Lyman-α Line Intensity Mapping

Measurements on Galactic to Cosmological Scales

Maja Lujan Niemeyer

München 2025

Lyman-α Line Intensity Mapping

Measurements on Galactic to Cosmological Scales

Maja Lujan Niemeyer

Dissertation
der Fakultät für Physik
der Ludwig-Maximilians-Universität
München

vorgelegt von
Maja Lujan Niemeyer
aus Chicago, Illinois, USA

München, den 10. Dezember 2025

Erstgutachter: Prof. Dr. Eiichiro Komatsu
Zweitgutachter: Prof. Dr. Jochen Weller
Tag der mündlichen Prüfung: 13. Februar 2026

Contents

List of Figures

List of Tables

Zusammenfassung

Linienintensitätskartierung (LIM) ist ein neues Werkzeug zur Erforschung von Astrophysik und Kosmologie im jungen Universum. LIM kartiert die Intensität einer atomaren oder molekularen Emissionslinie in einem dreidimensionalen Volumen, mitsamt des Lichts von anderenfalls schwer erfassbaren, lichtschwachen Galaxien und Gas, und bildet die Materieverteilung ab.

In dieser Dissertation führen wir zuerst eine LIM-Messung auf kleinen Skalen durch Stacking der Lyα-Emissionslinie um Lyα-lichtschwachen Galaxien in den Daten des Hobby-Eberly Telescope Dark Energy Experiments (HETDEX) durch und erforschen den dominanten Ursprung der Lyα-Photonen im zirkumgalaktischen Medium (CGM) von sternbildenden Galaxien. Wir stellen fest, dass die äußeren Bereiche der allgegenwärtigen Lyα-Halos für Lyα-helle und Lyα-lichtschwache Galaxien gleichermaßen hell sind. Dies deutet darauf hin, dass diese Photonen nicht von den gestackten Galaxien stammen, sondern von deren CGM oder von außerhalb ihrer Dunkle-Materie-Halos.

Als Vorbereitung für die Analyse und Interpretation der LIM-Daten entwickeln wir einen schnellen Simulationscode namens SIMPLE, der Galaxienkataloge und Linienintensitätskarten von beliebigen Emissionslinien mit Experiment-spezifischen Effekten erstellen kann. Nachdem wir den Code prozentgenau überprüfen, erstellen wir simulierte Daten für HETDEX und untersuchen, ob HETDEX das Kreuzleistungsspektrum von detektierten Lyα-emittierenden Galaxien (LAEs) und der Lyα-Intensität nachweisen kann.

Die vielfache Streuung von Lyα-Photonen kann einen Selektionseffekt von LAEs verursachen, der die Anisotropie des LAE-Leistungsspektrums verändert. Deshalb untersuchen wir den Einfluss der Lyα-Streuung auf LIM-Leistungsspektren der Lyα-Linie mithilfe eines linearen analytischen Modells und einer Umsetzung im SIMPLE-Code. Wir zeigen, dass das lineare Modell ungeeignet ist, da die meiste Streuung in stark überdichten, nichtlinearen Regionen nahe den LAEs erfolgt. Wir zeigen, dass das von HETDEX gemessene LAE-Leistungsspektrum von diesem Effekt stark beeinflusst wird, während das Kreuzleistungsspektrum von LAEs und der Lyα-Intensität zu verrauscht ist, um diesen Effekt wahrzunehmen.

Abschließend messen wir das Kreuzleistungsspektrum von detektierten LAEs mit der Lyα-Intensität undetektierter Quellen in HETDEX und detektieren das Signal deutlich. Wir säubern die Daten gründlich, um systematische Anteile zu entfernen. Wir modellieren die gesamte Datenverarbeitung und Messung mit simulierten Daten des SIMPLE-Codes und berechnen das erwartete Signal und Kovarianzmatrizen. Wir ermitteln anschließend die mittlere Lyα-Intensität im Universum. Unser ermittelter Wert is deutlich niedriger als diejenigen, die von Kreuzkorrelationen von Quasaren mit der Lyα-Intensität abgeleitet wurden, jedoch etwas höher als eine Obergrenze, die von einer Kreuzkorrelation von dem Lyα-Wald mit der Lyα-Intensität gewonnen wurde. Unsere Ergebnisse stimmen grob mit der Vorhersage einer kosmologischen Simulation mit Lyα-Strahlungstransport-Modellierung überein. Die mittlere Lyα-Intensität kann außerdem vollständig von Sternbildung in Galaxien produziert werden.

Zusammenfassend entwickeln wir einen vielseitigen, schnellen Simulationscode für LIM, untersuchen den Effekt von Lyα-Streuung auf LIM-Leistungsspektren und führen eine präzise Messung von Lyα-LIM in HETDEX-Daten durch. Diese Forschungsarbeit ebnet den Weg für die Einschränkung von Modellen zur Galaxienentstehung und -entwicklung mit zukünftigen, noch präziseren LIM-Messungen.

Abstract

Line intensity mapping (LIM) is a novel tool for studying astrophysics and cosmology in the young Universe. LIM maps the intensity of an atomic or molecular emission line in a three-dimensional volume, including the light from otherwise elusive faint galaxies and gas, and traces the matter distribution.

In this thesis, we first conduct a small-scale LIM measurement of the Lyα line by stacking Lyα-faint galaxies in the data of the Hobby-Eberly Telescope Dark Energy Experiment (HETDEX) and study the dominant origin of Lyα photons in the circumgalactic medium (CGM) of star-forming galaxies. We find that the outer parts of the ubiquitous Lyα halos is equally bright for Lyα-bright and Lyα-faint galaxies. This indicates that these photons do not originate from the stacked galaxies, but rather from the galaxies' CGM or from outside of their dark matter halos.

In preparation for the analysis and interpretation of LIM data, we develop a fast simulation code to create mock galaxy catalogs and line intensity maps of any emission line, called SIMPLE, assuming that the underlying mass distribution is lognormal and assuming an input power spectrum and line luminosity function. One can apply survey-specific effects to the intensity maps such as angular and spectral smoothing, a mask, and sky subtraction, to make the mock data more realistic. After validating the output of the code to percent-level precision, we create mock data for HETDEX and examine its ability to measure the cross-power spectrum monopole of detected Lyα-emitting galaxies (LAEs) with the Lyα intensity of undetected sources.

Lyα photons scatter many times within the interstellar medium, changing the shape and central wavelength of the Lyα emission line escaping the galaxy. Lyα photons can also scatter in the CGM and intergalactic medium (IGM). The scattering probability depends on the matter and velocity distribution, which can introduce a selection effect of LAEs that changes the anisotropy of the LAE auto-power spectrum. Therefore, we investigate the effect of Lyα scattering on Lyα LIM power spectra using a linear analytic model and an implementation in the SIMPLE code. We find that the linear analytic model is inadequate to describe this effect because most of the scattering takes place in highly overdense, nonlinear regions near the LAEs. We show that the LAE auto-power spectrum measured by HETDEX will be sensitive to this effect, while the LAE-Lyα intensity cross-power spectrum is too noisy to be affected.

Finally, we measure the cross-power spectrum of detected LAEs and the Lyα intensity of undetected sources in HETDEX. We thoroughly clean the spectra to remove systematic contributions. We forward-model the entire data set and data processing using mock data from the SIMPLE code and calculate the expected signal and covariance matrices. We then infer the mean intensity of Lyα photons permeating the Universe. Our inferred value is significantly smaller than those inferred from cross-correlations of quasars with the Lyα intensity, but slightly higher than the upper limit inferred from a cross-correlation of the Lyα forest with the Lyα intensity. Our results roughly agree with predictions from a cosmological simulation including Lyα radiative transfer modeling. The mean intensity is also consistent with being entirely sourced by star formation in galaxies.

In summary, we create a versatile mock simulation code for LIM, model the effect of Lyα radiative transfer on Lyα LIM measurements, and conduct a precise Lyα LIM measurement in HETDEX data. This research paves the way for constraining galaxy formation and evolution models with upcoming, even more precise LIM measurements.

Chapter 1

Introduction

1.1 Why Line Intensity Mapping?

Galaxy formation and evolution are still a mystery — highlighted by recent unexpected results from the James-Webb Space Telescope: bright, massive galaxies have formed at earlier times than state-of-the-art cosmological and galaxy formation models predict (see Stark et al. (2025) for a review of observational insight; and Dayal & Ferrara (2018) for a review on early galaxy formation theory). While it is important to study individual galaxies throughout cosmic time, this approach is limited to the brightest galaxies that rise above the noise level of deep observations, probing only the "tip of the iceberg" of the galaxy population. Faint galaxies and the abundant gas around and in between galaxies are extremely difficult to observe individually. The gas surrounding galaxies is vital for their evolution, and we do not fully understand how the interplay between galaxies and the circumgalactic (CGM) and intergalactic media (IGM) affects star formation (Tumlinson et al., 2017).

Instead of observing individual galaxies, we can map the intensity of an atomic or molecular emission line within a three-dimensional volume. This approach is called "line intensity mapping" (LIM; Kovetz et al., 2017; Bernal & Kovetz, 2022). LIM measures the intensity from all sources, including faint galaxies and gas, and thus probes the global history of star formation and the matter distribution. Different lines are emitted in different regions and from different types of gas, for example cold molecular clouds (CO and [C II]), the ionized interstellar medium (ISM; Lyα, Hα, and [O II]), and neutral hydrogen (21 cm). Combining the intensity maps of different emission lines thus promises to reveal a wealth of information on the physics of gas and galaxies.

Figure 1.1 illustrates the LIM observing method, showing all galaxies in a simulated volume, only those that are brigh enough to be detected, and the total Lyα intensity. While galaxy surveys can only find a small subset of the galaxies, LIM is sensitive to the intensity from all galaxies.

Because an intensity map is a biased tracer of the matter distribution, it can also be used for cosmology. Because LIM surveys do not require deep observations, they will be able to reach larger scales than galaxy surveys, at a lower cost. In the long term, LIM will constrain all cosmological parameters, including the nature of dark energy, modified

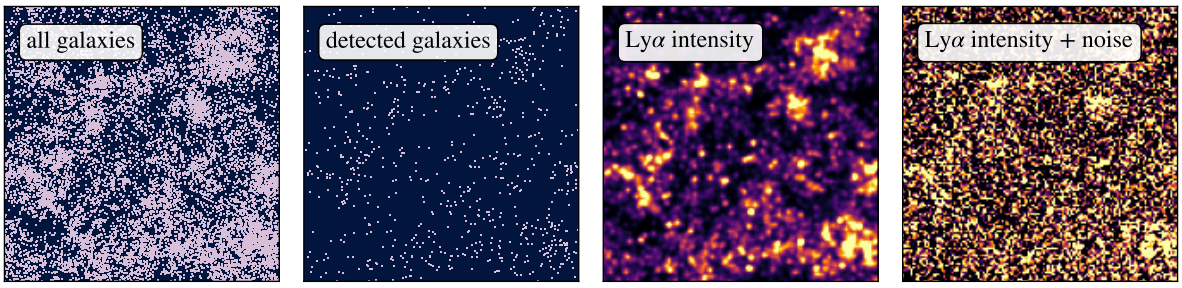

Figure 1.1: Simulated $(32\times128\times128)\,\mathrm{Mpc}^3h^{-3}$ comoving volume of galaxies with minimum Lyα luminosity $L_{\mathrm{min}} = 10^{41}\,\mathrm{erg\,s^{-1}}$ (first panel), galaxies with a flux above $5\times10^{-17}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$ that could be detected by a HETDEX-like survey (second panel), noise-less Lyα intensity map averaged over the first dimension (third panel) for a LIM survey similar to HETDEX, and Lyα intensity map including noise similar to HETDEX (fourth panel); using the SIMPLE code (see Chapter 5). With LIM, no photons are left behind.

gravity (Moradinezhad Dizgah et al., 2024), beyond-Λ-cold-dark-matter (ΛCDM) models (Sabla et al., 2024), neutrino physics (Moradinezhad Dizgah et al., 2022a), and primordial non-Gaussianity (Moradinezhad Dizgah & Keating, 2019). Due to its sensitivity to the total intensity of all sources, LIM will also reach the redshift range inaccessible to galaxy surveys and the cosmic microwave background (CMB). For instance, the universe was dark between recombination and the time when the first stars formed, except for the diffuse 21 cm glow of neutral hydrogen (e.g., Padmanabhan, 2024). Using space-based or lunar observatories, it will be possible to access these redshifts with 21 cm LIM for precision cosmology (Mondal & Barkana, 2023; Mondal et al., 2024).

LIM will also give insight into the Epoch of Reionization (EoR). Cross-correlating intensity maps of different emission lines, such as 21 cm, Lyα, CO, and [C II], will uncover the sources that drive reionization and study the evolution of the ionization state, morphology, and metal content of the IGM (Lidz et al., 2009, 2011; Gong et al., 2012; Dumitru et al., 2019; Cox et al., 2022; Ambrose et al., 2025). For these reasons, LIM is considered the next frontier of astrophysics and cosmology.

1.2 Why Lyman-α?

The Lyα emission line of the hydrogen atom is one of the brightest emission lines produced in the universe and an important observational probe at high redshift (Ouchi et al., 2020). Historically, Partridge & Peebles (1967) predicted that high-redshift galaxies emit Lyα photons through recombination radiation in the ISM, which had been ionized by young, massive stars. They predicted that the Lyα emission could make up $6-7\%$ of the total luminosity of a galaxy, bright enough to be detected from Earth. While their predicted luminosity was an overestimate, over a million Lyα-emitting galaxies (LAEs) have been detected by surveys such as the Hobby-Eberly Telescope Dark Energy Experiment (HETDEX; Gebhardt et al., 2021) and tens of thousands are publicly cataloged (e.g., Mentuch Cooper et al., 2023a).

Neutral hydrogen has a large cross-section around the Lyα line, making it prone to scattering (e.g., Dijkstra, 2019). As a result, Lyα photons produced in the ISM of a galaxy can scatter in its CGM and the IGM and light up the surrounding gas as Lyα halos (e.g., Wisotzki et al., 2018). Collisional excitation and recombinations of hydrogen ionized by the ultraviolet background contribute to the diffuse Lyα glow. Because hydrogen is the most abundant element, the entire universe glows in Lyα after the reionization of hydrogen (e.g., Wisotzki et al., 2018).

The abundance of LAEs at high redshift and the glow of the abundant diffuse hydrogen gas make the Lyα line an ideal target for LIM. However, few Lyα LIM measurements have been conducted, and theoretical predictions of the Lyα luminosity density vary widely. Figure 1.2 shows a comparison of theoretical predictions and observational constraints of the comoving Lyα luminosity density. Some predictions are obtained by integrating the (extrapolated) Lyα luminosity function within a luminosity range, $L_{\mathrm{min}} \leq L \leq L_{\mathrm{max}}$, which by construction does not contain diffuse IGM emission. I chose $L_{\mathrm{min}} = 10^{39}\,\mathrm{erg\,s^{-1}}$ to $L_{\mathrm{max}} = 10^{43}\,\mathrm{erg\,s^{-1}}$ for the luminosity functions of Ouchi et al. (2008), Konno et al. (2016), the Schechter components, i.e., excluding active galactic nuclei (AGNs), of Umeda et al. (2025), and Tornotti et al. (2025a). Comaschi & Ferrara (2016) used an analytic model to predict the luminosity density originating from the ISM (shown as a red star in the figure) and the IGM (shown as a red empty circle) at $z = 4$. Byrohl & Nelson (2023) post-process a cosmological hydrodynamic simulation at $z = 2$ with Lyα emission and radiative transfer (RT). They predict the mean Lyα luminosity density of the universe at that redshift (orange filled circle) and the contributions from galaxies (orange star), the CGM (orange diamond), and the IGM (orange empty circle). All of these values correspond to photons as they would be observed after scattering. This shows that most Lyα photons illuminate diffuse gas that is undetectable in galaxy surveys. Lyα LIM will not leave these photons behind.

The lowest-redshift observational constraint is the broadband intensity mapping by Chiang et al. (2019). At higher redshift, the constraints from the quasar-Lyα intensity cross-correlation of Croft et al. (2016), Croft et al. (2018), and Lin et al. (2022), and the upper limit from the Lyα forest-Lyα intensity cross-correlation of Croft et al. (2018) are shown, all assuming a fiducial Lyα intensity bias $b_I = 3$. The following section describes the observations in more detail.

1.3 Previous Lyman-α Intensity Mapping Studies

By stacking narrowband or pseudo-narrowband images around galaxies, many studies have shown the ubiquity of faint Lyα halos around LAEs and other star-forming galaxies (e.g., Steidel et al., 2011; Wisotzki et al., 2018; Kikuchihara et al., 2022a; Lujan Niemeyer et al., 2022a; Kikuta et al., 2023; Trainor et al., 2025). As we will show in Section 3.3, the stacked intensity profile of LAEs is similar to the angular cross-correlation of LAEs and the Lyα intensity. These measurements show that the intensity profile flattens at proper distances $\gtrsim 20\,\mathrm{kpc}$. The flattening has been predicted by Byrohl et al. (2021) using a cosmological hydrodynamic simulation post-processed with Lyα RT. In their simulation, the flattening is caused by photons originating from other galaxies in other dark matter

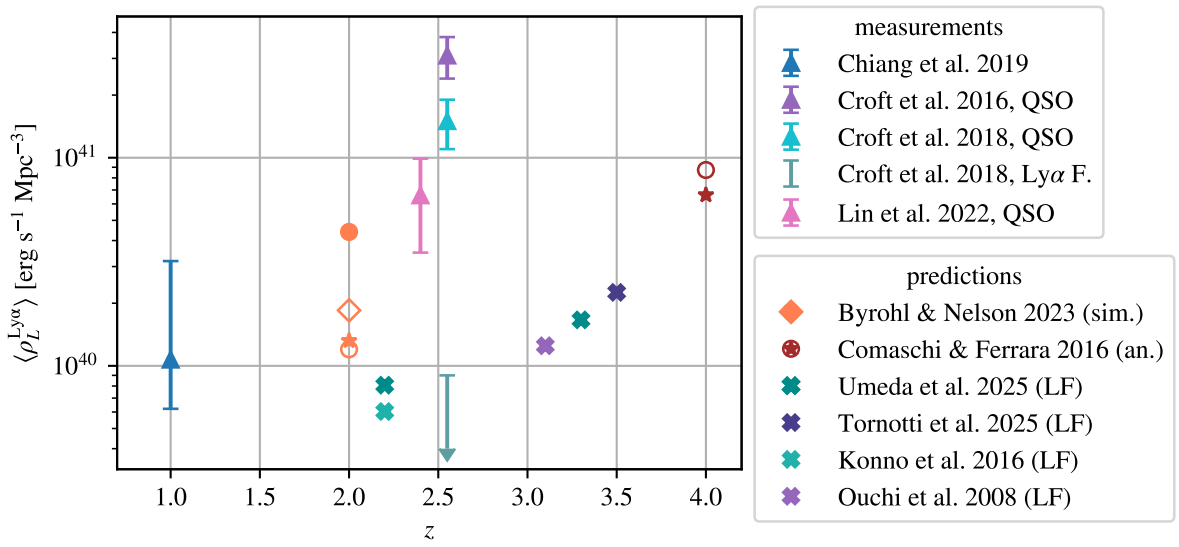

Figure 1.2: Measured (triangles) and predicted mean comoving Lyα luminosity densities. The measurements include the broad-band intensity mapping technique of Chiang et al. (2019), the QSO-Lyα intensity cross-correlation of Croft et al. (2016), Croft et al. (2018), and Lin et al. (2022), and the Lyα forest-Lyα intensity cross-correlation of Croft et al. (2018). The RT simulation of Byrohl & Nelson (2023) predicts the total luminosity density (orange filled circle), that of the CGM (diamond), the IGM (empty circle), and galaxies (star). The analytical model of Comaschi & Ferrara (2016) for the diffuse emission (empty circle) and for galaxies (star) is shown in red. The crosses show predictions of the luminosity density from galaxies obtained by integrating over the luminosity functions of Ouchi et al. (2008), Konno et al. (2016), Tornotti et al. (2025a), and the Schechter components of Umeda et al. (2025), from $L_{\mathrm{min}} = 10^{39}\,\mathrm{erg\,s^{-1}}$ to $L_{\mathrm{max}} = 10^{43}\,\mathrm{erg\,s^{-1}}$.

halos that scatter in the CGM of the stacked target sources. We will show in Chapter 4 that the Lyα intensity profiles at proper distances $\gtrsim 40\,\mathrm{kpc}$ is independent of the Lyα emission of the stacked sources, which supports this picture.

On larger scales, Croft et al. (2016) cross-correlate residual galaxy spectra with high-redshift quasar (QSO) positions to extract the QSO-Lyα cross-correlation signal at redshifts $2 < z < 3.5$. Lacking direct observations, the authors obtain the Lyα intensity from spectra of low-redshift luminous red galaxies (LRGs) after subtracting the best-fit galaxy spectrum. This process involves a delicate data analysis. Assuming a Lyα intensity bias of $b_I = 3$, their measured mean Lyα intensity is $\langle I_\lambda^{\mathrm{Ly}\alpha,\mathrm{C16}} \rangle = (3.9 \pm 0.9) \times 10^{-21}\,\mathrm{erg\,s^{-1}\,cm^{-2}\,\text{Å}^{-1}\,arcsec^{-2}}$, which is $21-35$ times higher than estimates using the luminosity function of individual LAEs. This implies that 97% of Lyα photons in the universe are produced by diffuse gas or faint galaxies below the detection limit of previous surveys. Croft et al. (2016) also find redshift-space distortions (RSD) of the QSO-Lyα cross-correlation that are inverted compared to the Kaiser effect (Kaiser, 1987) as predicted by Zheng et al. (2011) (see Section 2.4.2). While the Kaiser effect predicts $\beta_{\mathrm{Ly}\alpha}^{\mathrm{Kaiser}} = \Omega_{\mathrm{m}}^{0.55}(z = 2.55)/b_I \simeq 0.32$, they measure $\beta_{\mathrm{Ly}\alpha}^{\mathrm{C16}} = -0.76 \pm 0.36$.

However, in a more thorough analysis, Croft et al. (2018) find that Lyα emission

from the QSOs scattered within the camera into adjacent fibers, creating an artificially extended cross-correlation signal. After removing these fibers from the data, they measure a lower signal by 50%. Assuming a Lyα intensity bias of $b_I = 3$ and adopting $\beta_{\mathrm{Ly}\alpha}^{\mathrm{C16}}$, they find a mean intensity $\langle I_\lambda^{\mathrm{Ly}\alpha,\mathrm{C18}} \rangle = (1.9 \pm 0.5) \times 10^{-21} \mathrm{erg\,s^{-1}\,cm^{-2}\,\AA^{-1}\,arcsec^{-2}}$ from the QSO-Lyα emission cross-correlation. They do not detect the cross-correlation of Lyα emission with the Lyα forest, implying an upper limit on the mean Lyα intensity of $\langle I_\lambda^{\mathrm{Ly}\alpha,\mathrm{C18}} \rangle < 1.2 \times 10^{-22} \mathrm{erg\,s^{-1}\,cm^{-2}\,\AA^{-1}\,arcsec^{-2}}$ with 95% confidence. Because of the inconsistency between the two measurements, the authors argue that the strong QSO-Lyα intensity cross-correlation signal is not produced by star-forming galaxies but by the impact of QSOs on their surroundings, even on $\sim 20\,\mathrm{Mpc}$ scales. Croft et al. (2018) do not measure RSD in this work because of the lower signal-to-noise ratio.

Lin et al. (2022) use more recent data with more LRG spectra and a larger QSO catalog from the eBOSS survey (Ahumada et al., 2020) to repeat the analysis of Croft et al. (2018). They detect the QSO-Lyα cross-correlation with a slightly lower mean intensity than Croft et al. (2018): $\langle I_\lambda^{\mathrm{Ly}\alpha,\mathrm{L22}} \rangle = (1.44^{+0.45}_{-0.38}) \times 10^{-21} \mathrm{erg\,s^{-1}\,cm^{-2}\,\AA^{-1}\,arcsec^{-2}}$. While they rule out the scenario in which QSOs source most of these Lyα photons based on calculations of the required QSO luminosities, observed LAEs can still only account for $\sim 12\%$ of the measured Lyα photons, assuming a Lyα intensity bias of $b_I = 3$. They argue that most star-forming galaxies, including non-LAEs, produce Lyα photons, which are then scattered out of the galaxy into the halo, contributing to the mean intensity. Lin et al. (2022) find a large uncertainty of the RSD parameter $\beta_{\mathrm{Ly}\alpha}^{\mathrm{L22}} = 0.07^{+1.65}_{-0.73}$. This is consistent with the absence of an RT effect, which would imply $\beta_{\mathrm{Ly}\alpha}^{\mathrm{Kaiser}} = \Omega_{\mathrm{m}}^{0.55}(z = 2.4)/b_I \simeq 0.32$. While the authors do not detect the Lyα forest-Lyα intensity cross-correlation, their measurement is consistent with the model using the mean intensity derived from the QSO-Lyα intensity cross-correlation due to the large measurement uncertainty.

Chiang et al. (2019) employ a broad-band intensity mapping technique to constrain the Lyα luminosity density at $z \simeq 1$. They spatially cross-correlate the broad-band intensity from the GALEX far-UV (1500Å) and near-UV (2300Å) bands with a million spectroscopic objects in the Sloan Digital Sky Survey as a function of redshift. They measure the Lyman break strength, the non-ionizing UV continuum, and the Lyα equivalent width of the UV background. They constrain the Lyα luminosity density at $z \simeq 1$ to be $\rho_{\mathrm{Ly}\alpha}^{\mathrm{Ch19}}(z = 1) = 1.1^{+2.1}_{-0.5} \times 10^{40}\,\mathrm{erg\,s^{-1}\,Mpc^{-3}}$. The authors argue that it is consistent with expectations of galaxies and QSOs inferred from measurements of the Hα luminosity density, while the contribution from the IGM has to be small.

1.4 The Hobby-Eberly Telescope Dark Energy Experiment

The Hobby-Eberly Telescope Dark Energy Experiment (HETDEX) spectroscopically maps the locations of one million LAEs at $z \sim 1.88 - 3.52$ to measure their clustering and constrain cosmological parameters (Gebhardt et al., 2021). HETDEX aims to measure the Hubble expansion rate $H(z)$ and the angular diameter distance $D_A(z)$ at these redshifts to constrain the evolution of the dark energy equation of state parameter w_{de}.

The Hubble expansion rate $H = \dot{a}/a$ is described by the Friedmann equation,

$$\frac{H^2}{H_0^2} = \Omega_{\mathrm{r},0} a^{-4} + \Omega_{\mathrm{m},0} a^{-3} + \Omega_{\mathrm{de},0} e^{-3\int_0^a da' \frac{1+w_{\mathrm{de}}(a')}{a'}} + \Omega_{\kappa,0} a^{-2}, \tag{1.1}$$

where H_0 is the current expansion rate of the universe, $a = (1+z)^{-1}$ is the scale factor, the subscript "r" stands for radiation, "m" for matter, and κ is the spatial curvature of the universe. The Ω parameters are given by the ratio of the current (denoted by "0") energy densities and the critical density:

$$\Omega_{\mathrm{r},0} = \frac{\rho_{\mathrm{r},0}}{\rho_{\mathrm{crit},0}}\ , \qquad \Omega_{\mathrm{m},0} = \frac{\rho_{\mathrm{m},0}}{\rho_{\mathrm{crit},0}}\ , \qquad \Omega_{\mathrm{de},0} = \frac{\rho_{\mathrm{de},0}}{\rho_{\mathrm{crit},0}}\ , \quad \text{and} \quad \Omega_{\kappa,0} = -\frac{\kappa c^2}{H_0^2}. \tag{1.2}$$

The critical density is the density at which the spatial curvature of the universe vanishes,

$$\rho_{\mathrm{crit},0} = \frac{3H_0^2}{8\pi G}, \tag{1.3}$$

and G is the gravitational constant. The angular diameter distance in a spatially flat universe is given by

$$D_A(z) = \frac{1}{1+z}\int_0^z \frac{c dz'}{H(z')}. \tag{1.4}$$

HETDEX observes the sky with the Visible Integral-Field Replicable Unit Spectrograph (VIRUS; Hill et al., 2021) on the Hobby-Eberly Telescope (HET; Ramsey et al., 1994, 1998). The HET has an 11-m diameter primary mirror made from 91 identical 1-m hexagonal segments. VIRUS consists of 78 integral-field unit spectrographs (IFUs). Each IFU covers $51'' \times 51''$ on the sky and contains 448 optical fibers that are $1.5''$ in diameter. The IFUs are arranged on a $100''$ grid on the $18'$ wide field of view. The fibers from each IFU feed two spectrographs. The CCD camera of each spectrograph is read out by two amplifiers, so that the fibers from each IFU are split up into four amplifiers (see bottom left cutout of Figure 1.3). The resolving power of VIRUS is $R = \frac{\lambda}{\Delta\lambda} \simeq 800$, corresponding to a line full-width-at-half-maximum (FWHM) of $\Delta\lambda \simeq 5.6\,\text{Å}$ at $\lambda = 4500\,\text{Å}$. Figure 1.3 shows the IFU and fiber layouts of VIRUS.

Each HETDEX observation consists of three slightly dithered exposures to fill in the gaps between the fibers in each IFU. The gaps between the IFUs remain. Therefore, each HETDEX observation consists of $> 100,000$ fiber spectra. HETDEX is designed to detect roughly 2.5 LAEs per IFU observation on average to meet the science requirements. Each fiber spectrum has 1036 spectral pixels, so each IFU observation has $\simeq 1.4$ million voxels (3-dimensional pixels). Therefore, most of the spectra are noise-dominated. However, they are not empty, as faint galaxies below the detection limit and diffuse Lyα emission send photons into VIRUS. By cross-correlating the detected LAEs with the intensity in the "empty" spectra, we will recover the signal from these faint sources in Chapter 7.

HETDEX observes $540\,\text{deg}^2$ on the sky without pre-selecting targets. The total observed comoving volume comprises $10.9\,\text{Gpc}^3$. The observations are taken in two main fields - the Spring and Fall fields, and smaller science verification fields - the COSMOS, extended growth strip (EGS), and GOODS-N fields. The TESLA survey (Chávez Ortiz

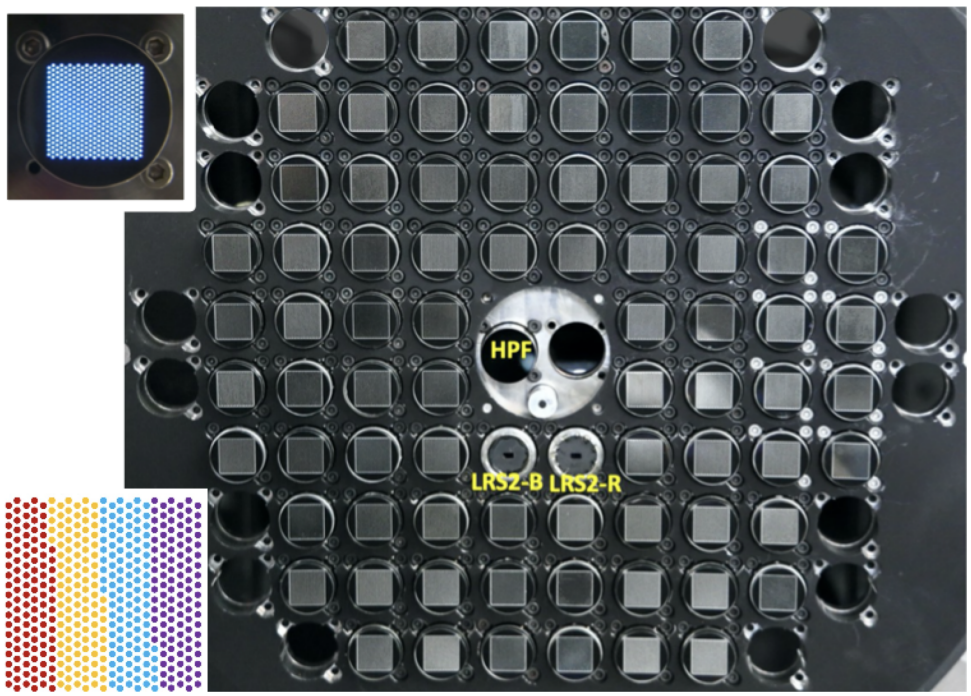

Figure 1.3: The focal plane of the HET. Each square is one IFU of VIRUS, with a zoom-in shown in the top left cutout. The bottom left cutout shows the fiber layout, where the color of the fibers shows which CCD amplifier reads out their spectra. The Habitable Zone Planet Finder (HPF) and the two IFUs for the Low Resolution Spectrograph (LRS2-B and LRS2-R) are located in the center of the focal plane. From Hill et al. (2021).

et al., 2023) uses HETDEX-like observations in the North Ecliptic Pole (NEP) field, which are also used in Chapter 7. Figure 1.4 shows the tiling of HETDEX observations in the Spring and Fall fields, along with a zoom-in to a smaller region, and the fiber layout including the three-exposure dithering pattern.

After observing, HETDEX processes the data in several steps, described in detail in Gebhardt et al. (2021). The sky subtraction, an especially important data processing step for LIM, estimates and removes the emission from airglow, zodiacal light, and other bright, but spatially smooth, foreground emissions. These emissions are many orders of magnitude brighter than the Lyα intensity, and residual sky emission is the dominant contaminator of the line intensity map. The sky subtraction also removes the mean Lyα intensity at each wavelength, thereby reducing the large-scale intensity fluctuations (see Section 3.5.1).

After subtracting the sky foreground, a line detection algorithm finds emission lines. A detected line can be a detector artifact, a false-positive detection, or an emission line from an astrophysical object. False positive detections include residuals from an imperfect sky subtraction, especially around bright sky emission or absorption lines, and malfunctioning detectors. Emission lines from astrophysical objects within the VIRUS wavelength range are predominantly Lyα emission at $1.88 < z < 3.52$, i.e., the target LAEs, and the [O II] emission line doublet at $\bar{\lambda}_{[\mathrm{OII}]} = 3727\,\text{Å}$ from galaxies at $z \lesssim 0.47$. A Bayesian analysis based on several indicators, such as the line equivalent width, is used

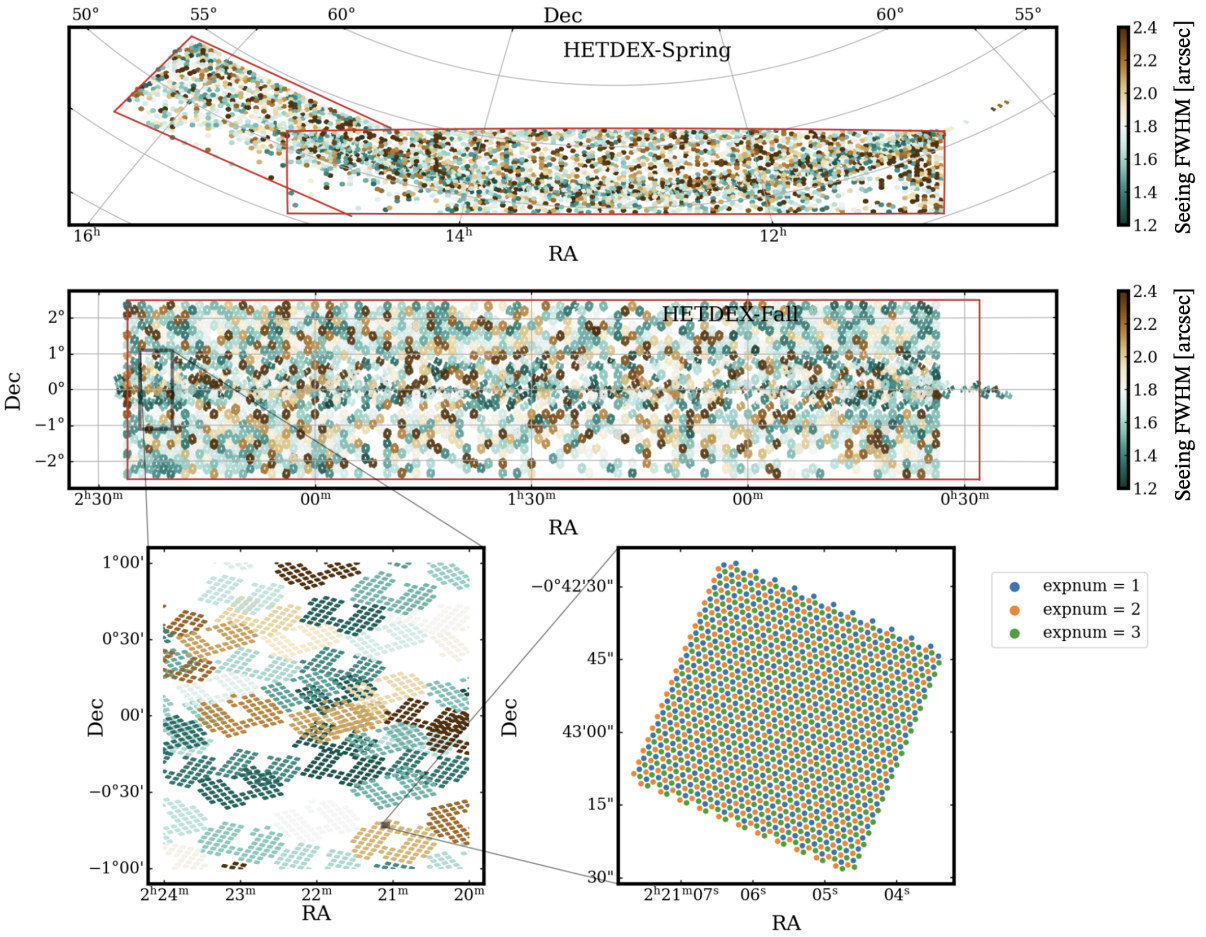

Figure 1.4: Observation tiling in the HETDEX Spring (top panel) and Fall (middle panel) fields (Mentuch Cooper et al. in preparation). The colors show the angular resolution FWHM in arcseconds, which is limited by the seeing. The bottom left panel shows a cutout of the Fall field with sometimes overlapping observations. The bottom right panel shows a cutout of one IFU, where the color of the fibers refers to the exposure number.

to determine whether a detected emission line is an LAE, an [O II]-emitting galaxy, or another object (Leung et al., 2017; Farrow et al., 2021; Davis et al., 2023a). Objects with strong continuum emission, such as stars and nearby galaxies, are collected in a separate catalog. A subset of the HETDEX continuum and emission line catalogs is published in Mentuch Cooper et al. (2023a).

1.5 Upcoming Lyman-α Intensity Mapping Studies

The Spectro-Photometer for the History of the Universe, EoR, and Ices Explorer (SPHEREX; Doré et al., 2014, 2016, 2018) maps the entire sky at infrared wavelengths between 0.75 μm and 5 μm with low spectral resolution ($R \simeq 35-130$, depending on wavelength). Because the SPHEREX spacecraft follows a polar orbit, the observations near the north and south ecliptic poles are deeper. The wavelength range of SPHEREX includes the Lyα line at

$z > 5.17$. Because the Universe is not yet completely ionized at these high redshifts, Lyα will be mostly absorbed in neutral regions. It is therefore an interesting probe of reionization (Ambrose et al., 2025). SPHEREX uses Linear Variable Filters (LVFs) to obtain spectra. LVFs transmit different wavelengths at different positions along one direction of the filter. With overlapping, but offset observations, every point in the survey is observed with every part of the filter to measure the low-resolution spectrum.

The proposed space mission called Cosmic Dawn Intensity Mapper (CDIM; Cooray et al., 2019) uses LVFs to map the intensity between $0.75\,\mu$m and $7.5\,\mu$m, which covers the Lyα line at $z > 5.17$, with a higher spectral resolution of $R \sim 300$ and a $9\,\mathrm{deg}^2$ field of view. While SPHEREX observes the entire sky, CDIM will observe a deep field ($15\,\mathrm{deg}^2$), a medium-depth field ($30\,\mathrm{deg}^2$), and a wide field ($300\,\mathrm{deg}^2$) overlapping with galaxy and 21 cm LIM surveys.

Padmanabhan & Loeb (2024a) predict a detection of scattered Lyα emission from LAEs in the surrounding IGM at redshifts $9 < z < 12$ both in autocorrelation and cross-correlation with 21 cm LIM. For the Lyα observations, they propose using SPHEREX and CDIM. Individual detections and detections using stacking with the James-Webb Space Telescope (JWST; Gardner et al., 2023) are also possible (Padmanabhan & Loeb, 2024b).

The Physics of the Accelerating Universe Survey (PAUS; Castander et al., 2012; Eriksen et al., 2019) is planning to measure the cross-correlation of the Lyα forest with Lyα emission at redshifts $2.7 < z < 3.3$, measured with 40 narrowband filters on the William Herschel Telescope with the PAU Camera (Renard et al., 2021). While a detection in this setup is predicted to be unlikely, Renard et al. (2024) predicts a detection of Lyα LIM through a cross-correlation between broad-band optical images and Lyα forest data. The narrowband surveys J-PAS (Benitez et al., 2014) and miniJPAS (Bonoli et al., 2021) could also be used Lyα LIM in cross-correlation with the Lyα forest (Renard et al., 2024).

The upcoming Ultraviolet Transient Astronomy Satellite (ULTRASAT; Shvartzvald et al., 2024) maps UV emission, including Lyα emission at $0.9 < z < 1.4$, in broad-band images. It is possible to recover the redshift dependence and constrain the Lyα emission using cross-correlations with a spectroscopic galaxy survey (Libanore & Kovetz, 2024) in a similar fashion as Chiang et al. (2019).

Chapter 2

Physics of Line Intensity Mapping

2.1 Intensity

The intensity, often also called "surface brightness", is the energy E transferred through a unit area dA_p (physical, not comoving), per unit time dt, per unit solid angle $d\Omega$,

$$I = \frac{d^3E}{dtdA_p d\Omega}. \tag{2.1}$$

The specific intensity is the intensity per unit wavelength or frequency

$$I_\lambda = \frac{dI}{d\lambda} \qquad \text{and} \qquad I_\nu = \frac{dI}{d\nu}. \tag{2.2}$$

The observed wavelength $\lambda_{\rm obs}$ and frequency $\nu_{\rm obs}$ of an emission line with rest-frame wavelength λ_0 and frequency ν_0 are determined by the redshift z,

$$\lambda_{\rm obs} = \lambda_0(1+z) \qquad \text{and} \qquad \nu_{\rm obs} = \frac{\nu_0}{(1+z)}, \tag{2.3}$$

where the redshift depends on the scale factor $a = (1+z)^{-1}$, neglecting peculiar velocities and Lyα line peak shifts (see Section 2.4.1).

Consider a source with luminosity $L = dE/dt$ in a comoving volume dV_c subtending a solid angle $d\Omega$. The intensity measured from this source is

$$I_{\lambda_{\rm obs}} = \frac{d^2L}{dA_p dV_c}\frac{d^2V_c}{d\lambda_{\rm obs}d\Omega}. \tag{2.4}$$

The flux from this source is

$$F = \frac{dL}{dA_p} = \frac{L}{4\pi D_L^2(z)}, \tag{2.5}$$

where D_L is the luminosity distance in a spatially flat universe,

$$D_L(z) = (1+z)\int_0^z \frac{cdz'}{H(z')} = (1+z)^2 D_A(z). \tag{2.6}$$

We can write the volume as the product of the comoving line-of-sight (LOS) distance and a comoving area, $dA_c = dA_p(1+z)^2$,

$$dV_c = d\chi dA_c = d\chi dA_p(1+z)^2. \tag{2.7}$$

A physical area perpendicular to the LOS is related to a solid angle by the square of the angular diameter distance D_A,

$$dA_p = D_A^2(z)d\Omega = \frac{D_L^2(z)}{(1+z)^4}d\Omega. \tag{2.8}$$

The infinitesimal LOS distance can be written in terms of the redshift or observed wavelength difference,

$$d\chi = \frac{c}{H(z)}dz = \frac{c}{\lambda_0 H(z)}d\lambda_{\rm obs}. \tag{2.9}$$

Inserting Equations (2.5), (2.8), and (2.9) into Equation (2.4), we obtain

$$I_{\lambda_{\rm obs}} = \frac{dL}{4\pi D_L^2 dV_c}\frac{cD_L^2}{(1+z)^2 H(z)\lambda_0} = \frac{c\rho_L}{4\pi\lambda_0(1+z)^2H(z)}, \tag{2.10}$$

where we defined the comoving luminosity volume density $\rho_L = dL/dV_c$. Similarly, using $|dz| = (1+z)^2\nu_0^{-1}|d\nu_{\rm obs}|$, we find

$$I_{\nu_{\rm obs}} = \frac{c\rho_L}{4\pi\nu_0 H(z)}. \tag{2.11}$$

Using the Rayleigh-Jeans relation of a blackbody spectrum, $T_{\nu_{\rm obs}} = c^2 I_\nu/(2k_{\rm B}\nu_{\rm obs}^2)$, with the Boltzmann constant $k_{\rm B}$, we can define the brightness temperature,

$$T_{\nu_{\rm obs}} = \frac{c^3(1+z)^2\rho_L}{8\pi k_{\rm B}\nu_{\rm obs}^3 H(z)}, \tag{2.12}$$

which is commonly used in radio and sub-mm LIM.

2.2 The Lyman-α Line

As described in Section 1.2, the Lyα line is one of the brightest emission lines in the Universe due to the abundance of hydrogen, its high transition energy, and the efficient emission through recombinations and collisional excitations.

High-redshift probes using Lyα are not limited to detecting individual LAEs. The faint Lyα glow of the CGM of LAEs and other high-redshift galaxies can be used to study the properties of the gas and its role in galaxy formation and evolution (e.g., Leclercq et al., 2017; Byrohl & Nelson, 2023). The Gunn-Peterson Lyα absorption troughs in continuum emission from high-redshift quasars constrain the timing of the end of hydrogen reionization (Gunn & Peterson, 1965; Scheuer, 1965). At lower redshifts, Lyα absorption of background quasar continuum produces the so-called Lyα forest, which can constrain the matter distribution between the quasar and the observer (Rauch, 1998). And damped Lyα absorption (DLA) systems point to dense neutral hydrogen pockets within galaxies even after reionization (Wolfe et al., 2005).

In this section, I will outline the emission mechanisms and the basic effects of the RT. Much of the content of this section is based on Dijkstra (2019).

2.2.1 Lyman-α Emission Mechanisms

The hydrogen atom consists of one proton and one electron. In the quantum-mechanical picture, the energy state of the electron is characterized by the quantum numbers n, l, m, and m_s. The energy level is given by the principal quantum number $n \in \{1, 2, ...\}$, where $n = 1$ is the ground state. The energy at the level n is

$$E_n = -\frac{q^2 m_e}{2n^2\hbar^2} = -13.6\,\mathrm{eV} \times \frac{1}{n^2}, \tag{2.13}$$

where q is the charge of the electron, m_e is its mass, and $\hbar$ is the reduced Planck constant. The orbital quantum number $l \in \{0, 1, ..., n-1\}$ quantifies the angular momentum of the electron. The magnetic quantum number $m \in \{-l, -l+1, ..., l-1, l\}$ characterizes the orientation of the orbital if it is not spherically symmetric, i.e., if $l \neq 0$. The spin quantum number $m_s \in \{-\frac{1}{2}, \frac{1}{2}\}$ denotes the spin of the electron.

Transitions from an energy state $|n, l, m, m_s\rangle$ to the $|n', l', m', m_s'\rangle$ state are only allowed if $l' - l \in \{-1, 1\}$ and $m' - m \in \{-1, 0, 1\}$. Lyα is the transition between the $|n = 2, l = 1\rangle$ state (2p) to the $|n = 1, l = 0\rangle$ state (1s). The energy difference is $E_{\mathrm{Ly}\alpha} = -E_0((1/2)^2 - 1) \simeq 10.2\,\mathrm{eV}$, which corresponds to a photon frequency and wavelength of

$$\nu_{\mathrm{Ly}\alpha} = \frac{E_{\mathrm{Ly}\alpha}}{h} \simeq 2.47 \times 10^{15}\,\mathrm{Hz} \tag{2.14}$$

$$\text{and} \qquad \lambda_{\mathrm{Ly}\alpha} = \frac{c}{\nu_{\mathrm{Ly}\alpha}} \simeq 1215.67\,\text{Å}, \tag{2.15}$$

where $h = 2\pi\hbar$ is the Planck constant. In order for the Lyα transition to occur, the electron has to get to the excited $|n = 2, l = 1\rangle$ state. The two most relevant ways for a hydrogen atom to get into an excited energy state are recombination and collisional excitation.

Recombination

Following the recombination of a proton and an electron to a hydrogen atom, the newly bound electron can be in any excited state $|n, l\rangle$ with the temperature-dependent probability $\mathcal{P}_{|n,l\rangle}(T)$. The probability of the electron to fall from the state $|n, l\rangle$ to the state $|n', l'\rangle$ is given by the branching ratio

$$\mathcal{P}\left(|n, l\rangle \to |n', l'\rangle\right) = \frac{A_{nl,n'l'}}{\sum_{n''=1}^{n-1} \sum_{l''=0}^{n''-1} A_{nl,n''l''}}, \tag{2.16}$$

where the Einstein coefficient $A_{nl,n'l'}$ quantifies the probability of a spontaneous transition from the $|n, l\rangle$ state to the $|n', l'\rangle$ state. The probability of the $|2, 1\rangle$ state to lead to a Lyα transition is $\mathcal{P}(|2, 1\rangle \rightsquigarrow \mathrm{Ly}\alpha) = 1$, while the $|2, 0\rangle$ state can only transition to the ground state through two-photon emission (Breit & Teller, 1940), so that $\mathcal{P}(|2, 0\rangle \rightsquigarrow \mathrm{Ly}\alpha) = 0$.

Knowing the Einstein coefficients, we can calculate the probability of any state $|n, l\rangle$ with $n > 2$ to lead to a Lyα transition,

$$\mathcal{P}(|n,l\rangle \rightsquigarrow \mathrm{Ly}\alpha) = \mathcal{P}(|n,l\rangle \to |2,1\rangle) + \sum_{n'=3}^{n-1} \sum_{l'=0}^{n'-1} \mathcal{P}(|n,l\rangle \to |n',l'\rangle)\mathcal{P}(|n',l'\rangle \to |2,1\rangle). \tag{2.17}$$

The probability of a recombination event in a gas with temperature T to result in a Lyα photon emission is then given by the sum of the probability of a recombination event ending up in state $|n, l\rangle$ multiplied by the probability of this state leading to a Lyα photon emission,

$$\mathcal{P}^{\mathrm{rec}}_{\mathrm{Ly}\alpha}(T) = \sum_{n=n_{\min}}^{\infty} \sum_{l=0}^{n-1} \mathcal{P}_{|n,l\rangle}(T)\mathcal{P}(|n,l\rangle \rightsquigarrow \mathrm{Ly}\alpha). \tag{2.18}$$

The value of $n_{\min}$ is determined by the optical depth of the medium. One typically distinguishes between two cases:

- "case-A recombination": The gas is optically thin at all Lyman series frequencies, i.e., all transitions leading to the ground state. Then we can set $n_{\min} = 1$.
- "case-B recombination": The gas is opaque at all Lyman series frequencies. Neutral hydrogen atoms immediately reabsorb all photons produced by a transition to the ground state or by a direct recombination into the ground state. This is modeled by setting $n_{\min} = 2$ and setting all Einstein coefficients of Lyman transitions to zero for the calculation of $\mathcal{P}_{|n,l\rangle}(T)$.

The Lyα luminosity density from recombination is

$$\rho^{\mathrm{rec}}_L(T) = \mathcal{P}^{\mathrm{rec}}_{\mathrm{Ly}\alpha}(T) n_e n_p \alpha(T) E_{\mathrm{Ly}\alpha}, \tag{2.19}$$

where n_e and n_p are the electron and proton number densities and $\alpha(T)$ is the recombination coefficient (see Draine, 2011). The Lyα emission probability for a recombination event $\mathcal{P}_{\mathrm{Ly}\alpha}(T)$ increases with decreasing gas temperature. At a temperature of $10^4\,\mathrm{K}$, it is $\mathcal{P}^{\mathrm{rec}}_{\mathrm{Ly}\alpha}(T = 10^4\,\mathrm{K}) = 0.68$. The remaining 32% of recombination events lead to radiative cascades that lead to the $|n = 2, l = 0\rangle$ (2s) state and therefore do not result in a Lyα transition.

Collisional Excitation

A collision of an electron and a neutral hydrogen atom can excite the electron. The probability of an excitation to the 2p state depends on the relative velocity between the hydrogen atom and the electron. The probability is also proportional to densities of neutral hydrogen n_{HI} and electrons n_{e} and therefore depends on the ionization state of the gas. The Lyα luminosity density of photons produced by collisional excitation is

$$\rho^{\mathrm{coll}}_L(T) = n_{\mathrm{e}} n_{\mathrm{HI}} \gamma_{1\mathrm{s},2\mathrm{p}}(T) E_{\mathrm{Ly}\alpha}, \tag{2.20}$$

where the rate coefficient $\gamma_{1s2p}(T)$ encodes the velocity dependence, assuming a thermal velocity distribution. Because at the energies of interest of the free electron, it spends a relatively long time around the target atom, and distorts the wavefunction of the bound electron (Dijkstra, 2019). This makes the calculation of γ_{1s2p} complicated, so that collisional excitation rates are known with an uncertainty of $10 - 20\%$.

Because the free electron loses energy, collisional excitation cools the gas. Assuming that the ionization state depends only on the temperature, one can calculate the cooling rate as a function of temperature and overall gas density. Under these assumptions, Lyα cooling is most efficient at $10^{4.2}\,\mathrm{K}$. At higher temperatures, collisional ionization removes neutral hydrogen that is necessary for this Lyα emission channel.

2.2.2 Lyman-α Radiative Transfer

The absorption cross-section of neutral hydrogen is very large around the Lyα line. Because of the large cross-section and the high neutral hydrogen column densities within the ISM, a Lyα photon is usually absorbed and re-emitted several times before escaping it. Because of the long path of a Lyα photon in the ISM, it is also prone to destruction by dust grains. The scattering and dust extinction significantly alter the observed Lyα line shape and luminosity compared to the intrinsic signal. Scattering also plays an important role on larger scales in the CGM and IGM (e.g., Byrohl & Nelson, 2023). Thus, one has to account for the RT when modeling and interpreting Lyα observations.

While the absorption cross-section of neutral hydrogen close to resonance with the Lyα line can be modeled analytically, the full quantum-mechanical calculation further away from resonance is complicated (e.g. Bach & Lee, 2015; Mortlock, 2016). However, one can model the cross-section of a hydrogen atom at rest with a Lorentzian function for simplicity.

In a gas with nonzero velocity dispersion, the photon frequency is Doppler-shifted differently for each atom in the gas, shifting its cross-section from the photon's point of view. To calculate the effective cross-section of the atoms in that gas, one has to convolve the rest-frame cross-section with the velocity distribution of the gas. For a Lorentzian rest-frame cross-section and a Maxwell-Boltzmann velocity distribution with temperature T and thermal velocity dispersion $v_{\mathrm{th}} = \sqrt{2k_{\mathrm{B}}T/m_{\mathrm{H}}}$, the Ly$\alpha$ cross-section becomes (Dijkstra, 2019):

$$\sigma_{\mathrm{Ly}\alpha}(x) = 5.9 \times 10^{-14} \left(\frac{T}{10^4\,\mathrm{K}}\right)^{-\frac{1}{2}} \phi(x)\,\mathrm{cm}^2. \tag{2.21}$$

Here, x is the dimensionless frequency

$$x = \frac{\nu - \nu_{\mathrm{Ly}\alpha}}{\Delta\nu_{\mathrm{Ly}\alpha}}, \text{ where} \qquad \Delta\nu_{\mathrm{Ly}\alpha} = \frac{\nu_{\mathrm{Ly}\alpha} v_{\mathrm{th}}}{c}, \tag{2.22}$$

and $\phi(x)$ is the Voigt function,

$$\phi(x) = \frac{a_\nu}{\pi} \int_{-\infty}^{\infty} dy \frac{e^{-y^2}}{(y-x)^2 + a_\nu^2} \sim \begin{cases} e^{-x^2} & \text{core} \\ \frac{a_\nu}{\sqrt{\pi} x^2} & \text{wing,} \end{cases} \tag{2.23}$$

with the Voigt parameter $a_\nu = 4.7 \times 10^{-4} (T/10^4\,\mathrm{K})^{-1/2}$. The cross-section is steep in the core at resonance with the Lyα line. However, it is also very high in the wings, making non-resonant scattering of Lyα very important (e.g., Chang et al., 2023).

The optical depth τ of a medium describes the ratio of the incoming flux, F_i, to the transmitted flux through the medium, F_t,

$$\tau = \log\left\{\frac{F_i}{F_t}\right\} = N_{\mathrm{HI}}\sigma_{\mathrm{Ly}\alpha}. \tag{2.24}$$

At typical astrophysical neutral hydrogen densities of $N_{\mathrm{HI}} \simeq 10^{17} - 10^{21}\,\mathrm{cm}^{-2}$ (physical, not comoving), the optical depth at line center reaches $\tau \simeq 10^4 - 10^8$. This high optical depth implies that a Lyα photon experiences many scattering events before escaping the gas.

The RT of Lyα photons can be described by the change of the specific intensity as a function of propagation direction $\mathbf{n}$ and distance traveled s:

$$\left(\frac{1}{c}\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) I_\nu = \underbrace{-\alpha_\nu I_\nu(\mathbf{n})}_{\text{absorption}} + \underbrace{j_\nu(\mathbf{n})}_{\text{emission}} + \underbrace{\int d\nu' \int d^3\mathbf{n}' \alpha_{\nu'} I_{\nu'}(\mathbf{n}') R(\nu, \hat{\mathbf{n}}|\nu', \hat{\mathbf{n}}')}_{\text{scattering}}. \tag{2.25}$$

Let us briefly outline the effects of each term.

Absorption

The absorption term describes the loss of photons from the line of sight, reducing the specific intensity. As it is proportional to the specific intensity, the solution for this term in the absence of emission and scattering is

$$I_\nu = I_{\nu,0} \exp\left\{-\int_0^s ds' \alpha_\nu(s')\right\} = I_{\nu,0} e^{-\tau}, \tag{2.26}$$

which is an exponential decay of the intensity and defines the optical depth τ. The attenuation coefficient is the sum of the absorption by neutral hydrogen and destruction,

$$\alpha_\nu = \alpha_\nu^{\mathrm{HI}} + \alpha_\nu^{\mathrm{destruction}}. \tag{2.27}$$

The attenuation coefficient for absorption by neutral hydrogen is simply $\alpha_\nu^{\mathrm{HI}} = n_{\mathrm{HI}}\sigma_{\mathrm{Ly}\alpha}$, neglecting a correction for stimulated emission. Lyα photons can be absorbed by dust grains, which re-emit lower-energy photons, effectively destroying the Lyα photons. Less importantly, molecular hydrogen can also absorb Lyα photons and re-emit them at lower energy. At high energies, an electron in the $|n = 2, l = 1\rangle$ (2p) state can interact with a free electron or proton and transition to the $|n = 2, l = 0\rangle$ (2s) state. From there, it has to emit two photons, not at the Lyα wavelength, to reach the ground state. Finally, Lyα photons can ionize or further excite already excited hydrogen. However, these destruction mechanisms are subdominant compared to absorption by dust (Laursen, 2010).

Emission

The emission is given by the sum of the recombination and collisional excitation luminosity densities given in equations (2.19) and (2.20)

$$j_\nu = \frac{\phi(\nu)}{4\pi^{\frac{3}{2}}\Delta\nu_{\mathrm{Ly}\alpha}} \left(\rho_L^{\mathrm{rec}} + \rho_L^{\mathrm{coll}}\right). \tag{2.28}$$

Here, $\phi(\nu)$ is the Lyα line shape given by the Voigt profile of equation (2.23). Continuum photons that are redshifted to the Lyα wavelength can also contribute to the Lyα source term.

Scattering

After absorbing a Lyα photon, the hydrogen atom emits it again after a short time. This can be modeled as scattering. While in the rest frame of the hydrogen atom, the frequency of the photon is the same at absorption and re-emission, the velocity of the hydrogen atom causes a frequency shift in the observer's frame of reference. Therefore, a scattering event changes the photon's direction and frequency. The redistribution function $R(\nu, \hat{\mathbf{n}}|\nu', \hat{\mathbf{n}}')$ in equation (2.25) is proportional to the probability distribution function of a photon with initial frequency ν' and direction $\hat{\mathbf{n}}'$ to scatter into the direction $\hat{\mathbf{n}}$ with frequency ν.

Repeated scatterings due to the typically high optical depth in the ISM lead to a substantial, non-trivial change in the spatial and spectral distribution of Lyα photons. While the emission and absorption terms have analytic solutions, one can only find analytical solutions for the RT equation including the scattering term in simple scenarios (Dijkstra, 2019). Instead, Monte-Carlo RT algorithms are used to model the RT in more realistic settings (e.g., Byrohl & Nelson, 2025).

2.3 Astrophysical Sources of Lyman-α Emission

Lyα photons are emitted through recombination of hydrogen in the ISM that has been ionized by young, massive stars or an AGN. Recombination and collisional excitation of hydrogen in the CGM and IGM can also produce Lyα photons. In this section, I will describe the astrophysical sources of Lyα emission and estimate typical Lyα intensities, often following Dijkstra (2019).

2.3.1 Lyman-α Emitting Galaxies

LAEs are characterized by their strong Lyα emission. The equivalent width $EW_{\mathrm{Ly}\alpha}$ quantifies the strength of the Lyα emission line compared to the continuum in the rest frame of the galaxy. It is the wavelength width one would need to integrate over the continuum flux density to obtain the Lyα line flux. A galaxy is classified as an LAE if its $EW_{\mathrm{Ly}\alpha} > 20\,\text{Å}$ (Ouchi et al., 2020). The Lyα emission of LAEs is either dominated by an AGN or star formation, which ionize the ISM and produce Lyα emission through

recombination. LAEs without an AGN are typically compact, metal-poor, star-forming galaxies with SFR of $\sim 1 - 10\,M_\odot\,\mathrm{yr}^{-1}$ and reside in dark matter halos with a mass around $10^{10} - 10^{11}\,M_\odot$ (Ouchi et al., 2020).

Young, massive stars emit highly energetic photons that ionize the ISM. As a result, recombining hydrogen emits Lyα photons and other recombination lines, such as Hα. In an ionization equilibrium, the rate of recombinations, $\dot{N}_{\mathrm{rec}}$, equals the ionization rate of hydrogen; $\Gamma_{\mathrm{ion}} = \dot{N}_{\mathrm{rec}}$. Because some ionizing photons escape the ISM, only a fraction $(1 - f_{\mathrm{esc}}^{\mathrm{ion}})$ ionizes the gas. The production rate of Lyα photons through recombination in the ISM is therefore

$$\dot{N}_{\mathrm{Ly}\alpha}^{\mathrm{rec}} = \mathcal{P}_{\mathrm{Ly}\alpha}^{\mathrm{rec}}(T)(1 - f_{\mathrm{esc}}^{\mathrm{ion}})\Gamma_{\mathrm{ion}} \simeq 0.68(1 - f_{\mathrm{esc}}^{\mathrm{ion}})\Gamma_{\mathrm{ion}}, \tag{2.29}$$

where we used $\mathcal{P}_{\mathrm{Ly}\alpha}^{\mathrm{rec}}(T \simeq 10^4\,\mathrm{K}) \simeq 0.68$ for case-B recombination. The ionization rate is governed by the abundance of hot, massive stars, and therefore depends on the star-formation rate (SFR), $\dot{M}_\star$. Assuming the initial stellar mass function of Salpeter (1955) at a solar metallicity and the mass limits $M_{\mathrm{min}} = 0.1M_\odot$ and $M_{\mathrm{max}} = 100M_\odot$, the ionization rate relates to the SFR as

$$\Gamma_{\mathrm{ion}} = 9.3 \times 10^{52}\,\mathrm{s}^{-1} \times \frac{\dot{M}_\star}{M_\odot \mathrm{yr}^{-1}}. \tag{2.30}$$

With $f_{\mathrm{esc}}^{\mathrm{ion}} = 0$ and using $L_{\mathrm{Ly}\alpha} = E_{\mathrm{Ly}\alpha}\dot{N}_{\mathrm{Ly}\alpha}^{\mathrm{rec}}$, we obtain the relationship between Lyα luminosity and SFR,

$$L_{\mathrm{Ly}\alpha} \simeq 10^{42}\,\mathrm{erg\,s}^{-1} \times \frac{\dot{M}_\star}{M_\odot \mathrm{yr}^{-1}}. \tag{2.31}$$

This is consistent with the relation between Hα emission and SFR of Kennicutt (1998), assuming a constant ratio of the Lyα and Hα emission rates of 8-10 (Hummer & Storey, 1987). However, the conversion can deviate by factors of $2 - 3$ with other assumptions (Raiter et al., 2010).

At a fixed mass, stars with lower metallicity are hotter (Tumlinson & Shull, 2000). The increased ionizing photon production also increases the Lyα photon production. At very low metallicities, $Z < 0.03Z_\odot$, the case-B approximation no longer holds and the Lyα production increases. This is due to a combination of two effects. Firstly, the hotter gas leads to more collisional excitations of hydrogen atoms. Secondly, the harder ionizing photons emitted by the stars can give enough energy to the electron after ionizing a hydrogen atom that this free electron can ionize or collisionally excite other hydrogen atoms (Shull & van Steenberg, 1985).

The luminosity of an LAE is determined by the escape fraction of Lyα photons from the ISM, $f_{\mathrm{esc}}^{\mathrm{Ly}\alpha}$, and its intrinsic luminosity, estimated by Equation (2.31). Because Lyα scatters many times before escaping the galaxy, it is prone to dust extinction, which lowers the escape fraction. By post-processing a hydrodynamic simulation of a disk galaxy with Lyα RT, Behrens & Braun (2014) show that the escape fraction depends on the observer's viewing angle with respect to the disk. The largest $EW_{\mathrm{Ly}\alpha}$ is observed perpendicular to the disk, while most Lyα is absorbed along the disk axis.

A fundamental observational quantity to describe galaxy properties is the luminosity function (LF). It is defined as the comoving number density of emitters per unit luminosity

dn/dL and therefore describes the number density of LAEs and the luminosity probability density function (PDF). A Schechter function is often fitted to the LF (Schechter, 1976),

$$\frac{dn}{dL_{\mathrm{Ly}\alpha}}(L_{\mathrm{Ly}\alpha}) = \frac{\phi^*}{L^*_{\mathrm{Ly}\alpha}} \left(\frac{L_{\mathrm{Ly}\alpha}}{L^*_{\mathrm{Ly}\alpha}}\right)^{-\alpha} \exp\left(-\frac{L_{\mathrm{Ly}\alpha}}{L^*_{\mathrm{Ly}\alpha}}\right), \tag{2.32}$$

with the characteristic number density ϕ^*, the characteristic luminosity $L^*_{\mathrm{Ly}\alpha}$, and the slope α. The mean luminosity density of the universe originating from LAEs is the first moment of the LF,

$$\rho_L^{\mathrm{LAE}} = \int_{L_{\mathrm{min}}}^{L_{\mathrm{max}}} \frac{dn}{dL_{\mathrm{Ly}\alpha}} L_{\mathrm{Ly}\alpha} dL_{\mathrm{Ly}\alpha}. \tag{2.33}$$

The LF increases monotonically from redshift $z = 0$ to $z \sim 3$ because of an increasing Lyα escape fraction from galaxies and an increasing ionizing photon abundance. It evolves little from $z \sim 3$ to $z \sim 6$, and drops rapidly at higher redshifts because the hydrogen reionization is incomplete at higher redshift, significantly increasing the optical depth in neutral hydrogen regions (Hayes et al., 2011; Ouchi et al., 2020; Umeda et al., 2025). At high luminosities of $L_{\mathrm{Ly}\alpha} \gtrsim 10^{43}\,\mathrm{erg\,s^{-1}}$, AGNs dominate the LF (Spinoso et al., 2020). While it is theoretically expected that the Lyα LF decreases at faint luminosities (Liu et al., 2016), the observations of LAEs have not been deep enough to detect the turnover of the LF. This turnover could be constrained using LIM because it measures the integrated luminosity of all sources, including bright and faint LAEs and diffuse Lyα emission ("no photons are left behind").

Figure 2.1 shows the mean intensity of the Universe originating from LAEs, calculated using equations (2.33) and (2.10), as a function of minimum luminosity L_{min}. The mean intensity flattens off at $L_{\mathrm{min}} \lesssim 10^{40}\,\mathrm{erg\,s^{-1}}$ to a value of $\langle I_\lambda^{\mathrm{LAE}} \rangle \simeq 10^{-22}\,\mathrm{erg\,s^{-1}\,cm^{-2}\,arcsec^{-2}\,\AA^{-1}}$ for all considered LFs.

Because of the high optical depth at the line center, Lyα photons scatter multiple times before the frequency shifts into the wing and they can escape the galaxy. Because of the complexity, Monte-Carlo methods are usually employed to model the RT (e.g., Byrohl & Nelson, 2025). Idealized models of galaxies, such as the expanding shell model, are used to interpret observed LAE line profiles (e.g. Verhamme et al., 2006a).

The expanding shell model consists of a central source of Lyα photons surrounded by a spherical gas shell with a given radial velocity, velocity dispersion, neutral hydrogen density, and temperature. If the shell is static, the RT leads to a double-peak profile symmetric around the line center. If the shell is expanding outwards, photons emitted in the center are redshifted with respect to the hydrogen atoms in the shell. Therefore, photons that were emitted bluewards of the Lyα line are redshifted into the optically thick line center in the hydrogen atoms' rest frame and cannot escape. On the other hand, photons that were emitted redwards of the Lyα line are redshifted far enough that they do not scatter in the hydrogen shell and escape. Because scattering events change the photon frequency, previously blue photons can also be redshifted out of the optically thick line center due to scattering and escape. As a result, the blue peak of the escaping profile is suppressed, while photons in the red peak escape. In an inflowing shell model, the red peak is suppressed instead. Most observed Lyα spectra have a dominant red peak (Ouchi et al., 2020).

