



The High-Redshift Voyage of Lyman α and Lyman Continuum Emission as told by MUSE

Josephine Victoria Kerutt

Leibniz-Institut für Astrophysik Potsdam (AIP)

Dissertation
zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)
in der Wissenschaftsdisziplin Astrophysik.
Eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Universität Potsdam.

October 29, 2019

1. Gutachter und Hauptbetreuer:
Prof. Dr. Lutz Wisotzki
Leibniz-Institut für Astrophysik Potsdam (AIP)

2. Gutachter und Zweitbetreuer:
Prof. Dr. Philipp Richter
Universität Potsdam (UP)

3. Gutachterin:
Prof. Dr. Dawn Erb
University of Wisconsin-Milwaukee (UWM)

Published online on the
Publication Server of the University of Potsdam:
<https://doi.org/10.25932/publishup-47881>
<https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-478816>



Yepun, where MUSE is installed, with its four Laser guide stars. Credit and licence: European Southern Observatory / Gerhard Hüdepohl (atacamaphoto.com).

"Imagination is the Discovering Faculty, pre-eminently. It is that which penetrates into the unseen worlds around us, the worlds of Science."
– Ada Lovelace

Abstract

Most of the matter in the universe consists of hydrogen. The hydrogen in the intergalactic medium (IGM), the matter between the galaxies, underwent a change of its ionisation state at the epoch of reionisation, at a redshift roughly between $6 \lesssim z \lesssim 10$, or $\sim 10^8$ years after the Big Bang. At this time, the mostly neutral hydrogen in the IGM was ionised but the source of the responsible hydrogen ionising emission remains unclear. In this thesis I discuss the most likely candidates for the emission of this ionising radiation, which are a type of galaxy called Lyman α emitters (LAEs). As implied by their name, they emit Lyman α radiation, produced after a hydrogen atom has been ionised and recombines with a free electron. The ionising radiation itself (also called Lyman continuum emission) which is needed for this process inside the LAEs could also be responsible for ionising the IGM around those galaxies at the epoch of reionisation, given that enough Lyman continuum escapes. Through this mechanism, Lyman α and Lyman continuum radiation are closely linked and are both studied to better understand the properties of high redshift galaxies and the reionisation state of the universe.

Before I can analyse their Lyman α emission lines and the escape of Lyman continuum emission from them, the first step is the detection and correct classification of LAEs in integral field spectroscopic data, specifically taken with the Multi-Unit Spectroscopic Explorer (MUSE). After detecting emission line objects in the MUSE data, the task of classifying them and determining their redshift is performed with the graphical user interface `QtClassify`, which I developed during the work on this thesis. It uses the strength of the combination of spectroscopic and photometric information that integral field spectroscopy offers to enable the user to quickly identify the nature of the detected emission lines. The reliable classification of LAEs and determination of their redshifts is a crucial first step towards an analysis of their properties.

Through radiative transfer processes, the properties of the neutral hydrogen clouds in and around LAEs are imprinted on the shape of the Lyman α line. Thus after identifying the LAEs in the MUSE data, I analyse the properties of the Lyman α emission line, such as the equivalent width (EW) distribution, the asymmetry and width of the line as well as the double peak fraction. I challenge the common method of displaying EW distributions as histograms without taking the limits of the survey into account and construct a more independent EW distribution function that better reflects the properties of the underlying population of galaxies. I illustrate this by comparing the fraction of high EW objects between the two surveys MUSE-Wide and MUSE-Deep, both consisting of MUSE pointings (each with the size of one square arcminute) of different depths. In the 60 MUSE-Wide fields of one hour exposure time I find a fraction of objects with extreme EWs above $EW_0 > 240 \text{ \AA}$ of $\sim 20\%$, while in the MUSE-Deep fields (9 fields with an exposure time of 10 hours and one with an exposure time of 31 hours) I find a fraction of only $\sim 1\%$, which is due to the differences in the limiting line flux of the surveys. The highest EW I measure is $EW_0 = 600.63 \pm 110 \text{ \AA}$, which hints at an unusual underlying stellar population, possibly with a very low metallicity.

With the knowledge of the redshifts and positions of the LAEs detected in the MUSE-Wide survey, I also look for Lyman continuum emission coming from these galaxies and analyse the connection between Lyman continuum emission and Lyman α emission. I use ancillary Hubble Space Telescope (HST) broadband photometry in the bands that contain the Lyman continuum and find six Lyman continuum leaker candidates. To test whether the Lyman continuum emission of LAEs is coming only from those individual objects or the

whole population, I select LAEs that are most promising for the detection of Lyman continuum emission, based on their rest-frame UV continuum and Lyman α line shape properties. After this selection, I stack the broadband data of the resulting sample and detect a signal in Lyman continuum with a significance of $S/N = 5.5$, pointing towards a Lyman continuum escape fraction of $\sim 80\%$. If the signal is reliable, it strongly favours LAEs as the providers of the hydrogen ionising emission at the epoch of reionisation and beyond.

Zusammenfassung

Die Materie zwischen den Galaxien im Universum, das sogenannte intergalaktische Medium, besteht zum größten Teil aus Wasserstoff, welcher sich zusammensetzt aus einem Proton und einem Elektron. Etwa $\sim 10^8$ Jahre nach dem Urknall änderte sich der Ionisationszustand des Wasserstoffs im intergalaktischen Medium. Durch Strahlung wurden Proton und Elektronen getrennt, der Wasserstoff wurde ionisiert. Die Strahlung mit der nötigen Energie um Wasserstoff zu ionisieren nennt man Lyman Kontinuum Strahlung. Woher diese stammt ist bis heute eine offene Frage. In der vorliegenden Arbeit diskutiere ich die wahrscheinlichsten Kandidaten für die Emission dieser ionisierenden Strahlung: die Lyman α Emittier (LAE), eine Gruppe von Galaxien die man im frühen Universum antrifft. Hat nämlich ein Lyman Kontinuum Photon ein Wasserstoffatom ionisiert, kommt es häufig vor, dass sich Proton und Elektron wieder zu einem neutralen Wasserstoffatom verbinden. Bei diesem Prozess wird dann ein Lyman α Photon freigesetzt, was den LAE ihren Namen gibt. Da wir in den Spektren dieser Galaxien die Lyman α Photonen beobachten können, deren Voraussetzung die Existenz von Lyman Kontinuum Emission in der Galaxie ist, liegt es nahe anzunehmen, dass die Lyman Kontinuum Emission auch in das intergalaktische Medium vordringen kann um den Wasserstoff dort ebenfalls zu ionisieren. Durch diesen Mechanismus sind Lyman α und Lyman-Kontinuumsstrahlung eng miteinander verbunden und werden in dieser Arbeit beide untersucht, um die Eigenschaften von Galaxien im frühen Universum und den Zustand des Wasserstoffs im intergalaktischen Medium besser zu verstehen.

Der erste Schritt ist jedoch die Detektion und korrekte Klassifizierung von LAE. Dafür benutze ich Daten des Multi-Unit Spectroscopic Explorer (MUSE), ein Instrument mit dem Beobachtungen möglich sind, die zwei räumliche Dimensionen mit einer spektralen Dimension verbinden. Somit habe ich in jedem Pixel der Daten ein Spektrum zur Verfügung, in dem nach Emissionslinien gesucht werden kann. Nach der Detektion von Emissionslinien in den MUSE-Daten wird die Aufgabe der Klassifizierung von der grafischen Benutzeroberfläche QtClassify unterstützt, die ich in dieser Arbeit entwickelt habe. Sie bietet dem Benutzer eine schnelle und übersichtliche Identifizierung der gefundenen Emissionslinien, was einen entscheidenden ersten Schritt zur Analyse der Eigenschaften der LAE darstellt.

Da Lyman α Photonen in neutralem Wasserstoff, der im Inneren von Galaxien in Gaswolken vorkommt, gestreut wird, verändert sich die Form der Lyman α Emissionslinie im Spektrum. Nachdem also die LAE mit Hilfe von QtClassify identifiziert werden konnten, analysiere ich die Lyman α Linie hinsichtlich ihrer Breite, ihrer Symmetrie und ihrer Stärke und suche nach Doppellinien, die Hinweise auf die Kinematik der Wasserstoff Gaswolken geben. Besonders wichtig dabei ist die Stärke der Linie im Vergleich zum restlichen Spektrum, die sogenannte Äquivalentbreite. Dabei zeige ich, dass die übliche Methode die Verteilung der Äquivalentbreiten als Histogramme darzustellen ohne die Detektionsgrenzen der Beobachtungen miteinzubeziehen keine Aussage über die Eigenschaften der LAEs, sondern lediglich über die Grenzen der Beobachtungen treffen. Daher konstruiere ich eine unabhängigere Verteilungsfunktion, die diese Grenzen mit einbezieht. Ich verdeutliche dies durch einen Vergleich des Anteils von Objekten mit hoher Äquivalentbreite in zwei verschiedenen Datensätzen, die auf Beobachtungen mit unterschiedlicher Belichtungszeit beruhen. Dabei finde ich in dem Datensatz mit der kürzeren Belichtungszeit einen Anteil von $\sim 20\%$ an Objekten mit hohen Äquivalentbreiten (über $> 240 \text{ \AA}$) und nur $\sim 1\%$ im Datensatz mit der zehnfach längeren Belichtungszeit obwohl die Eigenschaften der LAEs gleich sind. Zudem finde ich eine Galaxie mit einer Äquivalentbreite der Lyman α Linie von $600 \pm 110 \text{ \AA}$, was auf eine ungewöhnliche Sternpopulation hindeutet.

Als letzten Schritt suche ich in den klassifizierten LAE nach Lyman Kontinuum, was für die Ionisierung des Wasserstoffs im intergalaktischen Medium verantwortlich ist, und untersuche den Zusammenhang zwischen der Lyman α und der Lyman Kontinuumsstrahlung. Um das Lyman Kontinuum detektieren zu können benutze ich Daten des Hubble-Weltraumteleskops (HST) und finde sechs Lyman Kontinuum Kandidaten unter den LAE. Um zu testen ob die Lyman Kontinuumsstrahlung lediglich von diesen Objekten ausgeht oder auch im der Gesamtheit der LAE zu finden ist, addiere ich die Daten des HST und finde ein Signal, was darauf hindeutet, dass $\sim 80\%$ der Lyman Kontinuumsstrahlung aus den Galaxien entkommen und in das intergalaktische Medium vordringen kann. Wenn das gefundene Signal zuverlässig ist deutet es darauf hin, dass in der Tat LAE verantwortlich sein könnten für die Ionisierung des Wasserstoffs im Universum.

Contents

Abstract

1. Introduction	1
1.1. The Time Frame	2
1.1.1. Redshift	2
1.1.2. Developmental Stages of the Universe	3
1.2. Production of Ionising Emission	4
1.2.1. Star Formation Rate	5
1.2.2. Metallicity and Population III Stars	5
1.2.3. Properties of Stellar Populations at High Redshifts	8
1.3. The Path through the Interstellar Medium	8
1.3.1. Lyman α Emission	9
1.3.2. Lyman α Emitters	10
1.3.3. Radiative Transfer	16
1.3.4. Lyman α Emitters at Different Redshifts	23
1.4. Escaping the Circumgalactic Medium	25
1.4.1. Extended Lyman α Emission	26
1.4.2. Lyman Continuum Leakers	27
1.5. The Path through the Intergalactic Medium	33
1.5.1. Blocking the Way – Lyman α Absorbers	34
1.5.2. The Epoch of Reionisation	34
1.6. Structure of this Thesis	37
1.6.1. Open Questions	37
1.6.2. Summary of this Thesis	37
2. Graphical User Interface for the Classification of Emission Lines: QtClassify	39
2.1. Purpose	39
2.1.1. Input Data	40
2.1.2. Requirements	42
2.2. Inspiration and Applications	42
2.3. Functionality	42
2.3.1. Displaying Different Lines	43
2.3.2. Classification	43
2.3.3. Quality and Confidence	43
2.3.4. Accessing the data	44
2.3.5. Additional and Optional Functionality	45
2.3.6. Automatic Processes	46
2.3.7. Output catalogue	49

2.3.8. Customisation	49
2.4. Possible Improvements	49
2.4.1. Gamification	50
2.4.2. Identification of Counterparts	50
2.5. Access	51
2.6. Process	51
2.6.1. Conclusion	52
3. Equivalent Width Distribution of Lyman α Emitters with MUSE	53
3.1. Introduction	53
3.2. Data	54
3.2.1. Multi-Unit Spectroscopic Explorer (MUSE)-Wide	55
3.2.2. MUSE-Deep	55
3.2.3. Photometry	55
3.3. Sample Selection	55
3.3.1. Classification of Emission Lines	56
3.3.2. Counterparts	56
3.4. Determination of Equivalent Widths	57
3.4.1. Line Fluxes	58
3.4.2. Continuum Flux Densities	58
3.4.3. Distribution of Equivalent Widths	59
3.5. The Lyman α Equivalent Width Distribution Function	61
3.5.1. Luminosity Functions	62
3.5.2. Constructing the Equivalent Width Distribution Function	62
3.5.3. Lyman α Emitter Fractions	64
3.6. Summary and Discussion	65
4. Methods: Measuring Equivalent Widths	67
4.1. Fitting the UV Continuum	67
4.1.1. Counterparts and Galfit Models	67
4.1.2. The β -Parameter	70
4.2. Morphology of Lyman α Emitters	73
4.3. Fitting the Lyman α line	76
4.3.1. Blue Bump Fraction	77
4.3.2. The Lyman α Line Profile	79
4.4. Equivalent Widths	82
4.4.1. Distribution of Equivalent Widths	82
4.4.2. Ando Effect	85
5. Lyman Continuum Emission from MUSE Lyman α Emitters	87
5.1. Introduction	87
5.2. Data	89
5.2.1. MUSE Data	89
5.2.2. Photometry	91
5.3. Individual Lyman Continuum Leaker Candidates	92
5.3.1. Evaluation of Candidates	92
5.4. Lyman Continuum Escape	95
5.4.1. Lyman Continuum to UV Continuum Flux Ratios	95
5.4.2. Lyman Continuum Escape Fractions	95
5.5. Stacking the Selected Sample	98
5.5.1. Selecting the Lyman α Emitters	98
5.5.2. Masking	99
5.5.3. Stacking	99
5.6. Stacking Lyman α Emitter Subsets Based on the Lyman α Line	102
5.6.1. Spectral Properties	102
5.7. Interlopers, Superpositions and Error Estimation	106
5.7.1. Estimating the Reliability of Lyman α Emitter Classifications	107
5.7.2. Random Positions	107
5.7.3. Ruling out AGN Contamination	108

5.8. Discussion	109
5.9. Summary	110
6. Methods: Measuring Lyman Continuum Emission	111
6.1. Comparison with the HDUV Catalogue	111
6.1.1. Matching Catalogues	111
6.1.2. Determining Aperture Sizes	112
6.1.3. Flux Comparison	112
6.2. Correcting Background Offsets in Flux Measurements	117
6.3. Stellar Population Models	120
6.3.1. The Intrinsic UV to Lyman Continuum Flux Ratio	122
6.3.2. Correcting Flux Measurements Using Stellar Population Models	123
6.4. Noise Models	125
6.5. Spatial Offset	126
6.5.1. Shape and Offset of Stacks in Different Bands	128
6.5.2. Bootstrapping Analysis of Offset	131
7. Conclusion and Outlook	135
7.1. Summary	135
7.2. Outlook	137
A. Bibliography	139
B. Miscellaneous	147
B.1. List of Publications	147
B.2. Software and Packages	149
B.3. Abbreviations and Acronyms	149
B.4. Selbstständigkeitserklärung	151
B.5. Acknowledgements	152

”Once you have tasted flight, you will forever walk the earth with your eyes turned skyward, for there you have been, and there you will always long to return.”
– Leonardo da Vinci



Introduction

Lyman continuum emission, created in the star-forming regions of galaxies, transforms into Lyman α emission and makes its way through different kinds of matter until it ionises the neutral hydrogen on its path or is detected. In this thesis we will follow its journey and discuss how it interacts with the galaxy it is formed in and the intergalactic matter it encounters. The connection between hydrogen ionising and Lyman α emission will be examined and the question will be tackled if enough hydrogen ionising emission is escaping its galaxy of origin to account for the emission needed to ionise and keep the universe ionised. From the observation of galaxies at different redshifts we can learn about the path the Lyman continuum (LyC) and Lyman α photons are taking and how they contribute to the ionising background.

A photon that is able to ionise neutral hydrogen is called an ionising photon or ionising radiation in this thesis (always referring to the ionisation of hydrogen, H I, to ionised hydrogen H II¹). This means its energy is higher than the binding energy of hydrogen at 13.6 eV, which corresponds to a wavelength smaller than 911.75 Å. If such radiation encounters neutral hydrogen on its path, the hydrogen will be ionised and in the event of a recombination, will (in most cases) produce a Lyman α photon with a wavelength of 1215.67 Å.

On their way from production to detection or destruction, the ionising photons encounter various obstacles. The matter between the stars in the galaxy of their origin, the interstellar medium (ISM), is the first influence on the ionising emission, which can be converted to Lyman α photons or absorbed in the ISM. The Lyman α photons can then scatter in the ISM through radiative transfer processes that influence the shape of the Lyman α emission line. At the threshold to escaping the galaxy, the matter around the galaxy, the circumgalactic medium (CGM), can pose the next threat to the ionising and Lyman α emission and the latter can be used as a tool to study said CGM. Once the escape is successful, most of the journey takes place outside the galaxy of origin, in the medium between the galaxies, the intergalactic medium (IGM), which consists mostly of hydrogen. Thus the ionising emission can interact with the IGM by changing its ionisation state, which in turn reduces the ionising emission, while the Lyman α emission is less influenced. The final destination of the ionising and Lyman α emission is hopefully the detection by a suitable instrument. Although this doesn't influence the photons any more, the method of detection determines the way we categorise the galaxies that produced the emission in the first place.

The exact path the ionising emission takes also depends on the age of the universe and the properties of the galaxies and the different media (ISM, CGM and IGM) at that time (see figure 1.1 for an illustration). Especially the IGM underwent a significant transformation, changing its ionisation state due to ionising emission produced in the very first galaxies. This state is sustained by the continued input of ionising radiation, the source of which is still a matter of debate. Thus it is important to understand the different eras the universe went through and what it means to look back in space as well as in time.

¹In spectroscopic notation, the ionisation state of atoms is written in Roman numerals. Neutral atoms are denoted with an I, any ionisation is counted as one plus the ionisation state. Thus H II means singly ionised hydrogen, while e.g. the O III emission line results from two times ionised oxygen.

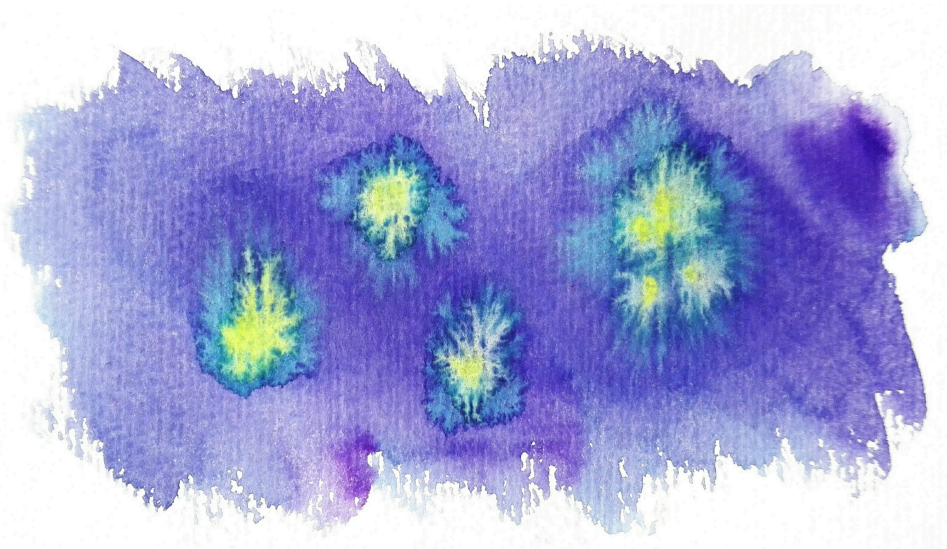


Figure 1.1.: Artist's impression of high redshift galaxies with ISM in yellow, CGM in green and IGM in purple.

1.1. The Time Frame

The universe underwent several major phases in its evolution and it is important to understand in which setting galaxies are to be placed. Depending on the state of the universe, light will be emitted and transmitted through different mechanisms, different elements will be available and the galaxies will be in different stages of their development. As the speed of light is constant and finite, a look into the depths of the universe is also a look back in time. This allows us to study not only the present-day cosmos, but also its past.

1.1.1. Redshift

To understand discussions about the processes at work in the high redshift universe, it is important to understand what redshift range we talk about exactly² and how redshift is defined.

In 1927 [Lemaître \(1927\)](#) proposed the expansion of the universe and [Hubble \(1929\)](#) later showed the relationship between the distances and velocities of galaxies, measuring the distances using the luminosity-period relation of Cepheids ([Leavitt & Pickering, 1912](#)) in 'extra-galactic nebulae' (now known to be other galaxies) and the velocities through spectroscopic redshifts interpreted as Doppler shifts. The result of this expansion is that galaxies in all directions (except for the closest whose relative velocities are influenced not by the expansion of the universe but by gravity) seem to move away from us, giving rise to a shift in their spectra to the red. This shift $\Delta\lambda$ in wavelength in the spectrum can be measured with respect to the rest-frame wavelength λ_0 of the spectral feature in question. Since the speed of light has a fixed value, the shift can be interpreted as a distance or redshift z using the following equation:

$$z = \frac{\Delta\lambda}{\lambda_0} = \frac{\lambda}{\lambda_0} - 1 \quad (1.1)$$

For the purpose of this thesis, the high-redshift universe is concerned with the formation and evolution of galaxies as well as the epoch of reionisation, which puts us in a redshift range of around $1 < z < 7$. Anything with a redshift below one is considered 'low redshift', anything above a redshift of 7 is considered 'extremely high redshift', as it deals with the first galaxies and exotic distance record holders (see figure 1.2 for an overview of the highest observed redshift objects over time).

²Once during a discussion in a group meeting, a colleague dismissed anything below redshift 6 as 'low redshift'. This shows it is best to be clear about the definitions, especially since the meaning of 'high' and 'low' redshift have changed over time and will likely continue to do so. Said colleague might well be more ahead of their time than we had given them credit for.

³Source: https://en.wikipedia.org/wiki/List_of_the_most_distant_astronomical_objects

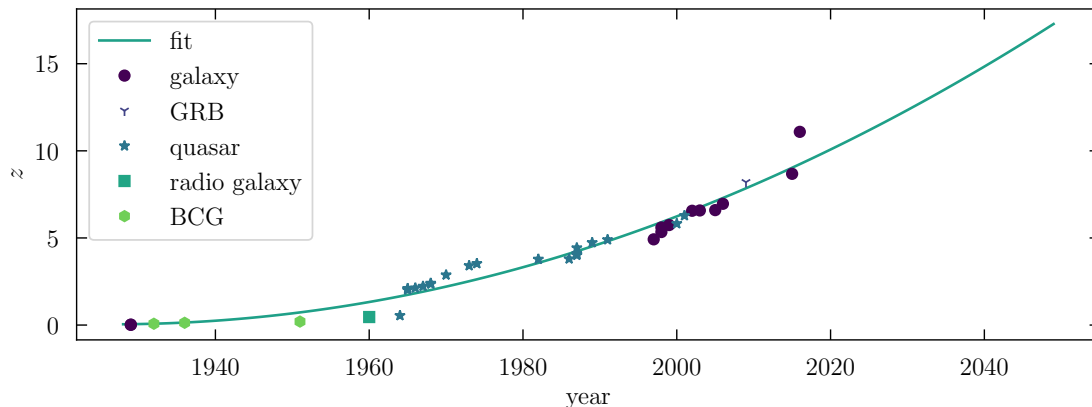


Figure 1.2.: Redshift records over time³. The different symbols and colours indicate the object types (brightest cluster galaxies (BCGs) are brightest cluster galaxies, gamma-ray bursts (GRBs) are gamma-ray bursts), the solid line is a quadratic fit to the redshift records. Most of the objects labelled as 'galaxy' are in fact Lyman α emitters (see section 1.3.2), except for the highest redshift and the lowest redshift objects.

1.1.2. Developmental Stages of the Universe

Lemaître (1927) was also the first to conclude that since the universe is currently expanding, it should be possible to estimate the moment in time when it was just a single point, a singularity. This concept is now known as the Big Bang, the origin time and space. It is aptly named, since the earliest stages of the universe are characterised by high temperatures (over 10^{32} K) and inflation at expansion speeds far exceeding the expansion rate of the present-day universe⁴.

When the universe had cooled down sufficiently, the first particles formed, from quarks over hadrons, neutrinos and leptons to finally baryons and the first atomic nuclei in the Big Bang nucleosynthesis, which produced mainly traces of lithium ($\sim 10^{-10}$ by number for ${}^7\text{Li}$), some deuterium and helium (with $\sim 10^{-5}$ by number for D and ${}^3\text{He}$ and $\sim 25\%$ ${}^4\text{He}$ by mass) and mostly hydrogen (see Cyburt et al., 2016). Since the dissertation of Cecilia Payne-Gaposchkin in 1925 (Payne, 1925), showing that hydrogen is much more abundant in the sun than any other element, we know that it is the most common element in the present-day universe. The next phase of the development of the universe is the origin of the first light, the photon epoch. However, the plasma consisting of atomic nuclei, electrons and photons was not transparent to radiation yet. Only when the universe had cooled down enough for electrons to be bound to the nuclei to recombine to neutral atoms (Peebles, 1968) could the photons finally travel freely, imprinting on them the structure and irregularities of the last scattering surface. The traces of these photons can still be detected today as the cosmic microwave background (CMB) at a temperature of 2.72548 ± 0.00057 K (Fixsen, 2009, which was ~ 4000 K at the epoch of recombination, Peebles, 1968). This point in time corresponds to a redshift of $z \sim 1100$ or 3.7×10^7 years after the Big Bang⁵.

Since the universe kept expanding and the CMB kept getting shifted to the infrared, the universe became dark again (at least in the visible wavelength range), the only light source being the 21-cm line of hydrogen. Thus, this epoch is called the Cosmic Dark Ages. Enlightenment followed, as the first dark matter halos formed with masses of $M \sim 10^8 M_{\odot}$ at redshifts of $z \sim 10 - 20$ (Rees, 1998), facilitating the creation of the first sources of radiation through providing the gravitational potentials for the baryons to gravitationally collapse into stars, eventually forming the first galaxies and some of the most distant objects observable today. The last major shift in the state of the universe took place at this time as well, as the light from the first stars started to ionise the neutral hydrogen between the galaxies in the epoch of reionisation (EoR) at redshifts possibly between $6 \lesssim z \lesssim 10$ or 10^8 to 5×10^8 years after the Big Bang (results from WMAP Hinshaw et al., 2013 and Planck Collaboration et al., 2018, suggest $z \sim 10$ as the redshift for the EoR, but the exact starting point of the process is undefined). The epoch of reionisation coincides with the formation of the first galaxies and the onset of star-formation in them. These early galaxies were likely the progenitors of the galaxies we see today, which makes them all the more interesting to study and important to understand. Through their production of hydrogen ionising emission they interact with their CGM and the IGM, which

⁴Note that the earliest stages of the universe are not directly observable and can only be inferred from the structure and temperature of the cosmic microwave background and theoretical models

⁵For conversions between redshift and age of the universe, see <http://www.astro.ucla.edu/~wright/CosmoCalc.html>

is why they are the main protagonists of this thesis.

