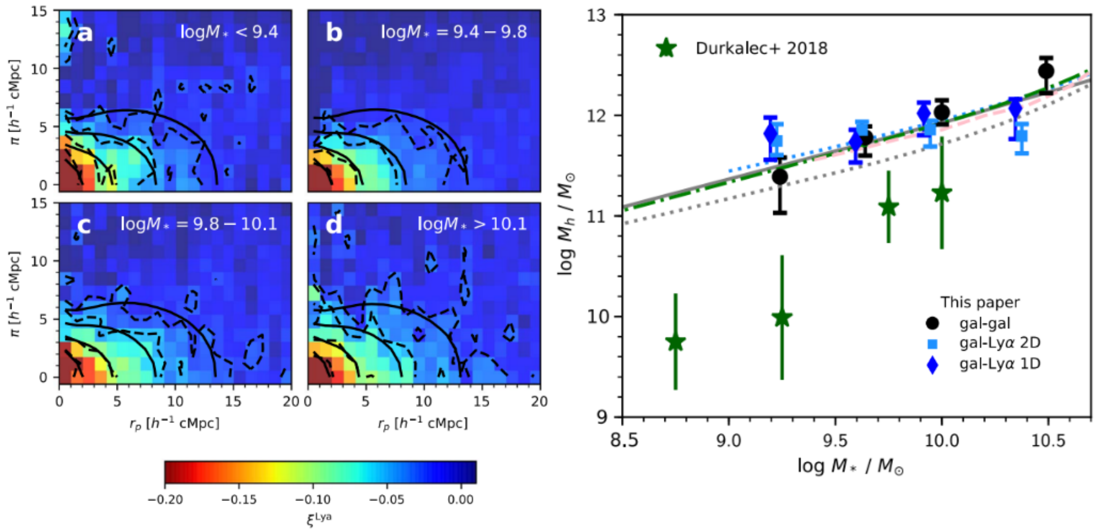


Figure 1.6: The Galaxy-halo connection inferred from LATIS Ly-α tomography. Figure from [136]. **Left Panel**: The cross-correlation between the Ly-α absorption and the distribution of Lyman Break Galaxies (LBGs) along (π) and transverse to the line of sight, r_p, binned into 4 stellar mass bins. The comparison between the observed (the colormaps and the dashed contours) and the simulated cross-correlation (the solid contours) confirms the reliability of our model in predicting the cross-correlation. **Right Panel**: The stellar-halo mass relation inferred from the cross-correlation compared to the theoretical models and the only other existing constraints at similar redshifts [51,].

ship between galaxy evolution and their surrounding environment. Our recently published paper [136,] centers on modeling the cross-correlation between Ly-α absorption and galaxy overdensities using the mock observations derived from the `ASTRID` simulation [161,]. This correlation, observed on intergalactic medium (IGM) scales, exhibits high sensitivity to the large-scale structure and primarily depends on the mass of the dark matter halo [131,]. As a result, it presents an innovative avenue for constraining the stellar-halo mass relationship for galaxies spanning the stellar mass range of 10^9 to $10^{11} M_\odot/h$ (or $M_h \sim 10^{11} - 10^{13} M_\odot/h$), refer to Figure 1.6.

1.5 Organisational Notes

This Thesis is structured as follows: In the following chapters, I listed the projects I led during my PhD program. The first two chapters of this thesis are published in peer-reviewed journals during the course of my PhD program. The last chapteris a draft in preparation to be submitted to the journals in the upcoming months.

Chapter 2: *Characterizing Protoclusters and Protogroups at* $z \sim 2.5$ *Using Ly-*α *Tomography*

This work was published in *The Astrophysical Journal*, [162,]. In this chapter, I discuss a novel method we developed to identify and characterize large matter overdensities at Cosmic Noon, which collapse into galaxy clusters and groups at the present time. This technique is vital for studying the connection between galaxy evolution and their environments.

Chapter 3: *Line Intensity Map and Ly-*α *Tomography*

This work was published as *Boosting Line Intensity Map Signal-to-Noise with the Ly-*α *Forest Cross-Correlation* in *MNRAS*, [161,]. In this chapter, for the first time, we explore the synergies between Ly-α tomography surveys and the upcoming Line Intensity Map (LIM) observations. We show that the cross-correlation between the Ly-α forest and the LIM can boost the signal-to-noise (S/N) of the Line Intensity by a factor of 10. This enhancement will facilitate the cosmological/Astrophysical studies with the LIM observations.

Chapter 4: *Cosmology/Astrophysics with 3D Ly-α Forest Power Spectrum*

This work is a draft planned to be expanded and published in a peer-reviewed journal. The authors will be M. Qezlou, S. Bird, M. Ho, and M. Fernandez, 2024, and is carried out in collaboration with the named co-authors.In this chapter, I will introduce an emulator for the 3D Ly-α forest power spectrum based on large hydrodynamical simulations. This work builds upon our previous paper in `PRIYA` collaboration [13,]. The robustness of this emulator is tested, and the cosmological/Astrophysical information encoded in the 3D Ly-α forest power spectrum is explored.

Chapter 2

Characterizing protoclusters and protogroups at $z \sim 2.5$ using Ly-α Tomography

Abstract

Ly-α tomography surveys have begun to produce three-dimensional (3D) maps of the intergalactic medium (IGM) opacity at $z \sim 2.5$ with Mpc resolution. These surveys provide an exciting new way to discover and characterize high-redshift overdensities, including the progenitors of today's massive groups and clusters of galaxies, known as protogroups and protoclusters. We use the IllustrisTNG-300 hydrodynamical simulation to build mock maps that realistically mimic those observed in the Ly-α Tomographic IMACS Survey (LATIS). We introduce a novel method for delineating the boundaries of structures detected in 3D

Ly-α flux maps by applying the watershed algorithm. We provide estimators for the dark matter masses of these structures (at $z \sim 2.5$), their descendant halo masses at $z = 0$, and the corresponding uncertainties. We also investigate the completeness of this method for the detection of protogroups and protoclusters. Compared to earlier work, we apply and characterize our method over a wider mass range that extends to massive protogroups. We also assess the widely used fluctuating Gunn-Peterson approximation (FGPA) applied to dark-matter-only simulations; we conclude while it is adequate for estimating the Ly-α absorption signal from moderate-to-massive protoclusters ($\gtrsim 10^{14.2}\, h^{-1}\mathrm{M}_{\odot}$), it artificially merges a minority of lower-mass structures with more massive neighbors. Our methods will be applied to current and future Ly-α tomography surveys to create catalogs of overdensities and study environment-dependent galaxy evolution in the Cosmic Noon era.

2.1 Introduction

The ancestors of massive galaxy groups and clusters are already evident as unvirialized large-scale structures at "cosmic noon," redshifts $z \sim 2-3$. These "protoclusters" and "protogroups" are moderate matter overdensities of $\delta_m = \rho/\bar{\rho} - 1 \approx 2-4$ that span $\sim 5-10\, h^{-1}$cMpc [30,] and form the nodes of the cosmic web. Locating and characterizing these overdensities is essential to understanding the onset of the relation between galaxy properties and large-scale environments. In the local universe, observations have established a correlation between galaxy properties, such as star formation and galaxy morphologies, and the local density [49, 159, 88, 148,]. Some recent studies suggest that these trends may persist to higher redshifts, $z \approx 2.5$ or even $z \approx 3.5$ [26, 112,]. However, the current sample of known protoclusters is small and may suffer from important biases; resolving them requires new methods (for a review see [143,]).

Unlike clusters at $z < 1.5$, protoclusters at $z \sim 2.5$ are unvirialized structures and have not yet developed sufficiently hot gas to be observed via the Sunyaev-Zel'dovich effect or X-ray observations. Thus the traditional approach for detecting protoclusters and groups at $z > 2$ is instead to search for overdensities of galaxies. However, using galaxies to locate protoclusters has some disadvantages. Most methods use a single type of galaxies to trace overdensities: sometimes typical star-forming galaxies, but often rare subtypes such as radio galaxies or extremely infrared-luminous systems [127, 154, 142,]. While such methods successfully identify overdensities, studying how galaxy evolution is affected by density is complicated when the galaxies used to locate overdense structures are required to have particular properties. Moreover, with current facilities, dense spectroscopic sampling of cosmological volumes is very challenging, while photometric surveys suffer from imprecise redshift estimates.

An exciting and complementary method for locating protoclusters and protogroups is Ly-α absorption, which is particularly advantageous as it does not depend on galaxies to trace the underlying density field. Rather, it relies on neutral hydrogen gas, which traces the underlying density on large scales with high fidelity [165, 121,]. Ly-α absorption has conventionally been observed in the spectra of bright background quasars and has enabled the detection of the H I reservoirs associated with several large-scale structures [**?**,]e.g.,][]MAMMOTH-I, MAMMOTH-II. However, protoclusters have a complex three-dimensional (3D) structure that cannot be fully characterized using 1D information from the spectra of one (or a few) bright quasars. Earlier theoretical works developed methods to optimally reconstruct the 3D structure of the intergalactic medium (IGM) from a set of spectra of closely spaced background sources [149,], a technique known as Ly-α (or IGM) tomography. By testing against numerical simulations, they proved these methods recover the cosmic web structure with a high fidelity [23, 111,]. On the observational side, [107,] showed that faint star-forming galaxies (Lyman-break galaxies, or LBGs) are suitable background sources due to their much higher abundance compared to QSOs. The current generation of telescopes is already sensitive enough to observe the Ly-α forest absorption within a large sample of these faint LBGs, sufficient to map the IGM in 3D with a resolution of few h^{-1}cMpc.

To date, there are only two tomography surveys dedicated to mapping the IGM with Mpc resolution at $z \sim 2.5$. The COSMOS Ly-α Mapping and Tomography (CLAMATO) project [109, 76,], the pioneering survey of this type, created the first Mpc-resolution large-scale structure map in the redshift range $z = 2.05 - 2.55$ comprising a survey volume of $4.1 \times 10^5 h^{-3}$ Mpc^3. This map led to the successful detection of a protocluster with an estimated mass of $9 \times 10^{13}\, h^{-1}\mathrm{M}_\odot$ at $z \sim 2.45$ within the COSMOS field [108,] along with a characterization of voids at the same redshift [100,]. These studies demonstrated the capacity of 3D Ly-α tomography as a systematic way to detect and study large-scale structures. A larger experiment, the Ly-α Tomography IMACS Survey (LATIS), is underway at the Magellan Baade telescope [138,]. LATIS achieves a similar, but slightly lower, sightline density to CLAMATO over a volume $\approx 10\times$ larger, sufficient to map a statistically representative sample of large-scale overdensities. [138,] detected 36 significant over- or under-dense regions within the redshift range of $z = 2.2$-2.8 using the first one-third of observations.

In this work, we are interested in developing methods to characterize the large set of structures that will be catalogued in the LATIS survey; however, our methods apply to other similar surveys. Using mock observations generated from cosmological hydrodynamical simulations, we improve on existing methods to detect and estimate the masses of structures detected in Ly-α flux maps. Previous works have mainly focused on characterizing the most prominent structures, i.e., protoclusters [188, 108,]. However, many significant overdensities will collapse into slightly less massive halos by $z = 0$, i.e., protogroups. We define protoclusters and protogroups to be the progenitors of $z = 0$ halos with masses $M_{\rm vir}/(h^{-1}{\rm M}_{\odot}) > 10^{14}$ and $10^{14} > M_{\rm vir}/(h^{-1}{\rm M}_{\odot}) > 10^{13.5}$, respectively. To maximize the utility of current surveys' tomographic maps, we extend earlier methods to enable the detection and characterization of the progenitors over a wider range of mass. We also introduce a new method to partition large structures composed of multiple substructures. This delineation is of great value for galaxy-environment relation studies.

To achieve a volume representative of the ongoing and near-future surveys, simulations of size over a few $100\,h^{-1}$cMpc are required. Previous works mainly employed dark-matter-only simulations to model the signal in the tomography maps and to tune models for detecting protoclusters [188, 108,] or voids [187, 100, 138,]. In this approach, the fluctuating Gunn-Peterson approximation (FGPA) is applied to the dark matter density to estimate the neutral hydrogen absorption. Occasionally, to assess the importance of hydrodynamical processes, smaller hydrodynamical simulations with or without feedback processes have been used (the Illustris simulation with a box size of $75\,h^{-1}$cMpc in [79,], and the Nyx simulation with a box size of $100\,h^{-1}$cMpc in [78,], [100,], and [114,]).

Simulations predict that some protoclusters host halos as massive as $\sim 10^{13.5}\,h^{-1}\mathrm{M}_{\odot}$; therefore, it is expected that energetic galactic feedback processes, particularly those associated with active galactic nuclei (AGN), will to some extent heat, ionize, and displace the gas within protoclusters. In this work, we take a step forward by using the state-of-the-art large hydrodynamical simulation TNG300-1, which incorporates detailed sub-grid models for feedback processes [150,]. We then compare our results with the FGPA prescription by applying it to the dark-matter-only counterpart of TNG300-1, which is identical except for its lack of treatment of hydrodynamical processes.

The organization of this paper is as follows. We describe the simulations and mock observations in section 2.2. In section 2.3, we introduce our method for detecting protoclusters/groups within the mock-observed tomography maps. We discuss the masses of these structures in section 2.4, including estimates of the structures' masses at $z = 2.5$ and the anticipated collapsed masses at $z = 0$. We conclude in section 2.5 by summarizing and discussing the applications of our findings.

2.2 Simulations and Mock Observations

2.2.1 Simulations

We use the state-of-the-art hydrodynamical simulation `TNG300-1` [185, 118, 133, 134, 151,], run with the `AREPO` code. The hydro solver is based on the moving mesh approach [182,], which evolves the gas particles on a moving mesh constructed from a Voronoi tessellation. The simulation box size is $205h^{-1}$ cMpc, and it contains 2500^3 dark matter particles and a similar number of gas cells, resulting in a dark matter particle mass of $5.9 \times 10^7\, h^{-1}\mathrm{M}_\odot$ and an average mass of $1.1 \times 10^7\, h^{-1}\mathrm{M}_\odot$ for the gas cells. The gravity is modeled using the `TreePM` algorithm with the gravitational force resolution set to 1 h^{-1}ckpc.

Supernovae and black hole (BH) feedback are constrained to reproduce observable properties of today's galaxy population such as star-formation histories, the stellar mass function, the stellar and gas versus halo mass relations, and the BH mass versus stellar mass relation. Due to the large spatial scales of interest in our work (larger than ≈ 1 physical Mpc, or pMpc), the most relevant type of feedback is the energetic feedback from BH accretion. BH feedback is implemented as thermal energy injection in the high-accretion mode and kinetic winds in the low-accretion mode [204,]. The simulations do not solve the radiative transfer equations; therefore, the local photo-ionization in the proximity zone of AGN is not modelled. However, local heating due to radiative BH feedback is enforced by modifying the cooling rates (section 2.6.4 in [202,]). To constrain the importance of hydrodynamical processes and galaxy feedback, we also apply the FGPA to a dark-matter-only companion run of `TNG300-1`, which has the same initial conditions and number of dark matter particles.

2.2.2 Simulated IGM absorption

We create mock-observed spectra by modelling the absorption from gas particles in the full-hydro `TNG300-1` simulation. The spectra are generated using the `fake_spectra`[1] package [14, 11,]. In this work, we added MPI support, allowing the analysis to scale to many nodes, and make it possible to post-process large cosmological simulations. Briefly, `fake_spectra` calculates the absorption spectra for any ion in the simulation, including the neutral hydrogen used in this work, along thousands of sightlines in the simulation box. Each particle is approximated as an individual absorber giving rise to absorption characterized by a Voigt profile. The internal physical quantities of these cells are smoothed by the kernel appropriate for the simulation. The supported interpolation kernels are SPH, top-hat, and partial Voronoi reconstruction. For the `Arepo` simulations, we used a top-hat kernel with the same volume as the Voronoi cell for each particle. This approximation has been tested to reproduce the spectra created by a full Voronoi mesh reconstruction for redshifts $z < 5$ [208,]. The neutral hydrogen fraction is obtained by assuming ionization equilibrium in the presence of a uniform ultraviolet background (UVB). We solve the collisional/photo-ionization and recombination rate network equations provided in [86,]. The self-shielding of dense neutral hydrogen is taken into account by equation A.8 in [163,], which is a fitting formula to a radiative transfer model. The assumed uniform photoionization rate dictates the mean absorbed flux in the simulated spectra. We rescale the optical depth in the spectra to enforce the empirical mean flux evolution, corrected for metal absorption [55]:

$$\bar{F} = \exp\left(-1.330 \times 10^{-3} \times (1+z)^{4.094}\right) \tag{2.1}$$

[1] GitHub: *fake_spectra*

The simulated spectra extend along the third dimension of the snapshot with a pixel width of $\approx 6.4\ km\ s^{-1}$ to capture the effect of saturation. Once the high-resolution spectrum is extracted, the absorption is averaged over every 20 adjacent pixels to mimic the detector's pixel size in the LATIS ($\approx 129\ km\ s^{-1}$). Further, we mimic the spectral broadening in LATIS data by smoothing each spectrum with a Gaussian kernel of $\sigma = 2.06$ Å in the observed frame.

To constrain the importance of simulated baryonic processes, we also apply the FGPA to a DM-only companion simulation with identical initial conditions to `TNG300-1`. This DM-only companion evolves cold dark matter (CDM) particles and baryons as a single collisionless fluid; therefore, it does not capture galaxy feedback effects. In this approach, by applying cloud-in-cell (CIC) interpolation, the dark matter density and line-of-sight peculiar velocity fields are deposited onto a uniform grid of length $\sqrt[3]{N_{\rm DM}}$, where $N_{\rm DM} = 2500^3$ is the number of dark matter particles in the simulation. We use the efficient NBODYKIT[2] code [74,] for this step. To mimic the baryonic pressure smoothing and estimate the baryonic density with this approach, it is often advised to smooth the DM density and velocity fields [180,]. However, we verified that for simulating the neutral hydrogen absorption on scales of 1 pMpc, this initial smoothing is unimportant, confirming the assumption in previous works [138, 188,]. The neutral hydrogen density is related to this approximate baryonic density by assuming ionization equilibrium with a uniform UVB. The recombination and photo-heating rates are functions of temperature, which in turn is a function of density. We connect density to temperature using the slope of the power-law measured by [172,]. After including the Hubble flow and peculiar velocity of the cells, we obtain the optical depth

[2] GitHub: Nvodykit

of the neutral hydrogen along the line-of-sight by convolving the density with a thermal Doppler broadening profile. Finally, we scale this optical depth to match the observed mean flux by [55,]. The `Python` implementation is accessible through our repository[3].

Moreover, in Appendix 2.6 we show both simulated full-hydro TNG300-1 and the FGPA method recover the 1D flux power spectra with a $\sim 10\%$ accuracy when compared to SDSS DR14 results[24,] .

2.2.3 Realistic mock spectra

From these noiseless spectra, we produce a set of realistic mock spectra that match the areal density of sightlines in LATIS and their observational noise, which are the key parameters determining the quality of the derived tomographic maps [188, 138,]. As a representative value, we consider the average sightline density in the COSMOS field (now fully observed) in the redshift range $z = 2.4 - 2.6$ (the middle of the survey), which is $n = 0.16\ h^2\ \mathrm{cMpc}^{-2}$. This corresponds to a mean sightline separation of $\langle d_\perp \rangle = n^{-1/2} \approx 2.70\ h^{-1}\mathrm{cMpc}$ and a total of 6739 mock spectra within the TNG300-1 simulation box. The variations due to redshift dependencies of these parameters within the observed range of LATIS is discussed in Appendix 2.8.

We randomly select these sightlines and assign each a continuum-to-noise ratio (CNR) drawn from the distribution in LATIS [138,]. The CNR distribution is different for Lyman-break galaxies (LBGs) and QSOs, which constitute 98% and 2% of sightlines, respectively. In the mock observations, we randomly divide the sightlines into these two

[3] [10.5281/zenodo.6130050]

Table 2.1: The continuum-to-noise ratio (CNR) distribution of LBGs and QSOs, parameterized as log-normal distributions. The mean and standard deviation of ln(CNR) are listed.

Source	Mean	σ
LBGs	$0.84 + 0.99(\mathrm{z} - 2.5) - 1.82(\mathrm{z} - 2.5)^2$	0.43
QSOs	2.3	1.2

populations. The CNR is modelled as a log-normal distribution with a constant mean and variance for QSO background sources, while for LBGs, the variation of the mean with redshift is modeled with a second-degree polynomial. These equations are summarized in Table 2.1. The error propagated by continuum fitting is also implemented as prescribed by [100,]:

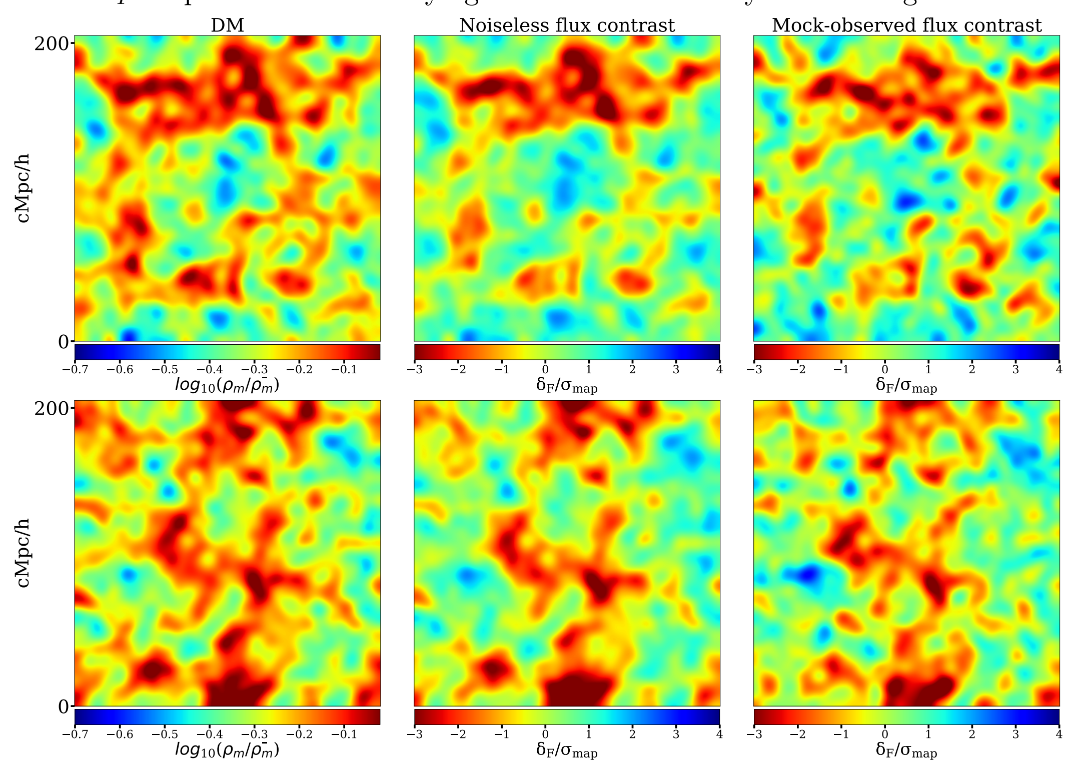
$$F_{\mathrm{obs}} = \frac{F_{\mathrm{sim}}}{1 + \delta_{\mathrm{cont}}}, \tag{2.2}$$

where δ_{cont}, which is constant along each sightline, is a realization of a normal distribution $\mathcal{N}(0, \sigma_{\mathrm{cont}})$ with $\sigma_{\mathrm{cont}} = 0.24 \times \mathrm{CNR}^{0.86}$ [138,]. Finally, we follow the prescription of [138,] and remove lines with an equivalent width EW $>$ 5 Å, which is intended to suppress damped absorption lines. We emphasize that rather than masking such lines explicitly based on the total neutral hydrogen column density along the sightline, we closely mimic the LATIS procedure by applying it to the mock spectra.

2.2.4 Generating 3D flux maps

To generate the 3D Ly-α mock absorption map, we first create a random set of mock-observed spectra at $z = 2.45$, close to the mid-redshift of the LATIS survey. In Appendix 2.8, we discuss our results for slightly different redshifts within the observed

Figure 2.1: Visualizations of two different slices (top and bottom rows) of the IllustrisTNG-300 hydro simulation at $z = 2.45$. All maps are smoothed by a gaussian with kernel size $\sigma_{\rm sm} = 4\ h^{-1}$ cMpc. Left column: The smoothed DM density, $\rho_{\rm DM}/\overline{\rho_{\rm DM}}$. Middle column: The smoothed noiseless Ly-α flux contrast $\delta_F/\sigma_{\rm map}$. Right column: The mock-observed flux contrast, constructed by Wiener interpolation of spectra that mimic LATIS observations. The mock δ_F maps trace the underlying DM distribution very well on large scales.



range in LATIS. However, the remainder of the paper is based on the snapshot at $z = 2.45$ (hereafter abbreviated $z = 2.5$). The quantity of interest is the flux contrast δ_F, defined as:

$$\delta_F = \frac{F}{\bar{F}} - 1, \tag{2.3}$$

where $\bar{F}$ is the mean flux. The flux contrasts in the mock spectra are interpolated to make a 3D map of δ_F using a Wiener filter. The Wiener filter is an optimal linear estimator of the 3D δ_F map from a set of observed noisy spectra and specified noise and signal covariance matrices. It requires the noise to be additive, hence the data vector is $\mathbf{d} = \mathbf{s}+\mathbf{n}$, where $\mathbf{s}$ and $\mathbf{n}$ are the vectors of signal and noise, respectively. By minimizing the mean squared error between the linear estimate, $\hat{s} = \mathbf{L}\mathbf{d}$, and the true signal, one obtains $\mathbf{L} = \mathbf{C}_{mp}(\mathbf{C}_{pp}+\mathbf{N})^{-1}$. $\mathbf{C}_{mp}$ and $\mathbf{C}_{pp}$ are the assumed signal covariance matrices, where the m and p indices indicate voxels in the 3D map and pixels along along the sightlines, respectively. $\mathbf{N}$ is the noise covariance, which is assumed to be diagonal. Following the *ad hoc* assumption in [23,] (also [188,]), we set $\mathbf{C}_{pp} = \mathbf{C}_{mp} = \mathbf{C}(\vec{r}_1, \vec{r}_2)$ as a Gaussian:

$$\mathbf{C}(\vec{r}_1, \vec{r}_2) = \sigma_F^2 \exp\left[-\frac{\Delta r_{||}^2}{2\sigma_{||}^2} - \frac{\Delta r_{\perp}^2}{2\sigma_{\perp}^2}\right], \tag{2.4}$$

where $\Delta r_{||}$ and $\Delta r_{\perp}$ are the components of $\vec{r}_1 - \vec{r}_2$ parallel and perpendicular, respectively, to the line of sight. The amplitude is set to $\sigma_F^2 = 0.05$, and the correlation lengths are fixed at $\sigma_{\perp} = 2.5\ h^{-1}$cMpc and $\sigma_{||}^2 = \sigma_{\perp}^2 - \sigma_{\rm inst}^2$, where $\sigma_{\rm inst} = 1.4\, h^{-1}$cMpc is the instrumental resolution of LATIS at $z = 2.5$. The diagonal entries of the noise covariance matrix, $\mathbf{N}$, are sourced by contributions from both random noise and continuum fitting error, $\sigma_\delta =$

$(\sigma_{\rm rand}^2 + \sigma_{\rm cont}^2)^{1/2}$, including a floor of $\sigma_\delta > 0.2$ on the total noise. All these parameters match those applied to the LATIS observations by [138,]. We use DACHSHUND[4], an efficient Wiener filtering implementation by [188,].

Figure 2.1 visually compares the quality of the flux map recovered from the realistic mock observations to that of a noiseless mock map and the underlying dark matter density field. The noiseless mock map (middle column) is produced by generating spectra on a regular grid with a fine spacing of 1 h^{-1}cMpc and adding no noise; the noiseless map is not Wiener filtered. All maps are smoothed with a Gaussian kernel with length of $\sigma = 4\ h^{-1}$ cMpc, the typical scales of protoclusters at $z = 2.5$ [30, 188,]. We follow the convention in previous works of normalizing δ_F by its standard deviation $\sigma_{\rm map}$ over the map volume; note that $\sigma_{\rm map}$ is not the measurement error. The visual similarity among these maps and the voxel-by-voxel comparison in the bottom panel of Figure 2.2 confirm the quality of the flux map recovered on Mpc scales.

The optimal reconstruction of the 3D absorption field from a discrete set of noisy spectra is an active field of research. [149,] compares a constrained Gaussian random field linear approach (or Wiener filter) with a general non-linear explicit Bayesian deconvolution method. They show that for low-resolution spectra with negligible measurement errors, and under the *ad hoc* assumption of Gaussian distributed flux, Wiener filtering produces the maximum *a posteriori* map of the Bayesian method. Using new optimization tools, [114,] loosens the constraint that the flux correlation be Gaussian and minimizes a loss function for a more general signal. In parallel, some other interesting methods under development aim at simultaneous reconstruction of matter density and velocity fields via reconstructing

[4] GitHub: caseywstark/dachshund

the likeliest initial conditions [158, 78, 79,]. However, in this work, we adopt the Wiener filter approach for reconstructing the Ly-α flux field, since it is used by current generation surveys, including LATIS. We defer evaluating other methods to future work.

Flux pdf

In order to verify that the δ_F fluctuations in the mock-observed maps are consistent with those seen in LATIS, we compared the probability distribution functions (pdfs). For this purpose we use the LATIS map in the COSMOS field, updated from the version presented in [138,] to include the fully observed volume. To account for cosmic variance, the mock map is cut into a few sub-boxes, each with roughly the same volume as LATIS. Figure 2.3 shows the flux pdf in LATIS (z = 2.2-2.8, dashed black) and mock-observed maps ($z = 2.45$, orange stripes enclosing 68% and 95% of the sub-boxes) for three different smoothing scales. The agreement is excellent. We verified that this agreement holds when the LATIS pdf is computed in several narrower redshift bins (z = 2.3-2.5, 2.4-2.5, 2.5-2.8) and compared to the nearest TNG snapshot. Figure 2.3 demonstrates the overall agreement between the mock-observed signal predicted by ΛCDM and observed in LATIS.

Correlation with the dark matter density

Quantifying the correlation between the mock-observed signal δ_F and the underlying density field on a voxel-by-voxel basis is an essential first step toward estimating the masses of structures [108,]. Since the flux map, which will be used to define the boundaries of structures, is reconstructed in velocity space, we computed the density of dark matter

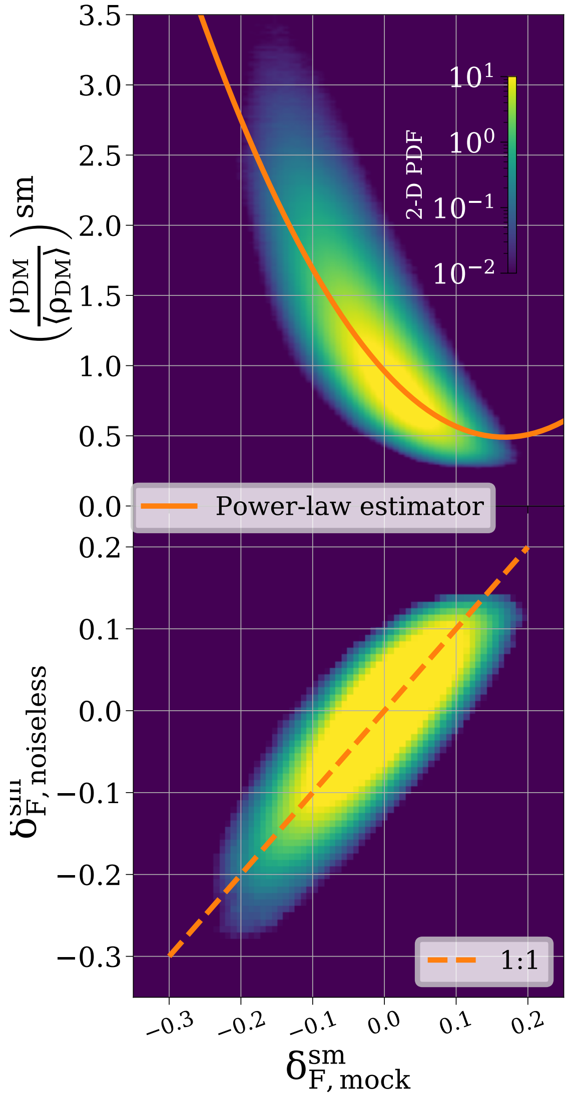
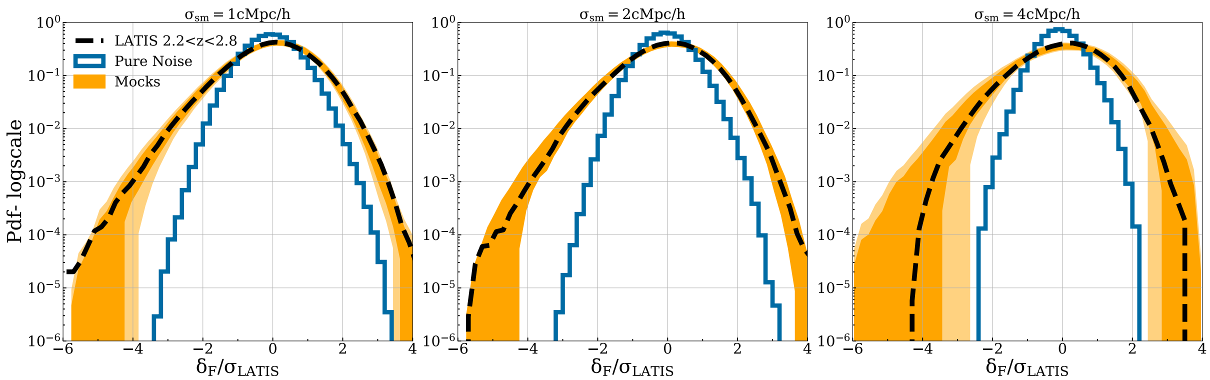


Figure 2.2: **Upper panel:** The median correlation between the smoothed ($\sigma_{\rm sm} = 4\ h^{-1}$ cMpc) DM density and the mock-observed flux contrast $\delta^{\rm sm}_{\rm F,mock}$ on a voxel-by-voxel basis in 20 mock maps at $z = 2.5$. The solid line is the median second-degree polynomial estimator (eq. 2.5), which we use to estimate the DM mass within structures detected in the tomographic maps. **Lower panel:** A voxel-by-voxel comparison between the noiseless and mock-observed flux contrasts, smoothed as in the upper panel. In both panels, median 2D histograms with logarithmic color scales are displayed.

Figure 2.3: The normalized distribution of the flux contrast, smoothed on three different scales, in LATIS and the mock maps. We normalize δ_F by its standard deviation $\sigma_{\rm LATIS}$ in the LATIS map. The black dashed lines show the pdfs of the observed map, and the orange stripes show those of mock-observed maps at $z = 2.45$. The width of the stripes represents the cosmic variance within 68% and 95% confidence levels. The solid blue curves show the pdfs of pure noise maps generated by Wiener filtering a set of pure noise spectra (see section 2.3.1). These plots confirm the accuracy of the mock maps in reproducing the basic statistics of the LATIS maps.



particles in velocity space. Specifically, the particles were displaced by their line-of-sight peculiar velocities and then interpolated with a CIC kernel onto a regular mesh grid with a resolution of 1 h^{-1}cMpc. The top panel in Figure 2.2 illustrates the correlation between these two fields, $\delta_F^{\rm sm}$ and $\rho_{\rm DM}^{\rm sm}$, both smoothed by a $\sigma_{\rm sm} = 4\,h^{-1}$cMpc Gaussian at $z = 2.5$. The solid curve is the median second-degree polynomial fit among 20 mock-observed maps. This estimator, along with its uncertainties, is

$$\begin{aligned} \left(\frac{\rho_{\rm DM}}{\langle\rho_{\rm DM}\rangle}\right)^{\rm sm}(z=2.5) &= 16.78(\pm 0.53)\ \delta_F^{\rm sm\ 2} \\ &- 5.62(\pm 0.07)\ \delta_F^{\rm sm} \\ &+ 0.96(\pm 0.01). \end{aligned} \tag{2.5}$$

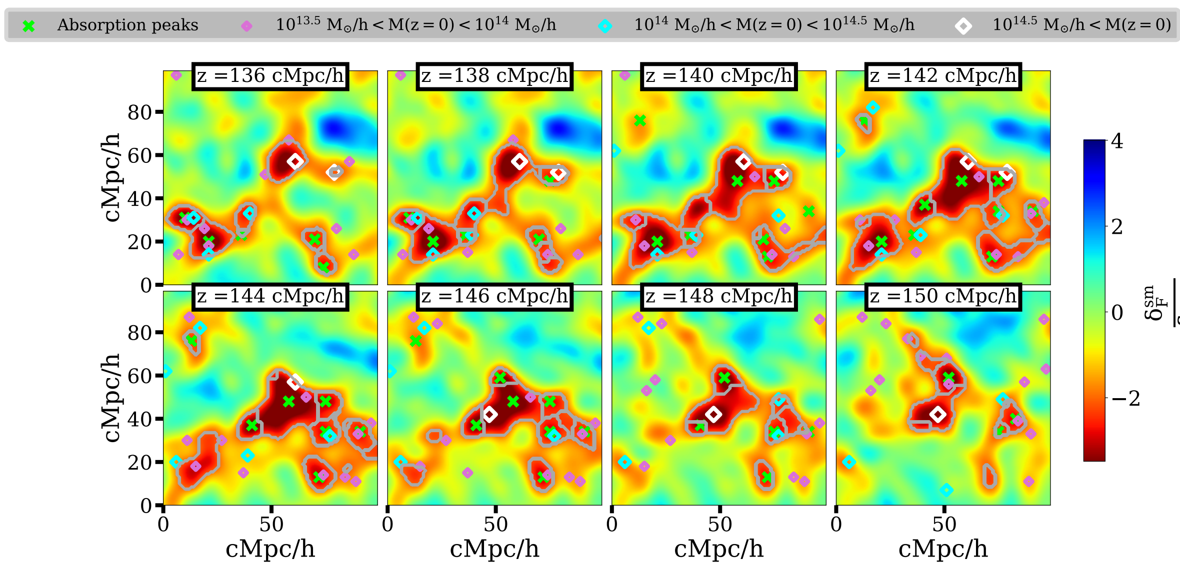
In Appendix 2.8 we show that although this relation evolves slightly within the redshift range of LATIS, its evolution is negligible for the purposes of this paper.

2.3 Characterizing structures in Lyman-α tomographic maps

Ly-α tomography provides continuous maps of the high-redshift H I absorption field and, via the correlation shown in the top panel of Figure 2.2, of the matter density field. However, it is often useful to identify and characterize discrete structures in these maps. Dark matter halos at $z \approx 2.5$ do not provide a functional route to define these structures, since due to hierarchical structure formation, the resolution of the maps is much coarser than the size of even the most massive halos yet formed. Instead, most previous works have focused on the volumes that will collapse into massive halos at $z = 0$, which are generally called protoclusters when the descendant mass $M_{z=0} > 10^{14}\, h^{-1}\mathrm{M}_{\odot}$. This perspective has the merit of supplying a rigorous way to define high-redshift "structures," but the connection between these structures and the observed $z = 2.5$ Ly-α field is complex due to the structures' varied assembly histories. In other words, variations in the kinematics and density structure of subcomponents within detected overdensities affect their evolution such that not all absorption-detected structures will form a galaxy cluster by $z = 0$, and not all galaxy clusters will have a progenitor detectable in absorption at $z = 2.5$. Furthermore, estimating density and environment at high redshifts, not only the structures' future evolution, is important for understanding the impact of environment on galaxy evolution.