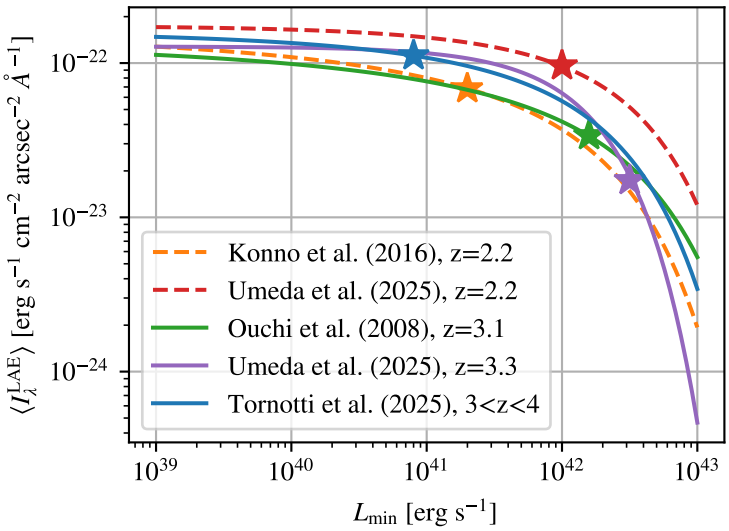

Figure 2.1: Mean specific Lyα intensity of the Universe from LAEs, obtained by integrating over the LFs from $L_{\rm min}$ to $L_{\rm max} = 10^{44}\,{\rm erg\,s^{-1}}$, as a function of $L_{\rm min}$. The LFs are the best-fit Schechter functions from Konno et al. (2016), Ouchi et al. (2008) (fit by Konno et al., 2016), the Schechter components of Umeda et al. (2025) at $z = 2.2$ and $z = 3.3$, and Tornotti et al. (2025a) at $3 < z < 4$. We used $z = 3.5$ for the Tornotti et al. (2025a) LF. The faintest luminosity bin of each LF measurement is shown as a star.

Expanding shell models can fit the majority of observed Lyα spectral profiles (Gronke, 2017). However, galaxies do not resemble expanding shells. Modeling slightly more complicated, asymmetric geometries, adding clumpiness to the gas, and allowing escape channels with low neutral hydrogen density illuminate the high variability of Lyα line profiles and the difficulty of interpreting them (Behrens et al., 2014; Chang et al., 2023; Almada Monter & Gronke, 2024).

2.3.2 Diffuse Lyman-α emission

Not only LAEs shine brightly in Lyα, but the surrounding CGM and IGM commonly exhibit a faint Lyα glow, called Lyα halo (e.g.; Steidel et al., 2011; Wisotzki et al., 2016). The diffuse Lyα emission around LAEs extends over such a large area that, statistically, every LOS passes through a Lyα halo at $z \sim 3 - 6$ (Wisotzki et al., 2018). Bright, extended Lyα emission is also found in galaxy-overdense regions independently of host LAEs or quasars (Kikuta et al., 2019). Depending on their luminosity, they are classified as Lyα blobs (e.g., Shibuya et al., 2018) or extremely luminous Lyα nebulae (ELAN; e.g., Cantalupo et al., 2014; Cai et al., 2018). Cosmic web filaments have also been observed in Lyα emission in the vicinity of quasars illuminating the surrounding cosmic web (Arrigoni Battaia et al., 2019), around protoclusters (Umehata et al., 2019), and around LAE overdensities (Bacon et al., 2021; Martin et al., 2023; Tornotti et al., 2025b,c). Possible emission mechanisms of diffuse Lyα emission include scattering of photons produced in the ISM of the galaxy into the CGM and IGM, recombination following photoionization

from an ionizing background ("fluorescence"), either in low-density gas or in the outer shell of a neutral dense cloud, and collisional excitation within such a dense cloud.

Recombination in the Low-Density Ionized IGM

Recombination in the low-density ionized IGM can be described by the luminosity density in equation (2.19). We can write the comoving proton density in terms of the mean proton density $\bar{n}_p$ and the proton density contrast δ_p,

$$n_p = \bar{n}_p(1+\delta_p) = \frac{0.75\Omega_{b,0}\rho_{\rm crit,0}}{m_p}(1+\delta_p), \tag{2.34}$$

where $\Omega_{b,0} = \rho_{b,0}/\rho_{\rm crit,0}$ is the ratio of the baryon density and the critical density. Assuming $n_e = n_p$ and substituting $H(z) = H_0\sqrt{\Omega_{\rm m,0}\,(1+z)^3}$ for the matter-dominated universe at $z \gtrsim 1$ in Equation (2.10), we obtain

$$I_\lambda^{\rm rec,IGM} = (1+\delta_p)^2\left(\frac{0.75\Omega_{b,0}\rho_{\rm crit,0}}{m_p}\right)^2 \frac{c\mathcal{P}_{{\rm Ly}\alpha}^{\rm rec}(T)\alpha_B^{\rm rec,com}(T)E_{{\rm Ly}\alpha}}{4\pi H_0\sqrt{\Omega_{\rm m,0}(1+z)^3}(1+z)^2\lambda_{{\rm Ly}\alpha}}. \tag{2.35}$$

Inserting the case-B recombination coefficient (comoving),

$$\alpha_B^{\rm rec,com}(T) = 2.6\times10^{-13}(T/10^4\,{\rm K})^{-0.7}(1+z)^3\,{\rm cm^3\,s^{-1}}, \tag{2.36}$$

and $\mathcal{P}_{{\rm Ly}\alpha}^{\rm rec}(T=10^4\,{\rm K}) \simeq 0.68$, the specific intensity from recombination in the low-density IGM becomes

$$I_\lambda^{\rm rec,IGM} \simeq 2\times10^{-24}\,(1+\delta_p)^2\left(\frac{1+z}{4}\right)^{-\frac{1}{2}}\left(\frac{T}{10^4\,{\rm K}}\right)^{-0.7}\frac{\rm erg}{\rm s\,cm^2\,arcsec^2\,\text{Å}}. \tag{2.37}$$

Recombination in Ionized, Dense Gas

If the emission region is smaller than the LOS resolution of the observing spectrograph, it is helpful to calculate the integrated intensity of this region. This is given by

$$I = \frac{\rho_L^p l}{4\pi(1+z)^4}, \tag{2.38}$$

where l is the physical (not comoving) size of the emission region along the LOS and ρ_L^p is the physical luminosity density (see Section 2.1). We can express the physical proton density as $n_p^p = N_p^p/l$, where N_p^p is the physical hydrogen column density. Then the intensity becomes

$$\begin{aligned} I^{\rm rec,dense} &= \frac{n_e^p N_p^p \mathcal{P}_{{\rm Ly}\alpha}^{\rm rec}(T)\alpha_B^{\rm rec}(T)E_{{\rm Ly}\alpha}}{4\pi(1+z)^4} \\ &\simeq 2\times10^{-21}\left(\frac{1+z}{4}\right)^{-4}\left(\frac{n_e^p}{10^{-3}\,{\rm cm^{-3}}}\right)\left(\frac{N_p^p}{10^{20}\,{\rm cm^{-2}}}\right)\left(\frac{T}{10^4\,{\rm K}}\right)^{-0.7}\frac{\rm erg}{\rm s\,cm^2\,arcsec^2}. \end{aligned} \tag{2.39}$$

The specific intensity measured by a spectrograph with LOS resolution $\Delta\lambda_{\rm obs}$ is $I_\lambda = I/\Delta\lambda_{\rm obs}$.

Recombination in the Skin of Self-Shielding, Dense Clouds

If a gas cloud is dense enough to shield itself against ionizing radiation, only its surface is ionized and can emit Lyα photons through recombination. The mean free path of a Lyα photon corresponds to the depth of the ionized skin,

$$\lambda_{\rm mfp} \simeq \frac{1}{n_{\rm HI}\sigma_{\rm ion}} = \frac{1}{n_{\rm H}x_{\rm HI}\sigma_{\rm ion}} \simeq \frac{(1-x_{\rm HI})\Gamma_{\rm ion}}{n_e^2\alpha_{\rm B}^{\rm rec}\sigma_{\rm ion}}, \tag{2.40}$$

where $\sigma_{\rm ion}$ is the ionization cross-section of neutral hydrogen, $n_{\rm HI}$ is the neutral hydrogen density, $n_{\rm H} = n_{\rm HI} + n_p$ is the density of neutral hydrogen and protons, $x_{\rm HI} = n_{\rm HI}/n_{\rm H}$ is the neutral hydrogen fraction, and $\Gamma_{\rm ion}$ is the ionization rate. In the last step, we assumed that most free electrons originate from hydrogen atoms, i.e., $n_e \approx n_p = (1-x_{\rm HI})n_{\rm H}$, and ionization equilibrium $x_{\rm HI}\Gamma_{\rm ion} = n_e\alpha_{\rm B}^{\rm rec}$. Dijkstra (2019) approximates the ionization cross-section and recombination coefficient, finding an intensity of

$$I^{\rm rec,shielded} \simeq 1.3\times10^{-20}(1-x_{\rm HI})\left(\frac{1+z}{4}\right)^{-4}\left(\frac{\Gamma_{\rm ion}}{10^{-12}\,{\rm s}^{-1}}\right)^{-4}\frac{\rm erg}{\rm s\,cm^2\,arcsec^2}. \tag{2.41}$$

Cooling by Dense, Neutral Gas

The neutral core of a self-shielding cloud will emit Lyα photons through collisional excitation, thereby cooling the gas (see Section 2.2.1). Calculating the intensity of this emission requires many assumptions and is highly uncertain (see Dijkstra, 2019; Ouchi et al., 2020; Byrohl, 2021). An attempt to quantify cooling emission rates is beyond the scope of this thesis. Nonetheless, cooling emission from collapsing halos can be very bright (e.g., Haiman et al., 2000; Dijkstra & Loeb, 2009).

Origin of Diffuse Lyman-α Emission

While diffuse Lyα halos around star-forming galaxies have been detected ubiquitously, their dominant emission mechanisms are still debated (see Ouchi et al., 2020, for a review). Lyα halos could be illuminated by scattered photons from the ISM of the galaxy, fluorescence, or collisional excitation. Given the modeling uncertainties, every emission channel described above can be bright enough near ionizing sources. A population of faint Lyα emitters below the individual detection limit can also appear like diffuse emission (Mas-Ribas et al., 2017).

Byrohl et al. (2021) and Byrohl & Nelson (2023) post-process a cosmological hydrodynamic simulation with Lyα RT to study the emission mechanisms of Lyα halos. Figure 2.2 shows the median-stacked Lyα surface brightness profile of simulated galaxies with solar masses $M_\star \in [10^{9.5}M_\odot, 10^{10}M_\odot]$ of Byrohl et al. (2021), separated into the contributions from different emission origins before scattering and convolved with the VIRUS PSF. Near the stacked galaxies, Lyα halo emission is dominated by scattered Lyα photons from within these galaxies. Most of these are sourced by star formation in the ISM, closely followed by collisional excitation. Photons at larger distances from the galaxy, at $r \gtrsim 20$ physical kpc ($\simeq 60$ kpc after smoothing), predominantly originate from the CGM of other dark matter halos and scatter in the IGM. This causes the surface brightness

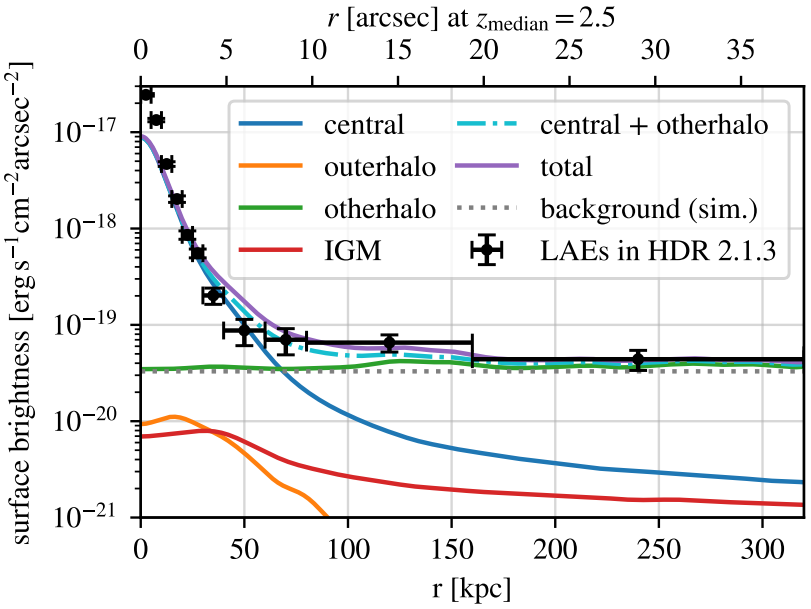

Figure 2.2: Median-stacked Lyα surface brightness profile of LAEs in HETDEX (black; Lujan Niemeyer et al., 2022a) as a function of physical projected distance, compared to the median surface brightness profiles of simulated galaxies with stellar masses $M_\star \in [10^{9.5} M_\odot, 10^{10} M_\odot]$ from Byrohl et al. (2021), convolved with the VIRUS PSF. The dashed gray line indicates the mean Lyα surface brightness of the simulation (background (sim.)), which was added to the measured Lyα surface brightness profile. Different colors show different emission origins of the simulated Lyα photons: the central galaxy in the target dark matter halo (central, blue), the outer parts of this dark matter halo (outerhalo, orange), another dark matter halo than the target halo (otherhalo, green), the IGM (red), and the sum of these (total, purple). This figure is taken from Lujan Niemeyer et al. (2022a).

profiles of Lyα halos to flatten at $r \gtrsim 15$ physical kpc ($\simeq 60$ kpc after smoothing). The predicted Lyα halo profiles agree with measured individual Lyα halo profiles observed by the MUSE spectrograph (Leclercq et al., 2017) and with stacked Lyα halo profiles of LAEs observed with HETDEX (Lujan Niemeyer et al., 2022a). While the former only probe the inner part of the Lyα halos, the latter detect a signal out to 160 kpc and confirm the flattening of the profile (see Figure 2.2).

As we will show in Chapter 4, star-forming galaxies without observable Lyα emission are also surrounded by faint Lyα halos. The outer regions of the Lyα halos is equally bright for these Lyα-faint galaxies as for LAEs, suggesting that Lyα halo properties are independent of the Lyα luminosity of the host galaxy. This supports the picture that these photons do not originate from the stacked galaxies.

Byrohl & Nelson (2023) explore the contribution of different emission mechanisms and origins, including scattering, to the total Lyα intensity and distribution. They find that a quarter of the Lyα photons originate from central galaxies. 98% of all Lyα photons originate from within halos, only 2% originate in the low-density IGM. Collisional excitation produces 48% of Lyα photons, making it the most important emission mechanism. Star

formation produces 29% of Lyα photons, and recombination outside of the ISM produces 23%. Scattering redistributes the Lyα photons significantly, so that the luminosity density of the IGM is boosted by an order of magnitude. After scattering, LAEs contribute 30% of the total Lyα photons in this simulation, 42% of photons last scatter in the CGM, and 28% of Lyα photons reach the observer from the IGM (see Figure 1.2).

2.4 Radiative Transfer Distortions of Galaxy Clustering Statistics

The complex RT in the ISM changes the Lyα line shape and the Lyα peak position of the LAEs. Inferring the redshift of LAEs from the Lyα line leads to a wrong redshift estimate and a Fingers-of-God-like damping of the small-scale power spectrum (Byrohl et al., 2019). Scattering of Lyα photons out of the LOS in the IGM causes the LAEs to appear fainter. Because the scattering probability depends on the gas density and velocity, it might be correlated with the large-scale structure, which would cause an anisotropic distortion of the power spectrum that is similar to the Kaiser effect (Kaiser, 1987), but inverted (Zheng et al., 2011). While these effects have not been observed unequivocally, they could significantly alter the LAE and Lyα LIM clustering statistics. This section provides an overview over these two effects.

2.4.1 Lyman-α Peak Shift

Because of the effective peak shift from RT as described in Section 2.3.1, the redshift of the Lyα line peak does not align with the cosmological redshift. This can be described as an additional peculiar velocity of the LAEs, causing a damping of the power spectrum on small scales, similar to the Fingers-of-God effect (FoG; Byrohl et al., 2019). Consider a galaxy at the location $\mathbf{x} = r(\sin\Theta\cos\phi, \sin\Theta\sin\phi, \cos\Theta)$, where the origin of the coordinate system is Earth, with the velocity $\mathbf{v}$. The observed ("redshift-space") position $\mathbf{s}$ of this galaxy is at

$$\mathbf{s} = \mathbf{x} + \frac{\mathbf{v}\cdot\hat{\mathbf{x}}}{aH}\hat{\mathbf{x}}, \tag{2.42}$$

where a is the scale factor, H is the Hubble expansion rate at the galaxy position, and $\hat{\mathbf{x}} = \mathbf{x}/r$ is the unit vector in the radial direction. The velocity only affects the observed position in the LOS direction, introducing an anisotropic distortion to observed galaxy clustering statistics, called RSD (see Hamilton, 1998; Saito, 2016, for reviews).

The LOS velocity $v_\parallel$ does not have to be a physical velocity of the galaxy. It can also be an effective velocity shift from Lyα RT. If the peak of the escaped Lyα photons from a galaxy is at a rest-frame wavelength $\lambda_{\rm peak}$, the apparent velocity of the galaxy due to RT is

$$v_\parallel^{\rm app} = c\frac{\lambda_{\rm peak} - \lambda_{\rm Ly\alpha}}{\lambda_{\rm Ly\alpha}}, \tag{2.43}$$

where c is the speed of light and $\lambda_{\rm Ly\alpha}$ is the Lyα rest-frame wavelength.

To linear order in the density contrast and velocity, the galaxy density contrast in redshift space becomes

$$\delta_{\rm g}^s(\mathbf{s}) \simeq \delta_{\rm g}(\mathbf{x}) - \partial_\parallel \frac{v_\parallel}{aH} = \delta_{\rm g}(\mathbf{x}) - f\partial_\parallel u_\parallel, \tag{2.44}$$

where $\partial_\parallel$ is the derivative with respect to the comoving LOS distance, $\mathbf{u} = \mathbf{v}/(faH)$ is the scaled velocity, $u_\parallel$ is its LOS component, and $f = d\ln D/d\ln a$ is the linear growth function. It can be shown that $f \simeq \Omega_{\rm m}^{0.55}$ in the ΛCDM model (see Appendix B of Wang & Steinhardt, 1998), which has also been found in simulations (e.g., Linder, 2005). In Fourier space, Equation (2.44) reads

$$\widetilde{\delta}_{\rm g}^s(\mathbf{k}) = \int {\rm d}^3\mathbf{x} \left(\delta_{\rm g}(\mathbf{x}) - f\partial_\parallel u_\parallel(\mathbf{x})\right) e^{i\mathbf{k}\cdot\mathbf{x} + ifk_\parallel u_\parallel(\mathbf{x})}. \tag{2.45}$$

The power spectrum in redshift space is therefore

$$P_{\rm g}^s(\mathbf{k}) = \int d^3\xi\, e^{i\mathbf{k}\cdot\xi} \langle e^{ifk_\parallel \Delta u_\parallel} \left(\delta_{\rm g}(\mathbf{x}) - f\partial_\parallel u_\parallel(\mathbf{x})\right) \left(\delta_{\rm g}(\mathbf{x}') - f\partial_\parallel u_\parallel(\mathbf{x}')\right)\rangle, \tag{2.46}$$

where $\xi = \mathbf{x} - \mathbf{x}'$ and $\Delta u_\parallel = u_\parallel(\mathbf{x}) - u_\parallel(\mathbf{x}')$. It can be shown that the factor $e^{ifk_\parallel \Delta u_\parallel}$ leads to the FoG damping of the form

$$D_{\rm FoG}(\mathbf{k}) = \left|\langle e^{ifk\mu u_\parallel}\rangle\right|^2 = \left|\int {\rm d}u_\parallel \phi(u_\parallel) e^{ifk\mu u_\parallel}\right|^2, \tag{2.47}$$

where $\mu = k_\parallel/k$ is the cosine of the angle between $\mathbf{k}$ and the LOS. $D_{\rm FoG}(\mathbf{k})$ is the square of the Fourier transform of the PDF of the velocity distribution $\phi(u_\parallel)$ (Hikage & Yamamoto, 2016). The FoG damping can be understood as the reduction of clustering along the LOS on small scales because of the uncorrelated peculiar velocities of galaxies. Similarly, the uncorrelated peak shift from Lyα RT reduces small-scale clustering along the LOS (Byrohl et al., 2019).

2.4.2 IGM Absorption

At linear order of the density contrast and for small velocities, the linearized continuity equation and equation (2.46) yield the so-called Kaiser effect (Kaiser, 1987),

$$P_{\rm g}^s(\mathbf{k}) = (b_{\rm g} + f\mu^2)^2 P_{\rm m}(\mathbf{k}) = b_{\rm g}^2(1 + \beta\mu^2)^2 P_{\rm m}(\mathbf{k}), \tag{2.48}$$

where $\beta = f/b_{\rm g}$ and $b_{\rm g} = \delta_{\rm g}/\delta_{\rm m}$ is the linear galaxy bias. We can write the anisotropic power spectrum in terms of its multipoles,

$$P_{\rm g}^s(k,\mu) = \sum_\ell P_{{\rm g},\ell}(k)\mathcal{L}_\ell(\mu), \tag{2.49}$$

where $\mathcal{L}_\ell(\mu)$ are the Legendre polynomials and

$$P_{{\rm g},\ell}(k) = \frac{2\ell+1}{2}\int_{-1}^{1} d\mu P_{\rm g}^s(k,\mu)\mathcal{L}_\ell(\mu). \tag{2.50}$$

For the linear RSD, only $\ell \in \{0, 2, 4\}$ are non-zero (e.g., Percival & White, 2009):

$$P_{\mathrm{g},\ell=0}(k) = \left(1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2\right) b_{\mathrm{g}}^2 P_{\mathrm{m}}(k), \tag{2.51}$$

$$P_{\mathrm{g},\ell=2}(k) = \left(\frac{4}{3}\beta + \frac{4}{7}\beta^2\right) b_{\mathrm{g}}^2 P_{\mathrm{m}}(k), \tag{2.52}$$

$$P_{\mathrm{g},\ell=4}(k) = \frac{8}{35}\beta^2 b_{\mathrm{g}}^2 P_{\mathrm{m}}(k). \tag{2.53}$$

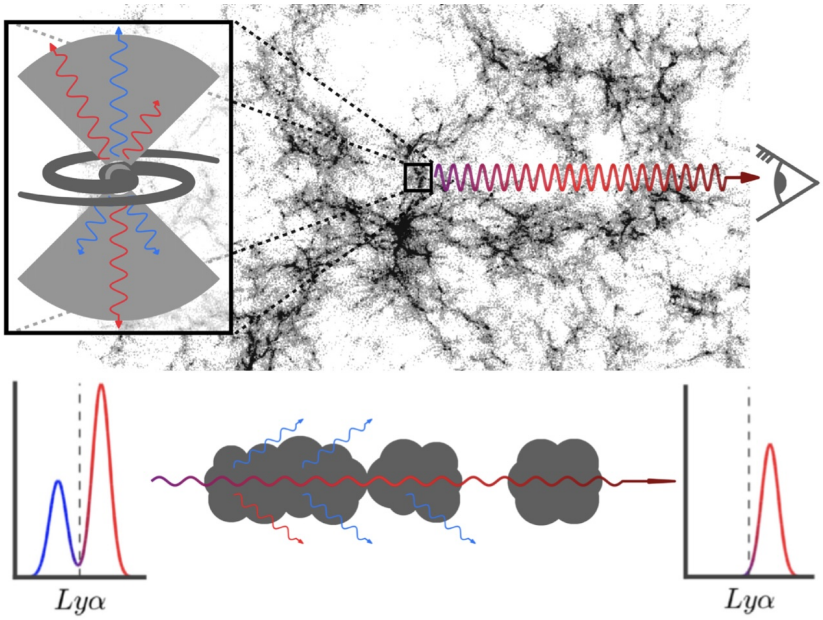

Figure 2.3: Illustration of the selection effect on LAEs due to Lyα RT, adapted from Gurung-López et al. (2020). Lyα photons are emitted in star-forming galaxies and travel toward the oserver through the IGM. Photons that are blue enough can scatter out of the line of sight, effectively being absorbed. The dependence of this process on the large-scale matter and velocity distributions can introduce a selection effect of LAE detections, which alters the clustering signal anisotropically.

Figure 2.3 shows an illustration of the effect of Lyα photons scattering out of the LOS in the IGM on the observed Lyα emission line. After escaping the ISM of the source galaxy, Lyα photons can scatter in the CGM and IGM on the way to an observer if their wavelengths are close enough to the Lyα line center or on its blue side. This reduces the observed Lyα flux of the galaxy. Because the scattering probability depends on the matter and velocity distributions, this can introduce an anisotropic selection effect of LAEs, changing the observed LAE clustering statistics.

The scattering of Lyα photons out of the LOS can be modeled as absorption. The optical depth for a photon with initial frequency ν on the way to an observer is

$$\tau(\nu) = \int_{r_{\mathrm{vir}}}^{\infty} \mathrm{d}r\, n_{\mathrm{H}}(r) x_{\mathrm{HI}}(r) \sigma_\alpha \left[\nu\left(1 + \frac{v_\parallel(r)}{c}\right)\right], \tag{2.54}$$

where $r_{\rm vir}$ is the virial radius of the dark-matter halo of the source galaxy, $n_{\rm H}$ is the hydrogen density, $x_{\rm HI}$ is the neutral fraction of hydrogen, and $v_\parallel$ is the LOS velocity of the gas at distance r from the galaxy, such that $v_\parallel > 0$ if it moves away from the observer (Wyithe & Dijkstra, 2011). Here, the coordinate system is centered on the galaxy ($r = 0$). $\sigma_{\rm Ly\alpha}(\nu)$ is the Lyα absorption cross-section at frequency ν. We can assume photoionization equilibrium, $x_{\rm HI} = n_{\rm H}\alpha^A_{\rm rec}(T)/\Gamma$, where $\alpha^A_{\rm rec}(T)$ is the temperature-dependent case-A recombination coefficient and Γ is the ionization rate. We assume case-A recombination because of the large ionization fraction and low density of the IGM. We obtain

$$\tau(\nu) = \int_{r_{\rm vir}}^{\infty} {\rm d}r\, \frac{n^2_{\rm H}(r)\alpha^{\rm A}_{\rm rec}\sigma_\alpha\left[\nu\left(1+\frac{v_\parallel(r)}{c}\right)\right]}{\Gamma(r)}. \tag{2.55}$$

We can approximate the absorption cross-section as a Dirac delta function,

$$\begin{aligned}\sigma_\alpha(\nu) &\approx f_\alpha \pi \frac{e^2}{m_e c}\delta(\nu-\nu_\alpha)\\ &= f_\alpha \pi r_e c \delta(\nu-\nu_\alpha) =: \sigma^{\rm tot}_\alpha \delta(\nu-\nu_\alpha),\end{aligned} \tag{2.56}$$

where ν_α is the Lyα rest-frame frequency, e is the electron charge, m_e is the electron mass, $r_e = 2.81\times10^{-13}\,$cm is the classical electron radius defined by $\frac{e^2}{r_e} = m_e c^2$, and $f_\alpha = 0.4167$ is the oscillator strength of the Lyα transition (e.g., Wyithe & Dijkstra, 2011; Bartelmann, 2021). Inserting this in equation (2.55) and integrating yields

$$\tau_\delta = \frac{\left(c+v_\parallel\right) n^2_{\rm H}\alpha^{\rm A}_{\rm rec}\sigma^{\rm tot}_\alpha}{\nu_\alpha \Gamma\left|\frac{{\rm d}v^p_\parallel}{{\rm d}r} - H(z)\right|}. \tag{2.57}$$

Here, $v^p_\parallel$ is the peculiar velocity, such that $v_\parallel = v^p_\parallel - H(z)r$. We used $\delta\left(g(r)\right) = \delta(r-r_0)\left|g'(r_0)\right|^{-1}$, where g' denotes the derivative of g with respect to r and $g(r_0) = 0$. Specifically, $g(r) = \nu(r) - \nu_\alpha = \nu\left(1+v_\parallel(r)/c\right) - \nu_\alpha$, so that $\delta\left(\nu(r)-\nu_\alpha\right) = \delta(r-r_0)\nu_\alpha^{-1}(c+v_\parallel)dv_\parallel/dr$. Equation (2.57) shows that the absorption of Lyα photons on the way to an observer increases with increasing hydrogen density, decreasing ionization rate, and decreasing LOS velocity gradient. Because the scattering of photons out of the LOS reduces the observed Lyα flux of a galaxy, it reduces the detectability of this galaxy through the Lyα line.

If the hydrogen density, the ionization rate, and the velocity gradient correlate with the large-scale structure, and if a large enough fraction of Lyα photons are subject to scattering after escaping the ISM and CGM, this could bias LAE correlation function and power spectrum measurements. Zheng et al. (2011) post-process a cosmological hydrodynamic simulation with Lyα RT and calculate the power spectrum of detected LAEs after RT and compare it to the galaxy power spectrum with random Lyα emission properties. They find that the selection effect of LAEs from RT in the IGM inverts the linear Kaiser RSD. This is because Lyα photons are less likely to escape high-density regions than low-density regions, changing the effective bias. Sources moving toward the observer are often on the far side of an overdensity, and are also less likely to be detected

because of the additional peculiar blueshift, and hence the stronger attenuation, of the Lyα photons. This velocity dependence affects the RSD. However, only photons that are close enough to the Lyα wavelength, or on its blue side, after escaping the ISM and CGM are subject to scattering in the IGM and can cause this effect. Zheng et al. (2011) assume that the escaped line shape is Gaussian centered on the Lyα wavelength, not taking into account the characteristic peak redshift and asymmetric line profile caused by the scattering within the ISM and CGM (see Section 2.3.1). A redshifted Lyα line is less prone to scattering than a Gaussian line centered on the Lyα wavelength.

We show in Chapter 6 that most of the Lyα RT takes place in very dense, nonlinear regions of the universe. However, the linearized model of Wyithe & Dijkstra (2011) can be helpful for a qualitative understanding of a potential effect of IGM absorption on the observed LAE clustering statistics. We can introduce a linearized optical depth following Wyithe & Dijkstra (2011),

$$\tau_{\rm lin} = \tau_0 \frac{1 + c_\gamma \delta_{\rm m}}{1 + \delta_\Gamma + \delta_v}, \tag{2.58}$$

where τ_0 is the mean optical depth in the IGM, $\delta_\Gamma = \frac{\Gamma}{\Gamma_0} - 1$ is the ionization rate perturbation, and $c_\gamma = (2.7 - 0.7\gamma) \simeq 1.72$ with the polytropic index $\gamma = 1.4$ (Hui & Gnedin, 1997) quantifies the dependence of the optical depth on dark matter density. The fluctuation in the LOS velocity is given by $\delta_v = \frac{{\rm d}v_z^p}{{\rm d}(ar_{\rm com})} \frac{1}{H}$, where $r_{\rm com}$ is the comoving LOS distance from the galaxy. We can further assume that a fraction $(1 - F_{\rm abs})$ of photons that escape the ISM and CGM are redshifted far enough out of resonance that they are not subject to scattering in the IGM. The transmittance of the IGM is therefore

$$\mathcal{T}(\delta_{\rm m}, \delta_\Gamma, \delta_v) = (1 - F_{\rm abs}) + F_{\rm abs} e^{-\tau_{\rm lin}}. \tag{2.59}$$

Because of the absorption, the observed flux is reduced by a factor of $\mathcal{T}$. Let $n_{g_\alpha}(\mathbf{x}|F > F_{\rm min})$ be the observed number density of LAEs at location $\mathbf{x}$ whose observed flux F is bright enough to be detected. Its dependence on $\delta_{\rm m}$, δ_Γ, and δ_v can be approximated as

$$\begin{aligned} n_{g_\alpha}(\mathbf{x}|F > F_{\rm min}) = {} & \bar{n}_{g_\alpha}(L_{\rm int} > L_{\rm min}) \left(1 + b_{g_\alpha} \delta_{\rm m}(\mathbf{x})\right) \\ & + \left[\Gamma(\mathbf{x}) - \Gamma_0\right] \left.\frac{\partial \mathcal{T}}{\partial \Gamma}\right|_{\mathcal{T}_0, \Gamma_0} \left.\frac{\partial \bar{n}_{g_\alpha}}{\partial \mathcal{T}}\right|_{\mathcal{T}_0} \\ & + \left[\rho(\mathbf{x}) - \rho_0\right] \left.\frac{\partial \mathcal{T}}{\partial \rho}\right|_{\mathcal{T}_0, \rho_0} \left.\frac{\partial \bar{n}_{g_\alpha}}{\partial \mathcal{T}}\right|_{\mathcal{T}_0} \\ & + \left[\frac{{\rm d}v_\parallel}{{\rm d}(ar_{\rm com})} - H\right] \left.\frac{\partial \mathcal{T}}{\partial \left(\frac{{\rm d}v_\parallel}{{\rm d}(ar_{\rm com})}\right)}\right|_{\mathcal{T}_0, H} \left.\frac{\partial \bar{n}_{g_\alpha}}{\partial \mathcal{T}}\right|_{\mathcal{T}_0}, \end{aligned} \tag{2.60}$$

where b_{g_α} is the LAE bias, $\bar{n}_{g_\alpha}(L_{\rm int} > L_{\rm min})$ is the mean number density of LAEs whose intrinsic luminosity, i.e., without IGM absorption, is larger than $L_{\rm min} = 4\pi D_L^2 F_{\rm min}$. We

can rewrite this expression as

$$\begin{aligned} n_{g_\alpha}(\mathbf{x}|F > F_{\rm min}) = \bar{n}_{g_\alpha}(L_{\rm int} > L_{\rm min})\,[1 + b_{g_\alpha}\delta_{\rm m}(\mathbf{x})] \\ + \delta_\Gamma \left.\frac{\partial \log \mathcal{T}}{\partial \log \Gamma}\right|_{\mathcal{T}_0,\Gamma_0} \left.\frac{\partial \bar{n}_{g_\alpha}}{\partial \log \mathcal{T}}\right|_{\mathcal{T}_0} \\ + \delta_{\rm m} \left.\frac{\partial \log \mathcal{T}}{\partial \log \rho}\right|_{\mathcal{T}_0,\rho_0} \left.\frac{\partial \bar{n}_{g_\alpha}}{\partial \log \mathcal{T}}\right|_{\mathcal{T}_0} \\ + \delta_v \left.\frac{\partial \log \mathcal{T}}{\partial \log\left(\frac{{\rm d}v_\parallel}{{\rm d}r_{\rm com}}\right)}\right|_{\mathcal{T}_0,H} \left.\frac{\partial \bar{n}_{g_\alpha}}{\partial \log \mathcal{T}}\right|_{\mathcal{T}_0}. \end{aligned} \tag{2.61}$$

We obtain the observed LAE overdensity,

$$\delta_{g_\alpha}(\mathbf{x}) = \frac{n_{g_\alpha}(\mathbf{x}|F > F_{\rm min})}{\bar{n}_{g_\alpha}(L > L_{\rm min})} - 1 = \delta_{\rm m}(b_{g_\alpha} + C_\rho^{g_\alpha}) + \delta_\Gamma C_\Gamma^{g_\alpha} + \delta_v C_v^{g_\alpha}, \tag{2.62}$$

where

$$C_\rho^{g_\alpha} = \frac{1}{\bar{n}_{g_\alpha}(L > L_{\rm min})} \left.\frac{\partial \log \mathcal{T}}{\partial \log \rho}\right|_{\mathcal{T}_0,\rho_0} \left.\frac{\partial \bar{n}_{g_\alpha}}{\partial \log \mathcal{T}}\right|_{\mathcal{T}_0}, \tag{2.63}$$

$$C_\Gamma^{g_\alpha} = \frac{1}{\bar{n}_{g_\alpha}(L > L_{\rm min})} \left.\frac{\partial \log \mathcal{T}}{\partial \log \Gamma}\right|_{\mathcal{T}_0,\Gamma_0} \left.\frac{\partial \bar{n}_{g_\alpha}}{\partial \log \mathcal{T}}\right|_{\mathcal{T}_0}, \tag{2.64}$$

$$\text{and} \quad C_v^{g_\alpha} = \frac{1}{\bar{n}_{g_\alpha}(L > L_{\rm min})} \left.\frac{\partial \log \mathcal{T}}{\partial \log\left(\frac{{\rm d}v_\parallel}{{\rm d}r_{\rm com}}\right)}\right|_{\mathcal{T}_0,H} \left.\frac{\partial \bar{n}_{g_\alpha}}{\partial \log \mathcal{T}}\right|_{\mathcal{T}_0}. \tag{2.65}$$

To calculate $C_\rho^{g_\alpha}$, $C_\Gamma^{g_\alpha}$, and $C_v^{g_\alpha}$, we can model the fluctuation of the ionization rate as a function of the matter overdensity,

$$\widetilde{\delta}_\Gamma(\mathbf{k}) = b_{\rm ion}\widetilde{\delta}_{\rm m}(\mathbf{k})\frac{\arctan(k\lambda_{\rm mfp})}{k\lambda_{\rm mfp}} = b_{\rm ion}\widetilde{\delta}_{\rm m}(\mathbf{k})K_\Gamma(k), \tag{2.66}$$

where $\lambda_{\rm mfp}$ is the mean free path of ionizing photons. The LAE overdensity in redshift space is given by $\delta^s_{g_\alpha} = \delta_{g_\alpha} - \delta_v$ (see Equation (2.44)), so that we obtain

$$\widetilde{\delta}^s_{g_\alpha}(\mathbf{k}) = \widetilde{\delta}_{\rm m}\left(b_{g_\alpha} + b_{\rm ion}C_\Gamma^{g_\alpha}K_\Gamma + C_\rho^{g_\alpha} + (1 - C_v^{g_\alpha})f\mu^2\right). \tag{2.67}$$

Here, we assumed linear RSD, i.e., $\tilde{\delta}_v(\mathbf{k}) = -f\mu^2\tilde{\delta}_m(\mathbf{k})$. If the luminosity function can be described as a power-law function with slope $-\alpha$ around $L_{\rm min}$, where $\alpha > 1$, the mean number density of LAEs with intrinsic luminosity above $L_{\rm min}$ is

$$\bar{n}_{g_\alpha}(L_{\rm int} > L) = \bar{n}_{g_\alpha}(L_{\rm int} > L_{\rm min})\left(\frac{L}{L_{\rm min}}\right)^{1-\alpha}. \tag{2.68}$$

After calculating the derivatives in equations (2.63), (2.64), and (2.65), we obtain

$$\begin{aligned} C_\Gamma^{g_\alpha} = C_v^{g_\alpha} = (\alpha - 1)\frac{F_{\rm abs}\tau_0 e^{-\tau_0}}{(1 - F_{\rm abs}) + F_{\rm abs}e^{-\tau_0}} \\ \text{and } C_\rho^{g_\alpha} = -c_\gamma C_\Gamma^{g_\alpha}. \end{aligned} \tag{2.69}$$

The power spectrum in redshift space after IGM absorption therefore reads

$$\begin{aligned} P_{g_\alpha}(\mathbf{k}) &= \left(b_{g_\alpha} + b_{\rm ion} C_\Gamma^{g_\alpha} K_\Gamma + C_\rho^{g_\alpha} + (1 - C_v^{g_\alpha}) f\mu^2\right)^2 P_{\rm m}(\mathbf{k}) \\ &= (b_{g_\alpha}^{\rm eff})^2 \left(1 + \beta_{g_\alpha}^{\rm eff}\mu^2\right)^2 P_{\rm m}(\mathbf{k}), \end{aligned} \tag{2.70}$$

where $P_{\rm m}(\mathbf{k})$ is the matter power spectrum and the effective bias is given by

$$b_{g_\alpha}^{\rm eff} = b_{g_\alpha} + b_{\rm ion} C_\Gamma^{g_\alpha} K_\Gamma(k) + C_\rho^{g_\alpha} \tag{2.71}$$

and the effective RSD parameter is given by

$$\beta_{g_\alpha}^{\rm eff} = \frac{(1 - C_v^{g_\alpha}) f}{b_{g_\alpha}^{\rm eff}} = \beta_{g_\alpha} \frac{(1 - C_v^{g_\alpha}) b_{g_\alpha}}{b_{g_\alpha}^{\rm eff}}. \tag{2.72}$$

This shows that the RSD can be reduced if $(1 - C_v^{g_\alpha}) b_{g_\alpha} / b_{g_\alpha}^{\rm eff} < 1$. While the above discussion only concerns the power spectrum of LAEs, we derive the corresponding model for the cross-power spectra of LAEs with non-LAE galaxies and the Lyα intensity, and for the Lyα intensity auto-power spectrum in Chapter 6. However, we find that the linear model is inadequate to describe the RT because most of the RT takes place in the nonlinear, high-density regime. This discussion should therefore only be used to build a qualitative intuition on the dependencies of IGM absorption on the environment and its potential effect on the power spectrum.

Behrens & Niemeyer (2013) and Behrens et al. (2018) repeat a similar study as Zheng et al. (2011), but using simulations with higher spatial resolution. They find that the Lyα observability is uncorrelated with the large-scale structure, while they can reproduce the effect of Zheng et al. (2011) when they artificially lower the resolution of their simulation before calculating the RT. The authors argue that the RT is dominated by the small scales that were not sufficiently resolved by Zheng et al. (2011). However, they find a strong influence of the spectra emerging from the ISM on the coupling between large-scale density and velocity fields with the scattering of Lyα photons in the IGM similar to Laursen et al. (2011).

This implies a large modeling uncertainty of the RT effect on clustering statistics, showing the need for observational constraints. At the time of writing, Croft et al. (2016) and Lin et al. (2022) constrained $\beta_{\rm Ly\alpha}^{\rm eff}$ by cross-correlating quasar positions with the diffuse Lyα intensity in residual spectra of low-redshift galaxies at redshifts $z = 2 - 3.5$ ($\bar{z} \simeq 2.5$, see Section 1.3). The expected RSD coefficient at redshift 2.5 from the Kaiser formula, assuming a Lyα intensity bias $b_I = 3$ is $\beta_{\rm Ly\alpha}^{\rm Kaiser} = \Omega_{\rm m}(z = 2.5)^{0.55}/b_I \simeq 0.32$. Croft et al. (2016) measure $\beta_{\rm Ly\alpha}^{\rm eff} = -0.76 \pm 0.36$, which is inconsistent with the linear Kaiser RSD and implies a strong Lyα RT effect. However, after improving the analysis, Croft et al. (2018) find a strong systematic contribution to the previous measurement. Because their signal is much lower, they do not constrain $\beta^{\rm eff}$. Using a larger quasar catalog and residual spectra, Lin et al. (2022) measure $\beta_{\rm Ly\alpha}^{\rm eff} = 0.07^{+1.65}_{-0.7}$. This is consistent with the linear Kaiser RSD.

2.5 Other Targeted Lines

Various emission lines can be used for LIM. Their emission mechanisms vary widely, and they probe different phases of the ISM, CGM, and IGM. In this section, I will briefly outline commonly targeted emission lines in order of increasing restframe frequency. We discussed the Lyα line in detail in Sections 2.2, 2.3, and 2.4 instead of this section. Much of this section is inspired by Bernal & Kovetz (2022).

2.5.1 21 cm

When the spin of the electron in a hydrogen atom flips from being parallel with the proton spin (triplet state) to being anti-parallel (singlet state), a photon with the wavelength of 21 cm is emitted (van de Hulst, 1945; Ewen & Purcell, 1951). This transition is forbidden with an Einstein A-coefficient of $A_{21\,\mathrm{cm}} = 2.85 \times 10^{15}\,\mathrm{s}^{-1}$. Therefore, the average lifetime of a photon in the triplet state is $\sim 10^8$ years, making this emission line extremely weak. However, the abundance of hydrogen atoms in the universe makes it bright enough for LIM.

Because the 21 cm line is sourced by neutral hydrogen and does not require the presence of star formation, a 21 cm signal is present after recombination at $z \sim 1100$. Before the formation of the first stars at $z \sim 20 - 30$, in the so-called Dark Ages, hydrogen was almost completely neutral and can be traced by 21 cm absorption against the CMB. During reionization, stars and galaxies ionized neutral hydrogen atoms around them, creating holes in the neutral hydrogen distribution. After reionization, the 21 cm line primarily traces neutral high-density regions in galaxies (Padmanabhan, 2024).

The 21 cm line falls into the radio wavelength range. Several surveys target the 21 cm line for LIM, pursuing different observing tactics and redshifts. LOFAR ($z \geq 4.9$; van Haarlem et al., 2013; Shimwell et al., 2019), nenuFAR ($z \gtrsim 16$; Bonnassieux et al., 2021), LWA ($7 \lesssim z \lesssim 20$; Eastwood et al., 2018), CHIME ($0.8 \leq z \leq 2.5$; Amiri et al., 2022), uGMRT ($2 \leq z \leq 3.6$; Chakraborty et al., 2021), Tianlai ($z \leq 0.068$; Perdereau et al., 2022), HIRAX ($0.8 \leq z \leq 2.5$; Newburgh et al., 2016), CHORD ($z \leq 3.7$; Vanderlinde et al., 2019), PUMA ($0.3 \leq z \leq 6$; Slosar et al., 2019), MWA ($4 \lesssim z \lesssim 17$; Tingay et al., 2013), HERA ($4.7 \leq z \leq 27$; DeBoer et al., 2017), SKAO ($z \leq 6$; Bacon et al., 2020), and its predecessor MeerKAT ($z \leq 1.45$; Santos et al., 2016) are interferometers, while FAST ($z \leq 0.35$; Bigot-Sazy et al., 2016) and BINGO ($0.13 \leq z \leq 0.48$; Wuensche & the BINGO Collaboration, 2019) are examples of single-dish telescopes. However, SKAO-MID and MeerKAT only calculate the auto-correlations of individual antennas for the cosmological survey, called "single-dish mode".

At the time of writing, 21 cm is the most studied LIM target. While there are no detections at the EoR or the Dark Ages, the 21 cm power spectrum has been detected in cross-correlation with galaxies at $z \lesssim 1.5$ (e.g., Chang et al., 2010; MeerKLASS Collaboration et al., 2025), in cross-correlation with the Lyα forest at $1.8 \leq z \leq 2.5$ (Amiri et al., 2024), and as an autopower spectrum at $z \simeq 0.32$ and $z \simeq 0.44$ (Paul et al., 2023).

2.5.2 Carbon Monoxide

The carbon monoxide (CO) molecule has quantized angular momentum from the carbon and oxygen atoms rotating around the center of mass. A change in this angular momentum can only occur through emission or absorption of a photon with the energy corresponding to the energy difference between the angular momentum states. The frequency of the emitted photon during the transition from the angular momentum J to $J-1$ is $\nu_{J\to J-1} = J \times 115.27\,\mathrm{GHz}$. The frequency falls into the mm or sub-mm wavelength range and can be observed with telescopes on Earth.

CO molecules exist in dense, cold molecular clouds within galaxies. Astrophysical CO emission was first discovered in the Orion nebula by Wilson et al. (1970). The CO intensity can therefore be used to estimate the cold gas content of galaxies; either by estimating the mass of a molecular cloud or by counting the number of molecular clouds, assuming that the counted objects have uniform properties (e.g., Kennicutt & Evans, 2012). Because stars form in cold molecular gas clouds, one can infer the SFR of the galaxies from the CO intensity (Tacconi et al., 2020). By measuring the cold gas density at higher redshifts, one can constrain the star-formation history of the universe. While most of our current understanding of the star-formation rate density of the universe comes from stellar light and emission lines from the hot ISM, CO emission is a more direct probe of the actual star-formation sites.

Current and future surveys observing the CO lines include COPSS (Keating et al., 2015, 2016), mmIME (Keating et al., 2020), COMAP (Cleary et al., 2022), CONCERTO (CONCERTO Collaboration et al., 2020), TIME (Crites et al., 2014; Sun et al., 2021), EXCLAIM (Switzer et al., 2021), FYST (Collaboration et al., 2022), and SPT-SLIM (Karkare et al., 2022).

The first detection of the CO power spectrum was measured from COPSS data at $2.3 \leq z \leq 3.3$ (Keating et al., 2015, 2016). Using data from mmIME, Keating et al. (2020) found measured the CO power spectrum even more precisely from $1 \lesssim z \lesssim 5$. Roy et al. (2024) detect the stacked CO(3-2) intensity at $z \sim 0.5$ of galaxies in eBOSS in the CO(2-1) map of the Planck CMB observations. The COMAP collaboration constrained the molecular gas density at $2.4 \leq z \leq 3.4$ by measuring upper limits on the CO power spectrum (Lunde et al., 2024; Stutzer et al., 2024; Chung et al., 2024). Dunne et al. (2024) calculated similar constraints from an upper limit on the stacked intensity on quasars.

2.5.3 Ionized Carbon

Singly ionized carbon ([C II]) has two fine-structure energy levels in the electronic ground state. The transition from the $^2P_{3/2}$ state to the $^2P_{1/2}$ state emits a photon with a wavelength of $\lambda_{[\mathrm{CII}]} = 157.74\mu\mathrm{m}$, corresponding to a frequency of $\nu_{[\mathrm{CII}]} = 1900.537\,\mathrm{GHz}$ (Cooksy et al., 1986). Astrophysical [C II] emission has first been detected in the Orion nebula (Russell et al., 1980). The [C II] fine-structure transition can be excited via collisions with hydrogen molecules, atoms, and electrons, and it is the primary cooling mechanism of the ISM in star-forming galaxies (e.g., Field et al., 1968; Dalgarno & McCray, 1972; Stacey et al., 1991). This frequency can be observed from the ground at $z \gtrsim 4.5$, where the Earth's atmosphere becomes transparent. Because the frequency

of the CO transitions is slightly lower than that of [C II], they act as line interlopers for [C II] observations (see Section 3.5.2).

The [C II] emission line traces multiple phases of the ISM: cold atomic clouds, the diffuse warm neutral and ionized media, and molecular clouds. It is difficult to distinguish between the different [C II] emission components (Lagache et al., 2018). The [C II] emission line can also be detected in the CGM (Fujimoto et al., 2019, 2020).

Surveys observing the [C II] line include CONCERTO (CONCERTO Collaboration et al., 2020), EXCLAIM (Switzer et al., 2021), FYST (Collaboration et al., 2022), TIM (Vieira et al., 2020), and TIME (Crites et al., 2014; Sun et al., 2021). The first [C II] LIM measurement detected the cross-correlation between the Planck High-Frequency Instrument and high-redshift catalogs of quasars and luminous red galaxies (LRGs) at a 2σ confidence level (Pullen et al., 2018), and later reached a 4σ detection using an improved methodology (Yang et al., 2019).

2.5.4 Hα

The Balmer-α or Hα line transition corresponds to the transition of the electron in a hydrogen atom from the energy state $n = 3$ to $n = 2$. A photon with a wavelength of $\lambda_{\mathrm{H}\alpha} = 6564.6\,\text{Å}$ is emitted during this transition (see Equation (2.13)). The emission follows the recombination of a hydrogen atom, in which the electron ends up in an excited state. Due to the small lifetime of these excited states, the electron "falls" through the lower energy levels in a radiative cascade, often involving an Hα transition. The Hα line therefore traces the ionized ISM and can be used to estimate the SFR (Kennicutt et al., 1994; Kennicutt, 1998),

$$\dot{M}_\star \simeq \frac{L_{\mathrm{H}\alpha}}{1.126 \times 10^{41}\,\mathrm{erg\,s^{-1}}} M_\odot\,\mathrm{yr}^{-1}. \tag{2.73}$$

However, this estimate is affected by systematic errors, including dust attenuation of the Hα emission and the sensitivity of the Hα emission to the abundance of high-mass stars (Kennicutt & Evans, 2012). The Hα emission line will be mapped by SPHEREX at $0.14 \leq z \leq 6.62$ (Doré et al., 2014, 2016, 2018, see also Section 1.5).

2.5.5 Cross-correlations

Cross-correlations between different emission lines or between an emission line and another tracer in the same volume can provide several benefits. While systematic errors are prominent in LIM autocorrelation measurements, instrumental systematics should be uncorrelated across different instruments, so that cross-correlation studies can be used to remove systematics and astrophysical foregrounds (e.g., Wolz et al., 2016; Fronenberg & Liu, 2024). Because line interlopers are uncorrelated with galaxy catalogs with the matching redshift of the target line, they also do not contribute to the cross-correlation signal.

Cross-correlations between different line intensity maps can be used to recover the LIM auto-power spectrum (Beane et al., 2019; McBride & Liu, 2024). Assuming a linear

bias model, one can estimate the autopower spectrum of line A from the crosspower spectra of line A with lines B and C and the cross-power spectrum of lines B and C,

$$\hat{P}_{AA} \propto \frac{P_{AB} P_{AC}}{P_{BC}}. \tag{2.74}$$

The proportionality factor depends on the biases of the lines.

Cross-correlating different line intensity maps can give insight into physical processes affecting the line emission. For example, Lyα and Hα are both emitted with hydrogen recombinations, but Lyα can also be emitted by collisional excitation. It is ongoing research to determine the dominant emission mechanism of Lyα halos (see Section 2.3.2) and the diffuse IGM (Bacon et al., 2021; Martin et al., 2023). Combining the Lyα and Hα emission at different scales, for example with HETDEX and SPHEREX data, could constrain the contribution of recombination to the Lyα glow of the CGM and IGM, and the escape fraction of Lyα photons from the ISM (see Schaan & White, 2021, for a formalism of decorrelation scales).

At the EoR, high-energy photons from the first stars, galaxies, and AGNs ionized the neutral hydrogen in the IGM. The size of the ionized regions as a function of time can be constrained with LIM cross-correlations (e.g., Lidz et al., 2011). While 21 cm photons are emitted in the neutral gas outside of the ionized bubbles, Lyα, CO, and [C II] photons originate from the star-forming regions and from within the ionized bubble. Therefore, a cross-correlation signal between these lines and the 21 cm line will be anticorrelated on scales larger than the bubble size. LIM can thus constrain the ionized bubble sizes at reionization.

Finally, by cross-correlating CMB lensing maps with a line intensity map, one can extract the contribution of the matter at a higher redshift than the line intensity map to the CMB lensing signal and use this to constrain the high-redshift matter power spectrum (Maniyar et al., 2022; Fronenberg et al., 2024b,a).

2.6 Modeling Line Intensity Maps

There are multiple approaches to modeling line intensity maps, ranging from high-resolution hydrodynamic simulations that are post-processed to include line emission and RT, to simple Gaussian or lognormal fields. While the former are more detailed than the latter, they are more computationally involved, and one can only create a small number of realizations. On the other hand, one can run many realizations of simple models to create covariance matrices of summary statistics, e.g., power spectra and bispectra. Because of their simplicity, they also enable more straightforward interpretation of measurements. However, they do not contain information about the emission mechanisms of the lines, which LIM aims to constrain.

Cosmological hydrodynamic simulations typically model cosmological structure formation as well as the hydrodynamics of gas. They employ semi-analytic sub-grid models for star formation and black hole growth, and their impact on the gas ionization state, metallicity and temperature (e.g., Somerville et al., 2015). The gas density, temperature,

metallicity, and ionized fraction are calculated in each cell. One can then infer an expected line luminosity for each cell (e.g., Moriwaki et al., 2019; Silva et al., 2021; Kannan et al., 2022; Byrohl & Nelson, 2023; Liang et al., 2023; Garcia et al., 2024). However, this type of simulation requires many assumptions, which can lead to vastly different outcomes (e.g. Bhagwat et al., 2024). They are also computationally demanding, so only a small number of realizations can be obtained with this amount of detail.

Because of the large cross-section of hydrogen atoms at the Lyα wavelength, simply painting the Lyα luminosity on simulated galaxies is insufficient. One needs to model the Lyα RT to accurately estimate the Lyα intensity of galaxies and diffuse gas. Byrohl et al. (2021) and Byrohl & Nelson (2023) post-process a cosmological hydrodynamic simulation with Lyα RT. First, they assign a Lyα luminosity to each cell, based on the hydrogen and electron densities, the temperature, and the ionization state of the gas, and the ionization rate from star-forming cells, using equations (2.19), (2.20), and (2.29). They then calculate the RT using a Monte-Carlo approach (Byrohl, 2021) and calibrate the emission from the ISM to the observed Lyα luminosity function to account for dust extinction. The RT significantly changes the photon budget, spatial distribution, and intensity of diffuse gas. For example, Byrohl & Nelson (2023) find that the IGM is illuminated predominantly by scattered Lyα photons originating from star-forming galaxies and their CGM, rather than diffuse emission of the IGM.

To lower the computational cost, one can simplify the modeling of the structure formation of matter. For example, one can simulate only gravitational interaction, removing the hydrodynamic and sub-grid modeling computations, in N-body simulations. Relations between the dark matter halo mass and the emitted line luminosities are used to paint the emission lines onto the simulation (Lidz et al., 2011; Gong et al., 2012; Li et al., 2016; Chung et al., 2019; Spina et al., 2021; Béthermin et al., 2022; Moradinezhad Dizgah et al., 2022b; Sato-Polito et al., 2023). To speed up the gravitational modeling even further, one can model the dark matter halos using Lagrangian perturbation theory or the mass-peak patch method (Mesinger et al., 2011; Silva et al., 2013, 2015; Mesinger et al., 2016; Heneka et al., 2017; Heneka & Mesinger, 2020; Mas-Ribas et al., 2023; Chung et al., 2022; Roy et al., 2023).

Relations between the dark matter halo mass and the line luminosities are obtained by relating the dark matter halo mass to its SFR and relating the SFR to the line luminosity. The SFR is often modeled as a double-power law function of the halo mass, and the relations between the line luminosities and the SFR are obtained from observations of galaxies (see Bernal & Kovetz, 2022, for a review). These relations are extrapolated to faint galaxies and higher redshifts to model line intensity maps. However, the low-redshift relations do not always hold at high redshift (Lagache et al., 2018; Chung et al., 2021). The relations are also calibrated to individual detected galaxies, while LIM is sensitive to the otherwise inaccessible faint galaxies by construction.

The simplest models for line intensity maps can be obtained by multiplying a random density field that follows a Gaussian or lognormal distribution with a given power spectrum by a bias and mean line intensity to obtain an intensity field (Alonso et al., 2014; Ramírez-Pérez et al., 2022; Rubiola et al., 2022). This approach is fast, but does not contain astrophysical information beyond the mean line intensity. While a Gaussian field

can output unphysical negative densities, a lognormal distribution is strictly positive and therefore preferable for modeling densities or intensities. The PDFs of the matter density in N-body simulations and the galaxy density in galaxy surveys also roughly follow a lognormal distribution (Kayo et al., 2001; Shin et al., 2017; Clerkin et al., 2017). If the intensity is sourced by discrete sources such as galaxies or halos, the intensity power spectrum also contains a shot noise term, which can be modeled by Poisson-sampling discrete sources from the lognormal or Gaussian field (e.g.; Agrawal et al., 2017). One can also model the intensity field directly based on effective field theory (Obuljen et al., 2023).

Chapter 5 presents a mock generation approach and public code based on a simple technique. Instead of multiplying a Gaussian or lognormal random field by a factor to obtain the intensity field, we Poisson-sample galaxies from the galaxy density field (Agrawal et al., 2017) and assign a line luminosity to each of these galaxies following a luminosity function. From these line luminosities, we calculate the line intensity map. This simple technique allows us to quickly simulate mock line intensity maps and include the shot noise contribution and observational effects. In Chapter 6, we apply basic Lyα RT effects to the line intensity maps to forecast the sensitivity of HETDEX to these effects. Because this thesis aims to measure the LAE-Lyα intensity cross-power spectrum of HETDEX, the primary purpose of these mocks is to estimate the covariance matrix of power spectra and provide a simple theoretical prediction for the signal. When the measurement becomes more precise in the future, we will need more detailed models for the interpretation.

Chapter 3

Statistics of Line Intensity Mapping

Because line intensity maps are typically noisy, summary statistics are necessary to extract physical information from them. One can measure one-point statistics such as the voxel intensity distribution (VID), two-point statistics such as the power spectrum and correlation function, higher-order statistics such as N-point correlation functions with $N \geq 3$, or topological statistics such as Minkowski functionals. These can be used in auto- or cross-correlation with other tracers. This chapter provides an overview of the most commonly used summary statistics for LIM. While the power spectrum, correlation function, and stacked signals have been measured, the VID and other summary statistics have only been explored theoretically at the time of writing.

3.1 Correlation Function and Power Spectrum

Similar to galaxy surveys, one can calculate the correlation function or its Fourier counterpart, the power spectrum, of the line intensity maps. One can calculate the auto-correlation or the cross-correlation with other large-scale structure tracers. This section follows Section 5.2, which is inspired by Blake (2019).

The cross-correlation between two fluctuation fields $\delta A(\mathbf{x}) = A(\mathbf{x}) - \langle A(\mathbf{x})\rangle$ and δB is defined as

$$\xi_{AB}(\mathbf{x}-\mathbf{y}) = \frac{\langle A(\mathbf{x})B(\mathbf{y})\rangle - \langle A(\mathbf{x})\rangle\langle B(\mathbf{y})\rangle}{\langle A(\mathbf{x})\rangle\langle B(\mathbf{y})\rangle} = \frac{\langle \delta A(\mathbf{x})\delta B(\mathbf{y})\rangle}{\langle A(\mathbf{x})\rangle\langle B(\mathbf{y})\rangle}\,, \tag{3.1}$$

where the angle brackets $\langle\cdot\rangle$ denote an average over many realizations. The corresponding power spectrum is defined as the Fourier transform of the correlation function,

$$P_{AB}(\mathbf{k}) = \int \mathrm{d}^3\mathbf{s}\, \xi^{AB}(\mathbf{s}) e^{i\mathbf{k}\cdot\mathbf{s}}. \tag{3.2}$$

Here, $\xi_{AB}(\mathbf{x})$ is dimensionless and $P_{AB}(\mathbf{k})$ has units of volume.

We can define the matter density contrast as

$$\delta_{\mathrm{m}}(\mathbf{x}) = \frac{\rho_{\mathrm{m}}(\mathbf{x}) - \langle\rho_{\mathrm{m}}(\mathbf{x})\rangle}{\langle\rho_{\mathrm{m}}(\mathbf{x})\rangle}, \tag{3.3}$$

where $\rho_\mathrm{m}(\mathbf{x})$ is the matter density at location $\mathbf{x}$ and $\langle\rho_\mathrm{m}\rangle$ is the mean matter density at that redshift. Similarly, a galaxy density contrast can be defined from the galaxy number density $n_\mathrm{g}(\mathbf{x})$ as

$$\delta_\mathrm{g}(\mathbf{x}) = \frac{n_\mathrm{g}(\mathbf{x}) - \langle n_\mathrm{g}(\mathbf{x})\rangle}{\langle n_\mathrm{g}(\mathbf{x})\rangle}, \tag{3.4}$$

where $\langle n_\mathrm{g}(\mathbf{x})\rangle$ is determined by the luminosity function of the observed galaxies and the detection limit of the survey,

$$\langle n_\mathrm{g}(\mathbf{x})\rangle = \int_{L_\mathrm{min}(\mathbf{x})}^{L_\mathrm{max}} dL \frac{dn}{dL}. \tag{3.5}$$

Here, $L_\mathrm{min}(\mathbf{x})$ is the minimum luminosity that can be detected at location $\mathbf{x}$. L_max is the maximum luminosity, for example, if one removes bright AGNs. Because the mean intensity is often unknown and is effectively subtracted from the maps (see Chapter 5), we define the intensity contrast without the division,

$$\delta I_\lambda(\mathbf{x}) = I_\lambda(\mathbf{x}) - \langle I_\lambda(\mathbf{x})\rangle, \tag{3.6}$$

where $I_\lambda(\mathbf{x})$ is the specific intensity at location $\mathbf{x}$ and $\langle I_\lambda(\mathbf{x})\rangle$ is the mean specific intensity at that redshift. The following equations are also valid for the specific intensity I_ν. On large scales, the galaxy density contrast and intensity contrast can be modeled as proportional to the matter density contrast with the proportionality factor known as the bias b (Desjacques et al., 2018),

$$\delta I_\lambda(\mathbf{x}) = b_{I_\lambda}\langle I_\lambda(\mathbf{x})\rangle\delta_\mathrm{m}(\mathbf{x}) \qquad \text{and} \qquad \delta_\mathrm{g}(\mathbf{x}) = b_\mathrm{g}\delta_\mathrm{m}(\mathbf{x}). \tag{3.7}$$

To model the LIM power spectrum, we assume that the intensity is sourced by galaxies, neglecting diffuse emission. Following Peebles (1980), we consider the intensity $I_\lambda(\mathbf{x})$ in an infinitesimal volume δV, so that only $N \in \{0, 1\}$ galaxies occupy this volume. The expectation value of N is $\bar{n}\,[1 + \delta_\mathrm{g}(\mathbf{x})]\,\delta V$, where $\bar{n}$ is the mean number density of galaxies in the universe at that redshift. Each galaxy has a line luminosity L_i, which is sampled from a luminosity function $\mathrm{d}n/\mathrm{d}L$. To convert the luminosity function into a PDF, $\mathcal{P}(L) := \frac{\mathrm{d}n}{\mathrm{d}L}\bar{n}^{-1}$, the integral $\int \mathrm{d}L\,\frac{\mathrm{d}n}{\mathrm{d}L} =: \bar{n}$ must converge. For realistic luminosity functions, setting a minimum luminosity is sufficient.

To get an expression for the LIM power spectra, we will need to know the expectation values $\langle I_\lambda\delta V\rangle$ and $\langle I_\lambda^2\delta V^2\rangle$. The expectation value of a function $f(A)$ of a continuous random variable A that is sampled from a PDF $\mathcal{P}(A)$ is $\langle f(A)\rangle = \int \mathrm{d}A\, f(A)\mathcal{P}(A)$. The PDF of $I_\lambda\delta V$ of a single source at position $\mathbf{x}$ is given by $\mathcal{P}'(I_\lambda\delta V) = \mathcal{P}(L)X_{I_\lambda}^{-1}(\mathbf{x})$, where (see Section 2.1)

$$X_{I_\lambda}(\mathbf{x}) = \frac{I_\lambda}{\rho_L} = \frac{c}{4\pi\lambda_0 H(z)(1+z)^2} \tag{3.8}$$

$$\text{and} \qquad X_{I_\nu}(\mathbf{x}) = \frac{I_\nu}{\rho_L} = \frac{c}{4\pi\nu_0 H(z)}. \tag{3.9}$$

The first moment of $I_\lambda \delta V$ is then given by

$$\begin{aligned}\langle I_\lambda(\mathbf{x})\delta V\rangle &= \langle X_{I_\lambda}(\mathbf{x})L\rangle \\ &= \bar{N} X_{I_\lambda}(\mathbf{x}) \int \mathrm{d}L\, \mathcal{P}(L) L \\ &= \delta V X_{I_\lambda}(\mathbf{x}) \int \mathrm{d}L \frac{\mathrm{d}n}{\mathrm{d}L} L,\end{aligned} \tag{3.10}$$

where we used $\bar{N} = \bar{n}\delta V$ in the last step. Similarly, the second moment is given by

$$\langle I_\lambda^2(\mathbf{x})\delta V^2\rangle = \bar{N} X_{I_\lambda}^2(\mathbf{x}) \int \mathrm{d}L\, \mathcal{P}(L) L^2 = \delta V X_{I_\lambda}^2(\mathbf{x}) \int \mathrm{d}L \frac{\mathrm{d}n}{\mathrm{d}L} L^2. \tag{3.11}$$

To get an expression for the intensity auto-correlation, we follow the integration approach of Feldman et al. (1994). Consider the expectation value of the integral

$$\int d^3\mathbf{x} \int d^3\mathbf{y}\, g(\mathbf{x},\mathbf{y}) I_\lambda(\mathbf{x}) I_\lambda(\mathbf{y}) \tag{3.12}$$

for an arbitrary function $g(\mathbf{x},\mathbf{y})$ over many realizations of the field values $I_\lambda(\mathbf{x})$. Using infinitesimal voxels with volume δV, we obtain

$$\begin{aligned}&\langle \int d^3\mathbf{x} \int d^3\mathbf{y}\, g(\mathbf{x},\mathbf{y}) I_\lambda(\mathbf{x}) I_\lambda(\mathbf{y})\rangle \\ &= \int d^3\mathbf{x} \int d^3\mathbf{y}\, g(\mathbf{x},\mathbf{y}) \langle I_\lambda(\mathbf{x}) I_\lambda(\mathbf{y})\rangle \\ &= \sum_{i,j} \delta V_i \delta V_j g(\mathbf{x}_i,\mathbf{x}_j) \langle I_\lambda(\mathbf{x}_i) I_\lambda(\mathbf{x}_j)\rangle \\ &= \sum_i \sum_{j\neq i} \delta V_i \delta V_j g(\mathbf{x}_i,\mathbf{x}_j) \langle I_\lambda(\mathbf{x}_i) I_\lambda(\mathbf{x}_j)\rangle + \sum_i g(\mathbf{x}_i,\mathbf{x}_i) \langle I_\lambda^2(\mathbf{x}_i)\delta V_i^2\rangle \\ &= \sum_i \sum_{j\neq i} \delta V_i \delta V_j g(\mathbf{x}_i,\mathbf{x}_j) \langle I_\lambda(\mathbf{x}_i)\rangle \langle I_\lambda(\mathbf{x}_j)\rangle \left(1 + \xi_{I_\lambda I_\lambda}(\mathbf{x}_i - \mathbf{x}_j)\right) \\ &\quad + \sum_i g(\mathbf{x}_i,\mathbf{x}_i) \delta V X_{I_\lambda}^2(\mathbf{x}) \int \mathrm{d}L \frac{\mathrm{d}n}{\mathrm{d}L} L^2 \\ &= \int d^3\mathbf{x} \int d^3\mathbf{y}\, g(\mathbf{x},\mathbf{y}) \langle I_\lambda(\mathbf{x})\rangle \langle I_\lambda(\mathbf{y})\rangle \left(1 + \xi_{I_\lambda I_\lambda}(\mathbf{x}-\mathbf{y})\right) \\ &\quad + \int d^3\mathbf{x}\, g(\mathbf{x},\mathbf{x}) X_{I_\lambda}^2(\mathbf{x}) \int \mathrm{d}L \frac{\mathrm{d}n}{\mathrm{d}L} L^2 \\ &= \int d^3\mathbf{x} \int d^3\mathbf{y}\, g(\mathbf{x},\mathbf{y}) \Bigg[\langle I_\lambda(\mathbf{x})\rangle \langle I_\lambda(\mathbf{y})\rangle \left(1 + \xi_{I_\lambda I_\lambda}(\mathbf{x}-\mathbf{y})\right) \\ &\quad + \delta_D(\mathbf{x}-\mathbf{y}) X_{I_\lambda}^2(\mathbf{x}) \int \mathrm{d}L \frac{\mathrm{d}n}{\mathrm{d}L} L^2 \Bigg],\end{aligned} \tag{3.13}$$

where δ_D is the Dirac delta function. Because this is valid for arbitrary functions $g(\mathbf{x},\mathbf{y})$, we can compare the first and last line and find

$$\langle I_\lambda(\mathbf{x}) I_\lambda(\mathbf{y})\rangle = \langle I_\lambda(\mathbf{x})\rangle \langle I_\lambda(\mathbf{y})\rangle \left[1 + \xi_{I_\lambda I_\lambda}(\mathbf{x}-\mathbf{y})\right] + \delta_\mathrm{D}(\mathbf{x}-\mathbf{y}) X_{I_\lambda}^2(\mathbf{x}) \int \mathrm{d}L \frac{\mathrm{d}n}{\mathrm{d}L} L^2\,. \tag{3.14}$$

The second term is the Poisson shot noise from discrete sources, which would not arise from smooth fields.