After this change was concluded, the galaxies continued to form and evolve, reaching a maximum in star formation rate (SFR) density at a redshift of $z \sim 3$ (see [Madau & Dickinson, 2014](#) for a review) and formed clusters and superclusters, following the trend of building ever larger structures. The time at a redshift of zero, some 13.8×10^9 years after the Big Bang, is the present day universe. Its future fate is subject to debate, with various possible scenarios ranging from a Big Crunch in which the universe will collapse in on itself again, over a Heat Death after the evaporation of all matter, to an ever expanding sphere, getting darker over time as fewer stars get formed and the last ones eventually burn out.

Most of the galaxies discussed in this thesis are at redshifts between $3 < z < 6$, in the early universe after the epoch of reionisation, $2.2 - 0.9 \times 10^9$ years after the Big Bang.

1.2. Production of Ionising Emission

As mentioned in Section 1.1.2, at a redshift between $z \sim 6 - 10$, the universe underwent its last major transition from containing mostly neutral to mostly ionised hydrogen in the IGM. While many observations support the time frame of the EoR and give insights into the processes (see Section 1.5.2), it is not yet clear what the sources of the required ionising photons are. In this section we will therefore look into the production mechanisms of ionising emission and what influences the production efficiency.

Photons that are able to ionise hydrogen need to be highly energetic, with energies over 13.6 eV (the binding energy of neutral hydrogen), corresponding to wavelengths below 912 Å via the equation

$$E = h\nu = \frac{hc}{\lambda} \quad (1.2)$$

where $\nu = c/\lambda$ is the frequency, λ is the wavelength, h is the Planck constant and c is the speed of light. This emission forms the extreme ultraviolet part of the spectrum of a galaxy and is called the LyC, implying that it does not consist of (a series of) emission lines.

Such ionising emission can be produced through different mechanisms:

- thermally inside (massive) stars in star-forming galaxies
- through potential energy, in the accretion discs of an active galactic nucleus (AGN), powered by supermassive black holes

Here we focus on ionising emission produced in stars. Their properties, including the amount of ionising photons they produce, is mainly dominated by their mass (as well as their metallicity). The mass of a star determines its temperature and lifetime, with higher masses leading to higher temperatures, but a shorter life expectancy, as the star burns through its fuel of hydrogen faster. If we assume stars to emit black body radiation, meaning their energy output is determined by their temperature, we can apply Wien's radiation law, according to which the maximum wavelength of the spectrum of a black body is given by its temperature, with higher temperatures shifting the maximum to smaller wavelengths (and thus higher energies). This means the higher the temperature of a star, the more flux it emits at high energies.

Taking all this together, the ideal candidates for the production of hydrogen ionising photons are stars with a high temperature due to their high mass, which results in a short lifetime and most importantly, in a high output of highly energetic emission. Such stars can be found in the star-forming regions of galaxies, since those regions contain young, massive, hot stars that would already be missing in non star-forming areas of a galaxy.

To form a coherent picture, we need to connect the photons needed for reionising and keeping the universe ionised with the photons available at and after the epoch of reionisation. In theory, for each IGM hydrogen atom we need one ionising photon. The ionising photon rate \dot{n}_{ion} from star-forming galaxies can be expressed as ([Ellis, 2014](#), [McCandliss et al., 2019](#)):

$$\dot{n}_{\text{ion}} = f_{\text{esc}} \xi_{\text{ion}} \rho_{\text{SFR}} \quad (1.3)$$

The parameter which is best known (as it's most directly observable, see [Madau & Dickinson, 2014](#) for a review and below in section 1.2.1) is the SFR density ρ_{SFR} . The ionising photon production efficiency ξ_{ion} is less well-known and relies on the stellar population, the initial mass function (IMF) (the initial mass distribution of stars in a star forming region) and the influence of the metallicity. There have been efforts to directly measure this parameter using H α luminosities ([Bouwens et al., 2016](#), but also recently from

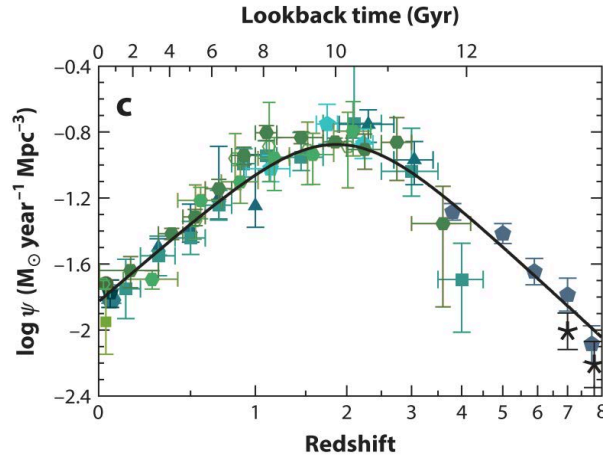


Figure 1.3.: Figure from [Madau & Dickinson \(2014\)](#), showing the SFR density over redshift using far ultra-violet (UV) and infrared data from different sources (see [Madau & Dickinson, 2014](#) for a discussion), all converted to instantaneous SFRs and valid for Salpeter IMFs. Although there are fewer data points at redshifts above $z = 3$, it can be seen that the SFR density peaks at $z \sim 2$.

galaxies at $z \sim 3.1$ by [Steidel et al., 2018](#)), and although equally uncertain, the fraction of ionising photons f_{esc} escaping the galaxy into the CGM and IGM can also be indirectly measured (see sections 1.4 and 1.5).

1.2.1. Star Formation Rate

The amount of ionising emission produced in a galaxy is influenced not only by the properties of the stars but also on the number of stars produced over time per volume (the SFR density ρ_{SFR}) and the distribution of masses of the stars when they are first produced, which is called the IMF. For high redshift galaxies, the stellar mass has to be inferred from the light they produce and the spectra of the galaxies are modelled using spectral energy distributions (SEDs), which uses our current knowledge on stellar formation and evolution to model the galaxies' observed light distribution.

The SFR density in the universe changes over time, as can be seen in figure 1.3, with a maximum between $z = 1.5-2$, which corresponds to an age of the universe of ~ 3.5 Gyrs ([Madau & Dickinson, 2014](#)). Looking at the development in figure 1.3, it is clear that at the epoch of reionisation only a fraction (around 1%) of the stars present today had already formed.

1.2.2. Metallicity and Population III Stars

Apart from the SFR and IMF, the metallicity can also influence the amount of ionising photons produced in stars. As we have seen in section 1.1.2, the Big Bang nucleosynthesis produced mostly hydrogen, some helium and traces of lithium. Heavier elements are produced in stars (up to iron) and supernova explosions, which means that the metallicity of stars and galaxies increases with time. At high redshifts the metallicities were lower and at very early stages of the universe stars with almost no metals⁶ must have existed. These are called population III stars (Pop III) and so far have not been detected conclusively. The question remains whether stars forming in a low metallicity environment will have a different IMF, maybe with more massive stars (top heavy IMF) produced than in current star-forming regions. These first stars were also the first to provide ionising emission to reionise the universe.

The ionising photon output of stars is linked to their metallicities. From stellar population models at low metallicities we know that Pop III stars can produce three times more LyC photons compared to solar metallicity stars (for a Salpeter IMF between $1 - 100 M_{\odot}$, [Schaerer, 2003](#), [Schaerer, 2002](#), see figure 1.4). This is due to their compactness at the zero age main sequence (ZAMS), which increases their temperatures (to effective temperatures of up to $\sim 10^5$ K, [Tumlinson et al., 2003](#), [Schaerer, 2002](#)). This can be seen in figure 1.5, as the Pop III star has a higher luminosity in the LyC than the Pop II star.

⁶In Astronomy, any element other than hydrogen and helium is called a metal. Thus, the metallicity of the ISM is measured usually relative to the abundance of hydrogen.

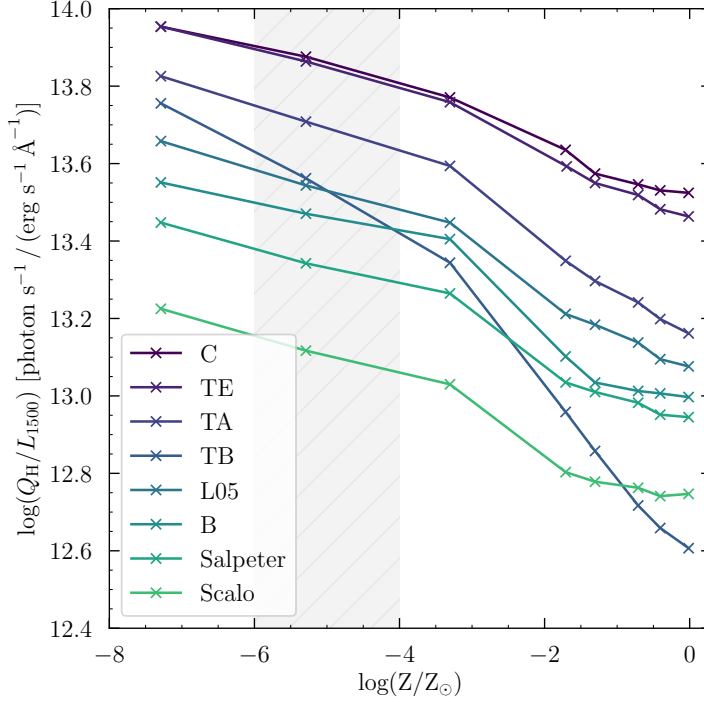


Figure 1.4.: Figure adapted from [Raiter et al. \(2010\)](#). The y-axis shows the logarithmic ionising photon output Q_H compared to the UV continuum luminosity L_{1500} measured at 1500 \AA . The x-axis shows the metallicity and the colour coding corresponds to models with different IMFs (see [Raiter et al., 2010](#) for a description) for a constant SFR at an age of 1 Gyr. The shaded area marks the critical metallicity between the Pop III regime on the left of the plot and the higher metallicities approaching the solar metallicity Z_\odot at the right end of the x-axis. For all IMFs, the metallicity is anti-correlated with the ratio of ionising over UV continuum photons.

Since the first Pop III stars are responsible for distributing the first metals in the universe as well as for providing the first LyC emission, the two properties are linked at high redshifts. Thus one can make a simple calculation of the number of available LyC photons per baryon n_{ion} while circumventing the need to understand the IMF at low metallicities ([Madau & Dickinson, 2014](#)), namely that

$$n_{\text{ion}} = \eta m_{\text{p}} c^2 Z_{\text{b}} Z_{*} / E_{\text{LyC}} \quad (1.4)$$

For this equation we need the rest proton mass $m_{\text{p}} c^2 = 938 \text{ MeV}$ which is multiplied by the conversion efficiency from rest mass to LyC emission, which is $\eta = 0.014$ for stars with a metallicity of $Z_{*} = Z_{\odot}/50$ ([Schaerer, 2002](#), [Madau & Shull, 1996](#) with Z_{\odot} being the solar metallicity). Extrapolating from literature measurements of the cosmic metallicity at lower redshifts, [Madau & Dickinson \(2014\)](#) derive a mean metallicity of the universe at redshift $z \sim 6$ of $Z_{\text{b}} = 7 \times 10^{-4} Z_{\odot}$. Dividing by the mean energy of a LyC photon $E_{\text{LyC}} = 22 \text{ eV}$ we get an estimated 8.4 LyC photons per baryon before a redshift of $z \sim 6$. In theory, only one LyC photon per hydrogen atom in the IGM is needed to reionise the universe and keep it ionised ([Madau et al., 1999](#)), $n_{\text{ion}} / \langle n_{\text{H}} \rangle > 1$, see the right panel in figure 1.6. However, this is further complicated by radiative transfer processes, absorption and radiative recombinations inside as well as outside of the galaxy.

In the right panel in figure 1.6 we can see the ionising photons per hydrogen atom $n_{\text{ion}} / \langle n_{\text{H}} \rangle$ over the redshift, calculated using the current best-fitting SFR. At the late stages of reionisation, at $z \simeq 6 - 7$, the available ionising photons range from 5 to 9 per hydrogen atom for the lowest metallicity. This means, if we need one ionising photon per hydrogen atom, the escape fraction of ionising photons has to be at least 20%. However, as mentioned above, assuming only one LyC photon for each hydrogen atom is an oversimplification. The hydrogen atoms in the IGM can recombine and to keep the universe ionised, closer to two LyC photons per hydrogen atom are needed ([Madau & Dickinson, 2014](#)), depending on the temperature of the IGM as well as the clumping factor.

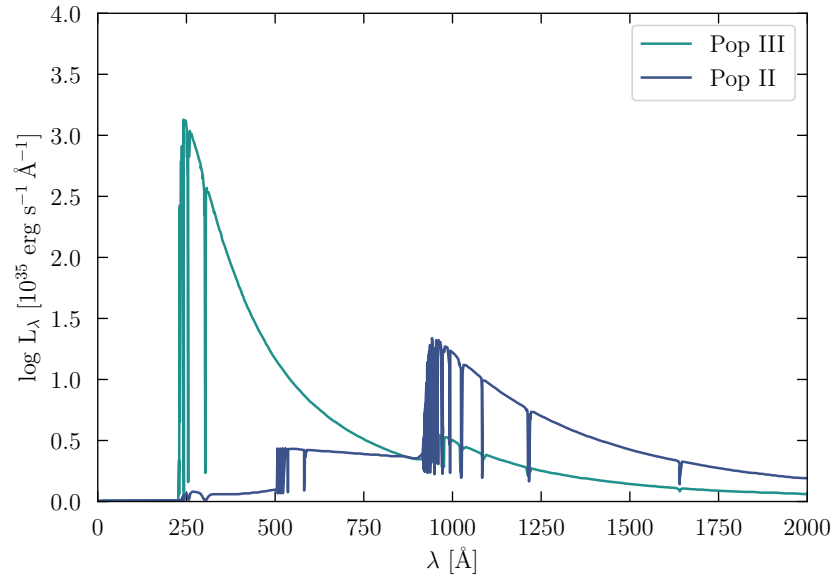


Figure 1.5.: Figure adapted from Tumlinson & Shull (2000), showing the spectrum of a Pop II star with a metallicity of $Z = 0.001$ and a temperature of $T_{\text{eff}} = 36\,000$ K and a Pop III star with a metallicity of $Z = 0$ and a temperature of $T_{\text{eff}} = 63\,000$ K, both with a mass of $M = 15 M_{\odot}$. The spectra only show H I, He I and He II lines.

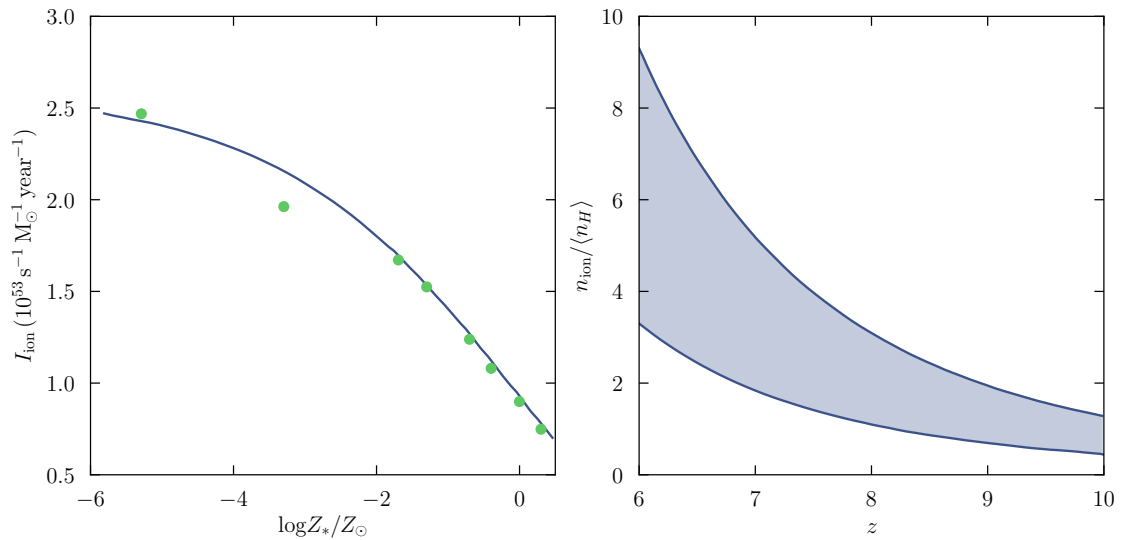


Figure 1.6.: Figure adapted from Madau & Dickinson (2014). The panel on the left shows the ionising photon yield I_{ion} over the metallicity, with data taken from Schaerer (2003) for a Salpeter IMF in a mass range of $0.1 - 100 M_{\odot}$. The lower the metallicity, the more ionising photons are produced per mass. The panel on the right shows the number of ionising photons per hydrogen atom emitted over redshift with the upper limit given by a metallicity of $Z_* = 0$ and the lower limit given by the solar metallicity $Z_* = Z_{\odot}$.

Masses of Pop III stars

The first stars at very high redshifts ($z > 20$) were not only metal-poor but also likely massive, which drives up their temperatures and thus ionising emission output.

Abel et al. (2002) use hydrodynamical simulations and find that in their pregalactic halo, only one star forms (with no further fragmentation). This star is metal-free and very massive ($M \gg 1 M_{\odot}$, likely above $100 M_{\odot}$). Cosmological simulations show that Pop III stars form through molecular hydrogen cooling (Abel et al., 2002, Yoshida et al., 2006) and have masses of $M_{\text{ZAMS}} \sim 100 M_{\odot}$ once they reach the ZAMS (Yoshida et al., 2006, Abel et al., 2002).

1.2.3. Properties of Stellar Populations at High Redshifts

For the escape fraction, it is important to know how much Lyman continuum is produced as compared to UV continuum, since the latter is easier to observe and can be used to infer the emitted Lyman continuum if their intrinsic luminosity ratio $(L_{\text{UV}}/L_{\text{LyC}})_{\text{int}}$ is known. We have seen already in figure 1.4 that the metallicity influences the ratio between LyC and UV continuum such that lower metallicity stars (which are expected in high redshift galaxies) have a higher ionising photon output. Other influences on the LyC to UV continuum ratio are discussed in detail in section 6.3 on page 120, where stellar population synthesis models are used to predict the intrinsic ratio $(L_{\text{UV}}/L_{\text{LyC}})_{\text{int}}$.

The properties of the stellar populations of one particular type of high redshift galaxies, the Lyman α Emitters (LAEs), are discussed further down in section 1.3.2. Generally speaking, high redshift galaxies have a higher SFR than present day galaxies, a lower metallicity, less dust, younger populations and smaller masses. In early galaxies, the fraction of high mass stars in stellar populations could have been higher, and stars with masses over $M > 9 M_{\odot}$ can have effective temperatures of $T_{\text{eff}} > 20\,000$ K, producing vast amounts of Lyman continuum emission. Additionally, for a fixed mass, the effective temperature of stars will increase for decreasing gas metallicity (Tumlinson & Shull, 2000, Schaerer, 2002).

The young ages, high SFR, low dust rate and low metallicity of star-forming high-redshift galaxies make them ideal sources of LyC emission, however their low masses and luminosities mean a lot of them are needed to provide enough ionising emission to ionise the universe. Before the LyC photons can reach the IGM however, they have to escape the vicinity of the stellar population in which they were formed and move through the ISM.

1.3. The Path through the Interstellar Medium

The stars in galaxies are surrounded by the ISM, which mostly consists of hydrogen and dust. There is a complex feedback mechanism between the ISM and the stars in a galaxy, as the gas clouds form the stars, which in return produce higher elements that enrich the ISM with metals.

The ISM consists of many different components apart from the hydrogen reservoir that fuels the star formation. There are dusty gas rich regions such as galactic bars or spiral arms in which the majority of stars form. There are also bubbles of ionised hydrogen created by star-forming regions or supernova explosions, which can be strong enough to blow the gas of the ISM outside of the galaxy, enriching the IGM with metals and tearing holes into the galaxy through which ionising emission can escape. Thus, the kinematics of the ISM can be just as complex as its morphology and influence the path of the ionising emission.

The ionising emission produced in the star-forming regions ionises the neutral hydrogen in the ISM. Star-forming regions can thus form bubbles of ionised hydrogen around them, called H II regions that let the Lyman continuum radiation through. Once a Lyman continuum photon ionises a hydrogen atom, the electron escapes the atom and recombines with the next hydrogen core (proton), either directly to the ground state through re-emitting another Lyman continuum photon or by cascading through downwards transitions, mainly resulting in a Lyman α photon, which can be seen in figure 1.7. We now have a new protagonist in our journey of Lyman photons through the universe, the Lyman α photon, derived from the Lyman continuum photon that started the journey. We will now follow both on their way through the ISM, CGM and eventually IGM.

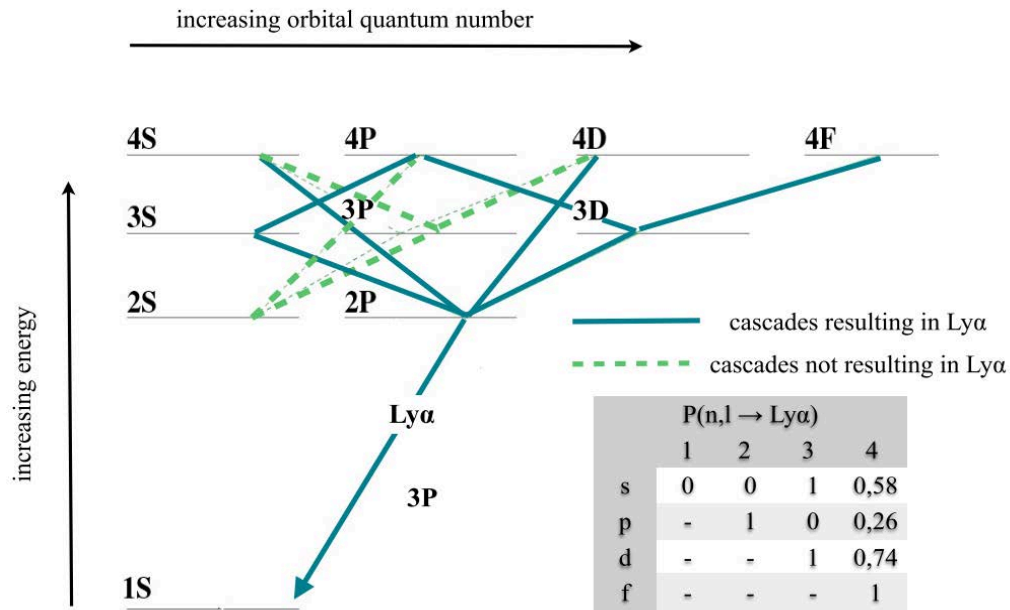


Figure 1.7.: Figure adapted from Dijkstra (2014). Shown are the different energy levels of a hydrogen atom with increasing principal number, and thus energy, along the y-axis. The x-axis shows the increasing orbital quantum numbers. Both numbers together define the quantum state of the atom. If an ionised hydrogen atom recombines with a free electron, its energy is emitted in a cascade of allowed transitions. Depending on the initial energy of the electron, it will populate one of the possible quantum states and has to follow the quantum selection rules to reach the lowest possible energy state. The orbital number l (here S, P, D, F, ...) has to change for each such transition. All possible transitions from the third excited state to the ground state are shown. The transitions that result in the emission of a Lyman α photon are shown in blue solid lines, while the transitions not resulting in a Lyman α photon are shown with green dashed lines. It can be seen that most of the possible cascades end in the emission of a Lyman α photon, as is shown in the inserted table to the bottom right of the schematic diagram.

1.3.1. Lyman α Emission

The most likely outcome of the series of transmissions shown in figure 1.7, as the hydrogen atom goes back to the ground state after recapturing an electron, is a Lyman α photon, the transition from the first ionised state to the ground state. Any emission line of any atom (not just hydrogen) from a higher energy level with a quantum number of $n \geq 2$ to the ground state is part of the Lyman series, discovered by Theodore Lyman in 1906 (Lyman, 1906). The Lyman series can be seen in figure 1.8, showing Lyman α at 1215.67 Å, Lyman β and Lyman γ lines at their wavelengths and energies, as well as a number of higher energy lines of the Lyman series, whose distances decrease until they reach the Lyman limit at 912 Å, which also corresponds to the binding energy of hydrogen. Any photon from a higher ionisation state is unbound.

There are two commonly assumed approximations to the recombination process in a hydrogen cloud surrounding a star formation region: Case A and case B (Baker & Menzel, 1938). They differ in the opacity of the gas cloud to Lyman photons and whether or not a recombination to the ground state is allowed (see Dijkstra, 2014 and Dijkstra, 2017 for reviews on Lyman α radiative transfer).

- case A: Any photon of the Lyman series is assumed to be able to escape the gas cloud without any further interaction with hydrogen atoms. The hydrogen cloud is optically thin to Lyman photons. Also, direct recombination to the ground state is possible, resulting in the emission of another ionising (Lyman continuum) photon (without cascading through the transitions in figure 1.7).
- case B: Any Lyman photon will be absorbed and re-emitted by a hydrogen atom immediately. Thus the Lyman photons scatter but do not contribute to the energy budget of the gas cloud, which is optically thick to Lyman photons. The direct recombination to the ground state is also inconsequential, as it is assumed that the ionising emission would ionise another neutral hydrogen atom immediately ('on-the-spot approximation'), creating a new free electron that will recombine with a hydrogen core.

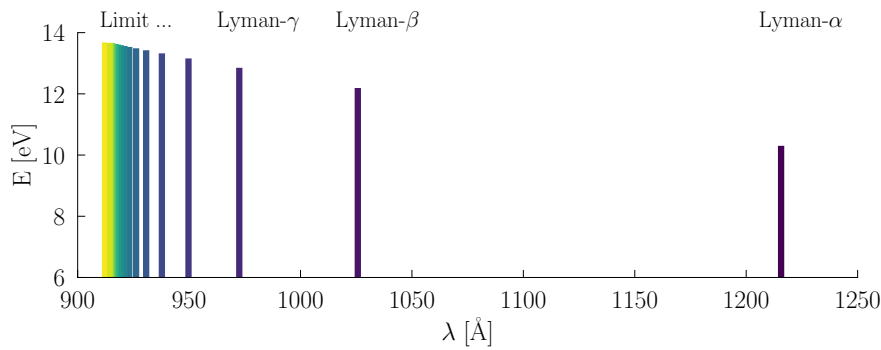


Figure 1.8.: Wavelengths (x -axis) and energies (y -axis) of the emission lines in the Lyman series, which consists of all transitions from higher energy levels ($n \geq 2$) to the ground state ($n = 1$).