In this section, we describe a method for identifying structures directly in Ly-α tomographic maps. The method extends previous techniques [108,] in two ways. First,

Figure 2.4: Illustration of the watersheds identified in a portion of the mock-observed map and their connection to the progenitors of massive halos at $z = 0$. The x, y and z coordinates are those of the simulation box. Crosses mark the absorption peaks in the mock-observed map and diamonds represent the centers of mass of the progenitors of massive halos at $z = 0$. For better visibility, these markers are copied along $\Delta z = \pm 4\, h^{-1}$cMpc of the actual location. The size and color of diamonds indicate the $z = 0$ halo mass. Silver contours illustrate the watersheds constructed using our method (section 2.3.2), which better isolates progenitors in highly-blended regions by using the gradient information in the flux map.



we include a larger selection of structures that extends to lower masses. Second, we introduce a method to delineate boundaries between overlapping structures. These improvements will aid future galaxy evolution studies by providing both larger numbers of structures and a simple way to associate galaxies with them. In Section 4, we will consider the masses of these structures and their descendants.

2.3.1 Defining detected structures as flux islands

Our basic unit for a structure in a smoothed tomographic map is a contiguous region in which the flux is below a threshold $\delta_F/\sigma_{\text{map}} < \nu$, which we call an "island". By default, we consider maps smoothed with a gaussian $\sigma_{\rm sm} = 4\,h^{-1}$cMpc since the map S/N degrades significantly when less smoothing is applied (see Figure 22 in [138,]). [188,] showed that this smoothing scale is nearly optimal for the detection of protoclusters. In Appendix 2.9, we show that some lower-mass structures leave significant signatures only in less-smoothed maps, but due to larger uncertainties in characterizing those structures, we will focus only on the $\sigma_{\rm sm} = 4\,h^{-1}$cMpc smoothed maps for the rest of this paper.

To determine the optimal value of ν, we need to balance completeness, which favors a high ν (less negative and less absorption) to include more structures, with purity, which favors a low ν (more significant and more absorption) to suppress the influence of noise. To estimate the level of noise in the maps, we generated a pure noise map by mock observing a structureless field (i.e., adding noise to $\delta_F = 0$ to create the spectra, then Wiener filtering as described in section 2.2.4). Figure 2.3 shows the flux pdf of the voxels in the mock-observed and the pure noise maps. The ratio of the two distributions at $\nu \approx -2$ is ≈ 0.01, indicating that this threshold selects voxels where absorption ($\delta_F < 0$) is present at $> 99\%$ confidence. Furthermore, we will show in the next section that, after appropriate subdivision, these islands are directly connected to the descendant massive halos at $z = 0$ only for a small range of ν, and any deviation from this leads to either the emergence of unrealistically large structures or missing some significant structures in the map.

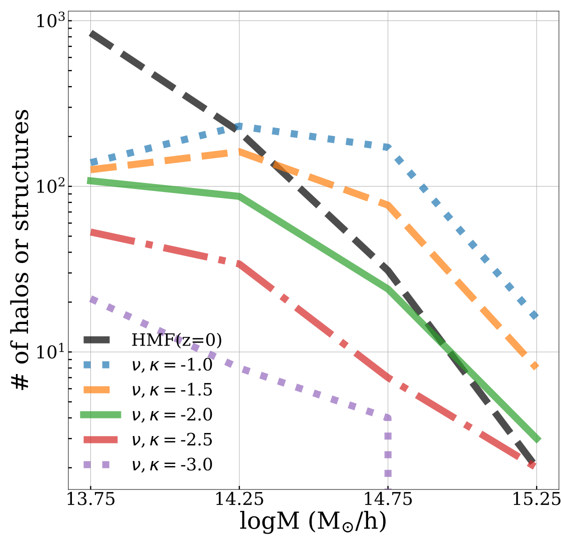


Figure 2.5: The distribution of the watersheds' tomographic masses $M_{\mathrm{tomo,raw}}$, for several choices of the contour level ν, is compared to the halo mass function at $z = 0$. For this comparison we define watersheds using all the absorption peaks and hence set $\kappa = \nu$. The tomographic masses are well matched to the masses of protoclusters when $\nu = -2$, which motivates our choice of this parameter.

2.3.2 Deblending substructures in flux islands with the watershed algorithm

Some flux islands extend over very large volumes, which may contain substructures (i.e., multiple flux minima) that will often collapse into several distinct halos. In these situations, it is useful to consider the island as being composed of multiple structures. Here we present a method to partition flux islands into smaller blocks, which we call "watersheds", each of which is associated with a distinct flux minimum. We will then analyze the connection between these watersheds and the progenitors of massive halos at $z = 0$.

Figure 2.4 shows a slice through our mock map at the location of a large contiguous region with $\delta_F/\sigma_{\rm map} \leq -2$ that contains several local minima. Diamonds show the centers of massive protogroups and protoclusters. These are computed by tracing back the DM particles located within R_{200} of a massive halo ($M_{\rm vir} > 10^{13.5}\,h^{-1}{\rm M}_{\odot}$) at $z = 0$; the center is then identified as the center of mass of these particles. As expected, the progenitors of massive halos cluster strongly; this flux island contains the progenitors of several massive halos. The flux minima do not always have a one-to-one correspondence with massive halo progenitors, both because of observational noise and the fact that the progenitors are blended at the map resolution. Thus, we do not expect that it will be possible to cleanly deblend all massive halo progenitors in our maps. Nonetheless, Figure 2.4 makes it clear that dividing islands into watersheds associated with local flux minima could provide a simple and effective way to deblend many of the massive protogroups and protoclusters.

To divide the flux islands into smaller blocks, we use the watershed algorithm [199,], hence naming the blocks "watersheds". This algorithm has been used previously in other cosmological applications [**?**,]e.g., for detecting voids,][]WVF-voide-detector. The watershed algorithm floods basins from a set of markers until basins associated with different markers meet on the watershed line. We flood flux islands and take a subset of the local flux minima as the markers. The subset is defined by a threshold on the flux contrast $\delta_F/\sigma_{\rm map} \leq \kappa$. Both κ and ν are selectable parameters in our method.

In Section 2.3.1, we motivated $\nu \approx -2$ based on the noise level in the mock-observed maps. An additional desirable property of the watersheds is that they approximately enclose the Lagrangian extent of the progenitors of $z = 0$ massive halos. To investigate this, we varied ν and subdivided islands into watersheds, considering all the local minima within each island, i.e., setting $\kappa = \nu$. We then integrated the estimated dark matter density $\rho_{\rm DM}$, as estimated from the flux contrast δ_F via equation 2.5, over each watershed's volume W to form a raw tomographic mass:

$$M_{\rm tomo,raw} = \int_W \left(\frac{\rho_{\rm DM}}{\langle\rho_{\rm DM}\rangle}\right)^{\rm sm} (\delta_F) \times \langle\rho_{\rm DM}\rangle(z)\, dV. \tag{2.6}$$

We refer to this as a "raw" tomographic mass to distinguish it from the calibrated mass $M_{\rm tomo}$ that we will define in Section 2.4. Figure 2.5 compares the distribution of $M_{\rm tomo,raw}$ to the halo mass function at $z = 0$. For $\nu > -2$, the watersheds become very large and contain much more mass than equally numerous halos at $z = 0$. On the other hand, setting $\nu < -2$ produces islands that do not contain all of the mass that will collapse. In addition to underestimating the volumes of protogroups and protoclusters, choosing $\nu < -2$ will also exclude some structures that produce significant absorption in the maps. Therefore, we define flux islands as regions with $\delta_F/\sigma_{\rm map} \leq \nu = -2$.

In order for the volume and mass of an island to be well defined, the absorption must reach a minimum depth below the contour level ν used to define the island. Otherwise, the volume can diverge to 0, and the uncertainties in the mass estimates become too large to be useful. We impose this margin by requiring flux islands to have a local minimum with $\delta_F/\sigma_{\rm map} \leq \kappa$. When an island has multiple such minima, we use only the minima with $\delta_F/\sigma_{\rm map} \leq \kappa$ as "markers" in the watershed algorithm. Thus, each watershed is associated with exactly one such local minimum.

In order to set the value of κ, we minimize the error in the tomographic masses of the watersheds. We compute $M_{\rm tomo,raw}$ over the watersheds obtained using $\nu = -2$ and several values for κ. The error is then defined as the difference between counterpart watersheds in the mock-observed and the noiseless flux maps. The counterpart of a mock-observed watershed is the one in the noiseless map with the largest overlap.[5] Finally, for a given value of κ, we calculate the rms of these errors over all watersheds. Figure shows the rms as a function of κ. Choosing $\kappa < \nu = -2$ reduces the rms mass error considerably. But as κ is progressively lowered, the improvement in the mass error diminishes, and more structures are excluded from the analysis. Balancing these considerations, we select $\kappa = -2.35$. Coincidentally, this value matches the threshold used by [138,] to select voxels with a 95% probability of lying in the bottom 10% of the δ_F distribution.

A visual example of the watersheds produced by this method is shown in Figure 2.4. The centers of mass of the progenitors of massive $z = 0$ halos are indicated with diamonds,

[5] The amplitude of the fluctuations in the mock-observed and noiseless flux maps are slightly different. In the mock-observed map, it is regulated by the signal covariance function supplied to the Wiener filter, σ_F in eq. 2.4; therefore, it is necessary to first scale the noiseless map to match the rms of the mock-observed one.

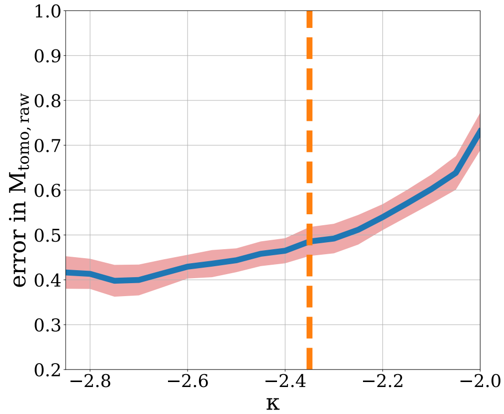


Figure 2.6: The rms error in the tomographic masses $M_{\text{tomo,raw}}$ of watersheds as a function of the parameter κ. The error is defined based on the difference between masses of watersheds in the mock observed map and their counterparts (based on maximal overlap) in the noiseless map. The solid line and shaded regions are the mean and 1σ scatter among different realizations of the mock observed map. Based on this analysis we select $\kappa = -2.35$.

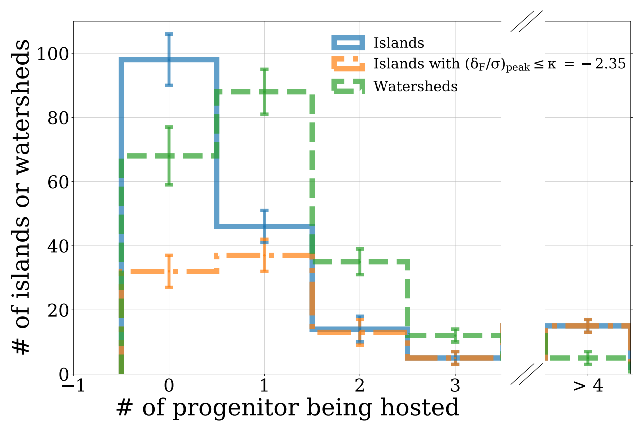


Figure 2.7: Comparing the utility of watersheds and simple flux contours (i.e., islands) in detecting the progenitors of massive $z = 0$ halos. Histograms show the mean distributions of the progenitor counts within islands or watersheds in 20 mock-observed flux maps. The 1-σ scatter around mean counts are also overlaid. Many islands, contours of $\delta_F/\sigma_{\text{map}} \leq \nu = -2$ (blue solid curve) host no progenitors exceeding $M_{z=0} > 10^{13.5}\ h^{-1}\text{M}_\odot$. Requiring these islands to contain at least one absorption peak with ($\delta_F/\sigma_{\text{map}} \leq \kappa = -2.35$) removes many of these vacant islands (orange dotted-dashed curve), but still leaves islands that contain as many as 40 progenitors. The watershed algorithm partitions those islands, significantly increasing the number of structures containing two or fewer progenitors (green dashed curve).

with size and color depending on the descendant halo mass. This visually confirms the capability of the method to deblend massive halo progenitors located within the same island. Figure 2.7 quantifies this by comparing the number of progenitors within each island or watershed. The distribution for islands, contours of $\delta_F/\sigma_{\rm map} \leq \nu = -2$, shows that many of them host no massive halo progenitor, at least in the mass range we consider here, $M_{z=0} > 10^{13.5}\, h^{-1}{\rm M}_{\odot}$. Many of these islands are removed after requiring the islands to have at least one minimum with $\delta_F/\sigma_{\rm map} \leq \kappa = -2.35$. However, there are some very large islands hosting up to 40 progenitors (the last bin of the histogram). By using the information on the gradient of the flux field, our method breaks these giant islands into watersheds that isolate progenitors individually, or at least in much smaller groups. This comparison shows that watersheds are clearly superior at isolating the progenitors of massive $z = 0$ halos compared to simple contours in the flux map. Nevertheless, some progenitors are extremely blended and cannot be distinguished even with our deblending method.

2.4 Results

In this section, we define and calibrate the tomographic masses $M_{\rm tomo}$ of watersheds, which can be estimated from a tomographic map. We then compare these tomographic masses to the underlying dark matter mass of the watershed at $z = 2.5$, and to the masses of halos at $z = 0$ that the watersheds will form. Our aim is to develop and characterize estimators for these masses that can be applied to tomographic surveys, including LATIS. Before discussing these estimators, we first review our notation for the various masses that will be discussed. $M_{\rm tomo}$ denotes a mass calculated from a flux map.

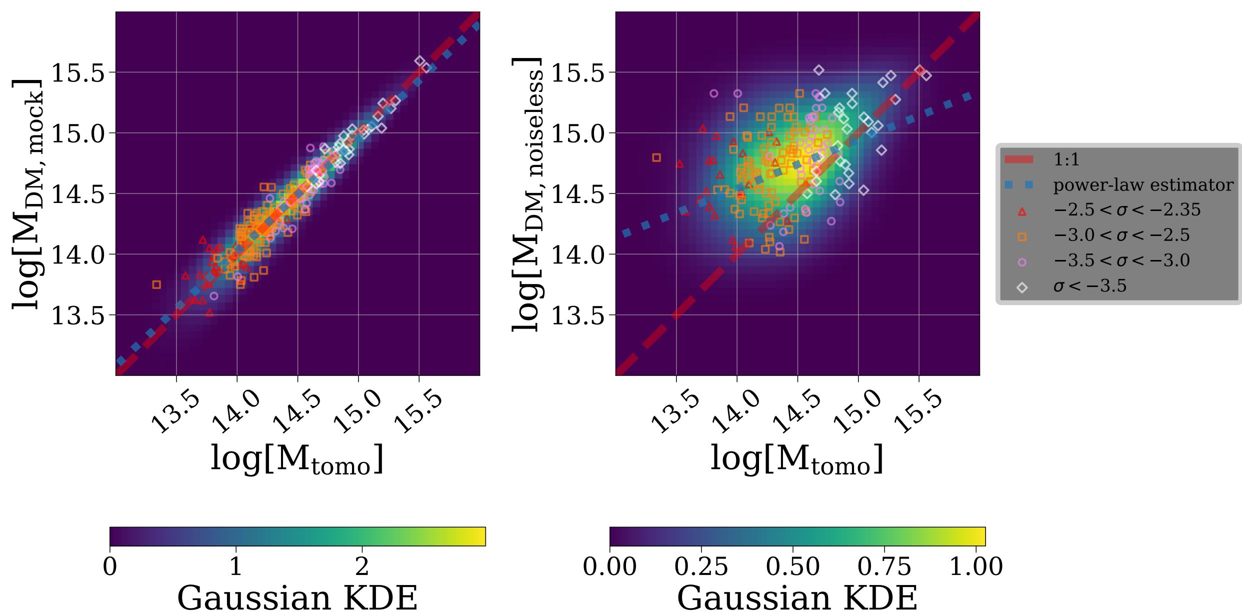


Figure 2.8: **Left Panel :** The tomographic masses (Equation 2.7) of watersheds in a mock-observed map are compared to the dark matter mass (i.e., the integrated unsmoothed dark matter density) within the same contours. The points are scattered around the one-to-one relation with a standard deviation of 0.12 dex. This scatter is sourced by scatter in the $\delta_F^{\rm sm}$-$\rho_{\rm DM}^{\rm sm}$ relation. **Right panel:** The tomographic mass is compared to the dark matter mass within companion structures in the noiseless flux map. The dotted line is the mean power-law fit $M_{\rm DM}^{\rm est}$ obtained from 16 mock-observed maps (Equation 2.8). The increased scatter and bias arise from noise in the boundaries of the watersheds. Points representing individual watersheds in one of the mock maps are overlaid on the mean Gaussian kernel density estimation (KDE) among 20 mock maps.

This could be a mock-observed map or a noiseless one, but for the rest of the paper, unless otherwise specified, we will consider $M_{\rm tomo}$ measured in mock-observed maps. When we consider the dark matter mass of a watershed at $z = 2.5$, we find that it is important to distinguish the mass within the watershed boundary derived from the mock-observed map, which we denote $M_{\rm DM,mock}$, from the mass within the boundary derived from a noiseless flux map, which we denote $M_{\rm DM,noiseless}$. $M_{\rm desc}$ will refer to the virial mass at $z = 0$ of the halo identified as the main descendant of a watershed. Finally, we use "est" to distinguish estimators, which are functions of $M_{\rm tomo}$. Thus $M^{\rm est}_{\rm DM}$ and $M^{\rm est}_{\rm desc}$ are estimators of $M_{\rm DM,noiseless}$ and $M_{\rm desc}$, respectively.

2.4.1 Calibrating tomographic masses at $z = 2.5$

As introduced in Section 2.3.2, $M_{\rm tomo,raw}$ is defined by estimating the dark matter density at each voxel, using the median $\delta_F^{\rm sm}$-$\rho_{\rm DM}^{\rm sm}$ relation (eqn. 2.5), and integrating over the watershed volume. As discussed by [108,], using a smoothed density field pushes some particles out of the watershed volume, which leads to a systematic underestimation of the DM mass within watersheds. We correct for this by applying an offset to our tomographic mass estimates, which are defined as

$$M_{\rm tomo} = 10^{0.14} \times M_{\rm tomo,raw}. \tag{2.7}$$

The motivation for the calibration factor $10^{0.14}$ is as follows. In the left panel of Figure 2.8, we compare $M_{\rm tomo}$ to the dark matter mass $M_{\rm DM,mock}$ (i.e., the integral of the unsmoothed

DM density) within the same watershed contours.[6] The slope of the linear fit is very close to one, and with our choice of calibration factor in Equation 2.7, there is no offset. The very small scatter of 0.12 dex originates from the scatter in the $\delta_F^{\rm sm}$-$\rho_{\rm DM}^{\rm sm}$ relation. This comparison shows that $M_{\rm tomo}$ provides a remarkably precise estimate of the dark matter mass within a specified volume.

Due to noise in the mock observations, which originates from the finite signal-to-noise and sightline density, the boundaries assigned to watersheds are distorted relative to those in the noiseless map. The right panel of Figure 2.8 compares the tomographic masses $M_{\rm tomo}$ of watersheds in the mock-observed map to the dark matter mass of companion watersheds in the noiseless map, $\rm M_{DM,noiseless}$. Again, the companion watershed is defined as the one that maximally overlaps a mock-observed one. Around 18% of watersheds in the mock-observed map overlap with no structure in the noiseless map, hence they are not shown on this figure. This comparison incorporates the additional uncertainty arising from the noisy boundaries of watersheds in the mock-observed map, which greatly inflates the scatter and produces a mass-dependent offset in the $M_{\rm tomo}$-$M_{\rm DM,noiseless}$ relation. In order to define an estimate for the expectation value of $M_{\rm DM,noiseless}$ given an observed $M_{\rm tomo}$, we fit a power law using ordinary least squares regression in logarithmic space. We generate 20 mock-observed maps with different sightlines and noise realizations. We use 16 of these maps as the training set to estimate the mean and uncertainties in the fit parameters; the

[6] We do not include the baryonic mass in $M_{\rm DM}$, but users who wish to estimate a total mass could simply scale by $\Omega_m/\Omega_{\rm DM}$, which is sufficiently accurate given the large scales we are considering.

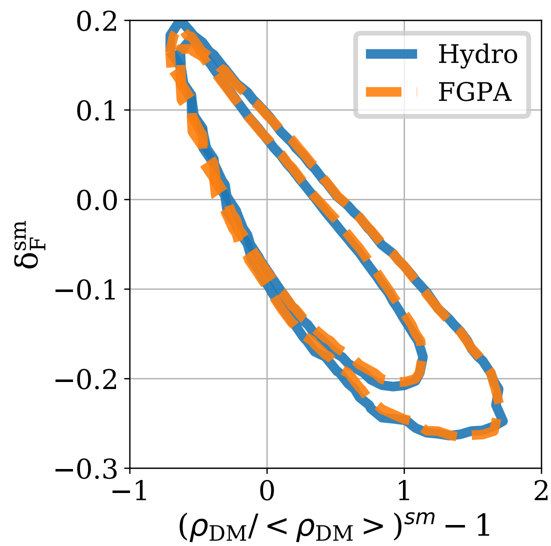


Figure 2.9: A comparison of the Ly-α flux fields simulated using the FGPA and the TNG300-1 simulation. The FGPA is applied to a dark-matter-only simulation, while TNG300-1 includes hydrodynamics and state-of-the-art feedback models. The noiseless flux contrast and DM overdensity maps have a voxel resolution of $1\,h^{-1}$cMpc and are smoothed with a gaussian kernel of size $\sigma_{sm} = 4\,h^{-1}$cMpc. The contours are the 0.02, 0.68 and 0.98 levels for the 2D PDF. The strikingly similar $\delta_F^{\rm sm}$-$\rho_{\rm DM}^{\rm sm}$ relation confirms the insensitivity of the Ly-α tomography signal, on Mpc scales, to the feedback and detailed hydro prescription.

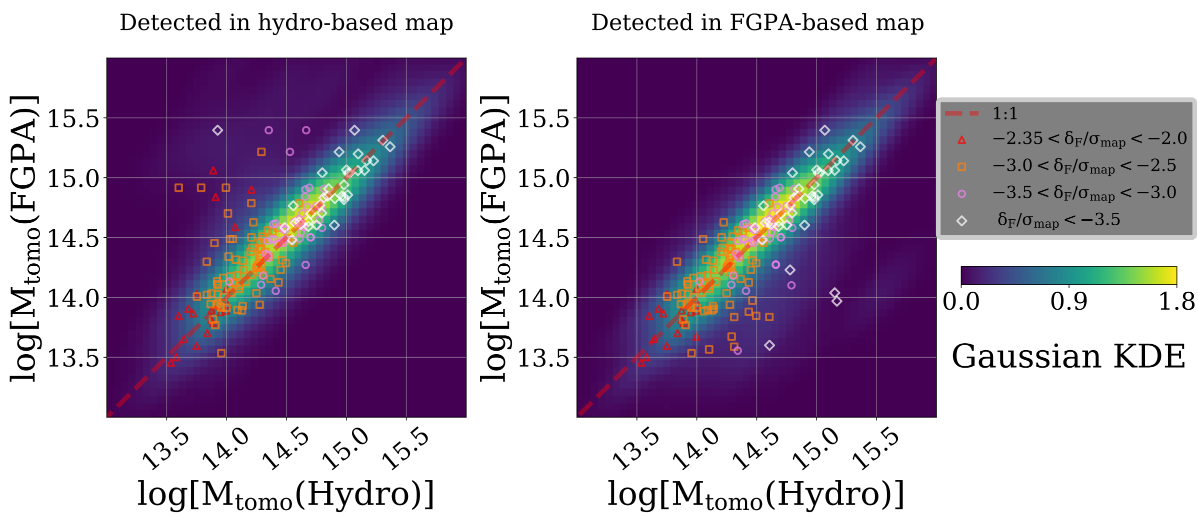


Figure 2.10: The accuracy of the FGPA, applied to a dark-matter-only simulation, for simulating the Ly-α flux field is assessed by evaluating the masses of watersheds. The same watershed finder is applied to noiseless maps derived from the full hydro simulation and from the dark-matter-only simulation using the FGPA. Both panels compare the tomographic masses of watersheds detected in one map to their counterpart watersheds (defined by maximal overlap) in the other. **Left panel:** Comparison of the masses of watersheds detected in the full-hydro map. There is a noticeable set of outliers discussed in the text. **Right panel:** Comparison of the masses of the watersheds detected in the FGPA map. We find that $\sim 18\%$ of lower-mass watersheds are prone to artificial blending with more massive structures in the FGPA-based map.

mean relation, shown as the dotted line in the right panel of Figure 2.8, is

$$M_{\rm DM}^{\rm est} = 10^{14.60\pm0.05} \times \left(\frac{M_{\rm tomo}}{10^{14}}\right)^{0.39\pm0.07} \quad h^{-1}{\rm M}_\odot. \tag{2.8}$$

We use the remaining 4 mock-observed maps to measure the scatter of the dark matter mass $\rm M_{DM,noiseless}$ around the estimator $M_{\rm DM}^{\rm est}$. The rms deviations within mass bins of $M_{\rm tomo}/\,h^{-1}{\rm M}_\odot = [10^{13}, 10^{14}], [10^{14}, 10^{14.5}], [10^{14.5}, 10^{15.5}]$ are 0.36, 0.33, and 0.29 dex. Therefore, the uncertainty in the mass estimate for watersheds is mostly driven by the noisy boundary of the watersheds rather than scatter in the $\delta_F^{\rm sm}$-$\rho_{\rm DM}^{\rm sm}$ relation.

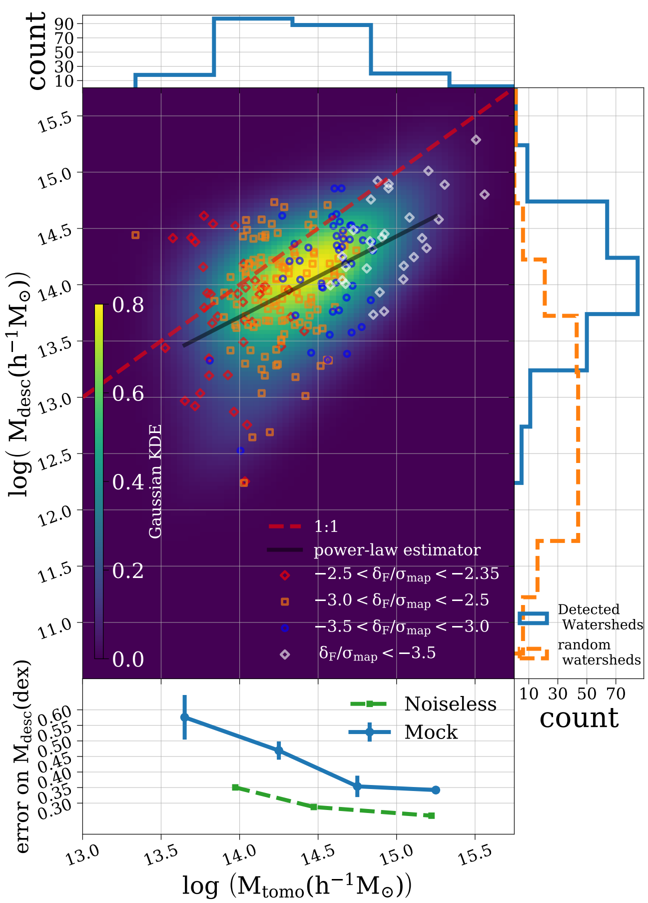
Collisionless versus Hydrodynamical Simulations

Collisionless simulations are often used to estimate the flux field, because only they have enough volume to enable many independent realizations of current-generation tomographic surveys. However, these simulations exclude physics that may be important for accurately estimating the Ly-α flux field on Mpc scales. To assess the accuracy of this method, we compare the noiseless δ_F maps obtained from FGPA and the full-hydro simulation. The contours in Figure 2.9 show the global distribution of the smoothed ($\sigma_{sm} = 4\,h^{-1}$cMpc) transmitted flux contrast vs the DM density, a diagnostic plot suggested in [96, 97,] to measure the effect of feedback on Ly-α tomographic maps. Contrary to [97,], Figure 2.9 shows a very small difference between these two methods on scales relevant to our analyses. In Appendix 2.7, we show how differences between our FGPA and Hydro maps are affected by the smoothing scale. We demonstrate that differences are slightly larger with a 3 Mpc/h smoothing scale, but still negligible compared to the current observational uncertainties in the $\delta_F^{\rm sm}$-$\rho_{\rm DM}^{\rm sm}$ relation.

Moreover, Figure 2.10 compares the tomographic masses of the watersheds detected in the FGPA and full-hydro noiseless maps. Counterpart watersheds are those with maximal overlap. The median difference in $M_{\rm tomo}$ within bins of $\log({\rm M_{tomo}}/\,{\rm h^{-1}M_\odot}) = [13.5, 14, 14.5, 15.5]$ is consistent with zero, confirming that the FGPA-based estimates are unbiased with respect to the full-hydro calculation. The scatter between the $M_{\rm tomo}$ estimates is generally small: $\sigma = 0.19,\ 0.18,\ 0.14$ dex in the same mass bins, where σ is the standard deviation obtained with an iterative sigma-clipped method. However, there is a noticeable set of outliers. These appear mainly in the left panel of Figure 2.10, which compares mass estimates for the watersheds detected in the full-hydro map, and are less numerous in the right panel, which shows the watersheds detected in the FGPA map. This comparison indicates that the main origin of the outliers are low-mass watersheds in the full-hydro map that are not present as separate structures in the FGPA map, because they are blended with larger watersheds (Left panel). We find that 32 watersheds (18% of the total) in the full-hydro map are over-blended in the FGPA map (i.e., they share a counterpart), while a smaller number (16, or 9%) are over-fragmented.

Comparing this scatter to the uncertainty in the tomographic mass estimation discussed earlier (Figure 2.8) and considering the fraction of failures, we find that the FGPA approach is adequate for simulating the signal from more massive structures, around $M_{\rm tomo} > 10^{14.2}{\rm M_\odot}$, but it becomes concerning in some lower-mass structures; in particular, those located near massive structures are prone to over-blending.

Figure 2.11: The tomographic masses $M_{\rm tomo}$ of watersheds are compared to the masses $M_{\rm desc}$ of their descendant halos at $z = 0$. **Main panel:** Watersheds are colored based on the significance of their peak absorption. The colormap illustrates the Gaussian KDE of the data points and the solid line shows the best-fit power-law estimator $M_{\rm desc}^{\rm est}$ (Equation 2.9). **Lower panel:** The rms error in the estimator $M_{\rm desc}^{\rm est}$, evaluated in several bins of $M_{\rm tomo}$, when $M_{\rm tomo}$ is measured in the mock-observed (solid line) or noiseless (dashed line) map. The comparison of these two curves show the scatter in this relation is mostly intrinsic (section 2.4.2). **Top and right panels:** The marginal distributions of $M_{\rm tomo}$ and $M_{\rm desc}$. In the right panel, the dashed histogram shows the distribution of $M_{\rm desc}$ when the watersheds are translated to random locations in the map.



2.4.2 The connection to massive halos at $z = 0$

To enable statistical studies of galaxy evolution over cosmic time, it is useful to relate the structures detected in tomography surveys to virialized massive halos in the late-time universe. Because the connection between the two is not one-to-one, tracing the descendants of high-z structures is not equivalent to tracing the progenitors of $z = 0$ halos, and the appropriate direction depends on the question. Here we consider both approaches in turn.

Descendants

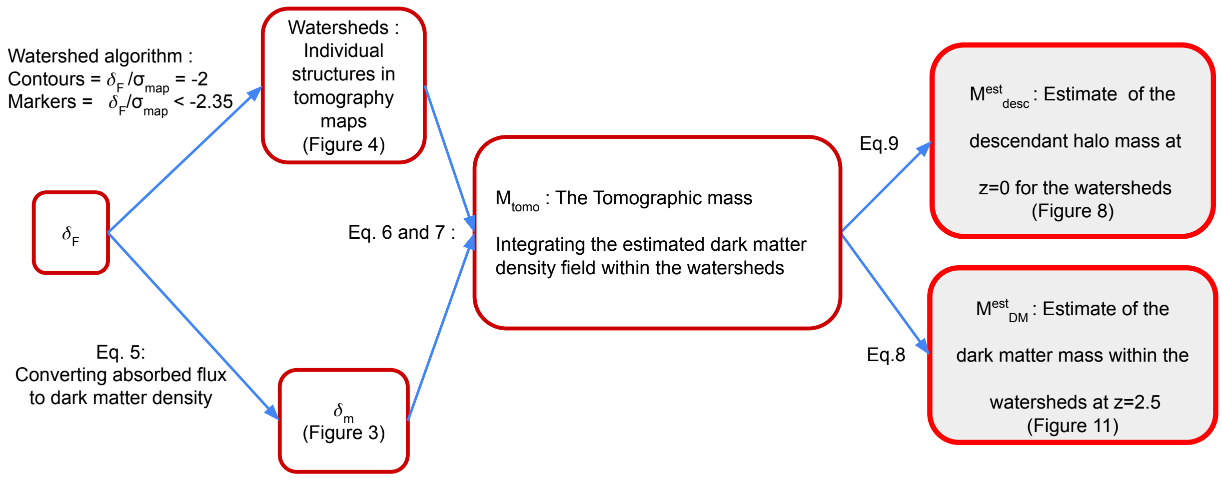
In this approach, we associate every watershed in a mock-observed tomographic map with a halo at $z = 0$. We aim to find the halo at $z = 0$ where the majority of the mass within the watershed structure will end up. To do so, we first link all halos (friends-of-fiends groups) at $z = 2.5$ to a halo at $z = 0$. Specifically, halo **B** at $z = 0$ is the descendant of halo **A** at $z = 2.5$ if the descendant of the most massive subhalo within halo **A** lies within the halo **B**. This approach allows us to use the SubLink trees provided with IllustrisTNG to connect subhalos across different redshifts. Then, to find the halo at $z = 0$ in which the majority of the collapsed mass within each tomographically mapped structure will end up, all halos within that structure vote for their descendant halo at $z = 0$, weighted by their halo mass. The mass of the descendant halo with the highest vote is chosen as the descendant of that watershed.

The main panel in Figure 2.11 compares the masses of these descendant halos, $M_{\rm desc}$, to the mock-observed tomographic masses, $M_{\rm tomo}$, of all individual watersheds detected in the mock-observed map. The power-law fit in that plot provides an estimate of the descendant mass of a watershed that can be applied to real observations. Again, we use 16 mock-observed maps to estimate the mean relation and the uncertainties in the parameters:

$$M_{\rm desc}^{\rm est} = 10^{13.72\pm0.09}\left(\frac{M_{\rm tomo}}{10^{14}h^{-1}M_\odot}\right)^{0.72\pm0.06} \quad h^{-1}{\rm M}_\odot\,. \tag{2.9}$$

The scatter around the mean relation is measured using the remaining four mock observations and is shown on the lower panel of Figure 2.11. The scatter ranges from 0.35 at the highest masses to 0.55 dex at the lowest. By applying the same method to the noiseless map, we characterize the intrinsic scatter in this relation. The lower panel of Figure 2.11 shows that this intrinsic error is around 0.30 dex, confirming the scatter is mostly intrinsic and likely originates from the variations in assembly history along with the fairly coarse resolution of the maps, which potentially blends tomographic structures with distinct descendants. [108,] also considered the descendant halos of structures in tomographic maps, and we compare our results in Section 2.5. Despite the increasing errors in estimating the descendant masses of the lower-$M_{\rm tomo}$ structures, we emphasize that these watersheds are still preferentially associated with massive $z = 0$ halos. To demonstrate this, we randomly displace the watershed contours within the mock-observed map and determine a new descendant halo mass for each. The distribution of $M_{\rm desc}$ in these randomly located watersheds is shifted to lower masses, as shown in the right panel of Figure 2.11. For a graphical summary of the procedures and definitions refer to Figure 2.12.

Figure 2.12: A step-by-step guide to our method for detecting structures in a Ly-α tomographic map, estimating the dark matter mass within them at $z = 2.5$, and estimating their descendant halo masses at $z = 0$.



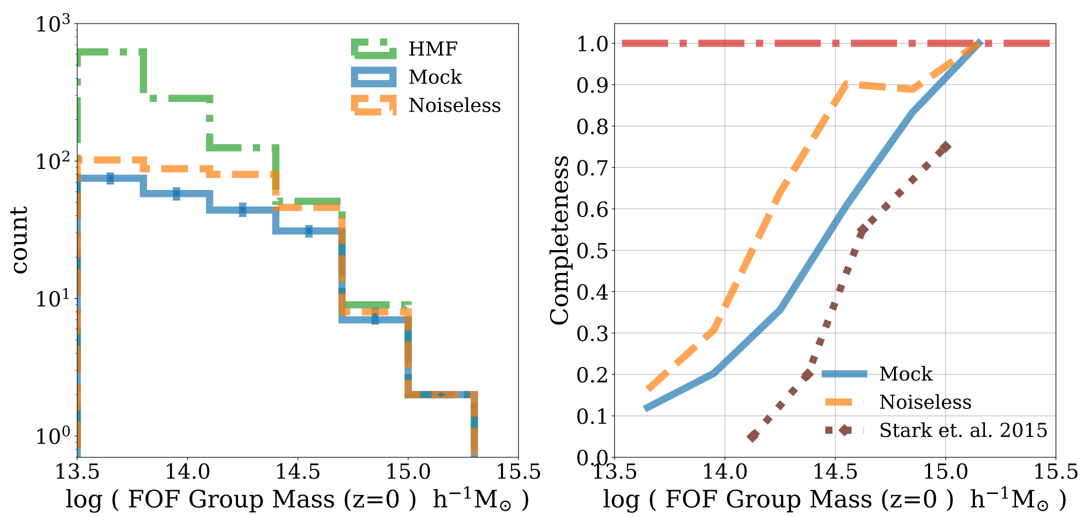


Figure 2.13: **Left panel:** The halo mass function at $z = 0$ (dotted dashed curve) is compared to the mass distribution of the halos whose progenitors are detected in the mock-observed (solid curve) and noiseless maps (dashed curve). For the mock-observed maps, the median distribution among 20 mock maps and the 1-σ scatter are shown. **Right panel**: The completeness, defined as the fraction of massive $z = 0$ halos whose progenitors are detected in the mock-observed (solid curve) or noiseless (dashed curve) maps. Our method detects progenitors of halos over about one decade of mass, from massive protogroups to protoclusters. The completeness reported by [188,] (dotted curve) is slightly lower due to their choice of a different detection threshold.

Progenitors

We now turn to computing the completeness of our method for detecting the progenitors of massive $z = 0$ halos. The method used in Section 2.4.2 is not suitable for this purpose since some watersheds host multiple progenitors of $z = 0$ clusters. For an accurate census of these progenitors, we follow a particle-based approach. We trace back to $z = 2.5$ all the dark matter particles within a distance R_{200} (where the mean density is $200 \times \rho_{\rm crit}$) of the centers of halos (friends-of-friends groups) at $z = 0$.[7] This method is computationally prohibitive to employ for the many lower-mass halos; therefore, we only consider the progenitors of halos with masses $M_{z=0} > 10^{13.5}\ h^{-1}{\rm M}_\odot$, which includes both protoclusters and protogroups. In TNG300-1, there are 1092 halos above this mass threshold. Any progenitor whose center of mass lies within any watershed is considered as "detected." The completeness is then quantified by comparing the total halo mass function at $z = 0$ to the subset with detected progenitors (Figure 2.13).