We follow the same line of reasoning for the galaxy-intensity cross-correlation,

$$\begin{aligned}
&\langle \int d^3\mathbf{x} \int d^3\mathbf{y}\, g(\mathbf{x},\mathbf{y})(1+\delta_{\rm g}(\mathbf{x}))I_\lambda(\mathbf{y})\rangle \\
&= \int d^3\mathbf{x} \int d^3\mathbf{y}\, g(\mathbf{x},\mathbf{y})\langle(1+\delta_{\rm g}(\mathbf{x}))I_\lambda(\mathbf{y})\rangle \\
&= \sum_{i,j} \delta V_i \delta V_j g(\mathbf{x}_i,\mathbf{x}_j)\langle(1+\delta_{\rm g}(\mathbf{x}_i))I_\lambda(\mathbf{x}_j)\rangle \\
&= \sum_i \sum_{j\neq i} \delta V_i \delta V_j g(\mathbf{x}_i,\mathbf{x}_j)\langle(1+\delta_{\rm g}(\mathbf{x}_i))I_\lambda(\mathbf{x}_j)\rangle + \sum_i \delta V_i^2 g(\mathbf{x}_i,\mathbf{x}_i)\langle(1+\delta_{\rm g}(\mathbf{x}_i))I_\lambda(\mathbf{x}_i)\rangle.
\end{aligned} \tag{3.15}$$

The second term is zero if there is no detected galaxy in this voxel or if the line luminosity of this galaxy is zero. If there is a galaxy in the voxel, the value of this is the intensity sourced from that galaxy. Therefore, the expectation value is

$$\begin{aligned}
\delta V\langle(1+\delta_{\rm g}(\mathbf{x}_i))I_\lambda(\mathbf{x}_i)\rangle &= X_{I_\lambda}(\mathbf{x}_i)\int dL \mathcal{P}_{\rm det}(L) L \\
&= \frac{X_{I_\lambda}(\mathbf{x}_i)}{\bar{n}_{\rm det}} \int dL \frac{dn_{\rm det}}{dL} L \\
&=: \left[\frac{X_{I_\lambda}(\mathbf{x}_i)}{\langle n_{\rm g}\rangle}\int dL \frac{dn}{dL} L\right]_{\mathcal{G}_{\rm g}\cap\mathcal{G}_\mathcal{I}},
\end{aligned} \tag{3.16}$$

where $dn_{\rm det}/dL$ is the luminosity function of only detected sources in the galaxy catalog and $\mathcal{P}_{\rm det}(L) = \bar{n}_{\rm det}^{-1} dn_{\rm det}/dL$ is the corresponding PDF. Here, $\mathcal{G}_{\rm g}$ is the set of galaxies contributing to the catalog and $\mathcal{G}_\mathcal{I}$ is the set of galaxies contributing to the intensity map. The integral becomes

$$\begin{aligned}
&\int d^3\mathbf{x} \int d^3\mathbf{y}\, g(\mathbf{x},\mathbf{y})\langle(1+\delta_{\rm g}(\mathbf{x}))I_\lambda(\mathbf{y})\rangle \\
&= \sum_i \sum_{j\neq i} \delta V_i \delta V_j g(\mathbf{x}_i,\mathbf{x}_j)\langle I_\lambda(\mathbf{x}_j)\rangle\,(1+\xi_{{\rm g}I_\lambda}(\mathbf{x}_i-\mathbf{x}_j)) \\
&\quad + \sum_i g(\mathbf{x}_i,\mathbf{x}_i)\delta V_i \left[\frac{X_{I_\lambda}(\mathbf{x}_i)}{\langle n_{\rm g}(\mathbf{x}_i)\rangle}\int dL\frac{dn}{dL}L\right]_{\mathcal{G}_{\rm g}\cap\mathcal{G}_\mathcal{I}} \\
&= \int d^3\mathbf{x}\int d^3\mathbf{y}\, g(\mathbf{x},\mathbf{y})\Bigg\{\langle I_\lambda(\mathbf{y})\rangle\,(1+\xi_{{\rm g}I_\lambda}(\mathbf{x}-\mathbf{y})) \\
&\quad + \delta_{\rm D}(\mathbf{x}-\mathbf{y})\left[\frac{X_{I_\lambda}(\mathbf{x}_i)}{\langle n_{\rm g}(\mathbf{x}_i)\rangle}\int dL\frac{dn}{dL}L\right]_{\mathcal{G}_{\rm g}\cap\mathcal{G}_\mathcal{I}}\Bigg\}.
\end{aligned} \tag{3.17}$$

Thus, the cross-correlation with the galaxy density contrast of detected galaxies is given by

$$\langle I_\lambda(\mathbf{x})(1+\delta_{\rm g}(\mathbf{y}))\rangle = \langle I_\lambda(\mathbf{x})\rangle\,[1+\xi_{I_\lambda{\rm g}}(\mathbf{x}-\mathbf{y})] + \delta_{\rm D}(\mathbf{x}-\mathbf{y})\left[\frac{X_{I_\lambda}(\mathbf{x})}{\langle n(\mathbf{x})\rangle}\int {\rm d}L\frac{{\rm d}n}{{\rm d}L}L\right]_{\mathcal{G}_{\rm g}\cap\mathcal{G}_\mathcal{I}}. \tag{3.18}$$

We define the weighted Fourier transform as

$$\widetilde{\delta I_\lambda}(\mathbf{k}) = \int \mathrm{d}^3\mathbf{x}\, w_{I_\lambda}(\mathbf{x})\delta I_\lambda(\mathbf{x}) e^{i\mathbf{k}\cdot\mathbf{x}} \tag{3.19}$$

with a dimensionless weight $w_{I_\lambda}(\mathbf{x})$, which can, for example, represent a survey footprint. We define the estimator for the intensity power spectrum,

$$\hat{P}_{I_\lambda I_\lambda}(\mathbf{k}) = V_{\mathrm{box}}^{-1}\langle|\widetilde{\delta I_\lambda}(\mathbf{k})|^2\rangle\,, \tag{3.20}$$

where V_{box} is the volume of a cuboid enclosing the survey used to compute the Fourier transform. $\hat{P}_{I_\lambda I_\lambda}(\mathbf{k})$ has the dimension of volume times specific intensity squared. Calculating the weighted Fourier transform of the intensity fluctuations and inserting it into equation (3.20) yields

$$\langle|\widetilde{\delta I_\lambda}(\mathbf{k})|^2\rangle = \int \frac{\mathrm{d}^3\mathbf{k}'}{(2\pi)^3} P_{II}(\mathbf{k}')|\widetilde{W}_{I_\lambda}(\mathbf{k}-\mathbf{k}')|^2 + \int \mathrm{d}^3\mathbf{x}\, w_{I_\lambda}^2(\mathbf{x}) X_{I_\lambda}^2(\mathbf{x}) \int \mathrm{d}L \frac{\mathrm{d}n}{\mathrm{d}L} L^2, \tag{3.21}$$

where the window function is defined as

$$\widetilde{W}_{I_\lambda}(\mathbf{k}) = \int \mathrm{d}^3\mathbf{x}\, e^{i\mathbf{k}\cdot\mathbf{x}} w_{I_\lambda}(\mathbf{x})\langle I_\lambda(\mathbf{x})\rangle \tag{3.22}$$

and $P_{I_\lambda I_\lambda}$ is the power spectrum defined in Eq. (3.2) with units of volume.

We can use the same weighted Fourier transform as defined in Equation (3.19) for the galaxy density contrast, using dimensionless weights w_{g}. We define the estimator for the cross-power spectrum,

$$\hat{P}_{\mathrm{g}I_\lambda}(\mathbf{k}) = V_{\mathrm{box}}^{-1}\,\mathrm{Re}\left\{\langle\widetilde{\delta I_\lambda}(\mathbf{k})\tilde{\delta}_{\mathrm{g}}^*(\mathbf{k})\rangle\right\}, \tag{3.23}$$

where $\mathrm{Re}\{\cdot\}$ denotes the real part of the expression within the curly brackets, and the asterisk is the complex conjugate operator. We find

$$\begin{aligned}\langle\widetilde{\delta I_\lambda}(\mathbf{k})\tilde{\delta}_{\mathrm{g}}^*(\mathbf{k})\rangle = &\int \frac{\mathrm{d}^3\mathbf{k}'}{(2\pi)^3} P_{I_\lambda \mathrm{g}}(\mathbf{k}')\widetilde{W}_{I_\lambda}(\mathbf{k}-\mathbf{k}')\widetilde{W}_{\mathrm{g}}^*(\mathbf{k}-\mathbf{k}') \\ &+ \int \mathrm{d}^3\mathbf{x}\, w_{I_\lambda}(\mathbf{x}) w_{\mathrm{g}}(\mathbf{x}) \left[\frac{X_{I_\lambda}(\mathbf{x})}{\langle n(\mathbf{x})\rangle}\int \mathrm{d}L \frac{\mathrm{d}n}{\mathrm{d}L} L\right]_{\mathcal{G}_{\mathrm{g}}\cap\mathcal{G}_{\mathcal{I}}},\end{aligned} \tag{3.24}$$

where $\widetilde{W}_{\mathrm{g}}(\mathbf{k}) = \int \mathrm{d}^3\mathbf{x}\, e^{i\mathbf{k}\cdot\mathbf{x}} w_{\mathrm{g}}(\mathbf{x})$ is the galaxy window function.

The galaxy power spectrum estimator can similarly be defined as

$$\hat{P}_{\mathrm{gg}}(\mathbf{k}) = V_{\mathrm{box}}^{-1}\langle|\tilde{\delta}_{\mathrm{g}}(\mathbf{k})|^2\rangle\,, \tag{3.25}$$

where

$$\langle|\tilde{\delta}_{\mathrm{g}}(\mathbf{k})|^2\rangle = \int \frac{\mathrm{d}^3\mathbf{k}'}{(2\pi)^3} P_{\mathrm{gg}}(\mathbf{k}')|\widetilde{W}_{\mathrm{g}}(\mathbf{k}-\mathbf{k}')|^2 + \int \mathrm{d}^3\mathbf{x}\, \frac{w_{\mathrm{g}}^2(\mathbf{x})}{\langle n(\mathbf{x})\rangle}\,. \tag{3.26}$$

In this formalism, the intensity units of the power spectrum estimators $\hat{P}_{I_\lambda I_\lambda}$ and $\hat{P}_{I_\lambda \mathrm{g}}$ are contained in the intensity window function, $\widetilde{W}_{I_\lambda}(\mathbf{k})$. Therefore, the power spectra

$P_{I_\lambda I_\lambda}$, $P_{I_\lambda \mathrm{g}}$, and P_{gg} within the integrals of equations (3.21), (3.24), and (3.26) are defined through Equation (3.2) and have units of volume. On large scales, the power spectra $P_{I_\lambda I_\lambda}$, $P_{I_\lambda \mathrm{g}}$, and P_{gg} can be modeled in terms of the matter power spectrum with a linear bias (see Equation (3.7)),

$$P_{AB} = b_A b_B P_{\mathrm{m}}, \tag{3.27}$$

where $P_{\mathrm{m}}(\mathbf{k}) = \langle |\delta_{\mathrm{m}}(\mathbf{k})|^2 \rangle$ is the matter power spectrum. If the intensity is sourced by galaxies with bias b_{source}, the intensity bias equals this bias, $b_{I_\lambda} = b_{\mathrm{source}}$. This assumption breaks down on small scales, where the intensity does not linearly trace the matter density. This can be modeled, for example, using the halo model (Cooray & Sheth, 2002; Wolz et al., 2019; Schaan & White, 2021). Different lines also are emitted by different types of sources, depending on their astrophysical properties, which affects the small-scale cross-correlation signal. This can be modeled in simulations (see Section 2.6).

3.1.1 Effect of Limited Resolution and Noise

LIM surveys typically have lower spatial and spectral resolution than galaxy surveys. Limited observational resolution can be modeled by smoothing the intensity map. If the intensity map $I_\lambda(\mathbf{x})$ is smoothed with a smoothing kernel $D(\mathbf{x})$, the smoothed intensity map in Fourier space is

$$\widetilde{I_{\lambda \mathrm{s}}}(\mathbf{k}) = \widetilde{I_\lambda}(\mathbf{k}) \tilde{D}(\mathbf{k}). \tag{3.28}$$

The factors in the power spectrum estimators become

$$\langle |\widetilde{\delta I_{\lambda \mathrm{s}}}(\mathbf{k})|^2 \rangle = \langle |\widetilde{\delta I_\lambda}(\mathbf{k})|^2 \rangle |\tilde{D}(\mathbf{k})|^2 \,, \tag{3.29}$$

$$\langle \widetilde{\delta I_{\lambda \mathrm{s}}}(\mathbf{k}) \tilde{\delta}^*_{\mathrm{g}}(\mathbf{k}) \rangle = \langle \widetilde{\delta I_\lambda}(\mathbf{k}) \tilde{\delta}^*_{\mathrm{g}}(\mathbf{k}) \rangle \tilde{D}(\mathbf{k}) \,. \tag{3.30}$$

Common smoothing kernels in the LOS direction are a Gaussian kernel, mimicking the line-spread function,

$$D_\parallel^{\mathrm{Gauss}} = \exp\left(-\frac{1}{2} k_\parallel^2 s_\parallel^2\right), \tag{3.31}$$

and top-hat smoothing, representing the binning of intensity into frequency channels,

$$D_\parallel^{\mathrm{top-hat}} = \mathrm{sinc}\left(\frac{1}{2} k_\parallel s_\parallel\right). \tag{3.32}$$

The beam smoothing or point-spread function (PSF) is often modeled as Gaussian smoothing in the angular direction,

$$D_\perp^{\mathrm{Gauss}} = \exp\left(-\frac{1}{2} k_\perp^2 s_\perp^2\right). \tag{3.33}$$

In Equations (3.31), (3.32), and (3.33), $s_\parallel$ and $s_\perp$ define the smoothing lengths parallel and perpendicular to the LOS, respectively, and $k_\parallel = \mu k$ and $k_\perp = \sqrt{1-\mu^2} k$ are the components of the wavevector parallel and perpendicular to the LOS.

Observed line intensity maps are contaminated with noise, $\Delta I_{\lambda,\mathrm{n}}$, containing noise from the instrument and the Poisson noise of individual photons. The noisy line intensity map is $I_{\lambda \mathrm{s,n}}(\mathbf{x}) = I_{\lambda \mathrm{s}}(\mathbf{x}) + \Delta I_{\lambda,\mathrm{n}}(\mathbf{x})$. Assuming that the noise is uncorrelated and has

zero mean, the noise can be characterized by its variance, $\langle \Delta I^2_{\lambda,\mathrm{n}}(\mathbf{x})\rangle = \sigma^2_{I_\lambda}(\mathbf{x})$, where σ_{I_λ} is the standard deviation of the noise in each voxel. This adds a term to the power spectrum estimator,

$$\langle|\widetilde{\delta I_{\lambda\mathrm{s,n}}}(\mathbf{k})|^2\rangle = \langle|\widetilde{\delta I}_{\lambda,\mathrm{s}}(\mathbf{k})|^2\rangle + \delta V \int \mathrm{d}^3\mathbf{x}\, w^2(\mathbf{x})\sigma^2_{I_\lambda}(\mathbf{x})\,. \tag{3.34}$$

The second term still holds for larger than infinitesimal voxel volumes δV.

3.2 Voxel Intensity Distribution

The VID, $\mathcal{B}_{I_\lambda}$, is an estimator for the PDF of line intensities $\mathcal{P}(I_\lambda)$ (e.g., Breysse et al., 2017; Breysse, 2022; Bernal & Kovetz, 2022; Bernal, 2024). The VID at intensity $I_{\lambda,i}$ with bins of width $\Delta I_{\lambda,i}$ in a survey with a total number of voxels N_{vox} is

$$\mathcal{B}_{I_\lambda}(I_{\lambda,i}) = N_{\mathrm{vox}} \int_{I_{\lambda,i}-\Delta I_{\lambda,i}/2}^{I_{\lambda,i}+\Delta I_{\lambda,i}/2} \mathcal{P}_{\mathrm{obs}}(I_\lambda) dI_\lambda. \tag{3.35}$$

Here, $\mathcal{P}_{\mathrm{obs}}(I_\lambda)$ is the observable PDF of intensities, which is a combination of noise of the survey $\mathcal{P}_{\mathrm{noise}}(I_\lambda)$ and the PDF of astrophysical sources $\mathcal{P}_{\mathrm{astro}}(I_\lambda)$.

The total intensity of a voxel is the sum of the intensities of all contributing sources in that voxel. For example, the intensity PDF of a voxel containing two sources with intensity PDFs $\mathcal{P}_{s_1}(I_\lambda)$ and $\mathcal{P}_{s_2}(I_\lambda)$ is

$$\mathcal{P}_{s_1+s_2}(I_\lambda) = \int_0^\infty dI'_\lambda \mathcal{P}_{s_1}(I'_\lambda)\mathcal{P}_{s_2}(I_\lambda - I'_\lambda) = (\mathcal{P}_{s_1} * \mathcal{P}_{s_2})(I_\lambda). \tag{3.36}$$

Hence, the PDF of the total intensity of N_{s} sources within a voxel is the convolution the PDFs of the N_{s} sources,

$$\mathcal{P}_{\sum_{j=1}^{N_{\mathrm{s}}} s_j}(I_\lambda) = (\mathcal{P}_{s_1} * ... * \mathcal{P}_{s_{N_s}})(I_\lambda). \tag{3.37}$$

Because this is easier to treat in Fourier space, we introduce its Fourier counterpart, the characteristic function,

$$\widetilde{\mathcal{P}}(\tau) = \int_0^\infty dI_\lambda \mathcal{P}(I_\lambda) e^{-iI_\lambda\tau} = \langle e^{-iI_\lambda\tau}\rangle. \tag{3.38}$$

Because of the integration limits from 0 to ∞, this is a one-sided Fourier transform, or a Laplace transform with a purely imaginary kernel. Then,

$$\widetilde{\mathcal{P}}_{\sum_{j=1}^{N_{\mathrm{s}}} s_j}(\tau) = \prod_{j=1}^{N_{\mathrm{s}}} \widetilde{\mathcal{P}_{s_j}}(\tau). \tag{3.39}$$

To account for the clustering of sources, we assume that the intensity is sourced only in dark matter halos and model the intensity as a function of the host dark matter halo mass M, following Breysse (2022). We also assume that dark matter halos are linearly biased tracers of the underlying matter density field. We can define a halo mass bin

$[M, M + dM]$ that is narrow enough that either zero or one halo within this mass range is in a voxel, i.e., $N \in \{0, 1\}$. A voxel containing one dark matter halo within this mass bin has the intensity PDF $\mathcal{P}_1(I_\lambda|M)$. The intensity PDF of a voxel containing either zero or one dark matter halo within the mass bin is

$$\mathcal{P}(I_\lambda|M) = \mathcal{P}(N = 0)\delta_D(I_\lambda) + \mathcal{P}(N = 1)\mathcal{P}_1(I_\lambda|M). \tag{3.40}$$

Writing this in terms of the characteristic functions, and using $\mathcal{P}(N = 0) = 1-\mathcal{P}(N = 1)$, we obtain

$$\widetilde{\mathcal{P}}(\tau|M) = 1 + \mathcal{P}(N = 1)\left(\widetilde{\mathcal{P}_1}(\tau|M) - 1\right). \tag{3.41}$$

Assuming that the distribution of the number of dark matter halos per voxel follows a Poisson distribution with mean $\bar{N}(M) \ll 1$, the probability of having one dark matter halo in the voxel is

$$\mathcal{P}(N = 1) \simeq \bar{N}(M) = \frac{dn}{dM} V_{\rm vox}\left(1 + b(M)\delta_m\right) dM, \tag{3.42}$$

where $\frac{dn}{dM}$ is the halo mass function, $V_{\rm vox}$ is the voxel volume, $b(M)$ is the dark matter halo bias, and δ_m is the matter density contrast at the voxel location. Using $\bar{N}(M) \ll 1$, we can approximate equation (3.41) as the Taylor expansion of the exponential function,

$$\widetilde{\mathcal{P}}(\tau|M) \simeq \exp\left[\frac{dn}{dM} V_{\rm vox}\left(1 + b(M)\delta_m\right)\left(\widetilde{\mathcal{P}_1}(\tau|M) - 1\right) dM\right]. \tag{3.43}$$

The total intensity is the sum of the intensities of the halos in all mass bins, so that

$$\begin{aligned}
\widetilde{\mathcal{P}}(\tau) &= \prod_M \widetilde{\mathcal{P}}(\tau|M) \\
&= \exp\left[\sum_M dM \frac{dn}{dM} V_{\rm vox}\left(1 + b(M)\delta_m\right)\left(\widetilde{\mathcal{P}_1}(\tau|M) - 1\right)\right] \\
&\simeq \exp\left[\int dM \frac{dn}{dM} V_{\rm vox}\left(1 + b(M)\delta_m\right)\left(\widetilde{\mathcal{P}_1}(\tau|M) - 1\right)\right],
\end{aligned} \tag{3.44}$$

where we take the limit $dM \to 0$ to approximate the sum as an integral in the last step. We can separate this result into two components. The first component gives the PDF in the absence of clustering,

$$\begin{aligned}
\widetilde{\mathcal{P}}_{\rm un}(\tau) &= \exp\left[\int dM \frac{dn}{dM} V_{\rm vox}\left(\widetilde{\mathcal{P}_1}(\tau|M) - 1\right)\right] \\
&=: \exp\left[\int dM p(\tau|M)\right].
\end{aligned} \tag{3.45}$$

The second component accounts for clustering,

$$\widetilde{\mathcal{P}}_{\rm cl}(\tau) = \exp\left[\delta_{\rm m} \int dM b(M) p(\tau|M)\right]. \tag{3.46}$$

To obtain the PDF of all voxel intensities, we need to average over all δ_m. Assuming that δ_m is Gaussian distributed, we can use $\langle \exp(x) \rangle = \exp(\langle x^2 \rangle / 2)$ to obtain

$$\widetilde{\mathcal{P}}_\mathrm{cl}(\tau) = \exp\left[\frac{\sigma_\mathrm{m}^2}{2}\left(\int dM b(M) p(\tau|M)\right)^2\right], \tag{3.47}$$

where

$$\sigma_\mathrm{m}^2 = \langle \delta_\mathrm{m}^2 \rangle = \int P_\mathrm{m}(\mathbf{k}) \left|W_\mathrm{vox}(\mathbf{k})\right|^2 \frac{d^3\mathbf{k}}{(2\pi)^3} \tag{3.48}$$

is the variance of the matter density contrast within a voxel with window function $W_\mathrm{vox}(\mathbf{k})$.

The observed intensity is the sum of the astrophysical intensity and the noise, so that the PDF of observed intensity can be written as the Fourier/Laplace transform of

$$\widetilde{\mathcal{P}}_\mathrm{obs} = \widetilde{\mathcal{P}}_\mathrm{astro}\widetilde{\mathcal{P}}_\mathrm{noise} = \widetilde{\mathcal{P}}_\mathrm{un}\widetilde{\mathcal{P}}_\mathrm{cl}\widetilde{\mathcal{P}}_\mathrm{noise}. \tag{3.49}$$

While the power spectrum only probes the first two moments of the luminosity function, the VID is sensitive to its entire shape and amplitude. Combining the VID and power spectrum can help break degeneracies between cosmology and astrophysics (Sato-Polito & Bernal, 2022; Sabla et al., 2024). Because $\mathcal{P}_\mathrm{cl}$ is directly related to the PDF of the matter distribution in the universe, the VID can constrain that PDF beyond the Gaussian information encoded in the power spectrum. See Bernal (2024) for a model of the VID including extended emission, line broadening, and limited resolution.

3.3 Stacking

A simple, but not necessarily easy, way of obtaining a detection of the astrophysical intensity in a noisy line intensity map is stacking (e.g., Dunne et al., 2025). This involves calculating the mean, median, or biweight location (Beers et al., 1990) of the line intensity map around the positions of detected galaxies, thereby reducing the noise. The stacked signal depends on the average line intensity of the stacked galaxies and the clustering and intensity of the surrounding dark matter halos.

Stacking is closely related to the cross-correlation function between galaxies and the line intensity. A mean stacking estimator in three dimensions can be written as

$$\hat{S}_{I_\lambda \mathrm{g}}(\Delta\mathbf{x}) = \frac{1}{N_\mathrm{g}} \sum_{i=1}^{N_\mathrm{g}} I_\lambda(\mathbf{x}_i + \Delta\mathbf{x}) - \langle I_\lambda \rangle, \tag{3.50}$$

where $\mathbf{x}_i$ is the location of the galaxy i and N_g is the number of galaxies in the stack. This is equivalent to

$$\hat{S}_{I_\lambda \mathrm{g}}(\Delta\mathbf{x}) = \frac{1}{N_\mathrm{g}} \sum_{i=1}^{N_\mathrm{g}} \int d^3\mathbf{x}\, I_\lambda(\mathbf{x} + \Delta\mathbf{x}) \delta_\mathrm{D}^{(3)}(\mathbf{x} - \mathbf{x}_i) - \langle I_\lambda \rangle. \tag{3.51}$$

Because the galaxy distribution is the sum of the Dirac delta functions at each galaxy position,

$$1 + \delta_\mathrm{g}(\mathbf{x}) = \frac{1}{N_\mathrm{g}} \sum_{i=1}^{N_\mathrm{g}} \delta_\mathrm{D}^{(3)}(\mathbf{x} - \mathbf{x}_i), \tag{3.52}$$

we can rewrite equation (3.51) as

$$\begin{aligned}\hat{S}_{I_\lambda \mathrm{g}}(\Delta\mathbf{x}) &= \int d^3\mathbf{x}\, I_\lambda\left(\mathbf{x}+\Delta\mathbf{x}\right)\left(1+\delta_{\mathrm{g}}(\mathbf{x})\right) - \langle I_\lambda\rangle \\ &= \int d^3\mathbf{x}\, I_\lambda\left(\mathbf{x}+\Delta\mathbf{x}\right)\delta_{\mathrm{g}}(\mathbf{x}).\end{aligned} \tag{3.53}$$

This is simply the cross-correlation of the intensity and galaxy density contrast.

Similarly, an angular stacking estimator using narrowband or pseudo-narrowband images with the intensity I can be constructed as

$$\hat{S}_{I\mathrm{g}}(\boldsymbol{\theta}) = \frac{1}{N_{\mathrm{g}}}\sum_{i=1}^{N_{\mathrm{g}}} I(\mathbf{y}_i+\boldsymbol{\theta}) - \langle I\rangle, \tag{3.54}$$

where $\mathbf{y}_i$ are the angular coordinates of the galaxies and $\langle I\rangle$ is the mean intensity within this narrowband image. Following the same steps as above, we find

$$\hat{S}_{I\mathrm{g}}(\boldsymbol{\theta}) = \int d^2\mathbf{y}\, I\left(\mathbf{y}+\boldsymbol{\theta}\right)\delta_{\mathrm{g}}(\mathbf{y}), \tag{3.55}$$

which is the angular cross-correlation function of the intensity and galaxy density contrast.

3.4 Other Summary Statistics

Minkowski functionals (Minkowski, 1903) describe the topology, specifically the size and connectivity, of subsets of a vector space. They can be used to describe the large-scale structure of matter, as probed by line intensity maps (Spina et al., 2021). One can define subsets of the space by drawing isocontour maps, and calculate the Minkowski functionals of these isocontour subsets as a function of the intensity threshold. Comparing this to simulations can, for example, reveal the relationship between dark matter halo mass and its intensity.

Line intensity maps can also be modeled directly at the map (or "field") level using an analytic forward model based on perturbation theory (Obuljen et al., 2023). This can be used for cosmological inference and for quickly generating mock line intensity maps. The difference between the modeled intensity field and the actual intensity map can yield additional information, such as on the presence of the FoG effect.

Using wavelet scattering transforms (WSTs; Mallat, 2012) for cosmology involves convolving the line intensity map with a family of wavelets, and calculating the spatial average of the complex modulus (Cheng & Ménard, 2021; Chung, 2022). These scattering coefficients provide information on the morphology of the underlying field beyond the Gaussian statistics of the power spectrum. They also promise much finer parameter constraints than the combination of the power spectrum and VID.

3.5 Observational Difficulties

Because LIM surveys span a wide frequency range and employ different observing strategies, the observational difficulties vary from survey to survey. However, there are common

problems across many surveys. For example, the line intensity maps contain not only photons from the target emission line, but also photons from continuum foreground sources, background sources, and other emission lines at different redshifts. Instrumental artifacts can also contaminate the line intensity maps. This section briefly describes foreground and interloper contamination in the context of HETDEX and a general method to clean spectra from foreground and systematic residuals.

3.5.1 Foregrounds

Line intensity maps observe the integrated intensity in each voxel, regardless of its origin. This includes sources between the target signal and the observer - called foregrounds, and background sources, such as the CMB in CO observations. The types of foreground and background sources depend on the observed wavelength or frequency range.

Continuum foregrounds, which contaminate 21 cm, CO, and [C II] observations, are spectrally smooth, while the target line intensity is expected to have strong spectral structure. Therefore, continuum foregrounds can be modeled by fitting a smooth curve to the intensity along the LOS, and removed from the map (Kovetz et al., 2017). However, some radio interferometers suffer from mode mixing, which moves spectrally smooth contamination into other regions of Fourier counterpart of the map. Most of this contamination is contained in a well-defined region in the $k_{\parallel} - k_{\perp}$ plane, which can be removed for the power spectrum analysis. While common in other LIM surveys, this type of bright, spectrally smooth foreground and mode mixing is less relevant in optical observations. Instead, the dominant foreground, called "sky", is smooth in the angular direction.

Sky Subtraction in Optical Observations

Optical observations are contaminated with "sky" emission, which is a combination of zodiacal light, aurora, airglow, diffuse Galactic light, and emission from the Galactic warm interstellar medium (e.g., Wyse & Gilmore, 1992). The sky spectrum in the optical wavelength range of VIRUS consists of a continuum component and emission lines. Some emission lines in the "sky" emission observed by HETDEX are caused by street lamps in West Texas, such as the 4358 Å and 5461 Å emission lines of mercury. Neutral nitrogen in the atmosphere also emits a line doublet at 5198 Å and 5200 Å.

Assuming that the sky foreground is homogeneous over the focal plane, one can estimate the sky spectrum by averaging the spectra of "empty" fibers without detected objects, using the biweight location (Beers et al., 1990) or a different robust estimator of the central location of the intensity distribution, The estimated sky spectrum, which is subtracted from all spectra, includes the mean intensity in each wavelength slice by construction. This removes large-scale fluctuations in the angular direction. The contribution of the line intensity to the estimated sky foreground can be modeled as a smoothed intensity map. The kernel corresponds to the region used for the sky spectrum estimation; for example, a two-dimensional spherical top-hat kernel in the plane perpendicular to the LOS. The size of the kernel is the size of the area used for the sky spectrum estimation, for example, a circle with radius s_{f}. Thus, the contribution of the line intensity to the

sky spectrum is

$$\tilde{I}_{\lambda,\mathrm{sky}}(\mathbf{k}) = \tilde{I}_{\lambda}(\mathbf{k})\tilde{D}_{\mathrm{sky}}(\mathbf{k}), \tag{3.56}$$

where $\tilde{D}_{\mathrm{sky}}(\mathbf{k})$ is the Fourier transform of the two-dimensional spherical top-hat kernel, given by

$$\tilde{D}_{\mathrm{sky}}(\mathbf{k}) = \frac{2J_1\left(s_{\mathrm{f}}\sqrt{k_a^2 + k_b^2}\right)}{s_{\mathrm{f}}\sqrt{k_a^2 + k_b^2}}\,, \tag{3.57}$$

where a and b denote the directions perpendicular to the LOS, and J_1 is the Bessel function of the first kind and first order. Because $\tilde{D}_{\mathrm{sky}}(\mathbf{k})$ is real-valued, we obtain

$$\langle|\tilde{I}_{\lambda}(\mathbf{k}) - \tilde{I}_{\lambda,\mathrm{sky}}(\mathbf{k})|^2\rangle = \langle|\tilde{I}_{\lambda}(\mathbf{k})|^2\rangle\left[1 - 2\tilde{D}_{\mathrm{sky}}(\mathbf{k}) + \tilde{D}^2_{\mathrm{sky}}(\mathbf{k})\right], \tag{3.58}$$

and

$$\left\langle\left[\tilde{I}_{\lambda}(\mathbf{k}) - \tilde{I}_{\lambda,\mathrm{sky}}(\mathbf{k})\right]\delta_{\mathrm{g}}^{*}\right\rangle = \langle\tilde{I}_{\lambda}(\mathbf{k})\delta_{\mathrm{g}}^{*}\rangle\left[1 - \tilde{D}_{\mathrm{sky}}(\mathbf{k})\right]. \tag{3.59}$$

This shows that the power spectrum is suppressed on scales larger than the scale used for the sky spectrum estimation s_{f}. Figure 3.1 shows an example sky spectrum of a HETDEX observation and an illustration of the effect of the sky subtraction on large-scale fluctuations of a simulated line intensity map.

For the main purposes of HETDEX, i.e., detecting LAEs, the sky subtraction is performed for each amplifier individually (see Figure 1.3 for a visualization of the layout of VIRUS). However, this effectively removes the mean spectrum, including the mean Lyα emission at each wavelength within a $\simeq 13'' \times 51''$ region. This corresponds to a comoving area of $\simeq 400\,\mathrm{kpc} \times 1600\,\mathrm{kpc}$ at $z = 3$. Therefore, LIM needs a different approach.

HETDEX employs a second sky subtraction technique - the "full-frame" sky subtraction, which we use for LIM (see Gebhardt et al., 2021; Lujan Niemeyer et al., 2022a, for details). Here, the sky spectrum is estimated from all "empty" fibers of the total $\sim$ 35k fibers in an exposure. Because every spectrograph, amplifier, and fiber has a different wavelength-dependent throughput, and some have slightly different spectral resolutions, one must accurately model these differences when estimating and subtracting the sky spectrum. The full-frame sky subtraction is less stable than the local, amplifier-based sky subtraction because of the differences between the amplifiers. Because of the high accuracy needed for LIM, it is necessary to apply empirical corrections to the sky-subtracted spectra, as described in Chapter 7.

3.5.2 Interloper Emission Lines

Similar to foregrounds, other emission lines at different redshifts can fall into the same observed frequency range as the target line. For example, [C II] observations also contain various rotational emission lines of CO (Kovetz et al., 2017). The sources of interloper lines are also clustered, so that an auto-correlation contains their clustering signal.

To mitigate the impact of interloper lines on the line intensity map, Breysse et al. (2015) suggest masking out the brightest voxels. This technique assumes that interloper sources reside at a lower redshift than the target sources, and are therefore more massive and brighter. Thus, they should produce the brightest voxels. Instead of blind masking,

one can also mask known objects from galaxy catalogs that could contaminate the signal (Sun et al., 2018; Van Cuyck et al., 2023). Bernal & Baleato Lizancos (2025) use the correlation between the interloper emission with complementary tracers at the interloper redshift to remove the contamination from the line intensity map. The same technique was previously suggested for analyzing galaxy clustering (Farrow et al., 2021).

Because the sources of interloper lines reside at a different redshift, their clustering is mapped onto wrong spatial scales if one assumes the redshift of the target line. Assuming a spatially flat universe, the apparent coordinates $\tilde{x}_\parallel$ and $\tilde{\mathbf{x}}_\perp$ of an interloper at redshift z_i are related to the true interloper coordinates $x_\parallel$ and $\mathbf{x}_\perp$ by

$$\tilde{x}_\parallel = \frac{H(z_i)}{H(z_t)}\frac{1+z_t}{1+z_i}x_\parallel \quad \text{and} \quad \tilde{\mathbf{x}}_\perp = \frac{D_A(z_t)}{D_A(z_i)}\mathbf{x}_\perp, \tag{3.60}$$

where z_t is the redshift of the target line (Lidz & Taylor, 2016; Leung et al., 2017). This produces a strongly anisotropic power spectrum or correlation function of the interloper contaminant. It is analogous to the Alcock-Paczyński effect, which arises from assuming the wrong cosmology rather than the wrong redshift (Alcock & Paczyński, 1979). In a sensitive enough experiment, the anisotropic part of the power spectrum can be fitted and subtracted to remove the interloper contamination (Lidz & Taylor, 2016).

The observed wavelength range of HETDEX at 3470 Å − 5540 Å observes the Lyα line at redshifts $z \simeq 2-3.5$ as well as the [O II] emission line doublet at a restframe wavelength of $\bar{\lambda}_{[\mathrm{OII}]} = 3727\,\text{Å}$ at $z \leq 0.47$. While other emission lines also fall into this wavelength range, [O II] emission is the dominant interloper for HETDEX. Because [O II] emitters are bright and most are identified by HETDEX, targeted masking of the affected spectra is an effective approach for HETDEX Lyα intensity mapping (Chapter 7). Because we measure the LAE-Lyα intensity cross-power spectrum, the interloper lines only contribute to the noise, not the signal.

3.5.3 Foreground and Systematics Cleaning

Foregrounds and other systematics often cannot fully be subtracted with the above-mentioned techniques. To remove residual contributions, Cunnington et al. (2023a) use a Principal Component Analysis (PCA), which we also apply in Chapter 7. Carucci et al. (2024) show that a more sophisticated technique, which involves applying a scale-dependent PCA, can clean their data with less signal loss. However, we limit this discussion to the standard PCA technique used in Chapter 7. This method assumes that the foreground residuals dominate the signal of the line intensity map and that they are correlated in frequency/wavelength. PCA calculates the orthonormal basis of the data such that the variance along the first basis vector is highest, and the following basis vectors are sorted in decreasing order of variance. Although removing the components of the data vectors along the highest-variance basis vectors should remove mostly foreground and other systematic residuals, some signal will also be removed. This must be quantified and corrected in the analysis.

We prepare the spectra for PCA by subtracting the mean spectrum and normalizing,

$$X_i(\lambda) = \frac{I_i(\lambda) - \langle I(\lambda)\rangle}{\sigma_I(\lambda)}, \tag{3.61}$$

where the subscript i refers to the individual spectra and $\sigma_I(\lambda)$ is the standard deviation of the intensity as a function of wavelength. A spectrum can be represented as a vector,

$$\mathbf{X}_i = \sum_{j=1}^{N_\lambda} X_i(\lambda_j)\hat{\mathbf{e}}_j^\lambda = \left(X_i(\lambda_1), X_i(\lambda_2), \ldots, X_i(\lambda_{N_\lambda})\right)^\top, \tag{3.62}$$

where N_λ is the number of wavelength bins per spectrum, and the normalized basis vector $\hat{\mathbf{e}}_j^\lambda$ is 1 in the jth wavelength bin and 0 otherwise. The jth value of $\mathbf{X}_i$ is the normalized intensity in the jth wavelength bin, $X_i(\lambda_j)$. We can define an $N_{\rm spec} \times N_\lambda$ matrix, $\mathbf{X}$, which contains the $N_{\rm spec}$ vectors $\mathbf{X}_i$ as its rows. PCA calculates its covariance matrix, $\mathbf{X}^\top\mathbf{X}$. The eigenvectors of the covariance matrix are called weight vectors, and form an orthonormal basis, $\hat{\mathbf{e}}_j^{\rm PCA}$, so that

$$\mathbf{X}_i = \sum_{j=1}^{N_\lambda} y_{ij}\hat{\mathbf{e}}_j^{\rm PCA}. \tag{3.63}$$

The weight vectors are arranged in an order of decreasing variance of y_{ij}, such that $\mathrm{var}(y_{i0}) > \mathrm{var}(y_{i1}) > ... > \mathrm{var}(y_{iN_\lambda})$. Assuming that the first $N_{\rm PC}$ principle components are dominated by systematics or foreground emission, i.e., that these cause most of the variance of the spectra, we can set these to zero,

$$\mathbf{X}_i^{\rm PCA} = \sum_{j=N_{\rm PC}+1}^{N_\lambda} y_{ij}\hat{\mathbf{e}}_j^{\rm PCA}. \tag{3.64}$$

To obtain the cleaned spectra, we perform an inverse coordinate transformation,

$$\mathbf{X}_i^{\rm PCA} = \sum_{k=1}^{N_\lambda}\sum_{j=N_{\rm PC}+1}^{N_\lambda} y_{ij}\left(\hat{\mathbf{e}}_k^{\lambda\top}\hat{\mathbf{e}}_j^{\rm PCA}\right)\hat{\mathbf{e}}_k^\lambda, \tag{3.65}$$

so that

$$X_i^{\rm PCA}(\lambda_k) = \sum_{j=N_{\rm PC}+1}^{N_\lambda} y_{ij}\left(\hat{\mathbf{e}}_k^{\lambda\top}\hat{\mathbf{e}}_j^{\rm PCA}\right), \tag{3.66}$$

and undo the normalization,

$$I_i^{\rm PCA}(\lambda) = X_i^{\rm PCA}(\lambda)\sigma_I(\lambda). \tag{3.67}$$

The first $N_{\rm PC}$ principal components likely contain some astrophysical signal from the target line. The power spectra calculated from the cleaned spectra $I_i^{\rm PCA}(\lambda)$ are therefore damped. The signal loss of the cross-power spectrum of a galaxy catalog with the cleaned intensity map can be quantified with the transfer function (e.g., Cunnington et al., 2023b; Chen, 2025),

$$\mathcal{T}(\mathbf{k}) := \frac{\hat{P}_{\rm clean}}{\hat{P}_{\rm true}} = \frac{\langle\widetilde{\delta I}^{\rm PCA}(\mathbf{k})\widetilde{\delta_{\rm g}}^*(\mathbf{k})\rangle}{\langle\widetilde{\delta I}^{\rm astro}(\mathbf{k})\widetilde{\delta_{\rm g}}^*(\mathbf{k})\rangle}, \tag{3.68}$$

where $\hat{P}_{\rm clean}$ is the cross-power spectrum estimator of the cleaned intensity map and the galaxy catalog, and $\hat{P}_{\rm true}$ is the underlying cross-power spectrum one would obtain in the

absence of foregrounds, systematics, and cleaning. The transfer function can be estimated using a mock intensity map and a corresponding correlated mock galaxy catalog. We start by inserting the real intensity into the mock data, so that the combined spectra are $I_i^{\rm mock+data}(\lambda) = I_i^{\rm data}(\lambda) + I_i^{\rm mock}(\lambda)$. One then performs the PCA on $I_i^{\rm mock+data}(\lambda)$ and calculates the PCA-cleaned, mock-only contribution,

$$I_i^{\rm mock,PCA}(\lambda) = \left[I_i^{\rm mock+data}(\lambda)\right]^{\rm PCA} - I_i^{\rm PCA}(\lambda). \tag{3.69}$$

Alternatively, one can calculate the weight vectors, $\hat{\mathbf{e}}_j^{\rm PCA}$ from $I_i^{\rm mock+data}(\lambda)$, and apply the coordinate transforms in Equations (3.63) and (3.65) to the mock-only data, $I_i^{\rm mock}(\lambda)$, to obtain $I_i^{\rm mock,PCA}(\lambda)$. We choose the second option in Chapter 7.

The transfer function can be estimated as

$$\hat{\mathcal{T}}(\mathbf{k}) = \frac{\langle \widetilde{\delta I}_{\rm mock}^{\rm PCA}(\mathbf{k}) \widetilde{\delta}_{\rm g,mock}^{*}(\mathbf{k}) \rangle}{\langle \widetilde{\delta I}_{\rm mock}(\mathbf{k}) \widetilde{\delta}_{\rm g,mock}^{*}(\mathbf{k}) \rangle}. \tag{3.70}$$

As $N_{\rm PC}$ increases, more astrophysical signal is removed, and the transfer function approaches zero. Cunnington et al. (2023b) show empirically that the signal loss of the intensity auto-power spectrum is the same as for the cross-power spectra, although the cleaned line intensity map contributes to the power spectrum twice. This is intuitively explained in MeerKLASS Collaboration et al. (2025) and formally derived in Chen (2025).

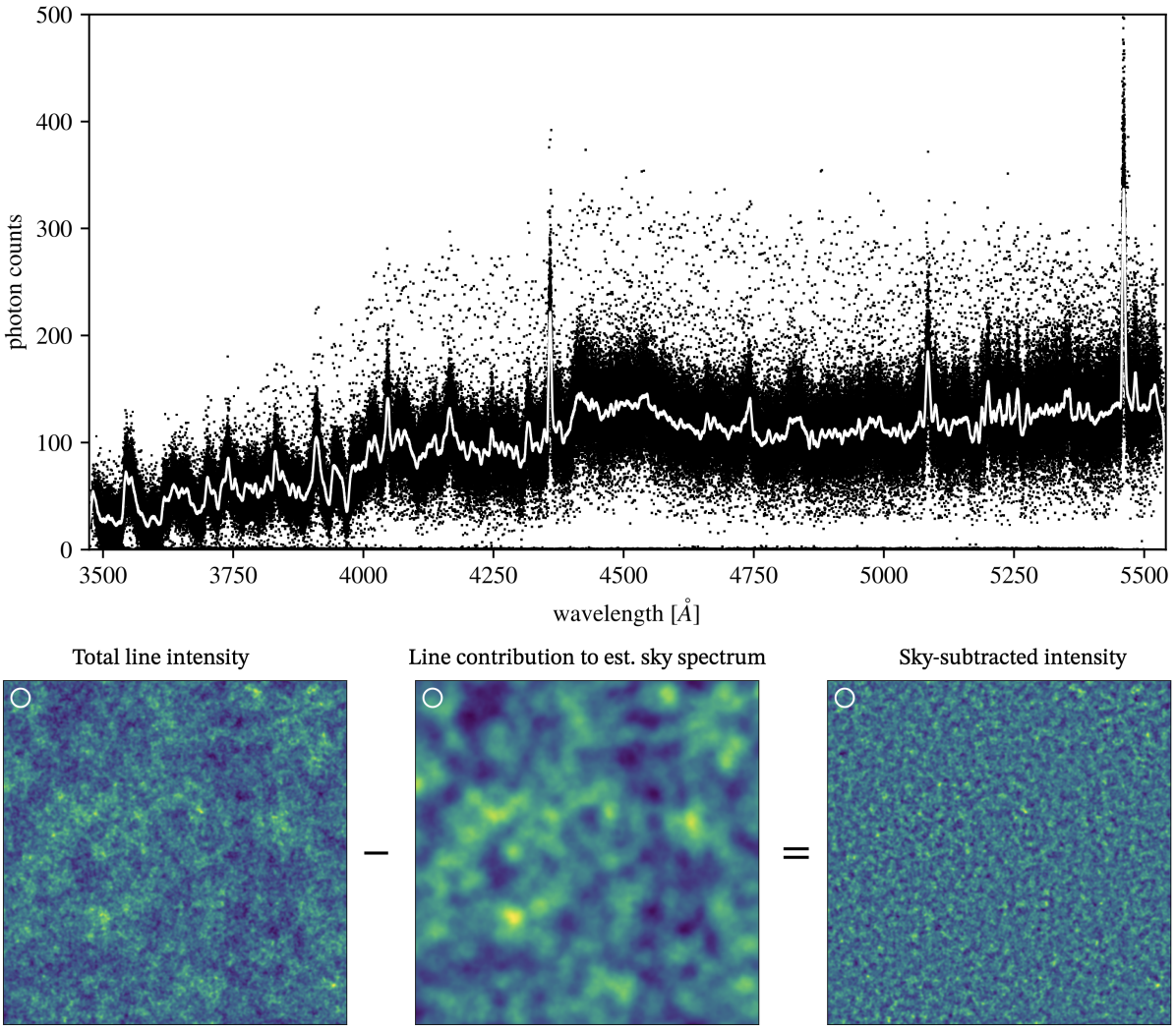

Figure 3.1: The top panel shows an example full-frame sky spectrum in white, estimated from all spectra of a HETDEX observation, a subset of which is shown as black dots. The bottom panels show the effect of the sky subtraction on a simulated redshift slice of a line intensity map. The bottom left panel shows the complete line intensity map and a circle showing the area over which the sky spectrum is estimated. The bottom middle panel shows the contribution of the signal line intensity to the estimated sky spectrum. The bottom right panel shows the sky-subtracted slice of the line intensity map. The fluctuations on larger scales than the sky subtraction are removed.

Chapter 4

Lyman-α Halos around [O III]-Selected Galaxies in HETDEX

The content of this chapter has been published in Lujan Niemeyer et al. (2022b).

We present extended Lyman-α (Lyα) emission out to 800 kpc of 1034 [O III]-selected galaxies at redshifts $1.9 < z < 2.35$ using the Hobby-Eberly Telescope Dark Energy Experiment (HETDEX). The locations and redshifts of the galaxies are taken from the 3D-HST survey. The median-stacked surface brightness profile of Lyα emission of the [O III]-selected galaxies agrees well with that of 968 bright Lyα-emitting galaxies (LAEs) at $r > 40$ kpc from the galaxy centers. The surface brightness in the inner parts ($r < 10$ kpc) around the [O III]-selected galaxies, however, is ten times fainter than that of the LAEs. Our results are consistent with the notion that photons dominating the outer regions of the Lyα halos are not produced in the central galaxies but originate outside of them.

4.1 Introduction

Extended Lyman-α (Lyα) emission around star-forming galaxies without an active galactic nucleus (AGN) has been found around Lyman-break galaxies (LBGs; e.g.; Steidel et al., 2011; Kusakabe et al., 2022; Ouchi, 2019, for a review on LAEs) and Lyα-emitting galaxies (LAEs; e.g.; Wisotzki et al., 2016; Kikuchihara et al., 2022a). One source of Lyα photons is the local recombination of hydrogen atoms ionized by photons from young, massive stars in star-forming regions. After their escape from the interstellar medium (ISM), Lyα photons will be scattered by neutral hydrogen atoms in the circumgalactic medium (CGM) and intergalactic medium (IGM). Hydrogen atoms in the CGM and IGM can also be ionized by photons from more distant AGN or star-forming regions, called the ultraviolet (UV) background, and recombine to emit Lyα photons ("fluorescence"; e.g.; Gould & Weinberg, 1996). Lyα photons from satellite galaxies (Mas-Ribas et al., 2017) and collisional excitation of hydrogen atoms in cooling gas ("cooling radiation";

e.g.; Haiman et al., 2000) can add to the extended Lyα emission.

Because the contribution of scattered photons from the central galaxy to the halo depends on the galaxy's Lyα emission, comparing the Lyα surface brightness (SB) profiles of galaxies with different intrinsic Lyα luminosities or escape fractions can probe the origin of Lyα halos. Because LAEs are selected using their large Lyα equivalent width (EW), they comprise a biased subset of high-redshift galaxies that have a large Lyα escape fraction along the line of sight (LOS). Galaxies selected via other methods such as LBGs or via their rest-frame optical emission lines may have similar physical properties, but with smaller Lyα escape fractions than LAEs. Erb et al. (2016), Hathi et al. (2016), Trainor et al. (2016, 2019), and Reddy et al. (2022) argue that LAEs have different properties than other star-forming galaxies, such as less dust and metal content, lower star-formation rates (SFR) and stellar masses, and higher H I covering fractions. Conversely, Hagen et al. (2016) and Shimakawa et al. (2017) report no statistical difference between the properties of the samples of LAEs and rest-frame-optical emission-line galaxies except at high stellar masses. Hence, comparing the Lyα halo profiles of LAEs with those of rest-frame-optical emission line galaxies can shed light on the emission sources and mechanisms of Lyα halos.

We compare the median-stacked Lyα SB profile of 1034 galaxies at $1.9 < z < 2.35$ selected via their rest-frame optical emission lines in the 3D-HST survey (Brammer et al., 2012; Momcheva et al., 2016; Bowman et al., 2019, 2020) with that of LAEs at $1.9 < z < 3.5$ detected in the Hobby-Eberly Telescope Dark Energy Experiment (HETDEX; Hill et al., 2021; Gebhardt et al., 2021; Lujan Niemeyer et al., 2022a, hereafter LN22). We use integral-field spectroscopic data from HETDEX to extract the Lyα SB profiles.

We adopt a flat Λ-cold-dark-matter cosmology with $H_0 = 67.37\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ and $\Omega_{\mathrm{m},0} = 0.3147$ (Planck Collaboration et al., 2020a). All distances are in units of physical kpc unless noted otherwise.

4.2 Data and Galaxy Samples

4.2.1 HETDEX Data

We use spectra from the HETDEX survey (Gebhardt et al., 2021), specifically the internal data release 3. The survey uses the VIRUS instrument on the 10−m Hobby-Eberly Telescope (HET). See Hill et al. (2021) for details.

VIRUS consists of up to 78 integral-field unit fiber arrays (IFUs), each of which contains 448 1.5″-diameter fibers and covers $51'' \times 51''$ on the sky. The fibers from each IFU are fed to a low-resolution ($R \simeq 800$) spectrograph covering $3500 - 5500\,\text{Å}$. The IFUs with $\simeq$ 35k total fibers are distributed on a grid with 100″ spacing throughout the 18′ diameter of the telescope's field of view. Each HETDEX observation comprises three 6-minute exposures, which are dithered to fill in gaps between the fibers. Because the gaps between the IFUs remain in an individual observation, the filling factor is $\simeq 1/4.6$.

We use the full-frame sky-subtracted data (details in Gebhardt et al. (2021); LN22). This sky-subtraction method measures the sky emission from the entire 18′-diameter field of view of VIRUS to ensure that extended emission on the scale of an IFU or larger is

not removed along with the sky model. The full-frame sky subtraction in the internal HETDEX DR 3 has some differences to that in DR 2, which is used in LN22. Instead of roughly 75% of the total fibers with the lowest continuum emission, only 50% are used for the sky estimate. This helps prevent the oversubtraction of continuum emission due to unresolved sources. To be more conservative, the smooth background subtraction within a six fibers by 600 Å window is omitted. These changes do not affect our measurement because we perform a local continuum subtraction. As expected, the Lyα SB profiles of the LAEs using the data from DR 3 and DR 2 and the same stacking procedure are very similar. We mask the wavelength regions around the brightest sky emission lines to avoid residuals associated with this component.

4.2.2 [O III]-Galaxy Sample

Our [O III]-galaxy sample is drawn from 3D-HST (Brammer et al., 2012; Momcheva et al., 2016), an *HST* Treasury program which used 2-orbit exposures with the WFC3 G141 grism to observe $\simeq 625$ arcmin^2 of sky within the Cosmic Assembly Near-IR Deep Extragalactic Legacy Survey (CANDELS; Grogin et al., 2011; Koekemoer et al., 2011) footprint. Bowman et al. (2019) vetted this dataset to define a sample of $\simeq$ 2000 optical-emission line galaxies with IR continuum magnitude $m_{\rm J+JH+H} < 26$, unambiguous emission-line redshifts between $1.90 < z < 2.35$, and a 50% line-flux completeness limit of $\sim 4\times10^{-17}$ ergs cm^{-2} s^{-1}. In over 90% of the sample, the brightest emission line in the spectral region surveyed by the grating is [O III] λ5007; in 90% of the remaining galaxies, [O II] dominates. Most AGN have been removed from this dataset via comparisons with X-ray source catalogs, and Bowman et al. (2019) estimate the fraction of remaining AGN to be less than 5%.

More than half of the Bowman et al. (2019) sample has been surveyed as part of science verification for the HETDEX survey (Gebhardt et al., 2021); this dataset includes over 900 galaxies that have been observed more than once, with some being observed up to 15 times. These repeat observations partly cover the gaps between IFUs and provide a better spatial sampling of the datacube. Weiss et al. (2021) measured the mean Lyα escape fraction of the subsample of these galaxies present in HETDEX DR2 ($6^{+0.6}_{-0.5}$%) and determined the systematic behavior of the Lyα escape versus stellar mass, SFR, internal extinction, half-light radius, and excitation.

We only include HETDEX observations with good seeing (point-spread function (PSF) full-width-at-half-maximum $< 1.7''$) and observing conditions (total system throughput > 0.1 = 13th percentile). These requirements are less strict than for the LAE sample because too few observations of [O III] galaxies meet these requirements. We inspect the remaining observations and exclude data with obvious artifacts such as interference patterns. We require that for an [O III] galaxy's halo to be included in our analysis, the center of the galaxy must lie within $3''$ of the center of a HETDEX fiber. A total of 1034 [O III] galaxies (in 44 HETDEX observations) meet our selection criteria, with 57 ($\simeq 6$%) having a Lyα detection in HETDEX (within $3''$ and 15 Å of the expected emission line). Each galaxy was observed in $1-15$ separate observations; thus there are 7401 individual observations of the 1034 galaxies. Their mean redshift is $\langle z\rangle = 2.1$.

Because of the abundance of imaging data in the CANDELS fields, the physical

properties of our [O III] sample have been well characterized, with stellar masses between $8.2 \lesssim \log_{10} M/M_\odot \lesssim 11.4$ (median mass of $\log_{10} M/M_\odot = 9.3$), SFRs between $0.02 \lesssim \mathrm{SFR} \lesssim 250\,M_\odot\ \mathrm{yr}^{-1}$ (median value of $1.9 M_\odot\ \mathrm{yr}^{-1}$), internal extinctions between $0 \lesssim E(B-V) \lesssim 0.6$ (median of $E(B-V) = 0.09$), and optical half-light radii $R_e \lesssim 5$ kpc (with a median of 1.4 kpc). The full distribution of these properties, along with their [O III] luminosity function and equivalent width distribution, can be found in Bowman et al. (2019, 2020, 2021).

To study the potential dependence of the Lyα SB profile on various galaxy properties, we form two sub-samples above and below the median observed $L_{[\mathrm{O\,III}]}$ ($41.3 \lesssim \log_{10} L\,\mathrm{erg}^{-1}\,\mathrm{s} \lesssim 43.1$, median 42.1, 517/517 sources above/below), SFR (517/517 sources above/below), stellar mass (517/517 sources above/below), Hβ flux ($\lesssim 8{\times}10^{-17}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$, median $10^{-17}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$, 516/518 sources above/below), [O II] flux ($\lesssim 1.7{\times}10^{-16}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$, median $2\times10^{-17}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$, 517/517 sources above/below), dust attenuation (515/519 sources above/below), and UV luminosity ($25.4 \lesssim \log_{10} L_{1600}\,\mathrm{erg}^{-1}\,\mathrm{s\,Hz} \lesssim 29.4$, median 28.5, 517/517 sources above/below). We also fit a line to the SFR as a function of stellar mass and create subsamples above and below this linear relation (459/575 sources above/below). We omitted unrealistic values from the spectral-energy-distribution fits in the property ranges above.

We estimate the virial radius of the host dark matter halos using the stellar mass-halo mass relation of Behroozi et al. (2019). Roughly 68% of the galaxies in our [O III] sample have stellar masses between $10^{8.8}\,M_\odot$ and $10^{9.9}\,M_\odot$ and therefore reside in $10^{11.4}\,M_\odot$ to $10^{11.9}\,M_\odot$ dark matter halos. Following the definition of r_{vir} of Bryan & Norman (1998), we obtain $r_{\mathrm{vir}} \simeq 59 - 105$ kpc.

4.2.3 LAE Sample

The LAE sample is selected from the HETDEX survey and is described in LN22. It consists of 968 LAEs at $1.9 < z < 3.5$ with narrow lines (Lyα line FWHM $< 1000\,\mathrm{km\,s^{-1}}$) and Ly$\alpha$ luminosities $10^{42.4}\,\mathrm{erg\,s^{-1}} \lesssim L_{\mathrm{Ly}\alpha} < 10^{43}\,\mathrm{erg\,s^{-1}}$. These conditions remove most AGN from the sample. Each LAE was observed once. 364 of these LAEs are at $z < 2.35$. The equivalent widths of the Lyα and other lines measured from the median-stacked rest-frame spectrum are consistent with star formation being the main powering mechanism of the Lyα emission. To resolve Lyα halos, LN22 chose LAEs observed with PSF FWHM $< 1.4''$ throughput $> 0.13 = 40$th percentile. While we do not know the SFR and stellar mass of this sample, LAEs at $z \simeq 2.2$ with slightly lower Lyα luminosity ($L_{\mathrm{Ly}\alpha} \simeq 10^{42.3}\,\mathrm{erg\,s^{-1}}$) typically have SFR $\simeq 14\,M_\odot\,\mathrm{yr}^{-1}$ and stellar mass $M_\star \simeq 5\times10^8\,M_\odot$ (Nakajima et al., 2012). HETDEX LAEs have similar SFR and stellar mass (McCarron et al. 2022, ApJ submitted).

4.3 Measurement of Lyα Halos

4.3.1 Extraction of Surface Brightness and Stacking

We measure the Lyα halo profiles of the [O III] galaxies and around our comparative sample of LAEs following a similar procedure to LN22. The LAE profiles are consistent with each other. Because we do not detect individual Lyα emission lines of most [O III] galaxies, we assume that the observed Lyα line lies at the 3D-HST redshift. First, we remove continuum emission from the spectra to isolate Lyα from the continuum flux and to mitigate the impact of continuum emission from projected neighbors. From each fiber spectrum, we subtract the median flux between 11.7 and 40 Å (observed) away from the Lyα line on the red and blue side. We then integrate the flux around the expected Lyα wavelength, obtaining a surface brightness for each fiber. We choose an integration window of $\lambda_{\mathrm{Ly}\alpha}^{\mathrm{obs}} \pm 10\,\text{Å}$ to account for the uncertainty of the expected observed Lyα wavelength due the limited spectral resolution of the grism ($R \simeq 130$).

The Lyα line can be redshifted by $\simeq 200\,\mathrm{km\,s^{-1}}$ from a galaxy's redshift because of radiative transfer effects (e.g., Nakajima et al., 2018a). We tested redshifting the integration window by $200\,\mathrm{km\,s^{-1}}$ and subtracting the continuum on the red side of the shifted Lyα line. We also tested subtracting only the red continuum without shifting the line. Both tests produce a Lyα SB profile consistent with our results.

We define two LAE samples, the entire sample and the low-redshift ($z < 2.35$) subsample. For the comparison with the entire LAE sample we correct for cosmological surface brightness dimming to the mean redshift (2.1) of the [O III] galaxy sample (factor $(1+z)^4 \times (1+2.1)^{-4}$). For the comparison with previous results (§ 4.4.1) we convert the surface brightness of each LAE and [O III] galaxy to surface luminosity to account for cosmological surface brightness dimming. The surface luminosity $\mathrm{SL_{Ly\alpha}}$ relates to the surface brightness $\mathrm{SB_{Ly\alpha}}$ as $\mathrm{SL_{Ly\alpha}} = dL/dA_{\mathrm{em}} = 4\pi\,(1+z)^4\,\mathrm{SB_{Ly\alpha}}$, where dA_{em} is the surface area at emission.

We sort the fibers around each galaxy in each observation by their distance from the galaxy center, and place them in radial bins with the bin edges at 5, 10, 15, 20, 30, 40, 60, 80, 160, 320 and 800 kpc.

We take the median of all fibers within each bin around each galaxy observation individually. Then we take the median of these radial profiles and estimate the uncertainty via a bootstrap algorithm. At least 355 (527) [O III] galaxies (galaxy observations) contribute to each bin.

4.3.2 Estimating Systematic Uncertainty

We estimate the systematic uncertainty in two parts following the approach of LN22. The first estimates the background surface brightness at the same wavelengths and in the same observations as the galaxies separately for each galaxy sample. This background includes physical emission, e.g. from interlopers, sky emission residuals, and systematics introduced in the continuum subtraction. We calculate the median surface brightness of all fibers farther away than 800 kpc from each galaxy observation. The median of these determines the background and a bootstrap algorithm determines the uncertainty. We

find $(1.51 \pm 0.02) \times 10^{-19}$erg s^{-1} cm^{-2} arcsec^{-2} for the [O III] galaxies. We subtract this background from the median profile of the galaxies.

The second part determines the systematic uncertainty of the median radial profile in proximity of the galaxies. We repeat the stacking procedure 40 times, but with the central wavelength shifted between 20 Å and 210 Å in increments of 10 Å in both wavelength directions. Some galaxies at the blue end of the covered wavelength range have fewer than 40 wavelength-shifted profiles. The standard deviation per bin of these 40 Lyα-free profiles determines the systematic uncertainty of the median Lyα SB profile. The median ratio of this systematic uncertainty to the uncertainty from the bootstrap algorithm is 1.4. In each bin, we choose the larger of the two estimates as the final uncertainty. The mean and median of the wavelength-shifted profiles are consistent with the background surface brightness.

We stack the radial profiles of stars from the Gaia DR 2 (Gaia Collaboration et al., 2018) in the same manner as LN22 out to 100″. The median profile plateaus at $10'' < r < 100''$, presumably because of unmasked continuum sources and the lack of a continuum and background subtraction. We subtract the mean value at $r > 10''$. We obtain a separate star profile for the observations of LAEs and [O III] emitters and scale them to match the flux within 2″ of the galaxy profiles. Both profiles are modeled well by a Moffat function with $\beta = 2.2$ with the mean seeing FWHM as the observations.

4.4 Results and Discussion

Figure 4.1 presents the median Lyα SB profile of the [O III] galaxies out to 800 kpc. The profile is significantly more extended than the star profile. Figure 4.1 also shows the median redshift-adjusted Lyα SB profile of the LAE sample at $1.9 < z < 3.5$. While the [O III]-galaxy profile is an order of magnitude fainter at $r < 10$ kpc, it reaches a consistent surface brightness at $r > 40$ kpc. The profiles of the entire LAE sample and the subsample at $z < 2.35$ are consistent at all radii.

Byrohl et al. (2021) find in their simulation that the photons in the core predominantly originate from the central galaxy, but those at large distances originate from other galaxies. Hence the central surface brightness should depend on the amount of photons escaping the central galaxy. At large distances, however, galaxies with similar CGM and clustering properties should have similar Lyα SB profiles.

While the intrinsic Lyα luminosities of the galaxy samples are unknown, the small Lyα escape fraction of the [O III] galaxies along the LOS can explain the lower surface brightness of the profile in the core. In contrast, Leclercq et al. (2017) find a weak positive correlation between the halo scale length and Lyα luminosity of the inner halo, implying that brighter LAEs have flatter halos. The similarity of the Lyα SB profiles of the two galaxy samples at $r > 40$ kpc supports the picture in which the outer parts of the profiles are dominated by photons not related to Lyα emission produced in the central galaxy. The profiles are modeled well by a PSF-plus-powerlaw model, with power law index -1.45, cut off at $r_{\rm min} \simeq 2$ kpc. While the power-law component is designed to be identical, the best-fit PSF component of the LAE profile is 40 times brighter than that of the [O III] galaxies.