Neither approximation fully describes any real H II region, but they provide the frame for possible scenarios and modelling, meaning anything in between both cases is a realistic assumption. Usually case B is assumed, but since we do observe both Lyman α and Lyman continuum emission from galaxies, those photons must at some point be able to leave the gas cloud. Thus, technically, the boundaries of a gas cloud should be considered as case A recombination.

The process above describes how a Lyman α photon is produced through recombination of a hydrogen core (proton) with a free electron. Another process that can produce Lyman α radiation is the excitation of hydrogen atoms by free electrons, who lose part of their kinetic energy in the process. Thus kinetic energy of the electrons in the gas cloud (which can be interpreted as its thermal energy or temperature) is converted into light, which cools the gas cloud. This process is also called 'cooling radiation' and can occur for example in regions of extended Lyman α emission (see section 1.4).

Whether the hydrogen atom was collisionally excited or the proton recombined with a free electron, afterwards the electron is in a higher energy quantum state and can either go directly to the ground state via emission of another Lyman continuum photon (only in case A) or cascade to the ground state through the emission of Lyman series photons. This only holds for a hydrogen gas that has a low enough density and temperature, otherwise the de-excitation would be triggered by collisions. Since the time between collisions is much longer (on the order of hundreds of years) than the lifetimes of excited states (milliseconds), the atom is de-excited through the intermediate states, as was shown in figure 1.7. Each transition has its own probability, but due to the quantum mechanical selection rules, some of the transitions are not allowed (the orbital number l has to change by one). Therefore the transitions in the table in figure 1.7 have a probability of 0 or 1. The higher the excited state, the more possible options and intermediate states are available. In the diagram, higher Lyman series transitions are not shown, as they can always result in Lyman α photons. A Lyman β photon starting in the 3D state could for example be absorbed and then either directly be re-emitted or cascade down as a Balmer α photon to 2P and then a Lyman α photon to 1S, the ground state. Since case B assumes that any photon from the Lyman series is immediately reabsorbed, this process continues until they end in Lyman α photons. The probability that a Lyman α photon is the result of such a recombination cascade is often given as $2/3$, but this value depends on whether case A or case B is assumed and also on the temperature of the gas. A more precise value for the probability would be $P_{\text{Ly}\alpha} = 0.68$, given case B recombination and a gas temperature of $T = 10^4$ K (Osterbrock, 1989, Dijkstra, 2014). With decreasing gas temperature, the probability can be even higher (see Cantalupo et al., 2008).

1.3.2. Lyman α Emitters

Some 50 years ago, Partridge & Peebles (1967) predicted that the Lyman α line should be useful for studying high-redshift galaxies. Not only is the line at a convenient wavelength, as it shifts into the optical part of the spectrum at a redshift of around $z \sim 3$, but it is also the most likely emission line from hydrogen recombination. As we have seen in section 1.2, star-forming regions of galaxies can produce a copious amount of Lyman continuum photons, which can then be transformed to Lyman α emission.

In the first prediction of the observability of LAEs Partridge & Peebles (1967) came to the conclusion that 6 – 7% of the total radiation of such a young galaxy could be converted to Lyman α emission (shown in their figure 1.9). However, they overestimated the observability of LAEs, which is why it took another 30 years before they became a useful tool to study the high-redshift universe. Before their discovery,

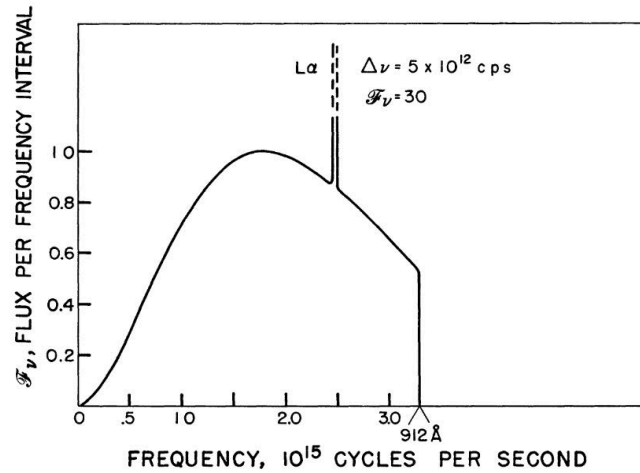


Figure 1.9.: Plot from Partridge & Peebles (1967) showing a simplified theoretical spectrum for a galaxy that converted all its Lyman continuum emission to Lyman series photons, specifically Lyman α . Note that the x -axis shows frequency, which means higher energies are to the right in this plot.

many studies tried to explain their absence, such as Neufeld (1990) who analysed radiative transfer and invoked absorption of the Lyman α line or Charlot & Fall (1993) who explain the null results in the searches for LAEs by the brief period of time in which a galaxy would be sufficiently dust-free to be bright in Lyman α . Eventually, observing technology became better and instruments more sensitive, and some of the first detections of LAEs were by Steidel et al. (1996), Hu et al. (1998), Rhoads et al. (2000) and others.

As we have learned in sections 1.3.1 and 1.2, the production of Lyman α photons is influenced by various properties of the galaxy, its ISM and stars: The initial production of the necessary ionising photons is determined by the temperature of the stars, their metallicity as well as their mass, which in turn is determined by the IMF. Since massive stars die young, the available number is determined by the SFR. The conversion from Lyman continuum photons to Lyman α photons needs neutral hydrogen which is either optically thin (case A) or optically thick (case B) to Lyman photons and its configuration (temperature and kinematics) also plays a role in the final Lyman α production. Therefore the strength of the Lyman α line can be a tool for studying the properties of star-forming regions in high-redshift galaxies.

Originally, the term LAE came from the observational techniques first used to find these objects, which were biased towards strong Lyman α lines. Traditionally, broad-band filters were compared to a narrow-band filter (see e.g. Finkelstein et al., 2015), which captures the excess flux in Lyman α , compared to the weaker flux (if any) in the broad-band filter(s), usually to the red side of the Lyman α line. This narrow-band technique allows to detect objects with strong Lyman α emission compared to the UV continuum and is most easily employed for galaxies with a strong contrast between Lyman α and the rest of the spectrum. At the epoch of reionisation itself (above redshifts $z \sim 6$), the IGM will absorb most of the Lyman α emission, making it difficult to observe. Therefore, the redshift range $3 \lesssim z \lesssim 6$, where Lyman α is shifted to the optical wavelength range where the sky background is lowest, is an ideal window to study LAEs.

In addition to the narrow-band technique which usually needs spectroscopic follow-up observations of (at least a fraction of) the detected objects to ascertain the redshift and confirm the existence of the Lyman α line, there are other possibilities of detecting LAEs, such as spectroscopic searches (e.g. Le Fèvre et al., 2015) and with the help of integral field units (IFUs), which combines the strength of photometry and spectroscopy, by providing a spectrum in each spatial pixel (e.g. Blanc et al., 2011, Bacon et al., 2017, Herenz et al., 2017, Urrutia et al., 2019, Inami et al., 2017). If the objects are detected using spectra or integral field spectroscopy, the definition of the LAE is usually fixed to the equivalent width of the Lyman α line: A somewhat arbitrary value of 20 \AA is often chosen to define LAEs (e.g. Kornei et al., 2010, Oteo et al., 2012) and distinguish them from other types of galaxies.

A technique similar to using narrow-band images is to find galaxies at high redshifts through their break in the spectrum at the Lyman α wavelength, which are called Lyman break galaxies (LBGs). In this case, multiple broad-band images are compared in order to find a break in their SED. Such galaxies can also have Lyman α in emission, but might not have a strong enough Lyman α line to be picked up with the narrow-band technique (possibly due to their low Lyman α equivalent widths). Therefore, the distinction between LAEs and LBGs is often only the detection technique that was used. Whether LAEs and LBGs

have fundamentally different properties or whether they are two incarnations of the same entity remains a topic of current scientific debate.

It could be that LAEs are just subsets of LBGs that have stronger Lyman α emission, as they share many properties (Jiang et al., 2013, Yuma et al., 2010, Verhamme et al., 2008). LAEs might be the later stages of the evolution of LBGs (Kornei et al., 2010), as the strength of the Lyman α line in LBGs correlates with the age of the stellar population and is anti-correlated with SFR and dust (Shapley et al., 2003, Kornei et al., 2010). The latter is derived from LAEs having bluer UV continuum slopes (or β -slopes (see section 4.1.2) than LBGs (Venemans et al., 2005 find $\beta = -1.76$ for LAEs and $\beta \sim -1.09$ for LBGs with weak Lyman α emission). Less dust would be an indicator of an earlier evolutionary stage, however the dust covering fraction could be reduced due to outflows created by supernovae (Kornei et al., 2010). An argument against the claim that LAEs are later stages of LBGs is the observation that they are typically smaller than LBGs (Venemans et al., 2005) at the same redshifts.

It has also been claimed that LAEs (at redshifts $z \sim 3$) might be the earlier evolutionary states of present-day Milky Way like galaxies. Most of their evolution might take place between redshifts of $z = 0 - 1$ (Cowie et al., 2011), with LAEs at higher redshifts representing the early stages of a starburst with small star-forming regions and low metallicities. Another hint at LAEs being the building blocks of present-day galaxies (Ono et al., 2010) is the fact that they are likely to evolve into galaxies with $L < 2.5 L_*$, (which has been found through the analysis of halo merger trees, see Gawiser et al., 2007), as opposed to other galaxies at this redshift, which evolve into more massive galaxies. The median halo mass of a resulting galaxy is close to that of an L_* galaxy. This makes LAEs not only interesting for understanding the ionisation history of the universe, but also for understanding the evolution of our own galaxy.

Properties of Lyman α Emitters

As described above, LAEs are often discovered using narrow-band surveys or IFU spectroscopy. Using multiple photometric bands or spectra, their SEDs can be modelled to derive properties like SFR (history), metallicity, age, mass and dust content.

Star formation rate history

There are four main scenarios for SFR histories (all of which are simplifications): constant SFR, (exponentially or linearly) increasing star formation, (exponentially or linearly) decreasing star formation and an instantaneous star burst. The real star formation history of a galaxy will be a mix of the four scenarios. Since as we learned in section 1.2 massive, young stars provide most of the LyC emission, the ionising emission output of a galaxy depends on the star formation history. If no new stars are formed, the LyC production quickly drops.

From SED fitting we know that high redshift star-forming galaxies such as LAEs can have various star formation history scenarios. McLinden et al. (2014) find that most of their 33 LAEs at a redshift of $z \sim 3.1$ have either instantaneous bursts or exponentially decreasing star formation, while only four are consistent with exponentially increasing star formation. Acquaviva et al. (2011) analyse LAEs at the same redshift and also come to the conclusion that an exponentially increasing SFR does not fit the SEDs better than a constant SFR. At higher redshifts and fainter magnitudes, the distinction between the different star formation histories becomes even more challenging (e.g. Karman et al., 2017 at redshifts $z = 3 - 6.1$).

Metallicity

High redshift galaxies have typically lower metallicities than present day galaxies. Samples of LAEs at redshifts of $z \sim 3.1$ have subsolar metallicities (Acquaviva et al., 2011, McLinden et al., 2014, Nilsson et al., 2007). Some of the brightest, most luminous LAEs at $z \sim 6.6$, at the end of the epoch of reionisation, possibly even contain metal poor Pop III stars. (Sobral et al., 2015). This claim has recently been put into doubt by e.g. Bowler et al. (2017), though, who do not find a strong evidence for Pop III stars in this particular galaxy.

Dust content

Although the dust in the galaxy is not directly connected to the stellar populations themselves, it can influence the transmission of Lyman α and Lyman continuum emission, as we will see in section 1.3.3. The dust content is also likely connected to the metallicity of the galaxy, as LAEs at high redshifts simply lack the time to go through the many cycles of star-formation and death necessary to enrich the ISM with metals. Typically, high redshift galaxies are low in dust (e.g. Gawiser et al., 2007, Gawiser et al., 2006, Ono et al.,

2010). Acquaviva et al. (2011) e.g. find no evidence for dust in their $z \sim 3.1$ LAE sample, Karman et al. (2017) find only few objects with $E(B - V) > 0.1$, which is also the median value in McLinden et al. (2014). At lower redshifts, the dust content of LAEs seems to be slightly larger (Vargas et al., 2014 find a median of $E(B - V) = 0.12$).

Ages and Masses

LAEs at high redshifts have ages that range from 1 Myr for faint, lensed LAEs (Karman et al., 2017), to ~ 10 Myrs (Gawiser et al., 2007 or Ono et al., 2010 for LAEs at redshifts 5.7 and 6.6) to ~ 1 Gyrs (Acquaviva et al., 2011, Finkelstein et al., 2015). Although LAEs typically have young stellar populations, the age and mass is often not well constrained and the values scatter significantly, with the mass ranging from $\sim 10^8$ to $10^{10} M_{\odot}$ (e.g. McLinden et al., 2014, Yuma et al., 2010). Faint ($L_{Ly\alpha} < \sim 10^{41.5}$) LAEs have even lower masses of $10^{6-8} M_{\odot}$ (Karman et al., 2017, Ono et al., 2010), which makes them comparable to massive globular clusters in the Milky Way. This dispersion in properties isn't captured in stacks of LAEs used for SED modelling (Vargas et al., 2014), which needs to be kept in mind for samples that are too faint to fit each object individually.

Equivalent Width as a Measure of Lyman α Line Strengths

The strength of the Lyman α line is usually measured by deriving the equivalent width, which is then used to classify a galaxy as an LAE, distinguish LAEs from LBGs or infer the stellar properties of the galaxy. In order to measure the equivalent width (EW) of an emission or absorption line such as Lyman α (in a way that is independent of the instrument and its resolution), the line flux is compared to the UV continuum flux density (which is the definition of the EW). Usually, the equivalent width is used for absorption lines, where it translates to the width the line would have in the continuum, but for emission lines this image works just as well. In this thesis, the EWs of emission lines are defined with a positive sign. Strictly speaking, the EW is defined as the integral over the difference between the flux density of the line $f_{\lambda, Ly\alpha}^{\text{line}}$ and the flux density of the continuum $f_{\lambda, Ly\alpha}^{\text{cont}}$, divided by the latter. However, if we assume the flux density of the continuum to be constant over the wavelength range of the Lyman α line, the equivalent width is simply the line flux divided by the continuum flux density. In this thesis, I give the equivalent widths in rest-frame, for which I need to take the redshift into account $EW = (1 + z)EW_0$.

$$EW = \int_{\lambda_0}^{\lambda_1} \frac{f_{\lambda, Ly\alpha}^{\text{line}} - f_{\lambda, Ly\alpha}^{\text{cont}}}{f_{\lambda, Ly\alpha}^{\text{cont}}} d\lambda \approx \frac{F_{\lambda, Ly\alpha}^{\text{line}}}{f_{\lambda, Ly\alpha}^{\text{cont}}} \quad (1.5)$$

Equivalent width distributions of samples of LAEs are used to compare their emission line strengths to models which take into account the ionising emission production and stellar population properties. An upper limit of the Lyman α rest-frame equivalent width of $\sim 240 \text{ \AA}$ has been suggested (first by Charlot & Fall, 1993 from stellar population models who claim that any $EW_0 \geq 200 \text{ \AA}$ has to be caused by AGN), above which the underlying stellar population has to have extreme properties such as low metallicities or an unusual, top heavy IMF (see also Laursen et al., 2013, Schaerer, 2003). Higher EWs would possibly indicate the presence of Pop III stars (Tumlinson et al., 2003) or a different behaviour of stellar populations at high redshifts. Hunting for such high EWs has subsequently become a goal for many studies, with varying results. Studies with a significant fraction of LAEs with $EW_0 > 200 \text{ \AA}$ include for example Malhotra & Rhoads (2002) (who find more than 60% of their 150 LAEs at $z = 4.5$ have high EW_0), Shimasaku et al. (2006) (who find 30 – 40% at $z = 5.7$), Zheng et al. (2014) (who find $\sim 44\%$ for LAEs at $z = 4.5$). A notable example of a study on a single high EW_0 LAE is Kashikawa et al., 2012, who find $EW_0 \sim 900 \text{ \AA}$ at $z = 6.5$, and conclude that this can be explained by a young stellar population with low-metallicity or even metal-free stars.

Examples for studies that do not find high EW_0 are Gawiser et al. (2006), Gronwall et al. (2007) and Guaita et al. (2011) with their sample of LAEs at $z \sim 3.1$ and $z \sim 2.1$. Indeed, they rule out a top-heavy IMF, but find low stellar masses and little dust, as is typical for LAEs. This could be due to a redshift evolution of EWs and thus LAEs. Figure 1.10 shows an example of distributions of EW_0 different redshifts.

Apart from the metallicity, other properties of the stellar population can increase EW_0 . Finkelstein et al. (2007) have found that younger stellar populations (~ 4 Myrs) have the highest equivalent widths, while the lowest EW_0 objects have ages around $\sim 40 - 200$ Myrs. Young stellar populations have more young, massive stars that produce Lyman continuum, the pre-requisite for Lyman α emission.

It should be noted though that the methods of measuring EW_0 are often not consistent across different

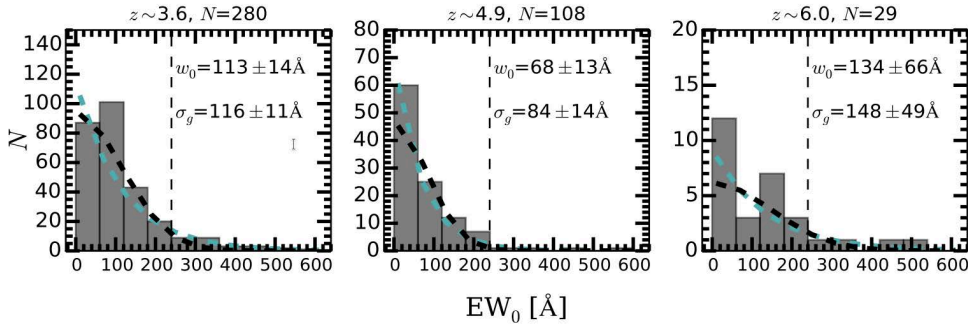


Figure 1.10.: Plot adapted from Hashimoto et al. (2017), showing equivalent width distributions for three different redshift ranges based on MUSE and Hubble Space Telescope (HST) observations of the Hubble Ultra Deep Field (HUDF). The black, vertical dashed lines indicates the $EW_0 = 240 \text{ \AA}$ limit, above which theoretical studies expect extreme stellar populations (or even AGN), the black and blue thick dashed lines show two different best-fit curves of the distributions (blue for an exponential law $N = N_0 \exp(-EW_0/w_0)$, with w_0 the scale factor and black for a Gaussian law $N = N_0 \exp(-EW_0^2/2\sigma_g^2)$ with σ_g representing the distribution width).

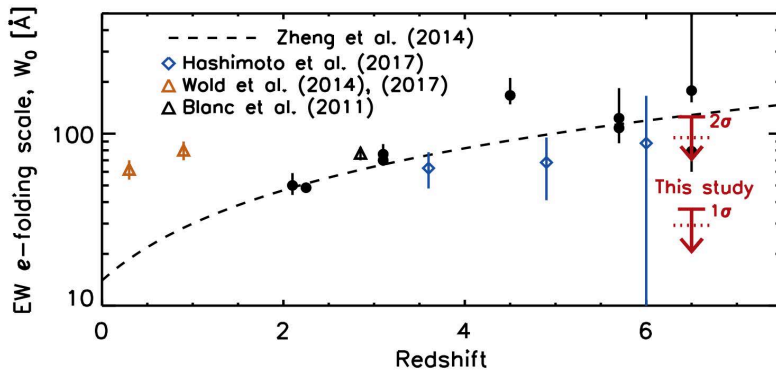


Figure 1.11.: Plot from Jung et al. (2018), showing a compilation of literature values of the redshift dependence of the e-folding scale W_0 , defined as $dN/dEW \propto \exp(-EW/W_0)$ (without IGM absorption correction).

studies, which can also explain the discrepancies. Especially with respect to if and how the IGM absorption of the Lyman α emission line is corrected for. As we will see in section 1.3.3, the Lyman α photons are scattered in neutral hydrogen, which increases the risk of hitting dust and being destroyed, but can also scatter them out of the line of sight and thus decrease the line flux. This can not only happen in the ISM, but also in the IGM at high redshifts, where a significant fraction of neutral hydrogen is still present (see section 1.5). Assuming a symmetric Lyman α line with a certain width, one could assume that the blue half of the line will be redshifted to the resonance wavelength of Lyman α while passing through the IGM and thus get scattered away. This would result in an asymmetric line profile with only the red part surviving, being far enough outside of resonance. And indeed, most Lyman α lines that have been observed spectroscopically are asymmetric, supporting the need to correct the intrinsic line flux for the IGM absorption. However, there are radiative transfer processes that can cause an asymmetric line profile even before the photons reach the IGM. This makes the process of determining the intrinsic rest-frame equivalent widths difficult and complex. Studies such as Zheng et al. (2014) or Malhotra & Rhoads (2002) apply an IGM correction by assuming a transmission of Lyman α of ~ 0.65 , which increases their measured EW_0 values when compared to studies that do not correct for the IGM. Due to the unknown effects of the IGM on the Lyman α line, Hashimoto et al. (2017) for example conclude that even though they only use MUSE data for their different redshift bins to stay self-consistent, they cannot state with absolute certainty that the equivalent width distribution does not change over redshift, as can be seen in figure 1.10. The same can be said for a recent comparison of equivalent width distributions over a larger redshift range by Jung et al. (2018) (see figure 1.11), although Zheng et al. (2014) claim to see a strong evolution in the Lyman α EW distribution with redshift (see the dashed line in figure 1.11).

Another possible explanation for the discrepancies in measuring equivalent widths is the fact that high equivalent widths mostly occur in objects with faint UV magnitudes. These objects are harder to detect,

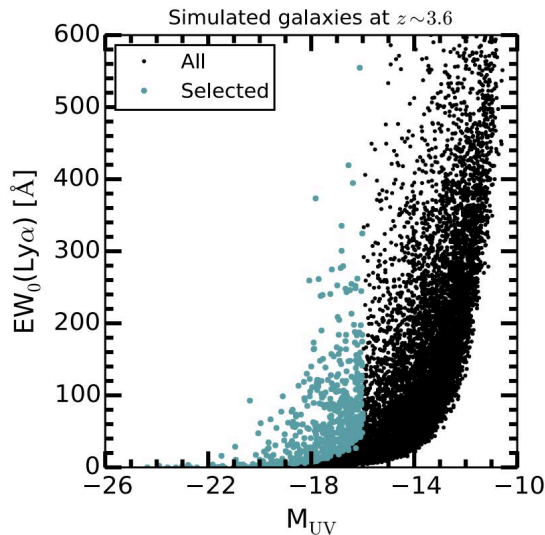


Figure 1.12.: Figure 11 adapted from Hashimoto et al. (2017), showing simulated galaxies at $z \sim 3.6$. The x -axis shows the UV magnitudes and the y -axis the Lyman α equivalent widths of the simulated galaxies. The black dots are the full sample of simulated galaxies, with Lyman α luminosities of $\log L_{\text{Ly}\alpha} = 40 - 44 \text{ erg s}^{-1}$, following the Lyman α luminosity function in Drake et al. (2017a) and Drake et al. (2017b). The blue dots show objects after applying selection cuts (to reproduce the observational constraints) with $L_{\text{Ly}\alpha} > 41 \text{ erg s}^{-1}$ and $M_{\text{UV}} < -16$. The lower boundary can be explained by the cut in Lyman α luminosity, while the apparent lack of high EW galaxies can be attributed to the selection in the UV magnitude. In section 4.4.2 I show a similar simulation and discuss the effects on the fraction of high EW objects. Since for these simulations no connection between the equivalent width and the UV magnitudes of the galaxies were assumed, as the EWs were drawn from an exponential distribution, the lack of high EW, UV bright galaxies is purely due to statistical effects, not underlying properties of UV bright galaxies that would prevent high EWs.

which means that the chosen significance cut in the sample can influence the resulting EW distribution. Similarly, for studies using broad- and narrow-band detections for the UV continuum and Lyman α line respectively, a none detection in the UV continuum band combined with a spurious detection in the Lyman α band can lead to a high lower limit on the equivalent widths. Gawiser et al. (2006) correct the Malhotra & Rhoads (2002) result to $\sim 10\%$ ($\sim 20\%$) of $\text{EW}_0 > 200 \text{ \AA}$ for the whole sample (only the spectroscopically confirmed) LAEs with a stricter cut on the continuum flux significance. These complications may be the explanation for the 'Ando effect' (see below).

The Ando Effect

When studying high redshift star-forming galaxies with high EWs of the Lyman α line, Ando et al. (2006) found a deficit in the number of high EW objects (both in LAEs and LBGs) with high UV luminosities for redshifts $z \sim 5 - 6$. This caused a debate on the cause of this lack of high EW objects, with suggestions on differences in the dust extinction, gas kinematics, age and neutral hydrogen content. The effect has been observed again and again in various studies, such as Shimasaku et al. (2006) for LAEs at $z = 5.7$, Ouchi et al. (2008) for LAEs with $z = 3.1, 3.7$ and 5.7 , Stark et al. (2010) for LBGs at $3 < z < 7$, Schaerer et al., 2011a for LAEs at $3 < z < 6$, Furusawa et al., 2016 who only find one Lyman α line in the spectra of 9 bright LBGs at redshift $z \sim 7$ or Ota et al. (2017) for LAEs at redshift $z = 7$, to name a few.

The effect can be explained by different properties of high and low UV luminosity galaxies (e.g. Forero-Romero et al., 2012, Verhamme et al., 2008, Verhamme et al., 2012). Garel et al. (2012) have found the Ando effect in their models and explain it by older populations in UV bright galaxies. Thus they have a smaller ratio of ionising emission over UV continuum, which leads to less Lyman α photons being produced. They also claim that UV bright galaxies have higher H I column densities, which could scatter and eventually absorb Lyman α photons (also seen in Verhamme et al., 2008). Another suggested explanation is that galaxies with low UV luminosities have less star formation, are thus more metal-poor and have less dust, which would facilitate the escape of Lyman α photons.

