UNIVERSITY COLLEGE LONDON

Faculty of Mathematics and Physical Sciences

Department of Physics & Astronomy

THE ROLE OF GALAXIES AND QUASARS IN REIONISING THE HIGH-REDSHIFT INTERGALACTIC MEDIUM

Thesis submitted for the Degree of

Doctor of Philosophy

by

Romain A. Meyer

Supervisors:

Prof. Richard S. Ellis

Prof. Andrew Pontzen

Examiners:

Prof. Max Pettini

Dr. Thomas Greve

October 12, 2020

A ma famille, à mon cabri,

and to the pursuit of pure knowledge

I, Romain A. Meyer, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Only one billion years after the Big Bang, the neutral hydrogen in the intergalactic medium had been completely ionised. This last phase transition of the Universe known as the epoch of reionisation is one of the frontiers in astrophysics. Despite our growing knowledge on the timing and topology of cosmic reionisation, the sources responsible for emitting the necessary ionising photons have remained elusive. Specifically, the escape fraction of ionising photons in reionisation-era galaxies and the role of quasars remain open and debated questions. This Thesis aims to answer these questions in order to understand the nature of the sources of reionisation.

Firstly, I present the discovery of a luminous galaxy whose double-peak Lyman-α emission profile indicates an escape fraction close to 100%. I show how this galaxy is the first evidence of an object self-ionising its own H II bubble deep into the reionisation era. Secondly, I measure the cross-correlation of $z \sim 5-6$ galaxies and metal absorbers with the IGM opacity to Lyman-α probed by high-redshift quasars. I extend the analytical model of the galaxy-IGM cross-correlation to derive average escape fractions for faint galaxies in the reionisation era. Thirdly, I investigate the evolution of quasars with redshift by measuring the relative offsets of broad emission lines in four hundred quasars at $1 < z < 7$. I discuss how quasar evolution and selection biases can explain the increased blueshift of the triply-ionised carbon (C IV) quasar broad emission line in the first billion years. I then present the first results of a programme to detect missing lensed $z \sim 6$ quasars. Finally, this Thesis concludes on the combination of the different results into a coherent picture of the nature of the sources of reionisation and prospects for future instruments and surveys.

Impact statement

Since the dawn of humankind, individuals and societies alike have looked at the sky in a bid to answer a most profound question: "Where do we come from?". With future telescopes, we will be able to observe the light from the first ever galaxies in which all elements heavier than helium were produced. Until then, we can study these first galaxies by looking at the impact they had on the intergalactic medium. This Thesis used state-of-the art deep spectroscopic data from the largest telescopes in the world to understand the fundamental role of galaxies and quasars in ionising the intergalactic medium in the first billion years of the Universe.

The evolution of quasars has been questioned for a long time due to scarce evidence and differing methods between studies. By studying hundreds of quasar spectra over six billion years of cosmic history, I demonstrate a significant difference between samples of early and late quasars. The escape fraction of high-redshift galaxies was deemed a fabled goal of reionisation studies. This Thesis has demonstrated that such escape fractions can be measured statistically, providing the first average measurement of the ionising properties of faint galaxies in the first billion years of the Universe. Meanwhile, the discovery of a galaxy with 100% escape fraction has confirmed the longtime unverified hypothesis that early galaxies could have extreme fractions in contrast to that in the present Universe. The fascination of the general public for this kind of fundamental research is exemplified by a press release on the work presented in Chapter 2 during the EAS 2020 June meeting that was reported on in various astronomy websites for the general audience.

This Thesis has achieved measurable academic impact through the publication of three

papers in the Monthly Notices of the Royal Astronomical Society which have altogether already been cited more than thirty times. A fourth paper is being peer-reviewed at the time of writing. The work produced in this Thesis was presented to more than a dozen scientific meetings with audiences ranging from forty to hundreds. Less than two years after publication, the new cross-correlation method developed in this work has been integrated as a core goal of large international observing programmes, and has inspired theorists to improve simulations which did not reproduce the observed signal. The work presented in Chapter 5 has been instrumental in preparing observations for a high-redshift quasar survey with ALMA and probably prevented tens of hours being lost on one of the most expensive ground-based telescope ever built.

Acknowledgements

I would like to thank my supervisor, Richard Ellis, for his perpetual enthusiasm and support in my endeavour to understand the reionisation era. He encouraged me, pushed me and showed me when I needed it. He has been an amazing mentor, teaching me what constitutes good science, how to write and communicate it, and above all sharing his passion and excitement when observing at the largest telescopes. I thank my second supervisor, Andrew Pontzen, for being a role model as a progressive scientist, committed to change and reform academia for the best.

I would like to thank the amazing trio of postdocs who mentored me in the First Light group since I arrived. Koki Kakiichi, for patiently explaining complex analytical work and rekindling a long-extinguished flame for theory in me, and his intransigence in matters of graph-making which has had a lasting imprint. Sarah Bosman, for saving the day with an enormous trove of semi-public reduced data when observations went wrong, and introducing me to the wild world of high-redshift quasars and those who study them. Finally, Nicolas Laporte, for giving me "les yeux de la foi" and his insightful and humorous perspective on the ethics and politics of modern scientists, and reminding me that at the end of the day, "c'est qu'un job".

A special thank you to Rebecca Martin, the manager of the First Light group we needed (but perhaps did not deserve?) without which observing and conferences trips would never have gone beyond the booking phase and the nightmarish UCL administrative paperwork would still be pending.

I'd like to thank those with whom I shared coffees, laughs and conversations at the

unofficial Group A Breakfast Club: Guido Roberts-Borsani, Isabella Lamperti, Maria Niculescu-Duvaz, Franziska Schmidt and Andrew Swan. I also thank the larger group of PhD students, postdoc, administrative and academic staff members in the Astronomy group at UCL who welcomed me for three years.

All my gratitude goes to the technicians, engineers, support astronomers and countless scientists who worked on the Keck, VLT, ALMA and NTT telescopes. Despite never having had access to an observing facility as an undergraduate student, or having possessed an amateur telescope as a child, I could embark on an amazing journey to explore the Universe with the most complex instruments built at the time of writing. This could not have been possible without their dedication to make these instruments and pipelines possible to handle without years of expertise and their help before, during and after observing runs.

I'd like to thank Elodie Savary et Alexandre Adler for proofreading the unpublished parts of this thesis, et surtout pour être restés de grands amis malgré la distance.

Je remercie ma famille, qui m'a toujours soutenu tout au long de mes études et dans mon parcours académique. Votre persévérance à me poser des questions sur mon travail et ma passion, malgré la complexité et l'abstraction grandissante au fil des années, est un geste qui me touche beaucoup. Une pensée spéciale va à mon grand-père, n'ayant pas eu le droit d'étudier l'astronomie dans sa jeunesse, qui espérait tant que je puisse réaliser ce rêve et mourut peu de temps avant que ce poste de doctorant me soit proposé.

Last but not least, je remercie Clémence, qui m'a accompagné à Londres et soutenu (supporté) ces trois dernières années, écoutant patiemment d'innombrables descriptions techniques, mes diatribes sur le monde académique, les répétitions de talks et mes plaintes outrées après le passage du referee ou du TAC. Sa volonté et sa capacité de comprendre ce qui me passionne sont admirables, et il n'y pas de mots pour exprimer combien je lui dois.

“ La recherche de la vérité doit être le but de notre activité; c’est la seule fin qui soit digne d’elle. Sans doute nous devons d’abord nous efforcer de soulager les souffrances humaines, mais pourquoi ? Ne pas souffrir, c’est un idéal négatif et qui serait plus sûrement atteint par l’anéantissement du monde. Si nous voulons de plus en plus affranchir l’homme des soucis matériels, c’est pour qu’il puisse employer sa liberté reconquise à l’étude et à la contemplation de la vérité. ”

La Valeur de la Science - Henri Poincaré

Contents

This page was intentionally left blank

List of Figures

This page was intentionally left blank

List of Tables

Chapter 1

Introduction

1.1 Prologue

A century ago the "Great Debate" between Shapley and Curtis took place in Washington, D.C., where both astronomers engaged in a heated exchange over the size of the Milky Way, and whether spiral "nebulae" were located within or outside the known "Universe" (Shapley & Curtis 1921). Indeed, the existence of other "island universes" - galaxies outside our own - would only become an accepted fact a few years later after the Cepheid distance measurements of Hubble (1926). Associated velocity measurements (e.g. Slipher 1915, 1917) also revealed that our Universe was expanding, providing the basis of modern cosmology. Meanwhile, the pioneering works of Zwicky (1933) and Rubin & Ford (1970) revealed that most of the matter in the Universe was so-called "Dark Matter" that only interacts gravitationally with baryonic matter of which the Earth, the Sun and humans are formed. The discovery of the Cosmic Microwave Background in 1964 consolidated the theory that our Universe was born in a "Big Bang" and expanding ever since (Penzias & Wilson 1965). It would take another few decades to discover that the expansion of the Universe was in fact accelerating (Perlmutter et al. 1998; Schmidt et al. 1998), leading to renewed research in what might be powering it - the mysterious "Dark Energy".

Extragalactic astronomy also grew exponentially over the past century. The field has moved from speculations about faint and distant "nebulaes" to the general public staring in wonder at the latest spectacular images of distant and nearby galaxies taken with the *Hubble Space Telescope*. Starting from the 1960s, astronomers have discovered more and more distant galaxies. Photometric and spectroscopic surveys have mapped large-scale structure in the Universe, revealing sheets, filaments and knots known as the Cosmic Web, and the evolution of galaxy properties through cosmic times. Even as these large datasets give us an ever more detailed understanding of the relatively nearby Universe, some astronomers are eagerly looking ahead and trying to push the high-redshift frontier.

The question of the birth and properties of the first galaxies is indeed fundamental. These very first "island universes" are key to understanding how the very first stars and black holes formed, how the circumgalactic medium (CGM) was enriched with the first metals and how the intergalactic medium (IGM) was ionised and reheated. These questions form the basis of early galaxy evolution, and, ultimately, are linked to the origin of the stars, planets and galaxies that surrounds us today, as well as the elements we are made of. Hence, the quest for the first billion years of the Universe - the so-called "last missing chapter" of cosmic history - is undoubtedly also the quest for our own origins.

1.2 Cosmic Reionisation: a brief overview

A few minutes after the Big Bang, the Universe had cooled to $\sim 10^9$ K and the high-energy physics processes responsible for baryonic and non-baryonic matter genesis had ended. What remained was a hot, primordial plasma of protons, neutrons and electrons interacting via Coulomb and Compton scattering. The photons remained tightly coupled to the baryons for much longer, until the temperature of the Universe dropped to $\sim 10^4$ K and recombination of electrons and protons into hydrogen could proceed. As the free electron fraction declined precipitously, the Compton scattering rate became smaller than the expansion rate of the Universe, and photons decoupled

from the baryons. Although helium recombination slightly lagged hydrogen recombination, all these events took place at $z \approx 1000 - 1100$ ($t \approx 400\,\mathrm{kyr}$). Immediately after, baryonic matter only consisted of neutral hydrogen, helium and a trace of light elements, and was bathed in the light of the decoupled photons - the Cosmic Microwave Background (CMB).

The growth of early quantum fluctuations into primordial density perturbations during the inflation phase is a standard prediction of inflation theory, confirmed by anisotropies observed in the CMB. As the Universe continued to expand at a slower rate, the gravitational collapse of these early overdensities led to the birth of the first stars, galaxies and larger structures. For the first time in the Universe, stars shone, ending the so-called Dark Ages at $z \sim 15 - 30$. The first galaxies and clusters grew on a Cosmic Web backbone of Dark Matter, until, half a billion years after recombination, the radiation of the first stars and quasars would reionise the diffuse intergalactic neutral gas. The bulk of this cosmic hydrogen reionisation occurred between $z \sim 6 - 12$, as will become clear in the coming sections. Helium reionisation would have to wait until $z \sim 3$ when the density of quasars peaks, and only hydrogen reionisation will be discussed in this thesis. In the pre-overlap phase, ionised bubbles around particularly efficient ionising sources or large overdensities grew in the neutral IGM. The reionisation process accelerated when the bubbles touched and merged, because the longer mean free path of photons means that isolated sources can contribute to ionising distant patches of neutral gas. We now live in the never-ending post-overlap phase which started at $z \sim 6$ where an overwhelming majority of the IGM is ionised, except for patches of self-shielded gas, called Lyman-Limit Systems (LLS).

The scenario presented above is mostly a working hypothesis guided by our current knowledge of physics. The main questions in cosmic reionisation are therefore i) when did reionisation start and end, and at what pace did it proceed? ii) what is the topology of reionisation (and its evolution)? iii) what are the sources of ionising photons? These questions are still largely open. In the following Section (1.3), I discuss the timing of reionisation, where most progress has been made in the past

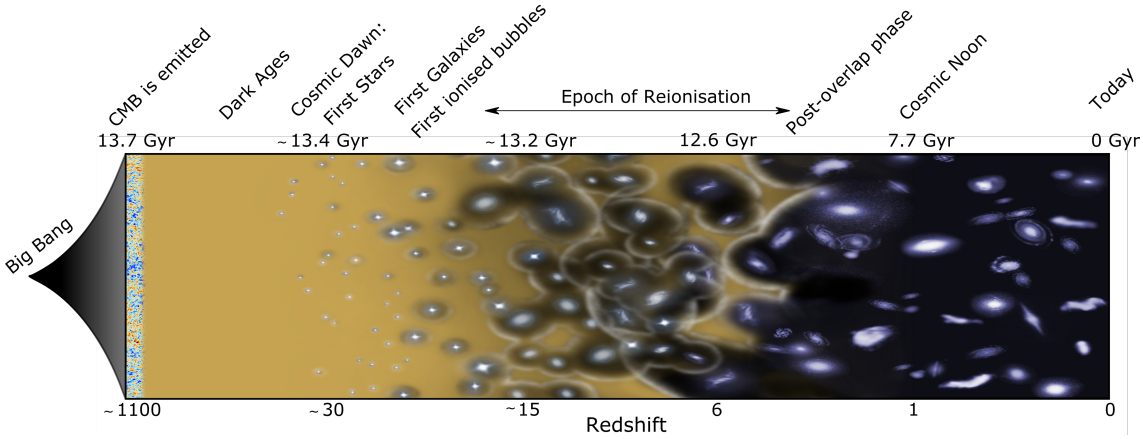

Figure 1.1: Illustration of the history of the Universe and important milestones of Cosmic Dawn and the epoch of reionisation. *Image credits: ESA*

decade. Section 1.4 introduces the current constraints on the nature of the sources of reionisation. Finally, I review the topology of reionisation and its link to the sources of ionising photons in Section 1.5.

1.3 The evolution of the neutral fraction in the first billion years

Fifty-five years ago, Gunn & Peterson (1965) noted that the detection of a continuum blueward of the Lyman-α line of a newly discovered $z \sim 2$ quasar implied that the intergalactic medium was already highly ionised 10.5 Gyr ago. The obvious corollary was that reionisation must have happened at $z > 2$, and that the continuum bluewards of Lyman-α should be absorbed in the spectra of quasars in the reionisation era. This absence of continuum bluewards of Lyman-α, called the Gunn-Peterson (GP) trough, is the best evidence for reionisation ending at $z \sim 6$, and one of the best probes of the evolving neutral fraction (Section 1.3.1). Another consequence of reionisation happening in the first billion years is that the $\gtrsim 10$ Gyr during which the Universe was ionised provides a long path for photons from the CMB to be scattered off free electrons, damping CMB anisotropies on small scales. The Thompson optical depth measured from the CMB provides the only constraint on the total duration of reionisation (Section 1.3.2). These two classic probes (GP

trough and Thomson optical depth) have been complemented by observations of the impact of the neutral IGM on the emission of galaxies and quasars, such as the fraction and clustering of Lyman Break galaxies emitting Lyman-α photons (Stark et al. 2010; Mason et al. 2018a) and the IGM damping wing in high-redshift quasars (e.g. Bañados et al. 2019), which will be discussed in Section 1.3.3. I conclude this section on the evolution of the IGM neutral fraction by a summary of analytical models combining multiple observational probes into a coherent global history of reionisation (Section 1.3.4).

1.3.1 The Gunn-Peterson Trough

As the light emitted from distant bright sources travels through the Universe, the continuum photons emitted bluewards of Lyman-α ($\lambda_\alpha = 1215.67$ Å) are redshifted to the wavelength of rest-frame Lyman-α and beyond. Patches of H I gas along the line of sight will absorb photons at the wavelength of Lyman-α (and other Lyman series). At low-redshift, this leads to an apparent forest of absorption bluewards of the Lyman-α emission of quasars (aptly named the Lyman-α forest). The forest-like pattern of absorptions reflects the projection of gas overdensities along the line sight, whereas the average absorption of the Lyman-α forest traces the evolving opacity of the diffuse IGM through time.

Following Gunn & Peterson (1965) and reviews by Dijkstra (2014); Becker et al. (2015a), I now compute the Gunn-Peterson opacity. I consider photons observed at frequency ν from a quasar at a redshift z_q. Photons with an observed frequency $\nu = \nu_\alpha/(1+z)$ ($\nu_\alpha = c/\lambda_\alpha = 2.47 \times 10^{15}$ Hz) were at the rest-frame frequency of Lyman-α at redshift $z < z_q$ and could be absorbed by the local neutral H I gas in the IGM. The total Lyman-α optical depth takes into account the redshift evolution of the neutral gas hydrogen density $n_{\rm HI}$ and the full scattering cross-section for Lyman-α photons $\sigma_s[\nu]$,

$$\tau_\alpha^{\rm GP} = \int_0^{z_{\rm q}} \sigma_s[\nu(1+z)] n_{\rm HI}(z) \frac{{\rm dl}}{{\rm dz}} {\rm dz} \tag{1.1}$$

where $\frac{\mathrm{dl}}{\mathrm{dz}} = -c/[H(z)(1+z)]$ is the proper path length element, and $H(z)$ is the Hubble parameter. Neglecting line-broadening effects, the scattering cross-section can be approximated by a Dirac delta

$$\sigma_s[\nu(1+z)] = \sigma_\alpha \nu_\alpha \delta[(1+z)\nu - \nu_\alpha] \quad , \tag{1.2}$$

where $\sigma_\alpha = \frac{\pi e^2 f_\alpha}{m_e c \nu_\alpha} = 4.48 \times 10^{18} \mathrm{cm}^{-2}$ is the Lyman-α cross-section (e.g. King 1980). Noting that $\frac{\mathrm{dl}}{\mathrm{d}\nu} = \lambda_\alpha (1+z)/H(z)$, a change of variable from z to ν gives

$$\tau_\alpha^{\mathrm{GP}} = \int_{\nu_\alpha}^{(1+z_\mathrm{q})\nu_\alpha} \sigma_\alpha \nu_\alpha \delta[(1+z)\nu - \nu_\alpha] n_{\mathrm{HI}}(z) \lambda_\alpha \frac{1+z}{H(z)} \mathrm{d}\nu \tag{1.3}$$

$$= \frac{\sigma_\alpha c n_{\mathrm{HI(z)}}}{H(z)} \int_{\nu_\alpha}^{(1+z_\mathrm{q})\nu_\alpha} \delta\left[\nu - \frac{\nu_\alpha}{1+z}\right] \mathrm{d}\nu \tag{1.4}$$

$$= \frac{\sigma_\alpha c}{H(z)} \langle x_{HI}(z) \rangle \overline{n_H} \tag{1.5}$$

where in the second equality the Dirac delta scaling property is used, and in the third one the neutral hydrogen density $n_{\mathrm{HI}}(z)$ is replaced by the volume-averaged average neutral fraction $\langle x_{HI}(z) \rangle$ times the hydrogen density $\overline{n_{\mathrm{H}}}$. The hydrogen density is related to the baryon density Ω_b and the primordial abundance of hydrogen $\overline{n_H} \simeq \rho_{\mathrm{crit}} \Omega_b (1-Y)(1+z)^3 / m_{\mathrm{H}}$, where Y is the primordial abundance of helium and ρ_{crit} the critical density. At high redshift, the approximation $H(z) \simeq H_0 \Omega_m^{1/2} (1+z)^{3/2}$ can be used to further simplify equation 1.5. The famous Gunn-Peterson optical depth is then

$$\tau_\alpha^{\mathrm{GP}} \simeq 2.3 \times 10^5 \langle x_{HI} \rangle \left(\frac{\Omega_\mathrm{b} h^2}{0.022}\right) \left(\frac{\Omega_\mathrm{m} h^2}{0.142}\right)^{-1/2} \left(\frac{1-Y}{0.76}\right) \left(\frac{1+z}{5}\right)^{3/2} \quad . \tag{1.6}$$

At $z \sim 5 < z < 7$, the last term in Eq. 1.6 only changes by a factor ~ 1.65. Therefore the evolution of the optical depth is mostly controlled by the average neutral fraction in the IGM. Even at very low neutral fractions $\approx 10^{-4.5}$, the transmission of Lyman-α photons $T = e^{-\tau_\alpha} \simeq e^{-7}$ is practically zero and therefore no flux is transmitted in the Lyman-α forest of $z \gtrsim 6$ quasars, resulting in the so-called Gunn-Peterson trough. Additionally, transmitted flux in the Lyman-α forest

of $z \sim 6$ quasars signals an IGM that is only $< 0.1 - 0.01\%$ neutral. The GP trough detection is therefore evidence for the very last stages of the post-overlap phase of reionisation.

The observational detection of Gunn-Peterson troughs (Figure 1.2) followed the discovery of $z > 6$ quasars in the Sloan Digital Sky Survey (SDSS, Fan et al. 2001). Becker et al. (2001) reported the first GP trough in the spectrum of $z = 6.28$ quasar J1030+0524. It was quickly pointed that the IGM neutral fraction evolved swiftly between $z \sim 5.8 - 6.3$, with a large scatter in Lyman-α opacities measured between sightlines (Fan et al. 2002, 2006; Songaila 2004). This scatter between sightlines, suggesting a patchy reionisation process, is confirmed with larger samples of intermediate resolution ($R \sim 9000$) quasar spectra (Bosman et al. 2018; Eilers et al. 2018a). While the detection of the GP trough and the Lyman-α opacity measurements have pinned down the timing of the end-stages of reionisation, the patchiness of reionisation is still a challenge for numerical simulations (see further Section 1.5).

At $z \gtrsim 6$, the opacity of the IGM to Lyman-α reaches $\tau \sim 3 - 4$ and is no longer measurable as the transmission is close to zero (e.g. Bosman et al. 2018; Eilers et al. 2018a; Mortlock et al. 2011; Bañados et al. 2018). The Lyman-β forest is visible at slightly higher redshifts as $\tau_\beta = \frac{\lambda_{Ly\alpha}}{\lambda_{Ly\beta}} \frac{f_{Ly\alpha}}{f_{Ly\beta}} \tau_\alpha \simeq 0.16\tau_\alpha$ and can be used to probe the neutral fraction as well (Eilers et al. 2019). However the steep evolution of the IGM opacity and the overlap of the Lyman-β forest with the lower-redshift Lyman-α absorptions complicates the measurement. Nonetheless, spectra of $z \gtrsim 6.5$ quasars can still provide valuable constraints on reionisation from the so-called "dark gaps". Dark gaps are defined as the uninterrupted sequence of dark pixels (e.g. where no flux is detected) in the Lyman-α forest. Since an absence of transmission can either be due to an extended diffuse neutral region or some residual dense H I clouds, this measurement gives a lower limit on the fraction of ionised gas (Fan et al. 2006). Deriving a neutral fraction constraint from the observed dark gap distribution requires a comparison to simulations (e.g. Mesinger 2010). Despite being model-dependent, dark gaps are still one of the best constraints on the neutral

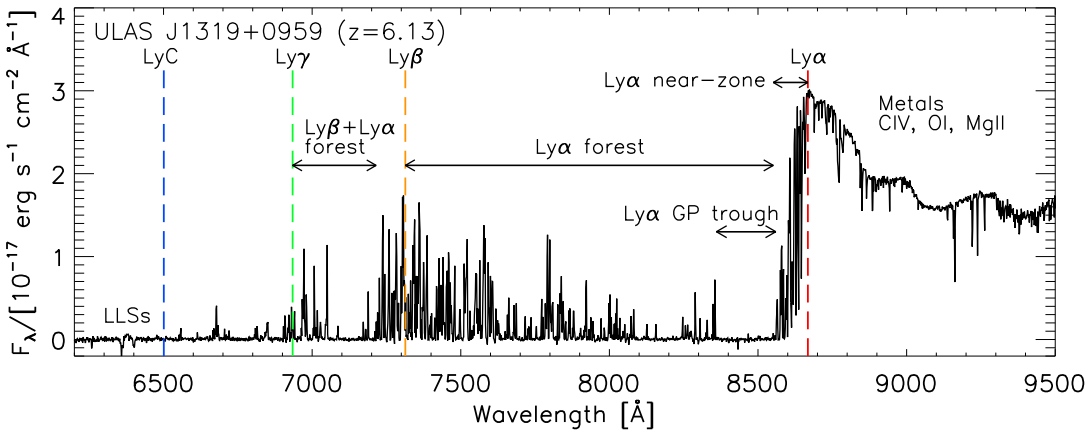

Figure 1.2: Illustration of the Lyman-series forests, Gunn-Peterson trough, quasar near-zone and metal absorbers in a $z > 6$ quasar spectrum. *Reproduced from Becker et al. (2015a).*

fraction at $z \sim 6.5 - 7$ (Greig et al. 2017).

1.3.2 The CMB Thomson optical depth

As photons from the CMB travel through the Universe, they scatter off free electrons, damping small scales anisotropies of the CMB power spectrum (e.g. Planck Collaboration et al. 2018). This suppression depends on the integrated optical depth due to free electrons

$$\tau^{\rm CMB} = n_{\rm H}(0) c \sigma_{\rm T} \int_0^{z_{\rm max}} x_e(z) \frac{(1+z)^2}{H(z)} {\rm dz} \tag{1.7}$$

where $x_e(z) = n_e(z)/n_{\rm H}(z)$ is the fraction of free electrons and $\sigma_{\rm T}$ the Thomson scattering cross-section. The suppression of the small scale anisotropies increases exponentially with the Thomson optical depth ($\propto e^{2\tau^{\rm CMB}}$). Additionally, CMB polarisation measurements constrain both the Thomson optical depth and the evolution of the free electron fraction $x_e(z)$ (e.g. Kaplinghat et al. 2003). This is important because a functional form of $x_e(z)$ must be assumed when $\tau^{\rm CMB}$ is constrained alongside the five cosmological parameters $(\theta_{\rm MC}, \Omega_b, \Omega_c, n_s, A_s)$ inferred from the CMB power spectrum, where $\theta_{\rm MC}$ is a numerical approximation to the acoustic scale angle, Ω_b the baryon density, Ω_c the dark matter density, n_s the scalar spectral index of the primordial fluctuation power spectrum, and A_s the primordial power

spectrum amplitude (see further,e.g., Planck Collaboration et al. 2018 and references therein). Polarisation measurements from the latest *Planck* results have therefore improved the inference on the history of reionisation compared to previous studies, regardless of the functional form assumed for $x_e(z)$ (Figure 1.3). Nonetheless, it is important to note that the reionisation histories showed on Figure 1.3 are entirely parametric. Observations of the neutral fraction at high-redshift (Section 1.3) or numerical simulations (Section 1.5.1) provide a more accurate picture of the timing and topology of reionisation.

It is interesting to note that the value of τ^{CMB} has been steadily decreasing in the past two decades as the CMB measurements have been refined. The previous tension has disappeared as the CMB constraints now match the Lyman-α forest and other probes that support a late reionisation. In what follows, the CMB constraint is taken to be the mid-point of reionisation (e.g. the redshift at which 50% of the IGM is ionised), which is largely independent from the modelling uncertainties discussed above. The current constraints (Planck Collaboration et al. 2018) are

$$\tau^{\rm CMB} = 0.054 \pm 0.007 \tag{1.8}$$

$$z_{\rm mid} = 7.7 \pm 0.7 \quad . \tag{1.9}$$

1.3.3 Additional observational probes of the neutral fraction

The opacity measurements in the Lyman-α forest of high-redshift quasars and the CMB power spectrum and polarisation constrain the start and end point of reionisation. During reionisation, the reduced transmission of Lyman-α photons from galaxies and quasars due to the partially neutral IGM leads to three observable probing the neutral fraction evolution with redshift.

The first test is the evolution of the so-called "Lyman-α emission fraction" (Figure 1.4, left). The Lyman-α fraction is the fraction of Lyman-α line detection in high-redshift Lyman Break galaxies (see further Section 1.4.2 for the Lyman Break technique). In a reionised Universe, the decreasing dust content and possibly H I

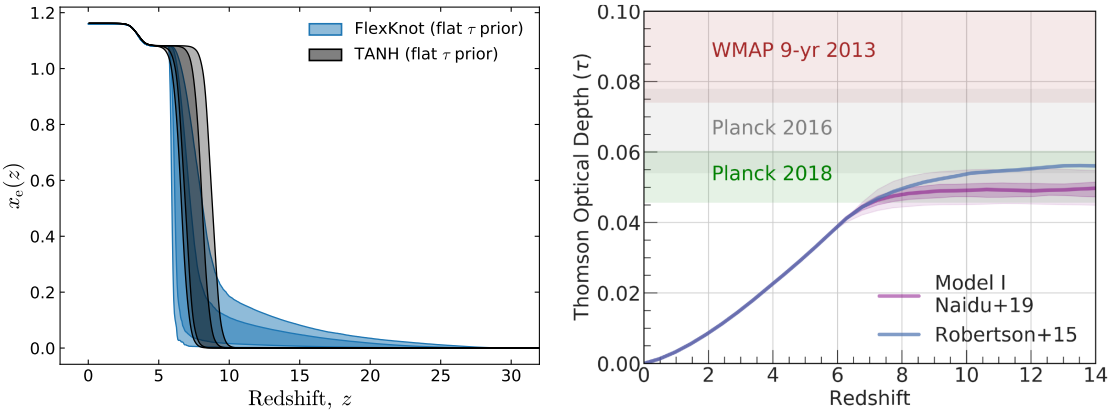

Figure 1.3: **Left:** Posterior probability density function of different neutral fraction histories parametrisations used in the *Planck* modelling. Regardless of the method used, the CMB constraints point to a rapid reionisation completed in under ~ 0.5 Gyr. *Reproduced from Planck Collaboration et al. (2018).* **Right:** The measured Thomson optical depth value has decreased over the years and the latest *Planck* values matches well analytical models of the neutral fraction (See further Section 1.3.4). *Reproduced from Naidu et al. (2020).*

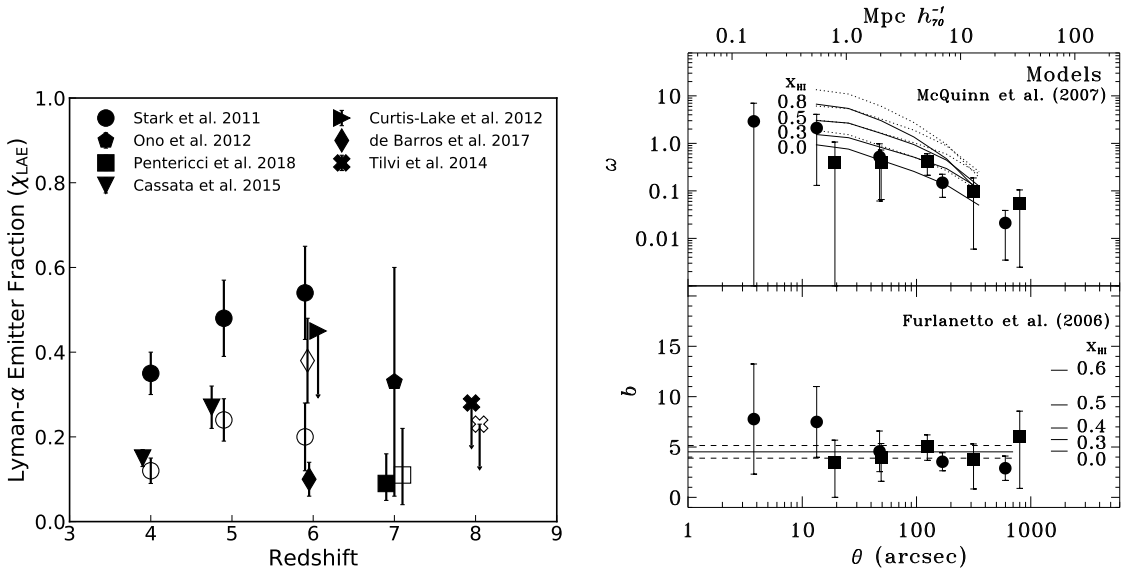

Figure 1.4: **Left:** Fraction of Lyman Break selected galaxies showing a Lyman-α line with an equivalent width EW > 25 Å. Full symbols represent the brighter population with $-21.75 < M_{\rm UV} < -20.25$ and the empty symbols the fainter galaxies with $-20.25 < M_{\rm UV} < -18.75$. **Right:** Observed and modelled angular clustering function (top) and bias (bottom) of $z \sim 6.6$ LAEs. *Reproduced from Ouchi et al. (2018).*

covering fraction with redshift explain the rising fraction of Lyman-α detections in Lyman Break Galaxies (LBG) at $3.7 \lesssim z \lesssim 5.7$ (see Stark 2016 , for a review). However, as the IGM becomes increasingly neutral, photons bluewards of fractionare absorbed by the $z > 5$ IGM, and the Lorentzian damping wing even suppresses photons redwards of fraction(Figure 1.5, left). The Lyman-α emission fraction is thus expected to decline sharply beyond $z \sim 5 - 6$, in agreement with observations (e.g. Stark et al. 2010; Ono et al. 2012; Caruana et al. 2012; Schenker et al. 2014; Tilvi et al. 2014; De Barros et al. 2017; Pentericci et al. 2018; Caruana et al. 2018; Arrabal Haro et al. 2018; Kusakabe et al. 2020). The exact turnover redshift is still debated, but a patchy reionisation process and small sample sizes could explain the discrepancy between studies. Stark et al. (2011); De Barros et al. (2017) have reported a differential evolution between faint and luminous galaxies, with the Lyman-α emission fraction declining earlier in the former than the latter. This could be partly due to the biased environment in which more luminous objects are found, although modelling by Mason et al. (2018b) suggest this is not sufficient to explain the boosted transmission of Lyman-α. Mesinger et al. (2015) have also argued that the decline is too steep to be accounted for solely by the increasingly neutral IGM, and might also come from co-evolving galaxy properties.

The absorption of Lyman-α photons also has consequences for the observed clustering of Lyman-α Emitters (LAEs). The effect is more subtle because it depends on the topology of reionisation, the interplay between the location of galaxies and residual H I patches. However, it is a more precise probe than the Lyman-α emission fraction because it is independent from an intrinsic change in the Lyman-α emission of galaxies with redshift. The angular correlation function of Lyman-α emitters has been measured at $z \sim 5.7$ and $z \sim 6.6$ using wide-field narrow-band surveys with the Subaru telescope (Ouchi et al. 2008, 2010, 2018). The absence of evolution of the angular clustering between $z \sim 5.7$ and $z \sim 6.6$ suggests it is not affected by patchy reionisation and the LAE clustering bias constrains $x_{HI} = 0.15 \pm 0.15$ at $z \sim 6.6$ (Figure 1.4). This constraint must be taken with caution because the interpretation relies on simulations that model the transfer of Lyman-α photons from

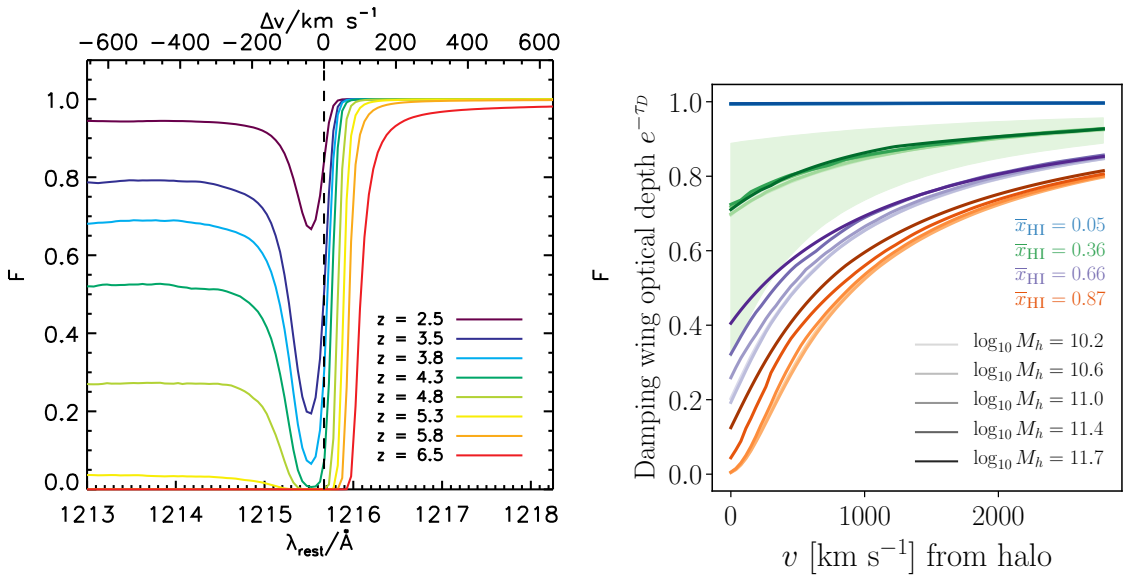

Figure 1.5: **Left:** Transmission of Lyman-α photons through the diffuse IGM. At $z \gtrsim 5.5$, the GP trough absorbs the majority photons bluewards of the line center. *Reproduced from Laursen et al. (2011).* **Right:** IGM damping wing models for a range of host halo masses and neutral fraction. More massive haloes are reionised earlier and therefore more Lyman-α photons are transmitted. *Reproduced from Mason & Gronke (2020).*

high-redshift galaxies through the reionising IGM (e.g Furlanetto et al. 2006; McQuinn et al. 2007; Kulkarni et al. 2019a). Adopting a broad range of such models, Greig et al. (2017) show that the LAE clustering mostly excludes early reionisation models starting before $z \sim 12 - 14$, but gives only a weak constraint on the neutral fraction at $z \sim 6.6$ compared to the dark gaps distribution measurement.

As discussed previously, the IGM only needs to have a neutral fraction of $10^{-4.5}$ for the transmission Lyman-α photons to be close to zero (Eq. 1.6). With larger neutral fractions ($x_{\rm HI} \gtrsim 0.1$), the Lorentzian wing of the Lyman-α absorption profile can no longer be neglected, and neutral patches can absorb photons up to $\sim 10^3$ km s^{-1} redwards of the line center (e.g. Miralda-Escude 1998). It is therefore expected that the Lyman-α emission from a quasar or a galaxy will be attenuated by the surrounding neutral IGM during reionisation. The IGM damping wing is even stronger when taking into account a patchy reionisation instead of a homogeneous ionised medium (Mesinger & Furlanetto 2008a). As the attenuation strength is dependent on the average density and neutral fraction of the gas, it can be used as a probe of the

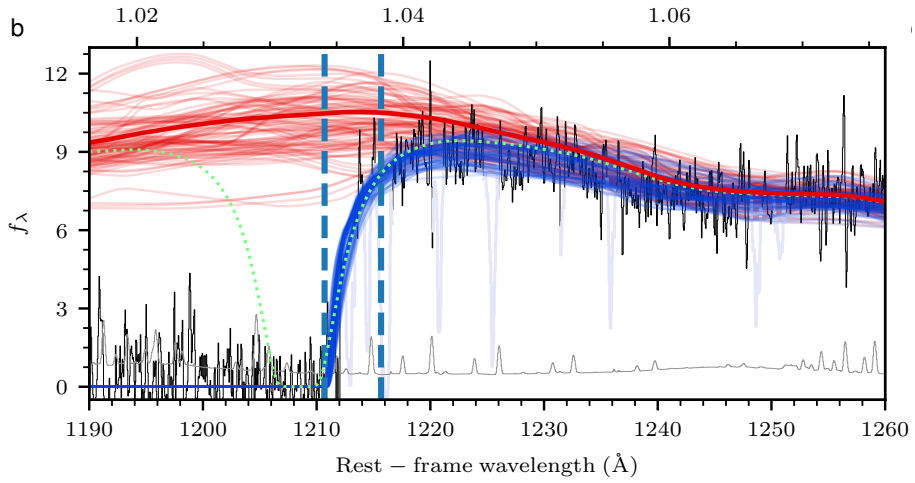

Figure 1.6: Lyman-α line of the $z = 7.54$ quasar J1342+0928, with different models (thin red lines) aggregated to produce the final intrinsic Lyman-α profile (thick red line) before being attenuated by the IGM damping wing. The IGM damping wing model is shown in blue, and the damping wing profile of a single DLA (dotted green), claimed to be incompatible with the observed transmission. *Reproduced from Bañados et al. (2018).*

neutral fraction (Figure 1.5, right).

The detection of a damping wing in $z \gtrsim 7$ quasars is a promising technique to measure the neutral fraction locally around the quasar (Mortlock et al. 2011; Bañados et al. 2018; Wang et al. 2020). There are however two concerns with this relatively new approach. Firstly, it presents a biased view of the neutral fraction history because it traces overdensities which are thought to reionise first. Secondly, it relies on the proper modelling of the intrinsic Lyman-α profile. This is usually done by using a Principal Component Analysis (PCA) spectrum model trained on the rest-frame UV continuum of low-redshift (usually SDSS) quasar spectra (e.g. Davies et al. 2018c; Ďurovčíková et al. 2020). However, these models often fail to reproduce the N V 1240 Å or C IV 1549 Å broad emission lines of $z > 7$ quasars, raising concerns about whether they therefore predict Lyman-α emission correctly. Several authors have argued that low-redshift training samples might not be comparable to observed high-redshift quasars as the latter might suffer from selection biases or evolve with redshift (e.g. Bolton et al. 2011; Bosman & Becker 2015; Greig et al. 2017).

1.3.4 Analytical modelling of the neutral fraction history

The evolution of the neutral fraction of the IGM with redshift has been measured with various probes, which collectively give a coherent picture of the neutral fraction history. (Figure 1.7). I now review how this neutral fraction history is modelled analytically and fitted to the observational constraints.

To first order, the evolution of IGM ionised volume filling fraction $\mathrm{Q_{HII}} = 1 - \mathrm{Q_{HI}}$ is driven by the ionising photon density n_{ion} and slowed by hydrogen recombination (e.g. Madau et al. 1999)

$$\dot{\mathrm{Q}}_{\mathrm{HII}} = \frac{\dot{n}_{ion}}{\langle n_{\mathrm{H}} \rangle} - \frac{\mathrm{Q_{HII}}}{t_{\mathrm{rec,HI}}} \quad , \tag{1.10}$$

where $t_{\mathrm{rec,HI}}$ is the hydrogen recombination timescale and $\langle n_{\mathrm{H}} \rangle$ is the mean cosmic density of hydrogen, which is set by the primordial fraction of hydrogen ($\approx 75\%$) and the mean baryon density ρ_b (see Eq. 1.6). The only variable is therefore the production rate density of ionising photons with energies > 13.6 eV, hereafter Lyman Continuum (LyC) photons, which can be elegantly expressed as

$$\langle \dot{n}_{\mathrm{ion}} \rangle = \langle \rho_{\mathrm{UV}} \xi_{\mathrm{ion}} f_{\mathrm{esc,LyC}} \rangle \quad , \tag{1.11}$$

where ρ_{UV} is the total UV luminosity density of sources producing ξ_{ion} LyC photons per UV magnitude (in units of $\mathrm{erg^{-1}Hz}$), and $f_{\mathrm{esc,LyC}}$ is the fraction of these photons that escape the galaxy (or quasar) and ionise the IGM.

The ionising photon budget of reionisation (Eq. 1.11) illustrates how hydrogen reionisation is a complex interplay between the number of the sources of reionisation, their emission properties and the interstellar medium and circumgalactic medium physics that control the escape fraction. The neutral fraction evolution is well-reproduced in analytical models of the neutral fraction if the photon budget is dominated by the contribution of faint galaxies $-16 \lesssim M_{\mathrm{UV}} \lesssim -10$, with $\log \xi_{\mathrm{ion}}/\mathrm{erg^{-1}Hz} \simeq 25.4$, and $f_{\mathrm{esc,LyC}} \gtrsim 10\%$ (e.g. Robertson et al. 2015; Ishigaki et al. 2018; Finkelstein et al. 2019; Naidu et al. 2020). Recently, Naidu et al. (2020)

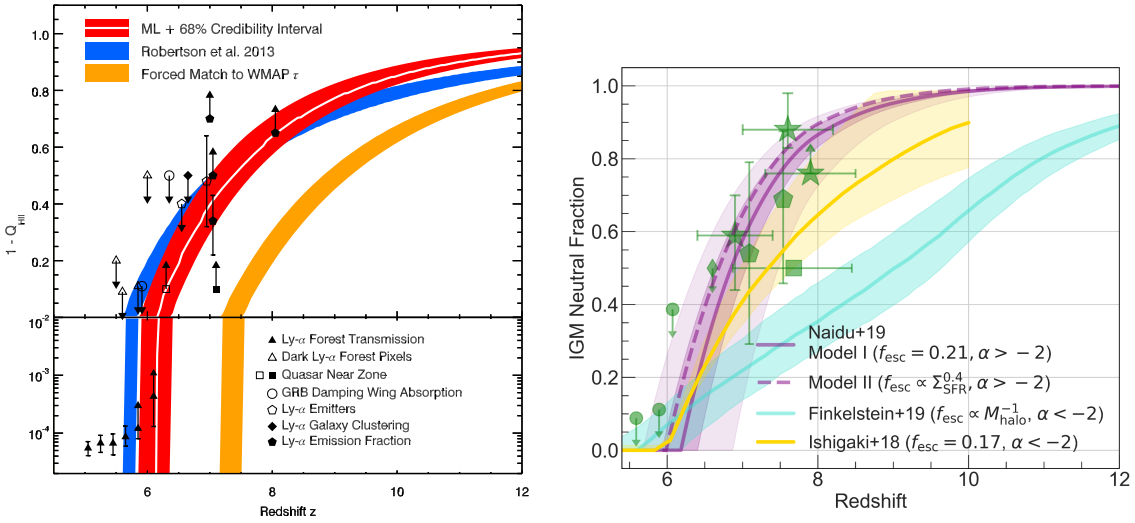

Figure 1.7: **Left:** Neutral fraction ($1 - Q_{HII}$) of the IGM evolution with redshift. The various observational constraint suggests that reionisation was mostly completed between $6 \lesssim z \lesssim 12$. The blue and red models assume $f_{esc,LyC} \simeq 0.2$, $\log \xi_{ion}/[erg^{-1}Hz] = 25.2$ and slightly different UV luminosity densities ρ_{UV} to predict the neutral fraction using Eq. 1.10 and 1.11. *Reproduced from Robertson et al. (2015).* **Right:** More recent models and measurements of the neutral fraction history. *Reproduced from Naidu et al. (2020).*

have suggested that the neutral history could be as well reproduced with an "oligarchy" of bright leakers with $M_{UV} \lesssim -21$, $f_{esc,LyC} \gtrsim 30\%$, but their proposition relies on the neutral fractions derived from the EW of Lyman-α emitters at $z \sim 7$ (Mason et al. 2018a) and the quasar damping wing measurement, which are both heavily model-dependent measurements. Regardless of the robustness of the different probes, which will increase in the coming years, the sources of reionisation are likely to be a heterogeneous ensemble of objects. In that sense, the average escape fraction and ionising efficiencies in Equation 1.11 must be considered as ensemble-average quantities. In the next two sections, I therefore review what is currently known of galaxies and quasars in the first billion years to see whether the photon budget can be balanced or not.

1.4 The sources of ionising photons

1.4.1 The role of quasars in cosmic reionisation

The number of known quasars in the first billion years has greatly increased since the detection of the first GP trough (see Section 1.3.1). If quasars were discovered first in the radio domain (e.g. Matthews & Sandage 1963), they are now found overwhelmingly using appropriate colour-colour cuts (e.g. Richards et al. 2002) in the large photometric catalogs of ground based surveys such as the Sloan Digital Sky Survey (SDSS, York et al. 2000), the Dark Energy Survey (DES, Abbott et al. 2018), the Panoramic Survey Telescope and Rapid Response System (PanSTARRS, Chambers et al. 2016), or the DESI Legacy Imaging surveys (DELS, Dey et al. 2019). In fifteen years, the number of $z \gtrsim 6$ quasars has increased from a handful to ~ 300 (Bosman 2020 for a constantly updated list), with the most distant quasars now found at $z \sim 7.5$ (Bañados et al. 2018; Yang et al. 2020).

As discussed in Section 1.3, quasars play a central role in probing the reionisation history of the Universe using various Lyman-α forest measurements and the IGM damping wing method. Their contribution to reionisation, however, has been periodically debated in the field. The LyC photon escape fraction in quasars is consistently measured to be close to unity at $z \sim 4$ (e.g. Cristiani et al. 2016; Grazian et al. 2018; Romano et al. 2019), and their emissivity of ionising photons is high due to their universal power law spectra (Lusso et al. 2015). The debate on the contribution of quasars to reionisation was reignited by Giallongo et al. (2015) who claimed that numerous faint quasars, unseen in wide imaging fields but detected in X-rays with *Chandra*, were present at $z > 5$. Soon after, Madau & Haardt (2015) presented a toy model where quasars could easily produce enough photons to ionise the Universe by $z \sim 6$. A simulation of reionisation driven by quasars (Chardin et al. 2017) was shown to reproduce the observed cosmic variance in the reionisation topology (Bosman et al. 2018 , see further Section 1.5).

However, the claims of numerous faint quasars at $z > 5$ are challenged by other studies (Onoue et al. 2017; Parsa et al. 2018) which cannot reproduce or match the

Chandra detections (but see also Giallongo et al. 2019). Kulkarni et al. (2019b) recently conducted review study of the quasar luminosity funtion (QLF) using all quasar samples available at $2 \lesssim z \lesssim 7$. They conclude that the parametrisation of the QLF can lead to important errors in the number of predicted quasars when extrapolating outside of the observed luminosity range, thus explaining the disagreement between previous studies. They also show that quasars still contribute $\sim 10\%$ of the photoionisation rate at $z \sim 6$, and very little further into reionisation (Figure 1.8). From a theoretical standpoint, Finkelstein et al. (2019) find that quasars may have a minor contribution to ionising photon budget at the end of reionisation, perhaps explaining patchy reionisation at $z \sim 5.5$. The role of quasars is considered minor in all analytical models of the neutral fraction history which either neglect their contribution (e.g. Robertson et al. 2015; Naidu et al. 2020) or include it (Finkelstein et al. 2019; Dayal et al. 2020), but find nonetheless similar results. Finally, Trebitsch et al. (2020) simulated a large cluster hosting galaxies with and without AGNs and conclude that, despite the larger escape fraction in AGN hosts, the total ionising contribution was still dominated by the UV faint galaxies in their local volume. Ideally, additional simulations are still needed to assess whether this is the case in different environments.

The best arguments against an important contribution of quasars to hydrogen reionisation ultimately come from external constraints: observations of helium reionisation at $z \sim 3$ and the temperature of the IGM.

Since quasars are expected to inject a lot of heat into the IGM and emit He II ionising photons, a surplus of quasars at $z \gtrsim 6$ would have an impact on the timing of helium reionisation and the IGM temperature. D'Aloisio et al. (2017) showed that the elevated number of quasars claimed by Giallongo et al. (2015, 2019) would reionise He II earlier than observed and lead to an overheated IGM by $z \sim 2$, in strong tension with observations (Worseck et al. 2016; Mitra et al. 2017). It therefore seems that the contribution of quasars to reionisation is subdominant, and that most of the ionising flux must come from high-redshift galaxies.

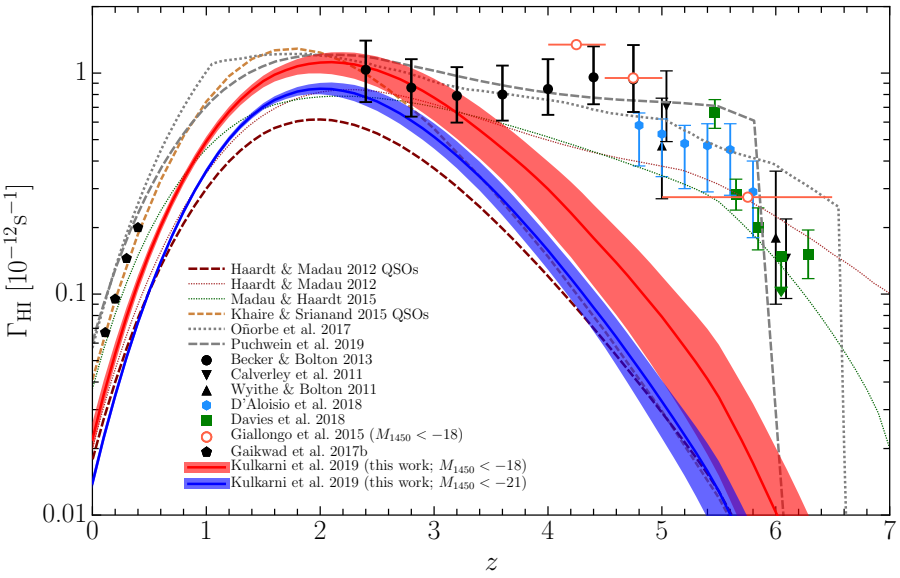

Figure 1.8: Contribution of quasars to the H I photoionisation rate throughout cosmic time. The derived photoionisation rate due to quasars Γ_{HI} (blue, red shaded regions) peaks at $z \sim 2{-}3$ (at the peak cosmic density of quasars) effectively completing He II reionisation. The observed photoionisation rates (coloured points and errorbars) are derived by fitting for the UVB necessary to match the IGM opacity to Lyman-α(-β) in hydrodynamical simulations to that measured in quasar spectra. At $z \gtrsim 5$ quasars are too rare and can provide only $\sim 10\%$ of the observed photoionisation rate. *Reproduced from Kulkarni et al. (2019b).*

1.4.2 Pushing the high-redshift frontier

The breakthrough in the discovery of numerous $z \gtrsim 2$ galaxies in the early 90s is mainly owed to the "Lyman Break Technique", which is now applied to galaxies up to $z \sim 10$. At redshift $z \gtrsim 2$, the UV continuum of galaxies is redshifted into the visible and is observable from the ground. Stellar population synthesis model predict that the typical break at the Lyman limit is $\sim 1.5 - 2$ mag (Bruzual & Charlot 1993). Once hydrogen absorption by external Lyman-limit systems and the Lyman-α forest is taken into account, the "Lyman Break" can reach five magnitudes or more (Madau 1995).

Steidel & Hamilton (1992, 1993); Steidel et al. (1995) pioneered the use of this strong Lyman Break in the spectra of distant galaxies to estimate the redshift of $z \sim 3$ galaxies. At $z \sim 3$, the Lyman Limit falls at the longer wavelength end of the U band, and most of the Lyman-α forest is still out of the G band. Assuming a flat spectrum $f_\nu \propto$ constant, at this redshift a Lyman break galaxy will reveal a red

U-G colour (or Lyman 'break') due to hydrogen absorption but a blue G-R colour representing its unattenuated UV continuum. Lower (higher) redshift candidates can be selected by requiring that the break happens in bluer (redder) bands. Such candidates are often often called "$u/g/r/i/z$-dropouts" if their flux falls in band $u/g/r/i/z$, respectively. The relative simplicity of this technique, combined with efficient Multi-Object Spectroscopy follow-up on 8-10m class telescopes such as Keck and VLT rapidly lead to the discovery of numerous $z \sim 2-4$ LBGs (see Shapley 2011 for a review). At $z \gtrsim 5$, the discontinuity shifts from the Lyman Limit to the wavelength of Lyman-α (1215.67 Å) owing to the saturation of the Lyman-α forest in the increasingly neutral intergalactic medium (see Section 1.3.1). This Lyman Break technique enabled the search for ever more distant galaxies, and is now commonly used to search for $z \sim 8$ galaxies and beyond (Stark 2016 , for a review).

A second breakthrough came with the *Hubble Space Telescope* (hereafter *HST*), which enabled observation of deep and relatively wide (202x202 arcsec^2) fields in the optical and infrared. Several surveys, such as the Great Observatories Origins Deep Survey (GOODS, Giavalisco et al. 2004), the Hubble Ultra Deep Field program (HUDF, Beckwith et al. 2006), the Beyond Ultra-deep Frontier Fields and Legacy Observations survey (BoRG, Trenti et al. 2011), the Cluster Lensing And Supernova Survey with Hubble (CLASH, Postman et al. 2012), the Hubble Ultra Deep Field 2012 campaign (UDF12 Ellis et al. 2013), the Frontier Fields survey of lensing clusters (Lotz et al. 2017), the Reionization Lensing Cluster Survey (RELICS, Coe et al. 2019), the Beyond Ultra-deep Frontier Fields and Legacy Observations (BUFFALO, Steinhardt et al. 2020) created a legacy of deep *HST* fields to find and study high-redshift galaxies. The various surveys altogether have pursued a tiered ("wedding-cake") strategy, with wider and shallower fields ($\sim 10^3$ arcmin^2, $m_{\rm AB} \sim 26$) as well as small and deep ones ($\sim 10^2$ arcmin^2, $m_{\rm AB} \sim 29$) imaged with the broad band filters between 0.6μm and 1.6μm the ACS/WFC3 camera. The aim of these tiered datasets is to be able both to detect faint and distant objects in the deepest fields, whilst constituting large samples and increasing the chance of finding rare luminous objects in wide fields. Additionally, some fields were also observed

with *Chandra* (e.g. *Chandra* Deep Field North/South) or *Spitzer* (e.g. Frontier Fields, RELICS) to increase the science legacy. Most $z > 7$ galaxies known today were found in these fields. At the end of reionisation ($z \sim 6$) where Lyman-α is still easily visible from the ground, ground-based wide-field surveys conducted with the Hyper Suprime-Cam (HSC) at the Subaru Telescope such as *GOLDRUSH* (> 100 degree^2, Ono et al. 2018) and *SILVERRUSH* (> 10 degree^2, Ouchi et al. 2018) are finding unprecedented numbers of luminous objects.

These imaging surveys can be searched for high-redshift galaxies by applying the Lyman Break technique to obtain large numbers of so-called "dropout" galaxies. Such samples are large enough to compute accurately the cosmic number density of galaxies ρ_{UV} at $z \sim 2$ to $z \sim 10$, a key quantity for reionisation which I will discuss in the following section (1.4.3). Depending on the redshift of the objects and the spectral coverage, the photometry can be used to determine star-formation rates (to compute the cosmic Star Formation rate density) and ionising efficiencies, which I discuss in Section 1.4.4. Additionally, the most distant galaxy candidates are often followed-up spectroscopically to confirm their redshift and study the rest-frame UV lines when present, or increasingly, confirmed via their far-infrared lines accessible with ALMA (e.g. Hashimoto et al. 2018a see for the most distant galaxy known at the time of writing, z=9.11).

1.4.3 UV luminosity function and cosmic star formation rate density

The UV luminosity function Φ_{UV} (the number density of galaxies per UV magnitude) is the first quantity that can be derived from the large samples of LBGs found in deep fields. In the context of reionisation, it determines how many luminous and faint galaxies can potentially act as sources of ionising photons. As a consequence of the efforts of several teams on various *HST* deep fields highlighted in Section 1.4.2, the UV LF is now known down to $z \sim 10$, providing useful constraints for relatively bright objects with $-16 \lesssim M_{\rm AB,UV} \lesssim -22$ (Figure 1.9, Finkelstein et al. 2015; Bouwens et al. 2015, 2017; Livermore et al. 2017; Ishigaki et al. 2018; Ono

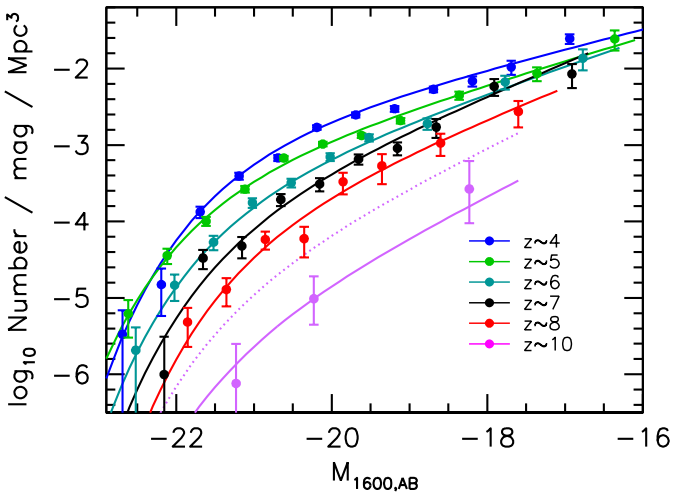

Figure 1.9: Evolution of the UV LF with redshift from the *HST* legacy fields (CANDELS, HUDF09, HUDF12, ERS, and BoRG/HIPPIES Bouwens et al. 2015). The expected decline in total number density of galaxies with redshift is accompanied by softening of the "knee" at $M_{1600,\mathrm{AB}} \sim -20$. *Reproduced from Bouwens et al. (2015).*

et al. 2018; Oesch et al. 2018; Atek et al. 2018; Bowler et al. 2018). Nonetheless, the following three issues are still debated: i) the slope and extent of the faint end of the LF, ii) a possible non-Schechter extension of the bright end and iii) a claimed precipitous decline in the integrated number density beyond $z \sim 8$.

The evolution of the UV LF has been claimed to accelerate towards higher-redshift (specifically, the number density of galaxies declines faster at $z > 8$), and accordingly the cosmic star formation rate (SFR) density (Figure 1.10, Bouwens et al. 2017; Oesch et al. 2018). This is however disputed by other studies using similar methods (e.g. McLeod et al. 2015; Morishita et al. 2018). Taken at face-value, the faster decline of the SFR density at $z > 8$ compared to lower redshift (Madau & Dickinson 2014) would suggest we are reaching the first galaxies in the Universe. However, recent simulations suggest that these tensions could be simply due to cosmic variance between overdense and underdense fields (Ucci et al. 2020). An older stellar population than commonly thought in the brightest objects at $z > 10$ could also reduce the discrepancy between the SFR density evolution predicted from low-redshift data and that observed at $z > 8$ (Roberts-Borsani et al. 2020). It is also worth noting that most galaxies used for the purpose of computing the UV LF and

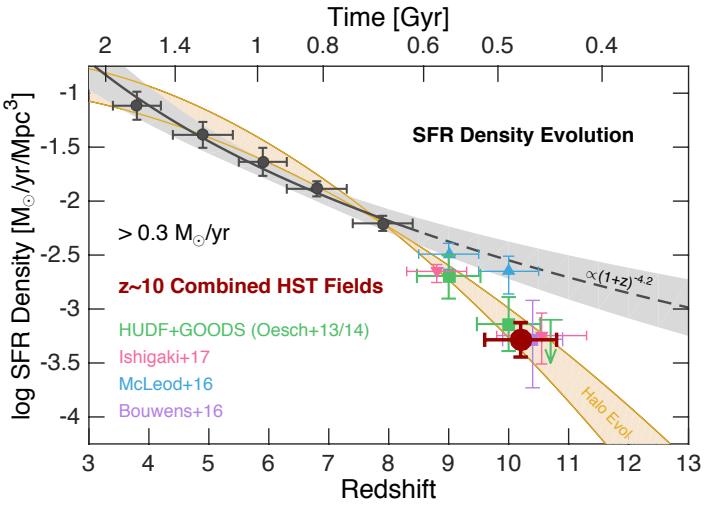

Figure 1.10: Cosmic SFR density at $3 < z < 13$ based on the most distant LBG candidates in the *HST* legacy fields. Assuming the same conversion factor from UV luminosity to SFR than at lower redshift (Madau & Dickinson 2014), the rapidly declining UV LF suggests a dramatic decline of the SFR density as we approach the first half billion of the Universe. *Reproduced from Oesch et al. (2018).*

SFR density at $z > 5$ are only "dropout" candidates and have not been confirmed spectroscopically.

The bright-end of the UV luminosity function is limited by cosmic variance as the brightest objects are by definition the rarest. Bowler et al. (2018, 2020); Adams et al. (2020) have claimed an excess of objects at the bright-end compared to the usual single Schechter function

$$\Phi(L_{\rm UV})\mathrm{d}L_{\rm UV} = \Phi^* \left(\frac{L_{\rm UV}}{L^*_{\rm UV}}\right)^{\alpha} \exp\left(-\frac{L_{\rm UV}}{L^*_{\rm UV}}\right) \mathrm{d}\left(\frac{L_{\rm UV}}{L^*_{\rm UV}}\right) \quad , \tag{1.12}$$

which they interpret as a lack of AGN feedback in the largest haloes at $z > 8$. However, the wider sky coverage of the HSC GOLDRUSH survey has shown that this excess could arise from contamination by low luminosity AGNs with $M_{\rm UV} \lesssim -23$ (Ono et al. 2018). This is of particular importance to reionisation because AGNs and quasars are thought to have much higher escape fractions and ionising efficiencies than star-forming galaxies (Section 1.4.1). Any contamination by AGNs at the bright-end of the galaxy UV LF would lead to an overestimate in the number of galaxies and their role in reionisation at the expense of quasars, and an underesti-

mate of the total number of ionising photons.

The determination of the faint-end of the UV LF is perhaps the issue most relevant to this thesis. Measurements at the faint-end are essentially limited by the depth of the observations. Lensing clusters can be used to detect fainter sources by taking advantage of gravitational lensing with magnifications up to $10-100$, pushing the nominal limit on the absolute magnitude detectable to $M_{\rm UV,AB} \sim -14$. However, Atek et al. (2018) have demonstrated that the different fields and lensing models used in different studies and the assumed shape of the UV LF (e.g. Livermore et al. 2017; Bouwens et al. 2017) affect the results greatly (see Figure 1.11). Simulations predict that the formation of dwarf galaxies should be suppressed in the reionisation era due to supernovae or radiative feedback which can eject or photoevaporate, respectively, the gas necessary for star formation(e.g. Wise et al. 2014; Gnedin 2016; Yue et al. 2016 and references whithin). The current data is insufficient to conclude on the possible turnover of the luminosity function at $M_{\rm UV} \gtrsim -10$. Even with *JWST*, which will improve on the *HST* magnitude limit by 2-3 mag (e.g. Mason et al. 2015), the turnover of the UV LF might not be constrained. The main consequence of this uncertainty is that all analytical models computing the UV luminosity density $\rho_{\rm UV}$ to solve the ionising photon budget (Eq. 1.11) must assume a minimum UV magnitude ($M_{\rm UV}^{\rm min}$) at which the UV LF effectively turns over to prevent the UV luminosity density from diverging

$$\rho_{\rm UV} = \int_{M_{\rm UV}^{\rm min}}^{\infty} L(M_{\rm UV})\Phi(M_{\rm UV}){\rm dM_{UV}} \quad . \tag{1.13}$$

Having described how the UV number density of galaxies $\rho_{\rm UV}$ is computed from deep fields, we now turn to the ionising emissivity of galaxies, the second key quantity in the ionising photon budget.

1.4.4 Ionising photon production efficiency of galaxies

The ionising efficiency $\xi_{\rm ion}$ is defined as the ratio between the production rate of LyC photons Q(LyC) in units of s^{-1} and the dust-corrected specific UV continuum

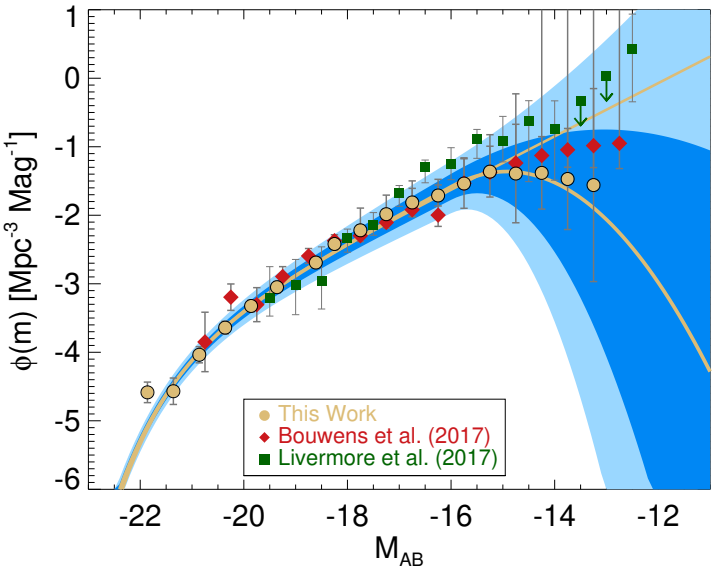

Figure 1.11: Recent constraints on the faint-end of the UV luminosity function of $z \sim 6$ galaxies. The two yellow lines show fits to the observed data (yellow, red and green points) assuming either a Schechter function or a modified Schechter function with a turnover factor at $M_{\rm UV} = -16$. The blue envelopes show the $1(2)\sigma$ contours of the fit to the yellow datapoints using a double-Schechter function, which is still consistent at $< 2\sigma$ with an absence of turnover of the UVLF. *Reproduced from (Atek et al. 2018).*

luminosity $L_{\rm UV}$ in units of erg s^{-1} Hz^{-1}, $\xi_{\rm ion} = \frac{\rm Q(LyC)}{L_{\rm UV}^{\rm int}}$. The UV luminosity is measured at $1500-1600$Å rest-frame depending on studies, and dust-corrected using a Calzetti (Calzetti et al. 2000) or SMC-like (Capak et al. 2015) dust extinction law. The production rate of ionising photons is frequently derived from Balmer emission lines.

Assuming Case B recombination (where the ISM is optically thick to Lyman-series photons) and zero escape fraction of ionising photons, every LyC photon eventually leads to a recombination cascade with a Balmer line emission. Assuming a temperature of 10^4 K, neglecting dust and escape of ionising photons, Leitherer & Heckman (1995) find in their evolutionary stellar synthesis models the following conversion factor between the Balmer line luminosity and the production rate of

LyC photons

$$\mathrm{Q(LyC)} = 7.35 \times 10^{11} \left(\frac{L(\mathrm{H}\alpha)}{[\mathrm{erg}]} \right) \quad , \tag{1.14}$$

$$\mathrm{Q(LyC)} = 2.10 \times 10^{12} \left(\frac{L(\mathrm{H}\beta)}{[\mathrm{erg}]} \right) \quad , \tag{1.15}$$

where $L(\mathrm{H}\alpha)(L(\mathrm{H}\beta))$ is the luminosity of the $\mathrm{H}\alpha(\mathrm{H}\beta)$ line in units of $\mathrm{erg\,s^{-1}}$. Kennicutt (1998); Schaerer (2003) give similar conversion factors, even though they can change slightly depending on the assumed IMF, temperature and metallicity.

If the number of ionising photons is derived from the *observed* $\mathrm{H}\alpha$ luminosity, it must be corrected by a factor $1/(1 - f_{\mathrm{esc,LyC}})$ as escaping LyC photons will not recombine and produce Balmer lines photons. The ionising efficiency is therefore

$$\xi_{\mathrm{ion}} = \frac{\mathrm{Q(LyC)}}{L_{\mathrm{UV}}} = \frac{\mathrm{Q}^{H\alpha}\mathrm{(LyC)}/(1 - \mathrm{f_{esc,LyC}})}{L_{\mathrm{UV}}} \quad ,c \tag{1.16}$$

where I have distinguished between the intrinsic LyC production rate from the quantity derived observationally from $\mathrm{H}\alpha$ with an appropriate superscript. In all studies detailed below, $f_{\mathrm{esc,LyC}}$ is assumed to be zero, and the ionising efficiency is therefore denoted ξ^0_{ion}.

Shivaei et al. (2018) applied this method to 673 $z \sim 2$ galaxies in the MOSDEF survey, the largest sample used so far for this purpose, with $\mathrm{H}\alpha$ and $\mathrm{H}\beta$ both observed with MOSFIRE/Keck. They find a mean ionising efficiency $\log_{10} \xi_{\mathrm{ion}}/[\mathrm{erg^{-1}Hz}] = 25.06(25.34)$ for a Calzetti (Calzetti et al. 2000) or SMC-like (Capak et al. 2015) dust extinction law.

At $z \sim 4 - 5$, $\mathrm{H}\alpha$ and $\mathrm{H}\beta$ are redshifted out of the range of ground-based spectrographs. Nonetheless, $L(\mathrm{H}\alpha)$ can be estimated from broad band *Spitzer* imaging and careful SED modelling (e.g. Smit et al. 2016). Using this method, Bouwens et al. (2016) derived ionising efficiencies for $z \sim 4 - 5$ galaxies, finding $\log_{10} \xi_{\mathrm{ion}}/[\mathrm{erg^{-1}Hz}] \sim 25.27^{+0.03}_{-0.03}(25.34^{+0.02}_{-0.02})$, for a Calzetti or SMC dust extinction law. They also report elevated ionising efficiencies for bluer galaxies (spectral slope $\beta < -2.3$).

Other studies suggest that specific populations might have higher ionising efficiencies, but all of them have smaller samples and cannot correct for dust obscuration due to insufficient coverage of either Hα and Hβ. For example, Nakajima et al. (2016, 2020) find $\log_{10}\xi_{\rm ion}/[{\rm erg^{-1}Hz}] \simeq 25.4$ in $z \sim 3-4$ extreme [OIII] emitters, and $\simeq 25.6$ for LAEs. The trend of higher $\xi_{\rm ion}$ with higher [OIII] is confirmed by Tang et al. (2019), whilst the trend of larger $\xi_{\rm ion}$ for fainter objects is supported by the Maseda et al. (2020) study of $z \sim 4-5$ LAEs with elevated equivalent widths (median ${\rm EW_{Ly\alpha}} = 249$Å, $\log_{10}\xi_{\rm ion}/[{\rm erg^{-1}Hz}] = 26.28^{+0.28}_{-0.40}$). Finally, Matthee et al. (2017) find a low ionising efficiency in Hα emitters (24.77 ± 0.04) compared to that of LAEs (24.77 ± 0.04) selected in a larger sample of $z \sim 2.2$ galaxies (Sobral et al. 2017).

I present in Figure 1.12 the different results obtained over wide range of different galaxy type at $2 \lesssim z \lesssim 7$. It appears clearly that particular sub-populations (brighter LAEs or [OIII] emitters) have higher $\xi_{\rm ion}$. Stark et al. (2017); Laporte et al. (2017b); Mainali et al. (2020) report the detection of rest-frame UV lines such as C IV, He II and C III] in individual galaxies at $z > 7$ which suggest harder radiation spectra ($\log_{10}\xi_{\rm ion}/[{\rm erg^{-1}Hz}] \gtrsim 25.6$) and/or AGN activity at high redshift. However, spectroscopic confirmations using the Lyman-α line might be biased towards exceptional objects which have already ionised their surrounding gas, and so these objects might not be representative of the larger population. Finally, I note that the intrinsic calculations above neglect the escape fraction of ionising photons, which indicates that the ionising efficiencies are probably underestimated. It follows that an evolution of the escape fraction, increasing with redshift, might thus steepen the evolution of $\xi_{\rm ion}$.

1.4.5 The escape fraction of ionising photons of galaxies

The escape fraction of LyC photons is perhaps the least understood and the most challenging factor in understanding the contribution of galaxies to hydrogen reionisation. It is naturally defined by the ratio of escaping LyC flux with respect to the

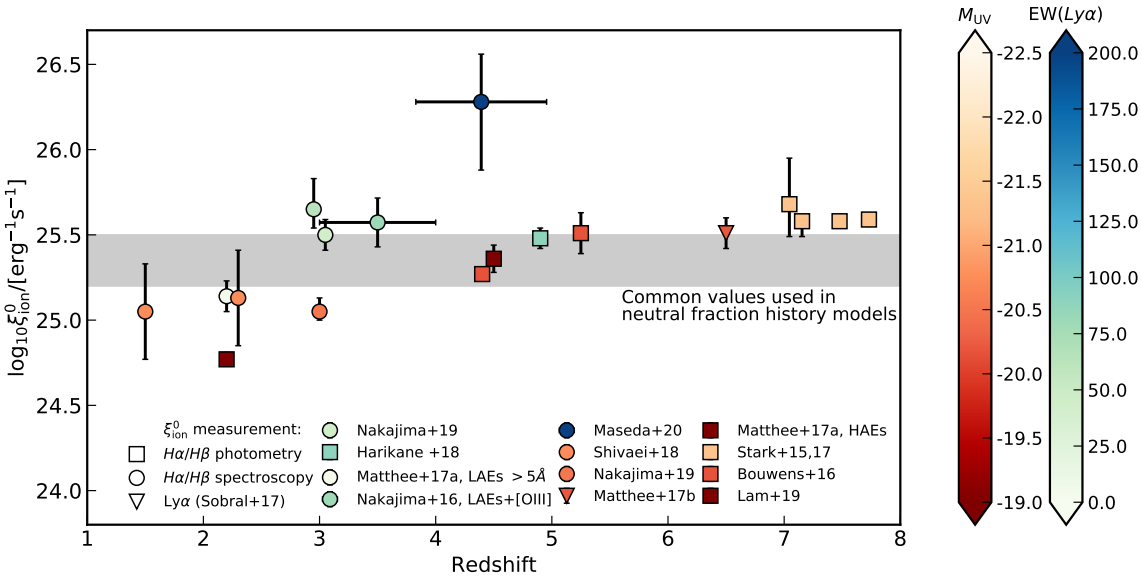

Figure 1.12: Ionising efficiency $\xi_{\rm ion}$ of galaxies throughout cosmic time. LAEs are displayed in shades of blues indicating the Lyman-α EW. Lyman Break selected galaxies (and usually confirmed with Lyman-α) are displayed in reds indicating the UV magnitude of the sample or the objects. The different symbols denote different methods of measuring the intrinsic production rate of LyC photons and thus infer $\xi_{\rm ion}$. The grey band illustrates the $\xi_{\rm ion}$ values commonly assumed in reionisation models.

intrinsic LyC flux. However, determining the exact number of LyC photons produced is difficult and dependent on model assumptions (see Section 1.4.4). Therefore, studies starting from Steidel et al. (2001); Shapley et al. (2006) measure a better defined quantity, the *relative escape fraction*

$$f^{\rm rel}_{\rm esc,LyC} = \frac{L_{1500}/L_{900}}{f_{1500}/f_{900}} \frac{1}{e^{-\tau_{\rm IGM,900}}} \quad , \tag{1.17}$$

where the flux ratio f_{1500}/f_{900} between 1500 Å and 900 Å is corrected for average IGM attenuation $e^{-\tau_{\rm IGM,900}}$ and normalized by the observed luminosity ratio L_{1500}/L_{900}. To recover the relative escape fraction, the intrinsic ratio of non-ionising to ionising luminosity must be determined by detailed SED fitting and in most cases is taken to be $L_{1500}/L_{900} = 3$ to facilitate comparison between studies. The absolute escape fraction can also be recovered if the dust attenuation at 1500 Å is known ($f_{\rm esc,LyC} = f^{\rm rel}_{\rm esc,LyC}/10^{-0.4A(1500)}$) (e.g. Inoue et al. 2005; Siana et al. 2007). In what follows we quote exclusively results in terms of absolute escape fractions for low-redshift studies.

The average escape fraction of LyC photons at $z \lesssim 4$ has long been a contentious issue. Notwithstanding the above modelling issues, sample selection appears to play an important role. The emerging picture is that, at $z \lesssim 4$, the average LBG population has low escape fractions $f_{\rm esc,LyC} \lesssim 10\%$ (e.g. Leitherer et al. 1995; Inoue et al. 2005; Shapley et al. 2006; Siana et al. 2007, 2010; Boutsia et al. 2011; Siana et al. 2015; Grazian et al. 2016). More encouragingly perhaps, Steidel et al. (2018) report an absolute escape fraction of $f_{\rm esc,LyC} = 0.09 \pm 0.01$ from an unprecedentedly large sample of LBGs at $z \sim 3$. Gamma-Ray Bursts hosts consistently have low escape fractions at all redshifts $z \lesssim 6$ (Tanvir et al. 2019). However, higher values ($f_{\rm esc,LyC} \sim 20 - 60\%$) have been found in some individual low-redshift galaxies (e.g. Mostardi et al. 2015; Vanzella et al. 2016; Izotov et al. 2016b,a, 2017).

These sources could conceivably be representative of sub-populations that have been overlooked and might play an important role in the early Universe. Steidel et al. (2001) already reported elevated escape fractions for a selection of some of the bluest LBGs available at the time, and Kusakabe et al. (2020) recently suggested that the traditional colour-cuts used to select LBGs are biased toward LAEs with high EW. Several studies have also noted a correlation between high [OIII]/[OII] ratios and the escape fraction of LyC photons (Faisst 2016; Fletcher et al. 2019; Izotov et al. 2018; Nakajima et al. 2020 but see also Bassett et al. (2019)). It is interesting to note that bluer spectra and high [OIII]/[OII] also correlate with high $\xi_{\rm ion}$ at $z \sim 4$ (Section 1.4.4), making extreme [OIII] emitters very efficient sources of ionising photons.

The apparent discrepancy between measurements could be due to the anisotropic nature of LyC escape. The commonly-held view is that channels of low-column density gas are created by supernovae (Wise et al. 2014; Sharma et al. 2016; Ma et al. 2020a,b). However, because young and massive stars drive the strongest winds and emit the majority of LyC, their short lifetime implies that few ionising photons are produced once the leakage channels have been created and the overall ionising output is low (Kimm & Cen 2014). Kakiichi & Gronke (2019) have suggested that the turbulent kinematics of the gas could also create those optically thin channels.