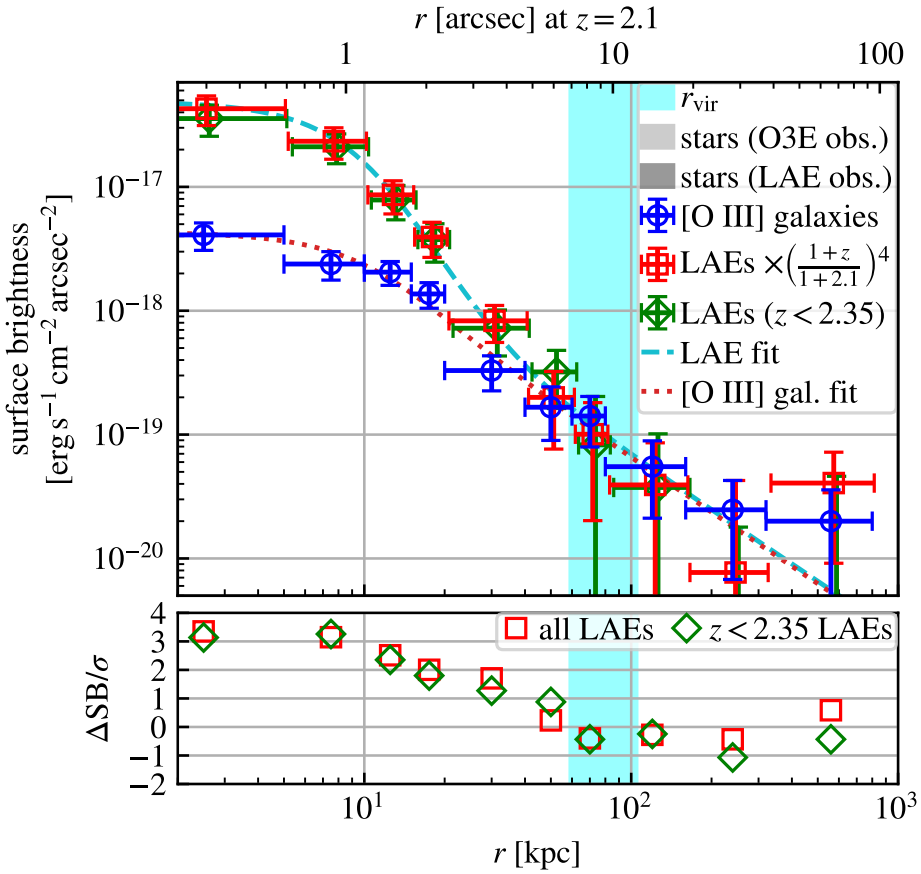

Figure 4.1: **Top**: Median Lyα SB profile of 1034 [O III] galaxies (blue circles) compared to the redshift-adjusted profile of all LAEs (red squares) and the profile of the LAEs at $z < 2.35$ (green diamonds). The LAE profiles are slightly shifted along the x axis for better visibility. The star profile in the [O III] galaxy observations at $z = 2.1$ and that in the LAE observations at $z = 2.5$ are shown as light gray and dark gray areas, respectively. The cyan area shows the estimated virial radius of the host dark matter halos of the [O III] galaxies. The dotted and dashed lines show the best-fit PSF-plus-powerlaw model. **Bottom**: Significance of the difference between the profiles, i.e. the difference between the [O III] profile and the LAE profiles divided by the uncertainties added in quadrature, with the same symbols as above.

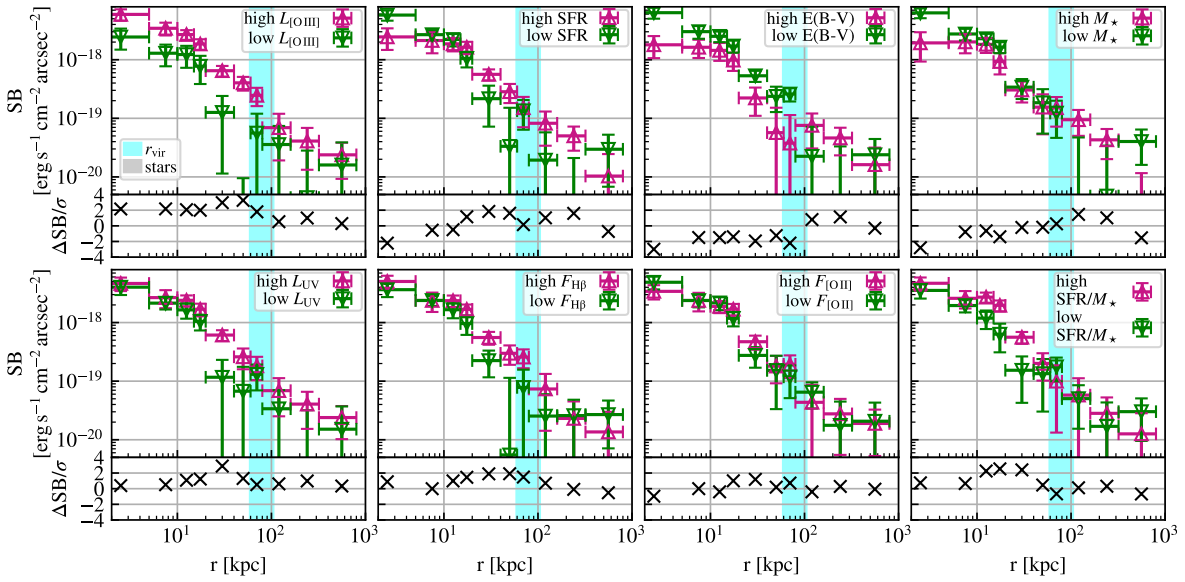

Figure 4.2: Median Lyα SB profiles of differently separated subsamples. Except for the bottom right, each panel includes the subsample above (upward-facing magenta triangles) and below (downward-facing green triangles) the median of one property. The bottom right panel shows the subsample above and below the linear SFR-$M_\star$ relation. The bottom part of each panel shows the significance of the difference between the two profiles.

Figure 4.2 shows the median Lyα SB profiles of several subsamples of the [O III] galaxies. Most profiles are similar ($< 2\sigma$ difference). The following differences are statistically significant ($> 2\sigma$). The Lyα surface brightness of the low-$L_{[O III]}$ sample is lower than that of the high-$L_{[O III]}$ sample at $r < 60$ kpc. While the low-$L_{[O III]}$ profile follows the star profile out to 40 kpc, it increases to match the high-$L_{[O III]}$ profile at $r > 60$ kpc. The subsamples with high dust attenuation, stellar mass, and SFR are fainter at $r < 5$ kpc than those with low dust attenuation, stellar mass, and SFR, but similar at larger distances. This can be explained by lower escape fractions for those subsamples, consistent with the notion that the escape fraction anti-correlates with dust extinction, stellar mass, and SFR (Runnholm et al., 2020; Weiss et al., 2021). The profiles of the subsamples with low UV luminosity or below the SFR-stellar mass relation are similar to those with high UV luminosity or above the SFR-stellar mass relation at most radii, but fainter at intermediate distances, similar to the low-$L_{[OIII]}$ subsample. This suggests that the surface brightness is independent of the properties of the central galaxies at large distances. However, the uncertainties at large radii are large and more data are necessary for a clear conclusion. The similarity of the Lyα SB profiles at different stellar masses appears to contradict the result of Byrohl et al. (2021). Their fiducial model, which does not account for destruction of Lyα photons by dust, indicates that the Lyα SB is higher for galaxies with higher stellar mass out to large distances. When including dust treatment (see Appendix A4 of Byrohl et al., 2021), the correlation between stellar mass and outer Lyα SB level weakens because massive galaxies are more strongly affected by dust attenuation. The resulting Lyα SB profiles are more similar across stellar masses, better matching our findings.

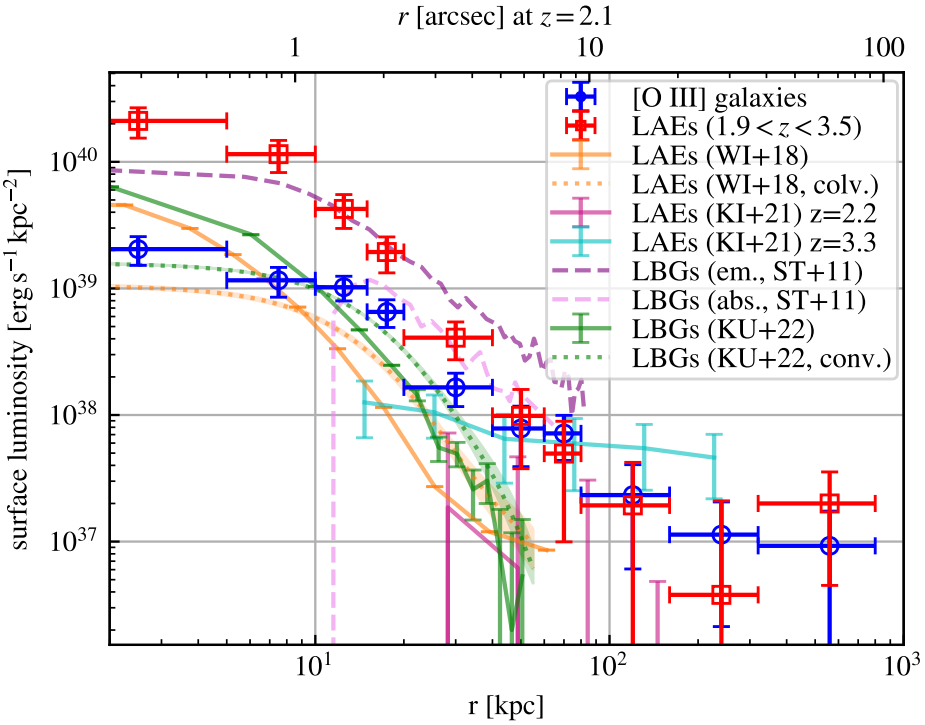

Figure 4.3: Comparison of the surface luminosity profile of the [O III] galaxies (blue circles) and that of the LAEs (red squares) with LAEs at $z = 3 - 4$ (WI+18; Wisotzki et al., 2018), LAEs at $z = 2.2$ and $z = 3.3$ (KI+21; Kikuchihara et al., 2022a), LBGs with net Lya emission and absorption at $z = 2.65$ (ST+11; Steidel et al., 2011), and LBGs at $z = 3.56$ (KU+22; Kusakabe et al., 2022). The dotted profiles are convolved with the VIRUS PSF.

4.4.1 Comparison with Previous Results

Figure 4.3 compares the surface luminosity profile of [O III] galaxies and that of the LAEs with previous results for LAEs and LBGs at redshifts $2 < z < 4$. We show the profiles as a function of physical distance because most of the data lies within the virial radii of the galaxies. The profiles that were given as a function of comoving distance or did not contain the $(1+z)^4$ factor were adjusted using one redshift for each sample (see caption of Figure 4.3). Because of the smaller PSF of MUSE, the profiles of Wisotzki et al. (2018) and Kusakabe et al. (2022) are steeper in the core than our profiles. We therefore show the convolved profiles with the PSF model following the stacked star profile in the [O III] galaxy observations (Moffat function with $\beta = 2.2$ and FWHM= 1.47 convolved with the VIRUS fiber profile), as though VIRUS observed these profiles at $z = 2.1$. Despite differences at small distances, all LAE and LBG profiles are similar at intermediate distances ($20\,\mathrm{kpc} \lesssim r \lesssim 80\,\mathrm{kpc}$).

4.4.2 Emission Mechanism

Star formation in other galaxies

To find out whether the measured Lyα emission can be powered by star formation alone, we estimate the required star formation rate density (SFRD) from the measured Lyα luminosity within 800 kpc of the [O III] sample ($L_{\mathrm{Ly}\alpha} = (2.3 \pm 1.3) \times 10^{43}\,\mathrm{erg\,s^{-1}}$). Using $L_{\mathrm{Ly}\alpha} = 10^{42}\,\mathrm{erg\,s^{-1}} \times \mathrm{SFR}\,\mathrm{M_\odot^{-1}\,yr}$ (see Dijkstra, 2019), we find $\mathrm{SFRD} = 0.05 \pm 0.03\,\mathrm{M_\odot\,yr^{-1}\,cMpc^{-3}}$. This value is smaller than those in the literature ($\simeq 0.1\,M_\odot\,\mathrm{yr^{-1}\,cMpc^{-3}}$; summarized by Rowan-Robinson et al., 2016), implying that star-formation-induced photons can account for the Lyα emission out to 800 kpc.

Star formation in the central galaxy

The median SFR of the [O III] galaxies is $10^{0.960}\,M_\odot\,\mathrm{yr^{-1}}$. Using the same $L_{\mathrm{Ly}\alpha}$-SFR relation as above, we expect an intrinsic Lyα luminosity $L^{\mathrm{int}}_{\mathrm{Ly}\alpha} \simeq 9.1 \times 10^{42}\,\mathrm{erg\,s^{-1}}$. Because this is consistent with the measured luminosity, the Lyα photons could originate from the central galaxies if the escape fraction is close to one. This scenario is disfavoured because of the small measured escape fraction of $6^{+0.6}_{-0.5}\%$ (Weiss et al., 2021), as well as the theoretically expected one. Using the standard relation for the optical depth due to dust extinction of Calzetti et al. (2000) and Verhamme et al. (2006b) yields $f^{\mathrm{Ly}\alpha}_{\mathrm{esc}} \leq \exp\left\{-\tau^{\mathrm{Ly}\alpha}_{\mathrm{dust}}\right\} \approx 0.21$ for the median E(B-V) of our sample.

Fluorescence

Cantalupo et al. (2005) predict a Lyα surface brightness through fluorescence from the UV background of $3.67 \times 10^{-20}\,\mathrm{erg\,s^{-1}\,cm^{-2}\,arcsec^{-2}}$ at $z \sim 3$. The photoionization rate of the UV background changes little from $z = 3$ to $z = 2$ (Faucher-Giguère, 2020). Accounting for cosmic dimming, this value would be $\simeq 10^{-19}\,\mathrm{erg\,s^{-1}\,cm^{-2}\,arcsec^{-2}}$ at $z = 2.1$, which is consistent with the intermediate and outer points of the radial profiles, but too low to explain levels at small distances.

Cooling radiation

Lyα photons can be emitted through collisional excitation and recombination in cooling gas flowing into a galaxy. The subsequent scattering in an inflowing medium can lead to a blue-shift of the Lyα line (Dijkstra et al., 2006). While the scattering and blue-shift may be negligible due to the low volume-filling factor of cold streams, we expect a filamentary morphology of the Lyα emission (Haiman et al., 2000; Dijkstra & Loeb, 2009). We cannot test whether the Lyα line is blue-shifted because of the low spectral resolution and the high redshift uncertainty of the [O III] galaxies. Detecting the filamentary structure requires deep observations of individual Lyα halos rather than stacking and circular averaging.

4.5 Summary

We measure the Lyα emission out to 800 kpc around 1034 [O III]-selected galaxies at $1.9 < z < 2.35$. While the central surface brightness in the core ($r < 10$ kpc) is fainter than that of the median redshift-adjusted Lyα SB profile of 968 LAEs at $1.9 < z < 3.5$ by an order of magnitude, the Lyα surface brightness in the outer parts ($r > 40$ kpc) reaches the same surface brightness as that of the LAEs.

This result supports the picture in which photons originating from outside of the central galaxies dominate the Lyα SB profiles at large radii. These photons either originate from other dark matter halos or satellite galaxies, or are emitted through fluorescence or cooling radiation in the CGM. While we cannot exclude any of these sources, star formation alone can account for the integrated Lyα emission out to 800 kpc, and fluorescence from the UV background is sufficient to explain the surface brightness at intermediate distances.

Acknowledgments

We thank C. Byrohl for helpful and interesting discussions and F. Arrigoni Battaia, V. González Lobos, and C. Peroux for comments on the draft. We also thank the anonymous referee for the helpful review. EK's work was supported in part by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC-2094 - 390783311.

HETDEX is led by the University of Texas at Austin, McDonald Observatory, and Department of Astronomy with participation from the Ludwig-Maximilians-Universität München, Max-Planck-Institut für Extraterrestrische Physik (MPE), Leibniz-Institut für Astrophysik Potsdam (AIP), Texas A&M University, Pennsylvania State University, Institut für Astrophysik Göttingen, The University of Oxford, Max-Planck-Institut für Astrophysik (MPA), The University of Tokyo, and Missouri University of Science and Technology. In addition to Institutional support, HETDEX is funded by the National Science Foundation (grant AST-0926815), the State of Texas, the US Air Force (AFRL FA9451-04-2-0355), and generous support from private individuals and foundations.

The observations were obtained with the Hobby-Eberly Telescope (HET), which is a joint project of the University of Texas at Austin, the Pennsylvania State University, Ludwig-Maximilians-Universität München, and Georg-August-Universität Göttingen. The HET is named in honor of its principal benefactors, William P. Hobby and Robert E. Eberly.

VIRUS is a joint project of the University of Texas at Austin, Leibniz-Institut für Astrophysik Potsdam (AIP), Texas A&M University (TAMU), Max-Planck-Institut für Extraterrestrische Physik (MPE), Ludwig-Maximilians-Universität München, Pennsylvania State University, Institut für Astrophysik Göttingen, University of Oxford, Max-Planck-Institut für Astrophysik (MPA), and The University of Tokyo.

The authors acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing high performance computing, visualization, and storage resources that have contributed to the research results reported within this paper. URL: http://www.tacc.utexas.edu.

This work is based on observations taken by the 3D-HST Treasury Program (GO 12177 and 12328) with the NASA/ESA HST, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555.

The Institute for Gravitation and the Cosmos is supported by the Eberly College of Science and the Office of the Senior Vice President for Research at the Pennsylvania State University.

This research made use of NASA's Astrophysics Data System Bibliographic Services.

We acknowledge the use of the astropy (Astropy Collaboration et al., 2013, 2018), matplotlib (Hunter, 2007), numpy (Harris et al., 2020), and scipy (Virtanen et al., 2020) software packages.

Chapter 5

SIMPLE: Simple Intensity Map Producer for Line Emission

The content of this chapter has been published in Lujan Niemeyer et al. (2023).

We present the Simple Intensity Map Producer for Line Emission (Simple), a public code for quickly simulating mock line-intensity maps, and an analytical framework for modeling intensity maps including observational effects. Simple can be applied to any spectral line sourced by galaxies. The Simple code is based on lognormal mock catalogs of galaxies including positions and velocities and assigns luminosities following the luminosity function. After applying a selection function to distinguish between detected and undetected galaxies, the code generates an intensity map, which can be modified with anisotropic smoothing, noise, a mask, and sky subtraction, and calculates the power spectrum multipoles. We show that the intensity autopower spectrum and the galaxy-intensity cross-power spectrum agree well with the analytical estimates in real space. We derive and show that the sky subtraction suppresses the intensity autopower spectrum and the cross-power spectrum on scales larger than the size of an individual observation. As an example application, we make forecasts for the sensitivity of an intensity mapping experiment similar to the Hobby-Eberly Telescope Dark Energy Experiment (Hetdex) to the cross-power spectrum of Lyα-emitting galaxies and the Lyα intensity. We predict that Hetdex will detect the galaxy-intensity cross-power spectrum on scales of $0.04\,h\,\mathrm{Mpc}^{-1} < k < 1\,h\,\mathrm{Mpc}^{-1}$.

5.1 Introduction

Line-intensity mapping (LIM) is a promising survey strategy for constraining cosmological parameters and studying astrophysics of galaxies and intergalactic gas (e.g., Kovetz et al., 2017; Bernal & Kovetz, 2022). Instead of relying on detections of individual galaxies to trace the large-scale structure of matter, LIM measures the integrated line emission from all galaxies and the intergalactic medium (IGM) as a biased tracer of the matter distribution. It collects light from all emitters, including those that are too faint for

individual detection at high redshift, and requires lower resolution and shorter integration times than traditional galaxy surveys.

Several LIM surveys are in operation (Santos et al., 2016; Keating et al., 2016, 2020; DeBoer et al., 2017; CONCERTO Collaboration et al., 2020; Gebhardt et al., 2021; Cleary et al., 2022) or in preparation (Doré et al., 2014; Vieira et al., 2020; Sun et al., 2021; Switzer et al., 2021; CCAT-Prime Collaboration et al., 2022). They target lines ranging from the 21 cm line in radio frequencies to the ultraviolet Lyman-α (Lyα) line, which trace atomic or molecular gas. Although the 21 cm line is emitted by neutral atomic hydrogen in the IGM and in neutral pockets within the interstellar medium (ISM) of galaxies, other lines targeted by LIM experiments are associated with star formation in the ISM (e.g., Bernal & Kovetz, 2022). Neglecting scattering and diffuse emission of the Lyα line, which can illuminate the circumgalactic medium (CGM) and IGM (see, e.g., Byrohl et al., 2021; Byrohl & Nelson, 2023), these line intensities are therefore sourced within galaxies and influenced by their astrophysical properties such as their star-formation rate.

Mock intensity maps are necessary to model the signal beyond the capabilities of analytical models, estimate statistical covariance, and explore observational effects such as foregrounds, sky subtraction, and survey footprints on summary statistics (e.g., Cunnington et al., 2023b). Modeling techniques of line-intensity maps have to compromise between astrophysical complexity, volume, and computational feasibility. On the one hand, cosmological hydrodynamical simulations can be post-processed using the astrophysical properties of galaxies and the gas to infer the line emission (e.g., Moriwaki et al., 2019; Silva et al., 2021; Kannan et al., 2022; Byrohl & Nelson, 2023; Liang et al., 2023). Halo catalogs from N-body simulations can also be combined with galaxy evolution models to predict line luminosities and produce intensity maps (e.g., Lidz et al., 2011; Gong et al., 2012; Li et al., 2016; Chung et al., 2019; Spina et al., 2021; Béthermin et al., 2022; Moradinezhad Dizgah et al., 2022b; Sato-Polito et al., 2023). Although these approaches provide accurate small-scale clustering and include astrophysical dependences of the line intensity, it is not feasible to produce enough realizations for covariance estimation or parameter inference. To speed up the calculation, one can model the underlying dark matter density field through Lagrangian perturbation theory or mass-peak patch and apply various post-processing steps to model the astrophysical dependence of line intensities (Mesinger et al., 2011; Silva et al., 2013, 2015; Mesinger et al., 2016; Heneka et al., 2017; Heneka & Mesinger, 2020; Mas-Ribas et al., 2023; Chung et al., 2022; Roy et al., 2023). The line intensity can also be modeled by multiplying the total matter density in a fast lognormal simulation by a bias (Alonso et al., 2014; Rubiola et al., 2022). These approaches are fast and enable simulating many realizations of large volumes, but do not account for the shot-noise contribution to the power spectrum from the discreteness of the line-emitting sources. Finally, Obuljen et al. (2023) use a field-level forward-modeling approach to simulate intensity maps based on effective field theory.

To include shot noise, one can generate a galaxy catalog from a lognormal galaxy number density field via Poisson sampling. Lognormal simulations take advantage of the roughly lognormal probability density function (PDF) of matter and galaxy density distributions, measured both in N-body simulations (e.g., Kayo et al., 2001; Shin et al., 2017) and in galaxy surveys (e.g., Clerkin et al., 2017). The lognormality of the PDF of

the density contrast δ implies that the logarithmic transformation field, $\ln(1+\delta)$, is a Gaussian random field whose statistics are defined entirely by its two-point correlation function or power spectrum. Agrawal et al. (2017) present a public lognormal mock generator for galaxy catalogs, including self-consistent velocities of the galaxies. The velocities enable redshift-space distortion (RSD) modeling beyond the linear Kaiser model (Kaiser, 1987), which is critical for any large-scale structure measurement in redshift space such as LIM.

The lognormal galaxy catalog generator of Agrawal et al. (2017) has been extended to generate weak-lensing fields (Makiya et al., 2021). In this paper, we extend it to quickly and self-consistently generate intensity maps and galaxy catalogs, which we call the Simple Intensity Map Producer for Line Emission (SIMPLE).[1] Given a luminosity function for any emission line, we assign line luminosities to galaxies in a lognormal galaxy simulation, apply a selection function to obtain a galaxy catalog, and calculate the intensity on a grid. We can add noise, smooth the intensity map, model the sky subtraction, and apply a mask before calculating the galaxy and LIM auto- and cross-power spectra. The simplicity of this approach enables us to quickly generate many mock intensity maps, e.g., to estimate the covariance matrices of the LIM power spectra.

As an example of its capabilities, we use the SIMPLE framework to forecast the sensitivity of the Hobby-Eberly Telescope Dark Energy Experiment (HETDEX; Gebhardt et al., 2021) to the cross correlation of Lyα-emitting galaxies (LAEs) and the Lyα intensity. An analytical forecast for the intensity autopower spectrum in HETDEX was presented in Fonseca et al. (2017).

This paper is structured as follows. Section 5.2 derives the power spectrum formalism. In Section 5.3 we describe the steps of the SIMPLE code to generate line-intensity mocks. Section 5.4 explains an example mock setup for HETDEX that is used in the rest of the paper. We validate the SIMPLE code by comparing the power spectrum multipoles with theoretical predictions in Section 5.5. Section 5.6 presents the forecast for the LAE-Lyα intensity cross-correlation of HETDEX. Section 5.7 discusses the limitations of the SIMPLE framework. We conclude in Section 5.8.

We use the following Fourier convention:

$$\begin{aligned} \tilde{f}(\mathbf{k}) &= \int \mathrm{d}^3\mathbf{x} f(\mathbf{x}) e^{i\mathbf{k}\cdot\mathbf{x}}\,, \\ f(\mathbf{x}) &= \int \frac{\mathrm{d}^3\mathbf{k}}{(2\pi)^3} \tilde{f}(\mathbf{k}) e^{-i\mathbf{k}\cdot\mathbf{x}}\,, \end{aligned} \tag{5.1}$$

where the tilde denotes quantities in Fourier space. We refer to *real* space in contrast to redshift space, and to *configuration* space in contrast to Fourier space.

Throughout this paper, we assume a flat Λ cold dark matter (ΛCDM) cosmology with $H_0 = 67.66\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_{\mathrm{b},0}h^2 = 0.022$, $\Omega_{\mathrm{m},0}h^2 = 0.142$, $\ln(10^{10}A_s) = 3.094$, and $n_s = 0.9645$.

[1] `https://github.com/mlujnie/simple` (Lujan Niemeyer et al., 2023)

5.2 Power Spectrum Modeling

Consider a fluctuation $\delta A(\mathbf{x}) = A(\mathbf{x}) - \langle A(\mathbf{x})\rangle$ of a field A, such as the intensity $A(\mathbf{x}) = I(\mathbf{x})$ and the normalized galaxy number density $A(\mathbf{x}) = n(\mathbf{x})/\langle n(\mathbf{x})\rangle = 1 + \delta_\mathrm{g}(\mathbf{x})$. Here, $\langle\cdot(\mathbf{x})\rangle$ denotes the mean field at location $\mathbf{x}$ over many realizations, for example, the mean intensity as a function of redshift, or the galaxy selection function as a function of position. The dimensionless correlation function of fields A and B in configuration space is defined as

$$\begin{aligned}\xi_{AB}(\mathbf{x}-\mathbf{y}) &= \frac{\langle A(\mathbf{x})B(\mathbf{y})\rangle - \langle A(\mathbf{x})\rangle\langle B(\mathbf{y})\rangle}{\langle A(\mathbf{x})\rangle\langle B(\mathbf{y})\rangle} \\ &= \frac{\langle \delta A(\mathbf{x})\delta B(\mathbf{y})\rangle}{\langle A(\mathbf{x})\rangle\langle B(\mathbf{y})\rangle}\,.\end{aligned} \tag{5.2}$$

The corresponding power spectrum with the dimension of volume is the Fourier transform of the correlation function,

$$P_{AB}(\mathbf{k}) = \int \mathrm{d}^3\mathbf{s}\, \xi^{AB}(\mathbf{s}) e^{i\mathbf{k}\cdot\mathbf{s}}. \tag{5.3}$$

5.2.1 Galaxy and Intensity Auto- and Cross-power Spectra

Following the standard approach of Peebles (1980), we model the galaxy number $N(\mathbf{x})$ as a Poisson point process in infinitesimal volume elements δV, so that the occupation number in each cell is $N_i \in \{0, 1\}$. The expectation value of N in one cell is $\bar{n}\,[1 + \delta_\mathrm{g}(\mathbf{x})]\,\delta V$, where $\bar{n}$ is the mean number density of galaxies in the entire volume, and δ_g is the galaxy overdensity. Let each galaxy have a line luminosity L_i, which is sampled from a luminosity function $\mathrm{d}n/\mathrm{d}L$. We require that the integral $\int \mathrm{d}L \frac{\mathrm{d}n}{\mathrm{d}L} =: \bar{n}$ converge so that $\phi(L) := \frac{\mathrm{d}n}{\mathrm{d}L}\bar{n}^{-1}$ is a PDF, for example, by setting a minimum luminosity.

The specific intensity in a cell with luminosity $L(\mathbf{x})$ is

$$I_\lambda(\mathbf{x}) = \frac{c}{4\pi\lambda_0 H(z)(1+z)^2}\frac{L(\mathbf{x})}{\delta V} = X_I^\lambda(\mathbf{x})\frac{L(\mathbf{x})}{\delta V}, \tag{5.4}$$

$$\text{or } I_\nu(\mathbf{x}) = \frac{c}{4\pi\nu_0 H(z)}\frac{L(\mathbf{x})}{\delta V} = X_I^\nu(\mathbf{x})\frac{L(\mathbf{x})}{\delta V}, \tag{5.5}$$

where c is the speed of light, $H(z)$ is the Hubble expansion rate, X_I is a redshift-dependent conversion factor, and λ_0 and ν_0 are the rest-frame wavelength and frequency of the line, respectively. For simplicity, we refer to the specific intensity as 'intensity' with the symbol I.

Given a function $f(A)$ of a continuous random variable A with PDF $\phi(A)$, we can calculate its expectation value as $\langle f(A)\rangle = \int \mathrm{d}A\, f(A)\phi(A)$. As the PDF of $I\delta V$ at position $\mathbf{x}$ is given by $\phi'(I\delta V) = \phi(L)X_I(\mathbf{x})$, the first and second moments of $I\delta V$ are

given by

$$\begin{aligned}\langle I(\mathbf{x})\delta V\rangle &= \bar{N}X_I(\mathbf{x})\int \mathrm{d}L\,\phi(L)L \\ &= \delta V X_I(\mathbf{x})\int \mathrm{d}L\,\frac{\mathrm{d}n}{\mathrm{d}L}L,\end{aligned} \tag{5.6}$$

$$\begin{aligned}\langle I^2(\mathbf{x})\delta V^2\rangle &= \bar{N}X_I^2(\mathbf{x})\int \mathrm{d}L\,\phi(L)L^2 \\ &= \delta V X_I^2(\mathbf{x})\int \mathrm{d}L\,\frac{\mathrm{d}n}{\mathrm{d}L}L^2,\end{aligned} \tag{5.7}$$

where $\bar{N} = \bar{n}\delta V$.

Following the integration approach of Feldman et al. (1994), we find

$$\begin{aligned}\langle I(\mathbf{x})I(\mathbf{y})\rangle = \langle I(\mathbf{x})\rangle\langle I(\mathbf{y})\rangle\,[1+\xi_{II}(\mathbf{x}-\mathbf{y})] \\ +\,\delta_{\mathrm{D}}(\mathbf{x}-\mathbf{y})X_I^2(\mathbf{x})\int \mathrm{d}L\frac{\mathrm{d}n}{\mathrm{d}L}L^2\,,\end{aligned} \tag{5.8}$$

where δ_{D} is the Dirac delta function. The second term in the previous expression assumes Poisson shot noise. Similarly, the cross correlation with the galaxy density contrast of detected galaxies, $\delta_{\mathrm{g}}(\mathbf{x}) = n(\mathbf{x})/\langle n(\mathbf{x})\rangle - 1$, is given by

$$\begin{aligned}\langle I(\mathbf{x})\delta_{\mathrm{g}}(\mathbf{y})\rangle = \langle I(\mathbf{x})\rangle\,[1+\xi_{I\mathrm{g}}(\mathbf{x}-\mathbf{y})] \\ +\,\delta_{\mathrm{D}}(\mathbf{x}-\mathbf{y})\left[\frac{X_I(\mathbf{x})}{\langle n(\mathbf{x})\rangle}\int \mathrm{d}L\frac{\mathrm{d}n}{\mathrm{d}L}L\right]_{\mathcal{G}_{\mathrm{g}}\cap\mathcal{G}_{\mathcal{I}}}\,,\end{aligned} \tag{5.9}$$

where the (second) shot-noise term only contains galaxies that contribute to the galaxy catalog and the intensity map, i.e., only detected galaxies with nonzero luminosity in the target line, denoted in the expression as $\mathcal{G}_{\mathrm{g}}\cap\mathcal{G}_{\mathcal{I}}$.

We define the weighted Fourier transform as $\widetilde{\delta I}(\mathbf{k}) = \int \mathrm{d}^3\mathbf{x}\,w_I(\mathbf{x})\delta I(\mathbf{x})e^{i\mathbf{k}\cdot\mathbf{x}}$ with a dimensionless weight $w_I(\mathbf{x})$, which can represent a survey footprint.[2] We define the estimator for the intensity power spectrum,

$$\hat{P}_{II}(\mathbf{k}) = V_{\mathrm{box}}^{-1}\langle|\widetilde{\delta I}(\mathbf{k})|^2\rangle\,, \tag{5.10}$$

where V_{box} is the volume of a cuboid enclosing the survey used to compute the Fourier transform, so that $\hat{P}_{II}(\mathbf{k})$ has the dimension of volume times intensity squared. We find that

$$\begin{aligned}\langle|\widetilde{\delta I}(\mathbf{k})|^2\rangle = \int\frac{\mathrm{d}^3\mathbf{k}'}{(2\pi)^3}P_{II}(\mathbf{k}')|\widetilde{W}_I(\mathbf{k}-\mathbf{k}')|^2 \\ +\int \mathrm{d}^3\mathbf{x}\,w_I^2(\mathbf{x})X_I^2(\mathbf{x})\int \mathrm{d}L\frac{\mathrm{d}n}{\mathrm{d}L}L^2,\end{aligned} \tag{5.11}$$

where P_{II} is the power spectrum defined in Eq. (5.3), and the window function is defined as $\widetilde{W}_I(\mathbf{k}) = \int \mathrm{d}^3\mathbf{x}\,e^{i\mathbf{k}\cdot\mathbf{x}}w_I(\mathbf{x})\langle I(\mathbf{x})\rangle$.

[2] Blake (2019) follow a similar approach, but use weights in units of (intensity)$^{-1}$.

Similary, we define the estimator for the cross-power spectrum as

$$\hat{P}_{\mathrm{g}I}(\mathbf{k}) = V_{\mathrm{box}}^{-1}\langle\widetilde{\delta I}(\mathbf{k})\tilde{\delta}_{\mathrm{g}}^{*}(\mathbf{k})\rangle\,, \tag{5.12}$$

with the dimension of volume times intensity, where the asterisk is the complex conjugate operator, and

$$\begin{aligned}\langle\widetilde{\delta I}(\mathbf{k})\tilde{\delta}_{\mathrm{g}}^{*}(\mathbf{k})\rangle = &\int\frac{\mathrm{d}^3\mathbf{k}'}{(2\pi)^3}P_{I\mathrm{g}}(\mathbf{k}')\widetilde{W}_I(\mathbf{k}-\mathbf{k}')\widetilde{W}_{\mathrm{g}}^{*}(\mathbf{k}-\mathbf{k}')\\ &+\int\mathrm{d}^3\mathbf{x}\,w_I(\mathbf{x})w_{\mathrm{g}}(\mathbf{x})\left[\frac{X_I(\mathbf{x})}{\langle n(\mathbf{x})\rangle}\int\mathrm{d}L\frac{\mathrm{d}n}{\mathrm{d}L}L\right]_{\mathcal{G}_{\mathrm{g}}\cap\mathcal{G}_{\mathcal{I}}}\,,\end{aligned} \tag{5.13}$$

where the galaxy window function is $\widetilde{W}_{\mathrm{g}}(\mathbf{k}) = \int\mathrm{d}^3\mathbf{x}\,e^{i\mathbf{k}\cdot\mathbf{x}}w_{\mathrm{g}}(\mathbf{x})$ with a dimensionless weight, $w_{\mathrm{g}}(\mathbf{x})$.

Finally, we define the galaxy power spectrum estimator as

$$\hat{P}_{\mathrm{gg}}(\mathbf{k}) = V_{\mathrm{box}}^{-1}\langle|\tilde{\delta}_{\mathrm{g}}(\mathbf{k})|^2\rangle\,, \tag{5.14}$$

where

$$\begin{aligned}\langle|\tilde{\delta}_{\mathrm{g}}(\mathbf{k})|^2\rangle = &\int\frac{\mathrm{d}^3\mathbf{k}'}{(2\pi)^3}P_{\mathrm{gg}}(\mathbf{k}')|\widetilde{W}_{\mathrm{g}}(\mathbf{k}-\mathbf{k}')|^2\\ &+\int\mathrm{d}^3\mathbf{x}\,\frac{w_{\mathrm{g}}^2(\mathbf{x})}{\langle n(\mathbf{x})\rangle}\,.\end{aligned} \tag{5.15}$$

5.2.2 Smoothing and Noise

Limited observational resolution can be modeled by smoothing the intensity map. Suppose that the intensity is smoothed with a smoothing kernel $D(\mathbf{x})$, i.e., $\tilde{I}_{\mathrm{s}}(\mathbf{k}) = \tilde{I}(\mathbf{k})\tilde{D}(\mathbf{k})$. Then the factors in the power spectrum estimators change as

$$\langle|\widetilde{\delta I}_{\mathrm{s}}(\mathbf{k})|^2\rangle = \langle|\widetilde{\delta I}(\mathbf{k})|^2\rangle|\tilde{D}(\mathbf{k})|^2\,, \tag{5.16}$$

$$\langle\widetilde{\delta I}_{\mathrm{s}}(\mathbf{k})\tilde{\delta}_{\mathrm{g}}^{*}(\mathbf{k})\rangle = \langle\widetilde{\delta I}(\mathbf{k})\tilde{\delta}_{\mathrm{g}}^{*}(\mathbf{k})\rangle\tilde{D}(\mathbf{k})\,. \tag{5.17}$$

Examples for smoothing kernels are a Gaussian or top-hat smoothing in the line-of-sight (LOS) direction, mimicking the line-spread function or binning of intensity into frequency channels,

$$D_{\parallel}^{\mathrm{Gauss}} = \exp\left(-\frac{1}{2}k_{\parallel}^2 s_{\parallel}^2\right), \tag{5.18}$$

$$D_{\parallel}^{\mathrm{top-hat}} = \mathrm{sinc}\left(\frac{1}{2}k_{\parallel}s_{\parallel}\right), \tag{5.19}$$

and a Gaussian smoothing in the angular direction, mimicking the beam smoothing or point-spread function (PSF),

$$D_{\perp}^{\mathrm{Gauss}} = \exp\left(-\frac{1}{2}k_{\perp}^2 s_{\perp}^2\right), \tag{5.20}$$

where $s_\parallel$ and $s_\perp$ define the smoothing lengths parallel and perpendicular to the LOS, respectively. Similarly, $k_\parallel = \mu k$ and $k_\perp = \sqrt{1-\mu^2}k$ are the components of the wavenumber parallel and perpendicular to the LOS, respectively, and μ is the cosine of the angle between the LOS and the wavevector $\mathbf{k}$.

Uncorrelated noise ΔI_n with zero mean can be added to the intensity to model instrumental and sky noise. Therefore, $I_\mathrm{s,n}(\mathbf{x}) = I_\mathrm{s}(\mathbf{x}) + \Delta I_\mathrm{n}(\mathbf{x})$, and the noise is characterized by its variance $\langle \Delta I_\mathrm{n}^2(\mathbf{x})\rangle = \sigma_I^2(\mathbf{x})$, where σ_I is the standard deviation of the noise in each voxel. This adds a term to the power spectrum estimator,

$$\langle |\widetilde{\delta I}_\mathrm{s,n}(\mathbf{k})|^2\rangle = \langle |\widetilde{\delta I}_\mathrm{s}(\mathbf{k})|^2\rangle + \delta V \int \mathrm{d}^3\mathbf{x}\, w^2(\mathbf{x})\sigma_I^2(\mathbf{x})\,. \tag{5.21}$$

The second term still holds for larger than infinitesimal voxel volumes δV.

5.2.3 Modeling the Sky Subtraction

The sky subtraction is a common step in the reduction pipeline of ground-based optical data to remove zodiacal light, aurora, airglow, diffuse Galactic light, and emission from the Galactic warm interstellar medium (WIM; e.g., Wyse & Gilmore, 1992). The optical sky spectrum consists of a continuum component and emission lines. With the assumption that the sky spectrum is homogeneous on scales of the focal plane size, one can estimate the sky foreground by taking an average of the spectra in 'empty' areas on the detector, i.e., areas that do not contain an object above a fiducial detection limit. The sky foreground spectrum is typically subtracted from the data.

In addition to sky subtraction, further foreground mitigation is required. For example, continuum subtraction of the spectra can be applied to remove small-scale continuum foreground fluctuations, or wavelength regions around strong optical emission lines from the WIM of our Galaxy can be masked to reduce line foreground fluctuations. Masking spectral channels would add structure to the mask along the LOS, but can be modeled using the formalism in this work. Continuum foreground substraction removes the longest-mode fluctuations along the LOS, which can be reconstructed modeling suitable transfer functions (see, e.g., Cunnington et al., 2023b). In this paper, we only model the sky subtraction for simplicity.

Sky subtraction also removes roughly the average intensity per redshift slice of the cosmological signal of interest within the scale used for the sky foreground estimation, for example, the focal plane radius (see, e.g., the discussion in Lujan Niemeyer et al., 2022a). Therefore, the sky subtraction decorrelates fluctuations on larger scales. This does not affect the galaxy clustering of detected galaxies because the detection of galaxies is not influenced by the zeropoint on larger scales. However, it does affect the intensity autopower and cross-power spectra even when all galaxies are detected because the intensity zeropoint is changed. While optical ground-based LIM experiments such as HETDEX suffer from a loss of large-scale power due to sky subtraction, ground-based LIM experiments in other wavelengths such as radio and submillimeter may also lose large-scale power due to similar issues.

We can model the effect of the sky subtraction on the intensity map by calculating the contribution of the line intensity to the estimated sky foreground. We smooth the

intensity map with a two-dimensional spherical top-hat kernel in the plane perpendicular to the LOS with the size of the area used for the sky spectrum estimation. Specifically, the estimated contribution to the sky foreground, i.e., the top-hat-smoothed intensity map, is

$$\tilde{I}_{\rm sky}(\mathbf{k}) = \tilde{I}(\mathbf{k})\tilde{D}_{\rm sky}(\mathbf{k}), \tag{5.22}$$

where $\tilde{D}_{\rm sky}(\mathbf{k})$ is the Fourier transform of the two-dimensional spherical top-hat kernel, given by

$$\tilde{D}_{\rm sky}(\mathbf{k}) = \frac{2J_1\left(s_{\rm f}\sqrt{k_a^2+k_b^2}\right)}{s_{\rm f}\sqrt{k_a^2+k_b^2}}\,, \tag{5.23}$$

where a and b denote the directions perpendicular to the LOS, and J_1 is the Bessel function of the first kind and of first order, and $s_{\rm f}$ is the radius of the area used for estimating the sky.

As $\tilde{D}_{\rm sky}(\mathbf{k})$ is real-valued, we obtain

$$\langle|\tilde{I}(\mathbf{k}) - \tilde{I}_{\rm sky}(\mathbf{k})|^2\rangle = \langle|\tilde{I}(\mathbf{k})|^2\rangle\left[1 - 2\tilde{D}_{\rm sky}(\mathbf{k}) + \tilde{D}^2_{\rm sky}(\mathbf{k})\right], \tag{5.24}$$

and

$$\langle\left[\tilde{I}(\mathbf{k}) - \tilde{I}_{\rm sky}(\mathbf{k})\right]\delta_{\rm g}^*\rangle = \langle\tilde{I}(\mathbf{k})\delta_{\rm g}^*\rangle\left[1 - \tilde{D}_{\rm sky}(\mathbf{k})\right]. \tag{5.25}$$

This shows that the power spectrum is suppressed on scales larger than $s_{\rm f}$.

5.3 Generating Mock Intensity Maps

This section describes the framework of the public code SIMPLE for generating mock intensity maps. In a nutshell, the code follows these steps:

1. Generate a galaxy catalog in real and redshift space using the lognormal code of Agrawal et al. (2017).

2. Randomly assign line luminosities to galaxies following an input luminosity function.

3. Assign redshift and flux to each galaxy. Apply the input selection function to distinguish between detected and undetected galaxies.

4. Paint an intensity map and a galaxy density map using detected, undetected, or all galaxies in each map, and optionally smooth the intensity map.

5. Optionally, generate and add an intensity noise map, apply a mask or weights, and model the sky subtraction.

6. Calculate the auto- and cross-power spectra.

The main input parameters to SIMPLE are cosmological parameters, a luminosity function, a linear galaxy bias for all galaxies, the central redshift of the box, and the rest-frame wavelength or frequency of the target line. If no tabulated matter power spectrum is provided as input, the cosmological parameters are used to generate the matter power spectrum using the Eisenstein & Hu fitting function (Eisenstein & Hu, 1998). They are also used to calculate the luminosity and angular diameter distances in later calculations.

Along with the luminosity function, one has to specify the minimum luminosity to obtain a finite number of galaxies. This defines the number of galaxies to simulate with the lognormal galaxy catalog generator of Agrawal et al. (2017), which produces their positions, velocities, and redshift-space positions. This procedure assumes a flat sky and a single redshift. We assign a luminosity to each galaxy by randomly drawing from the luminosity function. Unless otherwise specified, we define the first axis of the simulation box as the LOS and assign the redshift to each galaxy according to its distance from the observer, inferred from the position in this simulation axis. We use this redshift to convert luminosities into fluxes. Alternatively, a single redshift can be assigned to all galaxies.

We then apply a selection function based on an input flux limit above which a galaxy is detected, or on the target galaxy number density as a function of redshift. This produces a galaxy catalog with detection flags. We then calculate an intensity map and a galaxy number density map using the nearest-grid-point (NGP) assignment scheme. One can specify which galaxies contribute to the intensity and galaxy maps, i.e., detected, undetected, or all galaxies.

From the luminosities, we calculate the intensity map, either in terms of a specific intensity

$$\begin{aligned} I_\nu &= \frac{\mathrm{d}I}{\mathrm{d}\nu} = \frac{c\rho_L}{4\pi H(z)\nu_0}, \\ \text{or } I_\lambda &= \frac{\mathrm{d}I}{\mathrm{d}\lambda} = \frac{c\rho_L}{4\pi H(z)(1+z)^2\lambda_0}, \end{aligned} \tag{5.26}$$

or brightness temperature

$$T = \frac{\rho_L c^3 (1+z)^2}{8\pi k_\mathrm{B} \nu_0^3 H(z)}, \tag{5.27}$$

where $\rho_L = \sum_{i=0}^{N_\mathrm{g}} L_i/\delta V$ is the total emissivity in each voxel, i.e., the sum of the luminosities of all galaxies in that voxel divided by the voxel volume, and k_B is the Boltzmann constant. This step assumes that the emission line is narrow, i.e. a delta function.

If specified in the input, the intensity map is smoothed with a Gaussian kernel perpendicular to the LOS, imitating the beam smoothing. Along the LOS, one can apply Gaussian or top-hat smoothing, imitating a line-spread function or binning in redshift, wavelength, or frequency channels, respectively (see Section 5.2.2). The LOS smoothing can also be used to model broader lines than a voxel length. One can add random Gaussian noise with the provided standard deviation per voxel σ_I, and subtract the estimated sky foreground (see Section 5.2.3) from the intensity map. If a mask is specified in the input, the galaxy number density and intensity maps are multiplied by the mask. The mask is equivalent to the weights introduced in Section 5.2.1.

	$\bar{z} = 2.22$	$\bar{z} = 3.04$
$L_{\parallel}^{\text{survey}}$ $[h^{-1}$ Mpc]	622	624
N_{mock}	7	9
$S_{\perp}^{\text{survey}}$ $[h^{-2}$ Gpc$^2]$	2.38	3.23
f_V	1.14	1.09
N_{mesh}	311	312
$N_{\text{gal}}/10^6$	14.2	14.4
$\sigma_{\parallel}$ $[h^{-1}$ Mpc]	1.76	1.27
s_f $[h^{-1}$ Mpc]	10.0	11.6
σ_{I_λ} $[10^{-20}$erg s^{-1} cm^{-2} arcsec^{-2} Å$^{-1}]$	2.9	1.5

Table 5.1: Summary of the Differences between the Low-z and High-z HETDEX Mocks.

Finally, one can calculate the summary statistics, i.e. the autopower spectra of galaxies and the intensity and the cross-power spectrum. We compute the intensity and galaxy power spectra and the cross-power spectrum using the estimators defined in Section 5.2.1, where V_{box} is the volume of the simulation box. We use the fast Fourier transform (FFT) to calculate $\widetilde{\delta I}$ and $\tilde{\delta}_{\text{g}}$, keeping only the independent modes, and calculate the corresponding $\mathbf{k}$ and μ values of the cells. For the quadrupole, we multiply the mesh $\widetilde{\delta A}(\mathbf{k})\widetilde{\delta B}^*(\mathbf{k})$ by the second Legendre polynomial evaluated at the μ values of the mesh. For each k bin, we collect cells whose k values fall into the bin and calculate their mean. We calculate the mean k value of each bin in the same way.

5.4 Example Mock Setup: HETDEX

In this section, we introduce the setup for a HETDEX-like Lyα LIM experiment. HETDEX is primarily a cosmological galaxy survey that aims to map approximately one million LAEs through their Lyα emission line at $z \in [1.88, 3.52]$ (Gebhardt et al., 2021). HETDEX uses the Visible Integral-field Replicable Unit Spectrograph (VIRUS; Hill et al., 2021) on the Hobby-Eberly Telescope (HET), which consists of 78 integral-field unit spectrographs (IFUs), each of which contains 448 fibers that are $1.5''$ in diameter, with a spectral resolution of 5.6 Å.

HETDEX observes a total area of $540\,\text{deg}^2$ without target preselection, expecting $460,000$ IFU observations. Because each IFU spans $51'' \times 51''$ in the sky, this amounts to an effective fill factor of the survey of $f_{\text{survey}} \simeq 0.17$. The layout of the IFUs across the $18'$ diameter focal plane leaves an IFU-sized gap between adjacent IFUs and a hole in the center of the focal plane, which is used for other instruments (see Figure 5.1). Three six-minute exposures per HETDEX observation fill in the gaps between fibers, but not between IFUs, so that each individual HETDEX observation has a fill factor of $\simeq 1/4.6$. The HETDEX data reduction pipeline offers a sky-subtraction mode that estimates the sky spectrum from all IFUs simultaneously, i.e., roughly from a discontinuous circular area $\simeq 9'$ in radius.

We divided the redshift range of HETDEX into two parts with a similar LOS distance

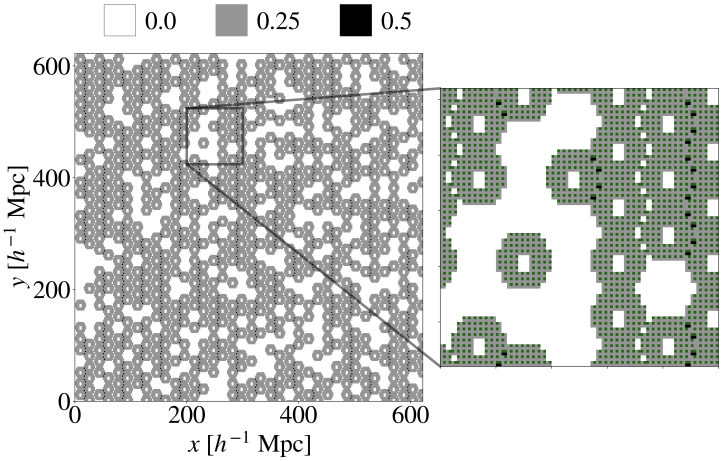

Figure 5.1: On-sky slice of a HETDEX-like mask applied to the HETDEX and validation mocks. The mask is constant in the LOS direction. The right panel is a zoom-in of a small area in the left panel. Each roughly hexagonal element with a hole in the center corresponds to one HETDEX observation with the VIRUS focal plane layout. The green squares in the right panel show the individual IFU positions in the higher-resolution mask that was used to generate the downsampled gray mask. The holes between observations show randomly removed individual observations to match the target fill factor of HETDEX.

of $622\,(624)\,h^{-1}$Mpc for the low-z (high-z) sample. The low-z (high-z) sample covers $z \in [1.88, 2.57]$ ($z \in [2.57, 3.52]$) and is centered around $\bar{z} = 2.22$ ($\bar{z} = 3.04$). At the mean redshift, the survey area of $540\,\mathrm{deg}^2$ translates into a comoving area of $2.38\,h^{-2}\mathrm{Gpc}^2$ ($3.23\,h^{-2}\mathrm{Gpc}^2$). We generated 7000 (9000) mocks of cubic volumes with side length $622\,h^{-1}$Mpc ($624\,h^{-1}$Mpc). By averaging the power spectra over 7 (9), cubic boxes, we effectively obtained 1000 power spectra of a box that is 1.14 (1.09) times the size of the low-z (high-z) volume. In the last step, we multiplied the covariance matrix of each redshift slice by the respective factor to correct for this oversampling. The largest accessible scale with this box size is $k_{\mathrm{min}} = 0.01\,h\mathrm{Mpc}^{-1}$.

We set the cosmological parameters to the fiducial cosmology (see Section 5.1). While the bias of detected LAEs is $\simeq 2$ (Gawiser et al., 2007; Guaita et al., 2010), the bias of LAEs fainter than the detection limit is unknown. To obtain a conservative estimate for the signal-to-noise ratios (S/N) of the HETDEX power spectra, we choose a smaller linear galaxy bias of $b = 1.5$. We simulate the Lyα intensity at the rest-frame wavelength of 1215.67 Å. In the following, we use the specific intensity I_λ. For both redshift sections, we adopt the Lyα luminosity function of the EWgt60 sample of Konno et al. (2016) for galaxies at $z = 2.2$, which is given by a Schechter function,

$$\frac{\mathrm{d}n}{\mathrm{d}L} = \frac{\phi^*}{L^*}\left(\frac{L}{L^*}\right)^{\alpha} e^{-L/L^*}\,, \tag{5.28}$$

with $L^* = 4.87 \times 10^{42}\,\mathrm{erg\,s^{-1}}$, $\phi^* = 3.37 \times 10^{-4}\,\mathrm{Mpc}^{-3}$, and $\alpha = -1.8$. We set the minimum luminosity to $4 \times 10^{40}\,\mathrm{erg\,s^{-1}}$, based on Figure 4 of Gronke et al. (2015), and do not set the maximum luminosity.

Chiang et al. (2013) investigated the impact of sparse sampling for galaxy redshift

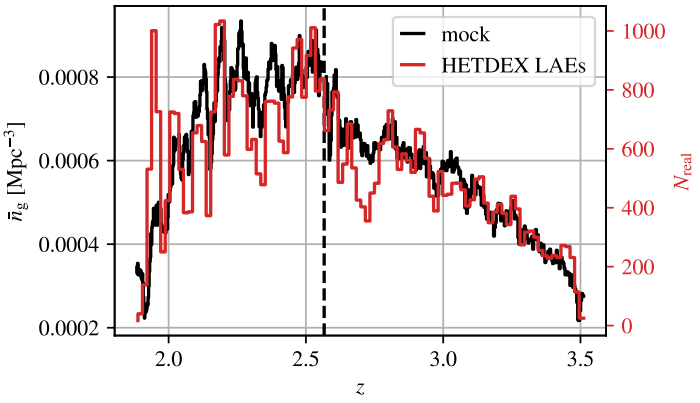

Figure 5.2: Mean detected galaxy number density as a function of redshift in one low-z and one high-z HETDEX-like mock (black). The dashed line shows the transition from the low-z to the high-z mock. The red histogram shows the distribution of detected LAEs in HETDEX with S/N $>$ 5.5 from the HETDEX public source catalog I (Mentuch Cooper et al., 2023b).

surveys, in particular HETDEX, and found that the voxel size used for the power spectrum calculation has to be at least twice as large as the separation between IFUs to obtain an unbiased power spectrum. For this reason, we grid the simulation box with a resolution of $2\,h^{-1}$Mpc, i.e. $N_{\rm mesh} = 311\,(312)$, which corresponds to a Nyquist frequency of $k_{\rm Ny} = 1.57\,h$Mpc^{-1}. Each voxel in our mock encompasses roughly two IFU side lengths (102″) perpendicular to the LOS and two spectral bins (4 Å) along the LOS, which is smaller than the spectral resolution. This means that in our HETDEX-like experiment, one has to average the intensity of the fibers in each of the larger voxels.

Each cubic box at low-z (high-z) contains roughly 14.2 (14.4) million galaxies, both detected and undetected. We use a wavelength-dependent flux limit, enforcing that there are 2.5 detected galaxies per IFU on average, summed over the entire HETDEX LOS range (see Gebhardt et al., 2021). This corresponds to a mean galaxy number density of $\bar{n}_{\rm g} \simeq 2.0 \times 10^{-3}\,h^3$Mpc^{-3}. We rescaled the measured flux noise standard deviation from Figure 18 in Hill et al. (2021) to obtain a wavelength-dependent flux limit that satisfies this condition.

Figure 5.2 shows the resulting mean $\bar{n}_{\rm g}(z)$ in one realization of low-z and high-z HETDEX mocks. We compare this to the distribution of LAEs from the HETDEX public source catalog I (Mentuch Cooper et al., 2023b), where we selected LAEs with the 'lae' flag in the 'source_type' column. The shapes agree well. A decrease in the detected LAEs at $z \simeq 2.7$ is caused by a mask applied at the center of 50% of the HETDEX detectors and an increase in night-sky emission. The real number density in the high-z volume can differ if the luminosity function at $z \simeq 3$ is different from that at $z \simeq 2$. However, because of the good agreement with the detected galaxies in HETDEX, we continue to use the luminosity function at $z = 2.2$.

The high angular resolution of HETDEX allows one to mask individual detected galaxies without masking entire voxels. Therefore, we focus on the intensity map of only undetected sources. This intensity map does not share galaxies with the catalog of detected galaxies, so there is no contribution of shot noise to the cross-power spectrum between galaxies and intensity.[3]

[3]Combining the catalog of detected galaxies with the full intensity map of all sources adds information with respect to the autopower spectrum of detected galaxies alone. This is true even if the intensity map is dominated by bright detected galaxies (i.e., if the line luminosity function flattens at the faint end)

We do not apply beam smoothing perpendicular to the LOS because the PSF of VIRUS is smaller than the size of an IFU/voxel. We smooth the intensity map along the LOS with a Gaussian kernel with standard deviation $\sigma = 2.38\,\text{Å}$(FWHM $= 5.6\,\text{Å}$) to imitate the line-spread function, i.e., spectral resolution, of VIRUS (Hill et al., 2021). This corresponds to $\sigma_\parallel = 1.76\,h^{-1}\text{Mpc}$ ($1.27\,h^{-1}\text{Mpc}$) for the low-z (high-z) volumes. Because of the low spectral resolution and the large size of the voxels along the LOS, we do not model the broadening of the Lyα line. For the sky subtraction, we set the focal plane radius to $9'$, which corresponds to a distance scale of $s_\text{f} = 10.0\,h^{-1}\text{Mpc}$ ($11.6\,h^{-1}\text{Mpc}$) for the low-z (high-z) part.

To add HETDEX-like noise, we transform the measured wavelength-dependent 5σ sky flux noise per resolution element per fiber in VIRUS (see Fig. 18 in Hill et al., 2021), $5\sigma_F$, into a specific intensity noise per fiber, σ_{I_λ}, by dividing by $\left(5 \times \pi(0.75'')^2 \times 5.6\,\text{Å}\right)$. Then we divide by $\sqrt{3 \times 448}$ to account for the averaging over fibers within an IFU in three dithers. The simulation voxels are shorter along the LOS than one spectral resolution element of VIRUS (5.6 Å; see Hill et al., 2021). Because the noise below this scale is correlated, the factor of $\sqrt{\frac{5.6\text{Å}}{\Delta\lambda}}$ to obtain the noise in the voxel with LOS length $\Delta\lambda$ is an overestimate of the correlated noise. We therefore do not apply this factor. Because the area of a voxel encompasses four IFU areas, we divide this noise level by $\sqrt{N_\text{IFU}(\mathbf{x})}$. Here, $N_\text{IFU}(\mathbf{x})$ is the number of IFUs observed in the voxel at position $\mathbf{x}$, i.e. where the supersampled mask described below is nonzero. This can in principle be an integer between zero and four; for our masks, it is ≤ 2. In summary, we convert the 5σ flux noise per resolution element $5\sigma_F$ into the intensity noise at position $\mathbf{x}$ by calculating

$$\sigma_{I_\lambda}(\mathbf{x}) = \frac{5\sigma_F}{5\pi(0.75'')^2\sqrt{3 \times 448 \times N_\text{IFU}(\mathbf{x})} \times 5.6\,\text{Å}}. \tag{5.29}$$

This results in $\sigma_{I_\lambda} = 2.9\,(1.5) \times 10^{-20}\,\text{erg}\,\text{s}^{-1}\,\text{cm}^{-2}\,\text{arcsec}^{-2}$ in each voxel in the low-z (high-z) boxes on average.

To obtain a mask on the coarse grid, we first generate a HETDEX-like mask with double resolution, so that each cell corresponds roughly to one IFU. We generate a mask of VIRUS-like tiles of ones on and zeros in between IFUs and keep the mask constant along the LOS. If we filled the entire area with observations, we would have a fill factor of $f_\text{obs} = 0.23$, similar to the focal plane fill factor of VIRUS $f_\text{obs}^\text{VIRUS} = 1/4.6 = 0.22$. To match the effective fill factor of $f_\text{survey} \simeq 0.17$ due to sparser observations, we randomly remove $\simeq 26\%$ of the individual observations after applying the VIRUS-like mask. Then we downsample this mask to the same resolution as our simulated maps by averaging eight adjacent cells, which is equivalent to NGP assignment. The result is shown in Figure 5.1. We generate 7 (9) separate masks for the low-z (high-z) boxes and then average the power spectra of the boxes with different masks.

We calculate the power spectrum monopoles and quadrupoles in linearly spaced bins from $k_\text{min} = 0.04\,h\text{Mpc}^{-1}$ to $k_\text{max} = 1\,h\text{Mpc}^{-1}$ with $\Delta k = 0.04\,h\text{Mpc}^{-1}$. We summarize the differences between the low-z and high-z HETDEX mocks in Table 5.1.

because intensity fluctuations have a different (luminosity-weighted) bias from number-count fluctuations. We leave the study of this case for future work.

5.5 Validation

In this section, we show that the results of our simulations in real space agree with the expected power spectra given in Section 5.2. We perform this comparison in real space because the lognormal algorithm precisely reproduces the input power spectra in real space, while the redshift-space power spectra deviate from the expectation in linear approximation (see Agrawal et al., 2017).

The setup of the validation mocks is almost identical to the low-z HETDEX-like mocks described in Section 5.4. However, we reduce the intensity noise by a factor of 300 and the flux limit by a factor of 5 to reduce the shot noise for validation, and do not apply sky subtraction. We apply one of the HETDEX-like masks to all intensity maps. We calculate the validation power spectra in real space and average over 1000 mocks with side length $622\,h^{-1}\,\mathrm{Mpc}$. All other input parameters are the same. We calculate the intensity-intensity, galaxy-galaxy, and galaxy-intensity power spectrum monopoles and quadrupoles, where undetected galaxies contribute to the intensity map.

To obtain the model, we evaluate the input power spectrum that includes the galaxy bias on a mesh with the same $\mathbf{k}$ and μ values as obtained from FFT of the mock maps. We also evaluate the damping functions from the intensity smoothing on this mesh. We calculate the window function by multiplying the mask by the mean expected intensity or the galaxy number density expected from the flux limit and the luminosity function as a function of redshift. We use FFT for the convolution of the power spectrum with the window function, add the shot noise, and multiply the result by the damping functions. Then we add the intensity noise power spectrum to the intensity autopower spectrum mesh. We calculate the power spectrum multipoles of these results in the same way as for the mock (see Section 5.3).

Figure 5.3 shows the power spectrum monopole and the quadrupole in real space measured from the validation mocks. The shot-noise (intensity noise) power spectrum is subtracted from the galaxy (intensity) autopower spectrum monopole. The shot noise is not subtracted from the intensity autopower spectrum because it contains information about the luminosity function. The quadrupole is affected by the anisotropic mask, the selection function, and the smoothing. It also shows the analytical prediction presented in Section 5.2, as well as the relative residuals between the power spectra of the mock and the model. The measured real-space power spectra from the mock agree with the model at all k modes. We also tested the validation in the cases where all or only detected galaxies contribute to the intensity map and find excellent agreement.

5.6 HETDEX Forecast

In this section, we forecast the sensitivity of HETDEX to LIM power spectra. We use the mock setup described in Section 5.4 in redshift space and calculate the galaxy and intensity autopower spectra, as well as their cross-power spectrum with and without sky subtraction. As explained earlier, only undetected sources contribute to the intensity map.

The upper panels of Figure 5.4 show the monopole and quadrupole power spectra in

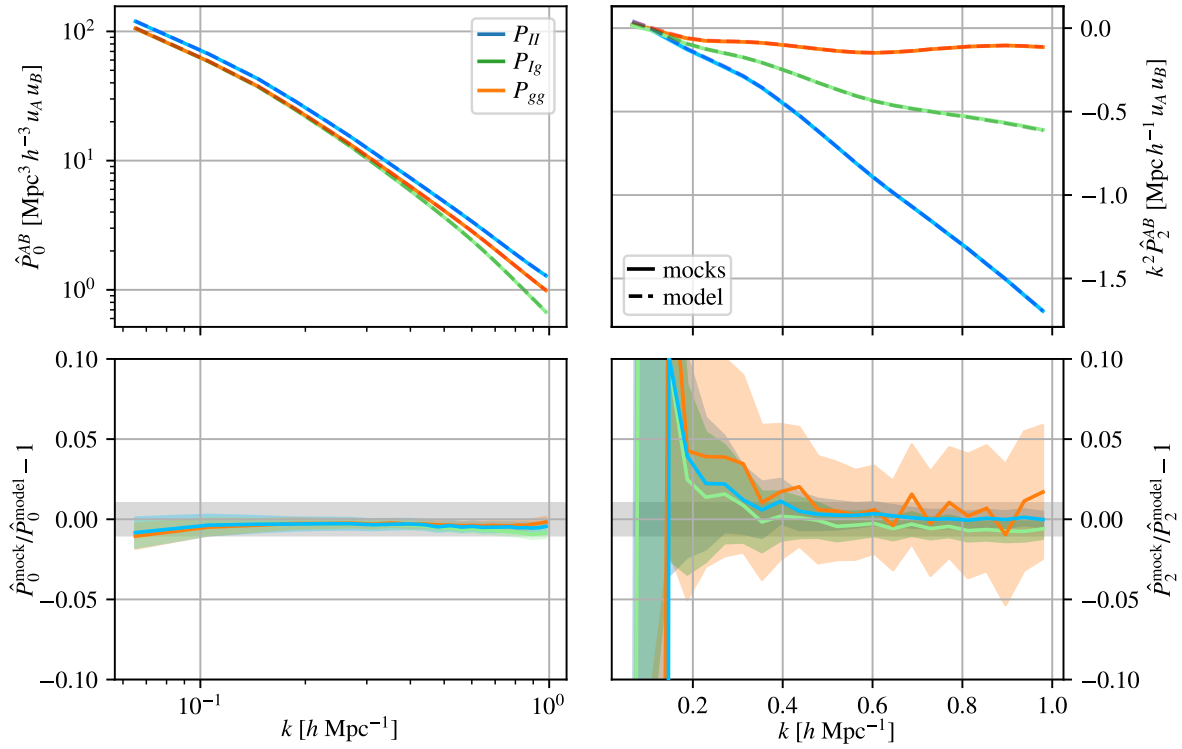

Figure 5.3: Comparison of the mock power spectra (solid lines) with the analytical model (dashed lines) of the monopole (left) and the quadrupole (right) power spectra in real space. The shot-noise (intensity noise) power spectrum is subtracted from the galaxy (intensity) autopower spectrum monopole. The bottom panels show the relative residuals between the mock and model power spectra. The units are $u_I = \langle I_\lambda \rangle \simeq 2.8 \times 10^{-23}\,\mathrm{erg\,s^{-1}\,cm^{-2}\,arcsec^{-2}}$ and $u_\mathrm{g} = 1$. The shaded areas show the 1σ error of the mean given by the standard deviation of the different realizations divided by the square root of the number of realizations. The gray area shows deviations within 1%.