The completeness for the smallest structures, protogroups with $10^{13.5} < M_{z=0}/\,h^{-1}{\rm M}_\odot < 10^{14.0}$, is $\approx 20\%$. However, the completeness increases rapidly with halo mass: more than 40% of protoclusters, $M_{z=0} > 10^{14.0}\,h^{-1}{\rm M}_\odot$, are detected, and the completeness for massive protoclusters, $M_{z=0} > 10^{14.5}\,h^{-1}{\rm M}_\odot$, is as large as 80%. Figure 2.13 also shows the completeness in our analysis is higher than reported by [188,]. This is because we include structures with smaller flux contrast than their threshold of $\delta_F/\sigma_{\rm map} < -3.5$, which increases completeness but also includes more progenitors of lower-mass halos (Figure 2.11).

[7] We neglect gas particles since it is complicated to trace their history using a moving mesh code such as Arepo.

Comparing the completeness in the mock and noiseless maps in Figure 2.13 suggests that observational noise (both from noise in the spectra and finite sightline sampling) cannot entirely account for the incompleteness. To understand the intrinsic difference between the progenitors that are detected and those that are not, we examine their densities in Figure 2.14. Colors encode the population of the detected and undetected protoclusters in the noiseless map, and density is defined as the spherically averaged dark matter density within 4 h^{-1}cMpc of the center of mass. This comparison shows that the tomographic selection is approximately unbiased for protoclusters with masses $M_{z=0} > 10^{14.2}\,h^{-1}\mathrm{M}_{\odot}$, while for lower-mass halos, it is biased towards the increasingly small subset of progenitors that were most dense at $z = 2.5$.

2.5 Summary and Discussion

Tomographically mapping the 3D structure of the Ly-α forest with a set of closely spaced sightlines toward background QSOs and LBGs provides a unique opportunity to detect and characterize the extended, massive structures at $z \sim 2.5$ that form today's clusters and groups. Numerical simulations predict these progenitors to be extended over $5 - 10\,h^{-1}$cMpc scales with overdensities of $\delta_m = \rho/\langle\rho\rangle - 1 \approx 2 - 4$ [30,]. In this work, we build upon the previous intuitive methods developed by [188,] and [108,] to detect massive structures in Ly-α tomographic maps and determine their mass and physical extent. This effort is complicated by significant observational noise in the maps, and by the frequent blending of the progenitors of today's clusters and groups at the resolution of these maps. Compared to earlier methods, we aim to include lower-mass structures while mitigating the

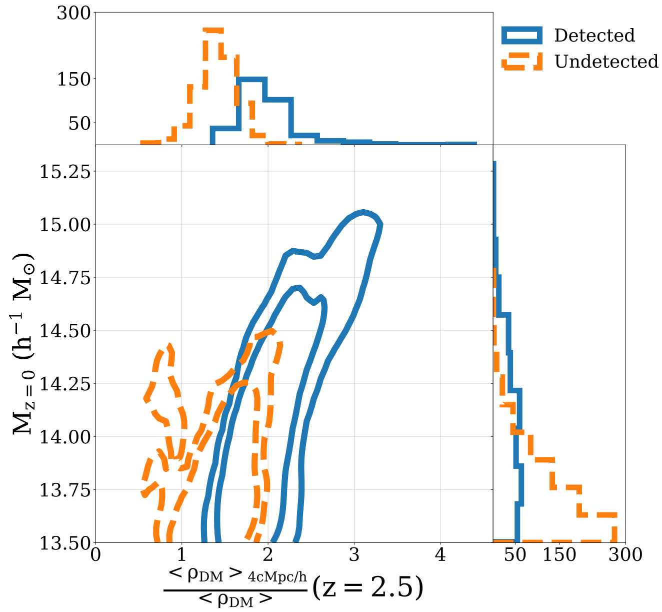


Figure 2.14: Comparison of the densities of the progenitors of massive $z = 0$ halos ($M_{z=0} > 10^{13.5}\,h^{-1}\mathrm{M}_\odot$) that are detected (solid curves) or undetected (dashed curves) in the noiseless $z = 2.5$ flux map. The x axis shows the ratio of the average dark matter density, measured within a sphere of radius 4 h^{-1}cMpc around the center of mass of each progenitor, to the global mean dark matter density at $z = 2.5$. The detected progenitors of massive halos are an approximately unbiased subset, but for lower-mass halos, the detection is biased towards the progenitors embedded in denser environments at $z = 2.5$. For each population, the contours show the 68 and 5 % levels of the 2D gaussian kernel density estimate.

associated increase of noise and blending. Below, we briefly summarize our method and its two tunable parameters, which we optimized for a LATIS-like survey at $z = 2.5$ by testing against a large hydrodynamical simulation. A graphical summary is presented in Figure 2.12.

Detecting structures:

1. Draw contours of $\delta_F/\sigma_{\rm map} \leq \nu = -2$ to define "islands" within the δ_F map, which is smoothed by a Gaussian kernel with $\sigma_{\rm sm} = 4\,h^{-1}$cMpc (Section 2.3.1).

2. Within each island, identify all local flux minima with $\delta_F/\sigma_{\rm map} \leq \kappa = -2.35$. If no such minima exists, remove that island. If multiple such minima are present, use the watershed algorithm to partition the island into watersheds associated with each minimum (Section 2.3.2 and Figure 2.4).

Estimating structures' masses at $z = 2.5$:

1. Use the second-degree polynomial fit in equation 2.5 to relate the measured δ_F to the dark matter density within each voxel of a watershed.(Section 2.2.4 and Figure 2.2)

2. Integrate these DM densities over the watershed's volume and apply the offset factor in equation 2.7 to estimate a tomographic mass $M_{\rm tomo}$ (Section 2.4.1 and Figure 2.8).

3. Use equation 2.8 to estimate the DM mass of the watershed from M_{tomo}. This estimator corrects for the bias from the noisy boundary. Uncertainties are provided in Section 2.4.1. (Figure 2.8)

Estimating structures' descendant masses at $z = 0$:

1. Use the measured M_{tomo} and the power law in equation 2.9 to estimate M_{desc}, the mass of the descendant halo at $z = 0$. Uncertainties are shown in Figure 2.11.

This method can be applied to LATIS-like tomographic maps to produce catalogs of structures (watersheds) that range from protogroups ($M_{z=0} = 10^{13.5} - 10^{14}\ h^{-1}\mathrm{M}_{\odot}$) to protoclusters ($M_{z=0} > 10^{14}\ h^{-1}\mathrm{M}_{\odot}$). We estimate the method's completeness is 80% for the most massive protoclusters (Figure 2.13). Less massive structures leave smaller signals within the absorption map; however, our method still achieves a 20% completeness for protogroups and is biased toward those embedded within denser regions on scales of $4\,h^{-1}$cMpc (Figure 2.14).

We find the scatter in the δ_F^{sm}-$\rho_{\text{DM}}^{\text{sm}}$ relation results in a remarkably small scatter of 0.12 dex (left panel of Figure 2.8) in the estimated mass of structures at $z \sim 2.5$, if the mass is measured within a fixed aperture, such as an observed flux contour. This result is similar to that shown in Figure 9 of [108,]. There are also Bayesian methods designed to infer the underlying DM density and velocity by comparing the approximate N-body simulations with tomography observations[78, 79, 76,]. They showed by adding information from the coeval galaxy distribution to the Ly-α tomography, their method reduces the scatter in protocluster mass estimation by a factor of two if one defines the mass as an integrated spherical overdensity within a radius around the Eulerian position of the corresponding cluster at z = 0 (i.e comparing the left panel in Figure 2.8 and Figure 10 in [79,]). However, we find that the dominant uncertainty in the mass estimate is due to the uncertainty in measuring the structures' volumes. When comparing each mock-observed structure to the overlapping structure in the noiseless map, the uncertainty in the mass estimate increases(right panel of Figure 2.8), particularly for lower-mass structures, and $M_{\rm tomo}$ becomes a biased estimator. We provide an unbiased estimator $M_{\rm DM}^{\rm est}$ and find that its uncertainty ranges from 0.29 to 0.36 dex for the highest to lowest mass bins.

This precision is competitive with the estimates from dynamical methods applied to galaxy spectroscopic surveys. For instance, [37,] estimate a median error of 0.35 dex for masses of components of the Hyperion proto-supercluster, which have a median mass of $10^{13.8}\,h^{-1}\mathrm{M}_{\odot}$. Similarly, [41,] spectroscopically confirm a protocluster in the COSMOS field at $z \sim 2.2$ and estimate the dynamical mass to be $1-2\times10^{14}\,h^{-1}\mathrm{M}_{\odot}$. On the other hand, dynamical mass estimates are strictly applicable only to virialized systems, which is not the case for protoclusters or protogroups. Ly-α tomographic mass estimates do not rely on this assumption. The other method commonly used to estimate the masses of structures at high redshifts is the galaxy overdensity [**?**,]e.g.,][]Steidel:1998. More recently, [7,] developed a Bayesian method to reconstruct the underlying DM density field for a combination of multiple independent spectroscopic galaxy surveys in the COSMOS field. However, all methods based on galaxy surveys require assuming that all the observed galaxy subpopulations trace the density field according to a simple model (i.e. same bias factor). Another key benefit of Ly-α tomographic masses is their independence from the galaxy content of a structure, which is valuable for galaxy-environment studies.

We also estimated the mass at $z = 0$ of the descendants of structures detected in tomographic maps (section 2.4.2). These estimates are significantly more uncertain, about 0.4 dex (Figure 2.11). We have verified the scatter is only slightly reduced in the noiseless map, which suggests the uncertainty is driven mostly by the scatter in assembly history, as concluded by [188,]. [108,] also considered the descendant halos of structures detected in tomographic maps. We find two significant differences with their results. The slope of the $M_{\rm tomo} - M_{\rm desc}$ relation (Equation 2.9) is closer to unity in our analysis (0.72 versus 0.27), and our estimate of the scatter is larger (0.3-0.45 dex versus ~ 0.2 dex). There are several explanations for these differences. First, when defining this relation, [108,] considered only the structures in the mock-observed map that are actually progenitors of cluster-scale halos. We find that censoring the 25% of structures with lower-mass descendants suppresses the slope and the scatter. Second, [108,] chose a different threshold ($\delta_F/\sigma_{\rm map} < -3$) to define the integration contours, which encloses a different fraction of the collapsed halo mass. Third, we impose a margin between the integration contour and the minimum δ_F within a watershed (i.e., $\kappa < \nu$), which effectively places a lower bound on $M_{\rm tomo}$ by eliminating very small contours. We tested the sensitivity of structure detection and mass estimates in mock tomographic maps to the detailed modelling of the gas and galaxy evolution (Section 2.4.1). We conclude that the fluctuating Gunn-Peterson approximation (FGPA) applied to dark-matter-only simulations imputes sizeable errors in characterizing lower-mass structures,

$M_{\rm tomo} < 10^{14.2}\,h^{-1}{\rm M}_{\odot}$. Compared to full-hydro simulations, FGPA confuses lower-mass neighboring structures by either merging them with more massive neighbors or fracturing them into few smaller watersheds. However, contrary to [97,], we find the global distribution of $\delta_F^{\rm sm}$-$\rho_{\rm DM}^{\rm sm}$ relation (Figure 2.9) at the relevant scales ($4\,h^{-1}$cMpc) and redshifts ($z \sim 2.5$) is insensitive to the galactic feedback and detailed hydro prescription in TNG simulations. The IllustrisTNG simulations also do not include all the physics that may be relevant. Although the galaxy formation model approximately accounts for the local heating due to black hole accretion (section 2.6.4 in [202,]), to fully account for the proximity effect, the local ionization should be modelled. This requires radiative transfer simulations, which are computationally demanding. However, [128,] placed an approximate bound on this effect by sampling the AGN luminosity function and assigning AGN to massive halos in a one-to-one rank order fashion (implying a duty cycle of one). They conclude local ionization usually has a small effect on the Ly-α absorption observed in tomographic maps, even within massive protoclusters.

Moreover, X-ray observations of galaxy clusters provide evidence for some feedback activities [52, 119,]. It is speculated a form of "preheating" started early on in the progenitors of these clusters [54,]. [96,] studied this effect in simulated tomography maps by implementing simple feedback models, i.e. injecting energy or entropy to all particles within protoclusters, using the broad constraints on the feedback strength from [27,] and [81,]. They predict significant ionization within protoclusters in cases with strong feedback. However, we do not find increased transparency in the overdense regions due to feedback when comparing the FGPA and full-hydro noiseless maps (Figure 2.9). The feedback prescriptions in the TNG300-1 simulation are state of the art (section 2.2.1 and [204,]). Additionally, our mock-observed tomography maps match the observed flux distribution in LATIS very well (Fig 2.3), while much stronger feedback could disturb the agreement for $\delta_F < 0$.

The methods developed in this work for detection and characterization of the structures in a LATIS-like tomography survey, as well as the realistic mock maps built here, will be used in future works. The catalogs of structures and their masses derived with these methods will be essential ingredients in studying the environmental dependencies of galaxy properties, e.g. star formation and metallicity, within statistically representative samples of overdensities at "cosmic noon." The two parameters in our method can be tuned for any other Ly-α tomography survey with a different spectroscopic noise properties or sightline density.

2.6 1D Flux Power spectrum

Figure 2.15 compares the noiseless 1D flux (i.e. δ_F) power spectrum simulated with the full-hydro TNG300-1 and the FGPA models to the observational constraints from SDSS DR14 [24,]. We estimate the power spectra by applying the Fast Fourier Transform (FFT) on ~ 8000 simulated spectra and correct for the finite pixel resolution effect. The convergence test is passed with half as many spectra. The oscillations in the observed power spectrum arising from Ly-α -SiIII correlations are not modelled in the simulated spectra. After correcting for that, we expect the simulated power spectra in both methods to be consistent with observations with 10% accuracy.

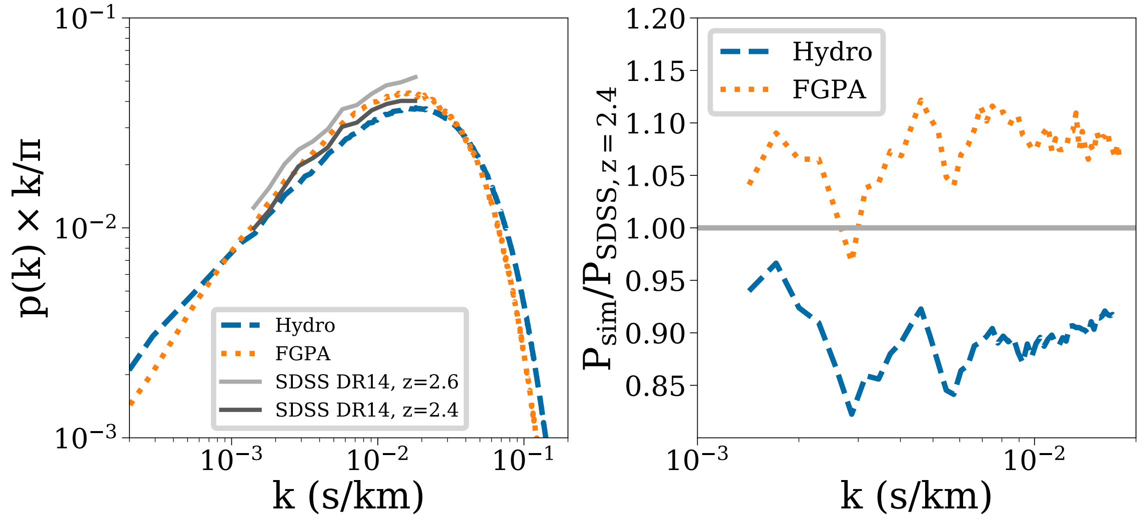


Figure 2.15: The noiseless 1D flux power spectra simulated with both the full-hydro and FGPA methods are compared to the observations in SDSS DR 14 [24,]. In the **right panel**, the ratio of the simulated power spectra to observations shows a $\sim 10\%$ agreement for both methods.

2.7 Sensitivity of FGPA Accuracy to Smoothing Scale

In this section we further examine the differences between the approximate FGPA approach and the full hydro simulations in detail. For a lateral comparison with the recent findings in [97,], we extract the 3D noiseless mock absorbed flux map from the TNG100-1 simulation at $z = 2.0$ using both hydro and FGPA prescriptions. The voxel resolution of each map is set to $V_{voxel} = 125\,h^{-1}$ckpc. On the left panel in Figure 2.16, the maps are smoothed with a kernel size of $\sigma_{sm} = 3\,h^{-1}$cMpc, similar to [97,]. Contrary to [97,], our comparison reveals more similarity between the two methods (compare the solid and dashed contours). The dotted contour however illustrates one could eliminate this small difference by adjusting the initial smoothing scale on the DM density and velocity in the FGPA method, set to mimic baryonic pressure smoothing, to the value suggested by previous

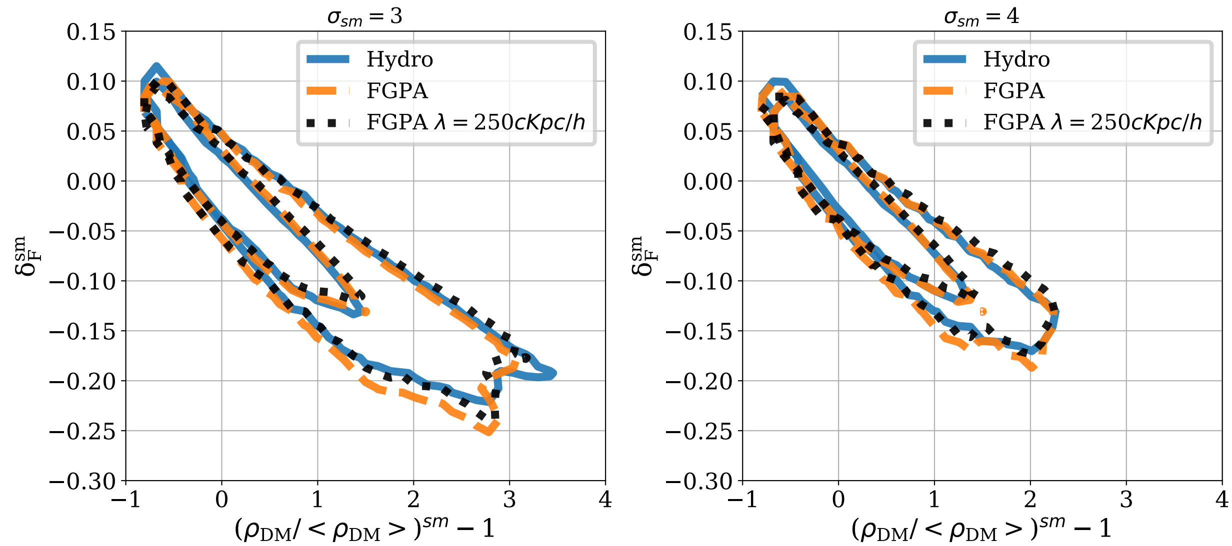


Figure 2.16: Similar to Figure 2.9, we compare the approximate FGPA method applied to DM-only simulations with spectra from a full hydro simulation. The noiseless mock maps are generated from TNG100-1 simulations at $z = 2.0$ and are smoothed with different kernel sizes of $\sigma_{sm} = 3,\ 4\ h^{-1}$cMpc on the left and right panels respectively. The FGPA map is increasingly similar to hydro as the smoothing scale is increased. The dotted line represents the FGPA map generated after an initial smoothing of the DM density and velocity to mimic baryonic pressure smoothing.

works [180,], i.e. $\lambda_{sm} \sim 250\,h^{-1}$ckpc. On the right panel, we instead smooth the mock maps with our fiducial kernel, i.e. $\sigma_{sm} = 4\,h^{-1}$cMpc. Comparison between the two panels shows that a slightly larger smoothing scale erases the small scale disparity between the two methods. This is in agreement with our conclusion in section 2.4.1 that using the FGPA method to generate maps smoothed with $\sigma_{sm} = 4\,h^{-1}$cMpc affects only a subdominant fraction of detected progenitors ($\sim 18\%$).

Table 2.2: The redshift evolution in $\delta_F^{\rm sm}$-$\rho_{\rm DM}^{\rm sm}$ relation $\left(\frac{\rho_{\rm DM}}{\langle\rho_{\rm DM}\rangle}\right)^{\rm sm} = {\rm a}_2\ \delta_{\rm F}^{\rm sm\ 2} + {\rm a}_1\ \delta_{\rm F}^{\rm sm} + {\rm a}_0$.

z	a_2	a_1	a_0
z=2.3	14.6 ± 0.6	-4.97 ± 0.06	0.97 ± 0.01
z=2.5	16.8 ± 0.6	-5.62 ± 0.07	0.96 ± 0.01
z=2.6	15.23 ± 0.7	-5.44 ± 0.06	0.98 ± 0.01

2.8 Redshift sensitivity

In a Ly-α tomographic survey that maps a significant redshift interval (for LATIS, $z = 2.2$-2.8), there will be both cosmological evolution (e.g., in the mean density and mean flux) as well a redshift dependence in the observational parameters (e.g., the sightline density and the noise in the observed spectra). Throughout this work, we fixed these parameters to those at the simulation snapshot $z = 2.45$, near the redshift midpoint of LATIS. To test the accuracy of our mass estimators at different redshifts within the LATIS map, we built 20 mock-observed maps at the $z = 2.3$ and $z = 2.6$ simulation snapshots using the LATIS sightline density and noise (Table 2.1) at the corresponding redshifts. This translates to 9017 and 4409 spectra for redshifts $z = 2.3$ and $z = 2.6$, respectively, compared to 6739 at $z = 2.5$. The estimator of the $\delta_F^{\rm sm}$-$\rho_{\rm DM}^{\rm sm}$ relation in eq. 2.5 changes slightly within this redshift range, as shown in Table 2.2. However, these slight changes result in a negligible offset, ~ 0.02 dex, in the tomographic mass measurements $M_{\rm tomo}$. After fixing the $\delta_F^{\rm sm}$-$\rho_{\rm DM}^{\rm sm}$ estimator to $z = 2.5$ and adopting the same watershed parameters for defining structures, i.e., $\nu = -2$ and $\kappa = -2.35$, we find that the best-fit parameters in the $M_{\rm DM}^{\rm est}$-$M_{\rm tomo}$ and $M_{\rm desc}^{\rm est}$-$M_{\rm tomo}$ estimators are consistent within this redshift range, as shown in Tables 2.3 and 2.4.

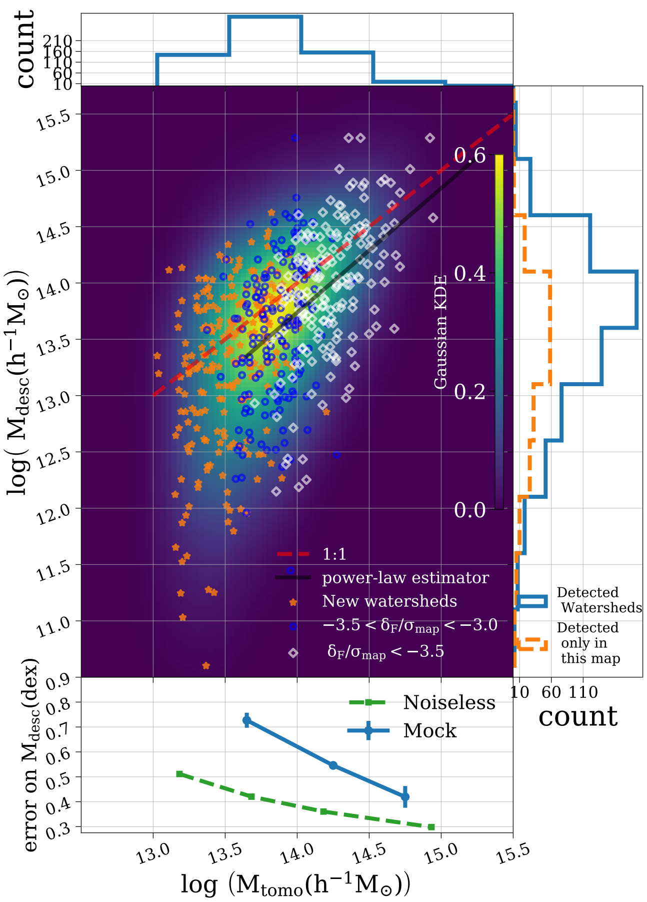


Figure 2.17: Same as Figure 2.11, but for the mock map smoothed with a smaller $\sigma_{\rm sm} = 2\ h^{-1}$cMpc Gaussian. Red stars indicate the new watersheds detected only in this less smoothed map. The other watersheds are colored based on their peak absorption significance.

Table 2.3: The redshift evolution in $M_{\rm DM}^{\rm est} = a \times \left(\frac{M_{\rm tomo}}{10^{14}}\right)^b$ in units of $h^{-1}{\rm M}_\odot$.

z	a	b
z=2.3	$10^{14.59\pm0.06}$	0.43 ± 0.09
z=2.5	$10^{14.54\pm0.04}$	0.39 ± 0.07
z=2.6	$10^{14.63\pm0.04}$	0.35 ± 0.05

Table 2.4: The redshift evolution in $M_{\rm desc}^{\rm est} = a \times \left(\frac{M_{\rm tomo}}{10^{14}}\right)^b$ in units of $h^{-1}{\rm M}_\odot$.

z	a	b
z=2.3	$10^{13.69\pm0.06}$	0.76 ± 0.09
z=2.5	$10^{13.71\pm0.06}$	0.72 ± 0.09
z=2.6	$10^{13.70\pm0.05}$	0.77 ± 0.09

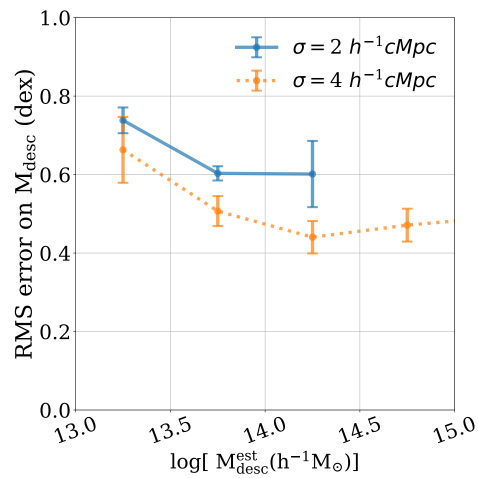


Figure 2.18: Comparing the uncertainty in descendant mass estimation between the mock-observed maps smoothed with $\sigma = 2\ h^{-1}$cMpc and $\sigma = 4\ h^{-1}$cMpc kernels. The x axis shows the descendant mass $M_{\rm desc}^{\rm est}$ estimated via the power-law estimator for each smoothing scale (Figures 2.11 and 2.17). At all masses, the precision of the prediction of the descendant mass is worse when using the less smoothed map.

2.9 Other smoothing scales

[188,] showed that smoothing the flux absorption map with a $\sigma_{\rm sm} = 4\,h^{-1}$cMpc Gaussian kernel behaves as a matching filter for detecting protoclusters, and our analysis throughout the paper is based on maps smoothed in this way. Since one goal of this paper is to expand the detections to smaller structures like protogroups, i.e., with $M_{z=0} < 10^{14}\,h^{-1}{\rm M}_{\odot}$, we test here whether smaller kernels are a better choice for detecting protogroups. Specifically, we investigate the watersheds that are only found in less smoothed maps and do not overlap with any watershed in maps smoothed with the standard kernel of $\sigma_{\rm sm} = 4\,h^{-1}$cMpc.

We follow the same steps discussed in sections 2.3.1 and 2.3.2 on the maps smoothed with a smaller $\sigma_{\rm sm} = 2\ h^{-1}$cMpc kernel, and find that the optimal parameters are $\nu = -2.30$ and $\kappa = -2.75$. Proceeding with the steps in section 2.4.2, except using the new parameters, we derive the $M_{\rm desc} - M_{\rm tomo}$ relation for the detected watersheds, which is shown in Figure 2.17. The orange stars indicate the watersheds that do not overlap with any of the watersheds within the map smoothed with $\sigma_{\rm sm} = 4\,h^{-1}$cMpc; these constitute 16% of the structures. The rest are colored based on their absorption peak significance. On one hand, the marginal histograms over $M_{\rm desc}$ (right panel) shows a fraction of the new structures found in the $\sigma_{\rm sm} = 2\ h^{-1}$ cMpc map are indeed significant overdensities, some associated with massive protoclusters. On the other hand, in comparison to the results for $\sigma_{\rm sm} = 4\ h^{-1}$cMpc map, the scatter in the estimated descendant mass is much larger. Figure 2.18 shows the rms error, similar to bottom panel of Figure 2.11, but plotted versus $M_{\rm desc}^{\rm est}$ since the $M_{\rm tomo}$ measured in maps smoothed differently are not directly comparable.

Due to the larger noise in less smoothed maps, the uncertainty in M_{desc} is larger at every mass scale. We conclude that maps with lower smoothing may be used to identify some lower-mass structures, but at the cost of larger noise contamination.

Chapter 3

Boosting Line Intensity Map Signal-to-Noise with the Ly-α Forest Cross-Correlation

Abstract

We forecast the prospects for cross-correlating future line intensity mapping (LIM) surveys with the current and future Ly-α forest measurements. Using large cosmological hydrodynamic simulations, we model the emission from the CO rotational transition in the COMAP LIM experiment at the 5-year benchmark and the Ly-α forest absorption signal for eBOSS, DESI, and PFS. We show that CO $\times$ Ly-α forest significantly enhances the detection signal-to-noise ratio of CO, with up to 300% improvement when correlated with the PFS Ly-α forest survey and a 50-75% enhancement with the available eBOSS

or the upcoming DESI observations. This is competitive with even CO $\times$ spectroscopic galaxy surveys. Furthermore, our study suggests that the clustering of CO emission is tightly constrained by CO $\times$ Ly-α forest due to the increased sensitivity and the simplicity of Ly-α absorption modeling. Foreground contamination or systematics are expected not to be shared between LIM and Ly-α forest observations, providing an unbiased inference. Ly-α forest will aid in detecting the first LIM signals. We also estimate that [CII] $\times$ Ly-α forest measurements from EXCLAIM and DESI/eBOSS should have a larger S/N ratio than planned [CII] $\times$ quasar observations by about an order of magnitude.

3.1 Introduction

Line intensity mapping (LIM) experiments have emerged as a powerful technique to study the interstellar medium (ISM) and the diffuse gas within the circumgalactic or intergalactic medium (CGM/IGM) by observing the aggregate atomic/molecular line emission [201, 99, 10,]. This observing strategy complements resolved observations with current or future flagship observatories such as JWST [66,], Roman [181,], HabEx, LUVOIR [195,], and Origins telescopes [126,]. LIM data is sensitive to the total line emission, including faint galaxies below the magnitude limits of even flagship observatories [98,]. LIM experiments can constrain the distribution of cold gas across cosmic time, which serves as the fuel for star formation in galaxies [91, 90, 190, 191, 31,]. LIM observations may also shed light on the role of early galaxy formation in reionizing the neutral gas [116, 115, 68, 84, 192,]. Additionally, LIM experiments measure the total integrated emission even from the faintest sources, making it easier to probe the large volumes accessible at higher redshifts compared to spectroscopic galaxy surveys at lower redshifts. This feature positions LIM to address questions in cosmology, such as constraining dark matter models [34,] and inflationary models of the early universe [132,].

Table 3.1: Summary of the key parameters in COMAP-Y5, Ly-α forest and galaxy surveys used in this work :

LIM Survey	area [deg^2]	z_{range}	timeline	Ref
COMAP-Y5	12	2.4-3.4	From 2022-2027	[33]

Ly-α Forest Survey	$d_\perp$	area [deg^2]	z_{range}	timeline	Ref
Ly-α , eBOSS	13	220	2.2-3.0	completed	[167]
Ly-α , DESI	10	14,000	2.2-3.0	started	[28]
Ly-α , PFS-bright	3.7	12.3	2.2-2.6	From 2024 to 2029	[71]
Ly-α , PFS-Faint	2.5	12.3	2.2-2.6	From 2024 to 2029	[71]

Spectroscopic Galaxy Survey	Targets (sampling rate)	$\frac{\sigma_z}{(1+z)}$	area [deg^2]	z_{range}	timeline	Ref
PFS	10870 (34 %)	7×10^{-4}	12.3	2.2-3.5	From 2024-2029	[71, 194]

Photometric Galaxy Survey	Selection	$\frac{\sigma_z}{(1+z)}$	area [deg^2]	z_{range}(for this work)	timeline	Ref
HSC- U+grizy+ YJHK (10-bands)	i ¡ 25 or $M_h > 10^{11.51}\,h^{-1}\mathrm{M}_\odot$	2×10^{-2}	5.5	2.0-3.0	completed	[45]
HSC- U+grizy+ YJHK (10-bands)	i ¡ 26 or $M_h > 10^{11.21}\,h^{-1}\mathrm{M}_\odot$	3×10^{-2}	5.5	2.0-3.0	completed	[45]
HSC- U+grizy (6-bands)	i ¡ 25 or $M_h > 10^{11.31}\,h^{-1}\mathrm{M}_\odot$	4×10^{-2}	18.6	2.0-3.0	completed	[45]
HSC- U+grizy (6-bands)	i ¡ 26 or $M_h > 10^{11.01}\,h^{-1}\mathrm{M}_\odot$	6×10^{-2}	18.6	2.0-3.0	completed	[45]

LIM has recently expanded from a focus on the 21cm emission of atomic hydrogen to measurements of other atomic or molecular line emissions, particularly at cosmic noon $z \sim 2-3$. For example, the CO Mapping Array Project (COMAP) [33,] started a 5-year Pathfinder program to observe the CO(1-0) rotational transition at z= 2.4-3.4 and CO(2-1) at $z = 6-8$ over 12.3 deg^2. In the next phase, COMAP-EOR will obtain a higher sensitivity to the same emission lines while expanding the detection to CO(1-0) emission at $z = 4.8 - 8.6$ [19,]. The Experiment for Cryogenic Large-Aperture Intensity Mapping (EXCLAIM), on the other hand, will observe the [CII] line emission at $z = 2.5 - 3.5$ and CO rotational lines at $z < 1$ over 305 deg^2 of the sky, rich with large extra-galactic data archives such as SDSS Baryon Oscillation Spectroscopic Survey (BOSS) [3,] and Hyper

Supreme Cam (HSC) photometric galaxy survey [4,].

However, it is challenging to detect the LIM signal owing to detector noise, and foreground contamination from our own galaxy, as well as extragalactic sources, including interloper emission lines in some cases. Interloper lines are emissions at different rest-frame frequencies but redshifted to the observed channel, which can lead to misinterpretations of the signal. One potential solution to overcome this issue is to cross-correlate the LIM signal with other known tracers of large-scale structure, which can improve the sensitivity of the measurement. For example, [64] show that the sensitivity of a 21-cm probe of the neutral hydrogen at the epoch of reionization improves by a factor of several when cross-correlated with a galaxy redshift survey. Similar cross-correlations have also been used to constrain the 21cm $\times$ galaxy [206,] and Ly-α emission $\times$ absorption [168,] at lower redshifts. [160] predict the signal in the EXCLAIM experiment can be detected through the cross-correlation between [CII] emission, at $z = 2.5 - 3.5$, and the quasars in stripe 82 of eBOSS. The first constraints on the small-scale power of emission from the CO molecule's rotational transitions at $z \sim 3$ have been obtained by cross-correlating with galaxies [92, 89, 93,]. Nevertheless, [32] demonstrate that the constraining power of the CO $\times$ galaxy is limited by the mass completeness and redshift uncertainty of the galaxy sample.

The frequent resonant scattering of the light by the neutral hydrogen within the intergalactic medium (IGM) leads to distinctive absorption features in the spectra of background galaxies. This large absorption field, commonly called the Ly-α forest, traces the large-scale distribution of the underlying matter with $\sim$Mpc resolution. A dense collection of these spectra can make a tomographic map of the large-scale structure [106, 77,], which provides a unique opportunity to study the relationship between the galaxy properties and their environment [139, 162, 130, 48,]. The largest such tomographic survey with a high density of sightlines covers around $\sim 1.7\ \ \mathrm{deg}^2$ of the sky at $z = [2.2 - 2.8]$ through spectroscopy of the Lyman Break Galaxies (LBG) and QSOs, [137,]. However, the upcoming prime focus spectrograph (PFS) will map $\sim 10\ \ \mathrm{deg}^2$ within the same cosmic time window, similar to the observed volume of the planned LIM pathfinders like COMAP [31,]. Quasar-based Ly-α forest surveys such as eBOSS [167,] and DESI [28,] map larger volumes ($\sim$ 220 and 14k deg^2 respectively) which facilitate overlap with upcoming LIM experiments; however, the sparsity of the sources results in lower transverse resolution.

In this study, we investigate the prospects of detecting molecular line emission through cross-correlation with the Ly-α forest, using particularly the COMAP-Y5 experiment [31,]. In Section 3.2, we introduce the hydrodynamic simulation we use. Section 3.3 describes how we generate synthetic observations from the simulated data. Section 3.4 details the summary statistics we use and how we model the noise. Our measurement forecasts for these statistics are provided in Section 3.5. We discuss the implications of our findings and the robustness of the results in Section 3.6. Finally, the main findings are summarized in Section 3.7.

3.2 Simulation

We create mock observations using the cosmological simulation `ASTRID` [16, 140,]. `ASTRID` is run using `MP-Gadget`, a smoothed particle hydrodynamics (SPH) code and a modified version of `Gadget-3` used for the `BlueTides` simulation. The simulation initially contains 2×5500^3 dark matter and gas particles within a periodic box of $250\,h^{-1}$cMpc. The gravity solver is based on a Tree-PM algorithm, which divides the force computation between a long-range particle mesh (PM) and a short-range hierarchical tree. The multi-phase prescription of star formation from [184] is adopted along with radiative and metal cooling of the gas particles [87, 203,]. A small correction to the star formation is applied to account for molecular hydrogen formation [101,]. Self-shielding of the dense gas is modeled using the fitting function of [164]. Newly formed star particles source hydrodynamically decoupled galactic winds, which recouple using a density change or time threshold. The simulation places the Super-Massive Black Hole (SMBH) seeds in halos with a Friend-of-Friend halo mass larger than $5 \times 10^9\,h^{-1}\mathrm{M}_\odot$ and a stellar mass larger than 2×10^6 by converting the densest gas particle in the halo. The SMBH seed mass is drawn from a power-law distribution between 3×10^4 to $3 \times 10^5\,h^{-1}\mathrm{M}_\odot$. SMBHs are kept near their host galaxy center with an effective dynamical friction force following [29]. ASTRID implements black hole accretion following a Bondi-Hoyle-Lyttleton-like prescription [46,] and thermal black hole feedback (AGN) with an efficiency of 5% [29,]. We refer to [140] for more details of the SMBH model used in ASTRID.

In ASTRID, the patchy reionization of hydrogen and helium occurring at $z > 6$ and $z > 2.8$, respectively, have significant implications for the intergalactic medium. Hydrogen reionization is modeled using a pre-computed reionization redshift map based on the large-scale smoothed overdensity, as described in [8]. On the other hand, helium reionization is implemented by randomly placing large bubbles at the probable location of quasars in massive halos, as outlined in [196]. The subhalos of dark matter used in this study are obtained using the SUBFIND algorithm [186,] in post-processing, and their corresponding subhalo masses are denoted as $M_{\mathtt{h}}$.

3.3 Mock observations

This section provides a brief overview of the methods used to generate mock observations and how these simulations can represent both current and upcoming observational surveys. Refer to Table 3.1 for the details of the surveys considered in this work.