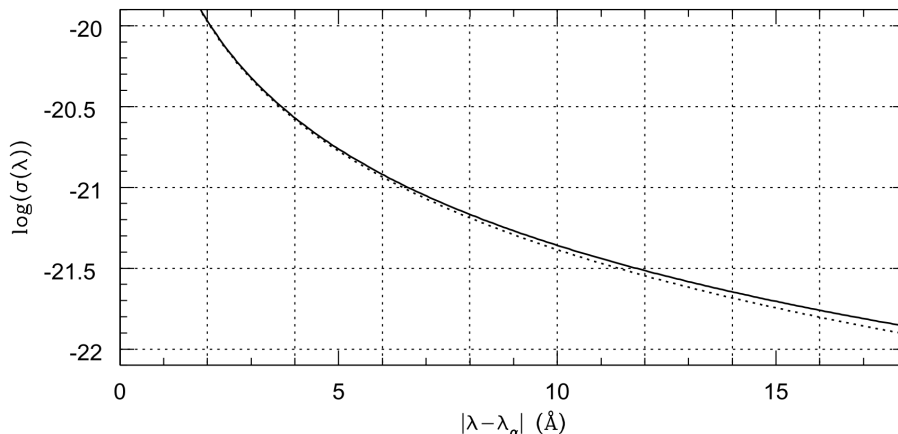


Figure 1.13.: Figure 3 from Lee (2013) showing the Lyman α scattering cross section σ [cm^{-2}] as a function of wavelength over the wavelength difference between the scattered photon and the Lyman α line centre at $\lambda_{\alpha} = 1215.67 \text{ \AA}$. The dashed line shows the cross section redward of Lyman α , the solid line shows the cross section blueward of Lyman α .

Although the Ando effect has been found in many studies, others have been unable to reproduce it (e.g. Kornei et al., 2010 find no evidence for LBGs at $z \sim 3$ and Zheng et al., 2014 for LAEs at $z \sim 4.5$). It has been proposed that rather than being a true deficit of high EW LAEs with higher UV luminosities, it could also simply be a selection effect. Nilsson et al., 2009 explain that the underlying bias could be caused by the limitations of the survey volumes. UV bright galaxies are rare, as are galaxies with strong Lyman α emission, which can lead to the apparent lack of a combination of both properties. Jiang et al., 2013 also discuss this problem in detail, showing that the apparent correlation can be traced back to their selection limits (which is demonstrated in Hashimoto et al., 2017, see figure 1.12). Jiang et al. (2013) also caution that the EW is both dependent on the line strength as well as on the UV continuum brightness, which causes both the values and the errors of the Ando effect (namely Lyman α EW and UV luminosity) to be correlated (Ciardullo et al., 2012). They propose to instead compare the line strength (instead of the EW) and continuum strength and find no correlation in such a plot (figure 6 in Ciardullo et al., 2012).

1.3.3. Radiative Transfer

Lyman α photons scatter resonantly in neutral hydrogen due to their large scattering cross section. This means their path through the ISM is influenced by the neutral hydrogen column density, the dust content and the kinematics of the ISM. All of those influences leave their imprint on the shape and strength of the Lyman α emission line compared to the UV continuum, which makes it a useful tool to infer the properties of galaxies at redshifts too high to observe any other spectral feature, if the processes of radiative transfer are properly understood.

A variety of different line shapes have been observed, from single peaks to multiple-peaked emission profiles (Kulas et al., 2012 find 30% of their Lyman α emitting star-forming galaxies show multiple peaks), which implies a variety of mechanisms that influence the radiative transfer. More complex Lyman α lines can have P Cygni profiles (e.g. in Kunth et al., 1998 and Mas-Hesse et al., 2003), double peaked profiles with asymmetric lines and a trough in the middle (e.g. Venemans et al., 2005, Tapken et al., 2004).

To explain the observed Lyman α line profiles, several radiation transfer models have been developed, most notably by Verhamme et al. (2006), Schaerer et al. (2011b) and Duval et al. (2014) (but see also Neufeld, 1990, one of the first to model radiative transfer in an attempt to explain the lack of LAE observations at the time). Such radiation transfer codes, often based on Monte Carlo simulations, take into account UV photon sources and velocity fields, hydrogen density, ionisation, dust distributions and temperature structures combined with several geometrical configurations. The most common radiative transfer models are based on a static plane (e.g. Neufeld, 1990), an expanding shell or an infalling shell of material that scatters the Lyman α photons. All these aspects can influence the path of the Lyman α photons and the shape of the Lyman α line.

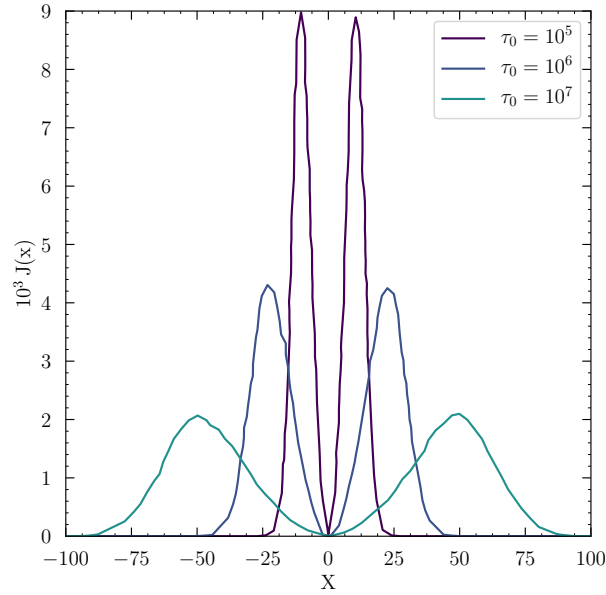


Figure 1.14.: Figure 1 from [Dijkstra et al. \(2006a\)](#) showing radiative transfer solutions for three different optical depths, $\tau_0 = 10^5$ for the narrowest line profile (purple), $\tau_0 = 10^6$ for the middle profile (blue) and $\tau_0 = 10^7$ for the broadest profile (teal). The x -axis is a measure of the frequency, the y -axis shows the strength of the line. The model shows line profiles for static, uniform, plane-parallel neutral hydrogen gas clouds (called 'slabs') with a source of Lyman α photons at the center.

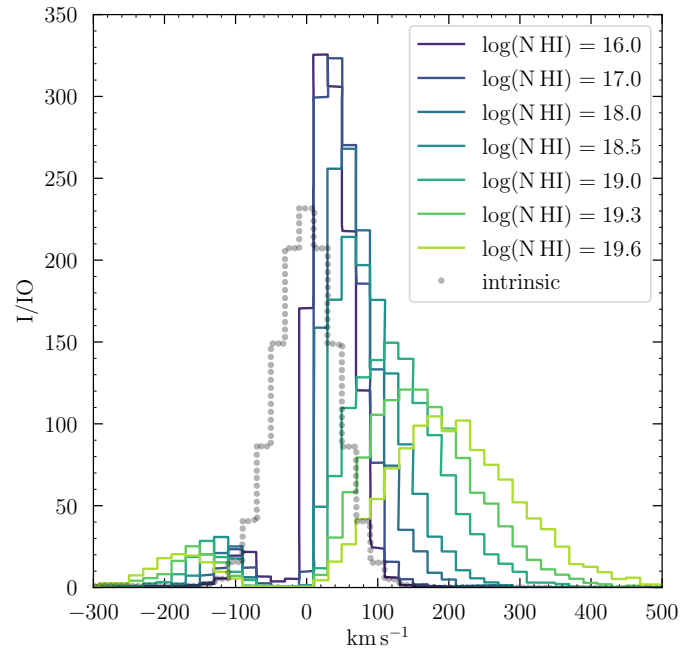


Figure 1.15.: Figure 1 from [Verhamme et al. \(2015\)](#) showing different Lyman α line profiles in an expanding shell model for varying neutral hydrogen column densities, from $\log(\text{N HI}) = 16$ (purple) to $\log(\text{N HI}) = 19.6$ (bright green) with the intrinsic line shape shown as the dotted grey line for a Gaussian emission line with a full width at half maximum of $\text{FWHM} = 100 \text{ km s}^{-1}$ and $\text{EW} = 100 \text{ \AA}$. Although the models have different neutral hydrogen column densities, they all have the same expansion velocity of $v_{\text{exp}} = 50 \text{ km s}^{-1}$ and a Doppler parameter of $b = 20 \text{ km s}^{-1}$. The effects of increasing neutral column hydrogen density are wider line separations, larger full width at half maximum (FWHM) and lower peak amplitude. The weakness of the blue part of the line compared to the red part is due to the outflow velocity, see section 1.3.3 on kinematics of the ISM.

Neutral Hydrogen Column Density

One of the most important influences on the Lyman α radiative transfer is the neutral hydrogen column density. It influences both the shape and indirectly also the strength of the Lyman α emission line. Since Lyman α is scattered in neutral hydrogen (see figure 1.13, showing the Lyman α scattering cross section) and under case B assumptions is immediately reabsorbed, Lyman α photons get trapped for a while in neutral hydrogen gas clouds. They follow a random walk through the neutral hydrogen until they eventually reach the edge of the gas cloud, where the neutral hydrogen column density is low enough to let them escape. This can lead to a longer time spend in the gas cloud, increasing the probability of hitting a dust grain and being absorbed, meaning the end of the Lyman α photon. However, apart from the increased risk of destruction, the scattering not only changes the direction of the Lyman α photon, each scattering event changes the photons frequency in the observed frame (due to the atoms' motion, see e.g. explanation in [Dijkstra et al., 2006a](#)).

This frequency scatter can result in the photon being shifted into the wing of the distribution, which increases its mean free path and the likelihood of being scattered by neutral hydrogen is reduced (as seen in figure 1.13 by the decreasing cross section with increasing wavelength shift). If it is reduced enough, the photon can escape the neutral hydrogen gas cloud. This results in the often observed doubled peaked line profile of the Lyman α line (see figure 1.14), as photons at the line centre are trapped inside the neutral hydrogen, while photons that are shifted either to the blue or red are able to escape. The depths of the trough in the middle and the distance of the two peaks is determined by the optical depth, given by the neutral hydrogen column density (see figure 1.15).

Geometry and Dust Content of the Interstellar Medium

Not only the neutral hydrogen column density or the presence of dust in the ISM can influence the escape of Lyman α photons. It has also been suggested that the geometry and distribution of the ISM can play a role in the escape and the strength of the Lyman α line. Moving on from static, uniform, plane-parallel 'slabs' as used by [Neufeld \(1990\)](#) and [Dijkstra et al. \(2006a\)](#), we now consider differently shaped neutral hydrogen gas clouds.

While the presence of dust can lead to the preferred destruction of Lyman α photons as compared to UV continuum photons due to the longer path lengths of the Lyman α photons when scattered in neutral hydrogen, there are also proposed scenarios in which dust localised in gas clouds can actually have the opposite effect and boost the Lyman α to UV continuum ratio, as suggested by [Neufeld \(1991\)](#) and thus called the 'Neufeld-scenario'. The path of a Lyman α photon and a UV continuum photon in such a scenario can be seen in figure 1.16. The Lyman α photon is scattered off of the surface of the high density gas clouds instead of penetrating them, thus reducing the probability of being absorbed by dust in the clouds. The inter-cloud medium is assumed to be nearly free of scattering neutral hydrogen or dust, allowing the Lyman α photons to move freely. The UV continuum will run a higher risk of dust absorption in the gas clouds, as it is not scattered away. The result would be increased Lyman α escape compared to the UV continuum, thus a higher equivalent width, but also a narrower Lyman α line ([Neufeld, 1991](#)). This scenario could explain the high equivalent widths measured in some LAEs at high redshift, which are above the maximum values expected from calculations (see section 1.3.2).

There are some caveats with the Neufeld scenario. For one it is unlikely to have a configuration where the scattering and non-scattering material is distributed in such a binary way as to make it possible to consider the gas clouds as single particles that have a possibility of scattering or absorbing Lyman α , as in [Neufeld \(1991\)](#). It is a strong simplification, necessary in the absence of better knowledge on the exact geometry of the ISM of high redshift galaxies, but to be taken with a grain of salt. [Finkelstein et al. \(2007\)](#) have tried to verify the theory that a clumpy ISM can boost the Lyman α equivalent width, but could mostly rule this out for their galaxies at $z \sim 4.5$. In a theoretical approach, [Gronke & Dijkstra \(2014\)](#) use their own radiative transfer code to analyse the Neufeld scenario. They do find a boost of Lyman α EW_0 in some sight lines, but only in models with Lyman α and UV continuum point sources. They conclude however that the Lyman α EW_0 can still be increased in certain directions, but more due to a reduced UV photon escape fraction than an increased Lyman α escape fraction. This can occur if the UV continuum sources are behind a dusty neutral hydrogen cloud, while the Lyman α emission can escape more isotropically. Similar to [Gronke & Dijkstra \(2014\)](#), [Duval et al. \(2014\)](#) find that the Neufeld scenario only boosts the Lyman α EW_0 under special circumstances, namely if the clumpiness is extremely high (density contrasts of $> 10^7$ between the inter-clump medium and the clumps), the clumps are extremely dusty (with $E(B - V) > 0.30$) and almost static. If those conditions are not met, the Lyman α line profile is not discernible from a homogeneous

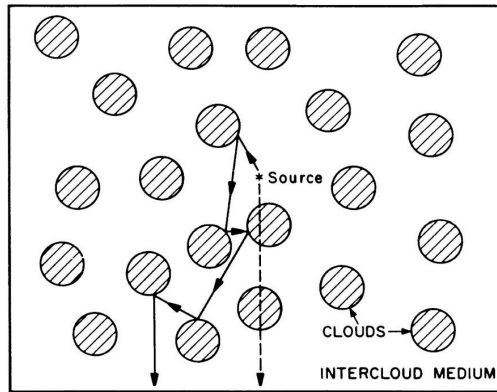


Figure 1.16.: Figure 1 from [Neufeld \(1991\)](#) showing the proposed mechanism of boosting Lyman α escape from the ISM through a clumpy medium. The gas clouds are shown as hatched circles and have larger absorption and scattering coefficients than the inter-cloud medium ([Neufeld, 1991](#)). The source of Lyman α and UV continuum photons is marked in the middle with a star, the paths of a Lyman α photon and a UV continuum photon are shown as a solid and a dashed line respectively. In this scenario, the Lyman α photons are scattered at the surfaces of the gas clouds, which increases their path lengths, but since the inter-cloud medium is assumed to be negligible, the Lyman α photons tend to not penetrate the clouds far enough to be absorbed, in contrast to the UV continuum. This could explain a boost of Lyman α escape.

medium.

Another way to boost the escape of Lyman α and even Lyman continuum photons more efficiently, is to clear their paths mechanically through supernova outflow-driven holes in the ISM of the galaxy. This can create channels through which the photons escape, increasing the Lyman α equivalent width ([Behrens & Braun, 2014](#)). In figure 1.17 we can see the influence such a cavity in the ISM can have on the Lyman α EW_0 .

Thinking on a larger scale, the escape of Lyman α and continuum photons can also depend on the viewing angle of the galaxy. If we assume a morphology of high-redshift LAEs that is similar to present-day spiral galaxies, there are regions in the LAE that contain more dust, such as the spiral arms. If the galaxy is viewed edge-on, the Lyman α photons need to travel through the disc of the galaxy and run the risk of getting absorbed in the dusty gas clouds before escaping. Viewed face-on, the photons can escape more directly, which is why some theoretical models have shown that the Lyman α escape and EW_0 is viewing-angle dependent ([Behrens & Braun, 2014](#), [Verhamme et al., 2012](#)), see figure 1.17.

Kinematics of the Interstellar Medium

Most observations of Lyman α emission line profiles are not as symmetric around line center as shown in the model from [Dijkstra et al. \(2006a\)](#) in figure 1.14 above. Usually, only one line is seen, which is asymmetric with a broad red wing and redshifted with respect to the systemic redshift (e.g. [Kunth et al., 1998](#)). In some cases ([Kulas et al., 2012](#) find 30%, [Yamada et al., 2012](#) even 50%), the blue part of the emission line is still visible, but mostly with a smaller line flux than the red main line. This blue part has been called the 'blue bump' in the literature and can have two causes:

- Not only is the ISM mainly composed of hydrogen, there are also copious amounts of it in the IGM. Since Lyman α is scattered in neutral hydrogen and the blue bump is eventually redshifted into resonance after leaving the galaxy, it can be scattered out of the line of sight in the IGM. This is discussed in section 1.5.
- The ISM is usually not static, it can expand, produce outflows (e.g. from supernova explosions) or have infalling gas. Depending on the kinematics of the ISM, the blue bump may be smaller intrinsically than the red, main peak.

Apart from static media or small systematic velocities, the Lyman α profile will not be symmetric and the blue bump will be smaller (stronger) for outflows (infalling gas). For a high column density ($N_H > 10^{20} \text{ cm}^{-2}$), the red main peak is shifted by approximately twice the expansion velocity of the outflowing gas cloud ([Verhamme et al., 2006](#)).

In an expanding shell model, the Lyman α emission is scattered at the far side of the expanding shell and

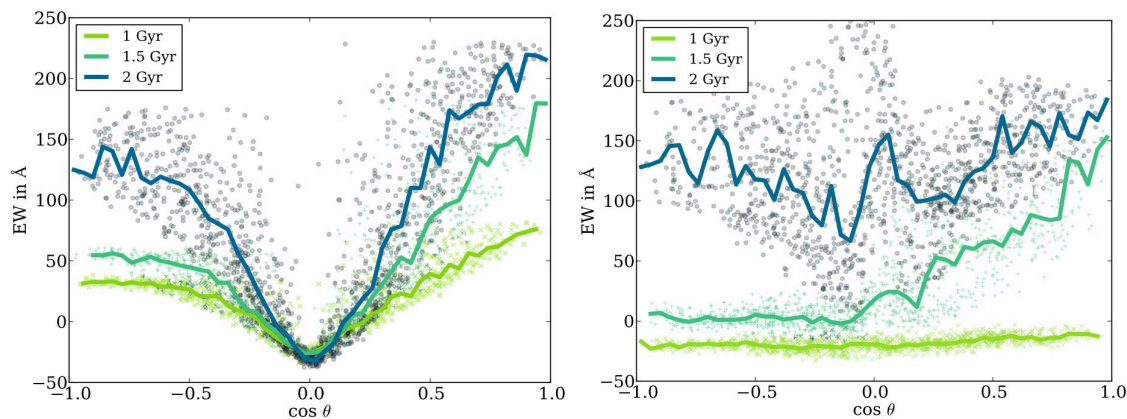


Figure 1.17.: Figures 4 and 7 from [Behrens & Braun \(2014\)](#). The results of a simulation of Lyman α radiative transfer is shown for three different ages of the stellar population (1 Gyr in light green, 1.5 Gyr in aqua and 2 Gyrs in blue). The dots show the Lyman α EW_0 for random lines-of-sight as a function of $\cos\theta$, with 0 being edge-on and 1 and -1 face-on. The solid lines are the binned median values. The two panels show different dust content realisations. The asymmetry in the simulation for 1.5 Gyrs in the right panel comes from a strong cavity that allows Lyman α to escape only to the top side of the galaxy, but not the bottom. This again illustrates the importance of outflows and holes in the ISM to facilitate the Lyman α (and Lyman continuum) escape.

scattered back into the line of sight, now redshifted for the observer. Thus the mean free path in the near side of the expanding shell will be large enough for the redshifted photons to escape ([Alexandroff et al., 2015](#)). Such a simple shell model has been used extensively in the literature to describe radiative transfer in LAEs (e.g. [Verhamme et al., 2008](#), [Hashimoto et al., 2013](#), [Hashimoto et al., 2015](#), [Karman et al., 2017](#)) and is a successful improvement of the static ‘slab’. In the presence of other emission lines in the spectrum, the Lyman α velocity shift can be compared directly to the systemic redshift obtained from a non-resonant line. [McLinden et al. \(2011\)](#) find velocity offset between [O III] and Lyman α in two galaxies at $z \sim 3.1$ of 342 and 125 km s^{-1} , hinting at strong outflows.

Still, as e.g. [Chonis et al. \(2013\)](#) have demonstrated by fitting three LAEs from [Adams et al. \(2011\)](#) (the pilot survey of the Hobby-Eberly Telescope Dark Energy Experiment, HETDEX), more complicated Lyman α line shape profiles containing more than two peaks, cannot easily be fit with a simple expanding shell model. Interestingly, [Marchi et al. \(2019\)](#) find that the Lyman α velocity shift is smaller for larger velocity shifts in interstellar absorption lines (a measure of the ISM outflow velocities), which is contradictory to the expanding shell model that would expect a positive correlation between the two. They explain their results with the influence of the neutral hydrogen column density, which can result in a small Lyman α velocity shift for large outflow velocities if the neutral hydrogen column density is small as well ($\sim 10^{19} \text{ cm}^{-2}$) and large Lyman α shifts for small outflow velocities if the neutral hydrogen density is large enough ($\sim 10^{20} - 10^{21} \text{ cm}^{-2}$). [Chonis et al. \(2013\)](#) also caution that the neutral hydrogen column density plays the dominating role on the Lyman α peak separation and that this correlation also depends on the spectral resolution. This was tested by [Gronke et al. \(2015\)](#) who tried to reproduce their input parameters from noisy, simulated spectra created with an expanding shell model. They use an automated line fitting algorithm and find that while they can indeed recover the outflow velocity and the neutral hydrogen column density well, other parameters such as dust content and effective temperature were not reproduced. This goes to show that the process of radiative transfer is indeed dominated by the amount and distribution of the neutral hydrogen as well as its kinematics.

Another degree of complexity of the kinematics of the ISM is the introduction of rotation of the expanding shell. As can be seen in figure 1.18 by [Remolina-Gutiérrez & Forero-Romero \(2019\)](#), the result of including rotation in models of radiative transfer is an increased flux at the line center, a broadening of the line and a further dependence on the viewing angle. Rotation could be detected in observational data by probing the line profile at two different locations (potentially possible for a large Lyman α halo, see section 1.4.1, a lensed LAE or a low-redshift Lyman α analogue, section 1.3.4). In addition to rotation of the ISM, a turbulent medium can also lead to a higher escape of Lyman α photons (e.g. [Herenz et al., 2016](#), [Tapken et al., 2004](#))

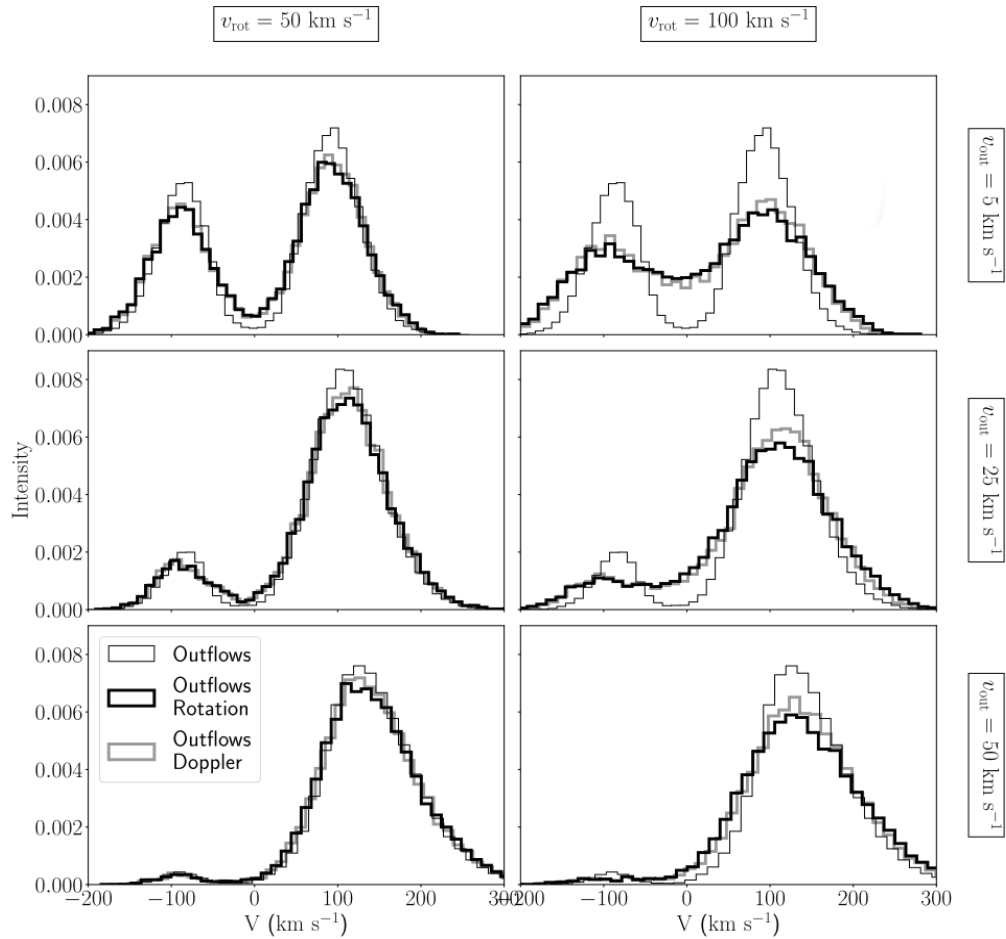


Figure 1.18.: Figure 1 from [Remolina-Gutiérrez & Forero-Romero \(2019\)](#) showing modelled Lyman α line profiles for expanding shell models that are rotating (viewed perpendicular to the rotation axis). The left panels have a rotation velocity of 50 km s^{-1} , the right panels have twice that at 100 km s^{-1} . From top to bottom, the outflow velocity increases (5 km s^{-1} , 25 km s^{-1} and 50 km s^{-1}). The thin black line shows the model without rotation but with an outflow, the thick black line shows the results for the model with both outflow and rotation and the thick gray line shows results for modelling only outflows and modifying the results by the Doppler shift of the rotation. For a viewing angle that is parallel to the rotation, the line profiles are almost identical to the case without any rotation.

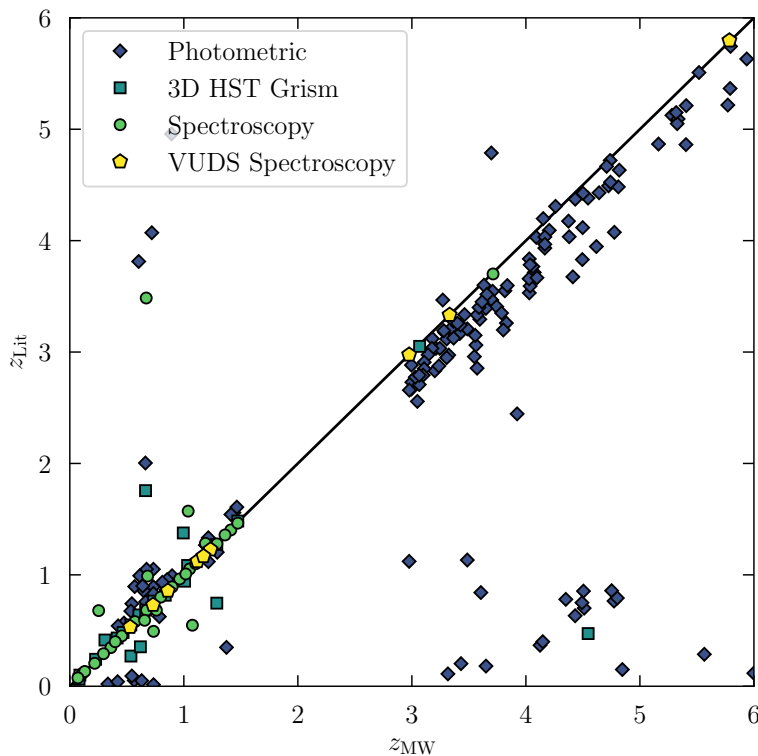


Figure 1.19.: Part of figure 12 from [Herenz et al. \(2017\)](#) showing the difference between spectroscopic redshifts from MUSE (x -axis) and literature redshifts from photometry (blue diamonds, [Skelton et al., 2014](#)) and spectroscopy (dark green squares, [Momcheva et al., 2016](#), green circles, [Wuyts et al., 2008](#) and yellow pentagons, [Tasca et al., 2017](#)). It can clearly be seen that for high redshift objects (mostly LAEs), there is a systematic offset between the MUSE redshifts and the photometric redshifts (even disregarding the ‘catastrophic failures’ where a completely different redshift was found).

