
Radiative transfer simulations of Lyman-alpha photons in the Universe: Application to cosmology and astrophysics

Chris Byrohl



München 2021

**Radiative transfer simulations of
Lyman-alpha photons in the Universe:
Application to cosmology and
astrophysics**

Chris Byrohl

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

vorgelegt von
Chris Byrohl
aus Hameln

München, den 31. August 2021

Erstgutachter: Eiichiro Komatsu

Zweitgutachter: Klaus Dolag

Tag der mündlichen Prüfung: 11. November 2021

ZUSAMMENFASSUNG / ABSTRACT

Die Lyman- α ($Ly\alpha$) Emissionslinie ist eine der hellsten Linien im Universum. Mit neuen Instrumenten werden zunehmend größere Bereiche des Himmels mit höherer spektraler und räumlicher Auflösung und Tiefe durchmustert. $Ly\alpha$ Beobachtungen werden dadurch zunehmend bedeutsame Erkenntnisse für unser astrophysikalisches und kosmologisches Verständnis liefern. Somit hat die $Ly\alpha$ Emissionslinie ein enormes Potential, jedoch ist die Interpretation jener Beobachtungen aufgrund des komplexen Strahlungstransports der $Ly\alpha$ Photonen schwierig. Die $Ly\alpha$ Emissionslinie ist eine resonante Emissionslinie und zugleich reichen kleine Mengen neutralen Wasserstoffs für hohe optische Tiefen aus. $Ly\alpha$ Photonen werden somit vielfach gestreut bevor sie ihrer Umgebung entkommen oder durch Staub absorbiert werden. Diese Streuprozesse modifizieren die beobachteten räumlichen und spektralen Eigenschaften der Emissionslinie in Beobachtungen grundlegend, was den Rückschluss auf die zugrunde liegenden physikalischen Eigenschaften des abstrahlenden Gases erschwert.

In dieser Dissertation präsentiere ich einen Code für $Ly\alpha$ Strahlungstransport, um die komplexen Streuvorgänge nachvollziehbar zu machen, die die Interpretation heutiger und zukünftiger $Ly\alpha$ Beobachtungen erschweren. Ich präsentiere drei Anwendungen jenes Codes in der Astrophysik und Kosmologie unter Zuhilfenahme existierender numerischer Simulationen der Galaxieentstehung.

Zuerst präsentiere ich den Effekt des $Ly\alpha$ Strahlungstransports auf das beobachtete Clustering von $Ly\alpha$ emittierenden Galaxien. Heutige und zukünftige spektroskopische Durchmusterungen des Himmels kartieren solche Galaxien im jungen Universum. Die räumliche Häufung solcher Galaxien kann genutzt werden um die Parameter des kosmologischen Modells genauer zu bestimmen, was etwa mit der HETDEX Durchmusterung derzeit vorangetrieben wird. Die räumliche Häufung jener Galaxien wird jedoch durch die Interaktion der $Ly\alpha$ Photonen mit dem neutralen Wasserstoff auf dem Weg zum Beobachter verfälscht. Ich präsentiere in dieser Arbeit wie die spektrale Form der beobachteten $Ly\alpha$ Emissionslinie die Statistik der räumlichen Häufung modifiziert und wie dieser Störungseffekt korrigiert werden kann.

Weiterhin zeige ich auf wie der Einfluss des Ly α Strahlungstransports auf die Form der Ly α Spektren durch verschiedene räumlicher Skalen auseinander gehalten werden kann. Die Form der Ly α Spektren wird sowohl durch den Strahlungstransport innerhalb des Galaxie als auch durch Streuungen im intergalaktischen Medium geprägt. Die Effekte von diesen verschiedenen räumlichen Skalen auf die Spektren ist schwierig auseinanderzuhalten. Die korrekte Interpretation der Stärke jener Aufprägung ermöglicht es wichtige Rückschlüsse etwa bezüglich der Kinematik innerhalb der Galaxien zu ziehen. Ich zeige verschiedene einfache statistische Methoden auf mit welchen die Aufprägung der verschiedenen räumlichen Skalen auseinandergehalten werden kann. Weiterhin mache ich einen Katalog der verwendeten Daten öffentlich, um weitere Studien zu ermöglichen.

Zum Ende untersuche ich mit dem Ly α Strahlungstransportcode die Natur sogenannter Ly α Halos (LAHs). Beobachtungen zeigen LAHs als diffuses Leuchten der Ly α Linie um Galaxien mit aktiver Sternentstehung. Durch Modellierung der gewöhnlich betrachteten Emissionsmechanismen und Nutzung der hochmodernen TNG50 Galaxieentstehungssimulationen stelle ich Voraussagen für die erwarteten LAHs um die simulierten Galaxien auf. In den Simulationen zeigt sich, dass sich die beobachteten Radialprofile von LAHs mit Streuungen von Photonen, welche den sternentstehenden Regionen innerhalb der Galaxien entkommen, erklären lassen. Dennoch kann die diffuse Emission im zirkumgalaktischen Medium für individuelle Galaxien erheblich sein. Weiterhin zeigen die simulierten Radialprofile der Ly α Halos eine vielversprechende Übereinstimmung mit neuesten Beobachtungsdaten bei Rotverschiebungen um $z \sim 3$ auf.

The Lyman- α ($\text{Ly}\alpha$) emission line is one of the brightest lines in the high-redshift Universe. With new instruments and surveys mapping larger areas of the sky with higher spectral resolving power, spatial resolution, and depth, Lyman- α observations are going to providing competitive constraints of astrophysical theories and cosmological models. Even though the Lyman- α line has huge potential, the interpretation of its observation is difficult given the complex radiative transfer (RT) $\text{Ly}\alpha$ photons experience. The $\text{Ly}\alpha$ line is resonant and even small amounts of neutral hydrogen lead to large optical depths. As a consequence, $\text{Ly}\alpha$ photons will scatter numerous times until they either escape from its source or destruction by dust. Hence, spectral and spatial information of $\text{Ly}\alpha$ emission is significantly altered and this modification needs to be accounted for in order to obtain the encoded physical information in its observation.

In this thesis, I present a framework for $\text{Ly}\alpha$ RT simulations to understand the complex RT involved in recent and upcoming $\text{Ly}\alpha$ observations. Primarily applied to cosmological hydrodynamic galaxy formation simulations, I present three applications of such framework in an astrophysical and cosmological context.

First, I analyze the effects of $\text{Ly}\alpha$ RT on the clustering of $\text{Ly}\alpha$ emitting galaxies in cosmological redshift surveys. Current and future blind spectroscopic surveys map out the distribution of such galaxies in the high-redshift Universe. The spatial clustering can be used to constrain cosmological models as planned with the HETDEX survey currently under way. However, the observed clustering signal can be distorted due to the RT $\text{Ly}\alpha$ photons experience before reaching the observer. I present a new phenomenological effect distorting the clustering signal due to the spectral shape of $\text{Ly}\alpha$ emitters. I demonstrate that such effect can be modelled analogously to the well-studied Fingers-of-God effect and discuss different methods to correct for this distortion.

Second, I statistically analyze and disentangle the shaping factors of the $\text{Ly}\alpha$ spectra. $\text{Ly}\alpha$ spectra are shaped by both the small-scale structure within the galaxy and the structure of the intergalactic medium along the line of sight. The imprinted spectral features from those two different spatial scales can appear degenerate. Properly distinguishing the features to arise on either scale allows us to infer important insights, such as the kinematic structure of the originating halo and the escape of ionizing photons. I sketch out different simple statistical measures to break the scale degeneracy, which might enable inferences on the underlying physical processes on both scales. By providing a public data set of $\text{Ly}\alpha$ transmission curves in the intergalactic medium, more dedicated studies by the astronomy community are made possible.

Third, I investigate the nature of so-called $\text{Ly}\alpha$ halos (LAHs). In observations, LAHs describe the diffuse $\text{Ly}\alpha$ glow around star-forming galaxies. Modelling the commonly considered emission mechanisms and using the state-of-the-art TNG50 galaxy forma-

tion simulation, I provide predictions for LAHs with an unprecedented combination of statistical sample and resolution. I find that scattering of Ly α photons from inner star-forming regions within the galaxies dominate the faint glow in the circumgalactic medium, but substantial contributions from diffuse emission can be present. The predicted surface brightness radial profiles show a promising agreement with latest observational constraints at $z \sim 3$.

ACKNOWLEDGEMENTS

My PhD thesis represents a collection of work I have undertaken over the last years in Garching. I am deeply thankful for the academic environment and the support of my friends and family that I had during this time. I would like to thank my supervisor Eiichiro Komatsu for his consistent guidance, enthusiasm and for cultivating such an open-minded and fun atmosphere in his group. I would also like to thank Shun Saito for his assistance throughout my first year at MPA.

As for most people, my research was a collaborative effort sparked by many discussions over the years. The three chapters 4, 5 and 6 were largely driven by collaboration with Shun Saito, Max Gronke, and Dylan Nelson respectively. These projects allowed me to lurk into different topics within the scope of the unifying topic of $\text{Ly}\alpha$ radiative transfer within this thesis. Thanks to all of you. Much of the work here spun off from Christoph Behrens' work and I thank Christoph for introducing and guiding me into the field of $\text{Ly}\alpha$ radiative transfer simulations over the last years. Furthermore, I would also like to thank Robert Fisher for giving me ample support since my visit as an undergraduate at UMass Dartmouth in 2015 and throughout my PhD.

Thank you to all members of the cosmology group at MPA for the great memories. I always felt among friends first (and incredibly talented colleagues second). Special thoughts go to Leila, Kaloian and Alex for endless coffee and lunch breaks getting out of hand, Fabian supplying me with beer for all the lost bets, Laura and Maja for sharing my confusion with Lyman-alpha, and Luisa for cheering me up during the last stressful months of thesis writing. I want to especially thank Elisa Ferreira for her continued academic and personal support during some of the rougher patches in the previous years.

Thanks to all the friends I made in Garching making MPA more than a workplace. Especially to Miha, Matteo, Timo, Ivan, Linda and Vlas for various late night and multi-day cycling adventures in Munich, southern Germany and beyond, Malin that I dearly miss as my officemate, and Ricard for making the weekly after work beer socials happen.

Thank you again to all my friends and colleagues at the Max Planck Institute for Astrophysics that I share so many fond memories and those that have been supporting me through the personal and academic challenges in the last years.

TABLE OF CONTENTS

| | |
|---|------------|
| List of figures | xv |
| List of tables | xix |
| 1 Introduction | 1 |
| 1.1 Motivation | 1 |
| 1.2 Observations | 4 |
| 1.2.1 Lyman-alpha emitters | 4 |
| 1.2.2 Lyman-alpha halos and blobs | 6 |
| 1.3 Tackled problems in this thesis | 7 |
| 1.3.1 Completing the distortion effects in LAE clustering | 7 |
| 1.3.2 Separating the imprint of spatial scales in LAE spectra | 8 |
| 1.3.3 Demystifying Lyman-alpha halos | 9 |
| 1.4 Structure of this thesis | 10 |
| 2 The Lyman-alpha line | 11 |
| 2.1 The hydrogen atom and the Ly α transition | 11 |
| 2.2 A quantum mechanical treatment of the hydrogen atom | 12 |
| 2.3 Photon absorption and emission | 15 |
| 2.4 Selection rules and radiative cascades | 17 |
| 2.5 Line profile and cross-section | 19 |
| 2.6 Emission | 22 |
| 2.6.1 Mechanisms | 22 |
| 2.6.2 Sources | 24 |
| 2.7 Absorption and dust | 29 |
| 2.8 Scattering | 30 |
| 2.8.1 Frequency shift | 31 |
| 2.8.2 Change of direction | 32 |
| 2.9 Ly α radiative transfer | 33 |

| | | |
|----------|--|-----------|
| 2.9.1 | Absorption term | 34 |
| 2.9.2 | Emission term | 34 |
| 2.9.3 | Scattering term | 35 |
| 2.10 | Building a Ly α radiative transfer intuition | 36 |
| 2.10.1 | Random walk | 36 |
| 2.10.2 | Insights from the scattering process | 38 |
| 2.10.3 | Toy models | 43 |
| 2.10.4 | Moving beyond toy models | 45 |
| 3 | voroiLTIS: A Lyα radiative transfer code for unstructured grids | 47 |
| 3.1 | Introduction | 47 |
| 3.2 | Monte Carlo procedure | 47 |
| 3.2.1 | Emission | 48 |
| 3.2.2 | Propagation, scattering and destruction. | 48 |
| 3.2.3 | Peeling-off photons | 50 |
| 3.3 | Numerical approximations | 50 |
| 3.3.1 | Cross-section | 50 |
| 3.3.2 | Scattering | 51 |
| 3.4 | Geometry handling and Voronoi tessellation | 53 |
| 3.4.1 | Preprocessing | 55 |
| 3.4.2 | Photon propagation | 55 |
| 3.5 | Code verification | 56 |
| 3.6 | Scaling and performance | 57 |
| 3.6.1 | Multithreading | 57 |
| 3.6.2 | Multiprocessing | 58 |
| 3.7 | Data analysis | 59 |
| 3.7.1 | Observables | 59 |
| 3.7.2 | Computation | 60 |
| 3.8 | Application to Illustris(TNG) | 61 |
| 3.8.1 | Overview | 61 |
| 3.8.2 | Gas state | 62 |
| 3.8.3 | Preparation | 63 |
| 4 | LAE Clustering: A new Fingers-of-God damping | 65 |
| 4.1 | Introduction | 66 |
| 4.2 | Theoretical background of RSD | 68 |
| 4.3 | Methods | 72 |

| | | |
|----------|---|------------|
| 4.3.1 | Radiative transfer simulations | 72 |
| 4.3.2 | Analysis of simulated LAEs in redshift space | 74 |
| 4.4 | Results | 77 |
| 4.4.1 | Spectra | 77 |
| 4.4.2 | Distinguishing distortion contributions | 80 |
| 4.4.3 | RT velocity PDF | 80 |
| 4.4.4 | Configuration space: TPCF and the pairwise velocity PDF | 83 |
| 4.4.5 | Fourier space: The FoG damping factor | 86 |
| 4.5 | Discussion | 90 |
| 4.5.1 | Spectra | 90 |
| 4.5.2 | Localization along the line of sight | 92 |
| 4.5.3 | Correction to mitigate the impact of RT | 93 |
| 4.6 | Conclusions | 96 |
| 4.7 | Supplementary material | 99 |
| 4.7.1 | Detection algorithm variations | 99 |
| 4.7.2 | Comparison to prior studies | 101 |
| 4.7.3 | Discrepancies in the damping factor modeling | 102 |
| 5 | Lyman-alpha spectra and the intergalactic medium | 105 |
| 5.1 | Introduction | 105 |
| 5.2 | Methodology | 107 |
| 5.2.1 | Simulations | 107 |
| 5.2.2 | Input spectra | 108 |
| 5.3 | Results | 109 |
| 5.3.1 | Averaged transmission curves | 109 |
| 5.3.2 | Variations in transmission curves | 111 |
| 5.4 | Conclusions | 115 |
| 5.5 | Supplementary material | 116 |
| 5.5.1 | Spectral resolution | 116 |
| 5.5.2 | Start of integration | 117 |
| 5.5.3 | Bimodality of the transmission curves | 118 |
| 6 | Simulating Lyman-alpha halos | 121 |
| 6.1 | Introduction | 122 |
| 6.2 | Methods | 125 |
| 6.2.1 | IllustrisTNG and TNG50 | 125 |
| 6.2.2 | voroILTIS | 125 |

| | | |
|----------|---|------------|
| 6.2.3 | Lyman-alpha radiative processes | 126 |
| 6.2.4 | Post-processing and observational realism | 128 |
| 6.2.5 | LAH sample and reduced statistics | 129 |
| 6.3 | Results | 130 |
| 6.3.1 | Radial profiles | 133 |
| 6.3.2 | Comparison of TNG50 and MUSE data | 136 |
| 6.3.3 | The origin and source of Lyman-alpha halo photons | 139 |
| 6.3.4 | Lyman-alpha halo sizes | 141 |
| 6.3.5 | LAH central brightness | 144 |
| 6.4 | Discussion | 145 |
| 6.4.1 | The shape and nature of Lyman-alpha halos | 145 |
| 6.4.2 | Current limitations and future outlook | 149 |
| 6.5 | Conclusions | 153 |
| 6.6 | Supplementary material | 156 |
| 6.6.1 | Modeling assumptions | 156 |
| 7 | Outlook and conclusions | 169 |
| 7.1 | Outlook | 169 |
| 7.1.1 | Follow-up on previous work | 169 |
| 7.1.2 | New applications of the radiative transfer code | 171 |
| 7.1.3 | Observational advances | 176 |
| 7.2 | Conclusions | 177 |
| | References | 179 |

LIST OF FIGURES

| | | |
|-----|--|----|
| 1.1 | Ly α emission as a multi-scale phenomenon. | 3 |
| 1.2 | Topics of this Ph.D. thesis. | 9 |
| 2.1 | Wave functions for the lowest principal and angular quantum numbers n and l | 13 |
| 2.2 | Ly α line profile | 21 |
| 2.3 | Ly α emission in primordial gas from collisional excitations and recombinations. | 23 |
| 2.4 | Ly α luminosity as a function of star-formation rate. | 25 |
| 2.5 | Frequency redistribution $R(x_{\text{out}} x_{\text{in}})$ by scattering. | 35 |
| 2.6 | Random walk and Ly α photon trajectory. | 36 |
| 2.7 | Illustration of Ly α RT for different densities and resulting spectra. | 41 |
| 2.8 | Illustration of Ly α RT for different velocity gradients and resulting spectra. | 42 |
| 2.9 | Illustration of Ly α RT for different dust structures and resulting spectra. | 44 |
| 3.1 | Voronoi tessellation and procedure. | 55 |
| 3.2 | Testing the Voronoi geometry with Ly α radiative transfer. | 56 |
| 3.3 | Spectrum in spherical setup. | 57 |
| 3.4 | Multithreaded performance. | 57 |
| 3.5 | Domain decomposition. | 58 |
| 3.6 | Neutral hydrogen fraction as a function of density and temperature | 63 |
| 4.1 | Lyman- α intensity map of the simulated volume. | 72 |
| 4.2 | Spectra of simulated LAEs. | 78 |
| 4.3 | Stack of simulated LAE spectra. | 79 |
| 4.4 | Contour plot of emitters' peculiar and radiative velocities. | 80 |
| 4.5 | RT velocity offset $v_{\text{RT,global}}/v_{\text{RT,red}}$ distributions for different n_{LAE} and z | 81 |
| 4.6 | TPCF for disentangled RSD of visible LAEs for $n_{\text{LAE}} = 10^{-2} \text{ h}^3\text{Mpc}^{-3}$ in redshift space. | 82 |

| | | |
|------|---|-----|
| 4.7 | The pairwise velocity distribution P for the peculiar and radiative transfer velocity component at $z = 3$ | 85 |
| 4.8 | The mean, standard deviation and kurtosis of the pairwise velocity distribution P for the peculiar and radiative transfer velocity component. | 86 |
| 4.9 | Damping factor as a function of the line-of-sight frequency $k_{\parallel} = k\mu$ due to Ly α RT velocity offset $v_{\text{RT,global}}$ | 87 |
| 4.10 | Damping factor as a function of the line-of-sight frequency $k_{\parallel} = k\mu$ due to Ly α RT velocity offset $v_{\text{RT,red}}$ | 89 |
| 4.11 | Sketch of the proposed RSD correction methods. | 94 |
| 4.12 | Scatter plot of $v_{\text{RT,red}}$ and $v_{\text{proxy}}^{\text{m2a}}$ for emitters detected at number density threshold $n_{\text{LAE}} = 10^{-2} \text{ h}^3 \text{ Mpc}^{-3}$ and redshift $z = 3.01$ | 95 |
| 4.13 | Radiative velocity distribution after different correction schemes at a number density threshold of $n_{\text{LAE}} = 10^{-2} \text{ h}^3 \text{ Mpc}^{-3}$ and redshift $z = 3.01$ | 96 |
| 4.14 | The radiative velocity distribution v_{RT} for varying spectral resolution in terms of $R = c/\Delta v$ at $z=3.01$ for all LAEs (no n_{LAE} restriction). | 99 |
| 4.15 | The radiative velocity distributions v_{RT} and stacked spectra for varying aperture sizes at $z = 3.01$ for all LAEs (no n_{LAE} restriction). | 99 |
| 4.16 | Example of refined detection of two individual emitters with given halo ID at $z=3.01$ | 100 |
| 4.17 | The halo mass function of detected LAEs at $z = 3.01$ | 101 |
| 4.18 | The radiative velocity distribution for varying halo mass range and underlying hydrodynamical resolution at redshift $z = 5.85$ | 102 |
| 4.19 | Damping factor as a function of the line-of-sight frequency $k_{\parallel} = k\mu$ due to Lyman-alpha radiative transfer for $z=3.01$ | 103 |
| 5.1 | Two different input (i.e., intrinsic) spectra (black lines) and how the IGM attenuation shapes the observed spectra along different LOS at $z = 3.0$ | 109 |
| 5.2 | PDF of the transmission, T , as a function input wavelength shift, $\Delta\lambda_e$, all emitters and the LOS at $z = 3$ | 110 |
| 5.3 | Evolution of detectable Ly α peaks assuming different intrinsic spectra. | 112 |
| 5.4 | Ratios L_{ratio} and F_{ratio} as a PDF for all lines of sight at a given redshift as shown by the colormap. | 113 |
| 5.5 | L_{ratio} and N_{peaks} statistics shown as a function of the large scale overdensity δ_{LSS} at redshift $z = 4.0$ | 114 |
| 5.6 | Impact of the spectral resolution on the PDFs of the transmission function T as a function of input wavelength shift $\Delta\lambda_e$ | 116 |
| 5.7 | Impact of the lower integration bound s_0 for mean transmission curves. | 118 |

| | | |
|------|--|-----|
| 5.8 | Distribution of transmission T at fixed input wavelength offset $\Delta\lambda_e$ at $z = 3.0$ across all emitters and LOS. | 118 |
| 6.1 | $\text{Ly}\alpha$ surface brightness map for the entire TNG50 cosmological simulation at $z = 3$, highlighting the large-scale structure of the cosmic web as seen in $\text{Ly}\alpha$ emission. | 131 |
| 6.2 | Two-dimensional $\text{Ly}\alpha$ surface brightness maps of nine Lyman- α halos at $z = 3.0$ in TNG50. | 132 |
| 6.3 | Radial surface brightness profiles of nine individual Lyman- α halos at $z = 3$ in TNG50. | 133 |
| 6.4 | Overview of the predicted median $\text{Ly}\alpha$ surface brightness profiles around TNG50 galaxies, as a function of mass and redshift. | 135 |
| 6.5 | Gallery of observed Lyman- α halos from from the MUSE UDF chosen as the 24 closest to redshift $z = 3$ | 138 |
| 6.6 | Gallery of observed Lyman- α halos from the MUSE UDF with best matching simulated profiles within 2 arcsec aperture. | 139 |
| 6.7 | Median stacked radial $\text{Ly}\alpha$ profile for galaxies with stellar masses $8.5 \leq \log_{10}(M_*/M_\odot) \leq 9.5$ at $z = 3$ in TNG50. | 140 |
| 6.8 | Median stacked $\text{Ly}\alpha$ surface brightness radial profiles at $z = 3$ for halos with $8.5 \leq \log_{10}(M_*/M_\odot) \leq 9.5$ decomposed into different emission sources. | 141 |
| 6.9 | The $\text{Ly}\alpha$ half-light radius $r_{1/2}$ as a function of stellar mass over redshift in TNG50. | 141 |
| 6.10 | The $\text{Ly}\alpha$ single exponential radius r_0 as a function of stellar mass and redshift. | 143 |
| 6.11 | The central surface brightness SB_0 as a function of stellar mass across the studied redshift range in TNG50. | 144 |
| 6.12 | Radial profiles decomposed into emission origin and extending to large-scales of 3000 pkpc radii in TNG50. | 147 |
| 6.13 | Median stacked $\text{Ly}\alpha$ surface brightness radial profiles at $z = 3$ at different resolution runs of TNG50. | 156 |
| 6.14 | Median stacked $\text{Ly}\alpha$ surface brightness radial profiles (solid lines) at $z = 3$ in TNG50 for a fixed stellar mass $8.5 \leq \log_{10}(M_*/M_\odot) \leq 9.5$ for varying Monte Carlo photon packages per star-forming cell. | 157 |
| 6.15 | Median stacked $\text{Ly}\alpha$ surface brightness radial profiles (solid lines) at $z = 3$ in TNG50 for a fixed halo mass bin of $\log(M_h/M_\odot) \in [10.5, 11.0]$ varying the spectral offset of injected $\text{Ly}\alpha$ photons. | 158 |

| | | |
|------|---|-----|
| 6.16 | The mean stacked Ly α flux spectra at $z = 3$ in TNG50 for a varied input spectrum. | 159 |
| 6.17 | Median stacked Ly α surface brightness radial profiles in various stellar mass bins at $z = 3$ in TNG50 with and without dust model. | 160 |
| 6.18 | Radial profiles of LAHs split by potential AGN activity. | 161 |
| 6.19 | Median stacked Ly α surface brightness radial profiles from intrinsic emission and scattered photons decomposed into different emission sources and into subsamples with and without AGN. | 163 |
| 6.20 | Median UV luminosity escaping the galaxy at a given stellar mass in TNG50. | 165 |
| 6.21 | Median stacked radial Ly α profile for galaxies with stellar masses $8.5 \leq \log_{10}(M_*/M_\odot) \leq 9.5$ at $z = 3$ in TNG50 for varying line-of-sight integration methods and depths. | 166 |
| 7.1 | Spectrum and IGM transmission of a simulated Ly α emitting galaxy in TNG100 at $z = 3$ | 170 |
| 7.2 | Correlating Ly α forest and LAEs. | 172 |
| 7.3 | Surface brightness maps for the MgII emission from the CGM of three different halos with stellar mass $M_* \approx 10^{10} M_\odot$ at $z = 0.7$ in TNG50. | 173 |
| 7.4 | The Ly α cosmic web as a radiative transfer simulation on top of TNG50 at $z = 3$ | 175 |
| 7.5 | Visualization of some recent and future instruments and telescopes for Ly α observations. | 176 |

LIST OF TABLES

| | | |
|-----|--|----|
| 3.1 | Simulations used by respective chapter/publication. | 64 |
| 4.1 | Key quantities of the redshift snapshots considered for the clustering analysis. | 73 |

UNITS

We primarily use the centimeter-gram-second system of metric units in the electrostatic units (CGS-ESU). We will deviate where astrophysical observations commonly differ, e.g. with wavelengths given in Angstrom. For distances, we express lengths as parsecs on astrophysical scales. A prefix “p” or “c” is commonly added to distances to explicitly specify physical and comoving coordinates respectively.

INTRODUCTION

1.1 Motivation

Starting from tiny density fluctuations nearly 14 billion years ago, which we observe through the temperature fluctuations of the cosmic microwave background, the Universe evolved into its cosmic structure we see today. Observations of nearby galaxies give us an impression of the complex physical processes shaping them. Given the finite speed of light, we can see the Universe's past when looking at distant galaxies. This allows us to observe the matter structure of the past and open an opportunity to study young galaxies as they accrete cold gas by cooling and gravitational infall, eject gas through feedback processes, enter and cease periods of intense star-formation, ionize and heat their surrounding, and potentially merge with each other. Hence, being able to observe these young galaxies helps us to understand the formative processes giving rise to the complex, rich set of galaxies that evolved from the small perturbations after the Big Bang.

Hydrogen is the most abundant element in the Universe and contributes most baryonic matter in both galaxies and the cosmic web within which galaxies reside. By measuring the emission lines of hydrogen, we can thus trace the structure of the Universe. The Lyman-alpha ($\text{Ly}\alpha$) emission line describes the transition from the hydrogen's first excited state to its ground state and is often the brightest emission line available in young star-forming galaxies. The $\text{Ly}\alpha$ line is frequently so bright that high-redshift galaxies ($z \gtrsim 2$) more than 10 billion years away can be solely detected through this emission line. Additionally, for $z \geq 2$ the cosmological expansion redshifts the ultraviolet $\text{Ly}\alpha$ line (1215.67 Å) to wavelengths in an atmospheric window reaching the Earth's surface unhindered. Its brightness and wavelength allow efficient and affordable surveys for the $\text{Ly}\alpha$ line with ground-based instruments. For example, the Hobby-Eberly Telescope Dark Energy Experiment (HETDEX) will map out next to a million galaxies between redshifts 1.8 and 3.5 by their $\text{Ly}\alpha$ emission over the next years. One of this survey's applications will

be to constrain the equation of state of dark energy, which characterizes the cosmological evolution of the component responsible for the recent accelerated expansion of our Universe, the dark energy (Hill *et al.*, 2008). Towards higher redshifts, beyond $z \gtrsim 6$, the Universe becomes increasingly neutral as much of the ultraviolet flux from stars and active galactic nuclei that eventually reionize their surrounding does not exist yet. This epoch of reionization and the earliest galaxy therein can be traced by their Ly α emission and absorption with upcoming instruments such as on the James Webb Space Telescope (Finkelstein *et al.*, 2019).

Notably, the Ly α emission is not just detectable from the galaxies themselves at these redshifts, but even from the diffuse gas surrounding them. This gas, the circumgalactic medium (CGM), is crucial in the galaxies' evolution as a gas reservoir for their star-formation. The CGM's kinematics and gas state imprint the interplay between galaxy and their surrounding through accretion and feedback processes (Tumlinson *et al.*, 2017). Being able to observe the CGM in Ly α emission in young galaxies thus has an immense potential to improve our understanding of galaxy evolution.

Fascinatingly, Ly α emission does not cease to be detected in the CGM, but radial profiles around star-forming galaxies show emission even at large radii with the profiles flattening out beyond the virial radius of the hosting halos (Wisotzki *et al.*, 2018). In fact, it appears that we are able to observe parts of the cosmic web, the overarching structure of filaments within which most galaxies reside, in Ly α emission (Umehata *et al.*, 2019; Bacon *et al.*, 2021).

Ultimately, we want to link the Ly α observations back to the underlying state of the gas. However, the Ly α observations' theoretical interpretation starts lagging behind, while the amount and quality of observational data continuously increase. Even small amounts of neutral hydrogen lead to a significant optical depth, making the medium opaque to Ly α photons. As a consequence, Ly α photons are quickly absorbed, but also re-emitted, leading to a series of scatterings before escaping their hosting structure. These scatterings cause changes in the spatial and spectral distribution of Ly α photons, a process that we call Ly α radiative transfer (RT). Ly α RT makes inferences from observations difficult: the position does not coincide with the Ly α photon's source (but rather its point of last scattering) and the frequency does not just reflect the velocity structure encoded by the Doppler shift but also the diffusion of Ly α photons in frequency space. Complicating things further, the Ly α photons emitted within galaxies are subject to the RT on multiple scales: initially shaped by the interstellar medium sourced through recombinations of ionized hydrogen, a fraction of the Ly α photons escape the galaxies into the CGM, illuminating it and upon experiencing more scatterings escape the halo within which they originated. From there, Ly α photons can be further scattered out of

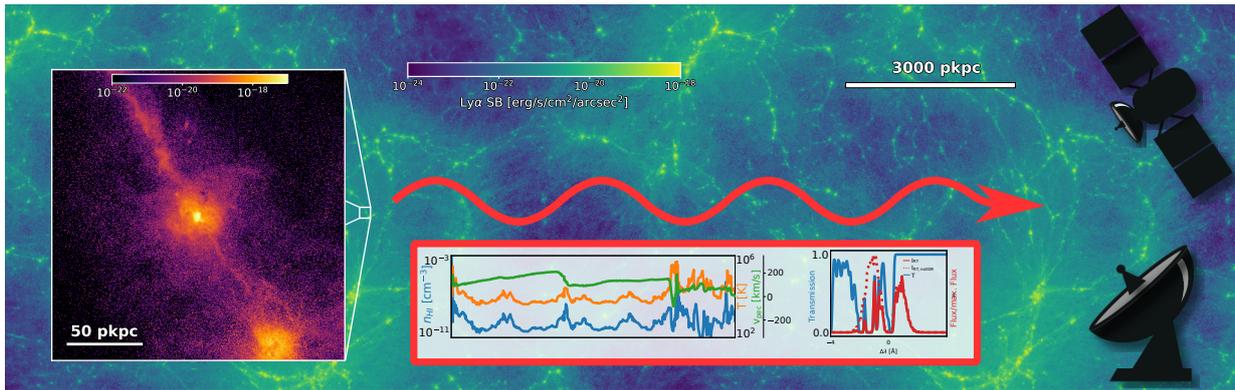


Fig. 1.1 $\text{Ly}\alpha$ emission as a multi-scale phenomenon. In the background, the $\text{Ly}\alpha$ surface brightness in a slice of $5.25 \text{ Mpc}/h$ at $z = 3.0$ predicted in the TNG50 simulation with the radiative transfer code described in this thesis. The large-scale structure is not just traced by $\text{Ly}\alpha$ emitters as a biased tracer, but also the diffuse emission/scatter of the filaments themselves. Around individual galaxies, an extended glow of $\text{Ly}\alpha$ can be observed and used to study the circumgalactic medium, see inset on the right. The shape of the $\text{Ly}\alpha$ spectra is strongly affected by the intervening intergalactic medium, see lower central inset. Thus, $\text{Ly}\alpha$ is a challenging but promising multi-scale tracer of the Universe's structure. Many ground-based instruments, such as KCWI, MUSE-VLT, Subaru-HSC and HET-VIRUS, provide us with the observational counterpart. In space, HST provides $\text{Ly}\alpha$ observations in the local Universe, while at high redshifts JWST will provide insights well into the epoch of reionization.

the line of sight on their way towards us. Figure 1.1 shows the $\text{Ly}\alpha$ emission and radiative transfer performed on the resolved scales from the scale of individual galaxies and to the large-scale structure.

To bridge the gap between the increasing amount of $\text{Ly}\alpha$ observations available and the latest theoretical and computational advancements in cosmological galaxy formation simulations, modeling of the complex $\text{Ly}\alpha$ radiative transfer is required. In this thesis, I present such radiative transfer framework. This framework allows us to compare current and future observational $\text{Ly}\alpha$ signatures with our current theoretical and computational understanding of structure formation. I subsequently apply this framework to state-of-the-art hydrodynamical galaxy formation simulations of cosmological volumes.

In the remainder of this Chapter, we are going to give a brief summary of the last decades' observational progress on $\text{Ly}\alpha$ observations in Section 1.2. Within this observational context, we sketch the theoretical challenges arising and how to address them in Section 1.3, which motivates the structure of this thesis outlined in Section 1.4.

1.2 Observations

The last decades brought an immense increase in observations of Ly α related phenomena. In 1967, Ly α emission from galaxies in the high redshift universe was only theorized (Partridge *et al.*, 1967). Since then, an increasing amount of Ly α emitting objects has been observed. Not only there was an increase in the amount and types of objects detected, but the quality of observations also rapidly improved with technological advances.

In the following, we introduce observations of the two most relevant groups of Ly α objects for this thesis: Ly α emitters and Ly α halos. For each group, we sketch out their observational characterization and progress of understanding, which will give us a better grasp of the open theoretical challenges we touch upon in this thesis.

1.2.1 Lyman-alpha emitters

Lyman- α emitters (LAEs) are a specific observational class of Lyman- α emitting galaxies that are detected through their Ly α emission. Typically this implies a large Ly α equivalent width (EW) relative to the continuum. With 1215.67 Å, Ly α is largely absorbed in the atmosphere and thus itself unavailable for ground-based observations at rest-frame wavelength. However, the cosmological expansion shifts its wavelength into the visible/near-IR window at redshifts $z \gtrsim 2$. Given their large luminosities, LAEs can thus be observed even at very high redshifts with ground-based instruments. For this thesis we adopt the most common definition that classifies those objects as LAEs that have been detected through the Ly α emission line with a rest-frame Ly α EW of $EW_0 \geq 20 \text{ \AA}$ (Ouchi, 2019).

Only in the late 90s, observations of Ly α emitters became frequent¹, starting with Hu *et al.* (1996), after earlier surveys remained unsuccessful in detecting LAEs given initially overestimated luminosities from Partridge *et al.* (1967). While early searches strategically targeted overdense regions in the proximity of active galactic nuclei (AGN), LAEs are nowadays blindly searched for and found by surveying large sky areas. To date, $\mathcal{O}(10^4)$ LAEs have been found, often with spectroscopic confirmation, and at the time of writing, this number is rapidly growing.

We will present the two major techniques used to find LAEs, narrowband imaging and blind spectroscopy, in the following.

For narrowband imaging, the flux for a targeted field is captured (and hence integrated) over a filter with a central wavelength λ_c and a full width at half maximum

¹Also, see Barnes *et al.* (2014) and Behrens (2014) for a compiled list of LAE detections from 1996 to 2012.

(FWHM) of $\Delta\lambda$. Typically λ_c is chosen accordingly to the targeted redshift $z = \lambda_c/\lambda_{\text{Ly}\alpha} - 1$ with the width $\Delta\lambda$ typically in the order of $\sim 100 \text{ \AA}$. At the high redshifts ($z \gtrsim 2$) that are surveyed for galaxies utilizing the Ly α line with ground-based instruments, narrowband observations allow efficient detection of LAEs given the line's high equivalent width. Even if no continuum is detected, a galaxy can be pinpointed by the Ly α emission line alone that dominates the flux in a given narrowband filter. The redshift of the LAE is then roughly set to the redshift that the narrowband filter is targeting. While there is potential for confusion with low redshift interloping lines, this method is robust, which can be verified by subsequently doing a spectroscopic follow-up for a subsample.

With blind spectroscopy, using slitless spectroscopy and integral field spectroscopy (IFS), spectra for a given source can be obtained in addition to their position. As the luminosity of the Ly α line is often uncontested by other emission lines in many young star-forming galaxies, LAEs can be efficiently detected by the presence of such bright emission line in spectra. Low-redshift galaxies can have equally large fluxes from other emission lines being misinterpreted as Ly α emission. Secondary emission lines can easily distinguish those low redshift galaxies from high redshift ones. However, as Ly α is the brightest line at high redshifts and surveys are designed to efficiently detect LAEs over large areas, if no other emission line is available, classification as an LAE or non-LAE needs to be done solely based on the detected emission line. The rest-frame equivalent width threshold of 20 \AA appears to be a decent discriminator to identify the Ly α line of high-redshift LAEs from low-redshift interlopers (Adams *et al.*, 2011; Leung *et al.*, 2017).

Early observations used narrowband imaging to detect larger samples of LAEs on various $\geq 5 \text{ m}$ telescopes (Hu *et al.*, 1998; Steidel *et al.*, 2000; Ouchi *et al.*, 2003; Hayashino *et al.*, 2004). While narrowband imaging remains an important tool to detect LAEs and observe the high-redshift Universe (Yamada *et al.*, 2012; Konno *et al.*, 2018), recent and ongoing surveys with integral field spectrographs allow a more detailed view of Ly α emitters such as with the Hubble Ultra Deep Field Survey with the Multi Unit Spectroscopic Explorer (MUSE-UDF; Bacon *et al.*, 2017) on the Very Large Telescope (VLT) and the Dark Energy Experiment on the Hobby Eberly Telescope (HETDEX; Hill *et al.*, 2008). Ly α observations of local galaxies are possible from space. Such observations have been performed with the Galaxy Evolution Explorer (GALEX) and Hubble Space Telescope (HST), allowing a more detailed study of potential analogs to those Ly α emitters at high redshifts (Cowie *et al.*, 2011; Hayes *et al.*, 2013; Henry *et al.*, 2015). At the same time, it remains debated to what extent these represent high-redshift LAEs.

Ly α emitters have since been characterized by analysis of their spectra. Using the continuum shape and nebular emission lines, if available, constraints on LAEs' typical properties can be derived. Ly α emitters typically have a stellar mass of 10^7 - $10^{10} M_{\odot}$, color

excess due to stellar extinction $E(B - V)_s$ of 0.0 to 0.2, star formation rate $1-100 M_\odot \text{yr}^{-1}$, stellar age 1-100 Myr and a relatively low metallicity $0.1-0.5 Z_\odot$ (Ouchi, 2019).

Apart from the interest in Ly α emitters themselves as astrophysical objects, they provide an opportunity to trace the large-scale structure and galaxy evolution over a large range of redshifts. Two of the common statistical measures are the luminosity function and the power spectrum.

The luminosity function (LF), describing the number density in a given luminosity interval, is one such tracer encoding information about the abundance of galaxies, their star-formation, the subsequent escape of Ly α emission, and the evolution of the intergalactic medium's ionization state. The Ly α LF has been measured in the local Universe up to the end of the epoch of reionization ($z \sim 0.3 - 7.3$; Ouchi *et al.*, 2008; Cowie *et al.*, 2010; Blanc *et al.*, 2011; Barger *et al.*, 2012; Konno *et al.*, 2014, 2018; Spinoso *et al.*, 2020). Its evolution substantially differs from that of the UV luminosity function. Looking at the respective luminosity densities from Ly α and UV, given as the luminosity weighted integral over the LFs, the luminosity rises significantly more rapidly for Ly α up to redshift $z \lesssim 3$. From there, at cosmic noon, where star-formation and the UV luminosity density peak, the UV luminosity density decreases while the Ly α luminosity function remains constant within its errorbars (Ouchi, 2019). Only for LF constraints at $z \sim 7.3$, we note an enormous drop in the Ly α luminosity function as an indicator of the rapidly increasing neutral hydrogen densities into the epoch of reionization (Konno *et al.*, 2014).

Clustering analysis of LAEs allows testing the expected structure formation paradigm and the underlying cosmological model. The clustering is commonly described by the LAEs' two-point correlation function/power spectrum. The HETDEX survey will be the largest blind LAE survey yet, spectroscopically mapping roughly a million LAEs between $z \sim 1.8$ and 3.5. As a biased tracer of the halo population and the underlying matter distribution, we can constrain the growth of structure and the cosmological model with detected LAEs. At high redshifts, LAEs are the most promising galaxy tracer given their brightness and abundance. Other suitable object types exist, such as Lyman break galaxies (LBGs) detected by their drop in rest-frame flux below $\lambda < 912 \text{ \AA}$, but require a spectroscopic follow-up for their redshift determination. (Hill *et al.*, 2004)

1.2.2 Lyman-alpha halos and blobs

Ly α emitters are commonly very compact with effective radii $r_e \sim 1 \text{ pkpc}$ (Paulino-Afonso *et al.*, 2018) and hence commonly below the resolution of modern LAE surveys². Nevertheless, we can find structures that are significantly more extended in deeper ob-

²For example, HETDEX has point spread function of around 1.2 arcseconds (Niemeyer, 2021).

servations. Spatially extended Ly α emission has been detected since the 1980s for the brightest objects with luminosities of 10^{44} to 10^{45} erg/s in the proximity of AGN (McCarthy *et al.*, 1987; Heckman *et al.*, 1991b) with a spatial extend of $\mathcal{O}(100$ pkpc). These objects, so-called Ly α nebulae, are the largest instances of Ly α blobs. Ly α blobs (LABs), defined by their extended Ly α emission on scales between $\sim 10 - 100$ pkpc, have since been observed in sizable numbers between redshift $\sim 2 - 7$ with narrowband imaging, slit spectroscopy, and integral field spectroscopy (see e.g. Heckman *et al.*, 1991a; Steidel *et al.*, 2000; Ouchi *et al.*, 2009; Borisova *et al.*, 2016).

In recent years, fainter extended Ly α emission around star-forming galaxies can be detected through narrowband image stacking and individually through integral field spectroscopy with extents of $\mathcal{O}(10$ pkpc). These objects are smaller than the larger LABs but surround a significantly larger population of galaxies and are potentially ubiquitous for deep observations of LAEs. Similar to its use for LAE detection, narrowband imaging allowed the detection of LAHs at targeted redshifts (Hayashino *et al.*, 2004; Steidel *et al.*, 2011; Matsuda *et al.*, 2012; Feldmeier *et al.*, 2013) and in substantial counts for narrowband surveys (Momose *et al.*, 2014, 2016; Kakuma *et al.*, 2021). While in latter surveys stacking is used and information about individual objects is hence lost, recent integral field spectroscopic surveys such as with the Multi Unit Spectroscopic Explorer (MUSE; Bacon *et al.*, 2010) and the Keck Cosmic Web Imager (KCWI; Morrissey *et al.*, 2018) allow individual identification and a more detailed characterization (Leclercq *et al.*, 2017).

IFS has huge potential to reveal additional information about the underlying circumgalactic medium's density structure and kinematics by utilizing the full information contained in the three-dimensional data cubes: two dimensions characterize the distribution perpendicular to the line of sight, while the third dimension tracks the spectral component that encodes the line of sight position, kinematics, and the complex Ly α radiative transfer. Separating the different encoded information pieces from the spectral component would allow for a tomographic view of the circumgalactic medium in those young distant galaxies. Only recently, this avenue of research picked up the pace with the availability and analysis of individual IFS data cubes (Leclercq *et al.*, 2020).

1.3 Tackled problems in this thesis

1.3.1 Completing the distortion effects in LAE clustering

As pointed out, LAEs are one of the most promising tracers of the high-redshift galaxies ($z \geq 2$) of our Universe. However, the two-point statistics of LAEs used to constrain the large-scale structure and the cosmological model are subject to additional redshift-

space distortions. In part, these additional distortions are caused due to a selection effect correlated with the large-scale structure. The attenuation of Ly α flux along the line of sight towards the observer could be correlated, e.g. with the line of sight velocity gradient. This would further modify the LAEs' bias as a tracer of the underlying matter field and add a non-isotropic distortion in the directionally split power spectrum (Wyithe *et al.*, 2011). Such effect has been the subject of multiple studies with cosmological simulations (Zheng *et al.*, 2011a; Behrens *et al.*, 2013, 2018; Gurung-López *et al.*, 2020) with varying inconsistent findings. The strength and presence of such effects thus remain open questions in the community.

Ly α radiative transfer can shift the observed LAE line-of-sight position from its systemic redshift, giving rise to yet another clustering distortion in addition to the above selection effects. For the first time, we study the clustering distortions from this spectral offset in the Ly α emission line in Byrohl *et al.* (2019), which we discuss in Chapter 4.

1.3.2 Separating the imprint of spatial scales in LAE spectra

The Ly α line of LAEs often shows complex spectral shapes from asymmetric single-peaked profiles to double-peaked and even being triple-peaked ones. These shapes result from the complex Ly α RT on the galaxies' scale and the attenuation from the intergalactic medium. While idealized shell models of neutral hydrogen can fit the observed Ly α spectra, simulations on top of hydrodynamic simulations of halos and galaxies commonly cannot reproduce the observational signatures. In particular, simulations overpredict the Ly α flux in the ("blue") peak at wavelengths lower than the line-center, potentially hinting at the lack of a volume-filling outflowing component on the scales of the galaxy (Mitchell *et al.*, 2021). An often invoked alternative explanation is the attenuation of those blue photons by the intergalactic medium as these photons are shifted into the Ly α line-center by the Hubble flow, reconciling simulations with observations. This apparent degeneracy of the Ly α spectrum's blue photons between the evolution of an outflowing gas component on the galaxies' scale and the attenuation by the IGM is important to be broken in order to evaluate the IGM's and galaxy evolution's respective redshift evolution.

In Chapter 5, we calculate the expected IGM attenuation while staying agnostic about the small-scales spectra by only proposing different generic mock spectra for those scales. Combining the small-scale spectra with the IGM attenuation, we calculate mocks of observed Ly α spectra. Using different simple statistical measures, we show how this degeneracy between different spatial scales on the spectral signature can be broken (Byrohl *et al.*, 2020b). We make our data sets of millions of Ly α attenuation curves publicly available to allow incorporation into zoom-in Ly α radiative transfer simulations

lacking IGM interaction and to enable more sophisticated exploration (Byrohl *et al.*, 2020a).

1.3.3 Demystifying Lyman-alpha halos

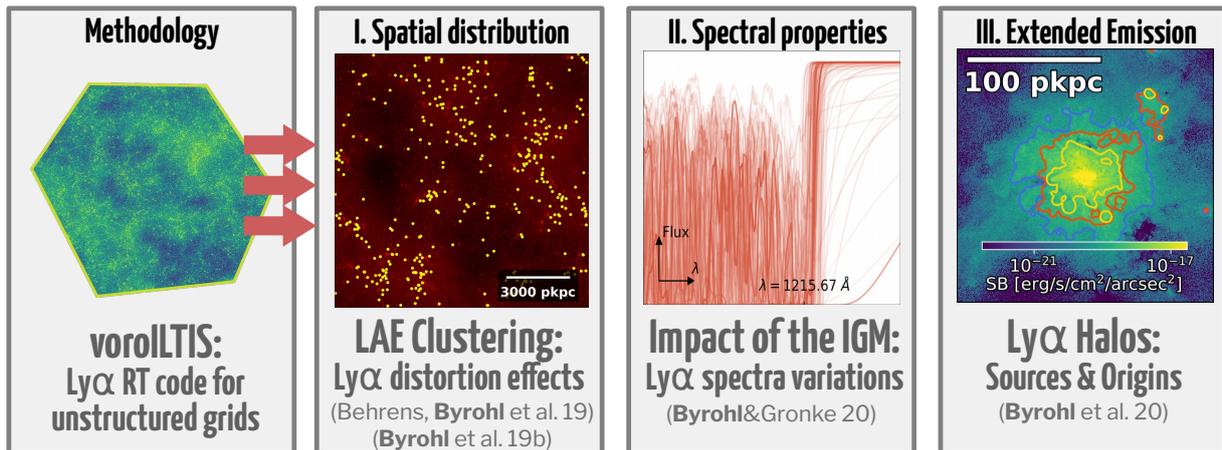


Fig. 1.2 In this Ph.D. thesis, I will present a Ly α radiative transfer code applied to the IllustrisTNG simulations (**leftmost panel**) and published results in an astrophysical and cosmological context (**other panels**).

As shown in Section 1.2.2, Ly α halos are frequently observed with narrowband imaging and integral field spectroscopy. Their surface brightness profiles often extend onto scales of the hosting halo's virial radius, making the circumgalactic gas around its galaxy visible even at high redshifts – more than 10 billion years away. This opens up huge potential for constraining and understanding an important epoch of galaxy formation through these young star-forming galaxies: can we trace the cold accretion streams providing gas to these galaxies for star-formation? When and for which masses do feedback processes disrupt inflows and halt star-formation? To answer any such question with Ly α halo observations, we need to demystify Ly α halos. In a first step, we need to establish if and how we can link and reproduce observations in theory and simulations. If this is possible, we can try to answer the most fundamental question about Ly α halos: what causes them? We were able to explain these halos by scattered photons from star-forming regions that we model on top of state-of-the-art hydrodynamical galaxy formation simulations in Byrohl *et al.* (2021). We present these results in Chapter 6.

1.4 Structure of this thesis

This thesis reflects my published work on Ly α radiative transfer simulations, which resulted in various publications on Ly α emission and reprocessing using a computational approach. Other than this underlying common denominator, the topics range from Ly α emitters' clustering in cosmology (Behrens *et al.*, 2018; Byrohl *et al.*, 2019) over spectral shapes shaped by the intergalactic medium (Byrohl *et al.*, 2020b) to extended Ly α halos around galaxies (Byrohl *et al.*, 2021). While these publications have been reworked for their own chapters in this dissertation, additional chapters that summarize the underlying context and methodology are introduced first.

In Chapter 2, we dive into the theoretical understanding of Ly α radiative transfer and interpretation of Ly α observations. In Chapter 3, we introduce the methods underlying the dissertation's research. This comprises of a short review of the implementation of a radiative transfer code for the Ly α line and application to state-of-the-art cosmological hydrodynamical galaxy formation simulations. Next, we present various applications of this methodology that resulted in first-author publications over the course of the Ph.D. Chapter 4 investigates the impact of the spectral shape of Ly α emitters on the clustering statistics. Chapter 5 examines how Ly α spectral shapes are affected by neutral hydrogen in the IGM, and Chapter 6 investigates the nature of Ly α halos using the TNG50 hydrodynamical simulations and the Ly α radiative transfer code. In Chapter 7, we give an outlook on related promising work based on the thesis' findings and summarize the presented work.

THE LYMAN-ALPHA LINE

2.1 The hydrogen atom and the Ly α transition

The neutral hydrogen atom, consisting of a proton as its nucleus and an electron, has a range of discrete energy states, which were first inferred from the atom's spectral lines. The visible spectral hydrogen lines were first mathematically related by Johann Balmer and later generalized by Johannes Rydberg (Balmer, 1885; Rydberg, 1890) as

$$\frac{1}{\lambda} = R \left(\frac{1}{n_1^2} - \frac{1}{n_2^2} \right) \quad (2.1)$$

where R is an element specific constant with $R_H = 109732.32 \text{ cm}^{-1}$ for hydrogen, $n_1 > 1$ is an emission line series' defining index and different indices n_2 with $n_2 > n_1$ are the series' members. Hydrogen's visible lines that Balmer found were given by $n_1 = 2$. Later, the ultraviolet Lyman series with $n_1 = 1$, named after Theodore Lyman first observing it in 1906, was discovered (Lyman, 1906). The Lyman- α transition ($n_2 = 2$) at the heart of this thesis is the least energetic emission line of the series with an energy/frequency/wavelength of

$$E_{\text{Ly}\alpha} = 10.2 \text{ eV}, \quad \lambda_{\text{Ly}\alpha} = 1215.67 \text{ \AA}, \quad \nu_{\text{Ly}\alpha} = 2.47 \cdot 10^{15} \text{ Hz}. \quad (2.2)$$

Rydberg's empirical relationship was soon explained by a theoretical model proposed by Niels Bohr. In this "Bohr model", electrons would orbit on circular orbits around the nucleus. Only discrete orbits at fixed energies are allowed and the frequency of emitted photons only depends on the energy of the initial and final state of the electron's orbit (Bohr, 1913). The Bohr model can be derived by assuming a quantized angular

momentum as

$$m_e v_e r_n = n\hbar \quad (2.3)$$

with m_e being the electron's mass, v_e being its velocity and r_n its radius for the quantized angular momentum $L = n\hbar$ described by the integer $n \geq 1$. $\hbar = \frac{h}{2\pi}$ is the reduced Planck constant. For a circular orbit, the electron experiences a constant Coloumb force from the central proton, such that the electron's motion is described by

$$\frac{m_e v_e^2}{r} = \frac{e^2}{r^2}. \quad (2.4)$$

Combining Equations (2.3) and (2.4), we find the allowed electron radii r_n to be

$$r_n = \frac{n^2 \hbar^2}{m_e e^2} = a_0 n^2, \quad (2.5)$$

where $a_0 \equiv \hbar^2/m_e e^2$ is the so-called Bohr radius. Further, the allowed energy levels are given by

$$E = \frac{m_e e^4}{2\hbar^2} \frac{1}{n^2}, \quad (2.6)$$

where m_e is the electron mass and e the electron charge. Hence, the ionization energy of hydrogen is 13.6 eV from the ground state ($n = 1$). Calculating the energy difference between to states n_1 and n_2 and converting this difference to an inverse wavelength as $1/\lambda = \Delta E/(h\nu)$, we arrive at the Rydberg formula (Eqn. (2.1)) including the correct Rydberg constant for hydrogen.

2.2 A quantum mechanical treatment of the hydrogen atom

Building on Bohr's model of quantized electron orbits, the refined the Bohr-Sommerfeld model (Sommerfeld, 1916, 1923), which introduced two additional major quantum numbers l and m for additional quantized degrees for elliptic orbits and projected momentum along a specified axis, was able to explain additional observed spectral features from the fine structure of the atomic structure easily seen in presence of external electromagnetic fields, the Stark and Zeeman effect. Nevertheless, these models of quantized spatially well-defined orbits derived with classical mechanics were soon phased out in favor of modern quantum mechanics. With its advent, the description of physical states was expressed in terms of an underlying wavefunction Ψ defining the probability density

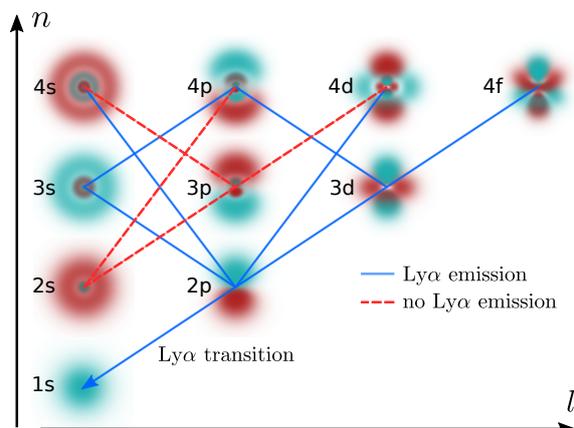


Fig. 2.1 Wave functions for the lowest principal and angular quantum numbers n and l . Only states with $m = 0$ are shown. In addition, the not forbidden transitions are shown. Those cascades leading to Ly α emission are shown in blue while cascades that end in the $2s$ state, not emitting Ly α . We do not show direct decay to the ground state for other lines of the Ly α series.

as its absolute square. Hence, this approach significantly differs from previous practise to assume well-defined, localized states: Instead of an atom being surrounded by one or multiple electrons on well-defined trajectories, the electrons are described by a wave function Ψ that sets a spatially extended probability for the electrons' position – not just as a measure of our limited knowledge of the underlying system, but as a measure of physical reality. The dynamics and states of a quantum mechanical system such as the energy states of an electron can be derived from the time-dependent Schrödinger equation

$$i\hbar \frac{\partial \Psi}{\partial t} = \hat{H} \Psi \quad (2.7)$$

with \hat{H} being the Hamiltonian operator closely related to the classical Hamiltonian function. Stationary solutions to this equation satisfy the time-independent Schrödinger equation $\hat{H}\Psi = E\Psi$. For a single electron system, the Hamiltonian operator consists of two terms for the electron's kinetic energy and the Coloumb potential of the nucleus, such that

$$\hat{H}^0 = -\frac{\hbar}{2m} \nabla^2 - \frac{e^2}{r}, \quad (2.8)$$

where we applied the standard quantum mechanical prescription $\vec{p} \rightarrow -i\hbar\nabla$ after expressing the kinetic energy in terms of the momentum \vec{p} . For multi electron systems, another Coloumb potential for the interacting electrons would occur – which however is not the case for the hydrogen atom considered here. As the Hamiltonian has a spherical symmetry, the solution of the time-independent Schrödinger equation for this Hamiltonian has

the form

$$\Psi(r, \theta, \phi) \propto R(r)Y(\theta, \phi) \quad (2.9)$$

with the solution for the radial component R depending on the two quantum numbers n and l with its energy state determined by n . The angular component is described by spherical harmonics, which are the eigenfunctions of the orbital angular momentum operator $\vec{L} = \vec{r} \times \vec{p}$

$$\vec{L}^2 Y_{lm} = l(l+1)Y_{lm}, \quad L_z Y_{lm} = mY_{lm}, \quad (2.10)$$

with the right hand sides measuring the squared total angular momentum and the angular momentum along the z-axis respectively. These two quantum numbers can take all integers sufficing

$$0 \leq l \leq n-1 \quad \text{and} \quad |m| \leq l. \quad (2.11)$$

We visualize the wavefunctions for the lowest n states in Figure 2.1. The functional forms for the radial and angular components can e.g. be found in Griffiths *et al.* (2018).

While above Hamiltonian leads to the state's energy being solely determined by the principal quantum number n , the degeneracy in l is broken when incorporating the Hamiltonian's perturbation by using a correction for the relativistic kinetic energy and spin-orbit coupling of the electron's magnetic moment with the proton's Coloumb field. For $l = 0$, we additionally have a perturbation due to the Coloumb potential being smeared out ("Darwin term"). These effects lead to the energy states n being split by an energy difference

$$\Delta E_{nj} = \frac{E_n \alpha^2}{n} \left(\frac{1}{j+1/2} - \frac{3}{4n} \right) \quad (2.12)$$

with $j \equiv |l \pm 1/2|$ representing the total angular momentum quantum number (Griffiths *et al.*, 2018).

Finally, electrons possess their own intrinsic momentum that can take the values $m_s \hbar$ with the quantum number $m_s = \pm 1/2$. This would help to explain the "anomalous" Zeeman effect unexplainable in the Bohr-Sommerfeld model. This leads to the so-called hyperfine structure that gives rise to the astrophysically and cosmologically promising 21cm line.

2.3 Photon absorption and emission

The absorption and emission of photons by the hydrogen atom can already be understood in a semi-classical theory of radiation, which uses a quantum mechanical treatment of the atom, while still adopting a classical description of the radiation. A proper description, particularly of the emission process, would require quantum electrodynamics (QED) not treated here. More details on the following summary can be found in Rybicki *et al.* (2004) and Griffiths *et al.* (2018).

We address this process by assuming a perturbation \hat{H}' , representing the incoming radiation. Adding this term to the hydrogen's time independent Hamiltonian \hat{H}^0 leads to

$$\hat{H} = \hat{H}^0 + \hat{H}'. \quad (2.13)$$

For absorption or emission within the hydrogen atom, we transition between two states $\Phi_1\rangle$ and $\Phi_2\rangle$ with their respective probabilities given by their absolute square amplitudes $|c_1|^2$ and $|c_2|^2$ with the wavefunction as their superposition

$$\Psi(t) = c_1 \Psi_1 \exp[-iE_1 t/\hbar] + c_2 \Psi_2 \exp[-iE_2 t/\hbar] \quad (2.14)$$

where we now allow for a time dependence of the amplitudes c_1 and c_2 . Using the Schrödinger equation (2.7) we can determine the amplitudes' evolution as

$$\dot{c}_1 = -\frac{i}{\hbar} H'_{1,2} \exp[-i\omega_0 t] c_2, \quad \dot{c}_2 = -\frac{i}{\hbar} H'_{2,1} \exp[i\omega_0 t] c_1 \quad (2.15)$$

with

$$H'_{i,j} \equiv \langle \Psi_i | \hat{H}' | \Psi_j \rangle \quad (2.16)$$

and $\omega_0 \equiv (E_2 - E_1)/\hbar$ for $E_b \geq E_a$. We now consider monochromatic radiation polarized along the z-axis:

$$\vec{E}(\vec{r}, t) = E_0 \cos(\vec{k} \cdot \vec{r} - \omega t) \hat{e}_z \quad (2.17)$$

If we assume a negligible spatial variation of the radiation field over the hydrogen atom's size, we can drop $\vec{k} \cdot \vec{r}$ from above equation, resulting the perturbing Hamiltonian to be

$$\hat{H}' = -eE_0 z \cos(\omega t). \quad (2.18)$$

This approximation represents electric dipole radiation and using a Taylor expansion of $\exp(-i\vec{k}\vec{r})$ allows to calculate higher order radiation such as the quadratic electric moment and dipole magnetic moment, which we neglect here.

Assuming $\omega_0 + \omega \gg |\omega_0 - \omega|$, which imposes frequencies ω close to that of the transition with ω_0 , we find

$$P_{1\rightarrow 2}(t) = \left(\frac{|\mathcal{P}| E_0}{\hbar} \right)^2 \frac{\sin^2 [(\omega_0 - \omega)t/2]}{(\omega_0 - \omega)^2} \quad (2.19)$$

representing the probability for absorption for a photon of energy $\hbar\omega_0$ where $\mathcal{P} \equiv e\langle\Psi_1|z\text{vecr}|\Psi_2\rangle$. Analogously, we obtain $P_{2\rightarrow 1}$ to be the same. Latter represents the case of stimulated emission, where the interacting photon leads to the emission of a secondary photon upon decay of the electron from state 2 to 1.

Generalizing to non-monochromatic radiation and arbitrary direction/polarization, we can determine the probability for absorption per unit time per unit energy density as

$$B_{12} = \frac{4\pi^2}{3\hbar^2} |\vec{\mathcal{P}}|^2, \quad (2.20)$$

where $\vec{\mathcal{P}} \equiv e\langle\Psi_1|\vec{r}|\Psi_2\rangle$.¹ The probability for stimulated emission B_{21} can be calculated analogously and we find $B_{12} = B_{21}$.

We cannot derive the spontaneous emission rate A within this framework: The energy states remain stable without external perturbation, which are present for a full treatment in QED. Instead, we start from detailed balance from the three potential processes of absorption, stimulated emission and spontaneous emission in thermal equilibrium such that

$$N_1 B_{12} \rho(\omega_0) = N_2 B_{21} \rho(\omega_0) + A_{21} \quad (2.21)$$

as considered by Einstein in 1916. $\rho(\omega_0)$ represents the energy density and N_1/N_2 represents the number of atoms in the states 1/2 respectively. B_{12} and B_{21} represent the absorption probability per unit time and unit energy for an atom as before. We now seek to determine the spontaneous emission rate A_{21} , representing the emission probability per unit time. Using Equation (2.21), determining the relative occupation of the states N_1/N_2 via the Boltzmann factor and comparing to Planck's blackbody formula, we obtain

¹Note that we define the probability B_{12} in terms per unit energy density rather than in terms of the mean intensity J_ν/J_ω , which lead to a factor $4\pi/c$ or $2/c$ respectively.

the following relations of the respective rates, called Einstein coefficients with

$$B_{12} = B_{21}, \quad A_{21} = \frac{\omega_0^3 \hbar}{\pi^2 c^3} B_{21}. \quad (2.22)$$

Note that the relation between the B coefficients only holds in our case of two non-degenerate energy states, otherwise the rates are multiplied by their respective degeneracies in above equation. We see that all rates, both for absorption and emission, immediately follow upon calculation of $|\vec{\mathcal{P}}|^2$.

2.4 Selection rules and radiative cascades

Given Equation (2.20) and (2.22), the determination of Einstein coefficients is primarily driven by the evaluation of the squared dipole moment operator $\vec{\mathcal{P}}$:

$$|\vec{\mathcal{P}}|^2 = e^2 |\langle \Psi_{n,l,m} | \vec{r} | \Psi_{n',l',m'} \rangle|^2 = e^2 \sum_{d \in \{x,y,z\}} |\langle \Psi_{n,l,m} | d | \Psi_{n',l',m'} \rangle|^2 \quad (2.23)$$

for the transition $(n, l, m) \rightarrow (n', l', m')$.

Importantly, many of the moments evaluate to zero, which can be generalized by the operators' properties. Using symmetry, we find the elements to evaluate to zero unless

$$\Delta l = l' - l = \pm 1, \quad (2.24)$$

$$\Delta m = m' - m \in \{-1, 0, 1\}. \quad (2.25)$$

We call transitions whose matrix element evaluates to zero forbidden. Nevertheless, these transitions can still occur, but are significantly more unlikely and outside of the approximations made here. Transitions are possible via multi-photon interaction or by incorporating spatial fluctuations that we started to neglect in Equation (2.18).

For the directional components, we find

$$\langle \Psi_{nlm} | x | \Psi_{n'l'm'} \rangle = \langle \Psi_{nlm} | y | \Psi_{n'l'm'} \rangle = 0 \quad \text{for } m' = m, \quad (2.26)$$

$$\langle \Psi_{nlm} | x | \Psi_{n'l'm'} \rangle = \pm i \langle \Psi_{nlm} | y | \Psi_{n'l'm'} \rangle \quad \text{for } m' = m \pm 1, \quad (2.27)$$

$$\langle \Psi_{nlm} | z | \Psi_{n'l'm'} \rangle = 0 \quad \text{for } m' = m \pm 1, \quad (2.28)$$

largely reducing the amount of elements to calculate.