Using high-resolution simulations of < 10 pc turbulent ISM gas, they show that low escape fractions are mostly due to an ionisation-bounded scenario where photons escape anisotropically through channels. In contrast, the highest escape fractions occur when the ISM is entirely ionised and the H II regions are density-bounded. Such spatial anisotropies in the LyC leakage have been observed in gravitationally-lensed $z \sim 4-5$ sources (Leethochawalit et al. 2016). Recently Gazagnes et al. (2020) also supported this hypothesis by showing that the covering fraction of H I, a proxy for the density of ionised channels, is tightly correlated with $f_{\rm esc,LyC}$ for values $< 80\%$, after which leakage is density-bounded, in agreement with Kakiichi & Gronke (2019). Finally, studies that determine average escape fractions made from aggregating best-fit values for individual sources also report a large scatter, with individual sources or sub-samples showing $f_{\rm esc,LyC} > 20\%$ for an ensemble-average that is much lower ($f_{\rm esc,LyC} < 2-3\%$) (e.g. Grazian et al. 2016; Fletcher et al. 2019).

For all the progress at low-redshift, the escape fraction of $z > 5$ galaxies remains virtually unknown. Indeed, Lyman Continuum has not been detected at these redshifts where the IGM opacity to LyC increases quickly. The suggestion from the ioinising photon budget calculations (Section 1.7) that early galaxies must have $f_{\rm esc,LyC} > 10\%$ (much higher than what is measured in stacking analyses at low-redshift) is therefore still unverified and the effective contribution of galaxies to reionisation cannot be assessed.

1.5 Linking the sources of ionising photons and the topology of reionisation

1.5.1 Modelling cosmic reionisation

Analytical models describing the evolution of the neutral fraction discussed in Section 1.3.4 only provide a one-dimensional view of reionisation. The volume filling fraction of neutral hydrogen Q_{HI} cannot be directly observed and must be forward-

modelled from its impact on various probes (Section 1.3). Cosmic variance in the Lyman-α forest of distant quasars suggests that the interplay between sources of ionising photons and the IGM produces a patchy reionisation. Reproducing its complex topology is a great challenge for simulations and suggests that understanding the 3D distribution of galaxies and ionised regions is necessary and can provide a path to identifying the sources of reionisation.

Simulating cosmic reionisation is a complex multi-scale problem. The ionising photons emitted around young stars and black holes must escape the ISM to ionise gas several tens of comoving Megaparsecs (cMpc) away. The boxes should ideally be large enough to be able to draw skewers comparable to the scales probed by quasar Lyman-α forests ($\sim 10^2$ cMpc) but with enough resolution to realistically model the escape of LyC photons through the ISM on parsec scales. Hydrodynamics and radiative transfer must be accounted for, and additionally metal enrichment, black hole seeding and growth, radiative cooling, turbulence or magnetic fields, which all affect widely different scales. Furthermore, meaningful comparison between the simulations and the observations can only be made if numerous realisations are produced to match the observed cosmic variance. Needless to say, such a herculean task will probably never be accomplished. Instead, approximations must be made and simulations designed to address specific questions at the expense of determining the holistic view.

Numerical simulations currently used to produce realistic topologies of the reionising IGM, as well as the temperature fluctuations and source locations in cosmological volumes ($10^1 - 10^2$ cMpc), can be broadly divided in two approaches. The first approach simulates reionisation in hydrodynamical simulations. The gas is assumed to be optically thin at every resolution element and the time-varying UV background and heat injection are supplied from 1D radiative transfer codes (e.g. Faucher-Giguère et al. 2009; Haardt & Madau 2012; D'Aloisio et al. 2017; Oñorbe et al. 2017; Puchwein et al. 2019). However, reionisation happens instantly at a given redshift in the box, after which the UVB is applied to compute the opacity fluctuations in the post-overlap phase. To simulate an inhomogeneous reionisation,

individual reionisation redshift, local UVBs and temperatures must be assigned to the each resolution element (e.g. Feng et al. 2016; D'Aloisio et al. 2017; Oñorbe et al. 2019). The prescription for heat injection and UV background can come from semi-numerical excursion set models (Nasir & D'Aloisio 2020) where the Press-Schechter formalism (Press & Schechter 1974) is used to provide an estimate of the amount of stars in haloes. The spatially varying UV background is computed by assuming the number of LyC photons produced by stars on average (Furlanetto et al. 2004; Mesinger et al. 2011). Alternatively, the entire hydrodynamical simulation can be post-processed with a radiative transfer code (e.g. Chardin et al. 2015, 2017; Kulkarni et al. 2019b; Keating et al. 2020). The relative low-cost of hydrodynamical simulations enables theorists to increase the resolution and test various UVB models at the cost of radiative feedback and cooling, and a somewhat unphysical treatment of the IGM opacity.

An alternative approach undertakes a full radiative transfer simulations that trade low resolutions and a limited volume for a more accurate treatment of the transmission of Lyman-α and Lyman Continuum photons. Early attempts by Gnedin & Kaurov (2014); Gnedin (2014) only managed to simulate 3-6 realisations of 20-40 cMpc wide boxes. They however match remarkably well the observed neutral fraction history by tuning only the emissivity of star particles in the simulation. Likewise, Finlator et al. (2018); Katz et al. (2019a) study relatively small volumes (5-12 Mpc) in order to study metal absorbers in the reionisation or the link between galaxy properties and local reionisation histories, respectively.

Before concluding, I will mention two orthogonal approaches that do not seek to improve the simulation physics but rather the comparison with observations. The first issue arising when comparing with observations is cosmic variance. Since large simulations cannot be re-run many times, faster codes can be used to post-process the output with thousands of different UVB prescriptions, facilitating the comparison with observed data (e.g. Chardin et al. 2015; Choudhury et al. 2020). Alternatively, Chardin et al. (2019) harness neural networks that learn to emulate simulations and can generate quickly new realisations with different initial param-

eters. Another interesting addition is the photon-tracing algorithm developed by Katz et al. (2019b,a). This photon-tracing algorithm keeps track of which sources reionised particular patches of gas and when. They find a trend where galaxies in lower mass haloes ($\lesssim 10^{9.5} M_{\odot}$) contribute most to the photoionisation rate (and thus the growing H II volume filling fraction) at $z \sim 9$, and high-mass haloes slowly dominating by the end of reionisation ($z \sim 6$). However, their resolution is not enough to resolve the ISM and they can only study the escape of LyC photons through the IGM. Any dependence of the escape fraction on the halo mass and/or the presence of optically thin channels due to supernova feedback or turbulence is not taken into account. Their simulation rather reflects the importance of the environmental dependence of the high-redshift LyC escape fraction. For example, photons from massive galaxies have difficulties to travel through the IGM before it is partially ionised because they are emitted in overdense regions.

The above simulations by design tune the instantaneous reionisation redshift, the emissivity of galaxies and/or AGN/supernova feedback to match the neutral fraction history. The insight they can offer on the global history of reionisation is therefore limited. Of greater interest is the 3D distribution of over- and under-ionised regions naturally arising in these simulations. I will now describe the observational constraints on the topology of reionisation and compare them to the simulations described above.

1.5.2 The patchy reionisation of hydrogen

The discovery of $z > 6$ quasars and the Gunn-Peterson trough led to multiple IGM opacity to Lyman-α measurements, which were instrumental in constraining the end of reionisation (Section 1.3.1). However, as the number of observed sightlines grew, it was quickly realised that the scatter in the measurements was too large to be explained solely by cosmic variance and might imply an inherently patchy reionisation process. The long GP trough seen in the quasar J0148 (Becker et al. 2015b) is challenging for all models to reproduce. Patchy reionisation was recently firmly

established by the cumulative distribution function measurement of the Lyman-α opacity in dozens of $z > 6$ (Bosman et al. 2018; Eilers et al. 2018a) which also appears challenging for simulations.

One notable exception is the model of Chardin et al. (2015, 2017), which has been able to reproduce long troughs and the opacity scatter if $\sim 50\%$ of the ionising budget is supplied by quasars at $z > 6$. Rare sources by definition add a lot of sightline-to-sightline scatter, but their model is ruled by observations of the QLF (Kulkarni et al. 2019a). Recently, Kulkarni et al. (2019b) presented a post-processed hydrodymical simulation in which a late (ending at $z \sim 5.5$) and patchy reionisation is achieved by a rapid evolution of the ionising emissivity of star-forming galaxies around $z \sim 6$. Their simulation reproduced the Lyman-α opacity scatter and also showed that long troughs similar to that of J0148+0600 could arise in underdense voids, which reionise last (Keating et al. 2020). Large transmission spikes are thought to appear when powerful sources of ionising photons switch on next to diffuse and underdense gas patches (Garaldi et al. 2019). On the observational side, Becker et al. (2018) show that the long trough of J0148 (Becker et al. 2015b) is surrounded by an under-density of narrow-band selected galaxies. Kashino et al. (2019) report a similar result for broad-band selected galaxies around a GP trough detected in another sightline. All these results show that the topology of reionisation is intrinsically linked to the distribution and nature of the sources of reionisation. In the next Section, I describe how the link between galaxies and ionised patches can be harnessed to measure escape fractions.

1.5.3 The cross-correlation of high-redshift galaxies with the IGM opacity to Lyman-α

A fundamental predicament of reionisation studies is the challenge of measuring escape fractions at high-redshift. Double-peaked Lyman-α are exceedingly rare at $z > 5$ (e.g. Hu et al. 2016; Songaila et al. 2018; Matthee et al. 2018; Bosman et al. 2020) and other methods measuring the covering fraction of H I gas require low-

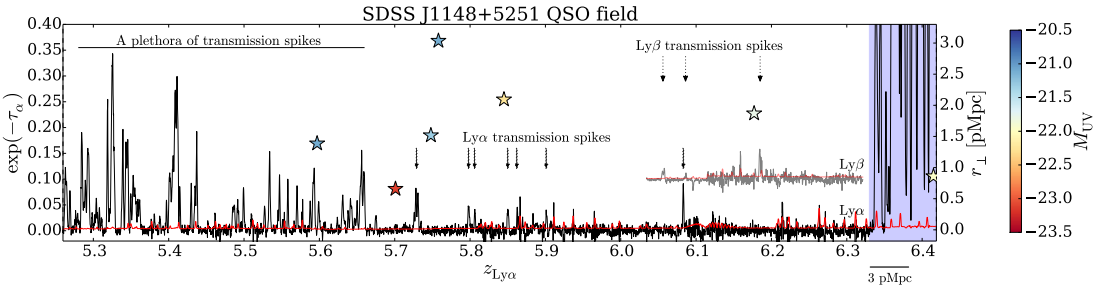

Figure 1.13: Demonstration of the Kakiichi et al. (2018) method to probe the average escape fraction of $z \sim 6$ galaxies. The Lyman-α forest transmission (black) of $z = 6.48$ quasar J1148+5251 is indicated on the left ordinate, with the transmission error in red. The abscissa indicates the redshift of Lyman-α transmission in the forest, and the Lyman-β forest is offset vertically for clarity. The stars spectroscopically confirmed in the field are plotted at their Lyman-α emission redshift and perpendicular distance from the quasar (right ordinate), and coloured by luminosity. *Reproduced from Kakiichi et al. (2018).*

ionisation absorption lines (e.g. Leethochawalit et al. 2016; Chisholm et al. 2018; Gazagnes et al. 2020) inaccessible at $z > 5$ until 30-40m class telescopes are available. Kakiichi et al. (2018) proposed a novel approach by investigating the correlation between overdensities of star-forming galaxies and the over-ionised gas in the post-overlap phase of reionisation. This breakthrough method seeks to determine statistically the effective ionising output of faint galaxies ($\langle f_{\rm esc,LyC}\xi_{\rm ion}\rangle$) in the reionisation era by correlating bright galaxies with the IGM opacity to Lyman-α probed by background quasars. The foundation of this idea is that an increased transmission close to galaxies should be expected as they ionise the IGM, and the amplitude of this effect can be used to measure their ionising output.

Kakiichi et al. (2018) investigated the correlation of galaxies and IGM transmission in the field of J1148+5251. They found several galaxies in the foreground of the quasar and a spectroscopic redshift within the range of the detected Lyman-α forest features (Figure 1.13). They determined the average transmission in a pixel at distance $r = \sqrt{r_\parallel^2 + r_\perp^2}$, where $r_\perp, r_\parallel$ are the distance perpendicular and parallel to the quasar sightline. An increased transmission next to spectroscopically confirmed LBGs was found (see Figure 1.14). They then determined whether a single LBG could be responsible for this enhanced transmission on scales of a few proper Mpc

(pMpc). The local photoionisation rate of a single galaxy is

$$\langle \Gamma_{\rm ion}^{\rm Gal}(r) \rangle = \frac{\alpha_g \sigma_{912}}{\alpha_g + 3} \frac{\langle \dot{N}_{\rm ion} \rangle}{4\pi r^2} e^{-r/\lambda_{\rm mfp}} \quad [{\rm s}^{-1}] \quad , \tag{1.18}$$

where r is the comoving distance from the LBG, α_g is the extreme UV ($>$ 13.6 eV) spectral slope, $\sigma_{912} = 6.35 \times 10^{-18}$ cm^2 the cross-section for LyC photons, $\langle \dot{N}_{\rm ion} \rangle = \langle f_{\rm esc} \xi_{\rm ion} \rangle \langle L_{\rm UV} \rangle$ the average ionising photon production rate for a fiducial UV luminosity $L_{\rm UV}$ [erg s^{-1}Hz^{-1}], $\lambda_{\rm mfp}$ the mean-free path of LyC photons, for which we use the value of $6.0[(1+z)/7]^{-5.4}$ pMpc (Worseck et al. 2014). Assuming $\alpha_g = 2$ (e.g. Kuhlen & Faucher-Giguère 2012; Becker & Bolton 2013) and a luminosity $L_{\rm UV} \simeq 2 \times 10^{29}$erg s^{-1}Hz^{-1} for a standard LBG, the H I photoionisation rate is

$$\Gamma_{\rm HI}^{\rm LBG} \approx 6.4 \times 10^{-15} \left(\frac{r}{[{\rm pMpc}]} \right) \left(\frac{f_{\rm esc,LyC} \times \log \xi_{\rm ion}}{0.1 \times 10^{25.2} [\,{\rm erg}^{-1}{\rm Hz}]} \right) \quad , \quad [{\rm s}^{-1}] \quad , \tag{1.19}$$

therefore a single LBG cannot be responsible for the H I photoionisation measured in the IGM at $z \sim 6$, $\overline{\Gamma}_{\rm HI} \approx 10^{-12} - 10^{-13}{\rm s}^{-1}$ (e.g. Davies et al. 2018a).

Kakiichi et al. (2018) therefore consider the enhanced photoionisation rate around luminous LBGs due to the presence of clustered faint galaxies. These clustered faint objects boost the photoionisation rate and provide in fact more ionising photons than the detected LBGs. To take into account the contribution of clustered faint objects, Kakiichi et al. (2018) convolve the 2-point correlation function of galaxies with the radiative transfer kernel ($\frac{e^{-|\boldsymbol{r}|/\lambda_{\rm mfp}}}{4\pi|\boldsymbol{r}|^2}$) such that

$$\langle \Gamma_{\rm HI}^{\rm CL}(\boldsymbol{r}) \rangle = \frac{\overline{\Gamma}_{\rm HI}}{\lambda_{\rm mfp}} \int \frac{e^{-|\boldsymbol{r}-\boldsymbol{r}'|/\lambda_{\rm mfp}}}{4\pi|\boldsymbol{r}-\boldsymbol{r}'|^2} \left[1 + \langle \xi_g(|\boldsymbol{r}'|) \rangle_L \right] d^3\boldsymbol{r}' \quad [{\rm s}^{-1}] \quad , \tag{1.20}$$

where $\overline{\Gamma}_{HI} = \frac{\alpha_g \sigma_{912}}{\alpha_g + 3} \lambda_{\rm mfp,912} \langle f_{\rm esc,LyC} \xi_{\rm ion} \rangle$ is the average H I photoionisation rate in the IGM, and $\langle \xi_g(|\boldsymbol{r}'|) \rangle_L$ is the luminosity-weighted correlation function of galaxies with luminosity $L_{\rm UV}$ contributing to the UVB,

$$\langle \xi_g(|\boldsymbol{r}|) \rangle_L = \frac{\int_{L_{\rm min}}^{\infty} L\Phi(L) \xi_g(|\boldsymbol{r}|, L)}{\int_{L_{\rm min}}^{\infty} L\Phi(L) {\rm dL}} \quad . \tag{1.21}$$

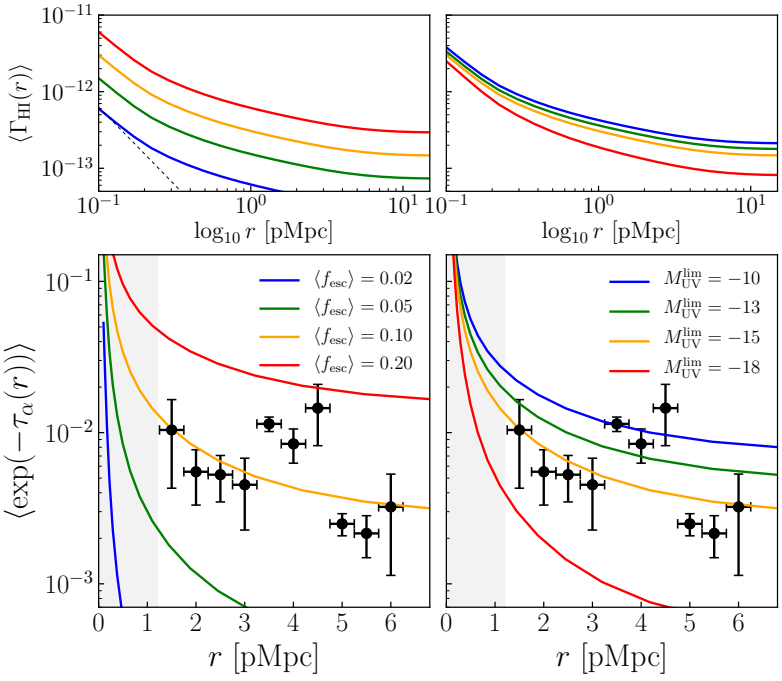

Figure 1.14: The bottom panels show the average Lyman-α transmission (black) around $z \sim 6$ galaxies in the field of $z = 6.48$ quasar J1148+5251. The abscissa indicates the 3D distance, dominated by the parallel component, to Lyman-α forest pixels in the spectra of the background quasar. Models of the expected enhanced photoionisation rate around galaxies (top panels) and corresponding enhanced transmission (bottom panels) are shown for various escape fraction and minimum UV magnitude of contributors. *Reproduced from Kakiichi et al. (2018).*

Equation 1.20 is the convolution of a transfer kernel function ($\frac{e^{-|\boldsymbol{r}|/\lambda_{\rm mfp}}}{4\pi|\boldsymbol{r}|^2}$) and the luminosity weighted cross-correlation. It is important to note that the integrals of the cross-correlation are taken between a minimum luminosity of contributors to reionisation and the brightest objects. This minimum luminosity of contributors is imposed to remedy the absence of a turnover in the best-fit UVLF at $z \sim 6$ (Figure 1.9). The minimum luminosity controls the steepness of enhanced photoionisation rate signal (Figure 1.14) and constrains whether the faintest galaxies do contribute or not to reionisation when fitting the data.

By using the convolution theorem, the integral in equation 1.20 can be simplified by taking the inverse Fourier transform of the product of the transforms of the kernel and the cross-correlation

$$\Gamma_{\rm HI}^{\rm CL}(r) = \bar{\Gamma}_{\rm HI}\left[1 + \int_0^\infty \frac{k^2 {\rm d}k}{2\pi^2} R(k\lambda_{\rm mfp}(z))\langle P_g(k)\rangle_L \frac{\sin kr}{kr}\right] \quad [{\rm s}^{-1}] \quad , \tag{1.22}$$

where $R(x) = \arctan(x)/x$ and

$$\langle P_g(k) \rangle_L = \frac{\int_{L_{\rm UV}^{\rm min}}^{\infty} L_{\rm UV} \Phi(L_{\rm UV}|z) P_g(k, L_{\rm UV}) {\rm d}L_{\rm UV}}{\int_{L_{\rm UV}^{\rm min}}^{\infty} L_{\rm UV} \Phi(L_{\rm UV}|z) {\rm d}L_{\rm UV}} \tag{1.23}$$

is the luminosity-weighted power-spectrum of galaxies at a given redshift. The power-spectrum of galaxies with a luminosity $L_{\rm UV}$, $P_g(k, L_{\rm UV})$, is estimated in Kakiichi et al. (2018) using a conditional luminosity function approach (CLF), constrained by the LBG angular correlation function (Harikane et al. 2018) and the $z \sim 6$ UV LF (Bouwens et al. 2015).

With this Fourier-transform trick, the integral of Equation 1.22 can be computed numerically and gives the overall radial dependence of the enhanced photoionisation rate signal which depends on the minimum UV luminosity of contributors to reionisation. The remaining variable is the effective ionising ouput of galaxies $\langle f_{\rm esc,LyC} \times \xi_{\rm ion} \rangle \propto \overline{\Gamma}_{HI}$ that effectively controls the amplitude of the signal. Therefore, three important parameters of reionisation ($f_{\rm esc,LyC}$, $\xi_{\rm ion}$, $M_{\rm UV}^{\rm min}$) are constrained by a measurement of the enhanced photoionisation rate around bright galaxies.

To transform the predicted enhanced photoionisation rate into an observed enhanced Lyman-α transmission, the Fluctuating Gunn-Peterson approximation is used,which I now derive following the review by Becker et al. (2015a). After reionisation, the IGM is assumed to be in ionisation equilibrium with the UVB. Hence, assuming A-case recombination and ignoring collisional excitation, the density of neutral and ionised hydrogen ($n_{\rm HI}, n_{\rm HII}$) are set by the hydrogen photoionisation rate $\Gamma_{\rm HI}$ and the hydrogen recombination coefficient $\alpha_{\rm HII}(T)$

$$n_{\rm HI}\Gamma_{\rm HI} = n_{\rm HII} n_e \alpha_{\rm HII}(T) \tag{1.24}$$

where n_e is the electron density. We have seen in Section 1.3.1 that transmitted flux in the Lyman-α forest indicates highly ionised gas ($\gtrsim 99.9\%$). Therefore at $z \lesssim 6$ we can safely assume $n_{\rm HII} \simeq n_{\rm H}, n_e \simeq n_{\rm H}(1 + n_{\rm He}/n_{\rm H})$. The ionisation equilibrium

(Eq. 1.24) can be rewritten to compute the neutral hydrogen fraction

$$x_{\rm HI} = \frac{n_{\rm HI}}{n_{\rm H}} \simeq 9.6 \times 10^{-6} \Delta_b \left(\frac{\Gamma_{\rm HI}}{10^{-12}{\rm s}^{-1}}\right)^{-1} \left(\frac{T}{10^4{\rm K}}\right)^{-0.72} \left(\frac{1+z}{5}\right)^3 \times (1+\chi_{\rm He}) \left(\frac{\Omega_b h^2}{0.022}\right) \left(\frac{1-Y}{0.76}\right) \tag{1.25}$$

where Δ_b is the baryon overdensity, $\Gamma_{\rm HI}$ is the hydrogen photo-ionisation rate, Y is the primordial helium mass fraction, T is the IGM temperature, and Ω_b the baryon density. The case-A recombination rate dependence on temperature $\alpha_{\rm HII}(T) = 4.063 \times 10^{-13}(T/10^{-4}{\rm K})^{-0.72}$ from Verner & Ferland (1996) is adopted. The factor $(1+\chi_{\rm He})$ accounts for the supplementary electrons released by ionised helium. By replacing the mean neutral fraction in the Gunn-Peterson optical depth (Eq. 1.6) and inserting a temperature-density relation ($T = T_0 \Delta_b^{\gamma}$), one gets the fluctuating Gunn-Peterson optical depth

$$\tau_\alpha \simeq 11 \Delta_b^{2-0.72(\gamma-1)} \left(\frac{\Gamma_{\rm HI}}{10^{-12}\ {\rm s}^{-1}}\right)^{-1} \left(\frac{T_0}{10^4\ {\rm K}}\right)^{-0.72} \left(\frac{1+z}{7}\right)^{9/2} \quad , \tag{1.26}$$

In this thesis, the fiducial values for the mean IGM temperature ($T_0 = 2 \times 10^4$K) and the exponent for the temperature-density relation ($\gamma = 1.3$) are adopted from lower redshift measurements and simulations (e.g. Boera et al. 2019; D'Aloisio et al. 2019). Once, as described above, the photoionisation rate is predicted at any distance from a detected galaxy ($\Gamma_{\rm HI}(r)$), then Equation 1.26 gives the corresponding decreased opacity (enhanced transmission) that should be observed.

By fitting their model to the enhanced transmission observed around LBGs, Kakiichi et al. (2018) concluded that faint $z \sim 6$ galaxies had $\langle f_{\rm esc,LyC} \rangle \geq 0.08$. Until then, studies of reionisation had relied on combining various observations to infer that galaxies *could collectively emit enough photons* to match the ionising budget *if they had the proper escape fraction*. Here the over-ionised gas is directly correlated with the presence of galaxies to constrain the average ionising output directly. Nonetheless, the potential of the method and the robustness of the signal

must be assessed with a large number of quasar fields and confirmed galaxies.

1.6 Goals of this thesis

Despite growing evidence on the timing and the topology of reionisation, the nature of the sources capable of bringing reionisation to a close before $z \sim 5.5$ remains a mystery. Early galaxies are thought to be capable of providing the bulk of the ionising photons necessary, but this hinges on the hypothesis that their escape fraction $f_{\rm esc,LyC}$ and/or ionising efficiency $\xi_{\rm ion}$ is much higher than what is observed at $z \lesssim 5$. Besides, framing reionisation in terms of the total ionising budget ignores the observed patchiness of the process, which suggests rare and efficient LyC leakers exist alongside weaker sources. Identifying and characterising the sources of reionisation is therefore an important research gap that must be addressed.

The major goals of this thesis are to measure the ionising contribution of galaxies to reionisation and to investigate the relative contribution of faint and luminous galaxies. To do so, I make use of two complementary approaches. Firstly, I study in Chapter 2 the double-peaked Lyman-α profile of an individual luminous galaxy to determine its escape fraction. Secondly, I develop further the statistical correlation of galaxies with the IGM opacity probed by background quasars presented in Section 1.5.3. In particular, I extend the cross-correlation to use metal absorbers instead of LBGs (Chapter 3) and demonstrate how the two-point cross-correlation with Lyman-α transmission spikes is superior to the average transmission measurement by studying LBGs and LAEs in eight $z \gtrsim 6$ quasar fields (Chapter 4). The cross-correlation approach provides the first measurements of the total ionising contribution of all galaxies to reionisation by measuring the impact of clustered galaxies on the $z \sim 6$ IGM, whereas the double-peak method is helpful to investigate which galaxies contribute the most to reionising local H II bubbles.

Finally, I investigate in Chapter 5 a potential evolution or selection bias of high-redshift quasars. Both have important consequences for reionisation studies. On the one hand, a selection bias in high-redshift quasar samples implies that their number density is underestimated, and thus their contribution to reionisation. On the other hand, an evolution of quasars has important consequence for reionisation

observables making use of their spectra, most notably the quasar damping wing measurement, changing our understanding of the timing of reionisation.

In Chapter 6, I aim to discuss the results of the four avenues of investigation to show how they form a coherent picture of the nature of the sources of reionisation. I will outline future work as well as the new observations that can be undertaken to extend the results of this thesis.

This page was intentionally left blank

Chapter 2

A reionisation-era galaxy self-ionising its local HII bubble

This Chapter (except Section 2.2) has been submitted for publication in MNRAS. The SED fitting was done by N. Laporte, and the shell models by T. Garel and A. Verhamme.

2.1 Introduction

The sources capable of emitting sufficient ionising photons to complete hydrogen reionisation by $z \sim 5.5$ still remain elusive (Section 1.4.5). A widely-held view is that intrinsically UV-faint galaxies are the primary contributors, typically leaking $\sim 10\%$ of their Lyman continuum (LyC) photons to the intergalactic medium (IGM, e.g. Robertson et al. 2015). However, to match the relative rapid decline of the neutral fraction at late times, rarer, luminous sources may play a significant role (e.g Naidu et al. 2020). The issue remains unsolved as there is yet no direct way of measuring the escape fraction, $f_{\rm esc,LyC}$, of LyC radiation at high redshift.

At $z < 4$, LyC leakers are being studied in detail providing new insight into the physical conditions under which ionising photons can escape. A picture is emerging where LyC leakage may be linked to the [OIII]/[OII] emission line ratio (Nakajima

et al. 2018), varies geometrically due to low-column density channels which allow the photons to escape (Fletcher et al. 2019) and is correlated with the Lyman-α emission line profile (e.g. Verhamme et al. 2015; Izotov et al. 2018). Of particular interest is the correlation with the velocity separation in double-peaked Lyman-α profiles (e.g. Izotov et al. 2018). As Lyman-α photons are scattered and Doppler-shifted in dense neutral gas before emerging out of resonance on either the blue or red side of the peak, the double peak separation is linked to the H I column density that controls the LyC escape fraction (Verhamme et al. 2015; Kakiichi & Gronke 2019). Moreover, after the Lyman-α photons escape, only a modest amount of neutral gas in the IGM would absorb the blue wing (Dijkstra 2014). Double-peaked Lyman-α emitters thus also constrain the size of any associated ionised bubble (e.g. Mason & Gronke 2020).

Thus far, only two galaxies at $z > 6$ (NEPLA4, $z = 6.54$ (Songaila et al. 2018) and COLA1, z=6.59 (Hu et al. 2016; Matthee et al. 2018)) are known to have a double-peaked Lyman-α profile. Bosman et al. (2020) also recently reported a double-peaked profile in a $z \sim 5.8$ Lyman-Break selected galaxy, Aerith B, in the near-zone of a quasar. The peak separation widths measured in the three Lyman-α profiles have provided providing useful estimates of $f_{\rm esc,LyC}$ in high-redshift galaxies.

In this Chapter, I start by reviewing in Section 2.2 the physics behind the emission of a double-peaked Lyman-α profile and its transfer through a patchy, partially reionised IGM. I then report the discovery of a new galaxy presenting a double-peaked Lyman-α profile at $z = 6.803$, deeper in the reionisation era than those above. Its Lyman-α profile indicates an escape fraction close to unity and is likely to self-ionise its local H II bubble, unlike previously discovered double-peaked LAEs. We discuss whether it is representative of those sources that ended cosmic reionisation. Throughout this Chapter, magnitudes are in the AB system (Oke 1974), and I use a concordance cosmology with $H_0 = 70, \Omega_M = 0.3, \Omega_L = 0.7$. We refer to proper (comoving) kiloparsecs and Megaparsecs as p(c)kpc and p(c)Mpc.

2.2 Physics of double-peaked Lyman-α emission line profiles

The H I Lyman-α line arises from the 2p $\rightarrow$ 1s transition of the hydrogen atom emitting a photon at $\lambda_\alpha = 1215.67\,\text{Å}$. Lyman-$\alpha$ photons can be either emitted as part of a recombination cascade following a photo-ionisation event, or de-excitation after collisional excitation. The Lyman-α line in distant galaxies is mostly due to the recombination of dense neutral gas ($\log N_{HI} \sim 10^{17} - 10^{19}\text{cm}^{-2}$) ionised by young, massive stars (Partridge & Peebles 1967). After emission, Lyman-α photons are resonantly scattered in the dense ISM and cannot escape the galaxy easily. Lyman-α photons are Doppler boosted after each scattering event, and therefore diffuse in frequency space as they travel through the ISM. Osterbrock (1962); Furlanetto & Pritchard (2006) show that Lyman-α photons with an initial offset Δv from line center are re-emitted on average with a velocity $\Delta v - \frac{1}{\Delta v}$, i.e. they are likely re-emitted closer to line center as atoms with increasingly high velocities become rarer. However, the root-mean-square shift is equal to one Doppler width $\Delta\nu_D = \sqrt{2k_BT/m_pc^2}/\nu_\alpha$, where k_B is the Boltzmann constant, T the temperature of the gas, m_p the mass of the proton and c the speed of light. Because the shift scatter is large, and the chance of a photon scattering back towards the line-center decreases as it diffuses outwards in frequency, photons have a non-negligible probability to escape. Moreover, Harrington (1973); Neufeld (1990); Dijkstra et al. (2006) show that they escape at a preferred (blue- or red-) shift from the line center controlled by the H I column density. The rationale is that photons below the escape frequency will be scattered back to line center, but photons above that threshold are increasingly unlikely to be further Doppler boosted out of resonance and will escape more or less at the escape frequency.

As a conclusion, the resonant scattering of Lyman-α photons in the ISM is expected to reprocess the Lyman-α profile to a double-peaked profile, because only photons that have been blue/redshifted enough from line center can escape the resonant absorption and scattering (e.g. Dijkstra 2014 for a review). If the medium is expanding/contracting, the red/blue peak is enhanced at the expense of the other

and the profile is often single peaked. A denser ISM increases the opacity to Lyman-α photons which must be more blue-/red-shifted to escape the Lorentzian wings of the scattering cross-section, leading to a larger peak separation. Izotov et al. (2018) found an empirical relation between the peak separation and the escape of LyC photons, which is set by the ISM H I column density

$$f_{\rm esc,LyC} = \exp(-\sigma_{912} N_{\rm HI}) \tag{2.1}$$

where σ_{912} is the H I cross-section at the Lyman-limit. A double-peaked Lyman-α line profile is therefore an excellent way to measure the escape of LyC photons.

At low-redshift, modelling the Lyman-α line with so-called "shell" models (e.g. Dijkstra et al. 2006; Verhamme et al. 2006, 2015; Gronke et al. 2015 and references therein) has been very successful in reproducing observed Lyman-α profiles (e.g. Verhamme et al. 2006, 2008; Orsi et al. 2012; Gronke 2017). These models compute the radiative transfer of a single-peaked Lyman-α line through a shell of neutral gas and can constrain the column density of the ISM, its Doppler width, the velocity of the expanding shell(s) and sometimes a few additional parameters such as the dust opacity. The most interesting of these quantities is the H I column density, which directly gives the $f_{\rm esc,LyC}$ as we have seen above. Low-redshift studies of so-called Green Pea galaxies have confirmed the effectiveness of modelling double-peak Lyman-α profiles to infer the LyC escape fraction (e.g. Verhamme et al. 2015). Further modelling with high resolution radiative transfer simulations (Kakiichi & Gronke 2019) reproduces the correlation between $f_{\rm esc,LyC}$ and the peak separation of Lyman-α profiles. At high-redshift however, a complication arises as the shape of the double-peaked profile is modified by the transfer of the photons through the neutral IGM during reionisation, complicating both the detection of double-peaks and their interpretation.

Once the Lyman-α photons have escaped the ISM, they are transmitted through the IGM where only a modest amount of neutral gas will absorb the blue wing (Dijkstra 2014 and references therein). If, however, the galaxy sits in an ionised

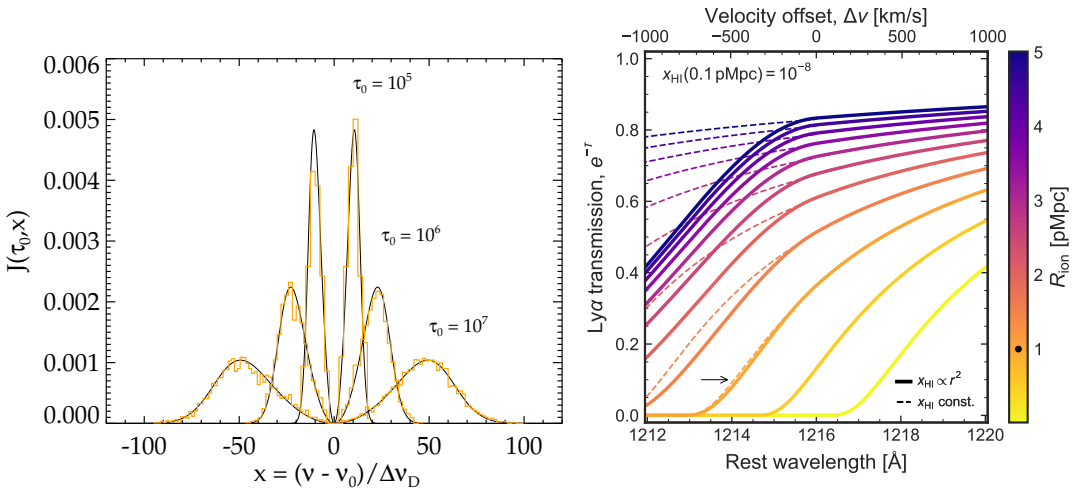

Figure 2.1: **Left:** Double peaked Lyman-α profiles in radiative transfer simulations. The photons are emitted within a static homogeneous sphere, whose increasing opacity τ_0 (controlled by the column density of N_{HI}) leads to higher velocity separations of the peaks. *Reproduced from Orsi et al. (2012)* **Right:** IGM damping wing transmission profile for various ionised bubble sizes. The damping wing (discussed in Section 1.3.3)) can absorb photons redwards of the rest-frame Lyman-α wavelength of the first ionised patch. If the ionised bubble is large enough ($\gtrsim$ 1 pMpc), the blue peak of Lyman-α is not absorbed. *Reproduced from Mason & Gronke (2020).*

bubble with radius r, photons emitted bluewards of Lyman-α are not absorbed by the IGM. Therefore, the Lyman-α profile is only truncated at $\Delta v_{blue} = rH(z)$. A double-peaked profile therefore indicates that the galaxy resides in an ionised bubble, and its radius is constrained by the extent of the blue wing flux (Mason & Gronke 2020).

In this idealised picture, I have so far neglected the IGM damping wing and considered only the opacity in a fully (or partially) homogeneously ionised medium. However, in a patchy reionisation, the photons encounter a succession of neutral and ionised patches and the total opacity is (e.g. Dijkstra 2014)

$$\tau_{\rm IGM}(z, \Delta v) = \tau_{\rm D}(z, \Delta v) + \tau_{\rm HII}(z, \Delta v) \tag{2.2}$$

where $\tau_{\rm HII}(z, \Delta v)$ is the opacity in a diffuse medium (Fig. 2.1, left) and $\tau_{\rm D}(z, \Delta v)$ is damping wing of the neutral IGM (2.1, right). Mesinger & Furlanetto (2008b) show that the opacity averaged over a succession of neutral and ionised gas patches can

be expressed as

$$\tau_{\rm D}(z, \Delta v) = \frac{\tau_{\rm GP}(z)}{\sqrt{\pi}} \overline{x}_D \frac{A_\alpha c}{4\pi\nu_\alpha} \frac{1}{\Delta\nu_{n1}} \tag{2.3}$$

where $\tau_{\rm GP}(z)$ is the Gunn-Peterson opacity (Eq.1.6), A_α is the Einstein A coefficient for the Lyman-α transition, and $\Delta\nu_{n1}$ is the frequency offset in the expanding Hubble flow to the first neutral gas patch encountered. As shown in Figure 2.1 (right), the damping wing can attenuate photons up to several thousands km s^{-1} redwards from the line center. The wavelength dependence, where blue photons are more affected than photons redwards of Lyman-α, explains the difficulty to find double-peaked LAEs at $z \gtrsim 6$ and the decline of the Lyman-α fraction and LAE LF at $z \gtrsim 6$ discussed in Section 1.3.3 (Dijkstra et al. 2011; Mesinger et al. 2015). Nonetheless, if detected these double-peaked profiles offer an unparalleled insight into reionisation era as the profiles can constrain both the size of the ionised bubbles and the escape fraction of the galaxy (Mason & Gronke 2020 and references therein).

2.3 Observations

The target of this study was originally observed as part of a search for rest-frame UV lines signalling AGN activity in bright $z \sim 7$ galaxies (X-Shooter/VLT, ID: 0100.A-0664(A), PI: Laporte). Following earlier detection of He II emission in a galaxy with evidence for strong [OIII] and Hβ emission lines (Laporte et al. 2017b), we searched for similar sources using data from the Hubble and Spitzer Space Telescopes in the Frontier Fields survey (Lotz et al. 2017), applying selection criteria defined in Bouwens et al. (2015). Possible evidence for intense [O III] and Hβ line emission was considered via excess emission in the appropriate IRAC bandpasses (see Labbé et al. 2013; Smit et al. 2014).

Spectroscopic follow-up was conducted with both X-Shooter/VLT and ALMA (Laporte et al. 2017b, Hashimoto et al. 2018b, Laporte et al. 2019) to determine the redshifts, star-formation rates (SFR) and other properties. Among this sample, one

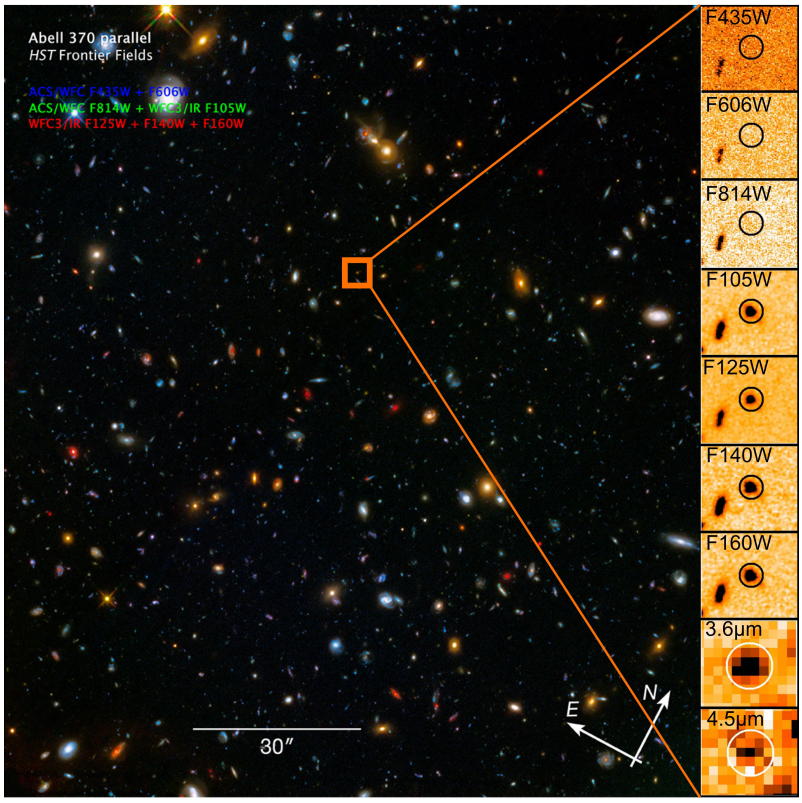

Figure 2.2: **Main panel:** *HST* Frontier Field composite image of the parallel field of cluster Abell 370, in which A370p_z1 was selected. (Composite image credits: NASA, ESA, Z. Levays (STSci)) **Left panels:** Frontier Field image stamps (*HST* + *Spitzer*, 3.5" × 3.5") of A370p z1, showing a clear drop in F105W-F814W, typical of $z > 6.5$ galaxies. The *Spitzer* channels have been decontaminated from the contribution of a southern object (see Section 2.3)

bright galaxy (F125W= 25.15, $z_{\rm phot}$=7.14±0.8, Figure 2.2), hereafter A370p_z1, was observed with X-Shooter/VLT in service mode in October 2018. Observing blocks were defined in order to maximise the exposure time in the NIR arm ($t_{\rm NIR}$ = 900s, $t_{\rm VIS}$ = 819s and $t_{\rm UVB}$ = 756s). The target was centred in a 0.9″ slit using a blind offset from a nearby bright star. After discarding time in poor seeing, the usable exposure time in the VIS arm was 6.3hrs.

The spectroscopic data were reduced using standard X-Shooter ESOREFLEX recipes (v3.3.5). Flux calibrated 2D spectra were stacked using IRAF's *imcombine* and visually inspected for emission lines by two authors (RAM, NL). Stacking with custom ESOREFLEX recipes produced similar results. The stacked 2D spectrum was optimally extracted (Horne 1986) with a boxcar aperture of 1.6" (10 pixels) revealing an emission line doublet at 9484, 9487 Å (Fig. 2.3). No other line was

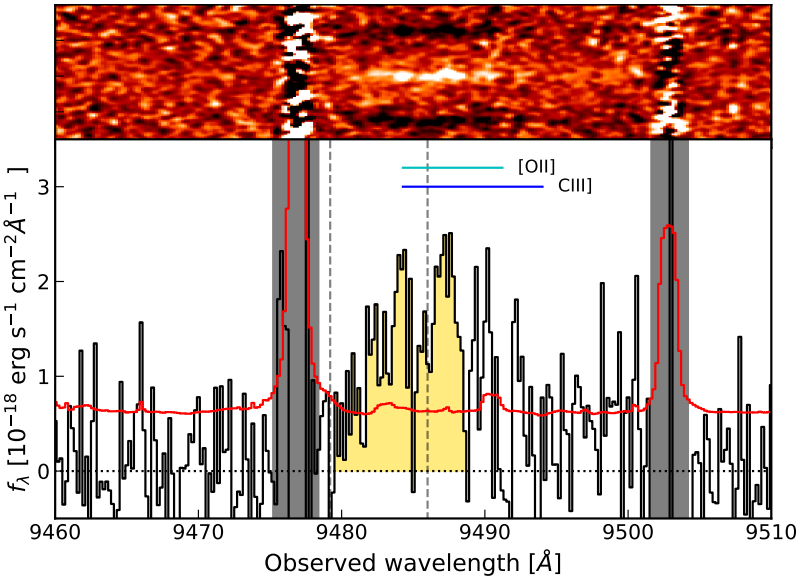

Figure 2.3: 2D spectra of A370p_z1, showing double-peaked Lyman-α emission at $z = 6.803$ with two negative counterparts arising from the telescope dither pattern. The lower panel shows the 1D spectrum (black) and error array (red) with OH sky lines masked in grey.The two peaks are highlighted in yellow. Vertical dotted lines show the maximum extent of the blue wing and the mid-point of the two peaks. The velocity separation of emission line doublets of potential low-redshift redshift interlopers is illustrated by cyan ([OII]$\lambda\lambda 3727, 3729$) and blue (CIII]$\lambda\lambda 1907, 1909$) horizontal lines.

found in the XShooter data.

The width of the trough between the two peaks is about twice the XShooter resolution for the adopted $0.9''$ slit (34 km s^{-1}). To verify that the inter-peak absorption is significant, we compute the residuals of the dip pixels with respect to the flux of the smaller peak (the blue peak). The χ^2 statistic gives a only $P(\chi^2) = 0.00013$ probability that the dip is consistent with Gaussian residuals around the blue peak maximum. The inter-peak absorption is therefore significant at 3.8σ. However, this statistic does not guarantee that the double-peaked profile would be selected by eye when inspecting the 2D spectra. In order to recognise a double-peak, observers look for a few significantly absorbed pixels, preferably consecutive, in-between the peaks. We resampled the spectrum between $\lambda = 9480, 9489$Å assuming a Gaussian noise distribution with variance drawn from the error array. We then identified the maximum pixels on either side of $\lambda = 9485$Å to find the profile peaks. Counting how many pixels are $> 2\sigma$ below the average of the peaks' maximum flux, we found that in $\sim 95\%$ (2σ) of the resampled spectra, there are at least four pixels satisfying this

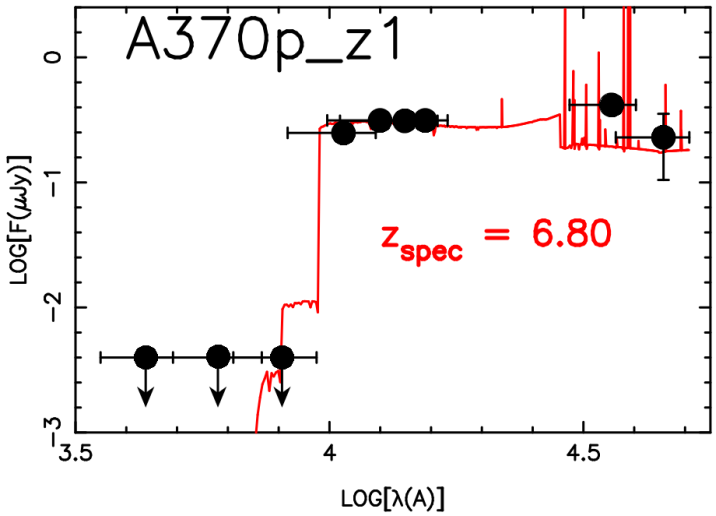

Figure 2.4: Spectral energy distribution based on the photometry (black) with a BAGPIPES fit (red) adopting a redshift $z = 6.80$ from the Lyman-α profile. Note a 3.6μm-4.5μm (leftmost two points) excess likely due to [OIII]+Hβ emission, claimed to be an indicator of high $f_{\rm esc,LyC}$ (see 2.5.2).

criteria, and at least two are contiguous. We used this bootstrap resampling technique to obtain robust errors on the peak velocity separation $\Delta v = 101^{+38}_{-19}$ km s^{-1}. We note that these errors might be slightly underestimated because the peak separation is measured from the maximum of the peaks. A more conservative error estimate based on the resolution of the spectra gives $\Delta v = 101 \pm 48$ km s^{-1}.

The observed peak separation rules out a $z \sim 1.54$ [OII] $\lambda\lambda$ 3727, 3729 Å interloper ($\Delta v = 225$ km s^{-1}) and a $z \sim 4.15$ C III]$\lambda\lambda$ 1907, 1909 Å doublet ($\Delta v = 314$ km s^{-1}). A low-redshift interpretation would imply other lines in the UV, VIS and NIR arms but none was found (e.g. [OIII] and [O II] are detectable with X-shooter to $z \sim 3.8$ and 5.4, respectively). A high redshift solution is also consistent with the Lyman break seen in the SED (Fig. 2.4); a dusty source with a Balmer break at $z \sim 1.5-2$ is inconsistent with the flat SED redwards of 1.5μm. Although it is possible the peaks come from different locations in a single galaxy or a merger, our 2D spectral data indicates both peaks are co-spatial.

We therefore conclude that A370p_z1 has a double-peaked Lyman-α profile at z=6.803 (taken as the mid-point of the two peaks[1]) with a peak velocity separation

[1]The bottom of the absorption trough is the closest to systemic in low redshift double-peaked Green

of $\Delta v = 101^{+38}_{-19}$ km s^{-1}. The Lyman-α rest-frame luminosity ($9.8 \pm 1.0 \times 10^{42}$erg s^{-1}) and equivalent width ($\mathrm{EW}_{\mathrm{Ly}\alpha} = 43 \pm 4$ Å) are similar to those seen in $z \sim 0$ Green Peas (Yang et al. 2017; Izotov et al. 2018) or $\sim 2-3$ double-peaked Lyman-α emitters (LAEs) (e.g. Yamada et al. 2012; Kulas et al. 2012; Hashimoto et al. 2015 see Fig. 2.6). Finally, Matthee et al. (2018) raised the possibility that high-redshift double-peaked Lyman-α could be potentially caused by a foreground absorber in a standard (red-wing only) Lyman-α line. However, the skewness of the red and (blue) peak is $S = 0.70 \pm 0.24 (-0.32 \pm 0.23)$ which is higher than the $S > 0.15$ threshold used for LAEs (Kashikawa et al. 2006). The skewness of the peaks also disfavours the merger interpretation. We searched for evidence of a hard ionisation spectrum or AGN activity but, at the expected location of N IV 1240 Å, C IV 1549 Å, He II 1640 Å, C III]$\lambda\lambda$1907, 1909 Å we do not find any significant emission lines (Fig. 2.5). Table 2.1 summarises the properties of A370p_z1. The uncertainties are derived using the spectral resolution ($R \sim 8900$) and the error array, except for the peak velocity separation which comes from bootstrapping.

2.4 Results

2.4.1 The nature of A370p_z1

We first characterised A370p_z1 by utilising the available deep *HST* and *Spitzer* Frontier Fields imaging (Fig. 2.4, upper panel). We extracted the spectral energy distribution (SED) following the method described in Finkelstein et al. (2013). We neglect any lensing magnification as A370p_z1 is in a parallel field and thus far from the cluster Abell 370. The *Spitzer* 3.6μm and 4.5μm images are contaminated by a point source 1.5" to the south-east. We used GALFIT (Peng et al. 2010) to remove its contaminating contribution and applied a standard aperture correction. We fit the SED using BAGPIPES (Carnall et al. 2018), experimenting with several star formation histories (SFH) adopting single (constant, exponential, burst) and

Peas (Gazagnes et al. 2020), but is rather difficult to measure here given the noise and resolution, hence we approximate it as the mid-point of the peaks.

Table 2.1: Properties of A370p_z1. Limits are quoted at the 2σ level

RA	02h40m14.1s
DEC	-01d37m14.3s
Emission lines	
$F_{\mathrm{Ly}\alpha}$	$(18.4 \pm 1.9) \times 10^{-18}\,\mathrm{erg\,cm^{-2}s^{-1}}$
F_{NV}	$< 1.9 \times 10^{-18}\,\mathrm{erg\,cm^{-2}s^{-1}}$
F_{CIV}	$< 1.3 \times 10^{-18}\,\mathrm{erg\,cm^{-2}s^{-1}}$
F_{HeII}	$< 2.6 \times 10^{-18}\,\mathrm{erg\,cm^{-2}s^{-1}}$
F_{CIII}	$< 0.7 \times 10^{-18}\,\mathrm{erg\,cm^{-2}s^{-1}}$
Lyman-alpha profile	
$z_{Ly\alpha}$	6.803
$\Delta v_{\mathrm{Ly}\alpha}$	101^{+38}_{-19} km s^{-1}.
$\mathrm{FWHM_{blue}}$	82 ± 48 km s^{-1}
$\mathrm{FWHM_{red}}$	120 ± 48 km s^{-1}
$\mathrm{f_{blue}}$	$(7.4 \pm 1.9) \times 10^{-18}\,\mathrm{erg\,cm^{-2}s^{-1}}$
$\mathrm{f_{red}}$	$(10.8 \pm 2.4) \times 10^{-18}\,\mathrm{erg\,cm^{-2}s^{-1}}$
Blue/red flux ratio	0.69 ± 0.24
Blue peak skewness	-0.32 ± 0.23
Red peak skewness	0.70 ± 0.24
$\mathrm{L_{Ly\alpha}}$ (rest-frame)	$(9.8 \pm 1.0) \times 10^{42}\mathrm{erg\,s^{-1}}$
$\mathrm{EW_{Ly\alpha}}$ (rest-frame)	43 ± 4 Å
$f_{\mathrm{esc,LyC}}$ (Izotov et al. 2018)	$> 0.59 (> 0.51)$
$f_{\mathrm{esc,LyC}}$ (RASCAS)	0.99
Photometry and SED fitting (BAGPIPES)	
m_{F435W}	< 29.90
m_{F606W}	< 29.80
m_{F814W}	< 30.00
m_{F105W}	25.41 ± 0.01
m_{F125W}	25.16 ± 0.01
m_{F140W}	25.17 ± 0.01
m_{F160W}	25.16 ± 0.01
$m_{3.6\mu}$	24.85 ± 0.14
$m_{4.5\mu}$	26.19 ± 0.50
$M_{\mathrm{UV}}(m_{\mathrm{F105W}})$	-21.5 ± 0.1
M_*	$(6.55^{+0.14}_{-0.10}) \times 10^9 \mathrm{M_\odot}$
SFR	$12 \pm 6\,\mathrm{M_\odot\,yr^{-1}}$
Age	$50 \pm 4\,\mathrm{Myr}$

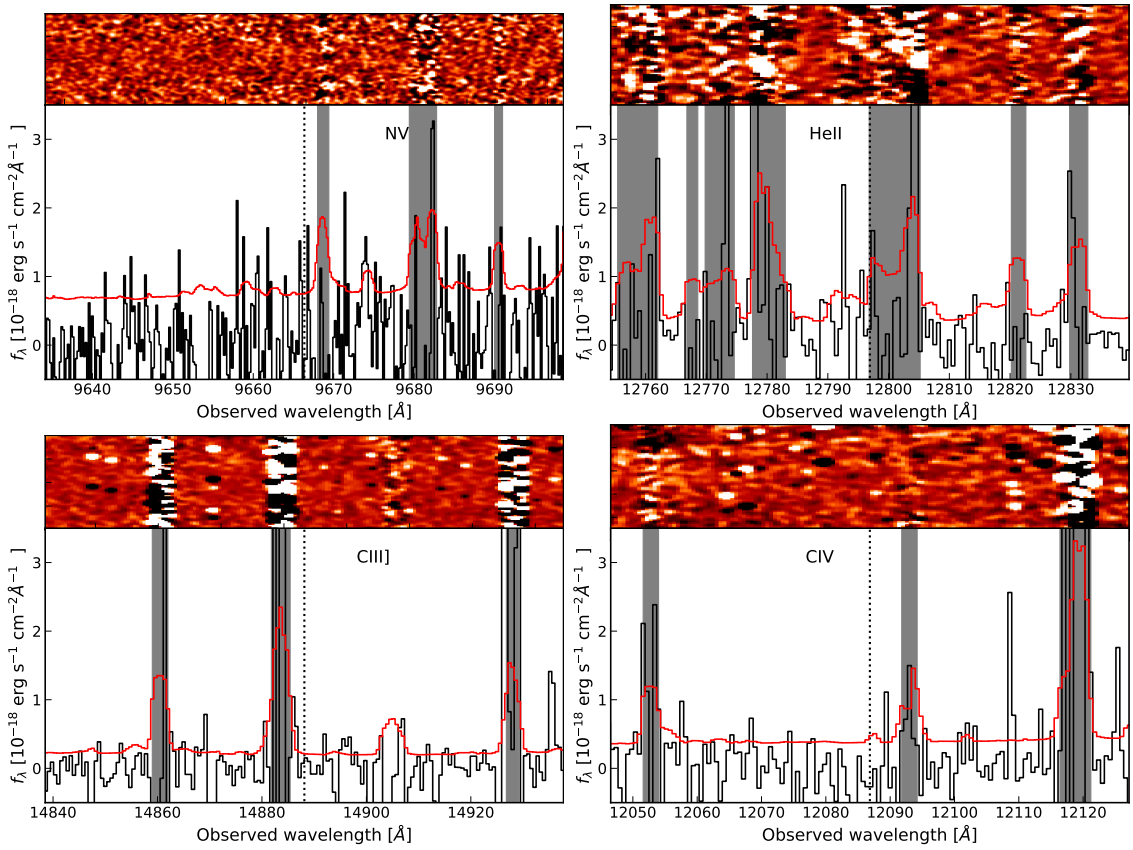

Figure 2.5: 2D and 1D spectrum of A370p_z1, showing the expected location of rest-frame UV lines at $z = 6.803$ (x-axis range: ± 1000 km s^{-1}). Real emission lines should show one bright line at the center of the 2D spectra (upper panel) with two negative counterparts (black) at the top and bottom of the spectrum. The colour scheme is identical to that of Figure 2.3, but the smoothing length is adjusted for the NIR arm. The lower panel shows the 1D spectrum (black) and error array (red) with sky lines masked in grey. Vertical dotted lines show the exact wavelength of the UV lines or the centroid for doublets.

two component models (constant + burst ; exponential + burst). The best-fit SED was a constant SFH model with the following properties : age $= (50 \pm 4)$ Myr, $M_\star = (6.55^{+0.14}_{-0.10}) \times 10^9\, M_\odot$ and a SFR $= (12 \pm 6)\, M_\odot yr^{-1}$ (Fig. 2.4, lower panel, and Table 2.1). We found no preference for an exponentially-declining SFH or a single-burst model. The flux limits at the rest-frame UV lines positions (see further Table 2.1) are consistent with the line flux from the best-fit SED.

As discussed in Section 2.2, a small separation for a double-peaked Lyman-α profile is a strong indicator of a high LyC escape fraction in low-redshift analogues (e.g Gronke 2017; Verhamme et al. 2017). However, Izotov et al. (2018) found their tight empirical relation may not apply for the range $v_{\rm peaks} \lesssim 150$ km s^{-1} which

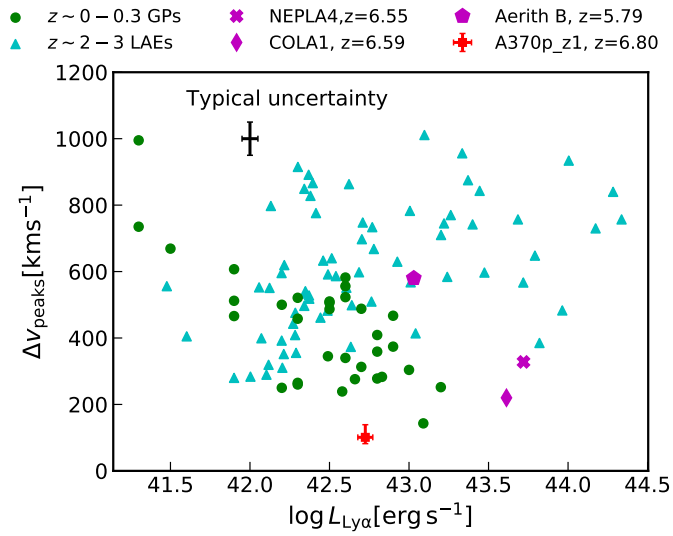

Figure 2.6: Lyman-α peak separation versus Lyman-α luminosities for $z\sim0$ Green Peas (green circles, Yamada et al. 2012) and $\sim 2-3$ double-peaked LAEs (cyan triangles, Yang et al. 2017; Kulas et al. 2012; Hashimoto et al. 2015), Aerith B (magenta pentagon, Bosman et al. 2020), NEPLA4 (magenta cross, Songaila et al. 2018), COLA1 (magenta diamond Matthee et al. 2018) and A370p_z1 (red square).

Table 2.2: RASCAS shell model parameter grid searched

b [km s^{-1}]	20, 80, 140
$v_{\rm exp}$ [km s^{-1}]	0, 20, 50
log N$_{\rm HI}$/[cm^{-2}]	15, 16, 17, 18
$\tau_{\rm d}$	0, 0.5, 1
FWHM$_{Ly\alpha}$ [km s^{-1}]	100, 200, 300, 400, 500

was not probed by their observations and where their relation would predict an unphysical $f_{\rm esc,LyC} > 100\%$. We therefore put a maximum of 100% to the polynomial function so it does not result in unphysical values. We then compute the escape fraction for each of the resampled spectra (see Section 2.3) to obtain a 2σ lower limit on A370p_z1 LyC escape fraction $f_{\rm esc,LyC} > 59\%$. Using the conservative error from the XShooter resolution gives $f_{\rm esc,LyC} > 51\%$ (2σ).

In order to better estimate the escape fraction, we compare the observed profile with double-peak shell models. We use the RASCAS 3D Monte-Carlo code (Michel-Dansac et al. 2020) to generate a grid of Lyman-alpha radiation transfer simulations in spherical geometries, allowing for static and expanding gas configurations. In these typical shell models (e.g. Dijkstra et al. 2006; Verhamme et al. 2008), H I gas and dust are distributed homogeneously around a central point source. The shell

is described by four physical parameters, namely the expanding velocity $v_{\rm exp}$, the HI column density N_{HI}, the dust opacity τ_d, and the Doppler parameter b which accounts for the thermal/turbulent gas motions (see Table 2.2 for the parameter grid used). The intrinsic emission is assumed to be a Gaussian line centred on the systemic redshift with a width set by the FWHM. Given the nearly symmetric double peak profile of A370p_z1 and the small peak separation, we restrict our analysis to relatively small N_{HI} and $v_{\rm exp}$ values because it is well-known that high column densities and shell velocities would significantly broaden the line and erase the blue peak respectively (Verhamme et al. 2006). We perform a quantitative comparison between the observed line profile and the models using the χ^2 statistics. We find that models minimising the reduced χ^2 preferentially select low N_{HI} ($\log \mathrm{N_{HI}}/[\mathrm{cm}^{-2}] = 15$), static geometries ($v_{exp} = 0$), low dust content, small ISM Doppler parameter values ($b = 20$) and relatively broad input lines ($\mathrm{FWHM}_{Ly\alpha} > 300$ km s^{-1}).

We show the best fit model in Figure 2.7 which corresponds to the following parameter set: $\log \mathrm{N_{HI}}/[\mathrm{cm}^{-2}] = 15, \mathrm{b} = 20$ km s^{-1}, $\mathrm{v_{exp}} = 0, \mathrm{FWHM} = 300$ km s^{-1}. We can derive the LyC escape fraction from the best-fit column density $f_{\rm esc,LyC} = \exp\left(-\sigma_{912}\mathrm{N_{HI}}\right) = 99\%$, where $\sigma_{912} = 6.35 \times 10^{-18}$ is the H I photoionisation cross-section at the Lyman limit. While several models could match the positions of peak emission within the errors, the $\log \mathrm{N_{HI}}/[\mathrm{cm}^{-2}] = 10^{15}$ model is the only one to reproduce the shallow central depression of the profile which is the important signature of a low H I opacity (Figure 2.7). Searching a finer parameter grid is beyond the scope of this work, but we note that even when adopting a higher column density $\log \mathrm{N_{HI}} = 10^{16}$ cm^{-2}, the escape fraction remains very high (94%).

2.4.2 Did A370p_z1 self-ionise its local H II bubble?

The detection of the blue peak in the Lyman-α indicates A370p_z1 sits in a large ionised bubble, otherwise the damping wing of even a partially neutral IGM would have absorbed it. Given its high escape fraction, we now consider whether A370p_z1 could have self-ionised its local H II bubble.

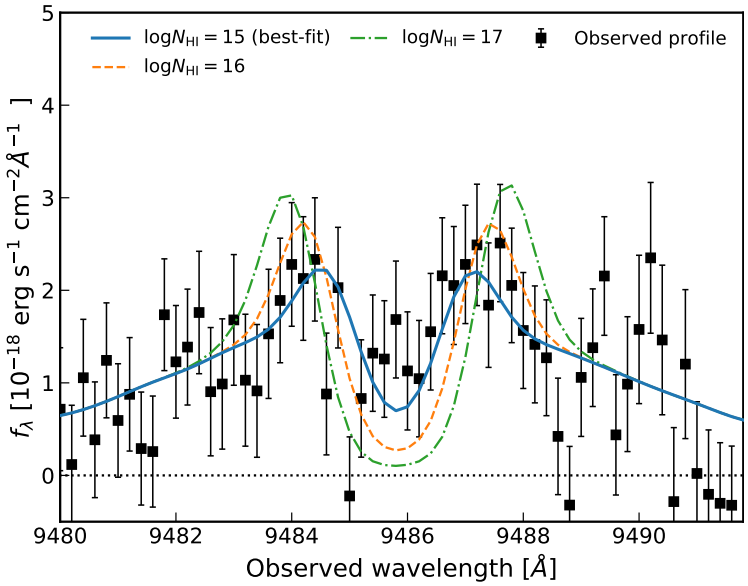

Figure 2.7: Comparison of RASCAS models, smoothed by the resolution of XShooter, of Lymanα transfer through non-expanding shells of homogeneous gas with the observed Lyman-α profile. The best-fit model (blue) has $\log \mathrm{N_{HI}}/[\mathrm{cm}^{-2}] = 10^{15}$ which correspond to a high LyC escape fractions $f_{\mathrm{esc,LyC}} = 99\%$. We also show examples of models with $\log \mathrm{N_{HI}}/[\mathrm{cm}^{-2}] = 10^{16}(10^{17})$ (dashed orange, dashed-dotted green), which would imply $f_{\mathrm{esc,LyC}} = 94\%(53\%)$ and have $\Delta\chi^2 \sim 0.8(2.1)$ compared to the best-fit model.

The blue wing extends to $\lambda = 9479.2$ Å, ≈ 215 km s^{-1} from the line centre, corresponding to a physical distance $r_{\mathrm{HII}} > 0.26 \pm 0.040$ pMpc. This estimate neglects any velocity offset between the Lyman-α absorption dip from which we have derived the redshift and the systemic redshift as defined by more reliable tracers such as nebular absorption lines. These velocities are however found to be $\lesssim 200$ km s^{-1} (Gazagnes et al. 2020), which would therefore require a larger ionised bubble for the blue wing to be transmitted ($\lesssim 0.5$ pMpc). We also neglect peculiar velocities of the galaxy with respect to the ionised bubble gas which would redshift(blueshift) the Lyman-α photons and decrease(increase) the bubble size needed for the blue wing to escape.

We now estimate the volume that could have been reionised by A370p_z1 by redshift $z = 6.803$ and whether its radiation is sufficient to reduce the opacity of the surrounding gas to permit the blue wing of Lyman-α to escape. Assuming no recombination, the ionising bubble created by a single galaxy in the reionisation era

is (e.g. Cen & Haiman 2000)

$$R_S \approx \left(\frac{4 f_{\mathrm{esc,LyC}} \xi_{\mathrm{ion}} L_{UV} t_{\mathrm{em}}}{3\pi \langle n_{HI} \rangle} \right)^{1/3} \quad [\mathrm{pMpc}] \tag{2.4}$$

where t_{em} is the duration of LyC leakage from a source with intrinsic ionising efficiency ξ_{ion} and escape fraction $f_{\mathrm{esc,LyC}}$ and $\langle n_{HI}(z) \rangle \approx 8.5 \times 10^{-5} \left(\frac{1+z}{8}\right)^3 \mathrm{cm}^{-3}$ is the mean hydrogen density of the IGM. The typical ionising efficiency of $M_{\mathrm{UV}} = -22$ galaxies at $z \sim 5$ is $\log \xi_{ion} \simeq 25.4$ cgs (Bouwens et al. 2015). However, recently it has been claimed that some $z > 7$ galaxies have enhanced ionising efficiencies Stark et al. (2015, 2017). We therefore derive an estimate of the ionising efficiency from the Lyman-α line following (Sobral & Matthee 2019). They relate the Lyman-α and Hα luminosities $L_{H\alpha} = \frac{L_{\mathrm{Ly}\alpha}}{8.7 f_{esc}(\mathrm{Ly}\alpha)}$ using the Lyman-α escape fraction $f_{esc}(\mathrm{Ly}\alpha)$. Combining this relation with Eq. 1.16 where we derived ξ_{ion} from the Hα luminosity, we obtain

$$\xi_{\mathrm{ion}}(\mathrm{Ly}\alpha) = \frac{\mathrm{L_{Ly\alpha}}/(1 - \mathrm{f_{esc,LyC}})}{\mathrm{L_{UV}\, c_{H\alpha}\, 8.7 f_{esc}(Ly\alpha)}} \quad [\mathrm{erg^{-1}Hz}] \tag{2.5}$$

where $c_{H\alpha}$ is the recombination rate of LyC photons into Hα lines (Kennicutt 1998). Sobral & Matthee (2019) further show that for $z \sim 2.6$ LAEs, $f_{esc}(\mathrm{Ly}\alpha) \simeq 0.0048 \mathrm{EW}_{\mathrm{Ly}\alpha}$, making the final expression

$$\xi_{\mathrm{ion}}(\mathrm{Ly}\alpha) = \frac{\mathrm{L_{Ly\alpha}}/(1 - \mathrm{f_{esc,LyC}})}{\mathrm{L_{UV}\, c_{H\alpha}\, (0.042 EW_{Ly\alpha})}} \quad [\mathrm{erg^{-1}Hz}] \tag{2.6}$$

Assuming case B recombination and the low-redshift trends of $f_{esc}(\mathrm{Ly}\alpha) \propto \mathrm{EW}_{\mathrm{Ly}\alpha}$ hold a higher redshift, we find $\log \xi_{\mathrm{ion}} \simeq 26.4$ cgs for A370p_z1. In the following, we indicate results based on the higher ionising efficiency in parenthesis. Assuming an $f_{\mathrm{esc,LyC}} \approx 0.9$, 50 Myr is a sufficient time for A370p_z1 to create an ionising bubble with radius $R_S \simeq 0.86(1.10)$ pMpc. This is more than three times larger than the distance at which the blue wing of Lyman-α is still transmitted. Therefore, it is plausible that A370p_z1 is able to self-ionise its surrounding bubble, even if the

escape fraction was $\approx 90\%$ only for a small fraction of its lifetime (e.g. $R_S(t_{\rm em} = 0.2t_{\rm age}) \approx 0.50(1.05)$ pMpc).

Being able to grow a Strömgren sphere larger than the distance required for the blue wing of Lyman$-\alpha$ to redshift out of resonance is a necessary but not sufficient condition for A370p_z1 to be solely responsible for its ionised bubble. This is because the Gunn-Peterson optical depth is virtually zero for neutral fractions as low as $10^{-4.5}$ (see Becker et al. 2015a for a review). Therefore, in the absence of an elevated photoionisation rate, the blue peak would readily be resonantly absorbed by even small pockets of neutral gas within the ionised bubble. We therefore examine whether A370p_z1 can maintain such a high photoionisation rate at the edge of its bubble or if an additional population of clustered UV-faint galaxies is required.

The local photoionisation rate due to A370p_z1 is

$$\begin{aligned}\Gamma_{\rm HI}^{\rm A370p\_z1}(r) &= \frac{\alpha_g \sigma_{912}}{\alpha_g + 3} \frac{f_{\rm esc,LyC} \xi_{\rm ion} L_{\rm UV}}{4\pi r^2} {\rm e}^{-{\rm r}/\lambda_{\rm mfp}} \\ &\simeq 0.8(7.4) \times 10^{-11} \left(\frac{r}{0.1{\rm pMpc}}\right)^{-2} {\rm s}^{-1} \end{aligned} \tag{2.7}$$

where α_g is the extreme UV spectral slope, and $\lambda_{\rm mfp}$ is the mean free path of LyC photons. We assume $\alpha_g = 2$ (e.g. Kuhlen & Faucher-Giguère 2012) and $\lambda_{\rm mfp} \simeq 6.0 \left(\frac{1+z}{7}\right)^{-5.4}$ pMpc (Worseck et al. 2014). The fluctuating Gunn-Peterson approximation links the photoionisation rate Γ to the Lyman-α opacity (see Section 1.5.3 and Becker et al. 2015a for a review)

$$\tau_\alpha \simeq 11 \Delta_b^{2-0.72(\gamma-1)} \left(\frac{\Gamma_{\rm HI}}{10^{-12}\ {\rm s}^{-1}}\right)^{-1} \left(\frac{T_0}{10^4\ {\rm K}}\right)^{-0.72} \left(\frac{1+z}{7}\right)^{9/2} \tag{2.8}$$

where Δ_b is the baryon overdensity and the temperature T_0 is assumed to be 10^4 K.

At a constant mean density of $\Delta_b = 1$, the photoionisation rate due to A370p_z1 is sufficient to have an average Lyman-α transmission in the bubble $\overline{\mathcal{T}_\alpha}({\rm blue\ wing}) = 0.25(0.69)$, and a transmission at the edge of the blue peak $\mathcal{T}_\alpha^{\rm blue\,peak} = 0.51(0.93)$. We note this does not take into account expansion in the Hubble flow, ignores the effect of the IGM damping wing or overdensities associated with the galaxy,

enhancements we consider beyond the scope of this discovery work. Recently, Mason & Gronke (2020) have laid out an extensive framework to model high-redshift double-peaked Lyman-α emitters. Their modelling suggests that a source with the luminosity of A370p_z1 and $f_{\rm esc,LyC} = 1$ could carve an ionised bubble with $r_{\rm ion} \sim 0.6$ pMpc sufficient to permit the blue peak to escape up to ~ 0.2 pMpc, in good agreement with our results.

Finally, we checked that A370p_z1 does not lie in an overdensity of $z \sim 6.8$ objects. We find 28 F105W-F814W dropout galaxies in the A370p field with $M_{\rm F125W} < 28$ whose 1σ photometric redshift is at least partially in the redshift interval $6.3 < z < 7.3$. This is in good agreement to that expected (28 ± 5) from the Bouwens et al. (2015) luminosity function. We thus conclude that A370p_z1 is very likely to have contributed to the totality or the large majority of the LyC photons in its surrounding ionised bubble.

2.5 Discussion

2.5.1 Differences and similarities between NEPLA4, COLA1 and A370p_z1

We now apply the methodology described in the previous section to determine which $z > 6$ double-peaks (A370p_z1, COLA1 and NEPLA4) can grow a H II bubble and ionise it sufficiently to permit blue peak photons to escape. We leave Aerith B aside as the ionised bubble created by the nearby quasar was studied in detail by Bosman et al. (2018). For this comparison, we assume that the redshift of all three objects is taken from the mid-point of the two Lyman-α peaks, which is closer to the central absorption and a better tracer of the nebular absorptions redshift (Gazagnes et al. 2020). To facilitate the comparison between objects, we assume an age of 10 Myr for each galaxy (which matches the estimate for COLA1 Matthee et al. (2018), but is lower than what we measure for A370p_z1). This only affects the Strömgren bubble radii which are proportional $\propto t_{em}^{1/3}$ and can be rescaled accordingly if needed. The escape fractions, UV magnitudes and extent of the blue wings r_α presented in Table 2.3 are taken from Matthee et al. (2018) and Songaila et al. (2018 and private

communication from A. Songaila.).

Interestingly, we find that all double-peaks can grow a Strömgren sphere as large as the bubble size derived from the blue wing maximum velocity offset. However, only A370p_z1 can grow a bubble that is $2-4$ times larger (depending on the ionising efficiency). This is important because the calculated Strömgren sphere radius is thought to be larger than the bubble size derived from the maximum extent of the Lyman-α blue wing (Mason & Gronke 2020). Indeed, as the resonant cross-section of Lyman-α is large, ionised gas can still be optically thick to Lyman-α photons and the optically thin region r_α is much smaller than the Strömgren sphere. Therefore, the most significant test of whether a galaxy is self-ionising its local bubble, or if additional faint sources are needed to let the blue peak escape, is to compute the opacity to Lyman-α photons. We find that COLA1 and NEPLA4 are unable to solely ionised the CGM/IGM sufficiently to allow blue peak photons to escape. The predicted opacity at the blue peak is 0.6% in the most favourable scenario for COLA1, and always zero for NEPLA4. However, the blue peaks are clearly detected, with an observed blue/red peak flux ratio of 0.31 for COLA1 and 0.6 for NEPLA4. We conclude that additional sources are needed to ionise their H II bubble. In contrast, A370p_z1 is able to maintain its bubble sufficiently ionised on its own in all scenarios within the large 1σ errors of the peak flux ratio. Collectively, the four currently known high-redshifts double-peaks present a large range of cases from a source not contributing to reionisation (Aerith B) to a powerful source ionising its H II bubble (A370p_z1), and intermediate cases with significant $f_{\rm esc,LyC}$ but probably surrounded with faint leakers which keep their H II bubble highly ionised (NEPLA4,COLA1).

2.5.2 Implications for reionisation

We have shown that A370p_z1 is possibly the first convincing example of a source capable, on its own, of creating a significant ionised bubble and maintaining this state so that photons escape bluewards of Lyman-α. A key question, therefore, is whether

Table 2.3: Comparison of the ionising properties of the three known $z \sim 6.5$ double-peaks. Strömgren radii are computed assuming ages of 10 Myr.

	NEPLA4	COLA1	A370p_z1
$M_{\rm UV}$	-21.8	-21.6	-21.5
$f_{\rm esc,LyC}(\Delta v)$	0.11	0.29	≈ 0.9
EW$_{Ly\alpha}$ [Å]	176	120	43
r_α [pMpc]	0.31	0.31	0.26
Assuming $\xi_{\rm ion} = 10^{25.4}$			
r_S [pMpc]	0.31	0.37	0.50
$\langle T_\alpha \rangle$(blue wing)	0.09	0.13	0.25
T_α(blue peak)	1×10^{-16}	7×10^{-5}	0.51
Deriving $\xi_{\rm ion}$ from Lyman-α (Sobral & Matthee 2019)			
$\xi_{\rm ion}$	$10^{25.38}$	$10^{25.66}$	$10^{26.38}$
r_S [pMpc]	0.28	0.45	1.05
$\langle T_\alpha \rangle$(blue wing)	0.09	0.17	0.69
T_α(blue peak)	3×10^{-17}	6×10^{-3}	0.93
Peak flux ratio	≈ 0.6	0.31 ± 0.03	0.93 ± 0.28

it is an exceptional source or representative of a larger population of luminous objects responsible for concluding cosmic reionisation. Although many luminous $z > 6$ galaxies have now been spectroscopically confirmed using the Lyman-α line (e.g. Zitrin et al. 2015; Laporte et al. 2017a,b; Stark et al. 2017; Songaila et al. 2018; Stark et al. 2018; Hashimoto et al. 2018b; Taylor et al. 2020), the majority did not have the spectral resolution nor the sensitivity to detect a double-peaked Lyman-α profile, especially if the blue wing were somewhat more absorbed than is the case in A370p_z1. Nonetheless, the unusually high confirmation rate of Lyman-α emission in the Roberts-Borsani et al. (2016) galaxies with strong IRAC 4.5μm excesses might be explained if they were efficient leakers that carved their own ionised bubbles (Zitrin et al. 2015; Stark et al. 2017). Alternatively, of course, there may be associated faint sources and/or AGN activity that contribute to the ionising flux.

Searching for a larger sample of $z > 6$ double-peaked Lyman-α emitters is therefore a promising way of studying both the sources of reionisation and their surrounding H II bubbles with the growing modelling capabilities highlighted above. The rest-frame optical lines of these luminous $z > 6$ sources will be detectable with *JWST*, enabling us to characterise, amongst other quantities, their intrinsic ionising output (see further Chapter 6).

Chapter 3

The cross-correlation of metal absorbers with the reionising IGM

This Chapter has been published in Meyer et al. (2019a MNRAS, 483, 19). The data reduction of the quasar spectra was done by S.E.I. Bosman. The analytical model was developed in collaboration with K. Kakiichi.

3.1 Introduction

As discussed in Chapter 1, the escape fraction of the sources of reionisation is the most uncertain physical quantity in the ionising photon budget. In the previous chapter, I showed how the escape fraction can be inferred from the Lyman-α profile of exceptional and bright galaxies such as A370p_z1. I have argued that more $z > 6$ double-peaked Lyman-α emitters can be found by selecting galaxies with strong [OIII] lines, but such samples will remain small for several years. Luminous galaxies selected in that way from broadband images are only the tip of the bright end of UV LF. However, multiple analytical models (Section 1.3.4) and simulations (Section 1.5.1) of reionisation suggest that faint galaxies ($M_{\rm UV} \gtrsim -12$) beyond the reach of *HST* and even *JWST* play a major, if not dominant, role in the ionising

photon budget. A better understanding of reionisation is therefore contingent on the determination of $f_{\rm esc,LyC}$ for all galaxies with ($-22 \gtrsim M_{\rm UV} \gtrsim -10$), or at least of the ensemble average escape fraction of reionisation-era galaxies.