Measuring the Redshifts of Lyman α Emitters

As we have seen in section 1.3.3, the Lyman α line can be asymmetric, double or multiple peaked, with one peak stronger than the other or symmetric, shifted or centered on the systemic redshifted Lyman α wavelength. This makes the line highly unreliable for measuring the redshift of the LAE ([Rakic et al., 2011](#) discuss the velocity offset of Lyman α found in foreground galaxies of quasi-stellar objects (QSOs) at $z \sim 2.5 - 3$, [Erb et al., 2014](#) find 30 of their 36 LAEs are redshifted with respect to the systemic redshift, [Song et al., 2014](#) show velocity offsets between Lyman α and H α and/or [O III] for LAEs at $z = 2.1 - 2.5$, [Hashimoto et al., 2015](#) find an offset for LAEs at $z \sim 2.2$) and if possible, other lines should be used to get the systemic redshift (e.g. [McLinden et al., 2011](#) showed the offset between Lyman α and [O III]). If present, the C III] doublet at 1907 and 1909 Å is the best candidate for measuring the systemic redshift (e.g. [Maseda et al., 2017](#)), as it originates from the same locations as Lyman α and is an optically thin nebular line. Often however, LAEs are selected by their prominent Lyman α line and have no visible continuum for absorption lines and also no other strong emission lines.

Thus, the Lyman α line is often the only spectral feature by which to measure the redshift. Even if there is photometric data available as well, SED fitting to get the redshift from photometry is less reliable than the spectroscopic redshift, even using Lyman α . This has been shown for example in [Brinchmann et al. \(2017\)](#) and in [Herenz et al. \(2017\)](#) (see figure 1.19) where a systematic offset between the redshifts of high-redshift LAEs in MUSE and photometric redshifts of the same objects ([Skelton et al., 2014](#)) was found, with the spectroscopic redshift from MUSE being larger with a median of $\Delta z = 0.26$ ([Oyarzún et al., 2016](#) find a similar offset).

Suggestions have been made on how to correct the redshift derived from the Lyman α emission line, taking into account various line properties such as equivalent width (as it relates to the shift from outflows, see [Adelberger et al., 2003](#) and [Shapley et al., 2003](#), but with significant scatter). From modelling radiative transfer processes and trying to reproduce observed Lyman α line shapes, we know that the shift with respect

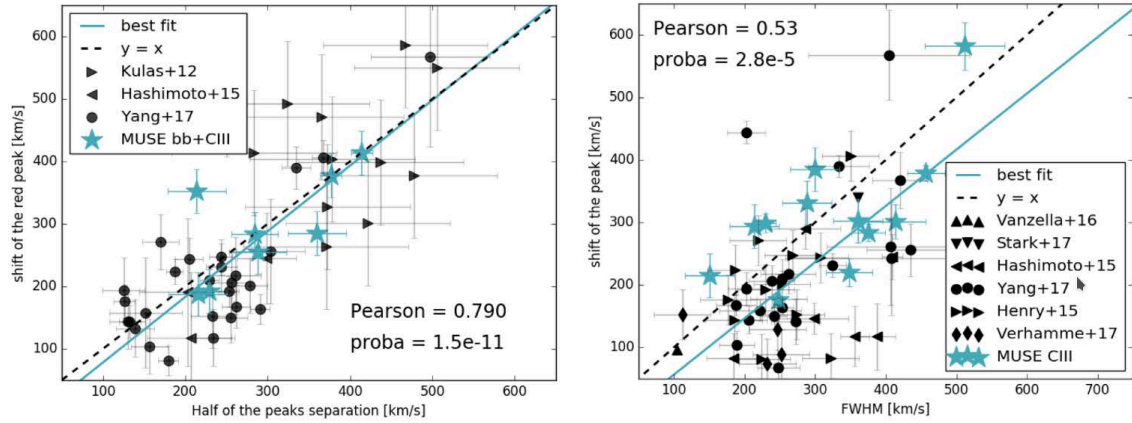


Figure 1.20.: Figure 2 from Verhamme et al. (2018) showing the correlation between the shift of the red peak and half of the separation between the blue bump and the red peak (left) and FWHM of the Lyman α line (right). For all LAEs, the systemic redshifts were obtained from other spectral features, mostly the C III] doublet. The blue lines show the fits to the data, while the black dashed lines show the one-to-one relation.

to the systemic redshift correlates with the neutral hydrogen column density, which influences not only the distance between the blue bump and the main Lyman α line, but also the FWHM of the line. Therefore, Verhamme et al. (2018) showed that the systemic redshift can be estimated from either, depending on whether or not a blue bump is visible. If the line is double peaked (Kulas et al., 2012 find almost 2/3 of their Lyman α emitting star-forming galaxies have double peaked profiles, Yamada et al., 2012 find 50%), the centre between the lines correlates with the shift of the red peak, if there is no double peak, the FWHM of the main, red peak can be used instead (see figure 1.20).

This correlation can be used to get a more precise estimate of the systemic redshift of LAEs that only have the Lyman α emission line as a redshift indicator. The relation between the velocity shift of the red line $V_{\text{peak}}^{\text{red}}$ and the double peak separation $\Delta V_{1/2}$, as derived from the fit in figure 1.20, is close to the one-to-one relation which is expected from radiative transfer modelling and is given by Verhamme et al. (2018) as:

$$V_{\text{peak}}^{\text{red}} = 1.05 (\pm 0.11) \Delta V_{1/2} - 12 (\pm 37) \text{ km s}^{-1} \quad (1.6)$$

In the absence of a blue bump, the relation between the velocity shift of the red line $V_{\text{peak}}^{\text{red}}$ and the FWHM can also be used (see also Zheng & Wallace, 2014), which is somewhat less significant, but still compatible with a one-to-one relationship. Verhamme et al. (2018) estimate this correlation to be:

$$V_{\text{peak}}^{\text{red}} = 0.9 (\pm 0.14) \text{FWHM}(\text{Ly}\alpha) - 34 (\pm 60) \text{ km s}^{-1} \quad (1.7)$$

The LAE redshifts in this thesis are all corrected with these relations to be as close to the systemic redshift as possible. Since the correlation between the shift of the red peak and the peak separation is slightly more robust, one should use this in case of double peaks, and the correlation with the FWHM in case of single peaked Lyman α lines. Thus the systemic redshift is recovered from the Lyman α line for redshifts between $z \sim 0 - 7$ with an uncertainty less than $\pm 100 \text{ km s}^{-1}$ (Verhamme et al., 2018).

1.3.4. Lyman α Emitters at Different Redshifts

Not only are LAEs (and star-forming galaxies in general) prime candidates for providing the LyC emission at the EoR, but they are also expected to keep the IGM ionised at later stages of the universe. How LAEs evolve over time is therefore an important question to ask. Additionally, some of their properties are difficult to observe at high redshifts, like their (rest-UV) morphology and spectral features (other than the Lyman α line). An additional challenge with faint high redshift galaxies is the surface brightness dimming with redshift, which decreases the surface brightness as $(1+z)^3$ for unresolved emission lines and $(1+z)^4$ for resolved emission. To mitigate this problem, stacks are used, but they don't preserve the huge variety in properties of LAEs (e.g. Vargas et al., 2014). This is why LAEs are not only studied at the EoR itself, but also both in intermediate redshift ranges and locally by searching for low-redshift LAE analogues.

Local Lyman α Emitter Analogues

There are several types of objects in the local universe that have properties similar to high redshift LAEs (LAE analogues). The Lyman α Reference Sample (LARS, Östlin et al., 2014) for example targets 14 galaxies at redshifts $z = 0.028 - 0.19$ with large $H\alpha$ equivalent widths to select galaxies with the potential of Lyman α emission. They find that the Lyman α emission is not as concentrated and more extended than the UV continuum or $H\alpha$ line (Hayes et al., 2013), implying that equivalent width measurements of Lyman α of high redshift LAEs might be underestimated if the extended nature of the Lyman α emission is not taken into account (Hayes et al., 2014). The morphologies of the LARS LAEs are consistent with irregular, starburst galaxies or merging systems (Guaita et al., 2015, Micheva et al., 2018b). Their varied morphologies suggest that they consist of multiple stellar populations (Micheva et al., 2018b), which would explain the difficulty of many high redshift studies to constrain properties such as metallicity or SFR with SED fitting.

Interestingly, one of the LARS galaxies has Lyman α in absorption at the center but a Lyman α halo in emission, suggesting that the Lyman α emission is produced in the disc of the galaxy and scattered in the outflowing H I halo of the galaxy instead of through the dusty central part (Duval et al., 2016). At low redshifts, it is also easier to study the effects of the kinematics of the ISM on Lyman α emission, but the results are inconclusive. While it looks like turbulence is useful (Herenz et al., 2016) and outflows are even necessary, they do not seem to be sufficient to produce Lyman α escape (Rivera-Thorsen et al., 2015).

The question of how dust distribution affects Lyman α escape can also be tackled with local LAE analogues. While various dust configurations have been proposed to either enhance or prohibit Lyman α radiative transfer, the reality might be a mix of different scenarios. Bridge et al. (2018) suggest a dusty inner core, a clumpy annulus and an outer layer where Lyman α emission can scatter.

Green Peas

Another type of objects that resemble high redshift LAEs, serendipitously discovered, are the so-called green peas (GPs, Cardamone et al., 2009). Originally discovered by volunteers of the citizen science project 'Galaxy Zoo' (Lintott et al., 2008), which aims at classifying Sloan Digital Sky Survey (SDSS) images of galaxies, these objects are unresolved in the SDSS images and distinctly green, which inspired their name. The colour comes from the strong [O III] $\lambda 5007 \text{ \AA}$ emission line and the galaxies were found to be low in mass, metallicity and dust as well as strongly star-forming ($\sim 10 M_{\odot} \text{ yr}^{-1}$, Cardamone et al., 2009). Upon closer inspection with HST data they appear to be clumpy and merging (Cardamone et al., 2009) and together with their other properties this makes them similar to high redshift LBGs or LAEs.

Due to their lower redshift, the $H\alpha$ line can be observed and compared to the Lyman α emission, which gives an estimate of the Lyman α escape. Henry et al. (2015) find Lyman α escape fractions between 5% – 60% (similar to results from Yang et al., 2016) and 9 of their 10 Green Peas have a double peaked Lyman α line profile. They also find a strong anti-correlation between Lyman α escape fraction and the peak separation, which leads them to conclude that the Lyman α escape is more influenced by the neutral hydrogen column density than outflow velocities (as already mentioned above in section 1.3.3 when discussing the kinematics of the ISM). But Green Peas are interesting not only because of their similarity to LAEs, but also because their defining feature, the strong [O III] emission and high ratios of [O III]/[O II] could be an indicator of Lyman continuum emission (Jaskot & Oey, 2013).

Apart from Green Peas and the LARS galaxies, there are other low-redshift LAE analogues. Among them are the 'Blueberry' galaxies (Yang et al., 2017), which are similar to Green Peas as they have high ionisation ([O III]/[O II] $\sim 10 - 60$) but are tiny ($< 1 \text{ kpc}$). Other galaxies similar to high-redshift LAEs are S II emitters, low metallicity C III emitters and extreme He II emitters (e.g. Berg et al., 2019).

Evolution of Lyman α Emitters over Time

Although Green Peas and other low redshift star-forming galaxies could be the analogues to high-redshift LAEs, it is not necessarily clear that LAEs at e.g. redshift $z \sim 2$ are the descendants of LAEs at redshift $z \sim 3$. Acquaviva et al. (2012) have found considerable differences in the properties of both populations based on SED fitting and stacks: The higher redshift LAEs are older and more metal poor when compared to LAEs at lower redshift.

An important property of the population of LAEs at different times in the universe is their number density per luminosity bin, the luminosity function (LF). This can be derived at different redshifts to compare the distribution of luminosities at different epochs. While the UV luminosity function of LAEs increases

between $z = 3 - 6$, the Lyman α luminosity function does not change much over the same redshift range, which would mean that LAEs are more frequent (compared to the LF of other galaxies, which declines, [Bouwens et al., 2015](#)) and Lyman α emission is escaping more easily in the earlier universe ([Ouchi et al., 2008](#)). The escape fraction went from around $f_{\text{esc,Ly}\alpha} \sim 20\%$ at redshift $z \sim 5$ down to $f_{\text{esc,Ly}\alpha} \sim 5\%$ at redshift $z \sim 2.5$ ([Cassata et al., 2015](#), [Ciardullo et al., 2014](#)).

Highest Redshifts

LAEs are more easily observed at high redshifts than other galaxies (since the strong Lyman α line shifts to the optical), which is why they are some of the objects with the highest redshifts known (see figure 1.2 in section 1.1.1 for redshift records). Some recent high redshift objects include an LBG at $z = 7.73$ with a Lyman α line equivalent width of $\text{EW}_0 = 21 \pm 4 \text{ \AA}$ ([Oesch et al., 2015](#)), a bright galaxy at redshift $z = 8.68$ with $\text{EW}_0 = 17 - 42 \text{ \AA}$ ([Zitrin et al., 2015](#)) and a gravitationally lensed dropout galaxy at $z = 8.38$ ([Laporte et al., 2017](#)). The LAE with the highest detected EW at $z > 7$ was discovered recently by [Larson et al. \(2018\)](#) with $\text{EW}_0 = 140.3 \pm 19 \text{ \AA}$. Although these high redshift galaxies have high SFRs ([Finkelstein et al., 2013](#)), some of them are unexpectedly dusty for such a young age ([Laporte et al., 2017](#), [Oesch et al., 2015](#)).

On the one hand, a redshifted, strong Lyman α line makes it easy to find LAEs at high redshifts, but on the other hand, many studies have shown that the number density of LAEs goes down at redshifts higher than $z \sim 7$ (e.g. [Shibuya et al., 2012](#), [Caruana et al., 2012](#), [Caruana et al., 2014](#)). One explanation would be a dimming of the LAEs. However, this has been examined by [Tilvi et al. \(2014\)](#), who find more evidence for a scenario where there are fewer LAEs instead of fainter LAEs at redshifts of $z \sim 8$, where the Lyman α fraction has dropped by more than a factor of 3 compared to $z \sim 6$. Thus, at such high redshifts the Lyman α emission would be suppressed by the intervening neutral hydrogen in the IGM.

While e.g. [Kashikawa et al. \(2011\)](#) find a stark drop in the normalisation of the Lyman α LF (and none in the UV LF of the same objects) towards higher redshifts, [Ouchi et al. \(2010\)](#) only find a drop of 30% in the Lyman α LF from $z = 5.7$ to $z = 6.6$, leading them to conclude that the main phase of the EoR started at $z > 7$ and the IGM is already largely ionised at $z = 6.6$. A sudden drop in the observed Lyman α emission at redshifts $z \sim 6$ or ~ 7 would require a fast change of the state of the IGM at the EoR, but there are other explanations (see [Dijkstra, 2014](#)): Instead of being an effect of a more neutral IGM at this redshift, it could be more opaque to Lyman α photons due to its increased density at the time. Another explanation might be the increased number density of Lyman limit systems (LLSs), which are absorption systems with a neutral hydrogen density of $N_{\text{H I}} > 10^{17} \text{ cm}^{-2}$ that can self-shield from ionising emission. The change in observed Lyman α emission from $z \sim 6$ to 7 could also be due to the enhanced gas accretion rate ([Papovich et al., 2011](#)) or the gas could have been distributed differently at that time (which would invoke a Neufeld-like scenario at decreasing redshifts, which I have described above is unlikely). Instead of an evolution of the gas distribution, it could also be due to an evolution of the escape fraction of ionising emission. It has been observed that the Lyman continuum escape increases with increasing redshift ([Mitra et al., 2013](#)) along lines with lower neutral hydrogen column densities of $N_{\text{H I}} < 10^{17} \text{ cm}^{-2}$. Lyman α emission could escape through the same paths without being broadened by scattering ([Behrens et al., 2014](#), [Verhamme et al., 2015](#)), only to be absorbed in the ionised gas of the IGM (as the line was not redshifted to begin with through ISM kinematics, see [Dijkstra, 2014](#)).

1.4. Escaping the Circumgalactic Medium

The ionising photons have now taken the first steps on their way out of the galaxy, through the ISM, possibly being transformed to Lyman α photons. The last barrier between them and the vastness of space (the IGM) is the CGM, which surrounds the galaxy. In recent years, the CGM has been the focus of an increasing number of studies (see white paper by [Cicone et al., 2019](#)), as this transition area is the potential stage for feedback between the galaxy and the IGM, with the galaxy enriching the IGM with metals from its star formation processes through supernova outflows, and the IGM in turn providing important fuel for the ongoing star formation, especially at high redshifts. As such, it is a relatively new concept and expression of the difficulties of defining the 'edge' of galaxies. This makes it also difficult to define exactly when photons have 'escaped' the galaxy: When they leave the ISM or the CGM (see [Steidel et al., 2018](#) for a discussion). And what happens with ionising radiation and Lyman α emission in this transitory zone?

Observing the Circumgalactic Medium

Observing the CGM is more difficult at low redshifts than it is at high redshifts. As it consists mostly of neutral hydrogen, it can be observed in 21cm emission, but this is only possible for a few galaxies and lacks high resolution. Another possibility is to observe it in X-rays, but again, the CGM so far has been observed this way only for a few galaxies. Optical emission lines have similar problems. However, as we have seen, Lyman α emission shifts into the optical wavelength range at high redshifts and is produced in hydrogen gas clouds, which makes it an ideal candidate to study the extent, shape and kinematics of the CGM.

The Lyman α photons produced in the star-forming regions of the galaxy are scattered in neutral hydrogen, which is the main component of the CGM. Through radiative transfer processes, their effective pathlengths can be drastically increased, potentially spanning a large fraction of the CGM. Another way that can make the CGM glow in Lyman α is by UV fluorescence, through ionising radiation that is not quite luminous enough to ionise the CGM, but still produces a number of Lyman α photons in the matter surrounding the galaxy. The Lyman α photons can also be produced through cooling radiation, after the hot ($\gg 10^4$ K) gas in the halo cools down rapidly and, being no longer pressure-supported, contracts and releases the binding energy as Lyman α photons in the process (Haiman et al., 2000). Another way to produce the Lyman α photons in the CGM directly is the presence of satellite galaxies, whose star-formation could ionise their neutral hydrogen reservoir.

All those phenomena have the same effect: The CGM glows in a halo of Lyman α light.

1.4.1. Extended Lyman α Emission

Lyman α nebulae around high redshift quasars, fueled by the central AGN, have been observed for a long time. But normal, star-forming galaxies have been shown to have similar extended Lyman α emission, in various shapes and sizes.

Lyman α Blobs

The first discovery of large, extended Lyman α nebulae around two LBGs at redshift $z = 3.09$ in a proto-cluster was by Steidel et al. (2000). These 'blobs' have physical sizes of $> 100 h^{-1}$ kpc and Lyman α line fluxes of $\sim 10^{-15}$ erg s $^{-1}$ cm $^{-2}$, similar to high-redshift radio galaxies, but without the radio emission. Since their total Lyman α luminosities reach 10^{43} erg s $^{-1}$, which is more than $10^{10} L_{\odot}$ (Scarlati et al., 2009), the source of the strong Lyman α line is unclear. Obscured AGN might be at play that photoionise the neutral hydrogen in the CGM of the galaxies (Steidel et al., 2000). It could also be shock ionisation from supernova outflows or cooling radiation (see above, Haiman et al., 2000, Furlanetto et al., 2005).

Lyman α Halos

A more recent discovery is that the CGM of almost all high redshift star-forming galaxies seems to be glowing in Lyman α , albeit fainter than their giant 'blob' counterparts. One of the first discoveries of Lyman α halos around high redshift LBGs was using stacking methods to make the faint halos visible (Steidel et al., 2011), even though only 55% of those LBGs have Lyman α in emission and only 20% have a Lyman α EW $_0$ of $> 20 \text{ \AA}$. The emission in the halo is most likely Lyman α emission originating in the galaxies' ionised hydrogen gas and scattered in the neutral hydrogen of the CGM (Steidel et al., 2011, Matsuda et al., 2012). An interesting conclusion of this observation and a possible answer to the question above about the differences between LBGs and LAEs is: Given a more sensitive instrument, all LBGs would be classified as LAEs and all would have extended Lyman α emission⁷. Evidence for the last claim was shown in Wisotzki et al. (2016) and Wisotzki et al. (2018) for individual LAEs (see figure 1.21⁸), proving that indeed the sky is lit up with Lyman α emission.

The scale lengths of such Lyman α halos are typically 10 times larger than the UV continuum (Steidel et al., 2011, Wisotzki et al., 2016) and five times larger than their low-redshift analogues (found in the LARS sample, Hayes et al., 2013, Guaita et al., 2015). The scale lengths stay constant over a redshift range of

⁷A note on the nomenclature: Originally, the term Lyman α 'blob' was reserved for large, extended Lyman α emission around LBGs or LAEs. Since the discovery that the existence of extended Lyman α emission is ubiquitous among LAEs and possibly also LBGs, the distinction is not clear anymore. Other names have been proposed, like Lyman α halos, or short LAHs, or giant Lyman α nebulae or even Lyman α rings. Here, I make no such distinctions and call extended Lyman α emission a Lyman α halo.

⁸Credit: NASA, ESA, and S. Beckwith (STScI) and the HUDF Team (CC BY 4.0; <https://creativecommons.org/licenses/by/4.0/>; crop, alignment, and application of blur filter by Wisotzki et al., 2018)

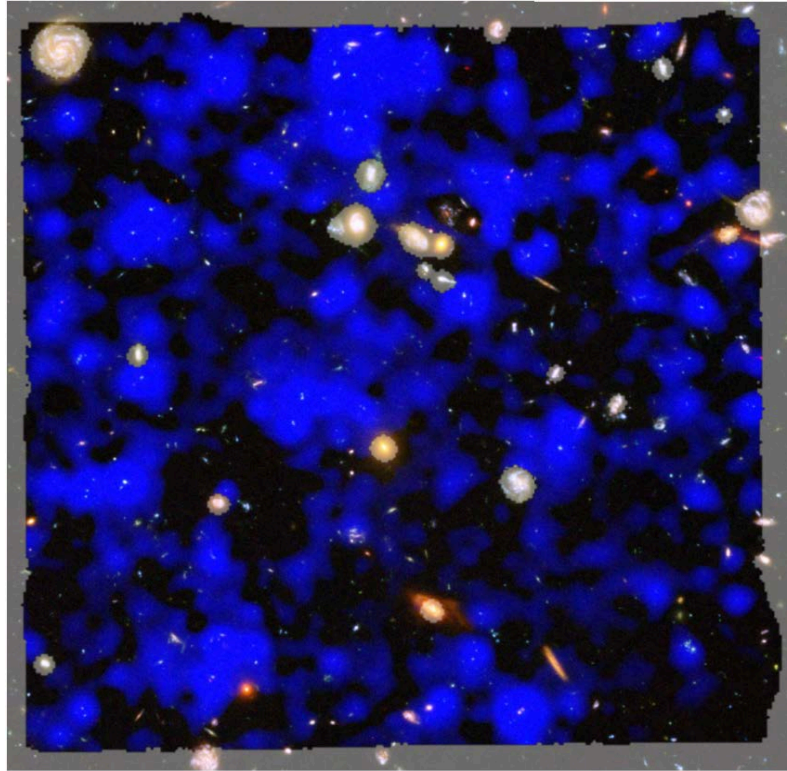


Figure 1.21.: Figure 1 from Wisotzki et al. (2018), showing an HST image of the HUDF superimposed with the Lyman α emission in that field in blue, observed by MUSE. The Lyman α emission is shown for LAEs with redshifts $3 < z < 6$, which is why the brighter galaxies at lower redshifts are not shown with Lyman α halos.

$z \sim 2 - 6$ (Leclercq et al., 2017a) with a possible increase at redshifts beyond $z \sim 6.6$ (Momose et al., 2014). Surprisingly, the scale lengths are anti-correlated with the Lyman α equivalent widths, which means larger Lyman α halos could be present in galaxies that would be classified as LBGs (Momose et al., 2016, Feldmeier et al., 2013) since their Lyman α emission is only visible in the halo. Even for LAEs, the majority of the Lyman α emission ($\sim 65\%$, Leclercq et al., 2017a or even up to $> 90\%$, Wisotzki et al., 2016) comes from the Lyman α halo. Thus it is important to include this extended emission in measurements of the Lyman α equivalent width, to avoid underestimating it.

Not only do the Lyman α halos contain most of the Lyman α emission of the galaxies, but the Lyman α emission line properties even vary across the halo (see Erb et al., 2018, who use integral field spectroscopy to study a $z = 2.3$ LAE). Both the peak ratio and the peak separation between the blue bump and the red main peak show significant variations with the red peak dominating the centre of the LAE (see figure 1.22 from Erb et al., 2018), while the outskirts show more equally distributed peaks. This is a sign of varying properties of the CGM when it comes to neutral hydrogen column density, covering fraction and velocity of outflows (Erb et al., 2018).

1.4.2. Lyman Continuum Leakers

As we have seen above, Lyman α emission can leave its galaxy of origin through outflows in the ISM and CGM, through scattering in the CGM (producing extended halos) or even after being produced in the CGM itself. In this section I want to focus on the escape of the emission that is the basis for the creation of Lyman α photons from star formation, the Lyman continuum emission. Since it is not scattered in neutral hydrogen like Lyman α , but instead destroyed (or converted to Lyman α), it has a harder time leaving its galaxy. Nevertheless, we know from observations of the epoch of reionisation that some of the ionising emission has to escape the ISM and CGM and reach the IGM to ionise it and keep it ionised. Therefore it is important to know where the ionising emission comes from and how it can escape into the IGM.