This finally allows us to calculate the emission coefficient $A_{\text{Ly}\alpha}$ for the Ly α transition $(2, 1, m) \rightarrow (1, 0, m')$. Using rules from Equations (2.26), we only need to evaluate

$\langle \Psi_{1,0,0} | x | \Psi_{2,1,\pm 1} \rangle = \pm 2^7/3^5 a_0$ and $\langle \Psi_{1,0,0} | x | \Psi_{2,1,0} \rangle = 2^{7.5}/3^5 a_0$ from which we obtain

$$A_{\text{Ly}\alpha} = \frac{2^{17} \omega_0^3 e^2 a_0^2}{3^{11} \hbar c^3} \approx 6.27 \cdot 10^8 \text{ s}^{-1} \quad (2.29)$$

for all m (Berestetskii *et al.*, 1971; Fitzpatrick, 2008) in more than sufficient agreement with full quantum electrodynamical result for our purposes (Wiese *et al.*, 2009).

Often, instead of an Einstein A coefficient, the (absorption) oscillator strength f is given. It can be interpreted as fractional (but often in the order of unity) quantum mechanical correction for a transition approximated by a classical electron oscillator. f can be defined as

$$f_{lu} = \frac{g_u}{g_l} \frac{mc^3 \hbar^2}{2e^2 \Delta E^2} A_{ul} \quad (2.30)$$

with subscripts l and u for the respective lower and upper state. g denotes the respective state's degeneracy. With $g_u/g_l = 3$ for Ly α , we obtain $f_{\text{Ly}\alpha} = 0.416$.

As an important consequence of the selection rules in Equations (2.24)/(2.25), the transition $2s \rightarrow 1s$ is forbidden and hence is very unlikely to emit a Ly α photon. If we have an excited state with $n \geq 2$ it is thus uncertain whether a Ly α photon will be emitted because either the electron ends up in the $2s$ state or the electron de-excites from a higher energy level $n \geq 2$. In Figure 2.1, blue lines indicate the allowed transitions that result in the emission of a Ly α photon, opposed to those states ending up in the $2s$ state not emitting a Ly α photon via red colored transitions.

The probability for a decay into a state (n', l') from (n, l) is given by the ratio of the considered decay over the sum of all other transitions from (n, l) :

$$P(nl \rightarrow n'l') = \frac{A_{nl,n'l'}}{\sum_{n'',l''} A_{nl,n''l''}} \quad (2.31)$$

With this, we can formulate the successive decay of an electron in (n, l) via an intermediate state (n', l') to the ground state under emission of a Ly α photon as

$$P(n, l \rightarrow \text{Ly}\alpha) = \sum_{n',l'} P(nl \rightarrow n'l') P(n'l' \rightarrow \text{Ly}\alpha). \quad (2.32)$$

Using Equation (2.32), we can iteratively solve the probability for an arbitrary initial state (n, l) by starting from the $n = 3$ levels and intermediate levels $n' = 2$. As a result, we will know the probabilities $P(3l \rightarrow \text{Ly}\alpha)$. Together with the already known $P(2l \rightarrow \text{Ly}\alpha) = \delta_{1l}$, we can solve the for $n = 4$ and so on.

The cascading will be important for the Ly α emission from high-level excitation and recombination events of hydrogen in Section 2.6.1.

2.5 Line profile and cross-section

So far, we have only considered the eigenstates without considering potential deviations that broaden the transition upon emission. The broadening is captured by the line profile $\Phi(\nu)$, which we will now discuss for the Ly α line. The line profile is defined as the probability density function (PDF) as a function of frequency ν normalized to unity:

$$\int_0^{\infty} \Phi(\nu) d\nu = 1 \quad (2.33)$$

Typically three broadening mechanisms around the transition frequency ν_0 exist:

- **Natural broadening:** Intrinsic frequency dispersion of the transition motivated by the uncertainty principle.
- **Collisional broadening:** The natural decay from an excited state is modified by collisions, adding additional broadening.
- **Doppler broadening:** Thermal and turbulent motion broaden the transition based on the gas' particles' velocity distribution.

The line profile resulting from natural broadening for Ly α is given as

$$\Phi(\nu) = \frac{A_{\text{Ly}\alpha}}{4\pi^2} \frac{1}{\left(\nu - \nu_{\text{Ly}\alpha}\right)^2 + \left(\frac{A_{\text{Ly}\alpha}}{4\pi}\right)^2}, \quad (2.34)$$

where $A_{\text{Ly}\alpha}$ is the Einstein A coefficient describing the transition probability per unit time for spontaneous emission (Dijkstra, 2019). The functional form, called Lorentzian profile, can easily be derived from an exponentially decaying electromagnetic field amplitude multiplied by the phase function and subsequent fourier transformation (see Rybicki *et al.* (2004)).

Collisional broadening affects the decay time and can be incorporated by adjusting the typical decay frequency $A_{\text{Ly}\alpha} \rightarrow A_{\text{Ly}\alpha} + 2\nu_{\text{col}}$, where $\nu_{\text{col}} \sim n\sigma_{\text{Ly}\alpha}v$. Hence, collisional broadening becomes important in dense ($n \uparrow$) and hot ($v \uparrow$) gas. In the following, we neglect broadening from collisions.

Doppler broadening arises from the velocity distribution relative to the bulk motion, implying a varying Doppler frequency shift for particles in the gas. In thermal equilibrium, we can describe the one dimensional² velocity distribution by a Doppler factor parametrized by the gas temperature T . We obtain a Gaussian profile

$$\Phi(\nu) = \frac{1}{\Delta\nu_D\sqrt{\pi}} \exp\left(-\frac{(\nu - \nu_{\text{Ly}\alpha})^2}{\Delta\nu_D^2}\right), \quad (2.35)$$

where $\Delta\nu_D$ is the Doppler frequency width set by the thermal velocity v_{th} with

$$\Delta\nu_D = \frac{\nu_{\text{Ly}\alpha}}{c} v_{th} = \frac{\nu_{\text{Ly}\alpha}}{c} \sqrt{\frac{2k_B T}{m_H}}. \quad (2.36)$$

In addition to thermal motion, small scale turbulence can be incorporated by adding the turbulent velocity to the thermal velocity below the considered spatial scale.

It is often useful to express the frequency in terms of the dimensionless frequency defined as

$$x \equiv \frac{\nu - \nu_0}{\Delta\nu_D}. \quad (2.37)$$

The combined impact of the Lorentzian and Gaussian broadening is given by convolution of both functional forms (Equations (2.34)/(2.35)) given that they are independent from one another: each particle shows an intrinsic broadening that may be modified by collisions. Additionally all particles within the gas show Doppler broadening. This results in the Voigt profile

$$\Phi(x) = \frac{1}{\sqrt{\pi}} H(a, x), \quad (2.38)$$

which we keep normalized as $\int_{-\infty}^{\infty} \Phi(x) dx = 1$. The profile's shape is given by the Voigt function

$$H(a, x) = \frac{a}{\pi} \int_{-\infty}^{\infty} \frac{\exp(-y^2)}{a^2 + (x - y)^2} dy \quad (2.39)$$

²As only the particle's velocity parallel to a photon's direction is relevant for the Doppler shift, we require the one dimensional form.

with the Voigt parameter

$$a \equiv \frac{A_{\text{Ly}\alpha}}{4\pi\Delta\nu_D}, \quad (2.40)$$

characterizing the ratio of Lorentzian and Gaussian broadening contributions.

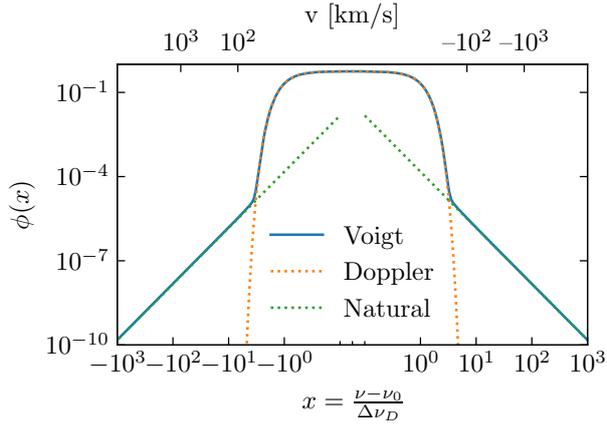


Fig. 2.2 The Ly α line profile normalized in dimensionless frequency space via $\int_{-\infty}^{\infty} \Phi(x)dx = 1$ shown for a temperature of $T = 10^4$ K. The profile is shaped by a Lorentzian (green) and Gaussian profile (orange), primarily due to intrinsic and thermal broadening respectively. The overall profile is given as a convolution resulting in a Voigt profile (blue).

In Figure 2.2 we show the Ly α line profile as a function of the dimensionless frequency split into shaping natural and Doppler broadening. Doppler broadening dominates near the line center, but – given the exponential form – falls off quickly, so that the line profile shows power-law wings from the natural broadening. The transition from the core to wings is approximately given by the respective power-law and Gaussian contributions being equal: $\exp(-x^2) \approx a/(\sqrt{\pi}x^2)$. Numerically solving gives a value of $x_c \sim 3.2$ at $T = 10^4$ K, which is usually a good approximation for typical temperatures encountered.

The Ly α cross-section describing the probability of a Ly α photon being absorbed by neutral hydrogen is given by

$$\sigma_{\text{Ly}\alpha}(\nu) = \frac{\pi e^2}{m_H c} f_{\text{Ly}\alpha} \phi_{\text{Ly}\alpha}(\nu) \quad (2.41)$$

with $\phi_{\text{Ly}\alpha}(\nu)$ being the Voigt profile from Equation (2.38). Note that the Voigt profile is an approximate form of the actual profile (for more details, see e.g. Mortlock, 2016), which however is sufficient for our purposes here.

With Equations (2.38) and (2.41), we find a Ly α line-center cross-section of

$$\sigma_{0,\text{Ly}\alpha} = 5.9 \cdot 10^{-14} \left(\frac{T}{10^4 \text{ K}} \right)^{-1/2} \text{ cm}^2. \quad (2.42)$$

The line-center optical depth of hydrogen, evaluated at 10^4 K, is thus multiple orders of magnitude larger than that of Lyman limit photons and more than 10 orders of magnitude larger than for Thomson scattering (Calzetti *et al.*, 1995). Relatively moderate column densities of 10^{17} cm $^{-2}$ and more will thus lead to countless scatterings of Ly α photons before escaping the hosting structures.

2.6 Emission

Photons at the wavelength of Lyman- α are primarily emitted by two channels after *recombination* and *collisional excitation*. We will call such channels *emission mechanisms* throughout this thesis. We can furthermore identify different physical origins that lead to emission through these channels, which we describe as *emission sources*.

2.6.1 Mechanisms

Recombinations

Ionized hydrogen can recombine with free electrons. Cascading down to the ground state, a significant fraction of recombinations includes emission of a Ly α photon. If we know the temperature dependent recombination coefficients α_{nl} resulting in a recombination leaving the electron in state (n, l) , we can determine the probability for the emission of a Ly α photon per recombination event (Dijkstra, 2019) as

$$P(\text{Ly}\alpha) = \sum_{n_{\min}}^{\infty} \sum_{l=0}^{n-1} \frac{\alpha_{nl}(T)}{\alpha_{\text{tot}}} P(nl \rightarrow \text{Ly}\alpha), \quad (2.43)$$

where $\alpha_{\text{tot}} = \sum_{n_{\min}}^{\infty} \sum_{l=0}^{n-1} \alpha_{nl}(T)$ and the probabilities for a Ly α decay from (n, l) are given by Equation (2.32).

We choose n_{\min} according to two limiting cases “case-A” and “case-B”. In former case, we assume the environment to be optically thin for Lyman series photons and direction recombinations to the ground state ($n_{\min} = 1$). In latter case, we assume the medium to be optically thick to all ionizing and Lyman series photons. These photons are undergoing subsequent emission and absorption until a different cascade to the ground state takes place. Hence, the medium effectively acts such that $A_{n1 \rightarrow 10} = 0$ and $n_{\min} = 2$. Note that this calculation is only straight forward given the limiting cases. Furthermore, we need to assume low densities. At higher densities the l number of excited electrons can change through collisions.

Calculating the respective cascades, we can calculate the average fraction of Ly α photons per recombination event $f_{\text{rec}}(T)$ and model the Ly α emissivity as

$$\epsilon_{\text{rec}} = f_{\text{rec}}(T) n_e n_{\text{HII}} \alpha(T) E_{\text{Ly}\alpha} \quad (2.44)$$

where n_e and n_{HII} are the electron and ionized hydrogen number densities, and $\alpha(T)$ is the recombination coefficient (from Draine, 2011). When explicitly modeling diffuse emission from recombinations in Chapter 6, we will assume case-B recombinations with the fit formula for f_{rec} given by Cantalupo *et al.* (2008). Note that $\alpha(T)$ and $f_{\text{rec}}(T)$ are both temperature dependent. Case-A can be formulated in a similar fashion while not discussed here (Dijkstra, 2019).

Collisional Excitations

In addition to recombinations, the de-excitation of excited hydrogen atoms (that have not been previously ionized) can lead to the emission of Ly α sourced by collisional excitations depending on the thermal state of the gas. The rate is proportional to the colliding species, n_e and n_{HI} . The luminosity density is then

$$\epsilon_{\text{coll}} = \gamma_{1s2p}(T) n_e n_{\text{HI}} E_{\text{Ly}\alpha}. \quad (2.45)$$

For our work here, we take the collisional excitation coefficient $\gamma_{1s2p}(T)$ from tabulated fits (Scholz *et al.*, 1990, 1991).

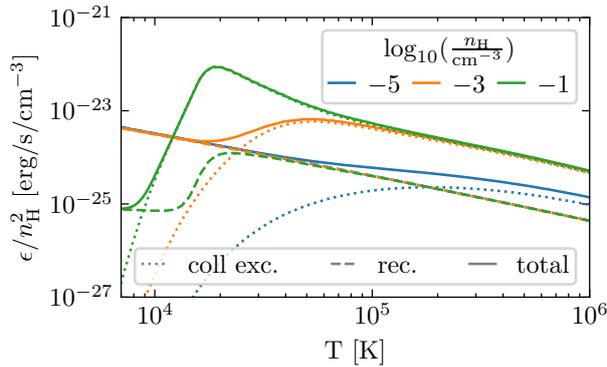


Fig. 2.3 Ly α emission in primordial gas from collisional excitations and recombinations as a function of temperature and for different densities. The gas state is calculated based on the recombination and collisional ionization rates in Katz *et al.* (1996) modified by a UVB from Faucher-Giguère *et al.* (2009) with a self-shielding description by Rahmati *et al.* (2013). For the Ly α emission rates calculated based on the gas state, see text.

In Figure 2.3, we show the Ly α emission for recombinations and collisional excitations normalized by the hydrogen density squared as a function of temperature and density. We use the model descriptions for these two processes as introduced above for a primordial gas of given temperature and optionally in presence of an ionizing background from the UVB at $z = 3$. Collisional excitations largely dominate over recombinations at high densities and intermediate temperatures where sufficient neutral hydrogen and free

electrons exist. Recombinations cease in the presence of a self-shielding description in high densities environments and low temperatures.

2.6.2 Sources

Ly α emission by the *mechanisms* listed above can have various emission *sources* that provide the required ionizing or heating contribution. We now present the major sources for Ly α emission grouped by whether they occur within the galaxies or in the more diffuse gas outside of them.

Sources inside galaxies

Sources inside of galaxies are primarily stellar populations and active galactic nuclei, which we will discuss in the following.

Stellar populations Young star-forming galaxies are often found to have a prominent Lyman- α emission line. Ionizing photons, predominantly emitted by massive short-lived stars, keep their surrounding gas ionized and heated. Through recombinations and de-excitations a substantial part of the stars' ionizing photon budget is subsequently converted to Ly α photons.

The amount of Ly α photons generated assuming case-B recombinations (also see Section 2.6.1) is

$$\dot{N}_{\text{Ly}\alpha} = P_{\text{Ly}\alpha}(T)\dot{N}_{\text{ion}} \quad (2.46)$$

with $P_{\text{Ly}\alpha}(T)$ being the temperature dependent probability for a recombination cascade leading to the emission of a Ly α photon and \dot{N}_{ion} being the hydrogen ionizing photon count (i.e. $\lambda < 912 \text{ \AA}$) emitted by the stellar population. The probability $P_{\text{Ly}\alpha}$ between 5000 – 20000 K is about 2/3 (Spitzer, 1978). Using stellar synthesis models for a chosen initial mass function and metallicity, we can compute the ionizing photon flux and thus the count of generated Ly α photons. Such calculation can then provide us with a linear relationship for the Ly α luminosity versus the star-formation rate.

Throughout this thesis, we adopt such linear model for the Ly α luminosity density ϵ_{SF} proportional to the star formation rate \dot{M}_* in a star-forming region of volume V_* :

$$\epsilon_{\text{SF}} = 10^{42} \left(\frac{\dot{M}_*}{\text{M}_{\odot}\text{yr}^{-1}} \right) \frac{\text{erg/s}}{V_*} \quad (2.47)$$

The proportionality factor in Equation (2.47) can also be motivated by observational studies. More specifically, the proportionality constant is consistent with the $H\alpha$ -SFR relation used in Kennicutt (1998) and an assumed $\text{Ly}\alpha$ to $H\alpha$ recombination ratio in the range of 8-10 (Hummer *et al.*, 1987).

The explicit integration over the ionizing flux for a given stellar population and use of Equation (2.46) has large modeling uncertainties (Furlanetto *et al.*, 2005) and individual stellar populations and galaxies show large deviations from this linear relation (Sobral *et al.*, 2019).

Nevertheless such relation is commonly used in simulations and theory as an estimate for the $\text{Ly}\alpha$ emission from stellar populations (see e.g. Furlanetto *et al.*, 2005; Zheng *et al.*, 2010; Behrens *et al.*, 2018). Equation (2.47) does not account for the escape from the interstellar medium and its potential destruction by dust. It also does not take into consideration the stellar populations' age and metallicity. Incorporating these factors would add suppression and scatter in the relation; we discuss the modeling uncertainties and their implications in Section 6.4.2 in the context of Lyman-alpha halos.

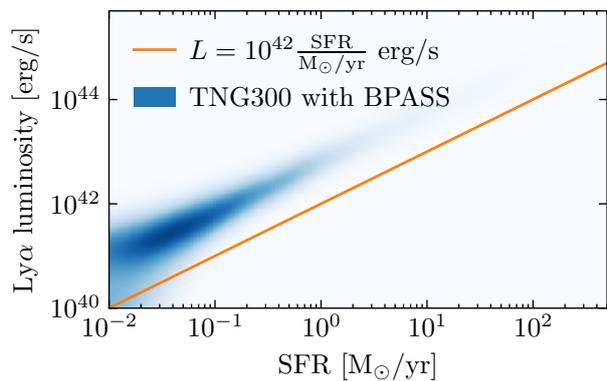


Fig. 2.4 The intrinsic $\text{Ly}\alpha$ luminosity for all galaxies within a halo as a function of total star-formation rate. The solid line shows the common linear relation between $\text{Ly}\alpha$ luminosity and SFR used as in Eqn. (2.47). The blue map shows the phase space density we obtain on a per halo basis by processing the halos and stellar populations following a Chabrier IMF (Chabrier, 2003) in TNG300 at $z=3.0$ using the BPASS spectral synthesis code (Stanway *et al.*, 2018) to estimate the ionizing photon rate from which we calculate the expected $\text{Ly}\alpha$ recombinations in the case-B scenario.

In Figure 2.4, we show the linear relation $L_{\text{Ly}\alpha}(\text{SFR})$ from Equation (2.47) contrasted with the $\text{Ly}\alpha$ luminosity from ionizing radiation using a spectral synthesis code incorporating age and metallicity distributions from the cosmological galaxy formation simulation TNG300. Beside obtaining higher luminosity estimates given the Chabrier IMF and assumption of a significant binary fraction, there is a considerable scatter for the halos' $\text{Ly}\alpha$ luminosity for a given star formation rate given the varying ages and metallicities.

When aiming to reproduce the luminosity functions of LAEs, an effective, parameterized treatment of dust (see e.g. the effective treatment through attenuation in Lake *et al.*, 2015; Inoue *et al.*, 2018) and the introduction of sources of scatter in Equation (2.47) are desirable for future work, and necessary at the high mass end.

Active galactic nuclei Studies find a sharp transition (~ 0.5 dex) at around $L_{\text{transition}} \sim 10^{43.2}$ erg/s for the fraction of LAEs showing AGN activity going from zero to unity with increasing luminosity (Sobral *et al.*, 2018; Spinoso *et al.*, 2020). In addition, the shape of the luminosity function seems to be well described by a composite model of two Schechter functions. The second Schechter function starts dominating around $L_{\text{transition}}$ (Spinoso *et al.*, 2020). Given this dichotomy of the LF along with the characteristic luminosity at which the second Schechter function dominates, it seems plausible that Ly α radiation from AGN starts to dominate the Ly α budget for LAEs at the bright end. But even at lower luminosities (and halo masses) the contribution of AGNs to Ly α luminosities remains unclear due to the complication of detecting obscured AGN (Hickox *et al.*, 2018).

A significant amount of the AGNs' ionizing flux can escape and ionize their surrounding onto large scales (Cen *et al.*, 2000), boosting the emission of Ly α photons via fluorescence, which we discuss next.

Diffuse emission

Outside of galaxies, in the diffuse gas of the circumgalactic and intergalactic medium, fluorescence and gravitational cooling are major sources of Ly α emission. We do not quantify their respective luminosities, which have significant uncertainties, but the resulting emission based on the ionization and temperature of the gas, which is shaped by the sources' ionization and heating terms.

Given the Ly α emission's spatial extent, it is sensible to formulate the magnitude of the expected Ly α emission in terms of surface brightnesses. The surface brightness is related to the emissivities $\epsilon_{\text{Ly}\alpha}$ introduced for recombinations and collisions in Equation (2.44) and (2.45) as

$$\text{SB} = \frac{\epsilon_{\text{Ly}\alpha} \Delta l}{4\pi(1+z)^4} \quad (2.48)$$

where Δl is the depth of the Ly α emitting region we look onto. For gravitationally bound structures, the depth is set by the structure's extent d . If the structure is not gravitationally bound, the Hubble flow shifts and stretches Ly α line for gas further away along the line of sight. For unbound gas the depth Δl is thus implicitly set by the spectral window we integrate over. In this case, the depth is given as $\Delta l = c\Delta\nu_{\text{Ly}\alpha}/\nu_{\text{Ly}\alpha}$. For a typical narrowband filter, we have a width of 100 Å and hence $\Delta\nu_{\text{Ly}\alpha}/\nu_{\text{Ly}\alpha} \approx 0.1$. Substituting this expression (or the gravitationally bound structure's size) into above equation, we can relate the emissivities to observed surface brightnesses. For example, we could find the surface brightnesses for the primordial gas' Ly α emissivities in Figure 2.3. In

particular, for the recombinations under the assumption of a fully ionized IGM, we can then derive a surface brightness of

$$SB_{\text{IGM}} = \frac{c\Delta E(1+z)^{0.5}\Omega_{b,0}^2}{4\pi H_0\Omega_{m,0}^{1/2}} \frac{\Delta\nu_{\text{Ly}\alpha}}{\nu_{\text{Ly}\alpha}} \left(\frac{0.76\rho_{c,0}}{m_H}\right)^2 (1+\delta)^2 f(T)\alpha(T) \quad (2.49)$$

where we have used $H(z) = H_0\sqrt{\Omega_{m,0}}(1+z)^{3/2}$ for the matter dominated era ($z > 1$) and expressed the electron/proton density as $n_e = n_p = 0.76\frac{\rho_{c,0}}{m_H}(1+\delta)\Omega_{b,0}(1+z)^3$. At $T = 10^4$ K, this results in a faint emission of $\sim 10^{-21}$ erg/s/cm²/arcsec² for the Universe's mean density at $z = 3.0$ too faint to detect with instruments such as HSC. At higher overdensities, given the $(1+\delta)^2$ scaling, we quickly get into the instruments' sensitivity range, but ionization and temperature cannot be upheld without the presence of a significant heating or ionization source.

Dedicated studies of diffuse Ly α emission in the IGM have lately been performed using cosmological hydrodynamical simulations (Elias *et al.*, 2020; Witstok *et al.*, 2021). However, the computational results by Witstok *et al.* (2021) have been contested by latest observations by Bacon *et al.* (2021) and thus such computational predictions have to be taken with some caution.

Fluorescence Fluorescence describes the spontaneous emission following an absorption of higher energy. In case of Ly α , we primarily consider ultraviolet radiation $\Delta\lambda < 912 \text{ \AA}$ emitting Ly α photons through a radiative cascade. Narrowing the definition here further, we only call such process fluorescence that occurs outside of the UV radiation's origin. The major sources for UV photons are stellar populations and AGNs upon escaping their immediate surrounding.

The ionizing photons are commonly artificially split into two contributions: The cosmic UV background (UVB) and nearby ionizing sources. At redshifts considered in this thesis ($z \lesssim 5$), well after reionization has completed, and on large scales, the ultraviolet radiation has huge mean free path. This motivates the introduction of the uniform UV background accounting for the averaged ionizing flux from distant sources, which has been successfully able to explain results from Ly α forest observations (Prochaska, 2019). On smaller scales, local ionizing sources from within the hosted galaxies can dominate.

Estimates for the two limiting cases of fluorescence in either a fully ionized gas or a dense, self-shielded gas have been given by Dijkstra (2019) using Equation (2.48). For a

fully ionized dense gas, the surface brightness can be estimated as

$$\begin{aligned} \text{SB}_{fl} &= \frac{0.68 E_{\text{Ly}\alpha} n_e \alpha(T) N_{\text{H}}}{4\pi(1+z)^4} \\ &\approx 2 \cdot 10^{-21} \cdot \left(\frac{1+z}{4}\right)^{-4} \left(\frac{n_e}{10^{-3} \text{cm}^{-3}}\right) \left(\frac{N_{\text{H}}}{10^{20} \text{cm}^{-2}}\right) T_4^{-0.7} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ arcsec}^{-2} \end{aligned} \quad (2.50)$$

with the column density $N_{\text{H}} = \Delta l n_p$ of hydrogen implicitly setting the line of sight length Δl . In above form, the ionization state is effectively captured by the density of free electrons n_e .

In very dense regions where self-shielding becomes important, Ly α radiation will only escape from the surface of depth λ_{mfp} that the ionizing radiation can penetrate (i.e. the length where $\tau_{UV} = 1$):

$$\lambda_{\text{mfp}} = \frac{1}{n_{\text{HI}} \sigma_{\text{ion}}} = \frac{1}{x_{\text{HI}} n_{\text{H}} \sigma_{\text{ion}}} \approx \frac{\Gamma_{\text{ion}}}{\sigma_{\text{ion}} \alpha_B(T) n_e^2} \quad (2.51)$$

for Γ_{ion} representing the ionization rate. Assuming equilibrium with recombinations, we use $x_{\text{HI}} \Gamma_{\text{ion}} = \alpha(T) n_e$ (compare with Equation (3.13)) and further assumed that the bulk of free electrons originates from hydrogen, i.e. $n_e \approx n_{\text{HI}}$. Approximating the cross-section and recombination coefficients, we obtain

$$\text{SB}_{\text{fl,shielded}} \approx 1.3 \cdot 10^{-20} \left(\frac{\Gamma_{\text{ion}}}{10^{-12} \text{s}^{-1}}\right) \left(\frac{1+z}{4}\right)^{-4} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ arcsec}^{-2} \quad (2.52)$$

with the respective temperature dependencies canceling (Dijkstra, 2019).

Gravitational heating Estimates for the Ly α emission from released gravitational potential energy during a structure's gravitational collapse have been given by Haiman *et al.* (2000). These estimates are to be understood as an order of magnitude argument and solely derive on the expected gravitational binding energy in need of dissipation for gas to accrete into the halos. Together with the expected dissipation time scales, this gives an estimate for the emitted Ly α luminosity under the assumption that the bulk of cooling occurs via the Ly α emission line. Estimates for high redshifts are of the orders of $10^{43} - 10^{45}$ erg/s. Faucher-Giguère *et al.*, 2010 find a Ly α cooling emission rate for halos as a function of mass M rate as

$$\langle \dot{E}_{\text{grav}} \rangle \approx 3.8 \cdot 10^{43} \text{ erg s}^{-1} f_{\text{eff}} \left(\frac{M}{10^{12} M_{\odot}}\right)^{1.8} \left(\frac{1+z}{4}\right)^{3.5} \left(\frac{f_{\text{gas}}}{0.165}\right) \quad (2.53)$$

for a NFW halo profile (Navarro *et al.*, 1997) with concentration parameter $c = 5$, where f_{gas} is the fraction of accreted mass that is gaseous and f_{eff} is the efficiency factor accounting for the amount of energy not being dissipated by contributing to a kinetic and thermal energy remaining in the formed structure.

Since this is a rough estimate, there has been a range of computational explorations using simulations of cosmological structure formation (Fardal *et al.*, 2001; Faucher-Giguère *et al.*, 2010) in which accretion of matter occurs in cold streams of $T \sim 10^4 - 10^5$ K primarily cooling through Ly α emission. Major uncertainties remain due to the insufficient capture of the relevant physical processes and resolution capturing the various feedback models and weak shocks that shape the kinematics, ionization and ultimately Ly α emission.

2.7 Absorption and dust

Ly α photons are easily scattered by neutral hydrogen given the large cross-section of the transition and abundance of neutral hydrogen. Scattering, i.e. the absorption and re-emission of Ly α photons is discussed in Section 2.8. Here, we discuss the simpler case of absorption of Ly α photons without subsequent re-emission at a similar wavelength, that we also call (Ly α) *destruction*.

Typically, destruction occurs by dust, but Ly α photons can also be destroyed by various other channels including molecular hydrogen, absorption by already excited hydrogen, and 2p-2s mixing by collision with ions before recombination (Spitzer *et al.*, 1951). These channels are not handled in this dissertation as they remain unimportant for the resolved regions in the cosmological galaxy formation simulations studied (Dijkstra, 2019).

Cosmic dust describes a range of small particles built up from metals. Upon absorption of a Ly α photon, their energy is commonly dissipated by vibrational modes into the infrared (Draine, 2011) and the Ly α photon is thus lost. While the composition of dust and the Universe's enrichment with it is a sizable research field and of particular uncertainties in the high redshift Universe, we are only interested in its impact on the Ly α line. Its impact can be effectively modelled without having to descend to cosmic dust's accurate description. In fact, the dust type and its choice of implementation have little impact on the results for the Ly α radiative transfer (Laursen, 2010b).

The cross-section σ_D for Ly α interacting with a dust grain can be expressed in terms of the typical dust grain size d and according absorption/scattering efficiencies Q_A/Q_S as

$$\sigma_d = \pi d^2 (Q_A^2 + Q_S^2) \quad (2.54)$$

following Verhamme *et al.* (2006) and Behrens (2014). The albedo

$$A \equiv \frac{Q_s}{Q_s + Q_A} \quad (2.55)$$

sets the probability to scatter (rather than being destructed with a probability $1 - A$). For Ly α an albedo of ~ 0.5 is appropriate (Verhamme *et al.*, 2006). While the cross-section for dust is strongly frequency dependent, it can be well approximated as flat for any reasonable shift near the Ly α line-center. The phase function for the direction of scattering is commonly approximated with the relatively simple Greenstein phase function (Henyey *et al.*, 1941) with a g -parameter of ~ 0.7 . We introduce phase functions, including the Greenstein phase function, when discussing the scattering process in Section 2.8.

In order to use observational measurements of dust extinction, it is convenient to define the cross-section $\sigma_{D,H}$ as the cross-section for Ly α photons per hydrogen atom (rather than per dust grain). Using observational constraints, we can then determine this cross-section from measured extinction coefficients for a given hydrogen column density (Laursen, 2010b).

The Ly α radiative transfer with dust can strongly depend on the clumpiness of hydrogen and dust. While observationally unclear, theoretical considerations estimate fragmentation of gas clouds on sub-parsec scales (McCourt *et al.*, 2018). The Ly α radiative transfer leads to vastly different outcomes depending on the small-scale structure (Gronke *et al.*, 2016b, 2017b). As resolving this structure remains computationally unfeasible for the dynamic range that would need to be covered for cosmological structure formation simulations considered in this thesis, we decided against incorporating dust in our fiducial models except for test purposes and qualitative trends. Alternatively, effective subgrid models can be incorporated (e.g. from Hansen *et al.*, 2006). A similar solution might be desirable in multiphase gas of star-forming regions, also see the discussion in Section 6.4.2.

2.8 Scattering

Upon absorption of a Ly α photon by a neutral hydrogen atom in the ground state, a photon of the same energy³ is emitted in the atom's rest-frame. Observed in the frame of the gas' bulk motion, absorption and reemission of a Ly α photon can nevertheless change the photon's energy due to the thermal motion of the interacting atom. Here, we discuss the change in frequency and direction of such Ly α scattering process.

³Minor changes in energy can occur due to a recoil effect and potential perturbation before decay on time-scale A_{21}^{-1} , see discussion below.

2.8.1 Frequency shift

To zeroth order, the scattering is coherent in the rest-frame of the atom: the frequency of the incoming photon equals that of the outgoing photon. Given the atom's velocity that the photon interacts with, we however have to account for the relativistic Doppler shift into and out of the atom's rest-frame. This implies that the photon's energy will in fact change for an observer moving with the gas' bulk motion. The relativistic Doppler shift for a photon's frequency ν with direction \vec{n} observed as frequency ν' in the atom's rest-frame moving with velocity \vec{v} is given as

$$\nu' = \nu \sqrt{\frac{1 - \frac{\vec{v} \cdot \vec{n}}{c}}{1 + \frac{\vec{v} \cdot \vec{n}}{c}}}, \quad (2.56)$$

which we can express in terms of the dimensionless frequency

$$x' = x - \frac{\vec{v} \cdot \vec{n}}{v_{\text{thermal}}}, \quad (2.57)$$

where we have assumed $(\vec{v} \cdot \vec{n})/c \ll 1$ and that the frequency shifts remain close the line-center. Shifting into and out of the atom's rest-frame results in two of the above transforms with $x'_{\text{in}} = x'_{\text{out}}$ for coherent scattering. This leads to

$$x_{\text{out}} = x_{\text{in}} + \frac{\vec{v}(\vec{n}_{\text{out}} - \vec{n}_{\text{in}})}{v_{\text{thermal}}}. \quad (2.58)$$

Coherence of the scattering in the atom's rest-frame is broken by the recoil effect due to a change in the photon's associated momentum. This subsequently also changes Equation (2.58). However, this effect should be negligible in astrophysical environments except for very low temperatures and we thus do not discuss this effect further (Adams, 1971; Zheng *et al.*, 2002).

The velocity \vec{v} entering Equation (2.58) is drawn from the gas' velocity distribution, which is well described by a Maxwell-Boltzmann distribution. For the individual dimensions, this corresponds to a simple Gaussian distribution for each dimension. Choosing the coordinate system to be aligned with the photon's direction, we can describe the atom's velocity with a parallel and two perpendicular components. While all three components follow a Gaussian velocity distribution, the probability distribution for the parallel component of a scattering atom is more complicated: Due to the Doppler shift set by the perpendicular velocity component, the probability of interaction with the photon is modified by the frequency dependent cross-section according to Equation (2.38). This

means that the perpendicular velocity component's PDF is given as the convolution of the line profile and the Gaussian velocity distribution

$$f(u_{\parallel}) = \frac{a}{\pi H(a, x)} \frac{e^{-u_{\parallel}^2}}{(x - u_{\parallel})^2 + a^2}, \quad (2.59)$$

while the perpendicular distributions ($i = 1, 2$) simply follow a Gaussian:

$$g(u_{\perp, i}) = \frac{1}{\pi} e^{-u_{\perp, i}^2}, \quad (2.60)$$

2.8.2 Change of direction

The phase function describes the angular probability distribution for the outgoing photon relative to the incoming photon. The phase function $P(\mu)$, with the angle $\mu \equiv \cos(\theta)$ parallel to the incoming photon, is normalized via

$$\frac{1}{4\pi} \int_0^{2\pi} \int_0^{\pi} P(\cos \theta) \sin \theta d\theta d\Phi = \frac{1}{2} \int_{-1}^1 P(\mu) d\mu = 1. \quad (2.61)$$

Isotropic re-emission is thus described by

$$P(\mu) = 1 \quad (2.62)$$

and commonly considered dipole scattering by

$$P(\mu) = \frac{3}{4}(1 + \mu^2). \quad (2.63)$$

Ly α photons are dipole scattered when the frequency is in the line's wings. For frequencies in the core, scattering occurs as a hybrid form due to the isotropic decay from the $2P_{1/2}$ state and dipole like decay from $2P_{3/2}$ (Hamilton, 1940; Stenflo, 1980; Laursen, 2010b).

For the description of dust, we use the Greenstein phase (Henyey *et al.*, 1941) function

$$P(\mu) = \frac{1}{2} \frac{1 - g^2}{(1 + g^2 - 2g\mu)^{3/2}}. \quad (2.64)$$

to approximate the dust's scattering process, where one commonly sets the characterizing factor $g = 0.7$ for dust (Li *et al.*, 2001).

Similar to the frequency shift in Equation (2.57), the direction of the photon will change when going from the gas' frame to the atom's rest-frame. In the gas' rest-frame, a photon traveling with an angle θ relative to the scattering atom's velocity v , will have a modified direction k' in the atom's rest-frame transformed by

$$k'_{\parallel} = \frac{k_{\parallel} - v/c}{1 - (v/c)k_{\parallel}} \quad (2.65)$$

where $k_{\parallel} \equiv \cos \theta$ is the parallel directional component relative to the atom's velocity (Rybicki *et al.*, 2004). The perpendicular component is then set by $k'_{\perp} = \sqrt{1 - k'^2_{\parallel}}$. The azimuthal angle, i.e. the direction within the plane perpendicular to the atom's velocity, remains unchanged. For example, a photon's direction remains unchanged in the atom's frame when solely moving parallel to the atom's velocity (i.e. $k'_{\parallel} = k_{\parallel} = 1$). In contrast, a photon reaching the atom in a right angle relative to its velocity component ($k_{\perp} = 1$) will have part of its component shifted into the parallel component ($k'_{\parallel} = v/c$).

2.9 Ly α radiative transfer

The radiative transfer equation (RTE) describes the propagation of radiation depending on the absorption, emission and scattering in the underlying medium.

The radiative transfer equation is commonly expressed in terms of the specific intensity I_{ν} , defined as the energy per unit time, area, solid angle and frequency. It is thus given by

$$I_{\nu} = \frac{dE}{dt dA d\Omega d\nu}. \quad (2.66)$$

The radiative transfer equation is given as

$$\left(\frac{1}{c} \frac{\partial}{\partial t} + \frac{\partial}{\partial s} \right) I_{\nu} = \underbrace{-\alpha_{\nu} I_{\nu}}_{\text{Absorption}} + \underbrace{j_{\nu}}_{\text{Emission}} + \underbrace{\int \int \alpha_{\nu'} I_{\nu'} R(\nu, \hat{n} | \nu', \hat{n}') d\nu' \frac{d\Omega}{4\pi}}_{\text{Scattering}}, \quad (2.67)$$

where I_{ν} , α_{ν} and j_{ν} are time-dependent fields that also depend on the radiation's direction \hat{n} and frequency ν . With the RTE being a complex integro-differential equation, analytical solutions are sparse even under simplifying assumptions, motivating our numerical Monte Carlo approach presented later in Chapter 3. We nevertheless shortly motivate the different terms of the RTE and link them to already introduced theory from above.

2.9.1 Absorption term

The absorption term describes the removal of photons from the beam and thus decreases the specific intensity. The absorption is quantified by the medium's attenuation coefficient α_ν . With the attenuation being proportional to the specific intensity, we obtain

$$\frac{dI_\nu}{ds} = -\alpha_\nu I_\nu \quad (2.68)$$

for specific intensity's change per path element ds . The attenuation coefficient is given by $\alpha_\nu = n\sigma_\nu$ with number density n and cross-section σ_ν of the respective medium. For Ly α photons, attenuation given by neutral hydrogen (and subsequent scattering) and destruction by dust with

$$\alpha = \alpha_\nu^{HI} + \alpha_\nu^{\text{dust}}. \quad (2.69)$$

Integrating Equation 2.68 yields

$$I = I_0 \exp \left[- \int_0^s \alpha(s') ds' \right] = I_0 \exp [-\tau]. \quad (2.70)$$

We thus obtain an exponential decay of the initial ray I_0 for the integrated attenuation coefficient over the distance it traverses. The latter defines the optical depth:

$$\tau \equiv \int_0^s \alpha(s') ds' \quad (2.71)$$

2.9.2 Emission term

The emission term is described by the emissivity j_ν in the radiative transfer equation. For a beam passing a segment ds , the specific intensity increases according to

$$\frac{dI_\nu}{ds} = j_\nu. \quad (2.72)$$

The emissivity in our case are the different presented emission mechanisms. If we fully capture the hydrogen's ionization structure, we get

$$j_\nu = \frac{\phi(\nu)}{4\pi} (\epsilon_{\text{rec}} + \epsilon_{\text{coll}}) \quad (2.73)$$

for the recombinations and collisional excitations from Equations (2.44) and (2.45). In above expression, we assume isotropic emission. As we do not resolve the hydrogen's

ionization and temperature structure properly in hydrodynamic simulations, we commonly add contributions such as from star-forming regions, see Equation (2.47). For the resolved gas structure the spectral distribution $\phi(\nu)$ for the emission is given by the Ly α line profile $\Phi(\nu)$ in the rest-frame of the gas. One needs to potentially adjust the spectral distribution $\phi(\nu)$ to account for the radiative transfer of the unresolved density and velocity structure at a given scale in simulations.

2.9.3 Scattering term

When Ly α photons are absorbed by neutral hydrogen, those excited hydrogen atoms quickly re-emit Ly α photons given the large $A_{\text{Ly}\alpha}$ leading to an average decay time of $t = \frac{1}{A_{\text{Ly}\alpha}} \approx 1$ ns. The scattering term is primarily characterized by the so-called redistribution function R . This function relates the probability $P(\nu, \hat{n}|\nu', \hat{n}')$ of an incoming photon from direction \hat{n}' and frequency ν' being scattered into a direction \hat{n} and a frequency ν via

$$P(\nu, \hat{n}|\nu', \hat{n}') = R(\nu, \hat{n}|\nu', \hat{n}') d\nu d\nu' \frac{d\Omega}{4\pi} \frac{d\Omega'}{4\pi}. \quad (2.74)$$

We can obtain the specific intensity by integrating over the directions and frequencies of the incoming photons:

$$\frac{dI_\nu}{ds} = \int \int \alpha'_\nu I_{\nu'} R(\nu, \hat{n}|\nu', \hat{n}') d\nu' \frac{d\Omega}{4\pi}. \quad (2.75)$$

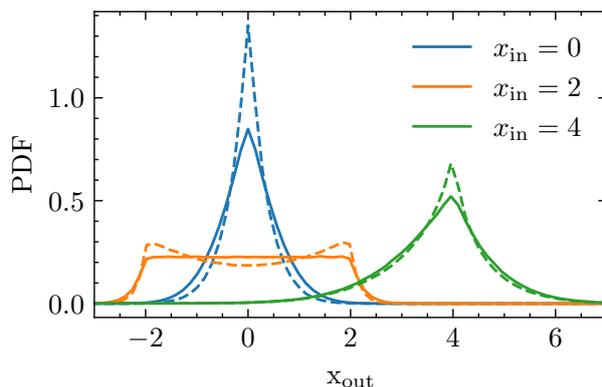


Fig. 2.5 Frequency redistribution $R(x_{\text{out}}|x_{\text{in}})$ by scattering. The PDF for outgoing photon frequencies is shown for different input frequencies (colors). We assume a isotropic (dipole) scattering phase function for the solid (dashed) line.

The redistribution processes is determined by the scattering behavior presented in Section 2.8. An interesting quantity we can obtain from it is the probability distribution $R(x_{\text{out}}|x_{\text{in}})$ for an outgoing frequency x_{out} following an ingoing photon's frequency x_{in} . This can be calculated by marginalizing over the angles of the incoming and outgoing photon (e.g. see Dijkstra, 2019). We show the resulting distribution in Figure 2.5. Differences

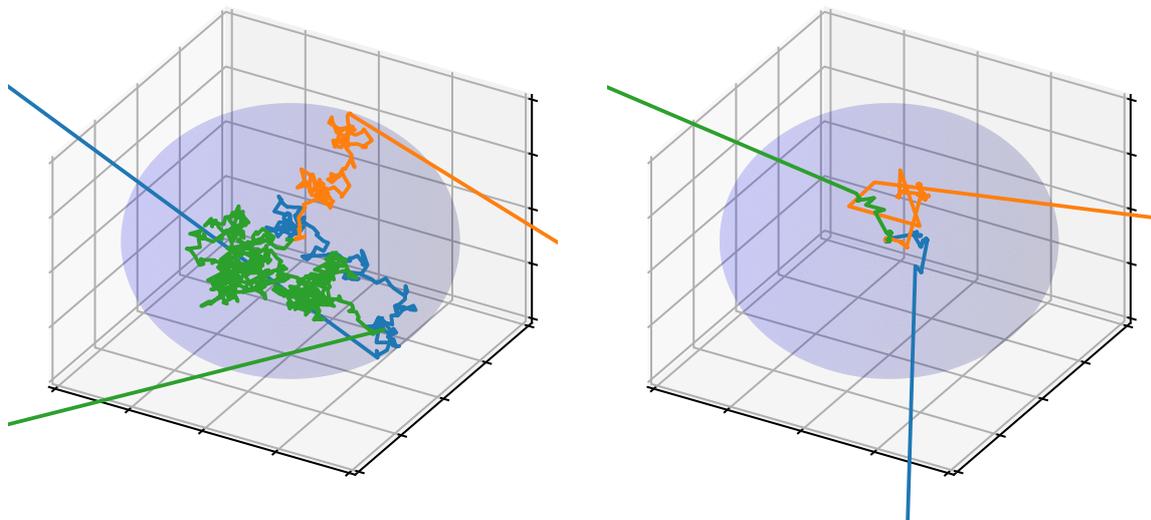


Fig. 2.6 A random walk (**left**) compared to the $\text{Ly}\alpha$ photon walk through a static uniform sphere (**right**). For the random walk, the directions are drawn isotropically, while for the $\text{Ly}\alpha$ photon trajectories photons experience changes in direction and frequency according to Section 2.8. Path lengths are drawn from an exponential distribution to reproduce an exponential intensity decay as in Eqn. (2.70). The optical depth in the line-center from the center of the sphere is $\tau = 10^1$ for the random walk and $\tau = 10^5$ for the $\text{Ly}\alpha$ photon trajectories.

between the isotropic and dipole scattering phase function are moderate. This motivates adopting either of those as an approximation for the $\text{Ly}\alpha$ scattering phase function in radiative transfer codes, while in reality both isotropic and dipole phase functions occur as discussed in Section 2.8.2.

2.10 Building a $\text{Ly}\alpha$ radiative transfer intuition

The radiative transfer solution for $\text{Ly}\alpha$ only holds analytic solutions for a few special cases, which motivates the use of a Monte Carlo radiative transfer code as we use and introduce, see Chapter 3. Nevertheless, we can gain insights into the scattering process by investigating the scattering term of the radiative transfer equation and by qualitatively studying simplified numerical and (semi-)analytic solutions.

2.10.1 Random walk

The radiative transfer for $\text{Ly}\alpha$ can appear peculiar at first. For a better understanding, we can start by investigating random walks. For random walks, we can follow individual photons, just as we will for the numerical radiative transfer simulations of $\text{Ly}\alpha$ photons

later on. We will shortly review the expected behavior for a random walk and then subsequently work out the differences for Ly α radiative transfer. We define a random walk as a sequence of n random steps with a respective displacement \vec{r}_i , such that the overall distance traveled is

$$\vec{R} = \sum_{1 \leq i \leq n} \vec{r}_i. \quad (2.76)$$

The mean displacement is hence zero for isotropic scattering for which $\langle \vec{r}_i \rangle = 0$. But while the mean displacement remains zero, the photons will stochastically traverse a finite distance L . Such distance could for example be the surface of an optically thick medium at which point a photon can escape. We can characterize this by the root mean square distance

$$R_{\text{RMS}} = \sqrt{\langle \vec{R}^2 \rangle} = \sqrt{\left\langle \left(\sum_{1 \leq i \leq n} \vec{r}_i \right)^2 \right\rangle} = \sqrt{\sum_{1 \leq i \leq n} \langle \vec{r}_i^2 \rangle} \quad (2.77)$$

where we used $\langle \vec{r}_i \cdot \vec{r}_j \rangle = 0$ ($i \neq j$) for independent steps. If we draw all steps from the same probability distribution, we thus find

$$R_{\text{RMS}} = \sqrt{N} \langle r \rangle \quad (2.78)$$

For a uniform medium with absorption coefficient α and extent L from its center, we can define its optical depth as $\tau = \alpha \cdot L$ and the mean free path $\lambda = \alpha^{-1}$. We can now estimate the number of scatterings needed to escape the medium by setting $\langle r \rangle = \lambda$ and $R_{\text{RMS}} = L$ and thus find

$$N = \tau^2 \quad (2.79)$$

for optically thick ($\tau \gg 1$) environments. In the optical thin case ($\tau \ll 1$), considerations of the random walks would be misguided here as less than a single step takes place. In such case, the number of scatterings scales linear with the optical depth (Rybicki *et al.*, 2004).

In Figure 2.6, we contrast a random walk (left panel) in a uniform sphere with trajectory of Ly α photons in such uniform sphere. As expected, the number of scatterings corresponds to $\mathcal{O}(10^2)$ for the random walks in the sphere of optical depth $\tau = 10$ from the sphere's center. Furthermore, the walks seem in line with the assumption of different steps to be independent from one another and isotropic. We draw the optical depths of steps from an exponential distribution in order to reproduce the behavior of Eqn. (2.70).

While step sizes can thus vary substantially, the escape of photons does not rely on a few large steps, but rather the overall ensemble of steps that happens to be anisotropic enough at some point of the random walk to escape the sphere.

In the right panel of Figure 2.6, we show the radiative transfer for three Ly α photons at a line-center optical depth of $\tau = 10^5$. The resulting trajectories look significantly different from those of the random walk even though the scattering is isotropic to zeroth order and the optical depths are drawn from an exponential distribution as for the random walk. In contrast to the random walk, the trajectories appear to have much larger and varying step length between the scattering events. The number of scatterings is in the order of $\mathcal{O}(10^5)$ per trajectory. This might appear surprising for two reasons: First, the amount of visible path elements appears much smaller, which hints that most scatterings appear after very small displacements, with photons spontaneously escaping in longer paths. Second, the expected number of scatterings would be $\mathcal{O}(10^{10})$ for a random walk – vastly larger than what we observe. In the next section, we discuss some important aspects of Ly α scattering which explain this behaviour.

2.10.2 Insights from the scattering process

Osterbrock (1962) calculated that a photon with an initial frequency shift x with a frequency shift x' after scattering in the Ly α profile's wings will have an expected shift difference $\Delta x = x' - x$ of

$$\langle \Delta x' | x \rangle = -\frac{1}{x}. \quad (2.80)$$

Photons are thus on average shifted back towards the line-center with each scattering in the Lorentzian wings. Furthermore, it was shown that the root mean square (rms)

$$\sqrt{\langle \Delta x^2 | x' \rangle} = 1 \quad (2.81)$$

is constant with a frequency displacement of one Doppler width per scattering (Osterbrock, 1962).

In the following, we will discuss the two optical thick regimes for a basic understanding of the escape of Ly α photon from astrophysical media and the characteristic double-peaked spectrum that arises. These scenarios are described by the single longest flight (Osterbrock, 1962) and the single longest excursion (Adams, 1972). In the former scenario, a Ly α photon will escape by a single flight upon being scattered so far into the line profile's wings that the medium appears optically thin. In the latter scenario, optical

depths are so large that a Ly α photon making a large frequency excursion can no longer escape in a single flight as the medium also appears optically thick in the line profile's wing. Instead, spatial diffusion of the photon sets in while gradually, but slowly, also wandering back into the line-center in frequency space.

Single longest flight

In an optically thick medium with a photon spawned in the line-center, seeing an optical depth $\tau = \tau_0$, the Ly α photon will scatter numerous times until scattering with a fast moving atom, changing its direction and subsequently its frequency. This can drastically change the photon's optical depth in the medium. For the Gaussian core of the line profile, the optical depth of the outgoing photon with frequency x_1 will be $\tau = \tau_0 \exp(-x_1^2)$. Hence, if the outgoing photon is scattered to

$$x_1 = (\ln \tau_0)^{1/2}, \quad (2.82)$$

where the medium is no longer appears optically thick (i.e. $\tau_1 = 1$), then the Ly α photon can escape in a single flight.

Assuming complete redistribution of the Ly α photon's frequency, i.e. no correlation between the incoming and outgoing photon's frequency (Zanstra, 1949), the frequency redistribution in the core is given by

$$p(x) = \sqrt{\pi}^{-1} \exp(-x^2) \quad (2.83)$$

and subsequently, the probability to scatter to a frequency $x \geq x_1$ where the photon can escape is given by

$$\omega(x_1) = 2 \int_{x_1}^{\infty} p(x) dx = 1 - \frac{2}{\sqrt{\pi}} \int_0^{x_1} \exp(-x^2) dx \quad (2.84)$$

and the number of scatterings is given by $N = 1/\omega(x_1)$. Approximating the error function from above integral with Chiani *et al.* (2003), we find

$$p(\tau_0) \approx \frac{1}{6} \exp(-x_1^2) + \frac{1}{6} \exp\left(-\frac{4}{3}x_1^2\right) = \frac{1}{6} \frac{1}{\tau_0} + \frac{1}{2} \frac{1}{\tau_0^{4/3}} \approx \frac{1}{6} \frac{1}{\tau_0} \quad (2.85)$$

for $\tau_0 \gg 1$. We thus find the number of scatterings to scale linearly with the optical depth. The frequency diffusion in the Gaussian core followed by a single flight thus substantially

decreases the amount of scatterings compared to a classical spatial diffusion with $N \sim \tau_0$ for a random walk in Section 2.10.1.

However, at high optical depths, part of the Lorentzian wings will eventually become optically thick such that above calculation is no longer a valid approximation. Osterbrock (1962) argued that scatterings should start to scale as $N \sim \tau_0^2$ but was however unable to numerically confirm it. Instead, the number of scatterings was found to remain at its $N \sim \tau_0$ scaling. Next, we will give an argument for this linear scaling incorporating the line profile's Lorentzian wings and for the position of the spectrum's peaks as given by Adams, 1972.

Single longest excursion

For $x \gg 1$, the mean shift is much smaller than the rms (see Equations 2.80/2.81) and thus the photon will undergo a random walk in frequency space, such that it will take $N \sim x^2$ to reach the line-center again. The mean free path scales inverse to the line profile: $l = 1/(n\sigma) \sim \Phi(x) \sim x^2/a$ where a is the Voigt parameter as before. As the line profile changes relatively slowly in the Lorentzian wings, we can hence estimate the rms spatial distance during the random walk in frequency space as $l_{\text{rms}} \sim x^2\sqrt{N}/a \sim x^3/a$. If the photon ought to escape within one such large frequency shift, we expect $l_{\text{rms}} \sim r_0$, where we can express the medium's size r_0 in terms of the line-center optical depth via $r_0 = \tau_0/(n\sigma)$. We hence expect that photons will effectively escape for frequency shifts x^* of

$$x^* \sim \left(\frac{a\tau_0}{\sqrt{\pi}} \right)^{1/3}. \quad (2.86)$$

We expect to x^* to be the peak of the emerging spectra as lower frequency shifts will return to the optically thick line-center before escaping and larger frequency shifts are generally more unlikely to occur.

Adams (1972) furthermore gave an argument for the average number of scatterings required for photons to escape. As the spectrum is relatively flat⁴, the probability to find a photon within $x \pm dx/2$ is approximately $\Phi(x)dx$ assuming complete redistribution. As seen, photons however scatter $\sim x^2$ times at a given frequency x . Hence, the probability to initially scatter *into* this frequency range is $\sim \Phi(x)/x^2 dx$. The cumulative escape probability P_{esc} is then given as the integration for shifting into any frequency that is

⁴Adams, 1972 found this numerically, while it is physically expected from the Fokker-Planck equation (Gronke *et al.*, 2017b).

further than x^* from the line center, i.e.

$$P_{\text{esc}} = \int_{-\infty}^{-x^*} \frac{\Phi(x)}{x^2} dx + \int_{x^*}^{\infty} \frac{\Phi(x)}{x^2} dx = 2 \int_{x^*}^{\infty} \frac{\Phi(x)}{x^2} dx = \frac{2a}{3\pi x^{*3}} \quad (2.87)$$

Photons are hence expected to escape after $N = 1/P_{\text{esc}}$ scattering events scaling as

$$N \sim \frac{3\sqrt{\pi}}{2} \tau_0 \quad (2.88)$$

using Equation (2.86). The number of scatterings thus remains proportional to the optical depth as in the case of the single longest flight rather than squared as for a classical random walk.

Illustrations of RT Dependencies

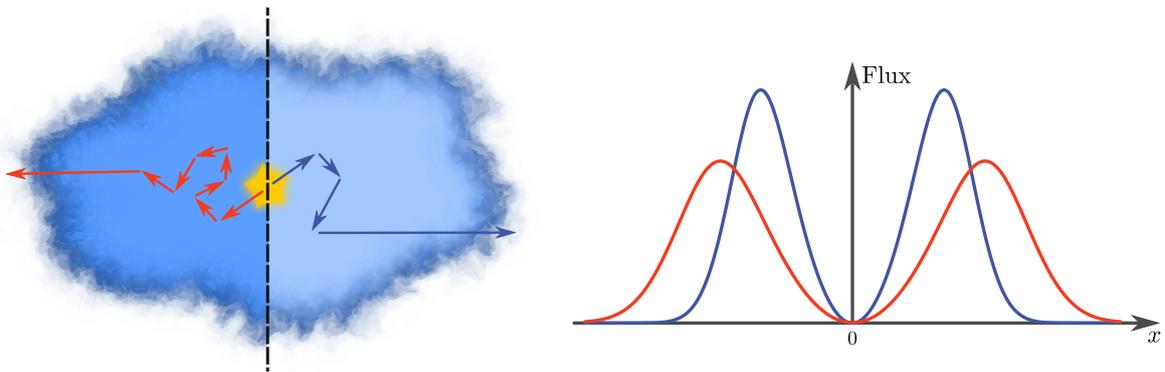


Fig. 2.7 Illustration of the Ly α radiative transfer for two different neutral hydrogen densities in a homogeneous medium (**left**). On the left side of the illustration, we show a high density environment, opposed to a lower density on its right side. Photons will scatter numerous times in this optically thick medium, until they happen to interact with a fast moving atom after which they can experience a frequency shift for which the medium occurs opaque. At higher densities, and hence optical depths, interactions with faster moving atoms are required to shift the photon to an optically thin frequency. Hence, more scatterings are required. We also show the resulting spectra (**right**). Given the larger frequency shifts at which the medium occurs optically thin at higher densities, the peak positions move further away from the Ly α line-center.

In the Figures 2.7, 2.8 and 2.9, we respectively visualize and explain the impact of a changing density, velocity and clumpiness on the photon trajectories and resulting spectra of escaping photons. In the left panel of the figures, we show photon trajectory sketches for two different setups to the left and right of respective sketch. In all figures, we assume very optically thick medium, where photons finally escape with large frequency shifts while omitting the required very high amount of scatterings required to statistically

enable such frequency shift. An analytic solution for a setup of homogeneous dense slabs/spheres exist, which we will present in Section 2.10.3.

Figure 2.7 shows the scattering process and resulting spectra in different dense homogeneous media. The spectrum is double peaked as photons need to acquire a frequency out of the optically thick Ly α line-center in order to escape. The peak position scales monotonously with the density as found in Equation (2.86).

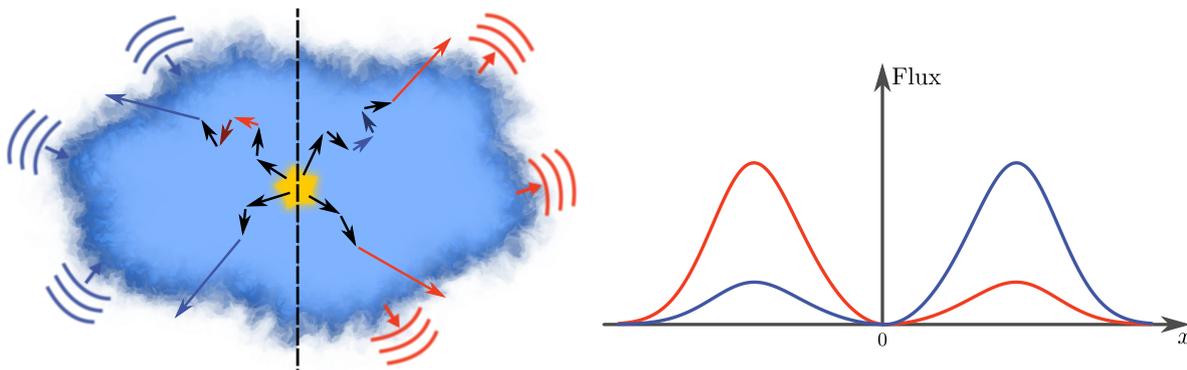


Fig. 2.8 Illustration of the Ly α radiative transfer for two different velocity gradients in a homogeneous medium (**left**). On the left side of the illustration, we show an inflowing structure, opposed to an outflowing structure on the right. Due to the velocity gradient, photons with a frequency offset will be shifted into or away from the line-center as they travel outwards according to their frequency shift relative to the velocity gradients sign. Photons that will be shifted into the line-center (red photons for infalling structures, blue photons for outflowing structures) will hence experience a high optical depth again. Only upon being shifted into a frequency that will be further shifted out of the line-center by the velocity gradient (blue photons for infalling structures, red photons for outflowing structure) photons can escape. We also show the resulting spectra (**right**). As the velocity will help or hinder Ly α photons to escape depending on their shift relative to the line-center, the spectra are accordingly asymmetric towards a blue or red peak depending on the sign of the velocity gradient.

Figure 2.8 shows the resulting spectra in the presence of an differential in- or outflow. For an inflow, photons will be increasingly blueshifted (increasing x) in the gas rest-frame as they move outwards. This means that photons already in the blue wing ($x > 0$) of the line-profile will experience lower optical depths, while photons in the red wing ($x < 0$) will be shifted towards the optically very thick line-center. As a result, we obtain a suppression of the red peak ($x < 0$) relative to the blue peak. For outflowing gas, the reasoning and result reverse.

The scenario of inflowing or outflowing gas and the resulting asymmetric spectrum is important for the understanding of spectral signatures of Ly α emitting galaxies. In physical environments in and around galaxies, inflows commonly arise from cold accretion streams and outflows signify feedback resulting from stars and AGN. Observations

commonly show a dominant blue peak⁵ ($x > 0$) interpreted as indicator for the presence of strong outflows in those objects (Gronke, 2017). However, there is a potential degeneracy in the interpretation of the blue peaks. Namely, the Hubble flow effectively acts as an outflow given the expanding space during the photons' flight. This potential degeneracy between galaxy feedback and Hubble flow for the Ly α spectra will be an important topic of Chapter 5.

In Figure 2.9, we show the impact of dust and clumping on the Ly α escape and spectra. In a homogeneous optically thick medium, tiny amount of dust will be sufficient to attenuate any Ly α flux due to the large integrated path that photons propagate between the numerous scatterings until interacting with very fast moving atoms. While the dust cross-section is approximately constant within the frequencies that Ly α photons are shifted into, the spectral shape (in addition to any attenuation) is still affected significantly by dust. Larger frequency shifts are more unlikely to occur. Larger shifts increase the photons' number of scatterings, travelled path, and thus chance to be attenuated by dust before being able to escape by a substantial frequency shift. As a consequence, dusty homogeneous environments will show a smaller peak separation than dust free environments (Laursen *et al.*, 2009). However, the Ly α spectral shape and photon escape can vastly differ for clumpy media depending on its structure with various regimes present characterized by the optical depth and clumping factor (Gronke *et al.*, 2017b).

2.10.3 Toy models

For large optical depths, the spectrum J emerging from a central source emitting at a resonant line frequency in a slab of infinite extent except for its perpendicular direction with an optical depth τ_0 from the slab's center and a Voigt parameter a can be derived to be

$$J(\pm\tau_0, x) = \frac{\sqrt{6}}{24} \frac{x^2}{a\tau_0} \frac{1}{\cosh \left[\sqrt{\pi^4/54} \frac{|x^3|}{a\tau_0} \right]} \quad (2.89)$$

normalized to $1/4\pi$. This result was first derived by Harrington (1973) and generalized by Neufeld, 1990 to allow for frequency shifts and source displacement. In Dijkstra *et al.*, 2006 a solution for a similar setup under spherical symmetry is provided, yielding a similar resulting shape but slight shift of the peak positions. Note that in the following we will call all results from such setups "Neufeld solution". It provides an easy to implement end-to-end test for Ly α Monte Carlo radiative transfer codes – such as is introduced in

⁵If no subdominant red peak is present, the single peak can still be identified as "blue" given its often measured systemic redshift from the line-center inferred from other emission lines.

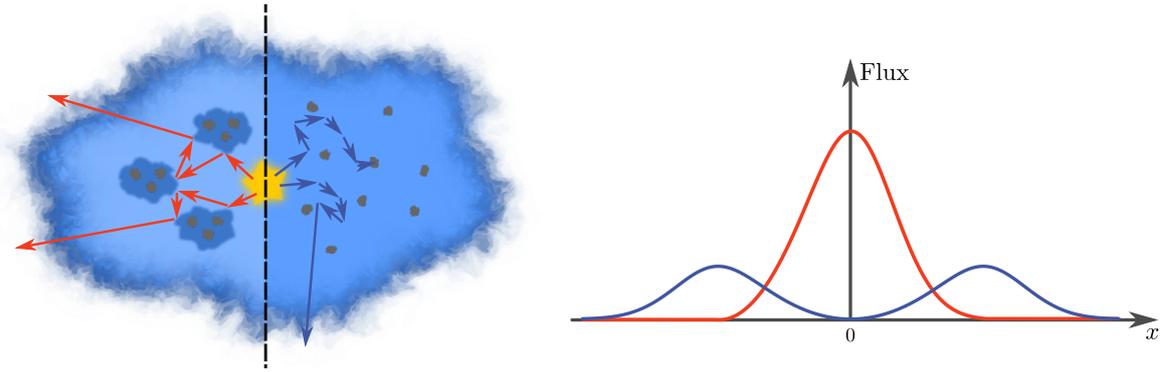


Fig. 2.9 Illustration of the Ly α radiative transfer for two different density distributions **right**. On the left side of the illustration, we show a clumped medium, opposed to a homogeneous medium on the right. Gray particles represent dust. While we assume the same average densities, the photon trajectories will qualitatively differ in the two scenarios. Ly α photons can effectively off-scatter at boundaries to much higher densities. As a consequence, dust that is clustered in clumped neutral hydrogen density regions is effectively dodged by Ly α photons. Furthermore the path lengths Ly α photons travel during their scattering process are effectively reduced compared to the homogeneous density case. In latter case, dust is furthermore not evaded and, as scatterings can largely increase the path lengths travelled, destruction of Ly α photons is largely increased at higher densities in homogeneous media. We also show the expected spectra (**right**).