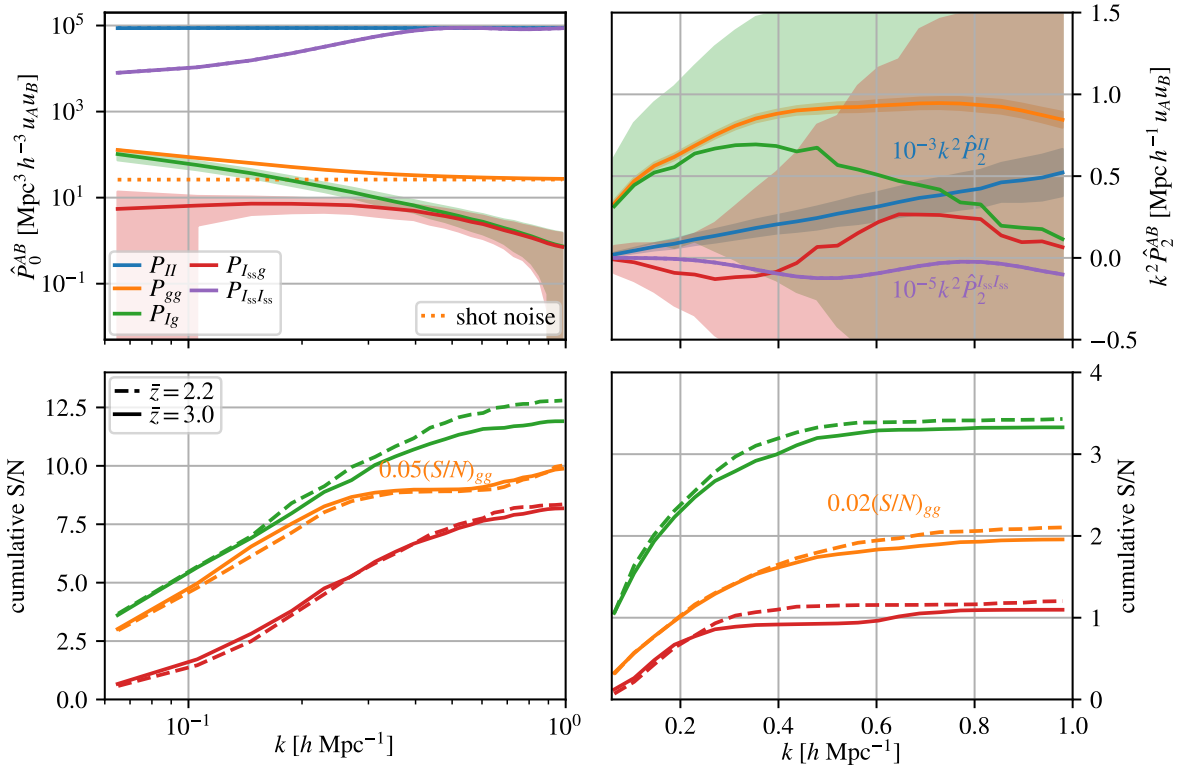

Figure 5.4: Power spectra of the HETDEX-like mocks in redshift space. The top left (right) panel shows the monopole (quadrupole) of the intensity (blue), galaxy (orange), and sky-subtracted intensity (purple) autopower spectra, and the intensity-galaxy cross-power spectra without (green) and with (red) sky subtraction in the high-z volume. The quadrupole of the normal and sky-subtracted intensity autopower spectra were multiplied by factors of 10^{-3} and 10^{-5}, respectively, for better visualization. The shaded areas show the square root of the diagonal elements of the corresponding covariance matrices. The units are $u_I = u_{I_{ss}} = \langle I_\lambda \rangle \simeq 2.9 \times 10^{-23}\,\mathrm{erg\,s^{-1}\,cm^{-2}\,arcsec^{-2}}$ and $u_g = 1$. The dotted lines show the intensity noise (indistinguishable from the total power spectra) and the galaxy shot-noise power spectra. The bottom panels show the cumulative S/N up to a given k bin after subtracting the shot noise from the galaxy autopower spectrum monopole and and quadrupole. The solid (dashed) lines correspond to the high-z (low-z) HETDEX-like mocks. The cumulative S/N of the intensity autopower spectra are not shown because they are zero. The cumulative S/N of the galaxy autopower spectrum monopole (quadrupole) is multiplied by 0.05 (0.02) for better visualization.

redshift space. The bottom panels show the cumulative S/N calculated as

$$\mathrm{S/N}^2_{AB,\ell}(k_N) = \bar{\mathbf{\Theta}}^T_{AB;\ell} \left(C^{\mathbf{\Theta}}_{AB,\ell}\right)^{-1} \bar{\mathbf{\Theta}}_{AB,\ell}, \tag{5.30}$$

where k_N denotes the maximum wavenumber considered. Here, $\bar{\mathbf{\Theta}}_{AB,\ell} = M^{-1} \sum_{i=1}^{M} \mathbf{\Theta}^{(i)}_{AB,\ell}$ is the mean of

$$\mathbf{\Theta}^{(i)}_{AB,\ell} = \left(\hat{P}^{(i)}_{AB,\ell}(k_0), \hat{P}^{(i)}_{AB,\ell}(k_1), \ldots, \hat{P}^{(i)}_{AB,\ell}(k_N)\right) - \mathbf{Q} \tag{5.31}$$

over M realizations, where $\hat{P}^{(i)}_{AB,\ell}$ is the estimator of the power spectrum monopole ($\ell = 0$) or quadrupole ($\ell = 2$) calculated from the ith realization of maps $A, B \in \{I, \delta_\mathrm{g}\}$.

Here, $\mathbf{Q}$ is the constant shot noise of the galaxy autopower spectrum monopole, the noise power spectrum for the intensity autopower spectrum monopole, and zero otherwise. The noise power spectra of the intensity autopower spectra were calculated by generating 7000 (9000) pure noise mocks in the low-z (high-z) bin, performing the sky subtraction, calculating their monopole and quadrupole power spectra, and averaging them over the realizations. For simplicity, we leave out the subscript AB, ℓ from now on. We do not subtract the shot noise from the intensity autopower spectrum because it contains information about the luminosity function. Because the intensity autopower spectra are indistinguishable from the intensity noise power spectra, the cumulative S/N are not shown in the figure.

$\left(C^{\mathbf{\Theta}}_{k_N}\right)^{-1}$ is the inverse of the covariance matrix, whose elements are defined as

$$C^{\mathbf{\Theta}}_{mn} = \frac{f_V}{M-1} \sum_{i=1}^{M} \left(\mathbf{\Theta}^{(i)}_m - \bar{\mathbf{\Theta}}^{(i)}_m\right) \left(\mathbf{\Theta}^{(i)}_n - \bar{\mathbf{\Theta}}^{(i)}_n\right). \tag{5.32}$$

Only elements of the covariance matrix up to the maximum k_N are considered for matrix inversion to obtain $\left(C^{\mathbf{\Theta}}_{k_N}\right)^{-1}$. As explained in Section 5.4, we multiply the covariance matrix by the factor $f_V = 1.14$ (1.09) at low-z (high-z) to correct for slightly larger simulated volumes than the HETDEX volume.

We predict that an ideal HETDEX-like experiment can detect the galaxy-intensity cross-power spectrum monopole at $\simeq 80\sigma$ in each redshift bin of the survey despite the significant loss of large-scale power from the sky subtraction. The quadrupole of the galaxy-intensity cross-power spectrum is not detectable (the S/N is 1). The intensity autopower spectrum also cannot be detected.

Figure 5.5 shows the correlation matrices of the data vectors

$$\mathbf{\Phi} = (P_0(k_0), P_0(k_1), \ldots, P_0(k_N), P_2(k_0), \ldots, P_2(k_N)) \tag{5.33}$$

for the different power spectra that include the shot noise in the high-z HETDEX mock. The correlation matrix R of a vector $\mathbf{\Phi}$ is given by $R^{\mathbf{\Phi}}_{mn} = C^{\mathbf{\Phi}}_{mn} / \sqrt{C^{\mathbf{\Phi}}_{mm} C^{\mathbf{\Phi}}_{nn}}$, where the covariance matrix is given in Eq. (5.32).

The galaxy autopower spectrum monopole has a large off-diagonal correlation that increases with increasing k. This is mainly due to the mode coupling introduced by the convolution with the complicated window function. The galaxy autopower spectrum monopole and quadrupole have a systematic low-level cross correlation that is visible as

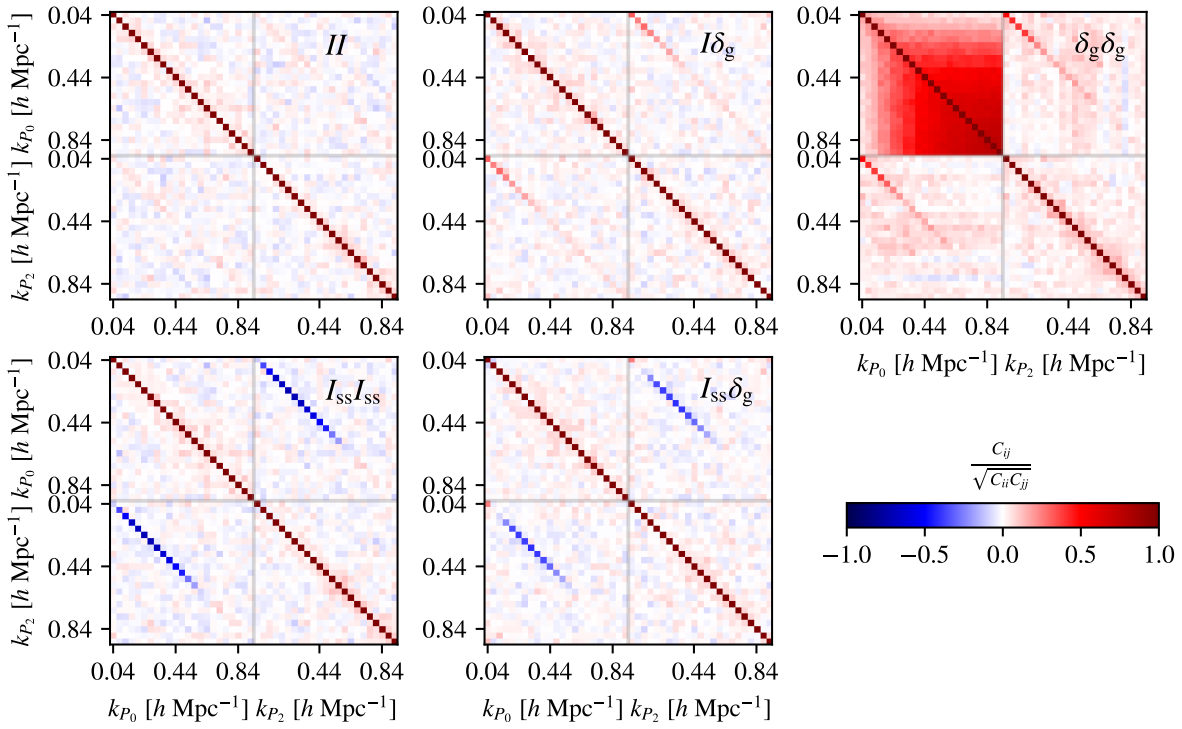

Figure 5.5: Correlation matrices of $\mathbf{\Phi} = (P_0(k_0), P_0(k_1), \ldots, P_0(k_N), P_2(k_0), \ldots, P_2(k_N))$ for the intensity autopower spectrum, cross-power spectrum, galaxy autopower spectrum, sky-subtracted intensity autopower spectrum, and sky-subtracted cross-power spectrum (from top left to bottom right) of the high-z HETDEX mock.

stripes at constant k of the quadrupole. This may result from the imperfect integration of the μ values, in which the Legendre polynomials are no longer orthogonal (see Appendix D.2.3 of Agrawal et al., 2017).

The correlation matrices of the intensity autopower spectrum and the cross-power spectrum without sky subtraction are dominated by diagonal elements. They have an off-diagonal correlation between the monopole and quadrupole power spectra at the same k, which is slightly higher than the noise around most off-diagonal elements. Although this monopole-quadrupole correlation is positive without sky subtraction, it is larger and negative with sky subtraction. The off-diagonal correlation for the intensity autopower spectrum monopole is negligible because the covariance is dominated by the uncorrelated intensity noise. Because of the strong non-Gaussianity of lognormal realizations, the off-diagonal elements of the covariance matrix are overestimated (Blot et al., 2019). SIMPLE therefore returns a conservative estimate of the S/N. This effect is most relevant for the galaxy power spectrum, the correlation matrix of which has the largest off-diagonal terms, as shown in Figure 5.5.

5.7 Discussion

The SIMPLE framework is a simple and fast LIM simulation scheme with two main approximations. First, the galaxy distribution is modeled by a lognormal random field following the input power spectrum that inculdes a linear bias (Agrawal et al., 2017). The galaxies are obtained using Poisson sampling, and the galaxy velocities follow the linear continuity equation. By construction, it is only accurate on the scales on which the input power spectrum is accurate, which, for example, is only true on linear scales for the Eisenstein & Hu fitting function (Eisenstein & Hu, 1998). We may also underestimate the small-scale power spectrum because we do not model the one-halo term, nonlinear bias, or assembly bias. It is possible to improve the density and velocity distribution of mocks at the expense of decreasing the speed of the code, for example by using other dark matter simulation methods described in the introduction.

The second approximation of the SIMPLE framework is that the luminosity of a galaxy is randomly sampled from a luminosity function. Therefore, it does not depend on galaxy properties such as the star-formation rate or on its environment. This procedure misses any nontrivial connection between a specific galaxy population and their clustering, limiting any astrophysical analysis to the luminosity function. For instance, SIMPLE can be used to study the performance of a given summary statistic to constrain the input luminosity function. Alternatives such as an empirical determination of the astrophysical properties, semi-analytic models of galaxy evolution, or hydrodynamical simulations defeat the purpose of SIMPLE, given their computational cost.

5.7.1 Limitations Specific to the Lyα Line

Our approach only models the emission from galaxies that is contained in the luminosity function. This does not capture all relevant physical emission processes of the Lyα line, as it neglects the recombination of ionized hydrogen and the collisional excitation of neutral hydrogen in the CGM and IGM (e.g., Dijkstra, 2019). Photons originating from a galaxy can also be observed far away from its source due to scattering in the CGM and IGM (Byrohl et al., 2021). Byrohl & Nelson (2023) find that while photons produced through diffuse emission in the IGM are negligible in the global Lyα luminosity budget, photons that originate from the ISM or CGM of galaxies and scatter in the IGM contribute substantially to the Lyα luminosity budget. One can test the significance of the scattered contributions for LIM by modifying the luminosity function to include emission from the CGM calibrated on observations or by smoothing a small fraction of the intensity map to imitate scattering on larger scales than the voxel size. One can also add an intensity component that is directly proportional to the matter distribution to mimic diffuse gas emission.

Scattering of Lyα photons in the ISM strongly affects the escape fraction of Lyα photons and the emission peak wavelength (e.g., Hashimoto et al., 2013; Blaizot et al., 2023). The luminosity function contains the observed Lyα luminosity of detected galaxies, which consists only of photons that escaped from the ISM. However, the SIMPLE framework does not account for the noncosmological redshift of the Lyα line due to radiative transfer within the ISM, which may cause an anisotropic effect similar to the

Fingers-of-God effect (Byrohl et al., 2019). It also does not account for radiative transfer in other environments, especially absorption, which may cause an anisotropic effect similar to the RSD, but with the opposite sign (e.g., Zheng et al., 2011; Behrens et al., 2018). These two effects can be modeled in an extension of the SIMPLE code by shifting the Lyα line and calculating the optical depth along the LOS as a function of density and velocity, which can be obtained from a lognormal simulation.

5.7.2 Limitations of the HETDEX Forecast

This HETDEX forecast is meant as an order-of-magnitude forecast for the detectability of the cross-power spectrum of sky-subtracted intensity and detected galaxies. There are several possible improvements for a more accurate forecast in addition to the above Lyα modeling improvements: one can account for the evolution of the luminosity function; include Galactic extinction; include interloper contamination of the intensity map, specifically [O II]-emitting galaxies at $z < 0.5$, and masking of bright interlopers; model the mask, selection function, and non-Gaussian noise more accurately using the HETDEX data; account for false-positive galaxy detections; and model the intensity data reduction more accurately. We leave these improvements for future work.

5.8 Summary and Conclusions

We have presented the publicly available SIMPLE code for quickly generating intensity maps that include observational effects such as noise, anisotropic smoothing, sky subtraction, and masking. It is based on a lognormal simulation of galaxies and random assignment of luminosities to these galaxies. Although this approach does not contain the dependence of the line luminosity on galaxy properties or of the galaxy properties on the cosmological environment, it provides a fast and versatile way to generate mock intensity maps. These can be used to study survey systematics and calculate covariance matrices of power spectra.

We also derived an analytical model and showed that sky subtraction suppresses the power spectrum on scales larger than the focal plane size of the telescope. We validated the SIMPLE code by showing that the power spectra in real space agree precisely with those of the analytical model. This is the advantage of a lognormal mock, where the output power spectrum is designed to agree with the input power spectrum in real space (Agrawal et al., 2017).

As an application, we generated mock intensity and galaxy number density maps for a HETDEX-like LIM survey in redshift space that included a realistic mask, selection function, and intensity noise. We calculated the cross-power spectra of detected galaxies and the intensity of undetected galaxies after subtracting the sky spectrum and the respective covariance matrix in two redshift regimes within the survey. We predict that HETDEX will detect the cross-power spectrum monopole.

In summary, SIMPLE has been designed to provide fast but reliable realizations that allow for changes in cosmology, luminosity function, and observational specifications with little effort. For example, one can numerically estimate covariance matrices for different

setups. Because lognormal realizations overestimate the off-diagonal elements of the covariance matrix due to strong non-Gaussianity, SIMPLE returns a conservative estimate especially for the galaxy autopower spectrum (Blot et al., 2019). Therefore, this code is an excellent complement to slower, more physical simulations with the potential to guide the analysis of LIM surveys.

Acknowledgments

We thank S. Saito for helpful comments on the draft. M.L.N. thanks L. Blot, M. Černetič, M. Fabricius, D. Farrow, G.J. Hill, D. Jeong, and J. Niemeyer for interesting and useful discussions. J.L.B. acknowledges funding from the Ramón y Cajal Grant RYC2021-033191-I, financed by MCIN/AEI/10.13039/501100011033 and by the European Union "NextGenerationEU"/PRTR.

We acknowledge the use of the astropy (Astropy Collaboration et al., 2013, 2018), matplotlib (Hunter, 2007), numpy (Harris et al., 2020), scipy (Virtanen et al., 2020), pmesh (Feng et al., 2017), dask (Rocklin, 2015), and h5py (Collette, 2013) software packages.

Chapter 6

Effect of Lyα Radiative Transfer on Intensity Mapping Power Spectra

The content of this chapter has been published in Lujan Niemeyer (2025).

Clustering of Lyα-emitting galaxies (LAEs) and Lyα line intensity mapping (LIM) are useful probes of cosmology. However, Lyα radiative transfer (RT) effects, such as absorption, line shift, and line broadening, and their dependence on the large-scale density and velocity fields can modify the measured LAE clustering and LIM statistics. We explore the effects of RT on the Lyα LIM power spectrum in two ways: using an analytic description based on linear approximations, and using lognormal mocks. The qualitative effects of intergalactic Lyα absorption on the LIM auto- and cross-power spectrum include a scale-dependent, reduced effective bias, reduced mean intensity, and modified redshift-space distortions. The linear absorption model does not describe the results of the lognormal simulations well. The random line shift suppresses the redshift-space power spectrum similar to the Fingers-of-God effect. In cross-correlation of LAEs or Lyα intensity with a non-Lyα tracer, the Lyα line shift leads to a phase shift of the complex power spectrum, i.e. a cosine damping of the real part. We study the impact of Lyα RT effects on the Hobby-Eberly Telescope Dark Energy Experiment (HETDEX) LAE and LIM power spectra using lognormal mocks. We find that even small amounts of IGM absorption will significantly change the measured LAE auto-power spectrum. The LAE-intensity cross-power spectrum stays within the measurement uncertainty. Therefore, HETDEX will be able to constrain Lyα RT effects.

6.1 Introduction

The Lyα emission line is an excellent tool for cosmology at high redshift (e.g., Partridge & Peebles, 1967). Detected Lyα-emitting galaxies (LAEs) are used to measure their clustering and constrain cosmological parameters (Ouchi et al., 2020; Gebhardt et al., 2021). Instead of detecting individual LAEs in deep observations with high resolution, one can also map the total Lyα intensity in noisy, low-resolution observations to constrain cos-

mological parameters, called line intensity mapping (LIM, e.g., Bernal & Kovetz, 2022).

Neutral hydrogen has a large scattering cross section around the Lyα line. The radiative transfer (RT) complicates the interpretation of measurements using Lyα emission. For example, using simulations, Zheng et al. (2011) find a strong correlation between the Lyα optical depth in the intergalactic medium (IGM) and the large-scale density and velocity structure. They predict that the anisotropic dependence of the observed fraction of LAEs suppresses line-of-sight (LOS) density fluctuations, makes the effective bias scale-dependent, and can even 'invert' the so-called Kaiser effect (Kaiser, 1987), the linear redshift-space distortions (RSD).

One can model the effect of IGM absorption on the LIM power spectra in various ways. Wyithe & Dijkstra (2011) (henceforth WD11) and Greig et al. (2013) derive an analytic model for this effect for the power spectrum and bispectrum. The analytic absorption model for the power spectrum in the first part of WD11 is based on linear approximations for the dependence of the Lyα transmittance on the matter, ionization rate, and velocity distributions. While this explains the qualitative effect of IGM absorption on the power spectrum, the amplitude is determined by three free parameters: the mean optical depth τ_0, the fraction of Lyα photons subject to IGM absorption $F_{\rm abs}$, and the smoothing kernel of the ionization rate with respect to the galaxy distribution. The linear approximations are also only valid when the matter overdensity $\delta_{\rm m}$, the ionization rate perturbations δ_Γ, and the velocity gradient perturbations δ_v are small, which is not the case in the immediate environments of galaxies. WD11 also present a more detailed analytic model that is based on assumptions on the density profile, ionization rate, temperature, and gas velocities in the environment of the LAEs.

Using a cosmological hydrodynamical simulation and post-processing it with Lyα RT, as done by Zheng et al. (2011), Behrens & Niemeyer (2013) and Behrens et al. (2018), may provide the most realistic estimate for the optical depth and the effect on the observed fluxes and the power spectrum, if it accurately simulates the matter and velocity structure within and outside of the galaxy halos. However, the results of these simulations are dependent on the resolution: using an RT simulation with higher resolution, Behrens et al. (2018) find little correlation between the large-scale environment and the observed fraction of LAEs, while they reproduce the results of Zheng et al. (2011) when they degrade the resolution of the simulation.

Gurung-López et al. (2020) develop a semi-analytic model for Lyα RT in the interstellar medium (ISM) and IGM. They find that, at low redshift ($z \in [2.2, 3]$), the spatial distribution of LAEs is independent of IGM properties. However, at $z = 5.7$, the LOS velocity and density gradients modify the clustering of LAEs on large scales in an isotropic fashion.

Another effect of the Lyα RT in the ISM and circumgalactic medium (CGM) is the broadening and the shift of the Lyα line, typically toward the red (e.g., Nakajima et al., 2018b). Byrohl et al. (2019) find in a cosmological RT simulation that this wavelength shift is independent of the large-scale velocities, and they show that the shift adds a Fingers-of-God-like damping to the LAE power spectrum.

Croft et al. (2016) observationally find a strong effect of RT on the cross-correlation of quasars with Lyα intensity. However, more recent studies are consistent with the absence

of RT effects (Croft et al., 2018; Lin et al., 2022). More observations of LAE clustering or Lyα LIM are necessary to constrain RT effects. The Hobby-Eberly Telescope Dark Energy Experiment (HETDEX; Gebhardt et al., 2021) uses integral-field spectrographs to find $\simeq 10^6$ LAEs without target preselection in a $10.9\,\mathrm{Gpc}^3$ comoving volume. Its primary goal is to use LAE clustering statistics such as the power spectrum to constrain cosmological parameters, especially the dark energy equation of state (Shoji et al., 2009). The blind nature of HETDEX enables Lyα LIM studies that may also be affected by Lyα RT effects. We therefore explore these effects on LIM power spectra and estimate the sensitivity of a HETDEX-like survey to these effects. Other Lyα LIM experiments (e.g., Doré et al., 2014, 2016, 2018; Renard et al., 2021; Renard et al., 2024) may also be affected by Lyα RT.

In this paper, we explore the effect of Lyα absorption in the IGM and the line shift and broadening on the Lyα intensity auto- and cross-power spectra. We use lognormal simulations for the forecast. SIMPLE[1] is a fast simulation tool for self-consistently generating galaxy catalogs and intensity maps in redshift space given an input power spectrum and luminosity function (Lujan Niemeyer et al., 2023). It is based on a lognormal galaxy catalog generator[2] (Agrawal et al., 2017), assigns luminosities, determines the detectability of each galaxy, and generates an intensity map. One can apply smoothing and a mask and add noise to make the mocks more realistic. Because the matter density and velocity fields are output by the lognormal galaxy simulations, one can self-consistently calculate the Lyα optical depth in each resolution element and attenuate the luminosities accordingly to simulate IGM absorption. Adding a random line shift and broadening is also straightforward.

A drawback of hydrodynamic simulations and semi-analytic models is their computational cost, which makes it unfeasible to generate enough realizations to calculate a covariance matrix and make sensitivity forecasts. Because lognormal simulations are fast, it is possible to generate enough mocks to calculate the covariance matrix for the HETDEX survey and make predictions for its sensitivity to Lyα RT effects.

The rest of the paper is organized as follows. Section 6.2 builds on the model by WD11 to develop an analytic model for the IGM absorption for LIM auto- and cross-power spectra. Section 6.3 extends the work of Byrohl et al. (2019) on the effect of the line shift and broadening of the Lyα line to LIM power spectra. Section 6.4 describes the modifications to the SIMPLE code to incorporate IGM absorption and a Lyα velocity shift, and shows the effects on the different power spectra using the lognormal simulations. Section 6.5 analyzes the sensitivity of a HETDEX-like experiment to these effects using SIMPLE mocks for the HETDEX survey. Section 6.6 discusses the shortcomings of this approach. We conclude in Section 6.7.

We use the following Fourier convention:

$$\begin{aligned} \tilde{f}(\mathbf{k}) &= \int \mathrm{d}^3\mathbf{x} f(\mathbf{x}) e^{i\mathbf{k}\cdot\mathbf{x}}\,, \\ f(\mathbf{x}) &= \int \frac{\mathrm{d}^3\mathbf{k}}{(2\pi)^3} \tilde{f}(\mathbf{k}) e^{-i\mathbf{k}\cdot\mathbf{x}}\,, \end{aligned} \tag{6.1}$$

[1] `https://github.com/mlujnie/simple`

[2] `https://bitbucket.org/komatsu5147/lognormal_galaxies`

where the tilde denotes quantities in Fourier space. We refer to real space in contrast to redshift space, and to configuration space in contrast to Fourier space.

Throughout this paper, we assume a flat Λ cold dark matter (ΛCDM) cosmology with $H_0 = 67.66\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_{\mathrm{b},0}h^2 = 0.022$, $\Omega_{\mathrm{m},0}h^2 = 0.142$, $\ln(10^{10}A_s) = 3.094$, and $n_s = 0.9645$.

6.2 Intergalactic Lyα Absorption

A Lyα photon escaping the CGM of a galaxy that is close enough to the Lyα line center can scatter off of neutral hydrogen on the IGM. Although the photon scatters out of the LOS, we refer to it as absorption in this work.

The optical depth for a photon with the initial frequency ν on its way from a galaxy's virial radius to the observer is

$$\tau(\nu) = \int_{r_{\mathrm{vir}}}^{\infty} \mathrm{d}r\, n_{\mathrm{H}}(r) x_{\mathrm{HI}}(r) \sigma_\alpha \left[\nu\left(1 + \frac{v_z(r)}{c}\right)\right], \tag{6.2}$$

where $n_{\mathrm{H}}(r)$ is the hydrogen number density, $x_{\mathrm{HI}}(r)$ is the neutral fraction of hydrogen, $v_z(r)$ is the LOS velocity of the gas at distance r from the galaxy, where $v_z > 0$ if it moves away from the observer, and $\sigma_\alpha(\nu)$ is the Lyα absorption cross section at frequency ν (see, e.g., WD11). Using $x_{\mathrm{HI}} = n_{\mathrm{H}}\alpha^{\mathrm{A}}_{\mathrm{rec}}/\Gamma$ at photoionization equilibrium, where Γ is the photoionization rate and $\alpha^{\mathrm{A}}_{\mathrm{rec}} \approx 4.18 \times 10^{-13}\,\mathrm{cm^3 s^{-1}}$ is the case-A recombination coefficient at temperature $T \approx 10^4\,\mathrm{K}$ (Burgess, 1965; Draine, 2011), we obtain

$$\tau(\nu) = \int_{r_{\mathrm{vir}}}^{\infty} \mathrm{d}r\, \frac{n^2_{\mathrm{H}}(r)\alpha^{\mathrm{A}}_{\mathrm{rec}}\sigma_\alpha\left[\nu\left(1 + \frac{v_z(r)}{c}\right)\right]}{\Gamma(r)}. \tag{6.3}$$

Because the Lyα cross section is within the integral, we can approximate it as a Dirac delta function:

$$\begin{aligned}\sigma_\alpha(\nu) &\approx f_\alpha \pi \frac{e^2}{m_e c}\delta(\nu - \nu_\alpha) \\ &= f_\alpha \pi r_e c \delta(\nu - \nu_\alpha) =: \sigma^{\mathrm{tot}}_\alpha \delta(\nu - \nu_\alpha),\end{aligned} \tag{6.4}$$

where ν_α is the Lyα rest-frame frequency, e is the electron charge, m_e is the electron mass, $r_e = 2.81 \times 10^{-13}\,\mathrm{cm}$ is the classical electron radius defined by $\frac{e^2}{r_e} = m_e c^2$, and $f_\alpha = 0.4167$ is the oscillator strength of the Lyα transition (e.g., WD11; Bartelmann, 2021). Inserting this in equation (6.3) and integrating yields

$$\tau_\delta = \frac{(c + v_z)\, n^2_{\mathrm{H}}\alpha^{\mathrm{A}}_{\mathrm{rec}}\sigma^{\mathrm{tot}}_\alpha}{\nu_\alpha \Gamma \left|\frac{\mathrm{d}v^p_z}{\mathrm{d}r} - H(z)\right|}, \tag{6.5}$$

where v^p_z denotes the peculiar velocity, such that $v_z = v^p_z - H(z)r$, and $H(z)$ is the Hubble parameter at redshift z. Here, we used $\delta(g(r)) = \delta(r - r_0)\,|g'(r_0)|^{-1}$, where g' denotes the derivative of g with respect to r and $g(r_0) = 0$. Specifically, $g(r) = \nu(r) - \nu_\alpha = \nu(1 + v_z(r)/c) - \nu_\alpha$.

Because equation (6.2) only considers the absorption at distances larger than the virial radius, we summarize the RT within the galactic halo by using an effective absorption fraction. Like WD11, we introduce the fraction $F_{\rm abs}$ of Lyα photons subject to absorption as a free parameter, such that the fraction $1 - F_{\rm abs}$ of photons travels freely. This is a simplified description of the line shape, where $1 - F_{\rm abs}$ of the photons are redshifted outside of the high Lyα cross-section region from the RT within the ISM and CGM. The fraction $F_{\rm abs}$ of photons are either on the blue side of the Lyα line center or close enough to be subject to absorption. The total transmittance is given by integrating over the spectral flux density profile of the Lyα line $J(\nu)$, which becomes

$$\mathcal{T} = \frac{\int \mathrm{d}\nu\, J(\nu) e^{-\tau(\nu)}}{\int \mathrm{d}\nu\, J(\nu)} \approx 1 - F_{\rm abs} + F_{\rm abs} e^{-\tau_\delta}. \tag{6.6}$$

6.2.1 Analytic Model for LIM Power Spectra

We modify the calculation of WD11 to derive a model for intergalactic Lyα absorption for the LIM power spectrum. Consider the Lyα intensity field as a biased tracer of the matter density with

$$\delta I(\mathbf{x}) = I(\mathbf{x}) - I_0(\mathbf{x}) = I_0(\mathbf{x}) b_I \delta_{\rm m}, \tag{6.7}$$

where b_I is the linear intensity bias, $\delta_{\rm m} = \rho(\mathbf{x})/\rho_0(\mathbf{x}) - 1$ is the matter density contrast, and $\rho(\mathbf{x})$ is the matter density in real space. The subscript $_0$ denotes the mean field over many realizations - for example, the mean intensity or matter density as a function of redshift. We use brackets $\langle\cdot(\mathbf{x})\rangle$ to denote the same when it is more convenient. For simplicity, we consider a single redshift, and therefore $I_0(\mathbf{x}) = I_0 = \text{const}$. Thus, the intensity power spectrum is

$$P_{II}(\mathbf{k}) = \langle |\widetilde{\delta I}(\mathbf{k})|^2 \rangle = b_I^2 I_0^2 P_{\rm m}(\mathbf{k}), \tag{6.8}$$

where $P_{\rm m}$ is the matter power spectrum and we have neglected the discreteness of the intensity sources and therefore the shot-noise contribution.

The intensity after IGM absorption can be approximated as

$$\begin{aligned} I^{\rm abs}(\mathbf{x}) = {} & I_0^{\rm abs} \left[1 + b_I \delta_{\rm m}(\mathbf{x})\right] \\ & + \left[\Gamma(\mathbf{x}) - \Gamma_0\right] \left.\frac{\partial \mathcal{T}}{\partial \Gamma}\right|_{\mathcal{T}_0,\Gamma_0} \left.\frac{\partial I^{\rm abs}}{\partial \mathcal{T}}\right|_{\mathcal{T}_0} \\ & + \left[\rho(\mathbf{x}) - \rho_0\right] \left.\frac{\partial \mathcal{T}}{\partial \rho}\right|_{\mathcal{T}_0,\rho_0} \left.\frac{\partial I^{\rm abs}}{\partial \mathcal{T}}\right|_{\mathcal{T}_0} \\ & + \left[\frac{\mathrm{d}v_z}{\mathrm{d}(a r_{\rm com})} - H\right] \left.\frac{\partial \mathcal{T}}{\partial \left(\frac{\mathrm{d}v_z}{\mathrm{d}(a r_{\rm com})}\right)}\right|_{\mathcal{T}_0,H} \left.\frac{\partial I^{\rm abs}}{\partial \mathcal{T}}\right|_{\mathcal{T}_0}. \end{aligned} \tag{6.9}$$

We adopt the linear model of WD11 for the transmittance:

$$\begin{aligned} \mathcal{T}(\delta_{\rm m}, \delta_\Gamma, \delta_v) = {} & (1 - F_{\rm abs}) \\ & + F_{\rm abs} \exp\left\{-\tau_0 \frac{1 + c_\gamma \delta_{\rm m}}{1 + \delta_\Gamma + \delta_v}\right\}, \end{aligned} \tag{6.10}$$

where τ_0 is the mean opacity in the IGM, $c_\gamma = (2.7 - 0.7\gamma) \simeq 1.72$ with the polytropic index $\gamma = 1.4$ (Hui & Gnedin, 1997) denotes the dependence of the optical depth on dark matter density, and $\delta_\Gamma = \frac{\Gamma}{\Gamma_0} - 1$ is the ionization rate perturbation. The expression $\delta_v = \frac{\mathrm{d}v_z^p}{\mathrm{d}(ar_{\mathrm{com}})}\frac{1}{H}$ represents the fluctuation in the LOS velocity. The dependence of the intensity on the transmittance is

$$\begin{aligned}\left.\frac{\partial I^{\mathrm{abs}}}{\partial \mathcal{T}}\right|_{\mathcal{T}_0} &= I_0 = I_0^{\mathrm{abs}}\mathcal{T}_0^{-1} \\ &\approx I_0^{\mathrm{abs}}\left(1 - F_{\mathrm{abs}} + F_{\mathrm{abs}}e^{-\tau_0}\right)^{-1}.\end{aligned} \tag{6.11}$$

We can rewrite equation (6.9):

$$\begin{aligned}\delta I^{\mathrm{abs}}(\mathbf{x}) &= I^{\mathrm{abs}}(\mathbf{x}) - I_0^{\mathrm{abs}} \\ &= I_0^{\mathrm{abs}}\left(b_I\delta_{\mathrm{m}} + \delta_{\mathrm{m}}C_\rho + \delta_\Gamma C_\Gamma + \delta_v C_v\right),\end{aligned} \tag{6.12}$$

where the constants C_x are given by

$$\begin{aligned}C_\Gamma &= \left.\frac{\partial \mathcal{T}}{\partial \log\Gamma}\right|_{\mathcal{T}_0,\Gamma_0}\left.\frac{\partial I^{\mathrm{abs}}}{\partial \mathcal{T}}\right|_{\mathcal{T}_0}(I_0^{\mathrm{abs}})^{-1} \\ C_\rho &= \left.\frac{\partial \mathcal{T}}{\partial \log\rho}\right|_{\mathcal{T}_0,\rho_0}\left.\frac{\partial I^{\mathrm{abs}}}{\partial \mathcal{T}}\right|_{\mathcal{T}_0}(I_0^{\mathrm{abs}})^{-1} \\ C_v &= \left.\frac{\partial \mathcal{T}}{\partial \log\left(\mathrm{d}v_z/\mathrm{d}r_{\mathrm{com}}\right)}\right|_{\mathcal{T}_0,H}\left.\frac{\partial I^{\mathrm{abs}}}{\partial \mathcal{T}}\right|_{\mathcal{T}_0}(I_0^{\mathrm{abs}})^{-1}.\end{aligned} \tag{6.13}$$

We find

$$C_\Gamma = C_v = \frac{F_{\mathrm{abs}}\tau_0 e^{-\tau_0}}{1 - F_{\mathrm{abs}} + F_{\mathrm{abs}}e^{-\tau_0}} =: C \tag{6.14}$$

and

$$C_\rho = -\frac{c_\gamma F_{\mathrm{abs}}\tau_0 e^{-\tau_0}}{1 - F_{\mathrm{abs}} + F_{\mathrm{abs}}e^{-\tau_0}} = -c_\gamma C. \tag{6.15}$$

The ionization rate fluctuations can be modeled by convolving the overdensity of ionizing sources with bias b_{ion} with a kernel $K_\lambda(k) = \arctan(k\lambda_{\mathrm{mfp}})/\left(k\lambda_{\mathrm{mfp}}\right)$, where λ_{mfp} is the mean free path of ionizing photons, so that

$$\tilde{\delta}_\Gamma(\mathbf{k}) = b_{\mathrm{ion}}\tilde{\delta}_{\mathrm{m}}(\mathbf{k})K_\lambda(k). \tag{6.16}$$

The fluctuations of intensity introduced by observing in redshift space, denoted by the superscript $^{\mathrm{s}}$, are

$$\delta I^{\mathrm{s}} = \delta I - I_0\frac{\mathrm{d}v_z}{\mathrm{d}r_{\mathrm{com}}}\frac{1}{Ha} = \delta I - I_0\delta_v. \tag{6.17}$$

Relating the velocity gradient fluctuations to the density fluctuations as $\tilde{\delta}_v(\mathbf{k}) = -f\mu^2\tilde{\delta}_m(\mathbf{k})$, where $f = \mathrm{d}\ln D/\mathrm{d}\ln a$ is the logarithmic growth factor, we can write

$$\begin{aligned}\widetilde{\delta I^{\mathrm{s}}_{\mathrm{abs}}}(\mathbf{k}) &= I_0^{\mathrm{abs}}\tilde{\delta}_{\mathrm{m}}(\mathbf{k}) \\ &\quad\times\left[b_I + b_{\mathrm{ion}}K_\lambda(k)C_\Gamma + C_\rho + (1 - C_v)f\mu^2\right] \\ &=: I_0^{\mathrm{abs}}\tilde{\delta}_{\mathrm{m}}(\mathbf{k})D_I(\mathbf{k}) = \mathcal{T}_{\mathrm{g}}I_0\tilde{\delta}_{\mathrm{m}}(\mathbf{k})D_I(\mathbf{k}).\end{aligned} \tag{6.18}$$

Here, we have assumed that the intrinsic luminosity of galaxies is uncorrelated with the local transmittance and $\mathcal{T}_{\rm g} = 1 - F_{\rm abs} + F_{\rm abs}\int {\rm d}\mathbf{x}\, n(\mathbf{x}) e^{-\tau(\mathbf{x})} / \int {\rm d}\mathbf{x}\, n(\mathbf{x})$ is the effective mean transmittance around galaxies. The intensity auto-power spectrum is then given by

$$P_{II}(\mathbf{k}) = \langle |\widetilde{\delta I^{\rm s}_{\rm abs}}(\mathbf{k})|^2 \rangle = I_0^2 \mathcal{T}_{\rm g}^2 P_{\rm m}(k) D_I^2(\mathbf{k}). \tag{6.19}$$

Taking a closer look at the intensity damping factor, we find

$$\begin{aligned} D_I &= b_I + b_{\rm ion} K_\lambda C - c_\gamma C + (1-C) f\mu^2 \\ &= (b_I + b_{\rm ion} C K_\lambda - c_\gamma C) \\ &\quad \times \left(1 + \frac{f}{b_I} \mu^2 \frac{b_I (1-C)}{b_I + b_{\rm ion} C K_\lambda - c_\gamma C} \right). \end{aligned} \tag{6.20}$$

The RSD-like effect of the IGM absorption is introduced because the RSD parameter f/b_I is effectively multiplied by the factor $(1-C)/(1 + CK_\lambda - c_\gamma C/b_I)$, assuming that $b_I = b_{\rm ion}$. This factor is smaller than one, i.e. the RSD is reduced, if $b_I > c_\gamma \simeq 1.72$ on small scales ($K_\lambda \simeq 0$), and $b_I > 0.5 c_\gamma \simeq 0.86$ on large scales ($K_\lambda \simeq 1$).

Following WD11, the LAE overdensity in redshift space is

$$\begin{aligned} \widetilde{\delta}^{\rm s}_{{\rm g}_\alpha} &= \widetilde{\delta}_{\rm m}(\mathbf{k}) \left[b_{{\rm g}_\alpha} + b_{\rm ion} C_\Gamma^{{\rm g}_\alpha} K_\lambda(k) + C_\rho^{{\rm g}_\alpha} + (1 - C_v^{{\rm g}_\alpha}) f\mu^2 \right] \\ &=: \widetilde{\delta}_{\rm m}(\mathbf{k}) D_{{\rm g}_\alpha}(\mathbf{k}), \end{aligned} \tag{6.21}$$

where

$$\begin{aligned} C_\Gamma^{{\rm g}_\alpha} &= C_v^{{\rm g}_\alpha} = (\beta_\phi - 1) \frac{F_{\rm abs} \tau_0 e^{-\tau_0}}{1 - F_{\rm abs} + F_{\rm abs} e^{-\tau_0}} =: C^{{\rm g}_\alpha}, \\ C_\rho^{{\rm g}_\alpha} &= -c_\gamma C^{{\rm g}_\alpha}, \end{aligned} \tag{6.22}$$

and $\beta_\phi > 0$ is -1 times the slope of the Lyα luminosity function, which is also negative. Note that, because $C_\rho^{{\rm g}_\alpha}$ is negative for $\beta_\phi > 1$, the effective LAE bias $b_{{\rm g}_\alpha} + b_{\rm ion} C_\Gamma^{{\rm g}_\alpha} K_\lambda(k) + C_\rho^{{\rm g}_\alpha}$ can become negative at large k (where $K_\lambda(k)$ becomes negligible) if $c_\gamma C > b_{{\rm g}_\alpha}$; for example, for $b_{{\rm g}_\alpha} \simeq 2$, $\beta_\phi \simeq 2.6$, and $\tau_0 \simeq 1$.

The cross-power spectrum of LAEs and Lyα intensity is given by

$$P_{{\rm g}_\alpha \times I_\alpha}(\mathbf{k}) = P_{\rm m}(k) I_0 \mathcal{T}_{\rm g} D_I(\mathbf{k}) D_{{\rm g}_\alpha}(\mathbf{k}). \tag{6.23}$$

The cross-power spectrum becomes negative under the same conditions as $D_{{\rm g}_\alpha}(\mathbf{k})$.

For galaxies detected through a different line than Lyα that are not affected by Lyα RT effects, denoted by the subscript or superscript g, we have $\widetilde{\delta}_{\rm g} = \widetilde{\delta}_{\rm m}\,(b_{\rm g} + f\mu^2)$, so that the cross-power spectrum of these galaxies with the Lyα intensity is

$$P_{{\rm g}\times{\rm I}}(\mathbf{k}) = P_{\rm m}(k) \left(b_{\rm g} + f\mu^2 \right) I_0 \mathcal{T}_{\rm g} D_I(\mathbf{k}). \tag{6.24}$$

Figures 6.1 and 6.2 show the effect of this model of Lyα absorption in the IGM on the power spectra with different bias values. We set the mean optical depth to $\tau_0 = 5$

and the negative slope of the luminosity function to $\beta_\phi = 1.8$ in Figure 6.1, so that the parameters match those of the lognormal simulation in Section 6.4.2. We set the mean free path of ionizing photons to $\lambda_{\rm mfp} = 300\,{\rm Mpc}$ (Bolton & Haehnelt, 2007). The first-order effect of both settings is that the amplitude decreases when including IGM absorption because of the smaller effective bias. This amplitude difference does not include the lower mean intensity, which will further decrease the amplitude of the LIM power spectra. The reason for the smaller effective bias is that the transmittance modeled in equation (6.10) is smaller at higher densities. While a larger ionization rate implies a larger transmittance and the ionization rate is higher in high-density regions, its influence is reduced by the smoothing kernel. The velocity gradient fluctuations are negative in overdensities, which also increases the transmittance.

The anisotropy of the suppression depends strongly on the input parameters. In the configuration of Figure 6.1 with the bias $b = 1.5$, large scales are more strongly suppressed perpendicular to the LOS. A higher bias of $b = 2$ inverts the RSD, leading to a stronger suppression along the LOS; see Figure 6.2.

The suppression of the monopole power spectrum shown in the right panels of Figures 6.1 and 6.2 shows that the effective bias is smaller at small scales than at large scales. This scale dependence is introduced by the smoothing kernel of the ionization rate parameterized by the mean free path of ionizing photons. A larger mean free path leads to a decrease at smaller $k \simeq \lambda_{\rm mfp}^{-1}$.

6.2.2 Shot Noise

We have ignored the shot-noise power spectrum in the previous section. Without absorption, and assuming constant redshift, the intensity auto-shot noise follows (e.g., Bernal & Kovetz, 2022)

$$P_{\rm shot} = \left(\frac{c}{4\pi(1+z)^2\lambda_\alpha H(z)}\right)^2 \int_{L_{\rm min}}^{L_{\rm max}} {\rm d}L\, \frac{{\rm d}n}{{\rm d}L} L^2, \tag{6.25}$$

where λ_α is the rest-frame Lyα wavelength, $\frac{{\rm d}n}{{\rm d}L}$ is the luminosity function, and $L_{\rm min}$ and $L_{\rm max}$ are the minimum and maximum luminosities of the galaxies contributing to the intensity map. The first factor assumes that we measure the specific intensity I_λ. Assuming that the intrinsic luminosity of a galaxy is independent of the matter density and therefore uncorrelated with the local transmittance, equation (6.25) turns into

$$P_{\rm shot}^{\rm abs} = \left(\frac{c}{4\pi(1+z)^2\lambda_\alpha H(z)}\right)^2 \int_{L'_{\rm min}}^{L'_{\rm max}} {\rm d}L\, \frac{{\rm d}n}{{\rm d}L} L^2 \mathcal{T}_{\rm g}^2. \tag{6.26}$$

If the galaxy sample changes - for example, because only undetected galaxies contribute to the intensity map through masking - the integration limits have to be changed: $L'_{\rm min/max} = \mathcal{T}_g^{-1} L_{\rm min/max}$.

6.3 Lyα Line Shift and Broadening

For Lyα photons to escape the ISM, they have to diffuse spatially and spectrally. In the absence of inflows or outflows, this gives rise to a symmetric, double-peaked spectrum,

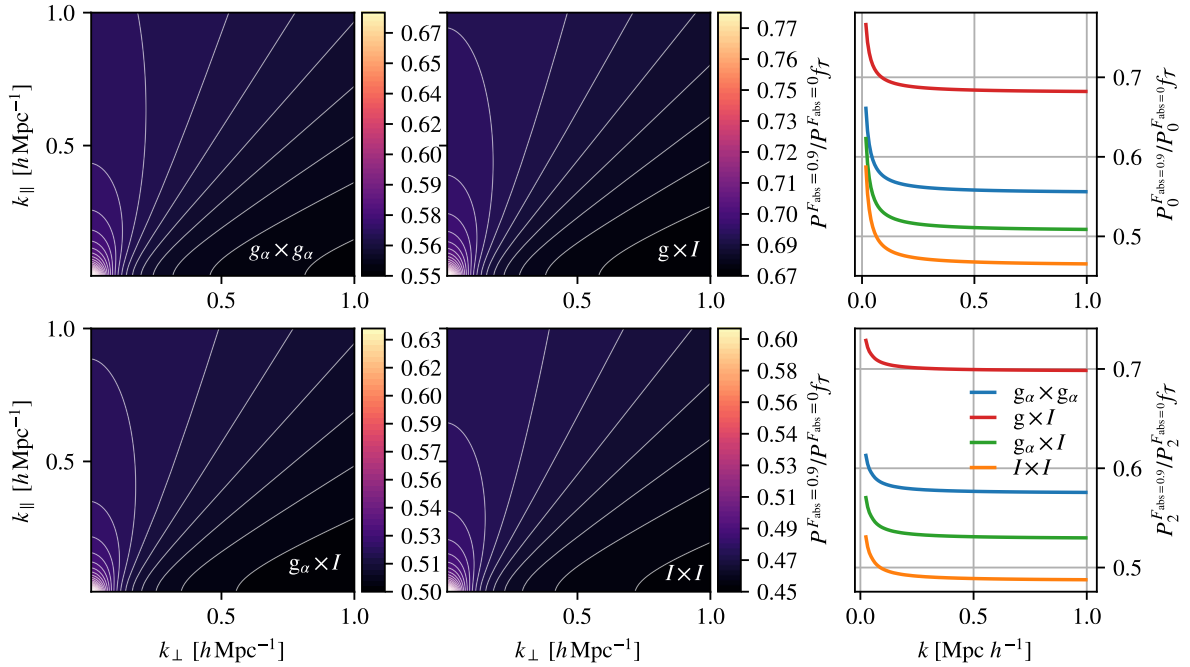

Figure 6.1: Effect of IGM absorption on the Lyα intensity power spectrum in the analytic model with $\tau_0 = 5$, $b_I = b_{g_\alpha} = b_g = 1.5$, $F_{\rm abs} = 0.9$, $\lambda_{\rm mfp} = 300\,{\rm Mpc}$, and $\beta_\phi = 1.8$. We show $P^{F_{\rm abs}=0.9}/P^{F_{\rm abs}=0.0}f_{\mathcal{T}}$, where $P^{F_{\rm abs}=0.9}$ is the power spectrum with $F_{\rm abs} = 0.9$ and $P^{F_{\rm abs}=0.0}$ is that without absorption. The factor $f_{\mathcal{T}}$ accounts for the amplitude change due to the lower mean intensity: $f_{\mathcal{T}} = I_0/I_0^{\rm abs}$ for the cross-power spectra ($g \times I$ and $g_\alpha \times I$), $f_{\mathcal{T}} = (I_0/I_0^{\rm abs})^2$ for the intensity auto-power spectrum ($I \times I$), and $f_{\mathcal{T}} = 1$ for the LAE auto-power spectrum ($g_\alpha \times g_\alpha$). The four left and middle panels show the power spectrum damping as a function of wavenumber perpendicular and parallel to the LOS. The top left panel shows the damping of the LAE auto-power spectrum, the top middle panel shows that of the cross-power spectrum of Lyα intensity with non-LAE galaxies. The bottom left panel shows the LAE-Lyα intensity cross-power spectrum damping. The bottom middle panel shows the Lyα intensity auto-power spectrum. The top right panel shows the damping of the monopole of the LAE auto-power spectrum (blue), the non-LAE-intensity cross-power spectrum (red), the LAE-intensity cross-power spectrum (green), and the intensity auto-power spectrum (orange). The bottom right panel shows the same for the quadrupoles. Note that the anisotropy of the suppression depends strongly on the input parameters, especially the bias.

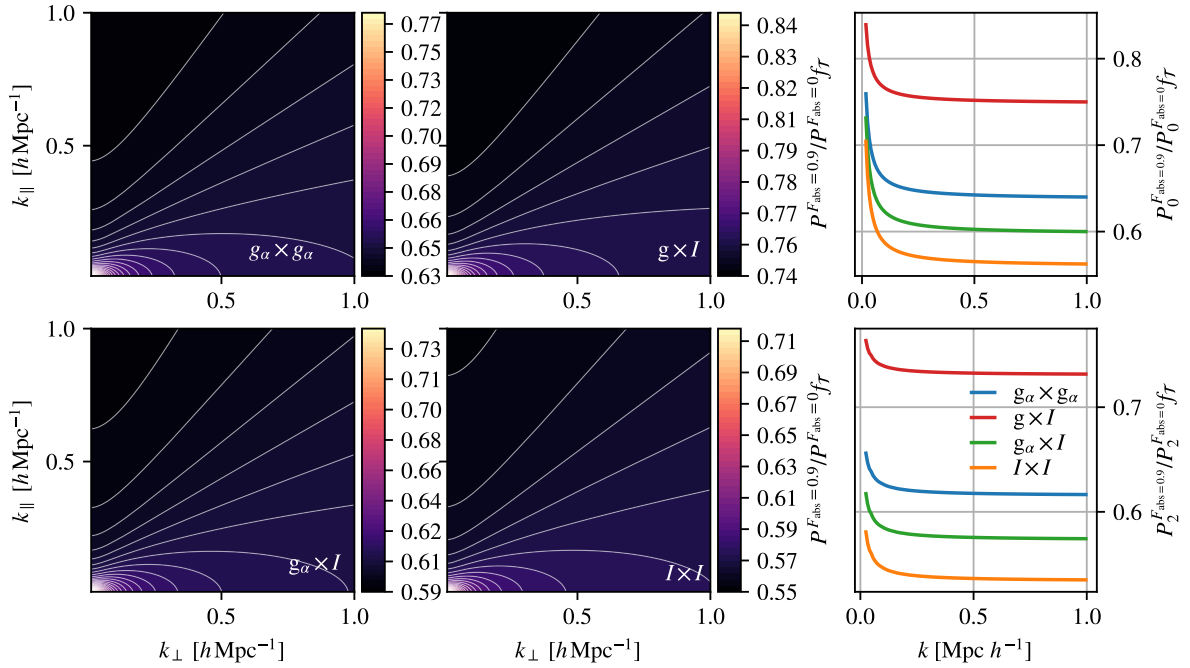

Figure 6.2: Same as figure 6.1, but with larger bias $b_I = b_{g_\alpha} = b_g = 2$.

while simple shell models show that inflows enhance and outflows suppress the blue peak (e.g., Verhamme et al., 2006a). Because a sufficiently redshifted peak is redshifted out of the Lyα cross section, only the blue part of the spectrum is subject to intergalactic absorption. At redshifts $z \gtrsim 2$, LAEs predominantly have red peaks (see, e.g., Ouchi et al., 2020).

The RT in the ISM also broadens the Lyα line. We model the effect on the intensity auto- and cross-power spectra in the same way as spectral smoothing of the intensity map (see, e.g., Lujan Niemeyer et al., 2023), assuming that the broadening is independent of galaxy properties. However, Chung et al. (2021) show that the dependence of the line width on halo mass or luminosity produces a different damping of the power spectrum than a mass-independent broadening. Li et al. (2024) show that the LIM power spectrum is mostly sensitive to the line width, but not the exact line shape. For modeling of the voxel intensity distribution including spectral broadening, see Bernal (2024).

When the redshift-space position of LAEs is determined from the Lyα line, it is affected by the line shift caused by RT as well as by the peculiar velocity of the galaxies. Following Byrohl et al. (2019), we consider the redshift-space galaxy density field that is exact under the assumption of one fixed global LOS direction $\hat{\mathbf{e}}_\parallel$ (Taruya et al., 2010),

$$\widetilde{\delta}^s_{\rm g}(\mathbf{k}) = \int {\rm d}^3\mathbf{x} \left[\delta_{\rm g}(\mathbf{x}) - \partial_\parallel u_\parallel(\mathbf{x})\right] e^{i\mathbf{k}\cdot\mathbf{x} + ik_\parallel u_\parallel(\mathbf{x})}. \tag{6.27}$$

Here, we have introduced a scaled velocity $\mathbf{u} = \mathbf{v}/(aH)$ and $\partial_\parallel$ denotes the derivative with respect to the LOS distance. The same equation can be written for $\widetilde{\delta I^s}(\mathbf{k})$. If we cross-correlate this galaxy overdensity with another field $\delta_{\rm g'}$ that is not affected by $u_\parallel(\mathbf{x})$,

and neglecting cross-shot noise, we can write the cross-power spectrum as

$$P^s_{\mathrm{gg}'}(\mathbf{k}) = \int \mathrm{d}^3\mathbf{r}\, e^{i\mathbf{k}\cdot\mathbf{r}} \langle e^{ik_\parallel u_\parallel(\mathbf{x})} \left[\delta_\mathrm{g}(\mathbf{x}) - \partial_\parallel u_\parallel(\mathbf{x})\right] \delta_{\mathrm{g}'}(\mathbf{x}')\rangle, \tag{6.28}$$

where we have set the condition that the expression within the angle brackets depends only on $\mathbf{r} = \mathbf{x}' - \mathbf{x}$. We can rewrite this in terms of the cumulants as (Scoccimarro, 2004; Taruya et al., 2010; Byrohl et al., 2019)

$$\begin{aligned} P^s_{\mathrm{gg}'}(\mathbf{k}) = \int \mathrm{d}^3\mathbf{r}\, e^{i\mathbf{k}\cdot\mathbf{r}} \exp\left\{\langle e^{ik_\parallel u_\parallel(\mathbf{x})}\rangle_c\right\} \\ \times [\langle e^{ik_\parallel u_\parallel(\mathbf{x})} A(\mathbf{x}) \delta_{\mathrm{g}'}(\mathbf{x}')\rangle_c \\ + \langle e^{ik_\parallel u_\parallel(\mathbf{x})} A(\mathbf{x})\rangle_c \langle e^{ik_\parallel u_\parallel(\mathbf{x})} \delta_{\mathrm{g}'}(\mathbf{x}')\rangle_c], \end{aligned} \tag{6.29}$$

where $A(\mathbf{x}) = \delta_\mathrm{g}(\mathbf{x}) - \partial_\parallel u_\parallel(\mathbf{x})$. The factor $\exp\left\{\langle e^{ik_\parallel u_\parallel(\mathbf{x})}\rangle_c\right\} = \langle e^{ik_\parallel u_\parallel(\mathbf{x})}\rangle$ can be taken out of the integration because it does not depend on $\mathbf{r}$. It constitutes a Fingers-of-God-like damping of the form

$$\begin{aligned} D^\mathrm{cross}_{v_\mathrm{RT}} &= \exp\left\{\langle e^{ik_\parallel u_\parallel}\rangle_c\right\} = \langle e^{ik\mu u_\parallel}\rangle \\ &= \int \mathrm{d}u_\parallel \phi(u_\parallel) e^{ik\mu u_\parallel}, \end{aligned} \tag{6.30}$$

where $\phi(u_\parallel)$ is the probability density function (PDF) of the LOS velocity shift $u_\parallel$. This factor is a one-dimensional Fourier transform of $\phi(u_\parallel)$ to the variable $k\mu$.

As an example, consider a Gaussian PDF with mean $\bar{u}_\parallel$ and standard deviation $\sigma_{u_\parallel}$. The cross-power spectrum damping factor is then

$$D^\mathrm{cross}_{v_\mathrm{RT}} = e^{ik\mu\bar{u}_\parallel - \frac{1}{2}k^2\mu^2\sigma^2_{u_\parallel}}, \tag{6.31}$$

which contains a phase shift due to $\bar{u}_\parallel$. The real component of the cross-power spectrum is therefore multiplied by

$$\mathrm{Re}\left\{\left(D^\mathrm{cross}_{v_\mathrm{RT}}\right)\right\} = \cos(k\mu\bar{u}_\parallel) e^{-\frac{1}{2}k^2\mu^2\sigma^2_{u_\parallel}}. \tag{6.32}$$

The imaginary part of the power spectrum is multiplied by the respective sine function. The cosine has a zero point at $k\mu = \pi/(2\bar{u}_\parallel) \approx 0.9\,h\mathrm{Mpc}^{-1}$ for $\bar{u}_\parallel = 200\,\mathrm{kms}^{-1}$ at $z = 3$. Note that an auto-power spectrum of δ^s_g will have a Fingers-of-God-like damping of the form

$$D^\mathrm{auto}_{v_\mathrm{RT}}(k,\mu) = \left|\int \mathrm{d}u_\parallel \phi(u_\parallel) e^{ik\mu u_\parallel}\right|^2, \tag{6.33}$$

which is unaffected by $\bar{u}_\parallel$ (see Byrohl et al., 2019).

The phase shift can also occur in a cross-power spectrum of two fields with different velocity distributions, such as the cross-correlation between the detected, bright LAEs with the intensity of undetected, faint LAEs as planned by the HETDEX collaboration (Lujan Niemeyer et al., 2023).

6.4 Lognormal Simulation

6.4.1 Modeling

The analytic model is limited to the linear approximation of the optical depth in equation (6.10), which is only expected to hold for small fluctuations in the matter density, ionization rate, and velocity. However, the Lyα absorption mostly happens in the immediate environment of the galaxies, where these fluctuations are large.

To introduce a model that is both fast and more accurate, we modify the SIMPLE code (Lujan Niemeyer et al., 2023) to include the effect of Lyα RT. SIMPLE is a tool for quickly generating mock intensity maps. It uses lognormal galaxy simulations (Agrawal et al., 2017) and randomly assigns a luminosity to each galaxy by sampling from the input luminosity function. One can smooth the map, add noise, and apply sky subtraction to make the mocks more realistic for observations. One can also apply a selection function to obtain a catalog of detected galaxies.

The lognormal simulations of Agrawal et al. (2017) calculate the velocity field from the linearized continuity equation. Together with the matter density field and a model for ionization and the IGM transmittance, we can build a model for IGM absorption.

We model the ionization rate as proportional to the galaxy number density field of all (detected and undetected) galaxies, smoothed with the kernel K_λ in equation (6.16). The amplitude of the ionization rate is chosen so that the mean ionization rate matches that of Khaire & Srianand (2019) in each redshift bin. The mean free path of ionizing photons λ_{mfp} is left as a free parameter.

We calculate the hydrogen number density field n_{H} to be proportional to the matter density field:

$$n_{\mathrm{H}}(\mathbf{x}) = \frac{\rho_{\mathrm{H}}(\mathbf{x})}{m_{\mathrm{P}}} = \frac{0.75\Omega_{\mathrm{b}}(z)\rho_c(z)}{m_{\mathrm{P}}}\left(1+\delta_{\mathrm{m}}(\mathbf{x})\right), \tag{6.34}$$

where ρ_{H} is the hydrogen mass density, m_{P} is the proton mass, Ω_{b} is the baryon density parameter, ρ_c is the critical density, and δ_{m} is the matter overdensity output from the lognormal simulation.

The velocity field is calculated by the lognormal simulation of Agrawal et al. (2017). We use *numpy.gradient*[3] to calculate the velocity gradient.

Finally, we calculate the local optical depth $\tau_\delta(\mathbf{x})$ and transmittance $\mathcal{T}(\mathbf{x})$ in each cell with Equations (6.5) and (6.6). The luminosity $L(\mathbf{x})$ of each galaxy at the position $\mathbf{x}$ is replaced with $\mathcal{T}(\mathbf{x})L(\mathbf{x}) = \left[1 - F_{\mathrm{abs}} + F_{\mathrm{abs}}e^{-\tau_\delta(\mathbf{x})}\right]L(\mathbf{x})$ before generating the intensity map. F_{abs} is equal for all galaxies. We use this transmitted intensity map to calculate the power spectra and the transmitted flux for the selection function for the detected galaxy catalog.

Using a cosmological RT simulation, Byrohl et al. (2019) find that the Lyα velocity shift from RT is independent of the peculiar velocity of the host halo. We therefore model this effect by adding a random velocity shift to the mock galaxies following an input PDF $\phi(u_\parallel)$.

[3] `https://numpy.org/doc/stable/reference/generated/numpy.gradient.html`

Because line broadening can be modeled in the same way as a limited spectral resolution, one can increase the LOS smoothing in the input to SIMPLE.

6.4.2 RT Effects in Lognormal Simulations

We set up a cubic box with length $L_{\rm box} = 512\,{\rm Mpc}\,h^{-1}$ and $N_{\rm mesh} = 256$ at mean redshift $\bar{z} = 2.2$ with galaxy bias $b = 1.5$ and the EWgt60 Lyα luminosity function of Konno et al. (2016), which includes photometrically selected LAEs with Lyα equivalent width larger than 60 Å. We adopt a constant flux limit $F_{\rm min} = 3 \times 10^{-18}\,{\rm erg\,s^{-1}\,cm^{-2}}$ for detection, no noise, and no smoothing of the intensity map. To remove the shot noise, we calculate the power spectrum using the half-sum-half-difference approach (HSHD; see Appendix and Ando et al., 2018; Wang et al., 2024). We study the IGM absorption and the line shift effects separately. Realistically, line shift and broadening affect the amount of Lyα photons subject to absorption, which we include in $F_{\rm abs}$.

To exaggerate the IGM absorption effect, we adopt a large absorption fraction $F_{\rm abs} = 0.9$ and use all galaxies to generate the intensity map. We set the mean free path of ionizing photons to $\lambda_{\rm mfp} = 300\,{\rm Mpc}$.

Figure 6.3 shows the distribution of the neutral hydrogen fraction, the optical depth, and the effective transmittance values (accounting for $F_{\rm abs}$) in the whole box compared to that in voxels containing galaxies and their mean values. The transmittance at galaxy positions is smaller than the overall mean transmittance in the simulation volume because galaxies lie in matter overdensities and therefore neutral hydrogen overdensities by construction. The mean galaxy-weighted transmittance is low, $\bar{\mathcal{T}}_{\rm g} \simeq 0.5$. The optical depth distribution has a long tail toward high optical depths. As a result, the mean optical depth is higher and inconsistent with the measurement of Turner et al. (2024), which is on the order of $\simeq 0.1$. The median optical depth in the lognormal simulations is lower at $\simeq 0.1$.

We calculate the intensity and LAE auto-power spectra, the LAE-Lyα intensity cross-power spectrum, and the cross-power spectrum of Lyα intensity with non-Lyα galaxies that have an uncorrelated luminosity function and are unaffected by IGM absorption. We subtract the shot noise using the HSHD method, and take the average power spectrum of 1000 mocks. Figure 6.4 shows the power spectrum ratios as a function of $k_\perp$ and $k_\parallel$ with and without IGM absorption. The main effect of the absorption is a suppression that is stronger at small scales. This is predicted by the analytic model, where the suppression is stronger at small scales where K_λ is small. However, the shape of the suppression differs from the analytic model in the setup with the same bias, luminosity function slope, and mean optical depth (see Figure 6.1). The suppression from IGM absorption in the lognormal simulations looks similar whether the bias is $b = 2$ or $b = 1.5$.

To explore the reason behind this discrepancy, we calculate the transmittance according to equation (6.10) using $\delta_{\rm m}$, δ_v, δ_Γ, and $\tau_0 = 5$ from the lognormal simulation. Figure 6.5 compares this transmittance to that directly calculated from the mocks. It shows that the linear approximation for the optical depth in equation (6.10) does not describe the results of the lognormal simulations well. The absorption is dominated by the immediate surroundings of the galaxies, where the δ values are too large for linear approximations to hold. However, lowering the value of τ_0 in the WD11 model by a factor of 10 leads to

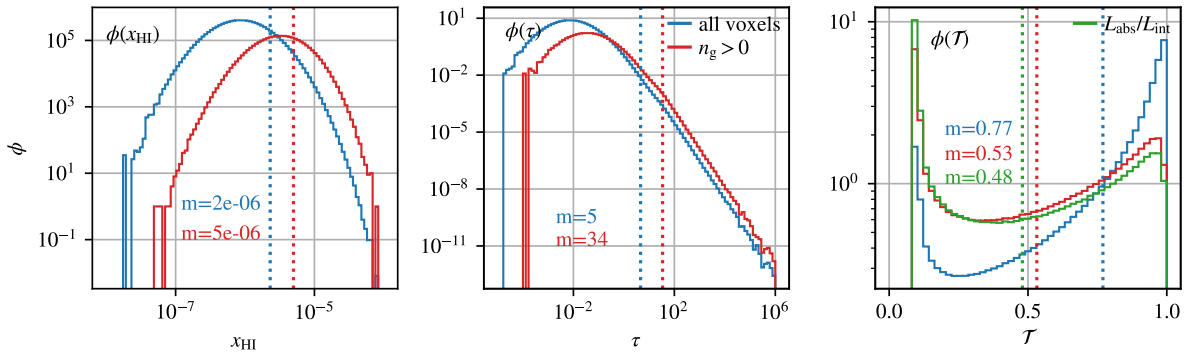

Figure 6.3: Distributions of the neutral hydrogen fraction $x_{\rm HI}$ (left panel), the optical depth (middle panel), and the effective transmittance with $F_{\rm abs} = 0.9$ (right panel) of the entire mock box (blue) of the test mock compared to those in voxels containing at least one galaxy (red). The green line in the right panel shows the effective transmittance of the galaxies calculated from ratio of the attenuated and original luminosity. The dotted lines and the text show the corresponding mean values.

a better agreement with the transmittance values of the lognormal simulation.

To model the Lyα line shift, we set $\phi(u_\parallel)$ to a Gaussian PDF with mean $\bar{v}_{\rm RT} = 639\,{\rm km\,s^{-1}}$ and standard deviation $\sigma_{v_{\rm RT}} = 169\,{\rm km\,s^{-1}}$. This line shift PDF is a best-fit Gaussian for the line shift distribution at $z = 3.01$ with a galaxy number density $\bar{n}_{\rm g} = 10^{-3}\,{\rm Mpc}^{-3}h^3$ considering only the red peak in the RT simulation of Byrohl et al. (2019). In order to see the phase shift of the cross-power spectrum in this test, we keep the redshift-space positions of the galaxies in the galaxy catalog unchanged, while we add the line shift to the galaxies to calculate the intensity map. We use all galaxies to generate the intensity map in order to see the effect of the line shift on the cross-shot noise. We calculate the shot-noise-subtracted 2D power spectrum from an average of 100 mocks using the HSHD method. We calculate the ratios between the power spectrum with and without the RT line shift and compute the mean damping along the LOS by averaging over $k_\perp$. We confirm that the cross- and auto-power spectra follow the expected damping in Equations (6.32) and (6.33).