As described in Section 1.5.3, the cross-correlation of high-redshift galaxies with the opacity of the surrounding IGM can be used to constrain the average escape fraction of $z \sim 6$ galaxies (as illustrated in Figures 1.13 and 1.14). The method of Kakiichi et al. (2018) offers for the first time the possibility to measure (indirectly) the average escape fraction of a population in a few cMpc^3 volume at $z \sim 6$. Because the cross-correlation essentially measures the local UVB enhancement due to the faint galaxy population clustered around a bright LBG, it represents a welcome and complementary approach to the search for rare double-peaked LAEs. This theoretical prediction is nonetheless difficult to exploit because it requires a large amount of observing time to obtain i) a medium-high SNR spectrum of the background quasar ii) deep photometry to select dropout candidates iii) confirmation spectra for all LBG candidates chosen. This suite of observations must then be repeated in multiple fields to help mitigate cosmic variance and obtain a robust result (see Chapter 4).

In this chapter, I aim at measuring a similar correlation of galaxies with the IGM transmission but using a different tracer of galaxy overdensities than the LBGs used in Kakiichi et al. (2018). Triply-ionised carbon (C IV) is the most common metal absorbing species in quasar spectra (Becker et al. 2009; D'Odorico et al. 2013 , thereafter DO13) and, at low redshifts, associated with the metal-enriched halos of galaxies (e.g. Adelberger et al. 2003, 2005; Steidel et al. 2010; Ford et al. 2013; Turner et al. 2014). Metal absorbers including C IV have been detected up to $z \sim 7$ (e.g. Pettini et al. 2003; Ryan-Weber et al. 2009; Becker et al. 2009; D'Odorico et al. 2010, 2013; Bosman et al. 2017; Codoreanu et al. 2018; Cooper et al. 2019) due to the significant increase in the number of high-SNR $z > 6$ quasar spectra (see Section 1.4.1). However, little is known about the appropriate host galaxies which are thought to have quite faint UV luminosities (Becker et al. 2015b). It is however expected that C IV absorbers lie $\lesssim 100$ pkpc from their host (e.g. Oppenheimer et al.

2009; Keating et al. 2016; Bird et al. 2017). More recently, (D'Odorico et al. 2018) reported the recent detection of a galaxy 40 pkpc away from a Damped Lyman-α Absorber (DLA) at $z \simeq 5.94$. Although there is no evidence for a consistent link between DLAs and C IV absorbers at that redshift, the authors also report the potential detection of a weak associated C IV absorption. This would support the idea that potential host galaxies can be found indeed very close to C IV absorbers. Finally, Diaz et al. (2020) recently reported the discovery of a subluminous LAE at 11 pkpc from a C IV absorber, supporting the above hypotheses.

Metal-tracing clustered faint sources of ionising photons should, in principle, allow us to probe the ionising capability of intrinsically faint galaxies well beyond reach of current spectroscopic facilities. Moreover, because metal absorbers lie directly on the quasar sightline, we can probe the IGM transmission around their hosts on the scales of $0.1 \lesssim r \lesssim 1$ pMpc unattainable in the approach introduced in Kakiichi et al. (2018) which uses foreground LBGs in the quasar field. In this Chapter, I take advantage of a large sample of $z > 5.4$ quasar spectra to study the abundance and distribution of C IV absorbers. We then study the 1D correlation of these absorbers with the IGM transmission measured in the Lyman-α forest of the quasar to assess their impact on the IGM. I focus on C IV absorbers at $4.3 < z < 6.2$, but future work will include other metals and potentially studies of the redshift evolution of such correlations.

The plan for this Chapter is as follows. Section 3.2.1 introduces our observational sample of quasar spectra and the initial data reduction. Section 3.2.3 details the methods and results of a semi-automated search for C IV absorbers. I present in Section 3.3 the new constraints on the $4.3 < z < 6.2$ C IV cosmic density derived from our large sample. The measurement of the C IV-IGM transmission 1D correlation is also presented, and Section 3.4 discussed two models of the said correlation. In Section 3.5 I then discuss the nature of the C IV absorbers host galaxies, our evidence for an enhanced transmission in the IGM surrounding C IV absorbers, and the derived constraint on the product of the escape fraction and the LyC photon production efficiency. I conclude in Section 3.6 with a brief summary of the findings

Table 3.1: Quasar sightlines used in this work. References: (1) McGreer et al. (2015); (2) Bosman et al. (2018); (3) Becker et al. (2015b); (4) Becker et al. (2006); (5) Eilers et al. (2017); (6) reduced from online archives in this work (see Table 3.2).

quasar name	z	Instrument	SNR	ref.
J1148+0702	6.419	HIRES	29.7	(1)
J0100+2802	6.30	XShooter	85.2	(2)
J1030+0524	6.28	XShooter	28.0	(1)
J0050+3445	6.25	ESI	24.4	(3)
J1048+4637	6.198	HIRES	29.2	(4)
J1319+0950	6.132	XShooter	96.8	(3)
J1509–1749	6.12	XShooter	88.9	(1)
J2315–0023	6.117	ESI	29.8	(3)
J1602+4228	6.09	ESI	33.3	(2)
J0353+0104	6.072	ESI	80.7	(3)
J0842+1218	6.07	ESI	18.0	(6)
J2054–0005	6.062	ESI	39.5	(3)
J1306+0356	6.016	XShooter	55.8	(1)
J1137+3549	6.01	ESI	31.7	(2)
J0818+1722	6.00	XShooter	114.0	(6)
J1411+1217	5.927	ESI	15.9	(1)
J0148+0600	5.923	XShooter	128.0	(3)
J0005–0006	5.85	ESI	28.8	(5)
J0840+5624	5.844	ESI	17.6	(1)
J0836+0054	5.81	XShooter	93.4	(1)
J0002+2550	5.80	ESI	121.0	(6)
J1044–0125	5.782	ESI	49.2	(3)
J0927+2001	5.772	XShooter	73.7	(3)
J1022+2252	5.47	ESI	19.0	(6)
J0231–0728	5.42	XShooter	115.0	(6)

of this study and future prospects for this new method measuring the escape fraction at the end of the reionisation era.

For this study I adopted the Planck 2015 cosmology $(\Omega_m, \Omega_\Lambda, \Omega_b, h, \sigma_8, n_s)$ =(0.3089, 0.6911, 0.04860, 0.6774,0.8159, 0.9667) (Planck Collaboration et al. 2016). I use pkpc and pMpc (ckpc and cMpc) to indicate distances in proper (comoving) units.

3.2 Methods

3.2.1 Observations

Table 3.2: Quasar sightlines in addition to those listed in Bosman et al. (2018). These spectra were reduced either from Keck Observatory Archive (KOA) data for the ESI spectrograph, or from the XShooter archive.

quasar name	z	Instrument	$t_{\rm total}(s)$	SNR	ref.	PID	P.I.
J0842+1218	6.07	ESI	2400	18	De Rosa et al. (2011)	U085E	R. Becker
J0002+2550	5.80	ESI	22200	121	Fan et al. (2004)	H31aE	A. Cowie
						H46aE	Kakazu
J1022+2252	5.47	ESI	6000	19	–	U130Ei	G. Becker
J0818+1722	6.00	XShooter	20750	114	Fan et al. (2006),	084A-0550	D'Odorico
					D'Odorico et al. (2011)	086A-0574	De Rosa
						088A-0897	De Rosa
J0231–0728	5.42	XShooter	21600	115	Becker et al. (2012)	084A-0574	G. Becker

Our sample consists of 25 optical spectra of quasars with $z_{\rm source} > 5.4$ originating from the Echellette Spectrograph and Imager (ESI) on the Keck II telescope (Sheinis et al. 2002), the XShooter instrument on Cassegrain UT2 (Vernet et al. 2011), and the High Resolution Echelle Spectrometer (HIRES, Vogt et al. 1994) as shown in Table 3.1. Out of these, 20 spectra are re-used from the quasar sample of Bosman et al. (2018). These spectra were selected for their high signal-to-noise ratios (SNR > 17) measured over the $1265 < \lambda < 1275$ Å range via

$$\mathrm{SNR} = \left\langle \frac{F}{\epsilon} \right\rangle \cdot \sqrt{N_{60}}\,, \tag{3.1}$$

where F is the flux, ϵ is the error, and N_{60} is the number of spectral bins covering 60 km s^{-1} [1]. An exception is the quasar J1411+1217 which is included despite its relatively poor SNR ($= 15.9$) due to the presence of a particularly broad C IV absorber. Out of these objects, 7 originate in a study from McGreer et al. (2015) (3 of which were independently re-reduced), 6 from Becker et al. (2015b), one from Becker et al. (2006), one from Eilers et al. (2017) (re-reduced), and 5 from Bosman et al. (2018) (of which 2 are archival).

Together with these 20 sightlines, we reduced 5 additional spectra from the Keck Observatory Archive[2] and the XShooter search tool for the ESO Science Archive Facility[3] as summarised in Table 3.2. The spectra were extracted optimally making use of the calibration files (flat fields and standard star exposures) available in the archives for each set of observations. After performing sky subtraction, different observations of the same sightline are combined when necessary. The implementations of optimal extraction, sky subtraction, and telluric correction used herein are outlined in more detail in Horne (1986), Kelson (2003) and Becker et al. (2012) respectively. Our final sample contains ESI spectra with either 55.9 km/s or 74.6 km/s resolution depending on the slit, while XShooter spectra have either a 28.0 km/s or 34.1 km/s resolution and HIRES spectra have 6 km/s resolution. We have

[1] The scale of 60 km/s is a convention chosen in Bosman et al. (2018) as a suitable intermediate scale for their large range of quasar spectra.

[2] https://koa.ipac.caltech.edu/cgi-bin/KOA/nph-KOAlogin

[3] http://archive.eso.org/wdb/wdb/eso/xshooter/form

re-binned the HIRES spectra by a factor 5 to match the XShooter resolution to facilitate the search for C IV absorbers redwards of Lyman-α.

To measure transmitted Lyman-α fluxes bluewards of the Lyman-α emission line, we fit each spectrum with a power-law continuum. This power-law is fitted over the relatively featureless wavelength interval 1270 Å–1450 Å in the rest frame. We exclude pixels affected by sky lines and use two rounds of sigma-clipping with thresholds of $|F_{\rm PL} - F_{\rm obs}| < 2\epsilon$ and 1.5ϵ, where ϵ is the observational error and $F_{\rm PL}, F_{\rm obs}$ are the values of the power-law fit and the quasar flux, respectively.

The Lyman-α forest of $z \gtrsim 5$ is characterised by high absorption at all wavelengths, making the continuum notoriously difficult to determine or model. Due to the sparsity of transmitted flux, we make no attempt at modelling the continuum bluewards of the Lyman-α emission line beyond a power-law. At lower redshift, more advanced techniques have been used and include subtracting profiles of weak intrinsic absorption and emission lines (e.g. Crighton et al. 2011). When comparing to such studies, a potential worry is a bias in the Lyman-α forest self-correlation on the scales of such features that are not removed by a simple power-law. We however verified that the self-correlation was not deviated from unity by more than 1σ on any scale (see Bosman et al. 2018).

3.2.2 Quasar Broad Emission Lines and continuum fitting

Previous systematic searches for C IV or metal absorptions in the continuum of $z \simeq 6$ quasar (e.g. Ryan-Weber et al. 2009; Becker et al. 2011; Simcoe et al. 2011; D'Odorico et al. 2013; Bosman et al. 2017; Codoreanu et al. 2018) proceed by removing first the power-law continuum of the quasar and secondly fitting the Broad Emission Lines (BEL). The BEL fit is usually performed with splines in an iterative "by-eye" process where the user supervises the selection of knot points. Since our sample contains 25 quasar spectra, we developed an automated quasar continuum spline fitting (redwards of Lyman-α only) further described below. The first step is to determine which part of the spectra are devoid of narrow emission lines or Broad Absorp-

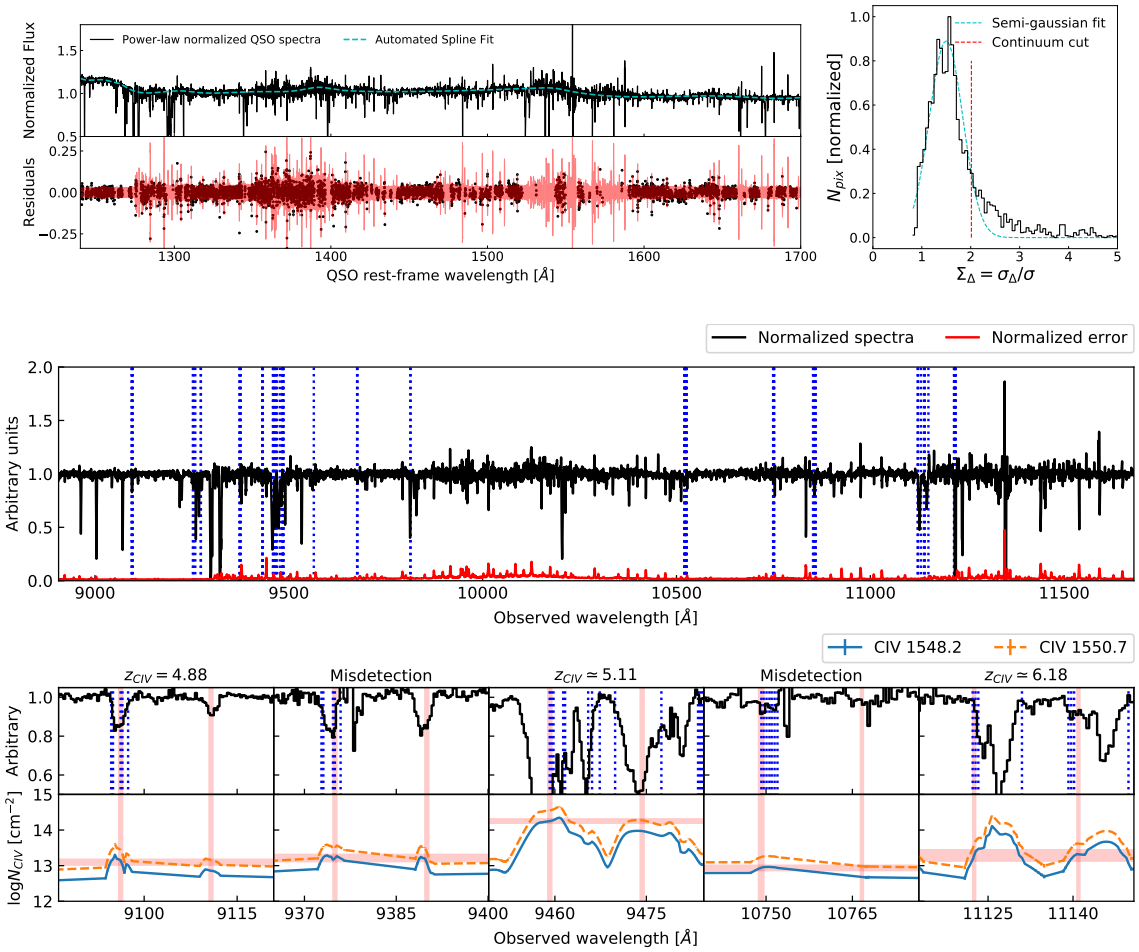

Figure 3.1: Illustration of our automated search procedure for C IV absorbers as demonstrated for the quasar J0100+2802. **Upper left panel:** Measured (power-law normalised) flux in black, fitted continuum in cyan. The residuals present very little to no slow-varying residual trend and are consistent within the 2σ errors (red). **Upper right panel:** Histogram (black) of the ΔF variance σ_Δ divided by the "error array" variance σ, and the associated threshold (red line) for non-continuum pixels, taken to be a 1.5σ cut away from the mean of a Gaussian (cyan) fitted to the values lower than the expected mean of $\sqrt{2}$. **Middle panel:** The intrinsic continuum normalised flux (black) and noise (red) are overlaid with the algorithm-identified C IV absorbers (indicated by blue dotted lines) (see Section 3.2.3 for the search strategy). **Lower panel:** Zoom-in on some flagged C IV candidates from the middle panel. The algorithm computes the column density for both C IV 1548 Å, 1550 Å at every pixel, and flags every pair with matching column density at the correct separation, with the tolerance indicated in red. False detections and misalignments due to sigma-clipping can be easily removed by eye or when fitting with VPFIT.

tion Lines. A "fit-by-eye" procedure would select such regions as representative of the quasar continuum. In fact, human users select regions where *the pixel-to-pixel flux variation is consistent with the error array.* Mathematically, we expect that for a slowly varying spectrum with high enough resolution, the pixel-to-pixel flux difference is distributed as

$$\Delta F_i = F_i - F_{i+1} \sim \mathcal{N}(0, 2\sigma_i^2) \ , \tag{3.2}$$

where F_i is the flux recorded at pixel i, σ_i the corresponding error, and $\mathcal{N}(\mu, \sigma^2)$ the normal distribution with mean μ and variance σ^2. We thus estimate the variance of the flux difference σ_Δ by simply computing a running variance on the flux variation ΔF_i. The running variance is taken as the square of the standard deviation of ΔF in a 40 pixel wide window centered on pixel i. We then take the ratio between the σ_Δ and the error array σ at each pixel i. The distribution of the resulting variable, $\Sigma_\Delta = \sigma_\Delta/\sigma$ is a Gaussian distribution of mean $\sqrt{2}$ with a tail of larger values, as expected (see upper right corner of Figure 3.1). We fit a Gaussian to the low-value wing of the distribution of Σ_Δ and exclude all pixels at $> N\sigma$ from the mean of the fitted Gaussian from the "continuum pixels". N is a parameter chosen by the user, and $1 \leq N \leq 3$ is used for all our spectra. We note however that in large, completely absorbed features, the pixel-to-pixel variation differs from the noise distribution array *only in the wings of the absorption.* To remove pixels at the bottom of these absorption troughs, we run a matched-filter with a Gaussian kernel with a width of 5 Å. Any pixel with $S/N_{\text{matched-filter}}$ greater than a chosen threshold (here > 3) should be rejected from the continuum.

The second step of the algorithm is based on the idea that the BEL that provide the complexity of the quasar continuum fitting are not a nuisance but rather a powerful indicator of where the splines knot points should be located. Based on a BEL rest-frame wavelength list given by the user, we assign a knot point on top of each BEL with some tolerance (provided by the user as well, but $\sim$ 10 Å is a reasonable choice), and another intermediate knot point in between each BEL

knots. We minimise the χ^2 of the spline fit on the previously selected continuum pixels by moving the knot points in the assigned tolerance regions to yield our final fit. An example of a resulting fit is shown in Figure 3.1. The automated continuum method successfully fits the BEL features as well as avoiding the regions contaminated by skylines. The QUICFIT (QUasar Intrinsic Continuum FITter) code is publicly available at https://github.com/rameyer/QUICFit.

3.2.3 C IV identification

It is possible to search for many metal absorber species redwards of the quasar Lyman-α emission at high-redshift, including amongst other O I, C II, Si II, N V, Si IV, Al II, Al III. However C IV is the most useful due to its ubiquity and the fact that it is reliably identifiable as a doublet. Once the BEL and power-law continuum of the quasar are subtracted, the processed spectrum is searched for C IV doublets. We use a semi-automated identification algorithm.

Following Bosman et al. (2017), we fit an inverted Gaussian profile to the optical depth $\tau = -\log F$ every $\Delta v \sim 10$ km s^{-1} interval, and apply 3 iterations of 2-σ clipping to fit the most prominent feature only. We then estimate the column density for each C IV transition following the apparent optical depth method (Savage & Sembach 1991)

$$N_{\rm CIV} = \frac{m_e c}{\pi e^2 f \lambda} \int \ln G(v) \mathrm{d}v \ , \tag{3.3}$$

where $G(v)$ is the fitted inverted Gaussian profile, $f = 0.095$ or 0.19 is the oscillator strength for the $\lambda = 1548.2, 1550.8$ Å transitions, respectively.

We select all pairs of Gaussian absorption profiles with $\log N_{\rm CIV} > 12.5$ and with a discrepancy in redshift $\Delta z < R/c$ (where R is the resolution in km s^{-1} and c the speed of light) and a discrepancy in column density $\Delta \log N_{\rm CIV}$ within 2σ of the Gaussian fitting errors. We demonstrate in Figure 3.1 the full fitting and search procedure on the sightline towards J0100+2802. The discrete nature of the search in wavelength space and the σ-clipping operations sometimes produces false detections

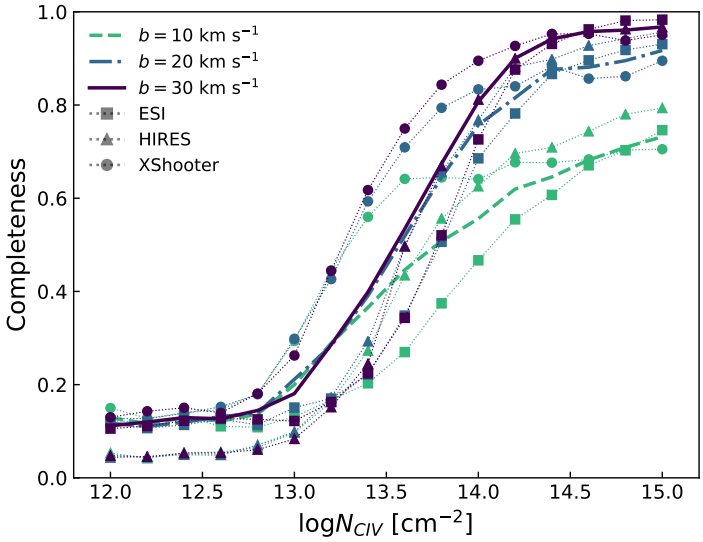

Figure 3.2: Completeness of our C IV semi-automated retrieval computed by the insertion of 1000 C IV doublets at random redshifts in the ESI (squares), HIRES (triangles) and XShooter (dots) for different Doppler parameters $b = 10, 20, 30$ km s^{-1} (green, blue, purple). The thin lines represent the average values per instrument, whereas the bold lines represent the weighted average values for our sample.

Table 3.3: Average completeness for the C IV search. The completeness is a weighted average of the completeness values of Figure 3.2 with $b = 20$km/s for our specific set of quasar sightlines (HIRES:2, ESI:12, XShooter:13).

$\log N_{\rm CIV}/[{\rm cm}^{-2}]$	13.0	13.2	13.4	13.6	13.8	14
Completeness	0.18	0.27	0.36	0.49	0.60	0.70
$\log N_{\rm CIV}/[{\rm cm}^{-2}]$	14.2	14.4	14.6	14.8	15.0	
Completeness	0.78	0.82	0.84	0.85	0.87	

or unreliable estimates of the column density of our absorbers. In order to account for these issues, these flagged candidates are then fitted with VPFIT (Carswell & Webb 2014) to confirm their nature and derive precise column densities.

To estimate the completeness of our search, we insert 1000 mock C IV absorbers for each of $b = 10, 20, 30$ km s^{-1} Doppler parameters and 0.2 dex increments in column density from $N_{\rm CIV} = 10^{12}$ to 10^{14} cm^{-2}. We achieve a 90% completeness level around $\log N_{\rm CIV} \simeq 14.0$ for all 3 instruments, assuming a Doppler parameter of $b = 30$ km s^{-1} (see Figure 3.2). This completeness is in good agreement with previous C IV searches cited beforehand given the resolution and SNR of the quasar spectra at hand.

An important issue is the possibility of false positives which would weaken the

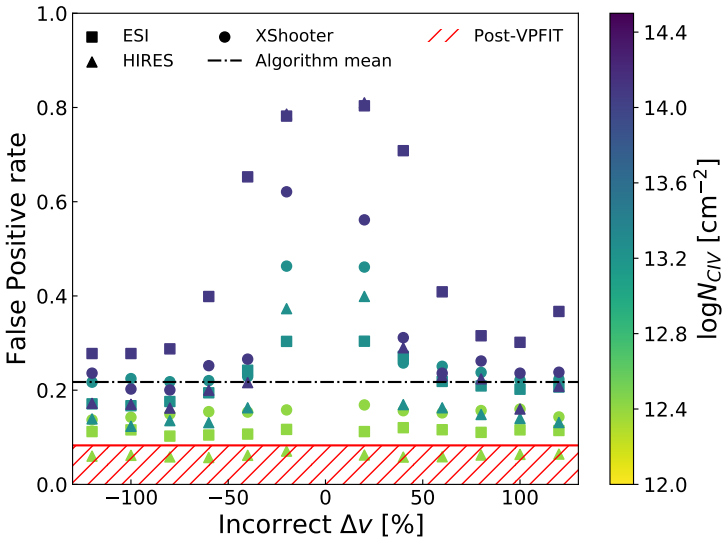

Figure 3.3: False positive rate of the algorithmic detection rate of C IV candidates for various velocity displacements, instruments and column densities (coloured dots according to the right hand scale bar). The higher densities are more contaminated due to the tolerance on the optical fitting and the broad absorption wings of theses systems. The weighted mean false positive rate for our particular sample is indicated in dotted black, whereas the allowed values of the final false positive rate in our sample after visual inspection is shown in hatched red.

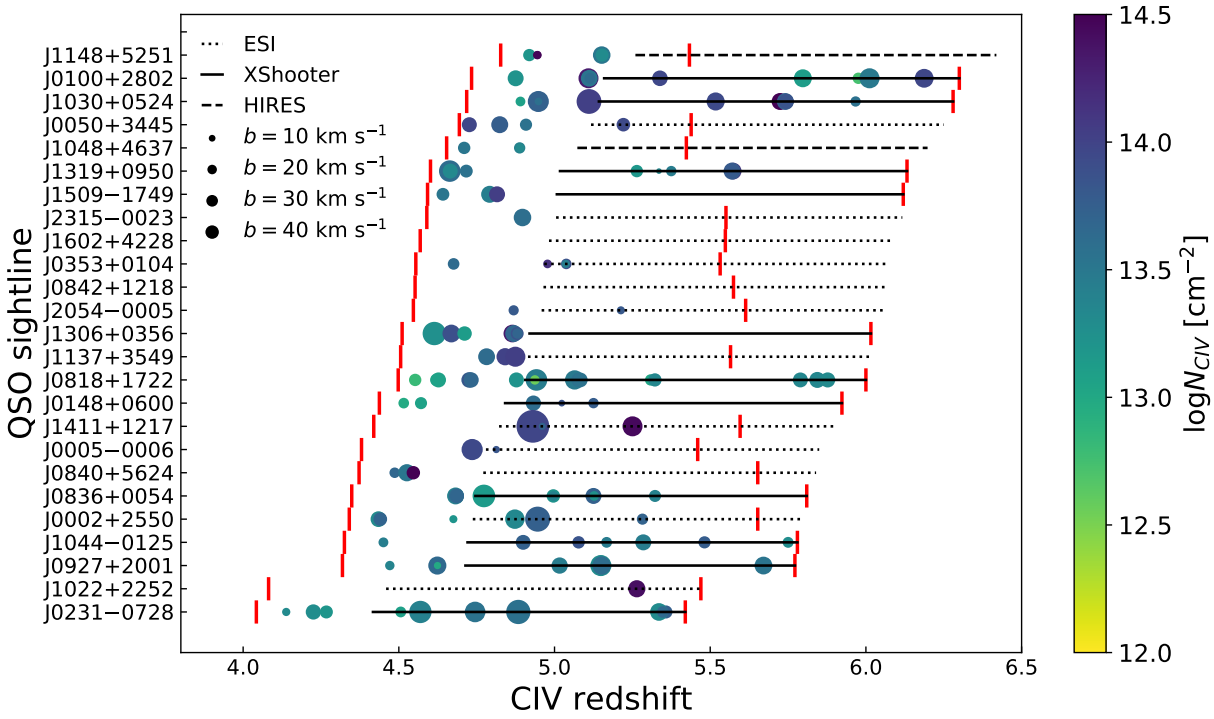

Figure 3.4: Redshift distribution of C IV absorbers for 25 quasar sightlines. The black lines show the full range of the Lyman-α forest for each quasar and the style of the line indicates the instrument. The red bars indicate the redshift range of our CIV search. The absorbers are shown as coloured dots whose size and colour coding reflect their Doppler b parameter and column density, respectively.

sought-after correlation. To assess the fraction of false positives in the candidates flagged by our algorithm, we first run the algorithm to search for doublet emission lines instead of absorption lines. This procedure should record no detections, but we sometimes recorded one or two detections per sightline due to glitches in the spectra or large residuals from continuum correction. Both kinds of false positives were rejected when fitting with a Voigt profile with VPFIT as they do not show the characteristic shape of genuine C IV absorbers. We also insert mock absorbers with incorrect velocity spacing to assess the sensitivity of our algorithm to spuriously aligned absorption lines. We show in Figure 3.3 the false positive rate for different velocity offsets of the two C IV transitions from 25% to 150% relative error on the correct velocity spacing Δv of the doublet. We find that the mean false positive rate is $\simeq 22\%$ for the algorithm search. We compute the mean false positive rate by weighting the false positive rates of each instrument by the number of quasar observed. Given that the candidates are inspected by eye while being fitted with VPFIT, we expect the final false positive rate to be lower because some false positives are discarded. However, assessing the efficiency of a visual inspection is difficult. We compare our search with DO13 on 6 matching sightlines where we found 56 C IV absorbers detections (see Section 3.2.4). We find that we were more cautious than DO13 and rejected some of their absorbers, the fact that we found only two additional detections also argues against our technique generating too many false positives. We can approximate the final (after visual inspection) false positive rate in two manners. Firstly, assuming that all absorbers found in our search and reported in DO13 are real, if the remainder of the 150 C IV sample is pure at the algorithm level ($< 22\%$), then the false positive rate of our final sample is expected to be $\lesssim 8\%$. We can reach an additional estimate by assuming that, in the worst case scenario, the two additional detections we report in the sightlines searched by DO13 are false positives. The purity of this large subsample (and by extrapolation of the total sample) is then 3.6%, in agreement with the upper limit derived previously.

Finally, our search results in 150 C IV absorber detections at $4.5 < z < 6.3$. Figure 3.4 shows a graphical representation of the C IV absorbers found. Appendix

A.1 presents a complete list of our detected C IV absorbers.

3.2.4 Comparison with previous C IV searches

We briefly compare our results for those quasar sightlines already searched for C IV by previous authors to assess the purity of our method. All other detections on other sightlines are new detections and are listed in Table A.1 as well as velocity plots in Appendix A.2 for the reduced sample lying in the Lyman-α forest. Based on this comparison, we estimate our search to be in agreement, if somewhat more conservative, than previous C IV searches. We note that small differences in the column densities (up to $\sim 0.2 - 0.3$) are easily explained by the continuum fitting differences. We also sometimes fit fewer components than previous searches. The excellent agreement in the total cosmic density of C IV between previous studies and our measurement shows these are minor issues driven by noise and different spectra.

J0818+1722 : We retrieve all C IV absorbers found previously by DO13 between $4.5 < z < 5.3$, with $\Delta z < 0.01$. We note however that we fitted one component less to the C IV system at $z \simeq 4.726$. This however has no impact on our analysis as we cluster systems with $\Delta z < 0.01$.

J0836+0054 : We detect the same systems at $z = 4.68, 4.99$,$z = 5.12$ and $z = 5.32$ as DO13. We fit two components less for the $z = 4.99$ to keep only the clear detections.

J0840+5624 : We report 3 new systems at $z = 4.49, 4.53, 4.55$, which were not in the redshift range searched by Ryan-Weber et al. (2009).

J1030+0524 / J1319+0950 : We detect the same systems between $4.65 < z < 5.6$ as DO13, to which we add new detections at $z = 5.11$ in the sightline of J1030+0524 and at $z = 5.34$ in the sightline of J1319+0950. We believe the detections were made possible by the quality of our XShooter spectra.

J1306+0356 : We detect the same systems between $4.5 < z < 5$, with the exception of $z = 4.723$ and $z \simeq 4.885$, which were both blended with sky lines in DO13 and that were not retained here.

J1509-1749 : We recover the same systems between $4.6 < z < 6$, with the exception of $z = 5.7690$ that we consider to be blended with a sky line in DO13's analysis. We also find that the $z = 5.915$ absorber is probably due a spurious alignment of lines.

3.3 Results

3.3.1 Cosmic mass density of C IV

The first physical result that can be readily derived from any sample of metal absorbers line is the comoving cosmic mass density as a function of redshift. This measurement provides valuable insight into the history of the metal enrichment of the Universe. Our large sample of C IV absorbers is used to place new constraints on the cosmic density of C IV at $5 < z < 6$. The comoving mass density of C IV is computed as

$$\Omega_{\rm CIV} = \frac{H_0 m_{\rm CIV}}{c\rho_{\rm crit}} \int N_{\rm CIV} f(N_{\rm CIV}) {\rm d}N_{\rm CIV} \approx \frac{H_0 m_{\rm CIV}}{c\rho_{\rm crit}} \frac{\sum N_{\rm CIV}}{\Delta X}, \tag{3.4}$$

where $f(N_{\rm CIV})$ is the C IV column density function, $m_{\rm CIV}$ is the mass of C IV ion and $\rho_{\rm crit} = 1.88 \times 10^{-29} h^2$ g cm^{-3} is the critical density of the Universe, ΔX is the total absorption path length searched by our survey. The summation runs over all C IV absorbers in the range of column densities of interest. The error is estimated as the fractional variance (Storrie-Lombardi et al. 1996)

$$\left(\frac{\delta\Omega_{\rm CIV}}{\Omega_{\rm CIV}}\right)^2 = \frac{\sum (N_{\rm CIV})^2}{(\sum N_{\rm CIV})^2} \ . \tag{3.5}$$

We present our measurement on Figure 3.5 alongside previous results in the literature. We chose a C IV absorber sample with selection criteria $13.8 \leq \log N_{\rm CIV} < 15.0$ at the two redshift intervals, $4.3 < z < 5.3$ and $5.3 < z < 6.2$, to facilitate a fair

Table 3.4: C IV number line densities and corresponding cosmic densities for the two redshift-selected samples. Columns refer to: (1) redshift interval of sample (2) C IV column density (3) Comoving path length (4) incidence rate per path length (5) cosmic mass density with completeness correction (see Figure 3.2 and Table 3.3).

Redshift	$\log N_{\rm CIV}$	ΔX	$d\mathcal{N}/dX$[a]	$\Omega_{\rm CIV}[\times 10^{-8}]$[a]
$4.3-5.3$	$13.8-15.0$	76.3	0.60 (0.39)	1.06 ± 0.18 (0.75 ± 0.16)
$5.3-6.2$	$13.8-15.0$	21.6	0.34 (0.28)	0.72 ± 0.32 (0.62 ± 0.31)

[a] The bracketed values are without completeness correction.

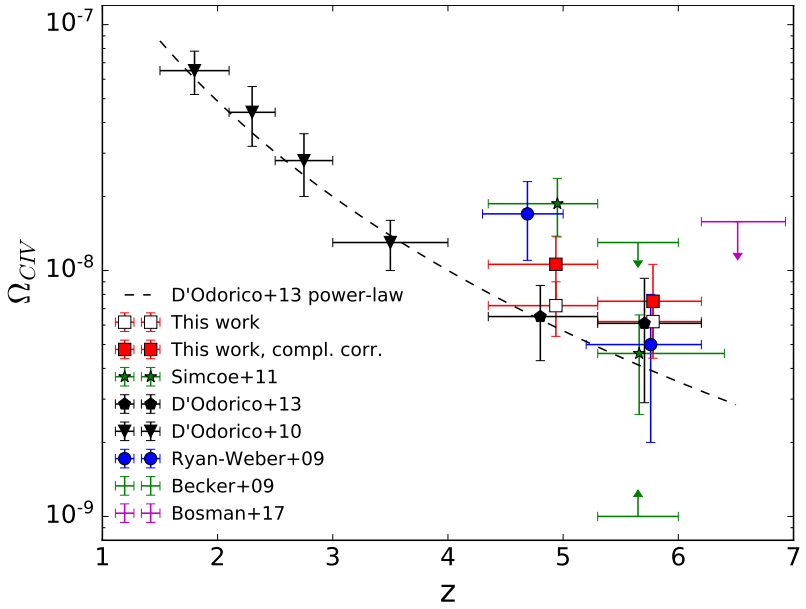

Figure 3.5: Evolution of the cosmic density of C IV at $1.5 \lesssim z \lesssim 6.2$ (D'Odorico et al. 2010, 2013 black triangles,black pentagons) and the corresponding power-law fit $\Omega_{\rm CIV} = (2\pm1)\times10^{-8}[(1+z)/4]^{-3.1\pm0.1}$ (black dotted line). Our measurements (filled and empty red squares, with and without completeness correction) are in good agreement with D'Odorico et al. (2013) as well as other previous studies: Ryan-Weber et al. (2009 blue dots), Becker et al. (2009 green 95% confidence interval), Simcoe et al. (2011 dark green stars) and Bosman et al. (2017 magenta upper limit). The density error bars indicate 1σ confidence intervals.

comparison with DO13's extensive dataset across all redshifts. We note the agreement with their evolution which indicates a general decline in cosmic density with redshift. We note that although different observations and recovery pipelines differ on the exact list of absorbers between studies, the cosmic density measurements are very similar. We list the values with and without complete correction (Table 3.3) for different redshift intervals in Table 3.4. The completeness correction does not significantly change the overall decline of the cosmic density in our redshift interval. Finally, we note that the results presented here could be slightly overestimated by a factor up to $\lesssim 4-8\%$, which is the maximum density-weighted fraction of false positives estimated in Section 3.2.3. Accounting for this effect would decrease (although very slightly) our resulting densities, pushing them even closer to that measured in previous studies.

This decline of $\Omega_{\rm CIV}$ reflects both (*i*) the build up of total carbon budget at decreasing redshifts, as more metal is ejected into the circum-/inter-galactic medium around star-forming galaxies by outflows, and (*ii*) the changing ionisation state of carbon due to the evolving spectral shape of the UV background (see e.g. Becker et al. 2015a for a review and references therein). We will discuss the chemical enrichment and other properties of C IV-host in Section 3.5 after presenting their 1D correlation with the IGM transmission.

3.3.2 The observed 1D correlation of C IV with the IGM transmission

The main aim of this study is to compute the 1D correlation between our C IV absorbers and the IGM transmission. We compute the correlation using Davis & Peebles (1983) estimator

$$\hat{\xi}_{\rm CIV-Ly\alpha}(r) = \frac{D_{\rm CIV} D_{\rm T}}{R_{\rm CIV} D_{\rm T}} - 1 = \frac{\langle T(r) \rangle_{\rm CIV}}{\overline{T}} - 1 \ , \tag{3.6}$$

where the $D_{\rm CIV}$ is the pixel corresponding to the H I Lyman-α absorption at the redshift of a detected C IV absorber, and $D_T = \langle T(r) \rangle_{\Delta r}$, is the transmission in

the forest of the quasars at a distance r from the C IV absorber. To make the comparison with previous studies easier, we provide a equivalent formulation where $\langle T(r)\rangle_{\rm CIV}$ is the average Lyman-α transmission T at a distance r of the redshift C IV absorbers, and $\overline{T}$ is the average transmission at the redshift z in the quasar line-of-sight (LOS) studied. Operationally, for each $D_{\rm CIV}$, we take the transmission in neighbouring pixels D_T and compute the LOS Hubble distance. Similarly, the average transmission is estimated from a random distribution of absorbers with redshifts $R_{\rm CIV}$. The random redshifts $R_{\rm CIV}$ are generated by oversampling 50 times the redshift interval $z_{\rm CIV} \pm 0.1$ around each observed C IV absorber detected at $z_{\rm CIV}$ in each LOS, so as to reproduce the observed redshift distribution and sightline-to-sightline transmission variance. We note that the conversion from velocity space to Hubble distance is subject to a caveat due to peculiar velocities on small scales further discussed in Section 3.5.3. We weight the pixels with the inverse variance as we perform the mean to bin the correlation function linearly or logarithmically depending on the analysis, and we bin the observed and random absorbers in a consistent manner. In order to have sensible values of the transmission and correct for any misfits of the power-law continuum (see Section 3.2.1), we remove pixel artifacts by excluding those with $T < -2e$ and $T - e > 1$, where T is the transmission and e the corresponding measurement error. We also only use pixels between 1045 and 1176 Å in the quasar rest-frame to avoid the quasar Lyman-β and -α intrinsic emission. To estimate the error we choose a Jackknife test given our modest sample of sightlines. We draw 500 subsets of half of the C IV sample, generate accordingly the random samples and compute the correlation for these subsets. The variance of the 500 draws is then used as an estimate of our errors. We note that this method is more conservative than Poisson or bootstrap errors, and converges with an increasing sample size.

Not all C IV absorbers detected in the sightline of the quasars are suitable for this measurement. We define here the sample of C IV absorbers redshifts used for the correlation with the IGM transmission, named Sample α. First and foremost, the corresponding H I Lyman-α should be between 1045 and 1176 Å in the quasar rest-

frame to avoid the quasar Lyman-β and -α intrinsic emission. Secondly, as we are using C IV as a tracer of galaxies we take only the redshift of the strongest absorber for systems with multiple components when they are $\delta v < 100$ km s^{-1} apart. This avoids systems with multiple components multiply sampling the same part of the quasar forest and thus biasing the measurement. Finally, as our estimator requires a proper measurement of the transmission in the Lyman-α, it cannot produce a sensible measurement where the average flux in the Lyman-α forest falls below the sensitivity of the spectrograph , producing lower limits on the correlation that are not easily interpreted. We hence remove C IV absorbers lying in saturated end regions of the Lyman$-\alpha$ forest where the average transmission over $\Delta z = 0.1$ is less than the average 1σ error level of the flux measurement. We emphasize that this last step only removes the end part of two quasar forests in which C IV absorbers sit, and exclude only $\sim 10\%$ of Sample α.

The selection described above leaves 37 C IV systems absorbers suitable for the correlation measurement to which we add the $z = 5.738$ absorber on the LOS of J1148+3549 from Ryan-Weber et al. (2009), which was detected in a NIRSPEC spectrum at wavelengths not covered in our ESI spectrum. The average redshift of Sample α is $\langle z_{\rm CIV} \rangle = 5.18$ and the average column density $\log \langle N_{\rm CIV} \rangle \simeq 13.8$. We show the 6 lower-redshift detections of Sample α on Figure 3.6. The whole sample is presented in Appendix A.2.

The 1D correlation signal (shown on Figure 3.7) reveals an excess absorption at $r_c \lesssim 5$ cMpc/h around C IV absorbers at $4.5 < z < 6$. This excess absorption is also found around $z \sim 3$ Lyman Break Galaxies (LBGs) (Adelberger et al. 2003, 2005; Bielby et al. 2017) (see Figure 3.7 for a comparison). The excess absorption is detected on the same scales (~ 5 cMpc/h) for both $z \sim 3$ LBGs and $z \sim 5.2$ C IVs, but the excess absorption seems much stronger in the latter objects. The absorption excess is perhaps emerging more clearly due the overall opaqueness Lyman-α forest at $z \sim 5.2$. The trough also shows the expected co-spatiality of C IV and Lyman-α absorption. There is also an excess of transmission at $10 \lesssim r \lesssim 30$ cMpc/h around C IV . This excess on large scales was detected around spectroscopically confirmed

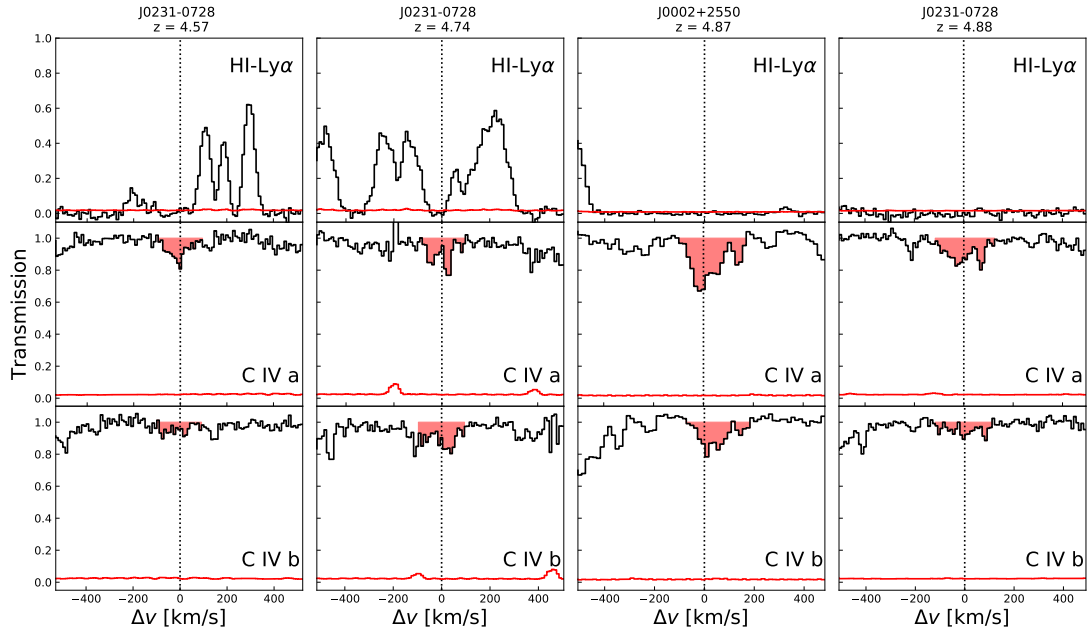

Figure 3.6: First lower-redshift detections of our C IV sample detected with a corresponding H I Lyman-α absorption in the quasar Lyman-α forest. The entire sample of 37 such absorbers is presented in Figure A.1.

LBGs in Kakiichi et al. (2018). The significance of the excess on $10 - 30$ cMpc/h scales is 2.7σ for the average transmission $T \simeq 0.14$ versus a null mean flux. We note that the signal goes to zero at large distances, indicating that the excess is unlikely to be caused by the wrong normalisation of the 1D C IV-IGM correlation. We discuss the physical implications in Section 3.5.

3.4 Modelling the C IV-IGM correlation

In order to interpret the data, we use the linearised version of the model introduced and discussed in detail in Kakiichi et al. (2018). The precepts for the LBG-IGM cross-correlation discussed in Kakiichi et al. (2018) can be easily modified for *any* tracer of galaxies, and the dense, ionised C IV gas is an ideal candidate.

Supposing that galaxies hosted by dark matter haloes eject their material by galactic winds, chemically enriching the surrounding IGM environment. C IV absorbers act as a tracer of galaxies. At $z \gtrsim 5$, patchy reionisation can produce the fluctuations in the UV background affecting the Lyα forest transmission around galaxies. Therefore, the C IV-IGM correlation will reflect both (i) the correlation

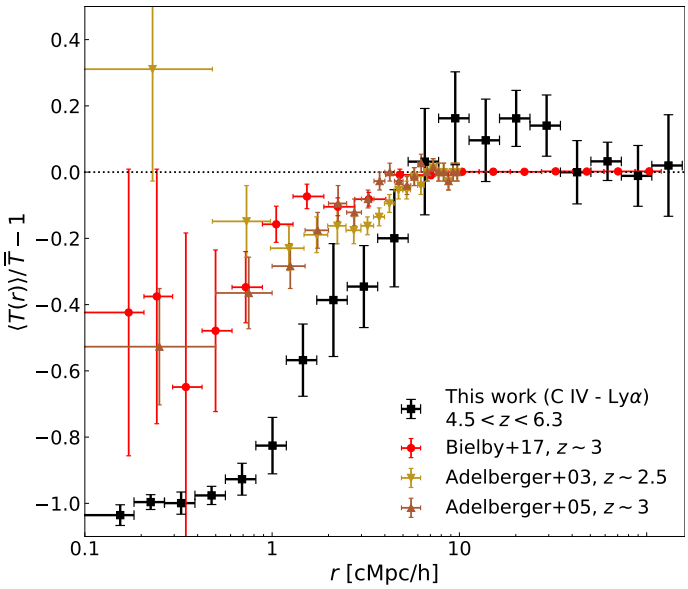

Figure 3.7: The measured 1D correlation between C IV absorbers and the IGM transmission (black) is compared to previous studies on LBG-IGM correlation: Adelberger et al. (2003, 2005 yellow, orange crosses), studies at $z \sim 2.5$ and $z \sim 3$ and the VLT LBG survey at $z \sim 3$ (Bielby et al. 2017 red crosses). The central trough is enhanced at high redshift and the excess on the large scales is in contrast to the flat profile seen in low-redshift data.

between matter and galaxies and (*ii*) the enhanced UV background around the C IV-host galaxies. Following Appendix B of Kakiichi et al. (2018) the expected flux transmission at a comoving distance r from the position of a C IV absorber is

$$T(r) = \exp(-\tau_\alpha(r)) = \overline{T}(1 + \xi_{\mathrm{CIV-Ly}\alpha}(r)) \ , \tag{3.7}$$

where $\overline{T}$ is the average transmitted flux in the Lyman-α forest at the redshift of interest, and the C IV-IGM correlation along the LOS is then given by

$$\xi_{\mathrm{CIV-Ly}\alpha}(r) \approx b_{\mathrm{CIV}} b_\alpha \xi_{\mathrm{m}}^{\mathrm{lin}}(r, \mu = 1) + b_{\mathrm{CIV}} b_\Gamma \xi_\Gamma^{\mathrm{lin}}(r) \ . \tag{3.8}$$

The contribution of the matter correlation around galaxies is quantified with the two bias factors of C IV-host galaxies b_{CIV} and the Lyman-α forest b_α. The redshift-space sightline linear matter correlation function $\xi_m^{\mathrm{lin}}(r, \mu = 1)$ follows from the real-space

matter correlation function (Hamilton 1992),

$$\xi_m^{\rm lin}(r,\mu=1) = \xi_0(r)P_0(\mu) + \xi_2(r)P_2(\mu) + \xi_4(r)P_4(\mu) \tag{3.9}$$

with $\xi_0(r) = [1 + (\beta_{\rm CIV} + \beta_\alpha)/3 + \beta_{\rm CIV}\beta_\alpha/5]\xi(r)$, $\xi_2(r) = [(2/3)(\beta_{\rm CIV} + \beta_\alpha) + (4/7)\beta_{\rm CIV}\beta_\alpha][\xi(r) - \bar{\xi}(r)]$, and $\xi_4(r) = (8/35)\beta_{\rm CIV}\beta_\alpha[\xi(r) + (5/2)\bar{\xi}(r) - (7/2)\bar{\bar{\xi}}(r)]$ where $\xi(r)$ is the linear matter correlation function in real space, and $\beta_{\rm CIV}$ and β_α are the redshift space distortion (RSD) parameters (Kaiser 1987). The RSD parameter of C IV is set to $\beta_{\rm CIV} \approx \Omega_m^{0.6}(z)/b_{\rm CIV} \approx 1/b_{\rm CIV}$. The Lyman-$\alpha$ forest RSD parameter β_α is found to be relatively constant at lower redshift from observations in the range $\beta \simeq 1.2 - 1.7$ (Slosar et al. 2011; Blomqvist et al. 2015; Bautista et al. 2017). We set $\beta_\alpha = 1.5$ as a fiducial value. The Lyα forest bias is chosen such that the model 1D Lyα forest power spectrum is consistent with observations from Viel et al. (2013), which leads to the fiducial value of $b_\alpha = -1.3$.

On large scales, the contribution of the enhanced UV background becomes increasingly important. The mean photoionisation rate $\bar{\Gamma}$ of the IGM from star-forming galaxies depends on the the population average of the product of LyC escape fraction and and LyC photon production efficiency $\langle f_{\rm esc}\xi_{\rm ion}\rangle$,

$$\bar{\Gamma} \propto \langle f_{\rm esc}\xi_{\rm ion}\rangle \lambda_{\rm mfp} \int_{-\infty}^{M_{\rm UV}^{\rm lim}} L_{\rm UV}(M_{\rm UV}) \frac{dn}{dM_{\rm UV}} dM_{\rm UV} \ , \tag{3.10}$$

where $dn/dM_{\rm UV}$ is the UV luminosity function (Bouwens et al. 2015) and $M_{\rm UV}^{\rm lim}$ is the limiting UV magnitude of galaxies that contribute to the UV background. We adopt the value of mean free path $\lambda_{\rm mfp}$ of Worseck et al. (2014). The effect of the UV background is then modelled through the bias factor b_Γ defined by

$$b_\Gamma = \frac{\int \Delta_b P_V(\Delta_b)\tau_\alpha(\bar{\Gamma},\Delta_b)e^{-\tau_\alpha(\bar{\Gamma},\Delta_b)}}{\int \Delta_b P_V(\Delta_b)e^{-\tau_\alpha(\bar{\Gamma},\Delta_b)}} \ , \tag{3.11}$$

where $\tau_\alpha(\bar{\Gamma},\Delta_b) \simeq 5.5\Delta_b^2(\bar{\Gamma}/10^{-12}{\rm s}^{-1})^{-1}(T/10^4{\rm K})^{-0.72}[(1+z)/6]^{9/2}$ is the Lyman-α optical depth at the mean photoionisation rate (we assume a uniform temperature of $T = 10^4$ K), $P_V(\Delta_b)$ is the volume-weighted probability distribution function of

baryon overdensity Δ_b (Pawlik et al. 2009). The correlation function of the UV background with galaxies is

$$\xi_{\Gamma}^{\rm lin}(r) = \langle b_g(< M_{\rm UV}^{\rm lim})\rangle_L \int \frac{k^2 {\rm d}k}{2\pi^2} R(k\lambda_{\rm mfp}) P_m^{\rm lin}(k) \frac{\sin kr}{kr} \ , \tag{3.12}$$

where $R(k\lambda_{\rm mfp}) = \arctan(k\lambda_{\rm mfp})/(k\lambda_{\rm mfp})$, $P_m^{\rm lin}(k)$ is the linear matter power spectrum, and $\langle b_g(< M_{\rm UV})\rangle_L$ is the luminosity-weighted bias of ionising sources above $M_{\rm UV}^{\rm lim}$, which is evaluated with the same procedure as in Kakiichi et al. (2018). We use the halo occupation number framework with the conditional luminosity function which parameters are fixed by simultaneously fitting the Bouwens et al. (2015) $z \sim 5$ luminosity function and the Harikane et al. (2016) LBG correlation functions, resulting in the best-fit parameters $(M_{\rm UV,0}, \log M_h^*, \gamma_1, \gamma_2, \sigma_c, \log \phi_0, \alpha_s, \beta_s)$ =(-22.29, 11.90, 2.39, -0.06, 0.2(fixed), -1.49,-1.26, 0.85) (see Kakiichi et al. (2018) for more details).

Overall, the linear model used to describe the observed C IV-IGM correlation contains five parameters; two to describe the UV background $(\langle f_{\rm esc}\xi_{\rm ion}\rangle, M_{\rm UV}^{\rm lim})$, one to describe the halo bias of C IV-host galaxies $b_{\rm CIV}$, and two to describe the matter fluctuations in the Lyman-α forest (b_α, β_α). Our fiducial linear model leaves the first three parameters free $(\langle f_{\rm esc}\xi_{\rm ion}\rangle, M_{\rm UV}^{\rm lim}, b_{\rm CIV})$, but the full five parameter model including (b_α, β_α) is examined in Appendix A.3.

We fit the linear model to the linearly binned correlation using the Markov chain Monte Carlo affine sampler from the EMCEE package (Foreman-Mackey et al. 2013). In doing so, we assume a flat prior in all three parameters in the following ranges: $23 < \log\langle f_{\rm esc}\xi_{\rm ion}\rangle/[{\rm erg}^{-1}\ {\rm Hz}] < 27$, $-20 < M_{\rm UV}^{\rm lim} < -8$, $0 < b_{\rm CIV} < 10^3$. The priors are quite broad and encompass all plausible physical values. We run the sampler for 5000 steps with 24 walkers, discarding 500 steps for burn-in and fixing the scale parameter to ensure the acceptance rate stays within $0.3 < r < 0.5$. The walkers are initialized in a Gaussian sphere with variance $\sigma = 0.1$ at different locations in the allowed parameter space, without any noticeable change to our results. We exclude the first datapoint at $r = 2.5 \pm 2.5$ cMpc/h from the fit and we cap the linear

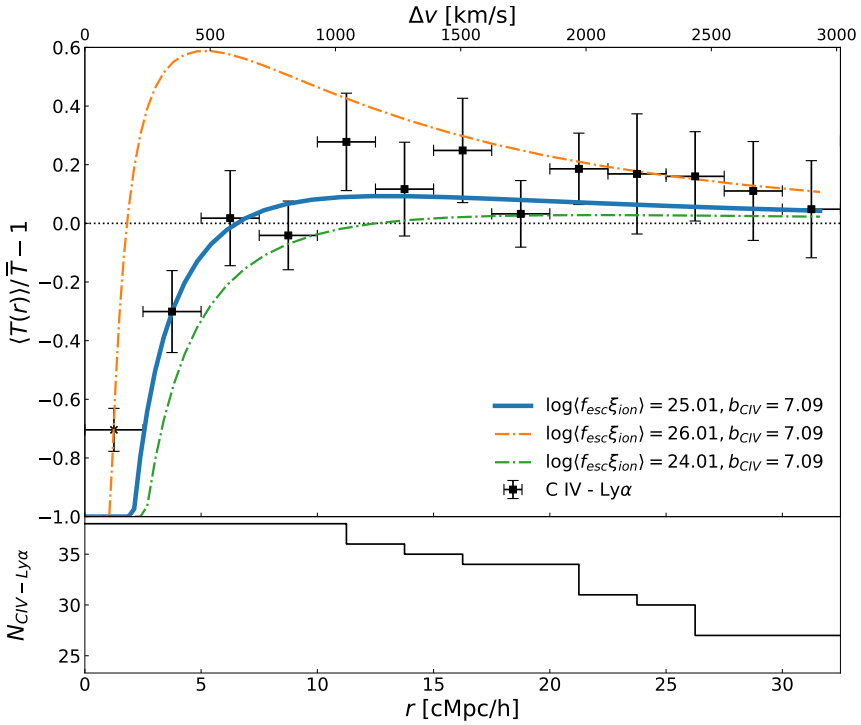

Figure 3.8: *Upper panel:* The measured correlation between C IV and the IGM transmission, linearly binned, compared to our models. We show various linear models of the correlation for illustrative purposes. Blue line (maximum likelihood model, see Figure 3.9): $(\log\langle f_{\rm esc}\xi_{\rm ion}\rangle/[{\rm erg}^{-1}\ {\rm Hz}] = 25.01, M_{\rm UV}^{\rm lim} = -10.82, b_{\rm CIV}, b_\alpha = -1.3, \beta_\alpha = 1.5)$. The blue curve is indeed the best-fit despite being significantly lower than most datapoints at $r > 10$cMpc. This is because the main parameter of our model, the ionising photon production, impacts mostly the knee of the curve at ~ 10cMpc and only marginally the large scales. Two additional models (orange semi-dotted line: $\log\langle f_{\rm esc}\xi_{\rm ion}\rangle/[{\rm erg}^{-1}\ {\rm Hz}] = 26.01$, $\log\langle f_{\rm esc}\xi_{\rm ion}\rangle/[{\rm erg}^{-1}\ {\rm Hz}] = 24.01$) show this effect over the whole dynamic range of $\langle f_{\rm esc}\xi_{\rm ion}\rangle$. The linear model produces unphysical values of the cross-correlation on small scales $\xi_{\rm CIV-Ly\alpha}(\lesssim 3{\rm cMpc/h}) < -1$. In order to limit the impact on the fit, we capped the value of the linear model at -1 and excluded the first point from the fit. *Lower panel:* Number of C IV absorbers in Sample α contributing to the measure at different scales.

model at -1 because it does not hold on very small scales, predicting unphysical correlation values $\xi(r \lesssim 2.5{\rm cMpc/h}) < -1$. The model is evaluated at $\langle z_{\rm CIV}\rangle = 5.18$. The resulting fit and the possible inference on the three parameters is discussed in the next section.

3.5 Physical implications

We discuss the physical implications of our measurements of the C IV-IGM 1D correlation at $z \simeq 5$. The two main features of the correlation are (i) an excess

Lyman-α forest absorption on small scales $r < 5$ cMpc/h suggestive of the gas overdensity around C IV absorbers and indicative evidence of the outskirt of the CGM around the $z > 5$ galaxies, and (*ii*) an excess IGM transmission on large-scale ($r \gtrsim 10$ cMpc/h) which is consistent with an enhanced UV background around C IV powered by galaxy clustering with a large ionising photon budget as predicted in Kakiichi et al. (2018). In Figure 3.8 we show the observed C IV-IGM correlation overlaid with our linear model presented before. The large scale excess transmission of the correlation is reasonably well captured by the model despite its simplicity, confirming the foregoing interpretation. Clearly any interpretation, and subsequent inference on the exact escape fraction and spectral hardness, will be subjected to the uncertainties due to our modest sample size and theoretical model. These are addressed in Section 3.5.3.

We show the posterior probability distribution of the parameters for our likelihood in Figure 3.9. We find that the observed large-scale IGM transmission excess requires a large population-averaged product of LyC escape fraction and spectral hardness parameter, $\log\langle f_{\rm esc}\xi_{\rm ion}\rangle/[{\rm erg}^{-1}\ {\rm Hz}] = 25.01^{+0.30}_{-0.19}$ (for the fiducial model fit) where the quoted error is the 1σ credibility interval. Even using a conservative modelling approach using the full five parameters with flat priors, the observed level of the large-scale excess seems to indicate a large value $\log\langle f_{\rm esc}\xi_{\rm ion}\rangle/[{\rm erg}^{-1}\ {\rm Hz}] \gtrsim 24.7$ (1σ limit) (see Appendix A.3 for full details). The limiting UV magnitude of the ionising sources is unconstrained in the prior range. At face value, the best-fit value of C IV bias $b_{\rm CIV} = 7.09^{+3.29}_{-2.89}$ appears somewhat large corresponding to a host halo mass of $11.3 \lesssim \log M_{\rm h}/{\rm M}_\odot \lesssim 12.6$. However, this value is degenerate with other Lyman-α forest parameters unknown at $z > 5.0$, permitting values as small as $b_{\rm CIV} = 3.5^{+2.0}_{-1.0}$. The corresponding host halo mass of $10.4 \lesssim \log M_{\rm h}/{\rm M}_\odot \lesssim 11.6$ is then consistent with the data. Therefore, the host halo mass of C IV absorbers is loosely constrained to lie between $10.4 \lesssim \log M_{\rm h}/{\rm M}_\odot \lesssim 12.6$.

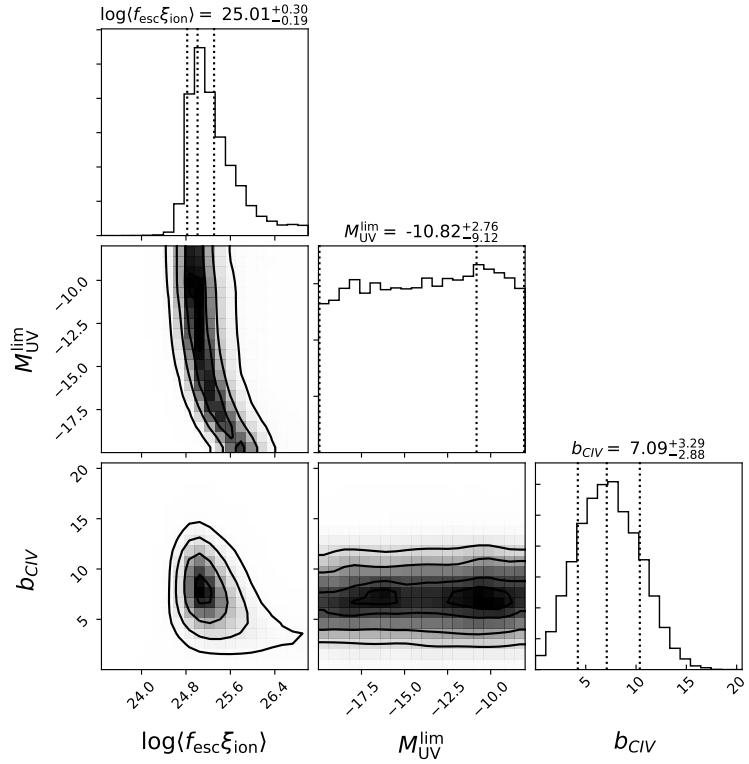

Figure 3.9: Posterior distribution of the parameters in our linear model of the correlation. The quoted numbers of the 1-D histogram give the 1σ credibility interval. Fitting the data with our linear model put constraints on the product $\log\langle f_{\rm esc}\xi_{\rm ion}\rangle/[{\rm erg}^{-1}\ {\rm Hz}] = 25.01^{+0.30}_{-0.19}$ and the bias of the C IV haloes $b_{\rm CIV} = 7.09^{+3.29}_{-2.86}$. The limiting UV magnitude of UV contributors is left unconstrained. It is indeed the parameter that affects the least the correlation.

3.5.1 The properties of C IV hosts, faint galaxies and feedback

We can use abundance matching to compute the halo mass of the C IV hosts. We find that the sightline number density of our sample of C IV absorbers with $\log N_{\rm CIV} > 13.0$ is $d\mathcal{N}/dX = 4.38 \pm 0.42$ (4.42 ± 0.85) for absorbers at $4.3 < z < 5.3\,(5.3 < z < 6.2)$, where the quoted errors are Poisson and we have applied a completeness correction.

We then compute the comoving density of the absorbers. As galactic outflows enrich the gas around galaxies out to a distance $R_{\rm CIV}$ and with a C IV covering fraction $f_{\rm c}$, assuming a one-to-one relation between C IV absorbers and dark matter haloes, the incidence rate is

$$\frac{d\mathcal{N}}{dX} = \frac{c}{H_0}\int_{M_{\rm h}}^{\infty}\langle f_{\rm c}\pi R_{\rm CIV}^2\rangle\frac{dn}{dM'_{\rm h}}dM'_{\rm h}\ , \tag{3.13}$$

where $\langle f_{\rm c}\pi R^2_{\rm CIV}\rangle$ is the population-averaged physical cross section of metal enriched gas, $dn/dM_{\rm h}$ is a halo mass function in comoving units, and the comoving density is given by $\phi \equiv \int_{M_{\rm h}}^{\infty} \frac{dn}{dM'_{\rm h}} dM'_{\rm h}$. A conservative maximal enrichment radius for C IV is $r \lesssim 100$ pkpc, which has been derived from simulations of high-redshift galaxies by Oppenheimer et al. (2009). Keating et al. (2016) have found that the maximal enrichment radius could be twice as small, with most metal-rich outflows traveling less than 50 pkpc from their host galaxy. Assuming a single physical cross-section for the whole sample, we derive likely UV luminosities and masses for C IV hosts with a conservative enrichment radius $25 < R_{\rm CIV} < 100$ pkpc with $f_{\rm c} = 1$. For our fiducial choice of $5.3 < z < 6.2$ absorbers, $d\mathcal{N}/dX = 4.42$, and a physical cross enrichment radius of $R_{\rm CIV} \simeq 100$ pkpc, we find $\phi \simeq 3.2 \times 10^{-2}$ cMpc^{-3}.

This translates to likely halo masses of $M_{\rm h} \gtrsim 10^{10} M_{\odot}$ (Murray et al. 2013). We can also translate the abundance to luminosities of $M_{\rm UV} \lesssim -16$ (e.g. Bouwens et al. 2015). Clearly a smaller enrichment radius by weaker outflow and/or clumpy distribution of metal enriched gas ($f_c < 1$) requires lower mass, more abundant haloes as the hosts of C IV. With improved measurements and analysis, in the future we hope to invert this argument such that, given an independent measure of C IV-host halo mass from the correlation, it will be possible to infer the cross-section of metal enriched gas, i.e. the properties of galactic outflows around near reionisation-era galaxies.

Our study offers an interesting new insight on C IV absorbers as we view them in absorption with respect to the Lyman-α forest at $z \sim 5$. We note that all C IV absorbers in our Sample α always fall into highly opaque troughs with no Lyman-α transmission, which is surprising given that velocity shifts can occur between the C IV and the expected associated Lyman-α line, and thus in principle flux could be detected in some cases (see Figure 3.6 and A.1). When multiple C IV absorbers are detected in the same trough, the separation is at least $0.002 < \Delta z < 0.005$. Pairs of C IV absorbers sharing the same trough could be two distinct C IV-enriched clouds, corresponding to separations of $200 - 400$ ckpc/h. Alternatively this might imply outflows with speeds around $\sim 25 - 100$ km s^{-1}. The absence of Lyman-

α transmission at the redshift of every C IV absorber and the average separation between absorbers within the same opaque region could potentially serve as a test for models which aim at reproducing the distribution and velocity of metals in the early Universe.

3.5.2 Escape fraction, spectral hardness and UV background

Our inference on the mean ionising photon production rate and escape fraction product of the galaxies clustered around C IV and likely contributing to reionisation is

$$\log\langle f_{\rm esc}\xi_{\rm ion}\rangle/[{\rm erg}^{-1}\ {\rm Hz}] = 25.01^{+0.30}_{-0.19}\ .$$

This would imply a $\gtrsim$ 50% escape fraction if adopting the "canonical" values of the LyC production efficiency of $\log\xi_{\rm ion}/[{\rm erg}^{-1}\ {\rm Hz}] = 25.2 - 25.4$ found for $z = 4 - 5$ LBGs (Bouwens et al. 2016). However, recent studies of intermediate redshift Lyman-α Emitters (LAEs) (Nakajima et al. 2016, 2018) and CIII] emitters (Stark et al. 2015, 2017) have found higher LyC production rates in the range at $\log\xi_{\rm ion}/[{\rm erg}^{-1}\ {\rm Hz}] \simeq 25.5 - 25.8$. Our estimated value for the product would still then imply a high mean escape fraction e.g. $\langle f_{\rm esc}\rangle \simeq 0.32$ (0.16) for a fiducial $\log\langle\xi_{\rm ion}\rangle/[{\rm erg}^{-1}\ {\rm Hz}] = 25.5$ (25.8). Conceivably, our sub-luminous sources clustered around C IV absorbers may have an even harder ionising spectrum if these sources are expected to have an average escape fraction of $\lesssim$ 10% as is found at lower redshift.

In Kakiichi et al. (2018) we found an escape fraction $\langle f_{\rm esc}\rangle \gtrsim 10\%$ with a fiducial $\log\langle\xi_{\rm ion}\rangle/[{\rm erg}^{-1}\ {\rm Hz}] = 25.2$ for faint galaxies clustered around LBGs, with a $\log\langle f_{\rm esc}\xi_{\rm ion}\rangle$ product 1 dex lower than our inference from the C IV-IGM 1D correlation. At face value both the escape fraction and the spectral hardness of the galaxies probed by C IV absorbers seems to be increased with respect to the population probed by LBGs in Kakiichi et al. (2018). Meanwhile, Kakiichi et al. (2018) confirmed a LBG in the vicinity of O I absorbers previously detected by (Becker et al. 2006), while none of our detected C IV absorbers had a confirmed bright

counterpart in the LOS towards J1148+5251. These two clues likely suggest that the population traced by LBGs and C IV absorbers is different.

If C IV systems correspond to $M_{\rm UV} \lesssim -16$ galaxies, then by using them as tracers we are likely selecting overdensities less massive and fainter than those traced by LBGs and LAEs. A harder spectrum might then be attributed to a faint clustered population, consistent with the trend of a harder ionising spectrum $\xi_{\rm ion}$ in fainter galaxies recently reported by Nakajima et al. (2016, 2018). We can however not exclude that a high average escape fraction is solely driving our high value of $\log\langle f_{\rm esc}\xi_{\rm ion}\rangle$.

Finally, our results have interesting implications for the observability of galaxies associated with C IV-hosting halos. C IV is a highly ionised ion, indicating the presence of radiation at the ∼4 Ryd level in the immediate vicinity of the hosts. While at low redshift this radiation is provided by the mean UVB, at $z = 5$ the large-scale excess transmission seems to indicate that the collective radiation by large-scale galaxy overdensities around C IV absorbers becomes important. Together with our relatively large halo masses for C IV hosts ($10.4 \lesssim \log M_{\rm h}/{\rm M}_{\odot} < 12.6$), this seems to indicate that C IV absorption should be tracing galaxy overdensities. However, the independent evidence from searches for emission counterparts to metal absorbers at high redshift is sparse and conflicting (Díaz et al. 2011; Cai et al. 2017). To date, no direct emission counterparts of C IV absorbers have been found at $z > 5.0$. In spite of this, Díaz et al. (2014); Diaz et al. (2015) found an overdensity of LAEs within 10 cMpc/h of two quasar sightlines containing C IV absorbers. This is broadly consistent with the picture in which the strongest ionisation takes place in small galaxies, implying the likely hosts of C IV would be fainter LAEs within overdensities of brighter objects.

3.5.3 Alternative interpretations and caveats

We have assumed for our analysis that C IV absorbers are good tracers of galaxies. Due to the C IV wind velocity and the spatial distance between the gas and the host

galaxy, this assumption is however only true on somewhat large scales. The typical outflow speed (e.g. Steidel et al. 2010) is $\sim$ 200 km/s, meaning that at $z \sim 5.5$, the maximum distance a metal can travel over the age of the Universe is about $\sim$ 0.2 pMpc. In simulations where a more careful modelling of the distribution of metals is done, C IV is expected to travel on average $\lesssim$ 100 pkpc away from the progenitor galaxy (Oppenheimer et al. 2009; Bird et al. 2016; Keating et al. 2016). The recent detection of a galaxy at $\sim$ 40 pkpc from a C IV absorber in the sightline of J2310+1855 further strengthen this point (D'Odorico et al. 2018) . Given the expected spatial offsets ($\lesssim$ 0.2 pMpc), and the wind speeds involved ($\sim$ 200 km/s), it is fair to argue that the redshift of C IV is a proxy for the redshift of the host galaxy with an error $<$ 400 km/s. We note that this is comparable with the typical difference between the systemic redshift and the one derived from Lyman-α emission for galaxies in Kakiichi et al. (2018). Thus C IV reasonably traces galaxies on scales $\gtrsim$ 0.6 pMpc at redshift $z \sim 5.5$. This impacts only the two innermost bins of the correlation in Figure 3.8, but the innermost bin is excluded from the fit for reasons exposed above. Hence we conclude C IV is therefore a suitable tracer of galaxies for the purpose of the 1-D correlation with the IGM transmission where a transmission excess is expected to show a positive signal on scales much greater than the redshift-space offset between C IV and its host ($10 - 30$ cMpc).

An alternative interpretation of the transmission excess seen at $r \gtrsim 10$ cMpc/h in Figure 3.8 is shifted Lyman-α flux from associated galaxies. This, however, would imply a mean velocity shift of $\Delta v_{\mathrm{Ly}\alpha} \gtrsim 1000$ km s^{-1}, which is a very high value considering the results of previous studies (e.g. Adelberger et al. 2003; Steidel et al. 2010; Erb et al. 2014; Stark et al. 2017). The addition of the physical offset between C IV gas and the progenitor galaxy could potentially add to this shift, but we have no reason to believe that spatial offsets of C IV and velocity offsets of Lyman-α should conspire to influence significantly the observed flux in the Lyman-α forests of quasars.

Measuring the correlation of any population with the IGM transmission is subject to uncertainties. First of all, the sample is subject to cosmic variance even with a

size of 38 objects. Indeed, some sightlines present up to ~ 5 C IV absorbers with $\log N_{\rm CIV} > 13$ where some are devoid of them in the redshift range searched (see Figure 3.4). This, in conjunction with the fact that the Lyman-α forest at $5 < z < 6$ can show large deviations from the mean opacity (Bosman et al. 2018), yields a noisy correlation even with our sample.

Two other sources of errors are the possible contamination of the Lyman-α forest of the quasar by weak emission and/or metal absorption lines from C IV host galaxies or nearby galaxies. The first should only contribute at most in a few bins of Figure 3.8, given the 200-250 km s^{-1} winds of C IV clouds. As all our observations were carried out with $0.5'' - 1'' \times 11''$ slits (XShooter) and $0.5'' - 1'' \times 20''$ slits (ESI), the C IV hosts most of the time do not fall in the slit if they are believed to be within < 100 pkpc of the C IV cloud. We hence expect no significant contamination from an associated Lyman-α emitting galaxy in the quasar Lyman-α forest. Metal absorption lines (e.g. Si III 1206) can only reduce the signal observed and thus would not affect our claimed excess transmission on large scales. The large redshift interval sampling is likely to smear the signal as there may be a rapid evolution in the population of C IV absorbers at $z \sim 5$. If C IV traces many distinct populations at once, the signal could be indeed mixed across species and redshifts, but the detection of an excess transmission still holds.

Although surprisingly effective for a first interpretation of the C IV-IGM correlation, our model has a number of shortcomings. It is firstly a linear model and thus the small scales may contain large modelling uncertainties due to nonlinear effects. This shortcoming on the small scales is probably best illustrated by the unphysical values derived in the $r \lesssim 2.5$ cMpc/h region. This model also requires a measurement of the bias and RSD parameter of the Lyman-α forest at $z \gtrsim 5$. To illustrate this, we have left b_α, β_α as a free parameter with flat prior in $-3 < b_\alpha < 0, 3 < \beta_\alpha < 0$ to see the effect on the inferred parameters (see Appendix A.3). We notice that b_α is in near perfect degeneracy with $b_{\rm CIV}$ and $\log\langle f_{\rm esc}\xi_{\rm ion}\rangle$. Although the inferred $\log\langle f_{\rm esc}\xi_{\rm ion}\rangle$ using a flat prior is consistent within 1σ of our result presented above, a substantial uncertainty still remains. This linear model would hence benefit from a

reliable measurement of the Lyman-α bias parameters at $z > 5$. Clearly one possible way to circumvent this issue is to directly compare the 1D correlation measurement with hydrodynamical (radiative transfer) simulations calibrated against Lyman-α forest observables at the same redshift. We are planning to investigate such approach in future work, but we here limit ourselves to the linear model for the sake of brevity.

3.6 Conclusion and future work

The 1D correlation of metals with the IGM transmission offers a promising tool to test different models of reionisation and requires high-resolution spectroscopy of a fair number of bright sources at the redshift of interest. This measurement enables the indirect study of objects aligned with and hence outshined by $z \sim 6$ quasars. We have therefore conducted a semi-automated search for C IV absorbers in order to study how these absorbers can trace potential sources of ionising photons and gathered the largest sample of C IV absorbers at $4.3 < z < 6.2$. We have updated the measurements of C IV cosmic density, confirming its rapid decline with redshift. Through abundance-matching arguments, we have identified C IV as being associated with $M_{\rm UV} \lesssim -16$ faint galaxies in $\log M_h/{\rm M}_\odot \gtrsim 10$ haloes.

We have detected excess H I absorption in the Lyman-α forest at the redshift of $z \sim 5-6$ C IV absorbers, at a similar scale to that of the IGM absorption around lower redshift LBGs. We have also detected an excess transmission at 2.7σ on larger scales in the correlation of C IV with the IGM transmission. We interpret this excess as a signal of the reionisation process driven by galaxies clustered around C IV absorbers. Using the model developed in Kakiichi et al. (2018), we have put constraints on the product of the escape fraction and the LyC photon production efficiency $\log\langle f_{\rm esc}\xi_{\rm ion}\rangle = 25.01^{+0.30}_{-0.19}$. Although caveats about the observation and the modelling remain, we have shown that C IV absorbers trace different galaxies than the ones clustered around LAEs (Kakiichi et al. 2018), with either higher spectral hardness or possibly larger escape fractions.

More quasar sightlines are needed to fully sample cosmic variance and provide a better measurement of the correlation. Larger numbers of sightlines and absorbers would not only improve the statistics but also allow a study of the redshift evolution of the escape fraction and LyC production efficiency of the probed galaxies. We point out that a decrease of the cosmic density of C IV makes it more difficult to trace the same objects at all redshifts. However at higher redshift different metal absorbers such as Mg II or Si IV could be used to trace galaxies. In doing so, we would probe as well different ionising environments and possibly different galaxy populations. Eventually, the degeneracy with the spectral hardness of our measurement can be broken by harvesting our large sample of aligned metal absorbers to probe their ionisation state using forward modeling. Radiative transfer simulations with non-uniform UVB including the tracking and modelling of the different metal gas phases could reproduce the correlation, provided large enough boxes and sightlines can be produced in a reasonable amount of time. This opens new avenues into the question driving this thesis: the nature of the sources of reionisation.

This page was intentionally left blank

Chapter 4

The cross-correlation of $z \sim 6$ galaxies with the reionising IGM

This Chapter has been published in Meyer et al. (2020 MNRAS, 494, 1560). The imaging data for the fields of J1030, J1148, J0836 was reduced by K. Kakiichi and K. Mawatari. The catalogues of sources in these fields were prepared by K. Kakiichi. The quasar spectra were reduced by S.E.I. Bosman.

4.1 Introduction

In the previous chapter, I extended the cross-correlation framework developed by Kakiichi et al. (2018) to enable us to correlate metal absorbers, considered to be hosted by sub-luminous LAEs/LBGs, with the IGM transmission measured in the Lyman-α forest of quasars. Kakiichi et al. (2018) was a pilot study that analysed only one quasar sightline, which raised the question of the statistical significance of the tantalising proximity effect detected. Indeed, it was shown in Chapter 3 that cosmic variance between sightlines is an important factor also noted independently in simulations (Garaldi et al. 2019). Though Chapter 3 overcame this issue by sampling C IV absorbers at $4.5 < z < 6$ in 25 quasar sightlines and detected an excess of

transmission around C IV absorbers, the assumption they were a proxy for massive galaxies raised questions regarding the nature of the CIV hosts. Nonetheless, both studies suggested that the faint population of galaxies at $z \sim 6$ has a high ensemble-averaged escape fraction and/or ionising efficiencies required to sustain reionisation (e.g. Robertson et al. 2015).