What makes observations of Lyman continuum emission from high redshift galaxies so difficult is the intervening IGM (see section 1.5 below), since the wavelength of the Lyman continuum (912 \AA) is only

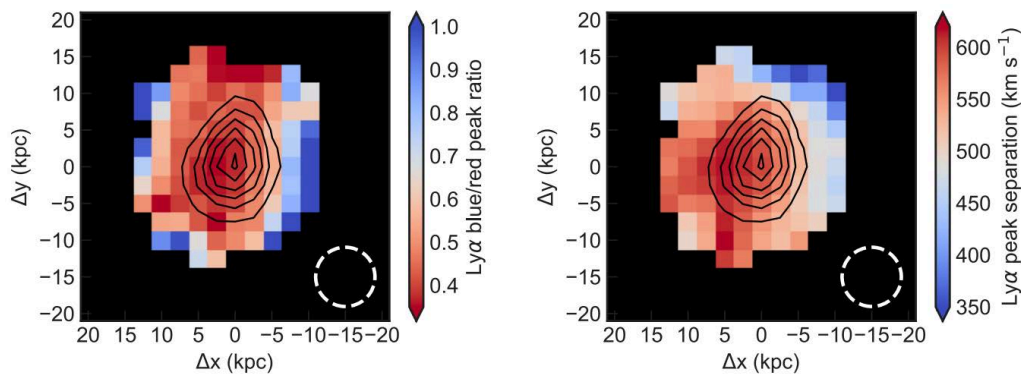


Figure 1.22.: Figure 5 from Erb et al. (2018) illustrating the variations in Lyman α line properties over the extended halo of an LAE at redshift $z = 2.3$. The left panel shows the ratio between the blue and the red part of the Lyman α line, the right panel shows the Lyman α peak separation in km s^{-1} .

shifted to the optical at redshifts where the IGM neutral fraction is already too high (at $z > 4$) to let enough Lyman continuum through. Therefore it has to be observed in UV bands (from space). What is more, if a galaxy is intrinsically bright in Lyman α , this could mean that a significant part of its Lyman continuum emission was transformed already to Lyman α (see figure 1.9 above by Partridge & Peebles, 1967). Then again, Lyman continuum emission could escape along the same channels as Lyman α emission, which would connect the escape fractions of both. These questions have been addressed in recent studies, as the first Lyman continuum leakers have been discovered in recent years. An overview of currently known Lyman continuum leakers at high redshifts and in the low redshift universe is given in tables 1.1 and 1.2 respectively.

Observations

After having observed Lyman continuum radiation from QSOs for some time, the first direct detection of Lyman continuum leakage from star-forming galaxies at high redshift was by Steidel et al. (2001). They stacked spectra of 29 LBGs at $z = 3.4$ and found an escape fraction of $f_{\text{esc}} \gtrsim 50$ (see below for a definition of the Lyman continuum escape fraction). They claim that if this value is typical for LBGs, they could produce 5 times more Lyman continuum than QSOs at that redshift (per unit comoving volume). As can be seen in table 1.1, later studies have found quite different (and usually lower) escape fractions at the same redshifts and the constraints vary widely.

When it comes to observing Lyman continuum from individual objects, the observations are even more challenging, as the Lyman continuum emission is already intrinsically fainter than the rest-frame UV continuum and can easily be attenuated by the intervening IGM. Therefore, most individual Lyman continuum leakers have been observed at lower redshifts (see table 1.2), with the first detection of Lyman continuum emission by Bergvall et al. (2006) in Haro 11 at $z = 0.021$ (and confirmed by Leitert et al., 2011). This object is controversial, though, as it is clumpy and has three distinct knots, which were analysed in detail by Keenan et al. (2017) who find no evidence for Lyman continuum emission from two of the knots, but instead find ionising emission at a distance of 1 kpc of one of them. If this is indeed the Lyman continuum source, they conclude that Lyman continuum escape might be independent of Lyman α escape. Another interesting example at low redshift is Tololo 1247-232 (e.g. Leitert et al., 2013, Leitherer et al., 2016), which seems to consist of stellar populations with two different ages, one older component of 12 Myrs which might have cleared the path for Lyman continuum escape and one younger component of 2 – 4 Myrs which could provide the bulk of the ionising emission (Micheva et al., 2018a). Recently, even the Antennae galaxies have been shown to be leaking in Lyman continuum (Weilbacher et al., 2018).

Other studies (Izotov et al., 2016a, Izotov et al., 2016b, Izotov et al., 2018a, Izotov et al., 2018b) targeted Green Peas (see section 1.3.4 above), due to their similarity to high-redshift star-forming galaxies. The proximity of these objects makes it easier to reliably classify them and avoid interlopers. At higher redshifts, even fewer objects have been found (e.g. Vanzella et al., 2016, de Barros et al., 2016, Shapley et al., 2016, Vanzella et al., 2018, Bian et al., 2017) and only one individual Lyman continuum leaker with a high escape fraction of $f_{\text{esc}}^{\text{LyC}} > 50\%$ is known thus far, Ion2 (Vanzella et al., 2015, de Barros et al., 2016).

object(s)	paper(s)	redshift	$f_{\text{esc,LyC}}$	notes
A2218-Flanking Ion2	Bian et al. (2017)	$z = 2.5$	$f_{\text{esc,abs}} > 28 - 57\%$	highest redshift LyC leaker candidate so far
Q1549-C25 1/102 LBGs	Vanzella et al. (2015), de Barros et al. (2016) Shapley et al. (2016) Vanzella et al. (2010a)	$z = 3.212$ $z = 3.212$ $z = 3.795$	$f_{\text{esc,rel}} = 64_{-0.1}^{+1.1}\%$ $f_{\text{esc,rel}} > 51\%$ $f_{\text{esc,abs}} > 15\%$	
29 LBGs	Steidel et al. (2001)	$z = 3.4$	$f_{\text{esc}} \gtrsim 50$	
7/73 LBGs, 10/125 LAEs 6/26 LBGs, 28/130 LAEs	Iwata et al. (2009) Nestor et al. (2011)	$z \approx 3.1$ $z \sim 3$	$> 4\%$ $\sim 10\%$ for LBGs	
4/49 LBGs, 7/91 LAEs 102 LBGs	Mostardi et al. (2013) Vanzella et al. (2010a)	$z \sim 2.85$ $z = 3.4 - 4.5$	$f_{\text{esc,rel}}^{\text{LBG}} = 5 - 8\%$, $f_{\text{esc,rel}}^{\text{LAE}} = 18 - 49\%$ $f_{\text{esc,abs}}^{\text{stack}} < 5 - 20\%$	low UV/LyC ratios of LAEs explained with orientation effects
11 LBGs 6 galaxies	Boutsia et al. (2011) Naidu et al. (2017)	$z \sim 3.3$ $z \sim 2$	$f_{\text{esc,rel}} < 5\%$ $f_{\text{esc}} > 60\%$	they also find one individual candidate smallest limit at that z

Table 1.1.: Overview of some of the currently known LyC emission at high redshift. The subscript rel. means I give the relative escape fraction, abs. means absolute and if the information is not given in the subscript it was not clear from the paper. As can be seen from the table, results on escape fractions vary, even for similar objects and selection criteria. It has to be noted though that the definition of escape fraction also is not uniform, which explains some of the discrepancies.

Contamination

One of the first claims of direct Lyman continuum emission detection from a high redshift star-forming galaxy at $z \sim 3$ was by Shapley et al. (2006) in spectra of 2 out of 14 of observed LBGs with a relative Lyman continuum escape fraction of $f_{\text{esc,rel}} = 14\%$. However, of those two detections, one has proven to be spurious (in the case of SSA22a-D3, see Nestor et al., 2011 and Iwata et al., 2009) and the other a foreground contamination (Nestor et al., 2013). The threat of foreground contaminants imposing as Lyman continuum emitters is ubiquitous, as e.g. Nestor et al. (2011) have demonstrated using Monte Carlo simulations to show that a fourth of their LAEs with potential Lyman continuum emission are possibly contaminated by foreground objects and 2 – 4 of their 6 LBGs. High resolution (usually space-based) observations are necessary to match the tentative Lyman continuum emission correctly to the UV continuum found in a different band or spectrum, as Mostardi et al. (2015) have shown, who found only one real Lyman continuum leaker in a sample of 16 candidates when using follow-up HST observations. Vanzella et al. (2010b) performed a detailed analysis of the possibility of foreground contamination, assuming ground-based observations that can only resolve objects at a distance of $0.5''$ or more. For galaxies at $z > 3$, they find foreground contamination in 2.1 – 3.2% and even a probability of 50% that at least one third of literature Lyman continuum leakers are affected by foreground contamination.

Especially for LAEs, a way to mitigate this problem is to obtain spectra to verify the redshift via the Lyman α emission line, although there could still be a chance alignment with a foreground object (see chapter 5 in this thesis).

Escape Fraction

In tables 1.1 and 1.2 above I have given escape fractions of Lyman continuum photons and now want to disentangle the different definitions and understand what the Lyman continuum escape fraction is used for. In its most basic interpretation, an escape fraction f_{esc} gives the fraction of photons (in this case the Lyman continuum) that escape from the galaxy compared to how many such photons were produced intrinsically

object(s)	paper(s)	distance	$f_{\text{esc,LyC}}$	notes
Haro 11	Leitet et al. (2011)	$z = 0.021$	$f_{\text{esc}} = 16.6^{+7.4}_{-6.5}\%$	First known low- z LyC leaking candidate. The relative escape fraction was taken from Leitert et al. (2013).
J1154+2443	Izotov et al. (2018a)	$z = 0.369$	$f_{\text{esc}} = 46\%$	Highest known LyC leakage in low- z galaxies
J0921+4509 Tol 1247-232	Borthakur et al. (2014) Leitet et al. (2013), Leitherer et al. (2016), Micheva et al. (2018a)	$z = 0.235$ 207 Mpc	$f_{\text{esc}} = 21\% \pm 5\%$ $f_{\text{esc,rel}} = 7.4^{+7.7}_{-6.7}\%$, $f_{\text{esc,rel}} = 21.6 \pm 5.9\%$	Lyman Break Analogue.
Mrk 54	Leitherer et al. (2016)	191 Mpc	$f_{\text{esc,rel}} = 20.8 \pm 6.1\%$	Chisholm et al. (2017) did not detect any LyC emission: the candidate is likely contaminated.
Tol 0440-381	Leitherer et al. (2016)	167 Mpc	$f_{\text{esc,rel}} = 59.8 \pm 13\%$	Might be contaminated by geocoronal lines.
5 objects	Izotov et al. (2016b)	$z \sim 0.3$	$f_{\text{esc,rel}} = 9 - 33\%$	high [O III] / [O II] ratio
6 objects	Izotov et al. (2018b)	$z = 0.3 - 0.43$	$f_{\text{esc}} = 2 - 72\%$	

Table 1.2.: Overview of currently known LyC leakers at low redshift with their escape fractions. The same subscript rules apply as for table 1.1. Note that Chisholm et al. (2017) found a lower LyC escape fraction in Tol 1247-232 and Tol 0440-381 than Leitherer et al. (2016) and no LyC emission from Mrk 54 at all, but they confirm J0921+4509 from Borthakur et al. (2014) as a leaker.

in the galaxy. We could write:

$$f_{\text{esc}} = \frac{f_{\text{LyC,out}}}{f_{\text{LyC,int}}} \quad (1.8)$$

Here, $f_{\text{LyC,out}}$ is the Lyman continuum flux coming out of the galaxy and $f_{\text{LyC,int}}$ is the intrinsic flux at a specific wavelength, which is usually set to 900 Å for Lyman continuum, just below the Lyman break. The problem with this form of the escape fraction, usually called the 'absolute' escape fraction, is that we do not know the intrinsic flux of Lyman continuum photons, as it is attenuated by dust (with the dust absorption coefficient A_{LyC} , which is uncertain) and absorbed in the neutral hydrogen of the ISM of the galaxy (with the optical depth $\tau_{\text{LyC,H I}}$, which is not known). In general, for any observed emission from the galaxy, we can write:

$$f_{\text{out}} = f_{\text{int}} e^{-\tau_{\text{LyC,H I}}} \times 10^{-0.4A_{\text{LyC}}} \quad (1.9)$$

Another way to write the absolute escape fraction would thus be:

$$f_{\text{esc,abs}} = e^{-\tau_{\text{LyC,H I}}} \times 10^{-0.4A_{\text{LyC}}} \quad (1.10)$$

In semi-analytic or purely theoretical approaches, the intrinsic Lyman continuum emission can be obtained from the star-formation rate, but is also relying on the assumed metallicity and IMF. In observations, we could obtain estimates of the properties of the stellar population from SED fitting (if available), which would give us the necessary assumptions on e.g. metallicity, IMF, dust extinction and population age, but relies on high quality data in the rest-UV of the galaxies. At lower redshifts, the intrinsic Lyman continuum emission can be inferred from the $\text{H}\alpha$ line flux, but this has to be corrected for dust reddening by the Balmer decrement, which is not always measurable and shifts to the near-IR at higher redshifts (see Siana et al., 2007 for a discussion).

Another way to estimate the intrinsic Lyman continuum emission is through its relation with the intrinsic UV continuum, $f_{\text{LyC,int}} = (f_{\text{LyC,int}}/f_{\text{UVc,int}}) f_{\text{UVc,int}}$. The UV continuum is a value easily measurable at high redshifts and has been measured for luminosity functions for LAEs and LBGs anyway. This makes

it even possible to predict the total available ionising emission from such galaxies directly from their UV continuum. However, we need to apply equation 1.9 again to express the intrinsic UV continuum as the observed UV continuum attenuated by the neutral hydrogen in the ISM and reddened by dust. We get:

$$f_{\text{LyC,int}} = (f_{\text{LyC,int}}/f_{\text{UVC,int}}) f_{\text{UVC,out}} e^{\tau_{\text{UVC,H I}}} \times 10^{0.4A_{\text{UVC}}} \quad (1.11)$$

If we pluck this into equation 1.8 above, we get the estimate of the absolute Lyman continuum escape fraction which is only dependent on the observed Lyman continuum and UV continuum, the intrinsic ratio between Lyman continuum and UV continuum and the dust attenuation of the UV continuum:

$$f_{\text{esc,abs}} = \frac{f_{\text{LyC,out}}}{(f_{\text{LyC,int}}/f_{\text{UVC,int}}) f_{\text{UVC,int}}} = \frac{f_{\text{LyC,out}}}{(f_{\text{LyC,int}}/f_{\text{UVC,int}}) f_{\text{UVC,out}} e^{\tau_{\text{UVC,H I}}} \times 10^{0.4A_{\text{UVC}}}} \quad (1.12)$$

We can rewrite this to the usual form of the escape fraction (assuming the UV continuum is not attenuated in the neutral hydrogen in the ISM):

$$f_{\text{esc,abs}} = \frac{f_{\text{LyC,out}}/f_{\text{UVC,out}}}{L_{\text{LyC,int}}/L_{\text{UVC,int}}} \times 10^{-0.4A_{\text{UVC}}} \quad (1.13)$$

This definition of the escape fraction is still dependent on knowledge of the dust attenuation in the UV continuum $A_{\text{UVC}} = k_{\text{UVC}} E(B - V)$ (e.g. with $k_{\text{UVC}} = 10.33$ for the Calzetti law, Calzetti et al., 1994). To avoid assumptions on the dust reddening, many studies use the 'relative' escape fraction, first introduced by Steidel et al. (2001) (but see also e.g. Steidel et al., 2018, Siana et al., 2007, Grazian et al., 2018, Bian et al., 2017), which is related to the absolute escape fraction by the dust:

$$f_{\text{esc,abs}} = f_{\text{esc,rel}} \times 10^{-0.4A_{\text{UVC}}} \quad (1.14)$$

$$f_{\text{esc,rel}} = \frac{f_{\text{LyC,out}}/f_{\text{UVC,out}}}{L_{\text{LyC,int}}/L_{\text{UVC,int}}} \quad (1.15)$$

The relative escape fraction and absolute escape fraction are the same (see Shapley et al., 2016) if the dust content in the galaxy is negligible, which is often a good assumption for high redshift LAEs. This overestimates the escape fraction or in other words, the relative escape fraction is always higher than the absolute escape fraction and can get over 100%. What is left unknown is the intrinsic Lyman continuum and intrinsic UV continuum $L_{\text{LyC,int}}/L_{\text{UVC,int}}$, which relies on assumptions on the stellar population (see section 6.3 in chapter 6 where I use different stellar population synthesis models to predict this fraction). Different values have been assumed in the literature and it is important to note if the intrinsic ratio is given in wavelength or in frequency (which is not the same due to $f_{\lambda} = f_{\nu} c/\lambda^2$). Values for $L_{\text{LyC,int}}/L_{\text{UVC,int}}$ in frequency range from 3 (see e.g. Steidel et al., 2001, Grazian et al., 2016, Marchi et al., 2018, Grazian et al., 2017, Japelj et al., 2017) to 5 (see e.g. Naidu et al., 2018) and even as high as 6 – 8 (see e.g. Siana et al., 2007), depending on what the models assume for star-formation history, metallicity, IMF and age.

To calculate the escape fraction, the Lyman continuum is usually measured at the wavelength 900 Å and the UV continuum at 1500 Å. If photometry is used, the continua cannot be measured at a fixed wavelength, but are rather measured over a wavelength range. Since the filter width is fixed and the redshifts of the objects vary, the rest-frame wavelength probed in this way changes depending on the redshifts of the objects. The measured flux can be assumed to be measured at the shifted effective wavelength of the band and from that the flux at either 900 Å or 1500 Å can be inferred, if the continuum slope is known (see the description of the β -slope for the UV continuum in section 4.1.2). The β -slope is connected to the dust attenuation as well and the empirical relation between the two is given by $A_{1600} = 4.43 + 1.99 \times \beta$ (Meurer et al., 1999 and also used e.g. by Bouwens et al., 2009 for a sample of high redshift LBGs). The UV slope β in this equation is measured between 1300 – 2600 Å.

It is unclear whether this relation holds for redshifts above $z > 5$, as the mechanisms producing dust change at earlier times in the universe. The asymptotic giant branch (AGB) stars in whose envelopes the dust forms are not around yet at such early times and it has been suggested that the dust could instead originate from supernovae explosions (e.g. Maiolino et al., 2004 for AGN). Galaxies with extreme properties also show a different relation between dust attenuation and β -slope. Nevertheless, the relation can be used as an approximation for the dust reddening needed for the absolute escape fraction in equation 1.13.

If we leave the dust reddening out of the equation, the question still remains how the Lyman continuum emission can escape the neutral hydrogen in the ISM. Similar mechanisms as for the escape of Lyman α

photons (see section 1.3.3 in radiative transfer) could be at work, since the Lyman α photons are scattered in the neutral hydrogen, while the Lyman continuum emission is absorbed. A low neutral hydrogen column density could facilitate the escape, as Lyman continuum photons could carve out ionised paths in the ISM. Another option are supernova induced outflows and holes. Often, theoretical models assume a uniform ISM with a single star-formation region in the center with only one supernova explosion, which highly underestimates the importance of supernovae increasing the porosity of the ISM (e.g. Clarke & Oey, 2002). If multiple supernovae work together to form supershells, they can clear paths through the ISM through which the ionising emission can escape (Fujita et al., 2003).

One important component is still missing from the considerations of the (relative) Lyman continuum escape fraction above: The IGM absorption of the Lyman continuum photons. The Lyman continuum flux that is observed is unfortunately not the same as that which escaped the galaxy, as the intervening neutral hydrogen in the IGM absorbs it. This can already be seen in table 1.1, where the highest observed Lyman continuum leaker is at redshift of $z \sim 3.8$ (Vanzella et al., 2010a). Therefore, we need to correct the observed Lyman continuum emission for the IGM transmission $T_{\text{IGM}} = e^{-\tau_{\text{LyC}}}$ where τ_{LyC} is the optical depth:

$$f_{\text{esc,rel}} = \frac{f_{\text{LyC,out}}/f_{\text{UVC,out}}}{L_{\text{LyC,int}}/L_{\text{UVC,int}}} e^{\tau_{\text{LyC}}} \quad (1.16)$$

The IGM transmission depends on the redshift of the objects and cannot be measured for individual objects. Often a mean or median value for a specific redshift is used (e.g. Grazian et al., 2016), but this ignores the fact that the IGM transmission also depends on the line-of-sight, since the distribution of neutral hydrogen is not perfectly homogeneous. I discuss the IGM transmission in section 1.5.

Connecting Properties to Leakage

As Lyman continuum emission is notoriously hard to observe at high redshifts due to the increasing influence of the IGM and at low redshifts can only be accessed from space with UV photometry, there have been attempts to predict the Lyman continuum escape from galaxy properties that are more accessible. This way, the most promising candidates can be targeted directly and the ionising UV background of the IGM can be estimated.

[O III]/[O II]

Such Lyman continuum proxies are usually concerned with the relative strengths of other emission lines in the spectrum or with properties of the Lyman α emission line. One of the most promising proxies is the ratio of the [O III] line(s) at 5007 and 4959 Å to [O II] at 3727 Å (proposed by Jaskot & Oey, 2013, but see also e.g. Nakajima & Ouchi, 2014, Nakajima et al., 2016, Faisst, 2016). This ratio can indicate a strong ionisation field and low optical depths and an overview of different studies can be seen in the left panel of figure 1.23 by Izotov et al. (2018a). Since the strong [O III] line is the reason for their green colour in SDSS images, Green Peas have been targeted for studies of the [O III]/[O II] ratio (Jaskot & Oey, 2013, Paalvast et al., 2018) and have been found to be good Lyman continuum emitter candidates (e.g. Izotov et al., 2016a, Izotov et al., 2016b). However, a high [O III]/[O II] ratio does not guarantee the escape of Lyman continuum photons (as Naidu et al., 2018 and Stasińska et al., 2015 have shown) and there is no clear correlation between the Lyman continuum escape fraction and the [O III]/[O II] (Bassett et al., 2019).

Other proxies based on lines or line ratios

Some other proposed indicators of Lyman continuum leakage based on emission lines, ratios or other spectral features:

- the ratio [Ne III] λ 3869/[O II] λ 3727 (Levesque & Richardson, 2014) indicates the ionisation parameter and could be an alternative for [O III]/[O II] as it is less influenced by dust reddening
- Mg II $\lambda\lambda$ 2796, 2803 (Henry et al., 2018) correlates strongly with Lyman α emission and could be used as a substitute if Lyman α is not observable
- a small [S II]/H α ratio and the residual intensity of the core of the absorption line Si II (Alexandroff et al., 2015) also indicates Lyman continuum leakage
- the ratio SFR_{UV}/SFR_{H α} (Alexandroff et al., 2015) is promising as well, but it does not seem to correlate with other indicators, though
- a high star formation surface density (e.g. Clarke & Oey, 2002, Verhamme et al., 2017, Verhamme et al., 2017) can produce high Lyman continuum emission

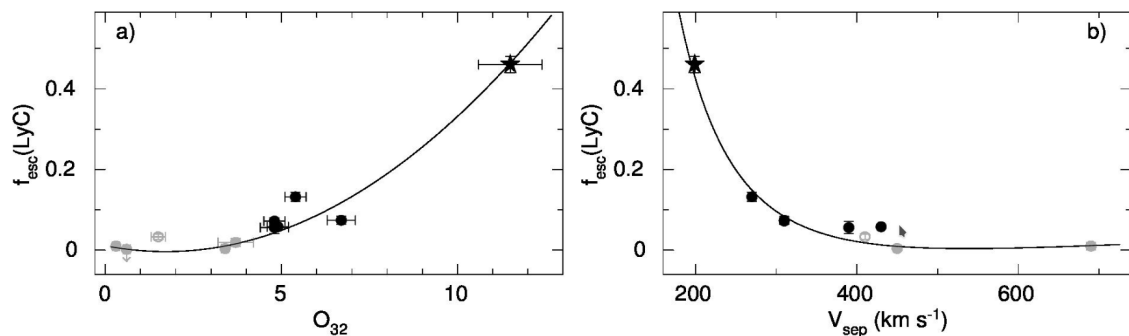


Figure 1.23.: Figure 6 from Izotov et al. (2018a) showing the Lyman continuum escape fraction $f_{\text{esc}}(\text{LyC})$ over the [O III] to [O II] ratio (left, panel a) and the Lyman α peak separation V_{sep} in km s^{-1} (right, panel b). The different symbols indicate various low redshift galaxies observed with FUSE (open circle) and HST/COS (filled symbols), from Izotov et al. (2016a) and Izotov et al. (2016b) (black filled circles), Borthakur et al. (2014), Leitherer et al. (2016), Chisholm et al. (2017) (gray filled circles), Leitert et al. (2013) (open circle) and Izotov et al. (2018a) (filled star). Except for the filled star, all peak separations are taken from Verhamme et al. (2017). The solid lines show the most likelihood regressions and indicate a correlation between Lyman continuum escape fraction and [O III]/[O II] ratio as well as an anti-correlation between Lyman continuum escape fraction and Lyman α peak separation.

- the $\text{EW}_{\text{C III}} \lambda\lambda 1907, 1909$ (Jaskot & Ravindranath, 2016) traces low metallicities and density-bounded regions
- a residual flux in saturated low-ionisation interstellar absorption lines (e.g. Borthakur et al., 2014) can indicate Lyman continuum emission
- a β -slope and $\text{EW}_{\text{H}\beta}$ (Zackrisson et al., 2013) can predict Lyman continuum emission of galaxies at very high redshifts ($z > 9$)

Lyman α line properties

Most other Lyman continuum emission proxies focus on the Lyman α emission. The two are correlated due to their similar origin but also because they are both prone to being destroyed in high neutral hydrogen column densities. Lyman α emission gets scattered in neutral hydrogen, especially in the halos of LAEs. This could mean that small Lyman α halos could facilitate the escape of Lyman continuum (Marchi et al., 2018 find a correlation between UV compactness and Lyman continuum leakage). Since most LAEs at high redshift have halos (Wisotzki et al., 2016, Leclercq et al., 2017b, Saust et al. in prep., Wisotzki et al., 2018, see section 1.4.1), their sizes could correlate with Lyman continuum escape.

Lyman α line shape properties that indicate a high Lyman α production in the galaxy are also candidates for predicting Lyman continuum leakage. A small velocity offset of the Lyman α line compared to the systemic redshift, a small peak separation, a large equivalent width and a narrow line could all be useful indicators (Verhamme et al., 2015) and have already been proven to work (Verhamme et al., 2017, Marchi et al., 2018, Izotov et al., 2018b, Dijkstra, 2014).

All these indicators can be used to select the most promising Lyman continuum leaker candidates or to estimate the ionising emission at high redshifts where the direct detection of Lyman continuum is not possible anymore, as the transmission in the IGM becomes too low towards the epoch of reionisation.

1.5. The Path through the Intergalactic Medium

The last part of the journey from the star-forming regions of high redshift galaxies is the path through the IGM. In this section I discuss what obstacles lie ahead and how Lyman α and Lyman continuum emission can overcome them. The most important aspect of the IGM for Lyman α and continuum photons is the amount of neutral hydrogen, as was the case for the ISM and CGM as well. The neutral hydrogen fraction in the IGM varies with redshift, though, as the universe was fully neutral before the epoch of reionisation, then became ionised again but retained neutral parts.

Lyman α emission is scattered in neutral hydrogen, causing Lyman α absorption signatures and the decline in observed LAEs traces the neutral hydrogen content in the IGM and the onset of the epoch of reionisation.

If Lyman α photons manage to avoid neutral hydrogen long enough, they are shifted out of resonance and can escape, which is not the case for Lyman continuum.

1.5.1. Blocking the Way – Lyman α Absorbers

There are different types of intervening neutral hydrogen components of the IGM in the line-of-sight towards an object. They are called the Lyman α forest, Lyman limit systems (LLSs) and damped Lyman α systems (DLAs) but only differ in the neutral hydrogen column density.