6.5 Sensitivity of a HETDEX-Like Experiment

We use the same HETDEX-like mocks as in Lujan Niemeyer et al. (2023) and include IGM absorption, a Lyα line shift, and Lyα line broadening to investigate the sensitivity of the power spectrum measured by a HETDEX-like survey to Lyα RT effects. We set the mean free path of ionizing photons to $\lambda_{\rm mfp} = 300\,{\rm Mpc}$. We set $\phi(u_\parallel)$ to a Gaussian PDF with mean $\bar{v}_{\rm RT} = 639\,{\rm km\,s^{-1}}$ and standard deviation $\sigma_{v_{\rm RT}} = 169\,{\rm km\,s^{-1}}$. Fig. 13 of Mentuch Cooper et al. (2023a) shows the observed line width distribution of the LAEs in HETDEX with a mean of $\sigma_\lambda = 3.54\,$Å, which includes the intrinsic Lyα line width of the LAEs and the smoothing of the spectrograph VIRUS ($\sigma_\lambda \approx 2.38\,$Å; see Hill et al., 2021). To model the line broadening through RT and the VIRUS resolution, we apply Gaussian smoothing of the intensity map along the LOS with $\sigma_\lambda = 3.54\,$Å in the case with Lyα

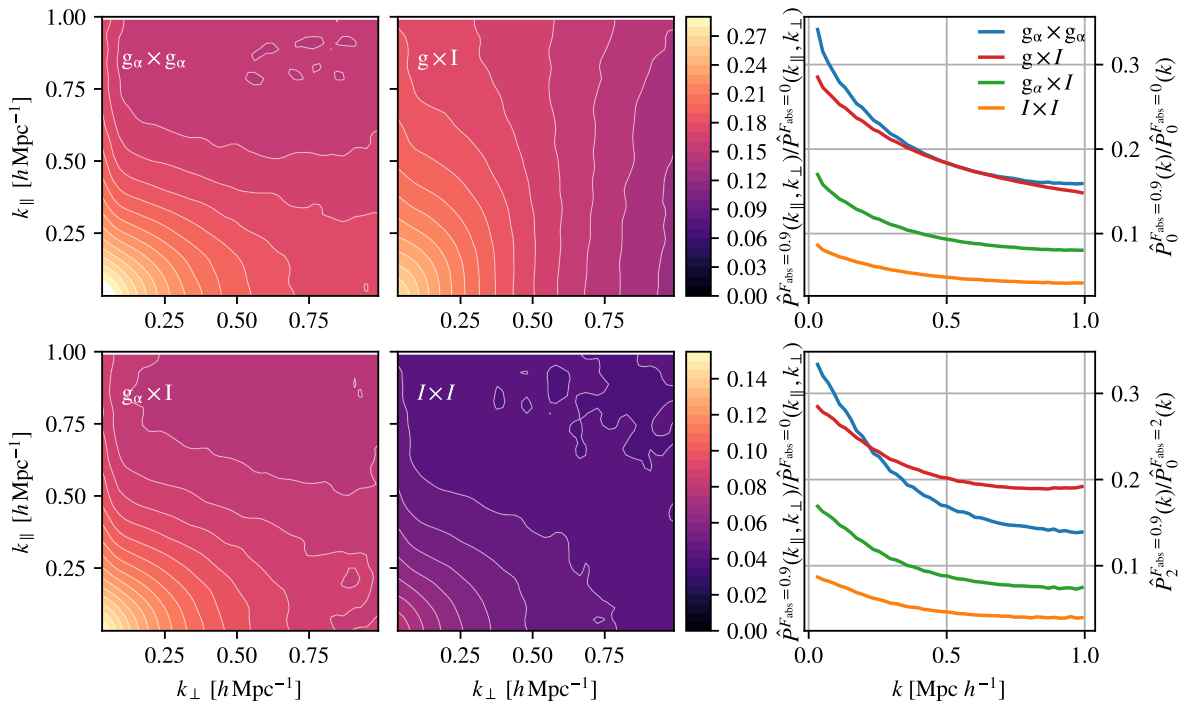

Figure 6.4: Similar to Figure 6.1, but for the lognormal mocks after shot-noise subtraction. These power spectrum ratios include the decreased mean intensity. The ratios of the power spectrum with absorption ($F_{\rm abs} = 0.9$) over that without absorption ($F_{\rm abs} = 0$) of the LAE auto-power spectrum (top left panel), the LAE-Lyα intensity cross-power spectrum (bottom left panel), the non-LAE-Lyα intensity cross-power spectrum (top middle panel), and the intensity auto-power spectrum (bottom middle panel) are shown. The right panels show the damping of the monopole and quadrupole of the LAE auto-power spectrum (blue), the non-LAE-Lyα intensity cross-power spectrum (red), the LAE-Lyα intensity cross-power spectrum (green), and the intensity auto-power spectrum (orange). The 2D damping maps were smoothed with a Gaussian kernel with a width of $\sigma \simeq 0.024\,h\mathrm{Mpc}^{-1}$ for better visualization.

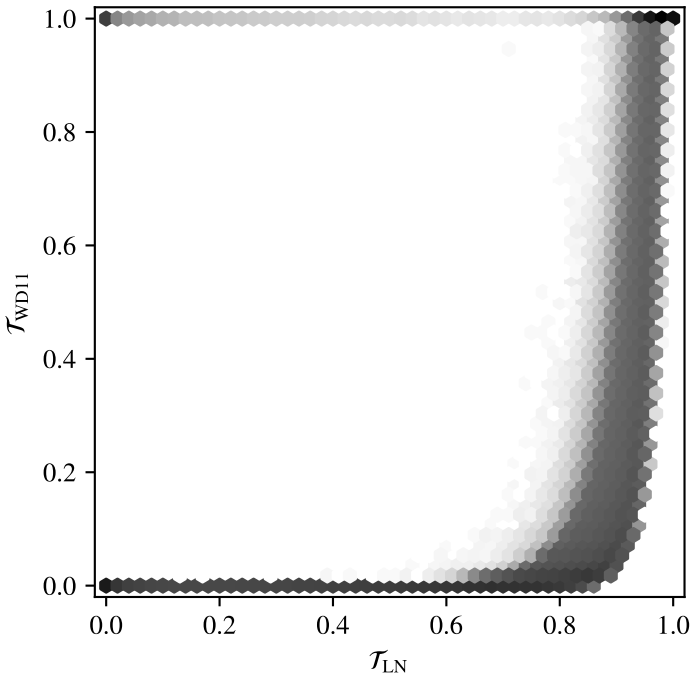

Figure 6.5: Transmittance calculated directly from the lognormal simulation $\mathcal{T}_{\rm LN}$ compared to that calculated with equation (6.10) using $\delta_{\rm m}$, δ_v, δ_Γ, and $\tau_0 = 5$ from the lognormal simulation ($\mathcal{T}_{\rm WD11}$), setting $F_{\rm abs} = 1$ for both transmittance values. Darker regions contain more points, shown with a logarithmic color scale.

RT effects, and $\sigma_\lambda = 2.38\,\text{Å}$ in the fiducial case without RT. We subtract the shot noise using the HSHD method.

Figure 6.6 shows the impact of the RT effects on the HETDEX power spectra compared to the fiducial case in dashed lines at $\bar{z} = 2.2$[4]. We obtain a similar result for $\bar{z} = 3.0$. The fiducial galaxy auto-power spectrum quadrupole looks slightly different than in Lujan Niemeyer et al. (2023) because the HSHD method removes some previously unaccounted-for shot noise in the quadrupole. The LAE-Lyα intensity cross-power spectrum is smaller and has a lower signal-to-noise ratio because of a missing factor of $(1+z)^{-2}$ in the equation (4) for the specific intensity I_λ in Lujan Niemeyer et al. (2023). This is corrected in this paper and the public SIMPLE code, such that the correct specific intensity is lower, while the noise level remains the same.

The amplitude, i.e., the effective bias, is lower for higher $F_{\rm abs}$. The RT line shift dispersion suppresses the power spectrum at small scales; this is also noticeable in the different shapes of the quadrupole. The effects are significant even at $F_{\rm abs} = 0.2$ for the LAE auto-power spectrum, while the LAE-Lyα intensity cross-power spectrum stays within the measurement uncertainty. Note that the covariance of the power spectra with $F_{\rm abs} > 0$ is overestimated because we do not change the input luminosity function, which is measured from observed fluxes, such that the number of observed galaxies is lower than for $F_{\rm abs} = 0$.

These results demonstrate that the HETDEX LAE power spectrum is sensitive to the presence of Lyα RT effects. However, because the main effect of the Lyα absorption is

[4]Note that, while the intensity unit contains $\text{arcsec}^{-2}\,\text{Å}^{-1}$, the intensity is not aperture-dependent. It is calculated using $I_\lambda(\mathbf{x}) = c\rho_L(\mathbf{x})/\left(4\pi(1+z)^2\lambda_0 H(z)\right)$, where $\rho_L(\mathbf{x})$ is the luminosity per unit comoving volume at location $\mathbf{x}$ and λ_0 is the rest-frame Lyα wavelength.

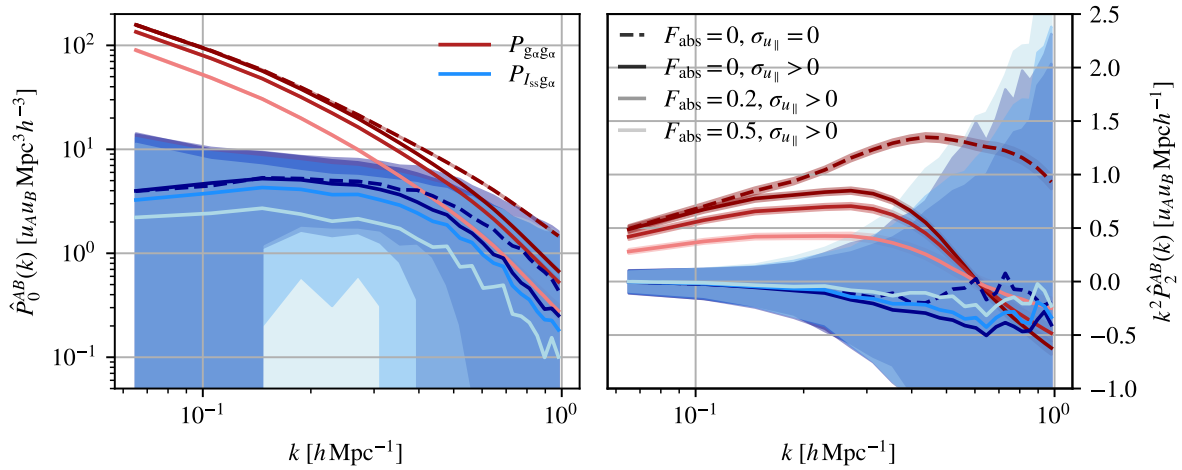

Figure 6.6: Power spectrum monopoles (left panel) and quadrupoles (right panel) with RT effects (solid lines) compared to the fiducial case without RT effects (dashed lines) in the $\bar{z} = 2.2$ bin. The red lines show the LAE auto-power spectra, and the blue lines show the cross-power spectra of LAEs with sky-subtracted intensity. The absorption fraction is shown with differently shaded lines, where lighter lines correspond to higher F_{abs}. The shaded areas show the standard deviation of the mocks. The units are $u_{g_\alpha} = 1$ and $u_I = 10^{-22}\,\mathrm{erg\,s^{-1}\,cm^{-2}\,arcsec^{-2}\,\AA^{-1}}$.

degenerate with the LAE bias, it can be difficult to isolate - for example, through its scale dependence. Greig et al. (2013) show that the bispectrum can help break degeneracies between gravitational and RT effects. Using the power spectrum, HETDEX can nonetheless constrain the Lyα line shift distribution (see Section 6.3). Because LAEs are mostly central halo galaxies and therefore unaffected by virial motion (Ouchi et al., 2020), a Fingers-of-God-like damping with a velocity dispersion of order $\simeq 100\,\mathrm{km\,s^{-1}}$ would likely stem from RT line shifts.

6.6 Discussion

Lognormal simulations directly produce the galaxy and matter distributions and linear velocities, which we use in this work to calculate the optical depth. This approach requires an assumption of the ionization rate smoothing kernel and the Lyα absorption fraction, but produces the mean optical depth as output. In this regard, there are fewer free parameters than in the analytic model. While the mean optical depth is dominated by a long tail toward high optical depths and inconsistent with the measurement of Turner et al. (2024), the median optical depth is lower and consistent with the measurement.

Because the lognormal simulations do not include galaxy-scale or CGM-scale physics, the optical depth is calculated from large-scale matter distributions and velocities, so a correlation between the optical depth and the galaxy distribution - and therefore an IGM absorption effect on the power spectrum - is inevitable. We showed that this approach produces high optical depths in voxels containing galaxies. As in the WD11 model, the parameter F_{abs} defines how much of the Lyα RT takes place on the scale of the resolution

of the simulations ($\simeq$Mpc), where $F_{\rm abs} = 0$ represents the case where no RT takes place on these scales. This means that we account for the shape of the Lyα line emerging from the CGM only effectively with $F_{\rm abs}$ (see equation (6.6)).

We approximate the RT outside of the virial radius as attenuation proportional to $\exp(-\tau_\delta)$. Zheng et al. (2011) compare the clustering of LAEs in their full RT simulation to $\exp(-\tau(\nu))$ attenuation from the centers of the galaxies. They find that, while the qualitative effects of RT on LAE clustering are captured by the $\exp(-\tau(\nu))$ model, they are quantitatively different, presumably because multiple scatterings are unaccounted for in the $\exp(-\tau(\nu))$ model. One important difference between our model and that used by Zheng et al. (2011) is that we calculate τ_δ only outside of the virial radius, similarly to Laursen et al. (2011), and we approximate the RT within the galactic halo through the parameter $F_{\rm abs}$. As shown in Figure 6.3, most of the τ_δ values in the lognormal simulations are smaller than 1, where we expect multiple scatterings to be rare.

We find that the linear analytic model for absorption does not describe the lognormal simulations well. In the lognormal simulations, the absorption takes place in the immediate environment of the galaxies, where the matter overdensity is large. In this regime, the linear approximations for the transmittance and the effect on power spectrum break down. Therefore, the linear model is an inadequate description of the effect of Lyα absorption on the LIM power spectra.

A minor shortcoming of the lognormal approach is that we used the same luminosity function as an input, such that fewer galaxies are detected and the mean observed intensity is lower after IGM absorption. To mitigate this, one could change the input luminosity function to match it to the observed one after absorption. One could also implement a distribution of absorption fractions, which we assume would not change the power spectra, but it would slightly increase the covariance.

To summarize, the correlations between Lyα transmission and the large-scale matter and velocity, and therefore the effects of IGM absorption on the LAE and LIM power spectra, are model-dependent. Observations of the LAE and Lyα LIM power spectra, such as from HETDEX, are clearly necessary for constraints. As shown in this paper, HETDEX could be strongly affected by Lyα RT, and this must be accounted for in the power spectrum modeling.

In this work, we have only accounted for Lyα absorption in the IGM and Lyα line shifts and broadening from RT. We have not attempted to study the impact of the Lyα absorption from background continuum sources on Lyα LIM power spectra. The Lyα forest in quasar spectra can easily be masked by masking quasar spectra for LIM. However, Weiss et al. (2024) find broad absorption troughs around LAEs in HETDEX through stacking, which will affect Lyα LIM studies.

This work also does not account for extended and diffuse Lyα emission, i.e. photons that scatter into the LOS or those that are produced in the CGM and IGM. Extended Lyα halos are ubiquitous around galaxies at $z \gtrsim 2$ (e.g. Leclercq et al., 2017; Lujan Niemeyer et al., 2022a,b). In this work, we assume that the flux originating from Lyα halos is included in the luminosity function and approximate the galaxies including their halos as point sources. Using high-resolution observations, Leclercq et al. (2017) find that $\simeq 65\%$ of the LAE flux comes from Lyα halos. However, Konno et al. (2016) use $2'' - 3''$

apertures to measure the Lyα flux of the LAEs for the luminosity function. The majority of the Lyα halo flux is contained within these distances (Leclercq et al., 2017). Lujan Niemeyer et al. (2022a) find that Lyα halos of LAEs can reach $\simeq 160$ kpc (proper), but with negligible intensity compared to the central emission. The scales of interest in this work are also much larger than the extent of Lyα halos. Neither lognormal simulations nor the $\exp(-\tau_\delta)$ attenuation model are adequate for studying Lyα LIM at kpc scales.

Furthermore, Byrohl et al. (2021) find that Lyα halo photons in the outer halo are scattered photons originating from galaxies outside of the host galaxy's dark matter halo, introducing a nonlocal component to Lyα halos. This environmental dependence of the intensity of the Lyα halo cannot be captured in this lognormal model.

6.7 Summary and Conclusion

We have presented an analytic model for the effect of Lyα absorption in the IGM on the Lyα LIM power spectra by adapting that of WD11. While the overall effect is similar to that of the LAE auto-power spectrum - a lower, scale-dependent effective bias and reduced RSD, the suppression of the LIM component does not depend on the slope of the luminosity function.

We have extended the model of Byrohl et al. (2019) of the effect of line shifts from Lyα RT on the galaxy power spectrum to LIM power spectra. The effect on the intensity auto-power spectrum is the same as for LAEs. In cross-correlations of one tracer affected by this line shift with another that is unaffected, a phase shift of the power spectrum is introduced, leading to a cosine-shaped damping of the real part of the power spectrum. This can be useful to measure the average line shift and its dispersion of different galaxy populations with or without LIM.

We have modified the SIMPLE code, a lognormal galaxy and intensity map simulator, to calculate the optical depth in each voxel from the matter and velocity distribution and attenuation of the intrinsic luminosities of galaxies in that voxel. In this model, a correlation between the optical depth and large-scale matter and velocity distributions is inevitable, but can be modulated with the effective absorption fraction $F_{\rm abs}$. We also add a random RT line shift to the peculiar velocities. The analytic model for the line shifts matches the results of the lognormal simulations.

While both the analytic and the lognormal models for IGM absorption predict a stronger power spectrum suppression at small scales, their predictions for the dependence of the suppression on $k_\parallel$ and $k_\perp$ differ because the linear approximations break down in the nonlinear environment of the galaxies.

Finally, we have implemented the modified SIMPLE code to model the effects of Lyα RT on the LAE and LAE-Lyα intensity power spectra for a HETDEX-like experiment. The line shift and broadening from RT significantly change the monopoles and quadrupoles of the LAE auto-power spectrum. The IGM absorption also changes the LAE auto-power spectrum significantly even at an absorption fraction of 0.2, while the LAE-Lyα intensity cross-power spectrum remains within the measurement uncertainty even with an absorption fraction of 0.5. Therefore, HETDEX will help constrain the interplay of Lyα RT and galaxy clustering. Our lognormal framework will be useful for the interpretation

of upcoming large-scale structure measurements using Lyα emission.

6.8 Appendix: Half-Sum-Half-Difference Method for LIM

Ando et al. (2018) introduce the half-sum-half-difference (HSHD) method for galaxy clustering to automatically remove shot noise. This is especially useful if the shot noise is anisotropic or scale-dependent. One randomly splits the galaxy sample into two halves and calculates the density contrast for each, $\delta_{\rm g1}$ and $\delta_{\rm g2}$. Then one calculates the half sum (HS) and half difference (HD) of the two fields:

$$\mathrm{HS} = \frac{1}{2}\left(\delta_{\rm g1} + \delta_{\rm g2}\right);\ \mathrm{HD} = \frac{1}{2}\left(\delta_{\rm g1} - \delta_{\rm g2}\right). \tag{6.35}$$

One can then calculate

$$\langle \widetilde{\mathrm{HS}}(\mathbf{k})\, \widetilde{\mathrm{HS}}^{*}(\mathbf{k}')\rangle - \langle \widetilde{\mathrm{HD}}(\mathbf{k})\, \widetilde{\mathrm{HD}}^{*}(\mathbf{k}')\rangle = (2\pi)^3\, \delta_D(\mathbf{k}-\mathbf{k}') P^{\rm auto}_{\rm HSHD}(\mathbf{k}), \tag{6.36}$$

where δ_D is the Dirac delta function. The auto-power spectrum of HS contains the signal and shot noise, while that of HD only contains the shot noise, so that equation (6.36) contains only the signal. As shorthand, we use the notation

$$\hat{P}^{\rm auto}_{\rm HSHD} = \langle \mathrm{HS}\, \mathrm{HS}^{*}\rangle - \langle \mathrm{HD}\, \mathrm{HD}^{*}\rangle. \tag{6.37}$$

Wang et al. (2024) extend this method to the cross-power spectrum of two galaxy catalogs A and B, where each galaxy in A corresponds to a galaxy in B. One splits the catalogs into two halves so that A_1 and B_1 (and A_2 and B_2) maintain the one-to-one correspondence, and then calculates $\mathrm{HS}_{A/B}$ and $\mathrm{HD}_{A/B}$. The shot-noise-free cross-power spectrum estimator takes the form

$$\hat{P}^{\rm cross}_{\rm HSHD} = \langle \mathrm{HS}_A \mathrm{HS}^{*}_B\rangle - \langle \mathrm{HD}_A \mathrm{HD}^{*}_B\rangle. \tag{6.38}$$

LIM surveys typically have low resolution, such that the separation into two galaxy samples is not possible. After all, measuring the integrated emission from all galaxies within a resolution element is the main concept of LIM. Instead of separating galaxy samples, one can still separate observations of the same volume at different times and cross-correlate these to remove the intensity noise power spectrum. The HETDEX survey, however, is not designed solely for LIM, but has a higher resolution in order to detect individual galaxies, which can be artificially decreased for LIM. In this case, it may be possible to separate galaxies within the same LIM voxel.

For modeling purposes, it is helpful to separate the clustering power spectrum from the shot noise and intensity noise. In the mock, we can randomly split the galaxies contributing to the intensity map into two separate samples and calculate $\delta I_{{\rm g}i} = I_{{\rm g}i} - \langle I_{{\rm g}i}\rangle$ for each sample i. Because only half of the galaxies contribute, the mean intensity is halved $\langle I_{{\rm g}i}\rangle = \frac{1}{2}\langle I_{\rm all}\rangle$. Let us assume that $I_{{\rm g}i}$ includes uncorrelated intensity noise with

the same variance $\langle I^2_{\rm noise}\rangle = \sigma^2_{\rm I}$, so that $\langle \delta I_{{\rm g}i} \delta I^*_{{\rm g}j}\rangle = \langle \delta I_{{\rm g}i,{\rm signal}} \delta I^*_{{\rm g}j,{\rm signal}}\rangle + \sigma^2_{\rm I}\delta_{ij}$, where δ_{ij} is the Kronecker delta. Defining

$$\mathrm{HS}_I = \delta I_{\rm g1} + \delta I_{\rm g2} \text{ and } \mathrm{HD}_I = \delta I_{\rm g1} - \delta I_{\rm g2}, \tag{6.39}$$

the power spectrum estimator becomes

$$\begin{aligned}\hat{P}_{II} &= \langle \mathrm{HS}_I \mathrm{HS}^*_I\rangle - \langle \mathrm{HD}_I \mathrm{HD}^*_I\rangle = 2\left(\langle \delta I_{\rm g1}\delta I^*_{\rm g2}\rangle + \langle \delta I_{\rm g2}\delta I^*_{\rm g1}\rangle\right) \\ &= \frac{1}{2}\frac{\langle I_{\rm all}\rangle^2}{\langle I_{{\rm g}i}\rangle^2}\left(\langle \delta I_{\rm g1}\delta I^*_{\rm g2}\rangle + \langle \delta I_{\rm g2}\delta I^*_{\rm g1}\rangle\right).\end{aligned} \tag{6.40}$$

The power spectrum estimator has the same normalization $\langle I_{\rm all}\rangle^2$ as the "normal" power spectrum estimator $\langle \delta I_{\rm all}\delta I^*_{\rm all}\rangle$. Note that HS_i and HD_i contain the intensity noise term $2\langle\sigma^2_I\rangle$, which cancels out in $\hat{P}_{II}$ as long as the intensity noise contributions of $\delta I_{\rm g1}$ and $\delta I_{\rm g2}$ are uncorrelated.

Similarly, the cross-power spectrum with a galaxy sample separated into $\delta_{\rm g1}$ and $\delta_{\rm g2}$ can be estimated as

$$\hat{P}_{I{\rm g}} = \langle \mathrm{HS}\,\mathrm{HS}^*_I\rangle - \langle \mathrm{HD}\,\mathrm{HD}^*_I\rangle, \tag{6.41}$$

which does not contain shot noise as long as $\delta_{\rm g1}$ and $\delta I_{\rm g2}$ (and $\delta_{\rm g2}$ and $\delta I_{\rm g1}$) do not share galaxies contributing to both fields. This can be achieved by using the same galaxy split for $\delta_{{\rm g}i}$ and $\delta I_{{\rm g}i}$ in the mock.

The treatment of intensity noise and sky subtraction for the mocks is not straightforward. To ensure that the noise power spectrum is removed along with the shot noise, the noise maps of $\delta I_{\rm g1}$ and $\delta I_{\rm g2}$ must be uncorrelated. Depending on the split in the real data, the effective σ_I of the subsamples is different than that of the total intensity. Staying agnostic, we add noise with the variance of the total intensity noise to each $\delta I_{{\rm g}i}$. Because the galaxy split cannot be done after the sky subtraction in the mocks, we apply the sky subtraction to each $\delta I_{{\rm g}i}$. In a realistic setting, the sky is subtracted before splitting the data.

Acknowledgments

I thank the anonymous referee for providing helpful comments. I thank E. Komatsu for helpful discussions and comments on the draft. I acknowledge J. Niemeyer, C. Byrohl, and J. Bernal for insightful comments on the draft, and M. Gronke for interesting discussions. Thank you to T. Niemeyer and V. Niemeyer for their help with a calculation.

Computations were performed on the HPC system Freya at the Max Planck Computing and Data Facility.

I acknowledge the use of the astropy (Astropy Collaboration et al., 2013, 2018), matplotlib (Hunter, 2007), numpy (Harris et al., 2020), scipy (Virtanen et al., 2020), pmesh (Feng et al., 2017), dask (Rocklin, 2015), and h5py (Collette, 2013) software packages.

Chapter 7

Lyα Intensity Mapping in HETDEX: Galaxy-Lyα Intensity Cross-Power Spectrum

The content of this chapter has been published in Lujan Niemeyer et al. (2026).

We present a measurement of the Lyman-α (Lyα) intensity mapping power spectrum from the Hobby-Eberly Telescope Dark Energy Experiment (HETDEX). We measure the cross-power spectrum of the Lyα intensity and Lyα-emitting galaxies (LAEs) in a redshift range of $1.9 \leq z \leq 3.5$. We calculate the intensity from HETDEX spectra that do not contain any detected LAEs above a signal-to-noise ratio of 5.5. To produce a power spectrum model and its covariance matrix, we simulate the data using lognormal mocks for the LAE catalog and Lyα intensity in redshift space. The simulations include the HETDEX sensitivity, selection function, and mask. The measurements yield the product of the LAE bias, the intensity bias, the mean intensity of undetected sources, and the ratio of the actual and fiducial redshift-space distortion parameters, $b_{\rm g} b_I \langle I \rangle \bar{F}_{\rm RSD} / \bar{F}^{\rm fid}_{\rm RSD}$ =(6.7$\pm$ 3.1), (11.7 $\pm$ 1.4), and (8.3 $\pm$ 1.5) $\times 10^{-22}$ erg s^{-1} cm^{-2} arcsec^{-2} Å^{-1} in three redshift bins centered at $\bar{z} = 2.1$, 2.6, and 3.2, respectively. The results are reasonably consistent with cosmological hydrodynamical simulations that include Lyα radiative transfer. They are, however, significantly smaller than previous results from cross-correlations of quasars with Lyα intensity. These results demonstrate the statistical power of HETDEX for Lyα intensity mapping and pave the way for a more comprehensive analysis. They will also be useful for constraining models of Lyα emission from galaxies used in modern cosmological simulations of galaxy formation and evolution.

7.1 Introduction

Line intensity mapping (LIM) is a novel tool for studying cosmology and the astrophysics of galaxies and intergalactic gas (see Kovetz et al., 2017; Bernal & Kovetz, 2022, for reviews). Intensity maps of one or more emission lines in large volumes can be used

as biased tracers of the underlying matter distribution. Instead of detecting galaxies as peaks in the intensity above the noise level, LIM can use noisy data to measure summary statistics of the intensity such as N-point correlation functions. It thus incorporates photons from all galaxies and diffuse gas within the survey volume that would otherwise remain undetected.

Several LIM surveys are currently in operation (Doré et al., 2014; Santos et al., 2016; DeBoer et al., 2017; CONCERTO Collaboration et al., 2020; Cleary et al., 2022; Karkare et al., 2022), or in preparation (Vieira et al., 2020; Sun et al., 2021; Switzer et al., 2021; CCAT-Prime Collaboration et al., 2022; Renard et al., 2024; Renard et al., 2025). The emission lines targeted by these surveys span from the ultraviolet (Lyman-α, or short, Lyα) to the radio (21 cm) and are emitted by atomic or molecular gas. LIM has produced detections of cross- and auto-power spectra for the 21 cm and CO lines (e.g., Chang et al., 2010; Keating et al., 2016, 2020; Cunnington et al., 2023c; Paul et al., 2023). Other surveys have provided upper limits on the auto-power spectrum (e.g., Cleary et al., 2022; Keenan et al., 2022) or employ stacking (e.g., Dunne et al., 2024, 2025; Chen et al., 2025).

The Lyα emission line of atomic hydrogen enables observations of high-redshift galaxies, called Lyα emitters (LAEs), and of diffuse gas in the circumgalactic and intergalactic media (CGM and IGM; e.g., Ouchi et al., 2020). Stacking of LAEs (Wisotzki et al., 2018; Kakuma et al., 2021; Kikuchihara et al., 2022b; Lujan Niemeyer et al., 2022a; Kikuta et al., 2023; Trainor et al., 2025, e.g.,) and other galaxies (Steidel et al., 2011; Kusakabe et al., 2022) has revealed the ubiquity of Lyα emission out to projected distances of $\simeq 1\,\mathrm{Mpc}$ from these galaxies. However, few Lyα LIM measurements have been performed on larger scales. All distance units refer to comoving distances unless otherwise specified.

Croft et al. (2016, 2018) and Lin et al. (2022) present a measurement of the cross-correlation of quasars (QSOs) from the Baryon Oscillation Spectroscopic Survey (BOSS) and the Extended Baryon Oscillation Spectroscopic Survey (eBOSS) with Lyα intensity from a projected radius of $r_\perp = 0.5\,h^{-1}\mathrm{Mpc}$ to $150\,h^{-1}\mathrm{Mpc}$ (see Lin et al., 2022, for a measurement from $r_\perp = 0.1\,h^{-1}\mathrm{Mpc}$). The Ly$\alpha$ intensity is measured by subtracting the best-fit galaxy spectrum from spectra of luminous red galaxies (LRGs), assuming that the residual consists of background Lyα intensity and noise. By measuring the QSO-Lyα intensity cross-correlation, they constrain the average cosmic Lyα luminosity density. However, these results are inconsistent with the upper limit on the cosmic Lyα luminosity density inferred from the Lyα forest, as reported in Croft et al. (2018). While Croft et al. (2018) argue that the intensity is dominated by Lyα emission from QSOs, Lin et al. (2022) conclude that the bulk of the Lyα luminosity originates from star-forming galaxies.

The Hobby-Eberly Telescope Dark Energy Experiment (HETDEX; Gebhardt et al., 2021) provides an ideal data set for Lyα LIM. HETDEX is a galaxy survey aimed at detecting $\sim 10^6$ LAEs at $1.88 < z < 3.52$ in a comoving volume of $10.9\,\mathrm{Gpc}^3$. HETDEX observes its survey area with the Visible Integral Field Replicable Unit Spectrograph (VIRUS; Hill et al., 2021) on the Hobby-Eberly Telescope (HET; Ramsey et al., 1998; Hill et al., 2021) without target preselection. Therefore, most fiber spectra do not contain sources that are bright enough to be detected, with detected LAEs comprising only

$\sim 0.01\%$ of the signal (Mentuch Cooper et al. in preparation). This makes LIM a unique scientific target for HETDEX.

Lujan Niemeyer et al. (2022a,b) stacked the Lyα emission in spectra around HETDEX LAEs and independently detected Lyα-faint galaxies selected by their [O III] emission, and detected Lyα emission out to $\simeq 100\,\mathrm{kpc}$ (proper) in both samples. This stacking signal is closely related to the small-scale angular cross-correlation of galaxies with Lyα intensity. Khanlari et al. (2025) increased the sample size and stacked HETDEX spectra as a function of line-of-sight (LOS) distance and angular separation, finding Lyα absorption halos (see Section 7.9.4 for a discussion). These studies confirm the suitability of HETDEX data for LIM. The simulation study presented in Lujan Niemeyer et al. (2023) (see also Fonseca et al., 2017) predicts a high-significance detection of the LAE-Lyα intensity cross-power spectrum at wavenumbers of $k \simeq 0.08 - 1\,h\,\mathrm{Mpc}^{-1}$ for an ideal HETDEX survey without systematics, while masking the spectra of detected LAEs.

In this paper, we present Lyα LIM results from HETDEX observations. Using the HETDEX LAE catalog and HETDEX spectra, we detect the LAE-Lyα intensity cross-power spectrum with high statistical significance. By masking LAEs that are detected above a certain signal-to-noise ratio (SNR) threshold, we ensure that the intensity contains only undetected sources. We generate mocks with the SIMPLE code (Lujan Niemeyer et al., 2023) to calculate the transfer function and estimate the power spectrum covariance matrix. By shuffling the on-sky positions of the HETDEX intensity, we calculate the uncertainty due to HETDEX noise. We constrain the product of the LAE bias, the intensity bias, the mean intensity of undetected sources, and the ratio of the actual and fiducial redshift-space distortion factors, $b_\mathrm{g} b_I \langle I \rangle \bar{F}_\mathrm{RSD} / \bar{F}^\mathrm{fid}_\mathrm{RSD}$. Finally, we compare our findings to the expected mean intensity from integrating extrapolated LAE luminosity functions, and to a cosmological Lyα radiative transfer (RT) simulation.

Our measurement is complementary to those of Croft et al. (2016, 2018) and Lin et al. (2022) because we cross-correlate the Lyα intensity with different sources, with an emphasis on larger scales, and with higher sensitivity. Because this work does not exclusively use QSOs, but LAEs, our results should not be strongly affected by Lyα emission around QSOs.

This work represents a LIM detection in HETDEX data. We expect to improve data processing and increase the sample size of LAEs and spectra in the future.

This paper is structured as follows. Section 7.2 describes the LAE catalog and spectra obtained by HETDEX and the data processing. It also describes the separation of the observations into smaller regions and the creation of galaxy and intensity maps. Section 7.3 presents lognormal mocks from the SIMPLE code. Section 7.4 explains how we estimate the LAE-Lyα intensity cross-power spectrum. Section 7.5 describes how we fit the power spectrum model to the data. We show the power spectra and constraints on $b_\mathrm{g} b_I \langle I \rangle \bar{F}_\mathrm{RSD} / \bar{F}^\mathrm{fid}_\mathrm{RSD}$ in Section 7.6. We compare our constraints with the QSO-Lyα intensity and Lyα forest-Lyα intensity cross-correlations in Section 7.7. We explore origins of the Lyα intensity in Section 7.8. Section 7.9 discusses potential improvements in data processing and modeling, and the possible effect of Lyα absorption on our measurement. We conclude in Section 7.10.

We use the following Fourier convention:

$$\begin{aligned} \tilde{f}(\mathbf{k}) &= \int \mathrm{d}^3\mathbf{x} f(\mathbf{x}) e^{i\mathbf{k}\cdot\mathbf{x}}\,, \\ f(\mathbf{x}) &= \int \frac{\mathrm{d}^3\mathbf{k}}{(2\pi)^3} \tilde{f}(\mathbf{k}) e^{-i\mathbf{k}\cdot\mathbf{x}}\,, \end{aligned} \tag{7.1}$$

where the tilde denotes quantities in Fourier space. We adopt a flat Λ-cold-dark-matter cosmology with $H_0 = 67.66\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_{\mathrm{b},0}h^2 = 0.022$, $\Omega_{\mathrm{m},0}h^2 = 0.142$, $\sum m_\nu = 0.06\,\mathrm{eV}$, $\ln(10^{10}A_s) = 3.047$, and $n_s = 0.9665$ (Planck Collaboration et al., 2020b). We refer to the specific intensity I_λ, the intensity per unit observed wavelength, as the intensity I for simplicity.

7.2 HETDEX Data

Our analysis uses the data of the internal HETDEX data release HDR5 (Mentuch Cooper et al. in preparation). These data are collected from 2017-01-01 to 2024-07-31 and comprise the full main HETDEX survey. The spectra cover $90\,\mathrm{deg}^2$ of sky, $87.7\mathrm{deg}^2$ of which meet the science quality criteria. These data include observations within an equatorial "Fall" field, whose footprint covers $150\,\mathrm{deg}^2$, a higher-declination "Spring" field, which covers $390\,\mathrm{deg}^2$, and several smaller fields, which were used for HETDEX science verification (Gebhardt et al., 2021). We also use observations of the Texas Euclid Survey for Lyα (TESLA), which covers a $10\,\mathrm{deg}^2$ region of the North Ecliptic Pole (NEP; Chávez Ortiz et al., 2023). The TESLA data were observed in the same manner, processed in the same way, and have the same exposure times as the HETDEX spectra.

The HETDEX spectra were obtained with VIRUS on the 10 m HET (Ramsey et al., 1998; Hill et al., 2021). VIRUS is composed of up to 78 integral field unit fiber arrays (IFUs), each of which spans $51'' \times 51''$ on the sky and contains 448 fibers that are $1.5''$ in diameter. The fibers from each IFU feed into a spectrograph with two spectral channels. The spectrographs have a resolution of $R \simeq 800$, which corresponds to a Full-Width-at-Half-Maximum (FWHM) of 5.6 Å at $\lambda = 4500$ Å, and cover the wavelengths between 3470 Å and 5540 Å. Each spectral channel is read out by a CCD with two amplifiers; thus, the fibers from each IFU are split into four amplifiers. Each HETDEX observation consists of three six-minute exposures in a triangular dithering pattern that fills the gaps between the fibers on each IFU.

The IFUs are spread out over the HET's $18'$ field of view on a $100''$-grid, so that the gap between each IFU is roughly equal to the size of the IFU. As a result, the gaps between the IFUs lead to a fill factor of $\sim 1/4.6$ for each observation. The observations in the NEP field are dithered to fill the gaps between the IFUs, while these gaps remain in the Spring and Fall fields.

For this work, we only include observations that are not masked by the HETDEX flag for low-quality observations. As HETDEX requires a minimum effective throughput of 0.08 at 4540 Å for an LAE to be included in the catalog, we only select spectra from observations that meet this requirement. Additionally, we exclude HETDEX observations before 2018 because these early data contain more artifacts and larger sky emission residuals. This approach leaves 1,949 observations in the Fall field, 3,797 in the Spring field,

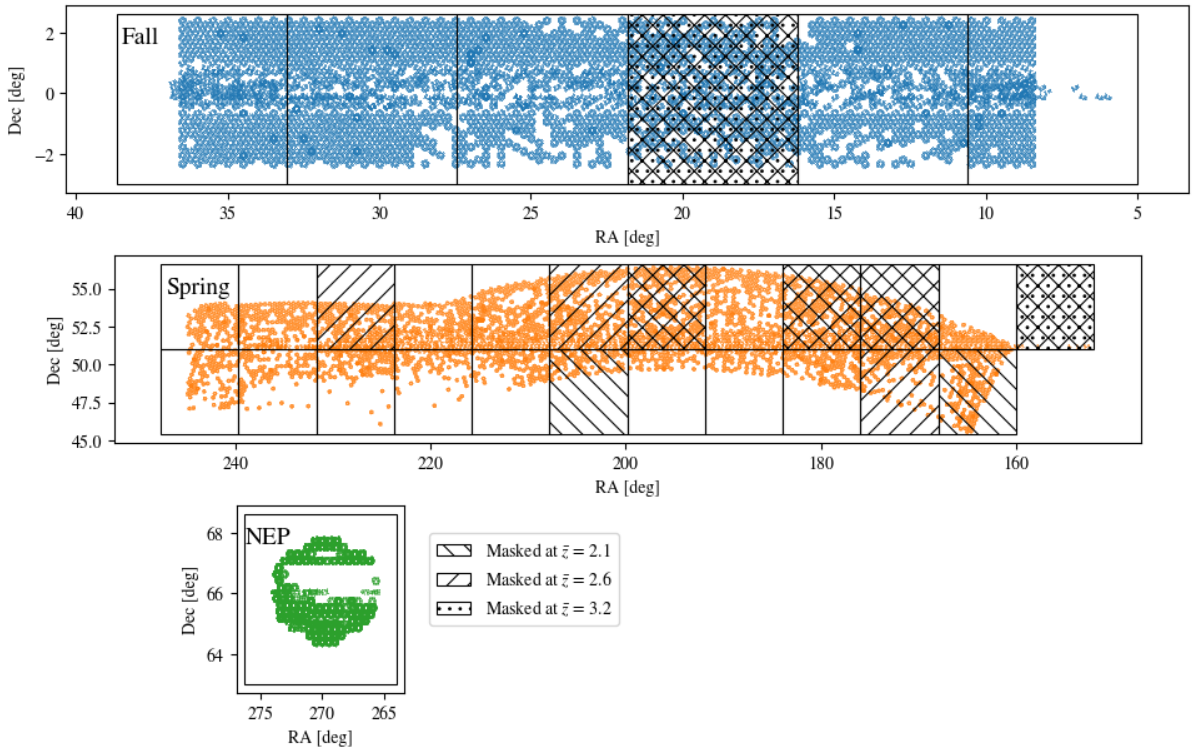

Figure 7.1: IFU coordinates in the Fall (top), Spring (middle), and NEP (bottom) fields. The fields are divided into 5.6° wide regions for the power spectrum, shown as black squares (see Section 7.2.3). Masked maps are shown as hatched regions (see Section 7.4.2). The backward-facing (forward-facing) diagonal hatchings represent boxes masked in the low-z (medium-z) bin, and the dotted hatchings indicate those masked in the high-z bin.

and 484 in the NEP field. Figure 7.1 shows the IFU coordinates used in the Fall, Spring, and NEP fields.

7.2.1 LAE Catalog

We select LAEs from catalog 5.0.1 of the internal HETDEX data release HDR5 (Mentuch Cooper et al. in preparation). A public release of an early subset of this sample is found in Mentuch Cooper et al. (2023a), including a description of source detection and classification. For each line detection, the catalog provides the central wavelength; the sky coordinates of the detection; the line width of the best-fit Gaussian to the emission line (σ); the line flux; the SNR of the detection; an aperture correction factor, which quantifies the fraction of a source's light that falls onto an IFU (as opposed to off an edge); and the source type, e.g., LAE or [O II]-emitting galaxy, and the confidence of this classification.

Because the LAE detections will be masked in the Lyα intensity map, there will be no shot noise in the cross-power spectrum. We restrict our LAE sample to high-confidence objects ('flag_best' detections with SNR > 5.5, see Mentuch Cooper et al. in preparation), preferring a smaller, higher-quality set of LAEs to a larger set with more

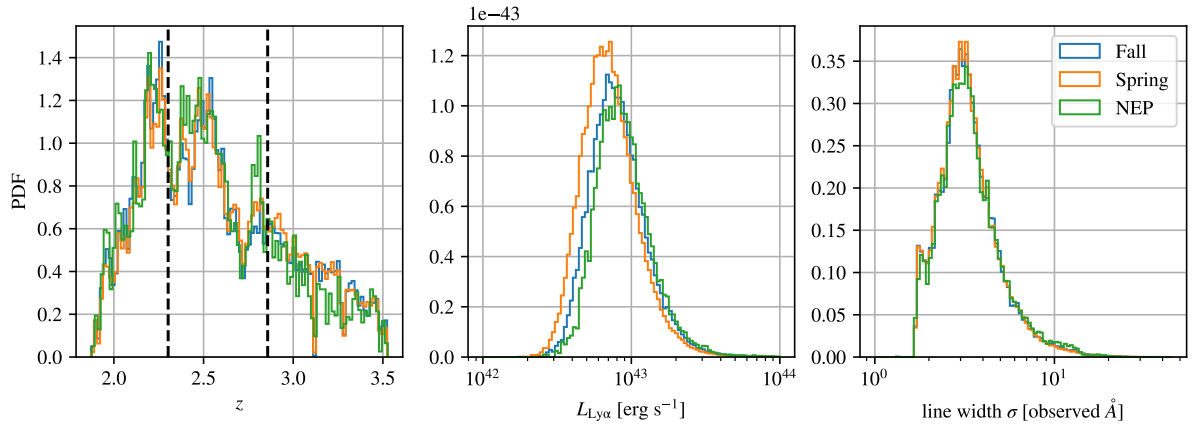

Figure 7.2: The normalized distributions of redshifts (left panel), luminosities (middle panel), and the best-fit Gaussian σ line widths in observed wavelength units (right panel) of HETDEX LAEs in the Fall (blue), Spring (orange), and NEP (green) fields. The black dashed lines in the left panel show the boundaries of the three redshift bins used to calculate the power spectra.

contamination. The 'flag_best' criterion of the HETDEX detection catalog summarizes quality criteria based on known bad data; masks of large galaxies, meteors, and satellites; goodness-of-fit of the line detection; detection confidence based on Machine Learning and citizen science (House et al., 2023, 2024); and other quality criteria (details can be found in Mentuch Cooper et al. 2023a and Mentuch Cooper et al. in preparation). We also select only those sources with an aperture correction factor > 0.4, and require the updated SNR, which incorporates systematics, to be $\mathrm{SNR}_{\mathrm{rres}} > 4.5$. The noise model used for $\mathrm{SNR}_{\mathrm{rres}}$ is the third model described in Section 6.20 of Gebhardt et al. (2021).

We further impose thresholds on the LAE detection confidence parameters derived from machine learning models. We require $p_{\mathrm{conf}} > 0.5$ from a Random Forest classifier based on emission-line measurements (Mentuch Cooper et al., in preparation) and a score > 0.4 from a convolutional neural network (CNN) trained on HETDEX spectra (Mukae et al., in preparation). These classifiers are developed to distinguish LAEs from false-positive detections due to systematic errors in the spectra and sky residuals. For the SNR > 5.5 sample, the adopted CNN threshold yields a precision of $\sim 96\%$ when compared with visual classifications as described in House et al. (2023), effectively classifying LAEs while minimizing false detections. Using these criteria, 90.4% of detections in the COSMOS science verification field are confirmed in a different HETDEX observation of the same field or a separate galaxy catalog. Thus, the fraction of false positive detections using these criteria is $< 9.6\%$.

We then remove detections within ± 10 Å of the [O III] emission line at 5007 Å to mitigate contamination from high-Galactic latitude planetary nebulae. We also exclude LAE candidates within ± 10 Å of the sky emission lines at 5200 Å and 5457 Å.

Figure 7.2 shows the redshift, luminosity, and line-width distributions of the LAE samples in the Fall, Spring, and NEP fields. The dip at $z \simeq 2.7$ in the redshift distributions is due to a mask that is applied at the center of 50% of the detectors as well as an increase in night sky emission (Mentuch Cooper et al., 2023a). The dip at $z \simeq 2.3$

is also due to bright sky emission. There remain 37,773 LAEs in the Fall field, 95,774 in the Spring field, and 9,372 in the NEP field. After removing detections in IFUs that we mask (see Section 7.2.2), 36,303 LAEs remain in the Fall field, 92,192 in the Spring field, and 8,932 in the NEP field. This number is much lower than the total number of detected LAEs in HETDEX because of our strict selection criteria.

We mask these LAEs in the intensity map for the power spectrum measurement, as described in Section 7.2.2.

7.2.2 Intensity Spectra

In this section, we describe the additional data reduction steps we apply to the HETDEX spectra. The intensity mapping results can be sensitive to reduction artifacts and our goal is to reduce those as much as possible.

Removing Sky Emission

Gebhardt et al. (2021) describe the HETDEX data processing pipeline. A crucial part of the processing for low-surface brightness measurements is sky subtraction. "Sky" emission, including airglow, zodiacal light, and foreground emission from the Milky Way, dominates the observed spectra and must be removed.

HETDEX employs two methods for sky subtraction. The first method uses a 'local' sky estimator, measured using the 112 fibers present on each CCD amplifier; these fibers cover a $\sim 13'' \times 51''$ field. This sky subtraction technique is optimized for LAE detection. The second method computes a 'full-frame' sky, derived from the photons detected on all $\sim$ 35k fibers of the $18'$ field of view of the VIRUS instrument within one exposure (see Lujan Niemeyer et al., 2022a, for details). As demonstrated in Lujan Niemeyer et al. (2023), the galaxy-intensity cross-power spectrum is suppressed on scales larger than the sky subtraction scale. To keep information on a scale as large as possible, we use the 'full-field' sky subtraction in this work.

Our IFU data can be visualized in a two-dimensional figure, where the x-axis is the wavelength, and the y-axis is the fiber number. When averaging the full-frame sky-subtracted spectra of each amplifier over all observations taken within a month, one would expect a mean of zero. However, we find systematic two-dimensional patterns that are unique to each amplifier. To remove these residuals, we first divide each fiber spectrum by the full-frame sky spectrum of the same fiber. We then calculate the biweight location (BL) of these relative residuals across observations within a month. The BL and biweight scale ($\sigma_{\rm bw}$) are estimators of the central location and scatter of a distribution that are robust to outliers (Beers et al., 1990). In each observation, we subtract this BL relative residual multiplied by the respective sky spectrum for each fiber.

Masking Bad Detectors, Bright Voxels, Continuum, and [O II] Detections

Processed HETDEX spectra are output in 2 Å bins between 3470 Å and 5540 Å. When all 78 IFUs are operational, $3 \times 34,944$ spectra are obtained from each observation. However, some of the CCD amplifiers and a few of the individual fibers produce bad data in certain

observations. The bad amplifiers, fibers, and pixels (including those affected by cosmic rays) are flagged by the HETDEX data processing software as described by criteria set in Table 2 of Mentuch Cooper et al. (2023a), and are masked in our analysis. We also mask spectra containing large galaxies and meteors using the fiber mask provided by the HETDEX Application Programming Interface (API; see Section 2 of Mentuch Cooper et al., 2023a)

For each observation, we also mask outlier pixels individually. First, we calculate the BL and $\sigma_{\rm bw}$ of all the spectra, and then mask all pixels with fluxes that deviate by $> 3\sigma_{\rm bw}$ from the BL of the fluxes.

Since HETDEX is an untargeted survey, the spectra of continuum objects, such as foreground stars and galaxies, are also in the HETDEX database, and these objects can contaminate the Lyα intensity map. We therefore remove all fiber spectra containing continuum emission. To do this, we estimate the continuum within each fiber by calculating the BL of its spectrum within the wavelength range of $4100 - 5100\,$Å, and mask out those fibers with a $3\sigma_{\rm bw}$ deviation from the BL of the continuum values within each observation.

Because light can scatter within the detector, we also mask out fibers adjacent to each masked fiber with continuum emission on the CCD. Spectra adjacent in the array obtained from the HETDEX Application Programming Interface (API) are adjacent on the CCD, unless they are the first or last fiber on the CCD. We mask all adjacent fibers in the array, which sometimes includes the first or last fiber on an adjacent amplifier.

We set the spectra of the fibers within $10''$ of all [O II]-emitting galaxies to zero. The detection criteria match those used for LAEs, as described in Section 7.2.1, except that the source type indicates [O II]-emission. For these objects, we do not apply cuts based on the CNN score, $p_{\rm conf}$, or ${\rm SNR}_{\rm rres}$.

Separating Intensity Contributions from LAEs and Undetected Sources

We separate the spectra into contributions from detected LAEs and undetected sources as described in this subsection. In this step, we consider all LAEs selected in Section 7.2.1 as 'detected.' We select voxels within $10''$ of an LAE's positional centroid and within 2.5 times the line width of its wavelength centroid. For the 'undetected' spectra, we set these voxels to zero. For the 'detected' spectra, we instead set all other voxels to zero. Adding the 'detected' and 'undetected' spectra therefore yields the total spectra. This process translates directly to the power spectra: adding the cross-power spectrum of LAEs with the 'detected' intensity to that of LAEs with the 'undetected' intensity has to be equal to the cross-power spectrum of LAEs with the 'total' intensity. We use this property as a check for the power spectrum pipeline. Henceforth, we apply the same data processing to the 'detected,' 'total,' and 'undetected' spectra, unless specified otherwise.

Lowering Angular Resolution

The angular resolution of the HET is much better than that of typical LIM experiments: the FWHM of the VIRUS point-spread function (PSF) is between $1.2''$ and $\simeq 4''$. For computational simplicity, we condense 3×448 spectra taken by each IFU during a single

HETDEX observation into one spectrum. We have 132,051 IFU spectra in the Fall field, 241,969 in the Spring field, and 34,239 in the NEP field. Each spectrum represents the average flux within $51'' \times 51'' \times 2\,\text{Å}$ voxels, which corresponds to $1.5\,\text{Mpc} \times 1.5\,\text{Mpc} \times 1.9\,\text{Mpc}$ at $z = 2.5$. We will henceforth refer to these as 'IFU spectral elements.'

We calculate the mean of the 'total,' 'detected,' and 'undetected' fiber spectra within each IFU. We also calculate the standard deviation of the 'total' fiber spectra within each IFU and save the mean coordinates of all the IFU's fibers.

Finally, we calculate the inverse variance of the 'total' fiber spectra of each pixel. Some voxels have much higher values than the typical distribution because few unmasked elements of the fiber spectra contribute to it. We therefore set all pixels with inverse variances of $> 3600/u_I^2$ to zero, where $u_I = 10^{-17}\,\text{erg}\,\text{s}^{-1}\,\text{cm}^{-2}\,\text{arcsec}^{-2}\,\text{Å}^{-1}$. This removes everything with a deviation $\gtrsim 5.7\sigma_{\text{bw}}$ from the BL of the inverse variance distribution.

Calculating the Intensity and Removing Extinction

The spectra are output as specific fluxes, f_λ, per 2 Å bin in each fiber. The intensity is obtained from f_λ as $I(z) = f_\lambda(z)/A_{\text{fiber}}$, where the fiber area is $A_{\text{fiber}} = \pi\,(0.75'')^2$.

To obtain the intensity as would be seen outside the Milky Way, we remove the reddening caused by Galactic extinction. We collect the color-excess values, $E(B-V)$, as measured by Schlegel et al. (1998) at the coordinates of each IFU, using the *dustmaps* API (Green, 2018). Following Schlafly & Finkbeiner (2011), we assume a $R_V = 3.1$ reddening model of Fitzpatrick (1999) and set $A_V = 2.742E(B-V)$. We calculate and apply the extinction correction using the open-source Python module extinction[1].

Removing More Systematics

We combine all IFU spectra within each of the three fields for the following steps.

First, we again mask those voxels that are much fainter or brighter than the rest of the distribution, i.e., those with values above $0.0125u_I$ or below $-0.0125u_I$. We determined this threshold empirically; this cut removes everything with a $\gtrsim 7\sigma_{\text{bw}}$ deviation from the BL of the pixel distribution, which is $\simeq 0$.

We expect that, after the subtraction of foreground sources and sky emission, the spectra of each IFU will be dominated by sky noise. However, some spectra exhibit systematic patterns, indicating imperfect sky subtraction or other detector problems. Since these spectra typically have a higher standard deviation along the wavelength direction, we mask out those spectra with a standard deviation along the wavelength direction $> 0.0035\,u_I$. This maximum value was determined empirically; the cut removes spectra whose standard deviation along the wavelength direction deviates $\gtrsim 3.3\sigma_{\text{bw}}$ from the BL of the distribution.

The residuals from sky subtraction are most prominent around the emission lines in the sky foreground, including light from street lamps. We therefore mask all pixels of the IFU spectra at $4350 - 4370\,\text{Å}$, $5190 - 5210\,\text{Å}$, and $5447 - 5467\,\text{Å}$. We also mask $\pm 10\,\text{Å}$ around the rest-frame [O III] emission line at $5007\,\text{Å}$ to mitigate contamination from planetary nebulae and diffuse gas within the Milky Way.

[1] `https://github.com/sncosmo/extinction/`

Principal Component Analysis

Despite the extensive masking described above, the spectra still contain sky emission residuals. We use a principal component analysis (PCA) to remove these sky emission residuals from the spectra.

We perform the PCA on the Fall, Spring, and NEP fields separately. We first subtract the mean spectrum of each field, $\langle I(\lambda)\rangle = N_{\rm IFU}^{-1}\sum_{i=1}^{N_{\rm IFU}} I_i(\lambda)$, and normalize,

$$X_i(\lambda) = \frac{I_i(\lambda) - \langle I(\lambda)\rangle}{\sigma_I(\lambda)}, \tag{7.2}$$

where the subscript i refers to the IFU and $\sigma_I(\lambda)$ is the standard deviation of the intensity within each wavelength slice. $N_{\rm IFU}$ is the number of IFU spectra used for the analysis, which is 132,051 in the Fall field, 241,969 in the Spring field, and 34,239 in the NEP field. Of these IFU spectra, 5,882 (Fall), 13,851 (Spring), and 1,803 (NEP) are fully masked in the steps described in Sections 7.2.2, 7.2.2, and 7.2.2. In the remaining fibers, $\simeq 7\%$ of the voxels are masked. We then set all masked intensity values to zero.

We define a vector of length N_λ as

$$\mathbf{X}_i = \sum_{j=1}^{N_\lambda} X_i(\lambda_j)\hat{\mathbf{e}}_j^\lambda = \left(X_i(\lambda_1), X_i(\lambda_2), \ldots, X_i(\lambda_{N_\lambda})\right)^\top, \tag{7.3}$$

where $N_\lambda = 1036$ is the number of wavelength bins per spectrum. The jth value of the vector $\mathbf{X}_i$ is the normalized intensity fluctuation in the jth wavelength bin, $X_i(\lambda_j)$. The value of the basis vector, $\hat{\mathbf{e}}_j^\lambda$, is equal to 1 in the jth wavelength bin and zero otherwise.

We define an $N_{\rm IFU}$-by-N_λ matrix, $\mathbf{X}$, which contains the vectors $\mathbf{X}_i$ as its rows. PCA then calculates an N_λ-by-N_λ covariance matrix, $\mathbf{X}^\top\mathbf{X}$. The eigenvectors of the covariance matrix are the weight vectors. These form an orthonormal basis $\hat{\mathbf{e}}_j^{\rm PCA}$, such that

$$\mathbf{X}_i = \sum_{j=1}^{N_\lambda} y_{ij}\hat{\mathbf{e}}_j^{\rm PCA}. \tag{7.4}$$

PCA arranges the eigenvectors $\hat{\mathbf{e}}_j^{\rm PCA}$ in order of decreasing variance of y_{ij}, with the variance being highest for $j = 1$. The variance of y_{ij} is proportional to the jth eigenvalue.

Figure 7.3 shows the first seven PCA weight vectors, $\hat{\mathbf{e}}_j^{\rm PCA}$ with $j = 1-7$, of the 'total' spectra in the fall field; an example 'full-field' estimated sky spectrum; and the eigenvalues corresponding to the first 200 weight vectors. The first weight vectors contain features of the sky spectrum, such as the K and H lines at 3934 Å and 3968 Å.

The high-variance weight vectors are dominated by systematics such as sky emission residuals. We therefore set the first $N_{\rm PC} \in \{10, 20, ..., 200\}$ of the principal components (PCs) to zero and work with

$$\mathbf{X}_i^{\rm PCA} = \sum_{j=N_{\rm PC}+1}^{N_\lambda} y_{ij}\hat{\mathbf{e}}_j^{\rm PCA}. \tag{7.5}$$

To obtain clean spectra, we perform an inverse coordinate transformation,

$$\mathbf{X}_i^{\rm PCA} = \sum_{k=1}^{N_\lambda}\sum_{j=N_{\rm PC}+1}^{N_\lambda} y_{ij}\left(\hat{\mathbf{e}}_k^{\lambda\top}\hat{\mathbf{e}}_j^{\rm PCA}\right)\hat{\mathbf{e}}_k^\lambda, \tag{7.6}$$

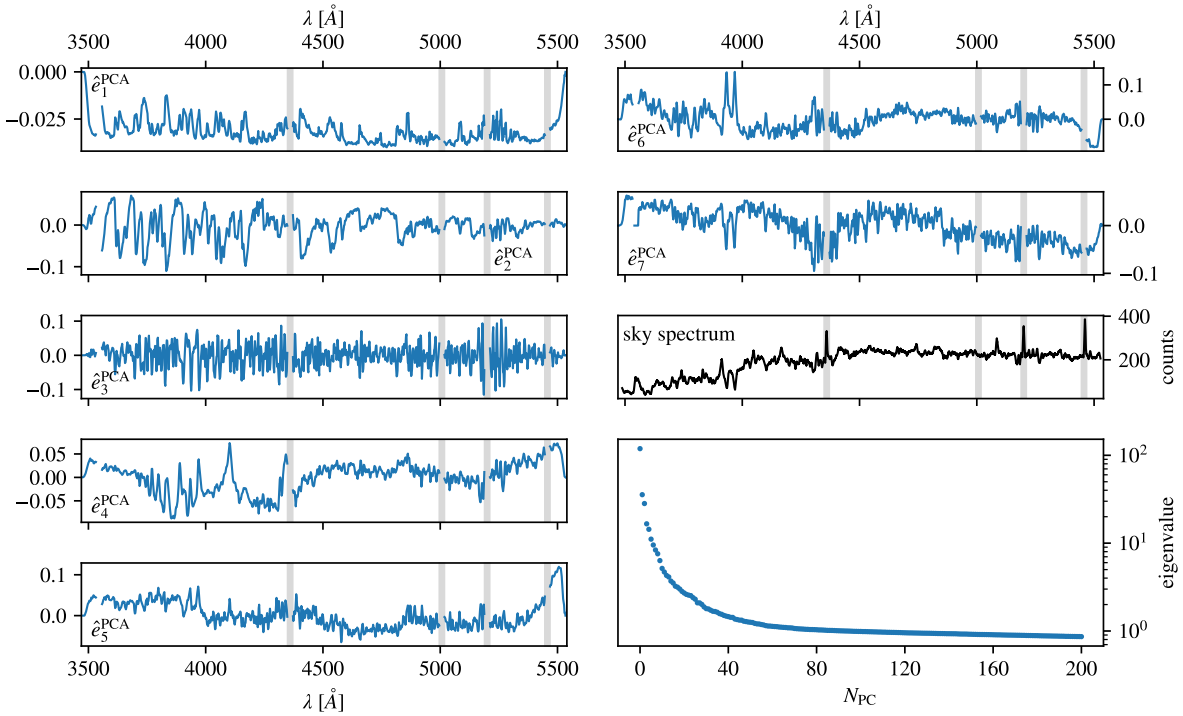

Figure 7.3: First seven weight vectors, $\hat{\mathbf{e}}_j^{\mathrm{PCA}}$ with $j = 1 - 7$, of the Fall field are displayed as functions of wavelength. All $\hat{\mathbf{e}}_j^{\mathrm{PCA}}$ are normalized vectors. For comparison, we also show an example 'full-field' sky spectrum in counts (third panel on the right). Masked wavelength regions are shown in gray. The bottom right panel shows the first 200 eigenvalues, which are proportional to the variance along the corresponding weight vectors.

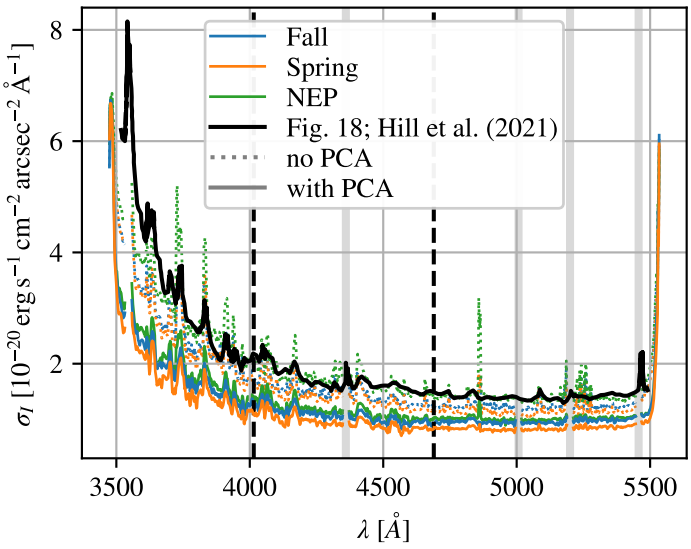

Figure 7.4: Standard deviation of the IFU spectra in each wavelength slice in the Fall (blue), Spring (orange), and NEP (green) fields compared to the pure-noise expectation inferred from Figure 18 of Hill et al. (2021) (black). The dashed colored lines show the standard deviation of the IFU spectra without PCA ($N_{\rm PC} = 0$). The solid colored lines are the standard deviation of the IFU spectra after PCA cleaning, where $N_{\rm PC} = 170$ in the low-z bin, $N_{\rm PC} = 130$ in the medium-z bin, and $N_{\rm PC} = 100$ in the high-z bin. The vertical black dashed lines indicate the boundaries of the three redshift bins used for power spectrum measurements. The gray shaded areas show masked wavelength values.

so that

$$X_i^{\rm PCA}(\lambda_k) = \sum_{j=N_{\rm PC}+1}^{N_\lambda} y_{ij} \left(\hat{\mathbf{e}}_k^{\lambda\top} \hat{\mathbf{e}}_j^{\rm PCA}\right), \tag{7.7}$$

and undo the normalization:

$$I_i^{\rm PCA}(\lambda) = X_i^{\rm PCA}(\lambda)\sigma_I(\lambda). \tag{7.8}$$

We then re-apply the mask.

We calculate the weight vectors, $\hat{\mathbf{e}}_j^{\rm PCA}$, from the covariance matrix $\mathbf{X}^\top\mathbf{X}$ of the normalized 'total' spectra. We then use these vectors to clean the 'total,' 'detected,' and 'undetected' spectra. This ensures that the sum of the 'detected' and 'undetected' spectra equals the 'total' spectra. Since most spectral elements do not contain a detected LAE, the 'total' and 'undetected' spectra are similar.

The PC removal also removes cosmological signal. We model this loss in the mocks in Sections 7.3.2 and 7.4.3.

Comparison with Noise

Figure 7.4 compares the measured and expected standard deviations of the IFU spectra in the absence of astrophysical sources, as a function of wavelength. We translate the measured 5σ sky and read noise per fiber per spectral resolution element (FWHM= 5.6Å)

of Figure 18 in Hill et al. (2021) to the intensity noise per fiber by dividing it by $(5 \times \pi(0.75'')^2 \times 5.6\text{Å})$. This assumes that the noise of the pixels within a resolution element is fully correlated, as expected if the noise is dominated by sky photons. The noise is indeed dominated by sky photons at $\lambda \gtrsim 4000\,\text{Å}$ (Hill et al., 2021). At shorter wavelengths, it is read-noise dominated, which is not correlated within a spectral resolution element. The expected intensity noise per IFU spectrum is obtained by dividing this value by $\sqrt{3 \times 448}$ for 3 dithers and 448 fibers per IFU per dither. This measured standard deviation of the IFU spectra without PCA cleaning is similar to the expectation inferred from Hill et al. (2021) in the NEP field, but slightly smaller at most wavelengths in the Fall and Spring fields. This discrepancy could be due to masking outlier pixels and high-variance spectra. After PCA cleaning, the standard deviation is smaller at most wavelengths. This is a natural consequence of our removal of PCs with the highest variance.

7.2.3 Creation of Intensity Maps

Due to the loss of power from sky subtraction (see Lujan Niemeyer et al., 2023), we measure the power spectrum at wavenumbers $k > 0.08\,h\text{Mpc}^{-1}$. To accommodate the required high angular resolution, we split the three fields into smaller regions with a side length of 5.6° for the maps.

We divide the redshift range into three redshift bins from 1.85 to 2.3 ($\bar{z} = 2.07$), from 2.3 to 2.86 ($\bar{z} = 2.58$), and from 2.86 to 3.56 ($\bar{z} = 3.21$). This binning yields $23 \times 3 = 69$ maps for the Spring field and $6 \times 3 = 18$ maps for the Fall field. The NEP field fits within one such map in each of the three redshift bins. Figure 7.1 shows the map division in the three fields.

Each map is contained within a cubic box with a side length of $432\,h^{-1}\,\text{Mpc}$ for the low-z and medium-z bins and $463\,h^{-1}\,\text{Mpc}$ for the high-z bin. The fundamental frequencies are $k_\text{F} = 0.015\,h\,\text{Mpc}^{-1}$ and $0.014\,h\,\text{Mpc}^{-1}$, respectively. We create cubic voxels that are $2\,h^{-1}\text{Mpc}$ long, corresponding to a Nyquist frequency of $k_\text{Ny} = 1.57\,h\,\text{Mpc}^{-1}$.

To create the intensity map, we collect all the IFU spectral elements contained within each region and redshift bin. Then, we transform the sky and redshift coordinates into Cartesian coordinates and collect the IFU spectral elements in each voxel.

For each voxel, we calculate the mean intensity $I(\mathbf{x})$ of the IFU spectral elements within. We then calculate the intensity fluctuation, $\delta I(\mathbf{x})$, in each voxel,

$$\delta I(\mathbf{x}) = I(\mathbf{x}) - \langle I \rangle(x_0), \tag{7.9}$$

where $\mathbf{x} = (x_0, x_1, x_2)$, x_0 is the position along the LOS axis of the map, x_1 and x_2 are the coordinates perpendicular to the LOS, and $\langle I \rangle(x_0)$ is the mean intensity in each slice along the LOS direction of the map. We calculate $\langle I \rangle(x_0)$ as

$$\langle I \rangle(x_0) = \frac{1}{N_\text{vox}^2} \sum_{i=1}^{N_\text{vox}} \sum_{j=1}^{N_\text{vox}} I\left(x_0, x_1^i, x_2^j\right), \tag{7.10}$$

where x_1^i and x_2^i are the coordinates of the ith voxel, and $N_\text{vox} = L_\text{box}/L_\text{vox}$ is the number of voxels in each direction of $\mathbf{x}$. Here, L_box is the length of the box and L_vox is the length of a voxel.

7.2.4 Mean Expected Number of LAEs per Voxel

We define the galaxy overdensity as

$$\delta_{\rm g}(\mathbf{x}) = \frac{N_{\rm LAE}(\mathbf{x}) - \bar{N}(\mathbf{x})}{\bar{N}(\mathbf{x})}. \tag{7.11}$$

The number of LAEs detected per voxel is $N_{\rm LAE}(\mathbf{x})$ and the expected number of LAEs detected per voxel at location $\mathbf{x}$ in the absence of clustering is $\bar{N}(\mathbf{x})$. This parameter can be calculated as

$$\bar{N}(\mathbf{x}) = \delta V(\mathbf{x}) \int_{L_{\rm min}}^{L_{\rm max}} {\rm d}L \, \frac{{\rm d}n}{{\rm d}L} C\left(\frac{L}{4\pi D_L^2(z)}, f_{50}(\mathbf{x})\right), \tag{7.12}$$

where $\delta V(\mathbf{x})$ is the volume covered by spectra within the voxel at position $\mathbf{x}$, $\frac{{\rm d}n}{{\rm d}L}$ the luminosity function of LAEs, D_L the luminosity distance, and f_{50} the flux limit at which 50% of LAEs will be detected. The completeness factor, C, is a function of LAE line flux and is characterized by the 50% completeness, f_{50}. Examples of completeness curves as a function of flux are shown in Figure 28 of Gebhardt et al. (2021).