In this fourth chapter, I present an improved study of the correlation between galaxies and the IGM at the end of reionisation and the resulting inference on the ionising capabilities of the sub-luminous population. We have gathered an extensive dataset of galaxies in the fields of 8 quasars at $z > 6$ through an additional observational campaign with DEIMOS/Keck as well as archival MUSE/VLT data. Moreover, I extend the analytical framework presented in Kakiichi et al. (2018 see 1.5.3 for a summary) to compute the enhanced photoionisation rate due to clustered galaxies. I include the effect of gas overdensities instead of assuming the average IGM density around LAEs and LBGs, investigate the impact of a spatially varying mean free path, peculiar velocities and finally forward model the effect of flux uncertainties on different galaxy-IGM cross-correlation statistics. Finally, I present a new probe of the galaxy-IGM connection by measuring and modelling the two-point cross-correlation function (2PCCF) between galaxies and transmission spikes in the Lyman-α forest of background quasars.

The plan of this chapter is as follows. Section 4.2 introduces our new dataset, starting with the 8 high-redshift quasar spectra used in this study. Section 4.2.2 presents DEIMOS/Keck spectroscopic data of Lyman-break galaxies in three quasar fields with multi-slit spectroscopy. Section 4.2.3 details our dataset drawn from MUSE archival observations, and our search for Lyman-α Emitters in the Integral Field Unit (IFU) datacubes. In Section 4.3, we present the galaxies detected in the field of background quasars with redshifts overlapping with the IGM probed by the Lyman-α forest, and the cross-correlations of galaxies with the surrounding IGM are detailed in Section 4.4. Section 4.5 presents an extension to the analytical framework of Kakiichi et al. (2018) necessary to interpret our new measurements. The final results and constraints on the ionising capabilities of galaxies at the end

of reionisation are presented in Section 4.6. We discuss the use of our measurement to differentiate between the relative contributions of faint and bright galaxies to reionisation and the difference between the cross-correlation statistics in Section 4.7 before concluding in Section 4.8.

Throughout this chapter, the magnitudes are quoted in the AB system (Oke 1974), we refer to proper (comoving) kiloparsecs as pkpc (ckpc) and megaparsecs as pMpc (cMpc), assuming a concordance cosmology with $H_0 = 70$ km s^{-1} Mpc^{-1}, $\Omega_{\rm M} = 0.3,\ \Omega_\Lambda = 0.7$.

4.2 Observations

This chapter focuses on the measurement and modelling of the correlations between galaxies and the surrounding IGM at the end of reionisation. In order to achieve that goal, we have gathered different datasets of luminous galaxies acting as signposts for overdensities of less luminous sources. We have continued the approach of Kakiichi et al. (2018) by confirming high redshift Lyman-break galaxies (LBG) via their Lyman-α emission with the DEep Imaging Multi-Object Spectrograph (DEIMOS, Faber et al. 2003) at Keck. For convenience, we refer to those as LBGs because of their selection technique (although formally they are all also Lyman-α emitters (LAEs) given they were confirmed with this line). Throughout this chapter, we thus reserve the terminology LAE for galaxies detected blindly in archival IFU data of the Multi Unit Spectroscopic Explorer (MUSE, Bacon et al. 2010) at the VLT. MUSE complements the early DEIMOS approach since we use a different selection method for galaxies and probe the smaller scales appropriate to the small MUSE Field of View (FoV) ($1'$ corresponding to ~ 360 pkpc at $z \sim 5.8$). These complementary datasets of galaxies were gathered in the field of $z \sim 6$ quasars with existing absorption spectroscopy of the Lyman-α forest, which probes the IGM transmission and ultimately enables us to compute the galaxy-IGM correlations. We now proceed to describe this rich observational dataset, starting with the quasar spectra and moving then to the DEIMOS and MUSE data.

4.2.1 Quasar spectroscopic observations

The 8 quasar fields studied in this work were chosen to have existing moderate or high Signal-to-Noise ratio (SNR) spectroscopy of the Lyman-α forests and either be accessible to Keck for DEIMOS follow-up or have archival MUSE data with adequate ($\geq$ 2h) exposure time. The quasar spectra used in this study were downloaded from the ESO XShooter Archive or the Keck Observatory Archive (ESI). We use the same ESI spectrum of J1148 as in Kakiichi et al. (2018), originally observed by Eilers et al. (2017). The quasars already presented in Bosman et al. (2018 J0836, J1030) were reduced using a custom pipeline based on the standard ESOREX XShooter recipes as detailed therein. The remaining quasars (J0305, J1526, J2032, J2100, J2329) were reduced with the open-source reduction package PYPEIT [1] (Prochaska et al. 2019). The quasars have a median SNR of $\sim$ 16 in the rest-frame UV. The spectra were finally normalised by a power-law $f(\lambda) = A\lambda^b$ fitted to the portion of the continuum redwards of Lyman-α relatively devoid of broad emission lines (1270 $-$ 1450 Å), as described in Bosman et al. (2018), to compute the transmission in the Lyman-α forest. Table 4.1 summarises the quasar spectroscopic data information alongside the galaxy detections.

[1] https://github.com/pypeit/PypeIt

Quasar	z	LAEs	LBGs	DEIMOS ID	MUSE ID	Total Exptime	ref	Spectrum	ref
J0305−3150	6.61	3	-	-	094.B-0893(A)	2h30m	(2)	XShooter	(3)
J0836+0054	5.81	-	1	U182	-	5h10m	(1)	XShooter	(5)
J1030+0524	6.28	7	8	C231, U182	095.A-0714(A)	10h50m / 6h40m	(1)/(4)	XShooter	(5)
J1148+5251	6.419	-	4	C231, U182	-	13h10m	(1,6)	ESI	(5)
J1526−2050[a]	6.586	2	-	-	099.A-0682(A)	3h20m	(7,8)	XShooter	(9)
J2032−2114[b]	6.24	3	-	-	099.A-0682(A)	5h	(7,8)	XShooter	(10)
J2100−1715	6.09	4	-	-	097.A-5054(A)	3h40m	(7,8)	XShooter	(11)
J2329−0301	6.43	2	-	-	060.A-9321(A)	2h	(7,8)	ESI	(12)

Table 4.1: Observational data summary of the quasar fields. Fields: 1) Quasar name (*a:* Also known as P231-20. *b:* Also known as P308-21) 2) Quasar redshift 3) Number of suitable LAEs in the Lyman-α forest redshift range of the nearby quasar detected with MUSE 4) Number of suitable LBGs detected with DEIMOS 5) DEIMOS Keck programme ID 6) MUSE ESO programme ID 7) Total exposure time of DEIMOS/MUSE in the field 8) Reference of the original published paper on the observational programme 9) Instrument used for the quasar spectrum 10) Reference of the discovery paper of the quasar. References: (1) This work (2) Farina et al. (2017) (3) Venemans et al. (2013) (4) Diaz et al. (2020) (5) McGreer et al. (2015) (6) Kakiichi et al. (2018) (7) Drake et al. (2019) (8) Farina et al. (2019) (9) Mazzucchelli et al. (2017) (10) Bañados et al. (2016) (11) Willott et al. (2007) (12) Bosman et al. (2018)

4.2.2 DEIMOS spectroscopy of LBGs in 3 quasar fields

As part of this study we have re-observed the field of quasar J1148+5251 explored in Kakiichi et al. (2018) to improve our selection of LBGs. As the slitmask design of DEIMOS does not allow small slit separations, only a selected subset of LBGs can be observed in any given mask. Accordingly, the detection of the proximity signal in Kakiichi et al. (2018) might be affected by the sampling of candidate LBGs in the field. We also include data for two new quasar fields: J1030+0524 ($z = 6.3$) and J0836+0054 ($z = 5.8$) (see Table 4.2).

Deep ground-based photometry of the three fields was used to construct catalogs of r' and i' drop-outs for follow-up. The fields of J1030 and J1148 have been observed in the SDSS r'-, i'-, z'-band filters with the Large Binocular Camera (LBC, Giallongo et al. 2008) at the Large Binocular Telescope (LBT, Hill & Salinari 2000), with the publicly available photometry reduced by Morselli et al. (2014) [2]. For the field of J0836, we used r'-, i'-, z'-band observations (Ajiki et al. 2006) taken with SuprimeCam on the 8.2m Subaru Telescope (Kaifu et al. 2000; Miyazaki et al. 2002). We chose the following colour cuts to select potential $z \sim 5-6$ LBG candidates (see Figure 4.1)

$$[r' - i' > 1.0] \wedge [i' - z' < 1.0] \wedge [z' < z'(3\sigma)] \tag{4.1}$$

for r'-drop-outs, and

$$[i' - z' > 1.0] \wedge [r' > r'(2\sigma) \vee r' - z' > 1.75] \wedge [z' < z'(3\sigma)] \tag{4.2}$$

for i-drop-outs, where r', i', z' indicates the magnitude in the corresponding SDSS band, and $r'(2\sigma), z'(3\sigma)$ the limiting $2, 3\sigma$ magnitude in the r', z'-band image respectively. All candidates were visually inspected to produce a final catalogue. In designing the DEIMOS masks, we prioritised drop-outs based on the strength of their colour drop ($i' - z'(r' - i') > 1.0, 1.3, 1.5$) or r'-band non-detection ($r' >$

[2] http://www.oabo.inaf.it/~LBTz6/1030/lbtz6.html

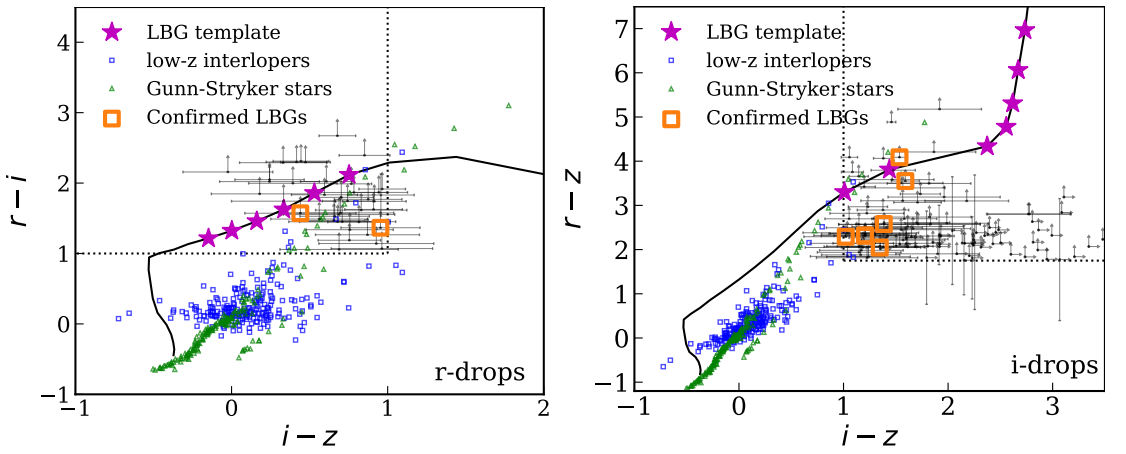

Figure 4.1: r-drop-out (top) and i-drop-out (bottom) selection of candidate LBGs in the fields of J0836, J1030 and J1148. A LBG template (magenta stars, black line) falls into the colour-colour cuts (dotted lines) at the redshift of interest $z \sim 5-6$. Galactic stars (green triangles, Gunn & Stryker 1983) and low-redshift interlopers (blue squares, VUDS-DR1 samples from the COSMOS field, Le Fèvre et al. 2015; Tasca et al. 2017) are however mostly rejected. The target candidates are in shown in black, and confirmed LBGs are highlighted with orange squares.

$r'(2\sigma, 3\sigma, 5\sigma)$) to optimise the chance of confirmation. Priority was always given to better candidates first, and then to i'-drop-outs over r'-drop-outs of the same quality. The masks contained $\sim 25-40$ slits for science targets and 5 or 6 square holes for alignment stars.

The candidates were observed with the DEIMOS instrument (Faber et al. 2003) at the Keck II 10-m telescope during two observing runs in 2017 March 26-27 (PI: A. Zitrin) and 2018 March 07-08 (PI: B. Robertson). We confirmed 13 LBGs in the 3 fields over the course of 4 nights in good conditions with a seeing of $0.7-0.9''$ except for one night at $0.9-1.5''$, as summarised in Table 4.2. The masks and the LBG detections are shown in the 3 fields in Figure 4.2. Surprisingly we could not confirm any new LBG in the field of J1148+5251, although Kakiichi et al. (2018) confirmed 5(+1 AGN) in a smaller exposure time. The other masks in J0836 and J1030, exposed for 1h30m to 5h10m, yielded 2 to 4 LBG confirmations each. The completeness of our search for LBGs in the relevant redshift range is not straightforward to estimate but fortunately not a major concern for our analysis. Whilst in principle the total number of galaxies in these fields can be estimated using the UV LF and the depth of the photometric data, we find no proportionality

Quasar	$N_{\rm LBG}$	$\#_{\rm Mask}$	Exptime	Seeing
J0836+0054	4[a]	K1	5h10m	$0.7'' - 0.9''$
J1030+0524	3	K1	4h00m	$0.9'' - 1.5''$
	3	K2	5h20m	$0.7'' - 0.9''$
	2	K3	1h30m	$0.7'' - 0.9''$
J1148+5251	4[b]	K1[b]	4h30m	$0.7'' - 1.5''$
	0	K2	8h40m	$0.7'' - 0.9''$

Table 4.2: Summary of the DEIMOS observations. The masks J1030-K1, J1030-K2 and J1148-K1 were observed in 26-27 March 2017 (PI: A. Zitrin, ID: C231) and the remainder in 07-08 March 2018 (PI: B. Robertson, ID: U182). The number of confirmed LBGs is only weakly correlated to the total exposure time on the mask. **a)** 3 LBGs in J0836+0054 are in the near-zone of the quasar and hence they do not appear in Table 4.1 as they are not suited for our purposes. They will be studied in greater detail Bosman et al. (2020). **b)** We remove the faint AGN as well as the least convincing LBG detection (ID = 022) presented in Kakiichi et al. (2018) to harmonize the LBG selection.

with the observed numbers of LBGs. The scatter from field to field detailed above is therefore mainly driven by the random selection of objects on each mask ($\lesssim 20$ 'prime' candidates) and the number of mask observed for each given field. Indeed, the number of objects confirmed per mask is roughly constant, regardless of the redshift and field. However, this does not affect our results since we are aiming to measure the average Ly-α transmission around the *detected* bright galaxies only. As we cross-correlate them with the Ly-α forest and do not measure their number density, we do not need to correct for completeness (see further Section 4.4).

The data were reduced using the DEEP2 pipeline (Cooper et al. 2012; Newman et al. 2013), and the 1D spectra were extracted using optimal extraction with a $1.2''$ boxcar aperture (Horne 1986). The 2D spectra were inspected visually by five collaborators (R.A. Meyer, K. Kakiich, S.E.I. Bosman, R.S. Ellis, N. Laporte) for line emitters. Candidate LBGs were retained if they were found by 3 authors or more in the 2D spectra. We show examples of a clear LBG detection and a less convincing one in Figure 4.3. The remaining LBG detections are presented individually in Appendix B.1. We also present serendipitous line emitters (without optical counterpart) which are not used in this study.

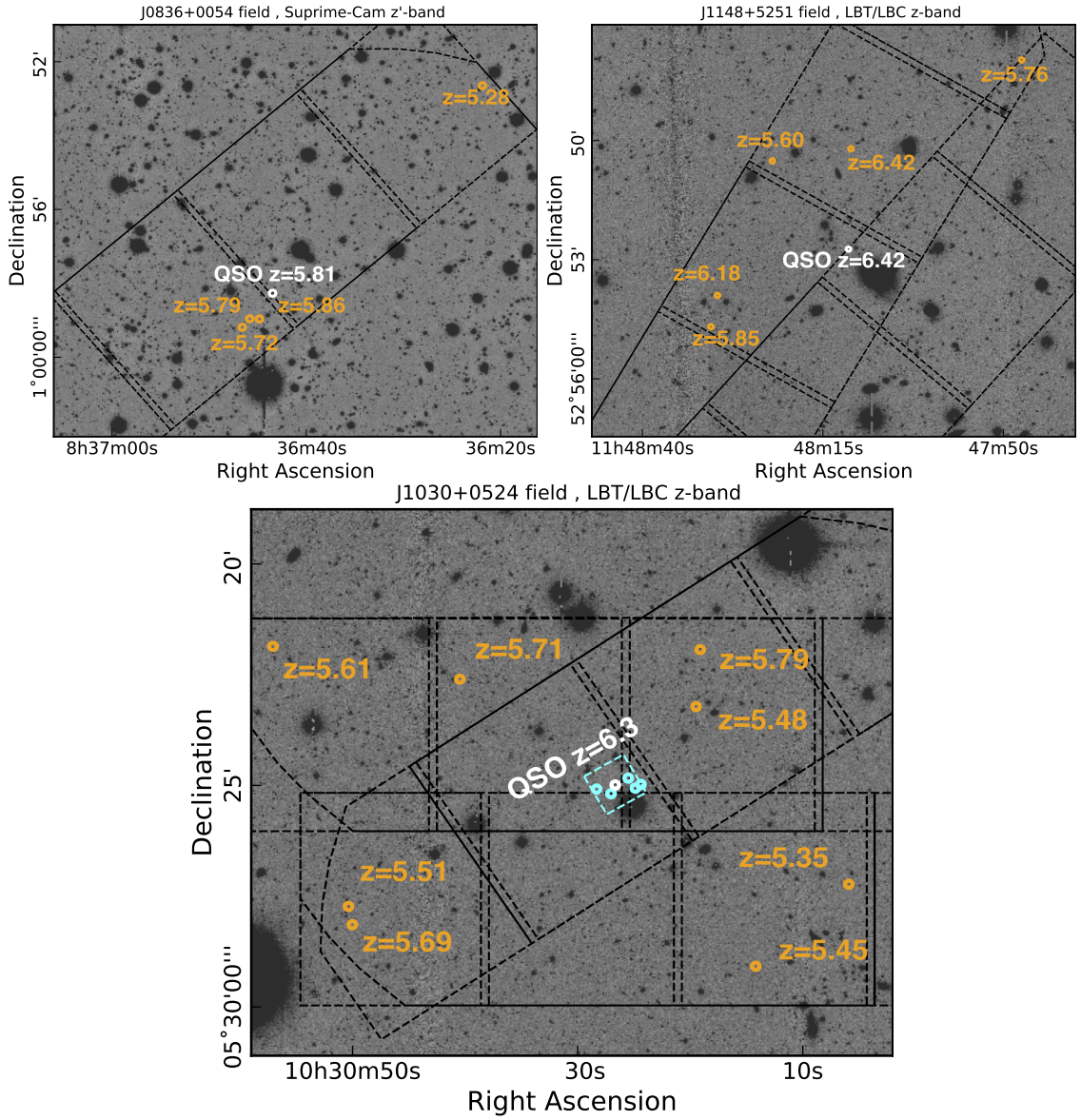

Figure 4.2: Spectroscopically confirmed high-redshift LBGs (orange squares) and potential LAEs (green circles) identified with DEIMOS in the quasar fields of J0836, J1030, J1148. We overlay the DEIMOS slitmask FoV in black. For J1030 (lower panel), we also add the MUSE FoV (cyan dashed square) and the LAEs detected in the MUSE datacube (cyan circles)

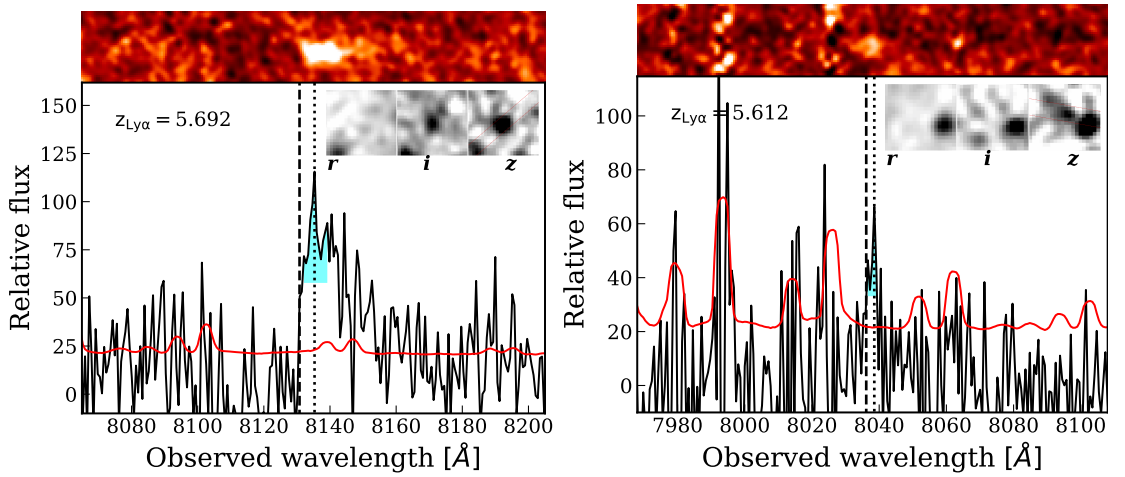

Figure 4.3: Examples of a clear and a marginal detection of Lyman-α emission for LBG-selected galaxies in the field of J1030. The top panels shows the 2D spectra from which the spectrum (black line) and noise (red) are optimally extracted using a boxcar aperture of 1.2". The peak of Lyman-α is identified with a dotted line, and the location of the systemic by a dashed line. The systemic redshift is found by applying a correction based on the FWHM of the Lyman-α emission (cyan, see Section 4.2.4). In the upper right corner is displayed the *riz* photometry used for the drop-out selection. The remaining detections are presented in Appendix B.1.

4.2.3 Archival MUSE quasar fields

We exploit 6 $z \sim 6$ quasar fields with deep ($\gtrsim$ 2h) archival MUSE data to search for galaxies close to the sightline. The MUSE quasar fields are listed in Table 4.1.

We reduce the archival MUSE data using the MUSE v2.6 pipeline (Weilbacher et al. 2012, 2015) with the standard parameters. We further clean the skylines emission using the Zurich Atmospheric Purge (ZAP) code (Soto et al. 2016). After masking the bright sources and the edges of the data cubes, we run MUSELET (Bacon et al. 2016) and LSDCat (Herenz & Wisotzki 2017) to extract line emitters. We find that both algorithms are complementary due to their different search strategy. MUSELET reduces the IFU cube to a series of narrow band images (6.25 Å width) and uses SExtractor to identify emission lines by subtracting a median continuum constructed from the adjacent wavelength planes. Detections in several narrow bands at similar positions can be grouped together to find a redshift solution. Whilst it is a robust technique, continuum absorptions or rapid continuum variations often produce spurious detections (when the adjacent narrow bands are subtracted). LSDCat improves the removal of foreground continuum objects by

utilizing a median filtering of the cube in the wavelength dimension. The emission lines are then detected with a 3-dimensional matched-filtering approach. LSDCat also allows one to mask brighter sources with custom masks. It however then fails to pick faint sources next to bright objects if the masking and/or the median filtering is too aggressive. Finally, the width of the narrow-bands in MUSELET and the convolution kernel sizes of the matched-filter in LSDCat can lead to different false positives or negatives. Therefore, we use both algorithms to generate a consolidated list of line emitter candidates which are then visually inspected to compile a robust sample of high-redshift LAEs. We check that candidates are present in the two datacubes produced with the two halves of the exposures to remove cosmic rays and other artifacts, and we remove double peaked emissions which are likely to be low-redshift [O II] $\lambda\lambda$ 3727 Å doublets as it would mimic a double-peaked $z \sim 5-6$ Lyman-α with a reasonable peak separation $\Delta v \sim 220$ km s^{-1}. Double-peaked Lyman-α profiles at $z > 5$ are exceedingly rare due to absorption by the IGM (Hu et al. 2016; Songaila et al. 2018; Matthee et al. 2018), so we expect to lose very few high-redhift LAEs in being so conservative. Finally, we produce a white light image of the MUSE cube and check that the line detection is not caused by a poor continuum subtraction of a bright foreground object or a nearby contaminant (see Figure B.9 for typical false positives). We show two representative detections in Figure 4.4 and the remainder in Appendix B.2.

We checked that the number of LAEs is consistent with expectations from the LAE luminosity function (LF) integrated down to the MUSE sensitivity limit. We first compute the number of LAEs in a given comoving volume using the LAE LF from de La Vieuville et al. (2019); Herenz et al. (2019) which measured the LAE LF in deep MUSE datacubes with MUSELET and LSDCat, respectively, i.e. the same algorithms that we use. By comparing the numbers of LAEs predicted with the LAE LF on a deep 27h field realised by the MUSE GTO team (Bacon et al. 2015) to the numbers of LAEs those authors recovered with LSDCat, we estimate that LSDCat has a recovery rate of $\simeq 32\,\%$ for LAEs at $z \sim 5.5$. This is a global recovery rate for all LAEs with peak flux density greater than $f > 4.8 \times 10^{-19}$erg s^{-1}cm^{-2},

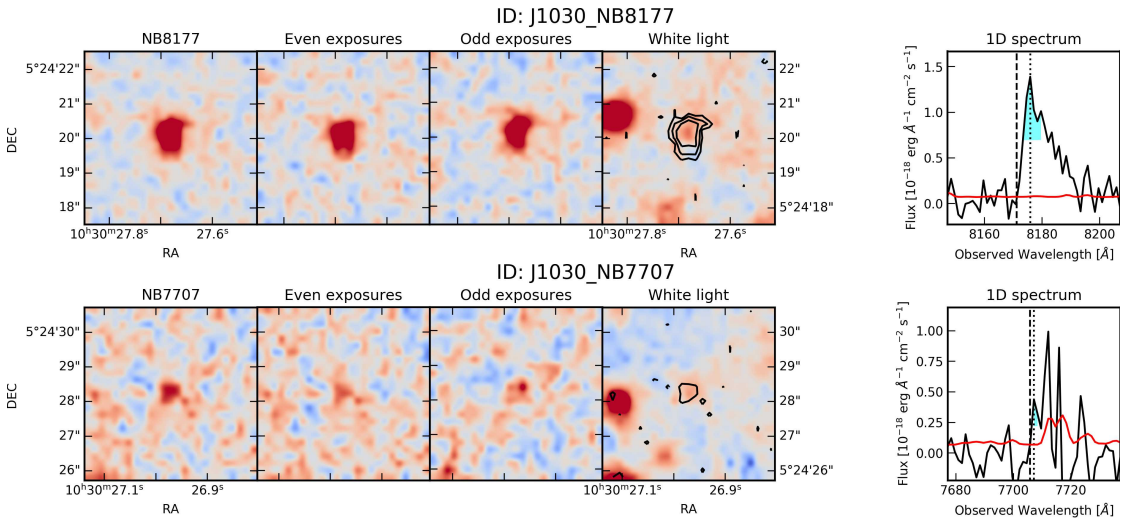

Figure 4.4: Two representative LAEs detected in MUSE archival data in the field of quasar J1030+0524. The first detection is a strong LAE in the field of J1030 already reported by Diaz et al. (2015). The four first panels show, in order, the detection in the narrow band centred on the detection in the full combined cube, followed by the same location in cubes with either half of the exposures. The fourth panel shows the white light image with superimposed black significance levels $(-3\sigma, 3\sigma, 5\sigma, 8\sigma)$ of the narrow band detection. The fifth panel shows the optimally extracted 1D spectra (black line) and the noise level (red line). The cyan shading highlights the FWHM of the line, used to correct the peak redshift (dotted vertical line) to systemic (dashed vertical line). The remainder of the LAE detections are summarised in Table B.2 and individual detections are presented in Appendix B.2.

which is below the sensitivity of all MUSE observations used in this study. We then predict the number of LAEs we expect to find in each of our MUSE quasar fields depending on the exposure time, effective survey area, and the redshift interval of the central quasar Lyman-α forest. We find good agreement between the predicted number (including LSDCat efficiency) and the number of retrieved LAEs (Figure 4.5), indicating a successful search for LAEs and low levels of contamination.

Table 4.1 summarises all the LBGs and LAEs detected in our quasar fields, alongside the reference of the MUSE and DEIMOS programmes, and the quasar discovery reference study.

4.2.4 Correcting the Lyman-α-based redshifts

The red peak of the Lyman-α emission line, commonly observed without its blue counterpart at high-redshift due to the increasingly neutral IGM, is often shifted

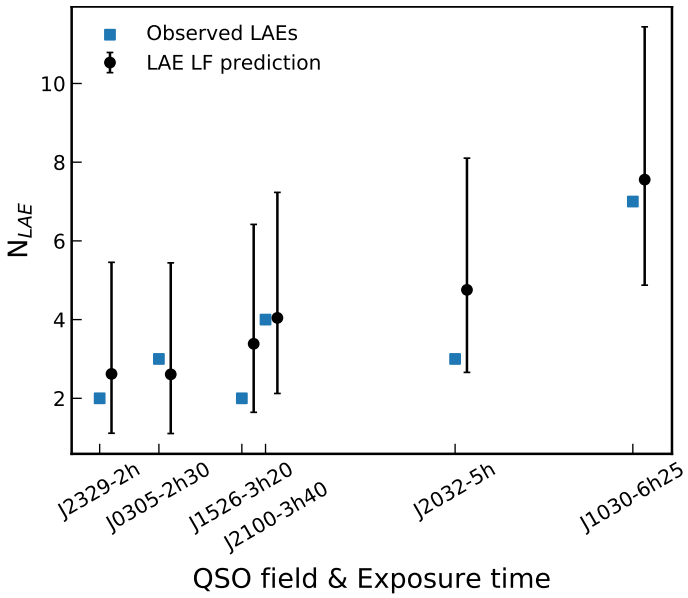

Figure 4.5: Predicted number of recovered LAEs in each MUSE cube (black 1-sigma Poisson ranges) compared to the number of retrieved LAEs (blue squares), with the quasar fields sorted (x-axis) by increasing exposure time of the MUSE data. The prediction is made from the LAE LF of (de La Vieuville et al. 2019; Herenz et al. 2019) in the redshift range of the Lyman-α forest of each central quasar and effective efficiencies of LSDCat and MUSE reduction. We apply an effective efficiency of the MUSE LAE search which is derived by comparison with the 27h GTO observation of a single field searched for LAEs at high-redshift with LSDCat in Bacon et al. (2015).

from the systemic redshift of the galaxy. Steidel et al. (2010) give a typical offset of $v_{\rm red}^{\rm peak} \sim 200\,{\rm km\ s^{-1}}$, but the range is large and can span $\sim 0 - 500$ km s^{-1} at high redshift (e.g. Vanzella et al. 2016; Stark et al. 2017; Hashimoto et al. 2018b). A velocity offset of ~ 200 km s^{-1} translates at $z \sim 6$ to ~ 2 cMpc (~ 280 pkpc), which is not negligible given the expected scale of ~ 10 cMpc for the peak of the cross-correlation signal. As the cross-correlation is computed in 3D space and then radially averaged, any offset would damp the sought-after signal.

In order to get a better estimate of the systemic redshift of the galaxy, we thus apply a correction to the Lyman-α redshift based on the Full-Width-Half-Maximum (FWHM) of the line. We follow the approach of Verhamme et al. (2018) who developed an empirical calibration using the FWHM directly measured from the data without modelling

$$v_{\rm red}^{\rm peak} = 0.9(\pm 0.14) \times {\rm FWHM(Ly}\alpha) - 34(\pm 60)\ {\rm km\ s^{-1}}\ . \tag{4.3}$$

The measured FWHM values of our LAEs (LBGs) all fall in the expected range 100 km s^{-1} $\lesssim$ FWHM $\lesssim$ 400 km s^{-1}, and are indicated for each LAE (LBG) in Table B.2. Throughout this Chapter, we use these corrected redshifts for the purpose of computing galaxy-IGM cross-correlations.

4.3 The apparent clustering of galaxies and Lyman-α transmission spikes from 8 quasar fields

Galaxies are usually thought to be responsible for reionising the Universe and driving the UVB fluctuations at $z \sim 6$. Having gathered a sample of 21 LAEs and 13 LBGs in the redshift range of the Lyman-α forest of nearby quasars, we are in a position to investigate the direct impact of galaxies on the surrounding IGM. The observational result of our work is summarised by Figure 4.6, where we overlay the detected LAEs and LBGs with the transmission features found at the same redshift in the Lyman-α forest of the background quasar.

The natural corollary to the large-scale underdensity of galaxies found around highly opaque sightlines (Becker et al. 2018; Kashino et al. 2019) would be a close association between overdensities of transmission spikes and detected galaxies. We find that LAEs and LBGs are found close to at least one transmission spike in all quasar sightlines, but it is difficult to conclude at first sight whether they trace local spike overdensities. Moreover, this is not a reciprocal relation: some large transmission spikes are not associated with any detected galaxy. Two of our quasars, J1030 and J2032, illustrate this complicated relationship very well. The two sightlines both have a transparent region at $z \sim 5.5$, followed by a relatively opaque one at $z \sim 6$, and a similar MUSE exposure time. The transparent region in J1030 is associated with a large overdensity of LAEs and LBGs. In contrast, it is the few transmission spikes in the high-redshift opaque region in the sightline of J2032 that are associated with detected LAEs, whereas only one LAE is detected in the transparent region at $z \sim 5.5$. The detection of LAEs across $5 \lesssim z \lesssim 6$ in both quasar fields implies that we do not miss existing LAEs in these sightlines.

Hence studying the correlations between galaxies and the IGM must be conducted in a more quantitative manner. In the following section, we compute the cross-correlation of the galaxies' positions with the transmission and the position of selected transmission spikes in the Lyman-α forest of the background quasar.

4.4 The correlation of galaxies with the surrounding IGM transmission

In this section, we introduce two indicators of the link between galaxies and the ionisation state of the surrounding IGM: the cross-correlation of galaxies with the transmitted flux, and the 2-point correlation function (2PCCF) of galaxies with selected transmission spikes. We present the mean transmitted flux around galaxies (the quantity used in Kakiichi et al. (2018)) in Appendix B.3 given that method has been superseded by the transmitted flux cross-correlation. Although the mean transmission around galaxies is the most intuitive measure of an enhanced UVB due to faint LyC leakers, in practice this measurement is dominated by the cosmic variance between sightlines and the redshift evolution of the IGM opacity, as noted in Chapter 3. For the purpose of these cross-correlation measurements, we only consider the Lyman-α forest between 1045 Å (to avoid the intrinsic Lyman-β/O VI quasar emission) and the start of the near-zone of the quasar, and consider only galaxies whose Lyman-α emission would fall in the same observed wavelength range[3]. These boundaries are plotted in dashed black lines in Figure 4.6.

4.4.1 The cross-correlation of the IGM transmission with field galaxies

We first compute the cross-correlation of the IGM transmission with galaxies. We measure the transmission in the Lyman-α forest at a comoving distance r of observed galaxies (DD) and of random mock galaxies (DR). The distance $r = (r_\perp + r_\parallel)^{1/2}$ is computed from the angular diameter distance $r_\perp$ of the galaxy to the quasar

[3]In the following, we loosely describe these galaxies as "being in the (redshift range) of the Lyman-α forest of the quasar".

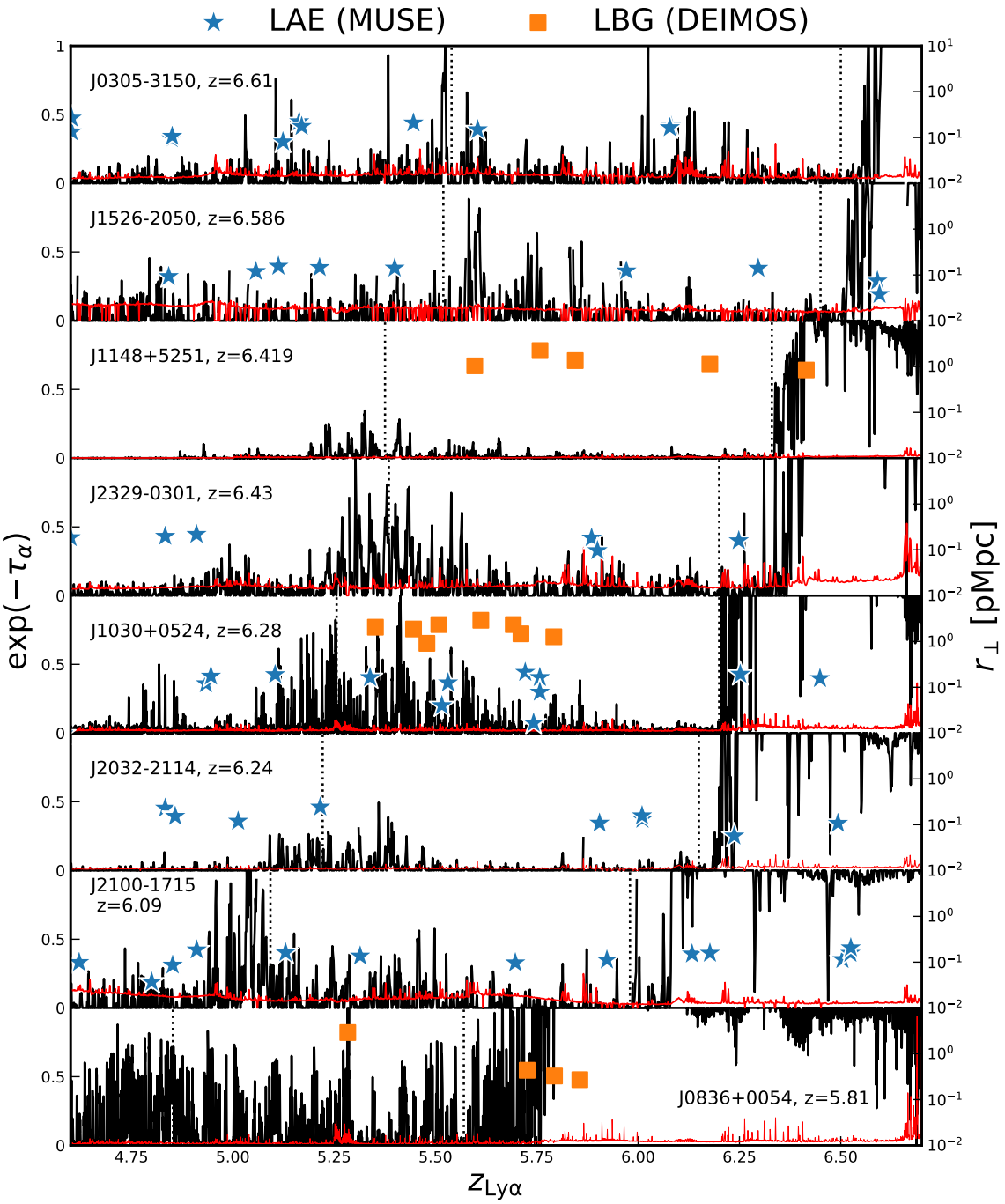

Figure 4.6: The Lyman-α forests of our high-redshift quasar sample and galaxies detected in the quasar fields. Whilst the average transmission is clearly decreasing with increasing redshift, the galaxies (LAEs in particular) seem to cluster with transmission spikes in some sightlines. The transmission in the quasar Lyman-α forest (black) is indicated on the left axis whilst the right axis indicates the transverse distance of foreground galaxies to the quasar sightline. LBGs detected with DEIMOS are indicated in orange squares. MUSE LAE detections are indicated with blue stars. Only J1030 displays both LAEs and LBGs as it is the only field with DEIMOS and MUSE data. The vertical black dotted lines indicate the redshift range of the Lyman-α forests as defined in Section 4.4.

sightline and the comoving distance parallel to the quasar line-of-sight $r_{\parallel}$. For each quasar field, we compute the probability of detecting a LAE at a given redshift in the quasar Lyman-α forest considering the LAE LF (de La Vieuville et al. 2019; Herenz et al. 2019) and the depth of the MUSE data. The redshifts of random galaxies are sampled from this probability distribution and the angular distances from the quasar sightlines are chosen uniformly up to $1'$ to mimic the MUSE FoV. For LBGs we sample the UV LF (Bouwens et al. 2015; Finkelstein et al. 2015; Bowler et al. 2015; Ono et al. 2018) at the 2σ depth of the photometry of our 3 fields with an appropriate k-correction ($2.5(\alpha - 1)\log_{10}(1 + z)$, with $\alpha = 2$) and the angular separation from the quasar is drawn uniformly in the $4' \times 16'$ DEIMOS FoV. As noted in Section 2.2, the number of LBGs detected in each field depends primarily on the number of slitmasks observed in each field, rather than the depth of the photometric data used for selection. If we sampled each field down to the 3σ detection limit in the z band, we would thus predict similar numbers of observable LBGs per field. However, a random sample created in this way would have a mean number of transmission spikes around the LBGs larger than is actually observed around our spectroscopically confirmed galaxies. Indeed, those observations which targeted higher redshift fields with reduced IGM transmission (e.g. J1030,J1148) have greater spectroscopic coverage (due to the use of more than one DEIMOS mask) than for the lower-redshift field of J0836. We therefore construct random LBGs by sampling the UV LF to the limiting depth (z') of each field matching the observed redshift distribution, but oversampling by the number of spectroscopic confirmations in each field. This procedure still reproduces the decline of the number of galaxies with redshift in each individual field.

The cross-correlation is then estimated using the standard estimator

$$\xi^{\exp(-\tau_\alpha)}_{\mathrm{Gal-Ly}\alpha}(r) = \frac{DD(r)}{DR(r)} - 1 \ . \tag{4.4}$$

Normalising the transmitted flux in this way removes the bias introduced by the evolving IGM opacity, and allows us to average sightlines without being biased by

the most transparent ones (see Appendix B.3).

We present the cross-correlation independently for LAEs and LBGs in Figure 4.7. We do not find any evidence for an excess transmission, unlike that seen around lower-redshift C IV absorbers in Chapter 3. The signal is still dominated by the small number of objects and sightlines as the large uncertainties show. The errors are estimated by bootstrapping the sample of detected galaxies, and thus they might be even underestimated given the small sample of sightlines and the large cosmic variance seen between Lyman-α forest at that redshift (Bosman et al. 2018). The issue is potentially more acute for LBGs as the selection is not complete down to a given luminosity as 1) only a fraction of drop-out candidates could be observed per field due to the instrument and time constraints 2) only a fraction of LBGs have a bright Lyman-α line (e.g. Stark et al. 2010; Ono et al. 2012; De Barros et al. 2017; Mason et al. 2019). Finally, we did not remove completely opaque parts of the Lyman-α where the flux is below the noise level unlike in Chapter 3, as it would greatly reduce our sample. The measured fluxes are therefore dominated by noise in some sections of the Lyman-α forest, which decreases the signal. Nonetheless, we find that the absorption on small scales $\lesssim 10$ cMpc around LAEs is similar to that seen for C IV absorbers in Chapter 3. The direct association or not of C IV with LAEs is outside the scope of this project and is studied in Diaz et al. (2020).

4.4.2 The 2-point correlation of galaxies with selected transmission spikes

At $z \gtrsim 5.5$, the opacity of the IGM has increased sufficiently that the Lyman-α forest resembles more a "savannah" than a forest: a barren landscape of opaque troughs occasionally interrupted by transmission spikes. At these redshifts the effective opacity can only be constrained with an upper limit ($\tau_{\rm eff} \gtrsim 3-4$), and the average opacity in large sections of the Lyman-α forest falls below this limit. It is thus not surprising that the previous transmission cross-correlation fails to capture the link between galaxies and the reionising IGM. Indeed, the normalisation term $DR(r)$ is often ill-defined at $z \gtrsim 6$ since the average flux measured is below or at the

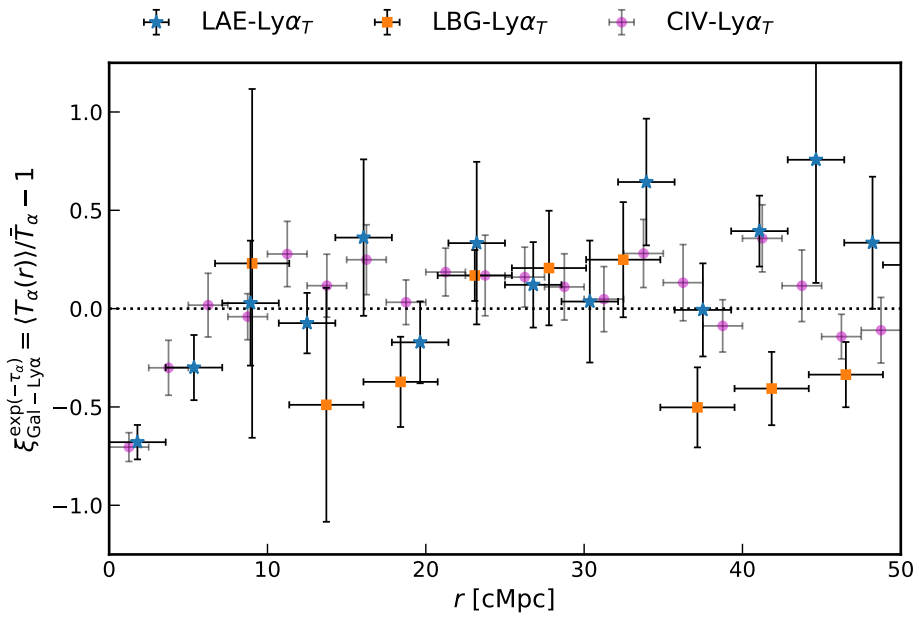

Figure 4.7: Cross-correlation of the position of LAEs (blue stars), LBGs (orange squares), with the IGM transmission in the Lyman-α forest of the background quasars. The errorbars are computed by bootstrap resampling of the sample of detected galaxies. The significant deficit of transmission in the first bin of the LAE transmission cross-correlation is consistent with that measured around high-redshift C IV absorbers in Chapter 3 (magenta dots).

level of the noise of the spectrograph. To circumvent this issue we examine the extrema of the opacity distribution rather than its mean by utilizing the *2-Point Cross-Correlation Function (2PCCF) of galaxies with selected transmission spikes* in the Lyman-α forest of quasars. We expect the observed Lyman-α transmission to be more sensitive to fluctuations of the extrema of the distribution, making the 2PCCF theoretically more suited to capturing small perturbations due to clustered faint contributors to reionisation.

We thus measure the 2-point cross-correlation function (2PCCF) between galaxies and selected transmission spikes in the Lyman-α forest. We identify transmission spikes with a Gaussian matched-filtering technique (e.g. Hewett et al. 1985; Bolton et al. 2004). We use Gaussian kernels with $\sigma = [10, 14, 20]$ km s^{-1} to pick individual small spikes and compute the SNR for each kernel at each pixel. We keep the best SNR at each pixel, and we select local peaks at SNR > 3, with $T > 0.02$ (corresponding to $\tau_\alpha \lesssim 4$) as the positions of our transmission spikes. The transmission threshold ($T > 0.02$) was chosen to balance recovery of the small transmission spikes

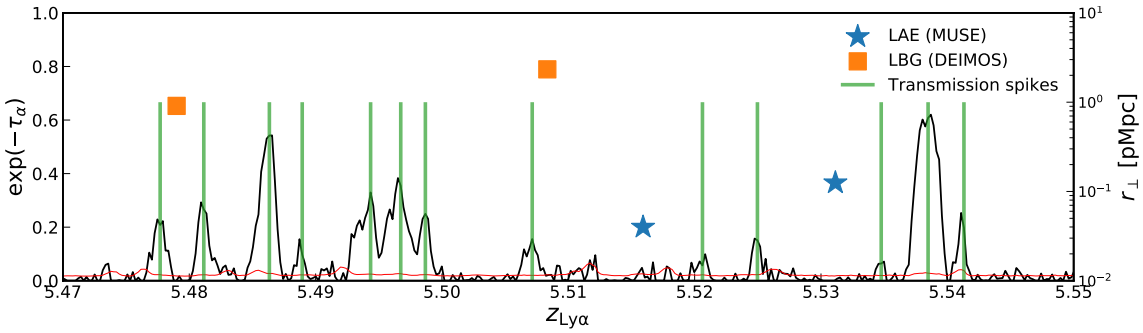

Figure 4.8: Zoom on the Lyman-α forest of J1030 (Figure 4.6). LAEs and LBGs are indicated with blue stars and orange squares, respectively, and the location of transmission spikes identified with a Gaussian-matched filter with green vertical lines.

in J1148 and J2032 whilst limiting spurious detections in sightlines with worse SNR (J1526, J2100), and the SNR threshold as a compromise between purity of the spike sample and large enough numbers to have reasonable bootstrap error estimates. We present an example of the successful recovery of transmission spikes in the high-redshift Lyman-α forest in Figure 4.8.

We then estimate the 2PCCF with the estimator

$$\xi^{\rm 2PCCF}_{\rm Gal-Ly\alpha} = \frac{D_G D_\perp(r)}{R_G D_\perp(r)} - 1 \ , \tag{4.5}$$

where $D_G D_\perp$ is the number of transmission spikes-galaxy pairs normalised by the number of pixels in each radial bin and $R_G D_\perp$ is the normalised number of transmission spikes - random galaxies pairs, and r the 3D comoving distance. As for the transmission cross-correlation, the redshift of random galaxies are sampled from the LAE or UV LF for LAEs and LBGs, respectively, the angular separation drawn from adequate uniform distributions, and the errors are computed by boostrapping the sample of detected galaxies.

We show in Figure 4.9 the 2PCCFs for both LAEs and LBGs. We detect a positive signal at 3.2σ as an excess of transmission spikes on $r \sim 10 - 60$ cMpc scales around LAEs and a deficit of transmission spikes at $r \lesssim 10$ cMpc. We also find an excess (1.9σ) of transmission spikes on large scales around LBGs. The significance of the LAE(LBG) 2PCCF excess is decreased by $-1.5\sigma(-0.2\sigma)$ if the redshift correction is not applied (Section 4.2.4). We compare in Figure B.12 the 2PCCFs with and

without correction, with the excesses being reduced in the latter case. The absence of correlation (or even an anti-correlation) on the smaller scales stems both from increased absorption by dense gas around the central LAE (which we model in Section 4.5.2) and the redshift error which dampens the signal (~ 200 km s^{-1} corresponding to ~ 1.8 cMpc at $z \sim 6$). The reduced significance of the LBG 2PCCF could stem from an inappropriate normalisation due to the difficulty of creating randomly samples of LBGs. As detailed previously, we have conservatively decided to scale the number of random galaxies to that of the observed ones. However if some of the fields *are* indeed slightly overdense, we would be overestimating the normalisation in the cross-correlation and thus decrease the significance of the signal. The difference in the strength of the signal between the transmission cross-correlation and the 2PCCF can be attributed to the uncertainty in the mean flux at high-redshift. We defer the discussion of this difference to Section 4.7 where we examine the impact of noise on our measurements of the flux transmission and transmission spikes cross-correlation with galaxies.

4.5 Modelling the galaxy-Lyman-α transmission and 2-point cross-correlations

In order to interpret the observed galaxy-Lyman-α forest cross-correlations, we use a radiative transfer model based on the halo occupation distribution (HOD) framework introduced in Kakiichi et al. (2018). Here we summarise the key ingredients and extensions used in this study.

Kakiichi et al. (2018) derived the average H I photoionisation rate at a distance r from a galaxy due to the clustered faint population

$$\begin{aligned}\langle \Gamma^{\rm CL}_{\rm HI}(r|M_h,z)\rangle &= \frac{\bar{\Gamma}_{\rm HI}}{\lambda_{\rm mfp}(z)} \int \frac{e^{-|\boldsymbol{r}-\boldsymbol{r}'|/\lambda_{\rm mfp}(z)}}{4\pi|\boldsymbol{r}-\boldsymbol{r}'|^2} \left[1+\langle \xi_g(|\boldsymbol{r}'|)\rangle_L\right] {\rm d}^3 r', \\ &= \bar{\Gamma}_{\rm HI}\left[1+\int_0^\infty \frac{k^2 {\rm d}k}{2\pi^2} R(k\lambda_{\rm mfp}(z)) \langle P_g(k|M_h,z)\rangle_L \frac{\sin kr}{kr}\right], \end{aligned} \tag{4.6}$$

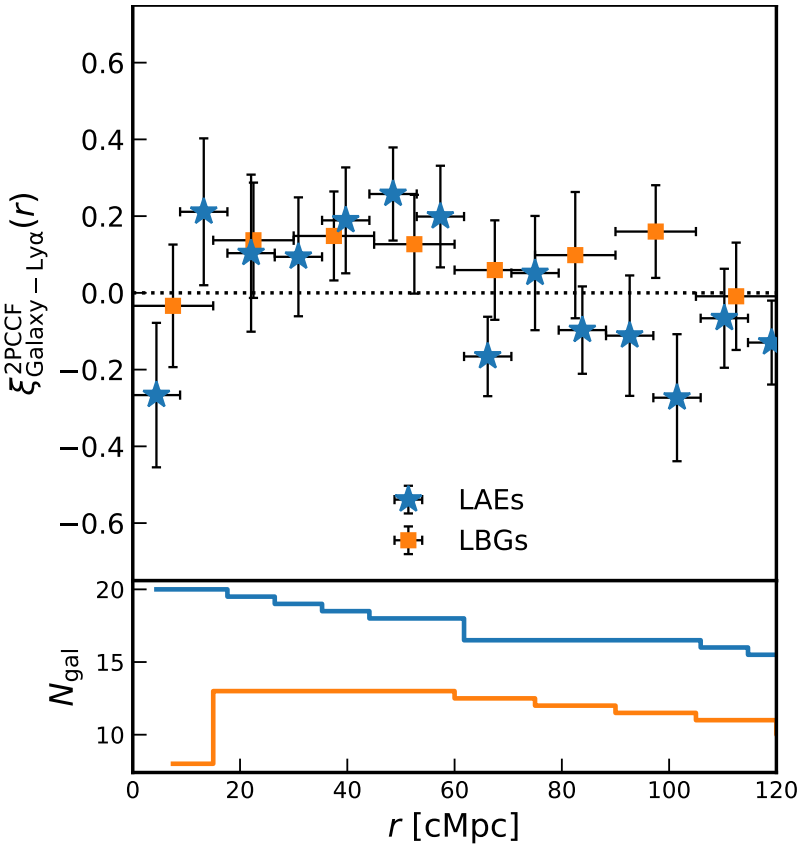

Figure 4.9: **Upper panel:** The two-point cross-correlation function of LBGs (orange squares) and LAEs (blue stars) with selected Lyman-α transmission spikes at $z \sim 6$. The errorbars are bootstrap errors on the number of detected galaxies. We find a significant excess of transmission spikes on scales $10 - 60$ cMpc around LAEs (3.2σ). The excess of transmission spikes around LBGs is significant at 1.9σ, but might extend to larger scales. We point out however that the LBG selection is less complete than the LAEs due to the DEIMOS mask design and that one of the three quasar fields (J0836) has only one detected LBG. On the smaller scales ($\lesssim 1$ pMpc or $\lesssim 7$ cMpc), a deficit of transmission spikes is possibly present. The scales of the excess and the deficit are in good agreement with the measurements of Kakiichi et al. (2018) and Chapter 3. **Lower panel:** Number of galaxies contributing to the 2PCCF in each radial bin. Note that due to the redshift distribution of galaxies and the limits of the Lyman-α forests, at larger distances some galaxies can only be correlated with transmission spikes at lower or higher redshift. In that case, we count these as contributing $N = 0.5$ instead of $N = 1$ to the total number of galaxies.

where $\lambda_{\rm mfp}(z) = 6.0\left(\frac{1+z}{7}\right)^{-5.4}$ [pMpc] (Worseck et al. 2014) is the mean free path of ionising photons and $R(k\lambda_{\rm mfp}) = \arctan(k\lambda_{\rm mfp})/(k\lambda_{\rm mfp})$ is the Fourier transform of the radiative transfer kernel $e^{-r/\lambda_{\rm mfp}}/(4\pi r^2\lambda_{\rm mfp})$. The luminosity-weighted galaxy power spectrum is

$$\langle P_g(k|M_h,z)\rangle_L = \frac{\int_{L_{\rm UV}^{\rm min}}^{\infty} L_{\rm UV}\Phi(L_{\rm UV}|z)P_g(k,L_{\rm UV}|M_h,z){\rm d}L_{\rm UV}}{\int_{L_{\rm UV}^{\rm min}}^{\infty} L_{\rm UV}\Phi(L_{\rm UV}|z){\rm d}L_{\rm UV}}, \tag{4.7}$$

where $P_g(k, L_{\rm UV}|M_h, z)$ is the Fourier transform of the correlation function between bright tracers (i.e. detected LBGs and LAEs) with host-halo mass $> M_h$ and galaxies with luminosity $L_{\rm UV}$. We assume only central galaxies will be detected as LBGs or LAEs and therefore populate each halo with a HOD using a step function, $\langle N|M_h\rangle = 1$ for halo mass $> M_h$ and zero otherwise. Fainter galaxies are populated using the conditional luminosity function pre-constrained by simultaneously fitting the $z \sim 6$ UV luminosity function (Bouwens et al. 2015; Finkelstein et al. 2015; Bowler et al. 2015; Ono et al. 2018) and the galaxy auto-correlation function (Harikane et al. 2016) as in Kakiichi et al. (2018).

4.5.1 From the cross-correlation of galaxies with the transmitted flux to the 2PCCF

As in Kakiichi et al. (2018) and Chapter 3, the enhanced UVB can be used to compute the mean Lyman-α forest transmission at a distance r of galaxy,

$$\langle \exp(-\tau_\alpha)(r|M_h,z)\rangle = \int \exp\left[-\tau_\alpha\left(\Delta_b, \langle\Gamma_{\rm HI}^{\rm CL}(r|M_h,z)\rangle\right)\right] \times P_V(\Delta_b|r,M_h){\rm d}\Delta_b \ , \tag{4.8}$$

where $P_V(\Delta_b|r, M_h)$ is the volume-averaged conditional PDF of the baryon overdensities Δ_b at a distance r from our galaxy tracer with a halo of mass M_h at redshift z, and $\langle\Gamma_{\rm HI}^{\rm CL}(r)\rangle$ is the clustering-enhanced photoionisation rate modelled previously. The optical depth τ_α is derived using the fluctuating Gunn-Peterson approximation

(see, e.g. Becker et al. 2015a for a review),

$$\tau_\alpha \simeq 11 \Delta_b^{2-0.72(\gamma-1)} \left(\frac{\Gamma_{\rm HI}}{10^{-12}\ {\rm s}^{-1}}\right)^{-1} \left(\frac{T_0}{10^4\ {\rm K}}\right)^{-0.72} \left(\frac{1+z}{7}\right)^{9/2} , \tag{4.9}$$

where Δ_b is the baryon overdensity and T_0 is the temperature of the IGM at mean density. We include thermal fluctuations of the IGM using the standard power-law scaling relation (Hui & Gnedin 1997; McQuinn & Upton Sanderbeck 2016),

$$T(\Delta_b) = T_0 \Delta_b^{\gamma-1}, \tag{4.10}$$

assuming the fiducial values $T_0 = 10^4$ K and $\gamma = 1.3$.

We now expand this framework to predict a new statistic: the probability of seeing a transmission spike in the Lyman-α forest. Given a transmission threshold over which a detection is considered secure, we can derive an equivalent optical depth threshold. We fix the transmission threshold at $\exp(-\tau_\alpha) \gtrsim 0.02$, corresponding to $\tau_\alpha^{\rm th} \simeq 4$, to match our measurement of the 2PCCF. By substituting the predicted clustering-enhanced photoionisation rate and the threshold optical depth in Eq. 4.9, we obtain the maximum baryon underdensity $\Delta_b^{\rm max}$ required to produce a detected transmission spike in the Lyman-α forest at a distance r of a tracer galaxy,

$$\Delta_b \leq \Delta_b^{\rm max}(\Gamma_{\rm HI}) \simeq 0.57 \left(\frac{\tau_\alpha^{\rm th}}{4}\right)^{0.56} \left(\frac{\Gamma_{\rm HI}}{10^{-12}\ {\rm s}^{-1}}\right)^{0.56} \left(\frac{T_0}{10^4\ {\rm K}}\right)^{0.4} \left(\frac{1+z}{7}\right)^{-2.52} . \tag{4.11}$$

Thus the occurrence probability of Lyman-α transmission spike at a location with H I photoionisation rate $\Gamma_{\rm HI}$ is given by the probability to reach such an underdensity:

$$P[< \Delta_b^{\rm max}(\Gamma_{\rm HI})\,|r, M_h] = \int_0^{\Delta_b^{\rm max}(\Gamma_{\rm HI})} P_V(\Delta_b|r, M_h){\rm d}\Delta_b \ . \tag{4.12}$$

The cross-correlation between galaxies and the Lyman-α transmission spikes can therefore be modelled as the excess occurrence probability, $P[< \Delta_b^{\rm max}(\langle\Gamma_{\rm HI}^{\rm CL}(r)\rangle)|r, M_h]$, of transmission spikes around an object with host halo mass M_h and an enhanced

photoionisation rate $\langle\Gamma_{\rm HI}^{\rm CL}\rangle$ relative to one at mean photoionisation rate $\bar{\Gamma}_{\rm HI}$ and average density fluctuations, i.e. $P[<\Delta_b^{\rm max}(\bar{\Gamma}_{\rm HI})|r\rightarrow\infty, M_h]$. It is then straightforward to deduce the cross-correlation between galaxies and the transmitted Lyα spikes as

$$\xi^{\rm 2PCCF}_{\rm Gal-Ly\alpha}(r) = \frac{P[<\Delta_b^{\rm max}(\langle\Gamma_{\rm HI}^{\rm CL}(r)\rangle)|r, M_h]}{P[<\Delta_b^{\rm max}(\bar{\Gamma}_{\rm HI})|r\rightarrow\infty, M_h]} - 1 \ , \tag{4.13}$$

The advantage of such a statistic over the transmission cross-correlation is that given the large number of pixels in high-resolution spectra of high-redshift quasars, a very low probability of transmission spikes can still be measured with acceptable significance, whereas often only an upper limit on the mean flux can be measured at $z \gtrsim 6$.

4.5.2 Extending our UVB model with varying mean free path and gas overdensities

We now proceed to extend the model of UVB enhancement due to galaxy clustering by adding a varying mean free path and taking into account the gas overdensities associated with LAEs and LBGs on scales of several cMpc.

We first consider the effect of gas overdensity using the relevant probability distribution function. We derive the conditional PDF of overdensities around suitable haloes $P_V(\Delta_b|r, M_h)$ from the IllustrisTNG simulations (Nelson et al. 2018). We utilise the TNG100-2 simulation for host halo masses $10^{10.5}\,{\rm M}_\odot < {\rm M_h} < 10^{11.7}\,{\rm M}_\odot$ whereas for larger host halo masses ($M_h > 10^{11.7}\,{\rm M}_\odot$) we use TNG300-3 in order to get higher number of such halos at the cost of larger gas and dark matter particle masses. We present in Appendix B.5 the extracted conditional PDFs for a range of halo masses and radii.

Following Miralda-Escude et al. (2000); Pawlik et al. (2009) we then fit each

conditional PDF with a parameterisation of the form

$$P_V(\Delta_b|r, M_h)\mathrm{d}\Delta_b = A(r, M_h) \exp\left[-\frac{(\Delta_b^{-2/3} - C_0(r, M_h))^2}{2(2\delta_0(r, M_h)/3)^2}\right] \Delta_b^{-\beta(r,M_h)} \mathrm{d}\Delta_b \ , \tag{4.14}$$

with the parameter $A(r, M_h)$ being determined by requiring that the integral of the PDF is unity ($\int P_V(\Delta_b|r, M_h)\mathrm{d}\Delta_b = 1$). The fitted values of A, C_0, δ_0 and β are listed in Table B.3 for a relevant choice of (r, M_h). We show in Figure 4.11 the good agreement of the fits with the simulated PDF in a snapshot at $z = 5.85$ and a chosen central halo mass $M_h \sim 10^{11.2\pm0.1}\,\mathrm{M}_\odot$ corresponding to the one derived from the clustering of LAEs (Ouchi et al. 2018).

Kakiichi et al. (2018) considered a constant mean free path for simplicity. Introducing a full self-consistency of the mean free path down to ckpc scales in Eq. 4.6 is the realm of numerical simulations if a real distribution of gas and discrete sources is to be considered (and not the average distribution we use here). However we can approximate variations of the mean free path to first order. Following Miralda-Escude et al. (2000); McQuinn et al. (2011); Davies & Furlanetto (2016); Chardin et al. (2017), the mean free path of ionising photons is dependent on the photoionisation rate and the mean baryon density,

$$\lambda_{\rm mfp}(r) = \lambda_0 \left(\frac{\langle\Gamma_{\rm HI}^{\rm CL}(r)\rangle}{\bar{\Gamma}_{\rm HI}}\right)^{\beta_{\rm mfp}} \left[\int \Delta_b P_V(\Delta_b|r, M_h)\mathrm{d}\Delta_b\right]^{-\gamma_{\rm mfp}} \ , \tag{4.15}$$

where $\beta_{\rm mfp}$ and $\gamma_{\rm mfp}$ reflects a simple parameterisation of the mean free path dependence on the local UVB and gas overdensity. In this work, we chose to use the values $\beta_{\rm mfp} = 2/3$, $\gamma_{\rm mfp} = -1$ for our fiducial model following previous works cited above.

Given the mutual dependence between $\langle\Gamma_{\rm HI}^{\rm CL}(r)\rangle$ and $\lambda_{\rm mfp}(r)$, we iterate the computation until $\langle\Gamma_{\rm HI}^{\rm CL}(r)\rangle$ is converged at the 1 % level at every distance r. As expected, a varying mean free path does not affect the photoionisation rate on large scales but decreases it by a factor 2-3 on scales $\lesssim 1$ cMpc. We show the impact on the pre-

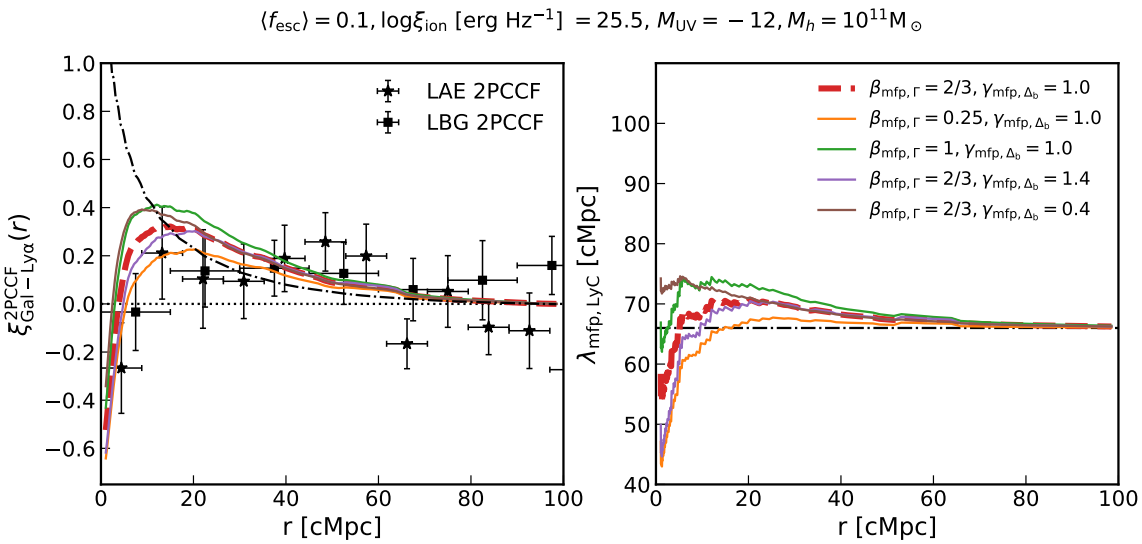

Figure 4.10: The impact of a spatially varying mean free path on the modelled 2PCCF of galaxies with transmission spikes in the Lyman-α forest. Variation of either the mean free path power-law dependence on the photoionisation rate ($\beta_{\rm mfp}$) or the gas overdensity ($\gamma_{\rm mfp}$) do not affect significantly the predicted 2PCCF. The models are generated with the fiducial parameters $\langle f_{\rm esc}\rangle = 0.1$, $M_h = 10^{11}\,{\rm M}_\odot$, $\log\xi_{\rm ion}/[{\rm erg}^{-1}{\rm Hz}] = 25.5$, and $\gamma_{\rm mfp} = 1.3$. The black dashed-dotted line in the left panel show a model with a fixed mean free path.

dicted 2PCCF in Figure 4.10. We find that any reasonable choice of $(\beta_{\rm mfp}, \gamma_{\rm mfp})$ modifies the 2PCCF only by a factor < 2 on scales $r < 10$ cMpc.

4.5.3 Peculiar velocities and the observed 2PCCF

We have so far only considered the cross-correlation in real space. However, the observed two-point correlation is distorted by peculiar velocities and infall velocities. We consider here only the impact of random velocities and redshift errors. Following Hawkins et al. (2003); Bielby et al. (2016), the real-space 2D correlation $\xi'(\sigma,\pi)$ is convolved with a distribution of peculiar velocities along the line of sight direction (π),

$$\xi(\sigma,\pi) = \int_{-\infty}^{+\infty} \xi'(\sigma, \pi - v/H(z)) f(v) {\rm d}v \ , \tag{4.16}$$

with an Gaussian kernel for the velocity distributions $f(v) = (2\pi\sigma_v^2)^{-1} \exp\left(-\frac{v^2}{2\sigma_v^2}\right)$. We use $\sigma_v = 200$ km s^{-1}, which is the observed scatter in the difference between Lyman-α and systemic redshifts at $z \sim 2-3$ (Steidel et al. 2010), encapsulating both

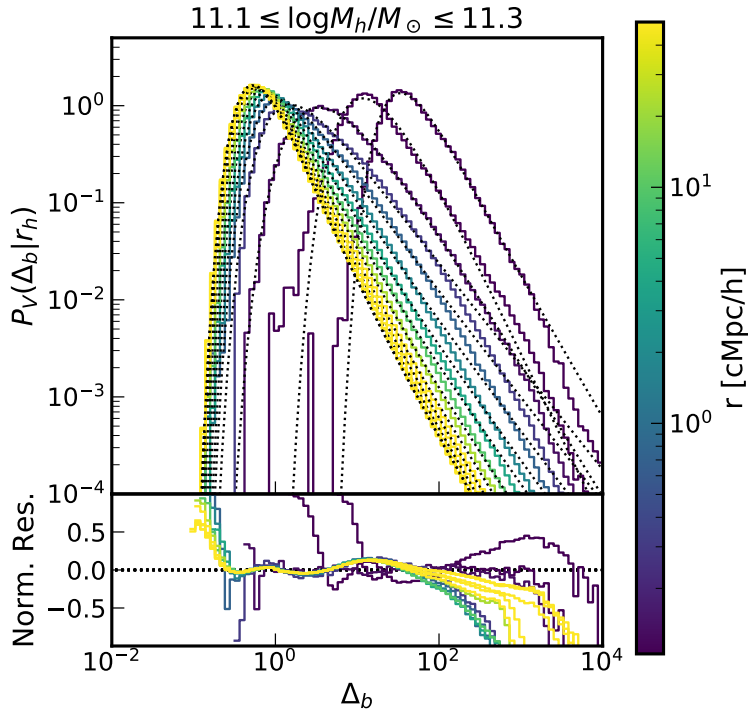

Figure 4.11: **Upper panel:** A subsample of the conditional PDFs from $r = 10^{-1}$ cMpc/h to $r = 10^{1.5}$ cMpc/h in increments of 0.045 dex. The fits with Eq. 4.14 are overlaid (dotted black) on top of the PDF extracted from the IllustrisTNG 100-2 simulation box in a snapshot at $z \sim 5.85$ (coloured histograms). **Lower panel:** Normalised residuals of the PDF fit, coloured by distance from the centre of the halo, showing good agreement in the validity limits of the prescribed analytical form between $10^{-1} \leq \Delta_b \leq 10^2$.

redshift errors and the random velocities of galaxies. We finally take the monopole of the 2D correlation function,

$$\xi_0(s) = \frac{1}{2} \int_{-1}^{-1} \xi(\sigma, \pi) P_0(\mu) \mathrm{d}\mu \ , \tag{4.17}$$

where $s = \sqrt{\sigma^2 + \pi^2}$, $\mu = \pi/s$, and $P_0(\mu) = 1$ is the zeroth order Legendre polynomial. As shown in Figure 4.12, the peculiar velocities reduce slightly the signal on small scales.

We show in Figure 4.12 various realisations of our model of the 2PCCF. We present here the impact of the modelling improvements that we described previously. The addition of gas overdensities decreases the correlation on the smallest scales ($r \lesssim 20\,$cMpc). The varying mean free path has little impact on the final shape of the predicted two-point correlation function, but boosts it slightly at $r > 20$ cMpc. Finally, the redshift errors and random velocities have a negligible impact on scales

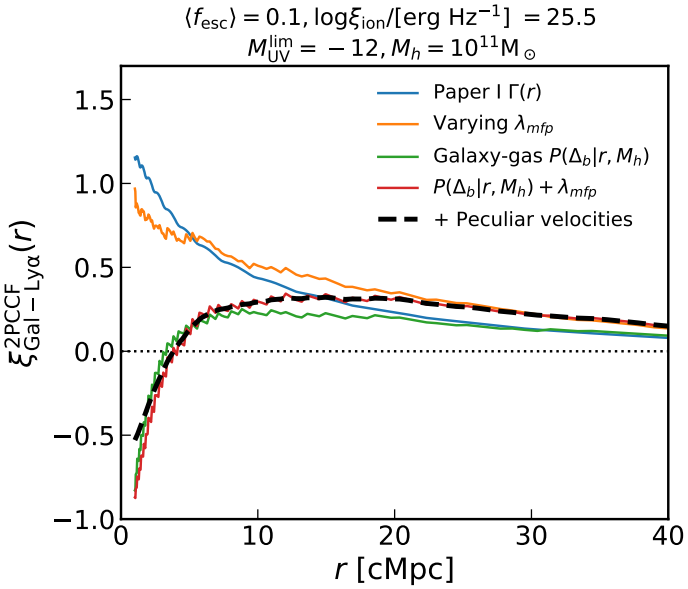

Figure 4.12: Successive versions of our model showcasing the increasing levels of refinement detailed throughout Section 4.5. The resulting predicted 2PCCF is mainly sensitive to the original UVB computed as in Kakiichi et al. (2018), but is dampened on small scales by the addition of a realistic gas overdensity PDF. The implementation of a variable mean free path enhances the signal on large scales.

larger than few cMpc.

We conclude this modelling section by comparing the 2PCCF to the data for various fiducial parameters of the limiting luminosity of contributors to reionisation $M_{\rm UV}^{\rm lim}$, escape fraction of the LyC photons $\langle f_{\rm esc}\rangle$ and host halo mass M_h of the detected bright galaxy in Figure 4.13. We adopt a fiducial $\log \xi_{\rm ion}/[\rm erg^{-1}Hz] = 25.5$, $\beta_{\rm mfp} = 2/3$, $\gamma_{\rm mfp} = -1$. Increasing the minimum UV luminosity increases the correlation as more sources contribute to the local photoionisation rate. We find that the host halo mass of the tracer galaxy is correlated positively with the 2PCCF signal strength, as they cluster more strongly with other galaxies. Finally, the escape fraction has a non-trivial effect on the cross-correlation: because it affects both the local and the overall UVB, an increase in the escape fraction decreases the total 2PCCF. Indeed, excesses of ionised gas close to clustered galaxies become harder to detect as the Universe becomes fully ionised and transmission spikes are ubiquitous. We defer to Appendix B.6 for a full mathematical derivation of the role of $f_{\rm esc}$ in our modelled 2PCCF.

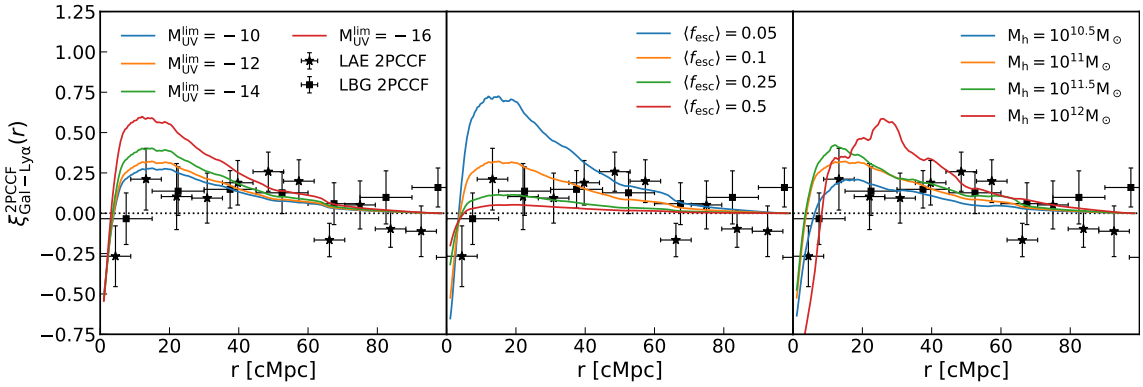

Figure 4.13: Examples of our model of the 2PCCF given a range of parameters (limiting luminosity of contributors to reionisation $M_{\rm UV}^{\rm lim}$, escape fraction of LyC photons $f_{\rm esc}$ and host halo mass of the detected LAE/LBG M_h). In each panel, one parameter is varying while the others are kept fixed at the fiducial values of $M_{\rm UV}^{\rm lim} = -12$, $\langle f_{\rm esc} \rangle = 0.1$, $M_h = 10^{11} {\rm M}_\odot$. All models plotted here assume a redshift of $z = 5.85$ and ionising efficiency $\log \xi_{\rm ion}/[{\rm erg}^{-1}{\rm Hz}] = 25.5$, and a mean free path dependence on the mean overdensity with exponent $\gamma_{\rm mfp} = -1$.

4.6 Constraints on the ionising capabilities of $z \sim 6$ contributors clustered around LAEs and LBGs

Our model of the statistical proximity effect of galaxies based on their correlation with Lyman-α transmission spikes can be applied at different redshifts, across absorbed and transparent sightlines, and to galaxy populations with different halo masses. We have detected a signal in the 2PCCF of high-redshift LAEs and LBGs with Lyman-α transmission spikes, which we will now proceed to fit.

The median redshift of our LAE (LBG) sample is $\langle z \rangle = 5.82(5.597)$. We therefore use the gas overdensity PDF from the Illustris TGN100-2 (TNG300-3 for LBGs) at $z = 5.85$ (the closest snapshot in redshift to the larger LAE sample), and we fix the redshift at the same value for the computation of the CLF and our 2PCCF model in general for consistency. We use the fiducial values of $\beta_{\rm mfp} = 2/3, \gamma_{\rm mfp} = -1$ for the mean free path of ionising photons, and a temperature density relation $T \propto \Delta_b^{\gamma-1}$ with $\gamma = 1.3$.

Our model constrains the number of ionising photons emitted around galaxies,

and thus the *luminosity-weighted-average contribution*[4] of sources to reionisation

$$\langle f_{\rm esc}\xi_{\rm ion}\rangle_L = \frac{\int_{M_{\rm UV}^{\rm lim}}^{\infty} f_{\rm esc}(L_{\rm UV})\xi_{\rm ion}(L_{\rm UV})L_{\rm UV}\Phi(L_{\rm UV})\mathrm{dL_{UV}}}{\int_{M_{\rm UV}^{\rm lim}}^{\infty} L_{\rm UV}\Phi(L_{\rm UV})\mathrm{dL_{UV}}} \ , \tag{4.18}$$

which for simplicity we have recast with a fixed $\log \xi_{\rm ion}/[\mathrm{erg^{-1}Hz}] = 25.5$, such that our main results will constrain the *luminosity-averaged escape fraction*. We emphasize that the limiting luminosity of contributing sources simply marks the truncation of the UV LF. A Gaussian prior on the turnover of the UV LF at $M_{\rm UV}^{\rm lim} \sim -12\pm1$ encompasses the scatter between different studies (e.g. Livermore et al. 2017; Bouwens et al. 2015; Atek et al. 2018) and the recent constraint via the extragalactic background light measurement (Fermi-LAT Collaboration 2018).

We fit the LAE 2PCCF signal with the EMCEE Monte Carlo sampler (Foreman-Mackey et al. 2013) using a flat prior in the range $0 \leq \langle f_{\rm esc}\rangle \leq 1$, a Gaussian prior over $M_{\rm UV}^{\rm lim} \simeq -12$ with $\sigma_{M_{\rm UV}} = 1$, and another Gaussian prior for the host halo masses based on the angular clustering measurements of LAEs (Ouchi et al. 2018). We use the values of $\log M_h^{\rm LAE}/[\mathrm{M_\odot}] = 11.1^{+0.2}_{-0.4}$ derived at $z = 5.7$ for all our LAE detections at $5.5 < z < 6.2$. We marginalise over LAE host mass and minimum luminosity priors to get our final constraint from the LAE-spike 2PCCF

$$\langle f_{\rm esc}\rangle_{M_{\rm UV}\lesssim-12} = 0.14^{+0.28}_{-0.05} \qquad (\log \xi_{\rm ion}/[\mathrm{erg^{-1}Hz}] = 25.5) \ , \tag{4.19}$$

where the errors represent a 1σ credible interval. The LBG-spike 2PCCF, where the host halo mass prior of LBGs at $z \sim 6$ ($M_h^{\rm LBG}/[\mathrm{M_\odot}] = 12.02^{+0.02}_{-0.01}$) is based on the clustering measurement with Hyper Suprime-Cam (HSC) at the Subaru telescope (Harikane et al. 2018), gives the following constraint

$$\langle f_{\rm esc}\rangle_{M_{\rm UV}\lesssim-12} = 0.23^{+0.46}_{-0.12} \qquad (\log \xi_{\rm ion}/[\mathrm{erg^{-1}Hz}] = 25.5) \ . \tag{4.20}$$

These average constraints on the entire luminosity range can be of course re-arranged to test any given functional form of the escape fraction and/or the ion-

[4] Which we shorten to *luminosity-averaged* for convenience.

ising efficiencies, and accommodate other fiducial values of $\xi_{\rm ion}$. For example, we present in Figure 4.14 the average escape fraction of galaxies as a function of the minimum UV luminosity of contributors between $-20 < M_{\rm UV}^{\rm min} < -10$. Our results are in good agreement with literature estimates derived from neutral fraction histories (e.g. Robertson et al. 2015; Naidu et al. 2020), especially for models invoking a substantive contribution of faint galaxies to reionisation. Both LAE-IGM and LBG-IGM 2PCCF constraints are in agreement with the C IV-IGM transmission cross-correlation of Chapter 3. Although the three measurements' maximum likelihood value differ, the uncertainties are still too large to conclude yet on any significant tension between the escape fraction of the galaxies traced by LAEs, LBGs and C IV absorbers.