Lyman α forest

Absorption lines in bright quasar spectra caused by intervening neutral hydrogen form the Lyman α forest (see McQuinn, 2016 and Meiksin, 2009 for a review). It consists of residuals from the cosmic structure formation (the 'trees') such as highly ionised filaments, halos or sheets. The neutral hydrogen column density in the Lyman α forest is low, by definition under $N < 10^{17} \text{ cm}^{-2}$. When the light from the quasar (or any bright source at high redshift) is obstructed by this forest, the part of the spectrum whose redshifted wavelength is the rest-frame Lyman α wavelength of the intervening neutral hydrogen clouds gets absorbed. The Lyman α forest can probe a region of $\sim 1 \text{ Gpc}$ adjacent to the bright background source, as the full Lyman continuum at wavelengths $\lambda < 912 \text{ \AA}$ gets absorbed by neutral hydrogen. There are also Lyman β , Lyman γ , ... forests, but since they all end at the Lyman limit, they become increasingly short. Thus the Lyman α forest corresponds to the low-density part of the IGM, with neutral hydrogen column density of $N < 10^{17} \text{ cm}^{-2}$.

Lyman Limit Systems (LLSs)

LLSs form the intermediate regime of the neutral hydrogen column density at $N = 10^{17} - 10^{19} \text{ cm}^{-2}$ with 'super' Lyman limit systems at $N = 10^{19} - 10^{20.3} \text{ cm}^{-2}$ (e.g. Fumagalli et al., 2016). Still, they consist of highly ionised clouds with neutral fractions of a few percent. If we combine this with their high neutral hydrogen column density, the LLSs trace dense parts of the IGM or even galaxy halos as well as inflows and outflows. Thus they might be the link between the CGM and the IGM. The imprint they leave on bright background spectra is a drop at the Lyman limit, at 912 \AA , which means any Lyman continuum emission is immediately absorbed by a Lyman limit system. They are more common than damped Lyman α systems, but less so than the Lyman α forest (e.g. Prochaska et al., 2010). At e.g. $z \sim 4.4$ they have an incidence rate of 2.6 ± 0.4 (Crighton et al., 2019). They seem to influence the reionisation process by halting (although not stopping) it by as much as $\Delta z \sim 0.8$ (Shukla et al., 2016) through slowing down the growth of ionised bubbles in the IGM.

Damped Lyman α systems (DLAs)

DLAs are dominated by neutral hydrogen with large neutral hydrogen column densities in excess of $N > 10^{20.3} \text{ cm}^{-2}$. They are called damped Lyman α systems (e.g. Prochaska & Wolfe, 1999, Wolfe et al., 2005) due to the damping wings that appear in the absorption lines in the quasar spectra where they are typically observed. They might contribute the star formation material in the time after the epoch of reionisation, but they are rare. An example of a DLA and an LLS in a quasar absorption spectra (with some other Lyman forest lines) can be seen in figure 1.24.

1.5.2. The Epoch of Reionisation

The epoch of reionisation, when the universe went from a mostly neutral to its present-day ionised state, is the last major change in the history of the universe, when it went from a radiation-bounded to a density-bounded state. The causes, time-frame and progress of this transformation are a topic of many current research studies, from the low-redshift regime concerned with analogues of high-redshift galaxies, to the intricacies of modern cosmology. What provides the photons for the metagalactic ionising background we still see today? Was the process of reionising the universe a fast, short event or more gradual? Was it uniform or did it happen in patches that grew and eventually overlapped? Here I want to provide a brief overview of the different observational techniques of the epoch of reionisation and the most likely candidates for its cause, with a focus on LAEs, what we can learn from them about the early universe and their possible contribution to the hydrogen ionising background. A summary of results from other means of observing the EoR can be found in table 1.3. From these studies, the EoR can be dated to a time of $z \sim 7$ with an unclear starting point and a more secure end at $z \sim 6$.

Method	Results
<p>Thomson scattering: elastic scattering of electromagnetic radiation by a free charged particle (in this case electrons). Planck Collaboration et al. (2016) measure the Thomson scattering optical depth τ through the correlation between temperature and polarization anisotropies of the cosmic microwave background (see Reichardt, 2016, Kogut et al., 2003). The optical depth τ parameterises the total column density of free electrons (in each line-of-sight).</p>	<p>The measured optical depth is $\tau = 0.058 \pm 0.012$ (Planck Collaboration et al., 2016) and best fits with an instantaneous reionisation with a duration of $\Delta z < 2.8$ and an average redshift of $z = 7.8 - 8.8$. At redshift $z \sim 10$ the universe is less than 10% ionised.</p>
<p>21 cm line: spin-flip transition of neutral hydrogen. Tomography of the 21 cm signal can map the distribution of neutral hydrogen in the universe at different redshifts, even before the EoR (Furlanetto & Briggs, 2004).</p>	<p>Stars produced Lyman α photons 1.8×10^8 years after the Big Bang Bowman et al. (2018). Interestingly, Giri et al. (2019) find large neutral island even at late stages of the reionisation (where the Universe was 90 – 95% reionised).</p>
<p>Kinematic Sunyaev Zel'dovich effect: Caused by a Doppler shift of scattered photons due to the bulk velocity of free electrons which move relative to the CMB (Sunyaev & Zeldovich, 1972, Reichardt, 2016). The resulting temperature shift scales with the free electron density.</p>	<p>The kinematic Sunyaev Zel'dovich effect can constrain the duration of the EoR to $\Delta z < 5.4$ (George et al., 2015).</p>
<p>Gunn-Peterson trough: Absorption trough in the spectra of high redshift quasars bluewards of Lyman α. Using this absorption signal to infer the neutral hydrogen density in the IGM was first proposed by Gunn & Peterson (1965).</p>	<p>Complete absorption troughs are seen at $z > 6$, with a neutral hydrogen fraction of 1 – 4% at redshift $z \sim 6.2$, indicating the end of the epoch of reionisation (Fan et al., 2006b, Fan et al., 2006a). Observations of damping wings in Gunn-Peterson troughs point towards a patchy or incomplete reionisation at $z \sim 6.2$, though (Schroeder et al., 2013).</p>
<p>Lyman α luminosity function: The luminosity function at different redshifts is a measure of the number of galaxies per luminosity bin.</p>	<p>Since the Lyman α line is scattered in neutral hydrogen, its luminosity function goes down at the epoch of reionisation at $z \sim 6$ (e.g. Caruana et al., 2012, Ouchi et al., 2008, Kashikawa et al., 2011).</p>

Table 1.3.: Overview of several means to study the epoch of reionisation, its nature, time and duration.

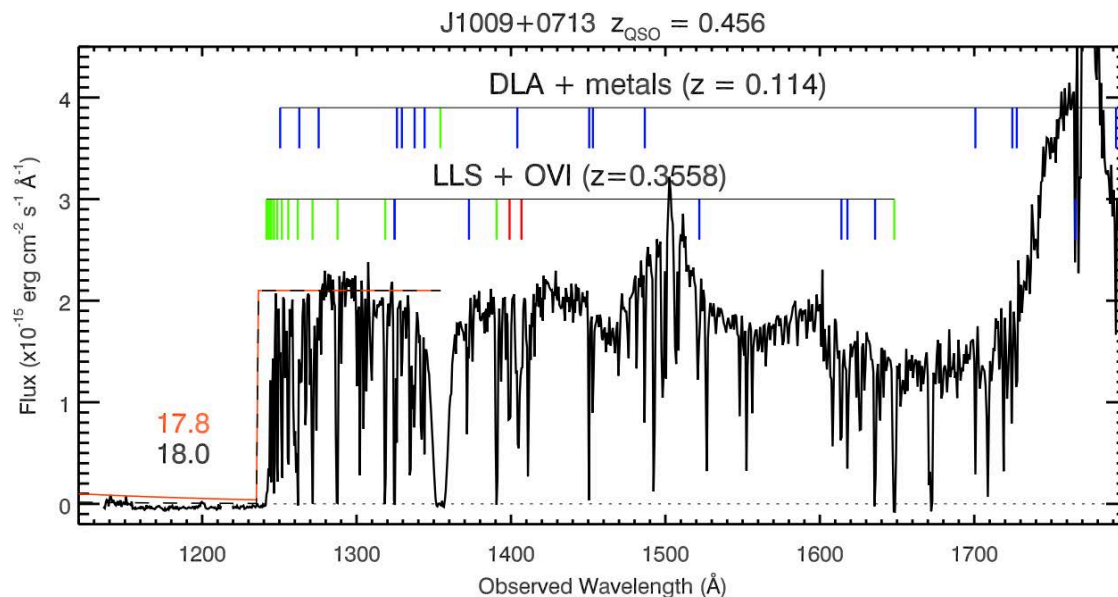


Figure 1.24.: Figure 1 from [Tumlinson et al. \(2011\)](#) showing the spectrum of a QSO at redshift $z = 0.456$. There is an LLS at redshift $z = 0.3558$ and a DLA at redshift $z = 0.114$ and their respective Lyman series lines (green), metal lines (blue) and O IV (red) are marked above. The steep drop of the spectrum at $\sim 1240 \text{ \AA}$ is caused by the LLS. The solid and dashed lines show the LLS opacity at $N_{\text{H I}} = 10^{17.8}$ and 10^{18} cm^{-2} .

Star-forming Galaxies vs. AGN

From the observations mentioned above we know approximately when the epoch of reionisation happened and also that the process was most likely not homogeneous but patchy. However, what remains a topic of debate is what type of galaxies provided the ionising emission. It could be either young, massive stars, releasing the necessary radiation through nuclear binding energy or quasars powered by black holes and thus by gravitational binding energy. The question is which of the two populations is numerous enough and has the necessary Lyman continuum escape fractions.

AGN

The case for AGN as the prime suppliers of ionising emission for the metagalactic ionising background is clear: They have a high escape fraction of ionising emission, usually assumed to be of order unity (although [Micheva et al., 2017a](#) have found the escape fraction to be a factor two or more smaller). In general however, AGN are not numerous enough at high redshift (e.g. [Masters et al., 2012](#), [Onoue et al., 2017](#)) to play the central part in the reionisation process, their number density reaching its maximum at $z \sim 2.5$ ([Fan et al., 2006a](#)). Therefore, their time to shine could be at intermediate redshifts, when their number density is higher. Nevertheless, many studies assume a large number of faint AGN at high redshifts that contribute to the ionising emission, based on the steep bright end of their luminosity function. [Giallongo et al. \(2015\)](#) have found a sufficient number of faint AGN at $z = 4 - 6$ in the CANDELS GOODS-South field by constructing a luminosity function. They claim that although their results hinge on uncertainties in the redshift determination and ionising photon escape fraction, AGN could indeed provide the ionising radiation needed at the EoR and even at lower redshifts. Their results have been put into question by [Parsa et al. \(2018\)](#), who find significantly fewer AGN after reanalysing the data. Yet, [Grazian et al. \(2018\)](#) measure Lyman continuum escape fractions of faint ($-25.1 < M_{1450} < -23.3$) AGN at $z \sim 4$ and find a mean value of 74%. Depending on which luminosity function is used, AGN could thus contribute the bulk of the ionising emission at that redshift ([Madau & Haardt, 2015](#)), but the case is not yet clear.

Star-forming Galaxies

An alternative or addition to AGN as sources of ionising emission are star-forming galaxies. They are numerous enough at high redshift, but they lack the high escape fractions that AGN have. It has been suggested that a combination of both could be responsible for the ionising emission, with AGN dominating at lower redshifts and star-forming galaxies at higher redshifts. The question remains where the turnover point is, as the escape fraction of ionising emission from star-forming galaxies seem to be too low at

redshifts $z \sim 3$ to maintain the ionisation level needed (e.g. [Grazian et al., 2016](#)) and even at higher redshifts, it has been claimed that their physical properties would have to change drastically to outperform faint AGN at the epoch of reionisation ([Grazian et al., 2016](#)). Still, other studies show that even at redshift $z \sim 7$, the AGN population could contribute only a third of the ionising flux ([Finkelstein et al., 2019](#)) and the bulk would be provided by faint star-forming galaxies (e.g. [Ouchi et al., 2009](#), [Mitra et al., 2013](#)). Just as with AGN however, bright star-forming galaxies are not numerous enough (e.g. [Steidel et al., 2001](#), [Cowie et al., 2009](#), [Iwata et al., 2009](#), [Robertson et al., 2013](#)).

This back and forth between different studies shows that there is no clear solution yet. The trend seems to go into a direction where AGN and star-forming galaxies share the production of ionising emission, with AGN dominating at lower redshifts (although the turnover is unclear) and star-forming galaxies at higher redshifts, with a possible contribution of faint AGN. This underlines the importance of understanding the properties of star-forming galaxies, both at the epoch of reionisation and at lower redshifts. There seems to be no strong evolution of the properties of LAEs and LBGs with redshift ([Ouchi et al., 2008](#), [Hu et al., 1998](#), [Blanc et al., 2011](#)), allowing us to infer their properties at the EoR without being able to directly observe them. A crucial element of the problem of reionisation is therefore the escape fraction of ionisation radiation from star-forming galaxies. If that is known with precision in dependence of other galaxy properties that are easier to measure, such as Lyman α equivalent widths or other emission line properties, we could constrain the ionising emission from star-forming galaxies and answer the question who ionised the universe.

1.6. Structure of this Thesis

Just as the introduction to this thesis tells the story of the paths of both the Lyman α and Lyman continuum emission from their origin in star-forming regions and hydrogen clouds in their galaxies, through the kinematics and morphologies of the ISM and CGM and eventually through the neutral hydrogen reservoirs in the IGM, the thesis itself is concerned with both Lyman α emission and Lyman continuum emission as well.

1.6.1. Open Questions

The properties of high redshift LAEs, Lyman continuum leakers and especially their connection is topic of this thesis. The question how many LAEs with extreme Lyman α equivalent widths can be found in integral field spectroscopic data from MUSE is addressed, as well as the connection between the equivalent width and other properties of the Lyman α line but also with the morphology and UV continuum. I try to find a connection between these properties and Lyman continuum leakage as well, using deep UV data from the Hubble Deep UV (HDUV) survey. Many proxies for Lyman continuum leakage have been suggested, to narrow down the promising candidates, such as a connection between Lyman continuum escape and the equivalent widths of Lyman α , the peak separation and the width of the line.

1.6.2. Summary of this Thesis

Here I give a short summary of this thesis with the most important results and methods that were used. The thesis is structured in three main parts: The first is the description of the graphical user interface (GUI) I created to classify emission lines found in integral field spectroscopy. The second (chapter 3 and 4) and third part (chapter 5 and 6) each use similar data-sets but with different scientific questions, the one addressing Lyman α equivalent widths, the other Lyman continuum emission. Therefore they overlap in some aspects of the data description and analysis, but can be viewed as individual projects that are planned to be published as a result of this thesis. Each of the two parts has a main chapter (chapters 3 and 5) that can be seen as a paper draft and a methods chapter associated with it (chapters 4 and 6) that contains further information and analysis that would not necessarily be included in a paper, but that are nonetheless important for understanding the process.

Chapter 2. Graphical User Interface for the Classification of Emission Lines: QtClassify

The first step is to find LAEs in the integral field spectroscopic data of MUSE. A catalogue containing emission line objects found in MUSE through a matched filtering approach is used as a basis. To help with

the visual classification of the objects and make use of the full information in the datacube, a graphical user interface was developed by me. The purpose and main functionalities are described in this chapter, as well as a brief introduction on how to use it most efficiently. This tool is useful not only to make the classification process more efficient, but it also has demonstrated its potential to detect new and unexpected objects and features in the data.

Chapter 3. Equivalent Width Distribution of Lyman α Emitters with MUSE

I construct the equivalent width distribution using data from the MUSE-Wide and -Deep surveys. To measure the UV continuum I use HST broad-band data and find that one third of our sample of LAEs have no visible counterparts, which demonstrates the power of integral field spectroscopic data in discovering LAEs. I find a fraction of 9% of our sample have high equivalent widths above the theoretical limit for normal stellar populations of 240 \AA , but I point out that this result is biased by the line flux limits of the survey. I therefore construct for the first time an equivalent width distribution function to better constrain the LAE population properties. I find one particular object with a measured equivalent width of $EW_0 \sim 600 \text{ \AA}$, a strong hint at an extreme underlying stellar population.

Chapter 4. Methods: Measuring Equivalent Widths

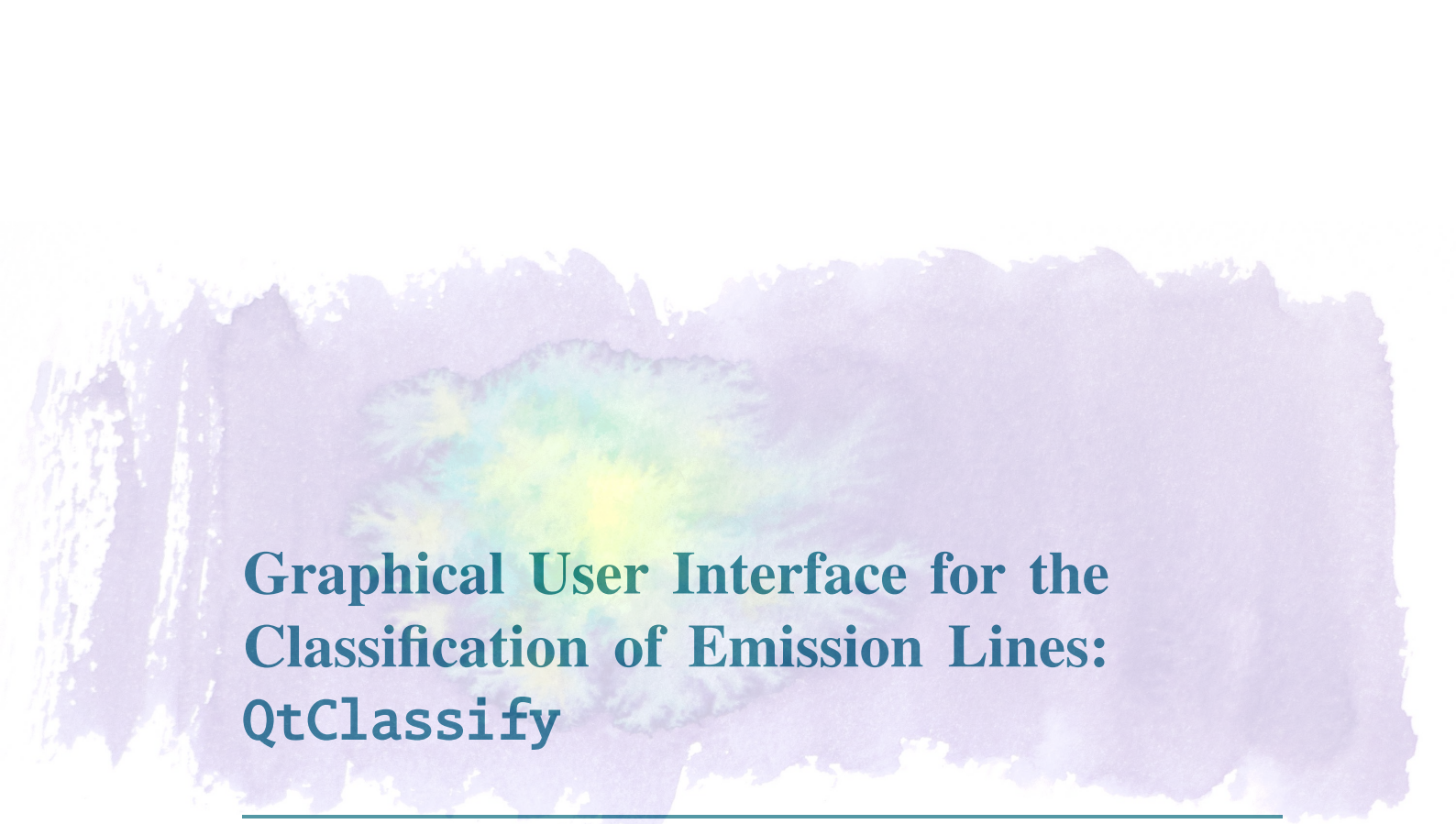
In this chapter I describe the steps taken to measure the equivalent width distribution shown in chapter 3. I show in detail the `GalFit` modeling process of the UV continuum counterparts in the HST broad-band data and discuss how I determine the β -slope that was used to correct the measured flux values to the Lyman α position. I find a median value of $\beta = -1.92$, slightly more red than previous studies. From our models I find that LAEs with smaller UV continua have higher EW_0 values and the distribution of axis ratios shows that the high redshift LAEs do not follow the same trend as local spiral galaxies, but are likely merging or irregular. The Lyman α line profile was fitted as well and I discuss the fraction ($\sim 30\%$) of objects with a blue bump. This fraction depends heavily on the S/N in the spectra and could be as high as $\sim 80\%$. I also show the expected correlations between e.g. the width of the line, the asymmetry and the peak separation.

Chapter 5. Lyman Continuum Emission from MUSE Lyman α Emitters

Moving from Lyman α emission to Lyman continuum emission, in this chapter I use UV data from the HDUV survey to probe the ionising emission from the sample of LAEs from the MUSE-Wide and -Deep surveys. I discuss the difficulties of identifying individual Lyman continuum leaker candidates and excluding the possibility of low redshift interlopers. I find two likely and four possible candidates and derive their Lyman continuum escape fractions. To determine the escape fraction in the complete sample, I select the 60 most promising candidates and stack them. To identify possible proxies for Lyman continuum emission, I also separate this sample into objects with high and low Lyman α equivalent widths, peak separation, line widths and asymmetry. I find a possible signal in the stack containing objects with a high asymmetry, which might favour the escape of Lyman continuum photons.

Chapter 6. Methods: Measuring Lyman Continuum Emission

In this chapter I describe the steps taken to find and measure the Lyman continuum from the broad-band data and discuss some issues with the analysis. First I compare the measured flux and standard deviation values to prior catalogues to ascertain that the aperture measurements give reasonable results. I find small background offsets in some of the HST bands, especially the MUSE-Wide area (not in the Ultra Deep Field (UDF), except for the band ACS F814W), which I correct for. To get the flux measurements at the correct wavelength positions for the Lyman continuum escape fraction, I use stellar population synthesis models and discuss the input parameters that were used. I find an intrinsic UV continuum to Lyman continuum ratio of $L_{1500}/L_{900} = 1.566$. I also discuss the spatial offset found in the stack of the selected LAEs but do not find a cause, which leaves the question whether the found signal is reliable or not.



Graphical User Interface for the Classification of Emission Lines: QtClassify

The task of classifying objects found in integral field spectroscopic data is challenging. The process of classification is facilitated by the amount of available information (spatial and spectral) but in order to save time and increase the reliability of the classification, the relevant information for each object needs to be easily accessible. Especially for cases where the classification is not obvious and has to be performed by a pair of human eyes, the size of the data-set and the subjectivity influence the efficiency of the classification process. Ideally, in order to decrease the influence of subjectivity based on personal biases, it is preferable to include multiple people in the classification of each object independently. To make a comparison of the results feasible, it is best if the people involved are using similar tools for the classification. For these reasons a graphical user interface (GUI) was developed, that combines the requirements of being easy to use, customisable and yet consistent.

Another aspect that should not be neglected in a process that involves human interaction is to keep the classifiers alert and motivated. Making the GUI not only helpful but genuinely enjoyable to use, with the aspect of gamification in mind, is therefore not only an added bonus but also a means to increase productivity. In this chapter I will introduce the main functions of QtClassify, which is publicly available via the Astrophysics Source Code Library at <http://ascl.net/1703.011> and the newest versions and updates can be found on Bitbucket at <https://bitbucket.org/Leviosa/qtclassify>. A basic description of its purpose and application was published in the appendix of [Herenz et al. \(2017\)](#).

2.1. Purpose

QtClassify was developed to make the task of classifying emission lines in integral field spectroscopy easy and comparable (when working in a team). It is not primarily an automatic redshift determination tool as it requires manual inspection of the data and makes use of the human eyes' natural ability to detect patterns. It has been tested for optical spectra from the MUSE (installed at the Very Large Telescope, VLT, [Bacon et al., 2010](#)), but might also be applicable for data from other instruments, other wavelength ranges and other purposes, as long as datacubes with three dimensions (two spatial and one spectral) are used as input, since using this information is the strength of QtClassify.

Its main use is to classify emission line objects in MUSE data that were previously found with a suitable detection software. It is meant for cases in which the emission lines of an object do not match any templates and prevent an automatic classification. This can happen if some of the emission lines in the object are hidden behind noise peaks, emission lines from other objects with different redshifts were combined into one object or if only one emission line is visible in the spectrum of the object. QtClassify can then be

used to visually detect additional emission lines that were below the detection threshold or to determine the nature of the line if no other lines are visible. For this thesis, the most important application for `QtClassify` was to find Lyman α lines and thus LAEs and distinguish them reliably from foreground objects. Therefore I will focus on separating Lyman α from other lines in the following description of `QtClassify`.

For a screenshot of `QtClassify` see figure 2.1 on page 41. In this example only one emission line was found in the data, but other emission lines are clearly visible, which makes it easy to classify the emission line ([O II] at a redshift of $z = 0.678$ in this case). Not only the type of emission line is important to know, but with the combined information of the observed wavelength, the redshift of the object can be determined. In the following the features and possibilities of `QtClassify` are explained.

2.1.1. Input Data

`QtClassify` was developed specifically for integral field spectroscopic data from MUSE (Bacon et al., 2010), installed at the Very Large Telescope (VLT) in Chile at the Utility Telescope 4. MUSE covers a wavelength range in the optical (from 4750 Å to 9350 Å), has a spatial sampling of 0.2'' and a field of view of one square arc minute. This results in 90 000 spectra for each datacube, which are however spatially correlated. Each spectrum has a spectral resolution of 1.25 Å, which means each of the 3.312×10^8 individual units of information, called volume pixels or voxels, has a resolution of $0.2'' \times 0.2'' \times 1.25 \text{ \AA}$. These numbers demonstrate the necessity for specialised tools to handle such a huge body of information.

The first step of identifying objects in a datacube is usually automated and has to be performed before using `QtClassify`, as it is mostly useful for classifying found objects, not so much for finding new ones. There are several approaches possible for detecting objects that might focus on different kinds of objects, emission line or absorption line or pure continuum objects. Going through the whole datacube by eye to find objects is tedious and difficult and requires a number of iterations by different people. It is also not ideal to find weak, low signal-to-noise emission lines or objects with a weak continuum. For a single deep field, manually going through the datacube to find objects might still be feasible, as was done partly for example for the UDF and Hubble Deep Field South (HDFS) (Bacon et al., 2017 and Bacon et al., 2015), but for a larger number of fields this is too time consuming.