Chapter 3. A generalized static solution for power-law density and emission structures have recently been calculated by Lao *et al.*, 2020.

The solution is characterized by a double-peaked, symmetric shape. The Neufeld solution is shown in Figure 2.7. As deduced from the arguments in Section 2.10.2, the peaks of the Neufeld solution indeed scale as

$$x_p = C(a\tau_0)^{1/3} \quad (2.90)$$

with $C = 0.92$ in the spherical geometry and $C = 1.06$ for the slab geometry (Dijkstra *et al.*, 2006).

In non-static media, early considerations for the Ly α line shift have been presented by Neufeld *et al.* (1988) and a solution for the resulting spatial and spectral distribution for a spherical symmetry with an outflow by the Hubble flow has been presented by Loeb *et al.*, 1999.

Today, Ly α spectra are often fit by shell models. Shell models consist of a central source emitting Ly α photons encircled by a moving shell of neutral hydrogen (Ahn, 2004; Verhamme *et al.*, 2006; Gronke *et al.*, 2015). The shells can be parametrized by relevant parameters such as their neutral hydrogen column density N_{HI} , dust optical depth τ_d , outflow velocity v_{exp} , intrinsic Ly α line width σ_i , and effective temperature T of the gas to reproduce observed Ly α spectra (Gronke, 2017).

The Ly α radiative transfer is sensitive to the kinematics and density structure. Clumpy multiphase media have been evaluated, which can heavily differ from the homogeneous case (Neufeld, 1991; Hansen *et al.*, 2006; Gronke *et al.*, 2016a; Gronke *et al.*, 2017b). This in particular has strong implications for the arising spectra and attenuation when incorporating dust. In uniform media, the optical depth for dust destruction vastly grows due to the distance travelled by Ly α photons until escape at optically thinner frequencies. However, if dust and hydrogen was bound in clumps, the scattering process, spectra and attenuation can vastly differ. As Ly α photons effectively start to trace out their environment in a succession of scatterings, they can “avoid” clumps and their dust. See Figure 2.9 for a visualization. There is a range of characteristic regimes set by the respective optical depths of hydrogen, dust and their clumpiness, see e.g. Gronke *et al.* (2016b, 2017b).

Similar to the clumpiness, anisotropies can be explored by the succession of scattering with Ly α photons “taking a path of least resistance”. Subsequently simple anisotropic toy models have been explored (Behrens *et al.*, 2014; Zheng *et al.*, 2014).

2.10.4 Moving beyond toy models

Moving beyond toy models, there has been a range of sophisticated computational explorations using hydrodynamic galaxy formation simulations to study Ly α emission. While early studies of Ly α emission did not treat the Ly α radiative transfer yet (Fardal *et al.*, 2001), full Ly α radiative transfer was soon incorporated (Cantalupo *et al.*, 2005; Tasitsiomi, 2006a; Laursen *et al.*, 2009).

Simulations have since covered a range of applications, such as Ly α emitters’ clustering and luminosity function (Zheng *et al.*, 2010, 2011a; Behrens *et al.*, 2018), the earliest galaxies (Smith *et al.*, 2019), exploration of Ly α halos (Faucher-Giguère *et al.*, 2010; Kollmeier *et al.*, 2010; Zheng *et al.*, 2011b; Mitchell *et al.*, 2021), and Ly α polarization signatures (Trebtsch *et al.*, 2016; Eide *et al.*, 2018).

In the future, we can hope for radiation-hydrodynamical galaxy formation simulations for larger representative samples to study LAEs/LAHs. Currently such simulations remain only feasible for individual zoom simulations (e.g. Rosdahl *et al.*, 2012; Mitchell *et al.*, 2021) as well as cosmological volumes, but only down to redshifts $z = 5-6$ (Gnedin *et al.*, 2017; Rosdahl *et al.*, 2018; Ocvirk *et al.*, 2020) above the best observed range for Ly α objects.

Another exciting area of research will be to run Ly α RHD simulations, coupling the hydrodynamics to the exerted Ly α radiation pressure, in the near future Smith *et al.* (2020). This will allow a better understanding of the Ly α radiation’s dynamical impact.

VOROILTIS: A $\text{Ly}\alpha$ RADIATIVE TRANSFER CODE FOR UNSTRUCTURED GRIDS

3.1 Introduction

$\text{Ly}\alpha$ Monte Carlo radiative transfer codes became increasingly popular in the last years. Similar to the scheme of Ahn *et al.* (2000), a variety of codes has been used in the last years, including IGMTRANSFER (Laursen, 2010a), TLAC (Gronke *et al.*, 2014), COLT (Smith *et al.*, 2015) and RASCAS (Michel-Dansac *et al.*, 2020). With the newest hydrodynamical codes for galaxy formation simulations such as AREPO (Springel, 2010) describing the gas distribution on unstructured grids, we aim to provide a $\text{Ly}\alpha$ radiative transfer code to provide such capabilities, being able to process large galaxy counts and cosmological volumes.

Our radiative transfer code VOROILTIS is based on an early version of ILTIS first used in Behrens *et al.* (2019). The latter code is based on stripped code initially presented in Behrens (2014). VOROILTIS is written in the C++ programming language using the Open Multi-Processing (OpenMP Board, 2020) and Message Passing Interface (MPI Forum, 2021) frameworks for multithreading and multiprocessing.

3.2 Monte Carlo procedure

We follow the common implementation of numerical radiative transfer in a Monte Carlo fashion (e.g. Tasitsiomi, 2006a). We describe the Monte Carlo procedure in three steps: First, photon packages are created. Second, these photon packages are propagated through the gas distribution and subsequently scattered until they are either destroyed by dust or another stop condition is reached. Third, at each scattering a peeling-off photon

is spawned that is propagated along the line-of-sight, integrating the optical depth that is encountered, until reaching an stop condition.

3.2.1 Emission

Photon packages are emitted from a list of predefined positions and luminosities. In earlier publications on LAEs (Behrens *et al.*, 2018; Byrohl *et al.*, 2019, 2020b), we emitted photons in the center of the LAE. Particularly for the latest publication on Ly α halos (Byrohl *et al.*, 2021), we assigned luminosities to each Voronoi cell by their defining global identifier. In the future, some fiducial emission descriptions can be implemented in the code for convenience, rendering the creation of pre-defined lists unnecessary.

Each photon is spawned in the rest-frame of the enclosing cell and its wavelength is set according to a chosen intrinsic spectrum. Here, we primarily use Gaussian and Delta distributions. We choose a random direction \vec{k} for propagation from an isotropic distribution.

Each photon package represents a luminosity of up to the luminosity specified in the input list. Depending on the problem to be tackled, different partitioning schemes for the luminosity across multiple photons are used. For example, as discussed in Section 6.2.3, we stick to an equal photon count irrespective of the cell's luminosity to properly trace the large dynamic range of surface brightnesses observed in simulated Ly α halos.

3.2.2 Propagation, scattering and destruction.

Upon initialization of each photon (and after each scattering), we assign an optical depth that the photon will travel upon interacting with dust or hydrogen. We draw the optical depth from the according exponential distribution

$$P = \exp[-\tau]. \quad (3.1)$$

We then propagate the photon within the underlying geometrical setup. For each step, we sum the optical depths $\Delta\tau$ for a photon stepsize Δs given by

$$\Delta\tau = n_{HI} \sigma(\nu, T) \Delta s, \quad (3.2)$$

where we calculate the frequency ν in the rest-frame of the volume element as

$$\nu = \nu \left(1 - \frac{\vec{k} \cdot \vec{v} + H(z)s}{c} \right) \quad (3.3)$$

accounting for the Doppler shift due to the element's bulk motion and the Hubble flow over the distance s since the last scattering of the photon.

The photon stepsize is commonly set by the underlying cells size. However, substepping can be involved within each cell. In particular, we use substepping when performing runs including the Hubble flow to linearly update the photons' frequency that can substantially change the cross-section in large cells.

The optical depth is integrated until the targeted optical depth is reached, after which an interaction event is calculated. In case the integrated optical depth exceeds that of the targeted depth after addition of a new step, we can linearly trace back to the point of interaction, which we adjust the photon's position to.

If dust is present in the simulation, we first determine whether a photon is destroyed by dust. First, we calculate the probability for interacting with dust as $f \equiv \Delta\tau_{0,dust}/(\Delta\tau_{0,dust} + \Delta\tau_{0,HI})$. The probability for destruction upon interacting with dust is given by $1 - A$, where A is the albedo, see Section 2.7. We draw two random numbers \mathcal{R}_1 and \mathcal{R}_2 from the uniform distribution within $[0, 1)$. If $\mathcal{R}_1 \leq f$ and $\mathcal{R}_2 > A$, destruction occurs and we terminate the photon. If $\mathcal{R}_1 < f$ but $\mathcal{R}_2 \leq A$, the photon is scattered by dust. The calculation of such scattering event follows the procedure for hydrogen scattering, albeit being simpler as we assume the cross-section for dust to be frequency independent over the relevant frequencies.

For the scattering event with hydrogen ($\mathcal{R}_1 > f$), we first need to determine the velocity component of the interacting hydrogen atom parallel to the photon's direction \vec{k} . The velocity distribution is determined by convolution of the line profile with the thermal one-dimensional Gaussian distribution. We will discuss the calculation of the interacting atom's parallel and perpendicular velocity components in Section 3.3.2.

Next, we shift the frequency into the atom's rest frame. We then draw an outgoing direction vector according to the chosen phase function. Finally, we shift the frequency out of the atom's rest frame. Optionally, we also perform a transform of the incoming and outgoing photon's direction upon switching frames.

After having obtained a new direction vector and a new target optical depth, the propagation of the photon continues, repeating the previous procedure. Photons that are not destroyed by dust will only finish their journey given a termination condition. In non-periodic volumes we commonly stop upon reaching its bounds. In periodic volumes, we terminate upon reaching a set distance from the photon's origin.

3.2.3 Peeling-off photons

As the probability for a photon to scatter and escape towards a distant observer situated in an angular element $d\Omega$ is infinitesimal, we spawn so-called “peeling-off photons” (Yusef-Zadeh *et al.*, 1984; Whitney, 2011). One or more peeling-off photons are spawned to directly propagate into predefined directions that we want to obtain a mock observation for. Each photon is propagated along its direction until hitting a non-periodic boundary or reaching a set distance from the scattering’s position. During propagation, we integrate the optical depth τ_{peel} that the photon experiences. Each “peeling-off photon” will calculate its luminosity upon termination as

$$L_{\text{output}} = L_{\text{input}} P(\mu) \exp[-\tau_{\text{peel}}] \quad (3.4)$$

where $P(\mu)$ is the phase function, L_{input} the scattering photon’s luminosity.

Each Monte Carlo tracer represents a photon package that carries relevant information for its propagation, scattering or destruction and can carry additional information fields such as initial frequency and the originating halo. Once a photon, whether scattering package or peeling-off package, reaches its end of lifetime in the simulation, this information can be written to harddrive as requested by the user.

3.3 Numerical approximations

3.3.1 Cross-section

The integration that defines the Voigt profile is not analytically integrable, which leads to approximative analytical forms being used. We rely on the approximation with accuracy to better than 1% at $T > 2$ K provided by Tasitsiomi (2006a). Different schemes with better accuracy and performance exist (e.g. see comparison in Michel-Dansac *et al.*, 2020). Particularly at high frequency shifts x , the speed-up can be significant with a factor of ~ 5 . As we find the code to be commonly memory bound, particularly for peeling-off photons being Hubble shifted in the IGM to those high frequency offsets, we expect performance improvements to be minor, but consider changes in the future.

3.3.2 Scattering

Redistribution

By default, we use either an isotropic or dipole phase function for the Ly α scattering to be set at compile time. For dust, we use the Greenstein phase function with $g = 0.7$, see Section 2.8.2. A more sophisticated phase function accounting for the different realistic behaviour for core and wing scatterings can easily be implemented, but for the optically thick environments we are interested in, we found little difference in the outcomes (e.g. also see Laursen, 2010b).

Atom velocity

Perpendicular components The perpendicular directions to the incoming photon's direction can be drawn from a Gaussian distribution each assuming thermal equilibrium. As commonly used, we can draw two independent random variates from a Gaussian distribution by calculating

$$u_{\perp,1} = (-\ln R_1)^{1/2} \cos(2\pi R_2) \quad (3.5)$$

$$u_{\perp,2} = (-\ln R_1)^{1/2} \sin(2\pi R_2) \quad (3.6)$$

for two random variates from a uniform distribution $R_i \in [0, 1)$ as demonstrated by Box *et al.* (1958).

Parallel component The parallel velocity component is given as convolution of a Gaussian with the Voigt line profile yielding Equation (2.59). Sampling this distribution is computationally challenging. Commonly, random variates are drawn by analytically integrating and inverting the PDF in question, the so-called transformation method (Press, 2007). By drawing a uniform number in the interval between zero and one and inserting this number into the inverse of the integrated PDF, we can draw a random variate according to the given PDF. As Equation (2.59) cannot be analytically integrated and inverted, this approach is not possible here. Instead, we use an approach commonly called rejection method (Press, 2007).

For a PDF $p(x)$ we want to draw from, we choose a comparison function $f(x)$ (i) whose integral we know analytically and which is invertible, and (ii) such that $f(x) > p(x)$. Using a uniform random number \mathcal{R}_1 in the interval $[0, 1)$, we can obtain a random variate x_1 from the normalized $f(x)$ via the transformation method. Drawing another random number \mathcal{R}_2 from $[0, 1)$, we obtain a random variate x_2 of $p(x)$ if latter random number

is not rejected. We reject and repeat the procedure for the case that $\mathcal{R}_2 > f(x_1)/p(x_1)$. See Press (2007) for more details and a visual explanation.

The efficiency of the rejection method strongly depends on the chosen comparison function $f(x)$. The closer $f(x)$ follows $p(x)$, the fewer rejections and thus computational time is needed. We adopt the rejection method using two rejection methods for the perpendicular direction as used by Zheng *et al.* (2002) with the boundary of the two functions' regime as chosen by Laursen (2010b) using numerical tests.

Acceleration scheme

In optically thick environments, the calculation of subsequent Ly α scatterings is responsible for the gross of the computational runtime. Different schemes have been proposed and implemented to reduce the number of scattering calculations in optically thick media. Methods include (1) constraining the probability density function for the drawn optical depth a Monte Carlo photon to lower values and biasing the photon weights to correct for this constraint, (2) replacing optically thick cells' radiative transfer with an analytic simulation similar to the Neufeld solution discussed in Section 2.10.3, (3) artificially increasing the frequency shift upon scattering to reduce core scatterings, and (4) solving the diffusion equation for the frequency and spatial redistribution (Cashwell *et al.*, 1959; Ahn *et al.*, 2002; Tasitsiomi, 2006b; Smith *et al.*, 2018).

While (4) might offer the best speed-up, we opt for approach (3), which achieves sufficient performance allowing us to perform the Ly α radiative transfer on galaxy formation simulations in cosmological volumes in post-processing.

This acceleration scheme, as suggested by Avery *et al.* (1968) and Ahn *et al.* (2002), skips interactions with low velocity atoms corresponding to those scatterings in the line profile's center. This is done by drawing the perpendicular velocity components from a truncated Gaussian distribution with a lower velocity limit. Given the numerical implementation in Equation (3.5), we obtain

$$u_{\perp,1} = (x_c^2 - \ln R_1)^{1/2} \cos(2\pi R_2), \quad (3.7)$$

$$u_{\perp,2} = (x_c^2 - \ln R_1)^{1/2} \sin(2\pi R_2), \quad (3.8)$$

with the critical dimensionless cutoff frequency x_c below which we are neglecting scatterings. We can use fixed values for x_c , which on the one hand can lead to inaccurate results in not very optically thick environments where the skipped scatterings would in reality lead to significant spatial diffusion. On the other hand, a fixed x_c can lead to an insuffi-

cient frequency shift at which the medium remains very optically thick. An alternative, adaptive scheme for x_c has been implemented by Laursen (2010b) who determined

$$x_c = \begin{cases} 0, & a\tau_0 \leq 1 \\ 0.02 \exp \left[0.6 \ln^{1.2}(a\tau_0) \right], & 1 < a\tau_0 \leq 60 \\ 0.02 \exp \left[1.4 \ln^{0.6}(a\tau_0) \right], & 60 < a\tau_0, \end{cases} \quad (3.9)$$

to fairly reproduce results with full radiative treatment. The criteria have been chosen relative to $a\tau_0$ as it is the shaping parameter describing the radiative transfer in homogeneous media in Section 2.10.3. A more aggressive, but seemingly still accurate choice has been proposed by Smith *et al.* (2015) with $x_c = 0.2(a\tau_0)^{1/3}$ for $a\tau_0 > 1$ for future review and improvement.

The acceleration scheme above only considers the line-center optical depth a photon experiences in a cell. In high resolution environments, many cells can trace very similar physical environments such that core skipping is applicable for a uniform environment across multiple cells. We follow the implementation by (Smith *et al.*, 2015): During initialization of the simulation, we integrate the line-center optical depth of evenly distributed rays in D directions from the cells' center starting from their boundary until reaching a threshold criterion $|\Delta n_{HI}/n_{HI}| < T$ fixing the gradient admissible. We follow the initial implementation's choice of $D = 6$ and $T = 0.5$. The optical depth entering the core acceleration scheme is then set as the sum of the local line-center optical depth of a photon to the nearest face and the minimum of the cells' D neighbouring contributions calculated during initialization.

3.4 Geometry handling and Voronoi tessellation

We aim to be agnostic concerning the underlying geometrical structure of the dataset the radiative transfer is performed on. This allows us to flexibly introduce support for new geometries.

The necessary functions of the geometry base class that need to be overwritten in a derived class are:

- **setup()**: Called during simulation initialization to load/compute required geometry. For example, this can include loading of data and balancing cell distribution onto different processes.
- **data_at()**: Passing a photon, this function returns three things: The cell properties such as density and temperature of the current cell, the status of the request (e.g.

informing us if the queried photon’s position is out of bounds), and the length that the photon continues to propagate in the current cell until reaching the next cell.

- **set_domain()**: For datasets that are distributed onto multiple processes, this routine will specify the unique process identifier for each photon that it belongs to. This routine is called on initial spawning of photons and in the communication of photons when they hit the boundary to a neighboring process’ domain.

There are various other functions that can be overwritten in the derived class to improve the code’s verbosity and performance. For example, implementation of **get_nearest_face()** allows a significant speed-up using a core-skipping algorithm (see Section 3.3.2).

Some of the latest generation hydrodynamical codes, such as AREPO (Springel, 2010), run on unstructured grids. In particular, the IllustrisTNG simulations of interest here use an unstructured Voronoi tessellation for representing the baryonic fields. Many of the current radiative transfer codes rely on classic approaches, e.g. by using an adaptive mesh refinement. Thus, the processing of many modern simulations requires an intermediate interpolation step. Given the potentially high resolution dependency for Ly α radiative transfer (e.g. Behrens *et al.*, 2018; Camps *et al.*, 2021), this requires sufficient sampling in the interpolation step. We found that the memory requirement drastically grows compared to original data set’s size when interpolating the Voronoi tessellation underlying the TNG simulations to a high-resolution adaptive mesh. For these reasons, we opted for implementing the radiative transfer directly on the unstructured grid, conserving the original resolution and yielding a significant reduction in memory overhead.

The Voronoi tessellation, named after Georges Voronoi (Voronoi, 1908), divides space¹ into a set of non-overlapping cells that for each point from a set of seed points the associated cell covers all space closest to said point. For running the radiative transfer on a Voronoi tessellated space, we start out with the seed points given in the snapshot of the simulation that we want to run the radiative transfer on. From this, we construct the list of direct neighbors, i.e. those seed points whose corresponding Voronoi cells share a face, see Section 3.4.1. This list constitutes all information necessary to run the radiative transfer on the Voronoi tessellated space: The cells’ faces are given as nearest surface elements of the equidistant planes to the direct neighbors. Any photon propagation from one Voronoi cell to another thus simply involves a loop of intersection operation over the equidistant planes to the direct neighbors, see Section 3.4.2. Figure 3.1 illustrates these two steps in preprocessing and during runtime starting from a set of seed points.

¹We restrict ourselves to three-dimensional space here, but the concept of the Voronoi tessellation is universally formulated for $n \geq 2$ dimensions.

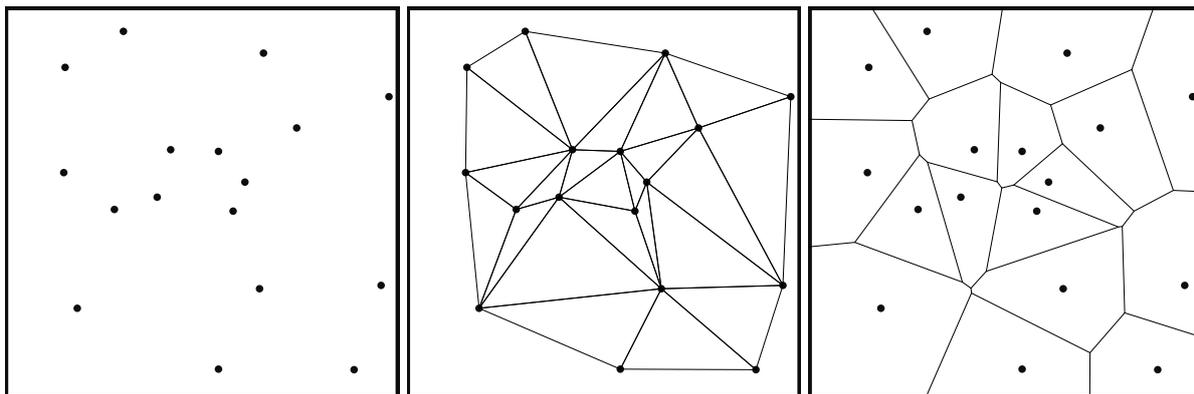


Fig. 3.1 Three different stages of the Voronoi tessellation for an example particle distribution in two dimensions are shown. **Left:** We show the seed points that the two dimensional rectangular area is to be partitioned with. **Middle:** For each seed point we construct the shown direct neighbors in a dedicated preprocessing step. **Right:** Saving only the direct neighbors allows a memory efficient reconstruction of the Voronoi faces from the orthogonal midpoint planes of the Delaunay links. The Voronoi faces are not explicitly calculated or held in memory. Instead, we only individually calculate the exit point of a ray propagating through a Voronoi cell from the Delaunay links.

3.4.1 Preprocessing

As the Voronoi tessellation needs to be only performed once for post-processing $\text{Ly}\alpha$ radiative transfer as done here, we wrote a stand-alone program to provide such tessellation. For this, we wrap functionality of the `voro++` library (Rycroft, 2009). The wrapper is necessary as latter library does not support domain decomposition across multiple nodes, which is required for the large simulations we aim to process. We use this wrapper, which we call `VOROPPP` (“`voro++` parallel”) to provide us with two lists: The first lists complements individual cell’s attributes such as position and density by adding the count of direct neighbors and a reference to the cell’s memory address in the second list. The second list contains the memory addresses of direct neighbors, that the cell shares a face with.

3.4.2 Photon propagation

For a photon’s propagation, in addition to the current floating-point position, we hold three additional integers for each photon: c_p, c_c, c_n holding the unique identifier of the prior, current and next Voronoi cell. There is a necessity for having at least one of those integers as we otherwise need to perform an expensive look-up every time we want to find the associated cell for a given floating point position. This is a consequence of unstructured meshes compared to structured ones, where we would not only know the relative position of cells in memory to the next neighbor, but can directly determine the

memory position from the floating point position. The additional two unique identifiers are necessary for speed-up and precision errors.

For optimal use of the processors' caches, the Voronoi cells are sorted using a Peano-Hilbert space filling curve (Peano, 1890; Hilbert, 1891) as used in Springel (2005, 2010).

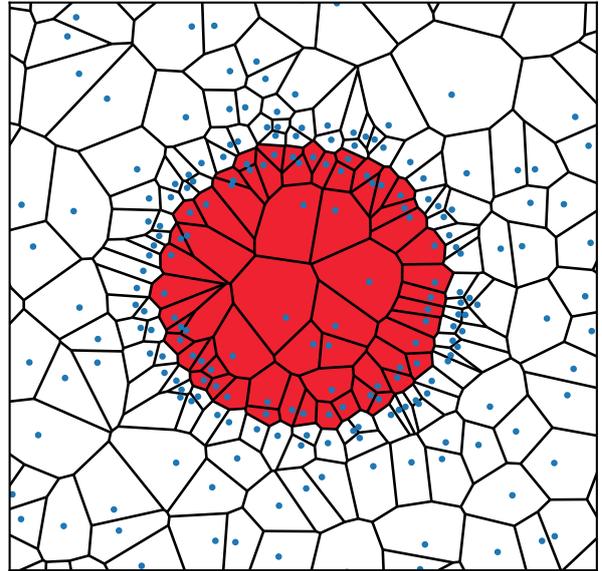
Given two Voronoi cells defined by their seeding positions p_1 and p_2 we specify the equidistant normal plane through the midpoint of the two seeding positions $p_0 = \frac{p_1+p_2}{2}$ and its normal vector $n = \frac{p_2-p_1}{|p_2-p_1|}$. For a photon at position l_0 in cell 1 propagating into direction k , the distance d needed for the photon to enter the cell 2 is

$$d = \frac{(p_0 - l_0) \cdot n}{l \cdot n}. \quad (3.10)$$

We determine the next cell seeded by p_2 by calculating d for all neighbors of c_c . The neighbor for which d is minimal but positive is the next neighbor. Once we calculated the next cell for the photon, we save the corresponding identifier as c_n . This will allow a faster look-up of the neighbor if the photon does not change direction.

3.5 Code verification

Fig. 3.2 We explicitly test the Ly α radiative transfer in the implemented Voronoi geometry by comparing to the analytical Neufeld solution, which describes the spectra to result from an homogeneous slab/sphere (Neufeld, 1990; Dijkstra *et al.*, 2006). For this, we uniformly draw random seed points. All random points are assigned a density ρ_0 inside of radius r_0 and density zero outside of it. Additionally, we add random points close to r_0 to better achieve better radial symmetry tracing the density jump. We show a two-dimensional Voronoi tessellation from a slice of the tested particle distribution.



As part of the code verification, we run an homogeneous sphere of radius r_0 with line-center optical depth τ_0 for which we know the analytic solution (Neufeld, 1990; Dijkstra *et al.*, 2006). We run the problem in a meshfree geometry and a reproduction of

the analytic setup in the Voronoi geometry as shown in Figure 3.2 and find an excellent match with the analytical solution.

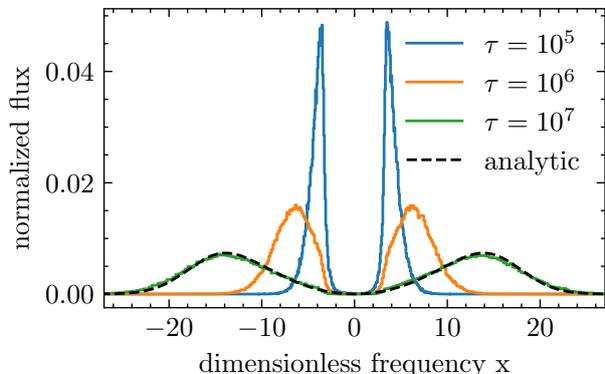


Fig. 3.3 Resulting spectrum for a homogeneous hydrogen distribution within a sphere of optical depth $\tau = \{10^5, 10^6, 10^7\}$ in the Ly α line center from the sphere's center and a constant temperature $T = 20000$ K, where the photons are emitted. The dashed lines shows the analytic solution derived in Dijkstra *et al.* (2006).

3.6 Scaling and performance

A hybrid parallelization, consisting of multithreading and multiprocessing, was adopted. Multithreading is done with OPENMP allowing a trivial shared-memory access to the underlying spatial geometry and propagated photons. Multiprocessing is done with MPI enabling communication not just between different processes but also across computing nodes. Each process holds a convex subvolume of the simulation's overall volume. Any tests have been performed on two socket systems with Intel Xeon Gold 6138 CPUs (20 cores each with hyperthreading deactivated; 27.5MB L3 cache) at 2.0 GHz and accompanied with 192GB of RAM.

3.6.1 Multithreading

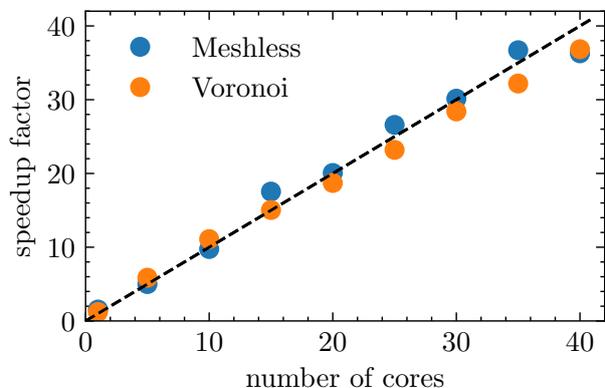


Fig. 3.4 Speedup factor as a function of used cores with one thread per core normalized to performance at 5 cores in the meshless geometry. The speedup initially drops to around 75% from a 1-to-1 scaling moving from single-core to $n \geq 2$ cores. Once multiple cores are used, scaling appears near perfect. We show the Neufeld test both on a meshless and Voronoi geometry. Performance appears equal.

```

1 #pragma omp parallel for schedule(dynamic, 100)
2   for (unsigned long i = 0; i < particles.size(); i++) {
3     // Propagate Photons
4   }

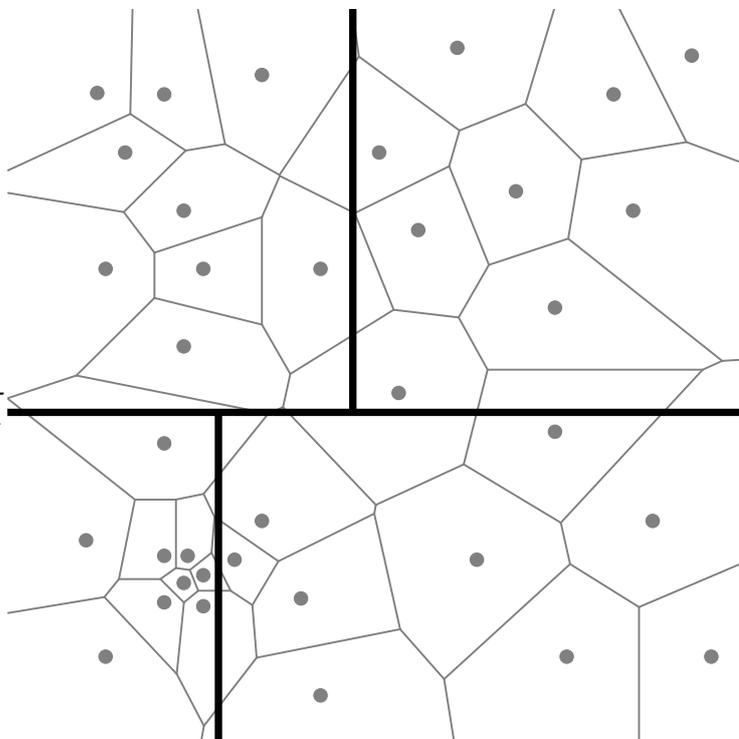
```

Listing 3.1 OpenMP parallelization of the main loop.

We achieve parallelization using a shared-memory model using OPENMP by simplistically chunking the iterations of the main loop onto different processes. Each iteration propagates an individual photon. This allows a trivial parallelization with the interface’s directives, see the schematic code in Listing 3.1. In Figure 3.6.1, we show the speedup factor for the Neufeld solution as a function of used cores with one thread per core. The solution has been calculated on a meshless spherical geometry and a Voronoi mesh distribution. The code scales well with the thread count. However, we find that in certain scenarios, the code is bound by the L3 cache such that the speedup appears limited for larger core counts.

3.6.2 Multiprocessing

Fig. 3.5 The Voronoi cells’ seed points are sorted into a balanced k-d tree. We descend the construct the tree levels until the total leaf cell count equals the amount of MPI ranks available. Alternatively, domains can be manually be specified rather than determined through the tree construction.



In order to process large, high-resolution volumes, we need to distribute the required memory for the underlying gas structure and the Monte Carlo photons. Each process is

allocated a convex subvolume upon the simulation's initialization, see Figure 3.5. For this volume, the process loads the required Voronoi cells. Initially, photons are spawned on the process that holds the containing subvolume. Upon reaching the subvolume's borders, the photon is put onto a communication stack. This communication stack is periodically exchanged with the neighboring subvolumes' processes. Currently, communication occurs synchronous with all processes. In the future, we will implement an asynchronous exchange between neighbouring processes to reduce idle time. We introduced a varying overlap between the subvolumes. Photons propagate up to the overlapping region before being communicated. This heavily reduces communication due to photons in optically thick regions close to the subvolumes' faces that would otherwise scatter forth and back between processes.

3.7 Data analysis

While on-the-fly reduction, e.g. compilation of spatial maps around halos, is technically possible, this is not guaranteed to reduce memory requirement given that we are effectively handling high-dimensional sparse data points. Using the underlying parallel filesystem and available harddrive memory on high-performance computing systems, we instead write the individual photon contributions directly to harddrive.

Working with the photon contributions directly has two advantages: (i) We can interactively modify the data products we try to analyze and thus explore the simulations in more detail and (ii) the simple data structure of lists with each list containing a certain property of all contributions allows efficient memory access and easy parallelization.

We first introduce the conversion from Monte Carlo contributions to observables and sketch the technical implementation of this conversion second.

3.7.1 Observables

Implicitly defining the luminosity distance d_L to recover the flux of an object with luminosity L at redshift z and the angular-diameter distance d_A giving the angular extent of an object of physical size D at redshift z as

$$F = \frac{L}{4\pi d_L^2(z)}, \quad \theta = \frac{D}{d_A(z)}, \quad (3.11)$$

we can calculate the surface brightness as

$$\text{SB} = \frac{F}{\theta^2} = \frac{L}{4\pi D^2} (1+z)^{-4}. \quad (3.12)$$

Note that while d_L and d_A require knowledge (and integration) of the cosmological model, the surface brightness only requires knowledge of the redshift z in addition to an observed patch's luminosity and physical size. The factor $(1+z)^{-4}$ is commonly referred to as cosmological surface brightness dimming (Mo *et al.*, 2010).

Each Monte Carlo photon carries its wavelength offset $\Delta\lambda$ from the Ly α line-center at the position of last scattering. All photons are redshifted according to the Hubble flow to the same line-of-sight distance. This allows us to create spectra and also to determine the respective redshift-space positions. With the spectral information, we can also calculate the specific intensity.

With the ability to calculate the fluxes or surface brightnesses and spectral offsets, we can calculate a range of increasingly reduced observables such as three-dimensional intensity cubes, two-dimensional surface brightness maps, surface brightness radial profiles, spectra, luminosities, luminosity functions, LAE clustering et cetera.

3.7.2 Computation

Often, the voroILTIS simulations and the underlying hydrodynamical simulations can span multiple terabytes of data. In addition, their reduction can take substantial computational time depending on the desired data product. An easy and generic approach for parallelization across multiple nodes is thus desirable. For fast prototyping and interactive exploration, we rely on the Python programming language. For Python, we found such desired approach in the Dask package (Dask Development Team, 2016). Dask is composed of two parts: Dynamic task scheduling for computational workload and integration for existing data interfaces such as numpy (Harris *et al.*, 2020) and pandas (McKinney, 2010; Pandas Development Team, 2020). However, workload is not limited to existing data interfaces, so that we utilize the task scheduling for custom workload using graphics processing units (GPUs) via numba and CuPy (Lam *et al.*, 2015; Okuta *et al.*, 2017).

Using Dask allows us to scale up the underlying computing power were needed. In addition, the analysis greatly benefits from Dask's use of task graphs: we can create a large amount of "recipes" by stringing together a range of operations to obtain physically relevant quantities. These recipes represent a graph of required IO operations and computations organized in a graph that allows to request the calculation when needed.

Also, we can easily adjust recipes, e.g. by replacing the peeling-off photons by the input photons when in need of the intrinsic emission.

3.8 Application to Illustris(TNG)

For this thesis, we apply the presented radiative transfer code to the Illustris and IllustrisTNG (“TNG”) cosmological hydrodynamical galaxy formation simulations. IllustrisTNG is the successor of the former. While some differences in the physics model exist, the overall physics remains similar. We will introduce the simulations in Section 3.8.1, shortly summarize the gas physics shaping the temperature and ionization state of hydrogen in Section 3.8.2, and the required pre-processing for application with the Ly α radiative transfer code in Section 3.8.3.

3.8.1 Overview

The Illustris simulations (Genel *et al.*, 2014; Vogelsberger *et al.*, 2014a; Vogelsberger *et al.*, 2014b; Sijacki *et al.*, 2015), completed by 2013 and made public in 2015, aimed at providing a sufficiently large galaxy sample at a competitive resolution and a comprehensive physics model to study a wide range of astrophysical processes and observational phenomena.

Illustris and TNG are run with the AREPO code (Springel, 2010), which solves the coupled equations of self-gravity and hydrodynamics with a ‘moving mesh’ discretization based on an unstructured Voronoi tessellation of space. Implemented physical processes has been described in Vogelsberger *et al.* (2013) and Torrey *et al.* (2014). All simulations include models for the physical processes most important for galaxy formation including primordial and metal-line cooling, ionization and heating from the metagalactic background radiation field, star-formation above a density threshold, stellar population evolution and chemical enrichment following supernovae Ia, II, and AGB stars, and the seedind, merging and growth via accretion of supermmasive black holes (SMBHs).

The IllustrisTNG simulations (hereafter, TNG; Marinacci *et al.*, 2018; Naiman *et al.*, 2018; Nelson *et al.*, 2018; Pillepich *et al.*, 2018a; Springel *et al.*, 2018) are a series of three large-volume magnetohydrodynamical cosmological simulations of galaxy formation succeeding Illustris. The simulation are named TNG50, TNG100 and TNG300 in line with their approximative comoving linear size in megaparsecs. Besides covering different volume sizes and including magnetohydrodynamics (Pakmor *et al.*, 2011; Pakmor *et al.*, 2013), an adjusted physics model (Weinberger *et al.*, 2017; Pillepich *et al.*, 2018b) has been run, including a number of important changes. There are two major changes to the existing Illustris small-scale physics model. First, the galactic-scale winds launched

by stellar feedback have been revised (Pillepich *et al.*, 2018b), which impacts the gas (and stellar) contents of low mass galaxies in both their ISM and CGM. Second, TNG includes a two-mode blackhole feedback operating in a thermal ‘quasar’ state at high accretion rates and a kinetic ‘wind’ state at low accretion rates. The latter is a new model for low-state SMBH feedback, in the form of a time stochastic, directionally variable, high-velocity kinetic wind (Weinberger *et al.*, 2017). The cosmological parameters have been updated from WMAP-9 (Hinshaw *et al.*, 2013) to a parameter set consistent with Planck collaboration results (Planck Collaboration *et al.*, 2016) with $\Omega_{\Lambda,0} = 0.6911$, $\Omega_{m,0} = 0.3089$, $\Omega_{b,0} = 0.0486$, $\sigma_8 = 0.8159$, $n_s = 0.9667$ and $h = 0.6774$. Unless specified different, as for Chapter 4 where we use the Illustris cosmology, we use these values given here.

3.8.2 Gas state

The temperature and ionization state of hydrogen – crucial to Ly α emission and scattering – is computed within Illustris(TNG) incorporating primordial cooling following Katz *et al.* (1996) with additional metal-line cooling from CLOUDY cooling tables. Both metal and primordial cooling are further modified by the assumption of a uniform, time-varying UV background using the intensities given in Faucher-Giguère *et al.* (2009) for photoionization and photoheating. The ionizing radiation is attenuated by a self-shielding factor based on Rahmati *et al.*, 2013. Given their large impact, an effective prescription for local ionizing flux from active galactic nuclei (AGN) is incorporated (Vogelsberger *et al.*, 2013). The AGN prescription and its impact onto the gas’ ionization state Ly α radiative transfer is discussed in detail in Section 6.6.1.

While radiation-hydrodynamical (RHD) simulations consistently incorporating ionizing radiation would be preferred to properly capture the hydrogen’s ionization and temperature state, this remains computationally unfeasible for the targeted redshifts requiring evolution of the simulations far beyond the end of the epoch of reionization. To date, only zoom-ins down to those redshifts, or cosmological volumes up to completion of the epoch of reionization exist. Hence, for a statistical sample, we need to rely on purely hydrodynamical simulations with approximative on-the-fly schemes such as in Illustris(TNG). We consider the impact of ionizing sources in Section 6.6.1.

The ionization of hydrogen is set by the respective recombination and ionization rates. In equilibrium, we have

$$\Gamma_{eH_0} n_e n_{H_0} + \Gamma_{\gamma H_0} n_{H_0} = \alpha_{H^+} n_e n_{H^+} \quad (3.13)$$

balancing the collisional ionization and photoionization with the recombinations, where n_e is the electron number density, n_{H_0} the neutral hydrogen number density, and n_{H^+} the ionized hydrogen number density. α_{H^+} is the recombination coefficient for ionized hydrogen, Γ_{eH_0} is the collisional ionization rate for neutral hydrogen, and $\Gamma_{\gamma\text{H}_0}$ is the photoionization rate in the presence of a radiation field. The temperature dependent form of the former two coefficients is approximated as in Cen (1992) for Illustris(TNG). The photoionization rate has been chosen according to description of the redshift dependent metagalactic background contribution described in Faucher-Giguère *et al.* (2009).

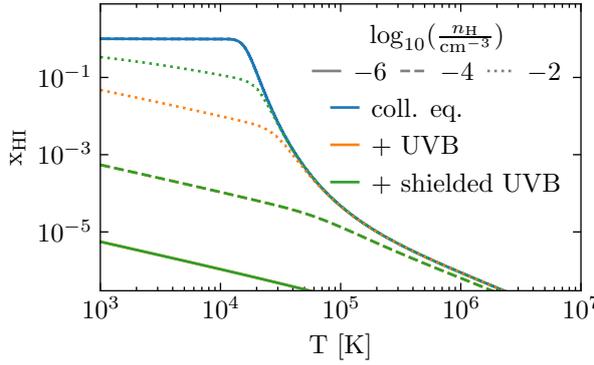


Fig. 3.6 Neutral hydrogen fraction as a function of temperature for different hydrogen densities (different line styles) at $z = 3$. We calculate the ionization in collisional equilibrium (blue), with added UVB contribution (orange), and added UVB contribution including self-shielding (green). Note that the UVB lines with and without self-shielding overlap at $n_{\text{H}} = 10^{-4} \text{ cm}^{-3}$ and $n_{\text{H}} = 10^{-6} \text{ cm}^{-3}$.

With Equation (3.13) we can calculate the neutral hydrogen fraction $x_{\text{HI}} = \frac{n_{\text{H}_0}}{n_{\text{H}_0} + n_{\text{H}^+}}$ as

$$x_{\text{HI}} = \frac{\alpha_{\text{H}^+}}{\alpha_{\text{H}^+} + \Gamma_{\text{eH}_0} + \Gamma_{\gamma\text{H}_0}/n_e}. \quad (3.14)$$

where the electron density for a primordial gas

$$n_e = n_{\text{H}^+} + n_{\text{He}^+} + 2n_{\text{He}^{++}}, \quad (3.15)$$

which enters Equation (3.14), relies on the respective ionization state of hydrogen and helium. Here, we omit the description of the helium ionization. The neutral hydrogen fraction can then be iteratively solved for with the according full set of equations (see Katz *et al.*, 1996). In Figure 3.6 we show the neutral hydrogen fraction as a function of temperature and density for collisional ionization equilibrium and optionally with a (self-shielded) ultraviolet background as present in TNG.

3.8.3 Preparation

In Figure 3.1, we list the simulations used for the respective chapters/publications. In Chapter 4, we rely on radiative transfer simulations that were not yet run using voroILTIS. An intermediate interpolation step from the Voronoi tessellation to an adaptive mesh

| Chapter (Publication) | Simulation |
|--|------------|
| Chapter 4 (Byrohl <i>et al.</i> , 2019) | Illustris |
| Chapter 5 (Byrohl <i>et al.</i> , 2020b) | TNG100 |
| Chapter 6 (Byrohl <i>et al.</i> , 2021) | TNG50 |

Table 3.1 Simulations used by respective chapter/publication.

refinement (AMR) structure was thus necessary to run with an existing radiative transfer code. For latter publications Byrohl *et al.* (2020b, 2021), a native Voronoi tessellation is used, which we reconstruct from the seeding points of cells from available snapshots. In addition, we define a list of positions for which we spawn photons. For Byrohl *et al.* (2019, 2020b), we compile a list of star-forming galaxies from whose center we spawn photons. This has been replaced for the procedure introduced in Byrohl *et al.*, 2021 that spawn photons for all Voronoi cells to incorporate diffuse emission.

LAE CLUSTERING: A NEW FINGERS-OF-GOD DAMPING

The content of this chapter has been published as Byrohl et al. (2019).

Context and summary

Complex radiative transfer (RT) of the Lyman- α photons poses a theoretical challenge to galaxy surveys which infer the large-scale structure with Lyman- α emitters (LAEs). Guided by RT simulations, prior studies investigated the impact of RT on the large-scale LAE clustering and claimed that RT induces a selection effect that results in an anisotropic distortion even in real space but in an otherwise negligible effect in redshift space. However, our previous study, which relies on a full radiative transfer code run on the Illustris simulations, shows that the anisotropic selection effect was drastically reduced with higher spatial resolution. Adopting the same simulation framework, we further study the impact of RT on the LAE clustering in redshift space. Since we measure the LAE's radial position through a spectral peak of the Lyman- α emission, the frequency shift due to RT contaminates the redshift measurement and hence the inferred radial position in redshift space. We demonstrate that this additional RT offset suppresses the LAE clustering along the line of sight, which can be interpreted as a novel Fingers-of-God (FoG) effect. To assess the FoG effect, we develop a theoretical framework modeling the impact of the RT similar to that of the small-scale peculiar velocity, which is commonly studied in the context of redshift-space distortions (RSD). Although our findings strongly encourage a more careful RSD modeling in LAE surveys, we also seek a method to mitigate this FoG effect due to RT by using additional information of the Lyman- α spectra.

4.1 Introduction

Current and future redshift surveys detecting galaxies with prominent Lyman- α emission, so-called Lyman- α emitters (LAEs), can give competitive astrophysical and cosmological constraints. For example, the Hobby-Eberly Telescope Dark Energy Experiment (Hill *et al.*, 2008; Adams *et al.*, 2011, hereafter HETDEX) is currently operating and will eventually detect close to a million of LAEs in a redshift range of $1.9 \leq z \leq 3.5$ over a sky patch of 450 deg^2 (Leung *et al.*, 2017). HETDEX will measure cosmic expansion history and growth history of matter fluctuations through baryon acoustic oscillations (BAOs) and redshift-space distortion (RSD), and also shed more light on the properties of star-forming galaxies by offering a range of statistical measures.

However, HETDEX and other future Lyman- α redshift surveys might have to deal with severe modifications of the detected spatial clustering signal due to the complex radiative transfer of Lyman- α given its resonant nature and high optical depth in astrophysical environments. Often Lyman- α photons, particularly those produced within the star-forming regions of galaxies, scatter many times in the interstellar medium (ISM) and the circumgalactic medium (CGM) before reaching the observer. Additionally, column depths of neutral hydrogen outside the galaxy's host halo might be sufficient to substantially attenuate the remaining flux in the intergalactic medium (IGM) by scattering photons out of the line of sight.

Scatterings of neutral hydrogen change both the frequency and position of photons before escaping towards the observer. This can introduce new distortion effects in the cosmological signal both in real space and redshift space. In real space, this corresponds to a selection effect favoring certain LAEs to be detected over others based on their environment, introducing both isotropic and anisotropic modifications to the two-point statistics. An isotropic distortion effectively corresponds to changing the bias due to emitters of different mass ranges being favorably detected. Similarly, the detection of emitters might be affected by their large-scale environment, which can also give rise to anisotropic distortions as demonstrated by Zheng *et al.* (2011a). While these real-space distortions automatically propagate into the redshift-space signal, there can be additional distortions purely arising in redshift space when the radiative transfer modifies the spectral features from which the line-of-sight position is inferred. Similar to the selection effects, this can cause both isotropic and anisotropic distortions in the two-point statistics.

The complexity of the radiative transfer limits analytic solutions to symmetric toy models. In contrast, realistic environments require explicit radiative transfer simulations, which are usually run as a post-processing step given the expensive numerical cost. Prior simulations of LAEs in their large-scale structure environment examining possible

distortion effects include Zheng *et al.* (2010)/Zheng *et al.*, 2011a, Behrens *et al.*, 2013 and Behrens *et al.*, 2018 running a radiative transfer code on top of hydrodynamical simulations tracing neutral hydrogen distribution down to scales of \sim kpc (for latter publication), but no further subgrid model is assumed below those scales. More studies of LAEs in large cosmological volumes such as Inoue *et al.* (2018) and Gurung-López *et al.*, 2019 exist that however use a different approach: A dark matter only N-body simulation is used to map the large-scale structure and identify halos that possibly contain LAEs. Different sophisticated semi-analytical models are used to approximate the spectra arising on the ISM/CGM scales that remain unresolved in the simulations themselves. After calibration of such a semi-analytical model, a good match with observations can be obtained.

Zheng *et al.* (2010)/Zheng *et al.*, 2011a find a strong anisotropic selection effect occurring in the real-space clustering signal caused by a correlation of the observed flux with the large-scale velocity gradient. This selection effect has been challenged by Behrens *et al.*, 2013 and Behrens *et al.*, 2018, who could not reproduce such effect. Behrens *et al.* (2018) show that prior findings might have been strongly overestimated due to a lack of spatial resolution and a simplified emitter model.

All these studies are primarily concerned with real-space distortions, either not evaluating (Behrens *et al.*, 2013, 2018) or not finding (Zheng *et al.*, 2010, 2011a) additional effects in redshift space. In this work, we revisit the idea of possible additional redshift space distortions of LAEs in redshift space. We reuse the radiative transfer simulations run by Behrens *et al.* (2018) for the analysis, covering a redshift range from $z = 2.0$ to $z = 5.85$ as described later.

The structure of this chapter is as follows. In Section 4.2, we describe the theoretical background needed to model the newly found distortion described later in the chapter. Afterwards, in Section 4.3, we introduce the radiative transfer simulations performed by Behrens *et al.* (2018) used here and how to reduce them to mock catalogs of LAEs. In Section 4.4, we present results of detected spectra and inferred positions before showing how these affect the two-point statistics. In Section 4.5, we summarize our findings along with a discussion of possible shortcomings in our findings and how future surveys might be affected by/corrected for radiative transfer redshift-space distortions before concluding in Section 4.6.

4.2 Theoretical background of RSD

In this section, we provide a brief review of the galaxy clustering in redshift space. In real space, which we considered in our previous work (Behrens *et al.*, 2018), the fluctuation in the number density of LAEs for a given sample is given by

$$1 + \delta_g(\vec{x}) = \frac{n_g(\vec{x})}{\bar{n}_g}. \quad (4.1)$$

Then we consider the two-point statistics to characterize the strength of the clustering signal, i.e., the correlation function, ξ_g , or the power spectrum, P_g , as its Fourier-counterpart mapped by the Fourier transform (FT)¹:

$$1 + \xi_g(\vec{r}) = \langle \{1 + \delta_g(\vec{x})\} \{1 + \delta_g(\vec{x} + \vec{r})\} \rangle, \quad (4.4)$$

$$(2\pi)^3 P_g(\vec{k}) \delta_D(\vec{k} + \vec{k}') = \langle \delta_g(\vec{k}) \delta_g(\vec{k}') \rangle, \quad (4.5)$$

where statistical homogeneity is implicitly assumed. In real space where statistical isotropy is given if no selection effect is present, the arguments of the two-point statistics depend only on the scale, i.e., $\xi_g(r)$ and $P_g(k)$.

Now let us consider contamination of *any form* of velocity, \vec{v} , along a line of sight (LOS) of a galaxy in measuring its redshift and hence its radial position. Mathematically this is equivalent to a mapping from real to redshift space:

$$\vec{s} = \vec{r} + \frac{\vec{v} \cdot \hat{r}}{aH(a)} \hat{r}, \quad (4.6)$$

where \hat{r} is a unit vector along the LOS direction, a is the scale factor of the Universe, and $H(a)$ is the Hubble expansion rate. It has been extensively discussed in the literature that the peculiar velocity of a galaxy contaminates its redshift-space position, making the clustering pattern anisotropic known as RSD (see e.g., Hamilton, 1998, for a review). This occurs simply because this is the effect only along the LOS, which breaks statistical isotropy.

¹We adopt the following convention for the FT:

$$A(\vec{k}) = \int d^3x e^{i\vec{k} \cdot \vec{x}} A(\vec{x}), \quad (4.2)$$

$$A(\vec{x}) = \int \frac{d^3k}{(2\pi)^3} e^{-i\vec{k} \cdot \vec{x}} A(\vec{k}). \quad (4.3)$$

The complexity to accurately model the two-point statistics in redshift space originates from the fact that the Jacobian of the mapping in Eqn. 4.6 is *nonlinear* in terms of the velocity, v . In the following, let us discuss the impact of the nonlinear mapping on the nonlinear power spectrum and the correlation function as general as possible. For this purpose, we begin with the following expression of the redshift-space density field (Taruya *et al.*, 2010) which is *exact* under the global plain-parallel approximation (i.e., the LOS is fixed with one global direction as $\hat{r} \approx \hat{z}$, see Beutler *et al.*, 2014):

$$\delta_g^s(\vec{k}) = \int d^3x \{ \delta_g(\vec{x}) - f \partial_z u_z(\vec{x}) \} e^{i\vec{k} \cdot \vec{x} + i f k_z u_z(\vec{x})}, \quad (4.7)$$

where we introduce a scaled velocity, $\vec{u} \equiv \vec{v}/(faH)$, and $f \equiv d \ln D / d \ln a$ is the linear growth function. We specifically denote a quantity in redshift space with a superscript ‘s’ throughout this chapter. We then find an expression for the redshift-space power spectrum as

$$P_g^s(\vec{k}) = \int d^3r e^{i\vec{k} \cdot \vec{r}} \langle e^{-i f k_z \Delta u_z} \times \{ \delta_g(\vec{x}) + f \partial_z u_z(\vec{x}) \} \{ \delta_g(\vec{x}') + f \partial_z u_z(\vec{x}') \} \rangle, \quad (4.8)$$

where $\vec{r} \equiv \vec{x} - \vec{x}'$ and $\Delta \vec{u} \equiv \vec{u}(\vec{x}) - \vec{u}(\vec{x}')$. Eqn. 4.8 apparently involves higher-order correlations between the density δ_g and the velocity field u_z . Linearizing Eqn. (4.8) in terms of δ_g and u_z yields

$$P_g^{s,L}(\vec{k}) = (1 + f\mu^2)^2 P_g(k), \quad (4.9)$$

where μ is the cosine of an angle between \vec{k} and the LOS, defined as $\mu \equiv k_z/k$. This equation, well known as the Kaiser formula (Kaiser, 1987), shows that the clustering in redshift space is more enhanced closer to the LOS direction, which is a valid picture on large scales and nothing but the main target of RSD measurements (see e.g., White *et al.*, 2009). On the other hand, Eqs. (4.7) and (4.8) imply that even random velocity suppresses the redshift-space clustering on small scales along the LOS, often quoted as the Finger-of-God (FoG) effect (Jackson, 1972). To see this more explicitly, let us rewrite Eqn. (4.8) in terms of the cumulants as (Scoccimarro, 2004; Taruya *et al.*, 2010)

$$P_g^s(\vec{k}) = \int d^3r e^{i\vec{k} \cdot \vec{r}} \exp \left\{ \langle e^{-i f k_z \Delta u_z} \rangle_c \right\} \times \left\{ \langle e^{-i f k_z \Delta u_z} \mathcal{A}(\vec{x}) \mathcal{A}(\vec{x}') \rangle_c + \langle e^{-i f k_z \Delta u_z} \mathcal{A}(\vec{x}) \rangle_c \langle e^{-i f k_z \Delta u_z} \mathcal{A}(\vec{x}') \rangle_c \right\}, \quad (4.10)$$

where $\mathcal{A}(\vec{x}) \equiv \delta_g(\vec{x}) + f\partial_z u_z(\vec{x})$ is used just to simplify the notation. As Zheng *et al.* (2016) pointed out, the overall exponential factor, $\exp \left\{ \langle e^{-ifk_z \Delta u_z} \rangle_c \right\}$, contains terms which depend only on the one-point cumulants. These terms survive even when two-point correlations such as $\langle \Delta u_z \mathcal{A} \rangle$ are zero and can be integrated out because it no longer depends on the scale. That is to say,

$$\begin{aligned} & \exp \left\{ \langle e^{-ifk_z \Delta u_z} \rangle_c \right\} = \exp \left\{ \sum_{n=1}^{\infty} (-ifk_z)^n \frac{\langle \Delta u_z^n \rangle_c}{n!} \right\} \\ = & \exp \left\{ \sum_{m=1}^{\infty} (-ifk_z)^{2m} \frac{2 \langle u_z(\vec{x})^{2m} \rangle_c}{(2m)!} \right\} \\ & \times \exp \left\{ \sum_{m=1}^{\infty} (-ifk_z)^{2m} \frac{\langle \Delta u_z^{2m} \rangle_c - \langle u_z(\vec{x})^{2m} \rangle_c - \langle u_z(\vec{x}')^{2m} \rangle_c}{(2m)!} \right\}. \end{aligned} \quad (4.11)$$

We have used the fact that the terms with odd power in the second and third lines vanish because of symmetry in a galaxy pair. The exponential factor in the second line of Eqn. (4.11) does not depend on the scale. For example, if u_z follows a Gaussian distribution with zero mean and variance of σ_u^2 , only $m = 1$ term survives in the exponential factor, corresponding to the FoG damping factor commonly assumed:

$$D_{\text{FoG}}^{\text{Gaussian}}(k, \mu) = e^{-f^2 k^2 \mu^2 \sigma_u^2}. \quad (4.12)$$

Notice that, since the FoG damping factor depends only on the one-point cumulants, it can be derived at the level of the density field, Eqn. (4.7). Namely, if the velocity field follows a probability distribution function (PDF), $P(u_z)$, we have

$$D_{\text{FoG}}(k, \mu) = |\langle e^{ifk\mu u_z} \rangle|^2 = \left| \int du_z P(u_z) e^{ifk\mu u_z} \right|^2 \quad (4.13)$$

that is the FT of the one-point PDF, $P(u_z)$ (e.g., Hikage *et al.*, 2016). Another common velocity PDF is an exponential distribution (e.g., Scoccimarro, 2004). The FT of such distribution of the $\exp(-\sqrt{2}|u_z|/\sigma_u)/(\sqrt{2}\sigma_u)$, is a Lorentzian damping function²

$$D_{\text{FoG}}^{\text{Lorentzian}}(k, \mu) = \left\{ \frac{1}{1 + f^2 k^2 \mu^2 \sigma_u^2} \right\}^2. \quad (4.14)$$

²We note that the exponential PDF is often adopted for the *pairwise* velocity PDF (see Eqn. (4.15)) rather than for the velocity PDF (Davis *et al.*, 1983; Ballinger *et al.*, 1996). In this case, there is no square factor in the damping factor in Eqn. (4.14). We avoid this choice because the pairwise velocity PDF is generally scale-dependent and hence its mean and dispersion are not necessarily constants.

Again, we note that the PDFs $P(u_z)$ are assumed for the one-point (rather than pairwise) velocity distributions. This results in a different interpretation of the σ for the Gaussian and the exponential distribution in the literature (e.g. Scoccimarro, 2004).

Here we stress that the FoG damping inevitably arises as long as the velocity field has a non-zero dispersion and higher-order moments. For instance, Agrawal *et al.* (2017) confirmed the damping due to the nonlinear mapping even assuming a linear velocity field. In addition, the two-point correlations between the density and velocity fields (i.e., the second and third lines in Eqn. (4.10)) are essential to model the nonlinear redshift-space power spectrum accurately, as several authors have shown (see e.g., Taruya *et al.*, 2010; Okumura *et al.*, 2012; Matsubara, 2014; Vlah *et al.*, 2019; Zheng *et al.*, 2019).

So far, we have shown that, in Fourier space, the FoG damping factor depending only on one-point PDF can be expressed as an *overall multiplicative* factor. In the following, let us instead discuss the configuration space as a complementary approach. The two-point correlation function (TPCF) in redshift space is generally written as (e.g., Scoccimarro, 2004)

$$1 + \xi_g^s(\vec{s}) = \int d\pi \{1 + \xi_g(r)\} \mathcal{P}(u_z; \vec{r}), \quad (4.15)$$

where $\pi \equiv s_z - u_z$ is the vector along the LOS direction, \hat{z} in configuration space. $\mathcal{P}(u_z; \vec{r})$ is the *pairwise* velocity PDF given by the FT of the pairwise velocity generating function $M(if\gamma, \vec{r})$:

$$\mathcal{P}(u_z; \vec{r}) = \int \frac{d\gamma}{2\pi} e^{-i\gamma u_z} \mathcal{M}(if\gamma; \vec{r}), \quad (4.16)$$

$$\mathcal{M}(if\gamma; \vec{r}) = \frac{\langle \exp(if\gamma \Delta u_z) [1 + \delta_g(\vec{x})] [1 + \delta_g(\vec{x}')] \rangle}{1 + \xi(r)}. \quad (4.17)$$

An advantage of Eqn. (4.15) is that the redshift-space correlation function can be expressed only in terms of quantities in real space. However, the complexity arises because of a convolution with the pairwise velocity PDF, which is weighted by density fields at two points and hence scale-dependent. At the linear level, the mean of the pairwise velocity PDF is related to coherent infall motion and hence the Kaiser factor, whereas its dispersion is related to the velocity power spectrum (Fisher, 1995). Similarly to Fourier space at nonlinear level, however, one has to take into consideration the correlation between the density and velocity fields as well as the contribution from the one-point PDF, as several authors have recently studied (see e.g., Scoccimarro, 2004; Reid *et al.*, 2011; Uhlemann *et al.*, 2015; Bianchi *et al.*, 2016).

As we will explain in detail in the next section, we investigate another velocity offset due to the RT effect in addition to the peculiar velocity of a galaxy. We will study the

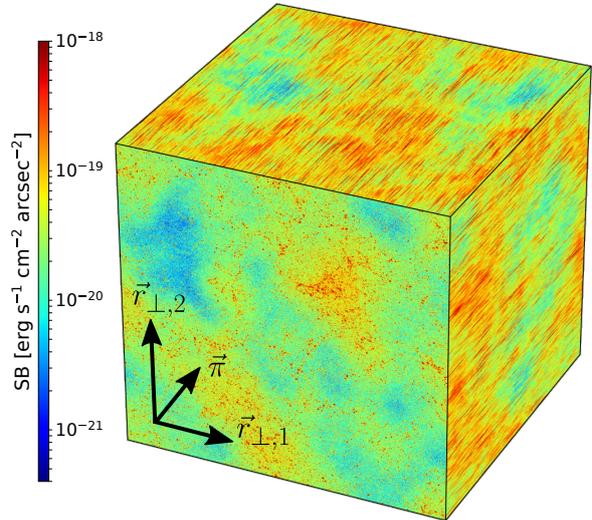
impact of the RT velocity component on the redshift-space clustering, mainly focusing on the FoG damping factor in Fourier space, $D_{\text{FoG}}(k, \mu)$, and the pairwise velocity PDF.

4.3 Methods

4.3.1 Radiative transfer simulations

We utilize previous radiative transfer simulations from Behrens *et al.* (2018) that have been run with a predecessor of the radiative transfer code presented in Chapter 4.3. We run a Monte Carlo radiative transfer code of Ly α photons on top of the Illustris simulations presented in Section 3.8 at redshift outputs of 2.00, 3.01, 4.01 and 5.85 with its box size of $L_{\text{box}} = 75 \text{ cMpc/h}$. The Illustris simulations provide a distribution of galaxies and their neutral hydrogen content in a context of the large-scale structure, necessary for studying the impact of Lyman- α radiative transfer onto statistics used in cosmology. Before running the radiative transfer, we convert the Voronoi tessellation in Illustris onto an octet-tree data structure with a maximal resolution of $\Delta = 3.3 \text{ ckpc}$. The refinement criterion is triggered for cells containing 32 or more Voronoi cells' defining positions. More details on the processing of the Illustris datasets can be found in Behrens *et al.* (2018).

Fig. 4.1 Lyman- α intensity map of the simulation box with a length of 75 Mpc/h at a redshift of 3.01 projected onto the cube's faces. $\vec{\pi}$ denotes the line-of-sight direction, while $\vec{r}_{\perp,1}$ and $\vec{r}_{\perp,2}$ indicate the perpendicular direction to the former. Thus, the top and right cube faces show a directional alignment stemming from redshift-space distortions. Most of the visible distortion is due to Lyman- α radiative transfer and the subject of this chapter. Individual spectra of LAEs are later reconstructed from such a cube.



We then explicitly place Ly α photons in the center of the dark matter halos and weight them by the halos' respective Ly α luminosity that is based on star formation rate (SFR) of each halo according to Eqn. (2.47) (see Section 2.6.2).

This equation only considers Lyman- α radiation after recombination locally sourced by ionizing radiation in the star-forming regions. The proportionality constant can vary by a factor of a few and depends on made assumptions such as the initial mass function,

Table 4.1 In this table, we summarize key quantities of the redshift snapshots considered relevant to our analysis: Spatial resolution Δ , number of LAEs considered N_{LAE} and median radius $r_{\text{crit},200}$ encompassing 200 times the critical density of the Universe for those emitters for the post-processed snapshots. Besides the radius' physical size, we also state the angular size as seen for an observer. For each redshift, we also state the average neutral fraction, f_{IGM} , at a characteristic hydrogen number density of 10^{-4} cm^{-3} . We also quote the conversion factor $(\text{aH})^{-1}$ from the peculiar velocity to the comoving distance at each redshift.

| redshift | 2.00 | 3.01 | 4.01 | 5.85 |
|---|--------|--------|--------|--------|
| $f_{\text{IGM}} [10^{-5}]$ | 2 | 3.7 | 6.8 | 35 |
| $\Delta [\text{pkpc}]$ | 1.2 | 0.8 | 0.7 | 0.5 |
| N_{LAE} | 45594 | 45434 | 39782 | 23114 |
| $r_{\text{crit},200} [\text{pkpc}]$ | 40.6 | 30.4 | 24.3 | 17.8 |
| $(\text{aH})^{-1} [\frac{\text{Mpc h}^{-1}}{\text{km s}^{-1}}]$ | 0.0105 | 0.0094 | 0.0085 | 0.0073 |

metallicity, and fraction of recombinations radiating Lyman- α photons (e.g. Furlanetto *et al.*, 2005). For this work, the value of the proportionality constant is irrelevant as we detect emitters at a fixed number density threshold. We use the SFR given by the Illustris simulation, which is given as its instantaneous value at given redshift and computed close to the multi-phase description in Springel *et al.* (2003). Note that we assume properties of the LAEs such as positions, velocities, and SFR only from host halo catalogs (i.e., 'group' catalogs) and hence ignore satellite galaxies. We impose a minimum threshold of $0.1 M_{\odot} \text{ yr}^{-1}$ on the SFR and $10^{10} M_{\odot}$ on the halo mass to limit ourselves to well-resolved halos and to limit the required computational resources. We summarize the characteristics of our simulated LAEs in Table 4.1.