4.7 Discussion

4.7.1 Relative contribution of sub-luminous sources

As the cross-correlation slope is sensitive to the minimum UV luminosity of contributing sources (Figure 4.13), it is theoretically possible to measure simultaneously the luminosity-averaged escape fraction of reionisation sources and their minimum or maximum luminosity. We now proceed to extend our analysis to test whether we can infer the relative contribution of bright and faint sources to reionisation. We examine two simple cases: a model where all galaxies fainter than a certain UV luminosity solely contribute to reionisation and, conversely, a model where such faint galaxies do not contribute at all. To do so, we treat the minimum/maximum UV luminosity as a parameter and fit the model with a flat prior on this quantity. We then fit these two models to the LAE/LBG-IGM 2PCCF.

We present the posterior distribution of our parameters in Figure 4.15 (left) for the model where bright galaxies dominate, and the inferred constraints in Table 4.3. The LAE/LBG 2PCCF were fitted with the parameters described above except for a flat prior on the minimum UV luminosity of contributing sources, $-10 < M_{\rm UV}^{\rm min} < -30$. In both cases, the minimum UV luminosity of the contribut-

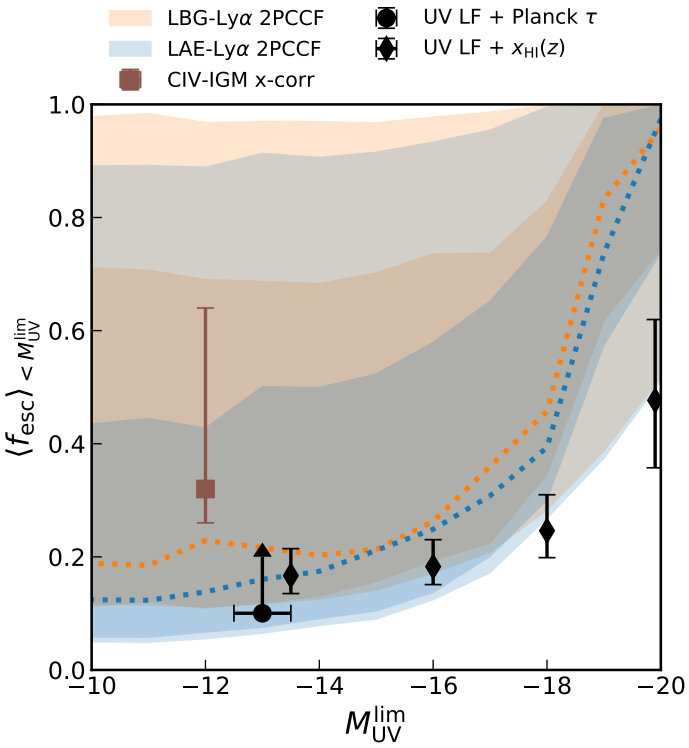

Figure 4.14: Luminosity-averaged escape fraction of galaxies contributing to reionisation as a function of their minimum UV luminosity. The two measurements of the 2PCCF from this work (LAE/LBG $1\sigma, 2\sigma$ blue/orange contours) are in agreement with the C IV-IGM transmission cross-correlation from Chapter 3 (brown squares). The irregular shape of the LBG 2PCCF posterior is due to the gas overdensities PDF which is poorly constrained by Illustris TNG due to the large mass and limited volume, and thus varies quickly from one mass bin to another, making the convergence of the MCMC chain difficult. The LAE-IGM 2PCCF is in better agreement with the average escape fraction derived from the UV LF and the Planck optical depth measurement (Robertson et al. 2015 black circle) or the neutral fraction history when the minimum UV luminosity of contributors is small (Naidu et al. 2020 black diamonds). The escape fractions are (re-)derived assuming $\log \xi_{ion}/[\mathrm{erg}^{-1}\mathrm{Hz}] = 25.5$.

ing sources is $M_{UV}^{min} < -20.0\,(2\sigma)$. In practice however, a model where *only* galaxies brighter than $M_{UV} = -20.0$ is implausible because it would require an overhelmingly high luminosity-averaged escape fraction of ≈ 1, contradicting existing $z \sim 6$ measurements (Matthee et al. 2018) and marking a stark departure from measurements at lower-redshift (e.g. Vanzella et al. 2016; Izotov et al. 2016b, 2018; Tanvir et al. 2019; Fletcher et al. 2019). It thus more likely that, *if only the brightest objects contribute*, they include at least relatively faint galaxies down to $M_{UV} \sim -18(-16)$.

We now examine our model where faint galaxies dominate. The LAE/LBG 2PCCF were fitted with the parameters described in Section 4.6 except for a flat prior

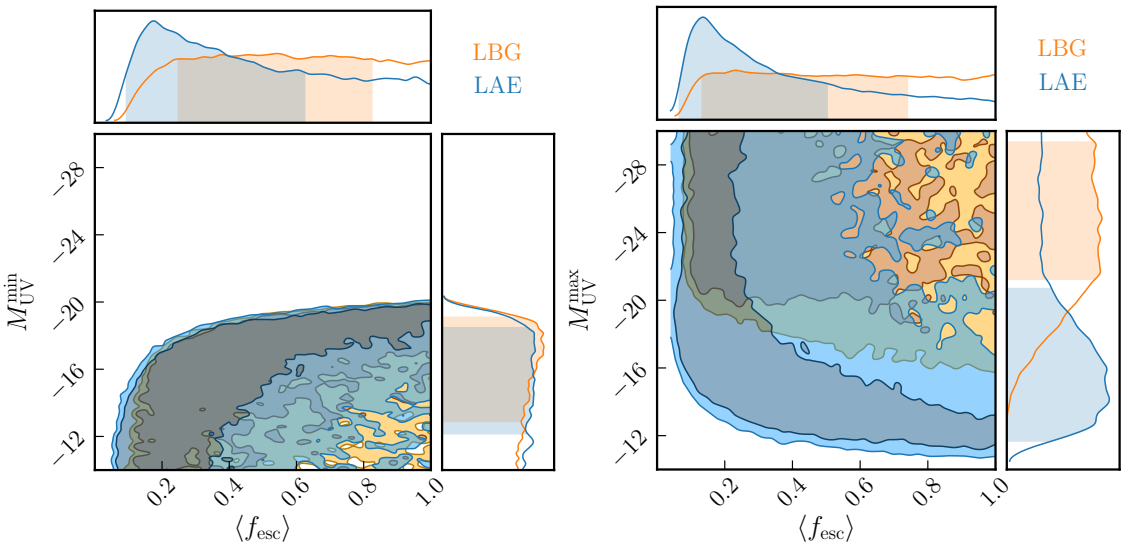

Figure 4.15: **Left:** Posterior distributions of our fiducial model parameters with reionisation dominated by more luminous galaxies fitted to the LAE/LBG-transmission spikes 2PCCF. The 2PCCF constrain the minimum UV luminosity of contributors to be at least $M_{\rm UV} \lesssim -20$. **Right:** Posterior distributions of our fiducial model parameters where faint galaxies dominate, fitted to the LAE/LBG-transmission spikes 2PCCF. The LAE and LBG constraints are in tension, with the LAE 2PCCF favouring a model where only low luminosity galaxies contribute ($-10 \gtrsim M_{\rm UV} \gtrsim -17$) and the LBG 2PCCF requiring the contribution of more luminous objects up to at least $M_{\rm UV} \lesssim -21$. Our fiducial models have the following fiducial parameters $\gamma = 1.3, \beta_{\rm mfp} = 2/3, \gamma_{\rm mfp} = -1$ and $\log \xi_{\rm ion}/[{\rm erg}^{-1}{\rm Hz}] = 25.5$.

on the maximum UV luminosity of contributing sources $-10 < M_{\rm UV}^{\rm max} < -30$, and the minimum UV luminosity of LyC contributing sources was fixed at $M_{\rm UV}^{\rm min} = -10$. We present the posterior distribution of our parameters in Figure 4.15 (right), and the inferred constraints in Table 4.3. The posteriors for the LAE and LBG 2PCCF are strikingly different: whereas the LAE signal is well fitted by a model where faint galaxies ($-17 \lesssim M_{UV} \lesssim -10$) drive reionisation, the LBG 2PCCF constrains the maximum luminosity of contributors to be at least $< -18.4\,(2\sigma)$. In other words, the 2PCCF signal around more luminous tracers (LBGs) of galaxies is consistent with a contribution of brighter objects, whereas faint tracers (LAEs) favour an ionising environment dominated by faint sources. Because clustering is already included in our model, this is not simply a consequence of LAEs likely sitting in smaller overdensities than LBGs, therefore tracing less massive and fainter objects. This result rather indicates that bright objects ($M \lesssim -20$) traced by LBGs have increased

Tracer	$\langle f_{\rm esc} \rangle$	$M_{\rm UV}^{\rm min}$	$M_{\rm UV}^{\rm max}$
LAE	$0.14^{+0.28}_{-0.05}$	$-12.1^{+1.1}_{-0.9}$	-30
	$0.18^{+0.44}_{-0.09}$	$> -19.0\,(2\sigma)$	-30
	$0.14^{+0.36}_{-0.07}$	-10	$-14.1^{+2.4}_{-6.6}$
LBG	$0.23^{+0.46}_{-0.12}$	$-12.0^{+0.9}_{-1.0}$	-30
	$> 0.17\,(2\sigma)$	$> -19.1\,(2\sigma)$	-30
	$> 0.14\,(2\sigma)$	-10	$< -17.3\,(2\sigma)$

Table 4.3: Summary of our constraints on the luminosity-weighted average escape fraction of galaxies at $z \sim 6$. For each galaxy overdensity tracer (LAEs or LBGs), we fit the galaxy-Lyman-α transmission spike 2PCCF for three different scenarios: a turnover of the luminosity function at $M_{\rm UV} = -12$ obtained by imposing a Gaussian prior (Section 4.6), reionisation dominated by luminous galaxies (Figure 4.15, left), and the reverse scenario where only low luminosity galaxies contribute (Figure 4.15, right). Our models assume the following IGM parameters: $\gamma = 1.3, \beta_{\rm mfp} = 2/3, \gamma_{\rm mfp} = -1$ and $\log \xi_{\rm ion}/[{\rm erg^{-1}Hz}] = 25.5$.

ionising efficiencies. One natural explanation is that they would create early ionised bubbles which would in turn enhance the confirmation rate *with a Lyman-α emission line detection* of such LBG candidates. This results is in agreement with the study by Mason et al. (2018b) which found that the boosted transmission around bright ($M_{\rm UV} < -22$) objects cannot only be explained by their biased environment, and thus they must have increased ionising efficiencies.

As a conclusion, it is interesting to note that the 2PCCFs can be fitted with two mutually exclusive populations of galaxies: the sources or reionisation can either be galaxies fainter or brighter than $M_{\rm UV} \sim -18$. These two results show how the 2PCCF is able to constrain the parameters of a *given specific model of escape fraction dependence on luminosity*. However, identifying which model is correct is difficult with the current data as the likelihood ratio of the two best (LAE 2PCCF) fits favours only very marginally (1.3σ) the faint-dominated scenario. Future measurements of galaxy-Lyman-α forest cross-correlations are required to distinguish between the two scenarios, as well the possible differences in the sub-populations traced by LAEs, LBGs and other potential overdensity tracers.

4.7.2 Comparison to low-redshift measurements of the escape fraction

The constraints on the escape fraction of high-redshift galaxies are in agreement with the predictions of analytical models of the neutral fraction evolution during reionisation (Section 1.3). By comparing to low-redshift studies of $f_{\rm esc,LyC}$, this Chapter average escape fraction measurements at $z \sim 6$ can give us some insight into galaxy evolution.

I have summarised escape fraction measurements in individual objects or extreme nebular line emitters in Section 1.4.5. However, they are difficult to compare to the average measurement of the whole galaxy population at high-redshift presented in this Chapter. Steidel et al. (2018) recently reported the average LyC escape fraction of a large sample of $z \sim 2-3$ galaxies selected with the Lyman-Break technique. They find an average escape fraction $f_{\rm esc,LyC} = 0.09 \pm 0.01$ in their sample of 124 galaxies with $-19.5 \leq M_{\rm UV} \leq -22.0$. Although these objects are much brighter than the faintest contributors invoked in some of our models ($M_{\rm UV} < -12$), we can draw two immediate conclusions.

Firstly, Steidel et al. (2018) $f_{\rm esc,LyC}$ value is only in mild (1σ) tension with that derived in this Chapter if all galaxies with $M_{\rm UV} \lesssim -12$ contribute to reionisation (Table 4.3). Therefore, if faint and bright $z \sim 3$ LBGs have a similar escape fraction, the escape fraction of galaxies does not (or only mildly) evolve with redshift, and faint galaxies do contribute significantly to reionisation. Secondly, if faint galaxies are assumed not to contribute significantly to reionisation, then bright $z \sim 3$ LBGs must differ significantly from $z \sim 6$ bright galaxies. Indeed, if only bright objects ($M_{\rm UV} \lesssim -19$) are assumed to contribute, our results show that they must have $f_{\rm esc,LyC} > 14(17)\%$ at the 2σ level, in strong tension with the $z \sim 3$ result. This could be either due to an evolution in the bright population or simply that the Lyman-Break technique selects objects at $z \sim 3$ that are poor analogues to galaxies in the first billion years.

Finally, these two arguments can be also be made for a varying ionising efficiency instead of a varying escape fraction, as the cross-correlation method only constrains

the product of these two quantities. Additionally, if the ionising efficiency at $z \sim 6$ is lower than the assumed value of $\log \xi_{\rm ion}/[{\rm erg}^{-1}{\rm Hz}] = 25.5$, then all escape fractions results should be corrected upwards and would be in stronger tension with the Steidel et al. (2018) results.

4.7.3 Impact of noise on the detection of transmission spikes and the non-detection of a transmission cross-correlation

We now investigate whether or not we can explain the apparently contradictory absence of a transmission cross-correlation but the detection of the transmission spike two-point correlation (Figure 4.7 and 4.9).

In order to do so, we use our improved model of the galaxy-IGM cross-correlation including the varying mean free path and the gas overdensities. We sample $P_V(\Delta_b|r, M_h)$ to generate 1000 values of transmission $\exp(-\tau_\alpha)$ for each distance r from the tracer LAE. We then sample the distribution of errors σ as measured in the Lyman-α forest pixels used in the cross-correlation measurement (i.e. after masking). We then add a flux error drawn from the normal distribution $\Delta_T \sim {\rm N}(0, \sigma)$ to every computed flux value to mimic the effect of noise. Finally, we bin the data to match the measurement binning using the same number of mock Lyman-α forest "pixel" points as the ones measured in the real quasar spectra. The transmission cross-correlation is computed as the mean flux value in each bin, whereas the 2PCCF is the fraction of transmission values above $T > 0.02$ (the same threshold used for the previous measurement).

The resulting mock observations are shown in Figure 4.16 alongside the original model without errors. Clearly, the mean flux or the transmission cross-correlation are difficult to measure with any certainty. This also explains why an increase in the mean transmission or a transmission cross-correlation is much harder to detect than the 2PCCF at $z \sim 6$, as we found in Section 4.4. The addition of noise is crucial because the noise level is comparable to the mean transmission ($T = 0.01 - 0.1$). It is thus no surprise that an increase in the average flux is difficult to measure. The

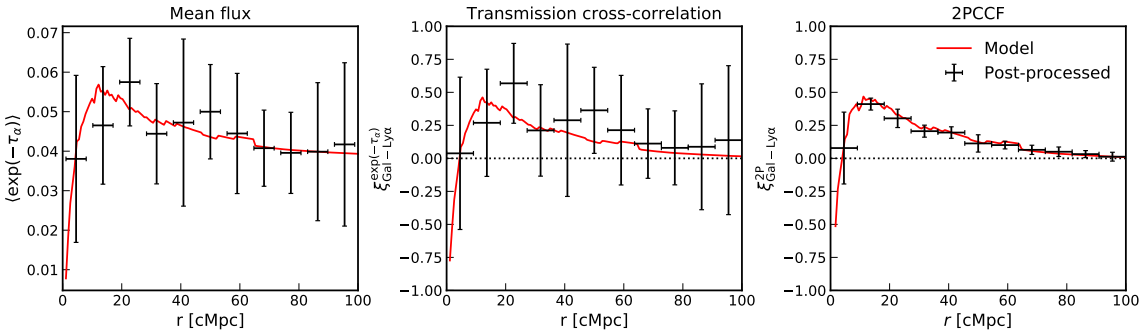

Figure 4.16: Post-processing of different statistics probing the Lyman-α forest - galaxy correlation from our improved model. The addition of noise directly drawn from the spectroscopic data of quasars used in this study highlights the extreme difficulty of measuring the mean flux in the Lyman-α forest around galaxies (left Kakiichi et al. 2018) or the transmission cross-correlation with galaxies/metal absorbers (Chapter 3, centre). The 2PCCF is however robust to such perturbations, hence its use in this study.

2PCCF however is shown to be rather unaffected by the addition of noise as the spikes we consider are at high enough SNR and transmission. Indeed, because the distribution of transmission pixels is log-normal (Bosman et al. 2018), there will be more pixels with intrinsic transmission below any given threshold ($T > 0.02$) than above. As the observational error is drawn from a normal distribution, there will be more pixels observed to have a higher transmission than the given threshold but with lower intrinsic transmission than the reverse, increasing the number of spurious spike detections. In practice, however, this only slightly decreases the 2PCCF and therefore the addition of noise is neglected in our modelling. We conclude that the 2PCCF is less biased by fluctuations of the mean opacity in different sightlines and should be less affected by continuum normalisation uncertainties.

We now conclude by examining whether the observed transmission cross-correlation (Figure 4.7) is consistent with the predicted uncertainty on the modelled signal generated using the best-fit physical parameters of the 2PCCF LAE-transmission spike detection. We find that our LAE transmission cross-correlation measurement is in agreement with the predicted 1σ uncertainty range of the model. (Figure 4.17). There appears to be a slight tension between the (LAE) post-processed model and the LBG measurement, but it is not very significant. The potential tension is more likely due to the smaller number of objects (and quasar sightlines) for the LBG

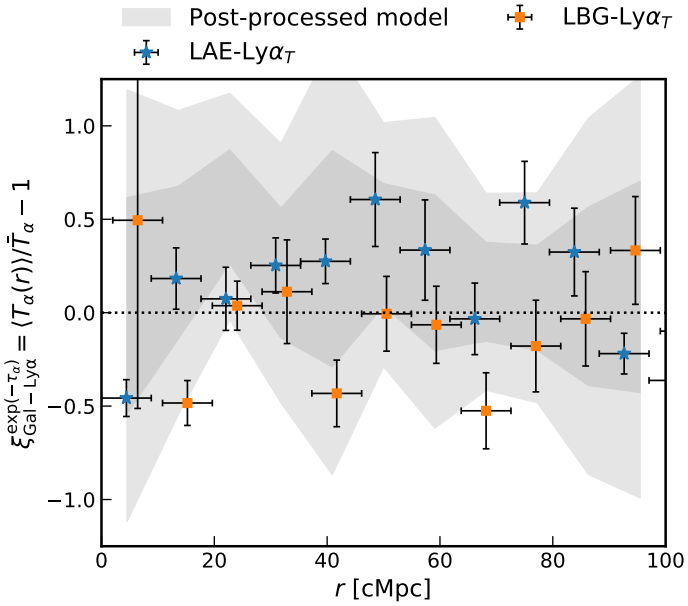

Figure 4.17: Comparison of the observed LAE/LBG-Lyα transmission cross-correlation (blue stars, orange square) with our post-processed (noisy) model (grey shaded areas, $1, 2\sigma$) generated using the best-fit parameters for the LAE-Lyα 2PCCF. The observation are in agreement with the predicted uncertainties stemming from the fact that the SNR of the Lyman-α at $z \sim 6$ is extremely low on average due to the IGM opacity, and that our quasars were primarily chosen to have MUSE coverage rather than deep spectroscopy.

transmission measurement which would lead us to underestimate the errors on the measurement. This is expected as the bootstrap uncertainties are primarily limited by cosmic variance and small sample size, and this measurement might be accurate with a larger sample of quasar sightlines and foreground objects (e.g. Chapter 3).

4.8 Summary

We have assembled a new sample of galaxies in the field of 8 high-redshift quasars in order to examine various correlations between galaxies and the fluctuations in the Lyman-α forest at the end of reionisation. We have extended the approach pioneered in Kakiichi et al. (2018) and Chapter 3 to model the galaxy-Lyman-α flux transmission and the two-point correlation with transmission spikes. We report the following key findings:

- We find a 3.2(1.9)σ-significant excess in the 2PCCF of transmission spikes with LAEs(LBGs) at on scales of $\sim 10 - 60$ cMpc. Our model of the LAE(LBG)

2PCCF constrains sources with $M_{\rm UV} < -12$ to contribute to reionisation with a luminosity-averaged escape fraction $\langle f_{\rm esc} \rangle_{M_{\rm UV}<-12} = 0.14^{+0.28}_{-0.05}(0.23^{+0.46}_{-0.12})$ assuming a fixed $\log \xi_{\rm ion}/[{\rm erg^{-1}Hz}] = 25.5$.

- We present a new model of the two-point cross-correlation function (2PCCF) of detected Lyman-α transmission spikes with LAEs which includes a consistent treatment of gas overdensities around detected LAEs and their peculiar velocities. We find that a spatially varying mean free path does not impact the 2PCCF significantly. We demonstrate that this model is more robust than the transmission cross-correlation at high-redshift.

- We show how parametric models of the escape fraction dependence on the galaxy luminosity can be constrained by the LAE-IGM 2PCCF. We find that the LAE 2PCCF is consistent with a local UVB enhanced either by *faint* galaxies with $M_{\rm UV}^{\rm max} = -14.1^{+2.4}_{-6.6}$ or *brighter* than $M_{\rm UV} < -19.0\,(2\sigma)$. The LBG 2PCCF favours brighter objects with at least $M_{\rm UV} < -19\,(2\sigma)$ contributing to reionisation. Differentiating between these hypotheses will however require a larger dataset of galaxies in quasar fields.

- We find no evidence for a correlation between the transmission in the Lyman-α forest and LAEs/LBGs at $z \sim 6$. We show how this absence of signal is consistent with scatter and noise of our quasar sightlines. Nonetheless, the deficit of transmission on scales up to ~ 10 cMpc is seen in the Lyman-α forest around LAEs as previously reported around C IV absorbers (Chapter 3).

Chapter 5

Evolving and missing quasars at the end of reionisation

The majority of this Chapter has been published in Meyer et al. (2019b MNRAS, 487, 3305). The high-redshift quasar spectra were reduced by S. E. I. Bosman or provided by colleagues as indicated in the text. Section 5.5 is based on the results of accepted ESO proposal (NTT/EFOSC2) ID : 0104.A-0662(A) (PI:Meyer).

5.1 Introduction

In this Chapter I focus on the potential evolution of quasars and the implications this might have for understanding their role in reionisation. Specifically, (i) can we safely assume quasar spectra do not evolve in making quantitative measures of the IGM opacity using the damping wing method and (ii) are some high redshift quasars missed in current surveys and thereby not included in the ionising budget.

The number of known quasars in the early Universe has recently increased sharply, mainly due to the availability of deep wide-field surveys in the optical and the near-infrared. At the time of writing, we count more than 300 quasars at $z > 5.7$. This progress is due to the concerted efforts of different teams selecting

and confirming candidates from diverse surveys such as the Sloan Digital Sky Survey (SDSS, Jiang et al. 2016), Panoramic Survey Telescope and Rapid Response System data (Pan-STARSS, Kaiser et al. 2010; Bañados et al. 2016), the Dark Energy Survey - Visible and Infrared Survey Telescope for Astronomy (VISTA) Hemisphere Survey (DES-VHS, Reed et al. 2015), the VISTA Kilo-Degree Infrared Galaxy Survey (VIKING, Venemans et al. 2013), the VLT ATLAS survey (Carnall et al. 2015), the UKIRT Infrared Deep Sky Survey (UKIDSS, Venemans et al. 2007; Mortlock et al. 2009, 2011), the DESI Legacy imaging Surveys (DELS) (Dey et al. 2019), the Wide-field Infrared Survey (WISE) survey, and, increasingly, from overlaps of some thereof (Bañados et al. 2018; Wang et al. 2019, 2018; Yang et al. 2018; Fan et al. 2019). The recent discovery of a gravitationally lensed quasar may suggest that we are missing further high-redshift quasars (Fan et al. 2019; Pacucci & Loeb 2019), and therefore it is possible their numbers will swell in the coming years.

This surge has enabled a large number of investigations into a variety of topics. Rising numbers of high-redshift quasars offer the possibility to study their central supermassive black holes. The high masses of the high-redshift black holes powering these quasars have proven a particular challenge to explain as they imply either high seed masses and/or continuous accretion at nearly the Eddington rate for most of their young life (e.g. Willott et al. 2010; Bañados et al. 2018). The study of the co-evolution of galaxies and quasars has also benefited from larger samples of younger quasars (e.g. Venemans et al. 2012; Decarli et al. 2017). Determination of the quasar luminosity function at $z \gtrsim 5.5$ has constrained the contribution of quasars to the cosmic reionisation of hydrogen (e.g. Onoue et al. 2017; Kulkarni et al. 2019b). Used as the most distant beacons of light in the Universe, quasars are also an invaluable tool to study the intervening material in absorption. Metal absorption studies have charted the cosmic densities of metals up to the first billion years (e.g. Becker et al. 2009; Ryan-Weber et al. 2009; D'Odorico et al. 2013; Bosman et al. 2017; Codoreanu et al. 2018; Meyer et al. 2019a; Becker et al. 2019). The scatter in the Lyman-α forest at the tail end of reionisation has highlighted the patchy nature of this process (Bosman et al. 2018; Eilers et al. 2018a). Finally, reduced absorption of the

intrinsic Lyman-α emission by neutral gas in the so-called quasar 'near-zone' has been explored as a promising test of the evolving neutral fraction of hydrogen at $z \gtrsim 6$ (Mortlock et al. 2011; Keating et al. 2015; Eilers et al. 2017; Bañados et al. 2018).

However, absorption studies rely on the accurate reconstruction of the quasar continuum. It is common practice to extrapolate the Lyman-α emission and the blueward continuum (both heavily absorbed in the reionisation era) from the redward continuum. The most common techniques are based on the fitting of spectral Principal Component Analysis (PCA) eigenvectors learnt from SDSS quasars (e.g. Suzuki 2006; Lee et al. 2012; Pâris et al. 2011; Davies et al. 2018c). Recent studies have debated whether or not these SDSS-PCA templates are suitable for early quasars (Mortlock et al. 2011; Bosman & Becker 2015; Bañados et al. 2018; Venemans et al. 2018; Davies et al. 2018b). A crucial question is hence whether the spectra of quasars evolve with redshift. On the one hand, absorption studies are based on the *a priori* characterisation of the intrinsic continuum at low-redshift and the assumption that it does not evolve. On the other hand, the intrinsic evolution of the spectra of quasars would provide a key observation to constrain the evolution of Active Galactic Nuclei (AGN), their physics and their co-evolution with their host galaxies.

A prime candidate for testing the evolution of quasar spectra is the broad C IV λ1549 Å emission line. The equivalent width of the C IV line strongly anti-correlates with the continuum luminosity - the so-called Baldwin effect (Baldwin 1977). The C IV line is also known to present a variety of asymmetrical shapes and blueshifts, measured with respect to the quasar' systemic redshift, are found to be correlated with its equivalent width (e.g. Richards et al. 2011 and references therein). The origin of these peculiarities is often attributed to radiation-driven outflows or poorly understood Broad Line Region physics (BLR). The bulk population of low-redshift quasars (e.g. SDSS, Shen et al. 2008; Pâris et al. 2017) present on average a mildly blueshifted C IV line ($\sim$ 800 km s^{-1}). Therefore, the extreme C IV blueshifts

($\gtrsim$ 3000 km s^{-1}) reported in the three first $z > 7$ quasar (Mortlock et al. 2011; Bañados et al. 2018; Wang et al. 2018) raised interesting questions about the possible evolution of quasar spectra (and the properties of the C IV line in particular).

It remains to be seen, though, if the objects detected so far at high-redshift are representative of the whole population of early quasars. Recent observational studies of the C IV line blueshift have provided contrasting evidence for such evolution (Mazzucchelli et al. 2017; Shen et al. 2019). Different methods for measuring the C IV blueshift have complicated comparison across authors and redshifts (e.g. Coatman et al. 2016). The determination of the systemic redshift against which C IV velocity shifts are measured has variously been done *via* the cross-correlation of a quasar template (Hewett & Wild 2010; Richards et al. 2011), Independent Component Analysis (ICA) decomposition (Allen & Hewett, in prep., used by e.g. Coatman et al. 2016; Rankine et al. 2020), fitting PCA eigenvectors (e.g. Mortlock et al. 2011), or using a reliable quasar Broad Emission Line (BEL) such as Mg II (e.g. Mazzucchelli et al. 2017; Shen et al. 2019; Reed et al. 2019). Similarly, the location of the C IV line has been variously determined via the line centroid (Coatman et al. 2016; Reed et al. 2019), the flux-weighted central wavelength (e.g. Mortlock et al. 2011) the peak of a multi-Gaussian (or Gauss-Hermite) profile (e.g. De Rosa et al. 2014; Shen et al. 2016, 2019) or the maximum of a PCA template (e.g. Pâris et al. 2011, 2017). Furthermore, quasar BELs can be broken down into broad and narrow components, which appear to shift differently (Greig et al. 2017), and absorption in the red or blue wing of the emission biases the measurement of the intrinsic shift. Some of these techniques may even hamper the opportunity to probe the possible evolution of quasar spectra as they explicitly rely on templates extracted solely from low-redshift objects. Such factors cast a shadow of uncertainty on these velocity shifts.

In this Chapter, I aim to address the question afresh and study the evolution of quasar spectra across cosmic time. In order to achieve this, we have gathered a comprehensive set of quasars at $1.5 < z < 7.5$ based on various samples of optical and near-infrared spectroscopic observations. Crucially, I take advantage of *QUICFit*

(Section 3.2.2), a lightly user-supervised, model-independent spline fitting method for the quasar continuum and broad rest-frame UV lines redwards of Lyman-α. This approach does not attempt to recover the systemic redshift, rather it simply retrieves in a uniform manner the observed peak of the BELs without assuming a particular model for these emissions. We can thus study the relative shift between C IV and Mg II, but also the relative shifts between O I, C II, C III], Si IV and the two previous lines as other potential markers of evolution. In doing so, we can probe the shifts of high-ionisation lines with respect to low-ionisation lines across cosmic time to evaluate on possible quasar evolution and provide insight into BLR physics.

The outline of this Chapter is as follows. In Section 5.2, we first detail our different datasets and the selection procedure of luminosity-matched control sample across our redshift range. We then assess the quality of our fits, and the different biases that could affect our measurement. Finally, we compare our method to PCA and Gaussian fitting on SDSS quasars. We present in Section 5.3 the evolution of the relative velocity shifts of all aforementioned BELs with respect to one another. In Section 5.4, we discuss our results in light of previous studies and we propose a novel interpretation of the increased C IV blueshift observed in early quasars.

Throughout this Chapter, we adopt a ΛCDM cosmology with $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$, $H_0 = 70$ km s^{-1}Mpc^{-1}. We adopt a positive sign for velocity redshifts and negative for blueshifts.

5.2 Methods

5.2.1 Observational samples

In order to study the velocity shifts of quasar UV BELs across cosmic time, we have gathered a large sample of quasar spectra in infrared and optical light from different surveys and archival data. We briefly describe here the original samples from which our quasars were drawn and how we construct our luminosity-matched control samples. The relevant parameters of the studied datasets are summarised in Table 5.1 and the luminosity distributions are presented on Figure 5.1. We also

Name	Δz	$\Delta L_{\lambda,1350}$	R	$\overline{\mathrm{SNR}}$	N
DR12Q[a]	$1.5-2.4$	40.00-44.5	1500-2500	3.4	99232
SDSS[b]	$1.5-4.5$	43.3-44.4	1500-2500	14	108
XQ100	$3.5-4.5$	43.0-44.2	6000-9000	30	100
GGG	$4.4-5.5$	42.7-44.3	800-900	20	163
z6	$5.4-6.3$	43.4-44.3	9000-11000	90	11
z7	$6.4-7.5$	42.9-43.8	~ 6000	6	12

Table 5.1: The different quasar samples with their redshift and luminosity ranges, resolution, median SNR and number of objects. (a) We use here the subset of DR12Q that have both tabulated C IV and Mg II-derived redshift. (b) Throughout the text, SDSS refers to a subsample of the DR7Q and DR12Q overlapping quasars, luminosity-matched to the z6 sample, described in Section 5.2.1

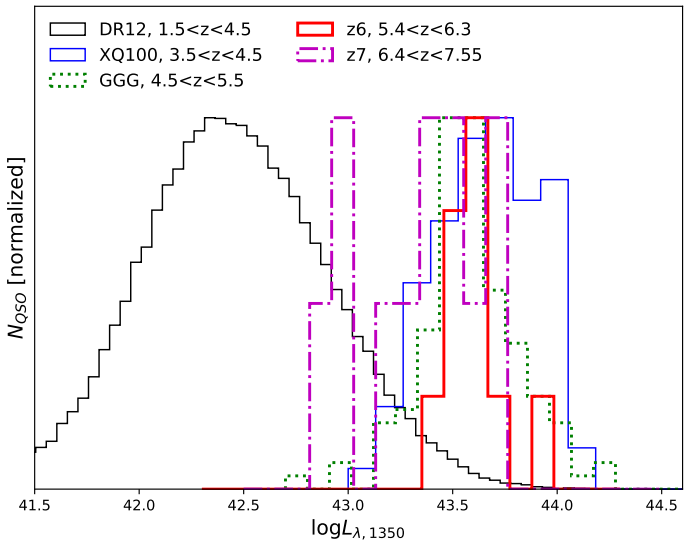

Figure 5.1: Specific luminosities (derived from the observed specific flux at 1350 Å $f_{\lambda,1350}$) of our different samples. Note the relative good match between the luminosities of the GGG (green), XQ100 (blue), z6 (red) and z7 (magenta) samples. They can then be readily compared to each other and to luminosity-matched subsamples of SDSS DR12 (black).

show the stacked spectra of all our samples on Figure 5.2. Throughout this study, we refer to the specific luminosity at 1350 Å, which we compute from the specific flux at 1350 Å rest-frame and the luminosity distance, simply as the *luminosity*. We note that although the 1350 Å luminosity is a practical choice given it can be measured for nearly all of our quasars, a different choice of luminosity matching (e.g. rest-frame X-ray or IR) might affect the results of this study due to dust correction effects. However, the emissivity at 1350 Å is more closely related to the UV emission lines of interest. In addition, the overlapping spectral coverage for our large sample of quasars is limited to the rest-frame UV.

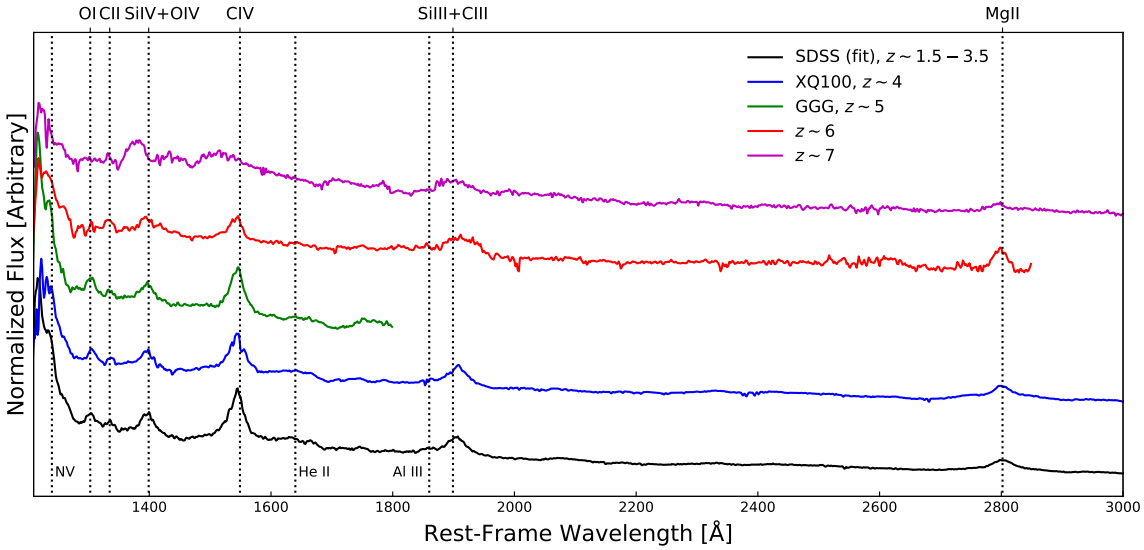

Figure 5.2: Rest-frame UV mean spectra of our different samples, presented with increasing redshift from bottom to top (black: SDSS luminosity-matched fitted subsample, blue: XQ100, green: GGG, red: z6 quasars, magenta: z7 quasars). The spectra are normalized by their flux at 1350 Å and the powerlaw is not removed before stacking. We indicate the six major BELs we fit with vertical dashed lines. The C IV emission can be seen to blueshift and to present a lower equivalent width in the high-redshift quasars. The mean C IV line of the $z \sim 7$ stacked spectra is extremely flat due to the huge spread of blueshifts between a relatively small number of objects, but the mean shift is still visible. Other noticeable features include the reduced Lyman-α emission at high-redshift due to an increased neutral fraction of hydrogen and the emergence of Si IV at the expense of O IV over $5 \lesssim z \lesssim 7$.

SDSS DR7 and DR12: SDSS-I/II, -III (York et al. 2000; Eisenstein et al. 2011) delivered nearly 5 million spectra and 32000 deg^2 of imaging in the *ugriz* filters. In particular, we are interested here in two major catalogues of quasars observed and identified within SDSS data: the DR7 quasar catalogue (DR7Q, Schneider et al. 2010) and the DR12 quasar catalogue (Pâris et al. 2017). For DR7Q, we make use of the value-added catalogue of Shen et al. (2011) which provides additional properties for the quasars listed in the original DR7 catalogue. Shen et al. (2011) report redshifts derived from C IV, C III] and Mg II, and the specific luminosities. In DR7Q, the line-based redshifts (i.e. redshifts derived from one emission line) are derived by fitting a multi-Gaussian template to each broad emission line and taking the template peak to derive the redshift. In DR12Q however, all quasars were inspected by eye and have thus a visual redshift, which is different from the line-based redshifts. The line-based redshifts in DR12 are derived from the peak of

a PCA template fitted locally for each emission line (Pâris et al. 2011). There is some overlap between DR7Q and DR12Q, although the latter sample is larger and contains quasars with slightly lower luminosities and in a broader redshift range. The SDSS optical spectrograph (Smee et al. 2012) covers the range $3600 - 1000$ Å, with a resolution going from $R \sim 1500$ to $R \sim 2500$. Hence the detection of both C IV and Mg II is limited to quasars at $1.5 \lesssim z \lesssim 2.4$. However, other UV BELs like O I, C II or Si IV can be detected simultaneously with C IV in the complementary and overlapping redshift range $1.8 \lesssim z \lesssim 5.1$.

XQ100: XQ100 (López et al. 2016) is a legacy European Southern Observatory (ESO) survey which observed 100 quasars at $3.5 \leq z \leq 4.5$ at medium resolution ($R \sim 6000 - 9000$) and a high median Signal-to-Noise Ratio SNR $\simeq 30$ with the XShooter instrument (Vernet et al. 2011) at the Very Large Telescope (VLT). Its primary aims included absorption lines studies, AGN science, Broad Absorption Lines (BAL) studies, Damped Lyman-α absorbers (DLAs) and Lyman-α forest-based cosmology. The combined wavelength range of XShooter's UVB, VIS and NIR arms ($3150 - 24800$ Å, or $3150 - 18000$ Å, depending on at which point in the survey the object was observed) is well suited to the aim of this study as all UV BELs can be retrieved over the entire redshift range $3.5 \leq z \leq 4.5$, except when they fall in telluric regions, spectrograph arm overlaps and parts impeded by the atmospheric transmission. XQ100 also overlaps partially with SDSS, making this subset ideal to study the impact of resolution on the observed spectra and the recovered fitted continuum. The XQ100 quasars were selected to be intrinsically bright, hence they are slightly brighter than the other samples used here

Giant Gemini GMOS Survey: The Giant Gemini GMOS survey (GGG, Worseck et al. 2014) obtained spectra of quasars over $4.4 \leq z \leq 5.5$ with the Gemini Multi Object Spectrometers (GMOS) to study the Lyman-continuum (LyC) flux at $z \sim 5$ and determine the LyC photon mean free path. The publicly available data release of the GGG survey is composed of 163 quasars at low-resolution ($R \sim 800 - 900$) with a medium-high SNR ~ 20, making it also suitable to capture the broad shapes of the UV BELs. The GMOS instrument covers the range $4800\mathring{A} \lesssim$

$\lambda \lesssim 10200$ Å, meaning that C III] and Mg II can only be observed for quasars at $z \leq 4.4$ and $z \leq 2.6$, respectively. Thus we cannot measure the C IV-Mg II relative shifts. However, since we do not restrict ourselves to this particular pair of BELs, we can use GGG quasars to compute velocity shifts of lower rest-frame wavelength BELs. GGG thus suitably completes the redshift ladder we have built by filling the gap between the SDSS datasets, XQ100 quasars and the highest-redshift samples at $z > 5.5$.

$z \sim 6$ **sample:** Our first high-redshift sample (hereafter 'z6') consists of a subset of quasars from the extensive list of $z \sim 6$ object from Bosman et al. (2018), of which some where re-reduced later in Chapter 3. We keep only those with near-infrared coverage and high SNR, limiting the sample to 11 objects. This approach ensures a comparison as complete as possible over $1.5 \leq z \leq 6.5$, including all the lines from O I to Mg II. This sample is critical to probe with additional spectra claims in the literature of (non-)evolution from the C IV-Mg II velocity shifts. The 11 quasars selected have both the highest-resolution ($R \sim 9000-11000$) and highest median SNR (90) of our samples (see Table 5.1).

$z \sim 7$ **sample:** Our second high-redshift sample is composed of the Mazzucchelli et al. (2017) quasars that are observed far enough in the infrared to detect at least C IV, to which we add the present record-holding $z = 7.54$ quasar (Bañados et al. 2018), kindly provided by private communication of the authors. This sample thus comprises 12 quasars observed with various instruments with $\langle z \rangle = 6.73$, median SNR ~ 6 (although individual spectra range from ~ 2 to ~ 100) and resolution $R \sim 6000$.

Since BEL shifts correlate with luminosity, studying their potential evolution with redshift requires doing so at a fixed luminosity. We therefore construct luminosity-matched control samples from the SDSS quasars for all our higher-redshifts samples by oversampling their luminosity distribution. For for each quasar in a test high-redshift sample (XQ100, GGG, z6, z7), we draw a random SDSS quasar with a similar luminosity ($\Delta L = 0.2$ dex). We then repeat the procedure until we reach 5000 SDSS quasars. DR7Q and DR12Q only contain information about the peak

location of C III], C IV and Mg II, derived with different methods highlighted above which cannot be compared directly with our measurements. To enable a fair and uniform comparison, we also construct a small luminosity-matched (to z6) control sample of 108 quasars spanning the redshift range $1.5 < z < 4.5$. This smaller control sample can be then fitted with our spline technique to control our systematics between different peak retrieval methods, resolution and instruments, and derive O I, C II and Si IV-based redshift for SDSS quasars as well.

5.2.2 Continuum fitting and line-derived redshifts

We now describe the continuum fitting method and the subsequent derivation of the line-based redshifts that we apply uniformly to the quasar samples described above.

To enable the algorithm to run on a globally flat spectrum, each quasar is fitted with a continuum powerlaw (e.g. Vanden Berk et al. 2001) with the following form

$$F(\lambda) = F_0 \left(\frac{\lambda}{2500\mathring{A}} \right)^{\alpha} \tag{5.1}$$

where F_0 [10^{-16} erg s^{-1} cm^{-2} Hz^{-1}] is the specific flux at 2500Å, λ the rest-frame wavelength in Å and α is the powerlaw exponent. The powerlaw is fitted to regions relatively devoid of features following Decarli et al. (2010); Mazzucchelli et al. (2017) at $1285-1295, 1315-1325, 1340-1375, 1425-1470, 1680-1710, 1975-2050, 2150-2250$ and $2950-2990$ Å rest-frame with a positive linear prior on F_0 and α.

Once each rest-frame UV spectrum spectrum is divided by the powerlaw, we fit it using QUICFIT, which we describe briefly here. Further details can be found in Section 3.2.2. QUICFIT is a lightly-supervised spline quasar continuum algorithm which aims to avoid the human selection of appropriate continuum portions of the quasar spectra. Instead, the algorithm first discards BALs (typically of equivalent width EW$> 1-2$ Å) by running a Gaussian matched-filter with a width of a few Ångströms. Smaller emission and metal absorption features, as well as bad pixels and cosmic rays are removed by searching for portions of the spectrum with excessive variance in the pixel-to-pixel flux differences. Indeed, if the spectral resolution

is much higher than the continuum typical variation scale, the pixel-to-pixel flux increments and decrements distribution should follow the error array distribution. By computing the empirical variance in small portions of the spectrum, QUICFIT can efficiently detect and removes structures (such as narrow absorbers, emitters or cosmic rays) that differ from a smooth continuum and expected uncertainty. Once this is done, the remaining parts of the spectrum can be considered representative of the intrinsic continuum and used to fit third order splines. We subsequently call these selected parts of the spectrum 'continuum pixels' later in this Chapter.

The spline knot points (where the polynomials of the spline join and the continuity conditions are enforced) can be initialized by the user, but it is found that one knot point at the location of each BEL, and one intermediate point in between, are sufficient to fit adequately third-order splines on the pre-selected continuum pixels. The process is lightly supervised and fast (a minute or two per quasar), as the user only has to adjust a threshold to reject pixels in narrow absorptions based on the expected uncertainty (see above) and choose whether or not to fit faint BELs by adding more spline knots. It is also found that complex lines such as C IV or C III] are better fitted with an extra knot point. Therefore we take particular care in accurately fitting the emission lines of interest: O I, C II, Si IV, C III], C IV and Mg II. This approach avoids any misfits occurring in pipelines for large datasets such as SDSS quasars, while being still applicable to smaller datasets with a range of SNR and resolution.

The resulting spline fits and the line-based redshift solutions for all the BELs of our quasars are presented in Appendix C.1. Before investigating the suitability of our scheme to recover unbiased measurements of the BELs peak, we first focus on the quality of the fits by looking at the residuals. In order to do so, we assume that our selection of continuum pixels is correct, and compute the stacked residuals on these points only. We present on Figure 5.3 (black points) the residuals of the continuum fits for all the different samples. The residuals are flat and show no particular trend, evidencing that we fit correctly the continuum of the selected continuum pixels.

Once the BELs have been fitted, we retrieve the peak of the spline fit to derive

the line-based redshifts. In doing so, we take care to select only significant relative maxim in the continuum according to the following criteria. We first take the maximum in a rest-frame window of 10 Å (O I, C II, Si IV, C III]) or 40 Å (C IV, Mg II) around the expected BEL location, computed using the reported redshift of the quasar in the literature. We then require that this maximum is significantly higher than the surrounding continuum. In order to do so, we take the relative minima of the continuum at lower and higher wavelength and we interpolate the continuum between the two minima. We compute the equivalent width (EW) of the line with respect to the interpolated continuum and we retain the line if it has EW $>$ 3Å. As the EW of the C IV line anti-correlates with its blueshift (e.g. Coatman et al. 2016), we expect such a threshold could potentially bias the measurement of the blueshifts towards lower values. However this should not affect a potential evolution of the average blueshift. If multiple such maxima are retained, as can be the case for the Si IV and C III] blended multi-line complexes, we retain only the lower and higher wavelength peak, respectively. This is because we are interested in the Si IV and C III] emissions, not the blended O IV λ1402 Å and Si III] λ1892 Å, respectively. This procedure disentangles the peaks when the two emission lines have similar equivalent widths and are resolved. Nonetheless, in some cases the line-based redshift simply tracks the peak of the blend, and is thus slightly biased towards larger (lower) values for Si IV (C III]). We further comment on the impact on our results and the suitability of these line complexes to study relative velocity shifts in Section 5.3.

5.2.3 Instrument and resolution biases

.

Our quasar samples (Table 5.1) have different resolutions and SNR, and have been observed at different redshifts. We have fitted the continuum with QUICFit in a uniform way for the SDSS, XQ100, GGG, z6 and z7 samples. However, we need to ensure that our method is not biased by the varying resolution of the different

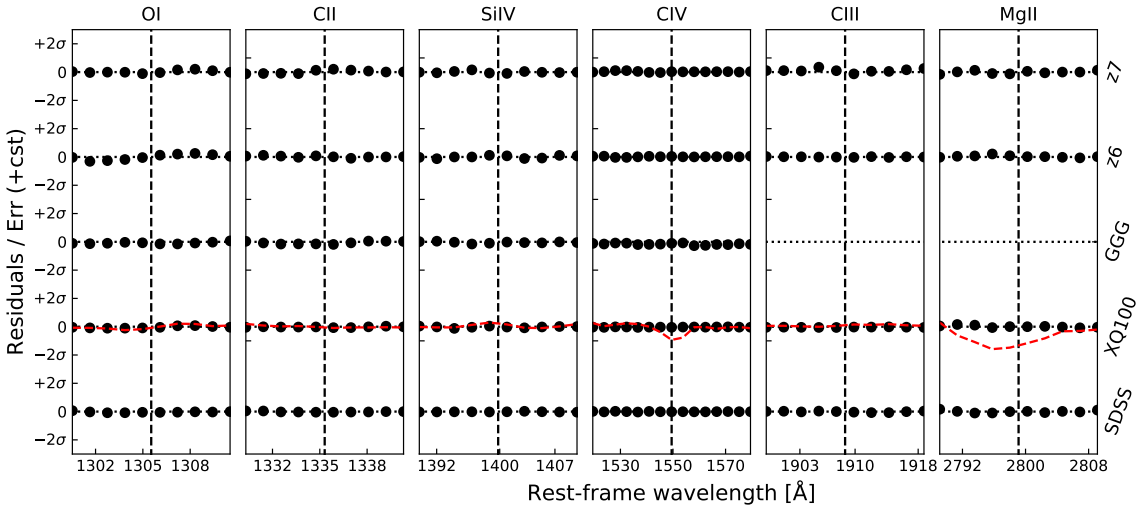

Figure 5.3: Stacked and binned residuals of the fitted splines (black points) for the six different UV BELs of interest in our five different samples. The residuals show a very good agreement between the continuum-selected pixels and the resulting spline fit. The SDSS sample is a 108 quasar-strong sub-sample of SDSS chosen to match the luminosity of the z6 sample (see Sec. 5.2.1). The red dashed line shows the residuals between the high-resolution XQ100 spectra and the continuum splines recovered from a degraded resolution version of the same spectra. The deviations at the $1-1.5\sigma$ for C IV and Mg II are expected from the widening of narrow absorbers damping the observed peak at low resolution.

samples towards lower or higher velocity shifts. This might be the case in the presence of a large number of unresolved associated narrow absorbers that would blend with the intrinsic profile and damp the blue or the red wing of the line. At higher resolution, these absorbers are resolved and thus successfully removed from the continuum to recover the intrinsic profile. This might create a marked difference between the high- and low- resolution samples if absorbers are found preferentially in the blue or red wing of BELs.

In order to address this point, we have degraded the XShooter spectra to SDSS-like resolution by re-binning them by a factor 3-5 (depending on the XShooter arm and the redshift of the object) and re-fitted the degraded quasar spectra with *QUICFit*. We demonstrate qualitatively the resilience of the fitting method by comparing the BEL spline fits recovered from the high- and low-resolution spectra in Figures 5.3 and 5.4. The recovered continua and residuals for O I, C II,Si IV and C III] are essentially identical. Nonetheless, the recovered continua for C IV and Mg II is slightly lower at the peak of the BEL. As expected, the lower resolution

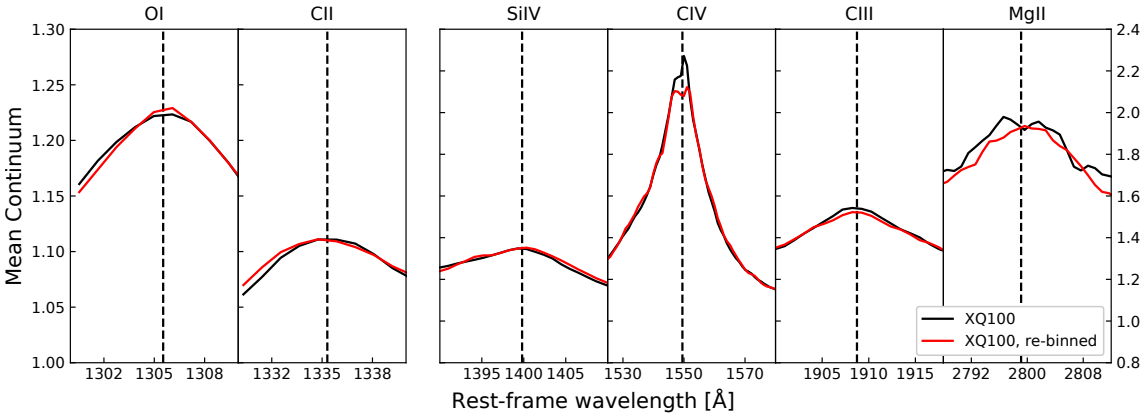

Figure 5.4: Comparison of the stacked emission profile for the six BELs of interest obtained from the regular ($R \sim 7500$,black) and degraded resolution ($R \sim 1800$,red) version for the XQ100 quasars. While the O I, C II, Si IV and C III] continuum are essentially identically recovered, C IV and Mg II are impacted by the widening of narrow absorbers at lower resolution. The emission profile recovery with *QUICFit* is thus quite robust to a change in resolution.

smoothes out sharp features such as the line peaks and narrow absorbers. This causes the residuals of the low-resolution continuum to be on average $1.0\sigma - 1.5\sigma$ lower at the peak value. However the peak dampening of C IV is highly symmetric, while Mg II presents a small offset to redder values. We now quantitatively qualify the potential small bias induced in the final measurement of the line-based redshifts and subsequent relative velocity shifts. We compare the redshifts derived from the fitted BELs at low and high resolution and we find that the line-based redshifts are not biased when the resolution is lowered (Figure 5.5 and 5.6). The average difference between high- and low- resolution is smaller than natural intrinsic scatter of the measured redshifts by a factor of $5-10$ (see Table 5.2). Specifically, the mean redshift error due to the resolution decrease is < 100 km s $^{-1}$, whereas the standard deviation is of the order of ~ 400 km s $^{-1}$.

To ensure that a change in both resolution and instrumental setup does not impact the measurement of the line-based redshift and relative velocity shifts, we apply our method to the 53 XQ100 quasars observed with the SDSS spectrograph. We fit the quasar continuum with *QUICFit* directly on the SDSS spectra and retrieve all possible BELs when they fall in the spectrograph coverage. We present as before the scatter in the measurement of the line-based redshifts on Figure 5.7 and the

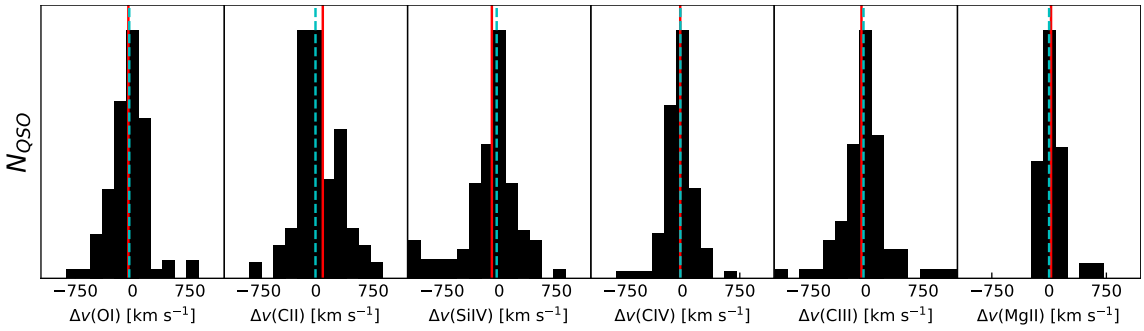

Figure 5.5: The impact of resolution degradation for the XQ100 spectra from $R \simeq 6000$ to $R \simeq 2000$ on derived BEL-based redshifts. The very tight distribution of velocity shifts shows how such a spectral resolution decrease does not affect the performance of *QUICFit* and does not bias the retrieved lines. The mean (red continuous line) and median (dotted blue line) offsets are much smaller than the standard deviation of the offset distribution for all species (see further Table. 5.2 for the tabulated value of the mean, median shift and the standard deviation of each distribution).

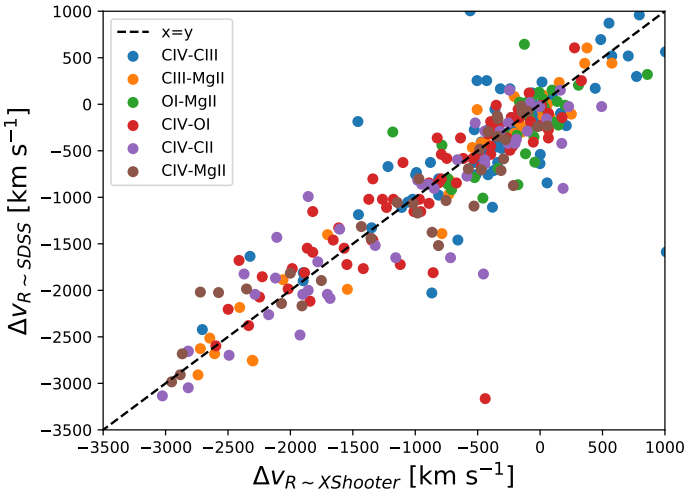

Figure 5.6: The impact of resolution degradation for the XQ100 spectra on derived BELs relative velocity shifts. Spectral resolution does not affect the performance of QUICFit as the figure demonstrates here the absence of bias towards higher or lower velocity shifts, despite a moderate spread (see Table. 5.2 for the tabulated values of the mean, median shift and 1σ error)

	Rebinned			SDSS		
Line	$\overline{\Delta v}$	$\langle\Delta v\rangle$	$\sigma_{\Delta v}$	$\overline{\Delta v}$	$\langle\Delta v\rangle$	$\sigma_{\Delta v}$
O I	−56	−44	250	−73	−38	349
C II	90	−6	538	−25	−6	179
Si IV	−96	−33	451	−53	−11	407
C IV	−30	−27	537	−30	−16	131
C III]	−57	−30	474	36	19	236
Mg II	25	−0	166	−35	−19	209

Table 5.2: Median ($\overline{\Delta v}$), mean ($\langle\Delta v\rangle$) and standard deviation ($\sigma_{\Delta v}$) of the velocity error Δv on of the recovered peak of each BEL with respect to the original value derived from XQ100 spectra. The columns one to four ('Rebinned') give the values when the resolution is lowered to SDSS-like levels by rebinning the data whereas the next three ('SDSS') show the comparison between redshifts retrieved from XShooter and SDSS spectra of the same quasars. All values are provided in km s^{-1}.

subsequent impact on the derived velocity shifts on Figure 5.8. The values of the mean, median and standard deviation of the line-based redshifts errors are presented in Table 5.2 alongside the previous results comparing the original XQ100 XShooter spectra to the degraded resolution version. As expected, we find no further bias by comparing XShooter and SDSS spectra than by degrading XShooter spectra. The standard deviation of the velocity shift difference between the XShooter and SDSS is often smaller than for the previous case, which we attribute to being pessimistic whilst degrading the resolution. We find no bias either in the line-based redshifts or the derived BELs relative velocity shifts. We are thus confident that our spline fitting and peak retrieval method will not bias the measured relative velocity shifts between broad lines and is independent of resolution, instrument and observational setup and date.

5.2.4 Comparing *QUICFit* with PCA and Gaussian template fitting methods

QUICFit provides a resolution- and instrument- resilient method to fit the UV continuum and BELs of quasar continua. In this section, we compare line-based redshifts derived with our method to PCA and Gaussian fitting techniques used in previous studies. In doing so, we can also check that a change in fitting method does not bias the measurement by examining XQ100 quasars also observed in SDSS. SDSS

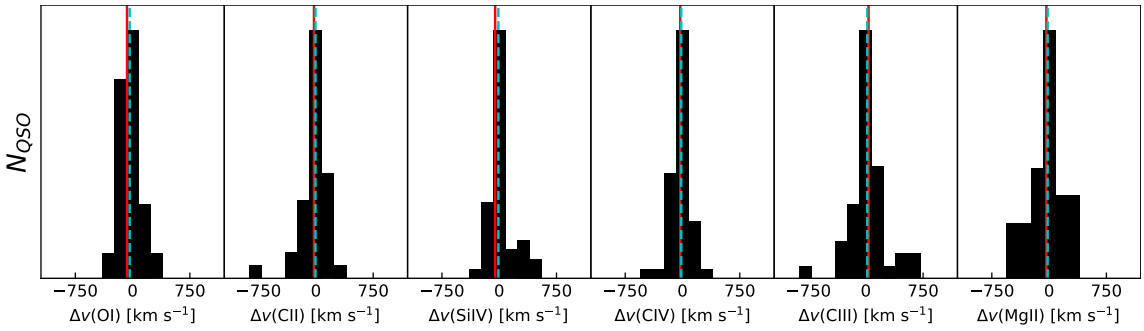

Figure 5.7: Comparison between derived BEL-based redshifts from XShooter and SDSS spectra of the same quasars. The very tight distribution of velocity shifts shows how a decrease in spectral resolution and different instrumental setups do not affect the performance of *QUICFit* and do not bias the retrieved lines. The mean (red continuous line) and median (dotted blue line) offsets are much smaller than the standard deviation of the offset distribution for all species (see further Table. 5.2 for the tabulated values of the mean, median shift and the standard deviation of each distribution).

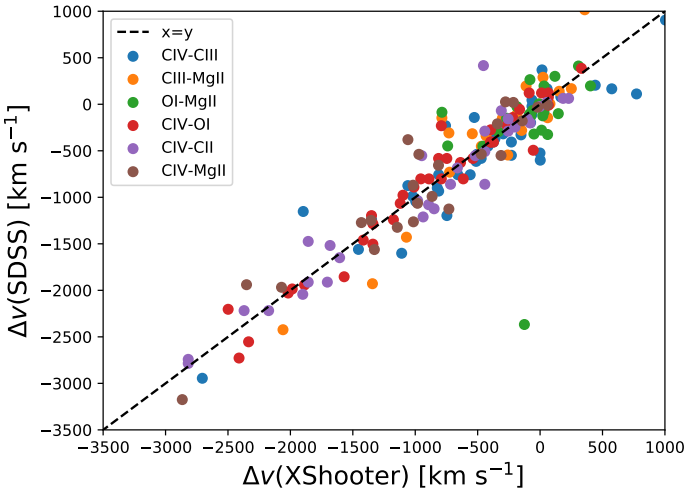

Figure 5.8: Comparison between derived BELs relative velocity shifts from XShooter and SDSS spectra of the same quasars. Spectral resolution does not affect the performance of QUICFit as the tight spread demonstrates here the absence of bias towards higher or lower velocity shifts (see Table. 5.2 for the tabulated value of the mean, median shift and 1σ error)

DR12 is too large a sample to be fitted by hand, and the DR12 quasar catalogue only contains redshifts for C IV, C III] and Mg II (Pâris et al. 2017). These redshifts are derived from the peak of the PCA template (Pâris et al. 2012). DR7 redshifts, however, are derived using a Gaussian template fitting method (Shen et al. 2011). Both methods have their drawbacks and advantages, but they agree on the average C IV-Mg II velocity shift at $1 \lesssim z \lesssim 3$. In order to harness the statistical power of the DR7Q/DR12Q catalogues which measured C IV-Mg II velocity shifts, we have to assess the possible difference between our measurements and the SDSS DR7/DR12 line-based redshifts. In order to do so, we compare the velocity shift derived from our method to the values provided by DR7Q and DR12Q where applicable in our luminosity-matched SDSS sample (Figure 5.9). We show that there is no significant bias between our values, DR7Q and DR12Q values. The standard deviation between the different methods is however quite large. We find that the velocity shift difference between our results and DR7Q is ($\Delta v = (-83 \pm 635)$ km s^{-1} and $\Delta v = (47 \pm 475)$ km s^{-1} (~ 500 km s^{-1}) for DR12Q. In both cases, the standard deviation is greater or consistent with the typical error expected from a change in resolution or a different instrumental setup as found in Section 5.2.3.

Finally, we can check that a concurrent change in resolution, SNR, instrument, and fitting method does not affect the measurement. For quasars present in both XQ100 and DR12Q, we compare the C IV- and C III]-derived redshifts from DR12Q to redshifts computed from our spline fits to the XShooter spectra. Because XQ100 contains quasars at $3.5 < z < 4.5$, the corresponding SDSS observations do not cover Mg II. We present on Figure 5.10 the distribution of the two methods' redshifts differences. We find that our solutions are similar to the SDSS ones with the usual ~ 500 km s^{-1} spread for the two BELs species. The median shift, mean shift and the standard deviation of the velocity shift distribution are $\overline{\Delta v_{\mathrm{CIV}}} = 53$ km s^{-1}, $\langle \Delta v_{\mathrm{CIV}} \rangle = 90$ km s^{-1}, $\sigma_{\Delta v_{\mathrm{CIV}}} = 392$ km s^{-1}, and $\overline{\Delta v_{\mathrm{CIII]}}} = -7$ km s^{-1}, $\langle \Delta v_{\mathrm{CIII]}} \rangle = -178.6$ km s^{-1}, $\sigma_{\Delta v_{\mathrm{CIII]}}} = 652$ km s^{-1} respectively. There are ~ 10 % of spectra presenting shifts above 500 km s^{-1} for C IV and ~ 30 % for C III].

For these 'catastrophic' outliers, we plot their SDSS spectra, the XShooter spec-

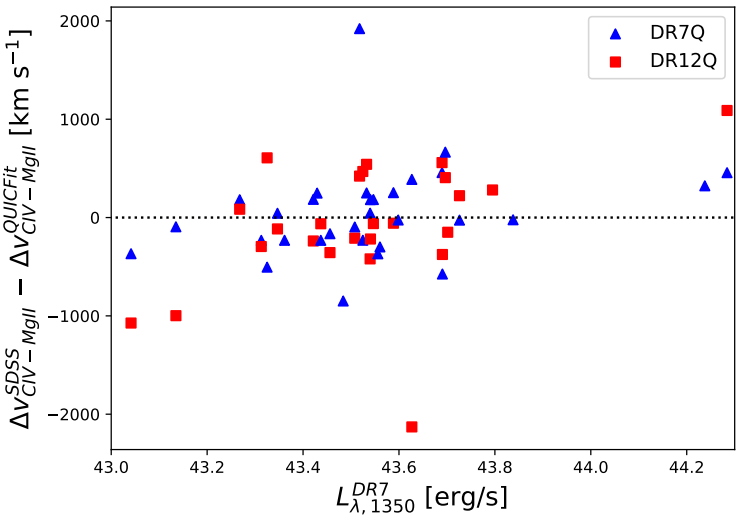

Figure 5.9: Comparison of the velocity shifts of C IV with respect to Mg II for our SDSS subsample (Table 5.1). We compare the redshifts derived from Gaussian fits (DR7, blue dots) and PCA template (DR12, red square) to the QUICFit solution when these DR7Q and DR12Q provide line redshifts for both C IV and Mg II. Although the trend is not significant, the DR12Q PCA method seems to underestimate slightly the blueshifts of more luminous quasars. In fact, the bias of the PCA towards smaller blueshifts for small C IV equivalent width, which in turn anti-correlate with luminosity, was already pointed out by Coatman et al. (2016). We find no (potential) systematic trend with redshift.

tra and the DR12Q and QUICFit line-based redshift solutions on Figures C.2 and C.3. Upon further inspection, we find that in most cases PCA seems to fail when multiple narrow absorptions, resolved with XShooter but not SDSS, artificially damp one wing of the emission. One could argue that PCA actually recovers the original feature because it is trained on unabsorbed spectra. However, the PCA emission line redshift solutions are determined by fitting a few eigenspectra to each UV broad line individually (Pâris et al. 2012). The PCA emission line redshifts are thus not derived from the overall spectral PCA fit and thus cannot reconstruct the 'true' emission from other features but rather perform a local best fit to the the emission shape. Another feature that the PCA seems unable to capture is a prominent contribution of Si III] λ1892 Å to the C III] complex.

5.3 Relative velocity shifts of broad UV emission lines

We have demonstrated that our method is suitable for application to a large compilation of quasars of different SNR, resolution and redshift. We have also shown

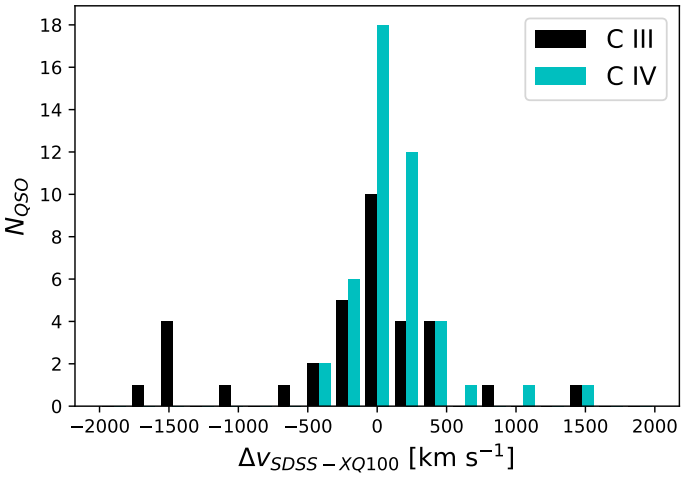

Figure 5.10: Comparison between the line-based (C III], C IV) redshift derived from our method on the high-resolution XQ100 quasars and the tabulated values in the DR12Q catalogue for the overlapping low-resolution SDSS quasars. The methods give redshift in agreement with a typical spread of ~ 500 km s $^{-1}$. BELs presenting a difference of $\gtrsim 500$ km s $^{-1}$ between our solution and the DR12Q tabulated values are presented in Figures C.2 and C.3.

that our method is comparable to two main methods for deriving line redshifts (PCA templates or Gaussian fitting), with an intrinsic spread between methods of $\Delta v \sim 500$ km s^{-1}. We can now harness the statistical power from our extended sample as well as our extended number of studied BELs.

5.3.1 Single species broad lines: O I, C II, C IV and Mg II

We start with the relative velocity shift of C IV with respect to Mg II, as it is the focus of most previous studies. We note that only a fraction of our z6 and z7 quasars can be used for this analysis, as it requires good coverage of both the Mg II and C IV regions. The GMOS wavelength coverage does not allow us to detect the Mg II line, therefore we discard the GGG sample for this particular analysis. We present on Figure 5.11 the results for both the z6 and z7 samples compared to appropriate DR12Q control samples. We find that, when luminosity is matched, there is a significant evolution of the shifts at $z \gtrsim 6 - 6.5$. Specifically, the mean C IV blueshift evolves from ~ 1000 km s^{-1} at $z \lesssim 6$ to ~ 2500 km s^{-1} at $z \sim 7$, as already reported by Mazzucchelli et al. (2017). The double-sided KS-test, which tests whether two samples are drawn from the same different distribution, gives

no evidence ($p = 0.69$) for the z6 blueshifts being different from their luminosity-matched DR12Q samples. This result is similar to that reported by Shen et al. (2019) . However, the p-value for the z7 sample ($p = 3 \cdot 10^{-6}$) is clear evidence for evolution compared to luminosity-matched SDSS objects. The evolution of the C IV line is also clearly visible in the stacked spectra of all our samples (Figure 5.2), despite the intra-object spread strongly smoothing the feature in the stacked spectrum.

We now turn to the relative velocity line shifts for four BELs (O I, C II, C IV, Mg II) in all our different samples. In order to do so, we compute the mean relative shift for every pair of species in all samples, which we show on Figure 5.12. The errors are computed by boostrapping using 1000 samples of the size of the smallest sample (z6, $N = 11$) to make them comparable across redshift and samples. The low-ionisation lines display no significant velocity offsets either between themselves at any redshift. The small relative blue- or red-shift of the lines are all consistent with bootstrapping errors. These low-ionisation line are not predicted to shift with respect to each other, therefore an absence of relative shifts is a good indicator of an unbiased measurement of the line shifts.

The C IV line, however, is markedly blueshifted from all the lower-ionisation lines (O I, C II, Mg II) at all redshift. The mean blueshift significantly increases at $z \gtrsim 6.5$ by a factor 2.5, irrespective of the choice of low-ionisation species used as a reference to measure the shift of C IV. This change is indicative of different environments for the emission (and possibly the absorption) of C IV and low-ionisation lines. It is probably linked to a dichotomy between high- and low- ionisation lines given that they are thought to originate in different parts of the BLR. To test this hypothesis, measuring the mean blueshift of another high-ionisation line is needed. However, the only other high-ionisation line in the rest-frame UV is N V λ1240 Å, which is often blended with Lyman-α and low-ionisation Si II λ1260 Å. Moreover Lyman-α is increasingly absorbed at high-redshift, making the deblending more difficult. Another potential high-ionisation line, He II λ1640 Å is often too faint to be detected in most high-redshift quasars. Hence, spectroscopy of the rest-frame optical will be

needed to conclude whether this blueshift is a particularity of the C IV line or a dichotomy between low- and high-ionisation lines.

5.3.2 Broad lines complexes: Si IV and C III]

We have left aside the Si IV and C III] broad lines from the previous analysis. These two intermediate-ionisation lines are known to be complexes: Si IV λ1397 Å is blended with O IV λ1402 Å whereas C III] λ1909 Å is blended with Al III] λ1857 Å and Si III] λ1892 Å. These complexes could affect the measurement and the interpretation of any potential velocity shifts, especially since the relative line strengths of the blended emission lines appear – qualitatively – to be evolving with redshift. Figure 5.2 illustrates this effect: the O IV line disappears between $z \sim 5$ and $z \sim 6$, whereas the Si IV BEL increases in strength at high-redshift, even though the samples are luminosity-matched. Notwithstanding this complication, we attempted to measure the velocity shifts of these complexes by taking the red/blue peak of the complex accordingly when the feature was double-peaked, and the overall peak of the complex when the lines were blended (see Section 5.2).

We find no evidence of velocity shifts of C III] and Si IV relative to low-ionisation lines, and the C IV line is seen to blueshift with respect to these lines in the same manner as to single low-ionisation lines (see Figure C.4). This result can be interpreted in two ways. If we were successful at disentangling the complexes, our findings indicate that C III] and Si IV are not blueshifted with respect to low-ionisation lines at any redshift. Even if we were not able to satisfactorily separate the blending, our results still provide some constraints on the velocity shifts of C III] and Si IV. As the C III] flux is dominant over Al III] and Si III] at $2 < z < 5$ (Nagao et al. 2006) and those blended lines have shorter wavelengths, a relative strengthening of Al III] and/or Si III] could only lead the peak of the blend to blueshift. Redshifting of the complex peak would then be attributable to C III]. Since neither is seen, we can plausibly conclude that on average C III] does not redshift with respect to low-ionisation single species. Moreover, any intrinsic blueshift would need to be carefully

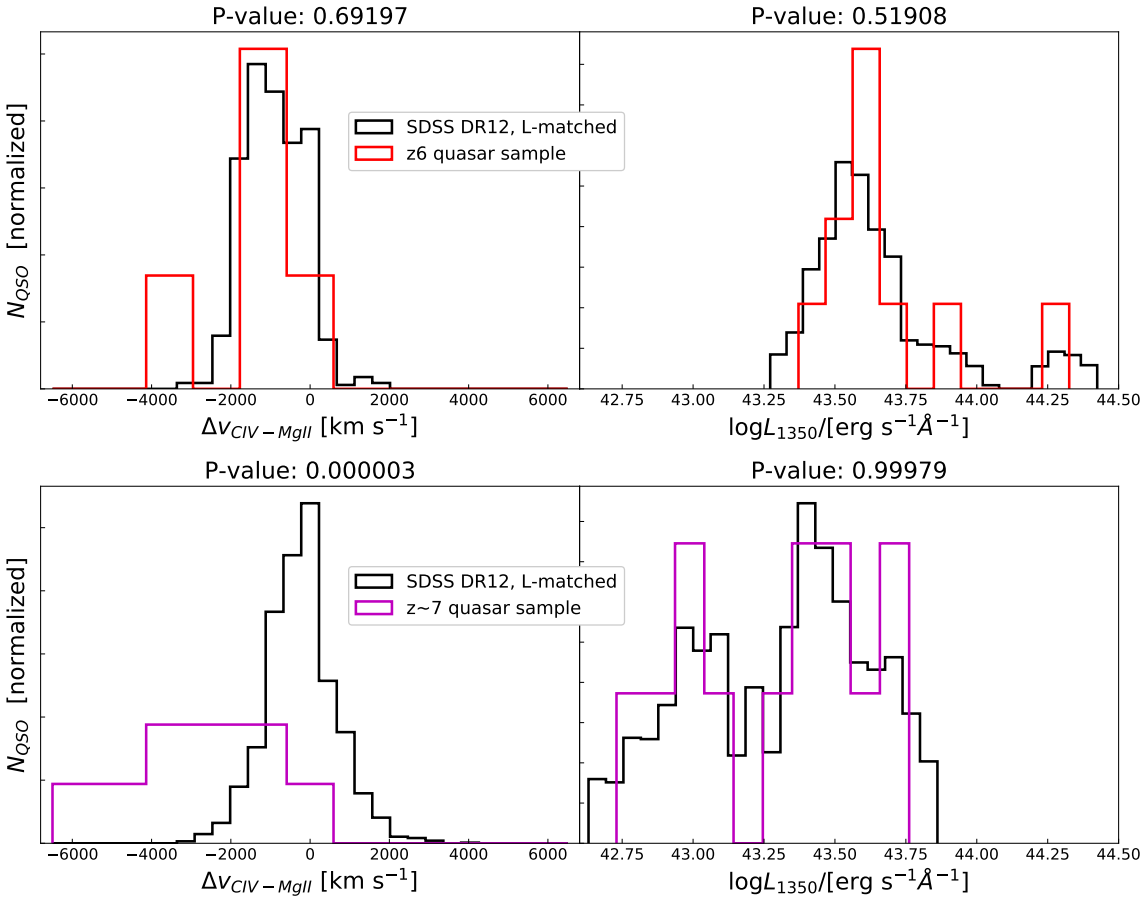

Figure 5.11: Comparison of the different C IV-Mg II velocity shifts between the z6 and z7 samples and their DR12Q luminosity-matched samples. The p-value of the double-sided KS test between the test and the control samples is shown above each histogram. The z7 quasars have significantly higher C IV-Mg II blueshifts than their luminosity-matched SDSS counterparts, whereas the z6 have comparable C IV blueshifts to the lower-redshift quasars, in agreement with Shen et al. (2019).

masked by evolution of the other blended lines in order to go unnoticed. The situation is reversed for Si IV, which is blended with lines of longer wavelength than itself. The lack of velocity shift strongly suggests no blueshifting of Si IV, with the only alternative being a simultaneous equal shifting of O IV coupled with a reversal in line strength balance.

Previous studies have already pointed at the lack of a strong Baldwin effect and blueshifts of Si IV (Osmer et al. 1994; Richards et al. 2002). If the similar lack of blueshift at high-z is confirmed, this would imply that C IV is seen to blueshift with respect to both low- and intermediate-ionisation lines, suggesting a different origin of the line within the BLR.