3.3.1 Mock Ly-α Tomography

We produce artificial Ly-α absorption sightlines using the fast fake_spectra[1] python package [15, 12, 162,]. Each gas particle is treated as a separate absorber, and the absorption from all gas particles along the line of sight is summed to calculate the absorption spectra. Absorption from a single gas particle is computed by convolving a Voigt profile with the particle's density kernel. The internal physical quantities in each SPH particle are smoothed using a quintic spline kernel[2]. At $z \sim 2.5$, the neutral hydrogen fraction is computed by solving a rate network assuming ionization equilibrium between the uniform ionizing UV background and the recombination rates from [87]. The observed mean flux imposed by the uniform UV background [56,] is enforced in the simulated spectra by scaling the overall optical depth, similar to [166, 36, 162,]. The simulated forest recovers the 1D flux power spectrum measured by the Sloan Digital Sky Survey (SDSS) DR14 [25,] at the

[1] GitHub: fake_spectra
[2] A top-hat for TNG300 simulations, see Appendix 3.8.

10% level. The final Ly-α absorption map generated for the power spectrum calculations is on a uniform grid with side-length of $250\,h^{-1}$ckpc.

In recent years, significant progress has been made in mapping the intergalactic medium (IGM) using Ly-α absorption tomography. We consider two classes of Ly-α survey with mean background source separations of $d_\perp$ = 10-13 or 2.5-3.7 h^{-1}cMpc. The eBOSS survey, covering 220 deg^2 of the sky at $z = 2.2 - 3.0$, achieved a mean sightline separation of $d_\perp \sim 13\,h^{-1}$cMpc using QSO spectra from the SDSS DR16 Stripe 82 region [167,]. Additionally, the DESI survey has started a five-year program to observe QSOs at $z > 2.1$ over 14 k deg^2 of the sky, achieving a mean separation of $d_\perp \sim 10\,h^{-1}$cMpc [28,]. For higher spatial resolution, spectroscopy of fainter sources like Lyman Break Galaxies (LBGs) over a smaller footprint has been pursued by COSMOS Lyman-Alpha Mapping And Tomography Observations (CLAMATO), which maps 0.2deg^2 of sky within $z = 2.05 - 2.55$, achieving a mean transverse resolution of $d_\perp \sim 2.5\,h^{-1}$cMpc [77, 110,]. The largest high-resolution tomography is performed with Lyman Alpha Tomography IMACS Survey (LATIS), which maintains similar mean transverse resolution across 13 times larger volume at $z = 2.2 - 2.8$ [137, 162, 139,]. A tomography survey with the multiplex spectroscopy on the Prime Focus Spectrograph (PFS) instrument is planned to obtain similar mapping quality, i.e. $d_\perp = 2.5 - 3.7$ $cMpch$, over ~ 12.3 deg^2 [71,]. This survey is large enough to potentially fully overlap with a pathfinder line intensity survey and is set to begin in 2024. However, the spatial resolution of the cross-correlated signal with CO emission in such dense tomographies will also be limited by the beam size and channel width of the COMAP-Y5 pathfinder, i.e. $(d_{||}, d_\perp) \sim (2.64, 4.66)\,h^{-1}$cMpc [31,]. For our mock observations, we adopt a realistic spectral signal-to-noise ratio of 2 per Å for all sources, consistent with typical Ly-α forest data from surveys such as eBOSS [104,], CLAMATO [110,], and LATIS [137,]. As noted by [125], longer exposure for individual spectra does not significantly improve the survey's sensitivity to the auto-correlation of the forest data. In Table 3.1, we summarize the details of the surveys considered in this work.

3.3.2 Mock galaxy surveys

To generate a mock galaxy survey for our analysis, we adopt a method similar to previous studies on CO $\times$ galaxy surveys [32, 113,] and 21 cm signal $\times$ galaxies [64,]. We construct a galaxy density map at $z \sim 2.5$ by displacing subhalos based on their peculiar velocities along the line-of-sight and then mapping them onto a uniform fine grid with a side-length of $250\,h^{-1}$ckpc using the cloud-in-cell kernel. The power spectrum of this density map in redshift space is the signal in our analysis.

In Section 3.4, we incorporate the effects of observed galaxy redshift uncertainties by accounting for them in the noise power spectrum, following the methodology developed in [64,] and [32,]. We consider two classes of galaxy redshift surveys: *photometric* and *spectroscopic*. An example of a wide and deep *photometry* survey at $z = 2-3$ is the deep layer of Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP; [4]), which imaged 36 deg^2 in 5 broad-band filters of *grizy*. The typical redshift uncertainties at the relevant redshifts, ignoring the outliers, are estimated to be $\frac{\sigma_z}{(1+z)} = 0.09$ by [141]. Further imaging of 18.6 deg^2 of these galaxies with additional filters in the u-band by the CLAUDES project [173,] reduces the redshift uncertainties to $\frac{\sigma_z}{(1+z)} \simeq 0.05$. Moreover, observations in auxiliary near-infrared bands (`YJHK`) can reduce the uncertainties to $\frac{\sigma_z}{(1+z)} = 0.03$ for a smaller subset of 5.5 $\deg^2$ of these galaxies [45,]. The *photoz* quality of the `SourceExtractor` catalog provided in [45] degrades for fainter galaxies; therefore, we consider two subsets with magnitude cuts of $i < 25$ or $i < 26$ (similar to those in [141]) for each of the 6-band (`ugrizy`) or 12-band (`ugrizy+YJHK`) observed catalogs. We match the source density of these catalog subsets with the abundance of the subhalos in `ASTRID` more massive than a halo mass (M_h) threshold, resulting in slightly different mass completeness for each subset, as summarized in Table 3.1. An accurate abundance matching technique, however, requires a better understanding of the selection function of the observed galaxy samples. We also consider an upcoming medium resolution *spectroscopic* galaxy survey over 12.3$\deg^2$ at $z = 2.2 - 3.5$ within the PFS galaxy evolution project [71,]. The redshift uncertainties of such a galaxy survey at $z \sim 2.5$ are expected to be about $\frac{\sigma_z}{(1+z)} = 7 \times 10^{-4}$, which is required for cosmological studies [194,]. [71] estimates ~ 10870 available targets in the 1.3 $\deg^2$ field of view when observing in the faint mode, of which $\sim 34\%$ will be targeted. We build mock spectroscopic galaxy surveys by random sampling from halos with $M_h > 10^{11.9}\, h^{-1}\mathrm{M}_{\odot}$, following the same subhalo abundance matching technique.

3.3.3 Mock CO LIM

We model the CO emission from galaxies using a power-law scaling relation between line luminosity and average star formation in halos. To this end, we adopt the prior provided by the COMAP-Early-Science project [31,], which links the scaling relation between CO(1-0) luminosity (L_{CO}) and star formation rate (SFR) to M_h through a double power-law parameterization modified from [144]. This uses an analytic fit between the average SFR and halo mass proposed by the UniverseMachine framework [9,]:

$$\begin{aligned} \frac{L'_{CO}}{K\ kms^{-1}pc^2} &= \frac{C}{(M_h/M)^A + (M_h/M)^B} \\ \frac{L_{CO}}{L_\odot} &\sim LogNorm(\mu = 4.9 \times 10^{-5} L'_{CO}, \sigma = \sigma)\,, \end{aligned} \tag{3.1}$$

where $LogNorm$ indicates a log-normal distribution and A, B, C, M, σ are the model parameters, which are broadly constrained by the observations from COLDz [170] and COPSS [92]. We adopt the fiducial model parameters from Table 5 in [31] and discuss the sensitivity of our results to variations in these parameters in section 3.6. Although the ASTRID simulation accurately reproduces the observed SFR [16,], we do not attempt to build a new CO emission model based on these predictions, aiming to be consistent with the COMAP analysis. We postpone this exploration to future work.

We construct the CO temperature map by integrating the emission from all subhalos ($M_{\texttt{SUBFIND}} > 10^7\,h^{-1}\mathrm{M}_\odot$) within voxels of side-length $250\,h^{-1}$ckpc using a cloud-in-cell interpolation. The CO luminosity to temperature conversion (L_{CO}-T_{CO}) is done using the standard conversions in units of μK, as described in Appendix B.1 in [32]:

$$T_{CO} = 3.1 \times 10^4 \mu K (1+z)^2 (\frac{\nu_{rest}}{GHz})^{-3} \tag{3.2}$$

$$\times \ (\frac{H(z)}{kms^{-1}Mpc^{-1}})(\frac{L_{CO,vox}}{L_\odot})(\frac{V_{vox}}{Mpc^3})^{-1} \tag{3.3}$$

Where $L_{CO,vox}$ is the total CO luminosity in each voxel with a volume of V_{vox}. COMAP pathfinder experiment [33,] observes the CO(1-0) rotational transition (rest frame frequency $\nu_{rest} = 115.27$ GHZ) at redshifts between 2.4 and 3.4 in three fields of $4deg^2$ each. At the five-year mark of the pathfinder experiment, the system temperature fluctuation amplitude is expected to be $\sim 17.8\ \mu K$ per map voxel, with a voxel size of $(d_{||}, d_{\perp}) = (2.64, 4.66)\ h^{-1}$cMpc at redshift of $z \simeq 2.5$ [31,]. We account for the instrumental noise and finite angular/spectral resolution in our model of the CO noise power spectrum (3.4). In our analysis, we assume COMAP-Y5-like volume fully overlaps with the other auxiliary surveys, either galaxy or Ly-α forest data.

3.4 Statistics

The primary summary statistic considered in this work is the spherically averaged power spectrum. To estimate the signal power spectrum, we perform a Fast Fourier Transform (FFT) of the noiseless signal on a uniform fine grid with a side length of $250\,h^{-1}$ckpc. However, due to asymmetric uncertainties along and transverse to the sightline, we first calculate the power spectrum in $k - \mu$ space using the following equation:

$$\sigma_{P_A}(k, \mu) = \frac{P_{A,s}(k, \mu) + P_{A,n}(k, \mu)/W_A^2(k, \mu)}{\sqrt{N_m(k, \mu)}}, \tag{3.4}$$

where μ is the cosine of the angle between the line of sight and k the wavevector $\vec{k}$. Here, $P_{A,s}$ and $P_{A,n}$ are the signal and noise power spectra for any mock observation A, respectively. $N_m(k, \mu)$ is the mode count in each bin, and $W_A^2(k, \mu)$ is the signal attenuation term due to the finite angular and spectral resolution of the survey. The first and second terms in Eq. 3.4 account for sample variance in the limited observed volume and the noise contribution in each survey, respectively.

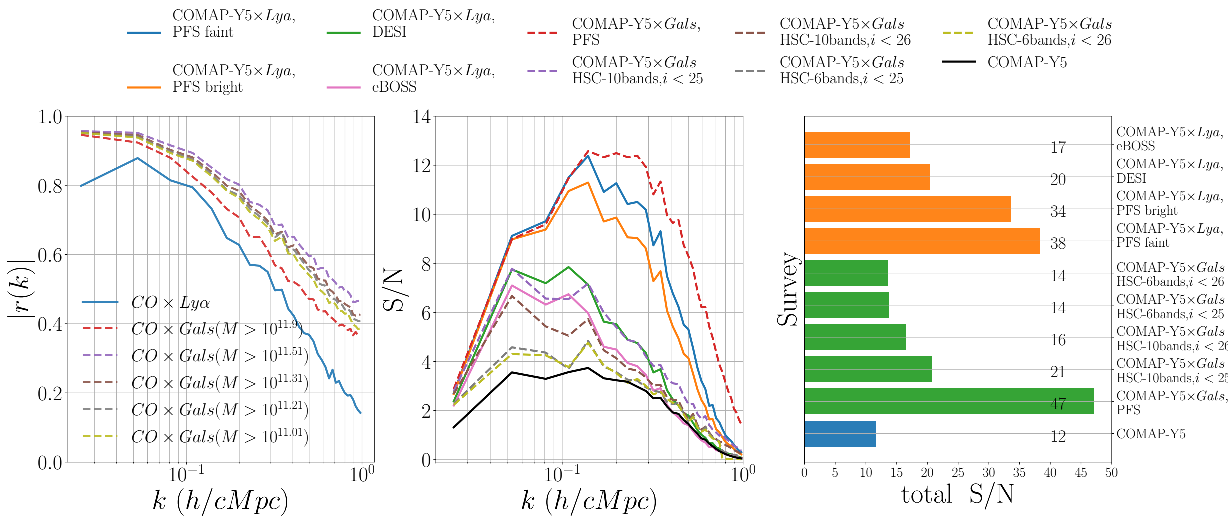


Figure 3.1: **Left Panel:** The plot shows the absolute value of the cross-correlation coefficients between auto CO power and CO $\times$ Ly-α or CO $\times$ galaxies power spectra. All signal maps are on a high-resolution grid of $(250\,h^{-1}\mathrm{ckpc})^3$. **Middle Panel:** The S/N forecast of the realistic mock observations within the COMAP-Y5 pathfinder's volume. The eBOSS observations are complete, while DESI data will be obtained in a 5-year program. The high-resolution PFS IGM map is planned to start observations in 2023. The spectroscopic galaxy survey is also planned with the PFS-GE program and the photometric galaxy surveys, HSC+CLAUDES and HSC+CLAUDES+NIR, are either completed or will be completed as part of PFS-GE. Refer to Table 3.1 for a summary of the survey parameters. **Right panel** illustrates the forecast total S/N for power over all k-bin, derived from Eq. 3.14. These values are shown on each bar for convenience.

For CO LIM, we adopt the noise model and survey parameters from the COMAP pathfinder survey [32, 31,]. Specifically, the noise power spectrum is given by:

$$\begin{aligned} P_{CO,noise}(k,\mu) &= \sigma_n^2 V_{vox,COMAP} \\ W_{CO}^2(k,\mu) &= e^{-k^2\sigma_\perp^2} e^{-\mu^2 k^2(\sigma_{||}^2-\sigma_\perp^2)}. \end{aligned} \tag{3.5}$$

where σ_n is the noise temperature in each voxel of volume $V_{vox,COMAP}$ and $(\sigma_{||}, \sigma_\perp)$ are the voxel size parallel and transverse to the sightline (section 3.3.3).

For Ly-α tomography, we estimate the uncertainties in the 3D power spectrum following [125]. They optimize the signal-to-noise ratio (S/N) in the power spectrum for a set of weights that quantify the contribution of each spectrum to the total signal. The noise power spectrum is then given by:

$$P_{Ly\alpha,n}(k,\mu) = P_{los}(k_{||})/\bar{n}_{2D,eff}\,, \tag{3.6}$$

where $\bar{n}_{2D,eff}$ is a noise-weighted projected density of the background source galaxies, given by

$$\bar{n}_{2D,eff} = \frac{1}{\mathcal{A}}\sum_{i=1}^{N}\nu_i\,, \tag{3.7}$$

$$\nu_i = \frac{P_{los}(k_{||})}{P_{los}(k_{||}) + P_{N,i}(k_{||})}\,. \tag{3.8}$$

$\mathcal{A}$ is the survey area and $P_{N,i}$ is the 1D noise power spectrum of the i'th source. In this work, we assume $P_{N,i}(k_{||})$ to be white noise with a signal-to-noise per angstrom of $(S/N) = 2$, typical of these surveys. Specifically, we have:

$$P_{N,i}(k_{||}) = (< F > /(S/N))^2 \Delta X\,. \tag{3.9}$$

$< F >$ is the mean absorption flux in the forest and $\Delta X = \lambda_{Lya\alpha} H(z)h/c$ is the conversion factor from Å to h^{-1}cMpc. Since this estimate is not convolved with any instrumental beam, we do not apply deconvolution to the noise power in equation 3.4, i.e. $W_{Lya}(k,\mu) = 1$. We note that the proposed noise power spectrum in Eq. 3.6 accounts only for the Poisson noise of the background sources. The clustering of the background sources becomes important only for tomography surveys with a higher sightline density than we consider here, $d_\perp \ll 2.5\,h^{-1}$cMpc [125,]. This is further discussed in section 3.6.

In galaxy surveys, the noise power spectrum depends on the number density of the galaxies, n_{3D}, amplified by the redshift uncertainty:

$$\begin{aligned} P_{Gal,noise} &= 1/n_{3D}\,, \\ W^2_{gal}(k,\mu) &= e^{-\mu^2 k^2 \sigma^2_{||}}\,, \end{aligned} \tag{3.10}$$

where $\sigma_{||}$ is the spatial resolution that is related to the redshift uncertainty, $\Delta_z = \frac{\sigma_z}{(1+z)}$, through:

$$\sigma_{||} = \frac{c\ \sigma_z}{H(z)} = \frac{c(1+z)}{H(z)}\Delta_z\,. \tag{3.11}$$

The uncertainties in the cross-power spectra between the CO signal and other tracers, such as Ly-α or galaxies, are estimated as:

$$\sigma^2_{P_{CO\times A}}(k,\mu) = \frac{\sigma_{P_{CO}}(k,\mu)\times\sigma_{P_A}(k,\mu)}{2} + \frac{P^2_{CO\times A}(k,\mu)}{2N_m(k,\mu)}\,. \tag{3.12}$$

A represents either a Ly-α or a galaxy survey. The uncertainties in the k-bins of the spherically averaged power spectra are then computed by summing the noise power over

μ-bins in inverse quadrature as shown in [64]:

$$\sigma_{P_A}^{-2}(k) = \sum_{\mu} \sigma_{P_A}^{-2}(k, \mu)\,. \tag{3.13}$$

The signal-to-noise ratio at each k-bin and the total signal-to-noise ratio are defined as:

$$\frac{S}{N} = \left[\sum_{k} \left(\frac{S}{N}(k)\right)^2\right]^{1/2} = \left[\sum_{k} \left(\frac{P_{noiseless}(k)}{\sigma_P(k)}\right)^2\right]^{1/2}\,. \tag{3.14}$$

The intrinsic correlation of the CO signal with other tracers, i.e. Ly-α forest or galaxies, is summarised by the cross-correlation coefficient between the noiseless fine-gridded signals:

$$r(k) = \frac{P_{Co\times A}(k)}{\sqrt{P_{CO}(k) P_A(k)}}\,. \tag{3.15}$$

3.5 Results

This section presents our findings on the cross-correlation signal between various current or planned Ly-α forest surveys and the COMAP-Y5 intensity map. We adopt the mean of the gaussianized prior provided in Table 5 of [31] as our fiducial set of CO emission model parameters. Based on the presented forecasts in this section, the cross-correlation of Ly-α tomography with line intensity mapping is expected to provide a competitive alterna-

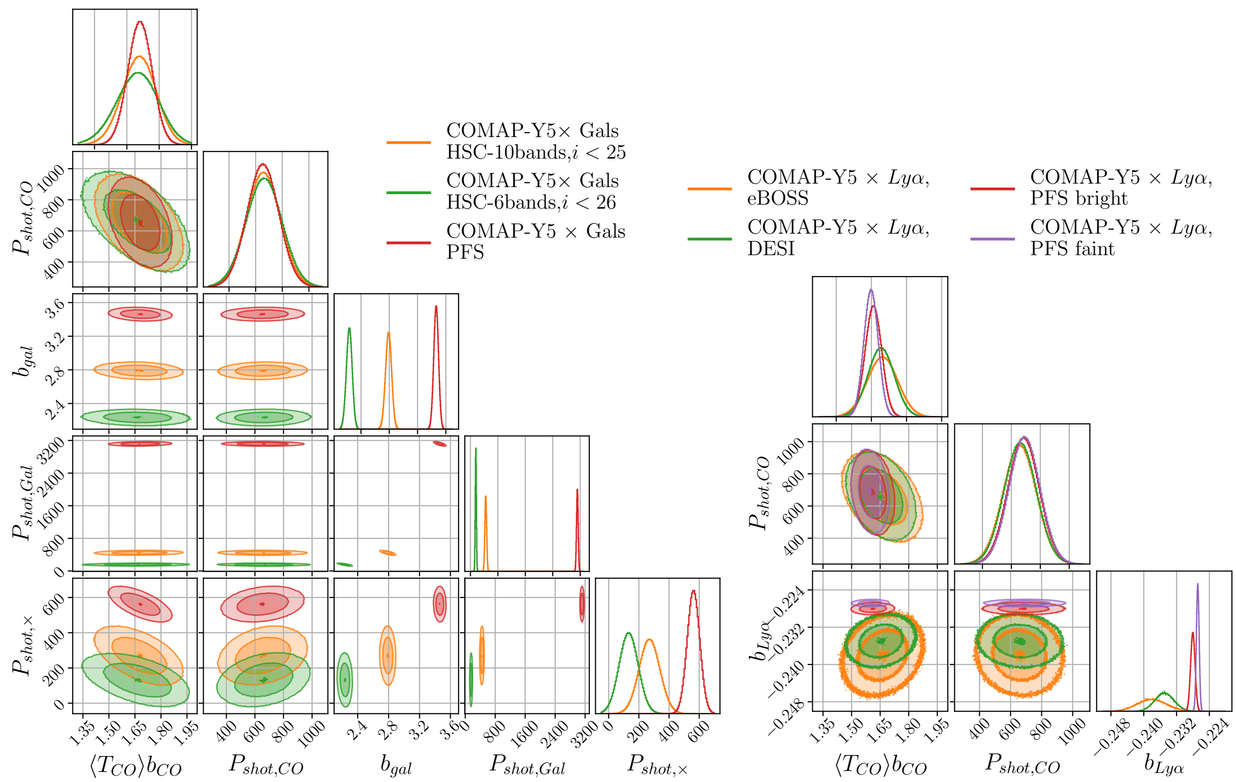


Figure 3.2: (**Left**) Forecasts for the inferred parameters in COMAP-Y5 × galaxy and (**Right**) COMAP-Y5 × Ly-α tomography. See Sections 3.5 and 3.6.

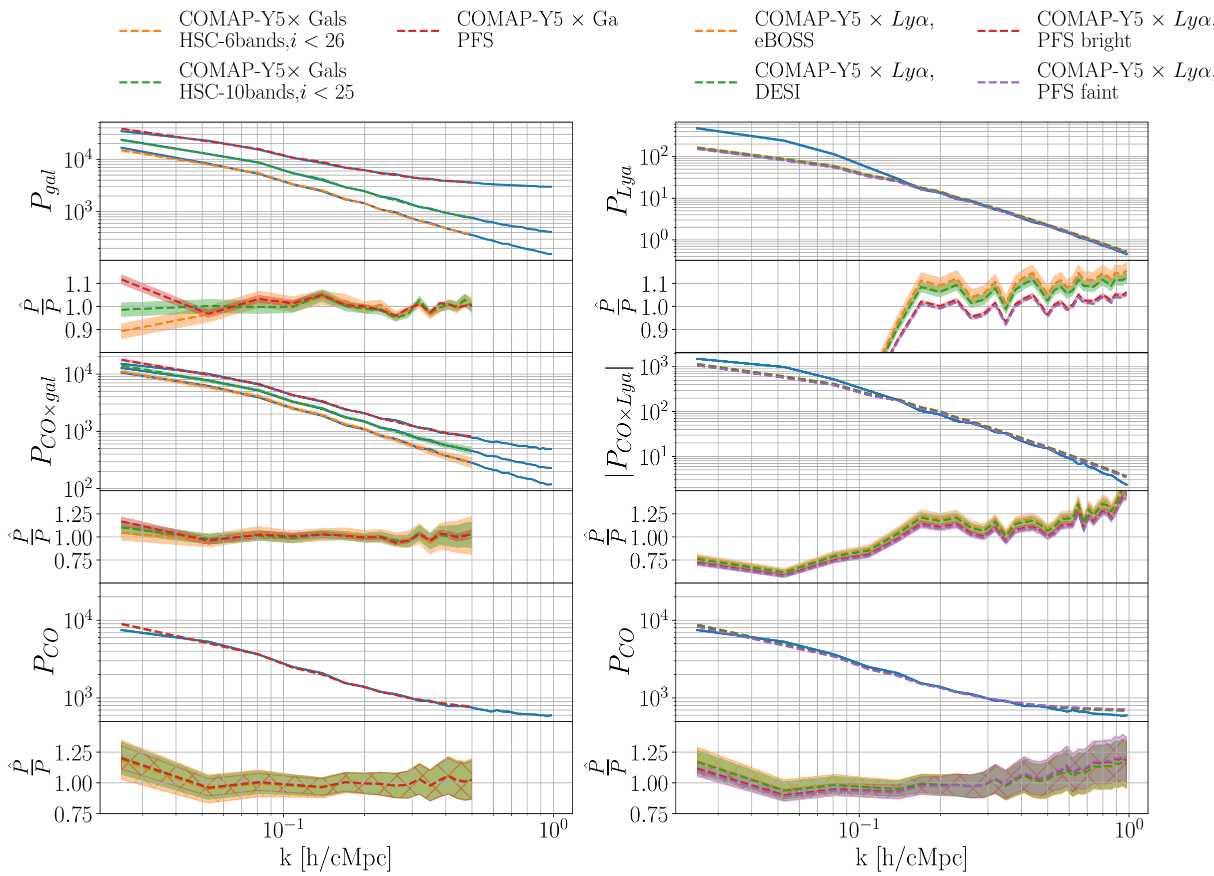


Figure 3.3: Comparison of the posterior prediction (dashed curves) and the signal power spectra (solid curves)with $1-\sigma$ uncertainties indicated by shaded regions. The left panel displays results for CO × galaxies with three separate curves for the power spectra signals due to varying mass completeness assumptions, which correspond to different magnitude cuts. The smallest scales, i.e. $k > 0.5\,h^{-1}$cMpc, are excluded from the inference due to the linear model inadequacies in describing the signal on the smallest scales. The right panel exhibits a significant deviation between the posterior prediction and signal for $P_{Ly\alpha}$ at $k < 0.1\,h^{-1}$cMpc attributed to bubbles of enhanced HII fraction formed during the HeII reionization. Further discussion is provided in Appendix 3.8.

tive to the cross-correlation with future galaxy redshift surveys for increasing the sensitivity to the CO signal. We assume all auxiliary surveys will fully overlap with the COMAP-Y5 volume. For a more detailed discussion and interpretation of the results, please refer to Section 3.6.

3.5.1 Signal to noise

Figure 3.1 illustrates the simulated signals of COMAP-Y5 cross-correlated against Ly-α forest or conventional photometric and spectroscopic galaxy surveys. In the left panel, two such noiseless signals are compared, that is, a Ly-α tomography map with a spatial resolution of $(250\,h^{-1}\mathrm{ckpc})^3$ and a few galaxy surveys with perfect redshift estimation and different mass completeness. Galaxies are better indicators of CO emission on smaller scales compared to Ly-α absorption. This could be because the CO lines originate from the interstellar medium (ISM) within galaxies, whereas Ly-α absorption traces larger scales in the intergalactic medium (IGM). On large scales with $k < 0.1\,h\mathrm{Mpc}^{-1}$, the forest becomes less correlated with the underlying matter density due to HeII reionization modeled in `ASTRID` [16,]. HeII reionization causes extra absorption on scales larger than $30\,h^{-1}\mathrm{cMpc}$ around the densest regions [124, 123, 156, 157, 69,]. For a detailed discussion on the HeII reionization effect, refer to Appendix 3.8.

Once realistic completeness and noise models are accounted for, Ly-α tomography surveys become competitive. The middle panel in Figure 3.1 compares the forecast S/N for various surveys with an observed volume similar to COMAP-Y5 Pathfinder. The simulated volume is scaled to match the COMAP-Y5 pathfinder by modifying the term $\sqrt{N_{modes}} \propto \sqrt{V}$ in Eq. 3.4. This scaling is justified since `ASTRID`'s volume has sufficient modes for power spectrum calculations on the scales to which COMAP-Y5 Pathfinder is most sensitive, i.e., $k = [0.05 - 0.6]\ h\mathrm{Mpc}^{-1}$ [80,]. The observed signal beyond this range is significantly contaminated and down-weighted during the pre-processing of the COMAP observations [62,].

In Figure 3.1, the total forecasted signal-to-noise ratio (S/N) is shown in the right panel, estimated using eq. 3.14. The sensitivity to the cross-correlated CO emission with a dense Ly-α tomography, where $d_\perp = 2.5 - 3.7\,h^{-1}$cMpc, is comparable to the cross-correlated field with a medium-resolution spectroscopic galaxy survey. Furthermore, coarse tomography surveys with $d_\perp = 10 - 13\,h^{-1}$cMpc show better sensitivity enhancement than most typical photometric galaxy surveys. The next section discusses the implications of a higher S/N for characterizing the clustering of cosmological line emission.

3.5.2 Forecast parameter inference

To forecast the inference on the clustering power of the CO emission, we model the simulated signal as biased tracers of the linear matter fluctuations in physical space:

$$\hat{P}_{CO} = (\langle T_{CO}\rangle b_{CO})^2\ P_m(k) + P_{shot,CO} \tag{3.16}$$

$$\hat{P}_{Gal} = b_{Gal}^2\ P_m(k) + P_{shot,gal} \tag{3.17}$$

$$\hat{P}_{Lya} = b_{Lya}^2\ P_m(k) \tag{3.18}$$

$$\begin{aligned}\hat{P}_{CO\times Gal} &= b_{Gal}\langle T_{CO}\rangle b_{CO}\ P_m(k) \\ &\quad + P_{shot,CO\times Gal}\end{aligned} \tag{3.19}$$

$$\hat{P}_{CO\times Lya} = b_{Lya}\langle T_{CO}\rangle b_{CO}\ P_m(k) \tag{3.20}$$

These equations do not account for redshift-space distortions, even though it is present in the simulated signal. The CO $\times$ Galaxy survey model requires five parameters, namely $\langle T_{CO}\rangle b_{CO}, P_{shot,CO}, b_{gal}, P_{shot,Gal}, P_{shot,CO\times Gal}$, while the CO $\times$ Ly-α tomography model requires three parameters, namely $\langle T_{CO}\rangle b_{CO}, P_{shot,CO}, b_{Lya}$. Shot noise terms are necessary for modelling the auto galaxy, CO, or their cross-power spectra due to the discrete nature of sources. The $P_{shot,CO\times Gal}$ term contains exclusive information on the CO emission strength of the sample galaxies in the cross-correlated survey compared to other shot noise terms [10,]. In contrast, due to the continuum nature of the HI gas density within the IGM, no such terms are necessary for modeling the forest power. A Gaussian likelihood is assumed for joint analysis in each scenario:

$$\mathcal{L} \sim \frac{1}{N_k}\sum_i\sum_k \mathcal{G}\left(P_i(k) - \hat{P}_i(k), \sigma_{P_i}(k)\right), \tag{3.21}$$

where $P_i(k)$ and $\sigma_{P_i}(k)$ are the simulated noiseless power spectrum and the estimated observational uncertainties, respectively (refer to Section 3.4). Here, i iterates over all available auto and cross-correlated signals in each scenario. We draw MCMC samples[3] from the posterior distribution obtained from the likelihood above, assuming flat priors.

The posterior constraints on all inferred parameters are shown in Figure 3.2. In Figure 3.3, the maximum a posteriori predictions for all modeled auto and cross-power spectra are compared to the simulated signal. Due to the different mass completeness thresholds for the spectroscopic and photometric surveys, three sets of simulated galaxy

[3] using `PyStan` python package [169, 22,]

power spectra are shown on the left panel of Figure 3.3. Tighter constraints on the model parameters from the Ly-α forest are observed, resulting in smaller uncertainties in the posterior power prediction compared to the galaxy surveys.

For galaxy surveys, we exclude non-linear scales (i.e. $k > 0.5\,h^{-1}$cMpc), from the inference as they are not modeled well by the linear theory in Eq. 3.17. Removing the smallest scales is necessary to avoid a systematic offset in the parameter inference; nevertheless, the width of the posterior remains unchanged after this cut since the S/N is already small on small scales (refer to Figure 3.1). The k-range of $[1/250 - 0.5]\,h\mathrm{Mpc}^{-1}$ adopted in our CO $\times$ galaxy analysis roughly matches the range constrained by the COMAP Early science results [31,]. For the inference from the CO $\times$ Ly-α signal, however, we use the full k-range of $[1/250 - 1.0]\,h\mathrm{Mpc}^{-1}$.

Moreover, we find that there is an excess of Ly-α absorption signal on the largest scales, $k < 0.1\,h\mathrm{Mpc}^{-1}$. We attribute this to the HeII reionization modeled in `ASTRID`, which enhances absorption on scales $L > 30\,h^{-1}$cMpc. This leads to a larger correlation in the signal at $k < 0.1\,h\mathrm{Mpc}^{-1}$[124, 123, 156, 157, 69,]. The linear matter power model in eq. 3.17 does not capture these effects in the Ly-α signal on large scales. Incorporating the model uncertainties into $\sigma_{P_{Ly\alpha}}$ is not straightforward, so we do not attempt to do so in this work. Consequently, this model insufficiency results in discrepant inferred values for the Ly-α bias parameter, $b_{Ly\alpha}$, across surveys with varying background source densities. We recommend using accurate simulation-based modeling, such as emulators, over biased linear perturbation for inferring from future observations. Emulator-based models [59,] offer a more reliable and precise approach to handling the complexities of large-scale structures.

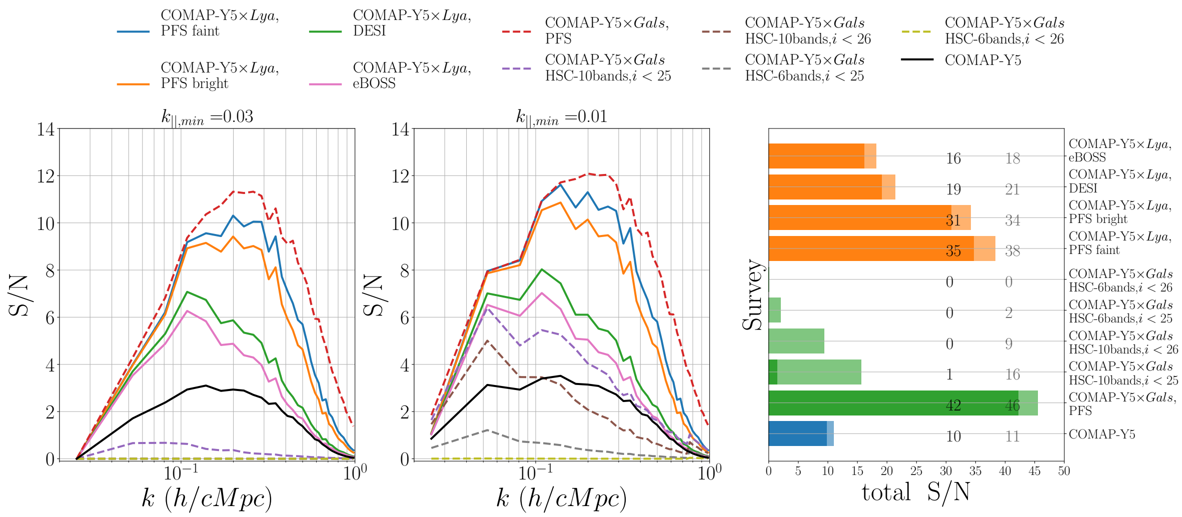


Figure 3.4: The forecast signal-to-noise ratio (S/N), taking into account contamination from continuum foreground emission along the line of sight. Two different values of the minimum parallel wave number, $k_{||,min}$, are considered (0.01, 0.03) $h\mathrm{Mpc}^{-1}$, above which the CO power is assumed to be contaminated by interlopers. The exact k-cut is observationally unconstrained. In broad-band photometric galaxy surveys, only the parallel modes $k_{||,min} < 0.02\,h\mathrm{Mpc}^{-1}$ are typically measured,and these modes are the most contaminated by continuum foreground emission. The right panel shows the total S/N, with different color intensities indicating different values of $k_{||,min}$.

We further discuss the details of the Ly-α power spectrum in Appendix 3.8.

3.6 Discussion

Figure 3.1 presents the predicted signal-to-noise ratios for CO $\times$ Ly-α and CO $\times$ galaxies. The surveys with high spatial resolution, such as the Ly-α tomography surveys with dense background sources and medium-resolution spectroscopic galaxy redshift surveys planned with the upcoming Prime Focus multiplex Spectroscopy survey (PFS), are expected to enhance the detection sensitivity of COMAP-Y5 by approximately 200-300%. Spectroscopic

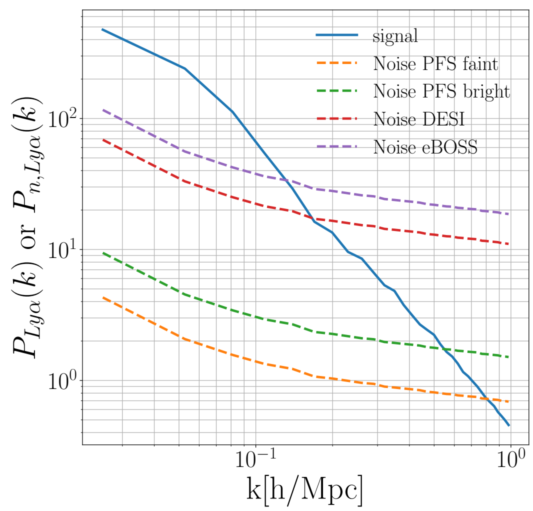


Figure 3.5: Comparing the impact of the cosmic variance and the finite background source density on the S/N of the Ly-α forest observations. For eBOSS and DESI, the S/N is primarily affected by the finite sightline density on scales $k > 0.15\,h^{-1}$cMpc. Meanwhile, for PFS observations, the S/N is mostly dominated by cosmic variance at scales where $k < 0.4$ or $0.5\,h^{-1}$cMpc.

galaxy surveys are slightly more effective in enhancing the detection S/N on smaller scales, probably because galaxies are inherently better tracers of the line emission from the ISM. This is further supported by the cross-correlation coefficient presented in the left panel of Figure 3.1. We defer a thorough examination of this behavior to future work.

The tomography surveys with a lower background source density, such as the present-day eBOSS observations or the forthcoming DESI survey, improve the sensitivity by $50-75\%$. On the other hand, a galaxy photometric catalog, such as CLAUDS observed in `u+grizy` bands [173,], only marginally improves the detection sensitivity by $\sim 15\%$. To match the S/N of COMAP-Y5 $\times$ galaxies to the CO $\times$ Ly-α forest in eBOSS or DESI, additional near-IR photometry in `YJHK` bands is required. Currently, `u+grizy+YJHK` photometric observations are limited to smaller areas (see Table 3.1). Moreover, the CO emission

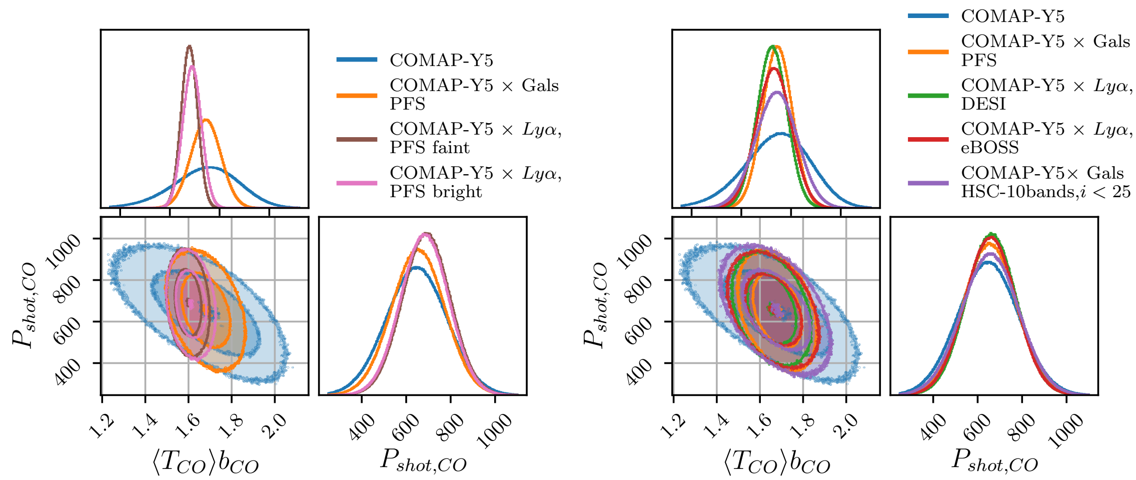


Figure 3.6: The constraints on CO bias and shot-noise in the linear bias power spectrum formalism. Joint analyses, including both Auto and cross-power spectra, provide tighter constraints. The left panel demonstrates that the constraints from CO × a PFS-like Ly-α tomography survey are tighter compared to a CO × PFS-like spectroscopy survey (refer to Section 3.6 for more details). The right panel shows that CO × an eBOSS or DESI-like Ly-α survey yields tighter constraints than CO × HSC-like photometric galaxy surveys. The medians and standard deviations of these distributions are tabulated in Table 3.2. Also, refer to Figure 3.2 for the posteriors on the complete set of parameters.

signal is susceptible to large-scale contamination caused by foreground continuum emission from Milky Way, which weakens the cross-correlated S/N. This contamination increases the CO noise power, $\sigma_{P_{CO,noise}}(k)$, on large parallel modes. The exact relevant scale where this contamination becomes prominent is not understood yet. The CO $\times$ photometric galaxy survey is, however, affected the most by foreground contamination as the signal originates mostly from the largest line-of-sight modes ($k_{||,min} < 0.02\,h\mathrm{Mpc}^{-1}$) due to the large galaxy redshift uncertainties. Fig 3.4 shows the impact of this contamination on the forecast S/N, where the contamination is modeled by expanding the CO noise power, $P_{CO,noise}$, on scales larger than $k_{||,min} = 0.01$ or $0.03\,h\mathrm{Mpc}^{-1}$ by a large factor of 10^9. These scale cuts are conservative compared to those proposed by [80], which constrains the CO auto power on scales $k = 0.051 - 0.062\,h^{-1}\mathrm{cMpc}$.