Additionally, we set the initial frequency profile emerging from the unresolved ISM to be a Gaussian whose width σ_i is set by the virial temperature of the halo, see Behrens *et al.* (2018). While the initial profile should have a significant impact on the observed properties, we lack a profound ISM sub-grid model for the scope of this chapter. For the fiducial sample of LAEs at a number density of $n_{\text{LAE}} = 10^{-2} \text{ Mpc}^{-3} \text{ h}^3$ in our survey, we find a mean input width $\sigma_i = 137 \text{ km}^{-1}$ at $z = 3.01$. This is roughly consistent with recent findings in shell models by Gronke (2017) finding $\sigma_i = 172_{-60}^{+75} \text{ km s}^{-1}$ fitted to an observed MUSE LAE sample at median redshift of $z = 3.83$.

Varying the Gaussian widths for $T < T_{\text{vir}}$, we found that this only has an insignificant impact on the spectra emerging after reprocessing on CGM scales. As cosmological simulations such as Illustris are unable to resolve the ISM regions, we explicitly cut out the unresolved ISM as defined by a hydrogen number density threshold of 0.13 cm^{-3} for the gas. Also, we ignore the impact of dust attenuation on the radiative transfer for

simplicity. Given these simplifications, we do not expect our simulations to agree well with the observed luminosity function as already discussed in Behrens *et al.* (2018).

After the luminosity-weighted photons are spawned with an isotropic angular distribution in the LAE's rest frame, the photons are propagated in a straight line until a scattering with a neutral hydrogen atom occurs. Then at each scattering point, the attenuated contribution along the line of sight towards the observer is computed ('peeling-off' photon) while the original photon is re-emitted and propagated/scattered subsequently.

In comparison to the simulations presented in Behrens *et al.* (2018), the only modification in our RT simulation stems from an increased initial Monte Carlo photon count: we increase this count from 100 to 1000 to properly sample the spectra as a function of wavelength. The requirement for the photon count was less important before as only the total flux was relevant for the analysis in Behrens *et al.* (2018).

As a result of those RT simulations, catalogs of attenuated photon contributions reaching the observer are created, including information such as the observed intensity as a function of wavelength and position perpendicular to the line of sight and the positions of photons' originating LAE. In Figure 4.1, we visualize the reprocessed Lyman- α photons escaping the simulation box and project the surface brightness onto the cube's faces. The cube's top and right face contain the line-of-sight direction and are shown in redshift space. One can easily notice a strong anisotropy in redshift space, which will be the focus of this study. The position of the individually observed LAE along the line of sight however depends upon a detection algorithm whose methodology we introduce in the next subsection.

4.3.2 Analysis of simulated LAEs in redshift space

LAE spectra and redshift-space positions

To determine the position of LAEs in redshift space, we calculate the flux and spectral information by applying a spherical aperture of 3 arcseconds radius (our default case) around a known LAE's position (from the halo catalogs), which also already fixes the angular position of the selected LAE. The aperture size is chosen to correspond roughly to the size of the host halos in our sample (4.8-3.0'' for redshifts 2.0 to 5.85; see Table 4.1). We consider contributions only from the targeted source, allowing us to separate out any issues due to source confusion (unlike a real observation). We stress that this detection algorithm is different from the one in our previous work in real space (Behrens *et al.*, 2018) where we adopted an adaptive Friends-of-Friends (FoF) grouping only along the directions perpendicular to the line of sight. The previous FoF algorithm naturally

leads to source confusion, making an interpretation of the redshift-space clustering more complicated.

The spectrum of an LAE is computed relative to the comoving frame at the line-of-sight distance of a given emitter. We impose a fixed spectral resolution in terms of the velocity shift as 24.7 km/s ($R \sim 12000$). We found this resolution to be sufficient for our analysis and the shot noise for chosen Monte Carlo photon count to be negligible. As we show later, the resulting spectra I_λ will have one or more peaks. In this chapter, we consider two localization methods for the line-of-sight position, s_{app} , of the LAEs: Either by using the global spectral maximum at λ_{max} or by the spectral maximum at $\lambda_{\text{max,red}}$ only considering the red wing relative to the LAE's rest frame, i.e., $\Delta\lambda > 0$.

Once a peak has been identified, this allows us to define a corresponding apparent line-of-sight velocity of the emitter as

$$v_{\text{app}} = c \cdot \frac{\lambda_{\text{max}} - \lambda_{\text{Ly}\alpha}}{\lambda_{\text{Ly}\alpha}}, \quad (4.18)$$

where c is the speed of light, $\lambda_{\text{Ly}\alpha}$ the rest frame Lyman- α line center wavelength ($\sim 1216\text{\AA}$). The comoving position in redshift space is given by Eqn. (4.6), i.e.,

$$\vec{s}_{\text{app}} = \vec{r} + \frac{v_{\text{app}}}{aH} \hat{r}. \quad (4.19)$$

The apparent line-of-sight position of the emitter is set by adding the apparent velocity to the position from the halo catalogs. This completes the localization of the emitters, and we construct mock LAE catalogs containing the position of detected emitters in redshift space, their velocity, and spectra.

While we focus on the distortions to the line-of-sight component of the position, we also tested the distortion that RT induces for the angular component. Except for source confusion, we found deviations from the LAEs' actual positions to be negligible compared to those arising in the line-of-sight direction.

Wavelengths and frequencies in this chapter are always evaluated at the emitters' cosmological redshift. It is important to note that two distinct velocities are contributing to v_{app} : the peculiar velocity of a halo, v_{pec} , and the velocity offset induced by the radiative transfer, v_{RT} . While v_{RT} is not a physical velocity as it only represents the shift of Lyman- α photons from the line center, we sometimes refer to it short-handedly as radiative velocity for the sake of brevity. We can estimate v_{RT} from measured v_{app} and v_{pec} from the halo

catalogs as

$$v_{\text{app}} = \vec{v}_{\text{pec}} \cdot \hat{r} + v_{\text{RT}} \quad (4.20)$$

Velocities and positions without vector notation correspond to the magnitude along the line of sight if not stated otherwise. The same decomposition is also done for $\lambda_{\text{max,red}}$, so that the radiative transfer velocity offset is inferred from the peak in the red part of the line center relative to the halo's rest frame.

While we choose the host halos' velocity for the peculiar velocity, we checked that the qualitative reasoning remains the same for other choices such as the linear velocity field or the velocity of the star-forming regions. We will show the convenience and significance of this velocity decomposition to determine v_{RT} in Section 4.4.2 and 4.4.3.

Measuring the two-point clustering statistics

Next, we use the LAE catalogs created in Section 4.3.2 to compute the two point statistics, $\xi_g(\vec{r})$ and $P_g(\vec{k})$.

We use `halotools` (Hearin *et al.*, 2017)³ to compute the TPCF of our LAE samples with the Landy-Szalay estimator (Landy *et al.*, 1993)

$$\xi_g(\vec{r}) = \frac{\text{DD}(\vec{r}) - 2\text{DR}(\vec{r}) + \text{RR}(\vec{r})}{\text{RR}(\vec{r})}, \quad (4.21)$$

where $\vec{r} = \vec{x} - \vec{x}'$ denotes the separation between a pair of emitters. DD, DR and RR represent LAE-LAE, LAE-random and random-random pair counts found at the given separation for a given spatial binning width $\Delta r = \frac{5}{6}$ Mpc/h. The pair separation can either be evaluated in real space or redshift space. In real space, we expect an isotropic clustering when ignoring RT, so that for such case we can characterize $\xi(\vec{r})$ as $\xi(r)$ with $r \equiv |\vec{r}|$. As the line-of-sight positions change in redshift space, we express the signal as a function of parallel (π) and perpendicular (r_{\parallel}) separation to the line-of-sight component.

To estimate $P_g(\vec{k})$, we make use of the Fast Fourier Transformation (FFT). For this purpose, we first assign LAEs to a three-dimensional grid with 512 cells in each direction (i.e., the Nyquist frequency is $k_{\text{Nyq}} \sim 21.4$ h/Mpc.) with the Triangle Shape Cloud interpolation. Next, we perform the FFT to obtain the LAE number density field on the grid in Fourier space and then measure the power spectrum as

$$P_g(\vec{k}) = \frac{1}{N_{\text{modes}}} \sum_{\vec{k} \text{ in bin}} \delta_g(\vec{k}) \delta_g(\vec{k})^*, \quad (4.22)$$

³<http://halotools.readthedocs.io/en/latest/>

where N_{modes} is the number of Fourier modes on the grid which fit within a range of a given bin, e.g., $[k - \Delta k/2, k + \Delta k/2]$ and $[\mu - \Delta\mu/2, \mu + \Delta\mu/2]$. We suppress the aliasing effect by applying the interlacing technique as presented in Sefusatti *et al.* (2016) and also subtract out the Poisson shot noise which properly takes the interlacing correction into account.

Another two-point statistics is the pairwise velocity distribution $\mathcal{P}(u_z, \vec{r})$, which encodes all the information in real space required to describe induced RSD (see Eqn. 4.15) on the clustering signal. We compute the distribution by

$$\mathcal{P}(u_z, \vec{r}) = \frac{\text{DD}(u_z, \vec{r})}{\text{DD}(\vec{r})}, \quad (4.23)$$

i.e. counting the direct LAE-pairs within at a given separation $r \in [r - \Delta r/2, r + \Delta r/2]$ and line-of-sight velocity $u_z = \hat{r} \cdot (\vec{v}_2 - \vec{v}_1) \in [u_z - \Delta u_z/2, u_z + \Delta u_z/2]$. The sign convention is chosen such that an infalling motion corresponds to a positive pairwise velocity (also across periodic boundaries). For the binning, we chose $\Delta r = 1.0$ Mpc/h and $\Delta u_z = 60$ km/s, because the number of pairs for central galaxies of interest here quickly goes down below these scales.

4.4 Results

4.4.1 Spectra

Figure 4.2 shows a random selection of LAE spectra for the simulated redshift range between $z = 2.0$ and $z = 5.85$. The spectra are evaluated with respect to the comoving rest frame at the emitters' position. Most emitters at redshifts $z = 2.00$ and $z = 3.01$ show a characteristic double-peaked spectrum, which is expected for optically thick environments as shown in fully homogeneous and isotropic analytic toy models (Adams, 1972; Harrington, 1973; Neufeld, 1990) or more recent simulated isotropic shell models for small offset velocities (Ahn *et al.*, 2003; Verhamme *et al.*, 2008).

At $z = 4.01$, double-peaked spectra become significantly sparser and are only exceptional cases at $z = 5.85$. This redshift evolution is mostly related to the decreasing transmissivity in the IGM at higher redshifts (Laursen *et al.*, 2011) due to the increasing neutral hydrogen density, explaining the disappearing of blue peaks at high redshift and a stronger trough around the line center. At the same time, the spatial luminosity profiles become more extended as scattering in the CGM increases with redshift. This trend shows up in Figure 4.2, as relative contributions to the flux at fixed finite aperture become lower at higher redshift.

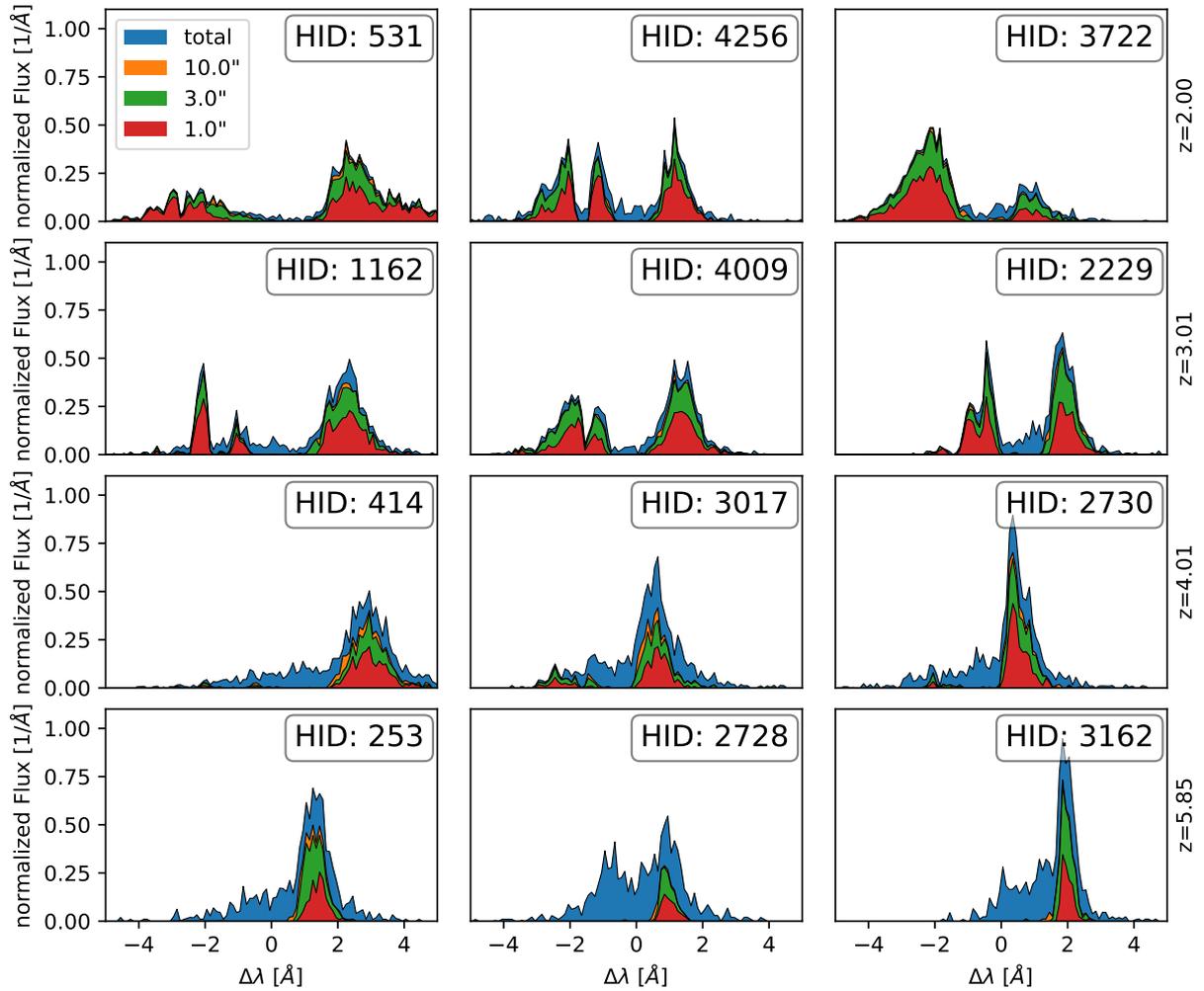


Fig. 4.2 Randomly selected spectra at different redshifts in the comoving rest frame at the emitters' position. The halo IDs (HIDs) are given with respect to the position in Illustris' halo catalogs for given redshift. Different colors show the flux for different aperture sizes. As larger apertures always enclose the flux of smaller apertures, only the excess flux over flux from smaller apertures is shown. The spectra show plenty of different characteristics, with double peaks being common for lower redshifts and rare at high redshifts. Emission becomes more diffuse at higher redshift as apparent from larger flux contributions from larger apertures.

Our simulations seem to overestimate the abundance of double-peaked emitters compared to observations at high redshifts $z = 2.0 - 3.0$, where the IGM interaction is limited. Typically observations find fractions of $\sim 30\%$ at these redshifts in star-forming galaxies (Kulas *et al.*, 2012; Herenz *et al.*, 2017), while most of our emitters show double peaked emitters. We shortly discuss the shortcomings in the modeled spectral shape and thus also the overprediction in the double-peaked profiles in Section 4.5.2.

More complex spectra are present depending on the underlying HI distribution and velocity structure, such as emitters with $n \geq 3$ peaks, particularly in the blue part of the spectrum. See HID 4256 at $z=2.0$ or HID 1162 at $z=3.01$ for such examples.

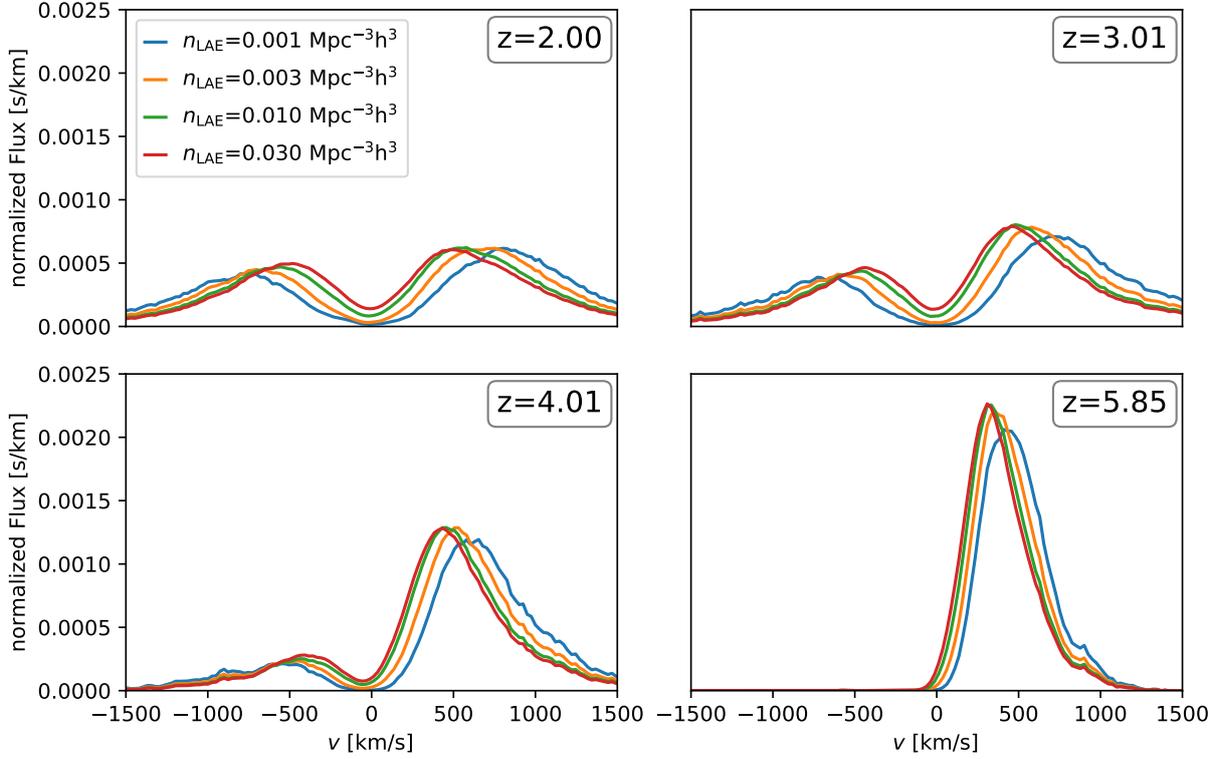


Fig. 4.3 Stacked spectra in the halos' frame for different number densities and redshifts for an aperture radius of 3 arcseconds. The overall flux is normalized with respect to the sample of chosen number density n_{LAE} .

The redshift evolution becomes particularly apparent when computing the stacked profiles I_{stacked} of the emitters as shown in Figure 4.3. Here we vary redshifts and LAE number densities n_{LAE} for which we consider only the $n_{\text{LAE}} L_{\text{box}}^3$ emitters with the highest apparent luminosity. Imposing a number density threshold has several advantages when compared to a surface brightness threshold, including independence from the proportionality constant in the luminosity model in Eqn. (2.47) and controlled shot noise behavior for measuring LAE power spectra. The spectra are stacked in the halos' rest frame. Stacking in the halo's rest frame reveals a trough at $v = 0$ km/s in the spectra caused by a combination of IGM attenuation and strong frequency diffusion into the wings due to high optical depths.

As the number density n_{LAE} is reduced, the peaks of the stacked profiles move to higher offsets. At the same time, the dispersion σ_{stacked} , the stacked profiles' second central moment for the respective red or blue spectral peak, also slightly increase but only

on the percent level. We motivate the significance of this dispersion in Section 4.5.2. The change in σ_{stacked} mostly stems from a significant correlation between brighter sources and more massive sources, which in turn have a larger trough/peak separation.

4.4.2 Distinguishing distortion contributions

Fig. 4.4 Contour plot of emitters' peculiar and radiative velocities. Detection of the global peak (**blue**) and the peak in the red wing of the emitters' rest frame (**orange**).

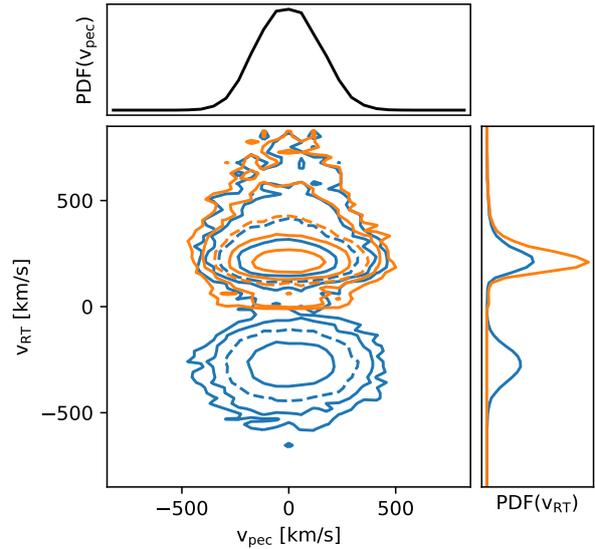


Figure 4.4 shows the one-point PDF, $P(v)$, decomposed into v_{RT} and v_{pec} for $z = 3.01$ as suggested in Eqn. (4.20). The projected PDFs onto the v_{pec} and v_{RT} axes give the one-point PDF, $P(v_{\text{pec}})$, and $P(v_{\text{RT}})$, respectively. There is no noticeable correlation between the two velocity components in Figure 4.4. This is also true for the other redshifts (not shown here). This result is expected to some extent since two physically distinct processes are responsible for each velocity component. As we discuss later, this independence allows us to model additional RSD due to the RT effect independently of the peculiar velocity.

4.4.3 RT velocity PDF

As discussed in Section 4.2, the velocity PDF, i.e. a probability distribution of the radial position of LAEs with respect to the real-space position (see e.g., Eqn. (4.13)), determines the redshift-space clustering. We show the PDF of the peaks for the brightest emitters detected by the observer for a given LAE number density n_{LAE} in Figure 4.5. The peak distributions show a strong redshift evolution: As we move to higher redshifts, the blue peaks are strongly suppressed and practically non-existent at $z=5.85$. When decreasing the number density, the selected emitters are restricted to the more luminous ones, which in turn have a higher average peak offset due to their higher optical depth.

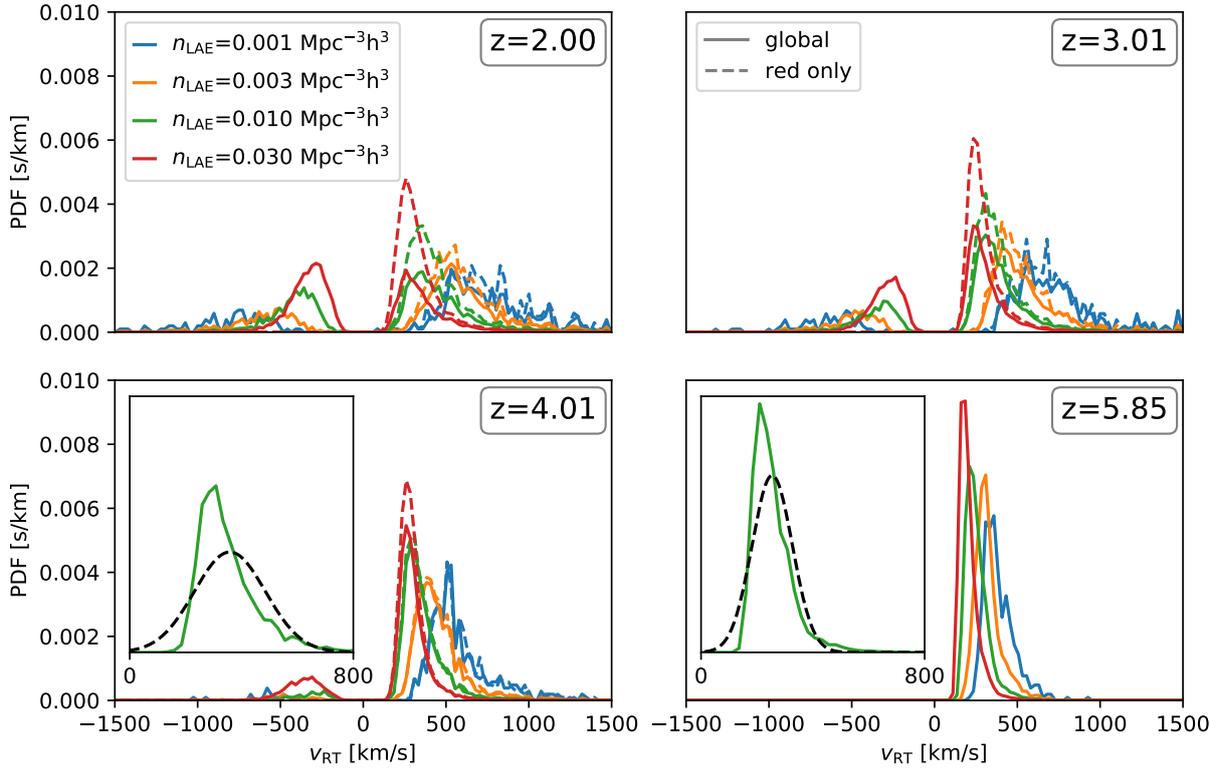


Fig. 4.5 Radiative transfer velocity offset $v_{\text{RT,global}}/v_{\text{RT,red}}$ distributions for different number densities n_{LAE} and redshifts. Solid lines show the velocity offset being deduced from the global peak of each LAE's spectra and dashed lines with the velocity offset deduced from the red peak only. For redshift $z = 3.01$ and 4.01 we show the distribution of $v_{\text{RT,global}}$ for the red peak at $n_{\text{LAE}} = 0.01 \text{ Mpc}^{-3} h^3$ along with a Gaussian of matching mean and standard deviation. The scaling of the y-axis remains the same as for the other subplots. The velocity distributions for the respective wings are positively skewed with respect to the velocity offset to the line center.

Although we have already seen similar trends in the stacked spectra as expected, it is important to notice that the velocity PDFs and the stacked spectra are *not* the same. We discuss a possible correlation between the distribution of stacked spectra and the velocity PDF in Section 4.5.2. Since only the stacked spectra are directly observable, such relation will prove itself crucial in estimating the additional radiative transfer effect in observational surveys stemming from the velocity PDF.

In Figure 4.5, we also provide a comparison of the v_{RT} -PDF with a Gaussian of the same mean and variance for the red peaks. We find that the distributions quickly fall off towards the line center while they extend more towards larger velocity offsets. This leads to a higher kurtosis than for the Gaussian and also a positive skewness, which will affect the quality of using a Gaussian approximation for the damping in the next sections.

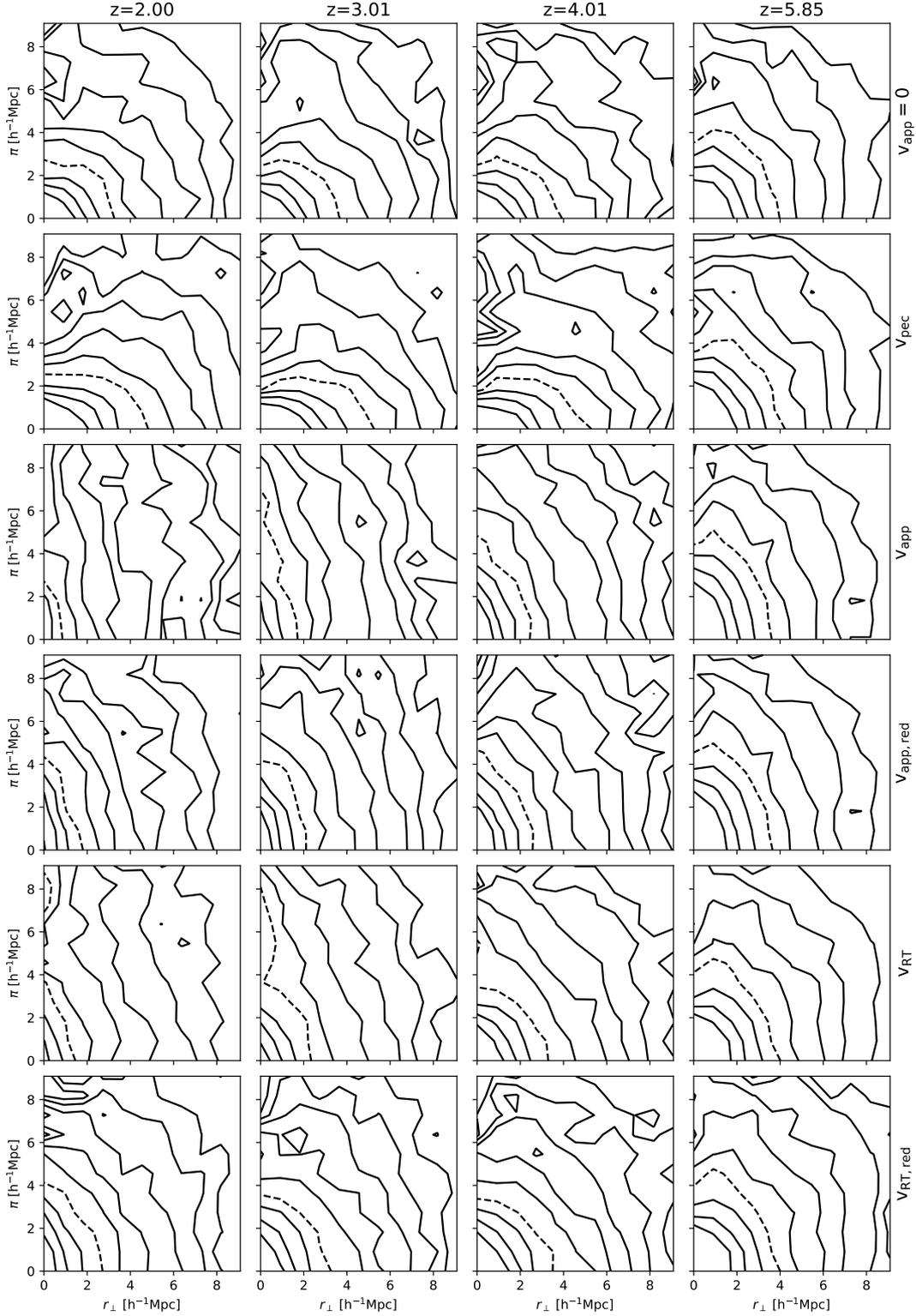


Fig. 4.6 TPCF for disentangled RSD of visible LAEs for $n_{\text{LAE}} = 10^{-2} \text{ h}^3 \text{ Mpc}^{-3}$ in redshift space. The dashed contour corresponds to $\xi = 1$ with contours decreasing by a factor of 1.4 further away from the origin. In the first row, we set the apparent velocity v_{app} to zero, i.e. we show the real-space clustering. In the second row, we only consider the peculiar velocity v_{pec} of the corresponding host halo. The third and fourth rows show the apparent overall velocity shifts v_{app} detected from the global peak and the red peak, respectively. The last two rows show the clustering when only the radiative transfer contribution is considered.

4.4.4 Configuration space: TPCF and the pairwise velocity PDF

In Figure 4.6 we show the correlation functions $\xi(\pi, r_{\parallel})$ measured from the mock observations with a LAE number density of $n_{\text{LAE}} = 0.01 \text{ h}^3/\text{Mpc}^3$. Different columns show the clustering at different redshifts in increasing order, while different rows show different velocity contributions (as defined in Section 4.3.2) added onto the real-space configuration to obtain the redshift-space clustering. In the first row, we set the velocity contribution v_{app} to zero so that we plot the real-space clustering. The second row shows the redshift-space result using the peculiar velocity v_{pec} from the halo catalogs and thus explicitly omitting the contribution from radiative transfer. We stress that the second row is often presented as the redshift-space clustering of LAEs in the literature (Zheng *et al.*, 2011a; Gurung-López *et al.*, 2019) but does not yet directly correspond to an observable radial position from the redshift measurement. Instead, the apparent v_{app} is the observable containing both contributions from the complex radiative transfer and the peculiar velocity. The third/fourth ($v_{\text{app}}/v_{\text{app,red}}$) row shows inferred overall position from the peaks in the spectra, which includes both peculiar velocity and radiative transfer effects. The fifth/sixth ($v_{\text{RT}}/v_{\text{RT,red}}$) rows show the radiative transfer component of the velocity only as the residual of the apparent and peculiar velocity.

Given the same simulation setup as in Behrens *et al.* (2018), we expect a very similar result to those shown for the real-space clustering in the first row. Only minor differences arise from an increased Monte Carlo photon count and the simplified detection algorithm. Note that there is a slight anisotropy in the clustering signal in real space. As we stressed in Behrens *et al.* (2018), the slight anisotropy of this dataset did not originate from a radiative transfer effect and was statistically consistent with zero.

As introduced in Section 4.2, we confirm two competing RSD effects in Figure 4.6. We see that on the shown scales ($1 - 10 \text{ h}^{-1} \text{ Mpc}$), the Kaiser effect dominates the redshift-space distortions from the peculiar velocity field v_{pec} over the usual FoG effect due to the random motion of the LAEs in the second row. However, once v_{RT} is added, the small-scale damping from v_{RT} is significant on these scales, and thus the overall redshift-space clustering with the apparent velocities v_{app} is elongated along the line of sight despite the squashing from the Kaiser effect (third and fourth row).

To investigate the distortions of the TPCF in redshift space more quantitatively, we discuss the pairwise velocity PDF, which encodes the full information of RSD (see Eqn. (4.15)). Note that we do not report the measurement of the multipole moment as we did in Behrens *et al.* (2018). Before showing the pairwise velocity PDF, let us first extend the discussion in 4.2 in the presence of two physically distinct velocity contaminations v_{pec} and v_{RT} (see Eqn. (4.20)). In Sec. 4.4.2, we show that there is no apparent correlation

between v_{pec} and v_{RT} at the level of the one-point PDF, i.e., $\langle v_{\text{pec}}(\vec{x})v_{\text{RT}}(\vec{x}) \rangle = 0$. We further assume that v_{RT} has no correlation with the density field or the peculiar velocity at scales of interest, i.e., $\langle u_{\text{RT}}(\vec{x})\delta_g(\vec{x}') \rangle = \langle u_{\text{RT}}(\vec{x})u_{\text{pec}}(\vec{x}') \rangle = 0$. Under this simple setting, Eqn. (4.17) follows that

$$\mathcal{M}(if\gamma; \vec{r}) = \frac{\langle e^{if\gamma\Delta u_{\text{pec},z}} [1 + \delta_g(\vec{x})] [1 + \delta_g(\vec{x}')] \rangle}{1 + \xi(r)} \langle e^{if\gamma\Delta u_{\text{RT},z}} \rangle. \quad (4.24)$$

Its FT, the pairwise velocity PDF, is written as

$$\mathcal{P}(u_z; \vec{r}) = (\mathcal{P}_{\text{pec}} * \mathcal{P}_{\text{RT}})(u_z; \vec{r}), \quad (4.25)$$

where $*$ denotes the convolution for simplicity with \mathcal{P}_{RT} given by

$$\begin{aligned} \mathcal{P}_{\text{RT}}(u_z) &= \int \frac{d\gamma}{2\pi} e^{-i\gamma u_z} \langle e^{if\gamma\Delta u_{\text{RT},z}} \rangle \\ &= \int \frac{d\gamma}{2\pi} e^{-i\gamma u_z} \left| \int du P_{\text{RT}}(u) e^{if\gamma u} \right|^2. \end{aligned} \quad (4.26)$$

Notice that the u_{RT} contribution in Eqn. (4.24) is not weighted by the density field at different scales and, as a result, the ensemble average becomes an integration over one-point PDF, $P_{\text{RT}}(u_z)$. In other words, the scale dependence in the pairwise velocity PDF in Eqn. (4.25) comes only from the peculiar velocity part, $\mathcal{P}_{\text{pec}}(u_z; \vec{r})$.

In Figure 4.7, we show the measured pairwise velocity PDFs at two different scales (solid and dashed lines for 1 cMpc/h and 10 cMpc/h, respectively) in cases of both the red-peak only (upper panel) and the global peak (lower panel) at $z = 3.01$. First of all, when we adopt v_{pec} (purple lines) as a velocity, we confirm a trend well known in the literature (see e.g., Scoccimarro, 2004); We see a positive peak, which corresponds to a coherent infall motion on large scales ($r_{\parallel} = 10$ cMpc/h), while the PDF follows a distribution with an exponential tail at small separation, $r_{\parallel} = 1$ Mpc/h. On the other hand, we do not confirm such a trend in the case of v_{RT} . In general, a large tail of the pairwise v_{RT} -PDF contributes to that of the pairwise v_{app} -PDF even at a large separation of 10 cMpc/h. This impact is clearly more significant for the global peak only case than for the red-only peak case.

To see the scale dependence of the pairwise velocity PDFs more explicitly, we show three low-order moments of the PDFs (mean, dispersion, and kurtosis) as a function of a separation scale in Figure 4.8, where we further confirm the aforementioned trends. The mean infall velocity (upper panel) is dominated by v_{pec} particularly on large scales, and the impact of v_{RT} on the mean is marginal even in the global peak case. This is the

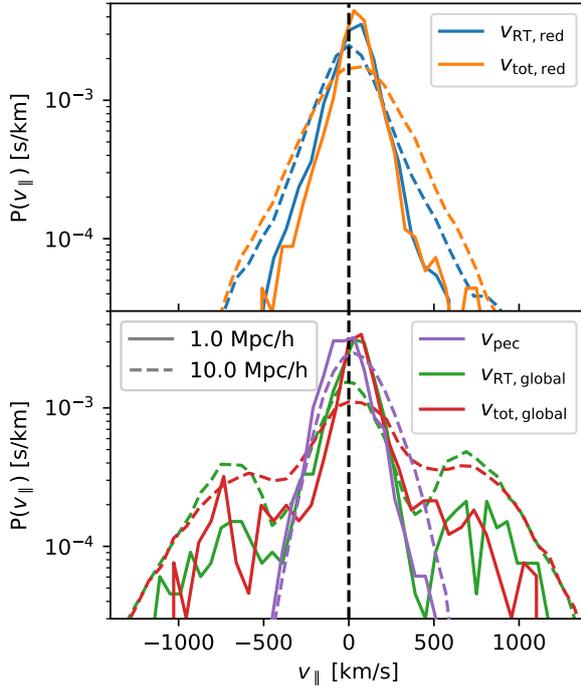
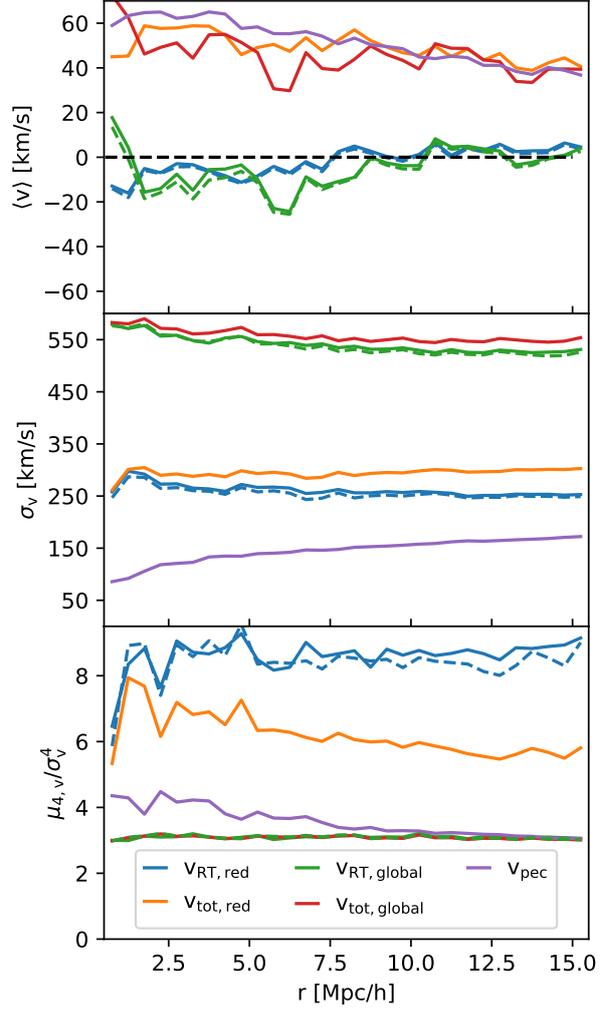


Fig. 4.7 The pairwise velocity distribution P for the peculiar and radiative transfer velocity component evaluated at different length scales for $z = 3.01$ with $n_{\text{LAE}} = 0.01 \text{ h}^3/\text{Mpc}^3$ along the line of sight. The upper panel shows the radiative transfer and total apparent velocities ($v_{\text{RT,red}}/v_{\text{tot,red}}$) based on the red peak only detection method. The lower panel shows the same for the global peak detection method plus the peculiar velocity distribution v_{pec} . Different line styles indicate different spatial separations \vec{r} .

reason why the Kaiser squashing effect is seen on large scales in Figure 4.6. Meanwhile, the dispersion (middle panel) is mainly dominated by v_{RT} for both detection algorithms and notably has little scale dependence. Interestingly, the dominant contribution to the kurtosis (lower panel) depends on the peak finding algorithm: As expected, the kurtosis of v_{pec} deviates from the Gaussian value of 3 at small scales due to its extended tail (e.g., Scoccimarro, 2004). In the case of the red peak detection method, a much wider distribution due to v_{RT} dominantly contributes to the kurtosis than for v_{pec} . On the other hand, both v_{pec} and v_{RT} roughly equally contribute to the kurtosis for the global peak detection method. We do not find a strong scale dependence of the kurtosis for both cases. Since the dispersion and higher-order moments of the pairwise velocity PDF contribute to the FoG elongation (Scoccimarro, 2004), this evidence suggests that random velocity following the one-point PDF plays a main role in the FoG elongation in Figure 4.6.

In Figure 4.8, we also show the three moments of the v_{RT} estimated from v_{tot} and v_{pec} , assuming that \mathcal{P}_{pec} and \mathcal{P}_{RT} are independent of each other (dashed lines compared with solid lines with the same colors). Although they are roughly consistent, we see small discrepancies between solid and dashed lines, implying that either $\langle v_{\text{RT}}(\vec{x})\delta_g(\vec{x}') \rangle$ or $\langle v_{\text{RT}}(\vec{x})v_{\text{pec}}(\vec{x}') \rangle$ is not exactly equal to zero.

Fig. 4.8 The mean, standard deviation, and kurtosis of the pairwise velocity distribution P for the peculiar and radiative transfer velocity component as a function of the total separation r of emitter pairs at $z = 3.01$ with $n_{\text{LAE}} = 0.01 \text{ h}^3/\text{Mpc}^3$ along the line of sight. The solid lines show the respective velocity contributions as listed in the legend, i.e. the apparent and radiative transfer velocity offset for each the global and red-only detection method along with the peculiar velocity. In the dashed lines, we show the expected radiative transfer velocity contribution v_{RT} at a given scale under the assumption that v_{RT} and v_{pec} are independent distributions. Thus, the difference between the solid and dashed line indicates the strength of any correlation between these two contributions present on a given scale. The color coding is chosen to be consistent with Figure 4.7.



4.4.5 Fourier space: The FoG damping factor

In the previous subsection, we show that the impact of v_{RT} on the pairwise velocity PDF can be roughly understood in terms of the one-point PDF, $P_{\text{RT}}(u_z)$. However, it is not straightforward to quantify its impact on the observable TPCF because of the involved convolution (see Eqn. (4.25)). It is simpler to work in Fourier space since the convolution becomes a multiplication after FT. Since Eqn. (4.25) follows

$$1 + \xi_g^s(r_\perp, \pi) = \left([1 + \xi_{g,\text{pec}}^s(r_\perp, \pi)] * P_{\text{RT}} \right) (u_z), \quad (4.27)$$

we obtain the FoG damping in Fourier space with a help from Eqn. (4.26),

$$D_{\text{FoG}}^{\text{RT}}(k, \mu) \equiv \frac{P_{g,\text{tot}}^s(\vec{k})}{P_{g,\text{pec}}^s(\vec{k})} = \left| \int du_z P_{\text{RT}}(u_z) e^{ifk\mu u_z} \right|^2, \quad (4.28)$$

where $P_{g,\text{tot}}^s(\vec{k})$ and $P_{g,\text{pec}}^s(\vec{k})$ denote the redshift-space power spectra when we adopt the total apparent velocity and the peculiar velocity only as a velocity contamination, respectively. This relation allows us to directly compare the line-of-sight damping from radiative transfer as found in the measured redshift-space power spectra with an expected damping from the underlying one-point PDF. We stress that the last equality holds only under the assumptions we made in Eqs. (4.24) and (4.25): namely, no correlation between v_{RT} and v_{pec} at the same point, and no spatial correlations between v_{RT} and δ_g (and v_{pec}). Furthermore, we will compare this damping with the two generic functions commonly adopted in the literature, i.e., the Gaussian and the Lorentzian damping functions (see Eqs. (4.12) and (4.14)). Here the second central moment σ is directly calculated from the one-point PDF.

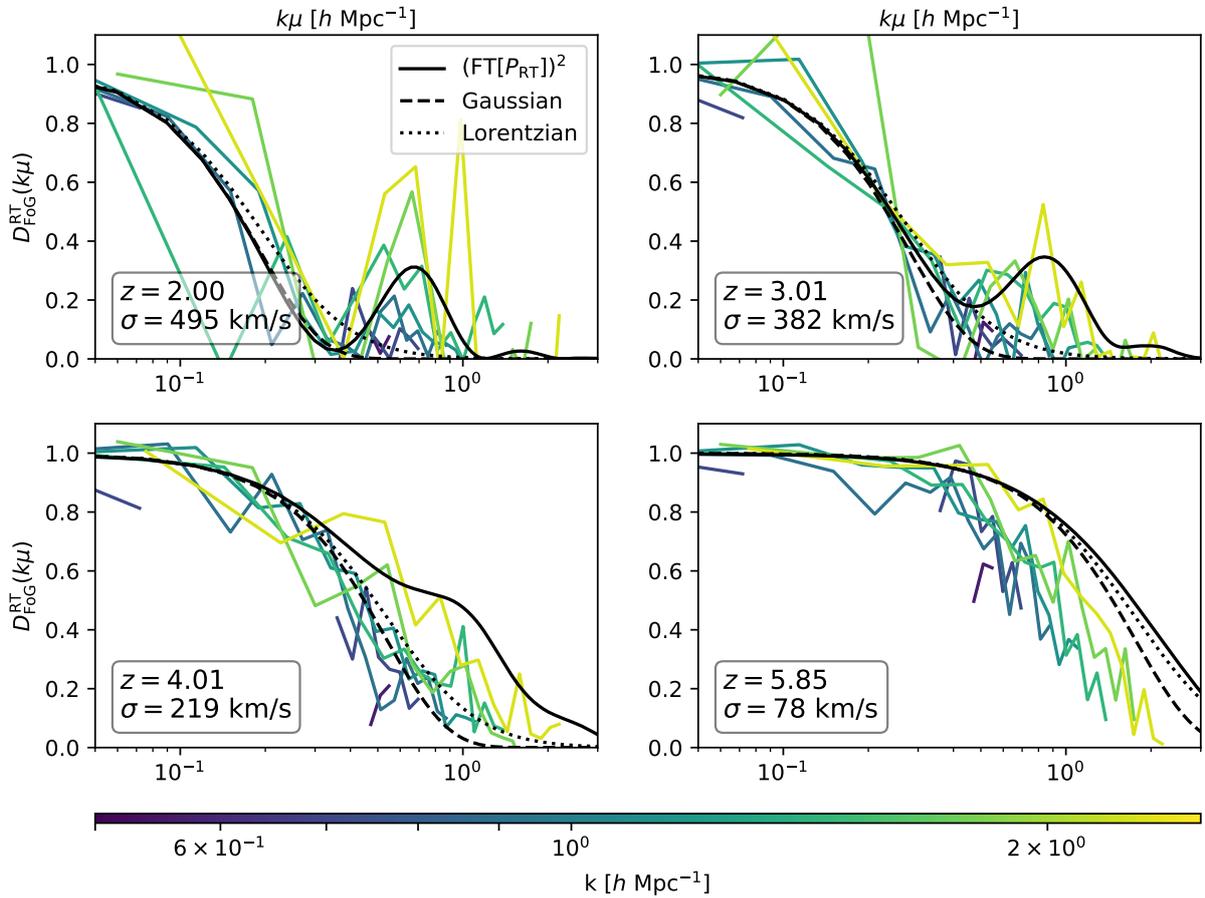


Fig. 4.9 Damping factor as a function of the line-of-sight frequency $k_{\parallel} = k\mu$ due to Lyman-alpha radiative transfer velocity offset $v_{\text{RT,global}}$. Emitter positions are assumed to coincide with the global peak. Different colored lines represent measurements of $D(k_{\parallel})$ from the mock catalogs at different total frequencies k . The black solid lines show the Fourier transform of the 1-point PDF (see Eqn. (4.28)), while the dashed and dotted lines show the Gaussian and the Lorentzian forms respectively (see Eq. (4.12)/(4.14)). **Left-to-right, top-to-bottom:** $z=2.00$, $z=3.01$, $z=4.01$, $z=5.85$

In Figures 4.9 and 4.10, we show such direct comparisons for the global-peak and the red-peak cases, respectively. We show the measurements of the ratio of the two redshift-space spectra as a function of the mode parallel to the line of sight, $k_{\parallel} = k\mu$, color-coding them by their absolute wavenumber, k . In addition, we plot the expected damping from the velocity PDF as implied by Eqn. (4.28) (solid lines) and the two generic fitting functions with the second central moment σ of the according velocity PDF (dashed and dotted lines for the Gaussian and the Lorentzian functions, respectively). In general, we find a strong damping even on relatively large scales, $k \gtrsim 0.1h/\text{Mpc}$, and strength of the damping depends on the peak detection algorithm and redshift. There is a typical redshift evolution with stronger damping at lower redshift, originating from two contributions. First, as seen in Figure 4.5, the velocity distribution widens at lower redshifts, translating to a larger damping scale. This is mainly because the neutral hydrogen density in CGM for the threshold sample becomes larger at lower redshift (see also Behrens *et al.*, 2018). Secondly, distance and velocity are related via the mapping from real to redshift space as given by Eqn. (4.19). As the factor of aH roughly scales as $a^{-1/2}$ in the analyzed redshift range, fixed velocity dispersions correspond to larger damping length scales at lower redshifts.

We also find that the measurements are in fair agreement with the squared direct FT of the one-point PDF (solid lines) at $k\mu \lesssim 1h/\text{Mpc}$, although we see some discrepancies in detail. We discuss some possibilities to cause the discrepancies in the supplementary material (Section 4.7.3), but do not provide a decisive reason. Comparing the solid lines with the dashed and dotted ones, we see that all of them can qualitatively trace the damping feature as a function of scale, although there are slight differences in detail. These differences are expected given the fact that the one-point PDFs follow neither Gaussian nor exponential distributions (see Fig. 4.4.3). It is apparent that our measurements are too noisy to conclude which model works the best. Nevertheless, this result suggests that the dispersion of the one-point PDF, σ , is a good proxy for the resulting FoG damping due to RT.

Strong evidence that our measurement is consistent with the direct FT of the one-point PDF comes from the oscillatory behavior in the solid lines. This can be understood as follows. Suppose that we model the double-peaked PDF as $P_{\text{RT}}(v) = f_1(v) + f_2(v + \Delta v)$ where $f_i(v)$ is a symmetric distribution with a peak and Δv denotes the separation of the two peaks. Then we have

$$\begin{aligned} D_{\text{FoG}}^{\text{toy}} &= |\text{FT}[P_{\text{RT}}]|^2 = |\text{FT}[f_1(v) + f_2(v + \Delta v)]|^2 \\ &= |\text{FT}[f_1]|^2 + |\text{FT}[f_2]|^2 + 2\text{FT}[f_1]\text{FT}[f_2]\cos[k\Delta v] \end{aligned} \quad (4.29)$$

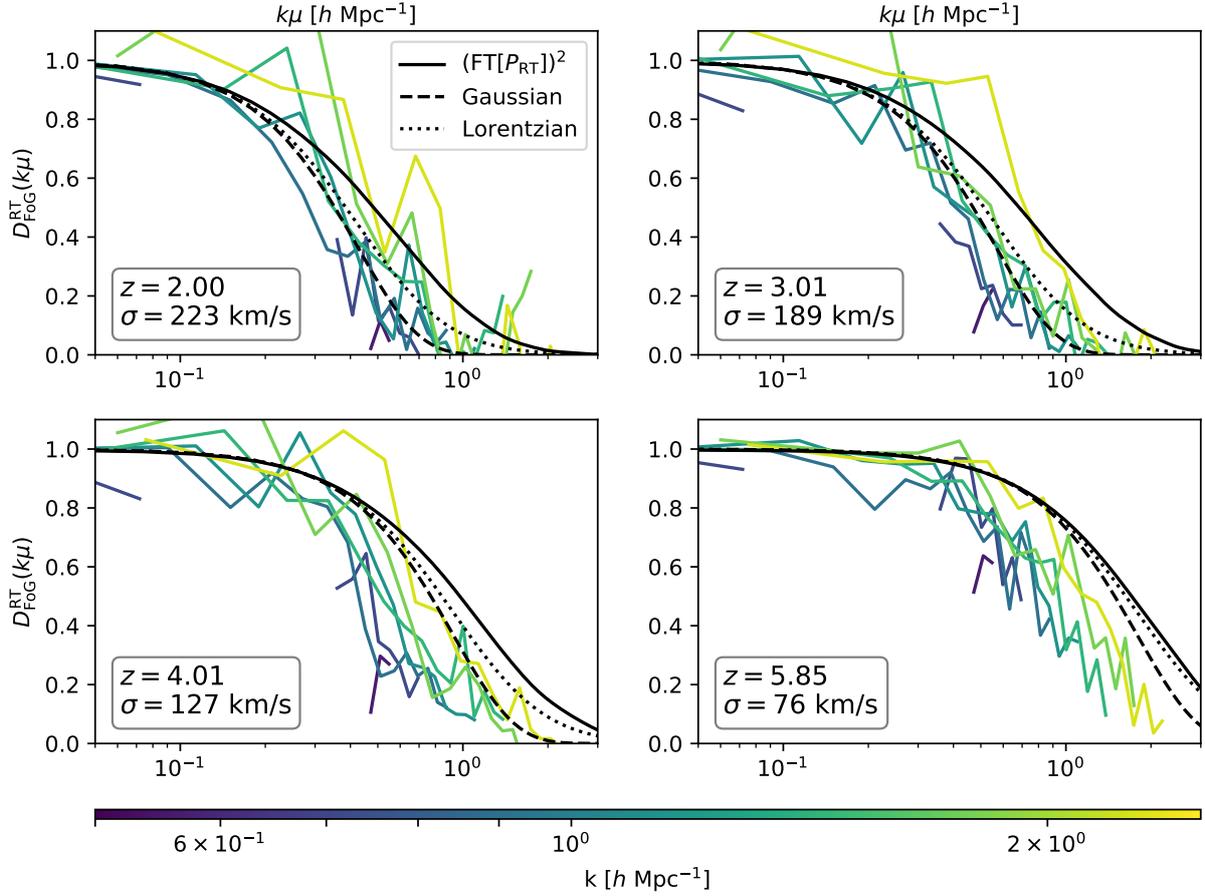


Fig. 4.10 Damping factor as a function of the line-of-sight frequency $k_{\parallel} = k\mu$ due to Lyman-alpha radiative transfer velocity offset $v_{\text{RT,red}}$. Emitter positions are assumed to coincide with the global peak in the red wing relative to the halo's frame. Different colored lines represent measurements of $D_{\text{FoG}}^{\text{RT}}(k_{\parallel})$ from the mock catalogs at different total frequencies k . The black lines show the Fourier transform of the 1-point PDF (see Eqn. (4.28)), the Gaussian and the Lorentzian form (see Eq. (4.12)/(4.14)). **Left-to-right, top-to-bottom:** $z=2.00$, $z=3.01$, $z=4.01$, $z=5.85$

This toy model shows that the first two terms give rise to the FoG-like damping due to the two single-peaked distributions individually, while the last term gives an oscillation due to their separation Δv . In other words, the oscillatory behavior originates from the double peak distribution in the global-peak case. The oscillations disappear at higher redshifts since the second peak in the PDF at a blue end is suppressed by the attenuation due to neutral hydrogen in IGM, as we visually confirm in Fig. (4.4.3). Furthermore, one can quantify when the oscillatory term becomes dominant by approximating both peaks by a Gaussian distribution with the same dispersion σ :

$$D_{\text{FoG}}^{\text{toy,Gaussian}} = \underbrace{\exp(-\sigma^2 k^2)}_{D_{\text{FoG}}^{\text{Gaussian}}} \underbrace{[w_1^2 + w_2^2 + 2w_1 w_2 \cos(k\Delta v)]}_{D_{\text{FoG}}^{\text{osci}}}, \quad (4.30)$$

where w_1 and w_2 are the relative contribution of the peaks to the PDF such that $w_1 + w_2 = 1$. By looking at the leading, second-order term of $D_{\text{FoG}}^{\text{Gaussian}}$ and $D_{\text{FoG}}^{\text{osci}}$ (assuming $k\Delta v \ll 1$ and $k\sigma \ll 1$), we find that the cosine term can dominate if

$$\frac{\Delta v}{\sigma} > \frac{1}{\sqrt{w_1 w_2}}. \quad (4.31)$$

In addition, notice that the impact of $D_{\text{FoG}}^{\text{osci}}$ is limited as it has a lower bound:

$$D_{\text{FoG}}^{\text{osci}} \geq (2w_1 - 1)^2. \quad (4.32)$$

Applying this double-peaked Gaussian to our mock data, we find reasonable fits that could be further improved by substituting the Gaussians with better fits for the peak PDFs. At all redshifts except for $z=5.85$, where the blue peak has completely vanished, the Gaussian dominates according to Eqn. 4.31. Looking at Fig. 4.9, we nevertheless see that the impact of the oscillatory part is weak. This can be attributed to the strong asymmetry in the height of the two peaks (e.g. $w_1 \sim 0.06$ at $z = 4.01$), which reduces the amplitude of $D_{\text{FoG}}^{\text{osci}}$ according to Eq. 4.32.

4.5 Discussion

We have shown that significant FoG-like damping arises in the large-scale LAE clustering in redshift space as the peak positions of observed LAE spectra are affected by RT. We now discuss the most important steps and caveats concerning our findings for the FoG-like damping factor, $D_{\text{FoG}}(k, \mu)$, for LAEs, and also investigate possibilities to mitigate the RT damping.

In Section 4.5.1, we discuss the shape of spectra arising in our simulations and possible shortcomings in our modeling. Next, we discuss the line-of-sight localization method, which reduces a spectrum to a radiative velocity offset v_{RT} in Section 4.5.2. Afterwards, we seek to use additional information from the spectra to reduce the damping effect in Section 4.5.3.

4.5.1 Spectra

We obtain individual spectra in our simulations by using an aperture method with a 3 arcseconds radius as introduced in Section 4.3. Using this simple detection algorithm, there are only three relevant parameters impacting the damping factor, which are related to instrumental specifications in a real observation: The spatial resolution (aperture size),

the spectral resolution, and the number density threshold. The number density threshold roughly corresponds to a flux threshold which should be determined by the signal-to-noise ratio in real observations. We show the moderate impacts of both spectral resolution and aperture size on the v_{RT} distribution in the supplementary material (Section 4.7.1 and 4.7.1). Similarly, we have already discussed the impact of the number density in Section 4.4.3.

In our simulations, we obtain a manifold of different spectral shapes as shown in Figure 4.2. Most prominently in this work, we seem to overpredict the amount of double-peaked emitters as we mentioned in Section 4.4. As the original Illustris simulations do not resolve the ISM, we effectively set the post-ISM spectrum to that of the input photons, which follow a Gaussian distribution. Thus, many of the photons, which are close to the line center, will be reprocessed on the CGM (rather than the ISM) scales to exhibit the rich dataset of spectra we obtain. It is not trivial to see how the lack of ISM modeling will affect the resulting spectra other than that it most likely reduces the fraction of double-peaked profiles when an ISM model with galactic outflows (Bonilha *et al.*, 1979) is chosen. In fact, more recent work attributes the frequency redistribution to the ISM scales rather than the CGM scales and includes IGM attenuation as an additional separated step (Inoue *et al.*, 2018; Gurung-López *et al.*, 2019). We do not expect that the addition of a sophisticated ISM model will change the qualitative nature of the newly modeled FoG-like damping: An ISM model, particularly with outflows, should not introduce additional large-scale correlations since the large-scale correlation is mainly driven by the IGM attenuation at a blue end of the input spectrum (Behrens *et al.*, 2018). Therefore, we expect the qualitative modeling as a FoG-like damping in redshift space to remain the same even when introducing a more sophisticated spectral modeling on ISM scales.

There is a series of other modeling shortcomings in our work, including the lack of dust (Laursen *et al.*, 2009), subgrid modeling (for sub-parsec clumps) (Gronke *et al.*, 2016b, 2017b), and spatial resolution. We only consider Lyman- α emission from recombination in the star-forming regions. Contributions from collisional excitation can be significant and show a bluer spectral signature (Smith *et al.*, 2019). As initial photons are spawned from a point source within the LAEs in our simulations and thus not reflecting different physical environments, such spectral modifications cannot be captured within our framework. As we are concerned with individual LAEs' spectra, we can neglect the small fluorescent contributions in the IGM (Dijkstra, 2017).

We note that the original Illustris simulations show a growing excess of neutral hydrogen (a factor of ~ 3 between $z = 2$ and 4) compared to observations (Diemer *et al.*, 2019) on galaxy scales where frequency diffusion of Lyman- α photons gives rise to the FoG-like damping. This leads to an overestimate of the dispersion in the v_{RT} -PDF. For

the Neufeld solution, (Neufeld, 1990) such a hydrogen excess would correspond to an overestimate in the peak offsets v_{RT} of roughly 31% ($\Delta v \propto \tau^{1/3} \propto n_{\text{HI}}^{1/3}$).

As found in Behrens *et al.* (2018) for the real-space clustering, large-scale correlations and LAE mock observables significantly change with the resolution of the underlying neutral hydrogen distribution in the RT simulations. Similarly, we give a possible explanation for the lack of detection of the additional redshift-space distortion in Zheng *et al.* (2011a) due to its limited hydrodynamical resolution in Section 4.7.2.

Overall, we thus expect the observational spectral shapes to differ from those found in our simulations. As a result, our simulations do not reproduce the observed Lyman- α luminosity function as addressed in Behrens *et al.* (2018). This problem of reproducing observables is common when not explicitly calibrating against Lyman- α specific observations such as the luminosity function (e.g. calibration against other redshifts, see Kakiichi *et al.* (2016), Inoue *et al.*, 2018).

From the spectra themselves, one can obtain stacked surface brightness profiles I_{stacked} as shown in Figure 4.3. However, it is important to realize that it is the *unobservable* one-point velocity PDF, P_{RT} , that mainly determines the additional FoG-like damping along the line-of-sight direction. Since the relation between I_{stacked} and P_{RT} is non-trivial, there is no *a priori* way of determining the damping factor from the stacked profiles. Nevertheless, we find an empirical relationship between the square root of the second central moment of I_{stacked} and P_{RT} :

$$\sigma_{\text{PDF}} \approx 0.4 \cdot \sigma_{\text{stacked}} \quad (4.33)$$

for $n_{\text{LAE}} = 10^{-2} \text{ h}^3 \text{Mpc}^{-3}$ at $z = 3$. This relation typically changes less than 10% among redshifts from 3.01 to 5.85. At redshift $z = 2.0$, this proportionality factor is consistently higher by a factor of $\sim 25\%$. The number density has a large impact on the resulting relation and rising from 0.4 to 0.55 when restricting the number density from 0.01 to $0.001 \text{ h}^3 \text{Mpc}^{-3}$. This change is caused by an increased dispersion in the velocity distribution, while the width of the stacked profiles is nearly constant independent of imposed number density threshold. This relation was found in the halos' rest frame, which is unknown unless a secondary emission line (e.g., $\text{H}\alpha$) is measured and its radial position is identified.

4.5.2 Localization along the line of sight

We proposed two simple localization methods to identify LAE's radial position in redshift space for given Lyman- α spectra: Either by identifying the position of the maximal

spectral flux with the LAE's position (referred to as 'the global peak') or by identifying the LAE position with the maximal spectral fluxes at wavelengths longward of the actual LAEs' radial positions as used for the $v_{\text{RT,red}}$ distribution (referred to as 'the red peak only'). Combining our simulated spectra with the choice of the localization methods gives us the one-point velocity PDF, $P(v)$, which is a key quantity to understand the resulting FoG damping. Nevertheless, let us briefly discuss potential issues in $P(v)$ regarding the two localization methods here.

The major issue with the global-peak method is closely related to the fact that we overestimate the number of double-peaked spectra in our simulations. As a consequence, the velocity distribution $P_{\text{RT}}(v)$ has a strong double-peak feature as well, as we see in Figure 4.5, which does not most likely represent real observations. Furthermore, as shown in Section 4.4.5, such a double-peaked velocity distribution induces a larger FoG damping as compared with the $v_{\text{RT,red}}$ distribution. In a simple toy model for a double-peaked velocity distribution, we see that two additional terms occur, strengthening the damping, where one is induced by the second peak's width and the second by the separation between the peaks.

The second detection method does not suffer from these additional damping contributions and can be applied to both the double-peaked spectra and single-peaked spectra. However, there remains a chance to get an oscillatory damping component in this localization method if some LAEs' visible peaks are blue of the line center, which is not distinguishable from a red peak in real observations without a complementary emission line. Nevertheless, it is ad hoc to completely ignore the blue part in the red-peak only method without any physically reasonable reasons. We thus stress that two localization methods and their velocity PDFs showcase two extremes.

The localization methods so far only use a single peak's frequency shift for the determination of the line-of-sight position. However, we can, in principle try to use other information available through the individual spectra to determine the line-of-sight position. In the coming section, we investigate methods to further reduce the radiative transfer damping in the clustering signal based solely on additional information of the features in the Lyman- α spectrum.

4.5.3 Correction to mitigate the impact of RT

We can try to correct for the presented Lyman- α radiative transfer distortion effect by utilizing the full spectral information available for the emission line. To do so, we investigate correlations of the peak offset to other characteristics of the spectra. These include: Half the separation between red and blue peak (**m1**), the full width at the half maximum

(FWHM) of the red peak (**m2**), twice the half-width at the half maximum (HWHM) of the red peak towards the line center (**m2a**) and twice the HWHM of the red peak away from the line center (**m2b**). Hence, **m2** is the average of **m2a** and **m2b**. A visualization of these methods is shown in Figure 4.11.