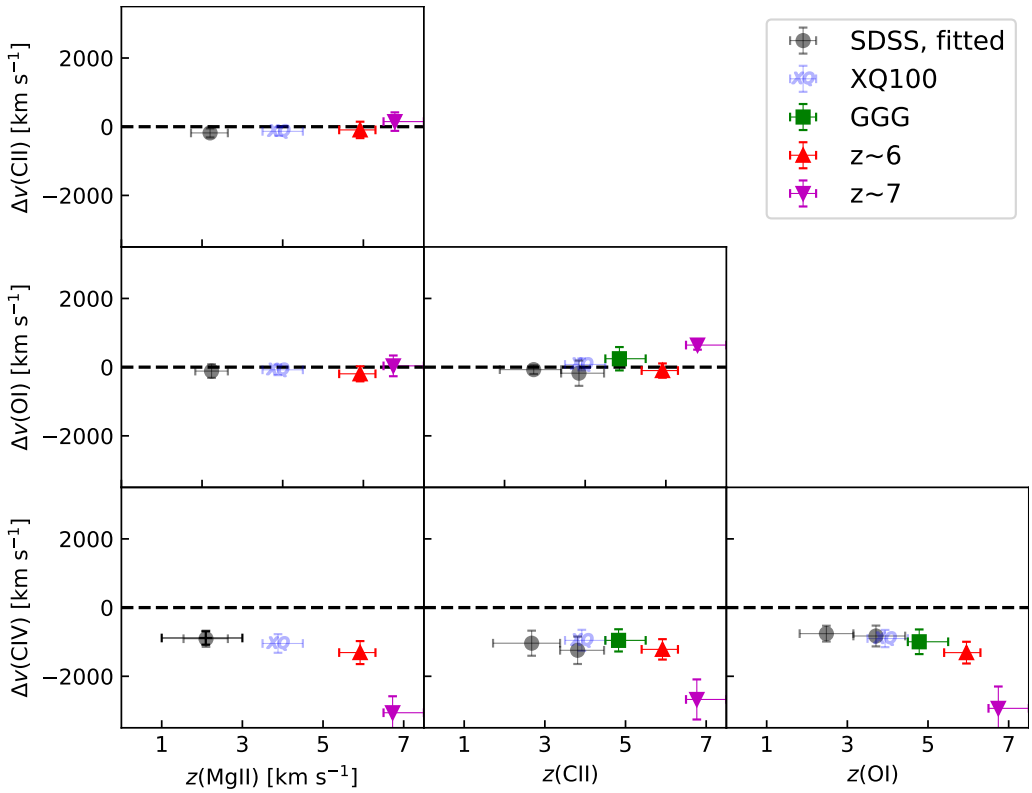

Figure 5.12: Velocity shifts of rest-frame UV BELs across redshift. The errors are computed by bootstrapping using samples size of the high-redshift sample. The O I, C II and Mg II BELs do not blueshift with respect to one another at all redshifts. However, the C IV line is increasingly more blueshifted with respect to these lines at high redshift.

5.3.3 Comparison to previous works

The most striking result of this analysis is the rapid evolution in the velocity blueshift of C IV in the restricted redshift range $5 < z < 7$ depicted in Figure 5.13. Possible evolutionary trends in velocity shifts, specifically in the blueshift of C IV with respect to Mg II, have been previously examined by several authors. Shen et al. (2019) found no significant evolution to $z \sim 6$ consistent with our results over $1.5 \lesssim z \lesssim 6$. If we restrict an evolutionary test to luminosity-matched samples drawn from the z6 and SDSS datasets, the probability that these are drawn from two different distributions has an inconclusive p-value of 0.69 (see Figure 5.11). Although the mean velocity offset at $z \simeq 6$ is slightly larger ($\Delta v_{\mathrm{CIV-MgII}} = (-1400 \pm 334)$ km s^{-1}), given the uncertainty due to the small sample size there is no convincing case for evolution.

However, with the addition of the $z \sim 7$ quasars from Mazzucchelli et al. (2017) and Bañados et al. (2018), we recover the larger mean blueshift of C IV found by

these authors. Since we have applied the same method to different samples over the redshift range $1.5 \leq z \leq 7.5$, we argue this evolution cannot be caused by resolution or other instrumental effects. Our method also allows us to extend the analysis by substituting O I or C II for Mg II. Given the similarity of the trends so revealed, we can aggregate over the low-ionisation lines to construct a C IV-low-ionisation velocity shift versus redshift trend that uses all datasets (Figure 5.13). When aggregating the low-ionisation species, we first construct a common 'systemic' redshift from the mean of all available low-ionisation line-based redshifts. We then derive the C IV blueshift and compute errors by bootstrapping as before. The sharp increase in the C IV blueshift between $z \sim 5 - 7$ is evident. We also show that the recently published C IV-Mg II velocity shifts for individual $z \gtrsim 6.5$ quasars are in agreement with the increasing mean C IV blueshift trend (Wang et al. 2019; Pons et al. 2019; Reed et al. 2019).

There has been some debate on the possibility that biases in the line-based redshift determination of SDSS quasars might remove the most blueshifted C IV lines. DR7Q derives C IV blueshifts up to a maximal value of 5000 km s^{-1}, making the comparison with some $z \sim 7$ quasars impossible. Coatman et al. (2016) showed that using a definition of the line center based on the peak emission or line centroid was strongly affecting the SDSS distribution of C IV blueshifts, potentially alleviating some concerns raised with the first reports of large blueshifts in early quasars (De Rosa et al. 2014). The recent results of Reed et al. (2019) showcase the use of a currently unpublished Independent Component Analysis (ICA) determination of the SDSS C IV blueshifts that identifies multiple extreme blueshifts > 4000 km s^{-1} in SDSS. Despite both studies extending the range of existing SDSS C IV blueshifts, the mean shift and the abundance of C IV lines with negative blueshifts, i.e. redshifted with respect to systemic, still does not match that for $z \sim 7$ quasars. We have tested this by independently measuring the line-base redshifts of C IV and different low-ionisation BELs in about a hundred luminosity-matched SDSS quasars. This result suggests that improvements in quasar template fitting will most likely not result in a dramatic increase of the average C IV blueshift at low redshift. Nonethe-

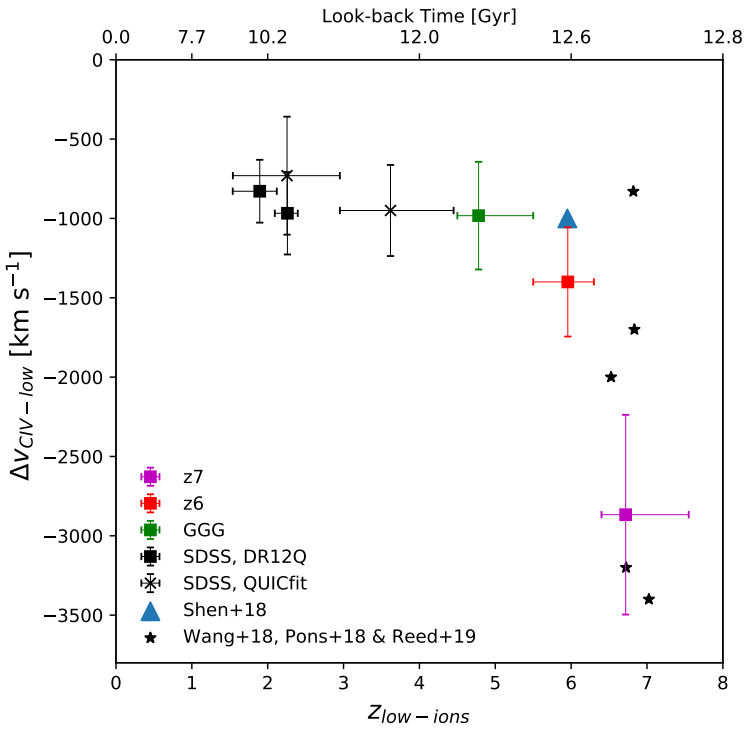

Figure 5.13: Velocity shifts of C IV with respect to the grouped low-ionisation (O I, C II, Mg II) rest-frame UV quasar lines over redshift. The errors are computed by bootstrapping over samples with the size of the smallest sample (z6) to make the errors comparable. The apparent drift in the blueshifts of the SDSS fitted subsample is due to the reference low-ionisation lines which include only Mg II at low-redshift, and then gradually move over to O Iand C II at high-redshift. The $\sim$ 500 km s $^{-1}$ velocity shift is consistent with the small relative shifts of these low-ionisation lines (see Figure 5.12). We also add the reported values of $\Delta v_{\mathrm{CIV-MgII}}$ of the Shen et al. (2019 blue triangles) sample and recent individual high-redshift quasars of similar luminosities (black stars, Wang et al. 2019; Pons et al. 2019; Reed et al. 2019).

less, these advances will enable the selection and study of low redshift quasars with extreme C IV blueshifts, which we argue below will be a crucial point to further our understanding of reionisation quasars.

5.4 Interpreting the high-redshift increased mean C IV blueshift

The most intriguing aspect of the trend in Figure 5.13 is that the inferred evolution occurs in less than 1 Gyr. At low redshift, discussion of the nature of C IV blueshifts has centered around radiation-driven winds originating in the central regions, or the BLR, of SMBH accretion discs. The broadening of quasar UV emission lines was first modelled through primarily planar outflows, in which material is stripped from

the accretion disc and accelerated outwards by X-ray radiation and/or resonant line radiation pressure from the inner regions (e.g. Krolik & Begelman 1986; Murray & Chiang 1995). Based on the singular behaviour of the C IV BEL, Wills et al. (1993) first suggested a possible separate origin of the line in a hotter, polar wind component (see also Denney 2012). Later models showed how gas could be entrained at a steeper angle from more central regions (e.g. via increased magnetisation of the outflows, Proga et al. 2000; Elvis 2000) and even give rise to outflows which are significantly polar by entraining gas from a dusty torus, rather than a thin accretion disc (e.g. Gallagher et al. 2015). Early results from sub-millimeter imaging have supported the possibility of significantly perpendicular winds, with detections of elongated dusty emission around AGN in Seyfert galaxies being preferentially *parallel* to their ionisation cones, rather than perpendicular (López-Gonzaga & Jaffe 2016 and references therein). If low-ionisation and high-ionisation broad lines do indeed arise through such related but different processes along different axis, with higher-ionisation outflows being more polar, it is possible to conceive scenarios in which one evolves but not the other.

The most commonly invoked explanation for the fast evolution of C IV blueshifts at early times involves (unusually) strong BLR outflows or winds to explain (extremely) blueshifted C IV emission (e.g. Richards et al. 2011; De Rosa et al. 2014; Mazzucchelli et al. 2017). This claim is supported by the widespread presence at $z \sim 7$ of extreme blueshifts with $\Delta v \gtrsim 3000$ km s $^{-1}$ which are exceedingly rare at $z < 5$. However, it is unclear what mechanism could power such an increase in wind speed in early quasars, or whether this can account for the whole effect. Such extreme blueshifts are found to be quite common in some lower-redshift populations such as the WISE/SDSS selected hyper-luminous quasars (WISSH, Bischetti et al. 2017; Vietri et al. 2018) at $z \sim 2-4$, and hence are not exclusive to $z \gtrsim 7$ quasars. Specifically, WISSH quasars with weaker O III λ5007 Å exhibit larger blueshifts and lower X-ray to optical luminosity ratio, which are also found in some high-redshift quasars (Bañados et al. 2018), and this trend seems to persist even in moderately bright SDSS quasars (Coatman et al. 2019). This suggests that mechanisms which

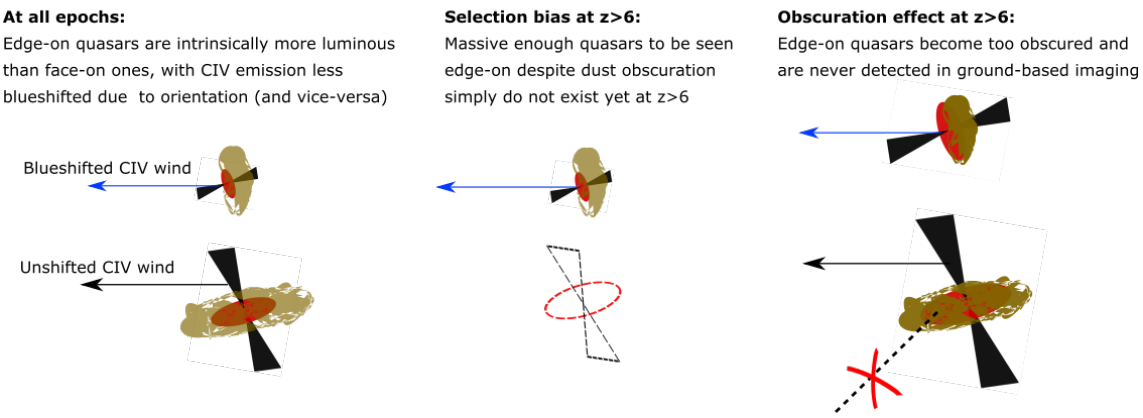

Figure 5.14: Illustration of the proposed interpretations of the increased blueshift of the C IV broad emission line in high-redshift quasar samples (for more details, see further Section 5.4.1 and 5.4.2).

increase C IV blueshift are still in place at much later times, and even dominant in some populations. Moreover, anomalously large C IV blueshifts at early times do not seem to correlate with higher accretion rates – as one might perhaps expect if they were caused primarily by anomalously strong winds. Mazzucchelli et al. (2017) reported that their $z \gtrsim 6.5$ sample had high accretion rates near the Eddington limit and their SDSS luminosity-matched sample did as well. Finally, a wind-only model would imply that some cases of no blueshift or even redshifted lines should always occur due to orientation with respect to the observer, unless the wind is increasingly spherical. In contrast, there is a total absence of redshifted C IV emission lines in luminous quasars beyond $z \gtrsim 6$. If increased wind speeds do power the increase in C IV blueshifts, they are then most likely accompanied by a change in morphology. This conclusion either suggests a fundamental change in the properties of quasars as they emerge from the reionisation era or perhaps some selection effect which would cause quasars with non-blueshifted C IV lines to be absent at $z \gtrsim 6$ despite our attempts to create luminosity-matched samples. We now outline two mechanisms, illustrated in Figure 5.14, which could separately or concurrently lead to an increase in the average C IV blueshift in high-redshift quasar samples.

5.4.1 Orientation selection bias

In a simple 'opaque torus' model of quasars (Denney 2012 and references therein), C IV radiation originates either in the broad line region or a surrounding intermediate line region and launches fast winds perpendicular to the plane of the accretion disk (or at least significantly more polar than low-ionisation winds). If the accretion disk is viewed edge-on where the line of sight is opaque, this will result in a broad C IV emission with no blueshift and a dimmed UV quasar luminosity. Viewed increasingly face-on, the C IV line will be blueshifted with a shape dependent on the wind velocity profile. This geometric model explains many observational features of broad quasar emission lines, such as the consistent, but moderate average blueshifts at low redshift.

Assuming that high-redshift quasars are obscured to the same extent as those at lower redshifts, the enhanced C IV velocity offsets would then likely be related to the relative youth of quasars at the end of the reionisation era. Although the $z > 6.5$ quasars in our sample have already massive ($\sim 10^9 M_\odot$) black holes accreting at nearly the Eddington rate (De Rosa et al. 2014; Mazzucchelli et al. 2017; Bañados et al. 2018), the UV luminosity of quasars would be expected to be attenuated when they are viewed edge-on, which may lead to them not being detectable at our luminosity threshold at $z \gtrsim 6.5$. In this picture, quasars with masses and accretion rates large enough to be UV-bright despite being edge-on might not have yet assembled as early in cosmic time. Sampling the top of the UV luminosity distribution at any cosmic time would then unwittingly bias observations towards face-on objects with large C IV blueshifts.

If this selection bias were the only effect at work, quasar continuum reconstruction models which successfully capture the variety of low-redshift quasars should be expected to work similarly well at high-redshift. Indeed, the physical origin of the C IV blueshift – outflows and orientation – would be intrinsically the same across cosmic time. One might expect a range of related observational consequences, for example:

- The most extremely blueshifted objects of a given luminosity – being face-on – should present roughly the same $\Delta v_{\rm CIV}$ at all redshifts. In our study, the maximum blueshift in the $z \sim 7$ sample is 2000 km s^{-1} greater than our DR12Q luminosity-matched sample extremum (Figure 9). As our samples are relatively small, it is possible that larger datasets of high-redshift quasars and new fitting techniques of low-redshift quasars could still change the values of the most extreme blueshifts.

- The shape and strength of the highly blueshifted C IV emission lines should be the same across redshift, since the evolution of the mean shift would be due only to the disappearance of quasars with non-blueshifted C IV. We present in Figure 5.15 the mean C IV line profiles of our quasars at different redshifts, stacked according to the blueshift of the C IV emission compared to low-ionisation lines. It can be seen that the higher-redshift quasars show a tentatively attenuated profile and smaller equivalent widths when C IV is not blueshifted, as reported in previous studies (Mazzucchelli et al. 2017; Shen et al. 2019).

- The effect should vanish for fainter high-redshift quasars. Perhaps the best test of this 'selection bias' would be to examine the C IV velocity shift at early times in less luminous objects. Such a strategy is obviously very challenging. However, the discovery of a faint $z \sim 7$ quasar with a significant C IV blueshift ($\sim$ 2500 km s^{-1}) in the Subaru High-z Exploration of Low-Luminosity Quasars survey (SHELLQS) provides marginal evidence against this interpretation (Matsuoka et al. 2019) while a recent small sample of six relatively faint quasars at $z \sim 6.3$ show mixed results including potentially the first C IV redshift with respect to Mg II at $z > 6$ (Onoue et al. 2019). Clearly a larger sample of fainter, possibly lensed quasars above $z \gtrsim 6$ can test this hypothesis.

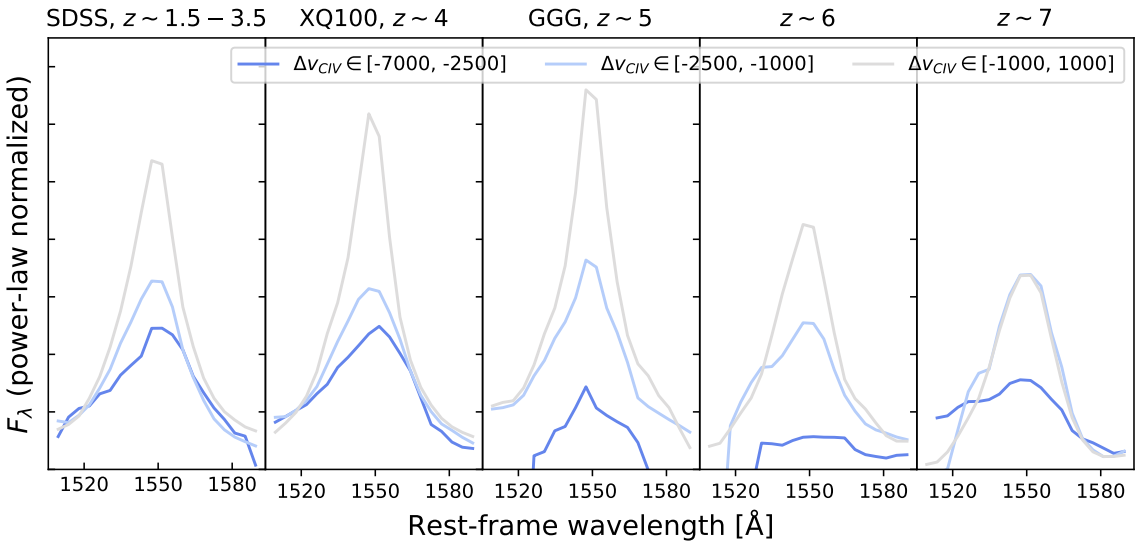

Figure 5.15: Profiles of the C IV BEL for quasars at $1.5 \lesssim z \lesssim 7.5$, stacked according by the C IV blueshift (gray: low blueshifts (< 1000 km s $^{-1}$), light blue: moderate (1000 km s^{-1} $\leq \Delta v_{\rm CIV} \leq$ 2500 km s^{-1}), dark blue: extreme blueshifts (> 2500 km s $^{-1}$) from low-ionisation lines (O I, C II, Mg II). When multiple low-ionisation lines are detected, we take the average redshift to compute the relative blueshift of C IV. The evolution of the blueshifts of C IV is matched by the evolution of the profiles. The emission is weaker at high-redshift, and the change is more prominent for the non-blueshifted objects. We attribute this to increased obscuration at high-redshift as the non-blueshifted C IV are originating in quasars seen edge-on in our interpretation (see Section 5.4).

5.4.2 Increased obscuration

A competing or complementary alternative is that dust obscuration is more intense in early quasars. In this case, those quasars viewed edge-on will again be (even more) strongly de-selected at high-redshift, increasing the average blueshift of the C IV line. Treister et al. (2011) reached the conclusion that black hole accretion is mostly obscured in the Early Universe by looking at the X-ray emission at $z \sim 6 - 8$ attributed to early quasars. Trebitsch et al. (2019) showed that increased quasar obscuration at high-redshift could perhaps be due to chaotic feeding and instabilities in the less mature host galaxies. The same authors also showed that early quasars accreting at higher Eddington ratios are preferentially surrounded by dense, obscuring gas and are only being visible with $M_{1450} < -23$ about 4% of the time. Given that the highest redshift quasars must have accreted at nearly Eddington rate for most of their life (e.g. Bañados et al. 2018), it is thus likely that they are more obscured than the lower-redshift ones. By selecting the most

UV-bright objects at high-redshift we would then only find quasars which have both very high accretion rates *and* are preferentially orientated, with potentially very narrow visibility channels. We speculate that chaotic feeding of the central black hole would affect the BLR and thus the BELs. Models and templates extracted from low redshift quasars would thus be expected to fare fairly poorly on high-redshift objects.

This second scenario might also explain some other intriguing recent finds in reionisation era studies. The small near-zones of some high-redshift quasars (Eilers et al. 2017, 2018b) could be due to intermittent obscuration of the quasar, blocking LyC photons up to $z \sim 6$, where the near-zone would start to grow steadily and C IV blueshift normalize. Depending on the type of obscuration at work, highly obscured and thus UV-faint quasars might in fact already be known as faint AGNs hosted in high-redshift galaxies showing N V emission lines (e.g. Stark et al. 2017).

Finally, both interpretations have potential consequences for the contribution of quasars to hydrogen reionisation (e.g. Madau & Haardt 2015). In a 'geometrical selection effect' scenario, only quasars which are 'sufficiently face-on' can be currently detected at high redshift. Indeed, $\sim 50\%$ of z7 quasars have C IV blueshifts ≥ 2500 km/s, while those make up $< 15\%$ (at 2σ) of the DR12Q sample. Since quasar ionising radiation output is calculated over all angles, this would imply that an average high-redshift quasar in our luminosity-matched sample contributes *less* to the IGM photon budget than its low-redshift counterpart. We deduce this because edge-on, non-C IV-blueshifted quasars on average are expected to have larger masses and larger accretion rates in an orientation-only model – and those are the ones missing at early times. By contrast, in a scenario where early quasars are more obscured than their low-redshift counterparts, it is likely that we are currently observationally missing a large fraction. We would expect to find an increased number of obscured early AGN – which are perhaps contributing to the bright end of the galaxy UV luminosity function (Ono et al. 2018) or to the AGN X-ray luminosity function, which should be less obscured (Giallongo et al. 2015). Using the rough value of $\sim 4\%$ UV-bright visibility fraction from Trebitsch et al. (2019), and if obscuration

is mostly geometric, we could be underestimating the number density of early massive quasars by factors of $3-4$. This deficit could be somewhat balanced by less ionising emission escaping through smaller channels. In any case, the C IV blueshift conundrum has far-reaching consequences on the role of quasars in reionisation.

5.5 A search for high-redshift lensed quasars

I now turn to the second aspect of this chapter, namely the question of whether some fraction of the high redshift quasar population is missed in current surveys. The observed evolution of C IV blueshifts in high-redshift quasars suggests that current samples of $z \gtrsim 5.5$ quasars could be incomplete and biased towards the brightest objects with the most actively accreting black holes. The main obstacle to the study of large samples of faint quasars is the observing time needed to confirm them and acquire a medium SNR spectrum.

For example, confirming a $m_z(\mathrm{AB}) \simeq 20$ quasar can take $10-20$ minutes (without overheads) on a 4m telescope. Pushing the quasar luminosity function (QLF) a mag or two lower is therefore difficult, especially if one takes into account that the spectroscopic confirmation rate of $z > 6$ quasar candidates is usually low ($20-30\%$) due to confusion with cool stars (e.g. Wang et al. 2019; Matsuoka et al. 2019). More sophisticated methods using SED fitting (Reed et al. 2017, 2019) or additional imaging of high-ranked dropouts before spectroscopic follow-up (Venemans et al. 2013, 2015) have much higher confirmation rates at the expense of requiring more imaging data. As telescope time is finite and shared between fields, getting fainter quasars confirmed and studied with current facilities might take several years.

Gravitational lensing can enable us to study fainter quasars without having to wait for 30-40m class telescopes. Moreover, the abundance of lensed quasars is directly linked to the number density of quasars and can thus independently constrain the $z > 5.5$ QLF (Pacucci & Loeb 2019). Since the faint end slope of the quasar LF is so poorly known, Pacucci & Loeb (2019) estimate that the missing fraction of $z \sim 6$ quasars could be anything from 5% to a factor 10. Gravitational lensing is

a well-used technique for probing the faint end of the galaxy luminosity function at high redshift (Bouwens et al. 2017; Atek et al. 2018) but, until recently, largely overlooked in quasar studies. In fact, the first gravitationally-lensed quasar at $z > 5$ was only discovered serendipitously and has a very high magnification ($\mu \simeq 50$, Fan et al. 2019). Crucially, the photometric selection techniques used for high redshift quasar searches are unlikely to find lensed quasars by design because of the contaminating presence of the foreground lens. As for galaxies, $z > 6$ quasars are selected on the basis of a sharp flux discontinuity bluewards of Lyman α emission, and in practice non-detection in the g, r bands. The presence of a contaminating foreground lens would induce a signal in the g, r bands and hence lead to the discarding of such a source using normal selection methods.

5.5.1 Selecting lensed quasar candidates

I started a programme to select and confirm lensed quasars. I aimed to address the hypothesis that a significant fraction of sub-luminous lensed quasars are present in current photometric surveys but have been overlooked due to the stringent dropout criteria. Standard colour cuts to select high-redshift quasar require strict non-detections in the (g,r) bands. As discussed above, the assumption of this work is that these cuts exclude lensed quasars because a foreground galaxy is expected to contaminate the bluer bands. I therefore applied standard colour-cuts for $z > 6$ quasars in the izY bands but disregarded the gr non-detection criteria (e.g. Venemans et al. 2013; Reed et al. 2015, 2017).

Specifically, the following colour-cuts were applied to the Dark Energy Survey-Vista Hemisphere Survey[1] catalog overlap: $z < 21.5, \sigma_z < 0.1, i - z > 1.694, z - Y < 0.5, Y - J < 1.0$, where i, z, Y, J are the "MAG_AUTO" magnitudes measured by *SExtractor* (Bertin & Arnouts 1996) in the DES (i, z, Y) and VHS (J) catalogues and σ_z the error on the z band magnitude. The cuts resulted in 18928 dropout candidates (Table 5.3) with 00h < RA < 06h (the search was limited to two-thirds of DES-VHS such that resulting candidates were all observable within a single observing run. The

[1]VHS, ESO Programme ID: 79.A-2010, PI R. McMahon

Table 5.3: Selection criteria and number of lensed quasar candidates in DES-VHS-WISE. The *grizY* magnitudes are taken from the DES DR1 catalogue (Abbott et al. 2018), the *J*-band magnitude from VHS DR5 and W1,W2 magnitudes from the AllWISE data release.

Selection	$N_{\rm Objects}$
$z < 21.5 \wedge \sigma_z < 0.1$	
$i - z > 1.694$	
$z - Y < 0.5$	
$Y - J < 1.0$	
00h < RA < 06h	18928
$g < 24 \vee r < 24$	654
Visual inspection	92
$g > 3\sigma \vee r > 3\sigma$	33
W1> $3\sigma\wedge$ W2> 3σ	26
WISE uncontaminated	21

nominal 5σ detection limits for DES DR1 (Abbott et al. 2018 MAG_AUTO) are $g(5\sigma) = 24.27$, $r(5\sigma) = 23.85$, and the 95% detection completeness is reached at $g = 23.72, r = 23.35$. I retained candidates only if they had at least a detection in either bands which is defined for simplicity as $g < 24$ or $r < 24$, resulting in 654 candidates.

These i-dropouts were inspected visually to establish a list of candidates. Most dropouts were spurious artifacts or moving asteroids. The final list contains 33 candidates all showing a break between the i and z bands, and a companion object offset by 1" − 3", which is brighter in the *gri* bands. To avoid the obvious contamination by low redshift objects that results in relaxing the g, r non-detections, only promising systems where a putative foreground lens is partially offset from the quasar but still contaminates the optical photometry were selected (e.g. Figure 5.16). The list was reduced to 21 to include only candidates with a clear detection in the WISE bands and no contamination from a nearby source. Table 5.3 summarises the selection procedure described above.

It is important to note that the visual inspection step is only feasible because the 5σ depth in the VHS band is $J = 21.41$ and only 5% of the DES i-dropouts have a J band detection [2]. Furthermore, the $z - Y$ cut (Table 5.3) requires a detection

[2] http://archive.eso.org/cms/eso-archive-news/new-data-release-dr5-of-the-eso-public-survey-vista-hemisphere-survey.html

in the DES Y band, which is shallow compared to the $griz$ bands ($Y(5\sigma) = 22.2$). For a example a quasar satisfying the cuts above with $z = 21.5, Y = 22, J = 22.5$ would not be selected.

Inspecting 8 images ($grizYJHK_s$) for each 654 candidates is very time-consuming. Moreover, the visual inspection must include single-epoch images to remove moving objects such as asteroids. In the future, additional data in the infrared bands will make visual inspection of all $i-z, z-Y$ dropouts impractical. I therefore set to develop a new machine learning selection technique of quasars to reduce the need for manual inspection of dropouts, and bypass the strict colour-cut selection which prevents selecting lensed quasar candidates. Briefly, I train a deep neural network to recognise quasars directly from DES 5-band images ($grizY$). The network was trained by injecting thousands of mock quasars randomly in DES tiles at RA$>$ 20h and was tasked to output the location of quasars and a confidence score for each detection. The score threshold adopted was chosen manually to maximise the purity of the final sample (rather than the completeness). The network can search an entire DES tile in a few minutes and only 10% of the best candidates returned were deemed to be artifacts after visual inspection. The full description of the architecture, training and hyperparameter fine-tuning of the neural netwrok is beyond the scope of this Chapter focused on high-redshift quasar observations. I therefore refer the interested reader to Appendix C.4 for more details.

5.5.2 Two confirmed $z > 6$ quasars selected with machine learning

The lensed quasar candidates were observed during a four-night observing run on EFOSC2/NTT at La Silla Observatory (ESO proposal ID : 0104.A-0662(A), PI: Meyer). The sky was clear throughout the four nights, but the seeing conditions varied rapidly through the nights from $\sim 0.8" - 1.8"$. The slit width was changed between $1.0", 1.5"$ and $2.0"$ to match the sky conditions. The first exposure time varied between 20 and 30 minutes depending on the apparent magnitude of the candidate, and additional exposures were taken if the candidate was not rejected as

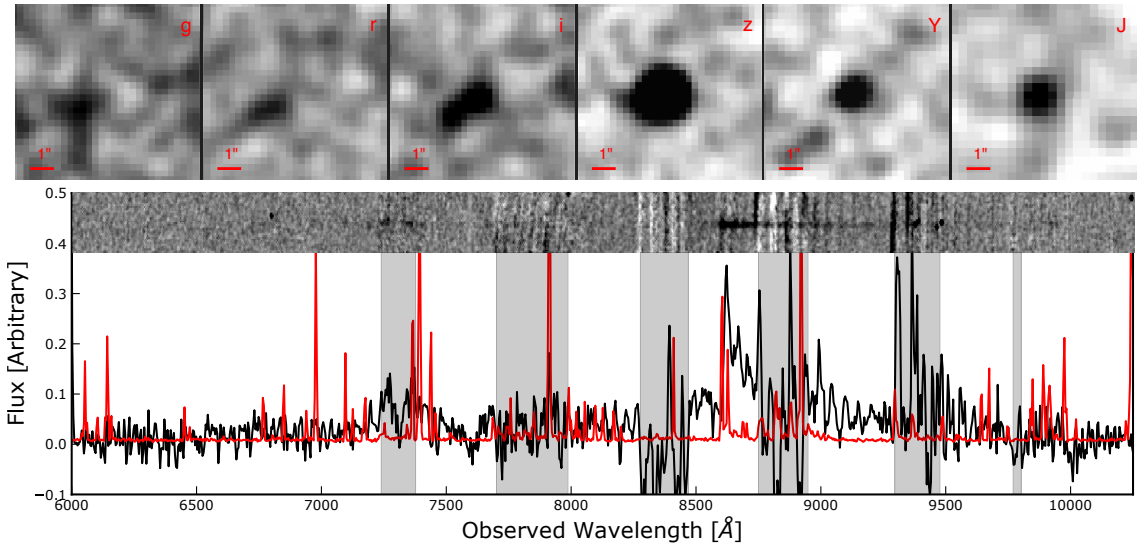

Figure 5.16: **Top panels:** Dark Energy Survey imaging of DESJ0200−1737 in the *grizy* bands, showing the central offset lens galaxy disappearing in the NIR bands, where the quasar emission dominates. Note the blending of the lens and the quasar images in the *iz* bands. **Lower panels:** 2D and 1D EFOSC/NTT confirmation spectrum showing a characteristic $z > 6$ quasar spectrum with a bright Lyman-α line. The 1D spectrum is unfluxed.

a nearby cool dwarf after inspection of the first exposure. 35 objects were observed in total: 7 from the relaxed colour-cuts (e.g. without the *gr* non-detections) and 28 from the machine learning selection (see Table C.2 for the full list). Most of the candidates were confirmed to be cool dwarfs serendipitously aligned with a background galaxy or another cool dwarf. However, two targets were confirmed as quasars: DESJ0200-1737 at $z = 6.09$ (Figure 5.16) and DESJ0252-0237 at $z = 6.16$ (Figure 5.17). These two quasars both came from the machine learning list, not the relaxed colour-cut selection.

Several conclusions can be drawn from this quasar selection experiment. First of all, removing the *gr* non-detections criteria does not introduce low-redshift galaxy contaminants into the candidate samples. As for quasars candidates selected with *gr* non-detections, the contaminants are still cool stars. If, in the future, another criterion can be used to separate cool star from quasars, the *gr* non-detection criteria could be dropped to avoid removing lensed quasars from the candidate selection. Secondly, this experiment shows that machine learning can reduce heavily the need for visual inspection and select $i - z$ dropouts directly from the imaging without

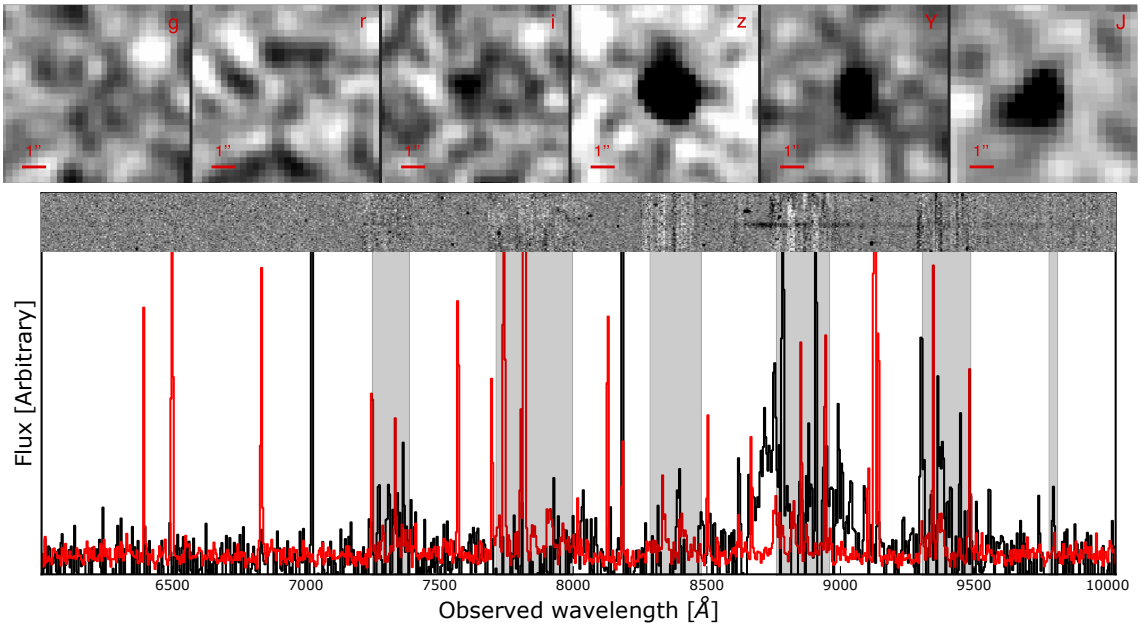

Figure 5.17: **Top panels:** Dark Energy Survey imaging of DESJ0252−0237 in the *grizy* bands. **Lower panels:** 2D and 1D EFOSC/NTT confirmation spectrum showing a characteristic $z > 6$ quasar spectrum with a bright Lyman-α line. The 1D spectrum is unfluxed.

utilising photometric catalogs. This is especially important since the magnitudes method (for example *MAG_AUTO* and 2" aperture forced photometry) can differ by $\sim 0.1 - 0.2$ mag depending on the method, significantly affecting the number of i-z dropouts. A neural network can bypass these issues and derive an homogeneous sample of candidates directly from the imaging. The high rate of cool dwarfs in the selected sample could be probably reduced by introducing cool dwarfs as well as quasars in the training dataset. Finally, this observing run confirmed what is likely a lensed quasar, the second only at $z > 5$.

5.5.3 DESJ0200-1737: A lensed quasar at $z \sim 6.09$?

J0200−1737 is a quasar of particular interest. It was confirmed to have a spectroscopic redshift of $z = 6.09$, and is seen at an angular separation $\Delta\theta = 1.1'' \pm 0.2''$ from a faint foreground galaxy (Figure 5.16). The observed angular separation between the quasar and the galaxy is measured between the centroids of *gr* and *Y* bands to avoid broadbands where the objects overlap and contaminate each other. The lens best-fit mass ($M_* \simeq 10^{9.4} M_\odot$) and redshift ($z = 0.90^{+1.09}_{-0.25}$) were derived

from SED fitting using LePhare (Figure 5.18, left Arnouts et al. 1999; Ilbert et al. 2006). This is the most likely redshift for a galaxy lensing a background $z \sim 6$ quasar (Pacucci & Loeb 2019), reducing the possibility of chance alignment without lensing. Unfortunately, LePhare does not provide an error on the inferred stellar mass as it only fits selected galaxy spectral templates. Other SED fitting codes which can constrain the stellar mass from detailed stellar libraries (Bruzual & Charlot 2003; Eldridge et al. 2017) were also used, but fitting of additional parameters to only three three datapoints (the putative lens is only detected in the *gri* bands) is inconclusive. Assuming a halo mass to stellar mass ratio of ~ 100 (Moster et al. 2010), and a Single Isothermal Sphere (SIS) lens model, the magnification of the primary image of J0200−1737 can be estimated to $\mu_+ \lesssim 10$ (Figure 5.18).

There are two routes to confirming the quasar magnification and thus its intrinsic magnitude. A better lensing model can be constructed with a spectroscopic redshift for the galaxy and a better separation measurement from space-based photometry. However, definitive evidence for lensing would be the detection of a second image, which can only be achieved with *HST*-like angular resolution. Indeed, in this configuration we expect only one demagnified counter-image very close to the galaxy. A suggested strategy is to use narrowband imaging of the Lyman-α line of the quasar to get maximum contrast between the lens and the source quasar. The *HST* exposure time calculator predicts that with only 5660s of exposure time, we would be able to detect a counter-image 20 times fainter than the primary at SNR = 10 in the FR853N filter (ACS/WFC3). Even in the case of a non-detection of the counter-image, the primary would be boosted by $30\% - 60\%$. A significant population of such modestly magnified quasars would still have a large impact on the quasar LF.

I finally show in Figure 5.19 the prospects for constraining the QLF bright-end. Whilst the confirmation of a single, mildly magnified $z \sim 6$ quasar cannot yet rule out existing constraints on the QLF, 10 additional lensed quasars would. An order of magnitude increase in the number of known lensed $z > 6$ quasars is therefore highly desirable but will require significant work to avoid the selection issues dis-

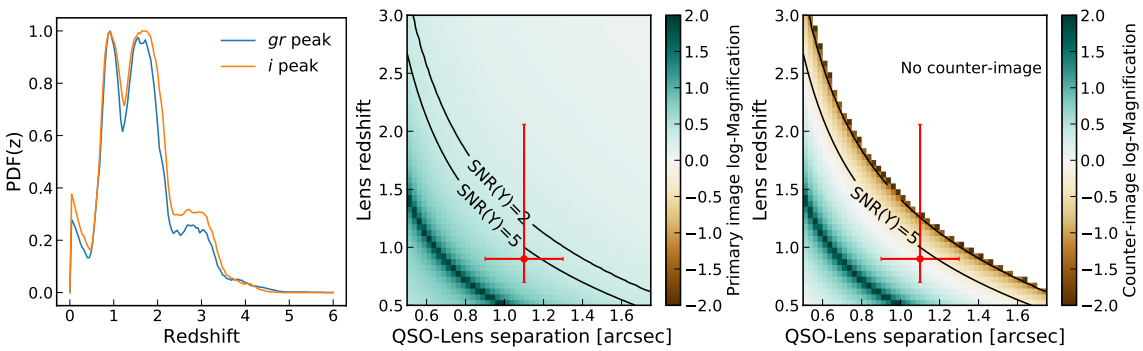

Figure 5.18: **Left:** *LePhare*-PEGASE2 SED redshift posterior for the foreground lensing galaxy of J0200−1737. The two curves show similar results for the forced aperture photometry centered on the *gr* or *i* band peak flux of the foreground galaxy. **Right:** Current constraints on the magnification of J0200−1737. The red crosses show the current 1σ uncertainty from ground-based data, and the black contour lines represent the constraints from the non-detection of J0200−1737 counter-image in the Y band. *Adapted from Pacucci & Loeb (2019)*

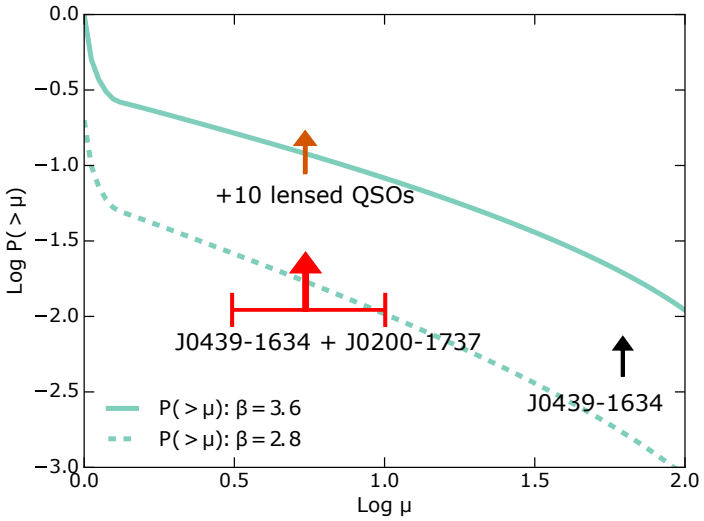

Figure 5.19: Predicted observable fraction of gravitationally-magnified $z > 6$ quasars (with magnification $> \mu$) for two different slopes β of the bright-end of the quasar LF ($\beta = 3.6$ (Yang et al. 2016), $\beta = 2.8$ (Jiang et al. 2016)). The discovery of 10 more lensed quasars in future surveys would rule out a shallow slope of the bright-end of the quasar LF. Adapted from (Pacucci & Loeb 2019)

cussed previously. This work has shown that quasars can be efficiently selected with machine learning without the strong gr non-detections usually required. However, this decreases the purity of the candidate samples to $\sim 5\%$. While this is not catastrophic compared to early searches for high-redshift quasars (e.g. Reed et al. 2015), further work must now focus on discriminating between quasars and cool dwarfs using other criteria that an absence of flux in the $gr(i)$ bands. This challenging step will unlock the scientific potential of statistically significant samples of lensed high-redshift quasars for reionisation and SMBH studies.

This page was intentionally left blank

Chapter 6

Conclusion and future prospects

6.1 Combining individual and statistical measurements of the escape fraction in reionisation-era galaxies

One of the major aims of this thesis was to measure the LyC escape fraction of galaxies in the epoch of reionisation. Two routes were employed: the analysis of a double-peaked Lyman alpha profile in an individual $z > 6$ galaxy and the statistical cross-correlation of galaxies with the IGM opacity probed by high-redshift quasars. I now combine the two different approaches to determine whether reionisation is dominated by luminous or faint sources. I show on Figure 6.1 a visual summary of the escape fraction constraints presented in this Thesis. The figure is very similar to that presented in the conclusion of Chapter 4: the galaxy-2PCCF average escape fractions as a function of the minimum UV luminosity of contributors are presented in blue and orange shades assuming a fixed $\xi_{\rm ion} = 10^{25.5}$. The 2PCCF indicate that, if faint galaxies ($M_{\rm UV} \sim -12$) contribute to reionisation, their escape fraction is $\sim 10\% - 20\%$. However, in a scenario where only luminous objects ($M_{\rm UV} < -21$) contribute, the LAE(LBG)-IGM 2PCCF constrain their average escape fraction to be $\langle f_{\rm esc,LyC} \rangle > 0.69(0.72)$ at 2σ. The C IV-IGM transmission cross-correlation at $z \sim$

5.5 suggests that C IV absorbers are surrounded with slightly more efficient leakers with $f_{\rm esc,LyC} = 0.32^{+0.32}_{-0.12}$ if galaxies as faint as $M_{\rm UV} \sim -12$ contribute. [1]. With the discovery of A370p_z1 (Chapter 2), the average escape fraction derived from four $z \sim 6-7$ luminous ($M_{\rm UV} \sim -22$) double-peaked LAEs is $\langle f_{\rm esc,LyC} \rangle \simeq 0.3 \pm 0.3$. If this modest sample is representative of the population of luminous $z > 6$ galaxies, it would suggest a reionisation history dominated by $M_{\rm UV} < -21$ systems can be rejected at the $\sim 2\sigma$ level by comparing to the LAE/LBG-2PCCF constraints. An ionising photon budget dominated by faint sources is also in agreement with most analytical models of the neutral fraction history, shown on Figure 6.1 in black (see also Section 1.3.4), and the conclusion that additional galaxies are probably needed to contribute to the ionised bubbles of COLA1 and NEPLA4 (Section 2.5.2).

Moreover, the sample of double-peaks is probably biased towards objects with high escape. Detecting a double-peak implies the existence of a large ionised bubble, enhancing the detection rate of efficient LyC leakers with high $f_{\rm esc,LyC}$ (or, alternatively, high $\xi_{\rm ion}$ and a low $f_{\rm esc,LyC}$). A similar effect is potentially present in the cross-correlation analysis. The detection of Lyman-α emission or C IV absorption implies an abundance of ionising photons, preferentially selecting environments with very efficient ionising sources. Considering these two effects, a reionisation process dominated by luminous galaxies is probably more unlikely than the 2σ rejection discussed above.

The double-peak method offers a simple estimate of the escape fraction that is straightforward to interpret. However, the exposure time necessary to secure a double-peak in $M_{\rm UV} \sim -21$ galaxies with Xshooter (~ 6 hours) indicates it is probably impractical to consider using this technique to explore the luminosity-dependent escape fraction, even with 30-40m class telescopes. Accordingly, combining double-peak measures for more luminous and possibly lensed galaxies with the integral constraint from the 2PCCF will remain the most productive way forward.

[1] Unfortunately the escape fraction constraints for the C IV-IGM cross-correlation were only computed for a fixed value of $M_{\rm UV} = -12$ in Chapter 3. It is expected that the constraints follow the shape of the LBG/LAE-2PCCF, with higher escape fraction for a higher minimum luminosity of contributors

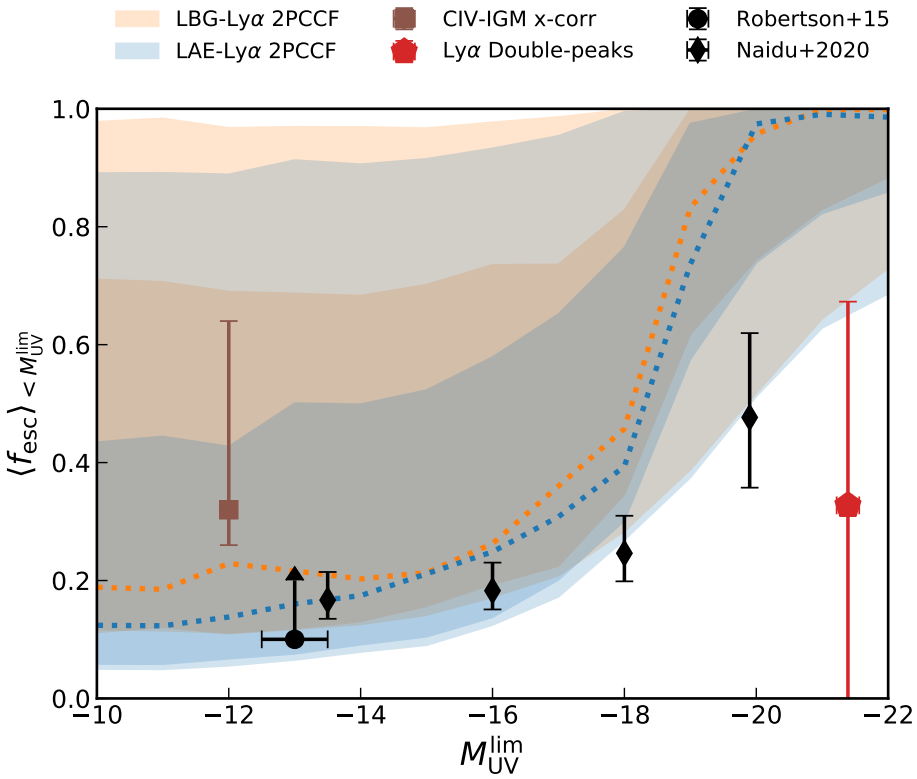

Figure 6.1: Summary of the $z \sim 6$ average escape fraction constraints as a function of the minimum UV luminosity of contributors presented in this thesis: average escape fraction of $z \gtrsim 6$ double-peak LAEs with A370p_z1 (Chapter 2, red pentagon), CIV-IGM transmission cross-correlation (Chapter 3, brown square), LBG/LAE-IGM 2PCCF (Chapter 4). I also show the results of analytical models of the neutral fraction history in black (Robertson et al. 2015; Naidu et al. 2020 circle and diamonds, respectively).

6.2 Evolution of high-redshift quasars and the neutral fraction history

In Chapter 5, I presented the evolution of C IV broad emission line blueshifts as evidence for quasar evolution and/or selection bias. Although my results were based on 17 $z > 6$ quasars, 17 newly-observed quasars of the XQR-30 ESO/VLT Large Program confirm the evolution observed (J.-T. Schindler, private communication). The latest $z > 7$ quasar (J1007+2115, $z = 7.515$ Yang et al. 2020) also follows the trend with a large C IV blueshift $\Delta v_{\rm CIV-MgII} = -3220 \pm 362$ km s^{-1}. This evolution may explain why quasar Lyman-α line reconstruction algorithms, which have proliferated in recent years (e.g. Mortlock et al. 2011; Davies et al. 2018c; Greig et al. 2017, 2019; Ďurovčíková et al. 2020; Reiman et al. 2020; Fathivavsari 2020;

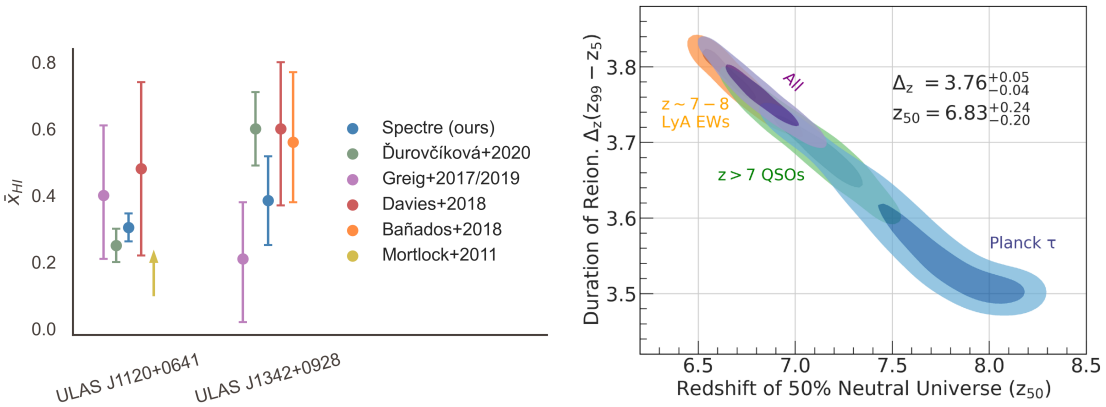

Figure 6.2: **Left:** Neutral fractions inferred from the damping wing of two $z > 7$ quasars with various algorithms. *(Reproduced from Reiman et al. (2020).)* **Right:** Posterior distribution of the mid-point and duration of reionisation for different neutral fraction probes. The quasar damping wing favors a faster and shorter reionisation process than most neutral fraction measurements. *(Reproduced from Naidu et al. (2020).)*

Liu & Bordoloi 2020), give different neutral fraction results (Figure 6.2, left). Any dissimilarity between the training sample (SDSS quasars) and the final application ($z > 7$ quasars) will affect the prediction according to the reconstruction method adopted. This is important in the context of this thesis because the high neutral fractions inferred from quasar damping wing measurements suggest a late and rapid end to reionisation, which could be driven by luminous LyC leakers at $z < 7$ (Figure 6.2, right, see also Greig & Mesinger 2017; Naidu et al. 2020), whereas a faint-galaxies powered ionising budget results in a longer reionisation (Ishigaki et al. 2018; Finkelstein et al. 2019). The evidence presented in Chapter 5, and the subsequent disagreement between quasar continuum reconstruction methods casts doubts on the neutral fraction derived from quasar damping wings. I therefore propose that the tension between a faint galaxy-dominated reionisation scenario, supported by the escape fraction measurements made in this thesis, and a slower reionisation process can be eased if the two current damping wing measurements are incorrect due to quasar evolution.

6.3 Future work

The method of cross-correlating the IGM opacity with metal absorbers (Chapter 3) can be extended to any additional sample of medium-high resolution $z \gtrsim 5.5$ quasar spectra. The XQR-30 ESO/VLT Large Programme has already observed half of its planned 29 $z > 6$ quasars, at a SNR comparable to those presented in Chapter 3. Not only will this double the number of high-redshift spectra available for the C IV-IGM cross-correlation, but it will also quadruple[2] the number of XShooter spectra available for this analysis. This will help address the thorny issue of cosmic variance discussed at length in Chapter 3 and 4. Near-infrared coverage with XShooter is important because Mg II absorbers can be detected. These lower-ionisation absorbers are still relatively numerous at $z \gtrsim 6$ and their numbers do not decline as is the case for C IV. Other absorbers such as O I, C II, Si II will also be detected in sufficient numbers for useful IGM cross-correlations. As the appropriate member of XQR-30, I expect progress in two respects. Firstly, it will improve the average escape fraction constraints by combining the different measurements. Secondly, it will probe the local environments of metal absorbers, and consequently different population of faint LyC leakers. This will highlight potential biases in the C IV-IGM cross-correlation results and maybe give evidence for subpopulations of reionisation era LyC leakers with varying $f_{\rm esc,LyC}$ around different metal absorbers, linking the evolution of the IGM and the CGM during the first billion years.

An obvious next step given the successful outcome of Chapter 4 would be a concerted effort, via a ESO Large Programme, exploiting deep MUSE data in numerous quasar fields to perform galaxy-IGM cross-correlations. Fortunately, a follow-up of XQR-30 with MUSE is beneficial to many areas (quasar environments, CGM-galaxy connection) besides the galaxy-IGM cross-correlation. It will be important to refine the analytical model of the 2PCCF to include luminosity-dependent escape fractions and ionising efficiencies. This is non-trivial because the enhanced photoionisation rate due to clustering is computed from the LyC-output-weighted correlation of

[2]The sample used in Chapter 3 contained only a dozen of XShooter spectra, the rest being ESI and HIRES spectra.

galaxies. Throughout this thesis, the weighting could be reduced to a luminosity-weighting since the escape fraction and ionising efficiencies were considered similar for all galaxies. Removing this assumption to make the model more flexible therefore demands additional theoretical work to rewrite some basic equations of the current model.

In the meantime, additional efforts on the numerical modelling of the IGM-galaxy cross-correlation will deepen our understanding of this new probe of reionisation.

The search for further examples of double-peaked Lyman-alpha emitters in the reionisation era is a topic with promising potential. A sensitive instrument with high-resolution (to resolve small peak separations) is required. At low- and high-redshift, several studies have shown that high EW([OIII]+Hβ) is correlated with higher ionising outputs and higher Lyman-α confirmation rates. A370p_z1 is the only double-peak with a $3.6-4.5\mu$m excess indicative of a likely intense [O III]+Hβ nebular contribution consistent with the low-redshift trend. I have therefore applied with collaborators for XShooter/VLT and already obtained GTC/MEGARA time to find double-peaks in further candidate [O III]+Hβ emitters. Although this approach is more likely to result in successful detections, it may result in samples biased high in terms of $f_{\rm esc,LyC}$ and/or $\xi_{\rm ion}$. Blind surveys for double-peaked Lyman-α must be conducted once the method is firmly established. Simulations are starting to produce double-peaks in cosmological volumes (e.g. Gronke et al. 2020), and statistical samples of $z > 6$ double-peaks are needed to make any comparison between observations and simulations fruitful.

Finally, the machine learning search for missing or lensed quasars outlined in Section 5 (see further Appendix C.4) is still at an elementary stage. Despite its early successes (two detections, one probably lensed), the YOLO network can be improved to distinguish between quasars and brown dwarfs, reducing the risk of confusion. I will also train it to predict various classes of quasars (such as lensed/unlensed, or $z > 5.8/z > 6.5$ quasars) to help prioritising targets before an observing run. The main obstacle here is the construction of a realistic training set: lensed $z > 6$ quasars remain very rare and their morphological configurations are challenging to

model. Additionally, the network was trained with an unphysically high density of mock quasars in the training images that can bias the network to predict too many quasars in real images. The network must learn to analyse thousands of images before detecting a quasar, and not flag an object in every 20" × 20" cutout. This can be easily corrected, but at the cost of a much longer training time and significant computer resources. Finally, other neural networks architectures can be adapted for this task, and complement the YOLO network. After all, image recognition and segmentation is one of the most thriving field of machine learning, generating more advanced models than what we will possibly have the time to adapt for astronomical purposes. I therefore predict that with the expanding datasets from all-sky surveys (LSST, *Euclid*, *Nancy Grace Roman Space Telescope*) and the concurrent increase in computer power, machine learning will be instrumental in detecting large samples of lensed quasars in the reionisation era.

6.4 Looking forward to the James Webb Space Telescope

The much awaited launch of the *James Webb Space Telescope* will undoubtedly revolutionise early galaxy evolution and reionisation studies as well as other fields. As a space observatory, *JWST* offers unrestricted access across the infrared spectrum which can be leveraged to detect Lyman alpha at $z > 6$ and rest-frame optical lines for $z > 5$. Nonetheless, the limited duration of the mission ($5 + 5$ years), the important fraction of GTO time, the inevitable competition for observing time and the large overheads limit the scale and scope of projects which must be planned accordingly. In this section, I outline possible plans for what *JWST* might accomplish in the near future to extend the projects presented in this thesis.

Double-peaked Lyman-α emitters such as A370p_z1 are prime candidates for *JWST* follow-up. Their escape fractions have been measured via the peak velocity separation, but the physical mechanisms involved in high escape fractions are still unknown. As the galaxy-IGM cross-correlation results suggest the average escape fraction from star-forming galaxies was higher in the first billion years, it is likely

that high escape fractions are not exclusive to the few double-peaked cases so far examined. Collectively these double-peaked examples provide an ideal sample to understand what governs the range of escape fractions at high-redshift. I have argued in Section 2.5.2 that high [OIII] emitters could be efficient leakers. Using the best-fit SED of A370p_z1, *JWST* can readily detect Hβ, [OIII], Hα at a SN$\simeq 4, 7, 4$ in $\sim$ 6h with NIRSpec in long slit mode. Besides determining whether the escape fraction in double-peaks correlate with the [OIII] strength, this programme will measure their ionising efficiencies precisely with Hβ and Hα (e.g. Shivaei et al. 2018). The other double-peaks are all more luminous and at lower redshift and hence their lines should be detected in a similar or smaller amount of time. A systematic exploration of the properties of reionisation-era LyC leakers is therefore whithin the reach of a medium *JWST* program. In all likelihood, some of these objects will be observed as part of a larger follow-up of the deep fields they were selected from, and only a small programme will be needed to provide the remaining observations necessary to conduct a comparison of the double peaks properties.

I have outlined above how I will realise new cross-correlations of metals with the IGM transmission using a new sample of $z > 6$ quasars, and how the galaxy-IGM cross-correlations can be measured with more deep MUSE observations of quasar fields. Nonetheless, *JWST* offers a complementary approach which addresses several caveats of the ground-based data. First of all, the selection of galaxies based on [O III] emission instead of Lyman-α or C IV absorption will ensure the galaxies are not tracing biased environments. Indeed, the strong decline at $z > 5.5$ of the Lyman-α fraction is attributed to the increasingly neutral gas (Section 1.4). Similarly, the simultaneous increase of O I,C II absorbers and decrease of C IV at $z \sim 6$ suggest metal absorbers trace specific sites in the reionising IGM. Secondly, the ground-based data only probed small transverse scales (< 0.1 pMpc for MUSE, < 4 pMpc for DEIMOS with high incompleteness). With the Wide-Field Slitless Spectroscopy (WFSS) of the NIRCam instrument onboard JWST, two nearby $2.2' \times 2.2'$ fields can be observed at once. If the background quasar is placed in the corner of one of the fields, we can probe distances up to $\sim$ 4 pMpc with high completeness. Finally,

the NIR coverage means we can push the measurement to $6 < z < 6.5$ where the detection of Lyman-α is hampered by sky lines in ground-based data.

Using the *JWST* ETC and following Maseda et al. (2018); Shen et al. (2020) to estimate the [OIII] luminosity function at $z \sim 6$, I evaluate that 11 galaxies suitable for the 2PCCF can be detected in 30min of NIRCam WFSS, for a total of 10 fields in a single GO1 small proposal. Each field would have more galaxy-transmission spike pairs than any presented in Chapter and 3 and 4, representing between one and two orders of magnitude in the number of galaxy-spike pairs and a significant reduction of the cosmic variance. These predictions do not include galaxies outside of the Lyman-α forest redshift range which will have high legacy value for quasar environments or the hosts of lower redshift metal absorbers.

Finally, the question of the evolution of quasars and the search for lensed quasars presented in Chapter 5 are perhaps the projects that will benefit least from *JWST*. In Chapter 5, I suggested fainter quasar spectra should be studied in order to assess whether the high C IV emission line blueshifts are a consequence of evolution or selection bias. *JWST* is not meant to be a quasar confirmation machine due to overheads and lower sensitivity at < 9000 Å. However, large extragalactic programmes with the NIRSpec MSA or the NIRISS grism might discover serendipitous faint quasars while surveying deep fields. Additionally, if more lensed quasars are found before the end of the mission, the NIRSpec IFU and MIRI IFU modes could be an excellent way to study the host galaxies of $z > 6$ lensed quasars, as the lensing shear and the diffraction-limited PSF might help resolve the host galaxy. Ultimately, searches for quasars will benefit more in the long-term from all-sky surveys in the infrared with *Euclid* and the *Nancy Grace Roman Space Telescope*.

6.5 Final thoughts

When I embarked on this PhD adventure, measuring the escape fraction of faint galaxies in the reionisation era was a fabled goal that seemed beyond the reach of current instruments and data. Three years after, this Thesis has shown that escape

fractions can be measured statistically at $z \sim 6$, and that galaxies indeed provided enough photons to reionise the Universe. Importantly, the development of the metal-IGM cross-correlation and the 2PCCF enables this measurement to be performed at even higher redshifts when galaxies and/or Lyman-α transmission features are scarce. Meanwhile, the search for lensed quasars has just begun, and the combination of machine learning with future large infrared imaging surveys will hopefully discover plenty more. I am confident that the new approaches presented in this Thesis will blossom in the future golden age of astronomy with *JWST,Euclid,Roman Space Telesope*, ELT, TMT, GMT and SKA coming online within the next decade. This is only the beginning of an amazing journey into the first billion years of the Universe.

Appendix A

Appendices to Chapter 3

A.1 Properties of all retrieved C IV systems

I hereby give the tabulated fitted redshift, column density, Doppler parameter and the corresponding errors of all our detected C IV absorbers in each QSO sightline studied.

Table A.1: Complete list of C IV absorbers with redshift, Doppler parameter b [km s^{-1}], column density $\log N$ [cm^{-2}] and associated errors.Fixed parameters in VPFIT are indicated with a †.

J0002+2550						
	4.434586	0.000101	43.31	10.31	13.317	0.054
	4.440465	0.000091	31.22	8.24	13.696	0.049
	4.441937	0.000077	21.93	7.62	13.708	0.042
	4.675162	0.000042	8.22	7.25	13.184	0.099
	4.870726	0.000066	21.65	7.59	13.579	0.031
	4.872056	0.000097	5.74	3.19	13.498	0.208
	4.873713	0.000102	6.00†	–	13.109	0.080
	4.941568†	-	10.87	9.71	13.387	0.117
	4.943826	0.000090	5.00	4.16	13.356	0.267
	4.945804 †	-	43.39	10.10	13.417	0.051

Continued on next page

Table A.1 – *Continued from previous page*

QSO	z	Δz	b	Δb	$\log N$	$\Delta \log N$
	5.282356	0.000070	19.67	8.62	13.709	0.069
J0005-0006						
	4.735463	0.000114	83.32	8.75	13.944	0.035
	4.812924	0.000088	6.64	3.67	13.832	0.400
J0050+3445						
	4.724040	0.000057	9.78	5.43	13.664	0.178
	4.726409	0.000043	38.82	4.01	13.979	0.024
	4.824044	0.000090	51.15	7.91	13.763	0.040
	4.908403	0.000140	24.06	16.53	13.559	0.087
	5.221126	0.000196	31.09	19.24	13.935	0.132
J0100+2802						
	4.875143	0.000119	44.59	9.08	13.218	0.065
	5.109066	0.000122	73.39	5.95	14.321	0.037
	5.111580	0.000412	49.73	30.36	13.530	0.255
	5.113629	0.000050	18.18	4.71	13.639	0.046
	5.338458	0.000059	42.39	3.94	13.955	0.033
	5.797501	0.000184	58.14	13.37	13.118	0.072
	5.975669	0.000158	20.63	13.10	12.742	0.120
	6.011654	0.000268	73.94	16.94	13.488	0.078
	6.184454	0.000362	31.35	22.12	13.061	0.256
	6.187095	0.000121	64.42	7.57	13.971	0.039
J0148+0600						
	4.515751	0.000067	17.57	7.02	12.982	0.078
	4.571214	0.000090	23.70	8.34	13.060	0.087
	4.932220	0.000915	40.41	28.81	13.614	1.591
	5.023249	0.000056	$6.00^{\dagger}$	-	13.811	0.189
	5.124792	0.000042	27.09	3.02	14.084	0.039
J0231-0728						
	4.138342	0.000017	9.32	2.88	13.315	0.037
	4.225519	0.000069	40.31	6.24	13.283	0.046
	4.267255	0.000063	28.12	6.14	13.200	0.053
	4.506147	0.000067	18.27	7.42	13.093	0.068
	4.569167	0.000224	93.48	16.61	13.407	0.066
	4.744957	0.000130	78.70	9.93	13.552	0.044

Continued on next page

Table A.1 – *Continued from previous page*

QSO	z	Δz	b	Δb	$\log N$	$\Delta \log N$
	4.883691	0.000205	111.96	14.74	13.566	0.048
	5.335394	0.000240	56.00	16.76	13.299	0.102
	5.357714	0.000044	28.37	3.55	13.749	0.031
	5.345812	0.000088	25.35	6.99	13.361	0.068
	5.348184	0.000043	21.49	3.75	13.647	0.038
J0353+0104						
	4.675911	0.000078	21.61	12.12	13.638	0.063
	4.977485	0.000124	10.81	6.33	14.158	0.481
	5.036105	0.000454	15.00†	-	13.465	0.232
	5.037794	0.000222	15.00†	-	13.933	0.158
J0818+1722						
	4.552517	0.000102	24.98	10.16	12.812	0.090
	4.620602	0.000073	25.99	6.99	13.048	0.062
	4.627338	0.000104	38.92	8.89	13.057	0.067
	4.727010	0.000068	31.70	3.07	14.005	0.065
	4.725780	0.000182	42.07	8.27	13.763	0.112
	4.731844	0.000033	41.16	2.57	13.672	0.020
	4.877661	0.000070	38.94	5.70	13.215	0.043
	4.936947	0.000107	32.24	9.12	12.666	0.075
	4.940176	0.000056	14.41	7.43	12.773	0.057
	4.941255	0.000071	10.00†	-	12.783	0.071
	4.942578	0.000060	40.13	5.22	13.188	0.036
	5.064643	0.000207	68.13	16.13	13.484	0.081
	5.076455	0.000082	33.61	6.51	13.563	0.056
	5.082651	0.000078	37.51	5.90	13.546	0.047
	5.308823	0.000061	16.45	6.50	13.031	0.054
	5.322429	0.000121	29.38	9.91	13.341	0.084
	5.789525	0.000111	38.43	7.49	13.339	0.062
	5.843998	0.000110	44.91	7.41	13.345	0.051
	5.877228	0.000106	39.41	7.26	13.315	0.055
J0836+0054						
	4.682130	0.000093	50.92	7.82	13.406	0.049
	4.684465	0.000046	40.42	4.38	13.608	0.033
	4.686503	0.000029	33.65	2.53	13.689	0.022

Continued on next page

Table A.1 – *Continued from previous page*

QSO	z	Δz	b	Δb	$\log N$	$\Delta \log N$
	4.773175	0.000389	96.43	31.27	13.135	0.113
	4.996702	0.000042	27.12	3.50	13.569	0.035
	5.125071	0.000126	44.86	7.84	13.729	0.067
	5.127266	0.000066	13.81	7.10	13.246	0.072
	5.322763	0.000070	23.54	5.46	13.400	0.059
J0840+5624						
	4.486724	0.000060	16.99	9.88	13.674	0.096
	4.525773	0.000144	52.80	14.54	13.519	0.058
	4.546597	0.000021	30.20	4.36	15.395	0.384
J0927+2001						
	4.471037	0.000031	13.02	3.60	13.369	0.042
	4.623510	0.000149	61.15	5.56	13.592	0.076
	4.624132	0.000066	20.94	8.47	13.154	0.190
	5.014445	0.000110	19.23	9.63	13.560	0.114
	5.016426	0.000216	46.85	16.71	13.449	0.116
	5.149170	0.000799	75.01	39.93	13.377	0.271
	5.151154	0.000204	41.88	13.54	13.498	0.219
	5.671001	0.000187	57.11	12.31	13.540	0.073
J1022+2252						
	5.264299	0.000152	52.66	12.47	14.410	0.078
J1030+0524						
	4.890614	0.000164	14.31	20.24	13.081	0.162
	4.947967	0.000222	82.22	15.03	13.738	0.075
	4.948468	0.000068	7.89	7.27	13.589	0.245
	5.110703	0.000664	117.36	34.50	14.023	0.143
	5.517461	0.000102	53.31	6.76	13.904	0.045
	5.724374	0.000044	51.29	2.53	14.543	0.029
	5.741183	0.000102	50.90	6.96	13.748	0.044
	5.744256	0.000062	39.49	4.31	13.853	0.033
	5.966889	0.000076	16.27	7.02	13.569	0.074
J1044-0125						
	4.450487	0.000027	14.28	3.52	13.456	0.035
	4.898284	0.000188	78.05	9.81	13.759	0.060
	4.899596	0.000035	13.63	5.40	13.612	0.060

Continued on next page

Table A.1 – *Continued from previous page*

QSO	z	Δz	b	Δb	$\log N$	$\Delta \log N$
	5.075408	0.000091	14.12	9.69	13.433	0.107
	5.077315	0.000042	26.74	3.42	13.847	0.033
	5.167313	0.000055	10.59	6.40	13.493	0.099
	5.285046	0.000147	36.47	11.42	13.477	0.089
	5.481678	0.000080	22.20	6.70	13.775	0.070
J1048+4637						
	4.709835	0.000057	25.08	5.33	13.520	0.050
	4.888006	0.000062	20.29	6.25	13.401	0.058
	4.889219	0.000058	10.29	8.04	13.318	0.083
J1137+3549						
	4.781802	0.000182	50.43	17.36	13.622	0.072
	4.841312	0.000108	53.00	10.08	14.005	0.042
	4.874166	0.000134	74.27	10.78	14.033	0.044
J1148+5251						
	4.919028	0.000058	24.72	4.66	13.192	0.049
	4.944227	0.000058	8.91	1.09	17.656	0.123
	4.945228	0.000067	9.85	4.14	14.480	0.526
	5.152113	0.000121	54.16	8.94	13.695	0.055
J1306+0356						
	4.613924	0.000790	103.30	31.35	13.250	0.225
	4.614603	0.000033	29.33	3.21	13.722	0.050
	4.615828	0.000045	23.36	4.28	13.465	0.075
	4.668109	0.000020	8.71	1.87	13.993	0.115
	4.668741	0.000081	56.99	4.05	13.892	0.035
	4.711108	0.000135	35.60	11.23	13.129	0.096
	4.860508	0.000044	35.43	3.46	13.952	0.031
	4.862256	0.000055	28.31	5.49	13.933	0.051
	4.863239	0.000043	10.18	3.83	13.881	0.091
	4.864687	0.000053	56.14	5.04	14.327	0.026
	4.866813	0.000025	20.73	2.13	14.009	0.029
	4.868951	0.000102	51.40	8.62	13.670	0.055
	4.880126	0.000061	27.59	4.35	13.873	0.047
	4.881271	0.000058	17.01	4.23	13.690	0.061
J1319+0950						

Continued on next page

Table A.1 – *Continued from previous page*

QSO	z	Δz	b	Δb	$\log N$	$\Delta \log N$
	4.653406	0.000076	33.18	6.20	12.989	0.058
	4.660650	0.000108	42.92	7.70	13.310	0.088
	4.663610	0.000734	86.53	46.22	13.625	0.283
	4.665107	0.000104	38.60	17.21	13.328	0.456
	4.716591	0.000030	27.60	2.58	13.554	0.026
	5.264302	0.000095	25.20	7.48	13.181	0.079
	5.335365	0.000056	3.30	2.64	13.333	0.367
	5.374931	0.000054	16.59	4.79	13.447	0.057
	5.570341	0.000425	51.95	27.03	13.793	0.183
	5.573751	0.000240	49.18	14.23	14.139	0.111
J1411+1217						
	4.930500	0.000439	211.93	26.04	13.982	0.048
	4.960547	0.000100	6.41	4.33	13.715	0.364
	5.250460	0.000063	73.55	4.62	14.442	0.025
J1509-1749						
	4.641643	0.000051	26.96	4.46	13.514	0.042
	4.791660	0.000166	51.38	12.64	13.439	0.082
	4.815429	0.000055	44.63	3.79	14.050	0.037
J2054-0005						
	4.868746	0.000079	15.81	8.89	13.804	0.143
	5.213101	0.000147	8.67	5.96	13.828	0.368
J2315-0023						
	4.897117	0.001163	55.02	64.70	13.598	0.478
	4.898876	0.000354	31.61	22.67	13.854	0.257

A.2 Velocity plots of C IV absorbers used in the correlation measurement

I present in Figure A.1 the velocity for the 37 C IV systems of Sample α. Note the consistent presence of at least of a few completely opaque pixels at the location where the Lyman-α absorption at the redshift of C IV is expected. Note that we have not plotted individual detections of systems with less than $\lesssim 100$ km s^{-1} separation as

Table A.2: Inferred model parameters for different choices of Lyman-α bias parameters (b_α, β_α). Notably, the choice of $b_\alpha = -1.3$ corresponds to our fiducial choice on the Lyman-α power spectrum measurement whereas $b_\alpha = -0.75$ corresponds to the extrapolation of the evolution of the bias from low-redshift values (McDonald et al. 2006; du Mas des Bourboux et al. 2017).

b_α	β_α	$\log\langle f_{\rm esc}\xi_{\rm ion}\rangle$	$M_{\rm UV}^{\rm lim}$	$b_{\rm CIV}$
-0.5	1.0	$25.55^{+0.14}_{-0.11}$	$-11.66^{+3.47}_{-8.16}$	$22.15^{+7.26}_{-8.29}$
-0.5	1.5	$25.66^{+0.21}_{-0.19}$	$-9.98^{+1.92}_{-9.96}$	$18.48^{+7.67}_{-5.75}$
-0.5	2.0	$25.66^{+0.32}_{-0.14}$	$-8.90^{+0.84}_{-10.92}$	$16.65^{+6.38}_{-5.92}$
-0.75	1.0	$25.25^{+0.19}_{-0.09}$	$-19.94^{+11.87}_{-0.00}$	$11.89^{+7.80}_{-4.96}$
-0.75	1.5	$25.35^{+0.23}_{-0.11}$	$-19.58^{+11.40}_{-0.36}$	$15.96^{+2.25}_{-8.04}$
-0.75	2.0	$25.54^{+0.22}_{-0.22}$	$-8.78^{+0.72}_{-11.16}$	$8.97^{+7.95}_{-2.84}$
-1.3	1.0	$24.92^{+0.23}_{-0.15}$	$-10.22^{+2.16}_{-9.72}$	$8.01^{+2.84}_{-3.72}$
-1.3	1.5	$25.01^{+0.30}_{-0.19}$	$-10.82^{+2.76}_{-9.12}$	$7.09^{+3.29}_{-2.86}$
-1.3	2.0	$25.17^{+0.28}_{-0.22}$	$-8.06^{+0.00}_{-11.88}$	$6.40^{+2.69}_{-2.15}$
-2.0	1.0	$24.63^{+0.15}_{-0.15}$	$-8.06^{+0.00}_{-11.87}$	$4.46^{+2.85}_{-2.07}$
-2.0	1.5	$24.80^{+0.22}_{-0.26}$	$-8.78^{+0.72}_{-11.16}$	$4.33^{+1.95}_{-2.20}$
-2.0	2.0	$24.74^{+0.55}_{-0.10}$	$-11.90^{+3.84}_{-8.04}$	$4.02^{+1.69}_{-1.58}$

only one redshift for the whole system was retained for the correlation measurement. However, multiple absorbers forming a system but can be easily spotted on some plots.

A.3 Relaxing the bias parameters of the Lyman-α forest

I present the posterior probability distribution of the parameters of our linear model including the bias of the Lyman-α and the associated RSD parameter as free parameters with flat priors in $-3 < b_\alpha < 0$ and $-3 < \beta_f < 0$, respectively. The result is presented in Figure A.2. The bias is in a degenerate state with all other parameters, parameters can be tuned to compensate the bias parameter changes and still produce the same fit to the data. I note that higher values of the bias than the one extrapolated from low-redshift measurements would yield in turn lower values of the host halo mass, and lower values of the escape fractions and LyC photon production product (see Table A.2).

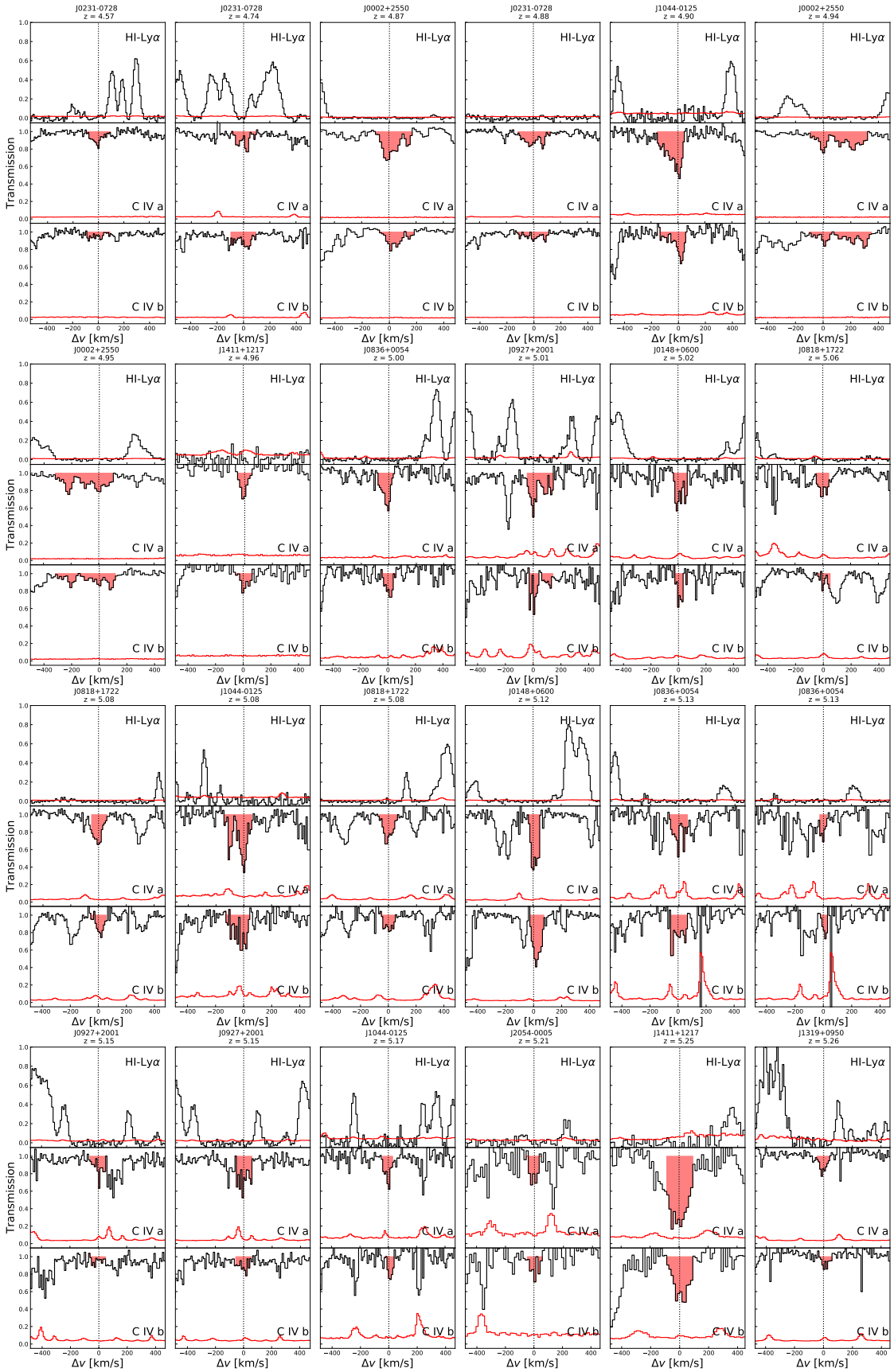

Figure A.1: Velocity plots of C IV absorbers in Sample α. For simplicity, absorbers with $\Delta v \lesssim 100$ km s^{-1} were not plotted twice. They can be however be easily spotted in the lower C IV panels. All C IV absorbers land in a Lyman-α opaque region, often enclosed by high transmission spikes.