We quantify the trade-off between finite sightline density and cosmic variance for the Ly-α forest observations by comparing the noise power spectrum and the signal for Ly-α tomography in Figure 3.5, i.e. $P_{Lya}(k)$ vs $P_{Lya,n}(k)$. Our results show that for eBOSS and DESI, the primary factor affecting the cross-power spectrum signal-to-noise ratio (S/N) at $k > 0.15\,h^{-1}\mathrm{cMpc}$ is the finite background source density. In contrast, for PFS forest observations, cosmic variance has the greatest impact at $k < 0.4$ or $0.5\,h^{-1}\mathrm{cMpc}$.

As detailed in Section 3.3.3, we adopt our fiducial CO emission model as the mean of the gaussianized covariance provided by COMAP-Early-Science [31,]. However, this line emission model is still poorly constrained. To assess the sensitivity of our findings, we varied the model parameters within the 1σ range reported in Table 5 of [31]. Interestingly, we observed that the rank ordering of the cross-power S/N forecast remained consistent with that of our fiducial emission model. Nevertheless, we noted that when the line emission is particularly strong, as seen with larger C, A or smaller B, M, the auto CO signal itself has a higher S/N, diminishing the utility of cross-correlation with coarse Ly-α tomographies or photometric galaxy surveys.

The EXCLAIM experiment [193,] will observe the [CII] emission at cosmic noon and is designed to overlap with BOSS stripe 82, in order to benefit from [CII] $\times$ BOSS QSO sample [160,]. However, on the scales where noise dominates over cosmic variance (refer to Figure 3.5), we can compare the detection signal-to-noise ratio of [CII] $\times$ QSO and [CII] $\times$ Ly-α forest data, that is eBOSS observations [167,], as:

$$\frac{(S/N)_{CII\times Ly\alpha}}{(S/N)_{CII\times QSO}} \quad \sim \quad \frac{b_{Ly\alpha}}{b_{QSO}}\left(\frac{P_{QSO,noise}}{P_{Ly\alpha,noise}}\right)^{1/2} . \tag{3.22}$$

Assuming a quasar bias factor and number density of $b_q = 3.64$ and $n_{QSO} = 10^{-6}(\,h^{-1}\mathrm{cMpc})^3$ [60, 53,], Ly-α bias factor of $b_{Ly\alpha} = -0.20$ [177,], we find the signal-to-noise ratio of [CII] $\times$ Ly-α forest wins over cross-correlation against quasars by a factor of 10 or larger.

[35] measure the Ly-α emission $\times$ Ly-α forest cross-correlation and the Ly-α emission $\times$ quasar cross-correlation function. The latter quantity is well detected, while the authors place an upper bound on the former cross-correlation. The upper bound limits the total Ly-α luminosity density from surrounding star-forming galaxies. On the other hand, [35] suggest that the *detection* of diffuse Ly-α emission around quasars (mostly from relatively close to the quasars at $1 - 15\,h^{-1}$cMpc) may largely result from reprocessed emission from the quasars themselves, rather than sourced by surrounding star-forming galaxies. On the

other hand, [117], a more recent reanalysis using SDSS DR16, proposes the bulk of this diffuse Ly-α emission originates from the star-forming center of galaxies or the diffuse gaseous halos around them and not from the QSOs. Similarly, our work argues that the CO $\times$ Ly-α forest could provide insights into the origin of the CO emission, which has not yet been tightly constrained.

Clustering of the background sources in the Ly-α forest observations, such as quasars for eBOSS/DESI or LBGs for PFS surveys, increases the noise power spectrum, $P_{Ly\alpha,noise}$, by a factor of $1 + C_q(\vec{k}\perp)n_{2D}$, where $C_q(\vec{k}\perp)$ and n_{2D} are the angular power spectrum and projected density of the sightlines contributing to the signal at a given redshift [125,]. To investigate this issue, we measured the sightline clustering in the largest observed dense Ly-α tomography map, LATIS [137,], which has a mean sightline separation comparable to PFS-faint. Our analysis reveals that the noise power spectrum in LATIS is only slightly elevated by a few percent at $k = 0.03 - 0.1\,h\mathrm{Mpc}^{-1}$, the scales relevant to our study. This increase is expected to be even smaller for Ly-α surveys with lower sightline densities, such as eBOSS and DESI. As a result, the source clustering is not expected to affect our forecast for the signal-to-noise ratio of the cross-correlated signal with CO emission.

The results from Section 3.5.2 are summarized in Figure 3.6 and Table 3.2. The left panel demonstrates that a PFS-like Ly-α tomography survey, with a transverse separation of $d_\perp = 2.5$-$3.7\,h^{-1}$cMpc, is more effective at constraining the line emission power spectrum model than a spectroscopic galaxy survey planned with the same instrument, with $\frac{\sigma_z}{(1+z)} = 7 \times 10^{-4}$. The right panel shows that coarser Ly-α tomography surveys, such as DESI with a larger mean sightline separation of $d_\perp = 10\,h^{-1}$cMpc, provide better constraints than

HSC-like photometric galaxy surveys observed in 10 bands (`u+grizy+YJHK`) with a redshift precision of $\frac{\sigma_z}{(1+z)} = 0.02$. These constraints are even comparable to those obtained from spectroscopic surveys. This is because the Ly-α power spectrum has a larger signal-to-noise ratio and lower dimensionality of parameter space compared to galaxy surveys, where there are additional $P_{shot,gal}$ and $P_{shot,\times}$ parameters absent in the CO $\times$ Ly-α model, as shown in equation 3.17. Figure 3.6 illustrates that *photometric* galaxy surveys offer only marginal improvements to the constraints from COMAP alone, which was expected based on the low forecast signal-to-noise ratio shown in Figure 3.1.

The density of background sources in Ly-α forest maps and the accuracy of redshift estimations for galaxy surveys both decrease with increasing redshift. The eBOSS Ly-α forest survey has a mean transverse sightline separation of $8 - 16\,h^{-1}$cMpc within the redshift range of $z = 2.2 - 3.0$ [167,]. The Ly-α tomography IMACS Survey (LATIS), which is similar to future PFS tomography surveys, has a mean sightline separation of $2.3 - 3.0\,h^{-1}$cMpc within the redshift range of $z = 2.2 - 2.8$[137,] . This study covers a range of sightline separations that falls broadly within this range. To compare with actual observed data, it is necessary to simulate the exact survey characteristics in the mock data, such as the mean source density in Ly-α forest observations or the redshift estimation errors for galaxy surveys. In this work, we acknowledge that the auxiliary surveys (refer to Table 3.1) considered are assumed to fully overlap with the COMAP-Y5 volume, although this may not be the case for all observations. Nonetheless, we believe that incorporating this factor is straightforward, and we hope that our forecast will inspire future decisions regarding observations.

3.7 Summary

The LIM technique is a recent development for measuring the collective emission of specific atomic or molecular lines from galaxies of varying masses. However, detecting these faint signals still poses a challenge due to the necessary sensitivity. To improve the detection signal-to-noise ratio, cross-correlation with other large-scale structure tracers has been found to be effective, such as combining LIM with galaxy redshift surveys or other LIM experiments. In this study, we explore a promising new survey method with exceptional spatial resolution, Ly-α tomography. Ly-α tomography uses a dense sample of background galaxies (or quasars) to create a 3D map of neutral hydrogen in the intergalactic medium [125, 106,].

In particular, using large cosmological hydrodynamic simulations, we model the anticipated signal for the COMAP LIM experiment at the 5-year benchmark (Section 3.3.3). Our findings are expected to apply to other molecular LIM experiments with similar instrumental noise; however, one should adopt the appropriate model for the emission line observed in each particular LIM survey. We also made mock observations of the Ly-α absorption signal for fully observed tomography surveys, such as eBOSS [167,], and those expected to be completed in the coming years, such as DESI [28,] or PFS [71,] (Section 3.3.1). The key variable for Ly-α tomography surveys is the mean separation between the observed background sources. The map area covered by these surveys is comparable to or exceeds the coverage for COMAP-Y5. In this work, we assume these auxiliary maps fully overlap with the observed volume of COMAP-Y5. For a comprehensive list of survey parameters, including the volume coverage, please consult Table 3.1.

The findings of this study highlight the potential benefits of utilizing the Ly-α forest to aid in the initial detection of signals in line intensity experiments. The enhancement of the signal-to-noise for the cross-correlated CO emission with any auxiliary survey depends on the spatial resolution and the noise in the auxiliary data [32,]. The cross-correlation between COMAP-Y5 and the PFS Ly-α tomography survey will enhance the detection signal-to-noise by ~ 200 to 300%, comparable to medium-resolution spectroscopic galaxy surveys planned with the same instrument (Fig 3.1). The cross-correlation signal with sparser Ly-α tomography surveys, such as eBOSS and DESI, still enhances the detection S/N by 50 to 75%. Our results can be readily applied to actual existing data thanks to the observed quasar spectra in eBOSS Stripe 82, which covers an extensive area of 220 deg^2. Additionally, we demonstrate that the clustering of CO emission sources can be tightly constrained by the Ly-α tomography surveys. This is possible as a result of the elevated signal-to-noise ratio in the cross-correlation, as well as the uncomplicated nature of Ly-α absorption power spectrum modeling compared to the galaxy redshift surveys in CO $\times$ galaxies. However, a joint constraint on the emission clustering from CO $\times$ Ly-α and CO $\times$ galaxies is expected to provide further consistency tests on the inferred parameters. Our findings are presented in section 3.5 and 3.6, depicted in Figure 3.6, and summarized in Table 3.2.

It should be noted that any foreground contamination, or other systematics, that are unique to CO and Ly-α forest surveys, will not lead to biases in the inferences from the cross-spectrum signal. For example, residual foreground contamination might strongly bias a CO auto-power spectrum but will not lead to a spurious correlation with the Ly-α forest.

Table 3.2: Comparison of forecast inference results for CO emission clustering across different surveys. The complete posterior distributions are shown in Figures 3.2 and 3.6. Physical units are in μK and $\mu K^2(cMpc/h)^3$ for $\langle T_{CO}\rangle b_{CO}$ and $P_{shot,CO}$ respectively.

Survey	$\langle T_{CO}\rangle b_{CO}$	$P_{shot,CO}$
COMAP-Y5	1.68 ± 0.17	647 ± 136
COMAP-Y5 × Gals HSC-6bands, i ¡ 26	1.67 ± 0.13	663 ± 135
COMAP-Y5 × Gals HSC-10bands, i ¡ 25	1.68 ± 0.10	657 ± 128
COMAP-Y5 × Gals PFS	1.68 ± 0.08	654 ± 119
COMAP-Y5 × $Ly\alpha$, eBOSS	1.66 ± 0.09	657 ± 114
COMAP-Y5 × $Ly\alpha$, DESI	1.65 ± 0.08	662 ± 112
COMAP-Y5 × $Ly\alpha$, PFS bright	1.61 ± 0.05	684 ± 108
COMAP-Y5 × $Ly\alpha$, PFS faint	1.60 ± 0.04	691 ± 107

Section 3.6 presents a simple order-of-magnitude calculation, which suggests that the cross-correlation with Lyman-alpha forest fluctuations would also be beneficial for the EXCLAIM survey [193,]. EXCLAIM will observe $[CII]$ (1900 GHz restframe) emissions at cosmic noon and has significant overlap with BOSS quasars [160,].

In Section 3.6 and Appendix 3.8, we emphasize that precise modeling of the galaxy line emission and the Ly-α absorption signals is necessary for an accurate inference from actual data. Consequently, we defer a thorough emulator-based inference using cosmological simulations similar to that in previous studies [59, 17, 75,] to future work.

3.8 Robustness of the results to Simulation Choice

To verify the robustness of our results to the cosmological simulation used, we conduct the forecast analysis using a second cosmological hydrodynamic simulation, TNG300-1 [135,]. `ASTRID` has 1.8 times larger volume and increased particle mass resolution, as well as models for patchy hydrogen and helium reionization. Inference on the mock signal generated from the TNG300 simulation is presented in Figure 3.7. The posteriors on $\langle T_{CO}\rangle b_{CO}$ and $P_{shot,CO}$ and the rank order of the signal-to-noise ratio predictions for CO $\times$ Ly-α and CO $\times$ galaxies have remained unchanged. Unlike the results shown in Figure 3.2, the inferred $b_{Ly\alpha}$ from multiple surveys mocked using `TNG300` are now in agreement. This agreement can be attributed to the following.

The difference between TNG300 and `ASTRID` is because TNG300 does not incorporate a model for HeII reionization, which boosts power on scales comparable to the size of the reionized bubble ($30\,h^{-1}$cMpc). In `ASTRID`, guided by the radiative transfer simulations of [125], HeII reionization is modelled by the creation of $30\,h^{-1}$cMpc ionized bubbles around potentially quasar-hosting halos [196,]. Comparing the 3D auto power-spectrum of the Ly-α absorption signal in `ASTRID` with TNG300-1 in Figure 3.8 shows `ASTRID` predicts roughly an order of magnitude larger power at scales $k < 0.1\,h\text{Mpc}^{-1}$. This large-scale power enhancement is expected in any patchy reionization model, as discussed in [156, 157, 69].

However, patchy reionization is not included in either the linear bias model outlined in Eq. 3.17, nor TNG300, which thus agree with each other, although neither would agree with the true expected signal.

Damped Ly-α absorbers (DLAs) and Lyman Limit Systems (LLS) produce substantial absorption in the spectrum. The effect of the damping wings in these absorbers resembles the observed overshooting of power on large scales [171,]. The contribution of DLAs/LLS to the power spectrum can be reduced by masking these absorbers, as is also done in observational surveys [137,]. Figure 3.8 shows that there is still a power spectrum excess on large scales, even after masking all the absorbers with equivalent width larger

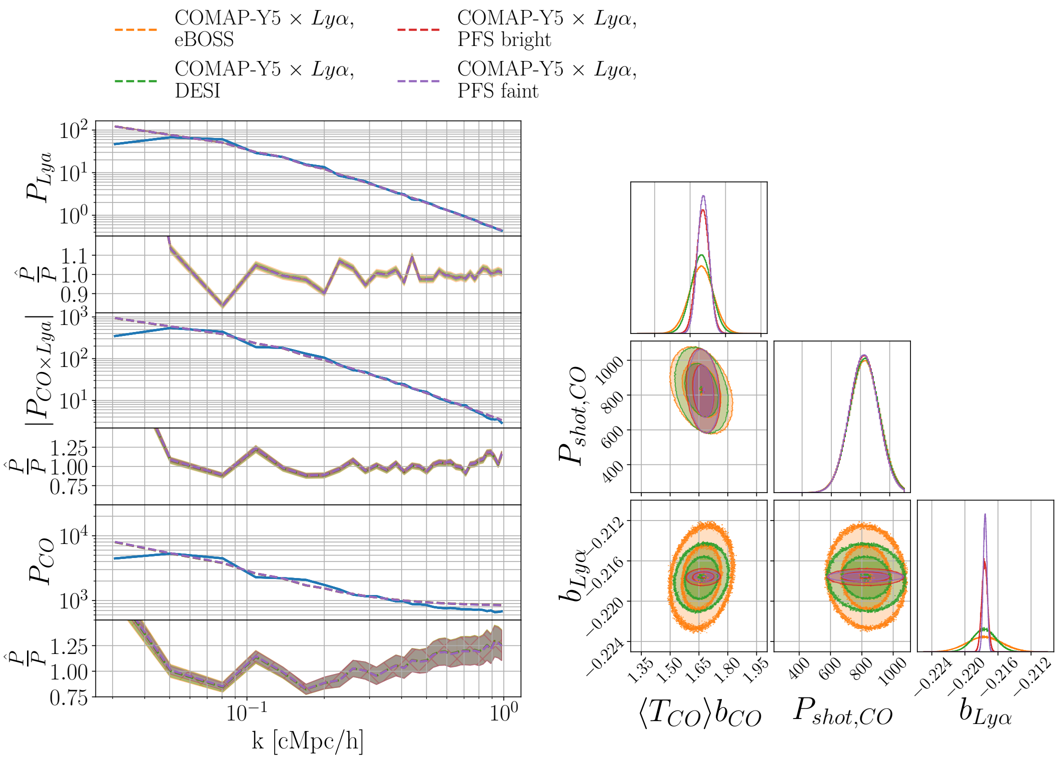


Figure 3.7: *Validation with an alternate cosmological simulation*: (**Left**) The maximum a posteriori power spectra. (**Right**) The inferred parameters for COMAP-Y × Ly-α tomography mock observations from the TNG300 hydrodynamical simulation. Compared to Fig 3.2, consistency in the inferred $b_{Ly\alpha}$ values is seen among surveys. The absence of HeII reionization in the TNG300 simulation allows a biased linear power spectrum to fit the largest scales well.

than $EW > 5$Å. This indicates that DLAs/LLS alone cannot fully account for the excessive power at large scales.

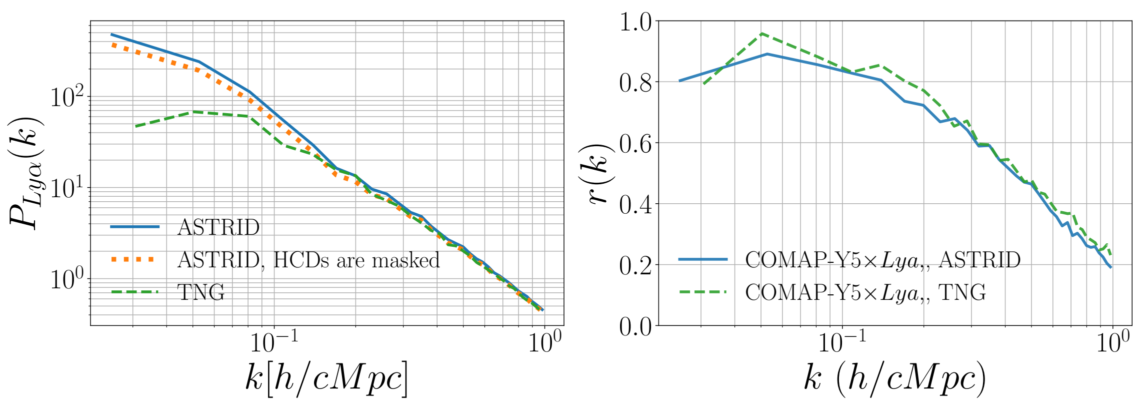


Figure 3.8: This figure demonstrates the effect of HeII reionization on the largest scales of the Ly-α power spectrum. **Left:** A higher power at $k < 0.1\,h\mathrm{Mpc}^{-1}$ is observed in ASTRID compared to TNG300. Masking the DLAs/LLS with $EW > 5$Å in ASTRID (dotted curve) does not substantially reduce the power on large scales. Thus, the excess power is due to the presence of ionized helium bubbles formed during HeII reionization. **Right:** The large-scale suppression of the cross-correlation coefficient due to HeII reionization is evident for ASTRID.

Chapter 4

An emulator for the 3D Ly-α forest power spectrum

Abstract

We present a multi-fidelity emulator for the non-linear 3D Ly-α forest power spectrum using the PRIYA simulation suite. This Gaussian Process (GP) emulator is trained on 48 low-fidelity and 3 high-fidelity simulations which vary 9 cosmological and astrophysical parameters. Our emulator achieves a 5% accuracy on scales $0.16 < k < 12\,h\mathrm{Mpc}^{-1}$ and redshifts $2.2 \leq z \leq 3.6$. We find that the 3D Ly-α power spectrum is significantly sensitive to the mean optical depth, $\bar{\tau}(z)$, and the primordial power spectrum parameters, A_p and n_p, in a non-degenerate manner. The HeII reionization parameters have a moderate impact on the 3D power, less than 8%, and are degenerate. Variations in the hydrogen reionization midpoint redshift, z_{Hi}, have a small effect on the power, $5-6\%$ only at higher redshifts

$z = 4$. Other cosmological parameters, the Hubble parameter, h, and the physical matter density, $\Omega_m h^2$, show much smaller impact, less than 5% on the power. The AGN feedback strength has the smallest impact on the 3D power spectrum. In a future work, using our emulator, we will infer these parameters from the publicly available BOSS Quasar survey. Our emulator code [1] and the training data[2] are publicly available.

4.1 Introduction

The absorption of the bright quasar light by the neutral hydrogen in the Intergalactic Medium (Ly-α forest, [72,]) is a tracer of the distribution of the underlying Dark Matter distribution, hence a powerful cosmological probe [36, 120, 200,]. The 3-dimensional correlation between the Ly-α absorption in the spectra of many QSOs has been used to measure the Baryon Acoustic Oscillations (BAO) peak to constrain the expansion history of the Universe and cosmological matter density, Ω_m [178, 179, 50,]. On smaller scales, averaging the Ly-α forest absorption power spectrum along many quasar spectra has been used to place powerful constraints on the primordial power spectrum by itself [58, 175,] or in combination with other probes including the Cosmic Microwave Background (CMB) [176, 145,]. The incorporation of both the large and small scale information from the Ly-α forest can improve the cosmological constraints due to access to larger data and the sensitivity to a larger set of parameters, i.e., the primordial power from smaller scales and Ω_m and H_0 from the BAO peak. This has been tested in [39] which improved the constraints on the matter

[1] GitHub: `https://github.com/sbird/lya_emulator_full`
[2] GitHub: `https://github.com/qezlou/3d_lya_power`

density Ω_m by a factor of two through incorporating the information from slightly wider scales ($25 < r < 180\, h^{-1}$cMpc) as opposed to the BAO peak alone.

The three-dimensional correlation function has primarily been used to estimate the correlations between observed QSO sightlines on large scales. Recent works have explored different methods for estimating the 3D power spectrum [61, 85,] from a discrete set of QSO sightlines with each spanning a different redshift range. For a consistent analysis of both the small and large scales, we consider the 3D power spectrum in this work which already contains the information from the 1D power spectrum along the line of sight.

Most previous large-scale forest correlation studies relied on perturbative models of the cosmological structure formation to model the Ly-α forest power spectrum. However, these models are not relieable on small scales. Using a suite of hydrodynamical simulations which varies a few cosmological parameters, [6] concluded that the non-linearities in modeling the 3D power spectrum of Ly-α forest are essential. They provided a parameterized fitting function between the linear power and the one estimated from the hydrodynamical simulations. Comparing these fits with other simulations, [67] finds the fits to be accurate at the 5% level. Testing these fits across a wider range of variations in cosmological and thermal history parameters than performed in [6] is required. In this work, for the first time, we build a simulations-based emulator for the non-linear 3D Ly-α power spectrum. The Gaussian Process (GP) emulator is trained on the `PRIYA` simulation suite [13,] which varies 4 cosmological parameters, 3 HeII reionization parameters, 1 hydrogen reionization, and 1 AGN heating strength feedback paramter. In light of recent success in constraining these parameters with the 1D Ly-α power spectrum [58,], we aim to provide a robust, percent-level accurate emulator for the 3D power spectrum applicable to BOSS [3, 50,] and DESI data [28, 70,]

The structure of this paper is as follows. In Section 4.2.1, we describe the PRIYA suite of simulations. Our methods for computing the 3D power spectrum from these simulation snapshots are described in Section 4.2.2. In section 4.2.3, we introduce our multi-fidelity emulator for the 3D power spectrum. We present our results in Section 4.3, including the accuracy of our emulator (Section 4.3.1) and the effect each cosmological and astrophysical parameter has on the 3d power spectrum (Section 4.3.2). We conclude in Section 4.4 and discuss the prospects for constraining the cosmological and astrophysical parameters with the existing and upcoming observed data.

4.2 Methods

4.2.1 Simulations

The work in this paper is based on the PRIYA simulation suite [13,]. The suite consists of 48 low-fidelity (LF) and 3 high-fidelity (HF) simulations. The LF simulations are run with 1536^3 particles in a $120\,h^{-1}$cMpc box size, and the HF simulations are run with 3072^3 particles with a similar box size. Using a Latin hypercube sampling, the simulations vary 9 cosmological and astrophysical parameters. The cosmological parameters include the amplitude and slope of the primordial power spectrum, A_p and n_p, the physical matter density, $\Omega_m h^2$, the Hubble parameter, h, the HI and HeII reionization parameters,and the strength of the heating from the AGN feedback, ϵ_{AGN}. The HeII reionization parameters include the redshift at which the HeII reionization starts and ends, z_{Hei} and z_{Hef}, and the strength of the heating from the sources, α_q. Moreover, midpoint redshift of the HI reionization, z_{HI}, is also varied. The range of parameter variations is listed in Table 4.1. The LF simulations are expected to capture the parameter sensitivities, and the resolution effects are corrected with the HF simulations.

The most relevant astrophysical processes at $z = 2 - 4$ are the temperature fluctuations from the ongoing HeII reionization. This is included through the patchy HeII reionization model implemented by [196] in the `MP-Gadget`[3] simulation code. Briefly, assuming the radiation from the QSOs is driving the HeII reionization, reionized bubbles are placed around randomly selected massive halos. The QSO's radiation spectral index, α_q, sets the amount of heating in the bubbles. The bubbles are of fixed size of $20\,h^{-1}$cMpc, and their counts at each redshift are determined such that the ionization fraction matches a linearly interpolated value between z_{Hei} and z_{Hef}.

The patchy HI reionization follows a semi-analytic model based on radiative transfer simulations [8,]. The outputof this model is a precomputed map of the reionization redshift at each voxel of size $1\,h^{-1}$cMpc in the simulation box. The gas in ach voxel, only after the corresponding reionization redshift is exposed to the Ultra-violet background radiation. To account for the effect from the reionization fronts, the temperature of the gas is boosted to $15,000K$ [40,]. As previously pointed out by [13], the inferred z_{HI} from the Ly-α forest at lower redshifts, $z < 6$, using this model, will be conditional on this temperature boost. For a detailed description of the `PRIYA` simulation suite, refer to [13].

[3] `https://github.com/MP-Gadget/MP-Gadget`

Table 4.1: The prior range of the cosmological and astrophysical parameters varied in the PRIYA simulation suite [13,]

Parameter	Prior range	Description
n_p	$[0.8, 0.995]$	Scalar spectral index
A_p	$[2.0, 3.0] \times 10^{-9}$	Primordial power amplitude at $k = 0.98\,h/\mathrm{Mpc}$
h	$[0.66, 0.75]$	Hubble parameter
$\Omega_m h^2$	$[0.141, 0.146]$	Physical matter density at $z = 0$
z_{HeI}	$[3.5, 4.1]$	HeII reionization start redshift
z_{HeII}	$[2.6, 3.2]$	HeII reionization end redshift
α_q	$[0.5, 1.5]$	Spectral index of QSOs driving the HeII reionization
z_{HI}	$[6.5, 8.0]$	HI reionization midpoint redshift
ϵ_{AGN}	$[0.03, 0.07]$	Thermal efficiently of black hole feedback
τ_0	$[0.75, 1.25]$	Mean optical depth at $z = 3$
$d\tau_0$	$[-0.25, 0.25]$	Mean optical depth redshift evolution

4.2.2 Computing the 3D Ly-α Forest Power Spectrum

In this work, we utilize the spectra generated from the `PRIYA` simulation suite in [13]. For a detailed description of the spectra generation process, we direct the reader to that paper. In summary, the Ly-α forest absorption signal from the simulations is generated using the publicly available `fake_spectra`[4] code. This code generates synthetic Ly-α forest spectra by computing the optical depth along sightlines through the simulation box. The contribution from all particles near the sightlines is accounted for through convolution with the Voigt profile estimated from the temperature, density, and velocity of each particle. Spectra are generated with sightlines across all three axes of the simulated box. The optical depth is scaled to match the mean flux constraints from [94]. Our model allows for variations

[4] `https://github.com/sbird/fake_spectra`

in these constraints by introducing two parameters, τ_0 and $d\tau_0$, which scale the mean optical depth at $z = 3$ and its redshift evolution, respectively:

$$\tau^{Kim}(z) = 0.0023 \times (1 + z)^{3.65} \tag{4.1}$$

$$\tau_{eff}(z) = \tau_o \times \left(\frac{1+z}{4}\right)^{d\tau_0} \tau_{Kim}(z) \tag{4.2}$$

The priors on these parameters are listed in Table 4.1. In practice, variations in $d\tau_0$ can be incorporated into the scaling of the mean optical depth at $z = 3$, τ_0; therefore, we only vary τ_0 within a slightly changed range of $0.70 - 1.28$.

The generated spectra are on regular grids with a resolution of $0.25\,h^{-1}$cMpc in the transverse and $\sim 10,$km/s along the line of sight. For the 3D power spectrum, we need the Ly-α absorption map to have a similar sampling rate along and transverse to the line of sight; therefore, we use Cloud-in-Cell (CIC) interpolation to downsample the spectra along the line-of-sight to $0.25\,h^{-1}$cMpc to match the transverse resolution. This absorption map on a regular grid is then fed to `Nbodykit` [74] software to compute the 3D power spectrum. The power is first computed with Fast Fourier Transform (FFT) in 3D dimensions and then binned in 2D (k, μ). We choose 10 uniformly sampled μ bins and adjust the k bins to make sure no bin is left with empty modes; hence, the k-bin size is $dk = 3 \times 2\pi/L_{\rm box}$, which is three times the fundamental mode of the box. The 3D power spectrum is then computed by averaging the power in each (k, μ) bin across all the sightlines. The entire pipeline for computing the 3D power spectrum is implemented in the `fake_spectra`[5] package.

[5]`https://github.com/sbird/fake_spectra`

4.2.3 Emulator

The structure of our multi-fidelity emulator for the 3D power spectrum is close to the emulator for the 1D power developed in [13] and [58]. We incorporated the 3D power spectrum emulator in the same package on GitHub[6]. All the (k, μ) bins are concatenated into a single vector, and each element of this vector is treated as a separate output. The input to the emulator is the 10 cosmological and astrophysical parameters varied in the `PRIYA` suite: 4 cosmological parameters, 3 HeII reionization parameters, 1 hydrogen reionization parameter, 1 for mean optical depth scaling factor, and 1 for the strength of AGN feedback.

One emulator for all (k, μ) bins accurately captures the correlation between the bins (Section 4.3.1), and training a separate emulator for each bin did not improve the accuracy of the emulator any further. Here, we briefly review the structure of the emulator and refer the reader to [13, 59, 75,] for a detailed description of the multi-fidelity emulator.

The emulator is trained on the low-fidelity simulations to capture the parameter sensitivities and then corrected with the high-fidelity simulations to capture the resolution effects. The low-fidelity emulator is a Gaussian Process (GP) drawing a vector from a multivariate normal distribution with a mean and covariance matrix:

$$P_{LF}^{3D}(k, \mu) \sim GP(\mu_\theta, k(\boldsymbol{\theta}, \boldsymbol{\theta'})) \tag{4.3}$$

[6]GitHub: `https://github.com/sbird/lya_emulator_full/tree/master`

Rescaling the 3D power to the median of the training sample helps set the prior on the mean of the GP, $\mu(\theta)$, to zero. The covariance is a combination of RBF and linear kernels:

$$k_{RBF}(\boldsymbol{\theta},\boldsymbol{\theta'};\sigma_0,\boldsymbol{l}) + k_{LIN}(\boldsymbol{\theta},\boldsymbol{\theta'};\sigma) = \sigma_0^2 \ exp\left(\sum_{i=1}^{d} -\frac{(\boldsymbol{\theta_i}-\boldsymbol{\theta'}_i)}{2l_i^2}\right) + \sum_{i=1}^{d}\sigma_i^2\boldsymbol{\theta_i}\boldsymbol{\theta'_i} \tag{4.4}$$

Where σ_0, σ_i, and l_i are hyperparameters to learn from the training samples. The length-scale l_i captures the sensitivity of the RBF kernel to the parameter θ_i, and we allow the emulator to find an independent value for each dimension. The multi-fidelity emulator is a linear combination of the low-fidelity emulator and another GP that incorporates parameter-dependent corrections to the low-fidelity emulator:

$$P_{HF}^{3D}(k,\mu,z;\boldsymbol{\theta}) = \rho(z).P_{LF}^{3D}(k,\mu,z;\boldsymbol{\theta}) + \delta(k,\mu,z;\boldsymbol{\theta}) \tag{4.5}$$

The GP $\delta(\boldsymbol{\theta})$ has a similar combination of RBF and linear kernels. The RBF, however, shares a single length scale, l, for all parameters. The scaling factor ρ_z and the shared length scale l are learned from the training samples. We train one separate emulator for each redshift bin from $z = 2.2$ to $z = 4.0$.

4.3 Results

The results of our multi-fidelity emulator for the 3D Ly-α power spectrum are presented in this section. In Section 4.3.1, we discuss the convergence of our simulations for the 3D Ly-α power spectrum. In particular, we compare the accuracy of our emulator (the

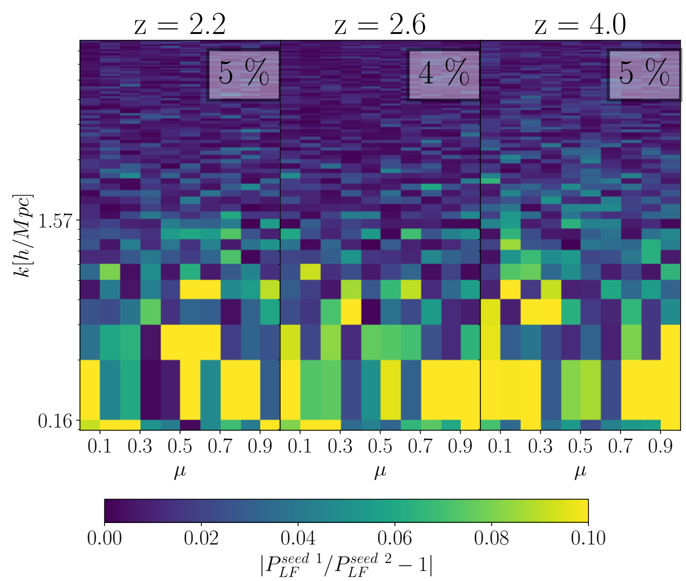


Figure 4.1: The cosmic variance effect on the 3D Ly-α forest power spectrum $P(k, \mu)$. The colormap shows the fractional difference between two low-fidelity simulations run with the same exact parameters but different random seeds, i.e., initial conditions. The 95th percentile of the fractional difference is printed in the top right corner of each panel. The largest scales $k < 0.5\,h\mathrm{Mpc}^{-1}$ are affected due to the different spatial distribution of the HeIII bubbles with each random seed.

leave-one-out errors) with the convergence level of the low-fidelity simulations and show our multi-fidelity emulator properly corrects for the majority of the differences between the low and high-fidelity simulations. In Section 4.3.2, we discuss the sensitivity of the 3D Ly-α power spectrum to the cosmological and astrophysical parameters and compare those to the accuracy of our emulator.

4.3.1 Convergence Test and the Accuracy of our Emulator

Figure 4.1 shows that the cosmic variance affects the 3D Ly-α power spectrum in the low-fidelity simulations by less than 5% across all (k, μ) bins. The panels compare the power in two LF simulations with the same cosmological/astrophysical parameters but run with different random seeds. Cosmic variance only affects the largest scales, $k < 0.5/hMpc$, and remains at the same level across redshift bins, $2.2 < z < 4.0$. The larger impact on the largest scales originates from different positioning of the HeIII bubbles in the two simulations. This effect was observed in the 1D Ly-α power spectrum to a lesser degree [13,]. This suggests better accuracy than 5% on the 3D power spectrum requires simulations with a larger volume to obtain a sufficient number of modes on large scales.

The comparison between the low-fidelity and high-fidelity simulations in Figure 4.2 shows that the particle resolution plays a more crucial role in the accuracy of the modeled 3D power spectrum. The LF power deviates from the HF simulations by $\sim 10 - 20\%$, depending on the parameters, at $z < 3.6$ and increases to $27 - 34\%$ at higher redshifts, $z \sim 3.6 - 4.0$. The power on the largest scales at high redshifts is the most affected by the resolution convergence, the scales and redshifts at which HeIII bubbles have a significant

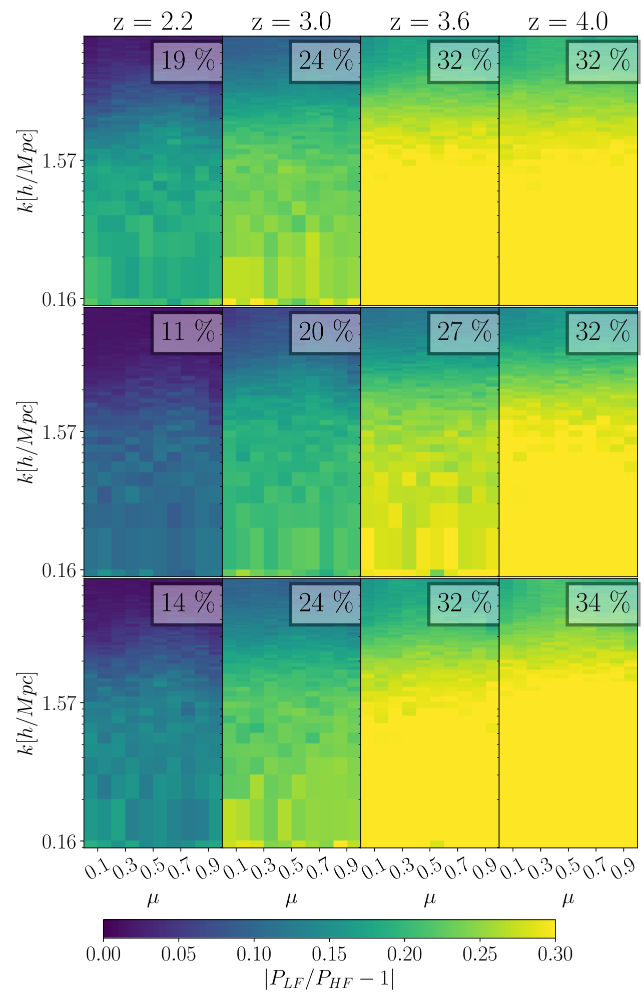


Figure 4.2: The convergence of the low-fidelity simulations for the 3D Ly-α power spectrum. Each row shows the difference between the high-fidelity and low-fidelity simulations at different redshifts. For simplicity, we only show the results for selected redshift snapshots. The low-fidelity simulations deviate the most on the largest scales at higher redshifts, where HeII reionization is happening. In Fig 4.3, we show that our multi-fidelity emulator properly learns to correct between LF and HF simulations.

impact on the signal (Appendix A in [161,] and [156, 69,]). This is a larger discrepancy between the low and high-fidelity simulations compared to the 1D Ly-α power spectrum studied in [13]. In Section 4.3.2, we will see that the 3D power is more sensitive to the modeling of the HeIII bubbles compared to the 1D power spectrum, where the latter is an averaged statistic across many sightlines. This may suggest that the proper modeling of the impact HeII reionization has on the 3D power requires high enough resolution to capture the small scale. We speculate the cooling of the gas in HeIII regions might require a higher resolution than the LF simulations. However, further investigation in a future work is required to confirm this hypothesis.