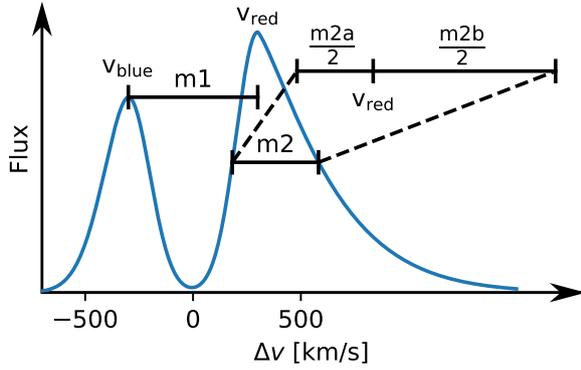


Fig. 4.11 Sketch of the proposed correction methods (see text) applied to an example double peak spectrum.

We choose to divide the second method (**m2**) into **m2a** and **m2b** as we expect different physical causes for their respective wing shape. The wing towards the line center should be strongly influenced by the IGM attenuation. This has been shown to hold true even at low redshifts very close to the line center (Laursen *et al.*, 2011). However, the IGM is transparent for the wing away from the line center and the red peak, and therefore should only be impacted by small-scale frequency diffusion. Indeed we will find large differences in the results from **m2a** and **m2b**.

We calculate slope f and offset v_{offset} for a linear regression of form

$$v_{\text{predict}} = f \cdot v_{\text{proxy}} + v_{\text{offset}} \quad (4.34)$$

and also the corresponding Pearson correlation coefficient p . v_{proxy} denotes the respective velocities by **m1**, **m2**, **m2a** and **m2b**. v_{predict} is the predicted correction for the corresponding peak position v_{RT} . With such prediction, we can correct the former v_{RT} distribution as

$$v_{\text{RT,corr}} = v_{\text{RT}} - v_{\text{predict}}. \quad (4.35)$$

When a blue peak is additionally available for **m1**, we use the detection algorithm introduced in Section 4.7.1: Peaks are identified as connected areas in $F_{\lambda}(\Delta\lambda)$ above such a threshold value that a given number density n_{LAE} is reached. Additionally, we require that the maximal brightness of a peak needs to exceed 10% of the maximal brightness of the brightest peak. If only one peak is available, the emitter is excluded when computing m_1 .

The linear regression in Eqn. (4.35) is motivated by Neufeld's solution for which both peaks' FWHM (i.e. **m2**) and peak offset scale with $\tau^{1/3}$ for an optically thick spherical HI distribution. From Neufeld's derivation, a slope of 1.26 is derived for the relationship

between offset and FWHM (Neufeld, 1990). We expect that anisotropies, dust, the velocity field, and IGM interaction introduce a significant scatter as well as a noticeable change to the slope parameter. Such a correlation has been found in observations (see e.g., Verhamme *et al.*, 2018).

In general, we find **m1** to perform the best with a correlation coefficient of $p \gtrsim 0.95$ across the studied redshift range. However, most of observed LAEs are not doubly peaked, thus only methods **m2**, **m2a**, **m2b** are available. For these, we find **m2a** with $0.82 \lesssim p \lesssim 0.85$ to perform the best. **m2** performs slightly worse with $0.77 \lesssim p \lesssim 0.78$ and **m2b** significantly worse with $0.40 \lesssim p \lesssim 0.65$. Note that we restrict the emitter sample for the regression to those with $v_{\text{RT}} < 800$ km/s in order to allow comparisons to observational studies and hinder the most massive emitters from dominating the fit due to their increased scatter.

An example of the correlation between v_{RT} and v_{proxy} is shown in Figure 4.12 for **m2a** at $z = 3.01$ and a number density threshold of $0.01 \text{ Mpc}^{-3} h^3$. The best fit in this case yields $v_{\text{predict}} = 1.47 \cdot v_{\text{proxy}}^{\text{m2a}} + 37$ km/s with a Pearson coefficient of 0.83. For other redshifts, we find a slope of 1.39, 1.43, 1.70 ($z = 2.0, z = 4.01, z = 5.85$) with similar or lower offsets. Note that the constant offset itself is irrelevant for the damping scale as it does not change the pairwise velocity distribution. We find that there is a strong dependence of the slope parameter on the chosen number density threshold. The slope parameter increases as the number density is decreased. This finding is related to the dependence of host halo mass: As shown in Figure 4.12, points with larger halo mass tend to have a slightly steeper slope with larger scatter.

For **m1**, we find that the slope is close to unity and in excellent agreement with observations (Verhamme *et al.*, 2018), while for methods using the peaks' width (**m2, m2a, m2b**) we find a slope above unity that makes it slightly higher than the Neufeld solution and considerably higher compared to observations and shell models (Zheng *et al.*, 2014; Ver-

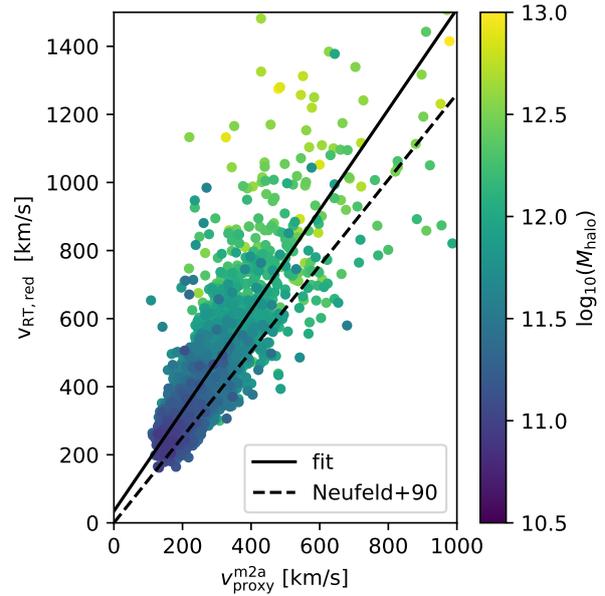
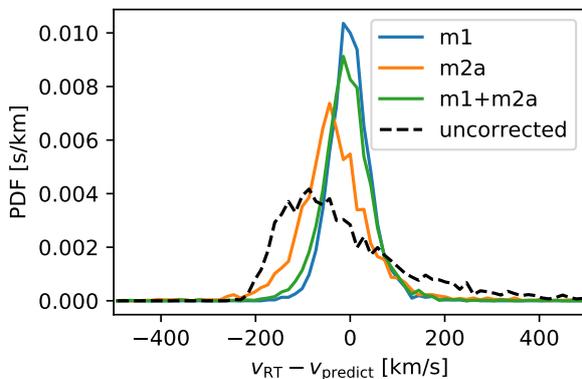


Fig. 4.12 Scatter plot of $v_{\text{RT,red}}$ and $v_{\text{proxy}}^{\text{m2a}}$ for emitters detected at number density threshold $n_{\text{LAE}} = 10^{-2} h^3 \text{Mpc}^{-3}$ and redshift $z = 3.01$. We also show a linear fit (see Eqn. (4.34)), the slope expected from an optically thick spherical HI distribution (Neufeld, 1990).

hamme *et al.*, 2018). In addition, we find a significant dependence on selection criteria and redshift, possibly explaining some of the discrepancies between our results and the literature apart from the mentioned modeling shortcomings in our simulations.

Fig. 4.13 Radiative velocity distribution after different correction schemes at a number density threshold of $n_{\text{LAE}} = 10^{-2} \text{ h}^3 \text{ Mpc}^{-3}$ and redshift $z = 3.01$. The mean velocity is subtracted from each distribution. The dashed line shows the uncorrected distribution. The dispersion for the three correction schemes reduce to 48 km/s, 92 kms/s, 65 km/s (m1, m2b, m1+m2a) compared to 189 km/s when not corrected.



Similarly to Verhamme *et al.* (2018), we can correct for the systemic redshift offset by subtracting the FWHM. Doing so consistently reduces the velocity dispersion by a factor of 2-3 across the simulated redshift range and accordingly shortens the damping scale. Figure 4.13 shows the corrected distribution according to Eqn. (4.35) at the same redshift and number density threshold. Also, the third and fourth central moments are reduced significantly: the normalized moments (skewness, kurtosis) shrink such that a Gaussian fit becomes feasible.

The fit parameters and correlation coefficient calculated here can be used as a first guess for applying the presented corrections to observations. However, variations from those in observational datasets are expected due to the modeling deficiencies discussed earlier. By measuring a second emission line for a subset of LAEs, the radiative transfer velocity offsets v_{RT} can be calculated, allowing a similar fit of different correction methods in observations that can then be applied to the entire cosmological volume of LAEs studied.

4.6 Conclusions

In this chapter, we have studied the clustering of LAEs in redshift space using a full RT simulation on Illustris. We find a new kind of the RSD effect due to RT, and our executive summary is the following:

- The additional redshift-space distortion stems from small-scale frequency diffusion of the Lyman- α line leading to a shift of the spectral peaks. The peak shifts can be

larger than those of the peculiar velocity field and thus impact the redshift-space clustering signal on larger spatial scales.

- We show that the peak shifts from Lyman- α RT damp the power spectrum along the line of sight at scales of $k_{\parallel} \gtrsim 0.1h/\text{Mpc}$. We also show that the shifts are mostly independent of the local density and velocity field such that independent modeling of this shift's impact can be done. This is similar to that of the Fingers-of-God effect due to random motion of galaxies, i.e., in terms of the one-point velocity PDF. However, the functional form of the damping can be more complex and even involve oscillations due to the double peak nature.
- The strength of the damping depends strongly on the chosen localization method of the Lyman- α emitters in its spectrum. We attempt two extreme scenarios where we find a peak from the entire spectrum (global peak) and only from the spectrum at the red end (red peak).
- We show that we can mitigate the impact of the distortion by applying a correction scheme of which we present two classes: In the case of double-peaked spectra, the midpoint between the two peaks is an excellent proxy for the emitter position. If only a single peak is present, the half-width-half-maximum on the wing towards the Lyman- α line center can be used as a mediocre proxy.

We do not attempt to quantify the exact amplitude of the RT effect on actual BAO and RSD galaxy surveys such as HETDEX for the following reasons. First, as we often mention, our simulated LAEs do not well reproduce a variety of observables, such as the luminosity function likely due to unresolved ISM physics in Illustris. Second, the actual impact should depend on the resolving power of a spectrum. For example, HETDEX has a spectral resolution roughly corresponding to $\Delta v \sim 400\text{km/s}$ where the resultant PDF is largely smeared out. Nevertheless, we stress that the frequency shift can be larger than $\Delta v \sim 500\text{km/s}$ as seen in Figure 4.4 and hence we expect that the FoG damping due to RT exists to some extent. We leave a detailed assessment for future work.

Although we focus on the clustering of LAEs, it would be straightforward to extend our analysis to the intensity mapping. In fact, we visually confirm strongly elongated feature in the Lyman- α intensity map in Figure 4.1: For intensity mapping, we expect a similar damping to that of the LAEs from their positional offset, but additionally from the width of the spectrum itself, which further strengthens the damping. Observationally, Croft *et al.* (2016) reported the large-scale elongation along the line of sight in the cross-correlation between the quasars and the Lyman- α intensity map at $z \sim 2$ in the Sloan Digital Sky Survey. Even though their more recent study argues that it is due to a special

environment around quasars given the lack of the cross-correlation signal between the intensity map and the Lyman- α forest (Croft *et al.*, 2018), the elongation might partly be due to the RT FoG effect.

As a concluding remark, we give the following general suggestions for cosmological LAE surveys:

- Having a second emission line as a tracer of the LAEs' radiative transfer velocity shift for a subset of emitters allows to calibrate the linear fit parameters of the correction methods based on the Lyman- α spectral features presented in Section 4.5.3 to mitigate the distortion effect.
- Furthermore, we found an empirical relationship between the second central moment for the red and blue peaks in stacked spectra in the halo's rest frame and the dispersion of those peaks in the one-point velocity distribution. Hence, the possible damping might be estimated by measuring the second central moments in the stacked spectra. Stacking these in the halo's rest frame also requires knowledge of a secondary emission line tracing the kinematics of the LAE.

Finally, concerning other target emission lines such as $H\alpha$ for e.g., Euclid (Laureijs *et al.*, 2011), Wide-Field InfraRed Survey Telescope-Astrophysics Focused Telescope Assets (WFIRST-AFTA, Spergel *et al.*, 2013), [OII] for the Subaru Prime Focus Spectrograph (PFS, Takada *et al.*, 2014), and Dark Energy Spectroscopic Instrument (DESI, DESI Collaboration *et al.*, 2016), the RT effect studied here has no impact on the BAO and RSD measurements, as the transition must be resonant and have high optical depths in the astrophysical environment.

Acknowledgements

We thank Max Gronke for useful discussions concerning the obtained spectra in our simulations for this chapter. We are also grateful to Eiichiro Komatsu for his comments on the manuscript submitted for Byrohl *et al.* (2019). The primary co-author of Byrohl *et al.*, 2019, Shun Saito, acknowledges support by JSPS KAKENHI Grant Number JP16J01890, and by World Premier International Research Center Initiative (WPI Initiative), MEXT, Japan. We acknowledge use of the Python programming language, the use of the Numpy (Harris *et al.*, 2020), IPython (Pérez *et al.*, 2007), and Matplotlib (Hunter, 2007) packages. Our analysis benefited from the use of the following astrophysical packages: astropy (Collaboration *et al.*, 2013; Astropy Collaboration *et al.*, 2018) and halotools (Hearin *et al.*, 2017).

4.7 Supplementary material

4.7.1 Detection algorithm variations

Spectral resolution

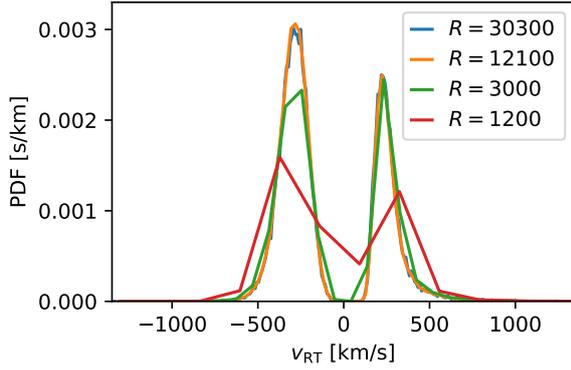


Fig. 4.14 The radiative velocity distribution v_{RT} for varying spectral resolution in terms of $R = c/\Delta v$ at $z=3.01$ for all LAEs (no n_{LAE} restriction).

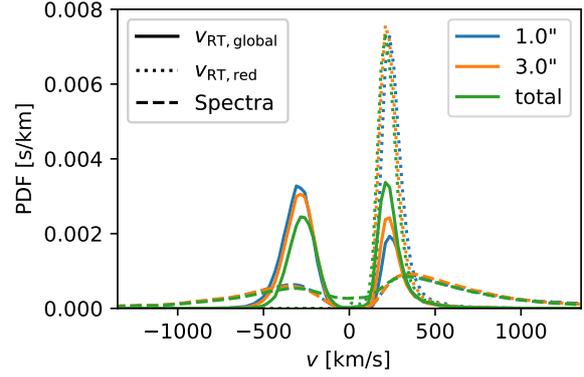


Fig. 4.15 The radiative velocity distributions v_{RT} and stacked spectra for varying aperture sizes at $z = 3.01$ for all LAEs (no n_{LAE} restriction).

In our study, we do not generally assume a spatial resolution match to a specific observation both in the spectral and the angular resolutions. For the spectral resolution, we infer the redshift-space position by adding the line shift to the real-space position known from the halo catalogs, while surveys directly infer the degenerate redshift space position from the line feature's position. Nevertheless, the spectral resolution in simulations and observations should be comparable quantities. In our fiducial case, velocities are resolved to 24.7 km/s corresponding to a spectral resolution of roughly $R \sim 12000$. This exceeds the resolution in HETDEX by more than an order of magnitude ($R \sim 800$, see e.g. Hill *et al.* (2008)). Even sophisticated spectroscopic instruments in search of LAEs such as MUSE only reach values up to $R \sim 3000$ (see e.g. Bacon *et al.* (2015)).

In Figure 4.14, we vary the spectral resolution for the fiducial case at $z = 3.01$ and find good reproduction of the characteristics of the distribution: If existent, the double-peaked structure is conserved and the second central moment of individual peaks vary by less than 10%. We have adopted the fiducial resolution of 24.7 km/s as we are interested in the physical impact of RT in theory, but simultaneously want to assure sufficient signal to noise in each bin.

Aperture size

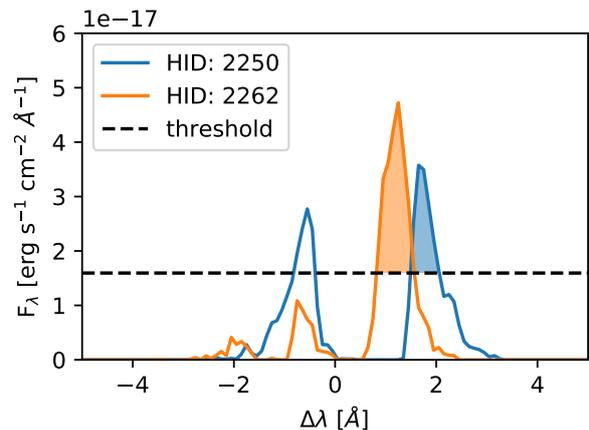
In Figure 4.15, we show the aperture size dependency of the v_{RT} -probability distribution and of the stacked profiles at $z=3.01$. The sample includes all simulated LAE rather than fixing the count to a number density threshold. This reduces the noise, while we find qualitatively similar results for a restricted sample.

Three qualitative changes occur when increasing the aperture: A larger aperture appears to favor the red peaks over the blue, the red peak slightly shifts towards the line center, and for large apertures the stacked spectra show emission in the otherwise deserted trough in the line center. The second central moments and maxima of the red peaks change by less than 10% so that the expected change in clustering signal should be similarly small.

Refined detection algorithm

In the detection algorithm applied so far, line-of-sight halo positions are inferred from the spectral peak, while the detectability is determined from the overall flux. Therefore, the same halo sample is considered as in prior real-space evaluation performed in Behrens *et al.* (2018). However, in redshift space some emitters' spectra might significantly diffuse while others do not based on varying small and large scale properties. Therefore we expect additional distortions when emitters are detected by the specific flux. Primarily the bias is expected to change.

Fig. 4.16 Example of refined detection of two individual emitters with given halo ID at $z=3.01$. Dashed line corresponds to the specific flux threshold of $1.7 \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1}$ (for number density $0.01 \text{ h}^3 \text{ Mpc}^{-3}$) and the shaded region the flux associated with the emitters given the refined detection algorithm. For shown emitters, the emitter with the highest total flux (which includes all peaks) is different from the emitter with the largest flux for the identified peak.



In order to capture the additional distortions, we refined our detection algorithm as follows: A specific flux threshold t_{flux} is imposed, and only peaks above this threshold are used to identify emitters. Once a peak (red or global) is identified as the emitter's position, the surface brightness is integrated around the peak for the spectral range that reaches t_{flux} .

Figure 4.16 shows an example of the refined algorithm for two emitters at $z=3.01$ and how the bias might change: While the spectrum of HID 2262 has a lower overall flux than HID 2250, it has a larger flux around the detected peak. Therefore, flux-sorted and number density limited samples will differ when sorted by total flux or flux associated with the peak.

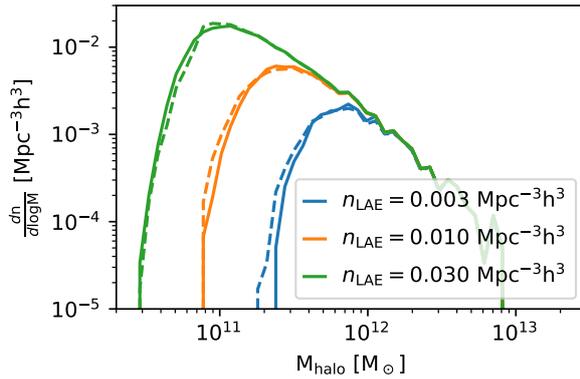


Fig. 4.17 The halo mass function of detected LAEs at $z = 3.01$. Solid lines represent the fiducial detection method and dashed lines the refined algorithm presented. There are only marginal changes in the function at the characteristic mass cut-off imposed by setting the number density threshold n_{LAE} .

Figure 4.17 shows the halo mass function of detected LAEs for the two discussed detection algorithms. Overall, changes only occur close to a characteristic mass cutoff that is implied by the number density threshold and due to the strong correlation between observed flux and halo mass. As we only find a small change in the mass distribution of detected LAEs, we do not expect an additional change in bias/isotropic distortion from radiative transfer in redshift space. We explicitly checked this by calculating the corresponding velocity distributions and power spectra.

Nevertheless, this detection method is useful in classifying individual peaks that we will use for identifying blue and red peaks in Section 4.5.3 concerning the correction of the localization error leading up to the damping.

4.7.2 Comparison to prior studies

While similar radiative transfer simulations have previously been performed, the shown redshift-space damping has not been observed by Zheng *et al.* (2010)/Zheng *et al.*, 2011a. For consistency, we try to reconcile our findings with these prior simulations.

As shown, the radiative transfer redshift-space distortions significantly increase at lower redshift: At $z = 2.00$, we detect a central second moment σ of the $v_{\text{RT,red}}$ distribution of 223 km/s, but only a moment of 76 km/s at $z = 5.85$. Furthermore, the significance for the redshift space clustering is reduced at higher redshifts due to the higher Hubble flow as discussed before.

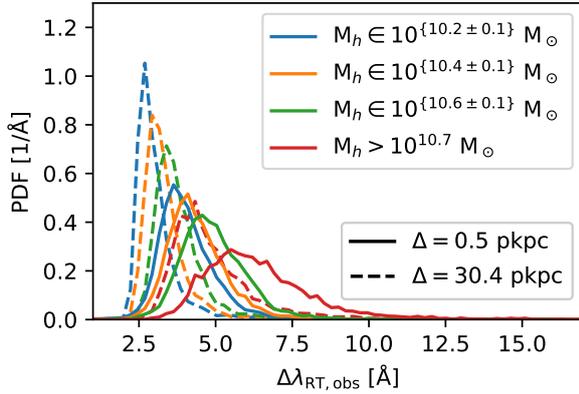


Fig. 4.18 The radiative velocity distribution for varying halo mass range and underlying hydrodynamical resolution at redshift $z = 5.85$. $\Delta\lambda_{\text{RT}}$ is the radiative velocity as observed wavelength shift at $z = 0$ with the peculiar velocity v_{pec} having been subtracted as usual.

redshift $z = 5.85$ of different underlying hydrodynamical resolutions for the radiative transfer post-processing. This plot can directly be compared to the results found in Zheng *et al.* (2010) showing a very similar plot (see Figure 6 there). Given the differences in the baryon modeling and detection method, we find a decent match between our low resolution run with a grid spacing of $\Delta = 30.4$ pkpc and the results found by Zheng *et al.* (2010) across all mass bins. Both shown simulations have been imported from Behrens *et al.*, 2018 and thus only have an initial photon count of 100 instead of 1000. Thus, minimal differences to other shown results might be expected.

We find a significant impact of the hydrodynamical resolution on the velocity distribution. First, the peak of the distribution is shifted towards larger spectral offsets from the line center at higher resolutions. If no blue peak is present, as is for $z = 5.85$ due to IGM interaction, this overall shift of the distribution should not affect the damping signal. Second, the distributions broaden at higher resolutions and thus increases the second central moment, which enlarges the damping length scale. For the shown resolutions, the second central moment increases by roughly 20% across all mass bins at higher resolution.

4.7.3 Discrepancies in the damping factor modeling

As we have shown in Section 4.4.5, we have made an attempt to understand the FoG suppression due to RT in terms of the one-point velocity PDF. In this appendix, however,

Additionally, as concluded by Behrens *et al.* (2018), the hydrodynamical resolution has a large impact on the radiative transfer’s results. In particular, the hydrodynamical resolution largely affects the photon diffusion in configuration and frequency space, determining whether a selection effect is detected in mock observations or not. Similarly, one might ask whether the hydrodynamical resolution impacts the radiative transfer redshift-space distortion presented here. We answer this question by comparing the results by analyzing a set of RT simulations at different resolutions.

Figure 4.18 shows the radiative velocity offset distribution as a function of mass at

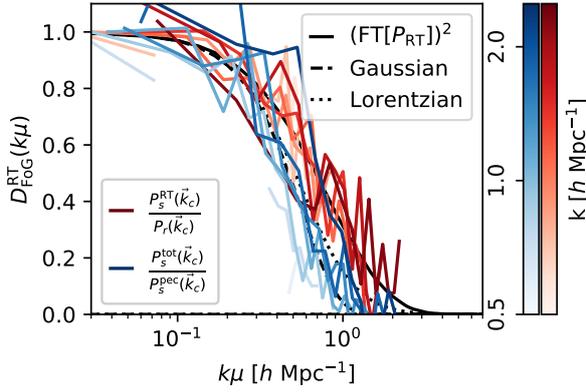


Fig. 4.19 Damping factor as a function of the line-of-sight frequency $k_{\parallel} = k\mu$ due to Lyman-alpha radiative transfer for $z=3.01$. Emitter positions are assumed to coincide with the red peak only. Two different ratios P_s^{RT}/P_r and $P_s^{\text{tot}}/P_s^{\text{pec}}$ are shown for the damping factor, see text for an explanation. Different colored lines represent measurements of $D(k_{\parallel})$ from the mock catalogs at different total frequencies k .

we show an indicator that it is not sufficient to know the one-point PDF to model the impact of RT on the redshift-space clustering fully.

In Figure 4.19, we present two ratios of the power spectra. The blue lines are the ratios of the total redshift-space spectra to ones only with the peculiar velocity component, i.e., $P_{g,\text{tot}}^s/P_{g,\text{pec}}^s$, which are the same results as we showed in Figure 4.10. We also present the ratios of the redshift-space spectra only with the RT velocity component to the real-space power spectra, i.e., $P_{g,\text{RT}}^s/P_g^r$ as red lines. We did not address the latter ratio in the main text since both the denominator and the numerator are not directly accessible in real observations. In Figure 4.10, it is apparent that the red curves are systematically higher than the blue ones even though both of them are noisy.

As we stressed in deriving the last equality of Eqn. (4.28), a necessary condition which makes the ratio, $P_{g,\text{tot}}^s/P_{g,\text{pec}}^s$, equivalent to the FT of the one-point velocity PDF is to satisfy

$$\langle v_{\text{pec}}(x)v_{\text{RT}}(x') \rangle = 0, \quad (4.36)$$

$$\langle v_{\text{RT}}(x)\delta_g(x') \rangle = 0, \quad (4.37)$$

$$\langle v_{\text{RT}}(x)v_{\text{RT}}(x') \rangle = 0. \quad (4.38)$$

We argued from Figure 4.4 that the condition (4.36) is satisfied at the level of one point, i.e., at $x = x'$. Similarly, we did not see any strong evidence which violates the conditions (4.37) and (4.38) in Figure 4.8. In contrast, the condition (4.36) is not necessary for $P_{g,\text{RT}}^s/P_g^r$, since it does not involve v_{pec} . It is encouraging to see that the red lines are more consistent than the blue ones with the direct FT of the one-point PDF (black solid line). We find this trend holds at other redshift snapshots for both the global-peak and the red-only cases. This fact implies that the conditions (4.37) and (4.38) are basically satisfied. We thus may attribute the discrepancy between the red and blue lines to the

residual correlation between v_{pec} and v_{RT} , but a conclusive statement with more careful investigation is left for future work.

LYMAN-ALPHA SPECTRA AND THE INTERGALACTIC MEDIUM

The content of this chapter has been published as Byrohl et al. (2020b).

Context and summary

Lyman- α ($\text{Ly}\alpha$) spectra provide insights into the small-scale structure and kinematics of neutral hydrogen (HI) within galaxies as well as the ionization state of the intergalactic medium (IGM). The former defines the intrinsic spectrum of a galaxy, which, in turn, is modified by the latter. These two effects are degenerate. Using the IllustrisTNG100 simulation, we studied the impact of the IGM on $\text{Ly}\alpha$ spectral shapes between $z \sim 0$ and 5. We computed the distribution of the expected $\text{Ly}\alpha$ peaks and the peak asymmetry for different intrinsic spectra, redshifts, and large-scale environments. We find that the averaged transmission curves that are commonly applied give a misleading perception of the observed spectral properties. We show that the distributions of peak counts and asymmetry can lift the degeneracy between the intrinsic spectrum and IGM absorption. For example, we expect a significant number of triple-peaked $\text{Ly}\alpha$ spectra (up to 30% at $z \sim 3$) if the galaxies' HI distribution becomes more porous at higher redshift, as predicted by cosmological simulations. We provide a public catalog of transmission curves for simulations and observations to allow for a more realistic IGM treatment in future studies.

5.1 Introduction

The Lyman- α ($\text{Ly}\alpha$) line is a promising astrophysical observable for the neutral hydrogen distribution, from the scale of parsecs in star-forming regions (e.g., Kunth *et al.*, 1998; Yang *et al.*, 2016) all the way to cosmological scales. As such, it is among the most powerful observables to constrain the cosmic neutral fraction during the Epoch of Reionization (e.g., Dijkstra, 2014; Mason *et al.*, 2019).

While Ly α observations allow us to tackle a wide range of astrophysical questions, probing different spatial scales is a challenge because of a possible degeneracy of the origin of spectral features in those observations. Ly α observables are shaped not only through the neutral hydrogen (HI) distribution and the kinematics internal to galaxies, that is, the interstellar medium (ISM), but also by HI residing in the circumgalactic medium (CGM) (Steidel *et al.*, 2011; Wisotzki *et al.*, 2016) and the intergalactic medium (IGM).

At low redshift ($z \lesssim 1$), the impact of the IGM is minimal due to low neutral hydrogen fractions, such that observations with the Hubble Space Telescope reveal the spectral features imprinted by the ISM and CGM only. Usually, Ly α spectra show very little flux at line-center ($\lambda_c \sim 1216 \text{ \AA}$), and emission mostly originates on the red ($\lambda > \lambda_c$) side of the spectrum, with a significant fraction of spectra showing essentially no flux on the blue ($\lambda < \lambda_c$) side (Hayes *et al.*, 2014; Östlin *et al.*, 2014; Henry *et al.*, 2015; Yang *et al.*, 2016; review by Hayes, 2015).

At higher redshifts, however, the impact of the IGM increases and the picture is less certain. Individual observations at $z \sim 5-7$ mainly show a single peak redshifted by a few hundred km s^{-1} (Matthee *et al.*, 2017). At intermediate redshifts, larger statistical samples do measure an asymmetry toward the red (Erb *et al.*, 2014). This spectral evolution leaves, in principle, two possibilities: either the "intrinsic" Ly α spectra emergent from the galaxies do not vary strongly with redshift or they vary, but the IGM transmission, which is also evolving, makes the observed spectral properties remain similar. While the former scenario is supported by the fact that the low- z samples are selected as "analogs" of higher redshift Ly α emitters (e.g., Yang *et al.*, 2016), the latter is suggested by modern radiative transfer simulations using galactic hydrodynamical simulations as input (e.g., Laursen *et al.*, 2011; Gronke *et al.*, 2018; Smith *et al.*, 2019). Due to strong feedback mechanisms, they produce a porous ISM at high- z , and thus, the predicted Ly α spectra exhibit relatively large flux at line-center and the blue side of the spectrum. Differentiating between these pathways is crucial to properly disentangle the Ly α line's use as a probe of galaxy and IGM evolution.

While recent studies have focused mainly on the intragalactic Ly α transfer (e.g., Smith *et al.*, 2019), less attention has been devoted to the effect of the IGM on the Ly α spectral shape. Even large-scale studies that include the IGM have focused primarily on global statistics, such as the Ly α emitter clustering or luminosity function (e.g., Iliev *et al.*, 2008; Zheng *et al.*, 2010; Behrens *et al.*, 2018; Byrohl *et al.*, 2019). This chapter seeks to clarify the IGM's impact on the Ly α spectra using a recent cosmological simulation in the redshift range of $z = 0 - 5$.

5.2 Methodology

5.2.1 Simulations

We analyzed the IGM attenuation using the IllustrisTNG100 simulations (Marinacci *et al.*, 2018; Naiman *et al.*, 2018; Nelson *et al.*, 2018; Pillepich *et al.*, 2018a; Springel *et al.*, 2018) with a box size of 106.5 comoving Mpc for redshifts 0.0, 1.0, 2.0, 3.0, 4.0, and 5.0. The attenuation of Ly α flux was calculated using the Ly α radiative transfer code introduced in Chapter 3. Each tracing Monte Carlo photon calculating the attenuation along a given line of sight integrates the optical depth encountered during its propagation. The integration occurs simultaneously for a linear grid of wavelengths close to the line-center. The calculations are natively carried out on the unstructured Voronoi tessellation used in TNG. The initial Voronoi tessellation was created with a parallelized wrapper of *voro++* (Rycroft, 2009) on IllustrisTNG’s particle distribution. IllustrisTNG uses a time variable UV background with self-shielding (Faucher-Giguère *et al.*, 2009; Rahmati *et al.*, 2013) as well as a mimicking of the effect of active galactic nuclei (AGN) on the local radiation field (Vogelsberger *et al.*, 2013) that is responsible for the hydrogen’s ionization state in the IGM. Due to the long mean-free path of ionizing photons at the considered redshifts ($z \leq 5$) and our exclusion of HI close to star-forming regions (see below), using a full radiation hydrodynamics simulation as input would not change the outcome of our study.

The optical depth, τ , is integrated over the intervening medium for given lines of sight and a given input wavelength, λ_i . The optical depth along the way can be expressed by

$$\tau(\lambda_i) = \int_{s_0}^{\infty} ds n_{\text{HI}}(s) \sigma(\lambda(\lambda_i, v, s), T_{\text{HI}}), \quad (5.1)$$

where n_{HI} is the neutral hydrogen density and $\sigma(\lambda)$ is the Ly α cross-section. In Equation (5.1), we integrate over the physical distance, s , from the source along the chosen lines of sight (LOS). The temperature, T_{HI} , sets the thermal broadening reshaping the cross-section profile σ . The cross-section is evaluated in the rest-frame of the gas and, thus, depends on the peculiar velocity v and Hubble flow $H(z)$ at redshift z . The wavelength is shifted as $\lambda = \lambda_i \left[1 + \frac{v(s) + H(z) \cdot s}{c} \right]$, where c is the speed of light. We commonly express the wavelength as its offset $\Delta\lambda_e = \lambda - \lambda_c$ from the Ly α line-center at the emitters’ redshift.

The input wavelengths λ_i are evaluated in the rest-frame of the halos, which we identify as the mass-weighted velocity. We compute the optical depth within the wavelength range $[\lambda_c - 5 \text{ \AA}, \lambda_c + 3 \text{ \AA}]$ with a resolution of 0.02 \AA ($R \sim 60000$). For the IGM attenuation, we start summing contributions to the optical depth from a distance $s_0 = f r_{\text{vir}}$, where r_{vir} the virial radius of the halo. For a comparison with Laursen *et al.* (2011), we chose $f = 1.5$.

This implies that while photons are redshifted by the Hubble flow, no contributions to the optical depth are considered for $r < s_0$. The robustness of the chosen s_0 is confirmed in the supplementary material in Section 5.5.2. We integrate to 30 cMpc/h using periodic boundary conditions for the box. This length corresponds to a Hubble shift of 2600 km/s (at $z = 1.0$) or more. We verified that for the chosen input wavelength range λ_i , all wavelengths have significantly redshifted beyond the $\text{Ly}\alpha$ line-center as facilitated by the Hubble flow and, thus, no attenuation contributions are expected beyond this upper integration limit.

In our analysis, we consider the centers of halos as possible $\text{Ly}\alpha$ -emitting galaxies if they contain regions of active star formation and have a total halo mass above $5 \cdot 10^9 M_\odot$, as provided by IllustrisTNG's friends-of-friends halo catalogs. For each emitter, we evaluate the optical depth along 1000 LOS. The LOS are constructed as normal vectors of a 1000-faced Fibonacci sphere, evenly tracing possible directions with the same normal vectors used throughout. A reduced public data set of our transmission curves has been published as Byrohl *et al.* (2020a) and a full data set can be made available upon request.

5.2.2 Input spectra

To demonstrate how the IGM attenuation affects the observed spectra in a statistical sample using the individual attenuation curves, we need to assume some input (intrinsic) spectra with flux density $I_{\lambda, \text{input}}$. Here, we use three different toy models:

First, we have a symmetric double-peaked profile resulting from a static neutral hydrogen sphere, also referred to as Neufeld solution (Neufeld, 1990; Dijkstra *et al.*, 2006). We set the temperature as $T_{\text{HI}} = 10^4$ K and the column density to $N_{\text{HI}} = 10^{20} \text{ cm}^{-2}$ for this model.

Second, we consider the red peak only that corresponds to the Neufeld solution under the assumption of a significant outflow. These two spectral shapes bracket the observed cases at low- z , consisting of a single or double peak dominant toward the red.

Third, we assume a Gaussian at $\text{Ly}\alpha$ line-center with a width of $\sigma = 200$ km/s as input spectrum. Such a setup with a significant line-center flux can be motivated for galaxies with a larger impact of stellar feedback at high redshifts, leaving a more porous HI distribution. $\text{Ly}\alpha$ photons then escape closer to the line-center and are less susceptible to the gas kinematics (Neufeld, 1991; Hansen *et al.*, 2006; Gronke *et al.*, 2016a). For this reason, recent galactic hydrodynamical models post-processed with $\text{Ly}\alpha$ radiative transfer show a wide, fairly symmetric intrinsic profile with little absorption at line-center emergent from these galaxies (e.g., Smith *et al.*, 2019). The width was chosen to match the predictions of those models approximately.

5.3 Results

5.3.1 Averaged transmission curves

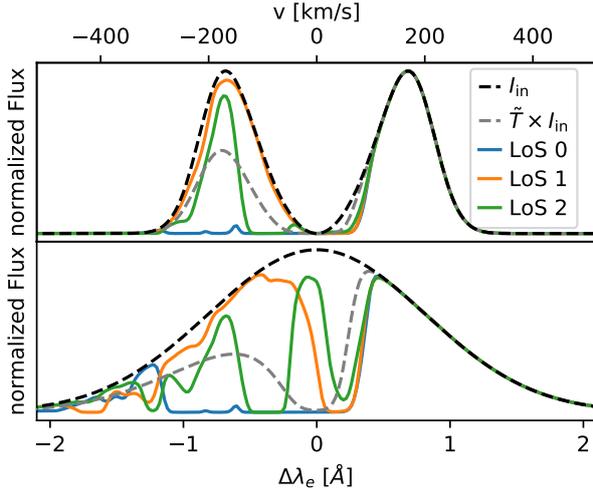


Fig. 5.1 Panels show two different input (i.e., intrinsic) spectra (black lines) and how the IGM attenuation shapes the observed spectra along different LOS at $z = 3.0$. All spectra are normalized by the respective underlying input spectrum's total flux. The gray dashed line shows the multiplication of the respective input spectrum with the median transmission curve, \tilde{T} , for all emitters and LOS at the given redshift. The colored solid lines show observed spectra for the same emitter but different LOS. Wavelengths are evaluated in the halo's rest-frame. **(top)** Input spectrum is obtained as analytic solution of a spherical hydrogen distribution with a column density $N = 10^{20} \text{ cm}^{-2}$ and temperature $T = 10^4 \text{ K}$ (Dijkstra *et al.*, 2006). **(bottom)** With a Gaussian intrinsic spectrum with standard deviation $\sigma = 200 \text{ km/s}$.

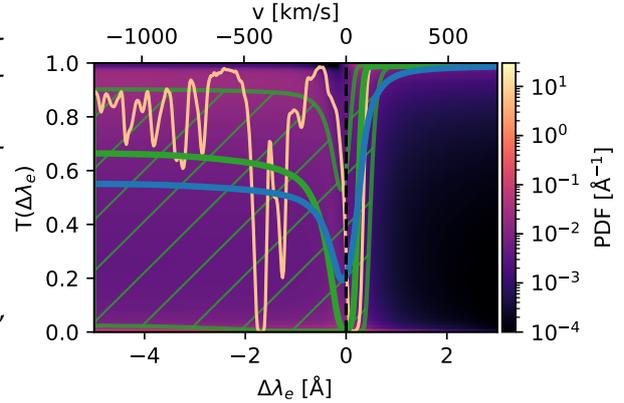
In Figure 5.1, we show the resulting spectrum along three LOS for the same origin with the Gaussian and double-peaked input spectra. In this plot, and in general, it is mostly the spectrum blueward of the line-center that is affected as those frequencies will eventually shift into the line-center by the Hubble flow. Figure 5.1 shows that the transmission blueward of the line-center can fluctuate strongly for different wavelengths of a given line of sight. In fact, the LOS in Figure 5.1 have been chosen such that the blue side of the observed spectrum exhibits a varying count of spectral peaks between zero and two for the Neufeld input spectrum. We formalize the count of peaks into a quantitative measure in Section 5.3.2. The transmission function is given as

$$T(\Delta\lambda_e) = \exp[-\tau(\Delta\lambda_e)] \quad (5.2)$$

and describes the fraction of the overall flux attenuated by the IGM between the emitter and the observer.

In Figure 5.2, we show the probability density function (PDF) $p(T, \Delta\lambda_e)$ of the transmission function T averaged over all emitters and LOS at $z = 3$. We also plot the mean and median curves (blue/green bold lines) along with the central 68 percentiles (hatched area). These curves show that the blue side is suppressed by roughly a factor of two,

Fig. 5.2 PDF of the transmission, T , as a function input wavelength shift, $\Delta\lambda_e$, for all emitters and the LOS at $z = 3$. The green line shows the median transmission with the hatched region enclosing the central 68% of all individual transmission curves. In contrast, the blue line shows the mean transmission curve over this PDF. The light orange curve shows an example of an individual LOS (corresponding to “LOS 1” in Figure 5.1).



while the red side is mostly unaffected. We find a trough around the line-center suppressing most of the flux. In general, we find these curves to be consistent with the results found by Laursen *et al.* (2011) and Gurung-López *et al.* (2020). Discrepancies in the asymptotic median value and the shaded 16th-84th percentile region are mostly due to these quantities being highly dependent on the spectral resolution (see supplementary material in Section 5.5.1). The spectral resolution in the literature is significantly lower with $R \lesssim 3000$ than $R \sim 60000$ used here.

For other redshifts, we find a similar qualitative trend over the shown wavelength range: the transmission on the blue side is increasingly suppressed at higher redshifts with a smaller asymptotic value toward $\Delta\lambda_e = -5 \text{ \AA}$ and a deeper trough around $\Delta\lambda_e = 0 \text{ \AA}$. We find the most likely offset for this central trough to be roughly 20 km/s at $z = 0.0$, monotonically increasing toward 70 km/s at $z = 5.0$. These velocity offsets and their redshift evolution are consistent with expected halo infall velocities at $r = 1.5r_{\text{vir}}$ (Barkana, 2004).

As illustrated by Figures 5.1 & 5.2, overall, the median curve is misleading and should be interpreted with caution: the underlying PDF $p(T|\Delta\lambda_e)$ of transmission $T(\Delta\lambda_e)$ is usually bimodal, that is, due to the large Ly α cross-section, it is mostly close to zero or unity – which is not well-represented by averaged transmission curves. For instance, in Figure 5.2, the bimodal distribution peaks around 0.0/0.9 (0.0/1.0) on the blue (red) side at $z = 3.0$. These features can be seen more clearly in the supplementary material in Section 5.5.3.

While this bimodality becomes more pronounced at lower redshifts, a unimodal distribution with $\langle T(\Delta\lambda_e) \rangle \sim 0$ forms at higher redshifts as the upper bimodal transmission peak disappears. This bimodal behavior has significant consequences for the observed spectra. Rather than blue peaks being uniformly suppressed along different LOS, some LOS will show a strong blue feature while others will show none. Similarly, there also is some variation for red peaks close to the line-center being suppressed given

the bimodality there. In Section 5.3.2, we investigate two different quantitative measures to characterize the spectral variations for different lines of sight as implied here.

5.3.2 Variations in transmission curves

After studying the averaged transmission curves, we proceed to quantify the variations of the transmission curves. Those variations are observable features of the emerging spectra after traversing the IGM.

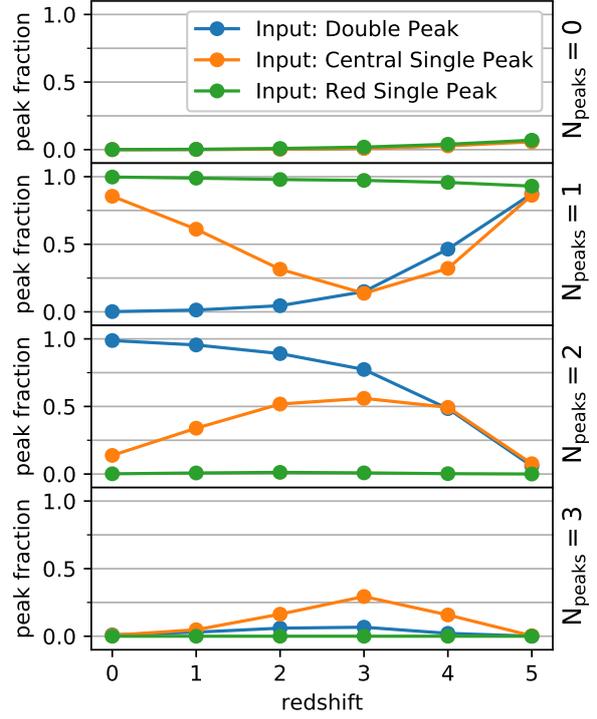
Peak distribution

As discussed in Section 5.1 and 5.2.2, the observed Ly α spectra at a low z usually exhibit a double or single red peaked spectrum. Attenuation in the IGM can modify the observed peak count in some LOS.

For an observed spectral flux density $I_\lambda(\Delta\lambda)$, we define a peak as connected flux density bins such that $I_\lambda(\Delta\lambda) > I_T$ for a threshold I_T . Here, we set $I_T = 0.01 \cdot \max_{0 < \lambda < \infty} (I_{\lambda, \text{input}})$. This criterion (while not its specific value) seeks to represent the flux sensitivity of a generic instrument. Furthermore, we require distinct connected areas to have a minimal separation of 0.2 \AA ($R \sim 6000$) from one another. This criterion aims to avoid any false identification of multiple peaks due to very small flux discontinuities that might additionally be below the spectral resolution of the spectrograph.

Figure 5.3 shows the distribution of the spectral peak count $0 \leq N_{\text{peaks}} \leq 3$ over the redshift range from 0 to 5 for the different input spectra. Hardly any LOS exist with $N_{\text{peaks}} > 3$ (not shown). The evolution with redshift is anchored by the intrinsic value of N_{peaks} at $z \sim 0$ (i.e., $N_{\text{peaks}} = 2$ and 1 for the double-peaked input spectrum and the other two, respectively) and a single red peak at $z = 5$, while in roughly 7% of all LOS, nearly all flux, and thus all peaks, are suppressed at $z = 5$. For the intermediate redshifts, the rugged transmission curve (cf. Section 5.3.1) causes small attenuation features that substantially increase the number of observable peaks for the wide, central input spectrum. Specifically, the number of resultant double and triple peaks increases to $\sim 50\%$ and 30% at $z \sim 2 - 4$, respectively. Given that at higher redshifts, attenuation features become wider and only the most redward peak survives, this gives a typical U-shaped evolution for N_{peaks} 1 to 3 in this case. For the same reason, there is a slight U-shape in the triple peak count of the double-peaked input spectrum. As the triple peak count is significantly at intermediate redshift, such a count is crucial in differentiating possible scenarios for the small-scale input spectra.

Fig. 5.3 Evolution of detectable Ly α peaks assuming the three different intrinsic spectra as described in Section 5.2.2. The peak detection algorithm is described in Section 5.3.2. The zero peak fraction rises to around 7% with redshift for the input spectra. The single and double-peak fractions differ for the input spectra but are anti-correlated. For the central and double-peak input spectrum, particularly the former, there is a notable triple peak fraction at around $z = 3$.



Blue peak flux

We introduce two observables to quantify the peak asymmetry. Namely, the flux ratio L_{ratio} between the integrated flux L_{blue} for wavelengths below the line-center and the total observed flux ($L_{\text{blue}} + L_{\text{red}}$):

$$L_{\text{ratio}} \equiv \frac{L_{\text{blue}}}{L_{\text{blue}} + L_{\text{red}}} = \frac{\int_0^{\lambda_c} T(\lambda) \cdot I(\lambda) d\lambda}{\int_0^{\infty} T(\lambda) \cdot I(\lambda) d\lambda}. \quad (5.3)$$

Analogously, we define the peak flux ratio F_{ratio} as the ratio of maximal flux blueward to the sum of the peak fluxes on both sides of the line-center, namely:

$$F_{\text{ratio}} \equiv \frac{F_{\text{blue}}}{F_{\text{blue}} + F_{\text{red}}} = \frac{\max_{0 < \lambda < \lambda_c} (T(\lambda) \cdot I)}{\max_{0 < \lambda < \lambda_c} (T(\lambda) \cdot I) + \max_{\lambda_c < \lambda < \infty} (T(\lambda) \cdot I)}. \quad (5.4)$$

We note that observational studies have used similar measures in the past (e.g., Erb *et al.*, 2014; Verhamme *et al.*, 2017).

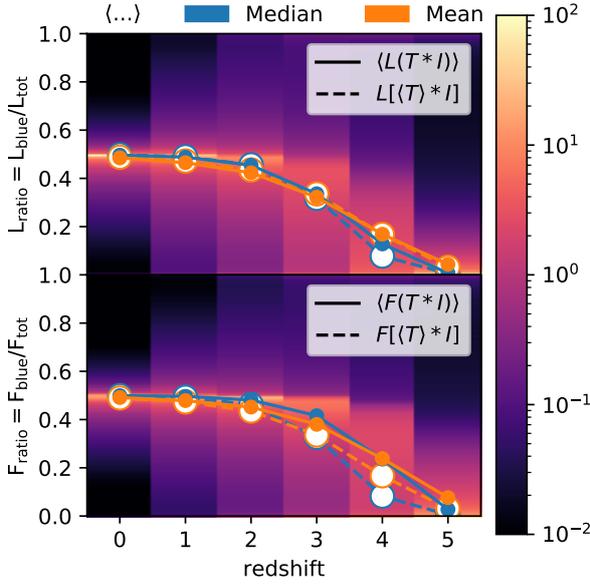


Fig. 5.4 Ratios L_{ratio} (**top**) and F_{ratio} (**bottom**) are shown as a PDF for all directions and emitters at a given redshift as shown by the colormap. We chose the double-peaked input spectrum, but we also find a very similar result for the centrally peaked input spectrum. The four overplotted lines in each panel show different averaging methods (see text). We assume that no peak can be detected (hence no ratio) when the flux remaining after passing the IGM is less than 1% of the intrinsic flux $I(\lambda)$.

In Figure 5.4, we show the distributions of L_{ratio} and F_{ratio} across all LOS (with both directions and emitters) using the double-peaked intrinsic spectrum introduced in Section 5.2.2. The distribution looks very similar when using the Gaussian input spectrum. We note that, in addition, we imposed a minimum flux on each side of the line-center as 1% of the flux of the input spectrum for a LOS to be deemed detectable. A ratio of 0.5 signifies an equal impact of the IGM on the blue and red side of the line-center. Thus, while Figure 5.4 predicts the observed distribution given the idealized input spectrum, it also represents the IGM’s impact on the asymmetry as such, indicating a favorable escape of blue (red) photons through the IGM for values over (below) 0.5.

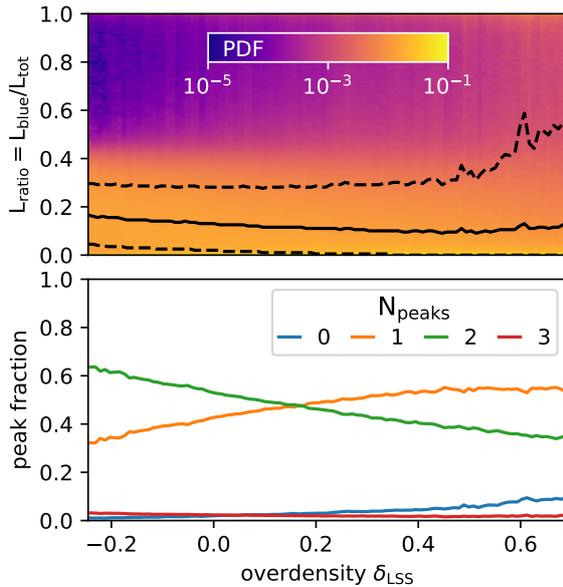
The solid lines show the mean and the median for the ratios, while the dashed lines show the ratio for the mean and median transmission curves $\langle T(\Delta\lambda) \rangle$ multiplied by the input spectra. Both ratios intuitively follow the expected redshift evolution for all lines: At redshift $z = 0$, the asymmetry is mostly unaffected by the IGM. At higher redshifts, the averaged ratios drop toward zero at $z = 5$ as the IGM becomes more opaque due to a higher HI density. A closer look at the peak asymmetry distribution at a given redshift reveals a more nuanced picture: At low to intermediate redshifts ($z \lesssim 3$), the distribution appears somewhat symmetric around the median with increasing variance and skewness for higher redshifts. Therefore, blue peaks can be prominent features of a fraction of observed spectra even at high redshifts. We find fractions of 13.5%, 17.5%, 14.3%, 8.5%, 4.1%, and 1.8% from redshift 0 to 5, enhancing the blue side of the spectrum (i.e., $L_{\text{ratio}} > 0.5$). Observational studies calculating the enhanced blue peak fraction exist (Erb *et al.*, 2014; Trainor *et al.*, 2015) with up to 16 – 22% at $z \sim 3$. However,

a fair comparison is not possible due to small number statistics, a varying definition of the flux ratio, and the poor spectral resolution. Future studies overcoming these challenges would be useful for constraining the (a)symmetry of the small-scale spectra.

A range of spectra will continue to contain significant blue contributions at high redshifts. This is another reason why the averaged transmission curves could be misleading concerning the underlying peak asymmetry distribution. For instance, for $z = 5.0$, the median transmission curve leads to a L_{ratio} on sub-percent level, leading to the perception that blue peaks are singularities at such redshift, when in reality we find roughly 4% of Ly α emitting galaxies that still show significant blue flux ($L_{\text{blue}} \geq 0.25 \cdot L_{\text{red}}$; analogously, for F , we even find 10%). For the other redshifts (0, 1, 2, 3, and 4) we find 99.5%, 97.0%, 91.8%, 75.5%, and 33.3% of LOS with significant blue flux (if present intrinsically).

Large-scale environment

Fig. 5.5 L_{ratio} and N_{peaks} statistics shown as a function of the large scale overdensity δ_{LSS} at redshift $z = 4.0$. The input spectrum is the Neufeld solution. The colormap in the upper panel shows the underlying PDF normalized at a given δ_{LSS} . Overdense regions overall decrease L_{ratio} . At a higher overdensity, this trend halts as the ratio becomes more fluctuating for different LOS. For the peak fractions, we find a gradual decrease of double peaks. This is caused by absorption features on the blue side of the line-center. Thus, the decrease of double peaks is strongly correlated with the increase of the fraction of single peaked spectra.



Besides the redshift and input spectrum, we find that our proposed statistics also depend on the emitters' large-scale environment. Most prominently, we find a correlation of the flux ratio and peak fraction with the linear overdensity as shown in Figure 5.5 for $z = 4.0$. We calculate the overdensity using a Gaussian smoothing kernel with $\sigma = 3$ cMpc/h. The flux ratio slightly decreases toward higher overdensities indicating more intervening HI. Additionally, the scatter of the flux ratio strongly increases as more varying matter structure is passed along the LOS. The fraction of double peaks, as present in the input spectrum, strongly decreases toward higher overdensities as more blue peaks are attenuated. In up to 10% of the cases, the red peak is additionally suppressed,

completely obscuring those Ly α emitting galaxies in high overdensity environments. This also affects the clustering signal of Ly α emitters, as has been studied by Zheng *et al.* (2011a) and Behrens *et al.* (2018) in more detail. The correlation with overdensity is much weaker for redshift between $z = 0.0$ and 2.0 , where the most likely outcome remains $L_{\text{ratio}} = 0.5$ over the overdensity range. This behavior changes for $z = 3.0$ and $z = 4.0$ in overdense regions, where the distribution starts to get skewed toward L_{ratio} close to zero. At higher redshifts, the correlation appears to be smaller as most blue peaks are already attenuated even in underdense regions.

5.4 Conclusions

The Ly α line can be used to constrain the neutral hydrogen distribution from galactic to cosmological scales. This versatility is, however, also problematic since degeneracies between these scales exist that come into play at $z \gtrsim 3$, when the more opaque IGM can compensate for the effects of a more porous ISM. Such a scenario – which is suggested by cosmological simulations – allows not only for the escape of Ly α photons closer to line-center but also the escape of ionizing photons that are susceptible to the same galactic HI distribution (e.g., Dijkstra *et al.*, 2016).

Using a large set of transmission curves that we calculated from the IllustrisTNG100 simulations, we quantified the impact of the IGM on spectra for $z = 0 - 5$. In doing so, we studied a new approach to breaking the degeneracy using two statistics, namely the peak count and peak asymmetry (Section 5.3.2/ 5.3.2). In particular, we found the fraction of triple peaks to be an important differentiator for different intrinsic spectra. In contrast, we show that the commonly used averaged transmission curves can be misleading with regard to the interpretation of the impact of the IGM on observed Ly α spectra and their redshift evolution.

Making our catalogs of transmission curves publicly available (Byrohl *et al.*, 2020a) opens this work up to other studies that look to incorporate an improved approach to IGM treatment and its impact on Ly α spectra. At the same time, it can provide flexibility when refining the statistics presented and the intrinsic spectral modeling in the future.

Our findings require a high spectral resolution (optimally $R \gtrsim 6000$), particularly for the peak count statistic. Hence, current samples of Ly α spectra at $z \gtrsim 3$ have too low a spectral resolution for testing our predictions (e.g., Erb *et al.*, 2014; Herenz *et al.*, 2017). However, individual triple-peaked spectra have been observed (e.g., Rivera-Thorsen *et al.*, 2017; Vanzella *et al.*, 2020) and future surveys will increase this count to a statistical sample that allows for the above-described degeneracy to be broken.

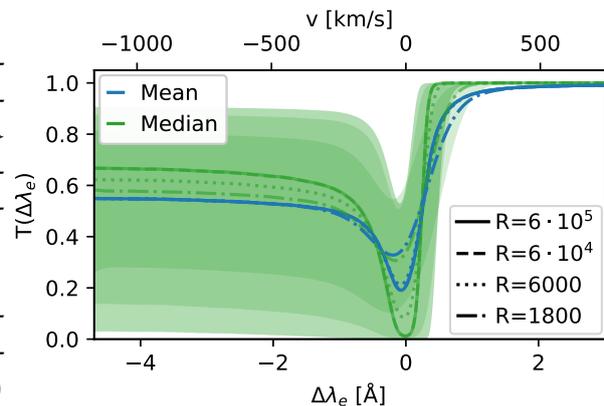
Acknowledgements

We thank Christoph Behrens for providing us with the early version of *ILTIS*, Eiichiro Komatsu and Shun Saito for useful discussions on the draft for Byrohl *et al.* (2020b), and Dylan Nelson for his help processing the IllustrisTNG data. The radiative transfer simulations and analysis were conducted on the supercomputers at the Max Planck Max Planck Computing and Data Facility (MPCDF). We acknowledge use of the Python programming language, and use of the Astropy (Collaboration *et al.*, 2013; Astropy Collaboration *et al.*, 2018), Numpy (Harris *et al.*, 2020), IPython (Pérez *et al.*, 2007), Dask (Dask Development Team, 2016), h5py (Collette, 2013) and Matplotlib (Hunter, 2007) packages for post-processing. The primary co-author of Byrohl *et al.* (2020b), Max Gronke, was supported by NASA through the NASA Hubble Fellowship grant HST-HF2-51409 and acknowledges support from HST grants HST-GO-15643.017-A, HST-AR-15039.003-A, and XSEDE grant TG-AST180036.

5.5 Supplementary material

5.5.1 Spectral resolution

Fig. 5.6 Impact of the spectral resolution on the PDFs of the transmission function T as a function of input wavelength shift $\Delta\lambda_e$. We show the PDF of T over all emitters and LOS at $z = 3$, similarly to Figure 5.2. In blue, we show the mean for a given $\Delta\lambda_e$, while in green, we show the median. We show the transmission at different spectral resolutions $R \in \{600000, 60000, 6000, 1800\}$ (solid, dashed, dotted, dash-dotted). Resolutions below 600000 are computed as convolution with a Gaussian of FWHM λ_{FWHM} that sets the spectral resolution as $R = \lambda_c / \lambda_{\text{FWHM}}$. The shaded regions enclose the central 68% of all individual transmission curves. The darkest (lightest) shade corresponds to the lowest (highest) spectral resolution. As the resolution increases, the median increases and the shaded region widens. At our fiducial resolution of ~ 60000 , these statistics have converged.



In Figure 5.6, we show the impact of the spectral resolution on the averaged transmission curve at $z = 3$. We show the median along with the 16th and 84th percentile

and the mean over the different emitters and LOS for at a given wavelength offsets $\Delta\lambda_e$. We find that the mean is nearly independent of the spectral resolution between 6000 and 600000, while the median and percentiles, in general, are strongly dependent on chosen resolution. We note that the result has converged at $R \sim 60000$, which is our fiducial resolution throughout this chapter. At $R = 1800$, the smoothing is larger than the size of the transmission curve's central trough, and both mean and median thus significantly change in its proximity.

Our simulations at $R \sim 60000$ are expected to be converged due to the Doppler broadening in the line profile. The full width at half maximum (FWHM) due to Doppler broadening is given as $\lambda_{\text{FWHM}} \approx 2.35\lambda_c \sqrt{\frac{k_B T}{m_p c^2}}$. Assuming an IGM temperature of $T \sim 10^4$ K, this implies a spectral resolution of $R \sim 14000$. Features imprinted by the hydrogen structure on smaller scales will thus be washed out.

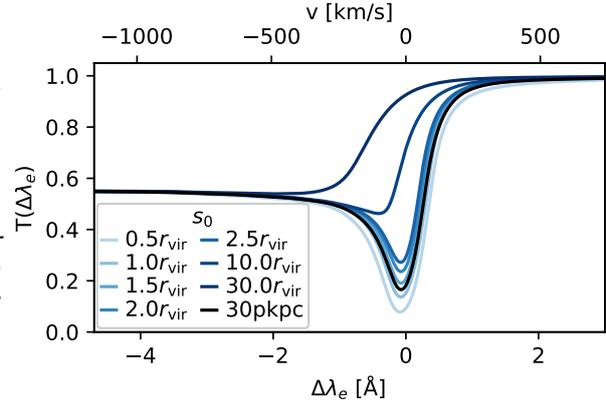
The difference in the behavior of median and percentiles compared to the mean is easily explained: Linearity, that is, $\text{STAT}[A + B] = \text{STAT}[A] + \text{STAT}[B]$ for two distributions, A and B , and a summary statistic, STAT , holds for the mean but not for percentiles such as the median. Take, for example, two PDFs, $A = P(T)_{\Delta\lambda_e=\lambda_i}$ and $B = P(T)_{\Delta\lambda_e=\lambda_j}$, where λ_i and λ_j describe two neighboring wavelength bins. The mean transmissivity T over those two bins can be determined from the mean of A and B alone, while this is not possible for the median. In particular, if two neighboring bins have the same mean, the mean of the sum of those distributions has to be the same, whereas this is not true for the median, as can be readily seen in Figure 5.6. This behavior is furthermore complicated for the median as the distributions in neighboring bins are strongly correlated. We recommend using the mean transmission curves over the median unless the spectral resolution is clearly stated to address this issue.

We obtain a triple peak fraction of 28.5%, 28.4%, 17.6%, and 0.1% for spectral resolutions of 600000, 60000, 6000, and 1800 given the Gaussian input spectrum. Thus, in addition, the peak count is converged at our fiducial resolution of 60000. At typical resolutions of spectrographs today ($R \sim 1800$), triple peaks are rare to non-existent for given small-scale spectrum.

5.5.2 Start of integration

In Figure 5.7, we vary the start s_0 of the integral for the optical depth and show the resulting mean transmission curves. For small changes around $s_0 = 1.5 \pm 1.0r_{\text{vir}}$, we find the central trough's depth to slightly change and becoming deeper for smaller values of s_0 . Going beyond $f = 10.0$ removes the relevant scale responsible for the mean attenuation

Fig. 5.7 Impact of the lower integration bound s_0 for mean transmission curves. We averaged over all LOS and Ly α -emitting galaxy candidates at $z = 3$. We show the cases of $f = 0.5, 1.0, 2.0, 2.5$ around our fiducial case of 1.5. These results are contrasted with cases $f = 10.0, 30.0$ for interaction on large scales only. In addition, we show the case of a fixed $s_0 = 30$ pkpc irrespective of the virial radius of the underlying halo.



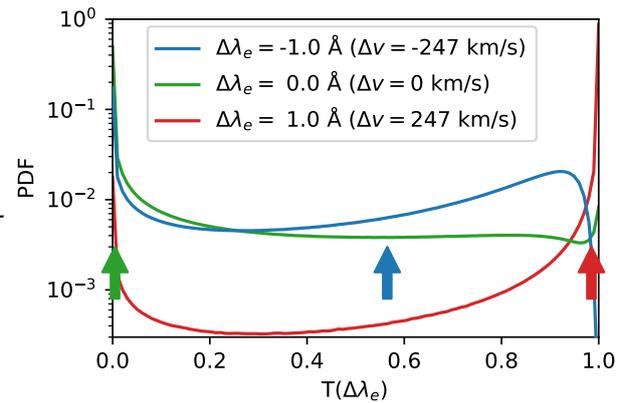
curve's central trough. On an even larger scale ($f = 30.0$), attenuation is shifted toward lower input wavelengths due to the Hubble flow.

The triple peak count remains stable around $\sim 30\%$ with 28.9%, 30.2%, 29.4%, 28.2%, and 27.1% for f equal to 0.5, 1.0, 1.5, 2.0, and 2.5 with $s_0 = fr_{\text{vir}}$. Similarly, when using a constant s_0 rather than depending on the virial radius we obtain 29.6% for $s_0 = 30$ pkpc. Thus, our results are robust for chosen $f = 1.5$.

The peak fraction is largely determined in the proximity of the targeted halo. The triple peak fraction significantly drops to 17.6% and 6.1% for $f = 10.0$ and $f = 30.0$, showing that spectra are shaped both in the proximity of emitters ($f \gtrsim 2.5$) and on large scales ($f \gtrsim 10.0$).

5.5.3 Bimodality of the transmission curves

Fig. 5.8 Distribution of transmission T at fixed input wavelength offset $\Delta\lambda_e$ at $z = 3.0$ across all emitters and LOS. The arrows show the accordingly color-coded median.



In Figure 5.8, we show the distribution of transmission coefficients T at fixed input wavelengths $\Delta\lambda_e$ at $z = 3.0$. We show three different wavelengths: Blue of the line-center (blue line), in the line-center itself (green line), and red of the line-center (red line). The bimodality for the line-center and larger wavelengths peaks at T zero or unity, while

shorter wavelengths show their second peak below unity (~ 0.9 here). This second peak's existence and position is redshift dependent (see main body). We also show the medians (arrows) color-coded for the corresponding distribution. For the line-center and redder wavelengths, one peak is commonly so dominant that the median is shifted to there accordingly. While in this case, the median is aligned with one of the peaks, this is not the case for blue wavelengths. for

SIMULATING LYMAN-ALPHA HALOS

The content of this chapter has been published as Byrohl et al. (2021).