We use the completeness model v4 of the HETDEX API[2] (Gebhardt et al., 2021, Farrow et al. in preparation). This completeness function depends on the ratio of the LAE flux to the flux at 50% completeness; $C(f, f_{50}) = C_{\rm v4}(f/f_{50})$, which is a non-parametric model to map f/f_{50} to detection completeness based on inserting simulated LAEs into the data. The value of f_{50} is obtained by multiplying the flux noise, $\sigma_f(\mathbf{\Theta}, \lambda)$, at the sky coordinate $\mathbf{\Theta}$ and wavelength λ by the minimum SNR for an emission-line detection (here 5.5, as described in Section 7.2.1), a polynomial dependence on the wavelength of the detection $a_\lambda(\lambda)$, and a line-width dependent adjustment (Farrow et al. in preparation). The value of $\sigma_f(\mathbf{\Theta}, \lambda)$ is calculated as a PSF-weighted sum of the spectral errors within an aperture $3.5''$ in diameter and 14 Å deep.

We collect the values of f_{50} in each IFU spectral element. We extract the flux noise $\sigma_f(\mathbf{\Theta}, \lambda)$ on a $2'' \times 2'' \times 2$ Å grid for each IFU using the HETDEX API. The redshift bins correspond to the 2 Å grid of wavelengths between 3470 Å and 5540 Å of the spectra. Because the API also provides a mask on the grid, we obtain the volume covered per IFU spectral element, $\delta V_i(z)$, for line detections. We transform these values to $f_{50}(\mathbf{\Theta}, \lambda) = 5.5\sigma_f(\mathbf{\Theta}, \lambda)a_\lambda(\lambda)$, not accounting for the line-width adjustment. For each IFU and wavelength, we calculate a histogram of f_{50} values in 5001 logarithmically spaced bins between 10^{-18} and $10^{-13}\,{\rm erg\,s^{-1}\,cm^{-2}}$.

We calculate the number density, $\bar{n}(f_{50}, z)$, for each bin center of the f_{50} histogram and for each redshift bin corresponding to the 2 Å wavelength grid. We use the best-fit Schechter function component (Schechter, 1976) of the $z = 2.2$ Lyα luminosity function measured by Umeda et al. (2025); with $L^* = 10^{42.8}\,{\rm erg\,s^{-1}}$, $\phi^* = 10^{-3.16}\,{\rm Mpc^{-3}}$ and $\alpha = -1.53$ for $z \leq 2.3$, corresponding to the low-z bin for the power spectrum measurement. In the medium-z and high-z bins ($z > 2.3$), we adopt the $z = 3.3$ luminosity function of Umeda et al. (2025) with $L^* = 10^{42.29}\,{\rm erg\,s^{-1}}$, $\phi^* = 10^{-2.13}\,{\rm Mpc^{-3}}$ and $\alpha = -1.19$.

[2] `https://github.com/HETDEX/hetdex_api`

We evaluate equation (7.12) by translating the f_{50} histogram, $N^{(i)}_{f_{50}}(z)$, in the ith IFU into the number of expected LAEs, $\bar{N}_i(z)$, in the absence of clustering:

$$\bar{N}_i(z) = \sum_{f_{50}} \delta V_i(z) \bar{n}(f_{50}, z) N^{(i)}_{f_{50}}(z). \tag{7.13}$$

We save $\bar{N}_i(z)$, the covered volume $\delta V_i(z)$, and the mean value of f_{50},

$$\bar{f}^{(i)}_{50}(z) = \frac{\sum_{f_{50}} f_{50} N^{(i)}_{f_{50}}(z)}{\sum_{f_{50}} N^{(i)}_{f_{50}}(z)}. \tag{7.14}$$

The value of $\bar{N}_i(\lambda)$ predominantly depends on the observing conditions and the sky spectrum. It therefore varies from IFU to IFU, although it is similar for different IFUs within the same observation.

Equation (7.13) overestimates the number of detected LAEs by a factor of $\simeq 2$. This could be due to an overestimated luminosity function, especially at the faint end, or an imperfect model for f_{50}. To avoid a systematic power spectrum signal resulting from this discrepancy, we ensure that the redshift distribution of $\bar{N}(z)$, averaged over the IFUs, matches that of the detected LAEs in each field. However, we still need $\bar{N}_i(z)$ to model the IFU-to-IFU variations. We therefore multiply $\bar{N}_i(z)$ by the ratio, $f_N(z) = N_{\mathrm{LAE}}(z)/\bar{N}(z)$. We calculate this ratio separately for the Spring and Fall fields. We use $f_N(z)$ from the Fall field for the NEP field due to the low number of LAEs. The values of $f_N(z)$ are mostly around 0.5, with individual spikes up to $\simeq 3$. After this correction, 84% (95%) of the values of $\bar{n}_i(z) = \bar{N}_i(z)/\delta V_i(z)$ are below 4, 6, and 4 (7, 10, and 6) $\times 10^{-4}\mathrm{Mpc}^{-3}h^3$ in the Fall, Spring, and NEP fields, respectively.

To create a map of the LAE overdensity, we first transform the number of LAEs, N_{LAE}, and the expected $\bar{N}$ into the same format as the spectra. If the mean total spectrum at an IFU spectral element is masked, then all other values, including N_{LAE} and $\bar{N}$, are also masked at this IFU spectral element. We then calculate the total number of detected LAEs $N_{\mathrm{LAE}}(\mathbf{x})$, the expected number of detected LAEs in the absence of clustering $\bar{N}(\mathbf{x})$, in each map voxel to create a map of δ_{g}. We also save the mean f_{50} in each voxel and use it for the mocks.

7.3 Mock Data

We generate mock data to compare the measured LIM power spectra with a theoretical expectation. These mocks allow us to quantify the effect of the data cleaning, and to calculate the covariance matrix of the power spectra.

7.3.1 General Setup

To create a model for the power spectra and estimate their covariance matrices, we generate mock data using the SIMPLE code[3] (Lujan Niemeyer et al., 2023). This package

[3] `https://github.com/mlujnie/simple`

generates galaxy and intensity maps with nonlinear redshift-space distortions (RSD) due to peculiar velocities (Agrawal et al., 2017).

We use the same box size, voxel size, and redshifts as in the HETDEX maps (see Section 7.2.3). The redshifts used to calculate the intensity of the mock data vary along one axis. We generate 100 mocks for each redshift bin and each region shown in Figure 7.1, henceforth called boxes, i.e., 6900 for the Spring field, 1800 for the Fall field, and 300 for the NEP field. We input the nonlinear matter power spectrum calculated at the three mean redshifts using the public CLASS code[4] (Blas et al., 2011) with the default Halofit model parameters (Smith et al., 2003; Takahashi et al., 2012; Bird et al., 2012). We compare the mock power spectra from the nonlinear and linear matter power spectra in Appendix 7.11.

We set the LAE bias to $b_{\rm mock} = 2$, which is consistent with the measurements of $b_{\rm LAE} = 1.7^{+0.3}_{-0.4}$ at $z = 3.1$ (Gawiser et al., 2007), 1.8 ± 0.3 at $z = 2.1$ (Guaita et al., 2010), and $1.72^{+0.26}_{-0.27}$ and $2.01^{+0.26}_{-0.29}$ at $z = 2.4$ and 3.1, respectively (Herrera et al., 2025). In this setup, detected and undetected LAEs in the mock have the same bias parameter; thus, the LAE bias is $b_{\rm g} = 2$ and the intensity bias is $b_I = 2$. Our simulations use the same luminosity functions of Umeda et al. (2025) as we did for the $\bar{N}$ calculation described in Section 7.2.4, and we set the minimum Lyα luminosity to $4 \times 10^{40}\,{\rm erg\,s^{-1}}$. This minimum luminosity mainly determines the number of simulated LAEs and the mean intensity. However, a smaller value barely changes the mean intensity. We apply a $\sigma_\lambda = 2.38\,$Å LOS Gaussian smoothing algorithm to the intensity maps in order to imitate the VIRUS spectral resolution. Since the VIRUS PSF is much smaller than the voxel size, we do not smooth in the angular direction.

To determine which simulated LAEs are detected, we randomly draw from the simulated LAEs. The probability that a simulated LAE with flux f at position $\mathbf{x}$ will be detected is given by the completeness function, $C(f, f_{50}(\mathbf{x}))$, described in Section 7.2.4, where $f_{50}(\mathbf{x})$ is the mean of the values of f_{50} within the voxel.

7.3.2 Inserting the Observed Data into the Mock Data

To create the same mask for the data and mocks, we convert the mock galaxy and intensity maps to the same format as the IFU spectra. For each IFU spectral element, we select the corresponding map and voxel in the mock. We save the mock 'total,' 'detected,' and 'undetected' intensities and the number of detected galaxies in the voxel.

The volume of a voxel, $V_{\rm vox}$, is larger than the volume covered by one IFU spectral element, $V_{{\rm IFU}\times\Delta\lambda}$. Therefore, the number of galaxies that can be detected in the spectra is less than the number calculated for the voxels by a factor of $f_{\rm cov} = V_{{\rm IFU}\times\Delta\lambda}/V_{\rm vox}$. Ideally, we could account for this effect through binomial sampling with a probability of $p = f_{\rm cov}$. However, the redshift distribution of the detected mock galaxies differs from the actual LAE distribution similarly to $\bar{N}$ (see Section 7.2.4). We therefore sample from the detected mock LAEs using a binomial distribution with a probability of $p = f_{\rm cov} f_N(z)$. We use the redshift of the voxel for z.

[4] `http://class-code.net`

We imitate the sky subtraction process by subtracting the mean mock intensity spectrum of each observation from the mock intensity. We calculate the mean mock "sky" spectrum for the mock 'total,' 'detected,' and 'undetected' intensity maps separately, and subtract it from the same mock intensity map. We refer to the mock intensity spectra as $I^{\rm m}(\lambda)$ and the actual data intensity spectra as $I^{\rm d}(\lambda)$. We add the mock intensity to the data intensity ($I^{\rm m+d}_{\rm undet}(\lambda) = I^{\rm m}_{\rm undet}(\lambda) + I^{\rm d}_{\rm undet}(\lambda)$, etc.), where the data have been processed until, and including, the step described in Section 7.2.2. The intensity noise is thus naturally included in the mock spectra, $I^{\rm m+d}_{\rm undet}(\lambda)$. Then, we perform the steps detailed in Section 7.2.2 (masking outlier pixels and high-variance spectra) and Section 7.2.2 (PCA) on $I^{\rm m}(\lambda)$ and $I^{\rm m+d}(\lambda)$. Finally, we calculate the PCA weight vectors of the $I^{\rm m+d}_{\rm tot}(\lambda)$ spectra and apply the PCA cleaning to the $I^{\rm m}_{\rm undet}(\lambda)$ and $I^{\rm m+d}_{\rm undet}(\lambda)$ spectra.

7.4 Power Spectrum Measurement

7.4.1 Power Spectrum Estimator

We calculate the cross-power spectrum of the detected LAE overdensity, $\delta_{\rm g}$, and the Lyα intensity fluctuation, δI, in each box. We use a fast Fourier transform (FFT) to calculate $\widetilde{\delta I}(\mathbf{k})$ and $\widetilde{\delta}_{\rm g}(\mathbf{k})$ and compute the power spectrum,

$$\hat{P}^{(i)}(\mathbf{k}) = V_{\rm box}\widetilde{\delta}^{*}_{\rm g}(\mathbf{k})\widetilde{\delta I}(\mathbf{k}), \tag{7.15}$$

where $V_{\rm box}$ is the volume of the box, the asterisk denotes the complex conjugate, and the superscript i identifies each box, as shown in Figure 7.1.

We calculate the power spectrum monopole, $\hat{P}^{(i)}_0(k)$, by averaging $\hat{P}^{(i)}(\mathbf{k})$ over angles,

$$\hat{P}^{(i)}_0(k) = \frac{1}{4\pi}\int d^2\hat{\mathbf{k}}\ \hat{P}^{(i)}(\mathbf{k}), \tag{7.16}$$

in 23 linearly spaced k bins from $k_{\rm min} = 0.08\,h{\rm Mpc}^{-1}$ to $k_{\rm max} = 1\,h{\rm Mpc}^{-1}$ with $\Delta k = 0.04\,h{\rm Mpc}^{-1}$. We calculate $\hat{P}^{(i)}_0(k)$ for the cross-power spectrum of LAEs with the intensity of undetected sources for $N_{\rm PC} \in \{0, 10, ..., 200\}$. We also calculate $\hat{P}^{(i)}_0(k)$ of detected mock galaxies and the mock intensity of undetected sources, $\delta I^{\rm m+d}_{\rm undet}$ and $\delta I^{\rm m}_{\rm undet}$.

We do not use weights for the intensity because we find that the generated inverse-variance weights do not improve statistical significance of the power spectrum and complicate the analysis.

$\bar{N}$ naturally includes duplicate detections in repeated observations. The effective mask, $m(\mathbf{x})$, for both the galaxy and intensity maps is therefore equal to 1 in the voxels that contain IFU spectral elements and equal to zero otherwise ($m \in \{0, 1\}$). We adjust the power spectra of the boxes by the fraction of observed voxels, $f_V = \langle m\rangle = \langle m^2\rangle$, and calculate the weighted mean in each redshift bin:

$$\hat{P}_0(k) = \frac{\sum_{i=1}^{N_{\rm boxes}} \hat{P}^{(i)}_0(k) f^{-1}_{V,i} w_i(k)}{\sum_{i=1}^{N_{\rm boxes}} w_i(k)}, \tag{7.17}$$

where $w_i(k)$ are the weights defined in the next section.

7.4.2 Covariance Matrix, Weights, and Masking

We calculate the covariance matrix of $\hat{P}_0(k)$ as the sum of two matrices: the first from shuffling the data ($C_{\rm shuffle}$), and the second from the mocks ($C_{\rm mock}$). $C_{\rm shuffle}$ quantifies the statistical covariance arising from noise in the data, while $C_{\rm mock}$ quantifies cosmic variance. The total covariance matrix is given by $C = C_{\rm shuffle} + C_{\rm mock}$.

For $C_{\rm shuffle}$ and the weights, $w_i(k)$, we randomly shuffle intensities of the IFU spectral elements in each of the three fields while keeping the redshift fixed, then calculate the power spectra, $\hat{P}_0^{s,(i)}(k)$. We repeat this process 100 times, and compute weights as the inverse variance of the power spectrum monopole of each box as a function of wavenumber,

$$w_i(k) = \frac{1}{\mathrm{var}\left(\hat{P}_0^{s,(i)}(k) f_{V,i}^{-1}\right)}. \tag{7.18}$$

For each shuffling instance, we calculate the weighted mean, $\hat{P}_0^{\rm s}(k)$, of the shuffled box power spectra, $\hat{P}_0^{s,(i)}(k)$. Finally, we construct the covariance matrix of the weighted mean power spectrum of each field in each redshift bin.

For $C_{\rm mock}$, we calculate the covariance matrix using the mock-only intensity, $I^{\rm m}_{\rm undet}(z)$, of 100 mocks per box. This covariance quantifies cosmic variance of the intensity and the LAE catalog.

The power spectra of some boxes exhibit a much higher variance along the wavenumbers than the rest, indicating the presence of residual systematic errors. To remain cautious, we exclude these boxes from the analysis. We also mask one box in the Spring field due to its small number of observations. The masked boxes are marked by the hatched regions in Figure 7.1.

7.4.3 Loss of Power from PCA

Figure 7.5 displays the effect of removing PCs on the cross-power spectrum of galaxies and intensity from the mocks. The left panel shows the cross-power spectrum of mock galaxies with $I^{\rm m}_{\rm undet}(z)$ after sky subtraction and removal of $N_{\rm PC} \in \{0, 10, ..., 200\}$ PCs in the Spring field in the medium-z bin. The suppression of power at $k \lesssim 0.4\,h\mathrm{Mpc}^{-1}$ is caused by sky subtraction. The right panel shows the ratio of the sky-subtracted mock power spectrum monopole with and without PCA.

Removing PCs suppresses power, but does not completely eliminate it. Mocks are useful for capturing the effect of PCA on the signal power spectrum. We will use this information when we compare theoretical predictions with observed power spectra.

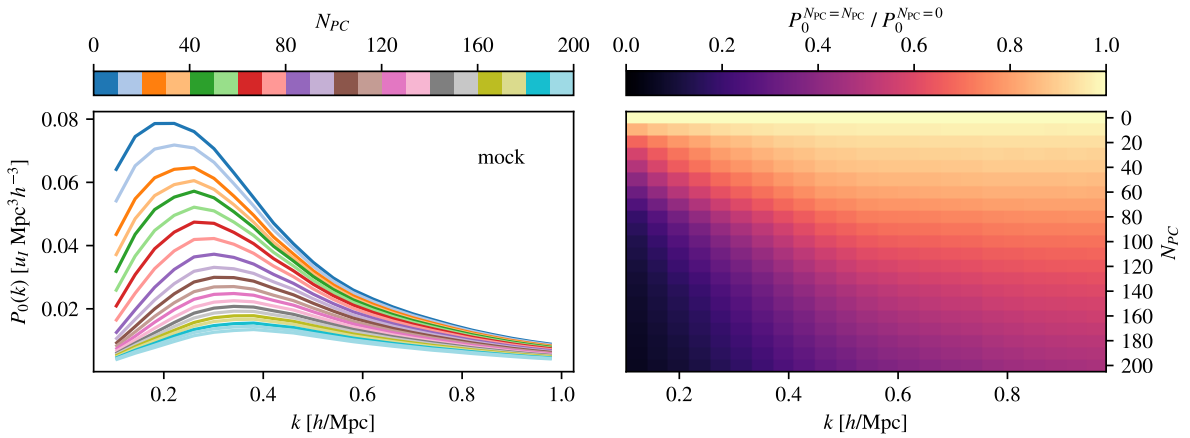

Figure 7.5: Left panel: suppression of the power spectrum monopole due to PCA cleaning. We show the mean power spectrum monopole obtained from the fiducial mocks in the Spring field in the medium-z bin. The different colors correspond to different values of $N_{\rm PC}$, as indicated in the color bar. Here, $u_I = 10^{-18}\,{\rm erg\,s^{-1}\,cm^{-2}\,arcsec^{-2}\,\AA^{-1}}$. Right panel: the ratio of the power spectrum monopole with and without $N_{\rm PC}$ PCs removed as a function of k and $N_{\rm PC}$.

7.5 Fitting the Power Spectrum Model

7.5.1 Model

We use mock power spectra to calculate theoretical predictions and compare them with observed power spectra. The mock power spectrum can be written as

$$\begin{aligned} P_{\rm mock}(\mathbf{k}) &= b_{\rm g} b_I \langle I \rangle \mathcal{T}(\mathbf{k}) \\ &\times \int \frac{d^3\mathbf{k}'}{(2\pi)^3} P_{\rm m}(\mathbf{k}') F_{\rm RSD}(\mathbf{k}') \tilde{D}(\mathbf{k}') \\ &\times \widetilde{W}_I^*(\mathbf{k}-\mathbf{k}') \widetilde{W}_g(\mathbf{k}-\mathbf{k}'), \end{aligned} \tag{7.19}$$

where $b_{\rm g}$ and b_I are the linear bias parameters of LAEs and intensity, respectively, $\langle I \rangle$ is the mean intensity of the mock, and $P_{\rm m}(\mathbf{k})$ is the matter power spectrum computed for the fiducial cosmological model.

The factor $F_{\rm RSD}(\mathbf{k})$ accounts for the effects of nonlinear RSD and a potential Lyα RT (Lujan Niemeyer, 2025). In the mock, we account for nonlinearity in the Jacobian of the coordinate transformation from real to redshift space non-perturbatively (Agrawal et al., 2017). In a linear model, it would be $F_{\rm RSD}^{\rm lin} = (1 + f\mu^2 b_{\rm g}^{-1})(1 + f\mu^2 b_I^{-1})$ (Kaiser, 1987), where $\mu = \hat{\mathbf{k}} \cdot \hat{\mathbf{k}}_\parallel$ is the cosine of the angle between the unit wavevector $\hat{\mathbf{k}}$ and the LOS $\hat{\mathbf{k}}_\parallel$.

In Appendix 7.12 we show a linear model for Lyα RT in $F_{\rm RSD}$. We do not, however, account for this effect in this paper because we do not detect the quadrupole power spectrum and the RT effect is degenerate with the bias parameters.

The factor $\tilde{D}(\mathbf{k})$ accounts for the LOS damping due to limited spectral resolution (see Section 7.3.1). The transfer function, $\mathcal{T}(\mathbf{k})$, accounts for the loss of power due to data

processing, especially sky subtraction and PCA (see the right panel of Figure 7.5 for the loss of power due to PCA). Note that we include the transfer function in our forward model and do not attempt to reconstruct a PCA-free power spectrum from the data. The window functions, $\widetilde{W}_I^*(\mathbf{k})$ and $\widetilde{W}_g(\mathbf{k})$, account for the observing mask and weights. Since the mask and weights are the same for the mocks and the data, we assume that the window functions are fully modeled by the mocks.

In our model for the monopole power spectrum, the only free parameter is the overall factor, $b_\mathrm{g} b_I \langle I \rangle \bar{F}_\mathrm{RSD}$, where $\bar{F}_\mathrm{RSD}$ is the effective nonlinear RSD factor from the integral given in Equation (7.19) and the angular averaging given in Equation (7.16).

In the absence of a mask, a linear model gives $\bar{F}^\mathrm{lin}_\mathrm{RSD} = 1 + \frac{f}{3}\left(b_\mathrm{g}^{-1} + b_I^{-1}\right) + \frac{f^2}{5} b_\mathrm{g}^{-1} b_I^{-1} \simeq 1.3$. Therefore, a slight mismatch in the values of b_g and b_I between the mock and the data only gives a minor correction to this factor.

In practice, we define the fitting parameter, $A = b_\mathrm{g} b_I \langle I \rangle \bar{F}_\mathrm{RSD} / (b_\mathrm{g}^\mathrm{fid} b_I^\mathrm{fid} \langle I^\mathrm{fid} \rangle \bar{F}^\mathrm{fid}_\mathrm{RSD})$, where $b_\mathrm{g}^\mathrm{fid}$, b_I^fid, and $\langle I^\mathrm{fid} \rangle$ are the fiducial LAE bias, the intensity bias, and the mean intensity of undetected sources assumed in the mocks. $\bar{F}^\mathrm{fid}_\mathrm{RSD}$ is the effective nonlinear RSD factor in the mocks. If the mocks were in perfect agreement with the data, we would find that $A = 1$.

Due to nonlinearity, $\bar{F}_\mathrm{RSD}$ depends on the wavenumber k. However, the ratio, $\bar{F}_\mathrm{RSD}/\bar{F}^\mathrm{fid}_\mathrm{RSD}$, depends on k only weakly if the mock and the data are in reasonable agreement, which is the case (see Section 7.6). Therefore, we neglect the scale dependence of $\bar{F}_\mathrm{RSD}/\bar{F}^\mathrm{fid}_\mathrm{RSD}$ and treat it as a constant factor. This is a good approximation given our limited ability in determining the shape of the power spectrum due to the large statistical uncertainty.

7.5.2 Parameter Inference

To determine A, we fit the fiducial mock power spectrum multiplied by an amplitude A to the data. We define $\chi^2 = \mathbf{Y}^\top \mathbf{C}^{-1} \mathbf{Y}$, where $\mathbf{Y} = \mathbf{Y}_\mathrm{data} - A\mathbf{Y}_\mathrm{fid}$ is the difference between the measured power spectrum and the model power spectrum,

$$\begin{aligned} \mathbf{Y}^\top_\mathrm{data} &= (P_0^\mathrm{data}(k_1), P_0^\mathrm{data}(k_2), \ldots, P_0^\mathrm{data}(k_{23})), \\ \mathbf{Y}^\top_\mathrm{fid} &= (P_0^\mathrm{fid}(k_1), P_0^\mathrm{fid}(k_2), \ldots, P_0^\mathrm{fid}(k_{23})), \end{aligned} \tag{7.20}$$

and $P_0^\mathrm{fid}(k)$ is the monopole of the mock power spectrum $P_\mathrm{mock}(\mathbf{k})$ (Equation (7.19)), calculated in the same 23 k bins as the data.

We also combine the power spectra of the fields and fit a single amplitude A to all of the fields within each redshift bin simultaneously. Specifically, we use

$$\begin{aligned} \mathbf{Y}^\top_\mathrm{data,all} = (&P_0^\mathrm{data,Fall}(k_1), \ldots, P_0^\mathrm{data,Fall}(k_{23}), \\ &P_0^\mathrm{data,Spring}(k_1), \ldots, P_0^\mathrm{data,Spring}(k_{23}), \\ &P_0^\mathrm{data,NEP}(k_1), \ldots, P_0^\mathrm{data,NEP}(k_{23})) \end{aligned} \tag{7.21}$$

and the equivalent vector for the model. For the covariance matrix, we combine the individual covariance matrices:

$$C_\mathrm{all} = \begin{pmatrix} C_\mathrm{Fall} & 0 & 0 \\ 0 & C_\mathrm{Spring} & 0 \\ 0 & 0 & C_\mathrm{NEP} \end{pmatrix}. \tag{7.22}$$

The best-fit amplitude that minimizes χ^2 for the individual fields is given by

$$A_{\rm best} = \frac{\mathbf{Y}_{\rm fid}^\top C^{-1} \mathbf{Y}_{\rm data}}{\mathbf{Y}_{\rm fid}^\top C^{-1} \mathbf{Y}_{\rm fid}}. \tag{7.23}$$

The best-fit amplitude that minimizes χ^2 for the combined fit is given by

$$A_{\rm best}^{\rm all} = \frac{\mathbf{Y}_{\rm fid,all}^\top C_{\rm all}^{-1} \mathbf{Y}_{\rm data,all}}{\mathbf{Y}_{\rm fid,all}^\top C_{\rm all}^{-1} \mathbf{Y}_{\rm fid,all}}. \tag{7.24}$$

7.5.3 Error Estimation and Goodness-of-fit

The uncertainty of the best-fit amplitude can be estimated as

$$\sigma_{\rm ls} = \frac{1}{\sqrt{\mathbf{Y}_{\rm fid}^\top C^{-1} \mathbf{Y}_{\rm fid}}}. \tag{7.25}$$

We calculate $A_{\rm best}$ and $\sigma_{\rm ls}$ for each field, as well as $A_{\rm best}^{\rm all}$ and $\sigma_{\rm ls}^{\rm all}$, for each $N_{\rm PC}$ and redshift bin.

To ensure that we do not underestimate the uncertainty of the measured power spectra, we empirically estimate the uncertainty, referred to as $\sigma_{\rm emp}$. This method assumes that the best-fit model is correct, but that the data contain an unaccounted-for systematic error that increases the point-to-point variance. We generate fake power spectra by randomly sampling from a Gaussian probability density function (PDF) at each wavenumber bin. The mean of the Gaussian is given by the power spectrum of the best-fit model, $\mathbf{Y}_{\rm model} = A_{\rm best}\mathbf{Y}_{\rm fid}$; the standard deviation of the point in the jth k bin is given by $\sqrt{N_{\rm RMS}^{-1}\sum_{i=j-4}^{j+4}(Y_{{\rm data},i} - Y_{{\rm model},i})^2}$, where we sum over the $N_{\rm RMS} = 9$ nearest wavenumber bins. We fit the model to the data using equations (7.23) and (7.24) and repeat this process 1000 times. The standard deviation $\sigma_{\rm emp}$ of the 1000 best-fit amplitudes is typically larger than $\sigma_{\rm ls}$ with the mean ratio $\langle\sigma_{\rm emp}/\sigma_{\rm ls}\rangle \simeq 1.6$. To remain conservative, we select the larger of the two estimated uncertainties,

$$\sigma_A = \max\{\sigma_{\rm emp}, \sigma_{\rm ls}\}. \tag{7.26}$$

To determine the goodness-of-fit, we calculate the reduced χ^2, $\chi^2_\nu = \chi^2_{\rm min}/\nu$, where $\chi^2_{\rm min}$ is the χ^2 minimum and $\nu = N_k - N_{\rm fit}$ is the number of degrees of freedom, assuming that the covariance matrix is diagonal (see Appendix 7.13 for the correlation matrices). Each individual field has $N_k = 23$ wavenumber bins per power spectrum. Since we are fitting a single parameter, $N_{\rm fit} = 1$, there are $\nu = 22$ degrees of freedom for the fit to individual fields. The combined fit has $\nu = 3\times23-1 = 68$ degrees of freedom per redshift bin.

We also calculate the probability-to-exceed (PTE, also called p-value), from the integral of a χ^2 PDF, $\mathcal{P}_\nu(\chi^2)$, from $\chi^2_{\rm min}$ of the best fit to infinity:

$${\rm PTE}(\chi^2_{\rm min}) = \int_{\chi^2_{\rm min}}^{\infty} \mathcal{P}_\nu(\hat{\chi}^2) d\hat{\chi}^2. \tag{7.27}$$

If a fit yields PTE > 0.05, the probability of obtaining a higher value of $\chi^2_{\rm min}$ than this fit is $> 5\%$ (assuming the model is correct), indicating consistency with the model.

7.6 Results

Choosing the optimal $N_{\rm PC}$ is a crucial decision in the analysis. We want $N_{\rm PC}$ to be high enough that most systematics are removed. We use goodness-of-fit as a guide: good agreement between the model and the data implies a low value of $\chi^2_\nu \simeq 1$ and a high value of PTE > 0.05. However, removal of PCs also reduces the signal. We therefore want $N_{\rm PC}$ to be small enough to retain a high SNR. Finally, we want the amplitudes inferred in the three fields to be consistent with each other at the chosen $N_{\rm PC}$.

Figure 7.6 shows χ^2_ν, the PTE, and the $SNR(A) = A_{\rm best}/\sigma_A$ as a function of $N_{\rm PC}$ for the three fields and the combined fit, for the three redshift bins. χ^2_ν decreases with increasing $N_{\rm PC}$, indicating a better agreement between the measured and model power spectra. In the low-z bin of the combined fit, χ^2_ν approaches $\simeq 1.5$ at $N_{\rm PC} \geq 100$; in the medium- and high-z bins of the combined fit, χ^2_ν is close to 1 at $N_{\rm PC} \geq 120$. The PTE generally also improves (increases) with increasing $N_{\rm PC}$. The SNR, however, decreases with $N_{\rm PC}$ above a threshold value of $N_{\rm PC} \simeq 50$.

The goodness-of-fit is acceptable in all cases except for the fit of the combined power spectrum in the low-z bin, which yields a low PTE < 0.05 at all $N_{\rm PC}$. This result could indicate the presence of systematic errors that are not quantified by the covariance matrix or an underestimation of the statistical covariance matrix of the power spectrum. Otherwise, the fit is excellent.

We also want to choose $N_{\rm PC}$ so that the inferred amplitudes in the three fields agree. Figure 7.7 displays the best-fit $b_{\rm g} b_I \langle I \rangle \bar{F}_{\rm RSD}/\bar{F}^{\rm fid}_{\rm RSD} = A_{\rm best} b^{\rm fid}_{\rm g} b^{\rm fid}_I \langle I^{\rm fid} \rangle$ and its uncertainty for the three fields and the combined fit as a function of $N_{\rm PC}$. The fiducial bias values used for the mocks are $b^{\rm fid}_I = b^{\rm fid}_{\rm g} = 2$ (Section 7.3.1). The fiducial mean intensity of undetected sources obtained from the mocks is $\langle I^{\rm fid} \rangle = 1.5$, 2.2, and 1.4×10^{-22} erg s^{-1} cm^{-2} arcsec^{-2} Å^{-1} in the redshift bins centered around $\bar{z} = 2.1$, 2.6, and 3.2, respectively.

In principle, we could choose different values of $N_{\rm PC}$ for each field and redshift bin because the PCA is performed independently in each field. However, for simplicity, we choose the same $N_{\rm PC}$ for all fields in each redshift bin. Based on the small χ^2_ν and high SNR shown in Figure 7.6, and the requirement that the best-fit amplitudes be consistent across fields, we choose $N_{\rm PC} = 170$, 130, and 100 in the low-, medium-, and high-z bins, respectively. The best-fit values of $b_{\rm g} b_I \langle I \rangle \bar{F}_{\rm RSD}/\bar{F}^{\rm fid}_{\rm RSD}$ at these $N_{\rm PC}$ are $b_{\rm g} b_I \langle I \rangle \bar{F}_{\rm RSD}/\bar{F}^{\rm fid}_{\rm RSD}$ =(6.7±3.1), (11.7±1.4), and (8.3±1.5) $\times 10^{-22}$ erg s^{-1} cm^{-2} arcsec^{-2} Å^{-1} in the low-, medium-, and high-z bins, respectively. Our main result is that we significantly detect the LAE-Lyα intensity cross-power spectrum.

Finally, Figure 7.8 shows the quality of the fits to the power spectrum monopoles from each field and redshift bin using the selected $N_{\rm PC}$. We compare the measured power spectra to the fiducial mock power spectrum and to the mock power spectrum multiplied by the best-fit amplitude. The error bars of the power spectra are given by the square root of the diagonal elements of the covariance matrix (see Section 7.4.2). The mocks accurately describe the data.

Figure 7.9 shows the weighted mean power spectrum monopole of the three fields and

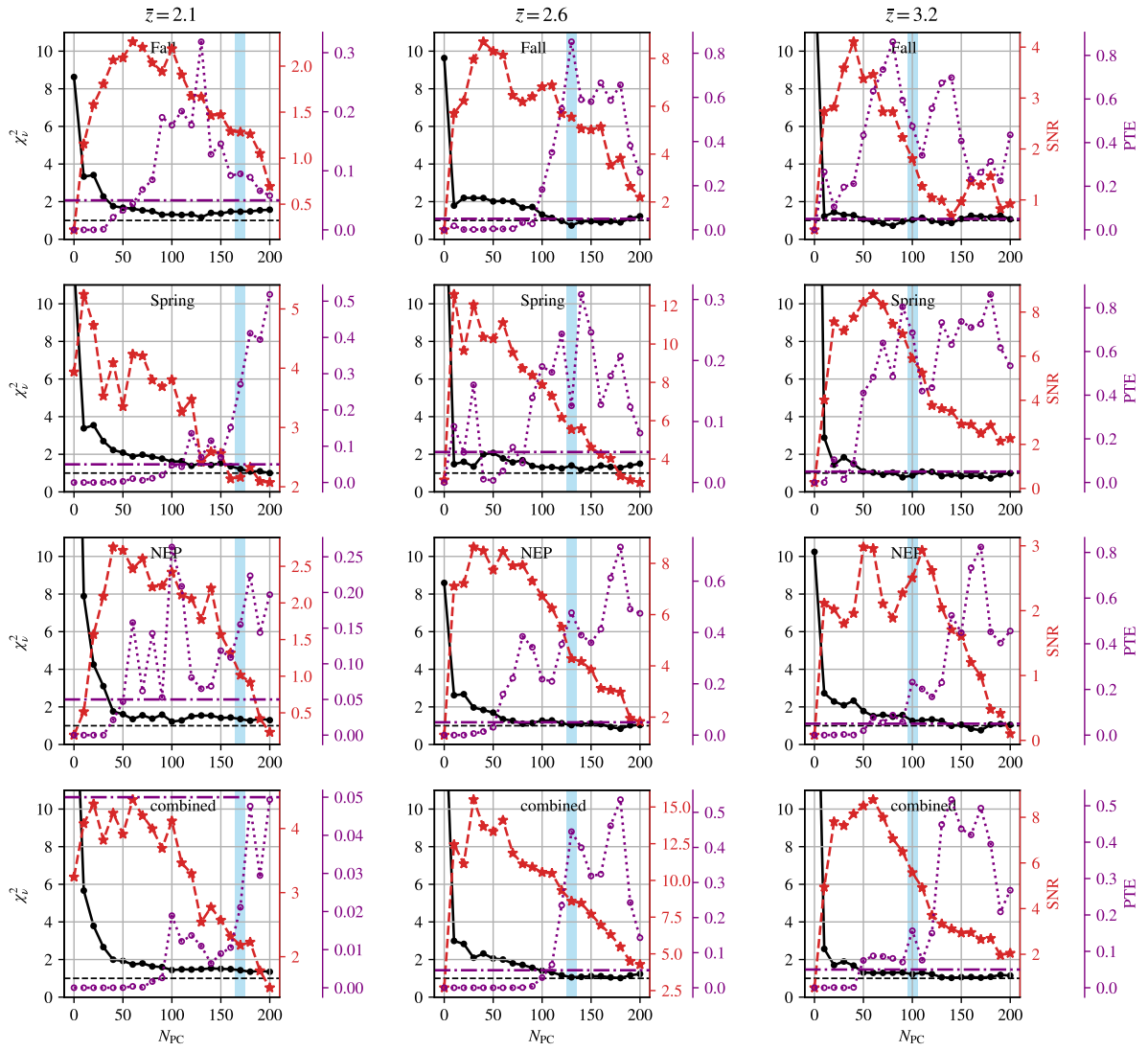

Figure 7.6: Reduced χ^2_ν (black dots), SNR (red stars), and PTE (purple empty circles) of the fit as a function of $N_{\rm PC}$ in the Fall (first row), Spring (second row), and NEP (third row) fields, and that combining the three fields (fourth row). The columns show the three redshift bins. The horizontal black dashed lines indicate $\chi^2_\nu = 1$. The horizontal purple dotted-dashed lines indicate PTE = 0.05.

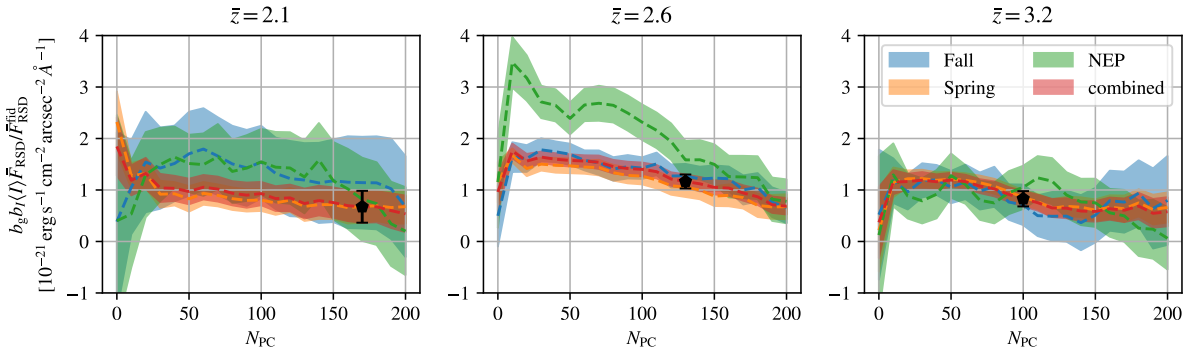

Figure 7.7: Constraints on $b_g b_I \langle I \rangle \bar{F}_{\rm RSD}/\bar{F}^{\rm fid}_{\rm RSD}$ inferred from the best-fit amplitude as a function of $N_{\rm PC}$ in the Fall (blue), Spring (orange), and NEP (green) fields, along with the combination of all three fields (red). The dashed lines show the mean, while the shaded areas show the 1σ uncertainties. The black polygons indicate the constraints from the combined fit, as determined from the best combination of χ^2_ν, SNR, and PTE values.

the three redshift bins shown in Figure 7.8. We use inverse variance weights,

$$\bar{P}_0(k_i) = \frac{\sum_{j=1}^{9} \hat{P}_0^{(j)}(k) \sigma^{-2}_{\hat{P}_0^{(j)}}(k_i)}{\sum_{j=1}^{9} \sigma^{-2}_{\hat{P}_0^{(j)}}(k_i)}, \tag{7.28}$$

where $\sigma_{\hat{P}_0^{(j)}}(k_i)$ is the square root of the diagonal element i of the covariance matrix of the power spectrum monopole $\hat{P}_0^{(j)}$, and the index j denotes a field and redshift bin. The propagated uncertainty of the weighted mean power spectrum monopole, $\bar{P}_0(k)$, is given by

$$\sigma_{\bar{P}_0}(k_i) = \frac{1}{\sqrt{\sum_{j=1}^{9} \sigma^{-2}_{\hat{P}_0^{(j)}}(k_i)}}. \tag{7.29}$$

We calculate the weighted mean of the best-fit mock power spectra using the same inverse weights of the data. We use the same $N_{\rm PC}$ as in Figure 7.8 for the data and mocks. As a test, we fit an amplitude times the weighted mean of the best-fit mock power spectrum monopole to the weighted mean power spectrum monopole. We find a best-fit amplitude of 0.97 ± 0.07, which is consistent with the expectation of 1. With $\nu = 22$ degrees of freedom, we find $\chi^2 = 21$, $\chi^2_\nu = 0.94$, and PTE $= 0.54$.

Although all quadrupole power spectra are consistent with zero, we do not use them to extract physical constraints, such as the Lyα RT effect (Lujan Niemeyer, 2025), as the shape and amplitude of the quadrupole power spectra are strongly affected by sky subtraction and PCA, and are therefore not yet stable. We leave detailed investigations of the quadrupole power spectra for future study.

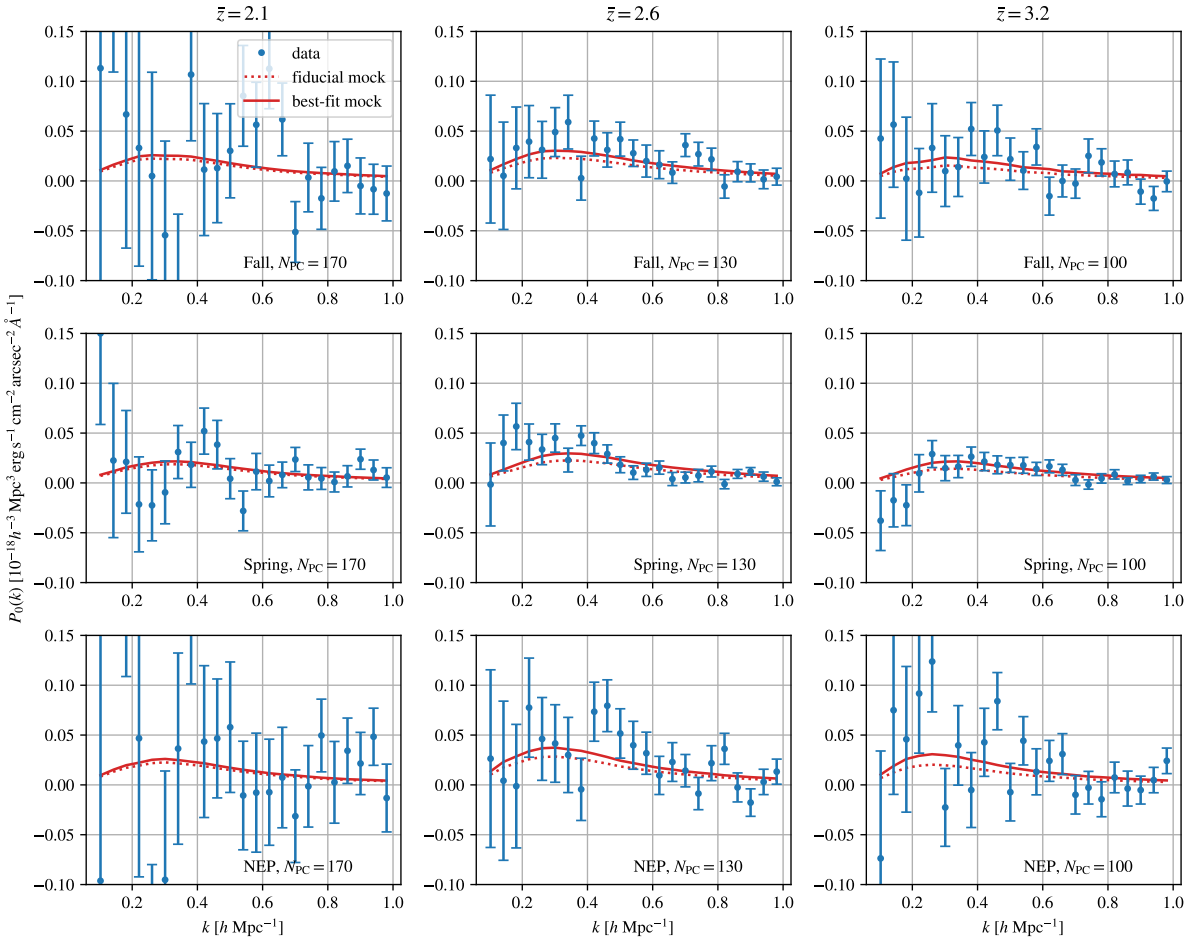

Figure 7.8: Power spectrum monopoles in the Fall (top row), Spring (middle row), and NEP (bottom row) fields and in the three redshift bins (different columns). Each panel shows the mean power spectrum monopole (Eq. (7.17)) in each field and redshift bin. We have removed $N_{\rm PC} = 170$, 130, and 100 PCs in the low-, medium-, and high-z bins, respectively. The dotted red lines indicate the fiducial mock power spectra using the same $N_{\rm PC}$. The solid red lines are the mock power spectra multiplied by the best-fit amplitude.

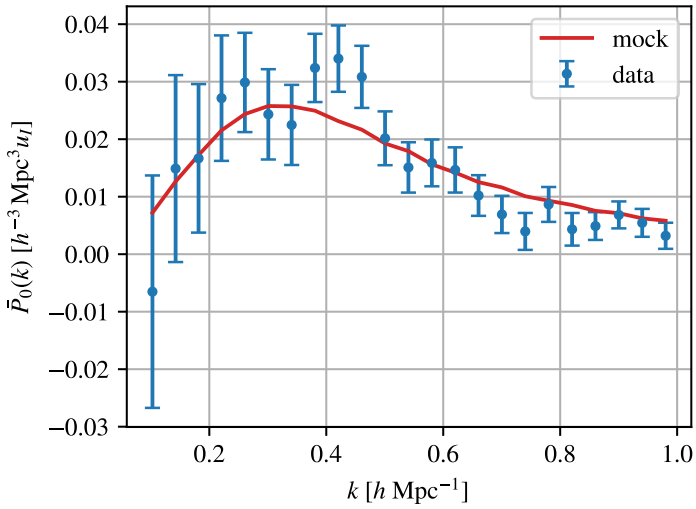

Figure 7.9: Weighted mean power spectrum monopole of the three fields and three redshift bins of the data (blue) and the best-fit mocks (red). Here, $u_I = 10^{-18}$ erg s^{-1} cm^{-2} arcsec^{-2} Å^{-1}.

7.7 Comparison to Previous Constraints

To compare our results to previous work, we convert our measurements to $b_I\langle I\rangle$ assuming a fiducial LAE bias of $b_g = 2$ (Section 7.3.1) and $\bar{F}_{\rm RSD} = \bar{F}^{\rm fid}_{\rm RSD}$.

Figure 7.10 compares our combined best-fit values of $b_I\langle I\rangle$, i.e., the product of the linear intensity bias and the intensity of only undetected sources, with those obtained from cross-correlations between QSOs and Lyα intensity (Croft et al., 2016, 2018; Lin et al., 2022), and the 95%-confidence upper limit on $b_I\langle I\rangle$ inferred from Lyα forest-Lyα intensity cross-correlation (Croft et al., 2018). These studies are all performed at redshifts from $z = 2$ to 3.5.

Since these analyses constrain the product of the intensity bias and total Lyα intensity, we also show $b_I^{\rm fid}(A_{\rm best}\langle I^{\rm fid}_{\rm undet}\rangle + \langle I^{\rm fid}_{\rm det}\rangle)$, where $A_{\rm best}$ is the best-fit amplitude, $b_I^{\rm fid} = 2$ is the fiducial intensity bias, and $\langle I^{\rm fid}_{\rm undet}\rangle$ and $\langle I^{\rm fid}_{\rm det}\rangle$ are the mean intensities of only undetected and only detected sources from the fiducial mocks, respectively. The luminosity function of the mocks does not include QSOs (see Section 7.3). These values are only slightly larger than $b_I\langle I_{\rm undet}\rangle$.

Both $b_I\langle I_{\rm undet}\rangle$ and $b_I^{\rm fid}(A_{\rm best}\langle I^{\rm fid}_{\rm undet}\rangle + \langle I^{\rm fid}_{\rm det}\rangle)$ are significantly smaller than the values inferred from QSO-Lyα intensity correlations. The low- and high-z best-fit values are consistent with the upper limit inferred from the Lyα forest-Lyα intensity cross-correlation, while the medium-z bin is 2.6σ higher.

This discrepancy could be due to the QSO-Lyα cross-correlations being dominated by small-scale physics that are affected by QSOs, such as an elevated radiation field surrounding them (e.g., Miller et al., 2021; Dong et al., 2023). Since our galaxy sample is dominated by star-forming galaxies, our measurement is less affected by QSO-related issues and may provide a more accurate representation of the Lyα radiation field in the IGM.

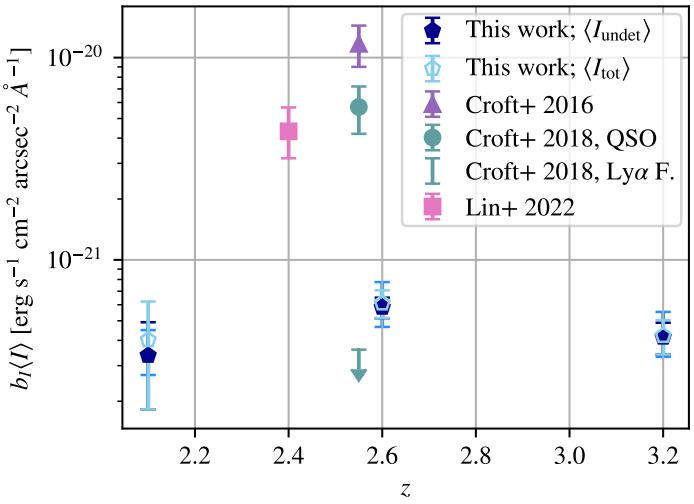

Figure 7.10: Constraints on $b_I\langle I\rangle$ obtained from this work (dark blue polygons). We assume $b_{\rm g}^{\rm fid} = 2$ and $\bar{F}_{\rm RSD} = \bar{F}_{\rm RSD}^{\rm fid}$. The medium blue error bars span the the constraints when assuming $b_{\rm g}^{\rm fid} = 2 \pm 0.5$. The light blue polygons show $b_I\langle I_{\rm tot}\rangle \simeq b_I^{\rm fid}(A_{\rm best}\langle I_{\rm undet}^{\rm fid}\rangle + \langle I_{\rm det}^{\rm fid}\rangle)$ (see text for more information). The purple triangle, olive circle, and pink square are the constraints from the QSO-Lyα intensity cross-correlations of Croft et al. (2016, 2018) and Lin et al. (2022), respectively. The upper limit from the Lyα forest-Lyα intensity cross-correlation (Croft et al., 2018) is displayed as an olive arrow.

7.8 Origins of the Lyα Emission

In this work, we masked the spectra within 10″ of the detected LAEs, which, in our redshift range, corresponds to a proper distance of $\simeq 75 - 85$ kpc. Beyond this distance, the photons contributing to the radial profiles of the simulated LAEs originate from outside the stacked LAEs' dark matter halos (Byrohl et al., 2021; Byrohl & Nelson, 2023). Therefore, the intensity should be dominated by photons unrelated to the detected LAEs. The intensity includes photons emitted within undetected galaxies, from detected LAEs that scatter many tens of kpc from the origin, and from the CGM and IGM. While we cannot directly constrain the origin and emission mechanism from our measurement, we can compare it to theoretical expectations.

We translate our constraint on $b_{\rm g} b_I \langle I\rangle \bar{F}_{\rm RSD}/\bar{F}_{\rm RSD}^{\rm fid}$ to a constraint on the comoving luminosity density

$$\rho_L = \frac{4\pi}{c}\langle I\rangle H(z)(1+z)^2\lambda_0, \tag{7.30}$$

where $\lambda_0 = 1215.67$ Å is the Lyα rest-frame wavelength. We assume $b_{\rm g}^{\rm fid} = b_I^{\rm fid} = 2$, as we do not know the bias of the detected LAEs or of the intensity. We will discuss a possible range of the intensity bias from simulation work in Section 7.8.3. In addition, we assume that $\bar{F}_{\rm RSD} = \bar{F}_{\rm RSD}^{\rm fid}$. We compare our constraint on ρ_L to theoretical expectations in Figure 7.11.

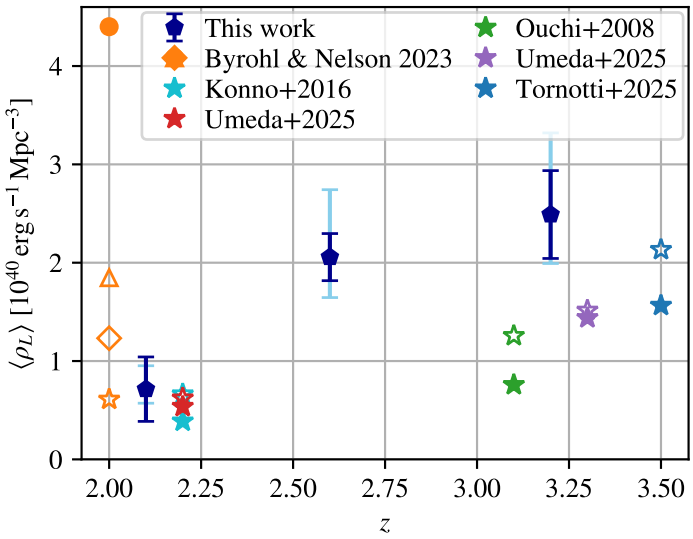

Figure 7.11: Constraints on the mean comoving luminosity density of undetected sources from this work, $\langle \rho_L \rangle$, assuming $b_{\rm g}^{\rm fid} = b_I^{\rm fid} = 2$ and $\bar{F}_{\rm RSD} = \bar{F}_{\rm RSD}^{\rm fid}$ (dark blue polygons). The dark blue error bars are the statistical uncertainty of the fit, whereas the light blue error bars indicate the best-fit values with $b_I^{\rm fid} = 2.0 \pm 0.5$ and $b_{\rm g}^{\rm fid} = 2$. The filled stars show the luminosity density obtained by integrating the luminosity functions (Konno et al., 2016; Umeda et al., 2025; Ouchi et al., 2008; Tornotti et al., 2025a) from $L_{\rm min} = 4 \times 10^{40}\,{\rm erg\,s^{-1}}$ to a typical HETDEX detection limit of $L_{\rm max} = 4 \times 10^{42}\,{\rm erg\,s^{-1}}$. The empty stars of the same colors show the luminosity density obtainted with $L_{\rm min} = 10^{36}\,{\rm erg\,s^{-1}}$, below which the integral barely changes. The orange symbols are predictions from cosmological simulations with Lyα RT (Byrohl & Nelson, 2023): the filled circle shows the total mean luminosity density; the empty orange star represents photons from galaxies with luminosity $L < 10^{41.75}\,{\rm erg\,s^{-1}}$; and the empty diamond and triangle denote photons that last scattered in the IGM and CGM, respectively.

7.8.1 Expectations from Luminosity Functions

The mock is based on our fiducial model, which only includes the intensity from faint, undetected galaxies, as expected from extrapolating the LAE luminosity function measured by Umeda et al. (2025). The mean comoving luminosity density from undetected galaxies can be calculated by integrating the luminosity function up to the HETDEX detection limit:

$$\langle \rho_L^{\rm undet} \rangle = \int_{L_{\rm min}}^{L_{\rm det}} {\rm d}L \frac{{\rm d}n}{{\rm d}L} L, \tag{7.31}$$

where $L_{\rm min}$ is given by the lowest Lyα luminosity of galaxies and $L_{\rm det}$ is the minimum luminosity detectable by HETDEX. The corresponding mean specific intensity at redshift z is then given by

$$\langle I_{\rm undet}^{\rm gal} \rangle = \frac{c \langle \rho_L^{\rm undet} \rangle}{4\pi H(z)(1+z)^2 \lambda_0}. \tag{7.32}$$

We integrate Equation (7.31) from $L_{\rm min} = 4 \times 10^{40}\,{\rm erg\,s^{-1}}$ to a typical HETDEX detection limit, $L_{\rm max} = 4 \times 10^{42}\,{\rm erg\,s^{-1}}$ (see Figure 7.2). To illustrate the uncertainty due to the unknown minimum luminosity, we also show the result for $L_{\rm min} = 10^{36}\,{\rm erg\,s^{-1}}$. Using a smaller value for $L_{\rm min}$ barely changes the integral. Figure 7.11 shows $\langle \rho_L^{\rm undet} \rangle$ obtained by integrating various luminosity functions from the literature (Ouchi et al., 2008; Konno et al., 2016; Umeda et al., 2025; Tornotti et al., 2025a). The luminosity densities of the different luminosity functions are similar to each other and slightly lower than that of our measurement. Note that the luminosity functions of Ouchi et al. (2008); Konno et al. (2016); Umeda et al. (2025) are obtained from narrowband surveys, which have different LAE selection criteria from spectroscopic surveys like HETDEX or Tornotti et al. (2025a). The value of ρ_L from integrating luminosity functions is also strongly dependent on the faint-end slope α of the luminosity function, which is degenerate with $L_\star$ and $\phi_\star$ (see Figure 3 of Tornotti et al., 2025a).

7.8.2 Star Formation

The comoving star-formation rate density at $z = 2-3.5$ is $\simeq 0.1\,M_\odot\,{\rm yr^{-1}\,Mpc^{-3}}$ (Rowan-Robinson et al., 2016). We use the conversion of the star-formation rate, $\dot{M}_\star$, to the intrinsic Lyα luminosity (Dijkstra, 2019)

$$L_{{\rm Ly}\alpha} \simeq 10^{42}\,{\rm erg\,s^{-1}} \left(\frac{\dot{M}_\star}{M_\odot\,{\rm yr^{-1}}} \right). \tag{7.33}$$

Note that this conversion factor strongly depends on the initial stellar mass function and the assumption of case-B recombination and can deviate by factors of $2-3$ with other assumptions (Raiter et al., 2010).

Using Equation (7.33), we obtain a comoving intrinsic luminosity density of $\rho_L \simeq 10^{41}\,{\rm erg\,s^{-1}\,Mpc^{-3}}$. The corresponding observed intensity at redshift $z = 2.1\,(2.6/3.2)$ is $\langle I_{\rm SF} \rangle \simeq f_{\rm esc} \times 2.4\,(1.4/0.8) \times 10^{-21}\,{\rm erg\,s^{-1}\,cm^{-2}\,arcsec^{-2}\,\AA^{-1}}$, where $f_{\rm esc}$ is the mean escape fraction of Lyα photons from star-forming regions.

Therefore, star formation can power the mean intensity inferred from our measurements if $b_I \langle I \rangle f_{\rm esc} / \langle I_{SF} \rangle = b_I f_{\rm esc} \simeq 0.1$, 0.4, and 0.5 in the low-, medium-, and high-z bins,

respectively, assuming $b_g = 2$ and $\bar{F}_{\rm RSD} = \bar{F}^{\rm fid}_{\rm RSD}$. These numbers are reasonable: much higher ($b_I f_{\rm esc} \gg 1$) and lower ($b_I f_{\rm esc} \ll 0.1$) values would not support the interpretation of Lyα emission powered by star formation.

7.8.3 Comparison to an RT Simulation

Byrohl & Nelson (2023) post-processed a cosmological hydrodynamic simulation in a cosmological volume at $z = 2$ with Lyα RT. The authors calibrated the escape fraction of Lyα emission from galaxies to an observed luminosity function, painted Lyα emission from recombination and collisional excitation into cells without star formation, and simulated the scattering process of the photons. In the simulation, 48% of Lyα photons are produced by collisional excitation, making it the dominant emission mechanism. Star formation produces 28% and recombination outside the interstellar medium (ISM) contributes 23% of the photons. The vast majority, 98%, of Lyα photons are produced within halos, with more than half produced within the CGM. Only 2% are produced in the IGM.

Figure 7.11 compares our constraints on the luminosity density with the total luminosity density of the simulation of Byrohl & Nelson (2023) and its contributions from photons that last scattered in faint galaxies, in the CGM, and in the IGM.

The total mean luminosity density of the Byrohl & Nelson (2023) simulation is $4.4 \times 10^{40}\,{\rm erg\,s^{-1}\,Mpc^{-3}}$ (filled circle). After scattering, 30% of the luminosity density originates from galaxies within an $1.5''$ aperture, 42% of photons last scatter in the CGM (empty triangle), and 28% of Lyα photons reach the observer from the IGM (empty diamond).

Of the photons that reach us from galaxies, $6.1 \times 10^{39}\,{\rm erg\,s^{-1}\,Mpc^{-3}}$ (empty star) are due to galaxies with a Lyα luminosity of $L < 10^{41.75}\,{\rm erg\,s^{-1}}$, which is fainter than the typical HETDEX detection limit. Given this luminosity density, the photons that reach us from faint galaxies, the CGM, and the IGM would produce an intensity of $9\,(5/3) \times 10^{-22}\,{\rm erg\,s^{-1}\,cm^{-2}\,arcsec^{-2}\,\AA^{-1}}$ at redshift $z = 2.1(2.6/3.2)$. Assuming that this value is the mean intensity of undetected sources, the intensity bias necessary for our measurement is $b_I \simeq 0.4\,(1.1/1.3)$. Note, however, that the approximation that $\bar{F}_{\rm RSD} = \bar{F}^{\rm fid}_{\rm RSD}$ breaks for values of b_I that differ significantly from $b_I^{\rm fid} = 2$.

We masked spectra containing continuum emission in our measurement, which may include galaxies with nonzero Lyα emission. When we include only emission that last scattered in the CGM or IGM, the necessary intensity bias to match the simulation results is $b_I \simeq 0.5$, 1.3, and 1.6 at $z = 2.1$, 2.6, and 3.2, respectively.

We can approximate the intensity bias b_I from the relation between the luminosity density as a function of dark-matter halo mass. As shown in Figure 3 in Byrohl & Nelson (2023), their simulation predicts a tight relationship between the Lyα luminosity and halo mass. From this relation, the intensity bias b_I can be estimated as

$$b_I = \frac{\int_{M_{\rm h,min}}^{M_{\rm h,max}} dM_{\rm h}\, Ldn/dM_{\rm h}\, b(M_{\rm h})}{\int_{M_{\rm h,min}}^{M_{\rm h,max}} dM_{\rm h}\, Ldn/dM_{\rm h}}, \tag{7.34}$$

where $M_{\rm h}$ is the halo mass and we adopt a fitting formula from Tinker et al. (2010) to compute the halo bias, $b(M_{\rm h})$. We obtain $b_I \simeq 1.4$ for $(\log_{10}(M_{\rm h,min}/M_\odot), \log_{10}(M_{\rm h,max}/M_\odot)) =$

$(9, 12)$ and $b_I \simeq 1.6$ for $(\log_{10}(M_{\rm h,min}/M_\odot), \log_{10}(M_{\rm h,max}/M_\odot)) = (9, \infty)$. This is a consequence of the intensity bias being weighted more toward small-mass halos given the relation in the simulation, which can be compared with $b_I \simeq 3$ assumed in Croft et al. (2016). Given the 10-20% accuracy of the fitting formula in Tinker et al. (2010), it is reasonable to estimate that $1.5 \lesssim b_I \lesssim 2$.

7.9 Discussion

7.9.1 Possible Data Processing Improvements

This work has demonstrated the feasibility of detecting the LAE-Lyα intensity cross-power spectrum in HETDEX data and provides a constraint on $b_{\rm g} b_I \langle I \rangle \bar{F}_{\rm RSD}/\bar{F}^{\rm fid}_{\rm RSD}$. We have used extensive masking and PCA to remove systematics from our results. This resulted in significant signal loss. This subsection describes three ideas for improving the HETDEX LIM measurement in the future.

- Implementing a better full-frame sky subtraction technique would reduce systematics. This includes identifying low-level systematic contributions associated with individual fiber spectra or entire IFUs. With these improvements, we would not have to mask vast regions of the observed HETDEX volume or apply the PCA.
- Using more robust weights for the intensity and galaxy catalogs could improve the statistical significance of the power spectrum detection.
- The completeness model and luminosity function used in this work could be updated to better match the redshift distribution of LAEs.

7.9.2 Possible Modeling Improvements

Within the statistical uncertainty, the measured LAE-Lyα intensity cross-power spectra are consistent with our mock based on a lognormal model (Agrawal et al., 2017; Lujan Niemeyer et al., 2023). This allows us to constrain $b_{\rm g} b_I \langle I_{\rm undet} \rangle \bar{F}_{\rm RSD}/\bar{F}^{\rm fid}_{\rm RSD}$. Fast lognormal mocks are therefore essential because they allow us to study many systematic effects in the data and enable us to faithfully forward-model the data.

However, future higher-fidelity measurements with lower statistical uncertainty will require more detailed models with better treatment of the relevant physics. Although cosmological simulations with Lyα RT are still computationally expensive, they are essential for a more accurate modeling of physics (see, e.g., Behrens et al., 2018; Byrohl et al., 2019; Gurung-López et al., 2021; Byrohl & Nelson, 2023; Khoraminezhad et al., 2025).

While our current model includes RSD, the LAE luminosity function, and the HETDEX sensitivity and masking function, it is otherwise simple. For example, all galaxies have the same bias, luminosity function, and emission line width, regardless of their environment. The model only includes Lyα emission sourced by galaxies expected from extrapolating the luminosity function of Umeda et al. (2025) to faint luminosities. Future models could

include diffuse Lyα emission of the CGM and IGM, and the scattering of Lyα photons into or out of the LOS. These emissions source Lyα halos around galaxies, which have been found ubiquitously at $z > 2$ (e.g., Steidel et al., 2011; Wisotzki et al., 2018; Lujan Niemeyer et al., 2022a,b). They could also model the Lyα absorption around LAEs, which we discuss further in Section 7.9.4.

7.9.3 Contamination from [O II]-emitting Galaxies

Despite thoroughly masking spectra with continuum emission and detected low-redshift sources, some interloper contamination may remain. If present in both the galaxy catalog and the intensity map, the signal of these interlopers would contaminate the cross-power spectrum of the detected LAEs and the Lyα intensity measurement.

However, we believe that any contamination in the HETDEX LAE catalog is likely to be small. Davis et al. (2023a) estimated that [O II]-emitting galaxies make up only $\simeq 1.2\% \pm 0.1\%$ of the HETDEX LAE catalog, and contamination from other sources is only $\simeq 0.8\% \pm 0.1\%$.

7.9.4 Lyα Absorption around LAEs

By stacking the spectra of tens of thousands of LAEs detected in HETDEX, Davis et al. (2023b) and Weiss et al. (2024) found broad, negative absorption troughs around the stacked Lyα emission lines.

Weiss et al. (2025) argued that these troughs are due to the extragalactic background light (EBL) that is absorbed in the CGM surrounding the LAEs. If HETDEX detects LAEs preferentially on the near side of overdensities – for example, because less Lyα emission is absorbed on the way to Earth – then the EBL is enhanced behind the LAEs and absorbed around the Lyα wavelength. The sky subtraction removes the mean EBL spectrum from all spectra, including those where the EBL is completely absorbed. Consequently, the absorption troughs become negative, explaining the data.

Khanlari et al. (2025) stacked the spectra as a function of the angular distance from the LAEs and find Lyα absorption out to $\simeq 350\,\mathrm{kpc}$ (proper). The Lyα absorption at LAE positions and in their surroundings is stronger for sources with a smaller detection significance, and is not observed for LAEs with SNR > 6.

The expected effect of this absorption on our measurement is twofold. First, the absorption around LAEs can be modeled as part of the emission line profile. This profile can be written as $\phi(\lambda) = \phi_{\mathrm{em}}(\lambda) + \phi_{\mathrm{abs}}(\lambda)$, where $\phi_{\mathrm{em}}(\lambda)$ is the shape of the emission line, and $\phi_{\mathrm{abs}}(\lambda)$ is the shape of the absorption feature (e.g., a Voigt profile). One can then model an intensity map from delta functions at the locations of point sources and convolve this map with $\phi(\lambda)$ (see, e.g., Lujan Niemeyer et al., 2023). This convolution dampens the power spectrum, $P(\mathbf{k})$, along the LOS. In this work, we have neglected $\phi_{\mathrm{abs}}(\lambda)$ and modeled $\phi(\lambda) \simeq \phi_{\mathrm{em}}(\lambda)$, including the LOS resolution of HETDEX. We have also not included continuum emission or the EBL in the model. The measured cross-power spectrum of detected LAEs with undetected Lyα intensity is consistent with the model without the absorption feature, suggesting that Lyα absorption found in Davis et al. (2023b) and Weiss et al. (2024) does not dominate the cross-correlation signal.

Second, we only considered sources with SNR > 5.5 as detected; thus, galaxies with lower detection significance and undetected galaxies contribute to the Lyα intensity map. If the intensity map were dominated by a continuum background that is absorbed in overdensities, the cross-correlation between the detected LAEs and the Lyα intensity map excluding detected LAEs would be negative. However, since we detect a positive signal, we conclude that Lyα emission dominates the intensity map.

More sophisticated modeling of and exploration of the absorption feature will be left for future work.

7.10 Summary and Conclusions

In this paper, we presented a Lyα LIM power spectrum measurement in HETDEX. We reported the detection of the LAE-Lyα intensity cross-power spectrum of LAEs with SNR > 5.5 and the intensity of undetected sources in three redshift bins centered around $\bar{z} = 2.1$, 2.6, and 3.2.