Nonetheless, our multi-fidelity emulator is capable of properly correcting for this resolution effect. In Figure 4.3, we show the leave-one-out errors of our multi-fidelity emulator. Each row compares the prediction of our MF-emulator trained on all but one high-fidelity simulations to the unseen high-fidelity simulation. The emulator is accurate to $1-6\%$ on all scales and redshifts $z < 3.8$ and slightly increases to $1-9\%$ at higher redshifts where the largest scales are harder to correct for due to the larger deviation between the low and high-fidelity simulations (compare to Figure 4.2). The three available HF simulations each have different cosmological/astrophysical parameters; therefore, the wide range of the leave-one-out errors suggests the resolution correction by the multi-fidelity emulator is parameter dependent. This accuracy can be improved by incorporating a few more HF simulations spanning different ranges of parameters.

In the next section, we examine the sensitivity of the 3D power spectrum to the cosmological/astrophysical parameters. Comparing those to the accuracy of our emulator

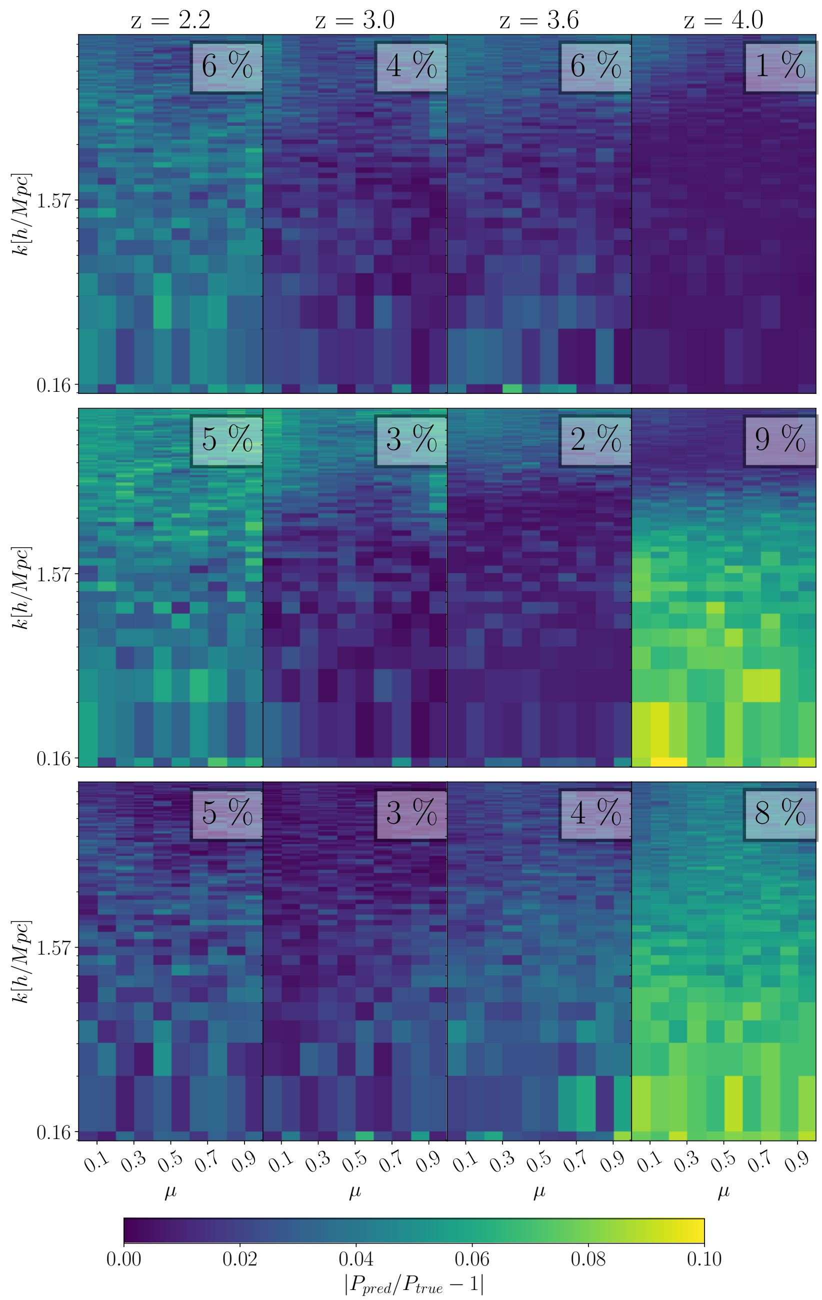


Figure 4.3: The accuracy of our multi-fidelity emulator for the 3D Ly-α power spectrum. Each row shows the leave-one-out errors obtained by comparing the multi-fidelity emulator's prediction and the unseen high-fidelity simulation. The box in the top right corner of each panel shows the 95th percentile of the fractional error. For redshifts $z < 3.8$, the emulator is accurate to approximately $1 - 6\%$ on all scales. At higher redshifts, the emulator's accuracy slightly drops due to the larger difference between the low and high-fidelity simulations (compare to Fig 4.2).

at each redshift will inform us if such sensitivities are reliably modeled and of the prospects for constraining these parameters with the upcoming observational surveys.

4.3.2 Effect of Parameters on 3D Ly-α Power Sepctrum

By training our multi-fidelity emulator on all the available low and high-fidelity simulations, we can now model the sensitivity of the 3D Ly-α power spectrum to the cosmological and astrophysical parameters. Each panel in Figures 4.4, 4.5, 4.6, and 4.7 illustrates the change in the power spectrum, $P(k, \mu)$, when we vary one parameter at a time. The colormap shows the fractional difference between the power estimated at a specific parameter value labeled in the top left corner to the fiducial values at the midpoint of the prior range. The 95th percentile of this fractional change across all (k, μ) bins is printed in the top right corner of each panel. The general trends are consistent with the 1D Ly-α power spectrum studied in [13]; however, the fractional changes introduced by these parameter variations are larger for the 3D power spectrum.

The impact of the mean Ly-α optical depth scaling is illustrated in the top panels of Figure 4.4. The fiducial value of $\tau_0 = 1$ is varied to 0.8 and 1.2. A larger optical depth scaling increases the power amplitude on all scales by approximately $38 - 66\%$. This effect is more pronounced at higher redshifts where there is a larger abundance of neutral hydrogen. The mean optical depth exhibits the most significant effect on the power spectrum compared to the other parameters and dictates the amplitude of the power with minimal scale-dependency. Varying the slope of the primordial power spectrum, $n_p = 0.82$ or 0.98 compared to the fiducial value of 0.898, results in a power change of about 10% (second row

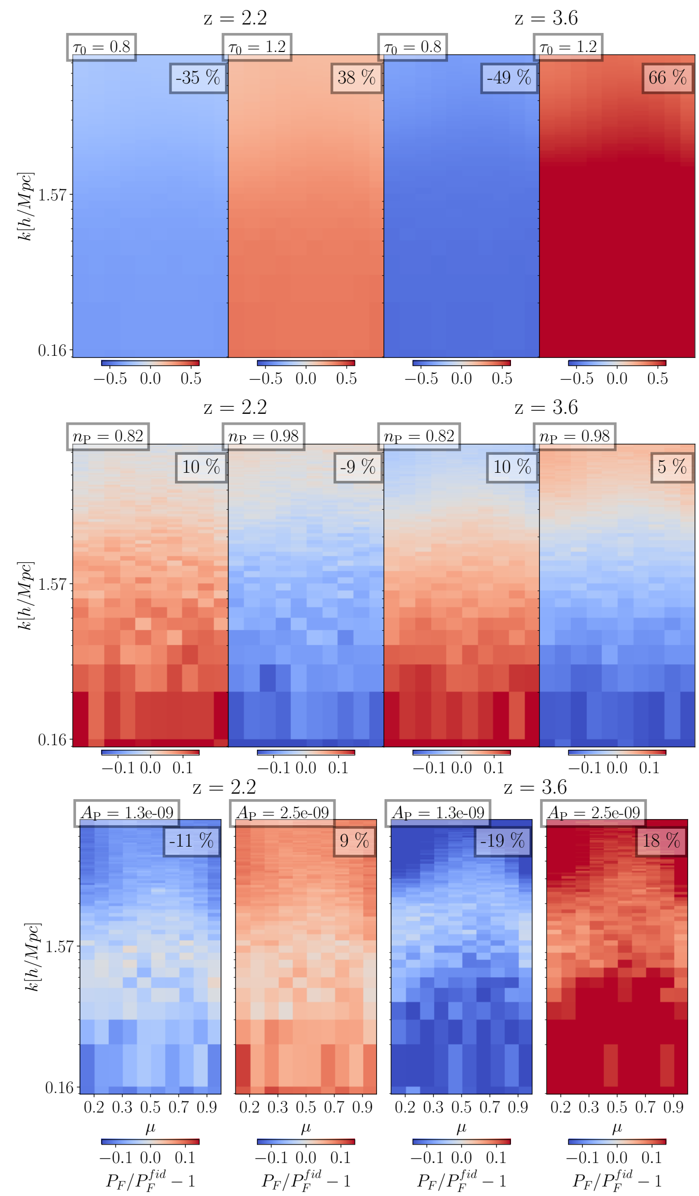


Figure 4.4: T The effect of cosmological parameters on the 3D Ly-α power spectrum is depicted. The colormap on each panel displays the fractional difference between the power estimated at a specific parameter value and the fiducial value (the midpoint of the prior range). The 95th percentile of this fractional change is printed on the top right corner of each panel. Two redshift samples are selected for each parameter. The mean optical depth scaling, τ_0, and the primordial power parameters have the largest impact on $P(k, \mu)$.

in Figure 4.4). This change is scale-dependent, where a smaller n_p increases the Ly-α power on larger scales and decreases it on smaller scales. Increasing the amplitude of the primordial power spectrum (the bottom panels of Figure 4.4) alters the power by approximately $\sim 20\%$, especially on the largest scales. As noted by [13], the scale-dependent effects of A_p and n_p on the power spectrum (both 1D and 3D) ensure they are non-degenerate with the mean optical depth, τ_0, which in turn determines the general scaling of the power spectrum. The effects from τ_0, A_p, and n_p are all larger than the $\sim 5\%$ accuracy of our emulator at similar redshifts, hence reliably modeled.

The variations of the Hubble parameter, $h = 0.66$ and 0.74 from the fiducial value of 0.7, and the physical matter density (the growth factor), $\Omega_m h^2 = 0.141$ and 0.145, compared to the fiducial value of 0.143, have a smaller impact on the power, as shown in Figure 4.5. The 5% and smaller changes are comparable to the accuracy of our emulator. [13] attributed the small sensitivity of the 1D power spectrum to changes in the h parameter to the unit conversion between comoving and observed coordinates, km/s. However, this cannot be the reason in our case since the computed 3D power spectrum in this work is in comoving coordinates, the native binning of the simulations. We speculate that the small sensitivity to h at higher redshifts could be attributed to the dependency of gas cooling rates on this parameter, which modulates the gas temperature; however, we postpone a detailed diognoisis to a future work.

The moderate effect of the HeII reionization parameters, less than 8%, is shown in Figure 4.6. Similar to the 1D power spectrum [13,], these effects are more subtle than those of τ_0, n_p, and A_p. These effects are largest at redshifts where HeII reionization is occurring, $z_{Hei} < z < z_{Hef}$. Variations in α_q and z_{Hef} primarily affect the largest scales, as previously observed for the 1D power spectrum [13,]. This scale-dependency originates from the temperature fluctuations across the box caused by the patchy HeII reionization [161, 157, 69,]. All these HeII reionization parameters impact the 3D power similarly, hence they are degenerate. One can break the degeneracies by incorporating extra information from the thermal history of the IGM, i.e., the mean temperature obtained from high-resolution spectra. This task is beyond the scope of modeling the 3D Ly-α power spectrum measurements but will be useful while performing inference on the observed data.

Similar scale-dependency is evident for the hydrogen reionization parameter, z_{Hi}. The top panel of Figure 4.7 shows that at higher redshifts, around $z \sim 4$, closer to the hydrogen reionization midpoint, z_{Hi}, the residual temperature fluctuations increase the power by about 6%, especially on the largest scales. This behavior was previously observed in the 1D power spectrum [13,] and radiative transfer simulations [129,]. However, observationally estimating power at these redshifts suffers from large statistical uncertainties due to the scarcity of observed QSOs at high redshifts. A detailed analysis of these observational factors is postponed to future work. Moreover, as discussed in Section 4.3.1 (Figure 4.2), our emulator's accuracy slightly drops at higher redshifts, $z = 3.8 - 4.0$, hence this effect may not be reliably modeled.

Finally, varying the AGN feedback strength (bottom panel of Figure 4.7) has the smallest impact on the 3D power spectrum, lower than the 5% accuracy of our emulator. For further discussion on its impact on the high column density absorbers, refer to [13].

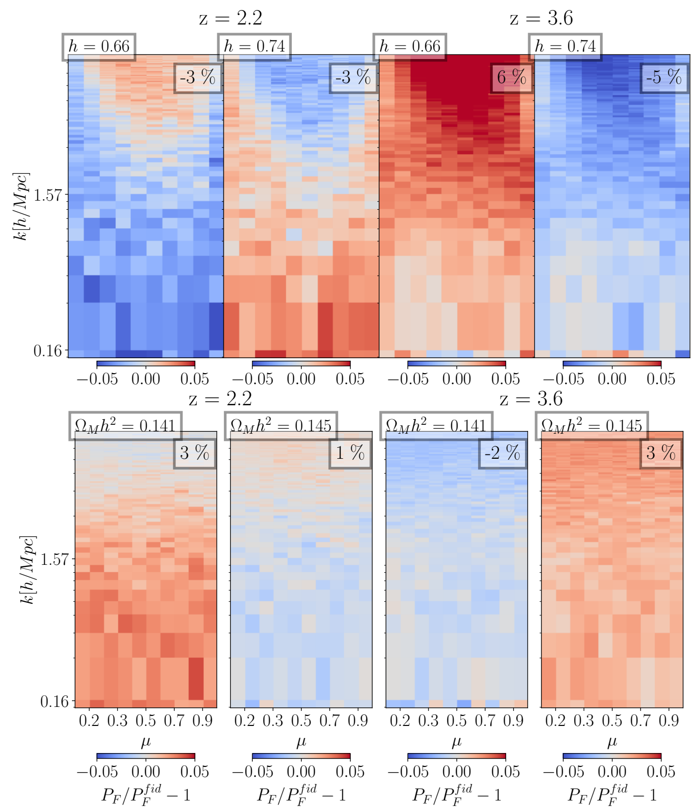


Figure 4.5: Smaller sensitivity of the 3D Ly-α power spectrum to other cosmological parameters, the Hubble parameter h, and physical matter density $\Omega_m h^2$. The $\sim 5\%$ and smaller sensitivity to these parameters is comparable to or smaller than the accuracy of our emulator.

However, a stronger AGN feedback model such as SIMBA may have a more pronounced impact on the 3D power spectrum [42,].

4.4 Conclusions

In this work, for the first time, we constructed a multi-fidelity emulator for the 3D Ly-α forest power spectrum. This Gaussian Process (GP) emulator was trained using the

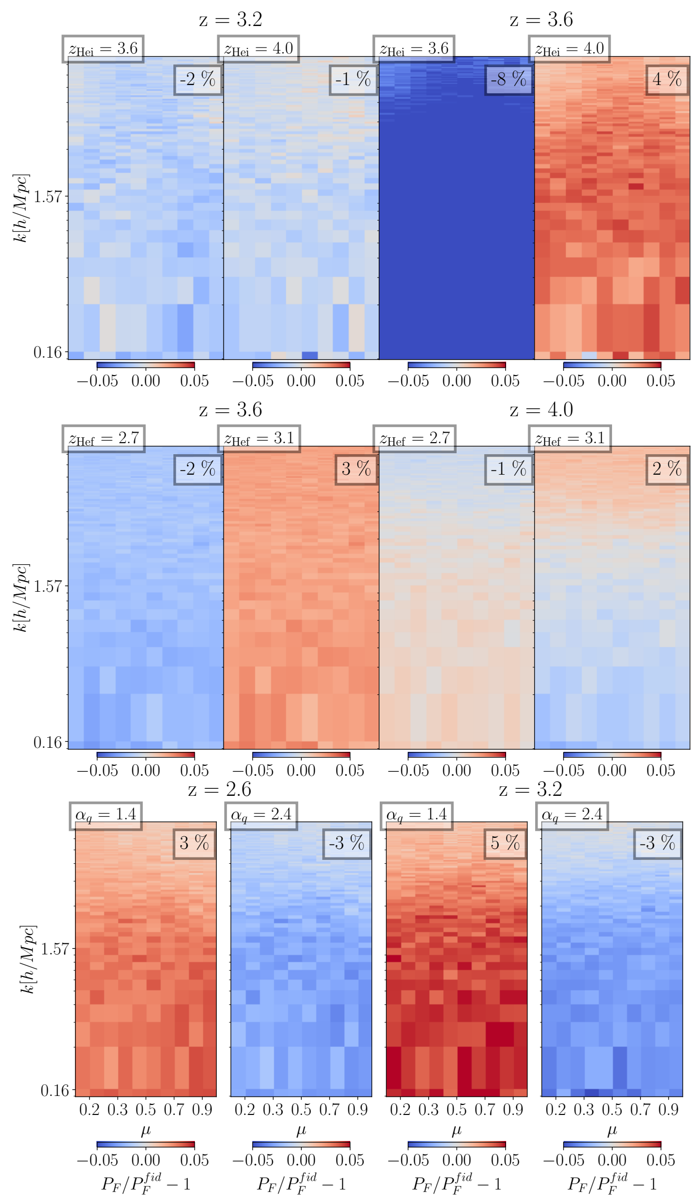


Figure 4.6: The moderate effect, $< 8\%$, of the HeII reionization parameters on the 3D Ly-α power spectrum. The redshift at which HeII reionization ends, z_{Hef}, and the strength of the heating from the sources, α_q, impact the largest scales the most.

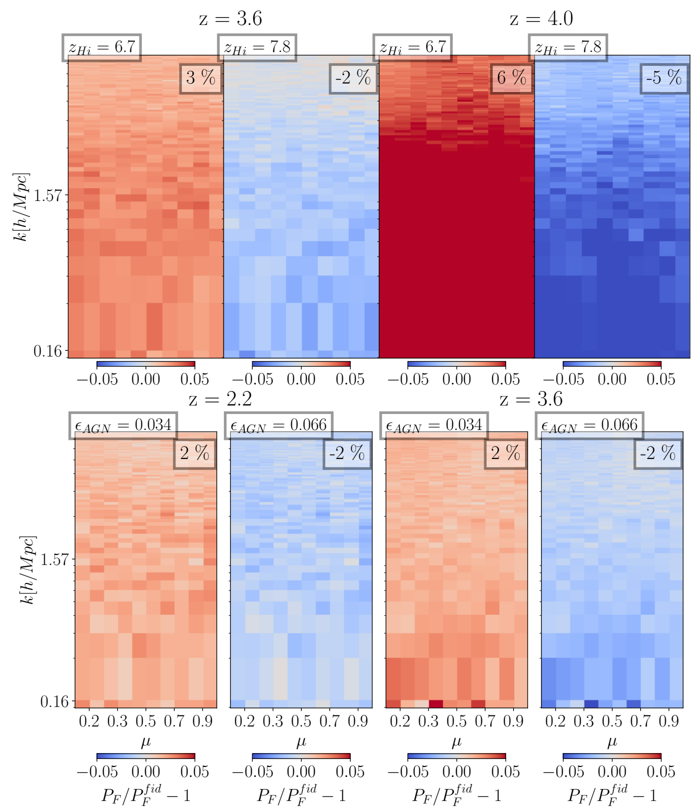


Figure 4.7: Small sensitivity of the 3D Ly-α forest power spectrum to the hydrogen reionization midpoint redshift, z_{Hi}, and the strength of AGN feedback, ϵ_{AGN}. At higher redshifts, $z \sim 4$, the residual temperature fluctuations from the hydrogen reionization increase the power by about 6%, especially on the largest scales. The AGN feedback has a marginal impact on the power.

PRIYA simulation suite [13], which comprises 48 low-fidelity and 3 high-fidelity simulations. These cosmological hydrodynamical simulations cover a 9-dimensional parameter space, incorporating 4 cosmological and 5 astrophysical parameters.

In Section 4.3.1 (Figure 4.1), we demonstrated the convergence of the 3D power spectrum, $P(k, \mu)$, obtained from simulations with a $120\,h^{-1}$cMpc box size to approximately 5% on the largest scales. This level of accuracy is adequate for measurements from the BOSS survey [50] and upcoming DESI data [70]. While running simulations with larger box sizes could help reduce the cosmic variance effect, the incorporation of another tier of simulations with different box sizes is deferred to future work. Additionally, we estimated the resolution convergence of the 3D Ly-α power estimated from low-fidelity (LF) simulations with 1536^3 particles by comparing to the power from the high-fidelity (HF) simulations with 3072^3 particles. The resolution convergence ranges from $10-20\%$ at lower redshifts ($z < 3.6$) to $27-34\%$ at higher redshifts (Figure 4.2). However, our multi-fidelity emulator effectively corrects for this resolution effect. The leave-one-out error of our emulator for the three available HF simulations achieves an accuracy ranging from $1-6\%$ at lower redshifts ($z < 3.8$) to slightly larger, $1-9\%$, at higher redshifts ($z = 3.8-4.0$) (Figure 4.3). These errors are only marginally larger than the 5% cosmic variance effect. The broad range of leave-one-out errors indicates that the map between the LF and HF simulations is parameter-dependent (cosmological and astrophysical), which could be improved by incorporating a few more HF simulations into the PRIYA suite in the future.

In Section 4.3.2, we investigated the effects of different parameters on the 3D Ly-α power spectrum. Our findings align with the 1D Ly-α power spectrum explored in [13], albeit with generally larger impacts for the 3D power spectrum. We found that the mean optical depth scaling, τ_0, determines the general scaling of the power spectrum. A 20% increase in the mean optical depth results in a 38% change in power at $z = 2.2$ and a 66% change at $z = 3.6$. The amplitude and slope of the primordial power spectrum, A_p and n_p, exhibit scale-dependent impacts on the 3D Ly-α power spectrum (Figure 4.4). Variations in these parameters from the midpoint of the prior range impact the 3D power by $10 - 20\%$. Similar to the 1D power spectrum [13,], this scale dependency ensures they are non-degenerate with the mean optical depth, τ_0. The HeII reionization parameters have a moderate impact on the 3D power, less than 8% (Figure 4.6), and are degenerate with each other. However, this degeneracy can be resolved by incorporating information from the thermal history of the IGM, such as the mean IGM temperature obtained from high-resolution spectra [65,], as previously done for the 1D power spectrum [58,]. Our emulator exhibits small sensitivity in the 3D Ly-α power to the hydrogen reionization midpoint redshift, z_{Hi}, especially at higher redshifts (Figure 4.7). This is attributed to the residual temperature fluctuations from patchy hydrogen reionization [129,]. The other cosmological parameters, the Hubble parameter, h, and the physical matter density, $\Omega_m h^2$, marginally affect the 3D power spectrum, with variations less than 5%. The strength of the heating form the AGN feedback has the smallest impact on the 3D power spectrum, less than 2%.

Our current emulator is sufficiently accurate to model the non-linear 3D Ly-α power spectrum observed in the BOSS and DESI datasets. We intend to conduct cosmological and astrophysical parameter inference using publicly available 3D power spectrum measurements from BOSS [50] and forthcoming DESI data [70,]. Our complete emulator, along with the training data and code, is openly accessible on GitHub.

Bibliography

[1] Elcio Abdalla and Guillermo Franco Abellán. Cosmology intertwined: A review of the particle physics, astrophysics, and cosmology associated with the cosmological tensions and anomalies. *Journal of High Energy Astrophysics*, 34:49–211, June 2022.

[2] Romina Ahumada, Carlos Allende Prieto, Andrés Almeida, Friedrich Anders, Scott F. Anderson, and Andrews. The 16th Data Release of the Sloan Digital Sky Surveys: First Release from the APOGEE-2 Southern Survey and Full Release of eBOSS Spectra. *ApJS*, 249(1):3, July 2020.

[3] Romina Ahumada, Carlos Allende Prieto, Andrés Almeida, Friedrich Anders, and Anderson. The 16th Data Release of the Sloan Digital Sky Surveys: First Release from the APOGEE-2 Southern Survey and Full Release of eBOSS Spectra. *ApJS*, 249(1):3, July 2020.

[4] Hiroaki Aihara, Yusra AlSayyad, Makoto Ando, Robert Armstrong, James Bosch, Eiichi Egami, Hisanori Furusawa, Junko Furusawa, Sumiko Harasawa, Yuichi Harikane, Bau-Ching Hsieh, Hiroyuki Ikeda, Kei Ito, Ikuru Iwata, Tadayuki Kodama, Michitaro Koike, Mitsuru Kokubo, Yutaka Komiyama, Xiangchong Li, Yongming Liang, Yen-Ting Lin, Robert H. Lupton, Nate B. Lust, Lauren A. MacArthur, Ken Mawatari, Sogo Mineo, Hironao Miyatake, Satoshi Miyazaki, Surhud More, Takahiro Morishima, Hitoshi Murayama, Kimihiko Nakajima, Fumiaki Nakata, Atsushi J. Nishizawa, Masamune Oguri, Nobuhiro Okabe, Yuki Okura, Yoshiaki Ono, Ken Osato, Masami Ouchi, Yen-Chen Pan, Andrés A. Plazas Malagón, Paul A. Price, Sophie L. Reed, Eli S. Rykoff, Takatoshi Shibuya, Mirko Simunovic, Michael A. Strauss, Kanako Sugimori, Yasushi Suto, Nao Suzuki, Masahiro Takada, Yuhei Takagi, Tadafumi Takata, Satoshi Takita, Masayuki Tanaka, Shenli Tang, Dan S. Taranu, Tsuyoshi Terai, Yoshiki Toba, Edwin L. Turner, Hisakazu Uchiyama, Bovornpratch Vijarnwannaluk, Christopher Z. Waters, Yoshihiko Yamada, Naoaki Yamamoto, and Takuji Yamashita. Third data release of the Hyper Suprime-Cam Subaru Strategic Program. *PASJ*, 74(2):247–272, April 2022.

[5] Yacine Ali-Haïmoud and Simeon Bird. An efficient implementation of massive neutrinos in non-linear structure formation simulations. *MNRAS*, 428(4):3375–3389, February 2013.

[6] Andreu Arinyo-i-Prats, Jordi Miralda-Escudé, Matteo Viel, and Renyue Cen. The non-linear power spectrum of the Lyman alpha forest. *JCAP*, 2015(12):017–017, December 2015.

[7] Metin Ata, Francisco-Shu Kitaura, Khee-Gan Lee, Brian C. Lemaux, Daichi Kashino, Olga Cucciati, Mónica Hernández-Sánchez, and Oliver Le Fèvre. BIRTH of the COSMOS field: primordial and evolved density reconstructions during cosmic high noon. *MNRAS*, 500(3):3194–3212, January 2021.

[8] N. Battaglia, H. Trac, R. Cen, and A. Loeb. Reionization on Large Scales. I. A Parametric Model Constructed from Radiation-hydrodynamic Simulations. *ApJ*, 776(2):81, October 2013.

[9] Peter Behroozi, Risa H. Wechsler, Andrew P. Hearin, and Charlie Conroy. UNIVERSEMACHINE: The correlation between galaxy growth and dark matter halo assembly from z = 0-10. *MNRAS*, 488(3):3143–3194, September 2019.

[10] José Luis Bernal and Ely D. Kovetz. Line-intensity mapping: theory review with a focus on star-formation lines. *A&A Rv*, 30(1):5, December 2022.

[11] Simeon Bird. FSFE: Fake Spectra Flux Extractor, October 2017.

[12] Simeon Bird. FSFE: Fake Spectra Flux Extractor. Astrophysics Source Code Library, record ascl:1710.012, October 2017.

[13] Simeon Bird, Martin Fernandez, Ming-Feng Ho, Mahdi Qezlou, Reza Monadi, Yueying Ni, Nianyi Chen, Rupert Croft, and Tiziana Di Matteo. PRIYA: a new suite of Lyman-α forest simulations for cosmology. *JCAP*, 2023(10):037, October 2023.

[14] Simeon Bird, Martin Haehnelt, Marcel Neeleman, Shy Genel, Mark Vogelsberger, and Lars Hernquist. Reproducing the kinematics of damped Lyman α systems. *MNRAS*, 447(2):1834–1846, February 2015.

[15] Simeon Bird, Martin Haehnelt, Marcel Neeleman, Shy Genel, Mark Vogelsberger, and Lars Hernquist. Reproducing the kinematics of damped Lyman α systems. *MNRAS*, 447(2):1834–1846, February 2015.

[16] Simeon Bird, Yueying Ni, Tiziana Di Matteo, Rupert Croft, Yu Feng, and Nianyi Chen. The ASTRID simulation: galaxy formation and reionization. *MNRAS*, 512(3):3703–3716, May 2022.

[17] Simeon Bird, Keir K. Rogers, Hiranya V. Peiris, Licia Verde, Andreu Font-Ribera, and Andrew Pontzen. An emulator for the Lyman-α forest. *JCAP*, 2019(2):050, February 2019.

[18] Michael Blomqvist, Hélion du Mas des Bourboux, Nicolás G. Busca, Victoria de Sainte Agathe, James Rich, Christophe Balland, Julian E. Bautista, Kyle Dawson, Andreu Font-Ribera, Julien Guy, Jean-Marc Le Goff, Nathalie Palanque-Delabrouille, Will J. Percival, Ignasi Pérez-Ràfols, Matthew M. Pieri, Donald P. Schneider, Anže Slosar,

and Christophe Yèche. Baryon acoustic oscillations from the cross-correlation of Lyα absorption and quasars in eBOSS DR14. *A&A*, 629:A86, September 2019.

[19] Patrick C. Breysse, Dongwoo T. Chung, Kieran A. Cleary, Håvard T. Ihle, Hamsa Padmanabhan, Marta B. Silva, J. Richard Bond, Jowita Borowska, Morgan Catha, Sarah E. Church, Delaney A. Dunne, Hans Kristian Eriksen, Marie Kristine Foss, Todd Gaier, Joshua Ott Gundersen, Andrew I. Harris, Richard Hobbs, Laura Keating, James W. Lamb, Charles R. Lawrence, Jonas G. S. Lunde, Norman Murray, Timothy J. Pearson, Liju Philip, Maren Rasmussen, Anthony C. S. Readhead, Thomas J. Rennie, Nils-Ole Stutzer, Marco P. Viero, Duncan J. Watts, Ingunn Kathrine Wehus, David P. Woody, and Comap Collaboration. COMAP Early Science. VII. Prospects for CO Intensity Mapping at Reionization. *ApJ*, 933(2):188, July 2022.

[20] Zheng Cai, Xiaohui Fan, Fuyan Bian, Ann Zabludoff, Yujin Yang, J. Xavier Prochaska, Ian McGreer, Zhen-Ya Zheng, Nobunari Kashikawa, Ran Wang, Brenda Frye, Richard Green, and Linhua Jiang. Mapping the Most Massive Overdensities through Hydrogen (MAMMOTH). II. Discovery of the Extremely Massive Overdensity BOSS1441 at z = 2.32. *ApJ*, 839(2):131, April 2017.

[21] Zheng Cai, Xiaohui Fan, Sebastien Peirani, Fuyan Bian, Brenda Frye, Ian McGreer, J. Xavier Prochaska, Marie Wingyee Lau, Nicolas Tejos, Shirley Ho, and Donald P. Schneider. Mapping the Most Massive Overdensity Through Hydrogen (MAMMOTH) I: Methodology. *ApJ*, 833(2):135, December 2016.

[22] Bob Carpenter, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1):1–32, 2017.

[23] S. Caucci, S. Colombi, C. Pichon, E. Rollinde, P. Petitjean, and T. Sousbie. Recovering the topology of the intergalactic medium at z ~2. *MNRAS*, 386(1):211–229, May 2008.

[24] Solène Chabanier, Nathalie Palanque-Delabrouille, Christophe Yèche, Jean-Marc Le Goff, Eric Armengaud, Julian Bautista, Michael Blomqvist, Nicolas Busca, Kyle Dawson, Thomas Etourneau, Andreu Font-Ribera, Youngbae Lee, Hélion du Mas des Bourboux, Matthew Pieri, James Rich, Graziano Rossi, Donald Schneider, and Anže Slosar. The one-dimensional power spectrum from the SDSS DR14 Lyα forests. *JCAP*, 2019(7):017, July 2019.

[25] Solène Chabanier, Nathalie Palanque-Delabrouille, Christophe Yèche, Jean-Marc Le Goff, Eric Armengaud, Julian Bautista, Michael Blomqvist, Nicolas Busca, Kyle Dawson, Thomas Etourneau, Andreu Font-Ribera, Youngbae Lee, Hélion du Mas des Bourboux, Matthew Pieri, James Rich, Graziano Rossi, Donald Schneider, and Anže Slosar. The one-dimensional power spectrum from the SDSS DR14 Lyα forests. *JCAP*, 2019(7):017, July 2019.

[26] Nima Chartab, Bahram Mobasher, Behnam Darvish, Steve Finkelstein, Yicheng Guo, Dritan Kodra, Kyoung-Soo Lee, Jeffrey A. Newman, Camilla Pacifici, Casey Papovich,

Zahra Sattari, Abtin Shahidi, Mark E. Dickinson, Sandra M. Faber, Henry C. Ferguson, Mauro Giavalisco, and Marziye Jafariyazani. Large-scale Structures in the CANDELS Fields: The Role of the Environment in Star Formation Activity. *ApJ*, 890(1):7, February 2020.

[27] Anya Chaudhuri, Subhabrata Majumdar, and Biman B. Nath. ACTIVE GALACTIC NUCLEUS FEEDBACK AND ENTROPY INJECTION IN GALAXY CLUSTER CORES. *The Astrophysical Journal*, 776(2):84, oct 2013.

[28] Edmond Chaussidon, Christophe Yèche, Nathalie Palanque-Delabrouille, David M. Alexander, Jinyi Yang, Steven Ahlen, Stephen. Bailey, David Brooks, Zheng Cai, Solène Chabanier, Tamara M. Davis, Kyle Dawson, Axel de la Macorra, Arjun Dey, Biprateep Dey, Sarah Eftekharzadeh, Daniel J. Eisenstein, Kevin Fanning, Andreu Font-Ribera, Enrique Gaztañaga, Satya Gontcho A Gontcho, Alma X. Gonzalez-Morales, Julien Guy, Hiram K. Herrera-Alcantar, Klaus Honscheid, Mustapha Ishak, Linhua Jiang, Stephanie Juneau, Robert Kehoe, Theodore Kisner, Andras Kovács, Anthony Kremin, Ting-Wen Lan, Martin Landriau, Laurent Le Guillou, Michael E. Levi, Christophe Magneville, Paul Martini, Aaron M. Meisner, John Moustakas, Andrea Muñoz-Gutiérrez, Adam D. Myers, Jeffrey A. Newman, Jundan Nie, Will J. Percival, Claire Poppett, Francisco Prada, Anand Raichoor, Corentin Ravoux@miscpystan, title = pystan (3.0.0), author = Riddell, Allen and Hartikainen, Ari and Carter, Matthew, year = 2021, month = mar, howpublished = PyPI , Ashley J. Ross, Edward Schlafly, David Schlegel, Ting Tan, Gregory Tarlé, Rongpu Zhou, Zhimin Zhou, and Hu Zou. Target Selection and Validation of DESI Quasars. *arXiv e-prints*, page arXiv:2208.08511, August 2022.

[29] Nianyi Chen, Yueying Ni, Michael Tremmel, Tiziana Di Matteo, Simeon Bird, Colin DeGraf, and Yu Feng. Dynamical friction modelling of massive black holes in cosmological simulations and effects on merger rate predictions. *MNRAS*, 510(1):531–550, February 2022.

[30] Yi-Kuan Chiang, Roderik Overzier, and Karl Gebhardt. Ancient Light from Young Cosmic Cities: Physical and Observational Signatures of Galaxy Proto-clusters. *ApJ*, 779(2):127, December 2013.

[31] Dongwoo T. Chung, Patrick C. Breysse, Kieran A. Cleary, Håvard T. Ihle, Hamsa Padmanabhan, Marta B. Silva, J. Richard Bond, Jowita Borowska, Morgan Catha, Sarah E. Church, Delaney A. Dunne, Hans Kristian Eriksen, Marie Kristine Foss, Todd Gaier, Joshua Ott Gundersen, Stuart E. Harper, Andrew I. Harris, Brandon Hensley, Richard Hobbs, Laura C. Keating, Junhan Kim, James W. Lamb, Charles R. Lawrence, Jonas Gahr Sturtzel Lunde, Norman Murray, Timothy J. Pearson, Liju Philip, Maren Rasmussen, Anthony C. S. Readhead, Thomas J. Rennie, Nils-Ole Stutzer, Bade D. Uzgil, Marco P. Viero, Duncan J. Watts, Risa H. Wechsler, Ingunn Kathrine Wehus, David P. Woody, and Comap Collaboration. COMAP Early Science. V. Constraints and Forecasts at z 3. *ApJ*, 933(2):186, July 2022.

[32] Dongwoo T. Chung, Marco P. Viero, Sarah E. Church, Risa H. Wechsler, Marcelo A. Alvarez, J. Richard Bond, Patrick C. Breysse, Kieran A. Cleary, Hans K. Eriksen, Marie K. Foss, Joshua O. Gundersen, Stuart E. Harper, Håvard T. Ihle, Laura C. Keating, Norman Murray, Hamsa Padmanabhan, George F. Stein, Ingunn K. Wehus, and COMAP Collaboration. Cross-correlating Carbon Monoxide Line-intensity Maps with Spectroscopic and Photometric Galaxy Surveys. *ApJ*, 872(2):186, February 2019.

[33] Kieran A. Cleary, Jowita Borowska, Patrick C. Breysse, Morgan Catha, Dongwoo T. Chung, Sarah E. Church, Clive Dickinson, Hans Kristian Eriksen, Marie Kristine Foss, Joshua Ott Gundersen, Stuart E. Harper, Andrew I. Harris, Richard Hobbs, Håvard T. Ihle, Junhan Kim, Jonathon Kocz, James W. Lamb, Jonas G. S. Lunde, Hamsa Padmanabhan, Timothy J. Pearson, Liju Philip, Travis W. Powell, Maren Rasmussen, Anthony C. S. Readhead, Thomas J. Rennie, Marta B. Silva, Nils-Ole Stutzer, Bade D. Uzgil, Duncan J. Watts, Ingunn Kathrine Wehus, David P. Woody, Lilian Basoalto, J. Richard Bond, Delaney A. Dunne, Todd Gaier, Brandon Hensley, Laura C. Keating, Charles R. Lawrence, Norman Murray, Roberta Paladini, Rodrigo Reeves, Marco P. Viero, Risa H. Wechsler, and Comap Collaboration. COMAP Early Science. I. Overview. *ApJ*, 933(2):182, July 2022.