Context and summary

Extended Lyman-alpha emission is now commonly detected around high redshift galaxies through stacking and even on individual basis. Despite recent observational advances, the physical origin of these Lyman-alpha halos (LAHs), as well as their relationships to galaxies, quasars, circumgalactic gas, and other environmental factors remains unclear. We present results from our new Lyman-alpha full radiative transfer code `VOROILTIS` which runs directly on the unstructured Voronoi tessellation of cosmological hydrodynamical simulations. We make use of the TNG50 simulation and simulate LAHs from redshift $z = 2$ to $z = 5$, focusing on star-forming galaxies with $8.0 < \log_{10}(M_*/M_\odot) < 10.5$. While TNG50 does not directly follow ionizing radiation, it includes an on-the-fly treatment for active galactic nuclei and ultraviolet background radiation with self-shielding, which are important processes impacting the cooling and ionization of the gas. Based on this model, we present the predictions for the stacked radial surface brightness profiles of $\text{Ly}\alpha$ as a function of galaxy mass and redshift. Comparison with data from the MUSE UDF at $z > 3$ reveals a promising level of agreement. We measure the correlations of LAH size and central brightness with galaxy properties, finding that at the masses of $8.5 \leq \log_{10}(M_*/M_\odot) \leq 9.5$, physical LAH sizes roughly double from $z = 2$ to $z = 5$. Finally, we decompose the profiles into contributions from diffuse emission and scattered photons from star-forming regions. In our simulations, we find rescattered photons from star-forming regions to be the major source in observed LAHs. Unexpectedly, we find that the flattening of LAH profiles at large radii becomes dominated by photons originating from other nearby halos rather than diffuse emission itself.

6.1 Introduction

The Lyman- α ($\text{Ly}\alpha$) line of hydrogen at 121.567 nm is one of the brightest emission lines in the Universe. It allows us to detect, and trace the distribution of, galaxies even out to very high redshifts $z > 5$. These Lyman- α emitters (LAEs; Partridge *et al.*, 1967) can be used to probe the physics of galaxy formation (Finkelstein *et al.*, 2009; Nagamine *et al.*, 2010; Erb *et al.*, 2014) as well as constrain cosmological parameters and large-scale structure (Hill *et al.*, 2008; Adams *et al.*, 2011). Starting in the 80s, spatially extended $\text{Ly}\alpha$ emission has been detected, often called Lyman- α blobs (LABs) and Lyman- α nebulae with large extents between $\sim 10 - 100$ pkpc (McCarthy *et al.*, 1987; Heckman *et al.*, 1991a; Steidel *et al.*, 2000).

More recently, Lyman- α halos (LAHs) have been discovered around star-forming galaxies that show $\text{Ly}\alpha$ emission far beyond the galaxies' optical bodies, tracing the circumgalactic rather than interstellar gas (e.g. Hayes *et al.*, 2013). While LAHs are fainter and smaller than LABs in their $\text{Ly}\alpha$ extent, they might be a generic feature around $\text{Ly}\alpha$ emitting galaxies.

Narrow-band imaging can efficiently detect LAHs at targeted redshifts through stacking (Hayashino *et al.*, 2004; Steidel *et al.*, 2011; Matsuda *et al.*, 2012; Feldmeier *et al.*, 2013), and narrow-band surveys enable ultra-deep, blind samples of LAHs around distant galaxies (Momose *et al.*, 2014, 2016; Kakuma *et al.*, 2021). In the last years modern surveys performed with integral field unit (IFU) spectrographs on 10m-class telescopes, such as the Multi Unit Spectroscopic Explorer (MUSE) and the Keck Cosmic Web Imager (KCWI), take place. These new instruments allow the study of individual, faint LAHs opposed to previous narrow-band stacks. Along with the IFUs' spectral resolution these recent surveys largely increase the information available from LAH observations. Hundreds of individually extended Lyman- α halos at $z \gtrsim 2$ have been revealed since (Wisotzki *et al.*, 2016). Many of these are specifically targeted samples which focus on bright quasars (Borisova *et al.*, 2016; Cai *et al.*, 2019; Guo *et al.*, 2020; O'Sullivan *et al.*, 2020), based on strong earlier evidence of enhanced $\text{Ly}\alpha$ emission around active galactic nuclei (AGN) (Cantalupo *et al.*, 2014; Arrigoni Battaia *et al.*, 2016, 2019; Farina *et al.*, 2019). Others exploit the large field of view of MUSE, in particular, to conduct blind surveys for LAHs around more typical, generally star-forming galaxies (Leclercq *et al.*, 2017; Wisotzki *et al.*, 2018; Leclercq *et al.*, 2020). At the same time, follow-up with other instruments such as ALMA reveals complementary views on other gas phases within LAHs including CO (Emons *et al.*, 2019).

Beyond the circumgalactic medium (CGM), first attempts have been made to detect cosmic web filaments in $\text{Ly}\alpha$ emission (Gallego *et al.*, 2018; Lusso *et al.*, 2019; Umehata

et al., 2019), as well as to study the interface between the IGM and CGM as gas flows feed galaxies (Martin *et al.*, 2019). Finally, Ly α emission is also a powerful cosmological tool. The Hobby-Eberly Telescope Dark Energy Experiment (HETDEX) is an IFU survey which will detect up to a million LAEs, as well as many extended LAHs, at more moderate spatial and spectral resolution (Hill *et al.*, 2008).

Despite extensive observational detection and characterization, the physical properties and nature of Ly α halos remains an open topic. It is unclear whether the extended profiles are mainly sourced by (i) diffuse emission outside of central galaxies, or whether (ii) scatterings of Ly α photons emitted from within central galaxies power observed LAHs. Diffuse emission is commonly considered to be sourced by gravitational cooling (Haiman *et al.*, 2000; Fardal *et al.*, 2001; Faucher-Giguère *et al.*, 2010) and fluorescence (Gould *et al.*, 1996; Cantalupo *et al.*, 2005; Kollmeier *et al.*, 2010; Mas-Ribas *et al.*, 2016), while star-formation and quasars can provide significant emission within galaxies (Dijkstra *et al.*, 2006; Zheng *et al.*, 2011b) that can scatter with neutral hydrogen in the CGM. Emission from orbiting satellite galaxies can also lead to extended Ly α profiles (Mas-Ribas *et al.*, 2017).

The difficulties in determining the powering source of LAHs are closely linked to the resonant nature of Ly α photons that can scatter many times in astrophysical environments before escaping towards the observer. This causes the observed frequency and angular position to significantly change due to radiative transfer (RT). Only in the simple, symmetric geometries, RT can be solved analytically (Harrington, 1973; Neufeld, 1990; Loeb *et al.*, 1999; Lao *et al.*, 2020). Moving to more realistic setups and in particular hydrodynamical simulations, RT has to be solved numerically.

Recently our theoretical understanding of the Ly α emission around galaxies has been pushed forward with the development of cosmological hydrodynamical simulations which are able to produce broadly realistic galaxy populations (Genel *et al.*, 2014; Vogelsberger *et al.*, 2014a; Vogelsberger *et al.*, 2014b; Schaye *et al.*, 2015; Dubois *et al.*, 2016; Pillepich *et al.*, 2018a; Davé *et al.*, 2019). Crucially, these simulations predict the full distribution of gas, including neutral hydrogen, in and around galaxies. This enables explicit RT calculations to solve for the propagation and scattering of Ly α photons through the interstellar, circumgalactic, and intergalactic media (Laursen *et al.*, 2009; Behrens *et al.*, 2018). Due to the computational expense, these calculations are done exclusively in post-processing.

There have been various recent efforts to understand the Ly α halos in emission by coupling RT with hydrodynamical simulations (e.g. Lake *et al.*, 2015; Smith *et al.*, 2019; Kimock *et al.*, 2020). These works typically consider one, to a few tens, of galaxies – rather than full cosmological volumes – making it difficult to draw conclusions about

the environmental dependencies of LAHs. Notable exceptions are Gronke *et al.* (2017a), who predicted LAH properties from the Illustris simulation, and Zheng *et al.* (2011b) who predicted low surface brightness Ly α emission from a cosmological reionization simulation ($z = 5.7$), although with a limited, ‘halo-scale’ hydrodynamical resolution ~ 30 pkpc.

In these theoretical works the most important emission origin(s) and source(s) remain disputed. For example, Lake *et al.* (2015) find good agreement for their set of 9 LAHs with mass $10^{11.5} M_{\odot}$ contrasted with data from Momose *et al.* (2014), stressing the importance of gravitational cooling in the outer halo to explain the observed profiles, while Gronke *et al.* (2017a) simulate Lyman- α nebulae with masses $10^{11.5} - 10^{13.5} M_{\odot}$ and find the simulation can produce halos as large and luminous as those observed, only using central emission from AGN and star-formation.

Recently, cosmological volumes from modern hydrodynamical simulations have been studied in the context of Ly α emission focusing on the detectability of the cosmic web (Elias *et al.*, 2020; Witstok *et al.*, 2021). Latest observations of the Ly α cosmic web in Bacon *et al.* (2021) might point at the importance of emission from (faint) galaxies that has been missing in former theoretical explorations.

In this work we improve on several aspects of previous computational studies, revisiting the nature of Lyman- α halos. Specifically, we couple the new, high-resolution cosmological magnetohydrodynamical simulation TNG50 (Nelson *et al.*, 2019a; Pillepich *et al.*, 2019) of the IllustrisTNG project to our new radiative transfer code voroILTIS. The former provides a competitive combination of volume (a statistically robust sample of $\sim 6,800$ galaxies with $M_{\star} > 10^7 M_{\odot}$ at $z = 2$) and resolution (~ 100 parsecs in the ISM, < 1 kpc in the CGM). The latter includes several Ly α emission models and a Monte Carlo treatment of the scattering process directly on the full Voronoi tessellation of the gas distribution of the entire simulation volume, enabling a self-consistent treatment of IGM attenuation (Byrohl *et al.*, 2020b). Our setup enables us to statistically contrast the simulation predictions to existing LAH observations, while also probing questions regarding the dominant origins, emission sources and relevance of rescattering for the existence of LAHs, and making future predictions in as of yet unobserved regimes.

The structure of this chapter is as follows. In Section 6.2, we describe our radiative transfer code voroILTIS, the Ly α emission model, and the analysis details of the underlying IllustrisTNG simulations on which the radiative transfer code is run. In Section 6.3 we present the results for the radial profiles and related reduced quantities from our simulations and a comparison to observational data. In Section 6.4 we discuss the radial profile shapes and reduced quantities in more detail. We summarize our findings in Section 6.5.

6.2 Methods

6.2.1 IllustrisTNG and TNG50

For this work, we use the outcome of the IllustrisTNG simulations – both the galaxy properties as well as gas distributions – as the basis for our radiative transfer simulations of Lyman- α halos that we describe introduce in Chapter 3.8. To shortly recap, the IllustrisTNG simulations (hereafter, TNG; Marinacci *et al.*, 2018; Naiman *et al.*, 2018; Nelson *et al.*, 2018; Pillepich *et al.*, 2018a; Springel *et al.*, 2018) are a series of three large-volume magnetohydrodynamical cosmological simulations of galaxy formation run with the AREPO code (Springel, 2010). AREPO solves the coupled equations of self-gravity and ideal, continuum magnetohydrodynamics (Pakmor *et al.*, 2011) with a ‘moving mesh’ discretization based on an unstructured Voronoi tessellation of space.

Of the three IllustrisTNG simulations, TNG50, TNG100, and TNG300 (Nelson *et al.*, 2019a) we here exclusively use the highest resolution box TNG50 (Nelson *et al.*, 2019b; Pillepich *et al.*, 2019) with a gas mass resolution of $m_{\text{bayron}} = 8.5 \times 10^4 M_{\odot}$ and a dark matter mass resolution of $m_{\text{DM}} = 4.5 \times 10^5 M_{\odot}$ which is ~ 15 (~ 120) times higher than TNG100 (TNG300). The corresponding spatial resolution of TNG50 is of order ~ 100 physical parsecs in the dense ISM, and this small-scale structure is useful given the strong resolution dependence of Ly α radiative transfer at lower resolutions as demonstrated in Behrens *et al.* (2018) and Camps *et al.* (2021), although a more sophisticated model for the cold phase of the ISM would be needed to better capture Ly α RT effects on small scales.

6.2.2 voroILTIS

The Ly α radiative transfer is calculated with an updated version of ILTIS¹, a light-weight line emission transfer code as presented in Behrens *et al.* (2019). ILTIS implements a Monte Carlo approach, spawning single-wavelength photon ‘packages’ (representing a large number of actual photons) at emission sites, and following their scattering as they traverse the underlying gas distribution. For efficiency, at each scattering event we output the attenuated luminosity contributions along specified lines of sight towards assumed observers (the ‘peeling-off’ algorithm; Whitney, 2011).

We have developed a new version of the code, voroILTIS, which runs directly on the unstructured mesh of a Voronoi tessellation. As IllustrisTNG uses this geometry to

¹The public version of the ILTIS code is currently available at github.com/cbehren/ILTis, where the Voronoi version will also be released in the future.

represent the gas distribution during the simulation and for the evolution of hydrodynamical quantities, no intermediate interpolation or re-sampling steps are required, and the density field in the RT calculation is self-consistent with the simulation. The mesh is re-created in postprocessing with a parallelized wrapper to the Voronoi tessellation code *voro++* (Rycroft, 2009). We then spawn photons for each cell in the mesh according to the local emissivity, an update which makes simulating diffuse emission feasible.

After emission, photons propagate through the simulation domain, and across the periodic boundaries of the box as appropriate. Given the amplitude of the Hubble flow (several 100 km/s/Mpc) in the simulated redshift range and typical gas velocities, photons are quickly shifted into the far wings of the line profile where they have negligible cross-section. Therefore, we do not need to construct light-cones and simply propagate all photons for the necessary propagation length l of order ~ 10 Mpc. Here, we chose $l = 28$ cMpc/h. An upcoming methods paper will document the code improvements used in the present work (Behrens & Byrohl, in prep). We present *voroILTIS* in Chapter 3. Additional information on the predecessor of *ILTIS*, that has been used in Behrens *et al.* (2018) and Byrohl *et al.*, 2019, can be found in those publications.

6.2.3 Lyman-alpha radiative processes

$\text{Ly}\alpha$ photons are dominantly created by recombination of ionized hydrogen atoms with electrons, and de-excitation of excited neutral hydrogen. Different physical processes can power their creation: For recombination, ionizing radiation originates from star-forming regions as well as from the metagalactic ultraviolet background. For collisional de-excitation, the thermal state of the gas provides the gas heating mechanism. We refer to those processes to emission sources and in figures call them ‘rec’, and ‘coll’, respectively. As we discuss in the next paragraphs, we use a special description for the recombinations in star-forming regions, which we abbreviate as ‘SF’ in figures.

We model recombinations after ionization by a spatially uniform background radiation field and AGN radiative feedback as implemented in TNG50. These effects can significantly change the ionization and temperature state of hydrogen that is reflected self-consistently in TNG50. We can then adopt the description in Equation (2.44) to calculate the recombinations. In addition to recombinations, the de-excitation of excited hydrogen atoms can lead to the emission of $\text{Ly}\alpha$ sourced by collisional excitations depending on the gas thermal state, which we incorporate according to Equation (2.45). In star-forming gas cells, where the TNG model invokes a sub-grid effective equation of state model for the two-phase ISM (Springel *et al.*, 2003), the simulation’s temperature and hydrogen density do not reflect their physical meaning entering Eqn. (2.44). Hence, we instead derive the

recombination rate from the star-formation rate as a proxy for the amount of ionizing flux recombining according to Equation (2.47). In condensed form, our emission model is thus a combination of the following three sources:

$$\epsilon_{\text{rec}} = f_{\text{rec}}(T) n_e n_{\text{HIII}} \alpha(T) E_{\text{Ly}\alpha} \quad (6.1)$$

$$\epsilon_{\text{coll}} = \gamma_{1s2p}(T) n_e n_{\text{HI}} E_{\text{Ly}\alpha} \quad (6.2)$$

$$\epsilon_{\text{SF}} = 10^{42} \left(\frac{\dot{M}_*}{M_{\odot} \text{yr}^{-1}} \right) \frac{\text{erg/s}}{V_*} \quad (6.3)$$

A more detailed description of these equations and the choice of coefficients is given in Section 2.6.1 and 2.6.2.

The star-formation rate is taken directly from the TNG output in each cell volume. The TNG star-formation model is described in Pillepich *et al.* (2018b). In short, a gas cell is star-forming if and only if its physical density exceeds 0.1 hydrogen atoms per cubic centimeter, in which case collisionless star particles are stochastically formed (also see Springel *et al.*, 2003; Vogelsberger *et al.*, 2013).

For recombination and de-excitations, we spawn one weighted photon package per Voronoi cell with the luminosity according to the models above. For star-formation, we only spawn photons in Voronoi cells bound to a halo of mass $M_{200,\text{crit}} > 10^{10} M_{\odot}$. In TNG50 this results in a sample of ~ 5000 to ~ 13000 LAHs at redshifts $z \in [2, 5]$.

We found that a constant photon count per cell for Equations (2.44) and (2.45) results in good convergence for the radial profiles. In such a scheme, photon packages can carry vastly varying luminosity weights. This however is desired as we need to also trace faint, optically relatively thin regions in the outer CGM where surface brightnesses are orders of magnitudes lower than in the central regions of the LAHs.

In contrast, a Monte Carlo sampling relying on a fixed luminosity weight per photon package appears faster for the emission from star-forming cells (see Eqn. (2.47)). For simplicity, we stick with the fixed Monte Carlo photon count per cell, but confirm that this photon count leads to a converged result (see supplementary material in Section 6.6.1) for star-forming cells.

We refer to these photon packets as ‘photons’ for brevity. Photons are always injected at the Ly α line-center in the rest-frame of the emitting gas cells. We have also explored different spectral emission distribution and generally find little impact – details are given in the supplementary material in Section 6.6.1.

In our model, we do not account for escaping ionizing radiation from star-forming regions that recombines outside of its emission region (“fluorescent radiation”). Fluorescence in our simulations hence only arises from the uniform UV background and

the radiation field from AGN incorporated in TNG. In the supplementary material in Section 6.6.1 we discuss the impact of ionizing sources by means of the AGN radiation implemented within the model underlying TNG50. The radiative feedback from AGN is accounted for in that it affects the ionization state, temperature and cooling rates of the gas and hence can boost recombinations and collisional excitations. We do not consider additional recombinations in gas cells due to the activity of SMBHs.

6.2.4 Post-processing and observational realism

The radiative transfer simulations provide us with large outputs of Monte Carlo photons. Each has a corresponding luminosity, frequency, emission source (see Section 6.2.3), and positions of initial emission and last scattering. We save two distinct sets of photons:

- **“intrinsic”**: $\text{Ly}\alpha$ photons as directly emitted from gas cells, neglecting any subsequent interactions with gas.
- **“processed”**: $\text{Ly}\alpha$ photons, emitted and propagated through the surrounding gas. This includes scatterings, IGM attenuation and potentially dust destruction.

Observed $\text{Ly}\alpha$ light corresponds to the ‘processed’ photons only. Comparing results for intrinsic and processed photons allows us to quantify the redistribution of photons since their initial emission and the overall importance of $\text{Ly}\alpha$ RT in our astrophysical setting.

Both intrinsic and processed photons can be filtered based on the originating Voronoi cell, which is recorded by the intrinsic photons, and inherited by any peeling-off contribution. Using this, we can classify the emission origin of each photon according to four distinct categories:

- **“central galaxy”**: The photon originates in the central subhalo (i.e., galaxy) of the targeted halo.
- **“outer halo”**: The photon originates within the targeted halo, but outside of the central galaxy.
- **“IGM”**: The photon originates in a region not associated with any halo.
- **“other halo”**: The photon originates from a region associated with a halo that is different from the targeted.

The definition of those categories relies on the halo and subhalo catalogues provided by IllustrisTNG, where halos and subhalos are defined via the Friends-of-Friends and Subfind algorithm respectively (see Nelson *et al.* (2019b)). Each photon falls into exactly one of these categories.

First, we compute 2D surface brightness maps for all galaxies centered on the host halo position, using a pixel size of 0.8 pkpc. We include all scattered (i.e. processed) photons irrespective of their origin, as would be seen observationally. As a result these maps include emission from diffuse gas around the halo and even emission from other nearby galaxies and halos. For the projection depth we include all scattered photons reaching the observer from within ± 100 pkpc around the galaxy along the line of sight. By adopting this simple prescription we effectively ignore Ly α frequency information (diffusion). Observational studies using integral field spectroscopy commonly adopt an adaptive wavelength window to incorporate all Ly α flux of the source based on the sources varying spectral width. In our approach, we can waive the need for such adaptive algorithm. In the supplementary material (Section 6.6.1), we show the impact when incorporating spectral information and adopting a fixed wavelength window around each emitter. Our quantitative results on LAH sizes and qualitative behaviour at large radii remains unchanged irrespective of the chosen method of depth integration.

Unless otherwise stated, we always apply a Gaussian point spread function (PSF) with a FWHM of 0.7 arcsec. This PSF roughly corresponds to that of MUSE UDF data we compare in the later part of the chapter (see Section 6.3.2).

6.2.5 LAH sample and reduced statistics

In our analysis we focus on halos with galaxy stellar masses of $8.0 < \log(M_*/M_\odot) < 10.5$. For those halos, one dimensional radial profiles are computed by averaging the pixel values for a given radial bin. We characterize these radial LAH profiles with a number of ‘reduced statistics’: two measurements of Lyman- α halo size, the half-light radius $r_{1/2}$ and the exponential scale length r_0 , and the ‘central’ surface brightness value SB_0 , which we take as the value of the surface brightness map pixel(s) closest to the halo center, after smoothing by the PSF.

The Ly α half-light radius $r_{1/2}$ is computed from the one-dimensional radial profile as the radius enclosing half of the total surface brightness contained within 50 pkpc. Because $SB(r)$ does not vanish at large distances due to contributions from other halos and diffuse gas, this measure depends on this chosen outer radius.

We also fit the one-dimensional radial profile with a single exponential $SB(r) = SB_{0,\text{fit}} \exp(-r/r_0)$ (as in Cai *et al.*, 2019). The two parameters are the normalization $SB_{0,\text{fit}}$

and the scale length r_0 . We fit the simulated profiles between 0.4 and 2 arcsec, taking a finite lower limit to exclude the impact of the PSF (as is common observationally; Wu *et al.*, 2020) and the upper limit where the profile commonly transitions from an exponential to more flattened shape. Some observational studies fit a sum of two exponentials (Leclercq *et al.*, 2017; Wu *et al.*, 2020), but this method sets one exponential scale length to that of the UV light, adding additional modeling uncertainties which we avoid with the simpler size measure. For comparison with observations, we impose a simple signal-to-noise criterion, considering only data points with $S/N \geq 5$. We derive the noise from a Gaussian standard deviation of 2×10^{-19} erg/s/cm²/arcsec² per pixel of the 2D SB maps. While definitions of LAH sizes vary, we impose the same fitting routine to all compared simulated and observed radial profiles for a fair comparison. Due to the variety of radial shapes for individual LAH, the exponential function can be a bad fit at times. We exclude such cases (<10%) by imposing a maximum relative error of 10% for either fit parameter as given from the estimated covariance matrix of the least square fit. We commonly specify the scale length of a given sample in the form of $\text{median}_{\text{low}}^{\text{high}}$ where "low" and "high" are the 16th and 84th percentile.

6.3 Results

We begin in Figure 6.1 with a visual overview of Ly α emission from a large cosmological region encompassing thousands of individual emitters. Here we show the large-scale structure of the entire TNG50 simulation at $z = 3$, with scattered Ly α photons illuminating not only LAHs, but also the cosmic web within which they reside. The zoom-in panels show two individual LAHs and their substructure (i.e. satellite galaxies), on the scale of the halo virial radius, for moderate halo masses of $M_{200} \simeq 5 \times 10^{10} M_{\odot}$ (top) and $M_{200} \simeq 1.2 \times 10^{11} M_{\odot}$ (bottom), where Ly α photons are predominantly emitted at the sites of star-formation in the central galaxy. These photons then resonantly scatter to illuminate extended gaseous halos of the circumgalactic medium (CGM), where the complex dynamics of high-redshift inflows mix with feedback-driven outflows.

In Figure 6.2 we show surface brightness maps for a collection of nine LAHs ordered by star-formation rate at $z = 3$, including the two LAHs from Figure 6.1. The three colored contours all trace a surface brightness of 10^{-19} erg/s/cm²/arcsec², differentiating between the observable flux (i.e. from processed photons; blue), intrinsic emission (red), and intrinsic emission due to star-formation alone (yellow).

Broadly, the Ly α scattering process increases the apparent sizes of LAHs beyond that of intrinsic emission. In some instances, such as the lower right panel, red contoured

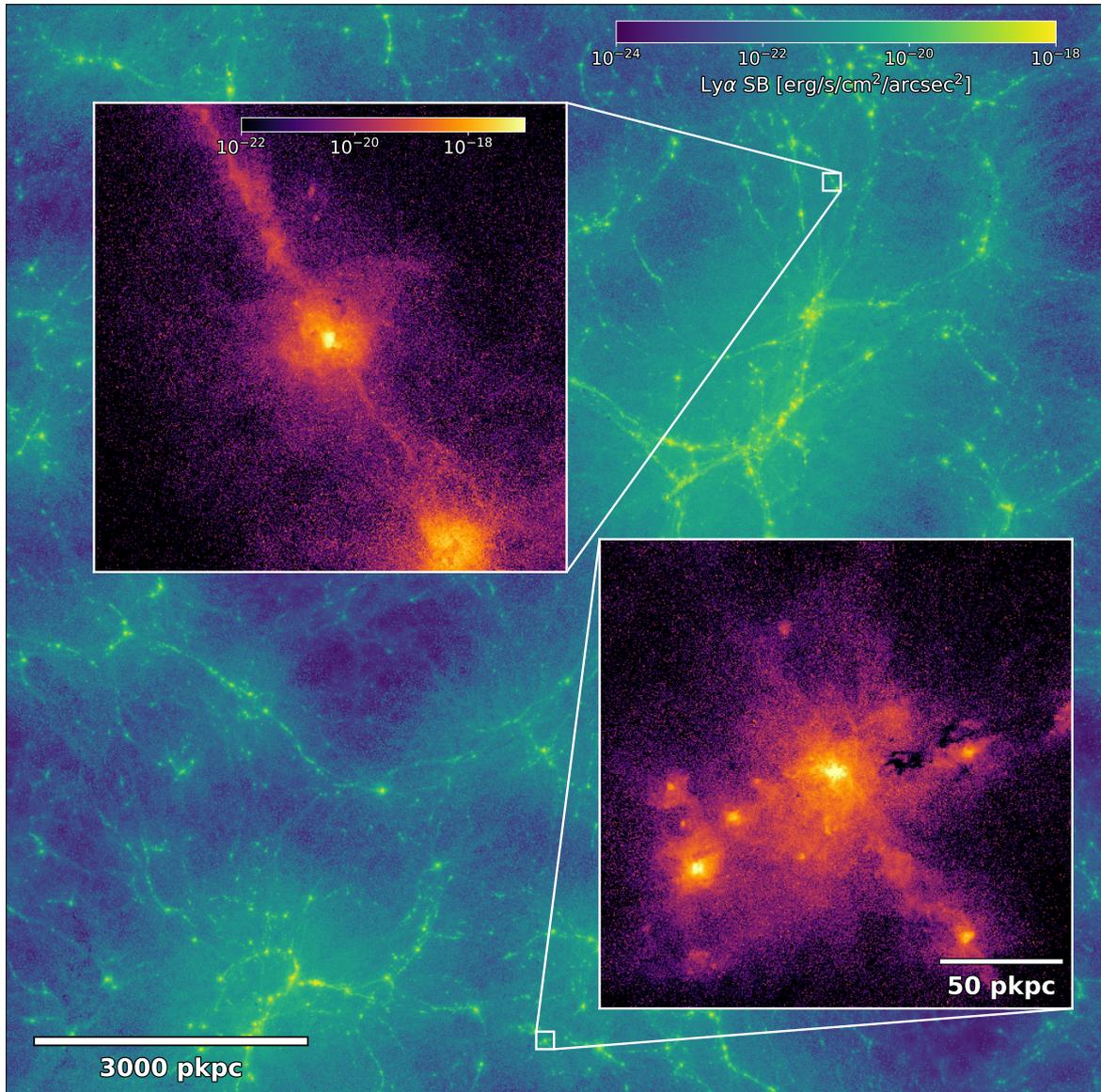


Fig. 6.1 Ly α surface brightness map for the entire TNG50 cosmological simulation at $z = 3$, highlighting the large-scale structure of the cosmic web as seen in Ly α emission. We impose a Gaussian point spread function (PSF) with a FWHM of 0.7 arcsec at a binning resolution Δ_{res} of 8.5 ckpc/h and project through a slice depth of 5.25 cMpc/h. The inset panels (no PSF, $\Delta_{\text{res}} = 0.4$ pkpc) show two individual Lyman- α halos, on the scale of the halo virial radii, for moderate mass objects: $M_{200} \simeq 5 \cdot 10^{10} M_{\odot}$ and $M_{200} \simeq 1.2 \cdot 10^{11} M_{\odot}$ (top and bottom, respectively). Ly α photons are predominantly emitted in the star-forming regions of the central galaxies, resonantly scatter and illuminate the more extended gaseous halos, including filamentary inflows. The more massive halo (lower right) has a number of star-forming satellites which also contribute Ly α emissivity and boost the local surface brightness.

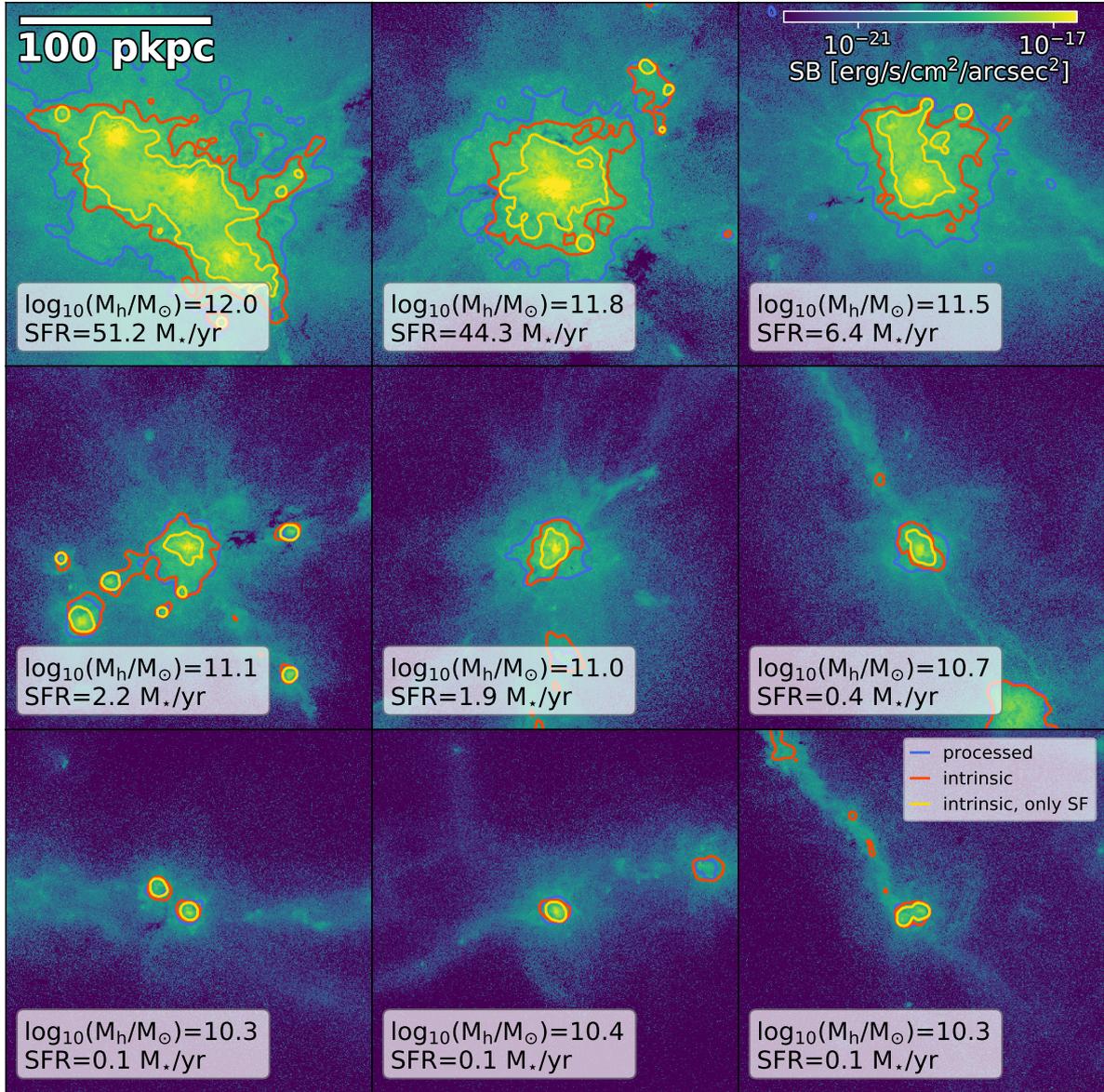


Fig. 6.2 Two-dimensional Ly α surface brightness maps of nine Lyman- α halos at $z = 3.0$ in TNG50. The LAHs are ordered by their star-formation rate in a mass range of roughly 10^{10} - $10^{12} M_{\odot}$. Contours highlight the surface brightness value of $10^{-19} \text{ erg/s/cm}^2/\text{arcsec}^2$, showing the final observable result (blue) contrasted against intrinsic photons (red) and intrinsic photons due only to star-formation (yellow). We commonly find spatially extended intrinsic emission from star-formation in the most massive halos ($\log_{10}(M_h/M_{\odot}) \geq 11.5$), while intrinsic emission from other sources is even more extended. The scattering of Ly α photons expands the extent of high surface brightness features, particularly in the more massive halos. Here we adopt a pixel size of $\Delta_{\text{res}} = 0.4 \text{ pkpc}$ (no PSF).

regions occur without corresponding yellow contours, indicating emission without the presence of star-formation in filamentary structures. We note that the surface brightness distributions of the scattered Ly α photons are significantly smoother and isotropic than the more complex structure of the underlying density field. Qualitatively this is compatible with the low eccentricities found for LAHs in observations (Wisotzki *et al.*, 2016). At distances beyond $\gtrsim 20$ pkpc from the central galaxies, LAHs become increasingly anisotropic, due to the combination of satellite galaxies (for the more massive halos) and anisotropic gas inflows.

6.3.1 Radial profiles

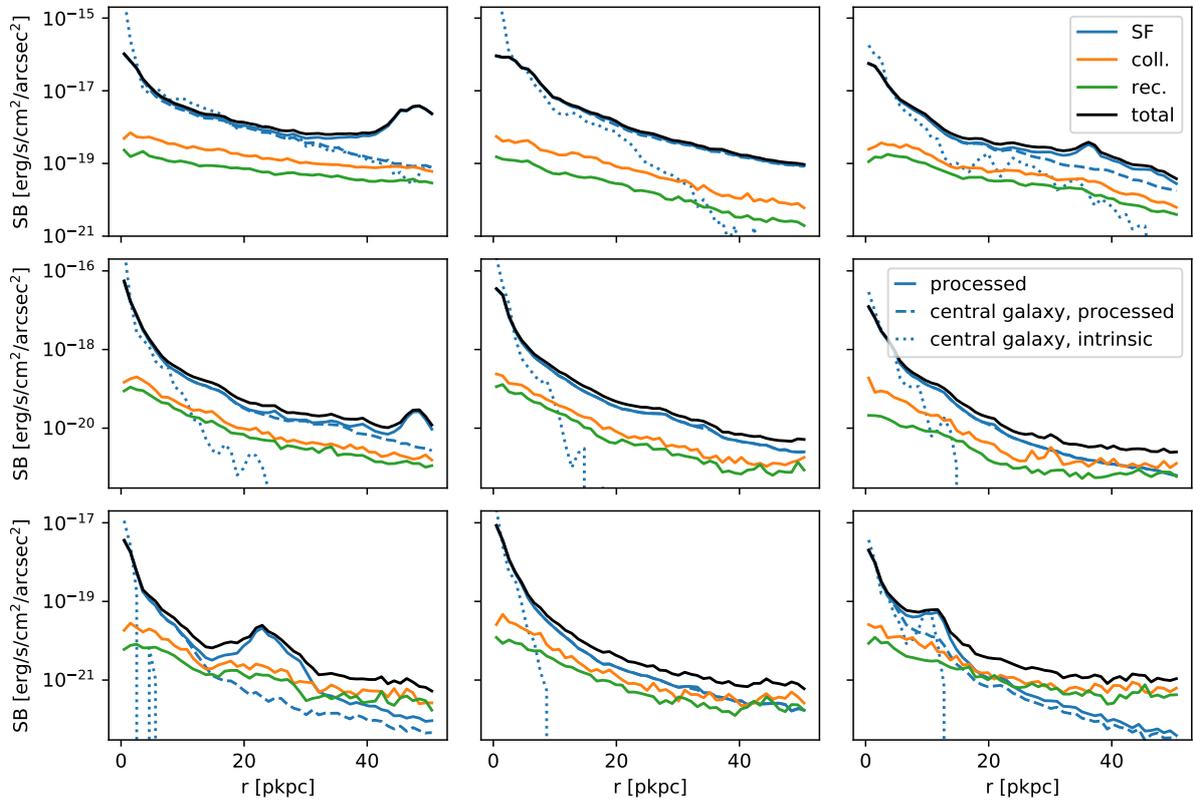


Fig. 6.3 Radial surface brightness profiles of nine individual Lyman- α halos at $z = 3$ in TNG50. The objects and plotting order are the same as in Figure 6.2. The black solid lines show the total, observable radial profiles extracted from the surface brightness maps of processed photons. We decompose these profiles into their three physical emission sources, namely star-formation (SF; blue), collisional excitations (coll.; orange) and recombinations (rec.; green). Generically, the radial profiles drop steeply within $\lesssim 5$ pkpc and flatten towards larger radii. For star-formation, we also split profiles into intrinsic (processed) photons from the central galaxy only as dotted (dashed) blue lines, emphasizing the important and non-negligible role of scattering in outer LAHs.

In Figure 6.3 we show the radial Ly α surface brightness profiles extracted from the same nine intensity maps of Figure 6.2. We decompose the total profile (including rescattering; black) into its contributions from the three emission sources: star-formation (solid blue), collisional de-excitation (orange), and recombination (green). For star-formation, we additionally show the emission neglecting subhalo/satellite contributions, both intrinsic (dotted blue) and processed (i.e. scattered, dashed blue).

In general, radial profiles are steep near the center of the halo and quickly flatten beyond $r \gtrsim 15$ pkpc. The inner regions are dominated by star-formation sourced recombination, and even at larger radii up to 50 pkpc, star-formation often remains the dominant contribution while excitations and recombinations (orange and green lines) reach similar magnitude. Except for the most massive halos in the upper row, intrinsic Ly α emission from star-forming regions (blue dotted line) quickly fades within $r \lesssim 15$ pkpc. At larger radii, the profiles are shaped by scattered photons. However, as expected, there are occasional bumps in the profile from star-formation in satellite galaxies of more massive halos.

The scattered photons contributions highlight two important effects of the radiative transfer. First, the central surface brightness is severely damped by photon rescattering (dashed blue vs. dotted blue curves). Second, large amounts of those Ly α photons are scattered further out, which provides an important contribution of star-formation to the extended radial profiles even though little in-situ star-formation may take place at those radii.

Collisional excitations and recombinations become important at radii $r \gtrsim 20$ pkpc for low mass halos, and the typical surface brightness contribution from collisional excitation exceeds that from recombination by a factor of ~ 2 . The most massive galaxies more frequently host nearby satellites, which result in the occasional bumps in the profile from star-formation in these subhalos.

We move from the case study of individual profiles to a quantitative exploration of the average predicted LAH profiles. In Figure 6.4 we stack galaxies based on stellar mass (main panel) and halo mass (lower left panel) at $z = 3$. We also show the evolution with redshift from $z = 2$ to $z = 5$ in a fixed stellar mass bin (lower right panel). Profiles always show the median stacked profile after radial binning, which we note is not the same as first median stacking the two-dimensional surface brightness maps. For the radial binning we calculate the mean surface brightness for a given annulus. Shaded regions show the central 68 percentiles. The dashed line in the main panel indicates the 0.7 arcsec FWHM Gaussian PSF we adopt, which dominates the smoothing of the radial profiles at small distances $r < 10$ pkpc. At larger radii, the surface brightness rapidly flattens, as we explore below.

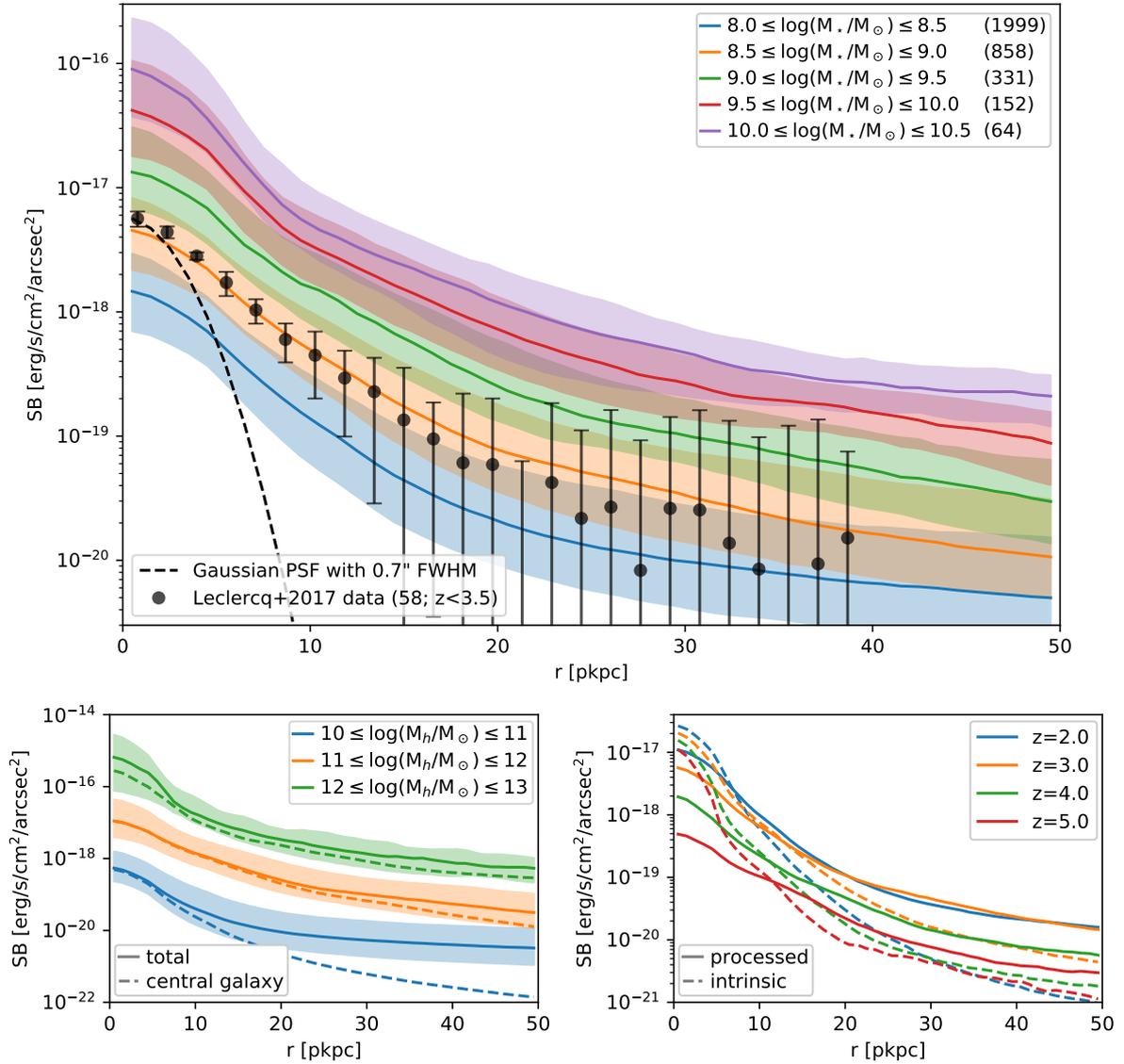


Fig. 6.4 Overview of the predicted median Ly α surface brightness profiles around TNG50 galaxies, as a function of mass and redshift. **Main panel:** Stacked radial profiles for five different stellar mass ranges from $M_* = 10^8 M_\odot$ to $10^{10.5} M_\odot$, at $z = 3$. The respective count of contributing halos to the stack is given in parentheses in the legend. At fixed cosmic time, the central Ly α surface brightness increases monotonically with stellar mass, roughly as $\text{SB}(r = 0) \propto M_*^{0.9}$. We overplot the median radial profile of 58 observed LAHs at redshifts between 2.9 and 3.5 presented in Leclercq *et al.* (2017) from the MUSE UDF (see text for details). **Lower left:** Stacked radial profiles as a function of dark matter halo mass, also at $z = 3$. The solid lines show the observed radial profiles, while the dashed lines only consider those photon contributions originating from the central halo. **Lower right:** Redshift evolution of stacked radial profiles for a fixed galaxy stellar mass bin of $8.5 \leq \log_{10}(M_*/M_\odot) \leq 9.5$. At fixed stellar mass, Ly α halos are more luminous towards lower redshift. We contrast the full radiative transfer result (RT; solid lines) with the intrinsic emission profiles (dashed lines). The scattering which occurs during the RT lowers the Ly α surface brightness at halo center ($\lesssim 5$ pkpc) while increasing it at larger radii, producing an overall flatter profile.

In the top panel, stacking surface brightness in stellar mass bins from low-mass galaxies with $M_\star = 10^8 M_\odot$ to Milky Way mass systems with $M_\star = 10^{10.5} M_\odot$, we see that $\text{Ly}\alpha$ surface brightness increases monotonically, at all radii, with increasing stellar mass. Despite this strong correlation between peak surface brightness and stellar mass, the overall shape of the median radial profiles is largely independent of the stellar mass. We note that a mass-dependent $\text{Ly}\alpha$ photon escape probability, i.e. due to dust physics, would impact the trend of overall luminosity and stellar mass, which we explore further in Section 6.3.5.

The lower left panel of Figure 6.4 shows the radial profiles in three halo mass bins. The central surface brightness of the $\text{Ly}\alpha$ profiles rises as a function of halo mass. In addition, we more clearly find a change in shape of the radial profiles as a function of halo mass, whereby flattening begins at smaller radii for lower mass halos. Considering central galaxy emission only (dashed lines), we observe that as star-formation decreases towards lower mass, the luminosity budget available for rescattering in the halo does likewise. Hence, diffuse emission outside of the central galaxy becomes dominant at smaller radii. Equivalently, since the central (or total) luminosities are lower at lower mass, external emission (i.e. from other halos) can take over more quickly.

In the lower right panel of Figure 6.4 we explore the redshift evolution of radial profiles in a fixed stellar mass bin of $8.5 \leq \log_{10}(M_\star/M_\odot) \leq 9.5$. Towards higher redshift, the central surface brightness drops significantly while changes to the overall shape are minor. Here we also show the radial profiles of the intrinsic photons (dashed lines), where $\text{Ly}\alpha$ photons are allowed to escape directly to the observer without scattering, contrasting against the processed emission (solid lines). The intrinsic profiles' lower brightnesses at higher redshifts are driven by surface brightness dimming, which is however largely countered by the higher specific star formation rates at fixed stellar mass. The intrinsic central SB luminosities decrease much more slowly towards higher redshift, implying that the photon redistribution due to resonant scattering is significantly more important at higher redshifts.

6.3.2 Comparison of TNG50 and MUSE data

In the top panel of Figure 6.4 we also overplot the median radial profile from the observational LAH dataset of Leclercq *et al.* (2017) for these objects with redshifts between 2.9 and 3.5. The data set is based on the MUSE Ultra Deep Field (UDF), which finds extended $\text{Ly}\alpha$ emission around 145 of 184 star-forming galaxies at $3 \lesssim z \lesssim 6$ with a median of $z \sim 3.7$. The observed galaxies extend in stellar mass down to $\sim 10^7 M_\odot$ with an average stellar mass of $M_\star \sim 10^{8.5} M_\odot$ (Boogaard *et al.*, 2018). As a result, the most

appropriate comparison is against the orange line, where we find an excellent agreement of the normalization and radial shape of the surface brightness profile between MUSE and TNG50. At larger radii $r > 20$ pkpc the observed stacked profile becomes uncertain given the large errorbars, so we cannot ascertain whether or not the strong flattening we observe in the TNG50 LAHs is also seen in the data. Additionally, this flattening is affected by MUSE's more extended Moffat PSF (Bacon *et al.*, 2017). The agreement of stacked radial profiles between observed and simulated samples degrades toward higher redshifts, particularly $z = 5.0$. This discrepancy is largely driven by the flattening of the simulated profiles at high redshifts (see bottom right panel) that is not found in the observed sample.

We note that quantitatively comparing TNG50 to Lyman- α halo observations has many challenges and subtleties. With respect to the MUSE data of Leclercq *et al.* (2017) in particular, although we address the issue of spatial smoothing and the PSF, remaining systematics could include details of (i) spectral smoothing, resolution, and the construction of the surface brightness map with a frequency-space integration; (ii) noise, including relevant surface brightness limits and sky backgrounds; (iii) the position chosen as the center of the halo; (iv) the derivation of galaxy properties, including mass (where SED fitting based on HST broad-band photometry lacks near-IR points, and neglecting emission lines including Ly α which can fall into the F606W filter (Feltre *et al.*, 2020); and (v) sample selection and selection biases, i.e. choosing appropriate analogs for comparison to the observed halos and/or matching the observed galaxy population in general.

Beyond the stacked profile comparison, we also contrast individual LAHs as observed in the MUSE UDF to those from TNG50. Figure 6.5 shows the 24 radial profiles from Leclercq *et al.* (2017) closest in redshift to $z = 3$ (blue data points with errorbars). For each, we search for the best 'match' from among our catalog of simulated LAHs at that redshift, and select the single LAH with the minimum least-squared difference² which are overplotted (orange).

In general, we are able to find excellent matches to the observed data, demonstrating that TNG50 can reproduce the observed diversity and variety of observed LAH profile shapes. For instance, we show good matches for more compact and more extended objects (e.g. id 2178, id 6521), and although noise starts to dominate at larger radii, we also find good matches for candidates with very flattened profiles (e.g. id 1226). The stellar masses of the observed galaxies, and the stellar mass of the selected TNG50 analog, are both shown in the legend for reference. For the subsample of observed LAHs for which we have a stellar mass estimate (Feltre *et al.*, 2020, from), we find a mean difference in

²Note that the low surface brightness measurements at large radii $r > 2''$, have only small weights in the linear least-square fit.

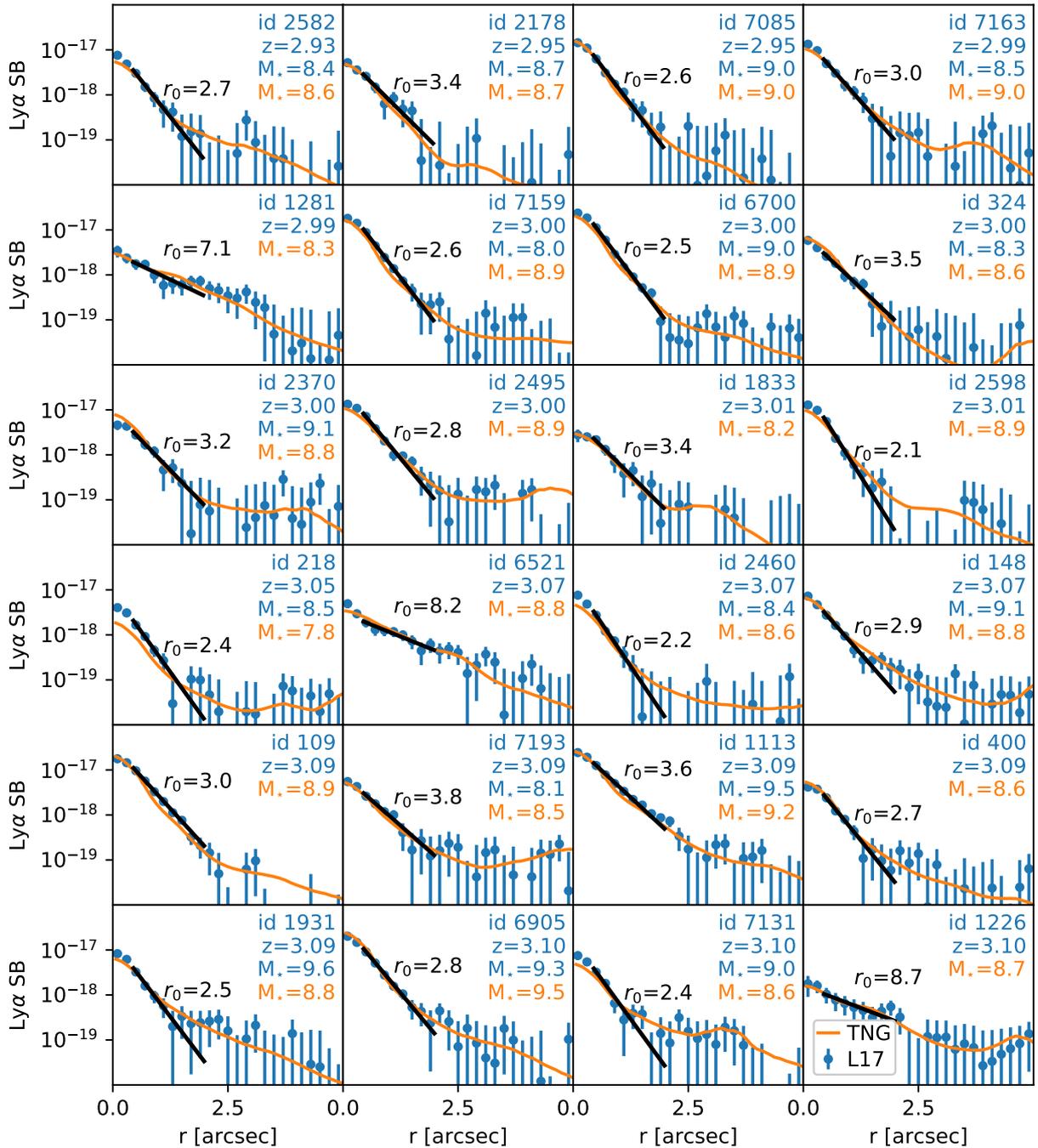


Fig. 6.5 Gallery of observed Lyman- α halos from from the MUSE UDF (blue points with uncertainties; Leclercq *et al.*, 2017), chosen as the 24 closest to redshift $z = 3$. Every observed LAH is matched to a simulated halo from TNG50 by choosing the best least-squares fit profile. All simulated radial profiles are smoothed with a Gaussian 0.7 arcsec FWHM PSF, and overplotted (orange lines). This demonstrates that the simulation has the diversity and sample statistics to recover excellent matches to all observed halos, and that TNG50 can successfully reproduce every observed profile, in both normalization and shape, with at least some simulated halo. For each halo, we include the ID and redshift of the observed MUSE object, as well as its stellar mass estimate if available. We also show the exponential scale length r_0 (in pkpc) fitted to the MUSE data. For the simulated profile we also include its galaxy stellar mass: observed LAHs and their simulated matches from TNG50 surround galaxies of similar mass.

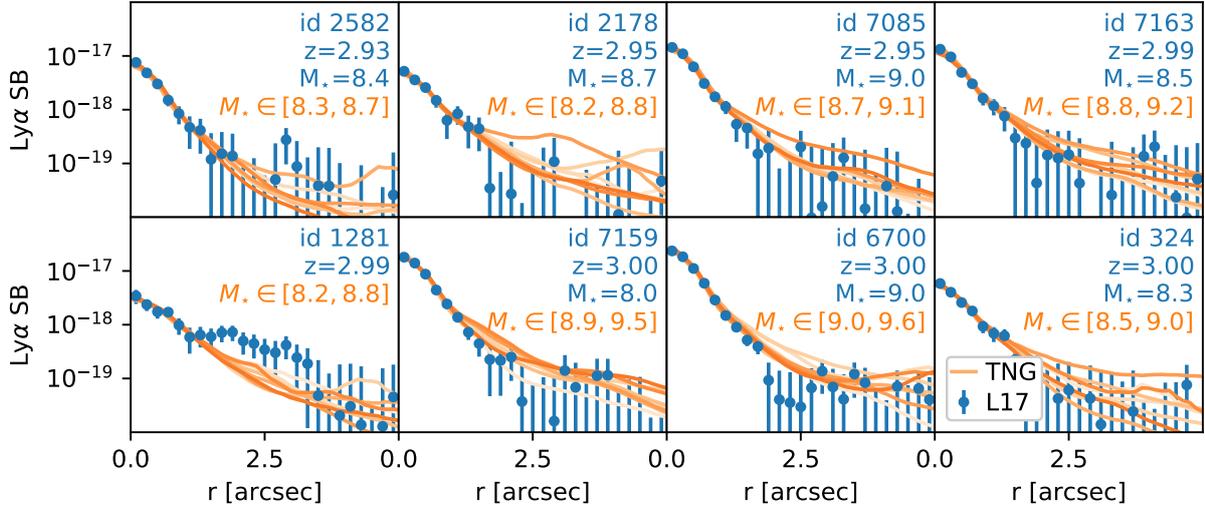


Fig. 6.6 Gallery of observed Lyman- α halos from the MUSE UDF (blue points with uncertainties; Leclercq *et al.*, 2017), where here we show the first 8 LAHs shown from Figure 6.5. As before, we search for the best matching LAH from TNG50, but now only fit data points for small radii $r < 2$ arcsec, and show the best five fits for each profile. As before, we contrast the MUSE stellar mass (if available) with the stellar mass range of the five simulated profiles, which are shaded from light to dark orange with increasing stellar mass. We find that although the simulated profiles show object-to-object variances at large distance, in most cases they are consistent with the observed profile within its errorbars.

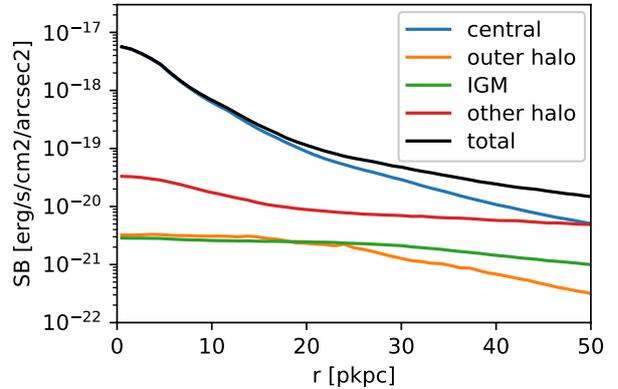
$M_{*,\text{MUSE}}/M_{*,\text{TNG}}$ of 0.11 dex with a standard deviation of 0.43 dex. This indicates that the simulated LAHs selected as good matches surround galaxies with comparable stellar masses as the observed systems.

Figure 6.6 shows a similar comparison of matched LAH profiles between the MUSE data and TNG50. However, we now restrict the least-square fit to those values at distances below $r \leq 2$ arcsec, and include the five matches in each case to highlight the range of predicted large radii behaviour from simulated profiles. Most of those fits are compatible with the observed radial profiles and their error-bars at large radii (see also Song *et al.*, 2020, for fits to L17 SB(r) profiles using 3D Ly α RT coupled to idealized shell models rather than cosmological simulations). A notable exception is the profile of MUSE-id 1281 which has an excess at $r \sim 2$ arcsec, possibly due to the existence of a satellite galaxy.

6.3.3 The origin and source of Lyman-alpha halo photons

Although LAHs are observed localized around galaxies and their dark matter halos, the photons which contribute to that emission can arise from a number of different origins. In Figure 6.7 we show the relative contribution to a stacked Ly α surface brightness profile, depending on the origin of emission, for galaxies in the stellar mass bin $8.5 \leq$

Fig. 6.7 Median stacked radial Ly α profile for galaxies with stellar masses $8.5 \leq \log_{10}(M_*/M_\odot) \leq 9.5$ at $z = 3$ in TNG50. We decompose this profile into the photons with differing emission origins: from the central galaxy (blue), outer halo (orange), the IGM (green), and other halos (red). At $r < 20$ pkpc emission from the central subhalo dominates, beyond which contributions from other halos start contributing to the overall shape. Beyond 40 pkpc the ‘other halo’ origin is critical and produces the flattening of the profiles towards large radii. The contributions originating in the outer parts of the halo and the IGM are negligible.



$\log_{10}(M_*/M_\odot) \leq 9.5$. We categorize the emission origin of each photon as one of the four categories previously introduced in Section 6.2.4.

We find that emission from the central galaxy (blue) clearly dominates the radial profile below $r \lesssim 20$ pkpc after which emission originating from other halos but scattering onto the targeted halo becomes increasingly important. Above 40 pkpc this ‘other halo’ origin (red) even dominates the radial profiles. The radial profile from rescattered photons originating in other halos has a very shallow slope, thus leading in large part to the flattening of the overall profiles at larger radii. In particular, we find that the contribution from other halos is significantly boosted if more massive halos are nearby, an effect we explore more in Section 6.4.1.

Emission originating in the IGM (green) and outside of the central subhalo (particularly from satellites; orange) is generally negligible, and never contributes more than a few percent to the total stacked profile. However, there can be infrequent radial profiles of individual halos with larger contributions from IGM and satellites than for the shown average. We note that the IGM contribution in particular will depend upon the line-of-sight integration depth.

For higher mass halos (not shown) we find that the emission from the central subhalo grows more rapidly than any of the other origins, pushing the observed flattening to larger radii. Analogously, lower mass halos flatten at smaller radii. Other halos start to significantly contribute ($\geq 10\%$ to the total stacked profile) at 7, 22 and ≥ 50 pkpc³ for the respective 1 dex stellar mass binned halos starting at 7.5, 8.5 and 9.5 $\log(M_*/M_\odot)$. There is very little redshift evolution for these radii from $z = 2$ to $z = 5$.

³In the highest mass bin, the radius lies outside of the 50 pkpc radius aperture.

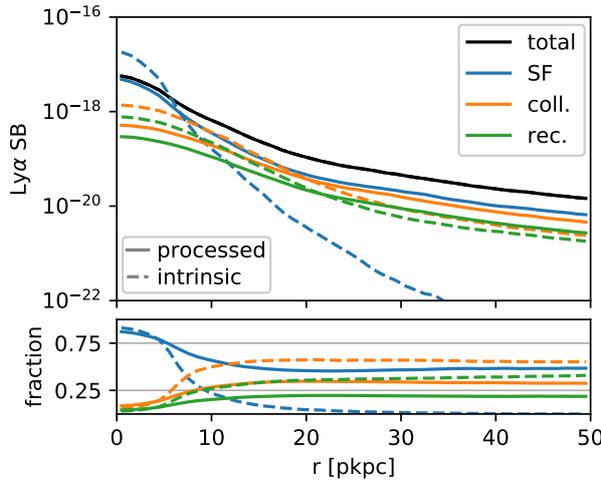


Fig. 6.8 Median stacked Ly α surface brightness radial profiles at $z = 3$ for halos with $8.5 \leq \log_{10}(M_*/M_\odot) \leq 9.5$ decomposed into different emission sources (upper panel), and the relative fraction of each (lower panel) in TNG50. Dashed lines show intrinsic emission, while solid lines show the processed (i.e. scattered) signal. Emission from star-forming regions typically dominates the intrinsic emission up to 10 pkpc after which collisional excitations start to dominate. However, radiative transfer redistributes this central emission towards the halo outskirts, such that star-formation remains the dominant source of Ly α emission for observed LAHs at all radii shown.

In Figure 6.8 we similarly decompose the stacked profile into the relative contributions of different emission *sources*: star-formation sourced rescattered photons (blue), collisional excitation (orange), and recombination (green). In addition to the ‘processed’ signal (solid lines) we also show the intrinsic emission signal (i.e. ignoring scattering effects; dashed lines). In both cases we find that star-formation makes up the bulk of the SB within the central 10 pkpc. At larger radii however rescattered photons from star-forming regions drop to a $\sim 50\%$ relative contribution as diffuse collisional excitations and recombinations rise to $\sim 30\%$ and $\sim 20\%$ respectively. These relative fractions, shown here at $z = 3$, are similar at other redshifts (not shown).

6.3.4 Lyman-alpha halo sizes

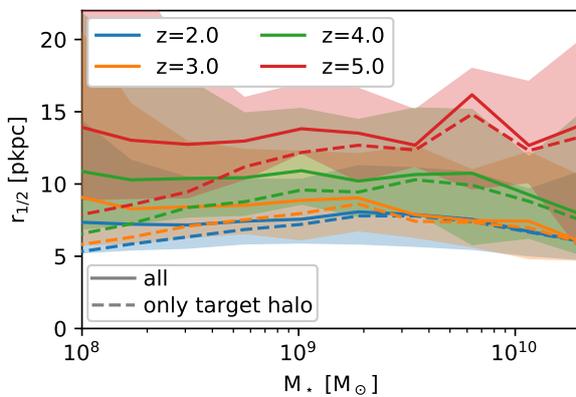


Fig. 6.9 The Ly α half-light radius $r_{1/2}$ as a function of stellar mass over redshift in TNG50. The colored solid lines show the median radius and the shaded regions enclose the radii within the central 68 percentiles. The typical $r_{1/2}$ size lies between 5 and 15 pkpc, and monotonically increases with redshift, while being mostly constant over the stellar mass range. For the colored dashed lines, contributions originating outside of the targeted halo have been ignored. The difference between solid and dashed lines therefore indicates an increasing impact of unbound and other halos’ contributions at low stellar masses.

To study the dependence of LAH profiles on galaxy/halo properties and redshift, we summarize each profile by a characteristic size and surface brightness, as defined in Section 6.2.5. Figure 6.9 shows the Ly α half-light radius $r_{1/2}$ as a function of stellar mass across the studied redshift range. We find that the half-light radii of our LAHs is typically between 5 and 15 pkpc. There is little correlation with stellar mass compared with the scatter. Half-light radii are systematically larger in physical kpc towards higher redshift, roughly doubling in size from $z = 2$ to $z = 5$ across the redshift range. This implies a stronger redistribution through scatterings into the outskirts of halos at higher redshifts, which is largely driven by a higher neutral hydrogen density at those redshifts.