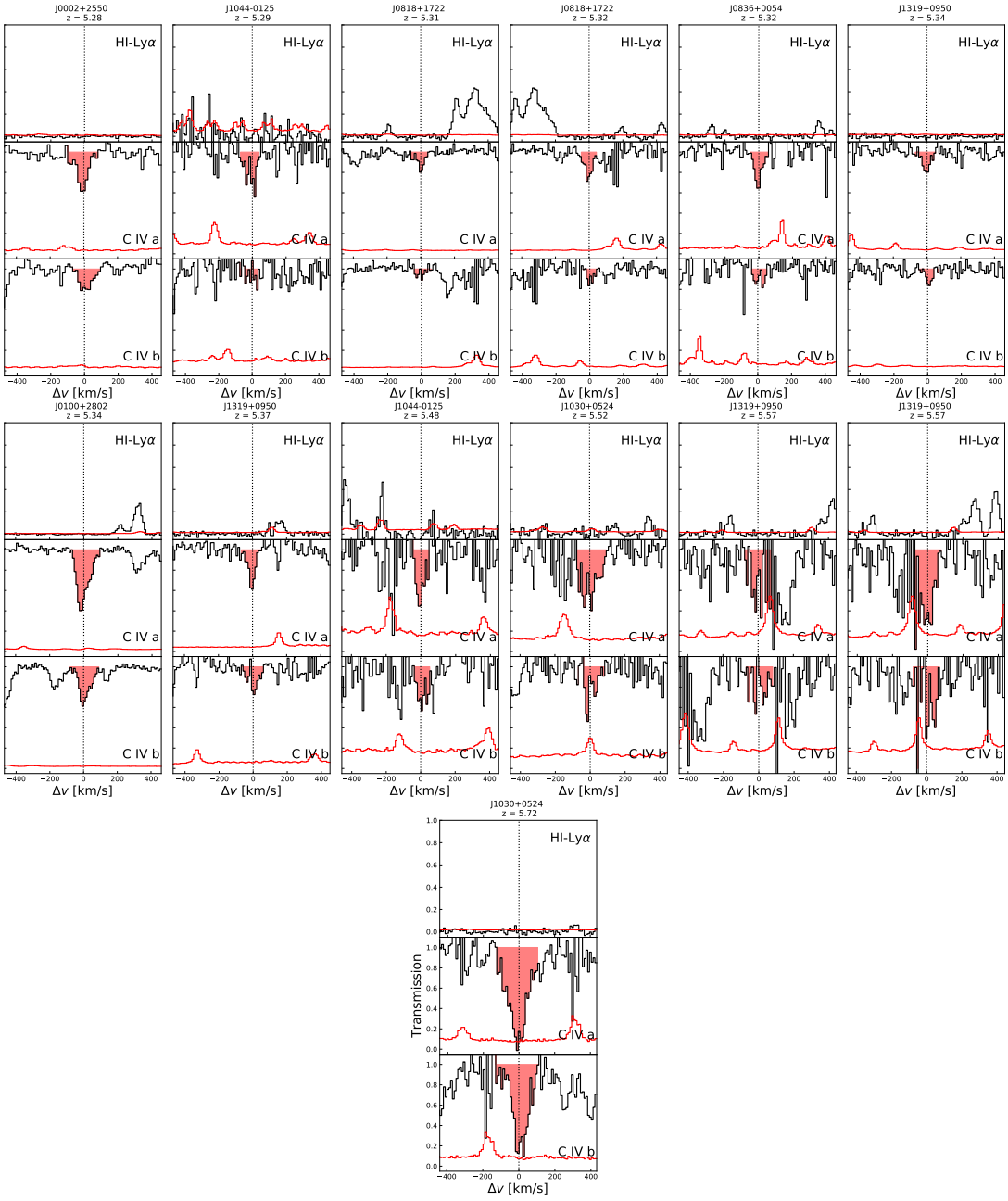

Figure A.1: (Continued)

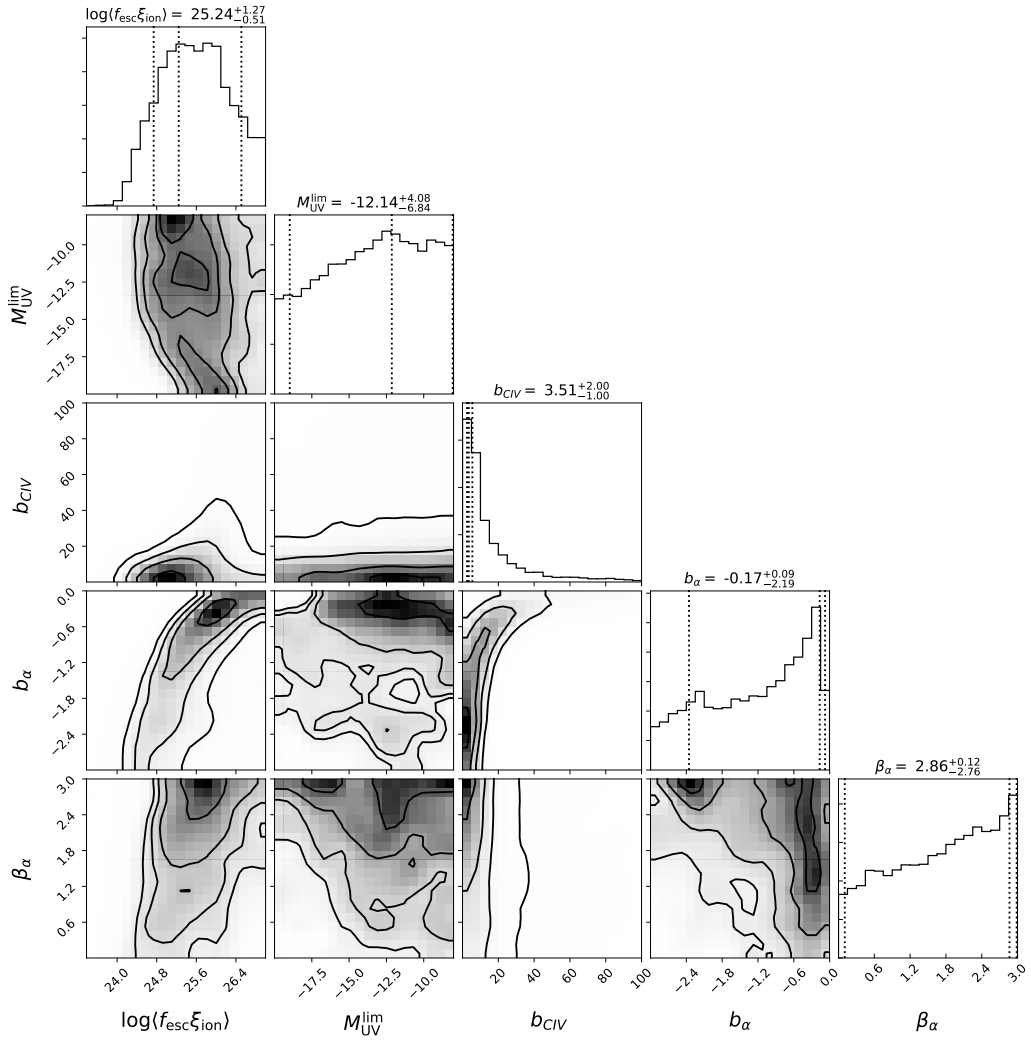

Figure A.2: Posterior distribution for the 3 original parameters of the linear model to which have been added the Lyman-α bias parameters (b_α, β_α) with appropriate priors. The quoted numbers give the best-fit (maximum likelihood) and the 1-σ credibility intervals on the marginalized distributions. The constraints on $\langle f_{\rm esc}\xi_{\rm ion}\rangle$ and $b_{\rm CIV}$ are in good agreement with the more restricted fit presented in the main text (Figure3.9). There is a clear degeneracy between these two parameters and b_α. A wide variety of combinations of the 4 parameters (excluding $M_{\rm UV}^{\rm lim}$) can produce a similar fit the data. Hence a choice of (b_α, β_α) must be made in order to infer the remaining parameters.

Appendix B

Appendices to Chapter 4

B.1 Summary of all LBG spectroscopic confirmation with DEIMOS

I present the DEIMOS spectroscopic confirmation of new LBGs in the field of J0836 (Figure A1, online material) and J1030 (Figure A2, online) used in this work for the cross-correlations. We leave the presentation and analysis of the 3 objects detected in the near-zone of J0836 to Bosman et al. (2020). Table B.1 lists the LBG detections with their coordinates, redshift, Lyman-α FWHM and corrected redshift.

Quasar	RA	DEC	z_{peak}	FWHM	z_{corr}	r (mag)	i (mag)	z (mag)	M_{UV}
J0836	129.09106	1.00954	5.283	95.8	5.281	> 27.62	26.35	25.33	-21.16
J1030	157.71105	5.36851	5.508	92.5	5.507	> 27.5	25.54	23.95	-22.61
	157.58161	5.46687	5.791	66.9	5.79	> 27.50	24.95	23.41	-23.23
	157.58308	5.44516	5.481	69.3	5.480	> 27.50	26.51	25.12	-21.43
	157.67004	5.45504	5.712	176.5	5.709	> 27.50	26.13	25.18	-21.44
	157.73887	5.46775	5.612	137.3	5.610	> 27.50	> 26.80	25.45	-21.13
	157.70962	5.36157	5.692	223.8	5.688	> 27.50	26.39	25.19	-21.42
	157.52691	5.37737	5.352	118.1	5.351	> 27.50	26.38	25.18	-21.33
	157.56116	5.34611	5.446	186.8	5.443	> 27.50	25.93	25.49	-21.05

Table B.1: Summary of the detected LBGs in the DEIMOS fields

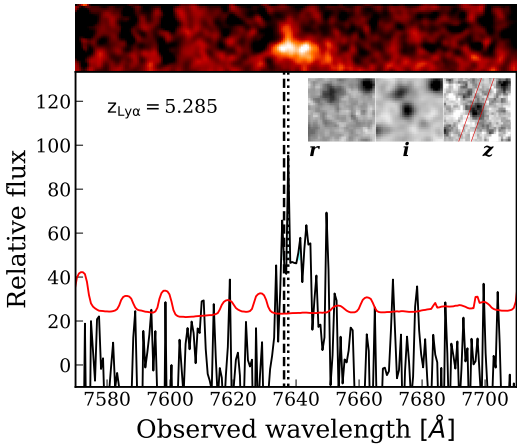

Figure B.1: Confirmed LBG observed with DEIMOS in the field of J0836 (see also Figure 4.3). The top panels show the 2D spectra from which the 1D spectrum (black line) and noise (red) are optimally extracted using a boxcar aperture of 1.2″. In the upper right corner is displayed the *riz* image used for the drop-out selection.

B.2 Individual LAE detections with MUSE in the Lyman-α forest of our quasars

I present a summary of all detected LAEs in the redshift range of the Lyman-α forest of the nearby quasar in Table B.2. We adopt an identification scheme where each LAE is named 'JXXXX_NBYYYY', where XXXX denotes the hours and minutes of the RA coordinates of the central quasar and YYYY the rounded wavelength of the narrowband (NB) frame in which MUSELET or LSDCat found the highest signal of the detection, which is often very close to the wavelength of the emission peak. Individual plots similar to Figure 4.4 for each LAE can be found online in Figure B1 (J0305, online), B2 (J1030, online), B3 (J1526, online), B4 (J2032, online), B5 (J2100, online), B6 (J2329, online). Finally, we provide an example of common misdetections that are removed by visual inspection in Figure B7 (online)such as low-redshift O II or continuum emitters, bright nearby foreground objects or defects or cosmic rays impacting only one of the exposures.

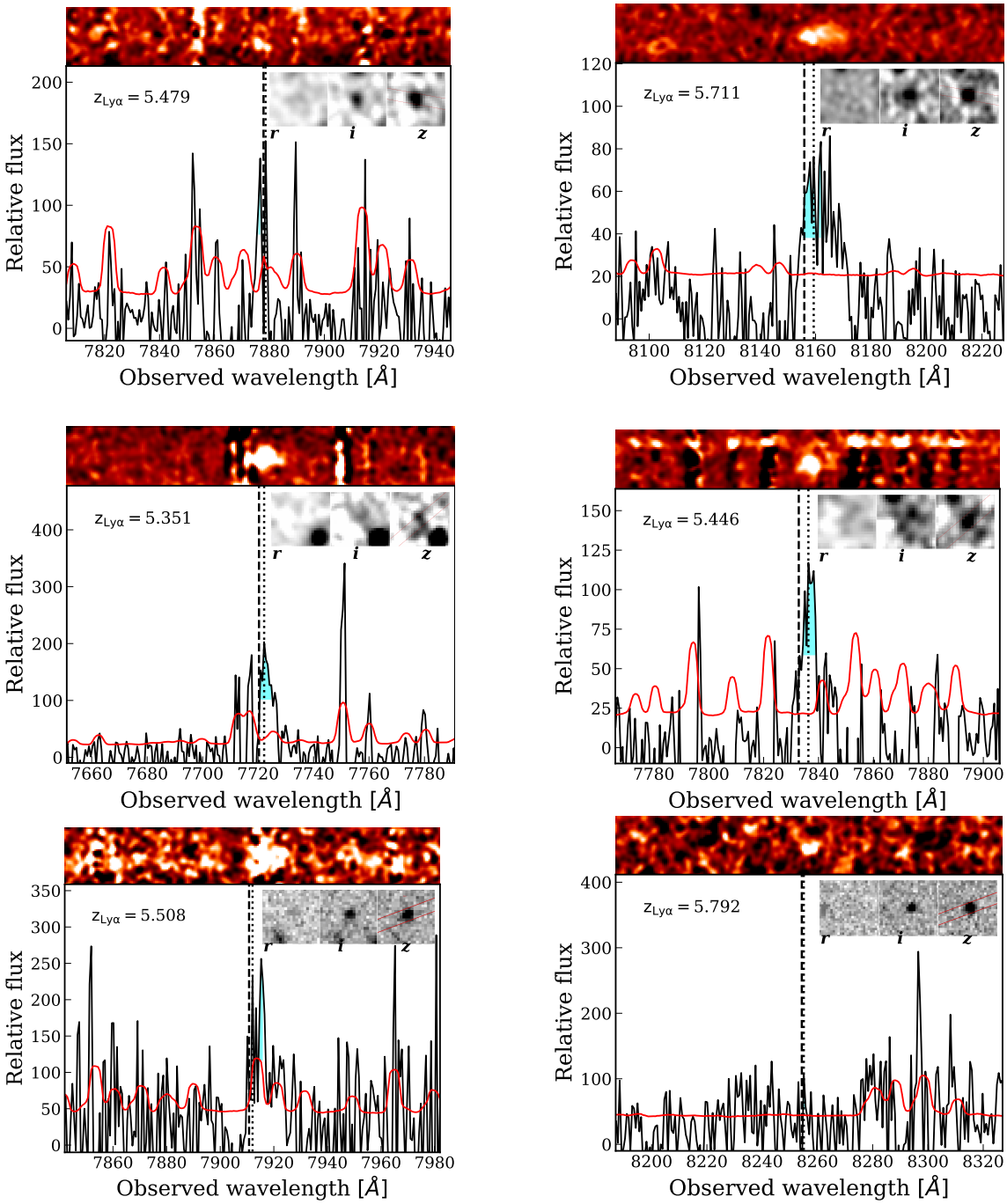

Figure B.2: Confirmed LBGs observed with DEIMOS in the field of J1030. The top panels shows the 2D spectra from which the 1D spectrum (black line) and noise (red) are optimally extracted using a boxcar aperture of 1.2″ . In the upper right corner is displayed the *riz* image used for the drop-out selection.

ID	RA	DEC	λ_{Lya}	z_{Lya}	FWHM [km/s]	z_{corr}
J0305_NB8032	46.32776	-31.84569	8034.7	5.607	186.7	5.604
J0305_NB8609	46.31154	-31.85152	8609.7	6.081	174.2	6.078
J0305_NB8612	46.31095	-31.85202	8612.2	6.084	174.2	6.081
J1030_NB7707	157.61238	5.40784	7707.2	5.340	97.3	5.339
J1030_NB7927	157.61109	5.41578	7927.2	5.520	236.5	5.516
J1030_NB7942	157.61054	5.40995	7942.2	5.533	141.6	5.531
J1030_NB8177	157.61534	5.40556	8177.2	5.725	229.3	5.722
J1030_NB8202	157.61366	5.41512	8202.2	5.746	228.6	5.742
J1030_NB8220a	157.62069	5.41484	8220.9	5.761	228.1	5.758
J1030_NB8220b	157.61321	5.41900	8220.9	5.761	228.1	5.758
J1526_NB8476	231.66377	-20.83180	8475.9	5.972	88.5	5.971
J1526_NB8874	231.65771	-20.82652	8874.7	6.299	169.0	6.296
J2032_NB8396	308.04785	-21.23293	8396.4	5.907	134.0	5.905
J2032_NB8524	308.04240	-21.22620	8523.9	6.012	132.0	6.010
J2032_NB8525	308.03598	-21.23630	8525.2	6.013	132.0	6.011
J2100_NB7454	315.23399	-17.26017	7454.8	5.132	150.9	5.130
J2100_NB7678	315.23219	-17.26062	7678.6	5.316	146.5	5.314
J2100_NB8146	315.22375	-17.25901	8146.1	5.701	230.2	5.697
J2100_NB8419	315.22404	-17.26045	8419.8	5.925	133.6	5.923
J2329_NB8372	352.28913	-3.04041	8372.7	5.887	134.4	5.885
J2329_NB8390	352.28769	-3.03636	8390.2	5.902	134.1	5.900

Table B.2: Summary of the detected LAEs in the MUSE fields (and in the suitable redshift range for the cross-correlation). The last column gives the corrected redshift using the method of Verhamme et al. (2018), as described in Section 4.2.4.

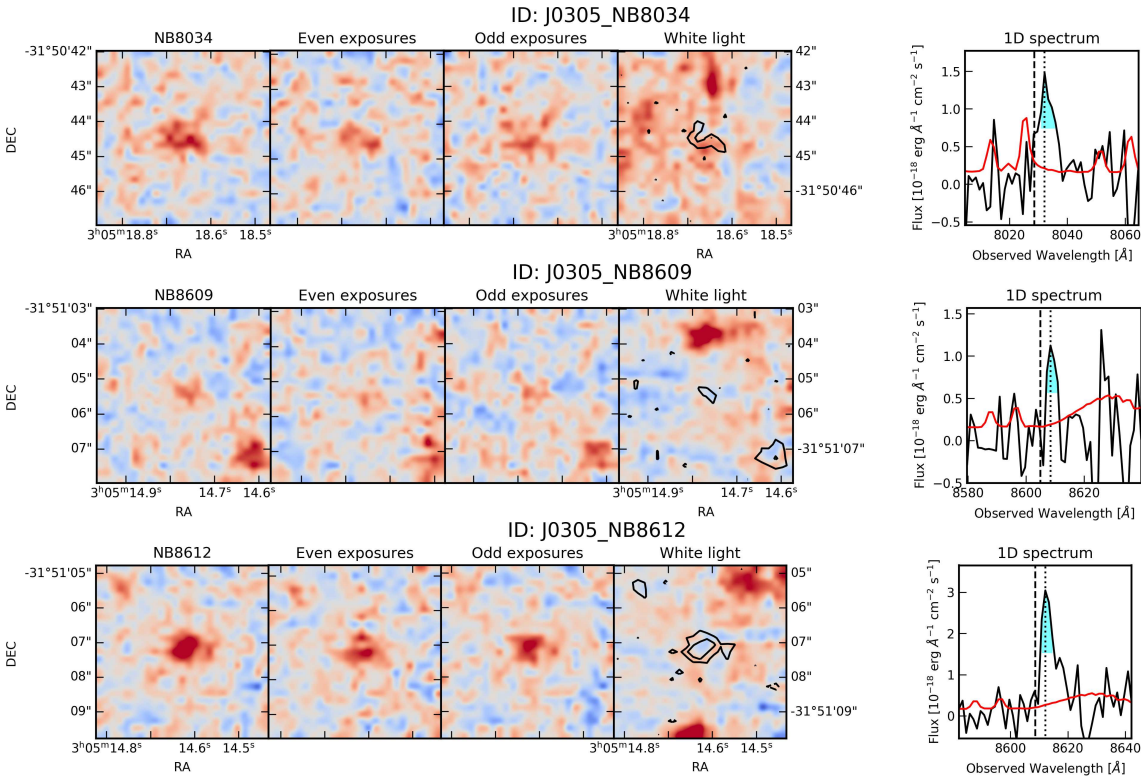

Figure B.3: LAEs detected in the field of J0305 used in this work

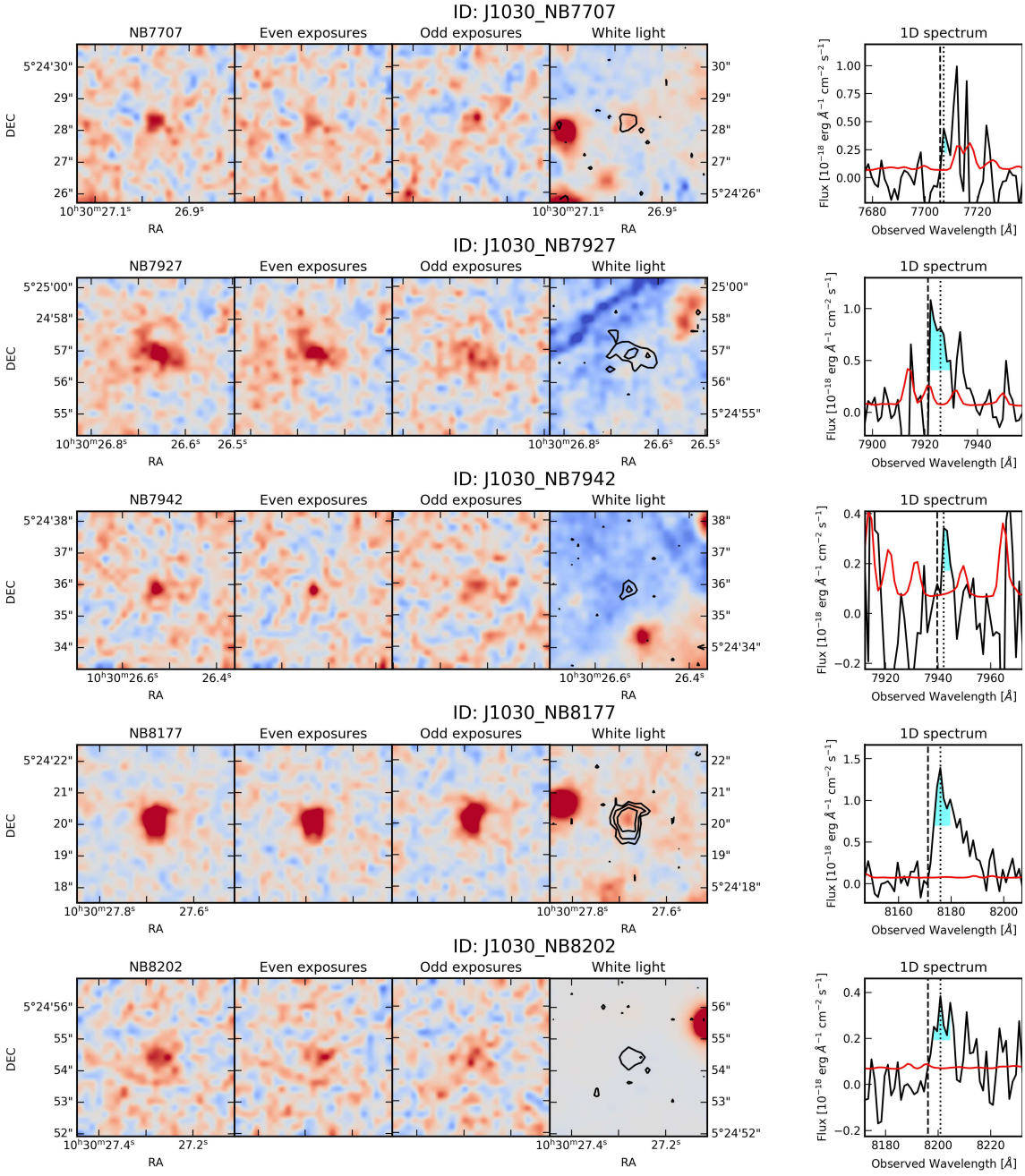

Figure B.4: LAEs detected in the field of J1030 used in this work

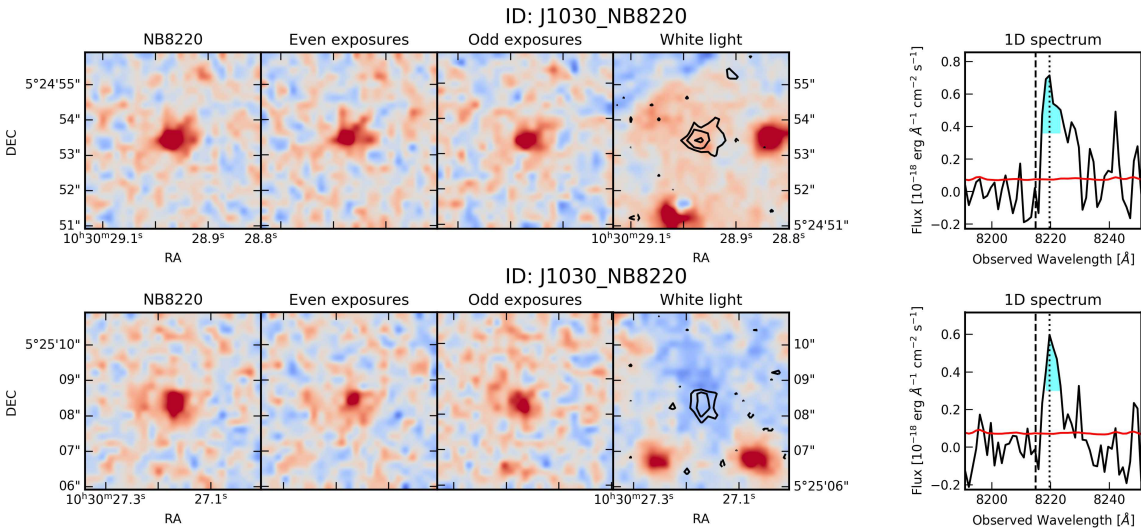

Figure B.4: (Continued) LAEs detected in the field of J1030 used in this work

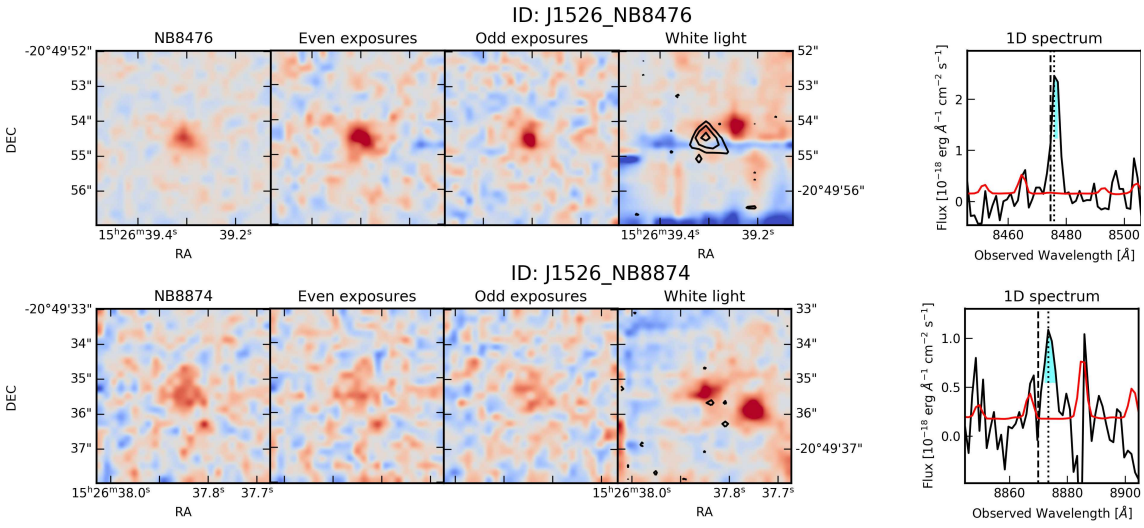

Figure B.5: LAEs detected in the field of J1526 used in this work

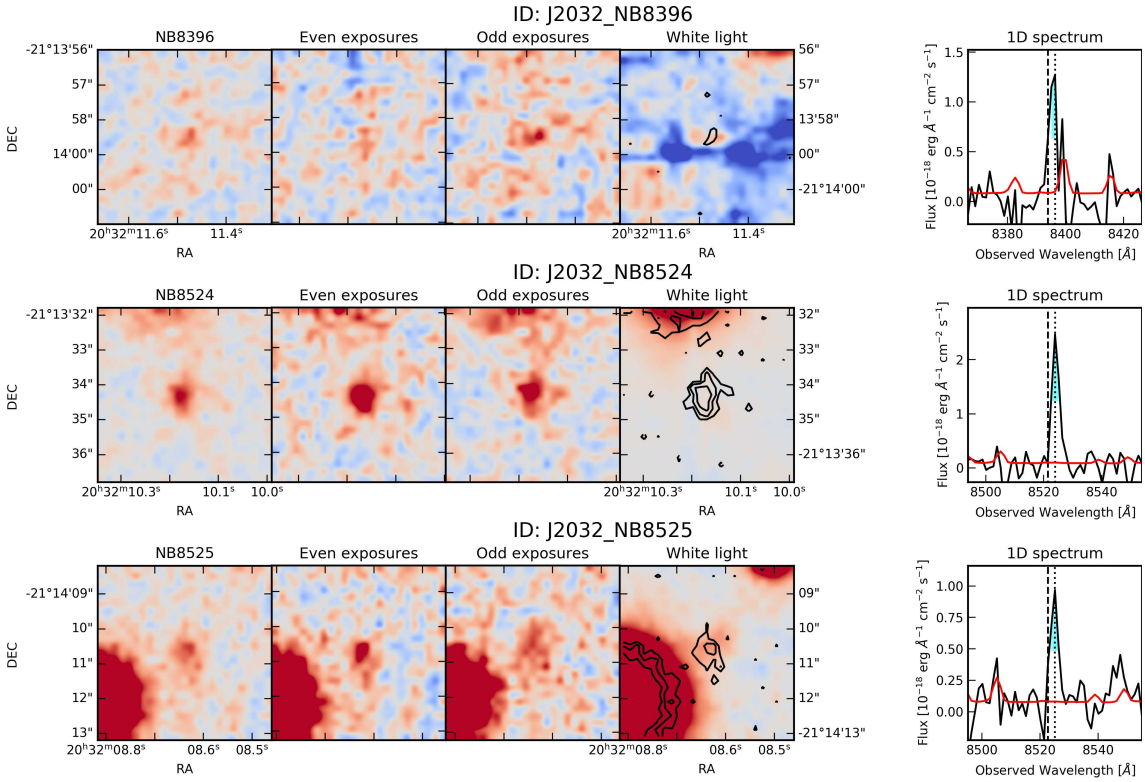

Figure B.6: LAEs detected in the field of J2032 used in this work

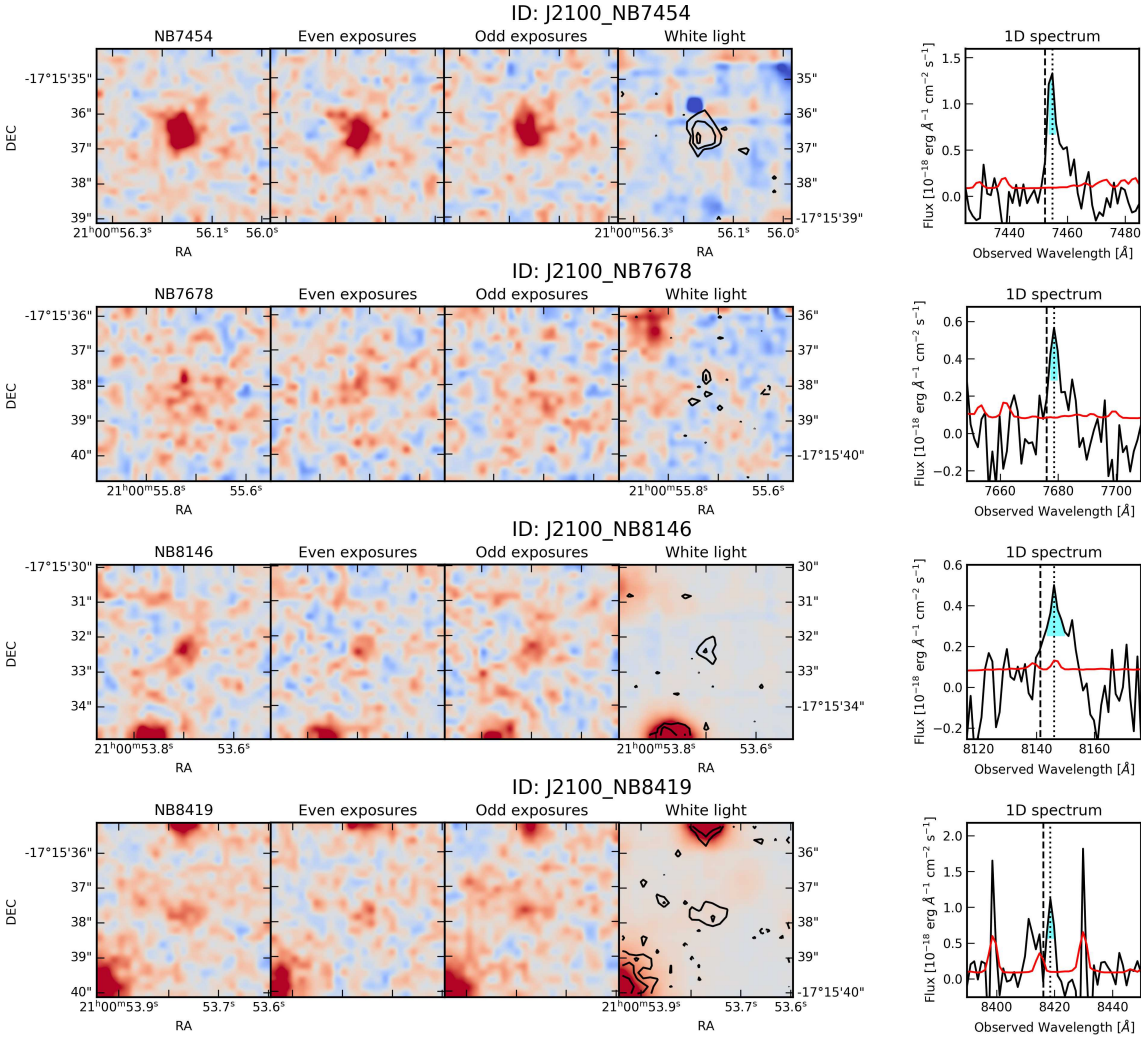

Figure B.7: LAEs detected in the field of J2100 used in this work

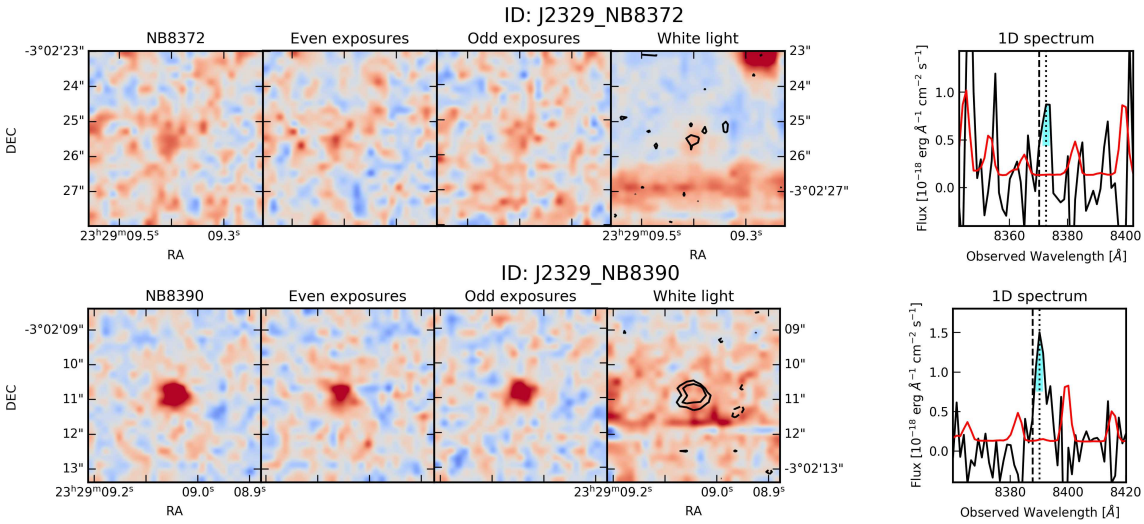

Figure B.8: LAEs detected in the field of J2329 used in this work

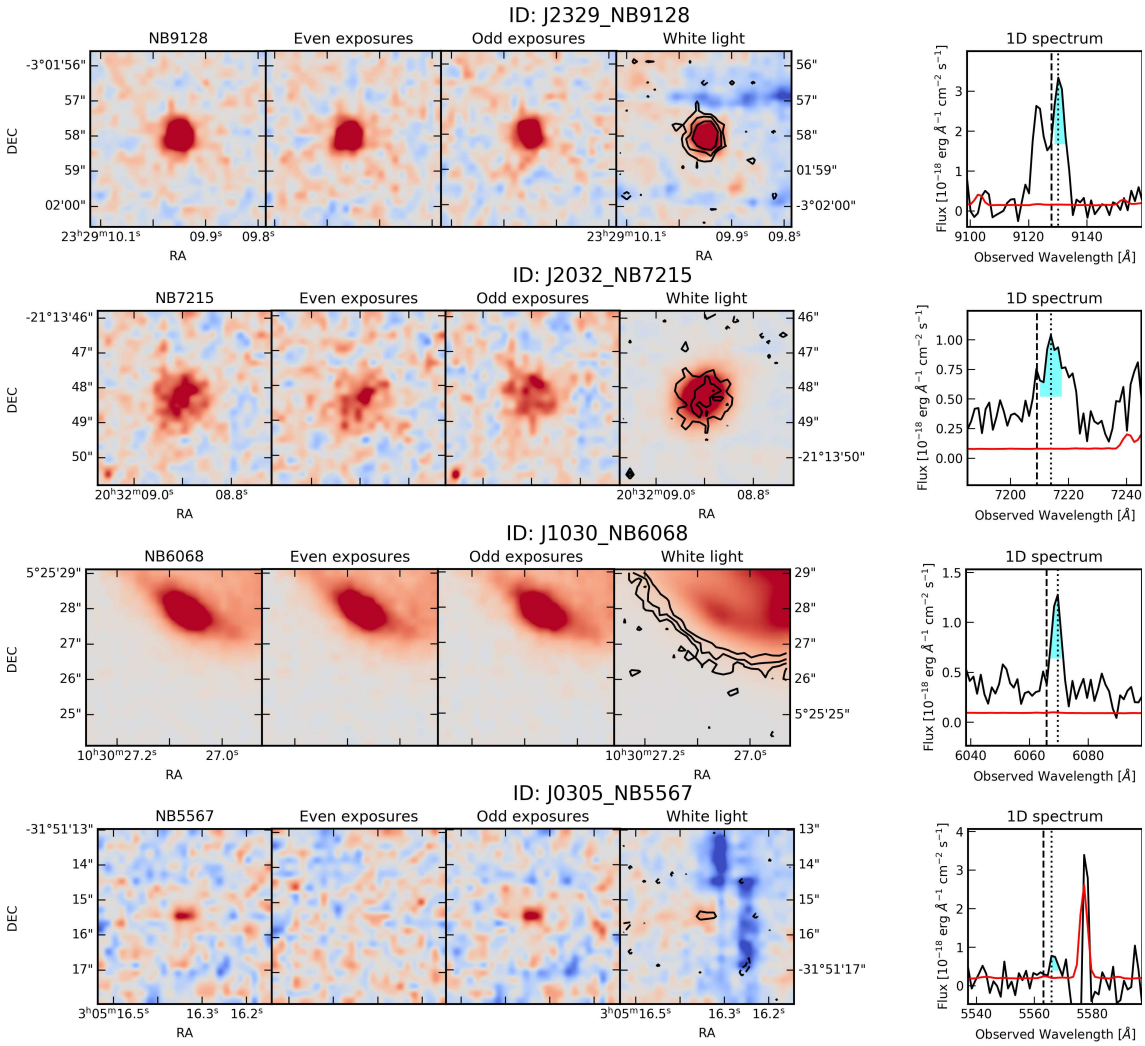

Figure B.9: Examples of typical LSDCat false positive detections of LAEs in MUSE cubes. **Top row:** Unsubtracted continuum visible both in the 1D spectra and the white light image. **2nd row:** [OII]$\lambda\lambda$3727 Å emitter with continuum **3rd row:** Detection due to a poorly substracted foreground object **Bottom row:** Faint misdetection due to an artifact or cosmic ray in one pixel in one exposure. The detection is subsequently only seen in the even or odd exposures cubes.

B.3 The mean transmission in the Lyman-α forest around LBGs and LAEs

The first measurement of the correlation between galaxies and the reionising IGM was performed in Kakiichi et al. (2018) by computing the mean flux at distance r from detected LBGs. We follow the same methodology here, computing the transverse and line-of-sight distance of every pixel in the Lyman-α forest of quasars to the detected LBGs and LAEs. We then bin the observed transmission in segment of a few cMpc, weighting each point by the inverse of the squared error. I present the results in Figure B.10 and Figure B.11.

We do not find any evidence for increased transmission close to LBGs with our stack of 3 sightlines and 13 LAEs, unlike the tantalising signal presented in Kakiichi et al. (2018). Although small number statistics might be biasing the measurement, we argue that we would not expect the observed average transmission to be enhanced near LBGs. The reason for this is twofold and demonstrated by Figure B.10. First of all, the signal is dominated by the sightlines with the largest number of LBGs and greatly affected by cosmic variance in these sightlines. The impact of cosmic variance was also demonstrated in theoretical work (Garaldi et al. 2019) with simulated Lyman-α skewers which concluded that normalising the transmission in each sightline as in Meyer et al. (2019a) was necessary to obtain a consistent signal across simulations boxes (or across the sky). Secondly, when the numbers of LBGs grows, the signal is dominated by the evolution of the mean opacity of the IGM with redshift. This is evidenced in the signal of the sightline with the largest number of LBGs (J1030, orange squares points in Figure B.10), which increased at negative separations from the LBG (defined as in the direction of the observer, i.e. towards lower redshift IGM), and conversely decreases at positive separations (towards the quasar, i.e. higher redshift IGM). It is therefore no surprising that the mean transmission around our large sample of LAEs does not show any enhancement of the transmission close to LAEs. It is worth noting that the absorption on small scales due to enhanced gas overdensities close to galaxies is still detected.

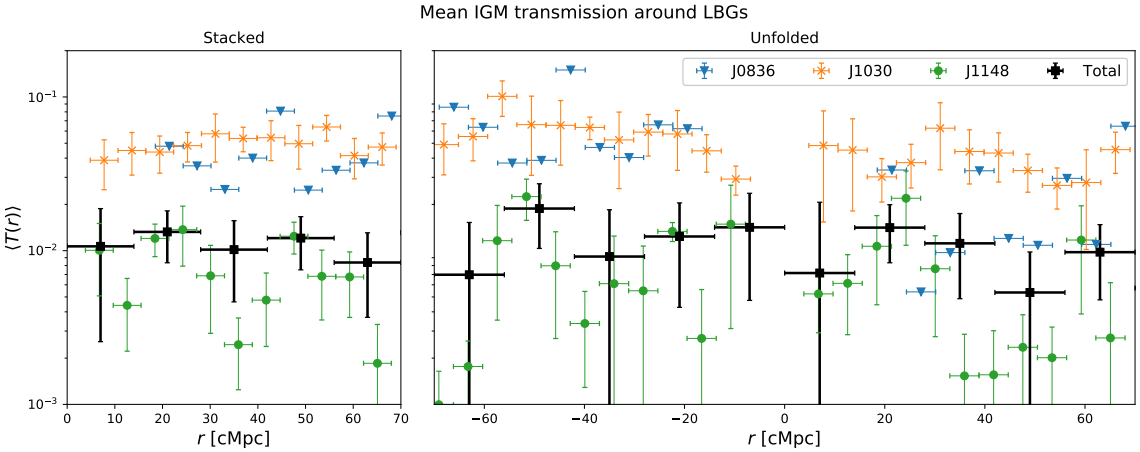

Figure B.10: Mean IGM transmission around LBGs in the three different quasar fields (blue triangles: J0836, orange squares: J1030, green circles: J1148) surveyed with DEIMOS to confirm LBGs with Lyman-α emission. The points are slightly offset in the x-axis direction for clarity. Error-bars on the transmission represent the 1σ confidence interval from bootstrapping, and thus bins to which only one LBG contributes have no uncertainty. The measurement introduced Kakiichi et al. (2018) presents a large scatter between sightlines as well as intrinsic scatter for poorly sampled sigthlines such as J0836.

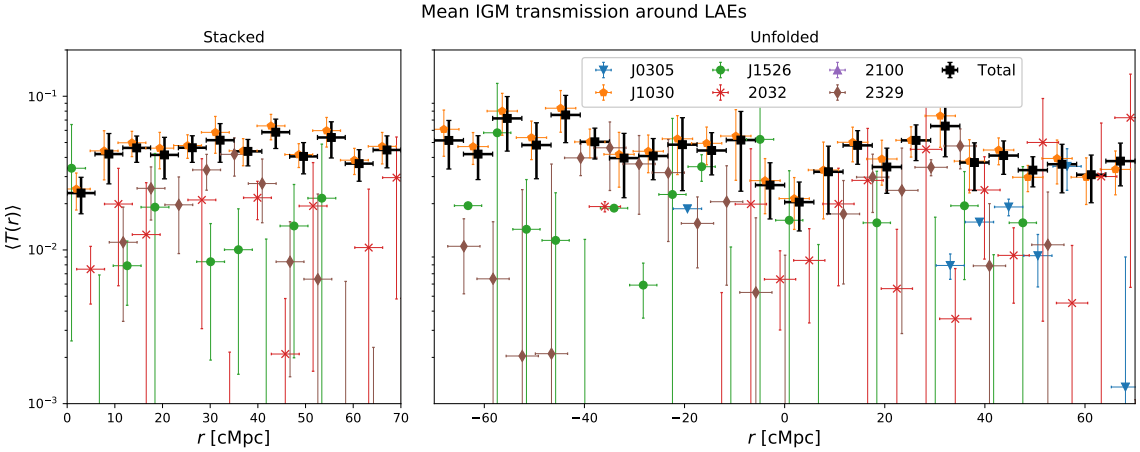

Figure B.11: Mean IGM transmission around LAEs averaged in 6 sightlines. Errorbars on the transmission represent the 1σ confidence interval from bootstrap resampling. The signal is dominated by the sightline to J1030, which contains two to three times more LAEs than other fields due to a significantly deeper MUSE observation and a potential overdensity of sources. The stacked flux only shows mild evidence for increased absorption on small scales, and no excess on larger scales of $\sim 10 - 20$ cMpc.

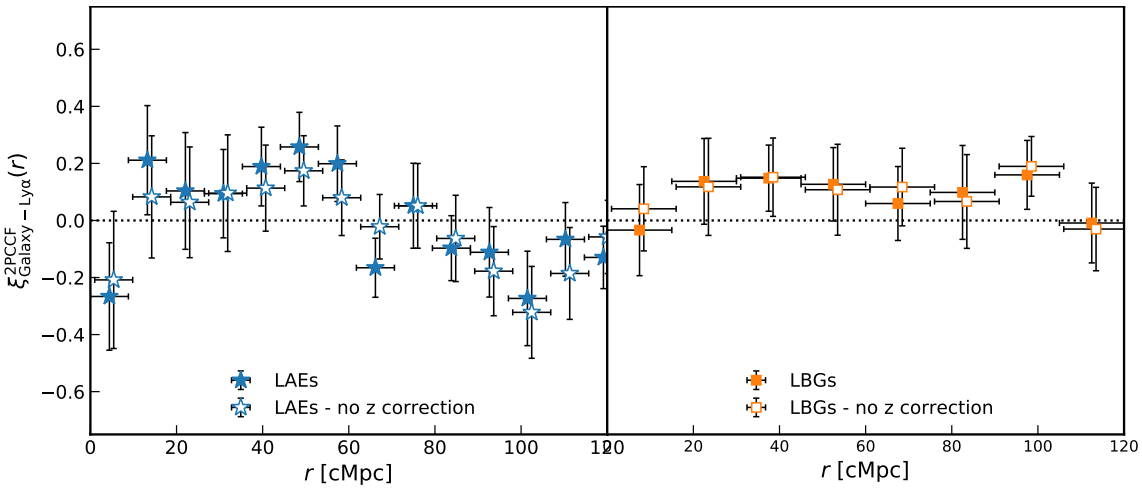

Figure B.12: 2PCCF of LAE/LBG with the Lyman-α transmission spike with (full symbols) and without (empty symbols) the redshift correction based on the FWHM of the Lyman-α line. Removing the redshift correction flattens significantly the LAE 2PCCF.

B.4 Impact of the redshift correction on the 2PCCF signal

I present in Figure B.12 the 2PCCF with and without the redshift correction detailed in Section 4.2.4. Not correcting for the Lyman-α offset to systemic reduces the signal by $-1.5\sigma(-0.2\sigma)$ for the LAE(LBG) 2PCCF. The lesser impact on the LBG 2PCCF can be explained by their larger transverse distance. Indeed, an error on the systemic redshift only affects the line-of-sight distance to transmission spikes. For LAEs, the line-of-sight direction dominates the 3D distance as $r_\perp \lesssim 7$cMpc whereas LBGs are further apart ($r_\perp \approx 10-20$cMpc), reducing the impact of an error in the line-of-sight direction.

B.5 IllustrisTNG gas overdensity PDF

I present in FigureB.13 fits of the IllustrisTNG gas PDF for additional masses. We contrast the effect of the host halo mass of on the PDF in FigureB.14. We also list the parameters of the analytical fit (Eq. 4.14) at all masses and at varied distances from the host halo centre in Table B.3. Additional parameterisation and quality plots are available upon request.

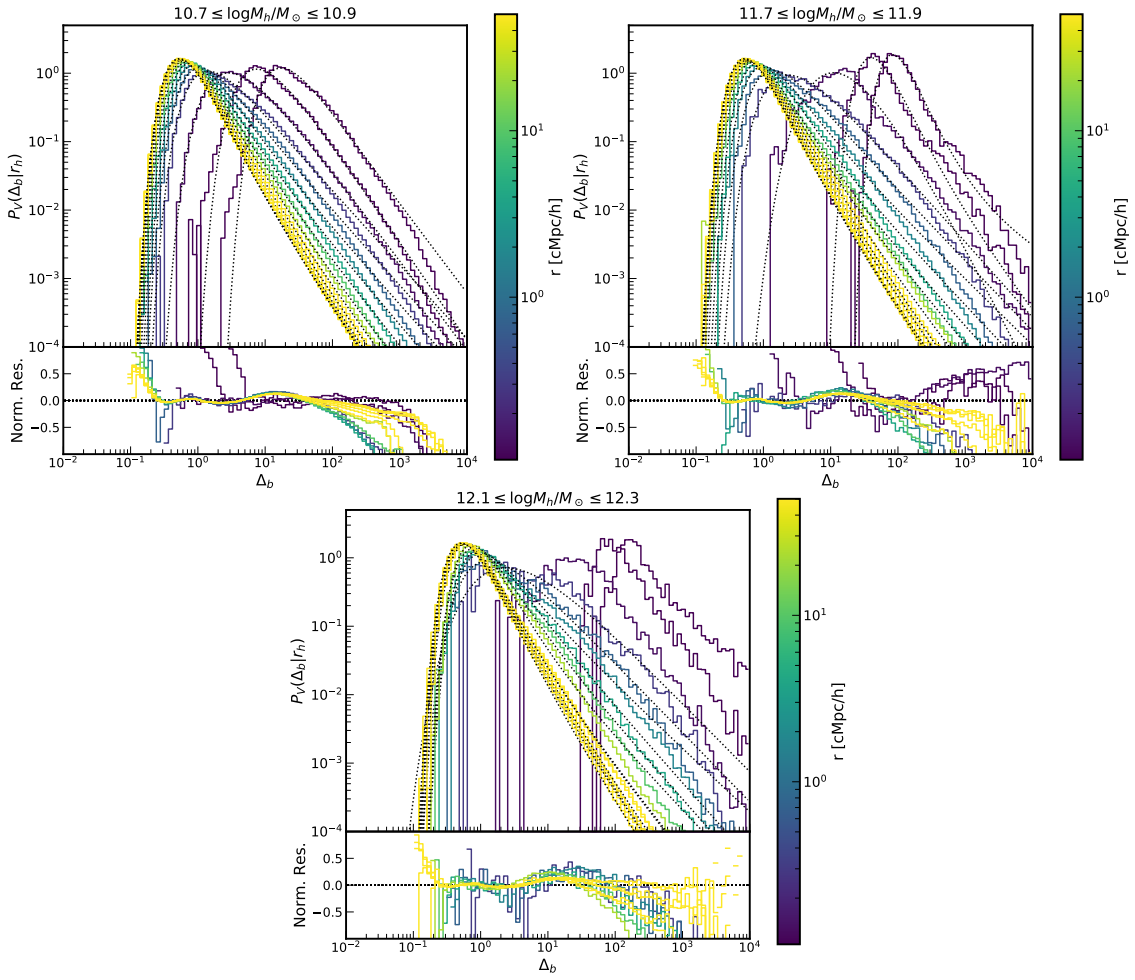

Figure B.13: **Upper panels:** The fits (dotted black) of our chosen analytical form (Eq. 4.14) are overlaid on the extracted PDF from the IllustrisTNG simulation boxes for one mass range and a given redshift, coloured by distance from the centre of the halo. **Lower panels:** Residuals of the PDF fit, coloured by distance from the centre of the halo, showing good agreement on the validity limit of the prescribed analytical form between $10^{-1} \leq \Delta_b \leq 10^2$. Only a quarter of the raw TNG PDF and associated fits are shown for clarity.

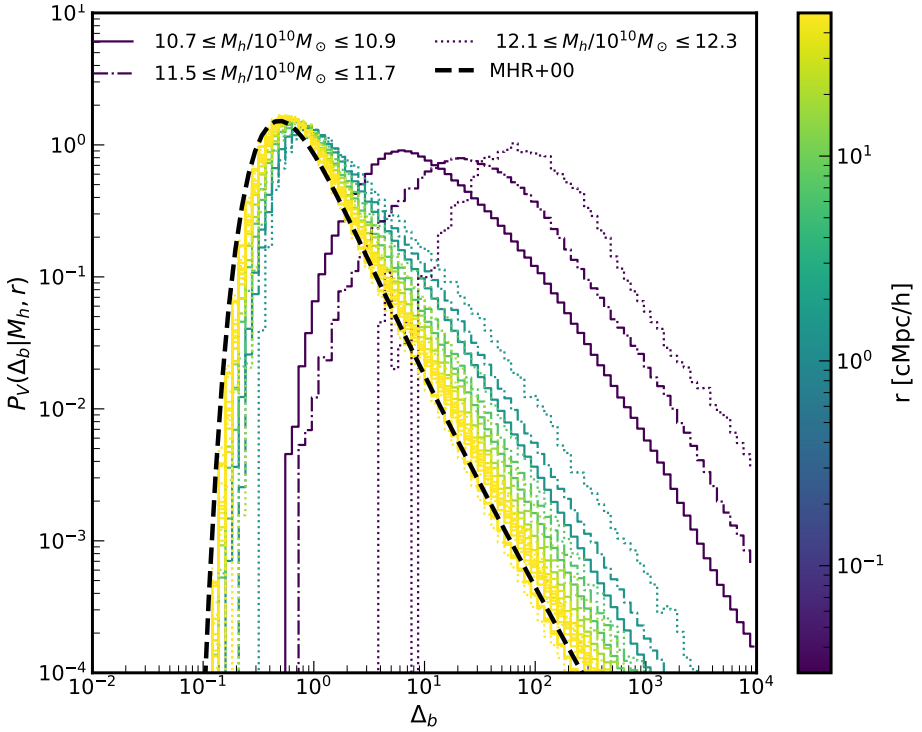

Figure B.14: Conditional baryon overdensity probability density function $P_V(\Delta_b|r, M_h)$ and its dependence on the mass and the radius from the nearest dark matter halo at $z \sim 5.85$ extracted from the IllustrisTNG100-2 cosmological simulation box. The PDFs are similar except close to the centre of the halo where gas overdensities are preferentially found. On large scales, the PDF matches the analytical fits of (Miralda-Escude et al. 2000) and (Pawlik et al. 2009 dotted black) used in Kakiichi et al. (2018) and Meyer et al. (2019a).

	$r = 1$ cMpc/h				$r = 15$ cMpc/h				$r = 50$ cMpc/h			
$\log M/M_\odot$	A	C0	δ_0	β	A	C0	δ_0	β	A	C0	δ_0	β
10.6	0.6671	0.838	0.659	2.131	0.424	0.920	0.976	2.433	0.418	0.923	0.982	2.447
10.8	0.6611	0.815	0.686	2.093	0.422	0.921	0.977	2.439	0.419	0.923	0.981	2.444
11.0	0.6699	0.814	0.677	2.101	0.423	0.922	0.975	2.437	0.419	0.922	0.981	2.444
11.2	0.6928	0.830	0.630	2.119	0.423	0.921	0.975	2.440	0.419	0.923	0.980	2.443
11.4	0.7180	0.835	0.585	2.098	0.427	0.920	0.971	2.428	0.418	0.923	0.983	2.448
11.6	0.7162	0.828	0.583	2.061	0.423	0.921	0.975	2.442	0.420	0.923	0.977	2.446
11.8	0.6681	-0.775	0.685	1.969	0.431	-0.915	0.971	2.422	0.418	-0.924	0.978	2.451
12.0	1.1519	-0.915	0.180	2.163	0.427	-0.916	0.973	2.440	0.420	-0.919	0.987	2.430
12.2	0.6338	-0.768	0.677	1.823	0.444	-0.899	0.974	2.375	0.414	-0.929	0.990	2.430

Table B.3: Parameters of the analytical fits (Eq. 4.14) to the volume-weighted gas overdensity PDF extracted from IllustrisTNG100-2 simulation box from a snapshot at $z = 5.85$. $\log M$ indicates the mid-point of logarithmic mass bin with $\Delta \log M = 0.2$.

B.6 Dependence of the cross-correlation on the escape fraction

We find that the 2PCCF is most sensitive to the escape fraction. Surprisingly perhaps, decreasing the escape fraction increases the correlation at the redshift and opacity considered here. We now investigate this behaviour by looking back at our cross-correlation model

$$\begin{aligned}\xi^{\rm 2PCCF}_{\rm Gal-Ly\alpha}(r) &= \frac{P[< \Delta_b^{\rm max}(\langle\Gamma^{\rm CL}_{\rm HI}(r)\rangle)|r, M_h]}{P[< \Delta_b^{\rm max}(\bar{\Gamma}_{\rm HI})|r \to \infty, M_h]} - 1 \\ &= \frac{\int_0^{\Delta_b^{\rm max,\ CL}(r)} P_V(\Delta_b|r, M_h){\rm d}\Delta_b}{\int_0^{\overline{\Delta_b^{\rm max}}} P_V(\Delta_b|r, M_h){\rm d}\Delta_b} - 1 \ , \end{aligned} \tag{B.1}$$

where we have substituted $\Delta_b^{\rm max,\ CL}(r) = \Delta_b^{\rm max}(\langle\Gamma^{\rm CL}_{\rm HI}(r)\rangle)$ and $\overline{\Delta_b^{\rm max}} = \Delta_b^{\rm max}(\overline{\Gamma}_{\rm HI})$. The maximum gas overdensity at which a transmission spike is detected depends on the photoionisation rate $\Delta_b \propto \Gamma^{0.56}$, and the enhanced photoionisation rate is proportional to the mean UVB $\Gamma^{\rm CL}_{\rm HI}(r) \propto \overline{\Gamma} \times (1+\zeta(r))$, where we have subsumed the boosting effect of the clustering faint sources in one function $\zeta(r)$ for convenience. Therefore the maximum underdensity permitted in order to get a spike around a galaxy is proportional, but slightly higher, than at a random position. The cross-correlation than can be simplified as

$$\begin{aligned}&\xi^{\rm 2PCCF}_{\rm Gal-Ly\alpha}(r) \simeq \\ &\frac{\int_0^{\overline{\Delta_b^{\rm max}}} P_V(\Delta_b|r, M_h){\rm d}\Delta_b + \int_{\overline{\Delta_b^{\rm max}}}^{\Delta_b^{\rm max,\ CL}(r)} P_V(\Delta_b|r, M_h){\rm d}\Delta_b}{\int_0^{\overline{\Delta_b^{\rm max}}} P_V(\Delta_b|r, M_h){\rm d}\Delta_b} - 1 \\ &= \frac{\int_{\overline{\Delta_b^{\rm max}}}^{\Delta_b^{\rm max,\ CL}(r)} P_V(\Delta_b|r, M_h){\rm d}\Delta_b}{\int_0^{\overline{\Delta_b^{\rm max}}} P_V(\Delta_b|r, M_h){\rm d}\Delta_b} \ . \end{aligned} \tag{B.2}$$

When the escape fraction increases, it increases the photoionisation rate, in turn increasing both $\overline{\Delta_b^{\rm max}}$ and $\Delta_b^{\rm max,\ CL}(r) \propto \overline{\Delta_b^{\rm max}}(1+\zeta(r))^{0.56}$. To understand how the escape fraction impacts the cross-correlation, we consider how the nominator and the denominator of eq. B.2 react to a small change in Δ_b. At first order,

$$\int_{\overline{\Delta_b^{\rm max}}+\delta\Delta_b}^{\Delta_b^{\rm max,\ CL}(r)+\delta\Delta_b} P_V(\Delta_b|r,M_h)\mathrm{d}\Delta_b - \int_{\overline{\Delta_b^{\rm max}}}^{\Delta_b^{\rm max,\ CL}(r)} P_V(\Delta_b|r,M_h)\mathrm{d}\Delta_b$$
$$\simeq P_V(\Delta_b^{\rm max,\ CL}|r,M_h)\delta\Delta_b - P_V(\overline{\Delta_b^{\rm max}}|r,M_h)\delta\Delta_b \tag{B.3}$$

At $\Delta_b \lesssim 1$, which is the regime we probe, P_V is an increasing function, and therefore increasing the escape fraction increases the nominator. However, it is easy to see that the denominator increases by a greater amount

$$\int_0^{\overline{\Delta_b^{\rm max}}+\delta\Delta_b} P_V(\Delta_b|r,M_h)\mathrm{d}\Delta_b - \int_0^{\overline{\Delta_b^{\rm max}}} P_V(\Delta_b|r,M_h)\mathrm{d}\Delta_b$$
$$\simeq P_V(\Delta_b^{\rm max,\ CL}|r,M_h)\delta\Delta_b\ , \tag{B.4}$$

therefore the cross-correlation decreases as the mean UVB increases. In fact, the decrease rate depends on the average maximum gas overdensity to detect a spike, meaning that the cross-correlation is maximally sensitive to changes in the ionisation background when $\overline{\Delta_b^{\rm max}} \sim 0.1 - 1$. At $z \sim 5.5 - 6$ we are roughly in that range, although this could be improved with next generation instrument to reach larger opacities. This also implies that as IGM temperatures and the UVB increase at lower redshift, $\overline{\Delta_b^{\rm max}}$ becomes greater than 1 and the cross-correlation is insensitive to UVB fluctuations.

Appendix C

Appendices to Chapter 5

C.1 Line-derived redshifts for all quasars

I present the continuum splines fitted around the broad emission lines for each quasar in this study. Panels from left to right show broad emission lines arranged by increasing rest-frame wavelength, i.e. : O I, C II, Si IV, C IV, C III], Mg II. For each quasar and line, we show the observed flux (black), the fitted continuum splines (blue line) and the location of the peak if it satisfies the criteria of Section 5.2 (red vertical liens). For XQ100 quasars, we also show the peak recovered from the low-resolution version of the spectra (green vertical lines) and the one derived from the SDSS spectra if the object is present in SDSS (blue dotted vertical lines). Each BEL and fitted continuum are rescaled to facilitate the inspection. We thus caution that the lines are all at a different scale except for O I and C II which are rescaled by the same amount as they are contiguous in rest-frame wavelength. For the XQ100 quasars, we also show the continuum splines fitted to the resolution-degraded flux, and the derived peak positions thereof. The quasar name and literature redshift are indicated on the left of all panels. We give only the fits for the first 6 objects of the XQ100 sample to limit the length of this manuscript. The remaining plots are available as online supplementary material to the published paper (Meyer et al.

2019b).

Eventually, I present a summary table of our measured quasar redshift based on each BEL in Table C.1. The entire tables for each samples are also available as supplementary material to the online published paper (Meyer et al. 2019b).

C.2 Catastrophic redshift errors between our method and DR12Q tabulated values

C.3 Relative velocity shifts including Si IV and C III]

I report on Figure C.4 the relative velocity shifts of all lines, including the Si IV and C III] complexes. It must be noted that, because shifts are derived from the peak of the complex, the shift of these two complexes are mostly driven by the relative amplitude and thus abundance of the underlying emission lines evolving with redshift.

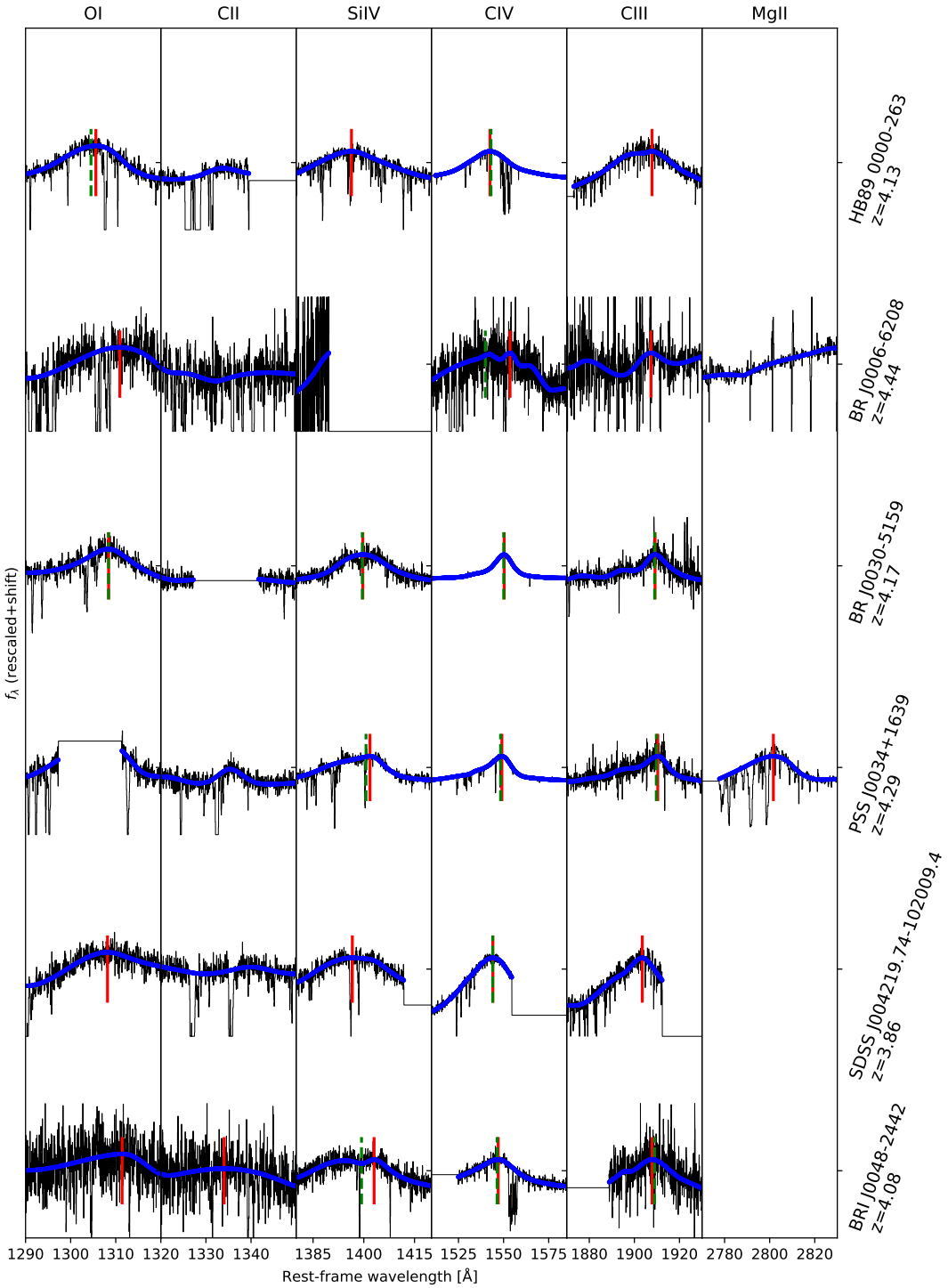

Figure C.1: XQ100 BELs fits given for the six lines of interest. The observed flux (black) is shown alongside the spline fits (blue) in the rest-frame computed from the literature redshift indicated on the vertical right axis. We indicate the location of peaks used to derived BEL-based redshifts by a vertical red line. We also indicate the peak solution derived from the low-resolution version of the spectra (vertical green line) and the SDSS spectra if available (vertical dotted blue).

Quasar	z_{OI}	z_{CII}	z_{SiIV}	z_{CIV}	$z_{\mathrm{CIII]}}$	z_{MgII}
SDSS Quasars						
5379-55986-0904	3.87	3.87	3.87	3.81	3.84	
4481-55630-0178	2.52		2.51	2.50	2.49	2.52
4193-55476-0376	2.28		2.26	2.26	2.25	2.27
⋮		⋮		⋮		⋮
XQ100 Quasars						
HB89 0000-263	4.13		4.11	4.10	4.12	
BR J0006-6208	4.46			4.45	4.44	
BR J0030-5159	4.18		4.17	4.18	4.17	
⋮		⋮		⋮		⋮
GGG Quasars						
SDSSJ0011+1446	4.96	4.97	4.95	4.94		
SDSSJ0822+1604			4.47	4.50		
SDSSJ1043+6506			4.48	4.46		
⋮		⋮		⋮		⋮
z6 Quasars						
J0148+0600	5.97	5.96	5.93	5.91	5.93	5.98
J0836+0054	5.80	5.80	5.81	5.78	5.76	5.77
J0927+2001	5.74	5.76	5.73	5.72	5.77	5.76
J1030+0524	6.30	6.27	6.30	6.28	6.31	6.30
J1306+0356	6.03	6.02	6.02	6.00	6.01	6.02
J1319+0950	6.10	6.11	6.17	6.03	6.14	6.12
J0100+2802	6.31		6.25	6.24	6.30	
J0818+1722					5.95	
J1509-1749	6.08	6.11	6.11	6.10	6.10	6.12
J1044-0125	5.78	5.80	5.77	5.76	5.79	5.78
J0231-0728	5.42	5.42	5.43	5.42	5.42	5.42
z7 Quasars						
J2348-3054	6.91	6.89	6.87	6.85	6.89	6.89
P231-20	6.57		6.68	6.47	6.54	6.60
P167-13	6.50	6.49	6.41	6.41	6.48	6.51
P036+03			6.45	6.42	6.50	6.53
J0305-3150	6.56		6.57	6.57		6.61
P183+05		6.48	6.40	6.35		6.42
J1120+0641		7.09	7.04	7.01	7.06	7.09
J1342+0928	7.56	7.55	7.41	7.35		7.52
P247+24	6.45	6.46	6.44	6.42		6.48
P338+29		6.64	6.70	6.63	6.64	
P323+12	6.61	6.58	6.59	6.58	6.59	6.58
J0109-3047	6.77	6.75	6.76	6.69		6.74

Table C.1: BEL-based redshifts for all our objects. SDSS quasars are designated by their plate-mjd-fiber identification to make their search easier in the SDSS database. The full table is published as an online supplementary material.

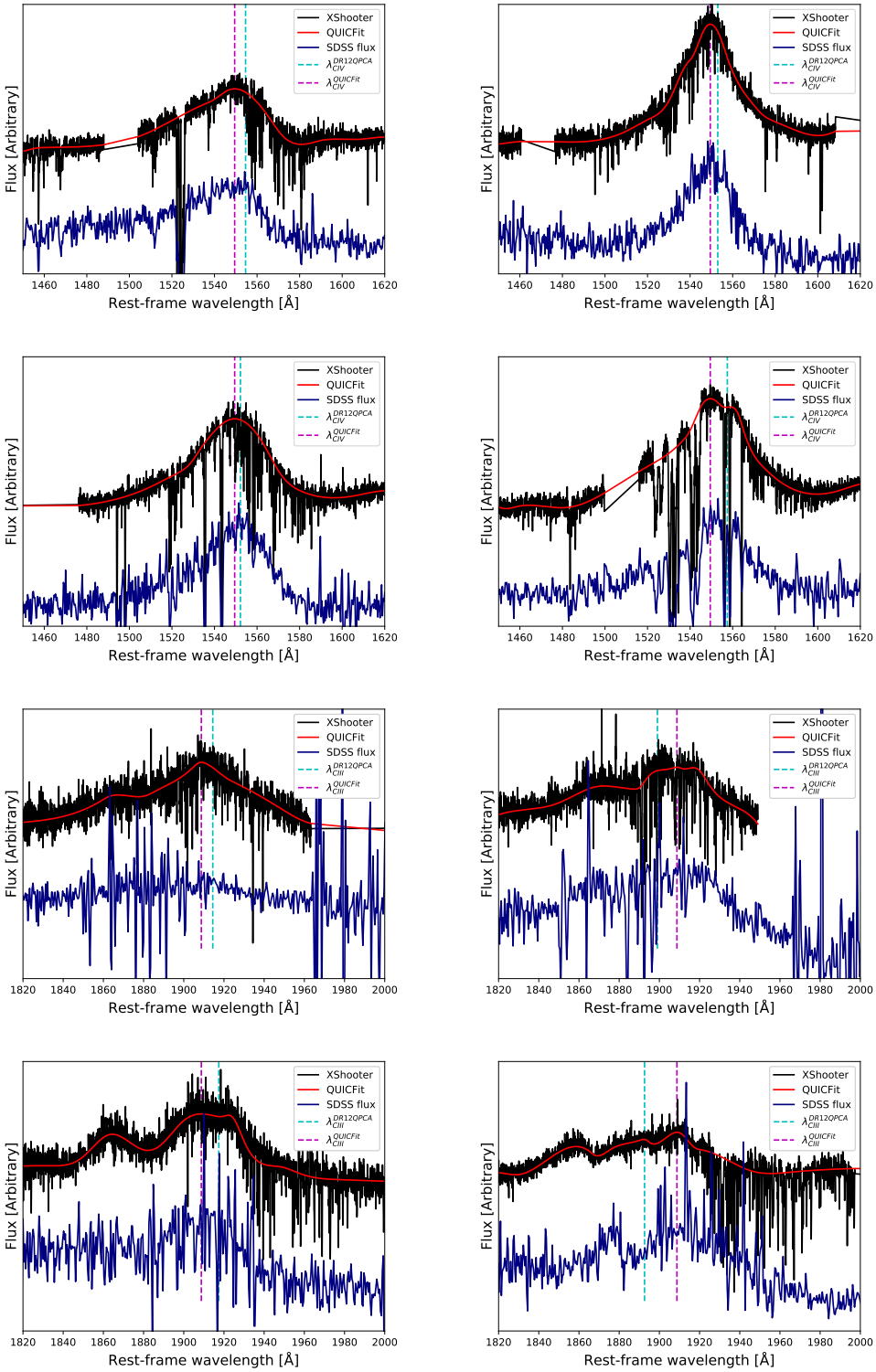

Figure C.2: Quasars in the XQ100-SDSS overlap for which the difference between the DR12Q PCA-based redshift and our solution for the C IV or the C III] line exceeds 500 km s $^{-1}$ (see Figure 5.10).

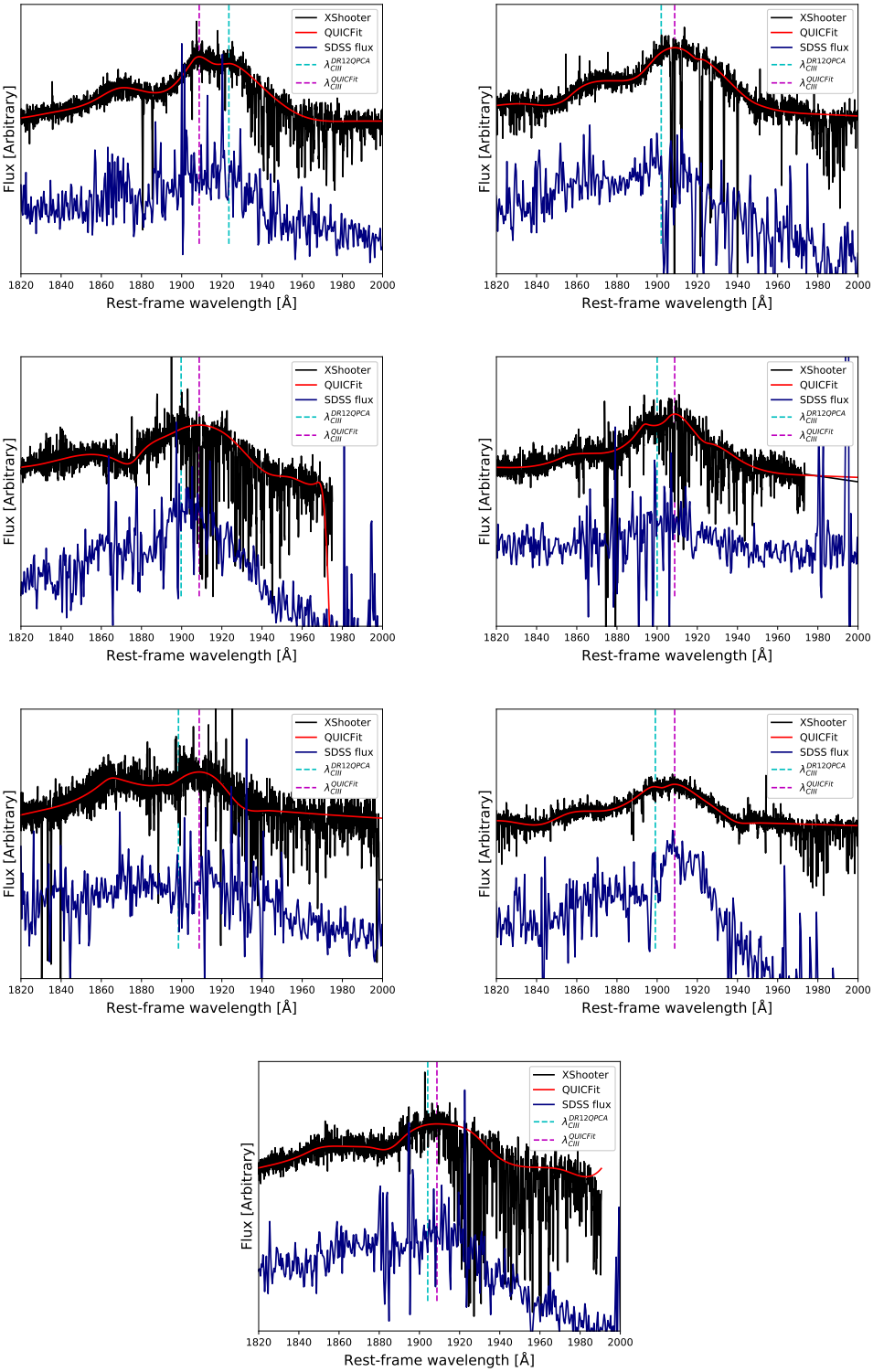

Figure C.3: Quasars in the XQ100-SDSS overlap for which the difference between the DR12Q PCA-based redshift and our solution for the C III]line exceeds 500 km s $^{-1}$ (see Figure 5.10).