[34] Cyril Creque-Sarbinowski and Marc Kamionkowski. Searching for decaying and annihilating dark matter with line intensity mapping. *PhRvD*, 98(6):063524, September 2018.

[35] Rupert A. C. Croft, Jordi Miralda-Escudé, Zheng Zheng, Michael Blomqvist, and Matthew Pieri. Intensity mapping with SDSS/BOSS Lyman-α emission, quasars, and their Lyman-α forest. *MNRAS*, 481(1):1320–1336, November 2018.

[36] Rupert A. C. Croft, David H. Weinberg, Neal Katz, and Lars Hernquist. Recovery of the Power Spectrum of Mass Fluctuations from Observations of the Lyα Forest. *ApJ*, 495(1):44–62, March 1998.

[37] O. Cucciati, B. C. Lemaux, G. Zamorani, O. Le Fèvre, L. A. M. Tasca, N. P. Hathi, K. G. Lee, S. Bardelli, P. Cassata, B. Garilli, V. Le Brun, D. Maccagni, L. Pentericci, R. Thomas, E. Vanzella, E. Zucca, L. M. Lubin, R. Amorin, L. P. Cassarà, A. Cimatti, M. Talia, D. Vergani, A. Koekemoer, J. Pforr, and M. Salvato. The progeny of a cosmic titan: a massive multi-component proto-supercluster in formation at z = 2.45 in VUDS. *A&A*, 619:A49, November 2018.

[38] Andrei Cuceu, James Farr, Pablo Lemos, and Andreu Font-Ribera. Baryon Acoustic Oscillations and the Hubble constant: past, present and future. *JCAP*, 2019(10):044, October 2019.

[39] Andrei Cuceu, Andreu Font-Ribera, Seshadri Nadathur, Benjamin Joachimi, and Paul Martini. Constraints on the Cosmic Expansion Rate at Redshift 2.3 from the Lyman-α Forest. *PhRvL*, 130(19):191003, May 2023.

[40] Anson D'Aloisio, Matthew McQuinn, Oliver Maupin, Frederick B. Davies, Hy Trac, Spencer Fuller, and Phoebe R. Upton Sanderbeck. Heating of the Intergalactic Medium by Hydrogen Reionization. *ApJ*, 874(2):154, April 2019.

[41] Behnam Darvish, Nick Z. Scoville, Christopher Martin, David Sobral, Bahram Mobasher, Alessandro Rettura, Jorryt Matthee, Peter Capak, Nima Chartab, Shoubaneh Hemmati, Daniel Masters, Hooshang Nayyeri, Donal O'Sullivan, Ana Paulino-Afonso, Zahra Sattari, Abtin Shahidi, Mara Salvato, Brian C. Lemaux, Olivier Le Fèvre, and Olga Cucciati. Spectroscopic Confirmation of a Coma Cluster Progenitor at z $\sim$ 2.2. *ApJ*, 892(1):8, March 2020.

[42] Romeel Davé, Daniel Anglés-Alcázar, Desika Narayanan, Qi Li, Mika H. Rafieferantsoa, and Sarah Appleby. SIMBA: Cosmological simulations with black hole growth and feedback. *MNRAS*, 486(2):2827–2849, June 2019.

[43] Victoria de Sainte Agathe, Christophe Balland, Hélion du Mas des Bourboux, Nicolás G. Busca, Michael Blomqvist, Julien Guy, James Rich, Andreu Font-Ribera, Matthew M. Pieri, Julian E. Bautista, Kyle Dawson, Jean-Marc Le Goff, Axel de la Macorra, Nathalie Palanque-Delabrouille, Will J. Percival, Ignasi Pérez-Ràfols, Donald P. Schneider, Anže Slosar, and Christophe Yèche. Baryon acoustic oscillations at z = 2.34 from the correlations of Lyα absorption in eBOSS DR14. *A&A*, 629:A85, September 2019.

[44] Timothée Delubac, Julian E. Bautista, Nicolás G. Busca, James Rich, David Kirkby, Stephen Bailey, Andreu Font-Ribera, Anže Slosar, Khee-Gan Lee, Matthew M. Pieri, Jean-Christophe Hamilton, Éric Aubourg, Michael Blomqvist, Jo Bovy, Jon Brinkmann, William Carithers, Kyle S. Dawson, Daniel J. Eisenstein, Satya Gontcho A. Gontcho, Jean-Paul Kneib, Jean-Marc Le Goff, Daniel Margala, Jordi Miralda-Escudé, Adam D. Myers, Robert C. Nichol, Pasquier Noterdaeme, Ross O'Connell, Matthew D. Olmstead, Nathalie Palanque-Delabrouille, Isabelle Pâris, Patrick Petitjean, Nicholas P. Ross, Graziano Rossi, David J. Schlegel, Donald P. Schneider, David H. Weinberg, Christophe Yèche, and Donald G. York. Baryon acoustic oscillations in the Lyα forest of BOSS DR11 quasars. *A&A*, 574:A59, February 2015.

[45] G. Desprez, V. Picouet, T. Moutard, S. Arnouts, M. Sawicki, J. Coupon, S. Gwyn, L. Chen, J. Huang, A. Golob, H. Furusawa, H. Ikeda, S. Paltani, C. Cheng, W. Hartley, B. C. Hsieh, O. Ilbert, O. B. Kauffmann, H. J. McCracken, M. Shuntov, M. Tanaka, S. Toft, L. Tresse, and J. R. Weaver. Combining the CLAUDS and HSC-SSP surveys. U + grizy(+YJHK$_s$) photometry and photometric redshifts for 18M galaxies in the 20 deg^2 of the HSC-SSP Deep and ultraDeep fields. *A&A*, 670:A82, February 2023.

[46] Tiziana Di Matteo, Volker Springel, and Lars Hernquist. Energy input from quasars regulates the growth and activity of black holes and their host galaxies. *Nature*, 433(7026):604–607, February 2005.

[47] Chenze Dong, Khee-Gan Lee, Metin Ata, Benjamin Horowitz, and Rieko Momose. Observational Evidence for Large-scale Gas Heating in a Galaxy Protocluster at z = 2.30. *ApJL*, 945(2):L28, March 2023.

[48] Chenze Dong, Khee-Gan Lee, Metin Ata, Benjamin Horowitz, and Rieko Momose. Observational Evidence for Large-scale Gas Heating in a Galaxy Protocluster at z = 2.30. *ApJL*, 945(2):L28, March 2023.

[49] A. Dressler. Galaxy morphology in rich clusters: implications for the formation and evolution of galaxies. *ApJ*, 236:351–365, March 1980.

[50] Hélion du Mas des Bourboux, James Rich, Andreu Font-Ribera, Victoria de Sainte Agathe, James Farr, Thomas Etourneau, Jean-Marc Le Goff, Andrei Cuceu, Christophe Balland, Julian E. Bautista, Michael Blomqvist, Jonathan Brinkmann, Joel R. Brownstein, Solène Chabanier, Edmond Chaussidon, Kyle Dawson, Alma X. González-Morales, Julien Guy, Brad W. Lyke, Axel de la Macorra, Eva-Maria Mueller, Adam D. Myers, Christian Nitschelm, Andrea Muñoz Gutiérrez, Nathalie Palanque-Delabrouille, James Parker, Will J. Percival, Ignasi Pérez-Ràfols, Patrick Petitjean, Matthew M. Pieri, Corentin Ravoux, Graziano Rossi, Donald P. Schneider, Hee-Jong Seo, Anže Slosar, Julianna Stermer, M. Vivek, Christophe Yèche, and Samantha Youles. The Completed SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Baryon Acoustic Oscillations with Lyα Forests. *ApJ*, 901(2):153, October 2020.

[51] A. Durkalec, O. Le Fèvre, A. Pollo, G. Zamorani, B. C. Lemaux, B. Garilli, S. Bardelli, N. Hathi, A. Koekemoer, J. Pforr, and E. Zucca. The VIMOS Ultra Deep Survey. Luminosity and stellar mass dependence of galaxy clustering at z 3. *A&A*, 612:A42, April 2018.

[52] A. C. Edge and G. C. Stewart. EXOSAT observations of clusters of galaxies - I. The X-ray data. *MNRAS*, 252:414, October 1991.

[53] Sarah Eftekharzadeh, Adam D. Myers, Martin White, David H. Weinberg, Donald P. Schneider, Yue Shen, Andreu Font-Ribera, Nicholas P. Ross, Isabelle Paris, and Alina Streblyanska. Clustering of intermediate redshift quasars using the final SDSS III-BOSS sample. *MNRAS*, 453(3):2779–2798, November 2015.

[54] A. E. Evrard and J. P. Henry. Expectations for X-Ray Cluster Observations by the ROSAT Satellite. *ApJ*, 383:95, December 1991.

[55] Claude-André Faucher-Giguère, Adam Lidz, Lars Hernquist, and Matias Zaldarriaga. Evolution of the Intergalactic Opacity: Implications for the Ionizing Background, Cosmic Star Formation, and Quasar Activity. *ApJ*, 688(1):85–107, November 2008.

[56] Claude-André Faucher-Giguère, Adam Lidz, Lars Hernquist, and Matias Zaldarriaga. Evolution of the Intergalactic Opacity: Implications for the Ionizing Background, Cosmic Star Formation, and Quasar Activity. *ApJ*, 688(1):85–107, November 2008.

[57] Yu Feng, Tiziana Di-Matteo, Rupert A. Croft, Simeon Bird, Nicholas Battaglia, and Stephen Wilkins. The BlueTides simulation: first galaxies and reionization. *MNRAS*, 455(3):2778–2791, January 2016.

[58] M. A. Fernandez, Simeon Bird, and Ming-Feng Ho. Cosmological Constraints from the eBOSS Lyman-α Forest using the PRIYA Simulations. *arXiv e-prints*, page arXiv:2309.03943, September 2023.

[59] M. A. Fernandez, Ming-Feng Ho, and Simeon Bird. A multifidelity emulator for the Lyman-α forest flux power spectrum. *MNRAS*, 517(3):3200–3211, December 2022.

[60] Andreu Font-Ribera, Eduard Arnau, Jordi Miralda-Escudé, Emmanuel Rollinde, J. Brinkmann, Joel R. Brownstein, Khee-Gan Lee, Adam D. Myers, Nathalie Palanque-Delabrouille, Isabelle Pâris, Patrick Petitjean, James Rich, Nicholas P. Ross, Donald P. Schneider, and Martin White. The large-scale quasar-Lyman α forest cross-correlation from BOSS. *JCAP*, 2013(5):018, May 2013.

[61] Andreu Font-Ribera, Patrick McDonald, and Anže Slosar. How to estimate the 3D power spectrum of the Lyman-α forest. *JCAP*, 2018(1):003, January 2018.

[62] Marie K. Foss, Håvard T. Ihle, Jowita Borowska, Kieran A. Cleary, Hans Kristian Eriksen, Stuart E. Harper, Junhan Kim, James W. Lamb, Jonas G. S. Lunde, Liju Philip, Maren Rasmussen, Nils-Ole Stutzer, Bade D. Uzgil, Duncan J. Watts, Ingunn K. Wehus, David P. Woody, J. Richard Bond, Patrick C. Breysse, Morgan Catha, Sarah E. Church, Dongwoo T. Chung, Clive Dickinson, Delaney A. Dunne, Todd Gaier, Joshua Ott Gundersen, Andrew I. Harris, Richard Hobbs, Charles R. Lawrence, Norman Murray, Anthony C. S. Readhead, Hamsa Padmanabhan, Timothy J. Pearson, Thomas J. Rennie, and Comap Collaboration. COMAP Early Science. III. CO Data Processing. *ApJ*, 933(2):184, July 2022.

[63] Steven R. Furlanetto and Adam Lidz. The Cross-Correlation of High-Redshift 21 cm and Galaxy Surveys. *ApJ*, 660(2):1030–1038, May 2007.

[64] Steven R. Furlanetto and Adam Lidz. The Cross-Correlation of High-Redshift 21 cm and Galaxy Surveys. *ApJ*, 660(2):1030–1038, May 2007.

[65] Prakash Gaikwad, Raghunathan Srianand, Martin G. Haehnelt, and Tirthankar Roy Choudhury. A consistent and robust measurement of the thermal state of the IGM at $2 \leq z \leq 4$ from a large sample of Ly α forest spectra: evidence for late and rapid He II reionization. *MNRAS*, 506(3):4389–4412, September 2021.

[66] Jonathan P. Gardner, John C. Mather, Mark Clampin, Rene Doyon, Matthew A. Greenhouse, Heidi B. Hammel, John B. Hutchings, Peter Jakobsen, Simon J. Lilly, Knox S. Long, Jonathan I. Lunine, Mark J. McCaughrean, Matt Mountain, John Nella, George H. Rieke, Marcia J. Rieke, Hans-Walter Rix, Eric P. Smith, George Sonneborn, Massimo Stiavelli, H. S. Stockman, Rogier A. Windhorst, and Gillian S. Wright. The James Webb Space Telescope. *SSRv*, 123(4):485–606, April 2006.

[67] Jahmour J. Givans, Andreu Font-Ribera, Anže Slosar, Louise Seeyave, Christian Pedersen, Keir K. Rogers, Mathias Garny, Diego Blas, and Vid Iršič. Non-linearities in the Lyman-α forest and in its cross-correlation with dark matter halos. *JCAP*, 2022(9):070, September 2022.

[68] Yan Gong, Asantha Cooray, Marta Silva, Mario G. Santos, James Bock, C. Matt Bradford, and Michael Zemcov. Intensity Mapping of the [C II] Fine Structure Line during the Epoch of Reionization. *ApJ*, 745(1):49, January 2012.

[69] Satya Gontcho A Gontcho, Jordi Miralda-Escudé, and Nicolás G. Busca. On the effect of the ionizing background on the Lyα forest autocorrelation function. *MNRAS*, 442(1):187–195, July 2014.

[70] C. Gordon, A. Cuceu, J. Chaves-Montero, A. Font-Ribera, A. X. González-Morales, J. Aguilar, S. Ahlen, E. Armengaud, S. Bailey, A. Bault, A. Brodzeller, D. Brooks, T. Claybaugh, R. de la Cruz, K. Dawson, P. Doel, J. E. Forero-Romero, S. Gontcho A. Gontcho, J. Guy, H. K. Herrera-Alcantar, V. Iršič, N. G. Karaçaylı, D. Kirkby, M. Landriau, L. Le Guillou, M. E. Levi, A. de la Macorra, M. Manera, P. Martini, A. Meisner, R. Miquel, P. Montero-Camacho, A. Muñoz-Gutiérrez, L. Napolitano, J. Nie, G. Niz, N. Palanque-Delabrouille, W. J. Percival, M. Pieri, C. Poppett, F. Prada, I. Pérez-Ràfols, C. Ramírez-Pérez, C. Ravoux, M. Rezaie, A. J. Ross, G. Rossi, E. Sanchez, D. Schlegel, M. Schubnell, H. Seo, F. Sinigaglia, T. Tan, G. Tarlé, M. Walther, B. A. Weaver, C. Yèche, Z. Zhou, and H. Zou. 3D correlations in the Lyman-α forest from early DESI data. *JCAP*, 2023(11):045, November 2023.

[71] Jenny Greene, Rachel Bezanson, Masami Ouchi, John Silverman, and the PFS Galaxy Evolution Working Group. The Prime Focus Spectrograph Galaxy Evolution Survey. *arXiv e-prints*, page arXiv:2206.14908, June 2022.

[72] James E. Gunn and Bruce A. Peterson. On the Density of Neutral Hydrogen in Intergalactic Space. *ApJ*, 142:1633–1636, November 1965.

[73] A. J. S. Hamilton. Linear Redshift Distortions: a Review. In Donald Hamilton, editor, *The Evolving Universe*, volume 231 of *Astrophysics and Space Science Library*, page 185, January 1998.

[74] Nick Hand, Yu Feng, adematti, Yin Li, and Chirag Modi. bccp/nbodykit: nbodykit v0.3.4, August 2018.

[75] Ming-Feng Ho, Simeon Bird, and Christian R. Shelton. Multifidelity emulation for the matter power spectrum using Gaussian processes. *MNRAS*, 509(2):2551–2565, January 2022.

[76] Benjamin Horowitz, Khee-Gan Lee, Metin Ata, Thomas Müller, Alex Krolewski, J. Xavier Prochaska, Joseph F. Hennawi, Martin White, David Schlegel, R. Michael Rich, Peter E. Nugent, Nao Suzuki, Daichi Kashino, Anton M. Koekemoer, and Brian C. Lemaux. Second Data Release of the COSMOS Lyman-alpha Mapping

and Tomographic Observation: The First 3D Maps of the Large-Scale Cosmic Web at 2.05¡z¡2.55. *arXiv e-prints*, page arXiv:2109.09660, September 2021.

[77] Benjamin Horowitz, Khee-Gan Lee, Metin Ata, Thomas Müller, Alex Krolewski, J. Xavier Prochaska, Joseph F. Hennawi, Martin White, David Schlegel, R. Michael Rich, Peter E. Nugent, Nao Suzuki, Daichi Kashino, Anton M. Koekemoer, and Brian C. Lemaux. Second Data Release of the COSMOS Lyα Mapping and Tomography Observations: The First 3D Maps of the Detailed Cosmic Web at 2.05 ¡ z ¡ 2.55. *ApJS*, 263(2):27, December 2022.

[78] Benjamin Horowitz, Khee-Gan Lee, Martin White, Alex Krolewski, and Metin Ata. TARDIS. I. A Constrained Reconstruction Approach to Modeling the z $\sim$ 2.5 Cosmic Web Probed by Lyα Forest Tomography. *ApJ*, 887(1):61, December 2019.

[79] Benjamin Horowitz, Benjamin Zhang, Khee-Gan Lee, and Robin Kooistra. TARDIS. II. Synergistic Density Reconstruction from Lyα Forest and Spectroscopic Galaxy Surveys with Applications to Protoclusters and the Cosmic Web. *ApJ*, 906(2):110, January 2021.

[80] Håvard T. Ihle, Jowita Borowska, Kieran A. Cleary, Hans Kristian Eriksen, Marie K. Foss, Stuart E. Harper, Junhan Kim, Jonas G. S. Lunde, Liju Philip, Maren Rasmussen, Nils-Ole Stutzer, Bade D. Uzgil, Duncan J. Watts, Ingunn Kathrine Wehus, J. Richard Bond, Patrick C. Breysse, Morgan Catha, Sarah E. Church, Dongwoo T. Chung, Clive Dickinson, Delaney A. Dunne, Todd Gaier, Joshua Ott Gundersen, Andrew I. Harris, Richard Hobbs, James W. Lamb, Charles R. Lawrence, Norman Murray, Anthony C. S. Readhead, Hamsa Padmanabhan, Timothy J. Pearson, Thomas J. Rennie, David P. Woody, and Compap Collaboration. COMAP Early Science. IV. Power Spectrum Methodology and Results. *ApJ*, 933(2):185, July 2022.

[81] Asif Iqbal, Subhabrata Majumdar, Biman B. Nath, Stefano Ettori, Dominique Eckert, and Manzoor A. Malik. Excess entropy and energy feedback from within cluster cores up to r200. *Monthly Notices of the Royal Astronomical Society*, 472(1):713–726, 08 2017.

[82] N. Kaiser. Evolution and clustering of rich clusters. *MNRAS*, 222:323–345, September 1986.

[83] Nick Kaiser. Clustering in real space and in redshift space. *MNRAS*, 227:1–21, July 1987.

[84] Rahul Kannan, Aaron Smith, Enrico Garaldi, Xuejian Shen, Mark Vogelsberger, Rüdiger Pakmor, Volker Springel, and Lars Hernquist. The THESAN project: predictions for multitracer line intensity mapping in the epoch of reionization. *MNRAS*, 514(3):3857–3878, August 2022.

[85] Marie Lynn Abdul Karim, Eric Armengaud, Guillaume Mention, Solène Chabanier, Corentin Ravoux, and Zarija Lukić. Measurement of the small-scale 3D Lyman-α forest power spectrum. *arXiv e-prints*, page arXiv:2310.09116, October 2023.

[86] Neal Katz, David H. Weinberg, and Lars Hernquist. Cosmological Simulations with TreeSPH. *ApJS*, 105:19, July 1996.

[87] Neal Katz, David H. Weinberg, and Lars Hernquist. Cosmological Simulations with TreeSPH. *ApJS*, 105:19, July 1996.

[88] Guinevere Kauffmann, Simon D. M. White, Timothy M. Heckman, Brice Ménard, Jarle Brinchmann, Stéphane Charlot, Christy Tremonti, and Jon Brinkmann. The environmental dependence of the relations between stellar mass, structure, star formation and nuclear activity in galaxies. *MNRAS*, 353(3):713–731, September 2004.

[89] Garrett K. Keating, Geoffrey C. Bower, Daniel P. Marrone, David R. DeBoer, Carl Heiles, Tzu-Ching Chang, John E. Carlstrom, Christopher H. Greer, David Hawkins, James W. Lamb, Erik Leitch, Amber D. Miller, Stephen Muchovej, and David P. Woody. First Results from COPSS: The CO Power Spectrum Survey. *ApJ*, 814(2):140, December 2015.

[90] Garrett K. Keating, Daniel P. Marrone, Geoffrey C. Bower, and Ryan P. Keenan. An Intensity Mapping Detection of Aggregate CO Line Emission at 3 mm. *ApJ*, 901(2):141, October 2020.

[91] Garrett K. Keating, Daniel P. Marrone, Geoffrey C. Bower, Erik Leitch, John E. Carlstrom, and David R. DeBoer. COPSS II: The Molecular Gas Content of Ten Million Cubic Megaparsecs at Redshift z $\sim$ 3. *ApJ*, 830(1):34, October 2016.

[92] Garrett K. Keating, Daniel P. Marrone, Geoffrey C. Bower, Erik Leitch, John E. Carlstrom, and David R. DeBoer. COPSS II: The Molecular Gas Content of Ten Million Cubic Megaparsecs at Redshift z $\sim$ 3. *ApJ*, 830(1):34, October 2016.

[93] Ryan P. Keenan, Garrett K. Keating, and Daniel P. Marrone. An Intensity Mapping Constraint on the CO-galaxy Cross-power Spectrum at Redshift 3. *ApJ*, 927(2):161, March 2022.

[94] T. S. Kim, J. S. Bolton, M. Viel, M. G. Haehnelt, and R. F. Carswell. An improved measurement of the flux distribution of the Lyα forest in QSO absorption spectra: the effect of continuum fitting, metal contamination and noise properties. *MNRAS*, 382(4):1657–1674, December 2007.

[95] L. Knox and M. Millea. Hubble constant hunter's guide. *PhRvD*, 101(4):043533, February 2020.

[96] Robin Kooistra, Shigeki Inoue, Khee-Gan Lee, Renyue Cen, and Naoki Yoshida. Detecting Preheating in Protoclusters with Lyα Forest Tomography. *ApJ*, 927(1):53, March 2022.

[97] Robin Kooistra, Khee-Gan Lee, and Benjamin Horowitz. Constraining the Fluctuating Gunn-Peterson Approximation \\Using Lyman-α Forest Tomography at $z = 2$. *arXiv e-prints*, page arXiv:2201.10169, January 2022.

[98] Ely Kovetz, Patrick C. Breysse, Adam Lidz, Jamie Bock, Charles M. Bradford, Tzu-Ching Chang, Simon Foreman, Hamsa Padmanabhan, Anthony Pullen, Dominik Riechers, Marta B. Silva, and Eric Switzer. Astrophysics and Cosmology with Line-Intensity Mapping. *BAAS*, 51(3):101, May 2019.

[99] Ely D. Kovetz, Marco P. Viero, Adam Lidz, Laura Newburgh, Mubdi Rahman, Eric Switzer, Marc Kamionkowski, James Aguirre, Marcelo Alvarez, James Bock, J. Richard Bond, Goeffry Bower, C. Matt Bradford, Patrick C. Breysse, Philip Bull, Tzu-Ching Chang, Yun-Ting Cheng, Dongwoo Chung, Kieran Cleary, Asantha Corray, Abigail Crites, Rupert Croft, Olivier Doré, Michael Eastwood, Andrea Ferrara, José Fonseca, Daniel Jacobs, Garrett K. Keating, Guilaine Lagache, Gunjan Lakhlani, Adrian Liu, Kavilan Moodley, Norm Murray, Aurélie Pénin, Gergö Popping, Anthony Pullen, Dominik Reichers, Shun Saito, Ben Saliwanchik, Mario Santos, Rachel Somerville, Gordon Stacey, George Stein, Francesco Villaescusa-Navarro, Eli Visbal, Amanda Weltman, Laura Wolz, and Micheal Zemcov. Line-Intensity Mapping: 2017 Status Report. *arXiv e-prints*, page arXiv:1709.09066, September 2017.

[100] Alex Krolewski, Khee-Gan Lee, Martin White, Joseph F. Hennawi, David J. Schlegel, Peter E. Nugent, Zarija Lukić, Casey W. Stark, Anton M. Koekemoer, Olivier Le Fèvre, Brian C. Lemaux, Christian Maier, R. Michael Rich, Mara Salvato, and Lidia Tasca. Detection of z $\sim$ 2.3 Cosmic Voids from 3D Lyα Forest Tomography in the COSMOS Field. *ApJ*, 861(1):60, July 2018.

[101] Mark R. Krumholz and Nickolay Y. Gnedin. A Comparison of Methods for Determining the Molecular Content of Model Galaxies. *ApJ*, 729(1):36, March 2011.

[102] Kamson Lai, Adam Lidz, Lars Hernquist, and Matias Zaldarriaga. The Impact of Temperature Fluctuations on the Lyα Forest Power Spectrum. *ApJ*, 644(1):61–70, June 2006.

[103] Johannes U. Lange, Andrew P. Hearin, Alexie Leauthaud, Frank C. van den Bosch, Enia Xhakaj, Hong Guo, Risa H. Wechsler, and Joseph DeRose. Constraints on S_8 from a full-scale and full-shape analysis of redshift-space clustering and galaxy-galaxy lensing in BOSS. *MNRAS*, 520(4):5373–5393, April 2023.

[104] Khee-Gan Lee, Stephen Bailey, Leslie E. Bartsch, William Carithers, Kyle S. Dawson, David Kirkby, Britt Lundgren, Daniel Margala, Nathalie Palanque-Delabrouille, Matthew M. Pieri, David J. Schlegel, David H. Weinberg, Christophe Yèche, Éric Aubourg, Julian Bautista, Dmitry Bizyaev, Michael Blomqvist, Adam S. Bolton, Arnaud Borde, Howard Brewington, Nicolás G. Busca, Rupert A. C. Croft, Timothée Delubac, Garrett Ebelke, Daniel J. Eisenstein, Andreu Font-Ribera, Jian Ge, Jean-Christophe Hamilton, Joseph F. Hennawi, Shirley Ho, Klaus Honscheid, Jean-Marc Le Goff, Elena Malanushenko, Viktor Malanushenko, Jordi Miralda-Escudé, Adam D. Myers, Pasquier Noterdaeme, Daniel Oravetz, Kaike Pan, Isabelle Pâris, Patrick Petitjean, James Rich, Emmanuel Rollinde, Nicholas P. Ross, Graziano Rossi, Donald P. Schneider, Audrey Simmons, Stephanie Snedden, Anže Slosar, David N. Spergel, Nao

Suzuki, Matteo Viel, and Benjamin A. Weaver. The BOSS Lyα Forest Sample from SDSS Data Release 9. *AJ*, 145(3):69, March 2013.

[105] Khee-Gan Lee, Joseph F. Hennawi, Casey Stark, J. Xavier Prochaska, Martin White, David J. Schlegel, Anna-Christina Eilers, Andreu Arinyo-i-Prats, Nao Suzuki, Rupert A. C. Croft, Karina I. Caputi, Paolo Cassata, Olivier Ilbert, Bianca Garilli, Anton M. Koekemoer, Vincent Le Brun, Olivier Le Fèvre, Dario Maccagni, Peter Nugent, Yoshiaki Taniguchi, Lidia A. M. Tasca, Laurence Tresse, Gianni Zamorani, and Elena Zucca. Lyα Forest Tomography from Background Galaxies: The First Megaparsec-resolution Large-scale Structure Map at z ¿ 2. *ApJL*, 795(1):L12, November 2014.

[106] Khee-Gan Lee, Joseph F. Hennawi, Martin White, Rupert A. C. Croft, and Melih Ozbek. Observational Requirements for Lyα Forest Tomographic Mapping of Large-scale Structure at z ˜2. *ApJ*, 788(1):49, June 2014.

[107] Khee-Gan Lee, Joseph F. Hennawi, Martin White, Rupert A. C. Croft, and Melih Ozbek. Observational Requirements for Lyα Forest Tomographic Mapping of Large-scale Structure at z ˜2. *ApJ*, 788(1):49, June 2014.

[108] Khee-Gan Lee, Joseph F. Hennawi, Martin White, J. Xavier Prochaska, Andreu Font-Ribera, David J. Schlegel, R. Michael Rich, Nao Suzuki, Casey W. Stark, Olivier Le Fèvre, Peter E. Nugent, Mara Salvato, and Gianni Zamorani. Shadow of a Colossus: A z = 2.44 Galaxy Protocluster Detected in 3D Lyα Forest Tomographic Mapping of the COSMOS Field. *ApJ*, 817(2):160, February 2016.

[109] Khee-Gan Lee, Alex Krolewski, Martin White, David Schlegel, Peter E. Nugent, Joseph F. Hennawi, Thomas Müller, Richard Pan, J. Xavier Prochaska, Andreu Font-Ribera, Nao Suzuki, Karl Glazebrook, Glenn G. Kacprzak, Jeyhan S. Kartaltepe, Anton M. Koekemoer, Olivier Le Fèvre, Brian C. Lemaux, Christian Maier, Themiya Nanayakkara, R. Michael Rich, D. B. Sanders, Mara Salvato, Lidia Tasca, and Kim-Vy H. Tran. First Data Release of the COSMOS Lyα Mapping and Tomography Observations: 3D Lyα Forest Tomography at 2.05 ¡ z ¡ 2.55. *ApJS*, 237(2):31, August 2018.

[110] Khee-Gan Lee, Alex Krolewski, Martin White, David Schlegel, Peter E. Nugent, Joseph F. Hennawi, Thomas Müller, Richard Pan, J. Xavier Prochaska, Andreu Font-Ribera, Nao Suzuki, Karl Glazebrook, Glenn G. Kacprzak, Jeyhan S. Kartaltepe, Anton M. Koekemoer, Olivier Le Fèvre, Brian C. Lemaux, Christian Maier, Themiya Nanayakkara, R. Michael Rich, D. B. Sanders, Mara Salvato, Lidia Tasca, and Kim-Vy H. Tran. First Data Release of the COSMOS Lyα Mapping and Tomography Observations: 3D Lyα Forest Tomography at 2.05 ¡ z ¡ 2.55. *ApJS*, 237(2):31, August 2018.

[111] Khee-Gan Lee and Martin White. Revealing the z ˜2.5 Cosmic Web with 3D Lyα Forest Tomography: a Deformation Tensor Approach. *ApJ*, 831(2):181, November 2016.

[112] B. C. Lemaux, O. Cucciati, O. Le Fèvre, G. Zamorani, L. M. Lubin, N. Hathi, O. Ilbert, D. Pelliccia, R. Amorín, S. Bardelli, P. Cassata, R. R. Gal, B. Garilli, L. Guaita, M. Giavalisco, D. Hung, A. Koekemoer, D. Maccagni, L. Pentericci, B. Ribeiro, D. Schaerer, L. Shen, M. Talia, A. R. Tomczak, E. Vanzella, D. Vergani, and E. Zucca. The VIMOS Ultra Deep Survey: The Reversal of the Star Formation Rate – Density Relation at $2 < z < 5$. *arXiv e-prints*, page arXiv:2009.03324, September 2020.

[113] Tony Y. Li, Risa H. Wechsler, Kiruthika Devaraj, and Sarah E. Church. Connecting CO Intensity Mapping to Molecular Gas and Star Formation in the Epoch of Galaxy Assembly. *ApJ*, 817(2):169, February 2016.

[114] Zihao Li, Benjamin Horowitz, and Zheng Cai. Improved Lyα Tomography Using Optimized Reconstruction with Constraints on Absorption (ORCA). *ApJ*, 916(1):20, July 2021.

[115] Adam Lidz, Steven R. Furlanetto, S. Peng Oh, James Aguirre, Tzu-Ching Chang, Olivier Doré, and Jonathan R. Pritchard. Intensity Mapping with Carbon Monoxide Emission Lines and the Redshifted 21 cm Line. *ApJ*, 741(2):70, November 2011.

[116] Adam Lidz, Oliver Zahn, Steven R. Furlanetto, Matthew McQuinn, Lars Hernquist, and Matias Zaldarriaga. Probing Reionization with the 21 cm Galaxy Cross-Power Spectrum. *ApJ*, 690(1):252–266, January 2009.

[117] Xiaojing Lin, Zheng Zheng, and Zheng Cai. Probing the Diffuse Lyα Emission on Cosmological Scales: Lyα Emission Intensity Mapping Using the Complete SDSS-IV eBOSS. *ApJS*, 262(2):38, October 2022.

[118] Federico Marinacci, Mark Vogelsberger, Rüdiger Pakmor, Paul Torrey, Volker Springel, Lars Hernquist, Dylan Nelson, Rainer Weinberger, Annalisa Pillepich, Jill Naiman, and Shy Genel. First results from the IllustrisTNG simulations: radio haloes and magnetic fields. *MNRAS*, 480(4):5113–5139, November 2018.

[119] Maxim Markevitch. The L_X-T Relation and Temperature Function for Nearby Clusters Revisited. *ApJ*, 504(1):27–34, September 1998.

[120] Patrick McDonald, Jordi Miralda-Escudé, Michael Rauch, Wallace L. W. Sargent, Tom A. Barlow, Renyue Cen, and Jeremiah P. Ostriker. The Observed Probability Distribution Function, Power Spectrum, and Correlation Function of the Transmitted Flux in the Lyα Forest. *ApJ*, 543(1):1–23, November 2000.

[121] Matthew McQuinn. The Evolution of the Intergalactic Medium. *ARA&A*, 54:313–362, September 2016.

[122] Matthew McQuinn, Lars Hernquist, Adam Lidz, and Matias Zaldarriaga. The signatures of large-scale temperature and intensity fluctuations in the Lyman α forest. *MNRAS*, 415(1):977–992, July 2011.

[123] Matthew McQuinn, Lars Hernquist, Adam Lidz, and Matias Zaldarriaga. The signatures of large-scale temperature and intensity fluctuations in the Lyman α forest. *MNRAS*, 415(1):977–992, July 2011.

[124] Matthew McQuinn, Adam Lidz, Matias Zaldarriaga, Lars Hernquist, Philip F. Hopkins, Suvendra Dutta, and Claude-André Faucher-Giguère. He II Reionization and its Effect on the Intergalactic Medium. *ApJ*, 694(2):842–866, April 2009.

[125] Matthew McQuinn and Martin White. On estimating Lyα forest correlations between multiple sightlines. *MNRAS*, 415(3):2257–2269, August 2011.

[126] M. Meixner, A. Cooray, D. Leisawitz, J. Staguhn, L. Armus, C. Battersby, J. Bauer, E. Bergin, C. M. Bradford, K. Ennico-Smith, J. Fortney, T. Kataria, G. Melnick, S. Milam, D. Narayanan, D. Padgett, K. Pontoppidan, A. Pope, T. Roellig, K. Sandstrom, K. Stevenson, K. Su, J. Vieira, E. Wright, J. Zmuidzinas, K. Sheth, D. Benford, E. E. Mamajek, S. Neff, E. De Beck, M. Gerin, F. Helmich, I. Sakon, D. Scott, R. Vavrek, M. Wiedner, S. Carey, D. Burgarella, S. H. Moseley, E. Amatucci, R. C. Carter, M. DiPirro, C. Wu, B. Beaman, P. Beltran, J. Bolognese, D. Bradley, J. Corsetti, T. D'Asto, K. Denis, C. Derkacz, C. P. Earle, L. G. Fantano, D. Folta, B. Gavares, J. Generie, L. Hilliard, J. M. Howard, A. Jamil, T. Jamison, C. Lynch, G. Martins, S. Petro, D. Ramspacher, A. Rao, C. Sandin, E. Stoneking, S. Tompkins, and C. Webster. Origins Space Telescope Mission Concept Study Report. *arXiv e-prints*, page arXiv:1912.06213, December 2019.

[127] George Miley and Carlos De Breuck. Distant radio galaxies and their environments. *A&A Rv*, 15(2):67–144, February 2008.

[128] Joel S. A. Miller, James S. Bolton, and Nina A. Hatch. Searching for the shadows of giants - II. The effect of local ionization on the Ly α absorption signatures of protoclusters at redshift z 2.4. *MNRAS*, 506(4):6001–6013, October 2021.

[129] Margherita Molaro, Vid Iršič, James S. Bolton, Laura C. Keating, Ewald Puchwein, Prakash Gaikwad, Martin G. Haehnelt, Girish Kulkarni, and Matteo Viel. The effect of inhomogeneous reionization on the Lyman α forest power spectrum at redshift z ¿ 4: implications for thermal parameter recovery. *MNRAS*, 509(4):6119–6137, February 2022.

[130] Rieko Momose, Khee-Gan Lee, Benjamin Horowitz, Metin Ata, and Jeyhan S. Kartaltepe. The Dependence of Galaxy Properties on the Underlying 3D Matter Density Field at 2.0¡z¡2.5. *arXiv e-prints*, page arXiv:2212.05984, December 2022.

[131] Rieko Momose, Ikkoh Shimizu, Kentaro Nagamine, Kazuhiro Shimasaku, Nobunari Kashikawa, and Haruka Kusakabe. Connection between Galaxies and H I in Circumgalactic and Intergalactic Media: Variation according to Galaxy Stellar Mass and Star Formation Activity. *ApJ*, 911(2):98, April 2021.

[132] Azadeh Moradinezhad Dizgah, Garrett K. Keating, and Anastasia Fialkov. Probing Cosmic Origins with CO and [C II] Emission Lines. *ApJL*, 870(1):L4, January 2019.

[133] Jill P. Naiman, Annalisa Pillepich, Volker Springel, Enrico Ramirez-Ruiz, Paul Torrey, Mark Vogelsberger, Rüdiger Pakmor, Dylan Nelson, Federico Marinacci, Lars Hernquist, Rainer Weinberger, and Shy Genel. First results from the IllustrisTNG simulations: a tale of two elements - chemical evolution of magnesium and europium. *MNRAS*, 477(1):1206–1224, June 2018.

[134] Dylan Nelson, Annalisa Pillepich, Volker Springel, Rainer Weinberger, Lars Hernquist, Rüdiger Pakmor, Shy Genel, Paul Torrey, Mark Vogelsberger, Guinevere Kauffmann, Federico Marinacci, and Jill Naiman. First results from the IllustrisTNG simulations: the galaxy colour bimodality. *MNRAS*, 475(1):624–647, March 2018.

[135] Dylan Nelson, Volker Springel, Annalisa Pillepich, Vicente Rodriguez-Gomez, Paul Torrey, Shy Genel, Mark Vogelsberger, Ruediger Pakmor, Federico Marinacci, Rainer Weinberger, Luke Kelley, Mark Lovell, Benedikt Diemer, and Lars Hernquist. The IllustrisTNG simulations: public data release. *Computational Astrophysics and Cosmology*, 6(1):2, May 2019.