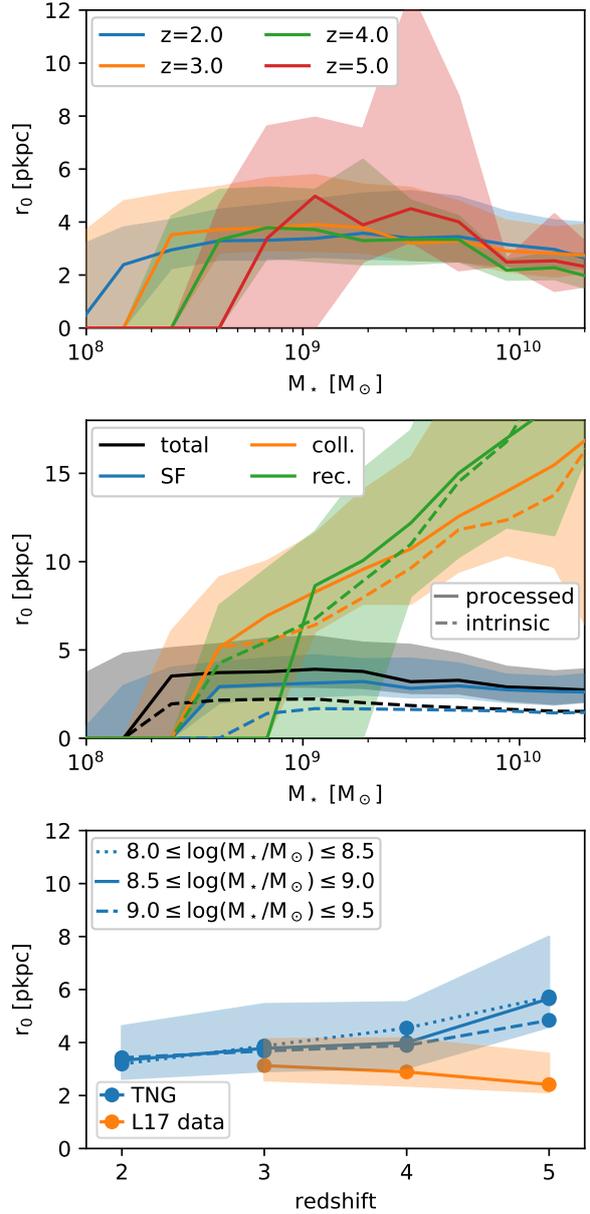
In Figure 6.10 we show the fitted single exponential scale length r_0 . In all panels we impose the noise modeling described in Section 6.2.5, which effectively imposes a sensitivity limit and induces a lower cut-off mass below which modelled LAHs show no *observable* extent. This limit starts to affect the median at a stellar mass below $\sim 2 \times 10^8 M_\odot$ ($\sim 10^9 M_\odot$) at $z = 2$ ($z = 5$), causing the artificial drop-offs at low M_* .

Except for this sensitivity limit, there are no clear trends of r_0 with stellar mass. Similarly, no clear redshift evolution is evident (upper panel). However, when looking at our fiducial stellar mass range of $8.5 \leq \log_{10}(M_*/M_\odot) \leq 9.5$ only, we do find that r_0 increases with redshift from $3.4_{-0.7}^{+1.3}$ pkpc at $z = 2$ to $3.7_{-0.9}^{+1.7}$ pkpc ($4.0_{-1.0}^{+1.7}$ pkpc, $5.5_{-1.6}^{+2.4}$ pkpc) at $z = 3$ ($z = 4$, $z = 5$).

The middle panel of Figure 6.10 shows the scale radius as a function of stellar mass at $z = 3$. Here we decompose the contribution to LAH size by emission source, by determining r_0 based on each of three respective emission sources alone (colored lines). The black solid line shows the median r_0 relation combining all three emission sources, as would be observable. We find that the overall scale radius r_0 is largely determined by the emission from star-formation due to its high surface brightness. Scale lengths from star-formation show no correlation with stellar mass, while both collisions and recombinations do show a strong positive correlation. Thus, r_0 follows the lack of evolution with mass seen in the star-formation source. In contrast, both diffuse collisions and recombinations show a strong positive correlation with mass. For halos between $8.5 \leq \log_{10}(M_*/M_\odot) \leq 9.5$ we find r_0 to be $3.7_{-0.9}^{+1.7}$ pkpc for the overall profile and $3.1_{-0.7}^{+1.4}$ pkpc, $8.0_{-1.7}^{+2.8}$ pkpc or $9.2_{-2.0}^{+3.2}$ pkpc for SFR, excitations, recombinations only respectively.

Solid lines indicate processed photons, while the dashed lines show r_0 based on the intrinsic photons only. Not surprisingly, intrinsic photons typically give rise to a smaller scale length than processed photons. Most importantly, intrinsic photons from star-formation give rise to a scale length close to a point-like source convolved with the PSF, while this value doubles from rescattering in the CGM.

Fig. 6.10 The Ly α single exponential radius r_0 as a function of stellar mass and redshift. The colored solid lines show medians, while shaded regions indicate the central 68 percentiles. **Top:** r_0 as a function of stellar mass, at four distinct redshifts. There is no clear trend of r_0 with either redshift or stellar mass. **Middle:** Decomposition of r_0 at $z = 3.0$ (black) into its three emission origins (SF, coll, rec), for both intrinsic and processed photons. As the luminosity budget in the proximity of the halos' center is dominated by star-formation, the latter effectively sets the scale length. **Bottom:** r_0 versus redshift compared to observational data from the MUSE UDF (Leclercq *et al.*, 2017). For simulations and observations, we fit the scale length r_0 using the same procedure. We show three different bins of fixed stellar mass for TNG50 (blue lines). At fixed stellar mass, LAH sizes are overall larger towards higher redshift. No clear redshift trend is evident in the observations which are consistent with no size evolution, although the galaxy stellar mass distribution as a function of redshift in the data is uncontrolled (see text for details).



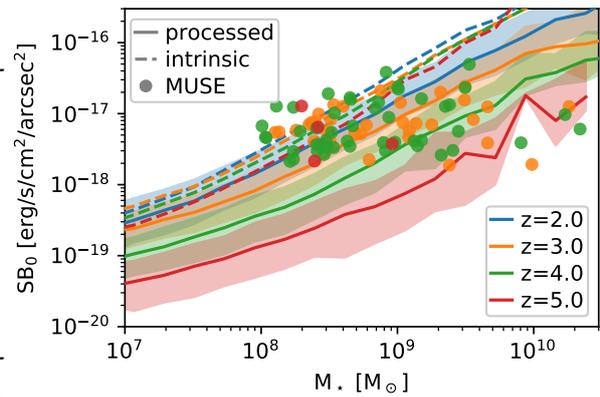
In the bottom panel of Figure 6.10 we show the explicit redshift evolution of LAH sizes, from $z = 2$ to $z = 5$, and compare the TNG50 result with that of Leclercq *et al.* (2017). For the comparison with MUSE, we show simulated halos in stellar mass bins $\log_{10}(M_*/M_\odot)$ from 8.0 to 9.5, which overlap the observed distribution. At fixed stellar mass, we find that the median r_0 increases towards higher redshift and the magnitude of this trend is similar to the scatter of r_0 at each redshift. For the MUSE UDF observations, we find $3.1^{+1.0}_{-0.6}$ pkpc, $2.9^{+1.3}_{-0.5}$ pkpc and $2.4^{+1.2}_{-0.3}$ pkpc for the redshift bins centered at $z = 3$, $z = 4$ and $z = 5$, respectively.

Our face value comparison with the MUSE data implies that LAHs in TNG50 are slightly more spatially extended than in observations, by roughly a factor of ~ 1.2 at redshift three, increasing to ~ 2.3 by redshift five with a simulated $r_0 \sim 5.5$ pkpc, the two size distributions being roughly one standard deviation apart. We caution, however, that the stellar mass distribution of the observed galaxies is not fixed as a function of redshift. More massive galaxies are more easily observed at higher redshift, and ideally we would match the joint (M_*, z) distribution to make this comparison. However, remaining methodological differences likely still dominate the uncertainty in this comparison, as sizes are not measured in exactly the same way in both the observational data and TNG50 simulation.

We note that observed sources of Ly α emission have been characterized by many different sizes ranging from roughly 2 pkpc to 9 pkpc at $z \sim 3$ (Bond *et al.*, 2009; Momose *et al.*, 2014). This large range of radii hints at the large diversity of different galaxy selection (LBGs/LAEs), functional fitting forms (single/double exponential) and methodologies (individual/stacking). We leave a quantitative comparison of Ly α sizes for future work, and do not try to explicitly compare LAE sizes to the observational literature here.

6.3.5 LAH central brightness

Fig. 6.11 The central surface brightness SB_0 as a function of stellar mass across the studied redshift range in TNG50. The solid lines show the median for the processed photons for a given mass bin and the shaded region shows the corresponding 68 central percentiles. In dashed, we show the median for the intrinsic photons. The dots show measured MUSE results color coded by simulated redshift they are closest to. The central surface brightness is a strong function of galaxy mass both for intrinsic and processed photons in our TNG50 simulation. There also is significant redshift evolution of SB_0 for the processed photons, roughly scaling with $(1+z)^4$. In the MUSE data, we find no significant correlation, neither with stellar mass nor redshift.



In addition to the two LAH size measures $r_{1/2}$ and r_0 we also calculate the central surface brightness SB_0 . Figure 6.11 shows the trend between SB_0 and galaxy stellar mass, from $z = 2$ to $z = 5$ (different line colors). We derive this value for both the intrinsic (dashed lines) and processed photons (solid lines). For the latter, we shade the central

68% of outcomes in a given mass bin. Additionally, we show the results from the MUSE UDF data set (circles), colored to match the nearest simulated redshift for comparison.

First, we see a strong correlation of the peak LAH brightness with galaxy stellar mass. This trend was previously noted in Figure 6.4 for the simulated halos. Over four orders of magnitude in stellar mass, the central surface brightness value increases by roughly two orders of magnitude. The evolution of SB_0 as a function of redshift is minimal for intrinsic photons. While the intrinsic surface brightnesses are subject to cosmological dimming, this seems to be countered by the increased specific star formation rate for halos at higher redshifts. For processed photons, CGM diffusion and IGM attenuation suppress SB_0 with a scaling of roughly $(1+z)^4$.

In comparison to the clear correlation in the simulations, the MUSE data does not show such a strong relationship between central surface brightness and either stellar mass or redshift. Although the SB_0 values are in reasonable agreement where the bulk of the observed systems reside, $10^8 < M_*/M_\odot < 10^9$, the flat trend in the data leads to lower inferred values at higher stellar masses, when compared to those in TNG50. This is certainly caused in part by systematic uncertainties in the observational determinations of stellar mass, which act to wash out this trend. Even more importantly, we speculate that this difference arises due to our omission of a model for unresolved dust attenuation and stochasticity on the smallest scales, as we discuss below.

6.4 Discussion

6.4.1 The shape and nature of Lyman-alpha halos

Based on our results we now discuss the implications for the shape and nature of Lyman- α halos. In particular, is there a typical shape for LAH radial profiles, is there a common cause for this shape, and what is the resulting interpretation of the observations? In Section 6.4.1, we focus on ‘small’ scales, of order of ~ 10 pkpc, where LAHs are detected around star-forming galaxies. In Section 6.4.1 we discuss larger scales, and profile flattening, as accessible in current and future intensity mapping studies.

Ly α profiles at small scales

We have shown that the radial profiles of our simulated LAHs are primarily shaped through rescattered photons which originate from star-forming regions in the central galaxy of a halo on scales around ~ 10 pkpc. This rescattering gives rise to a smoothing of the surface brightness maps and radial profiles that is larger than typical PSFs (0.7 arcsec

adopted herein). Beyond this distance median radial profiles tend to steepen rapidly before flattening on scales above ~ 20 pkpc.

We find typical exponential scale lengths r_0 of ~ 4 pkpc with little to no correlation with stellar mass. As radial profiles are largely dominated by emission from star-formation on these scales, the typical shape of rescattered photons from the central galaxy sets this typical r_0 and leads to the lack of correlation with stellar mass. If diffuse emission through collisions and recombinations were the dominant source of LAH photons, we would infer much larger scale radii r_0 and a strong correlation with stellar mass with scale lengths starting at ~ 5 pkpc growing to ~ 15 pkpc between 10^8 and $10^{10} M_\odot$. Thus, r_0 and its mass dependency can serve as a discriminator between rescattered photons from star-forming regions and diffuse emission in observations. This holds even if the relative importance of the different emission mechanisms is not precisely correct in our simulations.

An important modification of the emission mechanism stems from local ionizing sources. TNG's simplified on-the-fly treatment of radiation stemming from AGN is already reflected in our results. In particular, in our fiducial sample of LAH candidates with stellar masses of $8.5 \leq \log_{10}(M_*/M_\odot) \leq 9.5$ 1030 out of 1189 halos host a SMBH and incorporate their ionizing flux. While, we expect that additional ionizing flux from stellar populations could additionally boost diffuse emission, we only find a small impact of local ionizing radiation from AGN on the radial profiles on small scales measured by the exponential scale radius r_0 . As the AGN UV radiation (when present) dominates over that of the stellar population, we similarly do not expect r_0 results to significantly change if latter radiation was incorporated.

Interestingly, as the scale length r_0 keep growing toward higher masses for emission sourced by cooling and the UV background, high mass halos in TNG50 reach the lower end of extents observed in LABs. While a future dedicated study is required, this might strengthen the case of such diffuse emission sources (without local sources ionizing flux) causing observed LABs.

In comparison to data, our results for the stacked profiles are consistent with MUSE UDF observations presented in Leclercq *et al.* (2017). At the level of individual LAH profiles, we similarly find good agreement, such that there are numerous analogs in TNG50 which have compatible $\text{Ly}\alpha$ radial profiles. This agreement is noteworthy as we have no tuned or calibrated parameters in our $\text{Ly}\alpha$ modeling.

In the quantitative comparison of r_0 for individual LAHs we find up to a 20% mismatch at $z = 3$, which grows to a factor of two at $z = 5$. Similarly, the central surface brightness values from Leclercq *et al.* (2017) show significant scatter and tend to be below those obtained from TNG50 at the highest stellar masses. Despite these regimes of tension, the

observations show no clear mass or redshift evolution in either r_0 or SB_0 values, which together with the relatively small sizes implies that observed LAHs in the MUSE UDF sample are sourced by star-formation.

Although a rigorous comparison between simulated and observed LAHs is complicated by several systematic uncertainties, the SB_0 and r_0 tensions hint at possible shortcomings of our $Ly\alpha$ modeling. In Figure 6.10 we found that considering star-formation alone gives the smallest exponential scale lengths. At $z = 3$, we found $r_0 = 3.1_{-0.7}^{+1.4}$ pkpc in this case, which is very similar to the MUSE UDF estimate of $3.1_{-0.6}^{+1.0}$ pkpc. The simplest explanation is that we have underestimated the $Ly\alpha$ luminosity from star-forming regions, or overestimated the $Ly\alpha$ luminosity from other sources. Further, the strong correlation of SB_0 with stellar mass found in our models is not clearly present in the data. As we assume a fixed relation between star-formation and $Ly\alpha$ luminosity this outcome is not surprising, and can be alleviated by developing a more realistic model for the underlying relation between $Ly\alpha$ and SFR, as discussed in Section 6.4.2.

$Ly\alpha$ profiles at large scales

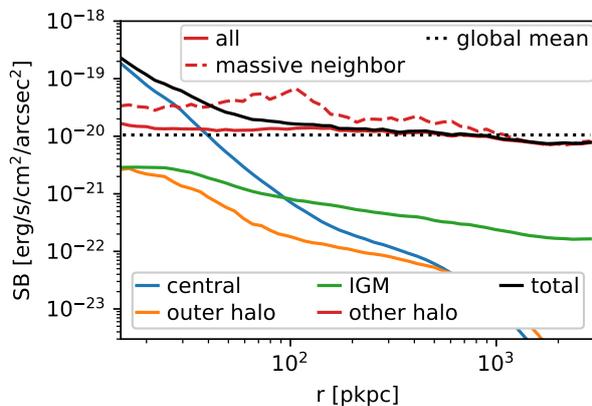


Fig. 6.12 Radial profiles for halos with $8.5 \leq \log_{10}(M_*/M_\odot) \leq 9.5$ at $z = 3$ decomposed into emission origin and extending to large-scales of 3000 pkpc radii in TNG50. Here we neglect PSF effects and integrate photon contributions from within ± 600 pkpc of each emitter along the line of sight. The overall profile (solid black) strongly flattens out as the very flat contribution from other halos starts to dominate after 40 pkpc. Contributions from the outer halo (orange) and the IGM (green) are negligible in the stacked profiles, while nearby neighbors strongly boost the signal (solid red). To emphasize this environmental effect, we also show the subsample of 127 (out of 1189) halos that have a more massive halo, by at least a factor of 10, within 0.5 pMpc (dashed red). The black dotted line shows the surface brightness based on the global mean of the luminosity density in the simulation.

Although much of our analysis has focused on the inner regions of LAH profiles, $\lesssim 20$ pkpc, we here consider the general shape of extended $Ly\alpha$ profiles where the impact of environment reveals itself.

The most prominent feature in the median stacked profiles of Figure 6.4, regardless of stellar mass, is the significant ‘flattening’ beyond the inner few 10s of pkpc. In Figure 6.12 we take advantage of our global, large-volume RT calculation and present the radial profiles of halos with $8.5 \leq \log_{10}(M_*/M_\odot) \leq 9.5$ out to 3000 pkpc. Here we neglect the PSF, and integrate through ± 600 pkpc from the emitter’s position along the line of sight, which roughly reflects the spectral resolution of the HETDEX survey.

We decompose this median surface brightness profile based on photon origin (different colored lines). For reference, the expected surface brightness given the average luminosity density in TNG50 is shown with the black dotted line. We find that the contribution from other halos (solid red line) is nearly constant with distance and reaches down to roughly match the mean global. This ‘other halo’ contribution starts to dominate the surface brightness profile beyond 100 pkpc by a factor of more than 10. To emphasize this environmental effect, we also show a stack of the subsample of halos that have a close (≤ 0.5 pMpc) massive neighboring halo, by a factor of ten or more, as the red dashed line. For those galaxies there is an elevated plateau above the global mean that only starts to drop around 100 pkpc.

This behavior is a proximity effect whereby nearby halos give rise to an effectively elevated background. As a result, emission from star-formation from nearby neighbors dominates the observed radial profiles over diffuse emission from the outer halo and the IGM. Zheng *et al.* (2011b) find a similar flattening effect from rescattered photons from other halos in their simulations, but are unable to assess the relative importance of diffuse emission, which remains untreated. In contrast, Lake *et al.*, 2015 do incorporate both central sources and diffuse emission, reporting a flattening up to scales of 80 pkpc that in equal parts stems from the two emission sources, in contrast to our results where emission from other halos dominates over IGM emission.

We note that modeling of this proximity effect remains tricky: off-scattered photons in a targeted halo originating in a neighboring halo could potentially not be modeled by a classic 2-halo term as this would merely capture the overlap between different halos’ profiles. Here – in addition to the possible overlap of profiles – we have Ly α photons freely traveling through the IGM to a neighboring halo at which point scatterings will trace out part of the targeted halo.

Our findings are qualitatively compatible with stacking results by Matsuda *et al.* (2012) showing a strong correlation of the flattening in the overdense regions. Wisotzki *et al.* (2018) find significant flattening up to scales of around ~ 50 pkpc, for which they consider UVB fluorescence as a potential source. In contrast, we find that towards large radii diffuse emission outside of halos contributes less than 10% to the surface brightness (see Figure 6.12), while we find that scattering of Ly α photons from galaxies dominate

the flattening. More recent Ly α intensity mapping results at $z = 5.7$ and 6.6 by Kakuma *et al.* (2021) found reasonable agreement with the MUSE stacking (see also Matthee *et al.*, 2020) and the simulated profiles of Zheng *et al.*, 2011b on small scales ($r < 150$ ckpc), they cannot confirm the proximity effect and flattening found here. We note a further complication, that this flattening could be removed in large part or entirely in narrow-band surveys due to the required background subtraction. However, a careful examination of upcoming data, such as from HETDEX, could reveal this proximity effect, particularly by stacking based on the presence of massive neighbor number or environmental density.

6.4.2 Current limitations and future outlook

Model limitations: TNG and its gas state

Our results depend inseparably on the outcome of the underlying cosmological simulation as well as our Ly α radiative transfer post-processing, inheriting limitations from each. The basis of our Ly α calculations is the TNG galaxy formation model (Weinberger *et al.*, 2017; Pillepich *et al.*, 2018b) and the TNG50 simulation in particular. Although the TNG model has been shown to reproduce a large variety of galaxy and CGM properties in broad agreement with observations (see Section 6.2.1), it does not treat local ionizing radiation self-consistently as done in radiation-hydrodynamical (RHD) simulations, that are now becoming increasingly feasible for individual zoom simulations (e.g. Rosdahl *et al.*, 2012; Mitchell *et al.*, 2021) as well as cosmological volumes down to redshifts $z = 5 - 6$ (Gnedin *et al.*, 2017; Rosdahl *et al.*, 2018; Ocvirk *et al.*, 2020). As a result, our knowledge of the gas ionization and temperature state is limited to the outcome of the physical modeling in our simulation, as is true in all galaxy formation simulations. As TNG50 remains unfeasible as RHD simulation, it substitutes such approach with an on-the-fly treatment of AGN as the dominant ionizing radiation source as studied in the supplementary material (Section 6.6.1). Towards higher redshifts, the relative importance of stellar ionizing sources increases, thus potentially hinting at larger uncertainties of TNG's model only incorporating AGN's radiative feedback.

While particularly emission from collisional excitation strongly depends on the temperature (Furlanetto *et al.*, 2005; Faucher-Giguère *et al.*, 2010) and subsequently can boost emission, we found that the diffuse emission remains subdominant to scattering from the central sources even if radiation from AGNs is present. If central Ly α emission from SMBHs themselves is incorporated, this finding will be further strengthened. Predicted central surface brightnesses when only considering collisional excitations from the CGM in TNG, even with AGN radiative feedback, are commonly an order of magnitude too

small compared to the MUSE observations. Apart from this normalization problem, collisional excitations from TNG's predicted temperature and ionization state show scale lengths significantly too extended for compatibility with observations (see Figure 6.10). Reconciliation with observations in case collisional excitations are the dominant source of LAHs thus requires significantly different ionization and temperature radial profiles from those found in TNG. Independent constraints on the density, ionization state and temperature from other observational probes would be useful to assess TNG's model and check resulting conclusions for predicted LAHs. In addition, these findings should be revisited with future RHD simulations that come closer TNG's sample size and studied redshift range.

In addition, the TNG50 simulation has a resolution of ~ 100 pc in the dense interstellar medium, which is furthermore modeled with an effective two-phase model. As above, this implies that a number of approximations and simplifications must be adopted in both the emission and transport of $\text{Ly}\alpha$ (see Section 6.2.3). Importantly, in the present work, we have adopted a simple mapping from the star-formation rate of a gas cell to its $\text{Ly}\alpha$ emissivity. We have also neglected dust in its entirety, as well as sub-resolution (unresolved) density structure during the radiative transfer step.

Model limitations: ISM emissivity

As discussed, we do not include the destruction of $\text{Ly}\alpha$ photons by dust in our $\text{Ly}\alpha$ radiative transfer (although we have investigated its impact, see Section 6.6.1). Such destruction would primarily take place in the ISM, where resolution is marginal for capturing the complex density and ionization structure relevant for the $\text{Ly}\alpha$ radiative transfer. If however we do not explicitly treat dust, we potentially overestimate the SF luminosities in Eqn. (6.3), where we do not explicitly model dust destruction. We review the emission model for stellar populations here, as such an overestimation would strongly affect our conclusions of central $\text{Ly}\alpha$ emission dominating observed LAH profiles through CGM scatterings.

Equation (6.3) is derived by integrating the ionizing flux from stellar population synthesis models, and a conversion of $\sim 2/3$ of ionizing photons into $\text{Ly}\alpha$ (Dijkstra, 2019). The proportionality factor between SFR and $\text{Ly}\alpha$ emission depends on a range of factors such as the stellar population synthesis model, the escape of ionizing flux, the stellar population's age and metallicity, and the initial mass function (IMF) (Furlanetto *et al.*, 2005). Overall, this relation should only be seen as a rough estimate with an uncertainty of a factor of a few. Furthermore, observations, such as from $\text{H}\alpha$, $\text{Ly}\alpha/\text{H}\alpha$ and $\text{Ly}\alpha/\text{H}\beta$

show a large scatter in the relation of hydrogen line emission to SFR (Kennicutt, 1998; Blanc *et al.*, 2011; Weiss *et al.*, 2021).

Given the large modeling uncertainties shown, we adopted this SFR-Ly α relation as a rough estimate, which could potentially be rescaled after comparison with observations. However, as we show in the following, such a calibration is not necessary for the mass range $8.5 \leq \log_{10}(M_*/M_\odot) \leq 9.5$, where we already obtain reasonable agreement with observations. This also implies that, while not used in the derivation of the relation, Ly α dust extinction is nevertheless effectively captured for the $8.5 \leq \log_{10}(M_*/M_\odot) \leq 9.5$. We note however, that at the high mass end luminosities can be significantly overestimated, and the role of dust becomes critical.

In Figure 6.11, we show the central luminosities as predicted in our simulations compared to MUSE observations in Leclercq *et al.* (2017). We find that the simulations show similar or slightly lower central surface brightnesses when compared with the observational data.

We have also compared the LAH luminosity function (LF) of our samples with observational LAE luminosity functions at $z = 2.2, 3.1$ and 5.7 (respectively based on Ouchi *et al.*, 2008; Konno *et al.*, 2016, 2018) (not shown). We calculate LAH luminosities from the photon contributions of the targeted halos falling within a 3 arcsecond aperture. We find very reasonable agreement of the observations with our intrinsic luminosity functions in the luminosity range $\sim 3 \times 10^{41}$ erg/s to $\sim 2 \times 10^{42}$ erg/s, which includes at least 68% of all halos with stellar masses $8.5 \leq \log_{10}(M_*/M_\odot) \leq 9.5$ for each redshift $z = 2$ to $z = 5$ in TNG50. Quantitatively comparing to Konno *et al.* (2016), we find that at $L = 1.4 \times 10^{42}$ erg/s, our intrinsic emission in fact underestimates the LF by just 0.16 (0.32, 0.17, 0.06) dex for $z = 2$ (3,4,5). For the processed photons, our LFs underestimate the observed LFs by 0.19 (0.10, -0.11, -0.63) dex at $z = 2$ (3,4,5). That is, our preliminary look already shows promising agreement, while a detailed analysis of the LFs remains a topic for future work.

We are thus confident that, on average, for the fiducial stellar mass range of $8.5 \leq \log_{10}(M_*/M_\odot) \leq 9.5$, the luminosities reasonably match observations. We thus adopt the simple SFR-Ly α relation without readjustment for dust attenuation. However, we overestimate the high luminosity end of the LF, and thus inaccurately capture the high-mass LAHs' shape. This might also affect the overall amplitude of the flattening from neighboring galaxies for the fiducial stellar mass range.

In observations there is a large scatter between the star-formation rate inferred from Ly α luminosity compared to UV based estimates (Runnholm *et al.*, 2020; Santos *et al.*, 2020). Typically, the median SFR estimate from Ly α exceeds that from UV measurements below $\sim 10 M_\odot/\text{yr}$, but decreases above this value. This systematic trend implies a larger

suppression of Ly α escape for emitters with higher SFR and stellar mass. In addition, there is non-negligible scatter between individual objects with similar properties. It is thought that Ly α emission traces the most recent star-formation (Santos *et al.*, 2020), which could be highly time variable, and is modified by the complex small-scale neutral hydrogen distribution and kinematics (Blanc *et al.*, 2011).

In our simulations we have instead assumed a strict proportionality between Ly α luminosity and SFR at the resolution scale, as given by Equation (6.3). We find that subsequent scattering only adds minor scatter to this correlation, implying that sub-resolution stochasticity may be required. As discussed in Section 6.4.1, we tend to slightly overestimate the r_0 metric of LAH sizes in comparison to the MUSE UDF sample, and generally find strong trends of both r_0 and SB_0 with stellar mass, which are less clear in the data. We suspect that adopting an observationally motivated scatter between Ly α luminosity and UV inferred SFR would also alleviate the tensions noted in the central surface brightness SB_0 trends between our simulations and the data. We ran a simple test to test the impact of including a naive dust treatment (Section 6.6.1), which demonstrates that the central brightness SB_0 is increasingly attenuated towards higher stellar masses, as dust counters the generally higher SFRs and thus Ly α luminosities of those halos.

While subgrid modeling and dust, along with intrinsic scatter between Ly α luminosity and SFR, should help reconcile the central surface brightness SB_0 comparison in Figure 6.11, we would additionally require the mean Ly α luminosity from star-formation to increase in order to obtain smaller LAH sizes. For the typical objects considered as LAH candidates with $8.5 \leq \log_{10}(M_*/M_\odot) \leq 9.5$ here, Santos *et al.* (2020) find $SFR_{Ly\alpha}/SFR_{UV} \sim 2$ for $SFR < 10 M_\odot/\text{yr}$ with significant scatter. Beside the simplistic model assumptions with significant uncertainties as previously discussed, we would need to incorporate Ly α emission from AGN in future work, which would boost emission and naturally induce scatter in the SFR-Ly α relation.

In conclusion, we demonstrated that the current ISM emissivities appear reasonable given the simplistic model, but future work on the effective modeling of dust and a modified SFR-Ly α relation (reflecting varying ISM environments and potential AGN presence) in both proportionality factor and its scatter might help to simultaneously diminish tensions of SB_0 and r_0 in future work.

Future directions

An important improvement for future Ly α radiative transfer modeling in large-scale cosmological simulations such as TNG50 will be a treatment of complexities below the resolution scale (~ 100 pc). Such a model could either implement explicit numerical

subgrid models (Hansen *et al.*, 2006; Gronke *et al.*, 2017b) and retain our current principle of a parameter-untuned model. Alternatively, we could adopt an effective parameter-based model, incorporating observational findings to motivate Ly α production and escape fractions from star-forming regions (Weiss *et al.*, 2021).

In future work we will also use our coupling of voroILTIS and TNG50 to examine two key areas: environmental imprints on large-scale observations, and the information content of the spectral dimension. We found a significant redistribution of photons from star-forming regions to large scales ($\gtrsim 100$ kpc) which might affect the interpretation of Ly α intensity mapping experiments. Furthermore, spectral modeling of Ly α emission in cosmological volumes remains challenging given its multi-scale nature (Byrohl *et al.*, 2020b; Song *et al.*, 2020). However, detailed and spatially resolved spectral information on Ly α -halos is increasingly becoming available (Claeyssens *et al.*, 2019; Leclercq *et al.*, 2020), and promises to offer significant insight into the kinematics and small-scale structure of hydrogen gas in the CGM of dark matter halos across cosmic time.

Finally, we have here focused on lower mass star-forming galaxies with $M_{\text{halo}} < 10^{12} M_{\odot}$. Above this threshold, the AGN in more massive galaxies are known to have a significant impact on the ionization state of the CGM and the Ly α scattering processes occurring therein. Our current radiative transfer methodology does not account for local radiation fields from AGN, but this is a natural extension of voroILTIS which will allow us to compare to the many rich observational data sets of Ly α emission around quasars.

6.5 Conclusions

In this chapter we develop a technique to perform full radiative transfer calculations to trace resonantly scattered Ly α emission, and the Ly α halos (LAHs) around galaxies, at $2 < z < 5$. We do so by post-processing the TNG50 cosmological magnetohydrodynamical simulation (Nelson *et al.*, 2019a; Pillepich *et al.*, 2019) of the IllustrisTNG project.

This large volume offers a powerful statistical sample of thousands of LAHs, together with a high resolution of order 100 physical parsecs in the dense interstellar medium. Furthermore, our new radiative transfer code voroILTIS (Behrens & Byrohl, in prep) operates natively on the global Voronoi tessellation of the TNG simulation volume, incorporates both diffuse and galaxy emission, and self-consistently accounts for attenuation within the intergalactic medium.

This allows us to carry out a detailed investigation into the origins, physical properties, and emission sources which shape Lyman- α halos. At the same time, the realism of the

underlying TNG50 simulation enables us to make quantitative connections between LAH and galaxy properties. Our main findings are:

- Star-forming galaxies with $10^7 < M_*/M_\odot < 10^{10.5}$ at $2 < z < 5$ emitting Ly α photons are surrounded by extended Lyman-alpha halos (LAHs). We present the stacked, median predictions for TNG50 LAHs as a function of galaxy mass and redshift. The radial surface brightness profiles of LAHs have a characteristic shape comprised of a rapid, exponential decline followed by a large-distance flattening. This flattening arises from the density structure probed by rescattering photons as well as in-situ diffuse emission.
- Scattered photons from star-forming regions are the dominant contributor to LAH profiles on typically observed scales $r \lesssim 20$ pkpc. Given the importance of scattered photons, we stress the need to use radiative transfer simulations or semi-analytic expressions capturing such behavior. At larger distances, contributions from diffuse emission via recombinations and de-excitations become equally important.
- On larger scales $r \gtrsim 30$ pkpc we find that the flattening of LAH profiles is actually dominated by rescattered photons that originate from other nearby massive halos, rather than the primary halo itself. This proximity effect is boosted in high density environments, and should be observable. A careful reproduction of a survey's background subtraction and wavelength window are needed for comparison of this flattening to observations.
- Characterizing LAH sizes, we find that their half-light radii $r_{1/2}$ are of order 5 pkpc at $z = 2$, increasing to ~ 15 pkpc at $z = 5$. This signposts a significant redistribution of photons due to higher neutral hydrogen densities at higher redshifts. The exponential scale lengths r_0 also increase slightly towards higher redshift for galaxy stellar masses of $8.5 \leq \log_{10}(M_*/M_\odot) \leq 9.5$. Neither $r_{1/2}$ and r_0 show clear trends with mass. In contrast, our fiducial model shows a strong positive correlation of the central surface brightness SB_0 with both stellar mass and redshift.
- While AGN radiative feedback adds significant heating and ionization to the surrounding CGM, subsequently boosting intrinsic emission from both collisional excitations and recombinations, we only find a marginal impact on the emission mechanisms' relative importance and the overall scale lengths r_0 for LAHs with AGN activity in TNG.
- We compare the stacked, median LAH radial profile between TNG50 and data for galaxies with $8.5 \leq \log_{10}(M_*/M_\odot) \leq 9.5$ from the MUSE UDF and find good

qualitative agreement. We also demonstrate, by finding statistically consistent simulated analogs for individual observed profiles, that the simulation successfully reproduces the diversity of observed LAHs.

- For the quantitative comparison to observational results of LAH sizes as measured by half-light radii $r_{1/2}$ and exponential scale lengths r_0 , we find agreement at the level of $\sim 20\%$ at $z = 3$, with the $r_0 \sim 4$ pkpc of our simulated profiles slightly above those of the MUSE UDF data set ($r_0 \sim 3$ pkpc). This difference increases with redshift, to a factor of two at $z = 5$. Similarly, we find that the central surface brightness SB_0 of LAHs is in good agreement at low stellar mass, but becomes progressively worse towards higher M_* . Both tensions arise because LAH properties tend to correlate strongly with both galaxy mass and redshift in the simulations, but less so in the data. We attribute this behavior to a number of simplifying assumptions in our Ly α modeling, and discuss future improvements.

The extended Ly α emission around galaxies and quasars at $z > 2$ provides an insightful window into many aspects of galaxy formation and evolution. Here we have demonstrated the power of coupling a large-volume cosmological hydrodynamical simulation with a global Ly α radiative transfer modeling. Future improvements in our treatment of unresolved small-scale gas structure and local radiation fields from AGN will enable interpretation of additional datasets and upcoming surveys, from large-scale intensity mapping experiments to highly detailed, spectral data from targeted IFU surveys.

Data Availability

Data directly related to this publication and its figures is available on request from the corresponding author. The IllustrisTNG simulations, including TNG50, are publicly available and accessible at www.tng-project.org/data (Nelson *et al.*, 2019b).

Acknowledgements

We thank Floriane Leclercq for providing the full Ly α SB(r) data shown in Leclercq *et al.* (2017). We thank Anna Feltre, Roland Bacon, and the MUSE UDF team for providing SED fitting results prior to their release. We thank Max Gronke for fruitful discussions on Ly α radiative transfer. The primary co-author of Byrohl *et al.*, 2021, Dylan Nelson, acknowledges funding from the Deutsche Forschungsgemeinschaft (DFG) through an Emmy Noether Research Group (grant number NE 2441/1-1).

The primary TNG50 simulation was carried out with compute time granted by the Gauss Centre for Supercomputing (GCS) under GCS Large-Scale Project GCS-DWAR (PIs: Nelson and Pillepich) on the GCS share of the supercomputer Hazel Hen at the High Performance Computing Center Stuttgart (HLRS). GCS is the alliance of the three national supercomputing centres HLRS (Universität Stuttgart), JSC (Forschungszentrum Jülich), and LRZ (Bayerische Akademie der Wissenschaften), funded by the German Federal Ministry of Education and Research (BMBF) and the German State Ministries for Research of Baden-Württemberg (MWK), Bayern (StMWFK) and Nordrhein-Westfalen (MIWF). Additional simulations and analysis were carried out on the supercomputers at the Max Planck Computing and Data Facility (MPCDF).

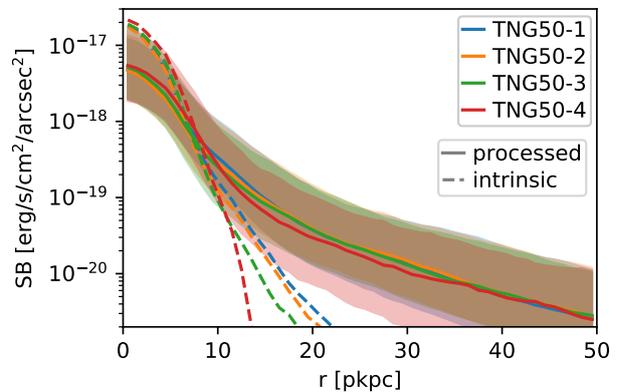
6.6 Supplementary material

6.6.1 Modeling assumptions

In this appendix, we assess various modeling assumptions along with numerical convergence convergence of our results. We focus on the star-formation emission source only, and derive radial profiles neglecting rescattered light from other halos, as we limit ourselves to rerunning the radiative transfer on halos with stellar masses $8.5 \leq \log_{10}(M_*/M_\odot) \leq 9.5$. However, in Sections 6.6.1/6.6.1, we only vary the analysis on the original simulations, hence the methodology there remains the same as in the main body in the Chapter 6 unless explicitly mentioned.

Hydrodynamical resolution

Fig. 6.13 Median stacked Ly α surface brightness radial profiles at $z = 3$ for a fixed stellar mass $8.5 \leq \log_{10}(M_*/M_\odot) \leq 9.5$ at different resolution runs of TNG50. TNG50-1 has a mass resolution of $m_{\text{baryon}} = 8.5 \times 10^4 M_\odot$, which increases by a factor of eight for each lower resolution run. Only star-formation is considered as an emission source, and scattered light contributions from other halos are ignored. The dashed lines correspond to the intrinsic emission radial profiles with a corresponding solid line for the profiles after radiative transfer. The shaded regions show the 16th to 84th percentiles. Differences between resolution levels are generally minor.



In Figure 6.13 we show the Ly α surface brightness radial profiles at different hydrodynamical resolutions of TNG50 runs stacked at fixed stellar mass $8.5 \leq \log_{10}(M_*/M_\odot) \leq 9.5$. Overall, we find the stacked profiles to be very robust over the different resolution levels shown. For fixed stellar bins, we find the intrinsic emission to slightly expand with increasing resolution. For the processed photons, there is a slight flattening of the stacked profiles between 10 and 20 pkpc.

Despite this the inferred r_0 values are effectively invariant with changing resolution: we find $3.1_{-0.6}^{+1.4}$ pkpc (TNG50-1), $3.0_{-0.7}^{+1.3}$ pkpc (TNG50-2), $2.7_{-0.5}^{+1.2}$ pkpc (TNG50-3) and $3.0_{-0.5}^{+0.8}$ pkpc (TNG50-4). However, there appears to be a decrease in the scatter, which may reflect the simpler density structure at lower resolution.

The resolution dependency appears to be significantly smaller than that in Zheng *et al.* (2011b). We note that even the lowest resolution run presented here has a higher resolution than that study in the proximity of halos.

Photon package count

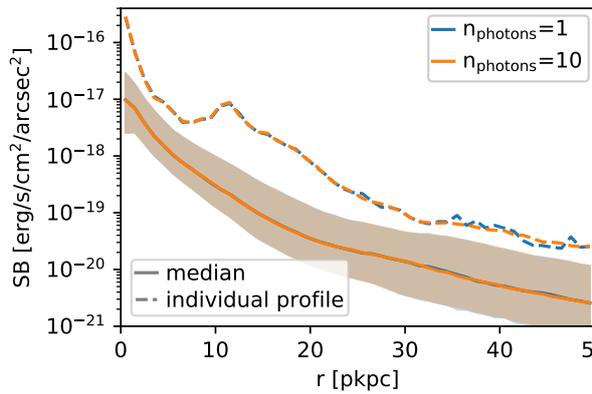


Fig. 6.14 Median stacked Ly α surface brightness radial profiles (solid lines) at $z = 3$ in TNG50 for a fixed stellar mass $8.5 \leq \log_{10}(M_*/M_\odot) \leq 9.5$ for 1 (fiducial, blue) and 10 Monte Carlo photon packages per star-forming cell (orange). To show the largest possible discrepancy, we also include the least-converged individual radial profile (dashed lines), which also demonstrates near perfect convergence in n_{count} . Our results are clearly insensitive to this parameter.

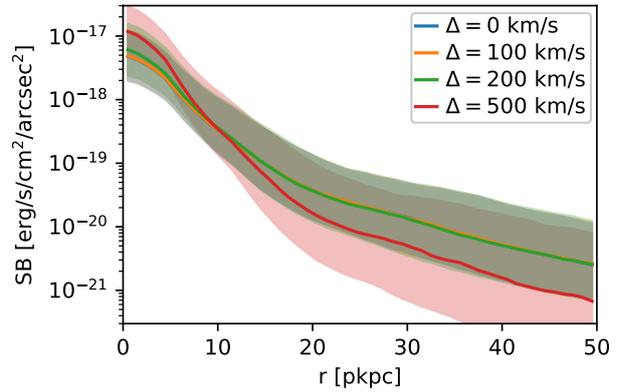
In our fiducial runs, we spawn one photon package per Voronoi cell and emission source (star-forming region). Particularly for photons originating from central star-forming regions, there might be a convergence issue given the large volumes that are to be traced out at larger radii. We thus focus on convergence checks for the photon package count from star-forming regions.

In Figure 6.14 we show the radial profiles with varying n_{count} of initial photon packages. The median stacked profiles (solid) are already fully converged at our fiducial choice of $n_{\text{photons}} = 1$. As the most stringent test, we also plot the radial profile of the single halo with the largest sum of squared errors between both these two test runs – the least converged individual profile. In this case we also find only minuscule deviations confirming that our choice of the photon count leads to a well converged result.

We note that for halos with stellar masses between $8.5 \leq \log_{10}(M_*/M_\odot) \leq 9.5$, our fiducial setup spawns ~ 35000 intrinsic photons: the intrinsically high resolution of TNG50-1 gives us already sufficient sampling.

Input spectrum

Fig. 6.15 Median stacked Ly α surface brightness radial profiles (solid lines) at $z = 3$ in TNG50 for a fixed halo mass bin of $\log(M_h/M_\odot) \in [10.5, 11.0]$. Here we contrast four different input photon wavelength offsets. In our fiducial model, photons are injected at the line-center within the rest-frame of the respective star-forming cell, indicated here as an offset of $\Delta = 0$ km/s from the line center (blue). Three non-zero offsets of $\Delta = 100$ km/s (orange), $\Delta = 200$ km/s (green) and $\Delta = 500$ km/s (red) are shown, where the shaded regions enclose the central 68 % of SB(r) at fixed r . We find little difference in the resultant LAH profiles between the different input spectra choices (Δ), except for the unrealistically large case of 500 km/s.



In our fiducial model, photons are injected at the Ly α line-center in the rest-frame of the respective hydrodynamical cell. For diffuse emission, Equations (2.44)/(2.45), this is a fair assumption given that optical depths are moderate, although subgrid clumpiness would add complexity. However, in dense star-forming cells radiative transfer is significantly more complex due to small-scale dust, clumpiness, and ionization. Our radiative transfer simulations do not self-consistently capture these details. While attenuation on these scales only changes the overall radial profile normalization of the attenuated component, the spectral shape of emitted photons might have an effect on the radial profile shape itself.

As shown in similar cosmological simulations by Byrohl *et al.* (2019), Ly α spectra often appear to have too much flux at wavelength larger than the line-center (too ‘blue’) compared to observations. We thus consider an additional redshift for the injected photons, which naturally leads to a more realistic ‘red’ spectrum. In Figure 6.15 we show the results of radiative transfer simulations where the redshift Δ from the line center has been varied between 0 km/s (fiducial model) and 500 km/s, considering emission from star-formation only. The corresponding intrinsic (dashed) and emergent (solid) Ly α spectra are shown in Figure 6.16.

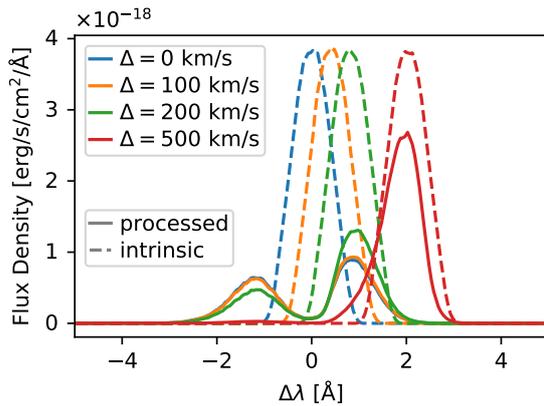


Fig. 6.16 The mean stacked Ly α flux spectra at $z = 3$ for a fixed halo mass $\log(M_{\text{h}}/M_{\odot}) \in [10.5, 11.0]$ in TNG50 for a varied injected photon wavelength offset $\Delta\lambda$ from the Ly α line-center, integrating within a 3 arcsecond radius aperture. Dashed lines show the intrinsic photon frequency distributions, while solid lines show the emergent (processed) spectra. This corresponds to the same sample for which we show the stacked profiles in Figure 6.15.

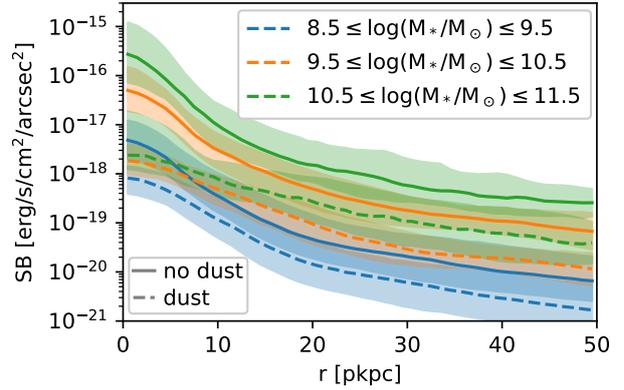
We find that the choice of Δ has virtually no impact, except for the unrealistically high offset of 500 km/s. We recover sizes of $3.1^{+1.4}_{-0.7}$ pkpc ($\Delta = 0$ km/s), $3.1^{+1.4}_{-0.7}$ pkpc ($\Delta = 100$ km/s), $3.0^{+1.4}_{-0.7}$ pkpc ($\Delta = 200$ km/s) and $2.5^{+1.0}_{-0.4}$ pkpc ($\Delta = 500$ km/s) for halos with $8.5 \leq \log_{10}(M_{\star}/M_{\odot}) \leq 9.5$. For large Δ the scale length r_0 decreases as less scatterings occur given the lower cross-section in the Ly α line profile wings, while interaction with the IGM also decreases as this is mostly driven by blue photons. Although the frequency distribution at injection will be important for future studies, it is not crucial for our study of LAH profiles.

Dust

In our fiducial model we have neglected dust. Comparison of luminosities with observations in Section 6.4.2 indicate that they are reasonable for the fiducial stellar mass range $8.5 \leq \log_{10}(M_{\star}/M_{\odot}) \leq 9.5$ without further explicit dust modeling, although this will be necessary for the high mass end. Dust modeling can either take the form of an additional attenuation factor of Ly α emission, particularly of star-forming gas cells, or an explicit dust treatment in the Ly α RT. In this section, we explore the option of including an explicit dust treatment in the Ly α RT, as a preliminary study. We neglect clumpiness, such that dust is smooth at the resolution scale of the simulation. We use the model for Milky Way like dust (Weingartner *et al.*, 2001) as implemented in Behrens *et al.* (2019), which relies on the metallicity field in TNG50.

We expect that dust will primarily modulate the overall normalization of the radial profiles for the emission from the ISM given in Equation (6.3). Nevertheless, this can boost the relative importance of other emission mechanisms, and change the overall radial profiles. As dust content is related to gas-phase metallicity and galaxy mass, the impact of dust could alter the trends of LAH properties with mass, and be particularly important at higher galaxy masses.

Fig. 6.17 Median stacked Ly α surface brightness radial profiles in various stellar mass bins at $z = 3$ in TNG50. In this plot, we only consider contributions from star-forming regions, where the impact from dust is most severe. We show the results with (dashed lines) and without (solid lines) dust modeling. We generally find a significant suppression of flux that significantly increases for higher stellar mass halos.



In Figure 6.17 we show the impact of dust in our simulations at $z = 3$ for emission from star-formation only. We contrast LAH radial profiles between the fiducial dust-free case (solid lines) and the dust included model (dashed lines). Our findings on the impact of dust are similar to those in Laursen *et al.* (2009). In particular, surface brightness is increasingly suppressed in overdense, dusty regions. In addition, regions of lower density are uniformly suppressed as dust limits the escaping (and then rescattering) contributions from star-forming regions.

Dust attenuation strongly scales with the stellar mass. For example, the median attenuation for the central surface brightnesses is less than one order of magnitude for $8.5 \leq \log_{10}(M_*/M_\odot) \leq 9.5$, but roughly two orders of magnitude for $10.5 \leq \log_{10}(M_*/M_\odot) \leq 11.5$.

The inner $\lesssim 10$ pkpc are particularly suppressed, decreasing the central surface brightness values and thus increasing the exponential scale lengths. For the scale lengths of halos with $8.5 \leq \log_{10}(M_*/M_\odot) \leq 9.5$, we find an average $r_0 \simeq 3.1_{-0.7}^{+1.4}$ pkpc without dust and $5.1_{-1.2}^{+2.4}$ pkpc with dust. Both the median scale length and its variance within the sample increase with dust. The increase in scale length with this dust modeling is even larger when incorporating other emission sources, as they become relatively more important (see Section 6.3). It is clear that future models will need to include at least a basic dust model.

Impact of local ionizing sources

Properly accounting for the impact of local ionizing sources on the temperature and ionization state of the CGM requires, ultimately, full radiation-hydrodynamical simulations. However, this remains computationally infeasible for the present combination of galaxy sample size, resolution and redshift.

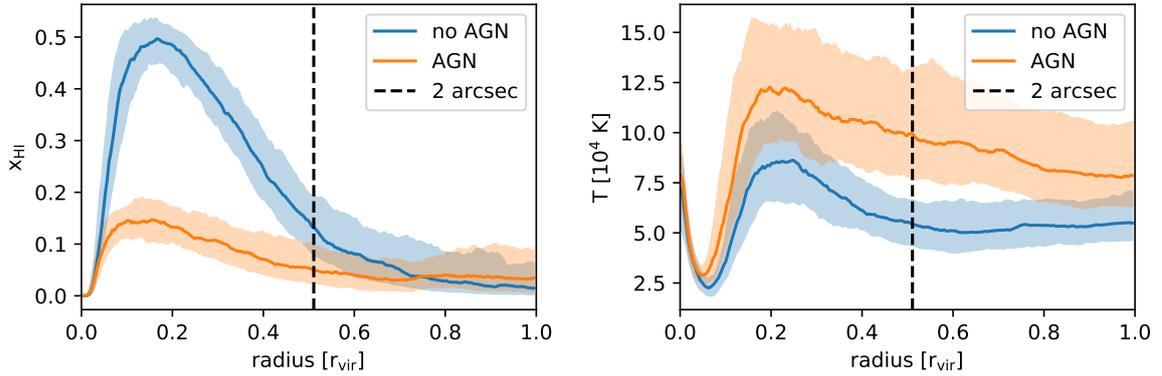


Fig. 6.18 Median radial profile at $z=3.0$ for $8.45 \leq \log_{10}(M_*/M_\odot) \leq 8.55$ split into halos with/without AGN in TNG50 (148/122 halos) for the neutral hydrogen fraction (**left**) and temperature (**right**). Angular averaging uses mass-weighting. The shaded regions show the 16th to 84th percentiles. The vertical dashed line shows, in the median, the radius corresponding to 2 arcseconds.

In this section, we assess the impact of local ionizing radiation from AGN and stellar populations. An on-the-fly treatment (i.e. self-consistently considering local sources' impact on cooling in each time step) is implemented for AGN radiative feedback in TNG. Hence, we can measure such ionizing sources' impact on the LAHs by comparing halo samples with and without SMBHs, as in Section 6.6.1. In the subsequent Section 6.6.1 we then discuss what impact ionizing radiation from stellar populations would have, given that this is not included in our models.

Active Galactic Nuclei Ionizing radiation from SMBH is modelled in TNG, which we will describe in the following. Note that we use the terms AGN and SMBH interchangeably here.

The intrinsic luminosity from black hole mass accretion at a rate \dot{M}_{BH} in TNG is taken to be

$$L_{\text{bol}}^{\text{SMBH}} = (1 - \epsilon_f) \tilde{\epsilon}_r \dot{M}_{\text{BH}} c^2. \quad (6.4)$$

according to the radiative efficiency $\tilde{\epsilon}_r$ and feedback energy fraction ϵ_f . The ionizing luminosity escaping the galaxy is then given by

$$L_{\text{UV,esc}}^{\text{SMBH}} = A f_{\text{esc}}^{\text{AGN}} L_{\text{bol}}^{\text{SMBH}}, \quad (6.5)$$

where A describes the bolometric correction factor of ionizing luminosity for the assumed spectral energy distribution and $f_{\text{esc}}^{\text{AGN}}$ incorporates the escape of ionizing flux from the

galaxy as

$$f_{\text{esc}}^{\text{AGN}} = \omega_1 \left(\frac{L_{\text{bol}}^{\text{SMBH}}}{10^{46} \text{ erg/s}} \right)^{\omega_2} \quad (6.6)$$

with $\omega_1 = 0.3$, $\omega_2 = 0.07$ (Hopkins *et al.*, 2007) that results in a median obscuration of $f_{\text{esc}}^{\text{AGN}} \sim 0.19_{-0.08}^{+0.10}$ for all AGN in TNG50 at $z = 3$. The sub- and superscript denote the interval of the central 99.73% of AGN. Higher obscuration factors typically occur for AGN in higher mass halos.

For AGN with low accretion $\dot{m}/\dot{m}_{\text{Eddington}} < 0.002$, the radiative feedback is set to zero in the TNG model. The radiative output of AGN differs substantially with redshift. While at $z = 0.0$ the vast majority of halos are inactive given this threshold, only $\lesssim 5\%$ are radiatively inactive for the studied redshift range $z \geq 2$.

More details of the implemented radiative feedback model in TNG can be found in Vogelsberger *et al.* (2013) and Weinberger *et al.* (2017, 2018).

TNG incorporates “on-the-fly” ionizing radiation from SMBH in each time step as given in Equation (6.5), which impacts gas cooling, temperature, and ionization state. For the gas cooling, these local sources’ photoionization and photoheating are incorporated. The model assumes the bolometric intensity as $J_{\text{UV,esc}}^{\text{SMBH}} \propto L_{\text{UV,esc}}^{\text{SMBH}}/r^2$ at a distance r of each nearby AGN under the assumption that the gas is optically thin to SMBH radiation. Photoionization and photoheating are calculated as a superposition of the local ionizing radiation from AGN and the uniform metagalactic background. For halos with a radiating black hole in TNG, the AGN field dominates over the UV background within the halo.

We will now assess the impact of the SMBHs’ ionizing radiation on the halos’ surrounding gas and the modifications to their Ly α halos. To do this, we split the sample in halos with and without SMBHs. In TNG50, this is equal to splitting the samples by AGN radiative feedback activity.

In Figure 6.18 we show the median radial profiles for the neutral hydrogen fraction x_{HI} and temperature T for the sample of halos in the stellar mass range $8.45 \leq \log_{10}(M_*/M_\odot) \leq 8.55$ at $z = 3.0$ in TNG50. For this mass range, we have a roughly equal sample of halos with (148) and without (122) SMBH activity (also see cutoff in Figure 6.20). In addition to the central 68 percentiles as shaded area, we also show the 2 median arcsecond radius given the virial radii, along the x-axis (vertical dotted line).

Note that only 1 of the 148 AGN is radiatively inactive. Hence, for our sample, AGN presence nearly always implies a significant ionizing flux in the surrounding CGM according to Equation (6.5). As a consequence of the SMBH (radiative) feedback, we find a strongly suppressed neutral hydrogen content in affected halos. At the same time,

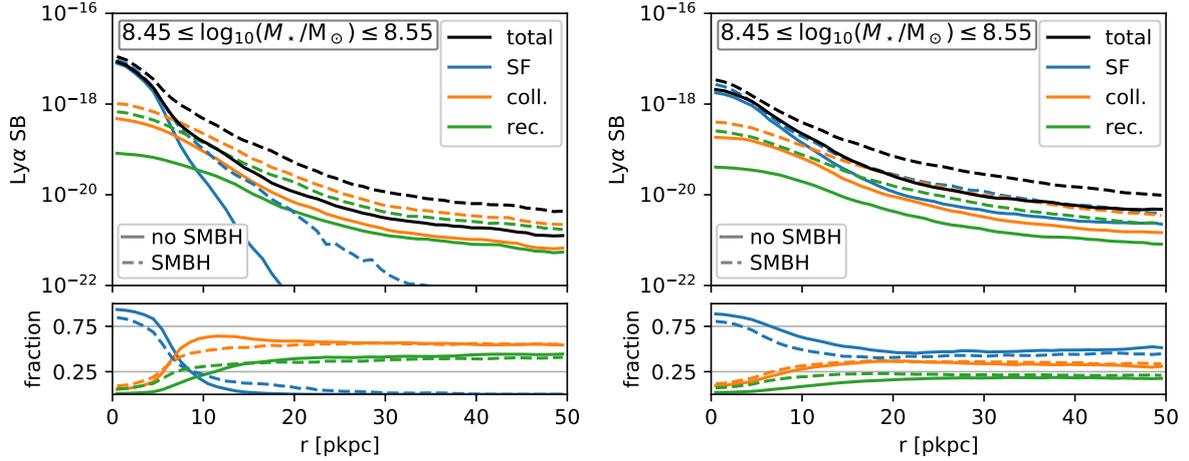


Fig. 6.19 Median stacked Ly α surface brightness radial profiles from **(left)** intrinsic emission (i.e. no scattering) and from **(right)** processed photons (i.e. with scattering) at $z = 3$ for halos with $8.5 \leq \log_{10}(M_*/M_\odot) \leq 9.0$ decomposed into different emission sources (upper panel), and the relative fraction of each (lower panel) in TNG50. Dashed lines show the halos with AGN activity (148 halos), while solid lines show the halos without AGN activity (122 halos). Without scatterings collisional excitations are dominating the radial profiles outside of the star-forming regions $r_0 \gtrsim 5$ pkpc irrespective of AGN activity. AGN activity however largely boosts emission from collisional excitations and recombinations.

the temperature is significantly higher outside the star-forming regions at $r > 0.2r_{\text{vir}}$. Given the decreased neutral hydrogen fraction, we expect a boosted Ly α emission from recombinations, and in particular a boosted Ly α emission from collisional excitations given the strong temperature dependency of the latter.

On the left of Figure 6.19 we show the *intrinsic* median Ly α surface brightness radial profiles for $z = 3.0$ and $8.45 \leq \log_{10}(M_*/M_\odot) \leq 8.55$ split into their respective emission mechanisms and divided into the sample with and without SMBH activity similar to the dashed lines in Figure 6.8. Without scattering, we generally find collisional excitations to be the dominant emission mechanisms for the majority of halos above 7 pkpc irrespective of SMBH activity. Particularly recombinations are strongly boosted in the presence of a SMBH for $r_0 \lesssim 20$ pkpc. Interestingly, while the overall emission from collisional excitations is boosted, the relative fraction decreases not just because of the higher recombination rates but also due to a larger fraction of emission from star-forming regions.

To understand latter point, we stress that the underlying samples differ. Particularly, the sample hosting SMBHs, even though with a similar stellar mass, are typically more massive with a median total halo mass $\log_{10}(M_h/M_\odot)$ higher by about 0.2 dex and a significant scatter towards higher masses. The high mass objects show a larger amount

of satellites contributing to the additional emission from star-formation in the halos' outskirts.

The right plot of Figure 6.19 shows the *processed* median radial profiles, i.e. after running the Ly α radiative transfer for the intrinsic emission. While all emission mechanisms remain boosted in absolute terms when SMBH activity is present, collisional excitations and particularly recombinations become more important in relative terms. However, irrespective of SMBH activity the radial profiles remain dominantly sourced by emission from star-forming cells. At $r_0 \gtrsim 20$ pkpc, contributions from collisional excitations become close to equally important to photons from star-forming regions. However, this trend is the same for both the AGN and no-AGN sample.

From the median profiles it appears that there is a slight additional flattening for the AGN hosting subsample compared to the no-AGN sample. To quantify the difference of the LAHs' shape, we calculate the individual exponential scale lengths r_0 and find an overall increase of roughly 11% from $3.6_{-0.9}^{+1.2}$ pkpc (no AGN) to $4.0_{-1.0}^{+1.4}$ pkpc (AGN).

This increase is however not solely driven by the ionizing AGN radiation but also by the different halo population, given the fixed stellar mass ranges of the sample. If we instead constrain the total halo mass to be within a fixed range of $10.6 \leq \log_{10}(M_h/M_\odot) \leq 10.8$, which corresponds to the central $\sim 68\%$ of the non-AGN $8.45 \leq \log_{10}(M_*/M_\odot) \leq 8.55$ sample, we obtain a reversed trend with r_0 being $3.8_{-0.9}^{+1.4}$ pkpc ($3.6_{-0.8}^{+1.7}$ pkpc) for the sample without (with) AGN.

In conclusion, the impact of ionizing radiation from SMBHs in TNG appears to be significant in terms of ionization state, temperature and thus intrinsic Ly α emission of the gas, but the findings on the LAH shape, particularly through the scale radius r_0 , remain unchanged irrespective of modelled SMBH activity. We therefore do not split our sample within the main body with respect to the presence of AGN.

Note that here, we only discuss the AGN's impact through its ionizing budget onto the surrounding. In particular, we do not discuss nor implement a description for the unresolved Ly α emission from AGN itself at this point. We would expect that this emission would scatter outwards and produce a contribution to the LAH of similar shape to that of the star-formation, given the concentrated emission source scattering into the surrounding CGM. As differences in the intrinsic emission in terms of spatial and spectral distribution exist, this would need to be explored in future investigations.

Stellar Populations

TNG does not incorporate the ionizing flux from local stellar populations as is done for SMBHs. Hence, the possible impact of those sources on our predicted LAH profiles

cannot be assessed within the existing simulations. However, we can derive an estimate of the upper limit of this effect, by comparing to the local ionization from AGN.

The ionizing luminosity escaping from the stellar populations, adopting a formulation consistent with our model, is given as:

$$L_{\text{UV,esc}}^{\text{SF}} = \frac{f_{\text{UV,esc}}}{f_{\text{B}}(1 - f_{\text{UV,esc}})} \frac{\langle E_{\gamma,\text{UV}} \rangle}{E_{\gamma,\text{Ly}\alpha}} \epsilon_{\text{SF}} V_{\star} \quad (6.7)$$

As in Equation (6.3), we assume no dust and *Case B* recombination with $f_{\text{B}} = 0.68$ being the conversion factor from ionizing to Ly α photons. $\langle E_{\gamma,\text{UV}} \rangle / E_{\gamma,\text{Ly}\alpha}$ is the ratio of the average ionizing photon energy in the population and the Ly α line transition energy. We assume $f_{\text{UV,esc}} = 0.1$ in Equation (17) of Dijkstra (2019).

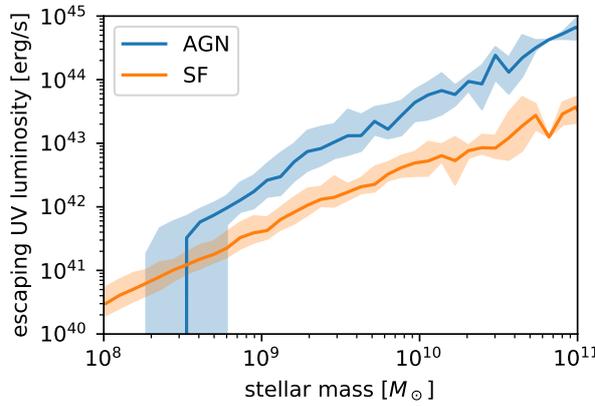


Fig. 6.20 We show the median of the UV luminosity escaping the galaxy at a given stellar mass of the host halo at $z = 3.0$ in TNG50. The UV luminosities for star-formation (SF) and SMBH are derived from Equations (6.5)/(6.7) respectively.

In Figure 6.20 we show the escaping UV radiation from stellar sources and AGN as a function of the stellar mass of the host halo at $z = 3.0$ in TNG50. We find AGN activity for the majority of halos with a stellar mass above $\log_{10}(M_{\star}/M_{\odot}) \gtrsim 8.5$. Above this mass, we find an approximate power law $L_{\text{UV,esc}} = L_0 \cdot \left(\frac{M_{\star}}{10^{8.5}M_{\odot}}\right)^{\alpha}$ with $L_0 = 10^{41.7}$ and $\alpha = 1.31$ ($L_0 = 10^{41.1}$, $\alpha = 0.99$) for AGN (SF). We thus find that $L_{\text{UV,esc}}^{\text{AGN}} \gtrsim 6 \cdot L_{\text{UV,esc}}^{\text{SF}}$ at $z = 3.0$ for the bulk of halos with AGN activity and a growing disparity between the luminosities with larger mass given the larger slope for AGN. The discrepancy grows (shrinks) at lower (higher) redshifts across the stellar mass range.

Changing $f_{\text{UV,esc}}$ or adopting different assumptions concerning e.g. metallicity, binary fraction or IMF can boost the stellar populations' escaping UV luminosity. Given the large margin at $z = 3.0$ between AGN and stellar luminosities, qualitative findings should be robust. However, such adjustments might lower the redshift at which stellar populations' UV luminosity becomes dominant into the upper studied redshift range. For the given UV luminosities, the AGN ionizing flux integrated over the halo population dominates over the ionizing flux from stellar populations. This ratio peaks at $z = 2$ in TNG50, where

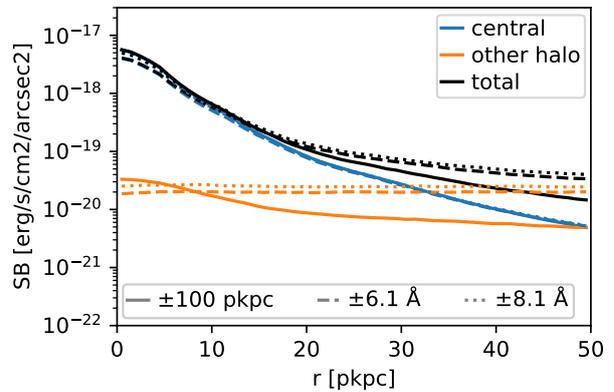
AGN are most relevant, while stellar populations start to dominate the overall ionizing flux at higher redshifts $z \gtrsim 6$.