To accomplish this measurement, we thoroughly cleaned the spectral data to remove systematic contributions. We created self-consistent lognormal mocks for the LAEs and the Lyα intensity, including RSD, using the HETDEX window function, sensitivity, and noise (Lujan Niemeyer et al., 2023; Lujan Niemeyer, 2025). We estimated the covariance matrix of the power spectra from the mocks and from the shuffled data.

By fitting the fiducial mock prediction times the overall amplitude to the measured power spectra, we constrained the product of the detected LAE bias, the intensity bias, the mean intensity of undetected sources, and the ratio of the real and fiducial RSD factors, $b_{\rm g} b_I \langle I \rangle \bar{F}_{\rm RSD} / \bar{F}^{\rm fid}_{\rm RSD}$.

Assuming a fiducial LAE bias $b^{\rm fid}_{\rm g} = 2$ and $\bar{F}_{\rm RSD} = \bar{F}^{\rm fid}_{\rm RSD}$, we inferred lower values of $b_I \langle I \rangle$ than those from the QSO-Lyα cross-correlations (Croft et al., 2016, 2018; Lin et al., 2022). Our constraint, however, is slightly higher than the upper limit inferred from the Lyα forest-Lyα intensity cross-correlation measurement (Croft et al., 2018).

Assuming fiducial LAE and intensity biases of $b^{\rm fid}_{\rm g} = 2$ and $b^{\rm fid}_I = 2$, and $\bar{F}_{\rm RSD} = \bar{F}^{\rm fid}_{\rm RSD}$, our constraints on the Lyα luminosity density are slightly larger than those obtained by integrating extrapolated LAE luminosity functions. Our constraints are on the same order of magnitude as the intensity of faint galaxies and of the CGM and IGM emission in the cosmological simulations with Lyα RT at $z = 2$ (Byrohl & Nelson, 2023). Our measured intensity is smaller than the total Lyα intensity expected from the star formation rate density at $z \simeq 2 - 3$, and is consistent with an escape fraction of Lyα photons from the ISM being $f_{\rm esc} < 1$. These results will be useful for constraining models of galaxy formation and evolution.

This work represents initial LIM results from the HETDEX data. We will improve the data processing and modeling of the signal in the future.

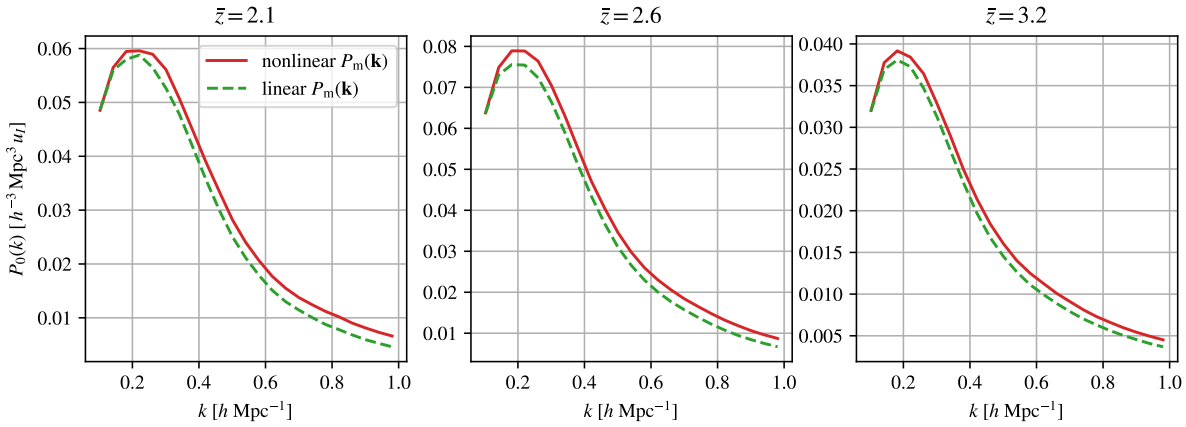

Figure 7.12: The red solid lines show the fiducial nonlinear mock cross-power spectra, while the green dashed lines show the linear mock cross-power spectra. These two power spectra are compared in the Spring field across three redshift bins. Here, $u_I = 10^{-18}\,\mathrm{erg\,s^{-1}\,cm^{-2}\,arcsec^{-2}\,\AA^{-1}}$.

7.11 Appendix: Effect of Nonlinearity on the Power Spectrum

To see the effects of nonlinear corrections to the matter power spectrum, we generate a new set of mocks with a linear input matter power spectrum generated by CLASS, i.e., not using Halofit. Otherwise, we keep the same settings as the fiducial mocks described in Section 7.3. In both cases, nonlinearity in RSD due to the Jacobian of the coordinate transformation from real space to redshift space is included (Agrawal et al., 2017). We also process the mock maps in the same way, including sky subtraction.

Figure 7.12 shows the LAE-Lyα intensity cross-power spectrum of the linear mock compared to that of the fiducial nonlinear mock in the Spring field. The power spectra are similar even at $k > 0.5\,h\,\mathrm{Mpc}^{-1}$ across all explored redshift bins from $z = 1.88$ to 3.52, as nonlinearity due to structure formation is mild at such high redshift (Jeong & Komatsu, 2006).

7.12 Appendix: F_{RSD} Including the Lyα RT Effect

When Lyα photons travel from a galaxy to an observer on Earth, they can be scattered out of the LOS by hydrogen atoms in the IGM, which can be modeled as an effective absorption. This reduces the observed Lyα luminosity of a galaxy, making it less likely to be detected, and the observed Lyα intensity in the intensity map. This effect depends on the density and velocity gradient of the gas between the observed galaxy and us. If a significant fraction of Lyα photons are subject to scattering in the IGM, this can affect the observed clustering statistics of LAEs and Lyα LIM (e.g., Zheng et al., 2011; Behrens & Niemeyer, 2013; Behrens et al., 2018). Specifically, it can change the anisotropy of the power spectrum in redshift space. In a linear model for RSD and Lyα RT following

Wyithe & Dijkstra (2011) and Lujan Niemeyer (2025), the RSD factor reads

$$F_{\rm RSD}(\mathbf{k}) = \left[b_I + b_{\rm ion}K_\lambda(k)C - c_\gamma C + (1-C)f\mu^2\right]\left[b_{\rm g} + b_{\rm ion}K_\lambda(k)C^{{\rm g}_\alpha} - c_\gamma C^{{\rm g}_\alpha} + (1-C^{{\rm g}_\alpha})\,f\mu^2\right], \tag{7.35}$$

where b_I and b_g are the intensity and LAE bias, respectively, $b_{\rm ion}$ is the bias of ionizing sources, and $c_\gamma \simeq 1.72$. The smoothing kernel $K_\lambda(k) = \arctan(k\lambda_{\rm mfp})/(k\lambda_{\rm mfp})$, with the mean free path of ionizing photons $\lambda_{\rm mfp}$, translates the locations of ionizing sources to a map of ionization rate fluctuations. The constants C and $C^{{\rm g}_\alpha}$ quantify the effect of the Lyα absorption in the IGM on the intensity fluctuation and the detected galaxy overdensity, respectively:

$$C = \frac{F_{\rm abs}\tau_0 e^{-\tau_0}}{1 - F_{\rm abs} + F_{\rm abs}e^{-\tau_0}}; \tag{7.36}$$

$$C^{{\rm g}_\alpha} = (\beta_\phi - 1)\frac{F_{\rm abs}\tau_0 e^{-\tau_0}}{1 - F_{\rm abs} + F_{\rm abs}e^{-\tau_0}}. \tag{7.37}$$

Here, $F_{\rm abs} \in [0,1]$ is the fraction of Lyα photons that can be absorbed in the IGM. The fraction $1 - F_{\rm abs}$ travels to the observer unobstructed, for example, because they have redshifted out of resonance before escaping the galaxy. τ_0 is the mean effective optical depth of the IGM; $\beta_\phi = -\alpha$ is -1 times the faint-end slope of the Lyα luminosity function. Note that the linear model for Lyα RT is only valid for small overdensities and breaks down in the nonlinear environment around galaxies.

7.13 Appendix: Correlation Matrices

We calculate the correlation matrices M_{ij} of the power spectrum monopoles shown in Figure 7.8 as $M_{ij} = C_{ij}/\sqrt{C_{ii}C_{jj}}$, where C_{ij} is the total covariance matrix described in Section 7.4.2. Figure 7.13 shows the correlation matrices. They are mostly diagonal.

Acknowledgements

HETDEX is led by the University of Texas at Austin McDonald Observatory and Department of Astronomy with participation from the Ludwig-Maximilians-Universität München, Max-Planck-Institut für Extraterrestrische Physik (MPE), Leibniz-Institut für Astrophysik Potsdam (AIP), Texas A&M University, Pennsylvania State University, Institut für Astrophysik Göttingen, The University of Oxford, Max-Planck-Institut für Astrophysik (MPA), The University of Tokyo and Missouri University of Science and Technology. In addition to Institutional support, HETDEX is funded by the National Science Foundation (grant AST-0926815), the State of Texas, the US Air Force (AFRL FA9451-04-2- 0355), and generous support from private individuals and foundations.

The observations were obtained with the Hobby-Eberly Telescope (HET), which is a joint project of the University of Texas at Austin, the Pennsylvania State University, Ludwig-Maximilians-Universität München, and Georg-August-Universität Göttingen. The HET is named in honor of its principal benefactors, William P. Hobby and Robert E. Eberly.

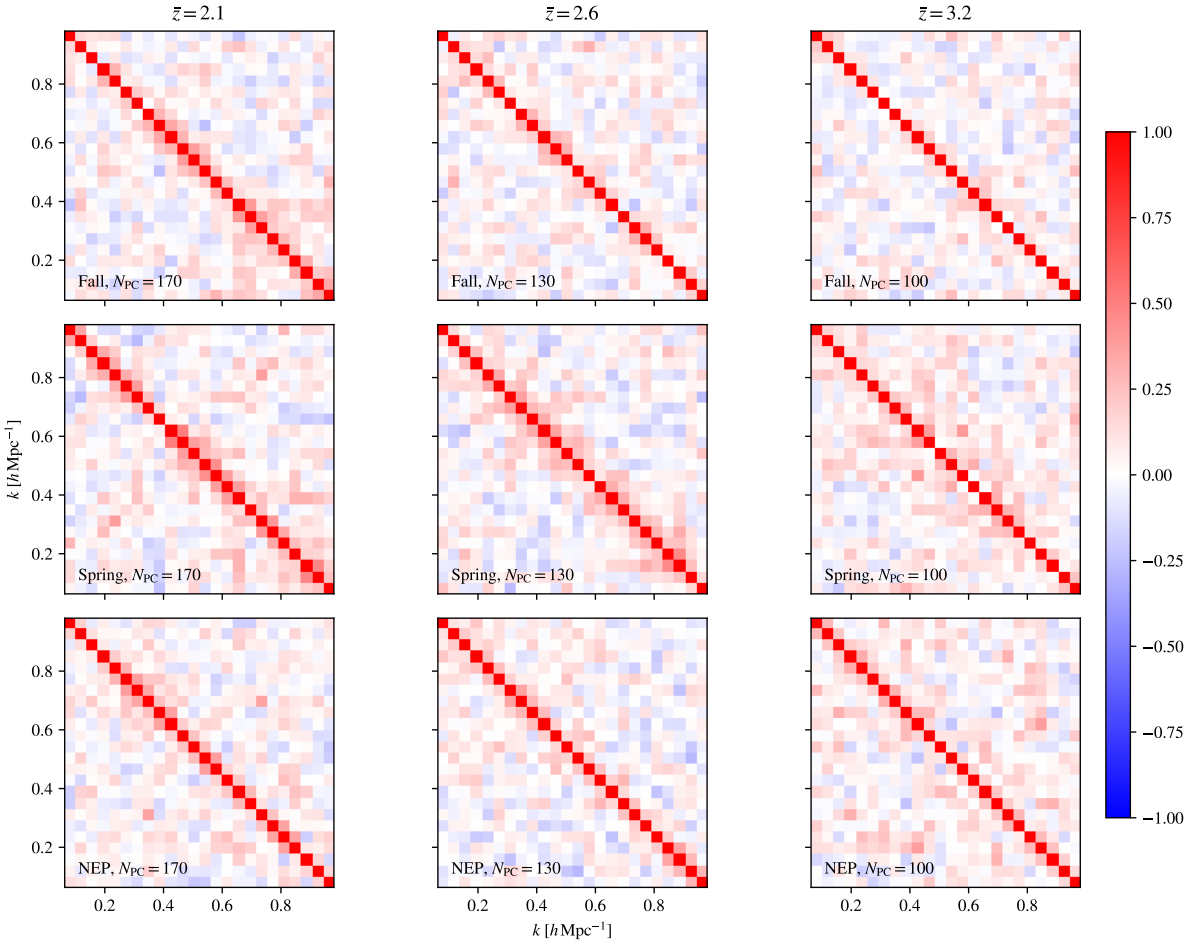

Figure 7.13: Correlation matrices of the power spectrum monopoles of Figure 7.8, in the same order.

VIRUS is a joint project of the University of Texas at Austin, Leibniz-Institut für Astrophysik Potsdam (AIP), Texas A&M University (TAMU), Max-Planck-Institut für Extraterrestrische Physik (MPE), Ludwig-Maximilians-Universität Muenchen, Pennsylvania State University, Institut für Astrophysik Göttingen, University of Oxford, and the Max-Planck-Institut für Astrophysik (MPA). In addition to Institutional support, VIRUS was partially funded by the National Science Foundation, the State of Texas, and generous support from private individuals and foundations.

The authors acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing high performance computing, visualization, and storage resources that have contributed to the research results reported within this paper. URL: `http://www.tacc.utexas.edu`.

This work was supported in part by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC-2094—390783311.

J.L.B. acknowledges funding from the Ramón y Cajal Grant RYC2021-033191-I, financed by MCIN/AEI/10.13039/501100011033 and by the European Union "NextGenerationEU"/PRTR, as well as the project UC-LIME (PID2022-140670NA-I00), financed by MCIN/AEI/10.13039/501100011033/FEDER, UE.

MJJ acknowledges the support of a UKRI Frontiers Research Grant [EP/X026639/1], which was selected by the European Research Council, and the STFC consolidated grants [ST/S000488/1] and [ST/W000903/1]. MJJ also acknowledges support from the Oxford Hintze Centre for Astrophysical Surveys which is funded through generous support from the Hintze Family Charitable Foundation.

The Institute for Gravitation and the Cosmos is supported by the Eberly College of Science and the Office of the Senior Vice President for Research at the Pennsylvania State University.

R.C. acknowledges support from the National Science Foundation under grant AST-2408358. S.S. acknowledges support from the National Science Foundation under grants NSF-2219212 and NSF-2511145.

We acknowledge the use of the astropy (Astropy Collaboration et al., 2013, 2018), dustmaps (Green, 2018), extinction `https://extinction.readthedocs.io/en/latest/`, matplotlib (Hunter, 2007), numpy (Harris et al., 2020), and scipy (Virtanen et al., 2020) software packages.

Chapter 8

Conclusions and Outlook

8.1 Summary

Lyα LIM is a versatile tool for observing the Universe at intermediate to high redshifts. We conducted small-scale Lyα LIM, i.e., stacking, to study the CGM of star-forming galaxies (Chapter 4), explored the possibility of large-scale Lyα LIM with HETDEX and created a fast mock simulation code for any LIM survey (Chapter 5), investigated the effect of Lyα RT on Lyα LIM measurements (Chapter 6), and finally measured and detected the Lyα LIM power spectrum in HETDEX data (Chapter 7).

In Chapter 4, we tested the hypothesis that Lyα halos around star-forming galaxies at large distances from the galaxies ($\gtrsim$ 60 proper kpc) are dominated by photons that originate from outside the target galaxies' dark matter halos and scatter in their CGM. This hypothesis implies that the surface brightness of the outer Lyα halos is independent of the Lyα luminosity of the central galaxy. To test this hypothesis, we stacked the Lyα emission around Lyα-faint galaxies in HETDEX, which were detected through their [O III] emission line in the 3D-HST survey. Indeed, while the central surface brightness in the core ($r < 10$ proper kpc) was found to be fainter than that of the stacked LAEs, i.e., Lyα-bright galaxies, in Lujan Niemeyer et al. (2022a) by an order of magnitude, the outer halos ($r > 40$ proper kpc) reached the same surface brightness as that of the LAEs. We found that these photons either originate from other dark matter halos or satellite galaxies, or are emitted through fluorescence or cooling radiation in the CGM. While we could not determine the dominant Lyα emission origin with this measurement, star formation alone can account for the integrated Lyα emission out to 800 proper kpc, and fluorescence from the UV background is sufficient to explain the surface brightness at intermediate distances.

The analysis in Chapter 4 and Lujan Niemeyer et al. (2022a) showed that the data quality of HETDEX is sufficient for low-surface brightness measurements, and opened the door for Lyα LIM. However, we required mock data to calculate the expected signal and covariance matrices for the LIM power spectrum measurement in HETDEX.

In Chapter 5, we developed a fast simulation code, called SIMPLE, to self-consistently generate galaxy catalogs and line intensity maps in redshift space. It is based on the assumption that the underlying mass distribution is lognormal and follows an input power

spectrum, and assigns random luminosities to the galaxies following an input line luminosity function. SIMPLE then calculates the line intensity maps and detected galaxy catalog. One can apply angular and LOS smoothing, a survey mask, and sky subtraction and finally calculate the LIM cross- and auto-power spectra. After validating the output of SIMPLE with an analytic model to percent-level precision, we created mock data for a HETDEX-like survey. Using our fiducial model, we forecasted that HETDEX would be able to detect the cross-power spectrum monopole of detected LAEs with the Lyα intensity of undetected galaxies with a signal-to-noise ratio of eight in two redshift bins.

The calculation in Chapter 5 did not include any effects of Lyα RT. Zheng et al. (2011) predict that the scattering of Lyα photons out of the LOS in the IGM introduces an anisotropic selection effect of LAEs due to the dependence of the scattering probability on the large-scale density and velocity distributions. In Chapter 6, we modeled this effect on Lyα LIM power spectra using two methods. We first derived a linear analytic model based on the approximation that the relevant matter overdensities and velocity fluctuations are small. We then implemented a simple model for IGM absorption in the SIMPLE code. We also modeled the effect of line broadening and spectral line shifts due to the scattering of Lyα photons before escaping the ISM, and revealed a new effect that arises in cross-correlations of lines with different spectral line shifts. We found that the linear model is insufficient to describe the Lyα RT effect on the LIM power spectra because most absorption takes place in highly nonlinear, overdense regions. We predicted that the LAE auto-power spectrum measured by HETDEX will be sensitive to this Lyα RT effect, while the LAE-Lyα intensity cross-power spectrum will be too noisy to be affected.

Lastly, in Chapter 7, we measured the cross-power spectrum of detected LAEs in HETDEX with the Lyα intensity of fibers excluding detected LAEs. This resulted in a breakthrough for LIM research by producing the most precise measurement of a LIM power spectrum to date. To achieve the LIM detection, we conducted an end-to-end analysis for measuring, modeling, and interpreting the HETDEX LIM power spectrum. After cleaning the spectra from systematic contributions, we measured the LAE-Lyα intensity cross-power spectrum. We created mock data using the SIMPLE code, including RSD, and using the HETDEX window function, sensitivity, and noise. Using these mock data, we calculated the expected signal, including the effect of the data cleaning and survey specifics, and covariance matrices. We estimated the total covariance matrix of the power spectra from the mocks and from the shuffled data. By fitting the fiducial mock prediction times an overall amplitude to the measured power spectra, we constrained the product of the detected LAE bias, the intensity bias, the mean intensity of undetected sources, and the ratio of the actual and fiducial RSD factors, $b_\mathrm{g} b_I \langle I \rangle \bar{F}_\mathrm{RSD} / \bar{F}^\mathrm{fid}_\mathrm{RSD}$. Assuming a fiducial LAE bias and $\bar{F}_\mathrm{RSD} = \bar{F}^\mathrm{fid}_\mathrm{RSD}$, we inferred lower values of $b_I \langle I \rangle$ than those from QSO-Lyα cross-correlations, but slightly higher values than the upper limit inferred from a Lyα forest-Lyα intensity cross-correlation measurement. Further assuming a fiducial intensity bias, our constraints on the Lyα luminosity density are slightly larger than those obtained by integrating extrapolated LAE luminosity functions, but on the same order of magnitude as the intensity of faint galaxies and of the CGM and IGM emission in a cosmological simulation with Lyα RT at $z = 2$. Finally, our results are consistent

with the total Lyα intensity being sourced by recombination in the ISM of star-forming galaxies. The philosophy of LIM is "no photons are left behind" and we have achieved exactly this for Lyα photons.

8.2 Outlook

This thesis showed the enormous potential of Lyα LIM with HETDEX to constrain models of galaxy formation and evolution at cosmic noon. In this section, I will outline possibilities for future research to extract even more information from Lyα photons in the HETDEX survey by optimizing the HETDEX data processing (Section 8.2.1) and combining the data with line intensity maps from other LIM observations (Section 8.2.2). It would also be useful to make the SIMPLE code more physically realistic by including effective models for line emission and RT, calibrated using high-fidelity RT simulations (Section 8.2.3). Finally, Section 8.2.4 will present ideas for desigining an optimal LIM survey. Figure 8.1 shows a summary of these research directions and a subset of scientific questions they will address.

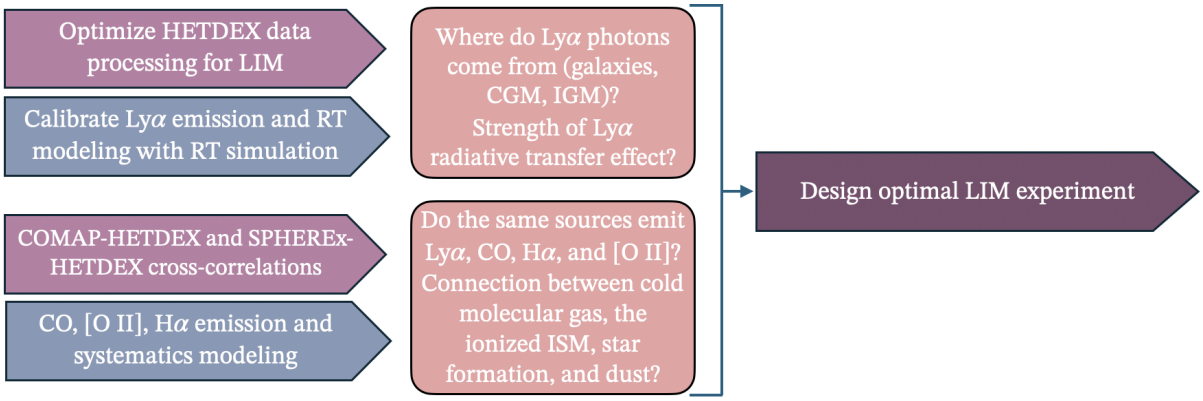

Figure 8.1: Future research directions informed by the results of this thesis and a subset of scientific questions to be answered next.

8.2.1 Optimizing LIM with HETDEX

While the LIM cross-power spectrum measured in Chapter 7 is very precise, it is still limited by the data quality and the analysis pipeline. It is currently necessary to remove large regions of the map and to clean the spectra using PCA, which reduces the cosmological signal and modifies the anisotropy of the measured power spectrum. We intend to optimize the HETDEX data processing for LIM measurements.

First, we will remove the source of the majority of systematic errors in the measurement by improving sky subtraction, using two methods. The first method builds on the current implementation of sky subtraction, but correcting for a known source of sky emission residuals. The second method is a new approach to sky subtraction altogether. The standard sky subtraction method estimates the sky spectrum model only from a single observation, which removes intensity fluctuations on large scales (see Section 3.5.1).

Instead, we will calculate the mutual information of the observed spectra with a global sky model, which combines all individual sky models from the thousands of HETDEX observations, using Machine Learning (ML) methods such as Non-negative Matrix Factorization (NMF). This new approach to sky subtraction will help LIM overcome the common limitation of large-scale signal loss.

Second, we will apply ML techniques to identify and correct for instrumental and observational systematics. For example, we will explore the performance of Rapid Automatic Image Categorization (RAIC) or an auto-encoder or U-network (similar to Zhong et al., 2025). This is already in place for the main HETDEX analysis pipeline (Mentuch Cooper et al. (2023a); Mentuch Cooper et al. in preparation). However, the sky subtraction method currently used for LIM introduces systematic errors that are not included in the main HETDEX mask. It would therefore be helpful to implement a similar ML systematics detection pipeline for the LIM analysis. Consequently, the necessary masking will be less extensive and much less PCA cleaning will be required, significantly increasing the measurement precision.

Additional, HETDEX-like data have been taken in parallel with other observations using the HET, which will double the map size when we include them in the LIM measurement. These improvements will significantly enhance the LIM signal detection, enabling us to extract valuable astrophysical and cosmological information. For instance, it will inform us about the origins of Lyα emission (see Section 8.2.3) and constrain the mean Lyα intensity of the Universe, the global volume density of the star-formation rate, and the escape fraction of Lyα photons from faint galaxies at cosmic noon more precisely. We could also be able to detect the anisotropy in the measured power spectrum caused by the peculiar velocity field and constrain the effect of Lyα RT on Lyα clustering measurements (see Section 2.4 and Chapter 6; Zheng et al., 2011; Lujan Niemeyer, 2025).

8.2.2 Measuring Cross-Correlations of Line Intensity Maps

Cross-correlating line intensity maps of different emission lines will give insight into the formation and evolution of galaxies throughout cosmic history. At the time of writing, there are no cross-correlation measurements of two different line intensity maps. We have the opportunity to measure the first LIM-LIM cross-correlations using data from HETDEX, the CO Mapping Pathfinder (COMAP; Cleary et al., 2022), and the SpectroPhotometer for the History of the Universe, Epoch of Reionization, and Ices Explorer (SPHEREX; Alibay et al., 2023).

One difficulty of LIM is that multiple emission lines at different redshifts contribute to the same line intensity map (see Section 3.5.2). For example, HETDEX maps photons from the Lyα emission line at $1.9 \leq z \leq 3.5$ and the [O II] emission line at $z \leq 0.47$. SPHEREX measures photons from the Hα ($0.14 \leq z \leq 6.6$), [O II] ($1.01 \leq z \leq 12.4$), and Lyα ($z \geq 5.17$) lines; and COMAP contains the CO (1-0) transition line at $2.4 \leq z \leq 3.4$ and the CO (2-1) line at $5.8 \leq z \leq 7.9$. In Chapter 7, the [O II] line does not contribute to the LAE-Lyα intensity cross-correlation because the [O II] line at $z \leq 0.47$ is uncorrelated with the LAEs at $1.9 \leq z \leq 3.5$. We will fully model the joint intensity maps and cross-correlations between the intensity maps using mocks data created in Section 8.2.3. It would also be interesting to test the performance of interloper line removal using

ancillary spectroscopic tracers of the large-scale structure, such as the HETDEX galaxy catalogs (Bernal & Baleato Lizancos, 2025).

The space-based SPHEREX mission maps the Hα and [O II] line intensities at redshifts overlapping with HETDEX. The SPHEREX deep field at the NEP is also observed by the TESLA survey (Chávez Ortiz et al., 2023), which uses the HETDEX observing strategy and data processing. We plan to measure the cross-correlation of the TESLA galaxy catalog and Lyα intensity map with the Hα and [O II] intensity maps of SPHEREX. Lyα and Hα are both emitted through hydrogen recombination in the ionized ISM; however, Lyα is much more prone to scattering and absorption by dust. Therefore, this measurement will constrain the Lyα escape fraction from faint star-forming galaxies. Because Lyα is also emitted through collisional excitation of neutral hydrogen, while Hα generally is not, the cross-correlation will also shed light on the relative contribution of different Lyα emission mechanisms to the total Lyα intensity. The cross-correlation with the [O II] intensity mapped by SPHEREX will yield even more information about the properties of faint galaxies, probing the connection between metallicity, ionization state, dust content, and star formation.

COMAP measures the intensity of the CO (1-0) transition line within two $4\,\mathrm{deg}^2$ fields overlapping with the HETDEX observations at redshifts $2.4 \leq z \leq 3.4$. CO traces cold molecular gas in galaxies, the fuel for star formation. The Lyα emission line is emitted by recombining hydrogen atoms following their ionization by young, massive stars. Lyα is thus associated with active star formation. Since CO emission is correlated with the presence of dust and Lyα photons are absorbed by dust, these two lines are expected to be anti-correlated. This anti-correlation of the CO and Lyα intensities encodes a wealth of information on the physics of star formation. We will be able to measure this cross-correlation using the data of HETDEX and COMAP at redshift $2.4 \leq z \leq 3.4$; the time when most stars in the Universe formed. This measurement will give insight into the sources emitting Lyα and CO and their connection to the dust content. It will connect the fuel for star formation, traced by CO emission, to its consumption, traced by the Lyα photons emitted in the ionized ISM.

8.2.3 Including Line Emission and RT Models in the SIMPLE Code

Mock data, such as those created with the SIMPLE code described in Chapter 5, are essential for quantifying the expected LIM signal and uncertainty, including the effect of data processing and survey specifics, and for the physical interpretation. However, the SIMPLE simulations do not contain astrophysical information beyond the line luminosity function, while LIM aims to study the physics behind the line emission and RT and their connection to the properties of the line-emitting galaxies. It would therefore be beneficial to create effective models for the line emission and RT of the most commonly targeted emission lines for LIM, such as Lyα, Hα, [O II], and CO, and include them in the modeling of the SIMPLE code. We plan to calibrate these models using high-fidelity cosmological simulations that include the physics of line emission and RT (e.g., Kannan et al., 2022; Byrohl & Nelson, 2023; Garcia et al., 2024). This will enable

SIMPLE to produce physically realistic intensity maps at minimal computational cost. As demonstrated in Chapter 7, the speed of these simulations is essential for forward-modeling all the complexities of the data. We also plan to create a framework to model the LIM-LIM cross-correlations described in Section 8.2.2.

These models will enable us to explore the dependence of Lyα LIM measurements on the relative contribution of star-forming galaxies and the CGM and IGM to the Lyα intensity and extract this information from the precise Lyα LIM measurement (see Section 8.2.1). We will also be able to investigate the imprint of the different Lyα emission mechanisms, i.e., collisional excitation and recombination of hydrogen atoms, on cross-correlations of Lyα intensity maps with other line intensity maps, such as Hα and [O II] (see Section 8.2.2). This will give insight into the dominant emission mechanisms of diffuse Lyα emission of the cosmic web (see Section 2.3.2). The effective line emission and RT models will also allow us to extract information on the connection between star formation activity, dust and metal contents, and molecular gas abundance from the LIM-LIM cross-correlation measurements in Section 8.2.2.

Finally, to establish SIMPLE as a robust and widely applicable tool for LIM experiments, it is important to ensure that the most important systematic effects of LIM surveys targeting emission lines across the electromagnetic spectrum can be accounted for.

8.2.4 Designing an Optimal LIM Experiment

Current optical and infrared LIM experiments are not specifically designed for LIM, but rely on data from other surveys such as HETDEX and SPHEREX. This section describes ideas for designing an optimal LIM experiment for studying galaxy formation and evolution in the young Universe. We intend to explore new observing strategies for LIM, especially in the optical and infrared wavelength regimes.

One limitation of HETDEX is the loss of intensity fluctuations on scales larger than a single observation due to the sky subtraction, a necessary step in the data processing (see Section 3.5.1). This is typical for deep observations of a single region, for example, for detecting galaxies. Instead, LIM surveys at lower frequencies, such as mm and sub-mm bands, move the telescope pointing during an observation, similar to the "nod-and-shuffle" technique for optical/near-infrared observations. As a result, the sky emission stays coherent over larger observed scales, and large-scale intensity fluctuations are preserved. It would be interesting to investigate whether a similar strategy is desirable and feasible for ground-based optical or infrared LIM experiments.

Space-based observations, such as those produced by SPHEREX, are not impacted by airglow, and therefore much less contaminated by foreground emission. This makes LIM measurements much cleaner, allowing for the preservation of large-scale intensity fluctuations. As a result, the power spectrum can be measured down to small wavenumbers, including the baryon acoustic oscillations, a powerful probe of cosmology. We plan to examine the benefits and drawbacks of space-based LIM surveys further.

Another consideration is the spectral resolution. Many LIM experiments have poor spectral resolution and are unable to measure the intensity fluctuations along the LOS. As demonstrated in Chapter 7, good spectral resolution improves the ability to identify and remove systematic contributions, e.g., the sky foreground emission, from the spec-

tra. Good spectral resolution also enables LIM surveys to measure the power spectrum anisotropy due to peculiar velocities and RT effects, revealing valuable information about cosmology and the physics of the line-emitting sources and the IGM. It would be interesting to quantify how much more can be learned from LIM surveys with good spectral resolution and to identify the ideal resolution.

8.3 Conclusions

In this thesis, we performed an end-to-end analysis for the LIM power spectrum measurement in HETDEX, first showing the potential for HETDEX to perform low-surface brightness measurements; then creating a fast mock simulation code to model line intensity maps of any emission line, including data cleaning and survey-specific effects; modeling the effects of Lyα RT on Lyα LIM power spectra; and finally precisely measuring the LIM power spectrum in HETDEX data. This is the most exciting time for LIM science, as the first observations are able to detect LIM power spectra and study otherwise elusive faint galaxies and gas. Upcoming LIM observations will bridge our knowledge of current-day galaxies to the first light in the Universe, and it is essential to develop robust data analysis and modeling techniques to extract the rich astrophysical and cosmological information encoded in the LIM data. This thesis has contributed to this endeavor by developing modeling and data processing techniques for LIM and I am excited to continue studying the frontier of astrophysics and cosmology.

Bibliography

Agrawal, A., Makiya, R., Chiang, C.-T., et al. 2017, JCAP, 2017, 003

Ahumada, R., Allende Prieto, C., Almeida, A., et al. 2020, ApJS, 249, 3

Alcock, C., & Paczyński, B. 1979, Nature, 281, 358

Alibay, F., Sindiy, O. V., Jansma, P. A. T., et al. 2023, in 2023 IEEE Aerospace Conference, 1–18

Almada Monter, S., & Gronke, M. 2024, MNRAS, 534, L7

Alonso, D., Ferreira, P. G., & Santos, M. G. 2014, MNRAS, 444, 3183

Ambrose, A. E., Visbal, E., Kulkarni, M., & McQuinn, M. 2025, JCAP, 2025, 080

Amiri, M., Bandura, K., Boskovic, A., et al. 2022, ApJS, 261, 29

Amiri, M., Bandura, K., Chakraborty, A., et al. 2024, ApJ, 963, 23

Ando, S., Benoit-Lévy, A., & Komatsu, E. 2018, MNRAS, 473, 4318

Arrigoni Battaia, F., Hennawi, J. F., Prochaska, J. X., et al. 2019, MNRAS, 482, 3162

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33

Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123

Bach, K., & Lee, H.-W. 2015, MNRAS, 446, 264

Bacon, D. J., Battye, R. A., Bull, P., et al. 2020, PASA, 37

Bacon, R., Mary, D., Garel, T., et al. 2021, A&A, 647, A107

Bartelmann, M. 2021, Theoretical Astrophysics: An Introduction, Lecture Notes (Heidelberg University Publishing)

Beane, A., Villaescusa-Navarro, F., & Lidz, A. 2019, ApJ, 874, 133

Beers, T. C., Flynn, K., & Gebhardt, K. 1990, AJ, 100, 32

Behrens, C., & Braun, H. 2014, A&A, 572, A74

Behrens, C., Byrohl, C., Saito, S., & Niemeyer, J. C. 2018, A&A, 614, A31

Behrens, C., Dijkstra, M., & Niemeyer, J. C. 2014, A&A, 563, A77

Behrens, C., & Niemeyer, J. 2013, A&A, 556, A5

Behroozi, P., Wechsler, R. H., Hearin, A. P., & Conroy, C. 2019, MNRAS, 488, 3143

Benitez, N., Dupke, R., Moles, M., et al. 2014, arXiv e-prints, arXiv:1403.5237

Bernal, J. L. 2024, Phys. Rev. D, 109, 043517

Bernal, J. L., & Baleato Lizancos, A. 2025, Phys. Rev. D, 111, 043539

Bernal, J. L., & Kovetz, E. D. 2022, A&A Rev., 30, 5

Béthermin, M., Gkogkou, A., Van Cuyck, M., et al. 2022, A&A, 667, A156

Bhagwat, A., Costa, T., Ciardi, B., Pakmor, R., & Garaldi, E. 2024, MNRAS, 531, 3406

Bigot-Sazy, M. A., Ma, Y. Z., Battye, R. A., et al. 2016, in Astronomical Society of the Pa-

cific Conference Series, Vol. 502, Frontiers in Radio Astronomy and FAST Early Sciences Symposium 2015, ed. L. Qain & D. Li, 41

Bird, S., Viel, M., & Haehnelt, M. G. 2012, MNRAS, 420, 2551

Blaizot, J., Garel, T., Verhamme, A., et al. 2023, MNRAS, 523, 3749

Blake, C. 2019, MNRAS, 489, 153

Blas, D., Lesgourgues, J., & Tram, T. 2011, JCAP, 07, 034

Blot, L., Crocce, M., Sefusatti, E., et al. 2019, MNRAS, 485, 2806

Bolton, J. S., & Haehnelt, M. G. 2007, MNRAS, 382, 325

Bonnassieux, E., Tremou, E., Girard, J. N., et al. 2021, Galaxies, 9, 105

Bonoli, S., Marín-Franch, A., Varela, J., et al. 2021, A&A, 653, A31

Bowman, W. P., Zeimann, G. R., Ciardullo, R., et al. 2019, ApJ, 875, 152

Bowman, W. P., Zeimann, G. R., Nagaraj, G., et al. 2020, ApJ, 899, 7

Bowman, W. P., Ciardullo, R., Zeimann, G. R., et al. 2021, ApJ, 920, 78

Brammer, G. B., van Dokkum, P. G., Franx, M., et al. 2012, ApJS, 200, 13

Breit, G., & Teller, E. 1940, ApJ, 91, 215

Breysse, P. C. 2022, arXiv e-prints, arXiv:2209.01223

Breysse, P. C., Kovetz, E. D., Behroozi, P. S., Dai, L., & Kamionkowski, M. 2017, MNRAS, 467, 2996

Breysse, P. C., Kovetz, E. D., & Kamionkowski, M. 2015, MNRAS, 452, 3408

Bryan, G. L., & Norman, M. L. 1998, ApJ, 495, 80

Burgess, A. 1965, MmRAS, 69, 1

Byrohl, C. 2021, PhD thesis, Ludwig-Maximilians-Universität München. `http://nbn-resolving.de/urn:nbn:de:bvb:19-308239`

Byrohl, C., & Nelson, D. 2023, MNRAS, 523, 5248

—. 2025, arXiv e-prints, arXiv:2507.11603

Byrohl, C., Saito, S., & Behrens, C. 2019, MNRAS, 489, 3472

Byrohl, C., Nelson, D., Behrens, C., et al. 2021, MNRAS, 506, 5129

Cai, Z., Hamden, E., Matuszewski, M., et al. 2018, ApJ, 861, L3

Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682

Cantalupo, S., Arrigoni-Battaia, F., Prochaska, J. X., Hennawi, J. F., & Madau, P. 2014, Nature, 506, 63

Cantalupo, S., Porciani, C., Lilly, S. J., & Miniati, F. 2005, ApJ, 628, 61

Carucci, I. P., Bernal, J. L., Cunnington, S., et al. 2024, arXiv e-prints, arXiv:2412.06750

Castander, F. J., Ballester, O., Bauer, A., et al. 2012, Proc. SPIE, 8446, 84466D

CCAT-Prime Collaboration, Aravena, M., Austermann, J. E., et al. 2022, ApJS, 264, 7

Chakraborty, A., Datta, A., Roy, N., et al. 2021, ApJ, 907, L7

Chang, S.-J., Yang, Y., Seon, K.-I., Zabludoff, A., & Lee, H.-W. 2023, ApJ, 945, 100

Chang, T.-C., Pen, U.-L., Bandura, K., & Peterson, J. B. 2010, Nature, 466, 463

Chávez Ortiz, Ó. A., Finkelstein, S. L., Davis, D., et al. 2023, ApJ, 952, 110

Chen, Z. 2025, MNRAS, 542, L1

Chen, Z., Cunnington, S., Pourtsidou, A., et al. 2025, ApJS, 279, 19

Cheng, S., & Ménard, B. 2021, arXiv e-prints, arXiv:2112.01288

Chiang, C.-T., Wullstein, P., Jeong, D., et al. 2013, JCAP, 2013, 030
Chiang, Y.-K., Ménard, B., & Schiminovich, D. 2019, ApJ, 877, 150
Chung, D. T. 2022, MNRAS, 517, 1625
Chung, D. T., Viero, M. P., Church, S. E., et al. 2019, ApJ, 872, 186
Chung, D. T., Breysse, P. C., Ihle, H. T., et al. 2021, ApJ, 923, 188
Chung, D. T., Breysse, P. C., Cleary, K. A., et al. 2022, ApJ, 933, 186
—. 2024, A&A, 691, A337
Cleary, K. A., Borowska, J., Breysse, P. C., et al. 2022, ApJ, 933, 182
Clerkin, L., Kirk, D., Manera, M., et al. 2017, MNRAS, 466, 1444
Collaboration, C.-P., Aravena, M., Austermann, J. E., et al. 2022, ApJS, 264, 7
Collette, A. 2013, Python and HDF5 (O'Reilly)
Comaschi, P., & Ferrara, A. 2016, MNRAS, 455, 725
CONCERTO Collaboration, Ade, P., Aravena, M., et al. 2020, A&A, 642, A60
Cooksy, A. L., Blake, G. A., & Saykally, R. J. 1986, ApJ, 305, L89
Cooray, A., & Sheth, R. 2002, Phys. Rep., 372, 1
Cooray, A., Chang, T.-C., Unwin, S., et al. 2019, in BAAS, Vol. 51, 23
Cox, T. A., Jacobs, D. C., & Murray, S. G. 2022, MNRAS, 512, 792
Crites, A. T., Bock, J. J., Bradford, C. M., et al. 2014, Proc. SPIE, 9153, 91531W
Croft, R. A. C., Miralda-Escudé, J., Zheng, Z., Blomqvist, M., & Pieri, M. 2018, MNRAS, 481, 1320
Croft, R. A. C., Miralda-Escudé, J., Zheng, Z., et al. 2016, MNRAS, 457, 3541
Cunnington, S., Li, Y., Santos, M. G., et al. 2023a, MNRAS, 518, 6262
Cunnington, S., Wolz, L., Bull, P., et al. 2023b, MNRAS, 523, 2453
Cunnington, S., Li, Y., Santos, M. G., et al. 2023c, MNRAS, 518, 6262
Dalgarno, A., & McCray, R. A. 1972, ARA&A, 10, 375
Davis, D., Gebhardt, K., Cooper, E. M., et al. 2023a, ApJ, 946, 86
—. 2023b, ApJ, 954, 209
Dayal, P., & Ferrara, A. 2018, Phys. Rep., 780, 1
DeBoer, D. R., Parsons, A. R., Aguirre, J. E., et al. 2017, PASP, 129, 045001
Desjacques, V., Jeong, D., & Schmidt, F. 2018, Phys. Rep., 733, 1
Dijkstra, M. 2019, Saas-Fee Advanced Course, 46, 1
Dijkstra, M., Haiman, Z., & Spaans, M. 2006, ApJ, 649, 14
Dijkstra, M., & Loeb, A. 2009, MNRAS, 400, 1109
Dong, C., Lee, K.-G., Ata, M., Horowitz, B., & Momose, R. 2023, ApJ, 945, L28
Doré, O., Bock, J., Ashby, M., et al. 2014, arXiv e-prints, arXiv:1412.4872
Doré, O., Werner, M. W., Ashby, M., et al. 2016, arXiv e-prints, arXiv:1606.07039
Doré, O., Werner, M. W., Ashby, M. L. N., et al. 2018, arXiv e-prints, arXiv:1805.05489
Draine, B. T. 2011, Physics of the Interstellar and Intergalactic Medium
Dumitru, S., Kulkarni, G., Lagache, G., & Haehnelt, M. G. 2019, MNRAS, 485, 3486
Dunne, D. A., Cleary, K. A., Breysse, P. C., et al. 2024, ApJ, 965, 7
—. 2025, A&A, 702, A247
Eastwood, M. W., Anderson, M. M., Monroe, R. M., et al. 2018, AJ, 156, 32

Eisenstein, D. J., & Hu, W. 1998, ApJ, 496, 605
Erb, D. K., Pettini, M., Steidel, C. C., et al. 2016, ApJ, 830, 52
Eriksen, M., Alarcon, A., Gaztanaga, E., et al. 2019, MNRAS, 484, 4200
Ewen, H. I., & Purcell, E. M. 1951, Nature, 168, 356
Farrow, D. J., Sánchez, A. G., Ciardullo, R., et al. 2021, MNRAS, 507, 3187
Faucher-Giguère, C.-A. 2020, MNRAS, 493, 1614
Feldman, H. A., Kaiser, N., & Peacock, J. A. 1994, ApJ, 426, 23
Feng, Y., Hand, N., & Biweidai. 2017, rainwoodman/pmesh 0.1.33, 0.1.33, Zenodo
Field, G. B., Rather, J. D. G., Aannestad, P. A., & Orszag, S. A. 1968, ApJ, 151, 953
Fitzpatrick, E. L. 1999, PASP, 111, 63
Fonseca, J., Silva, M. B., Santos, M. G., & Cooray, A. 2017, MNRAS, 464, 1948
Fronenberg, H., & Liu, A. 2024, ApJ, 975, 222
Fronenberg, H., Maniyar, A. S., Liu, A., & Pullen, A. R. 2024a, Phys. Rev. Lett., 132, 241001
Fronenberg, H., Maniyar, A. S., Pullen, A. R., & Liu, A. 2024b, Phys. Rev. D, 109, 123518
Fujimoto, S., Ouchi, M., Ferrara, A., et al. 2019, ApJ, 887, 107
Fujimoto, S., Silverman, J. D., Bethermin, M., et al. 2020, ApJ, 900, 1
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1
Garcia, K., Narayanan, D., Popping, G., et al. 2024, ApJ, 974, 197
Gardner, J. P., Mather, J. C., Abbott, R., et al. 2023, PASP, 135, 068001
Gawiser, E., Francke, H., Lai, K., et al. 2007, ApJ, 671, 278
Gebhardt, K., Mentuch Cooper, E., Ciardullo, R., et al. 2021, ApJ, 923, 217
Gong, Y., Cooray, A., Silva, M., et al. 2012, ApJ, 745, 49
Gould, A., & Weinberg, D. H. 1996, ApJ, 468, 462
Green, G. 2018, The Journal of Open Source Software, 3, 695
Greig, B., Komatsu, E., & Wyithe, J. S. B. 2013, MNRAS, 431, 1777
Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35
Gronke, M. 2017, A&A, 608, A139
Gronke, M., Dijkstra, M., Trenti, M., & Wyithe, S. 2015, MNRAS, 449, 1284
Guaita, L., Gawiser, E., Padilla, N., et al. 2010, ApJ, 714, 255
Gunn, J. E., & Peterson, B. A. 1965, ApJ, 142, 1633
Gurung-López, S., Orsi, Á. A., Bonoli, S., et al. 2020, MNRAS, 491, 3266
Gurung-López, S., Saito, S., Baugh, C. M., et al. 2021, MNRAS, 500, 603
Hagen, A., Zeimann, G. R., Behrens, C., et al. 2016, ApJ, 817, 79
Haiman, Z., Spaans, M., & Quataert, E. 2000, ApJ, 537, L5
Hamilton, A. J. S. 1998, Astrophysics and Space Science Library, 231, 185
Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Nature, 585, 357
Hashimoto, T., Ouchi, M., Shimasaku, K., et al. 2013, ApJ, 765, 70
Hathi, N. P., Le Fèvre, O., Ilbert, O., et al. 2016, A&A, 588, A26
Hayes, M., Schaerer, D., Östlin, G., et al. 2011, ApJ, 730, 8
Heneka, C., Cooray, A., & Feng, C. 2017, ApJ, 848, 52
Heneka, C., & Mesinger, A. 2020, MNRAS, 496, 581

Herrera, D., Gawiser, E., Benda, B., et al. 2025, ApJ, 988, L57
Hikage, C., & Yamamoto, K. 2016, MNRAS, 455, L77
Hill, G. J., Lee, H., MacQueen, P. J., et al. 2021, AJ, 162, 298
House, L. R., Gebhardt, K., Finkelstein, K., et al. 2024, ApJ, 975, 172
—. 2023, ApJ, 950, 82
Hui, L., & Gnedin, N. Y. 1997, MNRAS, 292, 27
Hummer, D. G., & Storey, P. J. 1987, MNRAS, 224, 801
Hunter, J. D. 2007, Computing in Science & Engineering, 9, 90
Jeong, D., & Komatsu, E. 2006, ApJ, 651, 619
Kaiser, N. 1987, MNRAS, 227, 1
Kakuma, R., Ouchi, M., Harikane, Y., et al. 2021, ApJ, 916, 22
Kannan, R., Smith, A., Garaldi, E., et al. 2022, MNRAS, 514, 3857
Karkare, K. S., Anderson, A. J., Barry, P. S., et al. 2022, Journal of Low Temperature Physics, 209, 758–765
Kayo, I., Taruya, A., & Suto, Y. 2001, ApJ, 561, 22
Keating, G. K., Marrone, D. P., Bower, G. C., & Keenan, R. P. 2020, ApJ, 901, 141
Keating, G. K., Marrone, D. P., Bower, G. C., et al. 2016, ApJ, 830, 34
Keating, G. K., Bower, G. C., Marrone, D. P., et al. 2015, ApJ, 814, 140
Keenan, R. P., Keating, G. K., & Marrone, D. P. 2022, ApJ, 927, 161
Kennicutt, R. C., & Evans, N. J. 2012, ARA&A, 50, 531–608
Kennicutt, R. C., & Evans, N. J. 2012, ARA&A, 50, 531
Kennicutt, Jr., R. C. 1998, ApJ, 498, 541
Kennicutt, Jr., R. C., Tamblyn, P., & Congdon, C. E. 1994, ApJ, 435, 22
Khaire, V., & Srianand, R. 2019, MNRAS, 484, 4174
Khanlari, M. M., Gebhardt, K., Weiss, L. H., et al. 2025, ApJ, 989, 169
Khoraminezhad, H., Saito, S., Gronke, M., & Byrohl, C. 2025, arXiv e-prints, arXiv:2507.16707
Kikuchihara, S., Harikane, Y., Ouchi, M., et al. 2022a, ApJ, 931, 97
—. 2022b, ApJ, 931, 97
Kikuta, S., Matsuda, Y., Cen, R., et al. 2019, PASJ, 71, L2
Kikuta, S., Matsuda, Y., Inoue, S., et al. 2023, ApJ, 947, 75
Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36
Konno, A., Ouchi, M., Nakajima, K., et al. 2016, ApJ, 823, 20
Kovetz, E. D., Viero, M. P., Lidz, A., et al. 2017, arXiv e-prints, arXiv:1709.09066
Kusakabe, H., Verhamme, A., Blaizot, J., et al. 2022, A&A, 660, A44
Lagache, G., Cousin, M., & Chatzikos, M. 2018, A&A, 609, A130
Laursen, P. 2010, PhD thesis, Niels Bohr Institute for Astronomy, Physics and Geophysics. `https://nbi.ku.dk/english/theses/phd-theses/peter-laursen`
Laursen, P., Sommer-Larsen, J., & Razoumov, A. O. 2011, ApJ, 728, 52
Leclercq, F., Bacon, R., Wisotzki, L., et al. 2017, A&A, 608, A8
Leung, A. S., Acquaviva, V., Gawiser, E., et al. 2017, ApJ, 843, 130
Li, T. Y., Wechsler, R. H., Devaraj, K., & Church, S. E. 2016, ApJ, 817, 169
Li, Z., Wolz, L., Guo, H., Cunnington, S., & Mao, Y. 2024, MNRAS, 534, 1801

Liang, L., Feldmann, R., Murray, N., et al. 2023, arXiv e-prints, arXiv:2301.04149

Libanore, S., & Kovetz, E. D. 2024, A&A, 687, A133

Lidz, A., Furlanetto, S. R., Oh, S. P., et al. 2011, ApJ, 741, 70

Lidz, A., & Taylor, J. 2016, ApJ, 825, 143

Lidz, A., Zahn, O., Furlanetto, S. R., et al. 2009, ApJ, 690, 252

Lin, X., Zheng, Z., & Cai, Z. 2022, ApJS, 262, 38

Linder, E. V. 2005, Phys. Rev. D, 72, 043529

Liu, C., Mutch, S. J., Angel, P. W., et al. 2016, MNRAS, 462, 235

Lujan Niemeyer, M. 2025, ApJ, 980, 250

Lujan Niemeyer, M., Bernal, J. L., & Komatsu, E. 2023, ApJ, 958, 4

Lujan Niemeyer, M., Bernal, J. L., & Komatsu, E. 2023, mlujnie/simple: v0.1.1, v0.1.1, Zenodo

Lujan Niemeyer, M., Komatsu, E., Byrohl, C., et al. 2022a, ApJ, 929, 90

Lujan Niemeyer, M., Bowman, W. P., Ciardullo, R., et al. 2022b, ApJ, 934, L26

Lujan Niemeyer, M., Komatsu, E., Bernal, J. L., et al. 2026, The Astrophysical Journal, 999, 177

Lunde, J. G. S., Stutzer, N. O., Breysse, P. C., et al. 2024, A&A, 691, A335

Makiya, R., Kayo, I., & Komatsu, E. 2021, JCAP, 2021, 095

Mallat, S. 2012, Communications on Pure and Applied Mathematics, 65, 1331

Maniyar, A. S., Schaan, E., & Pullen, A. R. 2022, Phys. Rev. D, 105, 083509

Martin, D. C., Darvish, B., Lin, Z., et al. 2023, Nature Astronomy, 7, 1390

Mas-Ribas, L., Dijkstra, M., Hennawi, J. F., et al. 2017, ApJ, 841, 19

Mas-Ribas, L., Sun, G., Chang, T.-C., Gonzalez, M. O., & Mebane, R. H. 2023, ApJ, 950, 39

McBride, L., & Liu, A. 2024, MNRAS, 533, 658–675

MeerKLASS Collaboration, Barberi-Squarotti, M., Bernal, J. L., et al. 2025, MNRAS, 537, 3632

Mentuch Cooper, E., Gebhardt, K., Davis, D., et al. 2023a, ApJ, 943, 177

—. 2023b, ApJ, 943, 177

Mesinger, A., Furlanetto, S., & Cen, R. 2011, MNRAS, 411, 955

Mesinger, A., Greig, B., & Sobacchi, E. 2016, MNRAS, 459, 2342

Miller, J. S. A., Bolton, J. S., & Hatch, N. A. 2021, MNRAS, 506, 6001

Minkowski, H. 1903, Mathematische Annalen, 57, 447. `http://eudml.org/doc/158108`

Momcheva, I. G., Brammer, G. B., van Dokkum, P. G., et al. 2016, ApJS, 225, 27

Mondal, R., & Barkana, R. 2023, Nature Astronomy, 7, 1025

Mondal, R., Barkana, R., & Fialkov, A. 2024, MNRAS, 527, 1461

Moradinezhad Dizgah, A., Bellini, E., & Keating, G. K. 2024, ApJ, 965, 19

Moradinezhad Dizgah, A., & Keating, G. K. 2019, ApJ, 872, 126

Moradinezhad Dizgah, A., Keating, G. K., Karkare, K. S., Crites, A., & Choudhury, S. R. 2022a, ApJ, 926, 137

Moradinezhad Dizgah, A., Nikakhtar, F., Keating, G. K., & Castorina, E. 2022b, JCAP, 2022, 026

Moriwaki, K., Yoshida, N., Eide, M. B., & Ciardi, B. 2019, MNRAS, 489, 2471

Mortlock, D. 2016, Astrophysics and Space Science Library, 423, 187

Nakajima, K., Fletcher, T., Ellis, R. S., Robertson, B. E., & Iwata, I. 2018a, MNRAS, 477, 2098

—. 2018b, MNRAS, 477, 2098

Nakajima, K., Ouchi, M., Shimasaku, K., et al. 2012, ApJ, 745, 12

Newburgh, L. B., et al. 2016, Proc. SPIE, 9906, 99065X

Obuljen, A., Simonović, M., Schneider, A., & Feldmann, R. 2023, Phys. Rev. D, 108, 083528

Ouchi, M. 2019, Saas-Fee Advanced Course, 46, 189

Ouchi, M., Ono, Y., & Shibuya, T. 2020, ARA&A, 58, 617

Ouchi, M., Shimasaku, K., Akiyama, M., et al. 2008, ApJS, 176, 301

Padmanabhan, H. 2024, arXiv e-prints, arXiv:2411.08113

Padmanabhan, H., & Loeb, A. 2024a, arXiv e-prints, arXiv:2408.16820

—. 2024b, JCAP, 2024, 059

Partridge, R. B., & Peebles, P. J. E. 1967, ApJ, 147, 868

Paul, S., Santos, M. G., Chen, Z., & Wolz, L. 2023, arXiv e-prints, arXiv:2301.11943

Peebles, P. J. E. 1980, The Large-scale Structure of the Universe (Princeton, NJ: Princeton Univ. Press)

Percival, W. J., & White, M. 2009, MNRAS, 393, 297

Perdereau, O., Ansari, R., Stebbins, A., et al. 2022, MNRAS, 517, 4637

Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020a, A&A, 641, A6

—. 2020b, A&A, 641, A6

Pullen, A. R., Serra, P., Chang, T.-C., Doré, O., & Ho, S. 2018, MNRAS, 478, 1911–1924

Raiter, A., Schaerer, D., & Fosbury, R. A. E. 2010, A&A, 523, A64

Ramírez-Pérez, C., Sanchez, J., Alonso, D., & Font-Ribera, A. 2022, JCAP, 2022, 002

Ramsey, L. W., Sebring, T. A., & Sneden, C. A. 1994, Proc. SPIE, 2199, 31

Ramsey, L. W., Adams, M. T., Barnes, T. G., et al. 1998, Proc. SPIE, 3352, 34

Rauch, M. 1998, ARA&A, 36, 267

Reddy, N. A., Topping, M. W., Shapley, A. E., et al. 2022, ApJ, 926, 31

Renard, P., Spinoso, D., Montero-Camacho, P., et al. 2024, MNRAS, 535, 826–852

Renard, P., Spinoso, D., Sun, Z., et al. 2025, in Highlights of Spanish Astrophysics XII, ed. M. Manteiga, F. González-Galindo, A. Labiano-Ortega, M. J. Martínez-González, N. Rea, M. Romero-Gómez, A. Ulla-Miguel, G. Yepes, C. Rodríguez-López, A. Gómez-García, & et al., 4

Renard, P., Gaztanaga, E., Croft, R., et al. 2021, MNRAS, 501, 3883

Rocklin, M. 2015, in Proceedings of the 14th Python in Science Conference, ed. Kathryn Huff & James Bergstra, 126 – 132

Rowan-Robinson, M., Oliver, S., Wang, L., et al. 2016, MNRAS, 461, 1100

Roy, A., Battaglia, N., & Pullen, A. R. 2024, arXiv e-prints, arXiv:2406.07861

Roy, A., Valentín-Martínez, D., Wang, K., Battaglia, N., & van Engelen, A. 2023, ApJ, 957, 87

Rubiola, A., Cunnington, S., & Camera, S. 2022, MNRAS, 516, 5454

Runnholm, A., Hayes, M., Melinder, J., et al. 2020, ApJ, 892, 48

Russell, R. W., Melnick, G., Gull, G. E., & Harwit, M. 1980, ApJ, 240, L99

Sabla, V. I., Bernal, J. L., Sato-Polito, G., & Kamionkowski, M. 2024, Phys. Rev. D, 110,

023507
Saito, S. 2016, Galaxy Clustering in Redshift Space, https://wwwmpa.mpa-garching.mpg.de/~komatsu/lecturenotes/Shun_Saito_on_RSD.pdf
Salpeter, E. E. 1955, ApJ, 121, 161
Santos, M., Bull, P., Camera, S., et al. 2016, in MeerKAT Science: On the Pathway to the SKA, 32
Sato-Polito, G., & Bernal, J. L. 2022, Phys. Rev. D, 106, 103534
Sato-Polito, G., Kokron, N., & Bernal, J. L. 2023, MNRAS, 526, 5883
Schaan, E., & White, M. 2021, JCAP, 2021, 068
Schechter, P. 1976, ApJ, 203, 297
Scheuer, P. A. G. 1965, Nature, 207, 963
Schlafly, E. F., & Finkbeiner, D. P. 2011, ApJ, 737, 103
Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525
Scoccimarro, R. 2004, Phys. Rev. D, 70, 083007
Shibuya, T., Ouchi, M., Konno, A., et al. 2018, PASJ, 70, S14
Shimakawa, R., Kodama, T., Shibuya, T., et al. 2017, MNRAS, 468, 1123
Shimwell, T. W., Tasse, C., Hardcastle, M. J., et al. 2019, A&A, 622, A1
Shin, J., Kim, J., Pichon, C., Jeong, D., & Park, C. 2017, ApJ, 843, 73
Shoji, M., Jeong, D., & Komatsu, E. 2009, ApJ, 693, 1404
Shull, J. M., & van Steenberg, M. E. 1985, ApJ, 298, 268
Shvartzvald, Y., Waxman, E., Gal-Yam, A., et al. 2024, ApJ, 964, 74
Silva, M., Santos, M. G., Cooray, A., & Gong, Y. 2015, ApJ, 806, 209
Silva, M. B., Santos, M. G., Gong, Y., Cooray, A., & Bock, J. 2013, ApJ, 763, 132
Silva, M. B., Baumschlager, B., Cleary, K. A., et al. 2021, arXiv e-prints, arXiv:2111.05354
Slosar, A., Ahmed, Z., Alonso, D., et al. 2019, in BAAS, Vol. 51, 53
Smith, R. E., Peacock, J. A., Jenkins, A., et al. 2003, MNRAS, 341, 1311
Somerville, R. S., Popping, G., & Trager, S. C. 2015, MNRAS, 453, 4337
Spina, B., Porciani, C., & Schimd, C. 2021, MNRAS, 505, 3492
Spinoso, D., Orsi, A., López-Sanjuan, C., et al. 2020, A&A, 643, A149
Stacey, G. J., Geis, N., Genzel, R., et al. 1991, ApJ, 373, 423
Stark, D. P., Topping, M. W., Endsley, R., & Tang, M. 2025, arXiv e-prints, arXiv:2501.17078
Steidel, C. C., Bogosavljević, M., Shapley, A. E., et al. 2011, ApJ, 736, 160
Stutzer, N. O., Lunde, J. G. S., Breysse, P. C., et al. 2024, A&A, 691, A336
Sun, G., Moncelsi, L., Viero, M. P., et al. 2018, ApJ, 856, 107
Sun, G., Chang, T. C., Uzgil, B. D., et al. 2021, ApJ, 915, 33
Switzer, E. R., Barrentine, E. M., Cataldo, G., et al. 2021, Journal of Astronomical Telescopes, Instruments, and Systems, 7, 044004
Tacconi, L. J., Genzel, R., & Sternberg, A. 2020, ARA&A, 58, 157–203
Takahashi, R., Sato, M., Nishimichi, T., Taruya, A., & Oguri, M. 2012, ApJ, 761, 152
Taruya, A., Nishimichi, T., & Saito, S. 2010, Phys. Rev. D, 82, 063522
Tingay, S. J., Goeke, R., Bowman, J. D., et al. 2013, PASA, 30, e007
Tinker, J. L., Robertson, B. E., Kravtsov, A. V., et al. 2010, ApJ, 724, 878

Tornotti, D., Fossati, M., Fumagalli, M., et al. 2025a, arXiv e-prints, arXiv:2506.10083
Tornotti, D., Fumagalli, M., Fossati, M., et al. 2025b, ApJ, 980, L43
—. 2025c, Nature Astronomy, 9, 577
Trainor, R. F., Lamb, N. R., Steidel, C. C., et al. 2025, arXiv e-prints, arXiv:2505.15881
Trainor, R. F., Strom, A. L., Steidel, C. C., & Rudie, G. C. 2016, ApJ, 832, 171
Trainor, R. F., Strom, A. L., Steidel, C. C., et al. 2019, ApJ, 887, 85
Tumlinson, J., Peeples, M. S., & Werk, J. K. 2017, ARA&A, 55, 389
Tumlinson, J., & Shull, J. M. 2000, ApJ, 528, L65
Turner, W., Martini, P., Karaçaylı, N. G., et al. 2024, ApJ, 976, 143
Umeda, H., Ouchi, M., Kikuta, S., et al. 2025, ApJS, 277, 37
Umehata, H., Fumagalli, M., Smail, I., et al. 2019, Science, 366, 97
Van Cuyck, M., Ponthieu, N., Lagache, G., et al. 2023, A&A, 676, A62
van de Hulst, H. C. 1945, Nederlandsch Tijdschrift voor Natuurkunde, 11, 210
van Haarlem, M. P., Wise, M. W., Gunst, A. W., et al. 2013, A&A, 556, A2
Vanderlinde, K., Liu, A., Gaensler, B., et al. 2019, in Canadian Long Range Plan for Astronomy and Astrophysics White Papers, Vol. 2020, 28
Verhamme, A., Schaerer, D., & Maselli, A. 2006a, A&A, 460, 397
—. 2006b, A&A, 460, 397
Vieira, J., Aguirre, J., Bradford, C. M., et al. 2020, arXiv e-prints, arXiv:2009.14340
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nature Methods, 17, 261
Wang, L., & Steinhardt, P. J. 1998, ApJ, 508, 483
Wang, Y., Zhao, G.-B., Koyama, K., et al. 2024, Communications Physics, 7, 130
Weiss, L. H., Bowman, W. P., Ciardullo, R., et al. 2021, ApJ, 912, 100
Weiss, L. H., Davis, D., Gebhardt, K., et al. 2024, ApJ, 962, 102
Weiss, L. H., Gebhardt, K., Davis, D., et al. 2025, ApJ, 983, 72
Wilson, R. W., Jefferts, K. B., & Penzias, A. A. 1970, ApJ, 161, L43
Wisotzki, L., Bacon, R., Blaizot, J., et al. 2016, A&A, 587, A98
Wisotzki, L., Bacon, R., Brinchmann, J., et al. 2018, Nature, 562, 229
Wolfe, A. M., Gawiser, E., & Prochaska, J. X. 2005, ARA&A, 43, 861
Wolz, L., Murray, S. G., Blake, C., & Wyithe, J. S. 2019, MNRAS, 484, 1007
Wolz, L., Tonini, C., Blake, C., & Wyithe, J. S. B. 2016, MNRAS, 458, 3399
Wuensche, C. A., & the BINGO Collaboration. 2019, in Journal of Physics Conference Series, Vol. 1269, Journal of Physics Conference Series (IOP), 012002
Wyithe, J. S. B., & Dijkstra, M. 2011, MNRAS, 415, 3929
Wyse, R. F. G., & Gilmore, G. 1992, MNRAS, 257, 1
Yang, S., Pullen, A. R., & Switzer, E. R. 2019, MNRAS: Letters, 489, L53–L57
Zheng, Z., Cen, R., Trac, H., & Miralda-Escudé, J. 2011, ApJ, 726, 38
Zhong, F., Napolitano, N. R., Heneka, C., et al. 2025, MNRAS, 543, 691–708

Acknowledgements

小松英一郎先生のご指導に心より感謝申し上げます。
I am deeply grateful to my supervisor, Eiichiro Komatsu. Thank you for giving me the opportunity to conduct research in astrophysics, for your continuous guidance and advice, for patiently answering all of my questions, and for meticulously editing my drafts. Thank you for letting me choose the projects I was most interested in and for always supporting me on my path.

I also want to express my deep gratitude towards Karl Gebhardt for showing me how much fun astronomy is, for his mentorship and friendship throughout the years, and for showing me where to find great music in Austin. Thank you very much for inviting me to Austin, I had an amazing and truly life-changing time.

I am grateful towards the group at MPA for making my time as a Master and PhD student immensely enjoyable. I have been blessed with many wonderful friends at MPA and I am glad to have learned from exceptionally talented scientists. Thank you to Laura Herold, Eirini Batziou, Jeongin Moon, and Ivana Nikolac, for being the best office mates in the secretly prettiest office at MPA. Thank you to José Bernal and Chris Byrohl for all of your patient explanations and for proof-reading my thesis. Thank you to Jakob Hein for the fun and productive library sessions and planetarium shows, and to Jakob Ehring, Robert Glas, Malte Heinlein, Daniel Kresse, and Johannes Ringler for playing table soccer with me. Many thanks to Marta Monelli and Vyoma Muralidhara for making me feel so welcome as a new student, to Julia Stadler, Drew Jamieson, and Max Grönke for patiently answering my physics questions, and to Miha Černetič for helping me with coding. And thank you very much to Toshiki Kurita for helping me with the Japanese sentence above.

Thank you to the HETDEX collaboration, especially Erin Mentuch Cooper, Robin Ciardullo, Dustin Davis, Maximilian Fabricius, Daniel Farrow, Steven Finkelstein, Eric Gawiser, Gary Hill, Matt Jarvis, Donghui Jeong, Mahan Mirza Khanlari, Masami Ouchi, Shun Saito, Don Schneider, Laurel Weiss, Lutz Wisotzki, and Greg Zeimann for your help with the HETDEX data analysis, paper writing and editing, and for making HETDEX such a fun and vibrant collaboration.

Finally, I want to express my heartfelt gratitude for my family. Thank you to my husband, Taylor, for your neverending encouragement and loving support. Thank you for celebrating the milestones with me and for helping me through the difficult times of the PhD. Thank you to Charlotte, for all the snuggles and games. Many thanks to my parents and my siblings, Tim and Veronika, for fun discussions about everything, and for your practical and mental support. Finally, thank you to my grandparents and my in-law family for taking an interest in my research and cheering me on. I am glad that I can always depend on all of you wonderful people.