[136] Andrew B. Newman, Mahdi Qezlou, Nima Chartab, Gwen C. Rudie, Guillermo A. Blanc, Simeon Bird, Andrew J. Benson, Daniel D. Kelson, and Brian C. Lemaux. LATIS: Constraints on the Galaxy-Halo Connection at z 2.5 from Galaxy-Galaxy and Galaxy-Lyα Clustering. *ApJ*, 961(1):27, January 2024.

[137] Andrew B. Newman, Gwen C. Rudie, Guillermo A. Blanc, Daniel D. Kelson, Sunny Rhoades, Tyson Hare, Victoria Pérez, Andrew J. Benson, Alan Dressler, Valentino Gonzalez, Juna A. Kollmeier, Nicholas P. Konidaris, John S. Mulchaey, Michael Rauch, Olivier Le Fèvre, Brian C. Lemaux, Olga Cucciati, and Simon J. Lilly. LATIS: The Lyα Tomography IMACS Survey. *ApJ*, 891(2):147, March 2020.

[138] Andrew B. Newman, Gwen C. Rudie, Guillermo A. Blanc, Daniel D. Kelson, Sunny Rhoades, Tyson Hare, Victoria Pérez, Andrew J. Benson, Alan Dressler, Valentino Gonzalez, Juna A. Kollmeier, Nicholas P. Konidaris, John S. Mulchaey, Michael Rauch, Olivier Le Fèvre, Brian C. Lemaux, Olga Cucciati, and Simon J. Lilly. LATIS: The Lyα Tomography IMACS Survey. *ApJ*, 891(2):147, March 2020.

[139] Andrew B. Newman, Gwen C. Rudie, Guillermo A. Blanc, Mahdi Qezlou, Simeon Bird, Daniel D. Kelson, Victoria Pérez, Enrico Congiu, Brian C. Lemaux, Alan Dressler, and John S. Mulchaey. A population of ultraviolet-dim protoclusters detected in absorption. *Nature*, 606(7914):475–478, June 2022.

[140] Yueying Ni, Tiziana Di Matteo, Simeon Bird, Rupert Croft, Yu Feng, Nianyi Chen, Michael Tremmel, Colin DeGraf, and Yin Li. The ASTRID simulation: the evolution of supermassive black holes. *MNRAS*, 513(1):670–692, June 2022.

[141] Atsushi J. Nishizawa, Bau-Ching Hsieh, Masayuki Tanaka, and Tadafumi Takata. Photometric Redshifts for the Hyper Suprime-Cam Subaru Strategic Program Data Release 2. *arXiv e-prints*, page arXiv:2003.01511, February 2020.

[142] Gaël Noirot, Daniel Stern, Simona Mei, Dominika Wylezalek, Elizabeth A. Cooke, Carlos De Breuck, Audrey Galametz, Nina A. Hatch, Joël Vernet, Mark Brodwin, Peter Eisenhardt, Anthony H. Gonzalez, Matt Jarvis, Alessandro Rettura, Nick Seymour, and S. A. Stanford. HST Grism Confirmation of 16 Structures at 1.4 ¡ z ¡ 2.8 from the Clusters Around Radio-Loud AGN (CARLA) Survey. *ApJ*, 859(1):38, May 2018.

[143] Roderik A. Overzier. The realm of the galaxy protoclusters. A review. *A&A Rv*, 24(1):14, November 2016.

[144] Hamsa Padmanabhan. Constraining the CO intensity mapping power spectrum at intermediate redshifts. *MNRAS*, 475(2):1477–1484, April 2018.

[145] Nathalie Palanque-Delabrouille, Christophe Yèche, Nils Schöneberg, Julien Lesgourgues, Michael Walther, Solène Chabanier, and Eric Armengaud. Hints, neutrino bounds, and WDM constraints from SDSS DR14 Lyman-α and Planck full-survey data. *JCAP*, 2020(4):038, April 2020.

[146] Christian Pedersen, Andreu Font-Ribera, Thomas D. Kitching, Patrick McDonald, Simeon Bird, Anže Slosar, Keir K. Rogers, and Andrew Pontzen. Massive neutrinos and degeneracies in Lyman-alpha forest simulations. *JCAP*, 2020(4):025, April 2020.

[147] P. J. E. Peebles and J. T. Yu. Primeval Adiabatic Perturbation in an Expanding Universe. *ApJ*, 162:815, December 1970.

[148] Ying-jie Peng, Simon J. Lilly, Katarina Kovač, Micol Bolzonella, Lucia Pozzetti, Alvio Renzini, Gianni Zamorani, Olivier Ilbert, Christian Knobel, Angela Iovino, Christian Maier, Olga Cucciati, Lidia Tasca, C. Marcella Carollo, John Silverman, Pawel Kampczyk, Loic de Ravel, David Sanders, Nicholas Scoville, Thierry Contini, Vincenzo Mainieri, Marco Scodeggio, Jean-Paul Kneib, Olivier Le Fèvre, Sandro Bardelli, Angela Bongiorno, Karina Caputi, Graziano Coppa, Sylvain de la Torre, Paolo Franzetti, Bianca Garilli, Fabrice Lamareille, Jean-Francois Le Borgne, Vincent Le Brun, Marco Mignoli, Enrique Perez Montero, Roser Pello, Elena Ricciardelli, Masayuki Tanaka, Laurence Tresse, Daniela Vergani, Niraj Welikala, Elena Zucca, Pascal Oesch, Ummi Abbas, Luke Barnes, Rongmon Bordoloi, Dario Bottini, Alberto Cappi, Paolo Cassata, Andrea Cimatti, Marco Fumana, Gunther Hasinger, Anton Koekemoer, Alexei Leauthaud, Dario Maccagni, Christian Marinoni, Henry McCracken, Pierdomenico Memeo, Baptiste Meneux, Preethi Nair, Cristiano Porciani, Valentina Presotto, and Roberto Scaramella. Mass and Environment as Drivers of Galaxy Evolution in SDSS and zCOSMOS and the Origin of the Schechter Function. *ApJ*, 721(1):193–221, September 2010.

[149] C. Pichon, J. L. Vergely, E. Rollinde, S. Colombi, and P. Petitjean. Inversion of the Lyman α forest: three-dimensional investigation of the intergalactic medium. *MNRAS*, 326(2):597–620, September 2001.

[150] Annalisa Pillepich, Dylan Nelson, Lars Hernquist, Volker Springel, Rüdiger Pakmor, Paul Torrey, Rainer Weinberger, Shy Genel, Jill P. Naiman, Federico Marinacci, and

Mark Vogelsberger. First results from the IllustrisTNG simulations: the stellar mass content of groups and clusters of galaxies. *MNRAS*, 475(1):648–675, March 2018.

[151] Annalisa Pillepich, Volker Springel, Dylan Nelson, Shy Genel, Jill Naiman, Rüdiger Pakmor, Lars Hernquist, Paul Torrey, Mark Vogelsberger, Rainer Weinberger, and Federico Marinacci. Simulating galaxy formation with the IllustrisTNG model. *MNRAS*, 473(3):4077–4106, January 2018.

[152] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, J. G. Bartlett, N. Bartolo, E. Battaner, R. Battye, K. Benabed, A. Benoît, A. Benoit-Lévy, J. P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, A. Bonaldi, L. Bonavera, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J. F. Cardoso, A. Catalano, A. Challinor, A. Chamballu, R. R. Chary, H. C. Chiang, J. Chluba, P. R. Christensen, S. Church, D. L. Clements, S. Colombi, L. P. L. Colombo, C. Combet, A. Coulais, B. P. Crill, A. Curto, F. Cuttaia, L. Danese, R. D. Davies, R. J. Davis, P. de Bernardis, A. de Rosa, G. de Zotti, J. Delabrouille, F. X. Désert, E. Di Valentino, C. Dickinson, J. M. Diego, K. Dolag, H. Dole, S. Donzelli, O. Doré, M. Douspis, A. Ducout, J. Dunkley, X. Dupac, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, M. Farhang, J. Fergusson, F. Finelli, O. Forni, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frejsel, S. Galeotta, S. Galli, K. Ganga, C. Gauthier, M. Gerbino, T. Ghosh, M. Giard, Y. Giraud-Héraud, E. Giusarma, E. Gjerløw, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gregorio, A. Gruppuso, J. E. Gudmundsson, J. Hamann, F. K. Hansen, D. Hanson, D. L. Harrison, G. Helou, S. Henrot-Versillé, C. Hernández-Monteagudo, D. Herranz, S. R. Hildebrandt, E. Hivon, M. Hobson, W. A. Holmes, A. Hornstrup, W. Hovest, Z. Huang, K. M. Huffenberger, G. Hurier, A. H. Jaffe, T. R. Jaffe, W. C. Jones, M. Juvela, E. Keihänen, R. Keskitalo, T. S. Kisner, R. Kneissl, J. Knoche, L. Knox, M. Kunz, H. Kurki-Suonio, G. Lagache, A. Lähteenmäki, J. M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, J. P. Leahy, R. Leonardi, J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. Linden-Vørnle, M. López-Caniego, P. M. Lubin, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marchini, M. Maris, P. G. Martin, M. Martinelli, E. Martínez-González, S. Masi, S. Matarrese, P. McGehee, P. R. Meinhold, A. Melchiorri, J. B. Melin, L. Mendes, A. Mennella, M. Migliaccio, M. Millea, S. Mitra, M. A. Miville-Deschênes, A. Moneti, L. Montier, G. Morgante, D. Mortlock, A. Moss, D. Munshi, J. A. Murphy, P. Naselsky, F. Nati, P. Natoli, C. B. Netterfield, H. U. Nørgaard-Nielsen, F. Noviello, D. Novikov, I. Novikov, C. A. Oxborrow, F. Paci, L. Pagano, F. Pajot, R. Paladini, D. Paoletti, B. Partridge, F. Pasian, G. Patanchon, T. J. Pearson, O. Perdereau, L. Perotto, F. Perrotta, V. Pettorino, F. Piacentini, M. Piat, E. Pierpaoli, D. Pietrobon, S. Plaszczynski, E. Pointecouteau, G. Polenta, L. Popa, G. W. Pratt, G. Prézeau, S. Prunet, J. L. Puget, J. P. Rachen, W. T. Reach, R. Rebolo, M. Reinecke, M. Remazeilles, C. Renault, A. Renzi, I. Ristorcelli, G. Rocha, C. Rosset, M. Rossetti, G. Roudier, B. Rouillé d'Orfeuil, M. Rowan-Robinson, J. A. Rubiño-Martín, B. Rusholme, N. Said, V. Salvatelli, L. Salvati, M. Sandri, D. Santos, M. Savelainen, G. Savini, D. Scott, M. D. Seiffert, P. Serra, E. P. S. Shellard, L. D.

Spencer, M. Spinelli, V. Stolyarov, R. Stompor, R. Sudiwala, R. Sunyaev, D. Sutton, A. S. Suur-Uski, J. F. Sygnet, J. A. Tauber, L. Terenzi, L. Toffolatti, M. Tomasi, M. Tristram, T. Trombetti, M. Tucci, J. Tuovinen, M. Türler, G. Umana, L. Valenziano, J. Valiviita, F. Van Tent, P. Vielva, F. Villa, L. A. Wade, B. D. Wandelt, I. K. Wehus, M. White, S. D. M. White, A. Wilkinson, D. Yvon, A. Zacchei, and A. Zonca. Planck 2015 results. XIII. Cosmological parameters. *A&A*, 594:A13, September 2016.

[153] Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, R. Battye, K. Benabed, J. P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J. F. Cardoso, J. Carron, A. Challinor, H. C. Chiang, J. Chluba, L. P. L. Colombo, C. Combet, D. Contreras, B. P. Crill, F. Cuttaia, P. de Bernardis, G. de Zotti, J. Delabrouille, J. M. Delouis, E. Di Valentino, J. M. Diego, O. Doré, M. Douspis, A. Ducout, X. Dupac, S. Dusini, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, Y. Fantaye, M. Farhang, J. Fergusson, R. Fernandez-Cobos, F. Finelli, F. Forastieri, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frolov, S. Galeotta, S. Galli, K. Ganga, R. T. Génova-Santos, M. Gerbino, T. Ghosh, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gruppuso, J. E. Gudmundsson, J. Hamann, W. Handley, F. K. Hansen, D. Herranz, S. R. Hildebrandt, E. Hivon, Z. Huang, A. H. Jaffe, W. C. Jones, A. Karakci, E. Keihänen, R. Keskitalo, K. Kiiveri, J. Kim, T. S. Kisner, L. Knox, N. Krachmalnicoff, M. Kunz, H. Kurki-Suonio, G. Lagache, J. M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, M. Le Jeune, P. Lemos, J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. Lilley, V. Lindholm, M. López-Caniego, P. M. Lubin, Y. Z. Ma, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marcos-Caballero, M. Maris, P. G. Martin, M. Martinelli, E. Martínez-González, S. Matarrese, N. Mauri, J. D. McEwen, P. R. Meinhold, A. Melchiorri, A. Mennella, M. Migliaccio, M. Millea, S. Mitra, M. A. Miville-Deschênes, D. Molinari, L. Montier, G. Morgante, A. Moss, P. Natoli, H. U. Nørgaard-Nielsen, L. Pagano, D. Paoletti, B. Partridge, G. Patanchon, H. V. Peiris, F. Perrotta, V. Pettorino, F. Piacentini, L. Polastri, G. Polenta, J. L. Puget, J. P. Rachen, M. Reinecke, M. Remazeilles, A. Renzi, G. Rocha, C. Rosset, G. Roudier, J. A. Rubiño-Martín, B. Ruiz-Granados, L. Salvati, M. Sandri, M. Savelainen, D. Scott, E. P. S. Shellard, C. Sirignano, G. Sirri, L. D. Spencer, R. Sunyaev, A. S. Suur-Uski, J. A. Tauber, D. Tavagnacco, M. Tenti, L. Toffolatti, M. Tomasi, T. Trombetti, L. Valenziano, J. Valiviita, B. Van Tent, L. Vibert, P. Vielva, F. Villa, N. Vittorio, B. D. Wandelt, I. K. Wehus, M. White, S. D. M. White, A. Zacchei, and A. Zonca. Planck 2018 results. VI. Cosmological parameters. *A&A*, 641:A6, September 2020.

[154] Planck Collaboration, N. Aghanim, B. Altieri, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, N. Bartolo, E. Battaner, A. Beelen, K. Benabed, A. Benoit-Lévy, J. P. Bernard, M. Bersanelli, M. Bethermin, P. Bielewicz, L. Bonavera, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, C. Burigana, E. Calabrese, R. Canameras, J. F. Cardoso, A. Catalano, A. Chamballu, R. R. Chary, H. C. Chiang, P. R. Christensen, D. L. Clements, S. Colombi, F. Couchot,

B. P. Crill, A. Curto, L. Danese, K. Dassas, R. D. Davies, R. J. Davis, P. de Bernardis, A. de Rosa, G. de Zotti, J. Delabrouille, J. M. Diego, H. Dole, S. Donzelli, O. Doré, M. Douspis, A. Ducout, X. Dupac, G. Efstathiou, F. Elsner, T. A. Enßlin, E. Falgarone, I. Flores-Cacho, O. Forni, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frejsel, B. Frye, S. Galeotta, S. Galli, K. Ganga, M. Giard, E. Gjerløw, J. González-Nuevo, K. M. Górski, A. Gregorio, A. Gruppuso, D. Guéry, F. K. Hansen, D. Hanson, D. L. Harrison, G. Helou, C. Hernández-Monteagudo, S. R. Hildebrandt, E. Hivon, M. Hobson, W. A. Holmes, W. Hovest, K. M. Huffenberger, G. Hurier, A. H. Jaffe, T. R. Jaffe, E. Keihänen, R. Keskitalo, T. S. Kisner, R. Kneissl, J. Knoche, M. Kunz, H. Kurki-Suonio, G. Lagache, J. M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, E. Le Floc'h, R. Leonardi, F. Levrier, M. Liguori, P. B. Lilje, M. Linden-Vørnle, M. López-Caniego, P. M. Lubin, J. F. Macías-Pérez, T. MacKenzie, B. Maffei, N. Mandolesi, M. Maris, P. G. Martin, C. Martinache, E. Martínez-González, S. Masi, S. Matarrese, P. Mazzotta, A. Melchiorri, A. Mennella, M. Migliaccio, A. Moneti, L. Montier, G. Morgante, D. Mortlock, D. Munshi, J. A. Murphy, P. Natoli, M. Negrello, N. P. H. Nesvadba, D. Novikov, I. Novikov, A. Omont, L. Pagano, F. Pajot, F. Pasian, O. Perdereau, L. Perotto, F. Perrotta, V. Pettorino, F. Piacentini, M. Piat, S. Plaszczynski, E. Pointecouteau, G. Polenta, L. Popa, G. W. Pratt, S. Prunet, J. L. Puget, J. P. Rachen, W. T. Reach, M. Reinecke, M. Remazeilles, C. Renault, I. Ristorcelli, G. Rocha, G. Roudier, B. Rusholme, M. Sandri, D. Santos, G. Savini, D. Scott, L. D. Spencer, V. Stolyarov, R. Sunyaev, D. Sutton, J. F. Sygnet, J. A. Tauber, L. Terenzi, L. Toffolatti, M. Tomasi, M. Tristram, M. Tucci, G. Umana, L. Valenziano, J. Valiviita, I. Valtchanov, B. Van Tent, J. D. Vieira, P. Vielva, L. A. Wade, B. D. Wandelt, I. K. Wehus, N. Welikala, A. Zacchei, and A. Zonca. Planck intermediate results. XXVII. High-redshift infrared galaxy overdensity candidates and lensed sources discovered by Planck and confirmed by Herschel-SPIRE. *A&A*, 582:A30, October 2015.

[155] Erwin Platen, Rien van de Weygaert, and Bernard J. T. Jones. A cosmic watershed: the WVF void detection technique. *MNRAS*, 380(2):551–570, September 2007.

[156] Andrew Pontzen. Scale-dependent bias in the baryonic-acoustic-oscillation-scale intergalactic neutral hydrogen. *PhRvD*, 89(8):083010, April 2014.

[157] Andrew Pontzen, Simeon Bird, Hiranya Peiris, and Licia Verde. Constraints on Ionizing Photon Production from the Large-scale Lyα Forest. *ApJL*, 792(2):L34, September 2014.

[158] Natalia Porqueres, Oliver Hahn, Jens Jasche, and Guilhem Lavaux. A hierarchical field-level inference approach to reconstruction from sparse Lyman-α forest data. *A&A*, 642:A139, October 2020.

[159] M. Postman and M. J. Geller. The morphology-density relation - The group connection. *ApJ*, 281:95–99, June 1984.

[160] Anthony R. Pullen, Patrick C. Breysse, Trevor Oxholm, Eric R. Switzer, Christopher J. Anderson, Emily Barrentine, Alberto D. Bolatto, Giuseppe Cataldo, Thomas

Essinger-Hileman, Abhishek Maniyar, Thomas Stevenson, Rachel S. Somerville, Carrie Volpert, Edward Wollack, Shengqi Yang, L. Y. Aaron Yung, and Zilu Zhou. Extragalactic Science with the Experiment for Cryogenic Large-aperture Intensity Mapping. *arXiv e-prints*, page arXiv:2209.02497, September 2022.

[161] Mahdi Qezlou, Simeon Bird, Adam Lidz, Guochao Sun, Andrew B. Newman, Gwen C. Rudie, Yueying Ni, Rupert Croft, and Tiziana Di Matteo. Boosting line intensity map signal-to-noise ratio with the Ly-α forest cross-correlation. *MNRAS*, 524(2):1933–1945, September 2023.

[162] Mahdi Qezlou, Andrew B. Newman, Gwen C. Rudie, and Simeon Bird. Characterizing Protoclusters and Protogroups at z 2.5 Using Lyα Tomography. *ApJ*, 930(2):109, May 2022.

[163] A. Rahmati, A. H. Pawlik, M. Raicevic, and J. Schaye. On the evolution of the H I column density distribution in cosmological simulations. *MNRAS*, 430:2427–2445, April 2013.

[164] Alireza Rahmati, Andreas H. Pawlik, Milan Raičević, and Joop Schaye. On the evolution of the H I column density distribution in cosmological simulations. *MNRAS*, 430(3):2427–2445, April 2013.

[165] Michael Rauch. The Lyman Alpha Forest in the Spectra of QSOs. *ARA&A*, 36:267–316, January 1998.

[166] Michael Rauch, Jordi Miralda-Escudé, Wallace L. W. Sargent, Tom A. Barlow, David H. Weinberg, Lars Hernquist, Neal Katz, Renyue Cen, and Jeremiah P. Ostriker. The Opacity of the Lyα Forest and Implications for Ω_b and the Ionizing Background. *ApJ*, 489(1):7–20, November 1997.

[167] C. Ravoux, E. Armengaud, M. Walther, T. Etourneau, D. Pomarède, N. Palanque-Delabrouille, C. Yèche, J. Bautista, H. du Mas des Bourboux, S. Chabanier, K. Dawson, J. M. Le Goff, B. Lyke, A. D. Myers, P. Petitjean, M. M. Pieri, J. Rich, G. Rossi, and D. P. Schneider. A tomographic map of the large-scale matter distribution using the eBOSS—Stripe 82 Lyα forest. *JCAP*, 2020(7):010, July 2020.

[168] Pablo Renard, Enrique Gaztanaga, Rupert Croft, Laura Cabayol, Jorge Carretero, Martin Eriksen, Enrique Fernandez, Juan García-Bellido, Ramon Miquel, Cristobal Padilla, Eusebio Sanchez, and Pau Tallada-Crespí. The PAU survey: Ly α intensity mapping forecast. *MNRAS*, 501(3):3883–3899, March 2021.

[169] Allen Riddell, Ari Hartikainen, and Matthew Carter. pystan (3.0.0). PyPI, March 2021.

[170] Dominik A. Riechers, Riccardo Pavesi, Chelsea E. Sharon, Jacqueline A. Hodge, Roberto Decarli, Fabian Walter, Christopher L. Carilli, Manuel Aravena, Elisabete da Cunha, Emanuele Daddi, Mark Dickinson, Ian Smail, Peter L. Capak, Rob J. Ivison,

Mark Sargent, Nicholas Z. Scoville, and Jeff Wagg. COLDz: Shape of the CO Luminosity Function at High Redshift and the Cold Gas History of the Universe. *ApJ*, 872(1):7, February 2019.

[171] Keir K. Rogers, Simeon Bird, Hiranya V. Peiris, Andrew Pontzen, Andreu Font-Ribera, and Boris Leistedt. Correlations in the three-dimensional Lyman-alpha forest contaminated by high column density absorbers. *MNRAS*, 476(3):3716–3728, May 2018.

[172] Gwen C. Rudie, Charles C. Steidel, and Max Pettini. The Temperature-Density Relation in the Intergalactic Medium at Redshift langzrang = 2.4. *ApJL*, 757(2):L30, October 2012.

[173] Marcin Sawicki, Stephane Arnouts, Jiasheng Huang, Jean Coupon, Anneya Golob, Stephen Gwyn, Sebastien Foucaud, Thibaud Moutard, Ikuru Iwata, Chengze Liu, Lingjian Chen, Guillaume Desprez, Yuichi Harikane, Yoshiaki Ono, Michael A. Strauss, Masayuki Tanaka, Nathalie Thibert, Michael Balogh, Kevin Bundy, Scott Chapman, James E. Gunn, Bau-Ching Hsieh, Olivier Ilbert, Yipeng Jing, Olivier LeFèvre, Cheng Li, Yuichi Matsuda, Satoshi Miyazaki, Tohru Nagao, Atsushi J. Nishizawa, Masami Ouchi, Kazuhiro Shimasaku, John Silverman, Sylvain de la Torre, Laurence Tresse, Wei-Hao Wang, Chris J. Willott, Toru Yamada, Xiaohu Yang, and Howard K. C. Yee. The CFHT large area U-band deep survey (CLAUDS). *MNRAS*, 489(4):5202–5217, November 2019.

[174] Joop Schaye, Tom Theuns, Michael Rauch, George Efstathiou, and Wallace L. W. Sargent. The thermal history of the intergalactic medium*. *MNRAS*, 318(3):817–826, November 2000.

[175] Uroš Seljak, Alexey Makarov, Patrick McDonald, Scott F. Anderson, Neta A. Bahcall, J. Brinkmann, Scott Burles, Renyue Cen, Mamoru Doi, James E. Gunn, Željko Ivezić, Stephen Kent, Jon Loveday, Robert H. Lupton, Jeffrey A. Munn, Robert C. Nichol, Jeremiah P. Ostriker, David J. Schlegel, Donald P. Schneider, Max Tegmark, Daniel E. Berk, David H. Weinberg, and Donald G. York. Cosmological parameter analysis including SDSS Lyα forest and galaxy bias: Constraints on the primordial spectrum of fluctuations, neutrino mass, and dark energy. *PhRvD*, 71(10):103515, May 2005.

[176] Uros Seljak, Anze Slosar, and Patrick McDonald. Cosmological parameters from combining the Lyman-α forest with CMB, galaxy clustering and SN constraints. *JCAP*, 2006(10):014, October 2006.

[177] Anže Slosar, Andreu Font-Ribera, Matthew M. Pieri, James Rich, Jean-Marc Le Goff, Éric Aubourg, Jon Brinkmann, Nicolas Busca, Bill Carithers, Romain Charlassier, Marina Cortês, Rupert Croft, Kyle S. Dawson, Daniel Eisenstein, Jean-Christophe Hamilton, Shirley Ho, Khee-Gan Lee, Robert Lupton, Patrick McDonald, Bumbarija Medolin, Demitri Muna, Jordi Miralda-Escudé, Adam D. Myers, Robert C. Nichol, Nathalie Palanque-Delabrouille, Isabelle Pâris, Patrick Petitjean, Yodovina Piškur, Emmanuel Rollinde, Nicholas P. Ross, David J. Schlegel, Donald P. Schneider, Erin

Sheldon, Benjamin A. Weaver, David H. Weinberg, Christophe Yeche, and Donald G. York. The Lyman-α forest in three dimensions: measurements of large scale flux correlations from BOSS 1st-year data. *JCAP*, 2011(9):001, September 2011.

[178] Anže Slosar, Andreu Font-Ribera, Matthew M. Pieri, James Rich, Jean-Marc Le Goff, Éric Aubourg, Jon Brinkmann, Nicolas Busca, Bill Carithers, Romain Charlassier, Marina Cortês, Rupert Croft, Kyle S. Dawson, Daniel Eisenstein, Jean-Christophe Hamilton, Shirley Ho, Khee-Gan Lee, Robert Lupton, Patrick McDonald, Bumbarija Medolin, Demitri Muna, Jordi Miralda-Escudé, Adam D. Myers, Robert C. Nichol, Nathalie Palanque-Delabrouille, Isabelle Pâris, Patrick Petitjean, Yodovina Piškur, Emmanuel Rollinde, Nicholas P. Ross, David J. Schlegel, Donald P. Schneider, Erin Sheldon, Benjamin A. Weaver, David H. Weinberg, Christophe Yeche, and Donald G. York. The Lyman-α forest in three dimensions: measurements of large scale flux correlations from BOSS 1st-year data. *JCAP*, 2011(9):001, September 2011.

[179] Anže Slosar, Vid Iršič, David Kirkby, Stephen Bailey, Nicolás G. Busca, Timothée Delubac, James Rich, Éric Aubourg, Julian E. Bautista, Vaishali Bhardwaj, Michael Blomqvist, Adam S. Bolton, Jo Bovy, Joel Brownstein, Bill Carithers, Rupert A. C. Croft, Kyle S. Dawson, Andreu Font-Ribera, J. M. Le Goff, Shirley Ho, Klaus Honscheid, Khee-Gan Lee, Daniel Margala, Patrick McDonald, Bumbarija Medolin, Jordi Miralda-Escudé, Adam D. Myers, Robert C. Nichol, Pasquier Noterdaeme, Nathalie Palanque-Delabrouille, Isabelle Pâris, Patrick Petitjean, Matthew M. Pieri, Yodovina Piškur, Natalie A. Roe, Nicholas P. Ross, Graziano Rossi, David J. Schlegel, Donald P. Schneider, Nao Suzuki, Erin S. Sheldon, Uroš Seljak, Matteo Viel, David H. Weinberg, and Christophe Yèche. Measurement of baryon acoustic oscillations in the Lyman-α forest fluctuations in BOSS data release 9. *JCAP*, 2013(4):026, April 2013.

[180] Daniele Sorini, José Oñorbe, Zarija Lukić, and Joseph F. Hennawi. Modeling the Lyα Forest in Collisionless Simulations. *ApJ*, 827(2):97, August 2016.

[181] D. Spergel, N. Gehrels, J. Breckinridge, M. Donahue, A. Dressler, B. S. Gaudi, T. Greene, O. Guyon, C. Hirata, J. Kalirai, N. J. Kasdin, W. Moos, S. Perlmutter, M. Postman, B. Rauscher, J. Rhodes, Y. Wang, D. Weinberg, J. Centrella, W. Traub, C. Baltay, J. Colbert, D. Bennett, A. Kiessling, B. Macintosh, J. Merten, M. Mortonson, M. Penny, E. Rozo, D. Savransky, K. Stapelfeldt, Y. Zu, C. Baker, E. Cheng, D. Content, J. Dooley, M. Foote, R. Goullioud, K. Grady, C. Jackson, J. Kruk, M. Levine, M. Melton, C. Peddie, J. Ruffa, and S. Shaklan. Wide-Field InfraRed Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA Final Report. *arXiv e-prints*, page arXiv:1305.5422, May 2013.

[182] V. Springel. E pur si muove: Galilean-invariant cosmological hydrodynamical simulations on a moving mesh. *MNRAS*, 401:791–851, January 2010.

[183] Volker Springel. The cosmological simulation code GADGET-2. *MNRAS*, 364(4):1105–1134, December 2005.

[184] Volker Springel and Lars Hernquist. Cosmological smoothed particle hydrodynamics simulations: a hybrid multiphase model for star formation. *MNRAS*, 339(2):289–311, February 2003.

[185] Volker Springel, Rüdiger Pakmor, Annalisa Pillepich, Rainer Weinberger, Dylan Nelson, Lars Hernquist, Mark Vogelsberger, Shy Genel, Paul Torrey, Federico Marinacci, and Jill Naiman. First results from the IllustrisTNG simulations: matter and galaxy clustering. *MNRAS*, 475(1):676–698, March 2018.

[186] Volker Springel, Simon D. M. White, Giuseppe Tormen, and Guinevere Kauffmann. Populating a cluster of galaxies - I. Results at [formmu2]z=0. *MNRAS*, 328(3):726–750, December 2001.

[187] Casey W. Stark, Andreu Font-Ribera, Martin White, and Khee-Gan Lee. Finding high-redshift voids using Lyman α forest tomography. *MNRAS*, 453(4):4311–4323, November 2015.

[188] Casey W. Stark, Martin White, Khee-Gan Lee, and Joseph F. Hennawi. Protocluster discovery in tomographic Ly α forest flux maps. *MNRAS*, 453(1):311–327, October 2015.

[189] Charles C. Steidel, Kurt L. Adelberger, Mark Dickinson, Mauro Giavalisco, Max Pettini, and Melinda Kellogg. A Large Structure of Galaxies at Redshift Z approximately 3 and Its Cosmological Implications. *ApJ*, 492(2):428–438, January 1998.

[190] G. Sun, T. C. Chang, B. D. Uzgil, J. J. Bock, C. M. Bradford, V. Butler, T. Caze-Cortes, Y. T. Cheng, A. Cooray, A. T. Crites, S. Hailey-Dunsheath, N. Emerson, C. Frez, B. L. Hoscheit, J. Hunacek, R. P. Keenan, C. T. Li, P. Madonia, D. P. Marrone, L. Moncelsi, C. Shiu, I. Trumper, A. Turner, A. Weber, T. S. Wei, and M. Zemcov. Probing Cosmic Reionization and Molecular Gas Growth with TIME. *ApJ*, 915(1):33, July 2021.

[191] Guochao Sun. Cosmological Constraints on the Global Star Formation Law of Galaxies: Insights from Baryon Acoustic Oscillation Intensity Mapping. *ApJL*, 931(2):L29, June 2022.

[192] Guochao Sun, Lluís Mas-Ribas, Tzu-Ching Chang, Steven R. Furlanetto, Richard H. Mebane, Michael O. Gonzalez, Jasmine Parsons, and A. C. Trapp. LIMFAST. II. Line Intensity Mapping as a Probe of High-Redshift Galaxy Formation. *arXiv e-prints*, page arXiv:2206.14186, June 2022.

[193] Eric R. Switzer, Emily M. Barrentine, Giuseppe Cataldo, Thomas Essinger-Hileman, Peter A. R. Ade, Christopher J. Anderson, Alyssa Barlis, Jeffrey Beeman, Nicholas Bellis, Alberto D. Bolatto, Patrick C. Breysse, Berhanu T. Bulcha, Lee-Roger Chevres-Fernanadez, Chullhee Cho, Jake A. Connors, Negar Ehsan, Jason Glenn, Joseph Golec, James P. Hays-Wehle, Larry A. Hess, Amir E. Jahromi, Trevian Jenkins, Mark O. Kimball, Alan J. Kogut, Luke N. Lowe, Philip Mauskopf, Jeffrey McMahon, Mona Mirzaei, Harvey Moseley, Jonas Mugge-Durum, Omid Noroozian,

Trevor M. Oxholm, Tatsat Parekh, Ue-Li Pen, Anthony R. Pullen, Maryam Rahmani, Mathias M. Ramirez, Florian Roselli, Konrad Shire, Gage Siebert, Adrian K. Sinclair, Rachel S. Somerville, Ryan Stephenson, Thomas R. Stevenson, Peter Timbie, Jared Termini, Justin Trenkamp, Carole Tucker, Elijah Visbal, Carolyn G. Volpert, Edward J. Wollack, Shengqi Yang, and L. Y. Aaron Yung. Experiment for cryogenic large-aperture intensity mapping: instrument design. *Journal of Astronomical Telescopes, Instruments, and Systems*, 7:044004, October 2021.

[194] Masahiro Takada, Richard S. Ellis, Masashi Chiba, Jenny E. Greene, Hiroaki Aihara, Nobuo Arimoto, Kevin Bundy, Judith Cohen, Olivier Doré, Genevieve Graves, James E. Gunn, Timothy Heckman, Christopher M. Hirata, Paul Ho, Jean-Paul Kneib, Olivier Le Fèvre, Lihwai Lin, Surhud More, Hitoshi Murayama, Tohru Nagao, Masami Ouchi, Michael Seiffert, John D. Silverman, Laerte Sodré, David N. Spergel, Michael A. Strauss, Hajime Sugai, Yasushi Suto, Hideki Takami, and Rosemary Wyse. Extragalactic science, cosmology, and Galactic archaeology with the Subaru Prime Focus Spectrograph. *PASJ*, 66(1):R1, February 2014.

[195] The LUVOIR Team. The LUVOIR Mission Concept Study Final Report. *arXiv e-prints*, page arXiv:1912.06219, December 2019.

[196] Phoebe Upton Sanderbeck and Simeon Bird. Inhomogeneous He II reionization in hydrodynamic simulations. *MNRAS*, 496(4):4372–4382, August 2020.

[197] Phoebe Upton Sanderbeck and Simeon Bird. Inhomogeneous He II reionization in hydrodynamic simulations. *MNRAS*, 496(4):4372–4382, August 2020.

[198] Phoebe R. Upton Sanderbeck, Anson D'Aloisio, and Matthew J. McQuinn. Models of the thermal evolution of the intergalactic medium after reionization. *MNRAS*, 460(2):1885–1897, August 2016.

[199] Stéfan van der Walt, Johannes L. Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D. Warner, Neil Yager, Emmanuelle Gouillart, Tony Yu, and the scikit-image contributors. scikit-image: image processing in Python. *PeerJ*, 2:e453, 6 2014.

[200] Matteo Viel and Martin G. Haehnelt. Cosmological and astrophysical parameters from the Sloan Digital Sky Survey flux power spectrum and hydrodynamical simulations of the Lyman α forest. *MNRAS*, 365(1):231–244, January 2006.

[201] Eli Visbal and Abraham Loeb. Measuring the 3D clustering of undetected galaxies through cross correlation of their cumulative flux fluctuations from multiple spectral lines. *JCAP*, 2010(11):016, November 2010.

[202] Mark Vogelsberger, Shy Genel, Debora Sijacki, Paul Torrey, Volker Springel, and Lars Hernquist. A model for cosmological simulations of galaxy formation physics. *MNRAS*, 436(4):3031–3067, December 2013.

[203] Mark Vogelsberger, Shy Genel, Volker Springel, Paul Torrey, Debora Sijacki, Dandan Xu, Greg Snyder, Dylan Nelson, and Lars Hernquist. Introducing the Illustris Project: simulating the coevolution of dark and visible matter in the Universe. *MNRAS*, 444(2):1518–1547, October 2014.

[204] Rainer Weinberger, Volker Springel, Lars Hernquist, Annalisa Pillepich, Federico Marinacci, Rüdiger Pakmor, Dylan Nelson, Shy Genel, Mark Vogelsberger, Jill Naiman, and Paul Torrey. Simulating galaxy formation with black hole driven thermal and kinetic feedback. *MNRAS*, 465(3):3291–3308, March 2017.

[205] Rainer Weinberger, Volker Springel, Lars Hernquist, Annalisa Pillepich, Federico Marinacci, Rüdiger Pakmor, Dylan Nelson, Shy Genel, Mark Vogelsberger, Jill Naiman, and Paul Torrey. Simulating galaxy formation with black hole driven thermal and kinetic feedback. *MNRAS*, 465(3):3291–3308, March 2017.

[206] Laura Wolz, Alkistis Pourtsidou, Kiyoshi W. Masui, Tzu-Ching Chang, Julian E. Bautista, Eva-Maria Müller, Santiago Avila, David Bacon, Will J. Percival, Steven Cunnington, Chris Anderson, Xuelei Chen, Jean-Paul Kneib, Yi-Chao Li, Yu-Wei Liao, Ue-Li Pen, Jeffrey B. Peterson, Graziano Rossi, Donald P. Schneider, Jaswant Yadav, and Gong-Bo Zhao. H I constraints from the cross-correlation of eBOSS galaxies and Green Bank Telescope intensity maps. *MNRAS*, 510(3):3495–3511, March 2022.

[207] Gábor Worseck, Frederick B. Davies, Joseph F. Hennawi, and J. Xavier Prochaska. The Evolution of the He II-ionizing Background at Redshifts 2.3 ¡ z ¡ 3.8 Inferred from a Statistical Sample of 24 HST/COS He II Lyα Absorption Spectra. *ApJ*, 875(2):111, April 2019.

[208] Xiaohan Wu, Matthew McQuinn, Rahul Kannan, Anson D'Aloisio, Simeon Bird, Federico Marinacci, Romeel Davé, and Lars Hernquist. Imprints of temperature fluctuations on the z ˜ 5 Lyman-α forest: a view from radiation-hydrodynamic simulations of reionization. *MNRAS*, 490(3):3177–3195, December 2019.

[209] Yongda Zhu, George D. Becker, Sarah E. I. Bosman, Laura C. Keating, Valentina D'Odorico, Rebecca L. Davies, Holly M. Christenson, Eduardo Bañados, Fuyan Bian, Manuela Bischetti, Huanqing Chen, Frederick B. Davies, Anna-Christina Eilers, Xiaohui Fan, Prakash Gaikwad, Bradley Greig, Martin G. Haehnelt, Girish Kulkarni, Samuel Lai, Andrea Pallottini, Yuxiang Qin, Emma V. Ryan-Weber, Fabian Walter, Feige Wang, and Jinyi Yang. Long Dark Gaps in the Lyβ Forest at z ¡ 6: Evidence of Ultra-late Reionization from XQR-30 Spectra. *ApJ*, 932(2):76, June 2022.