We found that the ionizing flux of SMBHs, if present, dominates over the ionizing flux from stellar populations in the same host halo. A consideration of the impact photoionization and photoheating thus needs to primarily address the impact of AGN, for which a simplified description is indeed implemented in TNG50.

In our fiducial sample with stellar masses of $8.5 \leq \log_{10}(M_*/M_\odot) \leq 9.5$, 87% of halos host at least one SMBH with a radiation field according to Eqn. 6.5. Therefore, our analysis contains the primary ionization source. As demonstrated in the supplementary material in Section 6.6.1, while the gas becomes significantly hotter and ionized, the LAH size measurements and our conclusions remain largely unaffected. As a consequence, we also find the inclusion of stellar ionizing sources, with overall weaker UV luminosities, would have a small impact on our overall findings.

Impact of spectral information

Fig. 6.21 Median stacked radial Ly α profile for galaxies with stellar masses $8.5 \leq \log_{10}(M_*/M_\odot) \leq 9.5$ at $z = 3$ in TNG50. We decompose the radial profiles into their two dominant origins from central galaxies (blue) and other halos (orange). For the solid line, we show the fiducial method throughout this chapter that ignores spectral information and integrates all photon contributions that scatter last in the ± 100 pkpc around the halo's position. The dashed and dotted lines show the contributions when incorporating spectral information in a ~ 12 Å (dashed) and ~ 16 Å (dotted) observed wavelength window around the Ly α line center. The central surface brightness is slightly suppressed when considering spectral information due to some emitters with a spectral diffusion from the central galaxies exceeding the imposed wavelength window. At large radii, the wavelength windows lead to a stronger flattening due to the larger physical depth.



For simplicity, we ignored spectral information throughout this chapter and instead sum all photons scattering last within ± 100 pkpc depth around the halos' center. Here, we show the difference when instead using a simple spectral bandwidth from the photons'

spectral information incorporating Hubble flow, peculiar velocity and spectral diffusion. We note that a fair comparison will need to carefully incorporate the varying methodology. For example Leclercq *et al.* (2017) uses an adaptive spectral bandwidth per emitter.

In Figure 6.21 we show the median stacked radial profiles using the fiducial depth integration based on the last scattering of photons (solid line) opposed to a fixed spectral bandwidth of $\pm 6.1 \text{ \AA}$ (dashed line) and $\pm 8.1 \text{ \AA}$ (solid line) for the overall radial profile and split into the dominant emission origins (central galaxies in blue, other halos in orange). Ignoring spectral diffusion, we effectively model an adaptive spectral bandwidth to capture potentially large spectral diffusion. Thus, a fixed spectral bandwidth leads to a slightly suppressed median radial profile in Figure 6.21.

The scale lengths $r_0 = 4.2_{-1.1}^{+2.0} \text{ pkpc}$ ($\pm 6.1 \text{ \AA}$) and $4.0_{-1.0}^{+1.9} \text{ pkpc}$ ($\pm 8.1 \text{ \AA}$) for the fixed bandwidth window slightly increase compared to the fiducial setup due to the exclusion of Ly α contributions that diffused outside of these windows.

In Figure 6.7 and 6.12 we demonstrate a flattening of the radial profiles that is dominated by contributions from other halos. This flattening will therefore be heavily influenced by the environment in a chosen field of view and the methodology for its detection and stacking. A fair comparison will thus require a thorough reproduction of factors such as field overdensity, source masking and chosen wavelength depth along the line of sight. Generally, a fixed bandwidth as in Figure 6.21 causes a larger flattening due to the larger relative background contribution from other halos which is expected to scale roughly linear with the integration depth. The integration depth from the differential Hubble flow corresponds to 2.4 pMpc and 3.2 pMpc respectively compared with the 0.2 pMpc in the fiducial setup.

OUTLOOK AND CONCLUSIONS

7.1 Outlook

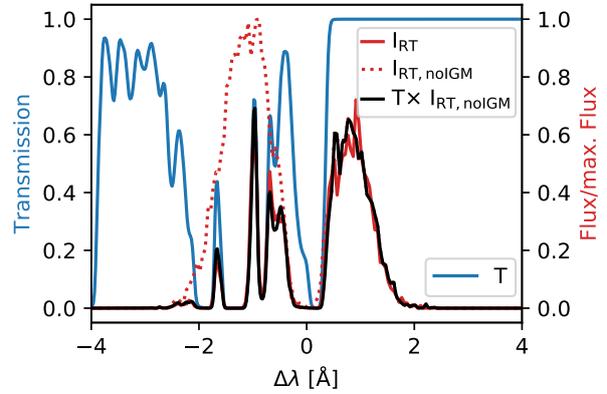
In Section 7.1.1, we will briefly summarize the advances and potential improvements on the results published from Chapters 4, 5 and 6. In Section 7.1.2, we will outline a few possible applications of the radiative transfer code in the near future.

7.1.1 Follow-up on previous work

In Chapter 4, we have shown new phenomenology in the clustering signal of LAEs due to Ly α spectral shifts. We have presented different schemes incorporating additional information of the Ly α spectra to partially correct for this distortion. Since then, an investigation by (Gurung-López *et al.*, 2021) showed further improvement by applying machine learning techniques to minimize the distortion effect.

As suggested in Byrohl *et al.* (2019) and Gurung-López *et al.*, 2021, calibration of those corrections to observations is desirable. Such calibration could be performed by obtaining deeper spectra of a subsample of LAEs that contain additional nebular lines, which can then be used to measure the true systemic redshift. The HETDEX survey is currently spectroscopically surveying more than 500 square degrees until 2023 (Gebhardt *et al.*, 2021). Upon completion, one of its primary targets will be to constrain the equation of state of dark energy. For competitive constraints, the two-point statistics need to be properly corrected for a range of distortions, such as interlopers (Farrow *et al.*, 2021). These corrections also have to account for Ly α radiative transfer effects both from potential selection effects (Wyithe *et al.*, 2011; Behrens *et al.*, 2018) and the distortion from Ly α spectral shapes presented in Byrohl *et al.*, 2019. As modeling uncertainties from our simulations are significant, calibration remains necessary, such as from suggested other emission line's systemic redshift.

Fig. 7.1 Spectrum and IGM transmission of a simulated Ly α emitting galaxy in TNG100 at $z = 3$. In blue, we show the transmission fraction of Ly α radiation for a given line of sight. The solid red line shows the spectrum resulting from the full Ly α radiative transfer. The solid dotted line shows the spectrum when we ignore IGM interaction at radii $r \geq 1.5 \cdot r_{\text{vir}}$ from the hosting halo. Finally, the black line shows the multiplication of the IGM transmission curve (blue line) with the galaxy spectrum without IGM interaction (red dotted line). From the good match between solid red and black lines, we see that we can separate the different scale's interactions and reproduce full spectra from the calculated IGM transmission curves.



In Chapter 5, we took a closer look at the IGM's impact on spectra from Ly α emitting galaxies. Introducing simple statistical measures, such as the fraction of triple-peaked spectra, we investigated how to break the degeneracies of the galaxy scale and IGM scale imprints. For more dedicated studies, e.g. using more realistic galaxy scale spectra, we published the catalog of transmission curves (Byrohl *et al.*, 2020a) that is being used by the community. In addition, this catalog might be useful for incorporating the IGM interaction in high-resolution zoom-in simulations, where the IGM interaction cannot be calculated self-consistently within the volume. In Figure 7.1, we show a proof of concept for incorporating the IGM interaction onto a galaxy-scale Ly α RT simulation: by multiplying the line-of-sight transmission curve with the small-scale spectrum, we reproduce the spectral shape that is arising from a full Ly α radiative transfer simulations including the intergalactic scale very well.

In Chapter 6, we presented predictions for the radial profiles of Ly α halos based on our radiative transfer code and the TNG simulations. We found good agreement with observed LAH radial profiles from the MUSE-UDF survey at $z \sim 3$. Within these simulations, we found that Ly α scattering is the dominant contribution to the extended Ly α halos, which motivates the need for Ly α radiative transfer simulations. In those previous simulations, we did not calibrate the luminosities to observations of the luminosity function and central surface brightnesses. Using such calibration in the future will allow us to make predictions in yet unobserved regimes such as the cosmic web, see Section 7.1.2. Also, the detection of a flattened radially averaged mean Ly α signal on cMpc scale (Kakuma *et al.*, 2021; Kikuchihara *et al.*, 2021) needs further exploration into its origin that we touched upon in this chapter.

With integral field spectroscopy, three-dimensional cubes of the Ly α halos become available. These cubes encode a range of the CGM's properties, such as ionization and kinematics. By employing our radiative transfer code and a suite of (radiation-)hydrodynamical simulations of galaxy formation, we can explore the Ly α signatures making the CGM visible at such high redshifts. Finding adequate statistical measures, we can then deduce some of the CGM's properties from observed IFS observations, whose interpretation is made difficult due to the complex radiative transfer. The nature of LAHs and the importance of scattering contributions to LAHs can be further constrained in future simulations of multiple emission lines. Complementary emissions lines have already been used in previous HST surveys for this purpose (Hayes *et al.*, 2014) and upcoming telescopes such as JWST will propel such studies.

7.1.2 New applications of the radiative transfer code

Within the presented framework developed around the radiative transfer code, there is a range of applications under investigation or open for future exploration. Here, we list a couple of those new avenues of research.

Correlation Lyman-alpha emission and absorption

The Ly α forest is a powerful tracer of the matter distribution yielding insights for the Universe's structure formation and underlying cosmology (Busca *et al.*, 2013). The Lyman- α forest arises from ionizing radiation that is redshifted in the expanding Universe and subsequently absorbed by the Ly α transition upon reaching the Ly α line-center. Observing bright spectra of galaxies and quasars in particular, these absorption features can be used to map out the neutral hydrogen distribution along the line-of-sight to such luminous objects. This allows to reconstruct the hydrogen distribution in three dimensions (Lee *et al.*, 2014, 2018) – even on Mpc scales given the sparsity of bright background sources.

Cross-correlation of the Ly α forest with LAEs can reveal additional information about the LAEs' environment, such as the surrounding overdensity they formed in and the subsequent escape of ionizing radiation back into it. First observational studies of such cross-correlation exist using HETDEX LAEs and Ly α forest information from eBOSS and the CLAMATO survey (Mukae *et al.*, 2020).

Analysis with Ly α radiative transfer is desirable. It enables a self-consistent treatment of the obscuration and modification of the spectral shape given the environment traced by the Ly α forest, whose impact we have seen in Chapter 5. Additionally, we can use the Ly α forest in cross-correlation to the Ly α intensity.

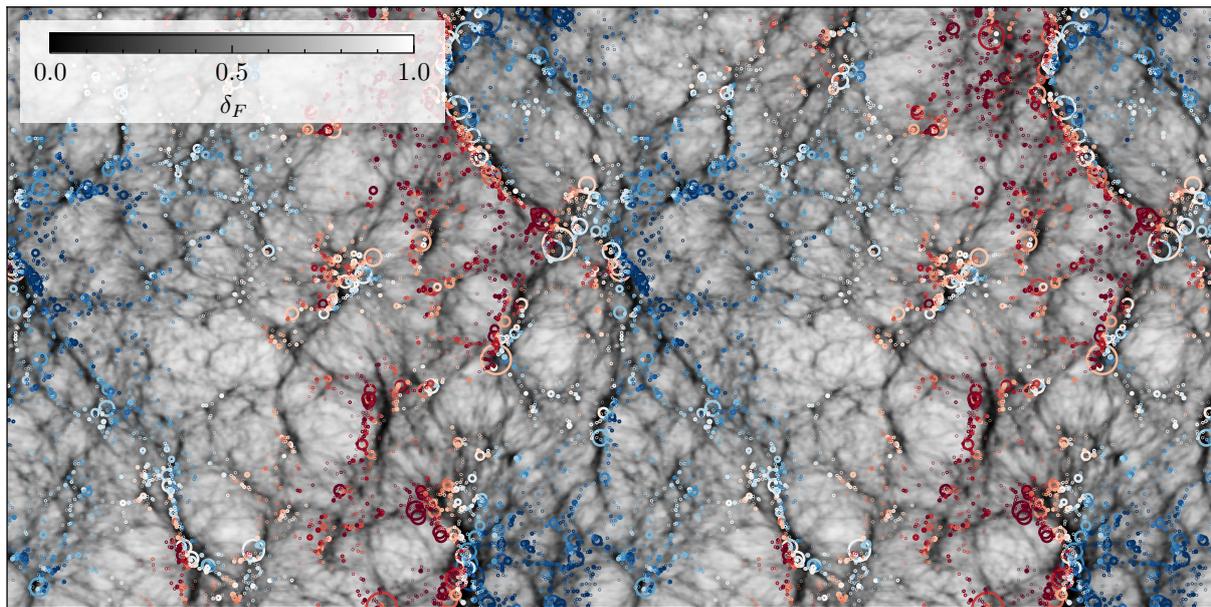


Fig. 7.2 Correlating the $\text{Ly}\alpha$ forest and LAEs can reveal additional information. We show the $\text{Ly}\alpha$ forest calculated with `VOROIL` on top of TNG100 at $z = 3$ with the line of sight along the x -axis. We project a slice with a depth of 2.1 pMpc and show the whole box of 26.6 pMpc (twice periodically along the line of sight). Additionally, the LAE (real space) positions are shown with circles with their sizes scaling with their star-formation rate. Additionally, color-coding of the LAE positions indicates the velocity along the line of sight.

Other resonant emission lines

The $\text{Ly}\alpha$ radiative transfer code can be easily expanded to other resonant emission lines whose scattering we need to capture to explain current and upcoming observations. Two promising resonant lines that have been observed with extended shapes are:

- **MgII:** Singly ionized magnesium, in the ground state with only one valence electron left, can be excited from $2p^63s$ to $2p^63p$ with next to 100% decay back into the ground state with the emission lines at 2796 and 2803 Å (Prochaska *et al.*, 2011). This enables resonant scattering. Spin-orbit coupling for the $2p^63p$ state leads to an emission line doublet with a moderate splitting of 780 km/s. While MgII optical depths are lower than for $\text{Ly}\alpha$ in the CGM, they can still be significant, leading to numerous scattering events depending on direction (Nelson *et al.*, 2021).
- **CIV:** The triply ionized carbon's $1s^22s$ to 1^22p resonant doublet with wavelengths of 1548 and 1550 Å traces highly ionized environments given the required ionization energy of ~ 48 eV (Draine, 2011). Complementary to $\text{Ly}\alpha$ as a potential indicator of Lyman continuum photon leakage, CIV could give further constraints on the

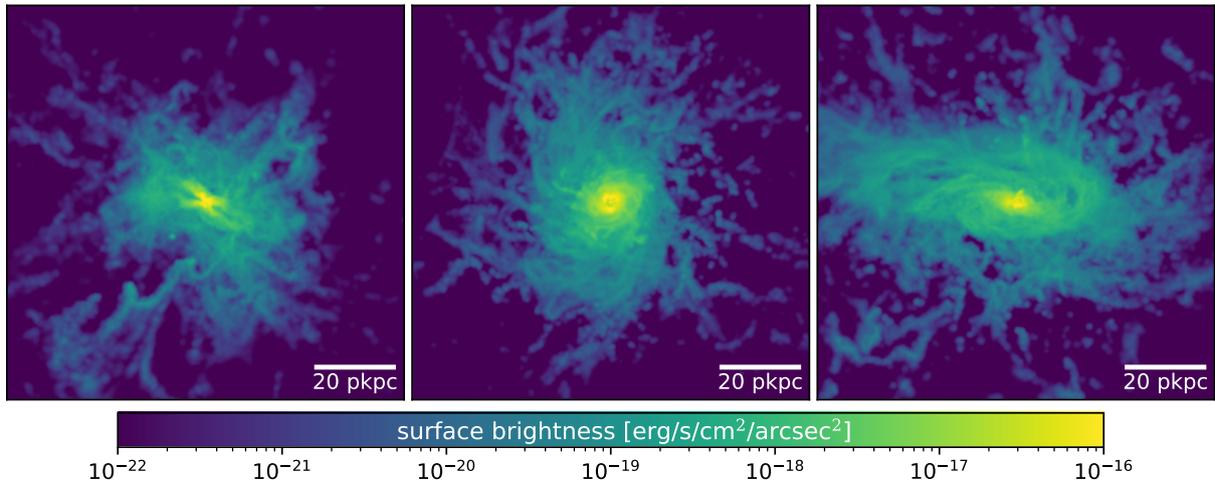


Fig. 7.3 Surface brightness maps for the MgII emission from the CGM of three different halos with stellar mass $M_{\star} \approx 10^{10} M_{\odot}$ at $z = 0.7$ in TNG50. The data was taken from the study published in Nelson *et al.* (2021). The maps only show the intrinsic emission, while in reality, MgII photons experience resonant scatterings due to moderate optical depths. With voroILTIS, we are able to simulate MgII and other resonant emission lines natively in AREPO based simulations such as TNG.

escape of higher ionizing photons in particular, which in turn can help to further pin the ionizing sources in the epoch of reionization (Berg *et al.*, 2019).

Extended emission from the resonant CIV and MgII lines has been observed in the last years (Cai *et al.*, 2017; Marques-Chaves *et al.*, 2019; Rupke *et al.*, 2019; Burchett *et al.*, 2021; Zabl *et al.*, 2021). However, detection of these lines is not guaranteed with deep observations (Arrigoni Battaia *et al.*, 2015). The presented radiative transfer code can be modularly extended to treat other resonant emission lines such as MgII in optically thick environments.

Combining different resonant emission lines, such as $\text{Ly}\alpha$, FeII, MgII, as well as other suitable candidates such as FeII and SiIV (Finley *et al.*, 2017a,b; Berg *et al.*, 2019), and non-resonant emission lines such as $\text{H}\alpha$ will enable a detailed characterization of the CGM's state.

In Figure 7.3, we show the surface brightness maps of MgII emission modeled on top of three halos in TNG50 from Nelson *et al.* (2021). The derived predictions for observational studies neglect the substantial optical depths the photons experience within the halos.

In the future, we will extend voroILTIS's capability to perform the radiative transfer for MgII and other resonant emission lines in a modular fashion to provide a link between the wealth of incoming observations with state-of-the-art simulations. The analysis of individual three-dimensional cubes from integral field spectrographs (e.g. for MUSE $\text{Ly}\alpha$

halos Leclercq *et al.*, 2020) could enable the study of the CGM's kinematics and ionization state.

Lyman-alpha intensity mapping and detection of the cosmic web

Line intensity mapping (LIM) measures spatial fluctuations of a line's luminosity without resolving individual emitters (Kovetz *et al.*, 2017; Kovetz *et al.*, 2019). Intensity mapping for Ly α enables us to use additional information of voxels that are individually below the signal-to-noise ratio to be identified as a source with confidence.

First measurements of Ly α intensity mapping have been performed using data sets from the Baryon Oscillation Spectroscopic Survey. After subtracting the best-fit model of low redshift luminous red galaxies, the residual flux at high wavelengths is cross-correlated with the positions of quasars at $z \sim 2.55$. The resulting signal is attributed to the Ly α emission around those quasars on large scales that had first been explained by emission from star-forming regions but later revised to primarily arise from emission associated with the quasars (Croft *et al.*, 2016, 2018). Additionally, the cross-correlations show distortions potentially related to the Ly α radiative transfer effects.

Intensity mapping has been explored theoretically for a range of emission lines (Schaan *et al.*, 2021a,b) with the goal of obtaining cosmological and astrophysical constraints and additional information beyond classical redshift surveys. Dedicated studies of Ly α intensity mapping into the epoch of reionization (Silva *et al.*, 2013, 2016) exist and have been complemented with Ly α radiative transfer run on top of semi-analytical reionization simulations (Visbal *et al.*, 2018). Nevertheless, an investigation on the grounds of the latest cosmological RHD simulations is lacking.

Moving beyond the circumgalactic medium that is traced by Ly α halos and blobs, the cosmic web that is already indirectly visible in absorption spectra of quasars might soon become visible through Ly α emission. A recent analysis of oriented LAH stacks by Gallego *et al.* (2018) was not able to detect a signal on intergalactic scales. Detection of filamentary structures on Megaparsec scale between large Ly α nebulae is challenging but appears achievable (Lusso *et al.*, 2019; Umehata *et al.*, 2019).

Recently Bacon *et al.* (2021) reported the discovery of Ly α filamentary structure on megaparsecs scale between redshifts $z = 3 - 4.5$. The authors argue that the majority of the filaments' flux does not seem to be associated with identified LAEs and cannot be explained by the UVB as ionizing sources. A favored explanation by the authors is the Ly α flux stemming from LAEs the sensitivity limits. In this case, the filaments would, in large parts, be powered by low luminosity emitters, i.e. the faint end of the luminosity. If

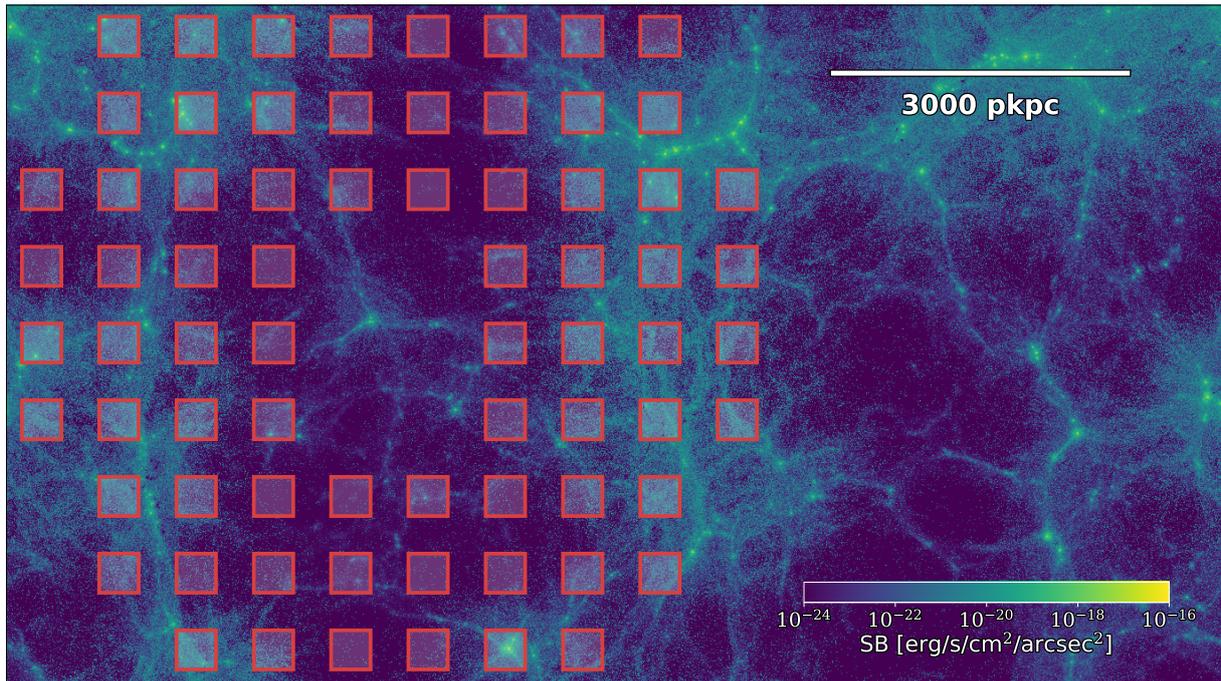


Fig. 7.4 The $\text{Ly}\alpha$ cosmic web as a radiative transfer simulation on top of TNG50 at $z = 3$. In addition, we show the HETDEX footprint for its 78 integral field units. With the TNG simulation suite, we can make predictions for the $\text{Ly}\alpha$ cosmic web’s observability.

true, this could provide constraints for the LAE luminosity function’s faint end, such as its slope and its cut-off.

Using the simulation setup from Chapter 6, we can explore the detectability of the $\text{Ly}\alpha$ cosmic web and the physical information that can be extracted from it. These explorations can be improved by calibrating the luminosities against the observed high luminosity LF end. In Figure 7.4, we show the HETDEX footprint on top of the TNG50 simulation, with which we can provide predictions for the detectability of the cosmic web with a dedicated HET survey.

Constraining the epoch of reionization

Astrophysicists long for the arrival of the James Webb Space Telescope (JWST) and the extremely large telescopes. With their arrival, large advances in the understanding of the epoch of reionization are expected. $\text{Ly}\alpha$ emission and absorption are to play a major role in these advances (Dunlop, 2013; Finkelstein *et al.*, 2019). Naturally, the $\text{Ly}\alpha$ observations’ interpretation is complicated, even more so given the more complex ionization state of the IGM. Applications to the epoch of reionization were primarily out of the scope of this thesis for two reasons. First, we would not be able to rely on the Illustris(TNG)

but would need a high-resolution radiation-hydrodynamics simulation to capture the IGM's evolution and progression of reionization. Second, the IGM's ionization state poses an additional complication of the $\text{Ly}\alpha$ radiative transfer. Instead, establishing a model at lower redshifts provides a stepping stone for future investigations into the epoch of reionization.

The work in Chapter 5 will be important to expand into the epoch of reionization to separate the impact from the IGM and the galaxy structure on the $\text{Ly}\alpha$ spectra. Both the galaxy structure and the IGM rapidly change in the EoR. Hence, disentangling the impact from galaxy structure and IGM will help to understand the EoR better. For example, the $\text{Ly}\alpha$ spectra contain information about the escape of Lyman-continuum photons from star-forming regions into the IGM (Chisholm *et al.*, 2018; Izotov *et al.*, 2018) driving (in part) the reionization.

7.1.3 Observational advances

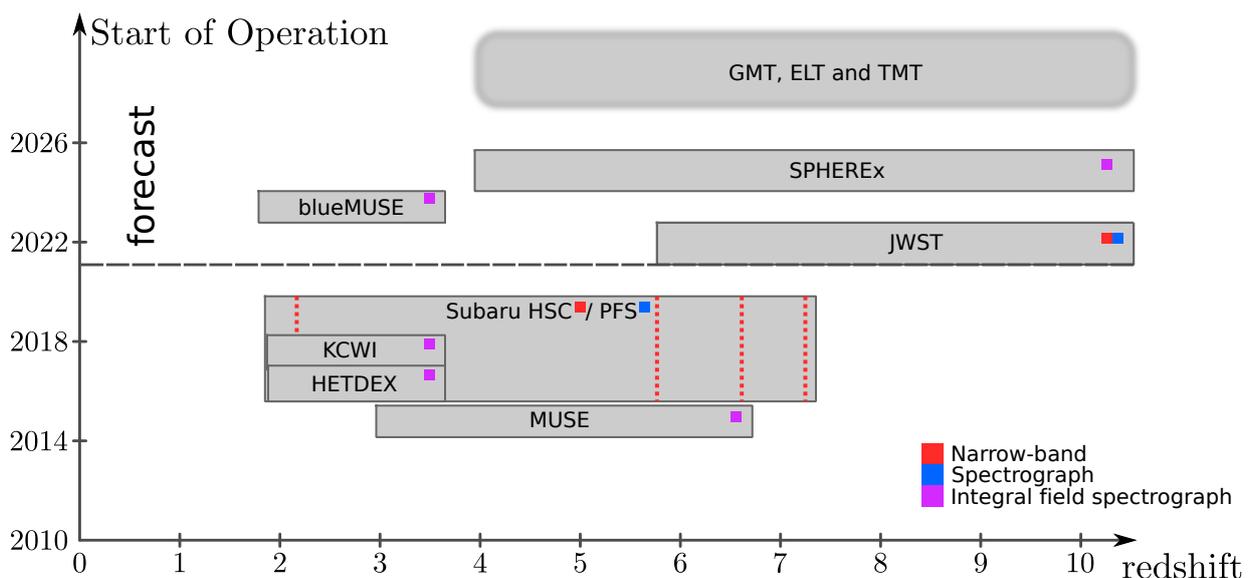


Fig. 7.5 Visualization of some recent and future instruments and telescopes for $\text{Ly}\alpha$ observations.

Many instruments are currently providing $\text{Ly}\alpha$ observations. Notably, data acquisition and reduction are underway for the HETDEX survey that will map out close to a million LAEs with spectroscopic redshift in a blind survey over the next years (Gebhardt *et al.*, 2021). MUSE (Bacon *et al.*, 2010) is providing yet another instrument beside HETDEX to map out $\text{Ly}\alpha$ emission with IFS at higher resolution with a smaller field of view. Soon, MUSE will be complemented by blueMUSE, enabling $\text{Ly}\alpha$ down to $z \geq 2$ instead of $z \geq 3$ facilitating observation of diffuse emission given the lower cosmological redshift

dimming. Lately, the KCWI has started its operation to map extended Ly α structures using IFS (Morrissey *et al.*, 2018; Cai *et al.*, 2019).

The HETDEX survey currently underway will allow an impressive first intensity mapping study with Ly α making use of the more than 500 square degrees to be spectroscopically mapped out. A low resolution Ly α intensity mapping survey at $z > 5.2$ can launch around 2025 with SPHEREx, which will be a promising cross-correlation probe to the 21cm-line signal (Doré *et al.*, 2016, 2018).

The James Webb Space Telescope will launch this fall, enabling us to observe the earliest formed galaxies in the Universe by their Ly α emission. Furthermore, at the end of this decade, multiple extremely large telescopes will come online to push to fainter Ly α sources than currently possible. See Figure 7.5 for current and upcoming instruments and telescopes that further improve sensitivities and quantity of Ly α observations. These extremely large telescopes with 25 to 38 m diameter will allow the observation of significantly fainter Ly α objects. Three of these telescopes, the Thirty Meter Telescope (TMT) in Hawaii, the Giant Magellan Telescope (GMT), and European Extremely Large Telescope (E-ELT) in Chile, are projected to see first light at the end of this decade. These telescopes and their respective instruments, along with JWST, will be able to push Ly α observations well into the epoch of reionization and yield novel constraints on the nature of these early Ly α emitting galaxies and the reionization's progression (Evans *et al.*, 2015; Sharp *et al.*, 2016; Finkelstein *et al.*, 2019).

7.2 Conclusions

In this thesis, we presented a Ly α radiative transfer code applied to galaxy formation simulations in cosmological volumes regarding different astrophysical and cosmological applications, namely the clustering of Ly α emitters, the encoded signature of the intergalactic medium in Ly α emitting galaxies' spectra, and the nature of Ly α halos:

- Beside previous studies of potential selection effects in LAE clustering analyses, we studied a new phenomenological effect originating from the shifted Ly α spectral signature for the line of sight localization. We find this new effect to be analogously modeled to the Fingers of God effect from the peculiar velocity field, albeit we find that scales of up to $k \sim 0.1 \text{ h Mpc}^{-1}$ can be affected. We presented multiple approaches to correct for such distortion (Chapter 4).
- We investigated the impact of the IGM on Ly α emitting galaxies' spectra. We propose going beyond the commonly used averaged transmission curves and stacked spectra. Using the distributions of new statistical measures such as the peak count

and peak asymmetry help to differentiate the redshift evolution from the evolution of the ISM of galaxies. We make our set of $\mathcal{O}(10^8)$ transmission curves available¹ for more dedicated studies (Chapter 5).

- The nature of Ly α halos remains debated since their first discovery. For the first time, we present the predicted signal from a statistical sample of star-forming galaxies for different halo masses over multiple orders of magnitude with resolutions down to ~ 100 pc. For intermediate redshifts also targeted by observational studies by e.g. MUSE and HETDEX, we find reasonable agreement of our fiducial emission model. Our radiative transfer simulations imply that extended Ly α halos are primarily sourced by scattering photons from the galaxy's inner star-forming regions (Chapter 6).

With the current and upcoming observations, we expect the quality and quantity of Ly α observations to grow substantially over the next years. The applications and advances in this thesis, combined with improvements in (radiation-)hydrodynamical simulations, will enable us to decode the physical information imprinted by the Ly α radiative transfer and thus to make Ly α radiation an even more important probe to study our cosmological and astrophysical models in the future.

¹A reduced data set has been published as Byrohl *et al.* (2020a) with the full set available on request.

REFERENCES

1. Adams, J. J., Blanc, G. A., Hill, G. J., *et al.* in *ApJ Supplement Series* **192**, 5.
2. Adams, T. F. in *ApJ* **168**, 575.
3. Adams, T. F. in *ApJ* **174**, 439.
4. Agrawal, A., Makiya, R., Chiang, C.-T., *et al.* in *JCAP* **10**, 003.
5. Ahn, S.-H. in *ApJ Letters* **601**, L25–L28.
6. Ahn, S.-H., Lee, H.-W. & Lee, H. M. in *JKAS* **33**, 29–36.
7. Ahn, S.-H., Lee, H.-W. & Lee, H. M. in *ApJ* **567**, 922–930.
8. Ahn, S.-H., Lee, H.-W. & Lee, H. M. in *MNRAS* **340**, 863–869.
9. Arrigoni Battaia, F., Hennawi, J. F., Cantalupo, S., *et al.* in *ApJ* **829**, 3.
10. Arrigoni Battaia, F., Hennawi, J. F., Prochaska, J. X., *et al.* in *ApJ* **809**, 163.
11. Arrigoni Battaia, F., Hennawi, J. F., Prochaska, J. X., *et al.* in *MNRAS* **482**, 3162–3205.
12. Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., *et al.* in *AJ* **156**, 123.
13. Avery, L. W. & House, L. L. in *ApJ* **152**, 493.
14. Bacon, R., Accardo, M., Adjali, L., *et al.* in **7735**, 773508.
15. Bacon, R., Brinchmann, J., Richard, J., *et al.* in *A&A* **575**, A75.
16. Bacon, R., Mary, D., Garel, T., *et al.* in *A&A* **647**, A107.
17. Bacon, R., Conseil, S., Mary, D., *et al.* in *A&A* **608**, A1.
18. Ballinger, W. E., Peacock, J. A. & Heavens, A. F. in *MNRAS* **282**, 877.
19. Balmer, J. J. in *Annalen der Physik* **261**, 80–87.
20. Barger, A. J., Cowie, L. L. & Wold, I. G. B. in *ApJ* **749**, 106.
21. Barkana, R. in *MNRAS, Volume 347, Issue 1, pp. 59–66.* **347**, 59–66.
22. Barnes, L. A., Garel, T. & Kacprzak, G. G. in *PASP* **126**, 969.
23. Behrens, C., Byrohl, C., Saito, S., *et al.* in *A&A* **614**, A31.
24. Behrens, C., Dijkstra, M. & Niemeyer, J. C. in *A&A* **563**, A77.
25. Behrens, C. & Niemeyer, J. in *A&A* **556**, A5.
26. Behrens, C., Pallottini, A., Ferrara, A., *et al.* in *MNRAS* **486**, 2197–2209.
27. Behrens, C. *The Influence of Small-Scale Anisotropies and the Large-Scale Environment on the Observed Properties of Lyman-Alpha Emitters* (Dec. 15, 2014).

28. Berestetskii, V. B., Lifshits, E. M. & Pitaevskii, L. P. *Relativistic Quantum Theory, Part 1* 1st edition. 616 pp. (Pergamon Press, Oxford, New York, Jan. 1, 1971).
29. Berg, D. A., Chisholm, J., Erb, D. K., *et al.* in *ApJ* **878**, L3.
30. Beutler, F., Saito, S., Seo, H.-J., *et al.* in *MNRAS* **443**, 1065–1089.
31. Bianchi, D., Percival, W. J. & Bel, J. in *MNRAS* **463**, 3783–3798.
32. Blanc, G. A., Adams, J. J., Gebhardt, K., *et al.* in *ApJ* **736**, 31.
33. Board, O. A. R. *OpenMP Application Programming Interface* 2020.
34. Bohr, N. in *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **26**, 1–25.
35. Bond, N. A., Gawiser, E., Gronwall, C., *et al.* in *ApJ* **705**, 639–649.
36. Bonilha, J. R. M., Ferch, R., Salpeter, E. E., *et al.* in *ApJ* **233**, 649–660.
37. Boogaard, L. A., Brinchmann, J., Bouché, N., *et al.* in *A&A* **619**, A27.
38. Borisova, E., Cantalupo, S., Lilly, S. J., *et al.* in *ApJ* **831**, 39.
39. Box, G. E. P. & Muller, M. E. in *The Annals of Mathematical Statistics* **29**, 610–611.
40. Burchett, J. N., Rubin, K. H. R., Prochaska, J. X., *et al.* in *ApJ* **909**, 151.
41. Busca, N. G., Delubac, T., Rich, J., *et al.* in *A&A* **552**, A96.
42. Byrohl, C. & Gronke, M. in *Zenodo*.
43. Byrohl, C. & Gronke, M. in *A&A* **642**, L16.
44. Byrohl, C., Nelson, D., Behrens, C., *et al.* in *MNRAS* **506**, 5129–5152.
45. Byrohl, C., Saito, S. & Behrens, C. in *MNRAS* **489**, 3472–3491.
46. Cai, Z., Cantalupo, S., Prochaska, J. X., *et al.* in *ApJ Supplement Series* **245**, 23.
47. Cai, Z., Fan, X., Yang, Y., *et al.* in *ApJ* **837**, 71.
48. Calzetti, D., Livio, M. & Madau, P. *Extragalactic Background Radiation* New edition. 307 pp. (Cambridge University Press, Cambridge ; New York, Jan. 26, 1995).
49. Camps, P., Behrens, C., Baes, M., *et al.* in *arXiv e-prints*, arXiv:2106.00281.
50. Cantalupo, S., Arrigoni-Battaia, F., Prochaska, J. X., *et al.* in *Nature* **506**, 63–66.
51. Cantalupo, S., Porciani, C. & Lilly, S. J. in *ApJ* **672**, 48–58.
52. Cantalupo, S., Porciani, C., Lilly, S. J., *et al.* in *ApJ* **628**, 61–75.
53. Cashwell, E. D. & Everett, C. J. *A Practical Manual on the Monte Carlo Method for Random Walk Problems* 172 pp. (London, 1959).
54. Cen, R. in *ApJ Supplement Series* **78**, 341–364.
55. Cen, R. & Haiman, Z. in *ApJ* **542**, L75–L78.
56. Chabrier, G. in *PASP* **115**, 763–795.
57. Chiani, M., Member, S., Dardari, D., *et al.* in *IEEE Transactions on Wireless Communications*, 840–845.
58. Chisholm, J., Gazagnes, S., Schaerer, D., *et al.* in *A&A* **616**, A30.
59. Claeysens, A., Richard, J., Blaizot, J., *et al.* in *MNRAS* **489**, 5022–5029.

60. Collaboration, A., Robitaille, T. P., Tollerud, E. J., *et al.* in *A&A* **558**, A33.
61. Collette, A. *Python and HDF5* (OReilly, 2013).
62. Cowie, L. L., Barger, A. J. & Hu, E. M. in *ApJ* **711**, 928–958.
63. Cowie, L. L., Barger, A. J. & Hu, E. M. in *ApJ* **738**, 136.
64. Croft, R. A. C., Miralda-Escudé, J., Zheng, Z., *et al.* in *MNRAS* **457**, 3541–3572.
65. Croft, R. A. C., Miralda-Escudé, J., Zheng, Z., *et al.* in *MNRAS* **481**, 1320–1336.
66. Dask Development Team. *Dask: Library for dynamic task scheduling* (2016).
67. Davé, R., Anglés-Alcázar, D., Narayanan, D., *et al.* in *MNRAS* **486**, 2827–2849.
68. Davis, M. & Peebles, P. J. E. in *ApJ* **267**, 465–482.
69. DESI Collaboration, Aghamousa, A., Aguilar, J., *et al.* in *arXiv e-prints*, arXiv:1611.00036.
70. Diemer, B., Stevens, A. R. H., Lagos, C. d. P., *et al.* in *MNRAS* **487**, 1529–1550.
71. Dijkstra, M. in *PASA, Volume 31, id.e040 26 pp.* **31**.
72. Dijkstra, M. in *arXiv e-prints* **1704**, arXiv:1704.03416.
73. Dijkstra, M. in *SaaS-Fee Advanced Course, Springer-Verlag* (eds Dijkstra, M., Prochaska, J. X., Ouchi, M., *et al.*) 1–109 (Springer, Berlin, Heidelberg, 2019).
74. Dijkstra, M., Gronke, M. & Venkatesan, A. in *ApJ* **828**, 71.
75. Dijkstra, M., Haiman, Z. & Spaans, M. in *ApJ* **649**, 14–36.
76. Doré, O., Werner, M. W., Ashby, M., *et al.* in *arXiv e-prints*.
77. Doré, O., Werner, M. W., Ashby, M. L. N., *et al.* in *arXiv e-prints*.
78. Draine, B. T. *Physics of the Interstellar and Intergalactic Medium* (Princeton University Press, Jan. 1, 2011).
79. Dubois, Y., Peirani, S., Pichon, C., *et al.* in *MNRAS* **463**, 3948–3964.
80. Dunlop, J. S. in *The First Galaxies* **396**, 223.
81. Eide, M. B., Gronke, M., Dijkstra, M., *et al.* in *ApJ* **856**, 156.
82. Elias, L. M., Genel, S., Sternberg, A., *et al.* in *MNRAS* **494**, 5439–5448.
83. Emonts, B. H. C., Cai, Z., Prochaska, J. X., *et al.* in *ApJ* **887**, 86.
84. Erb, D. K., Steidel, C. C., Trainor, R. F., *et al.* in *ApJ* **795**, 33.
85. Evans, C., Puech, M., Afonso, J., *et al.* in *arXiv e-prints*.
86. Fardal, M. A., Katz, N., Gardner, J. P., *et al.* in *ApJ* **562**, 605–617.
87. Farina, E. P., Arrigoni-Battaia, F., Costa, T., *et al.* in *ApJ* **887**, 196.
88. Farrow, D. J., Sánchez, A. G., Ciardullo, R., *et al.* in *MNRAS*.
89. Faucher-Giguère, C.-A., Kereš, D., Dijkstra, M., *et al.* in *ApJ* **725**, 633–657.
90. Faucher-Giguère, C.-A., Lidz, A., Zaldarriaga, M., *et al.* in *ApJ* **703**, 1416–1443.
91. Feldmeier, J. J., Hagen, A., Ciardullo, R., *et al.* in *ApJ* **776**, 75.
92. Feltre, A., Maseda, M. V., Bacon, R., *et al.* in *A&A* **641**, A118.
93. Finkelstein, S., Bradac, M., Casey, C., *et al.* in *BAAS* **51**, 221.

94. Finkelstein, S. L., Cohen, S. H., Malhotra, S., *et al.* in *ApJ* **700**, 276–283.
95. Finley, H., Bouché, N., Contini, T., *et al.* in *A&A* **605**, A118.
96. Finley, H., Bouché, N., Contini, T., *et al.* in *A&A* **608**, A7.
97. Fisher, K. B. in *ApJ* **448**, 494.
98. Fitzpatrick, R. *Quantum Mechanics* (2008).
99. Forum, M. P. I. in , 1139.
100. Furlanetto, S. R., Schaye, J., Springel, V., *et al.* in *ApJ* **622**, 7–27.
101. Gallego, S. G., Cantalupo, S., Lilly, S., *et al.* in *MNRAS* **475**, 3854–3869.
102. Gebhardt, K., Cooper, E. M., Ciardullo, R., *et al.* in *preprint*.
103. Genel, S., Vogelsberger, M., Springel, V., *et al.* in *MNRAS* **445**, 175–200.
104. Gnedin, N. Y., Becker, G. D. & Fan, X. in *ApJ* **841**, 26.
105. Gould, A. & Weinberg, D. H. in *ApJ* **468**, 462.
106. Griffiths, D. J. & Schroeter, D. F. *Introduction to Quantum Mechanics* Third edition. 495 pp. (Cambridge University Press, Cambridge ; New York, NY, 2018).
107. Gronke, M., Bull, P. & Dijkstra, M. in *ApJ* **812**, 123.
108. Gronke, M. & Dijkstra, M. in *MNRAS* **444**, 1095–1103.
109. Gronke, M. & Dijkstra, M. in *ApJ* **826**, 14.
110. Gronke, M. in *A&A* **608**, A139.
111. Gronke, M. & Bird, S. in *ApJ* **835**, 207.
112. Gronke, M., Dijkstra, M., McCourt, M., *et al.* in *ApJ Letters* **833**, L26.
113. Gronke, M., Dijkstra, M., McCourt, M., *et al.* in *A&A* **607**, A71.
114. Gronke, M., Girichidis, P., Naab, T., *et al.* in *ApJ Letters* **862**, L7.
115. Guo, Y., Maiolino, R., Jiang, L., *et al.* in *ApJ* **898**, 26.
116. Gurung-López, S., Orsi, Á. A., Bonoli, S., *et al.* in *MNRAS* **486**, 1882–1906.
117. Gurung-López, S., Orsi, Á. A., Bonoli, S., *et al.* in *MNRAS* **491**, 3266–3289.
118. Gurung-López, S., Saito, S., Baugh, C. M., *et al.* in *MNRAS* **500**, 603–626.
119. Haiman, Z., Spaans, M. & Quataert, E. in *ApJ Letters* **537**, L5–L8.
120. Hamilton, A. J. S. in *Ringberg Workshop on LSS* **231**, 185.
121. Hamilton, D. R. in *Physical Review* **58**, 122–131.
122. Hansen, M. & Oh, S. P. in *MNRAS* **367**, 979–1002.
123. Harrington, J. P. in *MNRAS* **162**, 43.
124. Harris, C. R., Millman, K. J., van der Walt, S. J., *et al.* in *Nature* **585**, 357–362.
125. Hayashino, T., Matsuda, Y., Tamura, H., *et al.* in *AJ* **128**, 2073–2079.
126. Hayes, M. in *PASA* **32**, e027.
127. Hayes, M., Östlin, G., Duval, F., *et al.* in *ApJ* **782**, 6.
128. Hayes, M., Östlin, G., Schaerer, D., *et al.* in *ApJ Letters* **765**, L27.

129. Hearin, A. P., Campbell, D., Tollerud, E., *et al.* in *AJ* **154**, 190.
130. Heckman, T. M., Lehnert, M. D., Miley, G. K., *et al.* in *ApJ* **381**, 373–385.
131. Heckman, T. M., Lehnert, M. D., van Breugel, W., *et al.* in *ApJ* **370**, 78.
132. Henry, A., Scarlata, C., Martin, C. L., *et al.* in *ApJ* **809**, 19.
133. Henyey, L. G. & Greenstein, J. L. in *ApJ* **93**, 70–83.
134. Herenz, E. C., Urrutia, T., Wisotzki, L., *et al.* in *A&A* **606**, A12.
135. Hickox, R. C. & Alexander, D. M. in *Annual Review of A&A* **56**, 625–671.
136. Hikage, C. & Yamamoto, K. in *MNRAS* **455**, L77–L81.
137. Hilbert, D. in *Mathematische Annalen* **38**, 460–460.
138. Hill, G. J., Gebhardt, K., Komatsu, E., *et al.* in **399**, 115.
139. Hill, G. J., Gebhardt, K., Komatsu, E., *et al.* in **743**, 224–233.
140. Hinshaw, G., Larson, D., Komatsu, E., *et al.* in *ApJ Supplement Series* **208**, 19.
141. Hopkins, P. F., Richards, G. T. & Hernquist, L. in *ApJ* **654**, 731–753.
142. Hu, E. M., Cowie, L. L. & McMahon, R. G. in *ApJ* **502**, L99–L103.
143. Hu, E. M. & McMahon, R. G. in *Nature* **382**, 231–233.
144. Hummer, D. G. & Storey, P. J. in *MNRAS* **224**, 801–820.
145. Hunter, J. D. in *Computing in Science & Engineering* **9**, 90–95.
146. Iliev, I. T., Shapiro, P. R., McDonald, P., *et al.* in *MNRAS* **391**, 63–83.
147. Inoue, A. K., Hasegawa, K., Ishiyama, T., *et al.* in *PASJ* **70**, 55.
148. Izotov, Y. I., Worseck, G., Schaerer, D., *et al.* in *MNRAS* **478**, 4851–4865.
149. Jackson, J. C. in *MNRAS* **156**, 1P.
150. Kaiser, N. in *MNRAS* **227**, 1–21.
151. Kakiichi, K., Dijkstra, M., Ciardi, B., *et al.* in *MNRAS* **463**, 4019–4040.
152. Kakuma, R., Ouchi, M., Harikane, Y., *et al.* in *ApJ* **916**, 22.
153. Katz, N., Weinberg, D. H. & Hernquist, L. in *ApJ Supplement Series* **105**, 19.
154. Kennicutt Jr., R. C. in *ApJ* **498**, 541–552.
155. Kikuchihara, S., Harikane, Y., Ouchi, M., *et al.* in *arXiv e-prints*.
156. Kimock, B., Narayanan, D., Smith, A., *et al.* in *arXiv e-prints*, arXiv:2004.08397.
157. Kollmeier, J. A., Zheng, Z., Davé, R., *et al.* in *ApJ* **708**, 1048–1075.
158. Konno, A., Ouchi, M., Nakajima, K., *et al.* in *ApJ* **823**, 20.
159. Konno, A., Ouchi, M., Ono, Y., *et al.* in *ApJ* **797**, 16.
160. Konno, A., Ouchi, M., Shibuya, T., *et al.* in *PASJ* **70**, S16.
161. Kovetz, E., Breyse, P. C., Lidz, A., *et al.* in *BAAS* **51**, 101.
162. Kovetz, E. D., Viero, M. P., Lidz, A., *et al.* in *arXiv e-prints*, arXiv:1709.09066.
163. Kulas, K. R., Shapley, A. E., Kollmeier, J. A., *et al.* in *ApJ* **745**, 33.
164. Kunth, D., Mas-Hesse, J. M., Terlevich, E., *et al.* in *A&A* **334**, 11–20.

165. Lake, E., Zheng, Z., Cen, R., *et al.* in *ApJ* **806**, 46.
166. Lam, S. K., Pitrou, A. & Seibert, S. in *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC* (Association for Computing Machinery, New York, NY, USA, Nov. 15, 2015), 1–6.
167. Landy, S. D. & Szalay, A. S. in *ApJ* **412**, 64–71.
168. Lao, B.-X. & Smith, A. in *MNRAS* **497**, 3925–3942.
169. Laureijs, R., Amiaux, J., Arduini, S., *et al.* in *arXiv e-prints*, arXiv:1110.3193.
170. Laursen, P. in *arXiv e-prints* **1012**, arXiv:1012.2886.
171. Laursen, P. *Interpreting Lyman α Radiation from Young, Dusty Galaxies* (Aug. 1, 2010).
172. Laursen, P., Sommer-Larsen, J. & Andersen, A. C. in *ApJ* **704**, 1640–1656.
173. Laursen, P., Sommer-Larsen, J. & Razoumov, A. O. in *ApJ* **728**, 52.
174. Leclercq, F., Bacon, R., Verhamme, A., *et al.* in *A&A* **635**, A82.
175. Leclercq, F., Bacon, R., Wisotzki, L., *et al.* in *A&A* **608**, A8.
176. Lee, K.-G., Hennawi, J. F., Stark, C., *et al.* in *ApJ* **795**, L12.
177. Lee, K.-G., Krolewski, A., White, M., *et al.* in *ApJ Supplement Series* **237**, 31.
178. Leung, A. S., Acquaviva, V., Gawiser, E., *et al.* in *ApJ* **843**, 130.
179. Li, A. & Draine, B. T. in *ApJ* **554**, 778–802.
180. Loeb, A. & Rybicki, G. B. in *ApJ* **524**, 527–535.
181. Lusso, E., Fumagalli, M., Fossati, M., *et al.* in *MNRAS* **485**, L62–L67.
182. Lyman, T. in *ApJ* **23**, 181.
183. Marinacci, F., Vogelsberger, M., Pakmor, R., *et al.* in *MNRAS* **480**, 5113–5139.
184. Marques-Chaves, R., Pérez-Fournon, I., Villar-Martín, M., *et al.* in *A&A* **629**, A23.
185. Martin, D. C., O’Sullivan, D., Matuszewski, M., *et al.* in *Nature Astronomy* **3**, 822–831.
186. Mas-Ribas, L. & Dijkstra, M. in *ApJ* **822**, 84.
187. Mas-Ribas, L., Dijkstra, M., Hennawi, J. F., *et al.* in *ApJ* **841**, 19.
188. Mason, C. A., Fontana, A., Treu, T., *et al.* in *MNRAS* **485**, 3947–3969.
189. Matsubara, T. in *Physical Review D* **90**, 043537.
190. Matsuda, Y., Yamada, T., Hayashino, T., *et al.* in *MNRAS* **425**, 878–883.
191. Matthee, J., Pezzulli, G., Mackenzie, R., *et al.* in *MNRAS* **498**, 3043–3059.
192. Matthee, J., Sobral, D., Darvish, B., *et al.* in *MNRAS* **472**, 772–787.
193. McCarthy, P. J., Spinrad, H., Djorgovski, S., *et al.* in *ApJ Letters* **319**, L39–L44.
194. McCourt, M., Oh, S. P., O’Leary, R., *et al.* in *MNRAS* **473**, 5407–5431.
195. McKinney, W. in *Proceedings of the 9th Python in Science Conference* (eds van der Walt, S. & Millman, J.) (2010), 56–61.
196. Michel-Dansac, L., Blaizot, J., Garel, T., *et al.* in *A&A* **635**, A154.
197. Mitchell, P. D., Blaizot, J., Cadiou, C., *et al.* in *MNRAS* **501**, 5757–5775.

198. Mo, H., Van den Bosch, F. & White, S. *Galaxy Formation and Evolution* 820 pp. (Cambridge University Press, Cambridge ; New York, 2010).
199. Momose, R., Ouchi, M., Nakajima, K., *et al.* in *MNRAS* **442**, 110–120.
200. Momose, R., Ouchi, M., Nakajima, K., *et al.* in *MNRAS* **457**, 2318–2330.
201. Morrissey, P., Matuszewski, M., Martin, D. C., *et al.* in *ApJ* **864**, 93.
202. Mortlock, D. in *Understanding the Epoch of Cosmic Reionization* **423**, 187.
203. Mukae, S., Ouchi, M., Hill, G. J., *et al.* in *ApJ* **903**, 24.
204. Nagamine, K., Ouchi, M., Springel, V., *et al.* in *PASJ* **62**, 1455–1472.
205. Naiman, J. P., Pillepich, A., Springel, V., *et al.* in *MNRAS* **477**, 1206–1224.
206. Navarro, J. F., Frenk, C. S. & White, S. D. M. in *ApJ* **490**, 493–508.
207. Nelson, D., Byrohl, C., Peroux, C., *et al.* in *arXiv e-prints*, arXiv:2106.09023.
208. Nelson, D., Pillepich, A., Springel, V., *et al.* in *MNRAS* **475**, 624–647.
209. Nelson, D., Pillepich, A., Springel, V., *et al.* in *MNRAS* **490**, 3234–3261.
210. Nelson, D., Springel, V., Pillepich, A., *et al.* in *Computational Astrophysics and Cosmology* **6**, 2.
211. Neufeld, D. A. in *Astrophysical Journal* **350**, 216–241.
212. Neufeld, D. A. in *ApJ Letters* **370**, L85–L88.
213. Neufeld, D. A. & McKee, C. F. in *ApJ* **331**, L87.
214. Niemeyer, M. L. *Lyman- α Halos around High-Redshift Lyman- α Emitting Galaxies* (2021).
215. O’Sullivan, D. B., Martin, C., Matuszewski, M., *et al.* in *ApJ* **894**, 3.
216. Ocvirk, P., Aubert, D., Sorce, J. G., *et al.* in *MNRAS* **496**, 4087–4107.
217. Okumura, T., Seljak, U. & Desjacques, V. in *JCAP* **11**, 014.
218. Okuta, R., Unno, Y., Nishino, D., *et al.* in *Proceedings of Workshop on Machine LearningSys in The Thirty-first Annual Conference on NIPS* (2017).
219. Osterbrock, D. E. in *ApJ* **135**, 195.
220. Östlin, G., Hayes, M., Duval, F., *et al.* in *ApJ* **797**, 11.
221. Ouchi, M. in *SaaS-Fee Advanced Course, Springer-Verlag* **46**, 189.
222. Ouchi, M., Ono, Y., Egami, E., *et al.* in *ApJ* **696**, 1164–1175.
223. Ouchi, M., Shimasaku, K., Akiyama, M., *et al.* in *ApJ Supplement Series* **176**, 301–330.
224. Ouchi, M., Shimasaku, K., Furusawa, H., *et al.* in *ApJ* **582**, 60–68.
225. Pakmor, R. & Springel, V. in *MNRAS* **432**, 176–193.
226. Pakmor, R., Bauer, A. & Springel, V. in *MNRAS* **418**, 1392–1401.
227. Pandas Development Team. *pandas-dev/pandas: Pandas version latest*. Feb. 2020.
228. Partridge, R. B. & Peebles, P. J. E. in *ApJ* **147**, 868.
229. Paulino-Afonso, A., Sobral, D., Ribeiro, B., *et al.* in *MNRAS* **476**, 5479–5501.
230. Peano. in *Mathematische Annalen* **36**, 157–160.

231. Pérez, F. & Granger, B. E. in *Computing in Science and Engineering* **9**, 21–29.
232. Pillepich, A., Nelson, D., Hernquist, L., *et al.* in *MNRAS* **475**, 648–675.
233. Pillepich, A., Nelson, D., Springel, V., *et al.* in *MNRAS* **490**, 3196–3233.
234. Pillepich, A., Springel, V., Nelson, D., *et al.* in *MNRAS* **473**, 4077–4106.
235. Planck Collaboration, Adam, R., Ade, P. A. R., *et al.* in *A&A* **594**, A1.
236. *Numerical Recipes: The Art of Scientific Computing* 3rd ed (ed Press, W. H.) 1235 pp. (Cambridge University Press, Cambridge, UK ; New York, 2007).
237. Prochaska, J. X. in *SaaS-Fee Advanced Course, Springer-Verlag* (eds Dijkstra, M., Prochaska, J. X., Ouchi, M., *et al.*) 111–188 (Springer, Berlin, Heidelberg, 2019).
238. Prochaska, J. X., Kasen, D. & Rubin, K. in *ApJ* **734**, 24.
239. Rahmati, A., Pawlik, A. H., Raičević, M., *et al.* in *MNRAS* **430**, 2427–2445.
240. Reid, B. A. & White, M. in *MNRAS* **417**, 1913–1927.
241. Rivera-Thorsen, T. E., Dahle, H., Gronke, M., *et al.* in *A&A* **608**, L4.
242. Rosdahl, J. & Blaizot, J. in *MNRAS* **423**, 344–366.
243. Rosdahl, J., Katz, H., Blaizot, J., *et al.* in *MNRAS* **479**, 994–1016.
244. Runnholm, A., Hayes, M., Melinder, J., *et al.* in *ApJ* **892**, 48.
245. Rupke, D. S. N., Coil, A., Geach, J. E., *et al.* in *Nature* **574**, 643–646.
246. Rybicki, G. B. & Lightman, A. P. *Radiative Processes in Astrophysics* 382 pp. (Wiley, Weinheim, 2004).
247. Rycroft, C. H. in *Chaos* **19**, 041111.
248. Rydberg, J. R. in *Kungliga Vetenskapsakademien Band 23 No 11*.
249. Santos, S., Sobral, D., Matthee, J., *et al.* in *MNRAS* **493**, 141–160.
250. Schaan, E. & White, M. in *JCAP* **2021**, 067.
251. Schaan, E. & White, M. in *JCAP* **2021**, 068.
252. Schaye, J., Crain, R. A., Bower, R. G., *et al.* in *MNRAS* **446**, 521–554.
253. Scholz, T. T. & Walters, H. R. J. in *ApJ* **380**, 302–306.
254. Scholz, T. T., Walters, H. R. J., Burke, P. J., *et al.* in *MNRAS* **242**, 692–697.
255. Scoccimarro, R. in *Physical Review D* **70**, 083007.
256. Sefusatti, E., Crocce, M., Scoccimarro, R., *et al.* in *MNRAS* **460**, 3624–3636.
257. Sharp, R., Bloxham, G., Boz, R., *et al.* in **9908**, 99081Y.
258. Sijacki, D., Vogelsberger, M., Genel, S., *et al.* in *MNRAS* **452**, 575–596.
259. Silva, M. B., Kooistra, R. & Zaroubi, S. in *MNRAS* **462**, 1961–1971.
260. Silva, M. B., Santos, M. G., Gong, Y., *et al.* in *ApJ* **763**, 132.
261. Smith, A., Kannan, R., Tsang, B. T.-H., *et al.* in *ApJ* **905**, 27.
262. Smith, A., Ma, X., Bromm, V., *et al.* in *MNRAS* **484**, 39–59.
263. Smith, A., Safranek-Shrader, C., Bromm, V., *et al.* in *MNRAS* **449**, 4336–4362.
264. Smith, A., Tsang, B. T.-H., Bromm, V., *et al.* in *MNRAS* **479**, 2065–2078.

265. Sobral, D. & Matthee, J. in *A&A* **623**, A157.
266. Sobral, D., Matthee, J., Darvish, B., *et al.* in *MNRAS* **477**, 2817–2840.
267. Sommerfeld, A. in *Annalen der Physik* **356**, 1–94.
268. Sommerfeld, A. in.
269. Song, H., Seon, K.-I. & Hwang, H. S. in *ApJ* **901**, 41.
270. Spergel, D., Gehrels, N., Breckinridge, J., *et al.* in *arXiv e-prints*, arXiv:1305.5422.
271. Spinoso, D., Orsi, A., López-Sanjuan, C., *et al.* in *A&A* **643**, A149.
272. Spitzer, L. *Physical Processes in the Interstellar Medium* (Jan. 1, 1978).
273. Spitzer Jr., L. & Greenstein, J. L. in *ApJ* **114**, 407.
274. Springel, V. in *MNRAS* **364**, 1105–1134.
275. Springel, V. in *MNRAS* **401**, 791–851.
276. Springel, V. & Hernquist, L. in *MNRAS* **339**, 289–311.
277. Springel, V., Pakmor, R., Pillepich, A., *et al.* in *MNRAS* **475**, 676–698.
278. Stanway, E. R. & Eldridge, J. J. in *MNRAS* **479**, 75–93.
279. Steidel, C. C., Adelberger, K. L., Shapley, A. E., *et al.* in *ApJ* **532**, 170–182.
280. Steidel, C. C., Bogosavljević, M., Shapley, A. E., *et al.* in *ApJ* **736**, 160.
281. Stenflo, J. O. in *A&A*, vol. 84, no. 1-2, Apr. 1980, p. 68-74. **84**, 68.
282. Takada, M., Ellis, R. S., Chiba, M., *et al.* in *PASJ* **66**, R1.
283. Taruya, A., Nishimichi, T. & Saito, S. in *Physical Review D* **82**, 063522.
284. Tasitsiomi, A. in *ApJ* **645**, 792–813.
285. Tasitsiomi, A. in *ApJ* **648**, 762–766.
286. Torrey, P., Vogelsberger, M., Genel, S., *et al.* in *MNRAS* **438**, 1985–2004.
287. Trainor, R. F., Steidel, C. C., Strom, A. L., *et al.* in *ApJ* **809**, 89.
288. Trebitsch, M., Verhamme, A., Blaizot, J., *et al.* in *A&A* **593**, A122.
289. Tumlinson, J., Peebles, M. S. & Werk, J. K. in *Annual Review of A&A* **55**, 389–432.
290. Uhlemann, C., Kopp, M. & Haugg, T. in *Physical Review D* **92**, 063004.
291. Umehata, H., Fumagalli, M., Smail, I., *et al.* in *Science* **366**, 97–100.
292. Vanzella, E., Caminha, G. B., Calura, F., *et al.* in *MNRAS* **491**, 1093–1103.
293. Verhamme, A., Garel, T., Ventou, E., *et al.* in *MNRAS* **478**, L60–L65.
294. Verhamme, A., Orlitová, I., Schaerer, D., *et al.* in *A&A* **597**, A13.
295. Verhamme, A., Schaerer, D., Atek, H., *et al.* in *A&A* **491**, 89–111.
296. Verhamme, A., Schaerer, D. & Maselli, A. in *A&A* **460**, 397–413.
297. Visbal, E. & McQuinn, M. in *ApJ* **863**, L6.
298. Vlah, Z. & White, M. in *JCAP* **03**, 007.
299. Vogelsberger, M., Genel, S., Springel, V., *et al.* in *Nature* **509**, 177–182.
300. Vogelsberger, M., Genel, S., Sijacki, D., *et al.* in *MNRAS* **436**, 3031–3067.

301. Vogelsberger, M., Genel, S., Springel, V., *et al.* in *MNRAS* **444**, 1518–1547.
302. Voronoi, G. in *Journal für die reine und angewandte Mathematik* **133**, 97–178.
303. Weinberger, R., Springel, V., Hernquist, L., *et al.* in *MNRAS* **465**, 3291–3308.
304. Weinberger, R., Springel, V., Pakmor, R., *et al.* in *MNRAS* **479**, 4056–4072.
305. Weingartner, J. C. & Draine, B. T. in *ApJ* **548**, 296–309.
306. Weiss, L. H., Bowman, W. P., Ciardullo, R., *et al.* in *ApJ* **912**, 100.
307. White, M., Song, Y.-S. & Percival, W. J. in *MNRAS* **397**, 1348–1354.
308. Whitney, B. A. in *Bulletin of the Astronomical Society of India* **39**, 101–127.
309. Wiese, W. L. & Fuhr, J. R. in *Journal of Physical and Chemical Reference Data* **38**, 565–720.
310. Wisotzki, L., Bacon, R., Blaizot, J., *et al.* in *A&A* **587**, A98.
311. Wisotzki, L., Bacon, R., Brinchmann, J., *et al.* in *Nature* **562**, 229–232.
312. Witstok, J., Puchwein, E., Kulkarni, G., *et al.* in *A&A* **650**, A98.
313. Wu, J., Jiang, L. & Ning, Y. in *ApJ* **891**, 105.
314. Wyithe, J. S. B. & Dijkstra, M. in *MNRAS* **415**, 3929–3950.
315. Yamada, T., Nakamura, Y., Matsuda, Y., *et al.* in *AJ* **143**, 79.
316. Yang, H., Malhotra, S., Gronke, M., *et al.* in *ApJ* **820**, 130.
317. Yusef-Zadeh, F., Morris, M. & White, R. L. in *ApJ* **278**, 186–194.
318. Zabl, J., Bouché, N. F., Wisotzki, L., *et al.* in *MNRAS*.
319. Zanstra, H. in *Bulletin of the Astronomical Institutes of the Netherlands* **11**, 1.
320. Zheng, Y. & Song, Y.-S. in *JCAP* **08**, 050.
321. Zheng, Y., Song, Y.-S. & Oh, M. in *JCAP* **06**, 013.
322. Zheng, Z., Cen, R., Trac, H., *et al.* in *ApJ* **716**, 574–598.
323. Zheng, Z., Cen, R., Trac, H., *et al.* in *ApJ* **726**, 38.
324. Zheng, Z., Cen, R., Weinberg, D., *et al.* in *ApJ* **739**, 62.
325. Zheng, Z. & Miralda-Escudé, J. in *ApJ* **578**, 33–42.
326. Zheng, Z. & Wallace, J. in *ApJ* **794**, 116.