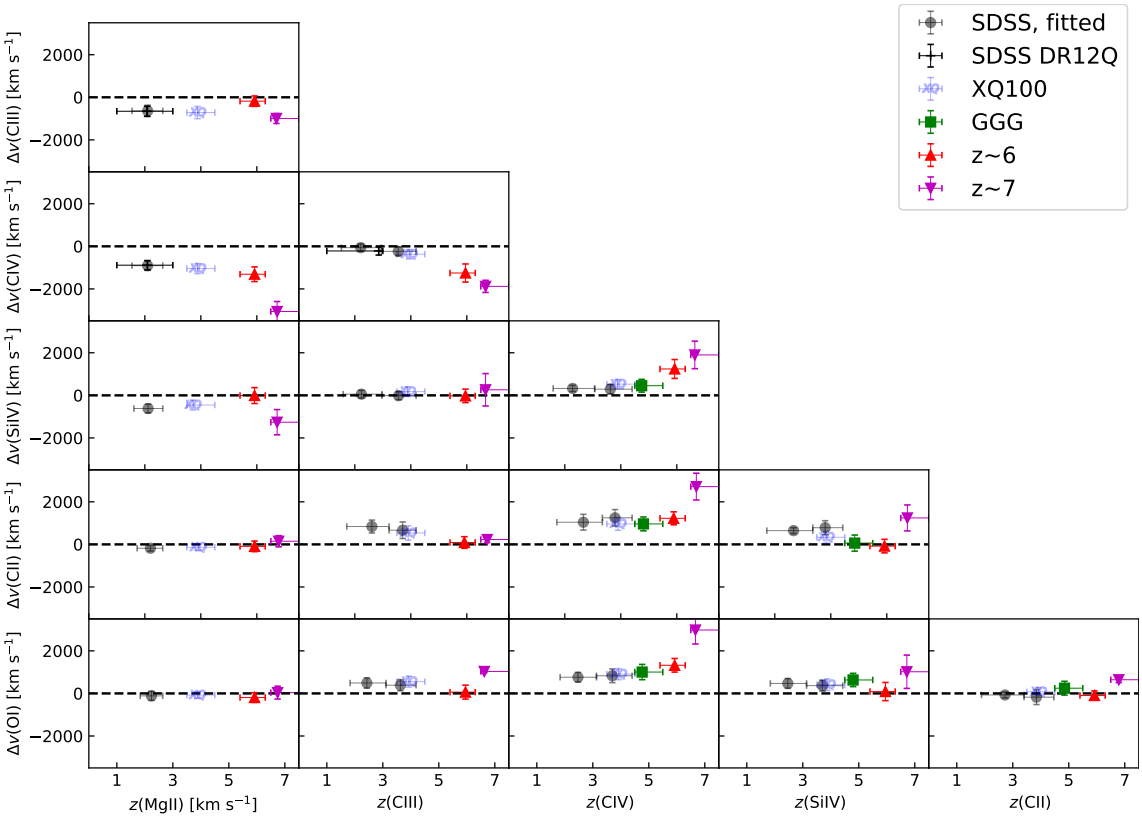

Figure C.4: Velocity shifts of rest-frame UV BELs across redshift. The errors are computed by bootstrapping using samples size of the z6 sample to make the errors comparable.

C.4 Selecting lensed quasar candidates with deep neural networks

In this appendix, I detail the architecture and training of the neural network used to search for lensed quasars in Section 5.5. The adopted model is an adaptation of the YOLO (You Only Look Once) neural network architecture (Redmon et al. 2015; Redmon & Farhadi 2016, 2018). This algorithm is designed for, and has been successfully applied to, fast labelling of objects in pictures, enabling near-instant labelling of multiple objects in videos. In this case, this fast application can be leveraged to scan a significant fraction of the DES imaging data in a few days. State-of-the art labelling neural networks such as Faster R-CNN (e.g. Ren et al. 2015) can process $\sim$ 1 $256 \times 256 \times 3$ (RGB) frame per second, whereas YOLO networks can process up to 100 frames per second depending on their architecture (Redmon & Farhadi 2016). Speed is crucial in the quasar candidate search problem because DES is composed of $> 10^4$ tiles which contain each $10^4 \times 10^4$ pixels times 5 colours. Data mining the entire DES imaging data with a R-CNN would take ~ 300 days, against a mere three with a YOLO network.

The principle of a YOLO network is to simultaneously detect objects in an image (image recognition), determine their extent (image segmentation) and label each object (image classification). These problems are often solved with separate networks, at the expense of longer training and testing time. The idea is that the network divides the input picture in $M \times M$ anchoring regions. In each region, it finds B bounding boxes for which it outputs the relative offset to the anchoring point x, y,the width of the box σ_x, σ_y, and the confidence that an object is in the box. Then, it determines the class probability in the region for each class C. The bounding boxes with high confidence are then retained and labelled according to the highest class probability averaged over all the regions they span. In summary, the YOLO network takes as input a $N \times N$ pixel colour image, and returns a $C+B*(2+2+1)$-dimensional vector in a $M \times M$ grid, with $(M < N)$. C is the number of classes in the problem, B is the number of allowed bounding boxes in each square defined by the $M \times M$

grid for which two coordinates, two widths estimates and one confidence score are given.

In this case, we can simplify the problem by using fixed size boxes, and only one class (quasar). To increase the speed, the input are 100×100 pixel cut-outs of a $10^4 \times 10^4$ pixels DES tiles in 5 colour filters ($grizY$), and two bounding boxes for each 10×10 pixels anchoring region ($M = 10$), for an output dimension $10 \times 10 \times 7$. The YOLO architecture adopted after a few numerical experiments is composed of convolutional layers with filter size 3×3 pixels and maxpooling layers[1] with pooling kernel size 2×2 pixels (Figure C.5). The architecture follows that of Redmon et al. (2015); Redmon & Farhadi (2016, 2018) but was shortened as much as possible to reduce training time. The convolutional layers use a LeakyReLU activation function (e.g. Maas et al. 2013)

$$\phi(x) = \begin{cases} \alpha x & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases}, \tag{C.1}$$

with $\alpha = 0.1$ and are followed by a batch normalisation layer which stabilises the training of the neural network. Dropout layers were tested as well, but resulted in poorer results. The last two convolutional layers are not followed by a batch normalisation and the last activation function is a sigmoid $\phi(x) = e^x/(e^x + 1)$. A residual connection is added between the penultimate block and the final two convolutional layers to improve the performance and stability of the deep neural network (e.g. He et al. 2015).

The loss function used to train the network is similar to that of Redmon et al.

[1]Maxpooling layers downsample the input images by taking the maximum pixel in kernel. In our case, the kernel is 2×2 pixels large so a $N \times N$ input images will be split in 2×2 squares, and only the maximum pixel will be retained, giving the output $N/2 \times N/2$ dimensions.

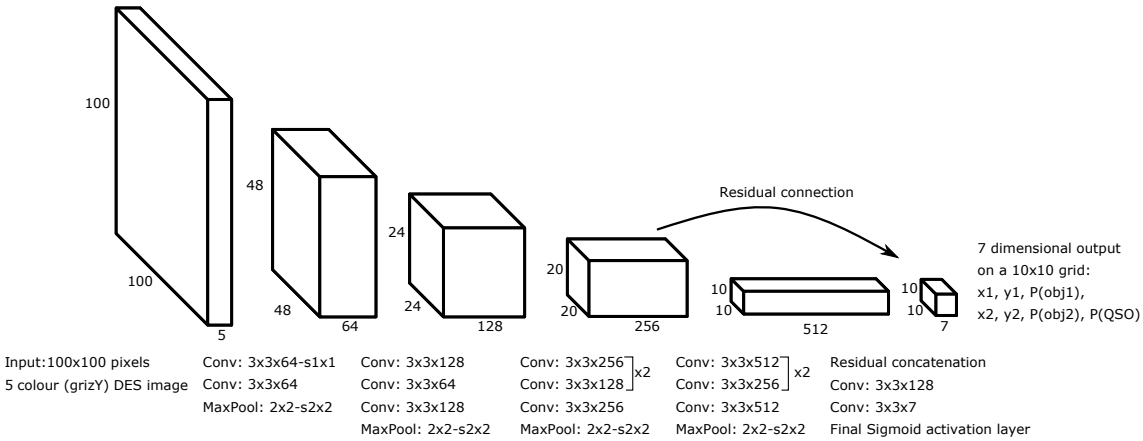

Figure C.5: Implemented YOLO network architecture for quasar searches. The convolutional layers are coded $A \times B \times C$ where A, B are the x, y dimensions of the convolutional filters, and C the number of such filters. $-sC \times C$ indicates that a stride is applied when performing the convolution of the data with the filters, with offset C pixels. The Maxpooling layers only have a pooling area and a striding pattern. Each convolution is followed by a LeakyReLU activation and a BatchNormalisation layer.

(2015)

$$\begin{aligned}\mathcal{L} = &\lambda_{\text{coord}} \Sigma_i \Sigma_j ||\mathbf{x}_{ij} - \hat{\mathbf{x}}_{ij}||^2 \\ &- \lambda_{\text{class}} \Sigma_i \mathbb{1}_i^{\text{obj}} \log \hat{p}(C_i) p(C_i) \\ &+ \lambda_{\text{obj}} \Sigma_i \Sigma_j \mathbb{1}_{ij}^{\text{obj}} (O_{ij} - \hat{O}_{ij})^2 + \lambda_{\text{noobj}} \Sigma_i \Sigma_j \mathbb{1}_{ij}^{\text{noobj}} (O_{ij} - \hat{O}_{ij})^2 \ , \end{aligned} \tag{C.2}$$

where the predicted parameters are denoted with a hat and the truth values without, and I now explain each term and the associated variables one by one. The first term is the quadratic error on the predicted coordinates of object j around anchoring point i. In this case, the sum on j only runs to $j = 1, 2$. The coordinates are normalised to $[0, 1]$ to be predicted by the sigmoid function. The second term in the loss function is the cross-entropy loss function for the class prediction (quasar or not) where $p(c_i)$ is the probability the 10x10 pixel region i contains a quasar (class c). This term is conditional on the presence of a object in region i ($\mathbb{1}_j^{\text{obj}} = 1$). The third term is a quadratic error on the confidence that an object is detected, $O_{ij} = 1$. This term is split in two parts when an object j is present in region i ($\mathbb{1}_{ij}^{\text{obj}} = 1$) or not ($\mathbb{1}_{ij}^{\text{noobj}} = 1$). The two separate coefficients $\lambda_{\text{obj}}, \lambda_{\text{noobj}}$ can be used to penalise more false positives or false negatives depending on the intended use of

the network. In this particular implementation, there is only one class (quasar), hence training the network to separately recognise an object and identify quasars might seem superfluous. However, the different loss functions (quadratic loss vs cross-entropy) are adapted to these specific tasks and I find that having both increase the performance.

The YOLO network was trained on a single nVidia Tesla P100 GPU on the Myriad cluster at UCL [2]. Different setups were tested by varying the coefficients of the loss functions and using different learning rate algorithms (also called optimisers).The training set was drawn from 50 DES tiles with RA¿20h since I aimed to select quasars with 00h $<$ RA $<$ 06h as backup targets for the relaxed dropouts described previously. The 50 DES tiles contain $10^4 \times 10^4$ pixels and were divided in 10^4 images of 100×100 pixels, for a final dataset size of 5×10^5 images. In each image, up to three quasars were randomly inserted so that they could be superposed with any foreground object (or none). The superimposed quasars were randomly drawn from the $z > 6$ quasars (Bosman et al. 2020) detected in the DES survey. The network was trained using a batch size of 64 and I achieved $\sim$ 150 epochs in an hour on the aforementioned GPU.

Neural networks are difficult to train, with multiple hyperparameters (number of layers, learning rate, loss function coefficients, activation functions) to fine-tune. For the number of layers and activations functions, I follow the proposed implementation of the original YOLO paper series (Redmon et al. 2015; Redmon & Farhadi 2016, 2018). Setting the learning rate during the training of a neural network is a complex problem: if the network updates its weights too much after each batch it is trained on, convergence will never be achieved. Decreasing the learning rate as time goes is thus necessary, but doing so too rapidly can leave the network unable to escape a local minimum of the loss function. Our first numerical experiment consists of testing various learning rate optimizers: Adam, Adamax (Kingma & Ba 2015), Adadelta (Zeiler 2012) & Nadam (Dozat 2016). I find that the network never converges when using Adam or Nadam, which are therefore excluded (Figure C.6, right). In the

[2] https://www.rc.ucl.ac.uk/docs/Clusters/Myriad/

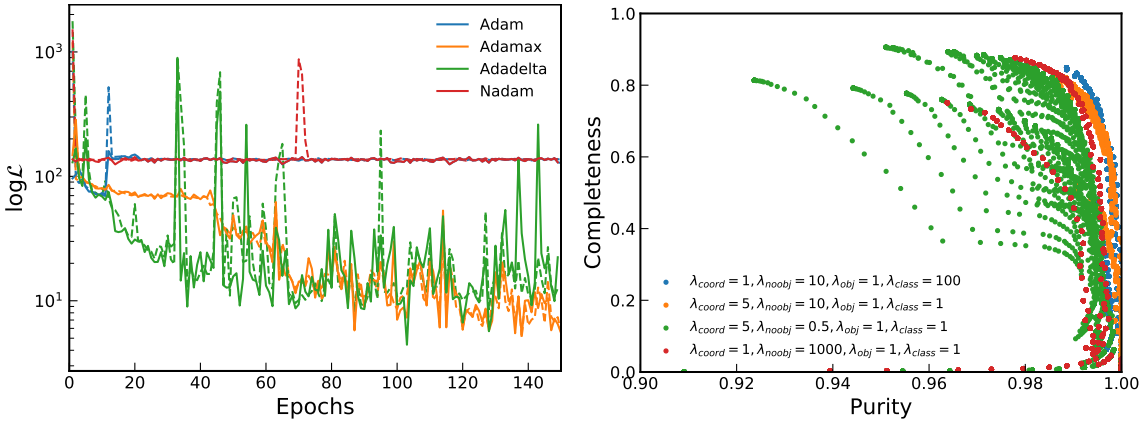

Figure C.6: Hyperparameter selection for the quasar-searching YOLO network **Left panel:** Training loss (dotted lines) and validation loss (full lines) of the YOLO network using four different optimiser, with only Adamax and Adadelta converging. **Right panel:** Completeness and purity of the selected sample using various loss function coefficients (the models are all optimized with Adamax for 150 epochs). The multiple points for each coefficients sample are obtained by varying the class and object probability threshold.

second numerical experiment, the loss coefficients are varied to see the impact on the performance of the trained network. The loss coefficients of Eq. C.2 can be varied to penalise the network more for mis-classifying objects ($\lambda_{\rm class}$), missing or finding too many objects ($\lambda_{\rm obj}, \lambda_{\rm noobj}$) or failing to precisely locate precisely the objects ($\lambda_{\rm coord}$). A full exploration of this 4-dimensional hyperparameter space is beyond the scope of this work, but various cases putting the emphasis on each of the above characteristics were tested. I assess the resulting network by computing the completeness (fraction of quasars selected) and purity (fraction of non-quasars selected) on three DES tiles in which 3×10^4 $z > 5.8$ quasars were added (see Figure C.6, right). The best loss coefficients (of those explored) are $\lambda_{\rm obj} = 10, \lambda_{\rm noobj} = 1, \lambda_{\rm coord} = 1, \lambda$class $= 100$, e.g. I penalize the network more for false negatives than false positives, which explains the high purity in the selected samples.

The final selection was made using an Adamax- and an Adadelta-trained network, and only candidates selected by both networks were retained. I checked that such configuration did detect the (Reed et al. 2017) quasars without too many contaminants ($0-5$ contaminants per DES tile containing one real quasar). Handling the entirety of the DES imaging data requires too much computer power for such an exploratory project designed initially to provide a back-up target list. The YOLO

network was therefore applied only to $i - z > 1.5$, $z - Y < 1$ dropouts in the DES catalog with limited cuts on the g, r bands ($g < 99, r < 99$) to allow for lensed quasars or foreground object contamination. The number of dropout examined by the network is $N_{\rm final} = 10043$, and the magnitudes were computed using fixed 2" aperture magnitudes. The first one hundred best-ranked candidates were visually inspected by two astronomers to remove artifacts, with 90% considered to be real objects.

C.5 Full list of observed candidates in the lensed quasar search observing run

I present in Table C.2 all the objects observed during the observing run at EFOSC/NTT in November 2019 to confirm lensed quasars. Coordinates, magnitudes, selection methods and type (cool dwarf or quasar) are given in the appropriate columns.

Table C.2: Name, coordinates and properties of the candidates observed in November 2019 with EFOSC2/NTT. As described in Section 5.5, the candidates were selected using relaxed colour cuts (method 1) and machine learning (method 2). Most candidates observed were cool dwarfs (CD), but two objects were confirmed as quasars.

Name	RA	DEC	$g_{\rm AB}$	σ_g	$r_{\rm AB}$	σ_r	$i_{\rm AB}$	σ_i	$z_{\rm AB}$	σ_z	$Y_{\rm AB}$	σ_Y	$J_{\rm Vega}$	σ_J	Type	OBS
DESJ000740.6-424539.6	1.91899	-42.760995	99	99	23.93	0.39	22.27	0.15	20.46	0.05	21.91	0.5	20.26	0.23	1	CD
DESJ004410.7-185722.2	11.044734	-18.956172	99	99	24.57	0.47	21.22	0.04	19.7	0.02	19.26	0.03	17.35	0.03	2	CD
DESJ004441.7-430217.5	11.173709	-43.038183	25.78	0.88	24.29	0.32	22.5	0.11	20.8	0.05	20.51	0.07	18.78	0.06	1	CD
DESJ005408.5-323237.5	13.535247	-32.54375	25.42	0.49	24.91	0.51	22.87	0.11	21.27	0.06	20.94	0.1			2	CD
DESJ010632.1-174107.5	16.633819	-17.685419	99	99	25.79	1.1	22.89	0.22	21.16	0.07	21.02	0.28	19.32	0.07	2	CD
DESJ010801.4-271307.5	17.00603	-27.218744	26.65	1.54	24.82	0.39	22.32	0.08	20.12	0.02	19.58	0.03			2	CD
DESJ011958.4-401932.6	19.993287	-40.325736	99	99	24.99	0.51	23.13	0.14	21.32	0.05	20.86	0.09	18.98	0.05	2	CD
DESJ014425.0-555225.4	26.10399	-55.873732	25.44	0.5	24.69	0.42	22.35	0.08	20.46	0.03	19.74	0.04	17.43	0.02	2	CD
DESJ015606.8-181651.6	29.028512	-18.280997	22.47	0.08	21.94	0.06	21.95	0.12	20.88	0.07	21.88	0.69			2	CD
DESJ020033.3-173725.9	30.138769	-17.623872	25.81	0.95	24.89	0.53	23.33	0.2	21	0.04	21.36	0.22	20.05	0.15	2	QSO
DESJ021204.9-445611.2	33.020593	-44.936437	25.81	0.87	25.08	0.64	21.78	0.05	20.3	0.02	20	0.06	18.14	0.03	2	CD
DESJ021551.8-002634.4	33.96599	-0.442891	24.7	0.41	26.31	3.29	22.89	0.21	21.18	0.09	21.1	0.2	19.59	0.1	1	CD
DESJ021824.6-502624.5	34.602433	-50.440131	26.16	0.75	27.81	4.67	25.57	1.08	21.23	0.04	20.22	0.05	17.65	0.03	2	CD
DESJ022653.8-230922.5	36.723976	-23.156247	25.64	0.59	23.85	0.14	21.89	0.04	20.45	0.02	19.99	0.05	17.78	0.02	2	CD
DESJ025226.7-023720.7	43.111128	-2.622415	25.78	0.74	27.32	4.05	24.5	0.57	21.16	0.04	21.27	0.16	19.9	0.15	2	QSO
DESJ025639.1-192000.7	44.16291	-19.33353	26.23	1.36	24.66	0.49	22.44	0.1	20.85	0.05	20.37	0.12	18.61	0.05	2	CD
DESJ030844.3-524731.6	47.184557	-52.792099	26.46	1.33	24.59	0.31	22.67	0.1	21.19	0.06	20.76	0.14	18.94	0.05	2	CD
DESJ033227.5-364038.2	53.114404	-36.677273	27.34	2.64	26	1.24	21.98	0.05	20.51	0.03	20.24	0.08	18.57	0.04	2	CD
DESJ040325.6-605346.7	60.856521	-60.896302	24.73	0.35	24.04	0.26	23.19	0.33	21.43	0.09	20.99	0.17	20.17	0.2	1	CD
DESJ041355.3-595138.9	63.480434	-59.860805	24.56	0.26	24.26	0.27	23.12	0.15	20.81	0.03	19.97	0.06	17.98	0.04	2	CD
DESJ041545.4-622918.1	63.939128	-62.488358	24.85	0.49	24.66	0.64	22.34	0.1	20.57	0.02	20.13	0.08	18.23	0.03	2	CD
DESJ041850.3-625328.1	64.70959	-62.891151	24.16	0.3	23.79	0.22	23.22	0.38	21.45	0.07	21.05	0.22	19.53	0.12	1	CD
DESJ042105.8-652834.0	65.274317	-65.476124	23.67	0.2	22.7	0.1	22.2	0.12	20.57	0.06	19.98	0.12	17.7	0.02	2	CD
DESJ042122.6-465247.0	65.343989	-46.879726	99	99	99	99	22.92	0.31	21.33	0.14	20.81	0.29	19.24	0.09	2	CD
DESJ043056.7-631242.0	67.736308	-63.211657	24.82	0.72	24.84	0.95	22.8	0.24	20.97	0.05	21.81	0.65	20.27	0.19	1	CD
DESJ043108.7-570441.8	67.786051	-57.078281	99	99	24.76	0.3	23.23	0.13	21.47	0.06	21.08	0.13	19.4	0.06	1	CD
DESJ044043.1-344613.4	70.179567	-34.770399	24.04	0.38	22.88	0.09	21.75	0.07	19.26	0.01	19.79	0.05	20.09	0.19	2	CD
DESJ045143.3-312331.3	72.930311	-31.392037	99	99	25.36	0.73	24.71	0.73	20.59	0.03	19.72	0.05	17.6	0.02	2	CD
DESJ050104.7-253234.2	75.269784	-25.542839	25.52	0.5	24.48	0.25	22.39	0.08	20.97	0.04	20.7	0.1	18.7	0.04	2	CD
DESJ050535.6-432922.5	76.398414	-43.489587	27.54	4.47	26.25	1.72	22.35	0.09	20.8	0.04	20.58	0.13	18.79	0.06	2	CD
DESJ051940.1-430328.9	79.917115	-43.058026	26.04	0.83	25.26	0.51	22.77	0.09	21.2	0.05	20.6	0.12	19.1	0.11	2	CD
DESJ052254.4-481015.6	80.726696	-48.171012	23.93	0.14	22.62	0.06	20.94	0.02	19.57	0.01	19.02	0.03	17.1	0.01	2	CD
DESJ053239.2-245528.2	83.163318	-24.924488	99	99	24.74	0.33	22.5	0.08	20.9	0.05	20.44	0.07			2	CD
DESJ054103.5-302047.5	85.264685	-30.346518	26.99	2.12	24.97	0.5	22.02	0.06	20.16	0.02	19.63	0.05	17.84	0.03	2	CD
DESJ060625.1-555814.1	91.604569	-55.970573	24.85	0.24	25.9	0.91	22.7	0.09	21.17	0.05	20.96	0.09			2	CD

Bibliography

Abbott T. M. C., et al., 2018, The Astrophysical Journal Supplement Series, 239, 18 1.4.1, 5.3, 5.5.1

Adams N. J., Bowler R. A. A., Jarvis M. J., Häußler B., McLure R. J., Bunker A., Dunlop J. S., Verma A., 2020, Monthly Notices of the Royal Astronomical Society, 494, 1771 1.4.3

Adelberger K. L., Steidel C. C., Shapley A. E., Pettini M., 2003, The Astrophysical Journal, 584, 45 3.1, 3.3.2, 3.7, 3.5.3

Adelberger K. L., Shapley A. E., Steidel C. C., Pettini M., Erb D. K., Reddy N. A., 2005, The Astrophysical Journal, 629, 636 3.1, 3.3.2, 3.7

Ajiki M., et al., 2006, Publications of the Astronomical Society of Japan, 58, 499 4.2.2

Arnouts S., Cristiani S., Moscardini L., Matarrese S., Lucchin F., Fontana A., Giallongo E., 1999, Monthly Notices of the Royal Astronomical Society, 310, 540 5.5.3

Arrabal Haro P., et al., 2018, Monthly Notices of the Royal Astronomical Society, 478, 3740 1.3.3

Atek H., Richard J., Kneib J.-P., Schaerer D., 2018, Monthly Notices of the Royal Astronomical Society, 479, 5184 1.4.3, 1.4.3, 1.11, 4.6, 5.5

Bacon R., et al., 2010, in McLean I. S., Ramsay S. K., Takami H., eds, Vol. 7735, Ground-based and Airborne Instrumentation for Astronomy III. International Society for Optics and Photonics, p. 773508, doi:10.1117/12.856027, http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.856027 4.2

Bacon R., et al., 2015, Astronomy & Astrophysics, 575, A75 4.2.3, 4.5

Bacon R., Piqueras L., Conseil S., Richard J., Shepherd M., 2016, Astrophysics Source Code Library, record ascl:1611.003 4.2.3

Baldwin J. A., 1977, The Astrophysical Journal, 214, 679 5.1

Bañados E., et al., 2016, The Astrophysical Journal Supplement Series, 227, 11 4.1, 5.1

Bañados E., et al., 2018, Nature, 553, 473 1.3.1, 1.6, 1.3.3, 1.4.1, 5.1, 5.2.1, 5.3.3, 5.4, 5.4.1, 5.4.2

Bañados E., et al., 2019, The Astrophysical Journal, 885, 59 1.3

Bassett R., et al., 2019, Monthly Notices of the Royal Astronomical Society, 483, 5223 1.4.5

Bautista J. E., et al., 2017, Astronomy & Astrophysics, 603, A12 3.4

Becker G. D., Bolton J. S., 2013, Monthly Notices of the Royal Astronomical Society, 436, 1023 1.5.3

Becker R. H., et al., 2001, The Astronomical Journal, 122, 2850 1.3.1

Becker G. D., Sargent W. L. W., Rauch M., Simcoe R. A., 2006, The Astrophysical Journal, 640, 69 3.1, 3.2.1, 3.5.2

Becker G. D., Rauch M., Sargent W. L. W., 2009, Astrophysical Journal, 698, 1010 3.1, 3.5, 5.1

Becker G. D., Sargent W. L., Rauch M., Calverley A. P., 2011, The Astrophysical Journal, 735, 93 3.2.2

Becker G. D., Sargent W. L. W., Rauch M., Carswell R. F., 2012, The Astrophysical Journal, 744, 91 3.2, 3.2.1

Becker G. D., Bolton J. S., Lidz A., 2015a, Publications of the Astronomical Society of Australia, 32, e045 1.3.1, 1.2, 1.5.3, 2.4.2, 2.4.2, 3.3.1, 4.5.1

Becker G. D., Bolton J. S., Madau P., Pettini M., Ryan-Weber E. V., Venemans B. P., 2015b, Monthly Notices of the Royal Astronomical Society, 447, 3402 1.5.2, 3.1, 3.1, 3.2.1

Becker G. D., Davies F. B., Furlanetto S. R., Malkan M. A., Boera E., Douglass C., 2018, The Astrophysical Journal, 863, 92 1.5.2, 4.3

Becker G. D., et al., 2019, The Astrophysical Journal, 883, 163 5.1

Beckwith S. V. W., et al., 2006, The Astronomical Journal, 132, 1729 1.4.2

Bertin E., Arnouts S., 1996, Astronomy and Astrophysics Supplement Series, 117, 393 5.5.1

Bielby R. M., et al., 2016, Monthly Notices of the Royal Astronomical Society, 456, 4061 4.5.3

Bielby R. M., et al., 2017, Monthly Notices of the Royal Astronomical Society, 471, 2174 3.3.2, 3.7

Bird S., Rubin K. H. R., Suresh J., Hernquist L., 2016, Monthly Notices of the Royal Astronomical Society, 462, 307 3.5.3

Bird S., Garnett R., Ho S., 2017, Monthly Notices of the Royal Astronomical Society, 466, 2111 3.1

Bischetti M., et al., 2017, Astronomy & Astrophysics, 598, A122 5.4

Blomqvist M., et al., 2015, Journal of Cosmology and Astroparticle Physics, 2015, 34 3.4

Boera E., Becker G. D., Bolton J. S., Nasir F., 2019, The Astrophysical Journal, 872, 101 1.5.3

Bolton A. S., Burles S., Schlegel D. J., Eisenstein D. J., Brinkmann J., 2004, The Astronomical Journal, 127, 1860 4.4.2

Bolton J. S., Haehnelt M. G., Warren S. J., Hewett P. C., Mortlock D. J., Venemans B. P., McMahon R. G., Simpson C., 2011, Monthly Notices of the Royal Astronomical Society: Letters, 416, L70 1.3.3

Bosman S. E. I., 2020, ZENODO, p. 10.5281/ZENODO.3634965 1.4.1

Bosman S. E. I., Becker G. D., 2015, Monthly Notices of the Royal Astronomical Society, 452, 1105 1.3.3, 5.1

Bosman S. E. I., Becker G. D., Haehnelt M. G., Hewett P. C., McMahon R. G., Mortlock D. J., Simpson C., Venemans B. P., 2017, Monthly Notices of the Royal Astronomical Society, 470, 1919 3.1, 3.2.2, 3.2.3, 3.5, 5.1

Bosman S. E., Fan X., Jiang L., Reed S., Matsuoka Y., Becker G., Haehnelt M., 2018, Monthly Notices of the Royal Astronomical Society, 479, 1055 (document), 1.3.1, 1.4.1, 1.5.2, 2.5.1, 3.1, 3.2, 3.2.1, 3.2.1, 1, 3.5.3, 4.2.1, 4.1, 4.4.1, 4.7.3, 5.1, 5.2.1

Bosman S. E. I., Kakiichi K., Meyer R. A., Gronke M., Laporte N., Ellis R. S., 2020, The Astrophysical Journal, 896, 49 1.5.3, 2.1, 2.6, 4.2, B.1, C.4

Boutsia K., et al., 2011, The Astrophysical Journal, Volume 736, Issue 1, article id. 41, 6 pp. (2011)., 736 1.4.5

Bouwens R. J., et al., 2015, The Astrophysical Journal, 803, 34 1.4.3, 1.9, 1.5.3, 2.3, 2.4.2, 2.4.2, 3.4, 3.4, 3.5.1, 4.4.1, 4.5, 4.6

Bouwens R. J., Smit R., Labbé I., Franx M., Caruana J., Oesch P. A., Stefanon M., Rasappu N., 2016, The Astrophysical Journal, 831, 176 1.4.4, 3.5.2

Bouwens R. J., Oesch P. A., Illingworth G. D., Ellis R. S., Stefanon M., 2017, The Astrophysical Journal, 843, 129 1.4.3, 1.4.3, 1.4.3, 5.5

Bowler R. A. A., et al., 2015, Monthly Notices of the Royal Astronomical Society, 452, 1817 4.4.1, 4.5

Bowler R. A. A., Bourne N., Dunlop J. S., McLure R. J., McLeod D. J., 2018, Monthly Notices of the Royal Astronomical Society, 481, 1631 1.4.3, 1.4.3

Bowler R., Jarvis M., Dunlop J., McLure R., McLeod D., Adams N., Milvang-Jensen B., McCracken H., 2020, Monthly Notices of the Royal Astronomical Society, 493, 2059 1.4.3

Bruzual G. A., Charlot S., 1993, The Astrophysical Journal, 405, 538 1.4.2

Bruzual G., Charlot S., 2003, Monthly Notices of the Royal Astronomical Society, 344, 1000 5.5.3

Cai Z., Fan X., Dave R., Finlator K., Oppenheimer B., 2017, The Astrophysical Journal, 849, L18 3.5.2

Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, The Astrophysical Journal, 533, 682 1.4.4, 1.4.4

Capak P. L., et al., 2015, Nature, 522, 455 1.4.4, 1.4.4

Carnall A. C., et al., 2015, Monthly Notices of the Royal Astronomical Society: Letters, 451, L16 5.1

Carnall A. C., McLure R. J., Dunlop J. S., Davé R., 2018, Monthly Notices of the Royal Astronomical Society, 480, 4379 2.4.1

Carswell R., Webb J., 2014, VPFIT: Voigt profile fitting program 3.2.3

Caruana J., Bunker A. J., Wilkins S. M., Stanway E. R., Lacy M., Jarvis M. J., Lorenzoni S., Hickey S., 2012, Monthly Notices of the Royal Astronomical Society, 427, 3055 1.3.3

Caruana J., et al., 2018, Monthly Notices of the Royal Astronomical Society, 473, 30 1.3.3

Cen R., Haiman Z., 2000, The Astrophysical Journal, 542, L75 2.4.2

Chambers K. C., et al., 2016, arXiv e-prints: 1612.05560 1.4.1

Chardin J., Haehnelt M. G., Aubert D., Puchwein E., 2015, Monthly Notices of the Royal Astronomical Society, 453, 2944 1.5.1, 1.5.2

Chardin J., Puchwein E., Haehnelt M. G., 2017, Monthly Notices of the Royal Astronomical Society, 465, 3429 1.4.1, 1.5.1, 1.5.2, 4.5.2

Chardin J., Uhlrich G., Aubert D., Deparis N., Gillet N., Ocvirk P., Lewis J., 2019, Monthly Notices of the Royal Astronomical Society, 490, 1055 1.5.1

Chisholm J., et al., 2018, Astronomy & Astrophysics, 616, A30 1.5.3

Choudhury T. R., Paranjape A., Bosman S. E., 2020, arXiv e-prints: 2003.08958 1.5.1

Coatman L., Hewett P. C., Banerji M., Richards G. T., 2016, Monthly Notices of the Royal Astronomical Society, 461, 647 5.1, 5.2.2, 5.9, 5.3.3

Coatman L., Hewett P. C., Banerji M., Richards G. T., Hennawi J. F., Prochaska J. X., 2019, Monthly Notices of the Royal Astronomical Society, 486, 5335 5.4

Codoreanu A., Ryan-Weber E. V., García L. Á., Crighton N. H. M., Becker G. D., Pettini M., Madau P., Venemans B., 2018, Monthly Notices of the Royal Astronomical Society, 481, 4940 3.1, 3.2.2, 5.1

Coe D., et al., 2019, The Astrophysical Journal, 884, 85 1.4.2

Cooper M. C., Newman J. A., Davis M., Finkbeiner D. P., Gerke B. F., 2012, Astrophysics Source Code Library 4.2.2

Cooper T. J., Simcoe R. A., Cooksey K. L., Bordoloi R., Miller D. R., Furesz G., Turner M. L., Bañados E., 2019, The Astrophysical Journal, 882, 77 3.1

Crighton N. H. M., et al., 2011, Monthly Notices of the Royal Astronomical Society, 414, 28 3.2.1

Cristiani S., Serrano L. M., Fontanot F., Vanzella E., Monaco P., 2016, Monthly Notices of the Royal Astronomical Society, 462, 2478 1.4.1

D'Aloisio A., Upton Sanderbeck P. R., McQuinn M., Trac H., Shapiro P. R., 2017, Monthly Notices of the Royal Astronomical Society, 468, 4691 1.4.1, 1.5.1

D'Aloisio A., McQuinn M., Maupin O., Davies F. B., Trac H., Fuller S., Upton Sanderbeck P. R., 2019, The Astrophysical Journal, 874, 154 1.5.3

D'Odorico V., Calura F., Cristiani S., Viel M., 2010, Monthly Notices of the Royal Astronomical Society, 401, 2715 3.1, 3.5

D'Odorico V., et al., 2011, Astronomische Nachrichten, 332, 315 3.2

D'Odorico V., et al., 2013, Monthly Notices of the Royal Astronomical Society, 435, 1198 3.1, 3.2.2, 3.5, 5.1

D'Odorico V., et al., 2018, The Astrophysical Journal, 863, L29 3.1, 3.5.3

Davies F. B., Furlanetto S. R., 2016, Monthly Notices of the Royal Astronomical Society, 460, 1328 4.5.2

Davies F. B., Hennawi J. F., Eilers A.-C., Lukić Z., 2018a, The Astrophysical Journal, 855, 106 1.5.3

Davies F. B., et al., 2018b, The Astrophysical Journal, 864, 142 5.1

Davies F. B., et al., 2018c, The Astrophysical Journal, 864, 143 1.3.3, 5.1, 6.2

Davis M., Peebles P., 1983, The Astrophysical Journal, 267, 465 3.3.2

Dayal P., et al., 2020, Monthly Notices of the Royal Astronomical Society, 495, 3065 1.4.1

De Barros S., et al., 2017, Astronomy & Astrophysics, 608, A123 1.3.3, 4.4.1

De Rosa G., Decarli R., Walter F., Fan X., Jiang L., Kurk J., Pasquali A., Rix H. W., 2011, The Astrophysical Journal, 739, 56 3.2

De Rosa G., et al., 2014, The Astrophysical Journal, 790, 145 5.1, 5.3.3, 5.4, 5.4.1

Decarli R., Falomo R., Treves A., Kotilainen J. K., Labita M., Scarpa R., 2010, Monthly Notices of the Royal Astronomical Society, 402, 2441 5.2.2

Decarli R., et al., 2017, Nature, 545, 457 5.1

Denney K. D., 2012, The Astrophysical Journal, 759, 44 5.4, 5.4.1

Dey A., et al., 2019, The Astronomical Journal, 157, 168 1.4.1, 5.1

Díaz C. G., Ryan-Weber E. V., Cooke J., Pettini M., Madau P., 2011, Monthly Notices of the Royal Astronomical Society, 418, 820 3.5.2

Díaz C. G., Koyama Y., Ryan-Weber E. V., Cooke J., Ouchi M., Shimasaku K., Nakata F., 2014, Monthly Notices of the Royal Astronomical Society, 442, 946 3.5.2

Diaz G. C., Ryan-Weber E. V., Cooke J., Koyama Y., Ouchi M., 2015, Monthly Notices of the Royal Astronomical Society, 448, 1240 3.5.2, 4.4

Diaz G. C., Ryan-Weber E., Karman W., Caputi K., Salvadori S., Crighton N., Ouchi M., Vanzella E., 2020, arXiv e-prints: 2001.04453 3.1, 4.1, 4.4.1

Dijkstra M., 2014, Publications of the Astronomical Society of Australia, 31 1.3.1, 2.1, 2.2, 2.2

Dijkstra M., Haiman Z., Spaans M., 2006, The Astrophysical Journal, 649, 14 2.2, 2.2, 2.4.1

Dijkstra M., Mesinger A., Wyithe J. S. B., 2011, Monthly Notices of the Royal Astronomical Society, 414, 2139 2.2

Dozat T., 2016, ICLR Workshop, pp 2013–2016 C.4

Drake A. B., Farina E. P., Neeleman M., Walter F., Venemans B., Banados E., Mazzucchelli C., Decarli R., 2019, The Astrophysical Journal, 881, 131 4.1

Eilers A.-C., Davies F. B., Hennawi J. F., Prochaska J. X., Lukić Z., Mazzucchelli C., 2017, The Astrophysical Journal, 840, 24 3.1, 3.2.1, 4.2.1, 5.1, 5.4.2

Eilers A.-C., Davies F. B., Hennawi J. F., 2018a, The Astrophysical Journal, 864, 53 1.3.1, 1.5.2, 5.1

Eilers A.-C., Hennawi J. F., Davies F. B., 2018b, The Astrophysical Journal, 867, 30 5.4.2

Eilers A.-C., Hennawi J. F., Davies F. B., Oñorbe J., 2019, The Astrophysical Journal, 881, 23 1.3.1

Eisenstein D. J., et al., 2011, The Astronomical Journal, Volume 142, Issue 3, article id. 72, 24 pp. (2011)., 142 5.2.1

Eldridge J. J., Stanway E. R., Xiao L., McClelland L. A. S., Taylor G., Ng M., Greis S. M. L., Bray J. C., 2017, Publications of the Astronomical Society of Australia, 34, e058 5.5.3

Ellis R. S., et al., 2013, The Astrophysical Journal Letters, 763, L7 1.4.2

Elvis M., 2000, The Astrophysical Journal, 545, 63 5.4

Erb D. K., et al., 2014, The Astrophysical Journal, 795, 33 3.5.3

Faber S. M., et al., 2003. International Society for Optics and Photonics, p. 1657, doi:10.1117/12.460346, http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.460346 4.2, 4.2.2

Faisst A. L., 2016, The Astrophysical Journal, 829, 99 1.4.5

Fan X., et al., 2001, The Astronomical Journal, 122, 2833 1.3.1

Fan X., Narayanan V. K., Strauss M. A., White R. L., Becker R. H., Pentericci L., Rix H. W., 2002, The Astronomical Journal, 123, 1247 1.3.1

Fan X., et al., 2004, The Astronomical Journal, 128, 515 3.2

Fan X., et al., 2006, The Astronomical Journal, 132, 117 1.3.1, 3.2

Fan X., et al., 2019, The Astrophysical Journal, 870, L11 5.1, 5.5

Farina E. P., et al., 2017, The Astrophysical Journal, 848, 78 4.1

Farina E. P., et al., 2019, The Astrophysical Journal, 887, 196 4.1

Fathivavsari H., 2020, arXiv e-prints: 2006.05124 6.2

Faucher-Giguère C. A., Lidz A., Zaldarriaga M., Hernquist L., 2009, Astrophysical Journal, 703, 1416 1.5.1

Feng Y., Di-Matteo T., Croft R. A., Bird S., Battaglia N., Wilkins S., 2016, Monthly Notices of the Royal Astronomical Society, 455, 2778 1.5.1

Fermi-LAT Collaboration T. F.-L., 2018, Science (New York, N.Y.), 362, 1031 4.6

Finkelstein S. L., et al., 2013, Nature, 502, 524 2.4.1

Finkelstein S. L., et al., 2015, Astrophysical Journal, 810, 71 1.4.3, 4.4.1, 4.5

Finkelstein S. L., et al., 2019, The Astrophysical Journal, 879, 36 1.3.4, 1.4.1, 6.2

Finlator K., Keating L., Oppenheimer B. D., Davé R., Zackrisson E., 2018, Monthly Notices of the Royal Astronomical Society, 480, 2628 1.5.1

Fletcher T. J., Tang M., Robertson B. E., Nakajima K., Ellis R. S., Stark D. P., Inoue A., 2019, The Astrophysical Journal, 878, 87 1.4.5, 2.1, 4.7.1

Ford A. B., Oppenheimer B. D., Davé R., Katz N., Kollmeier J. A., Weinberg D. H., 2013, Monthly Notices of the Royal Astronomical Society, 432, 89 3.1

Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, Publications of the Astronomical Society of the Pacific, 125, 306 3.4, 4.6

Furlanetto S. R., Pritchard J. R., 2006, Monthly Notices of the Royal Astronomical Society, 372, 1093 2.2

Furlanetto S., Zaldarriaga M., Hernquist L., 2004, The Astrophysical Journal, 613, 1 1.5.1

Furlanetto S. R., Zaldarriaga M., Hernquist L., 2006, Monthly Notices of the Royal Astronomical Society, 365, 1012 1.3.3

Gallagher S. C., Everett J. E., Abado M. M., Keating S. K., 2015, Monthly Notices of the Royal Astronomical Society, 451, 2991 5.4

Garaldi E., Gnedin N. Y., Madau P., 2019, The Astrophysical Journal, 876, 31 1.5.2, 4.1, B.3

Gazagnes S., Chisholm J., Schaerer D., Verhamme A., Izotov Y., 2020, Astronomy & Astrophysics, 639, A85 1.4.5, 1.5.3, 1, 2.4.2, 2.5.1

Giallongo E., et al., 2008, Astronomy & Astrophysics, 482, 349 4.2.2

Giallongo E., et al., 2015, Astronomy & Astrophysics, 578, A83 1.4.1, 1.4.1, 5.4.2

Giallongo E., et al., 2019, The Astrophysical Journal, 884, 19 1.4.1, 1.4.1

Giavalisco M., et al., 2004, The Astrophysical Journal, 600, L93 1.4.2

Gnedin N. Y., 2014, The Astrophysical Journal, 793, 29 1.5.1

Gnedin N. Y., 2016, The Astrophysical Journal, 833, 66 1.4.3

Gnedin N. Y., Kaurov A. A., 2014, The Astrophysical Journal, 793, 30 1.5.1

Grazian A., et al., 2016, Astronomy and Astrophysics, 585 1.4.5

Grazian A., et al., 2018, Astronomy & Astrophysics, 613, A44 1.4.1

Greig B., Mesinger A., 2017, Monthly Notices of the Royal Astronomical Society, 465, 4838 6.2

Greig B., Mesinger A., McGreer I. D., Gallerani S., Haiman Z., 2017, Monthly Notices of the Royal Astronomical Society, 466, 1814 1.3.1, 1.3.3, 1.3.3, 5.1, 6.2

Greig B., Mesinger A., Bañados E., 2019, Monthly Notices of the Royal Astronomical Society, 484, 5094 6.2

Gronke M., 2017, Astronomy & Astrophysics, 608, A139 2.2, 2.4.1

Gronke M., Bull P., Dijkstra M., 2015, Astrophysical Journal, 812 2.2

Gronke M., et al., 2020, arXiv e-prints: 2004.14496 6.3

Gunn J. E., Peterson B. A., 1965, The Astrophysical Journal, 142, 1633 1.3, 1.3.1

Gunn J. E., Stryker L. L., 1983, The Astrophysical Journal Supplement Series, 52, 121 4.1

Haardt F., Madau P., 2012, The Astrophysical Journal, 746, 125 1.5.1

Hamilton A. J. S., 1992, The Astrophysical Journal, 385, L5 3.4

Harikane Y., et al., 2016, The Astrophysical Journal, 821, 123 3.4, 4.5

Harikane Y., et al., 2018, Publications of the Astronomical Society of Japan, 70 1.5.3, 4.6

Harrington J. P., 1973, Monthly Notices of the Royal Astronomical Society, 162, 43 2.2

Hashimoto T., et al., 2015, Astrophysical Journal, 812 2.3, 2.6

Hashimoto T., et al., 2018a 1.4.2

Hashimoto T., et al., 2018b, Nature, 557, 392 2.3, 2.5.2, 4.2.4

Hawkins E., et al., 2003, Monthly Notices of the Royal Astronomical Society, 346, 78 4.5.3

He K., Zhang X., Ren S., Sun J., 2015, arXiv e-prints: 1512.03385 C.4

Herenz E. C., Wisotzki L., 2017, Astronomy & Astrophysics, 602, A111 4.2.3

Herenz E. C., et al., 2019, Astronomy & Astrophysics, 621, A107 4.2.3, 4.5, 4.4.1

Hewett P. C., Wild V., 2010, Monthly Notices of the Royal Astronomical Society, 405, 2302 5.1

Hewett P. C., Irwin M. J., Bunclark P., Bridgeland M. T., Kibblewhite E. J., He X. T., Smith M. G., 1985, Monthly Notices of the Royal Astronomical Society, 213, 971 4.4.2

Hill J. M., Salinari P., 2000. International Society for Optics and Photonics, pp 36–46, doi:10.1117/12.393947, http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=897593 4.2.2

Horne K., 1986, Publications of the Astronomical Society of the Pacific, 98, 609 2.3, 3.2.1, 4.2.2

Hu E. M., Cowie L. L., Songaila A., Barger A. J., Rosenwasser B., Wold I. G. B., 2016, The Astrophysical Journal, 825, L7 1.5.3, 2.1, 4.2.3

Hubble E. P., 1926, The Astrophysical Journal, 64, 321 1.1

Hui L., Gnedin N. Y., 1997, Monthly Notices of the Royal Astronomical Society, 292, 27 4.5.1

Ilbert O., et al., 2006, A&A, 457, 841 5.5.3

Inoue A. K., Iwata I., Deharveng J.-M., Buat V., Burgarella D., 2005, Astronomy & Astrophysics, 435, 471 1.4.5

Ishigaki M., Kawamata R., Ouchi M., Oguri M., Shimasaku K., Ono Y., 2018, The Astrophysical Journal, 854, 73 1.3.4, 1.4.3, 6.2

Izotov Y. I., Schaerer D., Thuan T. X., Worseck G., Guseva N. G., Orlitová I., Verhamme A., 2016a, Monthly Notices of the Royal Astronomical Society, 461, 3683 1.4.5

Izotov Y. I., Orlitová I., Schaerer D., Thuan T. X., Verhamme A., Guseva N. G., Worseck G., 2016b, Nature, 529, 178 1.4.5, 4.7.1

Izotov Y. I., Schaerer D., Worseck G., Guseva N. G., Thuan T. X., Verhamme A., Orlitová I., Fricke K. J., 2017, Monthly Notices of the Royal Astronomical Society, 474, 4514 1.4.5

Izotov Y. I., Worseck G., Schaerer D., Guseva N. G., Thuan T. X., Fricke Verhamme A., Orlitová I., 2018, Monthly Notices of the Royal Astronomical Society, 478, 4851 1.4.5, 2.1, 2.2, 2.3, 2.1, 2.4.1, 4.7.1

Jiang L., et al., 2016, The Astrophysical Journal, 833, 222 5.1, 5.19

Kaifu N., et al., 2000, Publications of the Astronomical Society of Japan, 52, 1 4.2.2

Kaiser N., 1987, Monthly Notices of the Royal Astronomical Society, 227, 1 3.4

Kaiser N., et al., 2010. International Society for Optics and Photonics, p. 77330E, doi:10.1117/12.859188, http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.859188 5.1

Kakiichi K., Gronke M., 2019, arXiv e-prints: 1905.02480 1.4.5, 2.1, 2.2

Kakiichi K., et al., 2018, Monthly Notices of the Royal Astronomical Society, 479, 43 1.13, 1.5.3, 1.5.3, 1.5.3, 1.14, 1.5.3, 1.5.3, 3.1, 3.3.2, 3.4, 3.4, 3.5, 3.5.2, 3.5.3, 3.6, 4.1, 4.2, 4.2.1, 4.1, 4.2.2, 4.2.2, 4.2, 4.4, 4.5, 4.9, 4.5, 4.5.1, 4.5.2, 4.12, 4.16, 4.8, B.3, B.10, B.14

Kaplinghat M., Chu M., Haiman Z., Holder G. P., Knox L., Skordis C., 2003, The Astrophysical Journal, 583, 24 1.3.2

Kashikawa N., et al., 2006, The Astrophysical Journal, 648, 7 2.3

Kashino D., Lilly S. J., Shibuya T., Ouchi M., Kashikawa N., 2019, The Astrophysical Journal, 888, 6 1.5.2, 4.3

Katz H., Kimm T., Haehnelt M. G., Sijacki D., Rosdahl J., Blaizot J., 2019a, Monthly Notices of the Royal Astronomical Society, 483, 1029 1.5.1

Katz H., Laporte N., Ellis R. S., Devriendt J., Slyz A., 2019b, Monthly Notices of the Royal Astronomical Society, 484, 4054 1.5.1

Keating L. C., Haehnelt M. G., Cantalupo S., Puchwein E., 2015, Monthly Notices of the Royal Astronomical Society, 454, 681 5.1

Keating L. C., Puchwein E., Haehnelt M. G., Bird S., Bolton J. S., 2016, Monthly Notices of the Royal Astronomical Society, 461, 606 3.1, 3.5.1, 3.5.3

Keating L. C., Weinberger L. H., Kulkarni G., Haehnelt M. G., Chardin J., Aubert D., 2020, Monthly Notices of the Royal Astronomical Society, 491, 1736 1.5.1, 1.5.2

Kelson D. D., 2003, PASP, 115, 688 3.2.1

Kennicutt R. C., 1998, Annual Review of Astronomy and Astrophysics, 36, 189 1.4.4, 2.4.2

Kimm T., Cen R., 2014, The Astrophysical Journal, 788, 121 1.4.5

King A., 1980, Physics Bulletin, 31, 359 1.3.1

Kingma D. P., Ba J. L., 2015, in 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings. International Conference on Learning Representations, ICLR (arXiv:1412.6980), https://arxiv.org/abs/1412.6980v9 C.4

Krolik J. H., Begelman M. C., 1986, The Astrophysical Journal, 308, L55 5.4

Kuhlen M., Faucher-Giguère C.-A., 2012, Monthly Notices of the Royal Astronomical Society, 423, 862 1.5.3, 2.4.2

Kulas K. R., Shapley A. E., Kollmeier J. A., Zheng Z., Steidel C. C., Hainline K. N., 2012, Astrophysical Journal, 745 2.3, 2.6

Kulkarni G., Keating L. C., Haehnelt M. G., Bosman S. E. I., Puchwein E., Chardin J., Aubert D., 2019a, Monthly Notices of the Royal Astronomical Society: Letters, 485, L24 1.3.3, 1.5.2

Kulkarni G., Worseck G., Hennawi J. F., 2019b, Monthly Notices of the Royal Astronomical Society, 488, 1035 1.4.1, 1.8, 1.5.1, 1.5.2, 5.1

Kusakabe H., et al., 2020, Astronomy & Astrophysics, 638, A12 1.3.3, 1.4.5

Labbé I., et al., 2013, Astrophysical Journal Letters, 777, L19 2.3

Laporte N., et al., 2017a, The Astrophysical Journal, 837, L21 2.5.2

Laporte N., Nakajima K., Ellis R. S., Zitrin A., Stark D. P., Mainali R., Roberts-Borsani G. W., 2017b, The Astrophysical Journal, 851, 40 1.4.4, 2.3, 2.5.2

Laporte N., et al., 2019, Monthly Notices of the Royal Astronomical Society: Letters, 487, L81 2.3

Laursen P., Sommer-Larsen J., Razoumov A. O., 2011, The Astrophysical Journal, 728, 52 1.5

Le Fèvre O., et al., 2015, Astronomy & Astrophysics, 576, A79 4.1

Lee K.-G., Suzuki N., Spergel D. N., 2012, The Astronomical Journal, 143, 51 5.1

Leethochawalit N., Jones T. A., Ellis R. S., Stark D. P., Zitrin A., 2016, The Astrophysical Journal, 831, 152 1.4.5, 1.5.3

Leitherer C., Heckman T. M., 1995, The Astrophysical Journal Supplement Series, 96, 9 1.4.4

Leitherer C., Ferguson H. C., Heckman T. M., Lowenthal J. D., 1995, The Astrophysical Journal, 454, L19 1.4.5

Liu B., Bordoloi R., 2020, arXiv e-prints: 2006.04814 6.2

Livermore R., Finkelstein S. L., Lotz J. M., 2017, The Astrophysical Journal, 835, 113 1.4.3, 1.4.3, 4.6

López-Gonzaga N., Jaffe W., 2016, Astronomy & Astrophysics, 591, A128 5.4

López S., et al., 2016, Astronomy & Astrophysics, 594, A91 5.2.1

Lotz J. M., et al., 2017, The Astrophysical Journal, 837, 97 1.4.2, 2.3

Lusso E., Worseck G., Hennawi J., Prochaska J., Vignali C., Stern J., O'Meara J., 2015, Monthly Notices of the Royal Astronomical Society, 449, 4204 1.4.1

Ma X., Quataert E., Wetzel A., Hopkins P. F., Faucher-Giguère C.-A., Kereš D., 2020a, arXiv e-prints: 2003.05945 1.4.5

Ma X., Quataert E., Wetzel A., Faucher-Giguère C.-A., Boylan-Kolchin M., 2020b, arXiv e-prints: 2006.10065 1.4.5

Maas A. L., Hannun A. Y., Ng A. Y., 2013, in Proc. ICML. p. 3 C.4

Madau P., 1995, The Astrophysical Journal, 441, 18 1.4.2

Madau P., Dickinson M., 2014, Annual Review of Astronomy and Astrophysics, 52, 415 1.4.3, 1.10

Madau P., Haardt F., 2015, The Astrophysical Journal, 813, L8 1.4.1, 5.4.2

Madau P., Haardt F., Rees M. J., 1999, The Astrophysical Journal, 514, 648 1.3.4

Mainali R., et al., 2020, Monthly Notices of the Royal Astronomical Society, 494, 719 1.4.4

Maseda M. V., et al., 2018, The Astrophysical Journal, 854, 29 6.4

Maseda M. V., et al., 2020, Monthly Notices of the Royal Astronomical Society, 493, 5120 1.4.4

Mason C. A., Gronke M., 2020, arXiv e-prints: 2004.13065 1.5, 2.1, 2.1, 2.2, 2.2, 2.4.2, 2.5.1

Mason C. A., Trenti M., Treu T., 2015, The Astrophysical Journal, 813, 21 1.4.3

Mason C. A., Treu T., Dijkstra M., Mesinger A., Trenti M., Pentericci L., de Barros S., Vanzella E., 2018a, The Astrophysical Journal, 856, 2 1.3, 1.3.4

Mason C. A., et al., 2018b, The Astrophysical Journal, 857, L11 1.3.3, 4.7.1

Mason C. A., et al., 2019, Monthly Notices of the Royal Astronomical Society, 485, 3947 4.4.1

Matsuoka Y., et al., 2019, The Astrophysical Journal, 883, 183 5.4.1, 5.5

Matthee J., Sobral D., Best P., Khostovan A. A., Oteo I., Bouwens R., Röttgering H., 2017, Monthly Notices of the Royal Astronomical Society, 465, 3637 1.4.4

Matthee J., Sobral D., Gronke M., Paulino-Afonso A., Stefanon M., Röttgering H., 2018, Astronomy & Astrophysics, 619, A136 1.5.3, 2.1, 2.3, 2.6, 2.5.1, 4.2.3, 4.7.1

Matthews T. A., Sandage A. R., 1963, The Astrophysical Journal, 138, 30 1.4.1

Mazzucchelli C., et al., 2017, The Astrophysical Journal, 849, 91 4.1, 5.1, 5.2.1, 5.2.2, 5.3.1, 5.3.3, 5.4, 5.4.1

McDonald P., et al., 2006, The Astrophysical Journal Supplement Series, 163, 80 A.2

McGreer I. D., Mesinger A., D'Odorico V., 2015, Monthly Notices of the Royal Astronomical Society, 447, 499 3.1, 3.2.1, 4.1

McLeod D. J., McLure R. J., Dunlop J. S., Robertson B. E., Ellis R. S., Targett T. A., 2015, Monthly Notices of the Royal Astronomical Society, 450, 3032 1.4.3

McQuinn M., Upton Sanderbeck P. R., 2016, Monthly Notices of the Royal Astronomical Society, 456, 47 4.5.1

McQuinn M., Hernquist L., Zaldarriaga M., Dutta S., 2007, Monthly Notices of the Royal Astronomical Society, 381, 75 1.3.3

McQuinn M., Peng Oh S., Faucher-Giguère C.-A., 2011, The Astrophysical Journal, 743, 82 4.5.2

Mesinger A., 2010, Monthly Notices of the Royal Astronomical Society, 407, 1328 1.3.1

Mesinger A., Furlanetto S. R., 2008a, Monthly Notices of the Royal Astronomical Society, 385, 1348 1.3.3

Mesinger A., Furlanetto S. R., 2008b, Monthly Notices of the Royal Astronomical Society, 386, 1990 2.2

Mesinger A., Furlanetto S., Cen R., 2011, Monthly Notices of the Royal Astronomical Society, 411, 955 1.5.1

Mesinger A., Aykutalp A., Vanzella E., Pentericci L., Ferrara A., Dijkstra M., 2015, Monthly Notices of the Royal Astronomical Society, 446, 566 1.3.3, 2.2

Meyer R. A., Bosman S. E., Kakiichi K., Ellis R. S., 2019a, Monthly Notices of the Royal Astronomical Society, 483, 19 3, 5.1, B.3, B.14

Meyer R. A., Bosman S. E. I., Ellis R. S., 2019b, Monthly Notices of the Royal Astronomical Society, 487, 3305 5, C.1

Meyer R. A., et al., 2020, Monthly Notices of the Royal Astronomical Society, 494, 1560 4

Michel-Dansac L., Blaizot J., Garel T., Verhamme A., Kimm T., Trebitsch M., 2020, Astronomy & Astrophysics, 635, A154 2.4.1

Miralda-Escude J., Haehnelt M., Rees M. J., 2000, The Astrophysical Journal, 530, 1 4.5.2, 4.5.2, B.14

Miralda-Escude J., 1998, The Astrophysical Journal, 501, 15 1.3.3

Mitra S., Choudhury T. R., Ferrara A., 2017, Monthly Notices of the Royal Astronomical Society, 473, 1416 1.4.1

Miyazaki S., et al., 2002, Publications of the Astronomical Society of Japan, 54, 833 4.2.2

Morishita T., et al., 2018, The Astrophysical Journal, 867, 150 1.4.3

Morselli L., et al., 2014, Astronomy & Astrophysics, 568, A1 4.2.2

Mortlock D. J., et al., 2009, Astronomy & Astrophysics, 505, 97 5.1

Mortlock D. J., et al., 2011, Nature, 474, 616 1.3.1, 1.3.3, 5.1, 6.2

Mostardi R. E., Shapley A. E., Steidel C. C., Trainor R. F., Reddy N. A., Siana B., 2015, The Astrophysical Journal, 810, 107 1.4.5

Moster B. P., Somerville R. S., Maulbetsch C., van den Bosch F. C., Macciò A. V., Naab T., Oser L., 2010, The Astrophysical Journal, 710, 903 5.5.3

Murray N., Chiang J., 1995, The Astrophysical Journal, 454, L105 5.4

Murray S. G., Power C., Robotham A. S., 2013, Astronomy and Computing, 3-4, 23 3.5.1

Nagao T., Marconi A., Maiolino R., 2006, Astronomy & Astrophysics, 447, 157 5.3.2

Naidu R. P., Tacchella S., Mason C. A., Bose S., Oesch P. A., Conroy C., 2020, The Astrophysical Journal, 892, 109 1.3, 1.3.4, 1.7, 1.4.1, 2.1, 4.6, 4.14, 6.1, 6.2, 6.2

Nakajima K., Ellis R. S., Iwata I., Inoue A. K., Kusakabe H., Ouchi M., Robertson B. E., 2016, The Astrophysical Journal, 831, L9 1.4.4, 3.5.2

Nakajima K., Fletcher T., Ellis R. S., Robertson B. E., Iwata I., 2018, Monthly Notices of the Royal Astronomical Society, 477, 2098 2.1, 3.5.2

Nakajima K., Ellis R. S., Robertson B. E., Tang M., Stark D. P., 2020, The Astrophysical Journal, 889, 161 1.4.4, 1.4.5

Nasir F., D'Aloisio A., 2020, Monthly Notices of the Royal Astronomical Society, 494, 3080 1.5.1

Nelson D., et al., 2018, Monthly Notices of the Royal Astronomical Society, 475, 624 4.5.2

Neufeld D. A., 1990, The Astrophysical Journal, 350, 216 2.2

Newman J. A., et al., 2013, Astrophysical Journal, Supplement Series, 208 4.2.2

Oesch P. A., Bouwens R. J., Illingworth G. D., Labbé I., Stefanon M., 2018, The Astrophysical Journal, 855, 105 1.4.3, 1.4.3, 1.10

Oke J. B., 1974, The Astrophysical Journal Supplement Series, 27, 21 2.1, 4.1

Ono Y., et al., 2012, The Astrophysical Journal, 744, 83 1.3.3, 4.4.1

Ono Y., et al., 2018, Publications of the Astronomical Society of Japan, 70 1.4.2, 1.4.3, 1.4.3, 4.4.1, 4.5, 5.4.2

Oñorbe J., Hennawi J. F., Lukić Z., 2017, The Astrophysical Journal, 837, 106 1.5.1

Oñorbe J., Davies F. B., Lukić Z., Hennawi J. F., Sorini D., 2019, Monthly Notices of the Royal Astronomical Society, 486, 4075 1.5.1

Onoue M., et al., 2017, The Astrophysical Journal, 847, L15 1.4.1, 5.1

Onoue M., et al., 2019, The Astrophysical Journal, 880, 77 5.4.1

Oppenheimer B. D., Davé R., Finlator K., 2009, Monthly Notices of the Royal Astronomical Society, 396, 729 3.1, 3.5.1, 3.5.3

Orsi A., Lacey C. G., Baugh C. M., 2012, Monthly Notices of the Royal Astronomical Society, 425, 87 2.2, 2.1

Osmer P. S., Porter A. C., Green R. F., 1994, The Astrophysical Journal, 436, 678 5.3.2

Osterbrock D. E., 1962, The Astrophysical Journal, 135, 195 2.2

Ouchi M., et al., 2008, The Astrophysical Journal Supplement Series, 176, 301 1.3.3

Ouchi M., et al., 2010, The Astrophysical Journal, 723, 869 1.3.3

Ouchi M., et al., 2018, Publications of the Astronomical Society of Japan, 70 1.4, 1.3.3, 1.4.2, 4.5.2, 4.6

Pacucci F., Loeb A., 2019, The Astrophysical Journal, 870, L12 5.1, 5.5, 5.5.3, 5.18, 5.19

Pâris I., et al., 2011, Astronomy & Astrophysics, 530, A50 5.1, 5.2.1

Pâris I., et al., 2012, Astronomy & Astrophysics, 548, A66 5.2.4, 5.2.4

Pâris I., et al., 2017, Astronomy & Astrophysics, 597, A79 5.1, 5.2.1, 5.2.4

Parsa S., Dunlop J. S., McLure R. J., 2018, Monthly Notices of the Royal Astronomical Society, 474, 2904 1.4.1

Partridge R. B., Peebles P. J. E., 1967, The Astrophysical Journal, 147, 868 2.2

Pawlik A. H., Schaye J., van Scherpenzeel E., 2009, Monthly Notices of the Royal Astronomical Society, 394, 1812 3.4, 4.5.2, B.14

Peng C. Y., Ho L. C., Impey C. D., Rix H. W., 2010, Astronomical Journal, 139, 2097 2.4.1

Pentericci L., et al., 2018, Astronomy & Astrophysics, 619, A147 1.3.3

Penzias A., Wilson R., 1965, The Astrophysical Journal, 142, 419 1.1

Perlmutter S., et al., 1998, The Astrophysical Journal, 517, 565 1.1

Pettini M., Madau P., Bolte M., Prochaska J. X., Ellison S. L., Fan X., 2003, The Astrophysical Journal, 594, 695 3.1

Planck Collaboration et al., 2016, Astronomy & Astrophysics, 594, A13 3.1

Planck Collaboration et al., 2018, preprint (arXiv:1807.06209) 1.3.2, 1.3.2, 1.3

Pons E., McMahon R. G., Simcoe R. A., Banerji M., Hewett P. C., Reed S. L., 2019, Monthly Notices of the Royal Astronomical Society, 484, 5142 5.3.3, 5.13

Postman M., et al., 2012, The Astrophysical Journal Supplement Series, 199, 25 1.4.2

Press W. H., Schechter P., 1974, The Astrophysical Journal, 187, 425 1.5.1

Prochaska J. X., et al., 2019,] 10.5281/ZENODO.3506873 4.2.1

Proga D., Stone J. M., Kallman T. R., 2000, The Astrophysical Journal, 543, 686 5.4

Puchwein E., Haardt F., Haehnelt M. G., Madau P., 2019, Monthly Notices of the Royal Astronomical Society, 485, 47 1.5.1

Rankine A. L., Hewett P. C., Banerji M., Richards G. T., 2020, Monthly Notices of the Royal Astronomical Society, 492, 4553 5.1

Redmon J., Farhadi A., 2016, arXiv e-prints: 1612.08242 C.4, C.4

Redmon J., Farhadi A., 2018, arXiv e-prints : 1804.02767 C.4, C.4

Redmon J., Divvala S., Girshick R., Farhadi A., 2015, arXiv e-prints: 1506.02640 C.4, C.4, C.4

Reed S., et al., 2015, Monthly Notices of the Royal Astronomical Society, 454, 3952 5.1, 5.5.1, 5.5.3

Reed S., et al., 2017, Monthly Notices of the Royal Astronomical Society, 468, 4702 5.5, 5.5.1, C.4

Reed S. L., et al., 2019, Monthly Notices of the Royal Astronomical Society, 487, 1874 5.1, 5.3.3, 5.3.3, 5.13, 5.5

Reiman D. M., Tamanas J., Prochaska J. X., Ďurovčíková D., 2020 6.2, 6.2

Ren S., He K., Girshick R., Sun J., 2015, arXiv e-prints: 1506.01497 C.4

Richards G. T., Vanden Berk D. E., Reichard T. A., Hall P. B., Schneider D. P., SubbaRao M., Thakar A. R., York D. G., 2002, The Astronomical Journal, 124, 1 1.4.1, 5.3.2

Richards G. T., et al., 2011, Astronomical Journal, 141, 167 5.1, 5.4

Roberts-Borsani G. W., et al., 2016, The Astrophysical Journal, 823, 143 2.5.2

Roberts-Borsani G., Ellis R. S., Laporte N., 2020, arXiv e-prints: 2002.02968 1.4.3

Robertson B. E., Ellis R. S., Furlanetto S. R., Dunlop J. S., 2015, The Astrophysical Journal, 802, L19 1.3.4, 1.7, 1.4.1, 2.1, 4.1, 4.6, 4.14, 6.1

Romano M., Grazian A., Giallongo E., Cristiani S., Fontanot F., Boutsia K., Fiore F., Menci N., 2019, Astronomy & Astrophysics, 632, A45 1.4.1

Rubin V. C., Ford W. K. J., 1970, The Astrophysical Journal, 159, 379 1.1

Ryan-Weber E. V., Pettini M., Madau P., Zych B. J., 2009, Monthly Notices of the Royal Astronomical Society, 395, 1476 3.1, 3.2.2, 3.2.4, 3.5, 3.3.2, 5.1

Savage B. D., Sembach K. R., 1991, The Astrophysical Journal, 379, 245 3.2.3

Schaerer D., 2003, Astronomy & Astrophysics, 397, 527 1.4.4

Schenker M. A., Ellis R. S., Konidaris N. P., Stark D. P., 2014, The Astrophysical Journal, 795, 20 1.3.3

Schmidt B. P., et al., 1998, The Astrophysical Journal, 507, 46 1.1

Schneider D. P., et al., 2010, The Astronomical Journal, 139, 2360 5.2.1

Shapley A. E., 2011, Annual Review of Astronomy and Astrophysics, 49, 525 1.4.2

Shapley H., Curtis H. D., 1921, Bulletin of the National Research Council, 2, 171 1.1

Shapley A. E., Steidel C. C., Pettini M., Adelberger K. L., Erb D. K., 2006, The Astrophysical Journal, 651, 688 1.4.5, 1.4.5

Sharma M., Theuns T., Frenk C., Bower R., Crain R., Schaller M., Schaye J., 2016, Monthly Notices of the Royal Astronomical Society: Letters, 458, L94 1.4.5

Sheinis A. I., Bolte M., Epps H. W., Kibrick R. I., Miller J. S., Radovan M. V., Bigelow B. C., Sutin B. M., 2002, Publications of the Astronomical Society of the Pacific, 114, 851 3.2.1

Shen Y., Greene J. E., Strauss M. A., Richards G. T., Schneider D. P., 2008, The Astrophysical Journal, 680, 169 5.1

Shen Y., et al., 2011, The Astrophysical Journal Supplement Series, 194, 45 5.2.1, 5.2.4

Shen Y., et al., 2016, The Astrophysical Journal, 831, 7 5.1

Shen Y., et al., 2019, The Astrophysical Journal, 873, 35 5.1, 5.3.1, 5.11, 5.3.3, 5.13, 5.4.1

Shen X., et al., 2020, Monthly Notices of the Royal Astronomical Society, 495, 4747 6.4

Shivaei I., et al., 2018, The Astrophysical Journal, 855, 42 1.4.4, 6.4

Siana B., et al., 2007, The Astrophysical Journal, 668, 62 1.4.5

Siana B., et al., 2010, The Astrophysical Journal, 723, 241 1.4.5

Siana B., et al., 2015, The Astrophysical Journal, 804, 17 1.4.5

Simcoe R. A., et al., 2011, The Astrophysical Journal, 743, 21 3.2.2, 3.5

Slipher V., 1915, Popular Astronomy, 23, 21 1.1

Slipher V., 1917, Proceedings of the American Philosophical Society, 56, 403 1.1

Slosar A., et al., 2011, Journal of Cosmology and Astroparticle Physics, 2011, 001 3.4

Smee S. A., et al., 2012, The Astronomical Journal, Volume 146, Issue 2, article id. 32, 40 pp. (2013)., 146 5.2.1

Smit R., et al., 2014, Astrophysical Journal, 784, 58 2.3

Smit R., Bouwens R. J., Labbé I., Franx M., Wilkins S. M., Oesch P. A., 2016, The Astrophysical Journal, 833, 254 1.4.4

Sobral D., Matthee J., 2019, Astronomy & Astrophysics, 623, A157 2.4.2, 2.4.2, 2.3

Sobral D., Santos S., Matthee J., Paulino-Afonso A., Ribeiro B., Calhau J., Khostovan A. A., 2017, Monthly Notices of the Royal Astronomical Society, 476, 4725 1.4.4

Songaila A., 2004, The Astronomical Journal, 127, 2598 1.3.1

Songaila A., Hu E. M., Barger A. J., Cowie L. L., Hasinger G., Rosenwasser B., Waters C., 2018, The Astrophysical Journal, 859, 91 1.5.3, 2.1, 2.6, 2.5.1, 2.5.2, 4.2.3

Soto K. T., Lilly S. J., Bacon R., Richard J., Conseil S., 2016, Monthly Notices of the Royal Astronomical Society, 458, 3210 4.2.3

Stark D. P., 2016, Annual Review of Astronomy and Astrophysics, 54, 761 1.3.3, 1.4.2

Stark D. P., Ellis R. S., Chiu K., Ouchi M., Bunker A., 2010, Monthly Notices of the Royal Astronomical Society, 408, 1628 1.3, 1.3.3, 4.4.1

Stark D. P., Ellis R. S., Ouchi M., 2011, The Astrophysical Journal Letters, 728, L2 1.3.3

Stark D. P., et al., 2015, Monthly Notices of the Royal Astronomical Society, 454, 1393 2.4.2, 3.5.2

Stark D. P., et al., 2017, Monthly Notices of the Royal Astronomical Society, 464, 469 1.4.4, 2.4.2, 2.5.2, 3.5.2, 3.5.3, 4.2.4, 5.4.2

Stark D., et al., 2018, Monthly Notices of the Royal Astronomical Society, 477, 2513 2.5.2

Steidel C. C., Hamilton D., 1992, The Astronomical Journal, 104, 941 1.4.2

Steidel C. C., Hamilton D., 1993, The Astronomical Journal, 105, 2017 1.4.2

Steidel C., Pettini M., Hamilton D., 1995, The Astronomical Journal, 110, 2519 1.4.2

Steidel C. C., Pettini M., Adelberger K. L., 2001, The Astrophysical Journal, 546, 665 1.4.5, 1.4.5

Steidel C. C., Erb D. K., Shapley A. E., Pettini M., Reddy N. A., Bogosavljević M., Rudie G. C., Rakic O., 2010, The Astrophysical Journal, 717, 289 3.1, 3.5.3, 4.2.4, 4.5.3

Steidel C. C., Bogosavljević M., Shapley A. E., Reddy N. A., Rudie G. C., Pettini M., Trainor R. F., Strom A. L., 2018, The Astrophysical Journal, 869, 123 1.4.5, 4.7.2

Steinhardt C. L., et al., 2020, The Astrophysical Journal Supplement Series, 247, 64 1.4.2

Storrie-Lombardi L., McMahon R., Irwin M., Hazard C., 1996, The Astrophysical Journal, 468, 121 3.3.1

Suzuki N., 2006, The Astrophysical Journal Supplement Series, 163, 110 5.1

Tang M., Stark D. P., Chevallard J., Charlot S., 2019, Monthly Notices of the Royal Astronomical Society, 489, 2572 1.4.4

Tanvir N. R., et al., 2019, Monthly Notices of the Royal Astronomical Society, 483, 5380 1.4.5, 4.7.1

Tasca L. A. M., et al., 2017, Astronomy & Astrophysics, 600, A110 4.1

Taylor A. J., Barger A. J., Cowie L. L., Hu E. M., Songaila A., 2020, The Astrophysical Journal, 895, 132 2.5.2

Tilvi V., et al., 2014, Astrophysical Journal, 794, 5 1.3.3

Trebitsch M., Volonteri M., Dubois Y., 2019, Monthly Notices of the Royal Astronomical Society, 487, 819 5.4.2

Trebitsch M., et al., 2020, arXiv e-prints: 2002.04045 1.4.1

Treister E., Schawinski K., Volonteri M., Natarajan P., Gawiser E., 2011, Nature, 474, 356 5.4.2

Trenti M., et al., 2011, The Astrophysical Journal Letters, 727, L39 1.4.2

Turner M. L., Schaye J., Steidel C. C., Rudie G. C., Strom A. L., 2014, Monthly Notices of the Royal Astronomical Society, 445, 794 3.1

Ucci G., et al., 2020, arXiv e-prints: 2004.11096 1.4.3

Vanden Berk D. E., et al., 2001, The Astronomical Journal, 122, 549 5.2.2

Vanzella E., et al., 2016, The Astrophysical Journal, 821, L27 1.4.5, 4.2.4, 4.7.1

Venemans B. P., McMahon R. G., Warren S. J., Gonzáles-Solares E. A., Hewett P. C., Mortlock D. J., Dye S., Sharp R. G., 2007, Monthly Notices of the Royal Astronomical Society: Letters, 376, L76 5.1

Venemans B. P., et al., 2012, The Astrophysical Journal, 751, L25 5.1

Venemans B. P., et al., 2013, Astrophysical Journal, 779 4.1, 5.1, 5.5, 5.5.1

Venemans B. P., et al., 2015, The Astrophysical Journal Letters, 801, L11 5.5

Venemans B. P., et al., 2018, The Astrophysical Journal, 866, 159 5.1

Verhamme A., Schaerer D., Maselli A., 2006, Astronomy and Astrophysics, 460, 397 2.2, 2.4.1

Verhamme A., Schaerer D., Atek H., Tapken C., 2008, Astronomy and Astrophysics, 491, 89 2.2, 2.4.1

Verhamme A., Orlitová I., Schaerer D., Hayes M., 2015, Astronomy & Astrophysics, 578, A7 2.1, 2.2

Verhamme A., Orlitová I., Schaerer D., Izotov Y., Worseck G., Thuan T. X., Guseva N., 2017, Astronomy & Astrophysics, 597, A13 2.4.1

Verhamme A., et al., 2018, Monthly Notices of the Royal Astronomical Society: Letters, 478, L60 4.2.4, B.2

Verner D. A., Ferland G. J., 1996, ApJSS, 103, 467 1.5.3

Vernet J., et al., 2011, Astronomy & Astrophysics, Volume 536, id.A105, 15 pp., 536 3.2.1, 5.2.1

Viel M., Becker G. D., Bolton J. S., Haehnelt M. G., 2013, Physical Review D, 88, 043502 3.4

Vietri G., et al., 2018, Astronomy & Astrophysics, 617, A81 5.4

Vogt S. S., et al., 1994, in Instrumentation in Astronomy VIII. SPIE, pp 362–375, doi:10.1117/12.176725, https://ui.adsabs.harvard.edu/abs/1994SPIE.2198..362V/abstract 3.2.1

Wang F., et al., 2018, The Astrophysical Journal, 869, L9 5.1

Wang F., et al., 2019, The Astrophysical Journal, 884, 30 5.1, 5.3.3, 5.13, 5.5

Wang F., et al., 2020 1.3.3

Weilbacher P. M., Streicher O., Urrutia T., Jarno A., Pécontal-Rousset A., Bacon R., Böhm P., 2012, in Radziwill N. M., Chiozzi G., eds, Vol. 8451, Software and Cyberinfrastructure for Astronomy II. Proceedings of the SPIE, Volume 8451, article id. 84510B, 9 pp. (2012).. p. 84510B, doi:10.1117/12.925114, http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.925114 4.2.3

Weilbacher P. M., Streicher O., Urrutia T., Pécontal-Rousset A., Jarno A., Bacon R., 2015, Astronomical Data Analysis Software and Systems XXIII. Proceedings of a meeting held 29 September - 3 October 2013 at Waikoloa Beach Marriott, Hawaii, USA. Edited by N. Manset and P. Forshay ASP conference series, vol. 485, 2014, p.451, 485, 451 4.2.3

Willott C. J., et al., 2007, The Astronomical Journal, 134, 2435 4.1

Willott C. J., et al., 2010, The Astronomical Journal, 140, 546 5.1

Wills B. J., Brotherton M. S., Fang D., Steidel C. C., Sargent W. L. W., 1993, The Astrophysical Journal, 415, 563 5.4

Wise J. H., Demchenko V. G., Halicek M. T., Norman M. L., Turk M. J., Abel T., Smith B. D., 2014, Monthly Notices of the Royal Astronomical Society, 442, 2560 1.4.3, 1.4.5

Worseck G., et al., 2014, Monthly Notices of the Royal Astronomical Society, 445, 1745 1.5.3, 2.4.2, 3.4, 4.5, 5.2.1

Worseck G., Prochaska J. X., Hennawi J. F., McQuinn M., 2016, The Astrophysical Journal, 825, 144 1.4.1

Yamada T., Matsuda Y., Kousai K., Hayashino T., Morimoto N., Umemura M., 2012, Astrophysical Journal, 751 2.3, 2.6

Yang J., et al., 2016, The Astrophysical Journal, 829, 33 5.19

Yang H., et al., 2017, The Astrophysical Journal, 844, 171 2.3, 2.6

Yang J., et al., 2018, arXiv e-prints: 1810.11927 5.1

Yang J., et al., 2020, The Astrophysical Journal, 897, L14 1.4.1, 6.2

York D. G., et al., 2000, The Astronomical Journal, Volume 120, Issue 3, pp. 1579-1587., 120, 1579 1.4.1, 5.2.1

Yue B., Ferrara A., Xu Y., 2016, Monthly Notices of the Royal Astronomical Society, 463, 1968 1.4.3

Zeiler M. D., 2012, arXiv e-prints: 1212.5701 C.4

Zitrin A., et al., 2015, The Astrophysical Journal, 810, L12 2.5.2

Zwicky F., 1933, Helvetica Physica Acta, 6, 110 1.1

de La Vieuville G., et al., 2019, Astronomy & Astrophysics, 628, A3 4.2.3, 4.5, 4.4.1

du Mas des Bourboux H., et al., 2017, Astronomy & Astrophysics, 608, A130 A.2

Ďurovčíková D., Katz H., Bosman S. E. I., Davies F. B., Devriendt J., Slyz A., 2020, Monthly Notices of the Royal Astronomical Society, 493, 4256 1.3.3, 6.2