For bright, low-redshift objects one possibility of identifying them in the datacube is to use ancillary broad-band data of the same field, e.g. from the Hubble Space Telescope (HST) or to make use of existing catalogues as priors if possible. The objects can then be matched to their position in the MUSE datacube and the spectra can be extracted directly, if the redshift is already known. The full discovering potential of integral field spectroscopy lies in its power to detect faint emission lines from objects that don't have a strong continuum and might not have been previously found and are not or only barely detectable in broad-band data. Several approaches to finding such objects are possible. One can create a running narrow-band from the MUSE data by collapsing (combining) a number of adjacent wavelength layers to increase the signal of potential objects. Then, an automated detection software for photometry can be run on the data (similar to SExtractor, Bertin & Arnouts, 1996), as has been done in the MUSE-Deep survey (see Bacon et al., 2017 and MUSELET Piqueras et al., 2017 which is based on SExtractor). Another approach that does not rely on photometry tools, is to use a three dimensional matched filtering method. This was used for example in the MUSE survey MUSE-Wide (Herenz et al., 2017 and Urrutia et al., 2019), where we applied the Line Source Detection and Cataloguing Tool (LSDCat, Herenz & Wisotzki, 2017) to create a catalogue of emission lines found in the datacube. As a useful by-product, it produces signal-to-noise datacubes by filtering the data with a template based on the point spread function (PSF) of the data. This signal-to-noise cube is useful in finding low signal-to-noise emission lines by eye, that are only barely visible in the actual data. After running LSDCat on the data, a catalogue of emission lines is produced, which can be read into `QtClassify`. The emission lines are sorted by object, which is important for `QtClassify`, since ideally every found emission line belonging to an object should be examined. That way it is also possible to disentangle objects that are spatially close and were thus grouped as one object, but have distinct emission lines, by which they can be separated. The same can be done in reverse for objects that were separated by LSDCat because of their spatial extent (which often happens for near-by galaxies) but have clearly the same emission lines at different positions (for example in spiral arms). For such approaches of finding objects it is useful to first subtract the continuum of objects, to only detect the emission lines. This can be done by using a median filter on the data and subtracting it from the datacube.

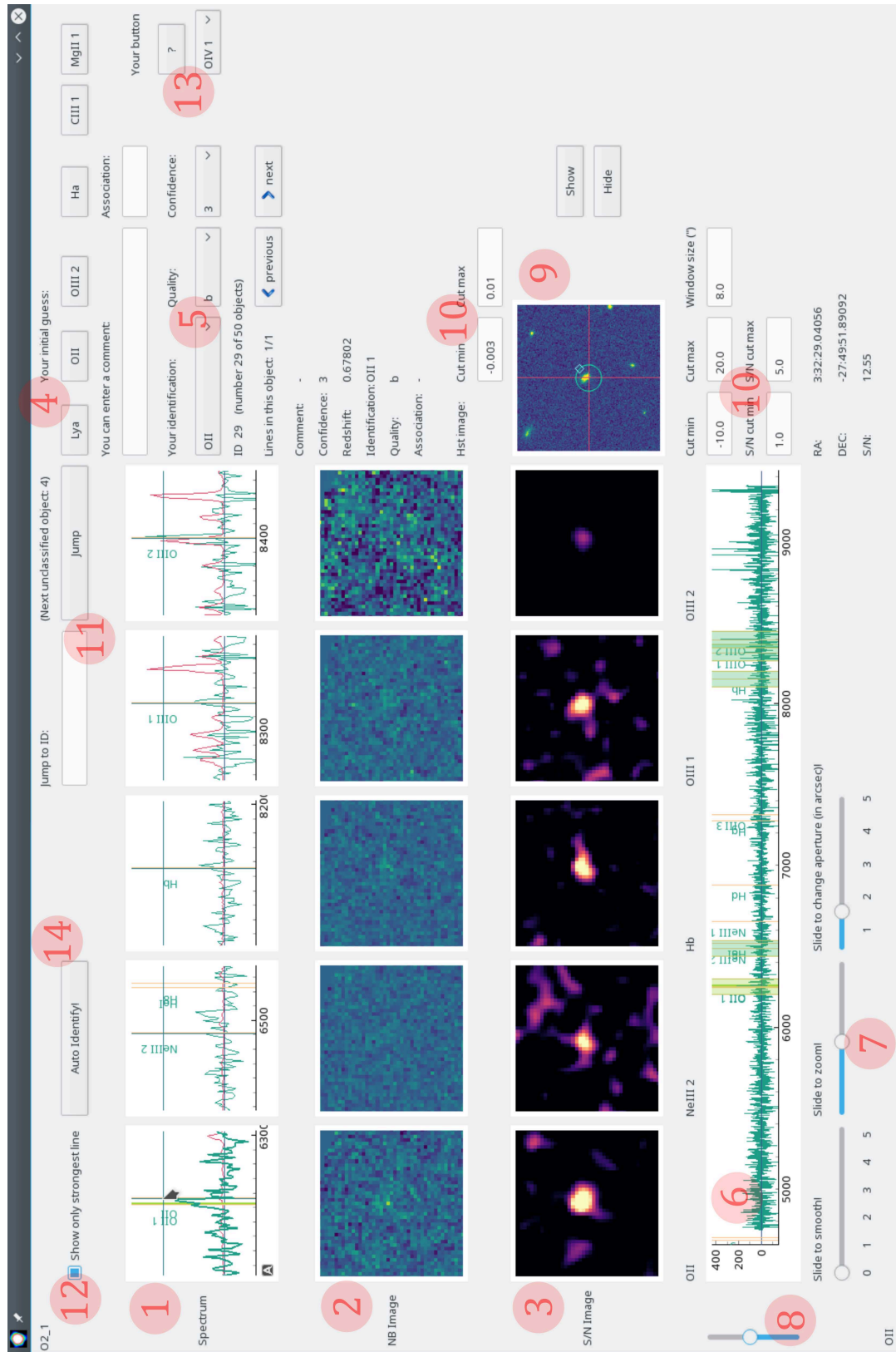


Figure 2.1.: Screenshot of QtClassify.

No matter which approach is used to find emission lines in the data, the goal is eventually to produce a catalogue not only with positions (both spatial and spectral) but also containing information on the redshift and nature of the found objects. This classification process can be done using `QtClassify`. As input data it needs a catalogue containing all found emission lines grouped spatially into objects with information on their right ascension and declination (or pixel positions) and wavelength. Additional information that is needed is the signal-to-noise (S/N) ratio of each emission line, to identify the lead line (with the strongest signal). The datacube itself is needed of course (ideally continuum subtracted) along with information on the variance in the data, as well as a cube containing the significance (a signal-to-noise cube), which is provided by `LSDCat` automatically. In the following, I assume that such input data is available and show examples of applying `QtClassify` to the MUSE-Wide survey (see chapter 3 section 3.2 for a more in-depth introduction to MUSE-Wide). As an additional help it is useful to also have a broad-band image at hand, e.g. from HST, to use as input as well.

2.1.2. Requirements

For `QtClassify` to run you will need python version 2.7.x or python 3 (the updated version is available on Bitbucket at <https://bitbucket.org/Leviosa/qtclassify>) and at least the following packages:

- numpy
- astropy
- scipy
- pylab
- PySide (or PyQt5 or PyQt4)
- SIP
- pyqtgraph
- matplotlib

You can download pyqtgraph here: <http://pyqtgraph.org/>. Pyqtgraph provides the main functionality of `QtClassify` and is based on PyQt4/PySide and numpy. Depending on the size of the datacubes that are loaded, the user should make sure there is enough RAM available. Usually MUSE datacubes of a single pointing will be around 4 GB.

2.2. Inspiration and Applications

The challenge of classifying emission (and absorption) lines in spectroscopic data, especially in integral field spectroscopy, arises in many different research areas. Whenever the observation was not aimed at a specific, already known target but instead directed at a point in the sky in a kind of 'spectroscopy of everything' approach (which is often the case for surveys such as MUSE-Wide survey, [Herenz et al., 2017](#) and [Urrutia et al., 2019](#), for which `QtClassify` was developed), after the first step of identifying objects in the data, the objects need to be classified, which means their redshift and object type has to be determined. In the case of the MUSE-Wide survey, the objects were first found purely based on their emission lines, which were grouped into objects by their positions. This can lead to superpositions of different objects, where one object was ascribed emission lines that in reality belong to a different object. However, identifying these superpositions is another task that can be facilitated using `QtClassify`.

There are several similar approaches to object classification in the literature, for example `SpecPro` by [Masters & Capak \(2011\)](#) (which they used e.g. in [Masters et al., 2012](#) for identifying quasars in the COSMOS survey field), which is an interactive IDL program using spectra and photometry, displays SEDs and lets the user classify objects based on this information. This program is not suited for integral field spectroscopy, however, which is the distinguishing trait of `QtClassify`. It uses all available three dimensional information of the datacube in a unique way to help classify objects, identify superpositions and find extended Lyman α emission or other objects the detection algorithm might have overlooked.

2.3. Functionality

In this section the functionality of `QtClassify` is described in detail and it can be seen as a rough guide on how to use the tool. Please refer to figure 2.1 to look at the individual features, which are marked by red

numbers. The following explanations are based on using an input catalogue for QtClassify that contains the detected emission lines grouped into objects with information on each line's signal-to-noise ratio (the catalogue used in the following examples was created by LSDCat).

2.3.1. Displaying Different Lines

The core functionality of QtClassify is the display of different lines at positions in the data where they would be expected given a specific guess for the main line (marked number 1 in the screenshot in figure 2.1). Most objects have multiple lines in their spectra, but some might be too weak to be detected with the emission line detection tool that was used. However, with the prior of knowing the positions (both in spatial and spectral dimension) of the expected additional lines, one can try to pick signal up by eye that might be too faint or too hidden in the noise for an automated process. Basically, the tool lets you ask the question 'If the detected emission line was X, do I see lines Y and Z as well?' This is both a guess about the classification of the emission line as well as about the redshift of the entire object.

At the top right corner under 'Your initial guess:' (4) you can choose your guess for the line that you want to classify. The default options for guesses are Lyman α (1215.67 Å), O II (3729.875 Å), [O III] 2 (5008.240 Å) and H α (6564.61 Å). The strongest line (in the sense of the highest signal-to-noise ratio) in most objects is one of those four, which makes them a good first guess. Additional options that are less frequent are C III 1 (1906.683 Å) and Mg II 1 (2796.35 Å).

Depending on which line is your first guess, other lines are expected to be visible in the spectrum of the object. These are displayed in the five by three array of images that take up most of the GUI (1, 2 and 3). The top row (1) shows parts of the spectrum of the object that should contain the additional lines. The middle row (2) shows the part of the datacube where the object is located and the wavelength layer that would contain the additional emission lines while the bottom row (3) shows the same but for the signal-to-noise cube (produced by LSDCat). See section 2.3.4 for more information on the possible interaction with these panels.

The columns display the different lines sorted by wavelength with the guessed line highlighted by a bold line in the spectrum (top row, 1). If one of the additionally expected lines would be outside of the wavelength range of MUSE for the redshift the object would have for the guess of the detected main line, the column is left empty.

2.3.2. Classification

The classification options are stored in a list in an additional file, which can be modified to accommodate different needs. The options can be found in the drop down menu to the right below 'Your identification:' (5). The lines available as first guesses are among the possible classifications, as well as other options in case those don't fit or the signal is not actually an emission line (but instead noise or a continuum residual from a bright star). In some cases spurious detections were still left in the catalogue, which can be marked as 'Crap' in our classification scheme. This is meant for a signal that is definitely not real. In other cases the line might be real but it is not possible to classify it as it is close to the edge of the field, which can be marked as 'Edge'. In rare cases continuum objects such as stars or bright quasars mimic emission lines after the continuum is subtracted with a median filter. Thus only the strongest variations of the continuum are left behind, which can look like an emission line to LSDCat. This falls under the category of 'Continuum residual'. Should no option be adequate, the line (if it is real) can be classified as 'Something else', which should be accompanied by a comment in the comment line above the drop-down menu (5) explaining which line is seen. If it is not possible to say whether the line is real or not and if so what line it is, it can be marked as 'Revisit', but this option should be used with caution. Since there is also the possibility to flag an emission line with a confidence (see below section 2.3.3) it makes sense to make an educated guess and then flag the object as uncertain.

2.3.3. Quality and Confidence

There are two measures of how secure the classification of an object is: Quality and confidence. The former is mostly objective, the latter subjective and may therefore vary from classifier to classifier. A similar scheme to distinguish the quality of a classification was used e.g. by Drozdovsky et al., 2005.

Quality

The quality of a classification is determined by the presence of other identified emission lines (the drop-down menu is to the right of the classification drop-down, 5). If the lead line is classified, `QtClassify` automatically tries to match all other detected lines of that object to the redshift the object would have if the lead line were indeed the line it was classified as. If all other lines match one of the lines in the internal line list, the quality is set to 'a'. This is done automatically and usually needs no manual intervention (see section 2.3.6 on other automatic processes). In cases where the object contains one or more lines which are not in the internal line list `QtClassify` uses to compare with, the quality will be set to 'c'. This tells the classifier that there are unclassified lines in the object, which could either be missing in the line list or contamination from another object, which needs to be noted in the comment line. The quality will also automatically be set to 'c' if there is only one detected line in the object in question, which will be the case for most LAE (since the other possibly visible line is `C iv`, which is rare). This means the classification hinges on that one line, not that the classification is wrong. In rare cases where only one line was detected by `LSDCat` but other lines are visible to the classifier in the MUSE data or in the S/N cube data, the classifier should select the quality 'b'. This is the main reason why `QtClassify` was developed, since objects like these are not easily classified automatically and require the capacity of pattern recognition of the human eye as well as the knowledge of the classifier. Again, in these cases it is useful to write a comment containing the nature of the possible additional lines that are visible, but whose S/N did not exceed the detection cut of the emission line detection software that was used.

Confidence

Unlike the quality flag, the confidence flag is mostly subjective and has to be selected by the user for each lead line (while the quality is set automatically and just needs to be adjusted in certain cases). The default value for the confidence is '0', which means there is no classification yet. This flag should not be used by the classifier, since this would mean the classification is not possible. The lowest confidence is '1', which means the classifier is uncertain about the classification. Confidence '2' means the classification is probable and confidence '3' means the classification is certain. The latter is automatically selected by `QtClassify` in cases where the quality is set to 'a' (either because all other lines match the ascribed redshift or because it was set to 'a' by the user), since this means the classification is secure. For quality 'b' objects the confidence should also mostly be set to '3'. For all other cases the confidence flag depends on the classifier but can be influenced by the shape and strength of the line and also (if available) on ancillary broad-band data from HST. This aspect of the classification process is subjective and not only varies by person but also by day or mood (which has been observed during our own classification process for MUSE-Wide). It is useful therefore to have multiple people do the classification for the same object or even decide together after a discussion.

2.3.4. Accessing the data

So far I have discussed the static aspects of `QtClassify`, but since it is a GUI, it lets the user interact with the data in different ways in order to access the data more easily.

Moving through the datacube

In the top row of the main window relevant parts of the spectra are shown where other lines are expected (1). When moving the mouse over these sections, the wavelengths of the cut-outs in the second (narrow-band, 2) and third row (signal-to-noise image, 3) change to match the mouse position in the spectra in the top panels. This happens simultaneously for all shown lines (all columns in the main window), making it easy to see if emission lines appear or disappear when moving through the datacube. This is extremely useful when determining whether a line is real or a noise feature (which would only be visible in one of the panels, meaning at one wavelength).

Another interesting option is to shift the line which indicates the wavelength at which the emission line was found in the column where the guess is displayed (where the mouse is positioned in the screenshot 2.1). This will change the wavelength that is used to determine the redshift. For accurate redshift determination, it is advisable to fit the line after the classification process (with a normal Gaussian or an asymmetric Gaussian for Lyman α lines). However, to get a first by-eye improvement of the redshift or if the assigned wavelength is off due to a double peak or a noise peak, the line can be shifted to the correct position.

The full spectrum

Under the three rows with panels showing the spectra and datacube cut-outs, the full spectrum is displayed (6). All lines that are internally used to match found emission lines (not just the guesses) are

displayed with orange vertical lines annotated with their respective elements (just as in the first row, 1). The spectral ranges shown in the panels of the first row (1) can also be seen in the full spectrum (6) as light green shaded areas. They can be moved, which will in turn move the spectral cut-out of the first row (1) to the mouse position. That way, other parts of the spectrum can be viewed. The wavelength range of the individual panels can also be changed by clicking and dragging the edges of the shaded areas in the full spectrum (6).

Using the Sliders

There are four sliders available (7 and 8) which influence the spectra that are displayed. The left, vertical slider (8) changes the y -axis scaling of the full spectrum and the top row of spectral cut-outs. This is useful for faint emission lines or for datacubes with different exposure times or flux values. The leftmost horizontal slider (7, first) under the full spectrum can smooth the spectra using a Gaussian function with a standard deviation as displayed on the slider. This is useful if the spectrum is noisy, since the smoothed spectrum makes it easier to see if a possible line is simply a noise peak (and vanishes when smoothed) or whether it persists. The slider in the middle (7, second) changes the wavelength range that is displayed in the top row of spectral cut-outs for all cut-outs simultaneously (as opposed to dragging the shaded areas in the full spectrum, which will only change the wavelength range of that specific line). The rightmost slider (7, third) is probably the most useful one. It changes the aperture size used for extracting the spectra. The size can be seen as a circle in the HST (or other broad-band data) cut-out at the right of the main window (9). Changing the aperture can be helpful for distinguishing between Lyman α and a weak O II doublet, since LAE usually have extended halos (Wisotzki et al., 2016, Wisotzki et al., 2018). This means that for an increased aperture the line in the spectrum will grow if it is indeed Lyman α and stay the same or get more noisy if it is not.

Changing the cuts

In the bottom right corner there are five lines that can be edited (10, bottom), four for the cuts of the second and third row (narrow-band, 2, and signal-to-noise data panels, 3) and one for the window size. The cuts of the broad-band data can be changed above the broad-band data panel (10, top). The cuts determine the colours of the values in the panels such that the brightest colour will correspond to the 'cut max' value and the darkest colour will correspond to the 'cut min' value. All the cuts should be set such that faint objects are still clearly visible and should be adjusted using an object where the classification is easy.

The most important cuts are those for the signal-to-noise data. They should be matched to the signal-to-noise cuts used by the source detection software (e.g. LSDCat) that was used to create the emission line table. The most useful way is to set the maximum cut to the detection limit, which makes it easy to see if a line was detected or not (in that case it is the lightest colour).

2.3.5. Additional and Optional Functionality

QtClassify offers some additional functionality that can further help with the classification process.

Broad-band data

Looking at broad-band data (if available) can be useful in several cases (9):

- If there is only one emission line in the object and no other line can be seen in the spectrum, not seeing a counterpart in a broad-band filter image that is to the blue side of the potential Lyman α line can be a strong hint at the object being an LAE. The size of the object in the broad-band data (in a filter to the red side of the line) can also be indicative, as LAEs are typically small in the UV continuum but have extended Lyman α halos (see for example Wisotzki et al., 2016), which means the Lyman α emission is more extended than the broad-band counterpart (keeping the PSF in mind).
- Sometimes there are lines in an object that cannot be identified and that don't match the redshift determined by the other lines in the spectrum. This can happen if the emission line belongs to an entirely different object but was falsely grouped. This can be determined by checking if there are multiple objects visible in the broad-band data that could not be clearly resolved with MUSE. Often the unidentified lines in question are slightly offset to the other lines in the object, which might hint at a background LAE that is not visible in the broad-band data, either.
- Large galaxies are often fragmented into multiple objects by LSDCat, which becomes clear by looking at the broad-band data again. Especially near-by spiral galaxies can be prone to fragmentation due to separated emission line regions in their spiral arms.
- Spurious detections, though rare, can also be more easily identified if there is no counterpart in the broad-band data. However, this could also be caused by a low-luminosity LAE.

Cuts and window size

Both the window sizes and the cuts (10) for the displayed colours are important to adjust in a way that makes it easy to distinguish signal from noise. Setting the cuts for the signal-to-noise datacube (3) should reflect the signal-to-noise cut used as a detection threshold. That way, you can directly determine the S/N ratio of the line you are looking at. If the maximum is set to the detection threshold it will be the lightest colour of the chosen colour map. If any lines are visible as bright dots in the flux or S/N panels that were below the detection threshold, their colour and brightness can give a feeling for the significance of the additional line(s). You can also invert the colours by simply switching the cut values.

Jump to object

Enter the ID of an object to jump directly to it (11). This way it is easy to revisit specific objects with emission lines that are difficult to classify or that might belong to another object (in cases where objects were incorrectly grouped). Next to the 'Jump' button the ID of the next unidentified object is given, which can be jumped to so as not to go through all the already identified objects and lines. For this information any objects containing lines that are either not classified at all or that were marked as 'revisit' are considered. This is especially helpful in combination with 'Auto identify' (14, see below), which takes care of the most obvious cases and the classifier can then jump to the objects that were not classified automatically and that require manual intervention.

Only show strongest line

By default, when clicking on 'next' or 'previous', each emission line in an object is shown for all objects in the input catalogue. Since QtClassify automatically tries to match the other lines in an object once one line was classified, in most cases it is sufficient to only classify the strongest line manually. If the check box is ticked (12), only the strongest line will be displayed. In case the redshift derived from classifying this line does not match all the other lines of the object, the confidence will be automatically set to '0' and the quality to 'c' as an indicator that the other lines in the object need to be looked at individually. This can happen if the classification was wrong but also if there are lines in the object that are not in the internal line list or if there are lines that were grouped to the object but belong to another, close-by object.

The custom button

When selecting the 'Show only strongest line' option, most emission lines that need to be classified in order to determine the redshift of a given object are covered by the standard options to choose from: Lyman α , O II, [O III] 2 or H α . Other common lines that might be of interest are C III 1 and Mg II 1, which are also options. If neither of those options match the strongest detected line of the object, this might be because the real strongest line is hidden behind a sky or noise line in the MUSE spectrum and is thus either not detected at all or lost a significant amount of flux in the data reduction process. In this case, a different line will be detected as strongest and the custom button can be used to test if this line is another emission line. To use the custom button (13), select the emission line you want to probe in the drop-down menu below the custom button. This will then turn green (signaling that it can now be used) and display the selected emission line (see the screenshot in figure 2.2, where the custom button was set to H β). It can now be used like any of the default buttons for line guesses. It can also be changed again at any time using the drop-down menu.

The custom button is not only useful for classifying the main emission line in an object, but also to test other, weaker lines. For example if the strongest line is not easily classified, it might be better to look at other lines. This can happen in case of superpositions, where the emission lines won't match to the same redshift, or in cases where one galaxy was split into multiple objects (corresponding to different parts of the galaxy), not all of which have the same emission lines.

2.3.6. Automatic Processes

To help speed up the classification process, there are a few automatic processes implemented in QtClassify.

Auto identify!

Since the detected emission lines are usually combined into objects by their proximity (for example when using LSDCat for detecting the emission lines in MUSE datacubes, as is recommended), for most cases where there are multiple emission lines and no superpositions, the classification is straight forward and does not need any manual intervention. The button 'Auto identify!' (14) uses a simple algorithm to cross-match the list of found emission line wavelengths to the list of possible emission lines provided by QtClassify (which can also be customised, see 'Customisation' in 2.3.8 on page 49). This is done by determining the line with the highest S/N ratio as the anchor and iterating over the possible classifications for this line. For

each possibility, the redshift of the object is determined and each of the vacuum wavelengths in the list of possible lines is shifted accordingly. The wavelengths of the rest of the detected emission lines in the object are then compared to the shifted list of lines and each match (to within 5 \AA) is counted. If there are matches for all detected emission lines in the object, the redshift is assumed to be correct, if any one line does not have a match, the process is repeated by assuming the strongest emission line of the object to be a different line and thus a different redshift and repeating the process. Once a redshift is determined by matching the wavelengths of all found lines to a list of possible lines, the process is halted. This means that the procedure does not work for superpositions, where not all lines can be matched to a single redshift, as well as for objects with only one emission line. The automatic redshift determination is similar to tools like e.g. EZ, by Garilli et al. (2010).

For matching the lines, a margin of error (the default is 5 \AA) is assumed for the wavelength of the detected emission lines. The initial wavelength measurement of the found line can be influenced by the width of the line, the line spread function, the noise in the spectrum or intrinsic line shifts (as is the case for the Lyman α line) and should be measured again after the classification process with an extracted spectrum.

It is recommended that after using the automatic identification, which will usually classify most of the objects that have more than one emission line, the objects are still checked visually for accuracy, as the algorithm is not infallible. Especially for objects with only a few lines that might be superpositions or not all real, an unreasonable redshift might be determined. A known problem is the separation of the O II and Mg II lines. If both lines of the O II doublet are detected in the object, but nothing else, the automatic classification will usually identify the lines wrongly as the Mg II doublet at 2795.528 \AA and 2802.705 \AA instead of the O II doublet at 3727.092 \AA and 3729.875 \AA . Although the former have a wider separation, the margin of error assumed for the wavelength position makes the Mg II doublet a viable option for the algorithm. This could be improved in future versions by sorting lines with similar separations by their likelihood of being detected.

Excluding single line objects (which cannot be automatically classified), roughly three of four objects can be classified that way, with an error rate of around 10% (which only consists of objects with two emission lines, where there are more possible combinations). This means the automatic classification is extremely useful and speeds up the classification process, but needs a quick manual check for dubious objects. The remaining objects where the automatic classification fails are usually stars (where a stellar continuum was interpreted by the emission line detection tool as emission lines), bright galaxies with more lines than in the default line list used by the automatic classification or objects where noise peaks were identified as emission lines. Cases where the automatic classification gives a wrong answer are objects with two emission lines, where either the O II doublet was mistaken for the Mg II doublet (see above) or there was a double peaked Lyman α line with both parts of the line detected individually (rare but interesting cases). The latter are usually also either identified as the O II doublet or the Mg II doublet, but since the blue bump needs to be bright to be detected as an individual emission line, the Lyman α line will display the typical asymmetric shape and can easily be picked up by a visual inspection.

Note that in cases where the redshift could be determined successfully and all lines match, the confidence is automatically set to '3' and the quality to 'a'. Another thing to keep in mind is that the automatic identification will overwrite any prior manual classification, even in cases where it can't find a solution (but not for single emission line objects). Therefore it is advisable to use this button only at the beginning of the classification process and leave it untouched afterwards.

Automatic filling in of quality and confidence

Automatically adding the quality and confidence flags is somewhat redundant if the automatic identification button is used (see above). In cases where a line is selected as a classification and the other lines found in the object match the redshift determined by the classification, the quality will automatically be set to 'a' and the confidence to '3'. If any of the other lines do not match, the quality is left at 'c' and the confidence at '0', indicating no confidence level was set yet.

Matching of other lines

Similarly, if an emission line is classified, all other lines of the object will automatically be matched to the redshift and compared to the internal list of possible emission lines. However, if there is one detected line that can not be matched, the quality and confidence will not be set to 'a' and '3', and the classification of that particular emission line will be set to 'no match'. If the classifier is certain that the choice was correct and the line that could not be matched is noise, a superposition or real but simply not one of the lines in the default list of possible lines, they have to manually set the quality to 'a' and the confidence to '3'.

If a line is classified and the corresponding redshift would be below '0', a pop-up window appears notifying the classifier of their error.

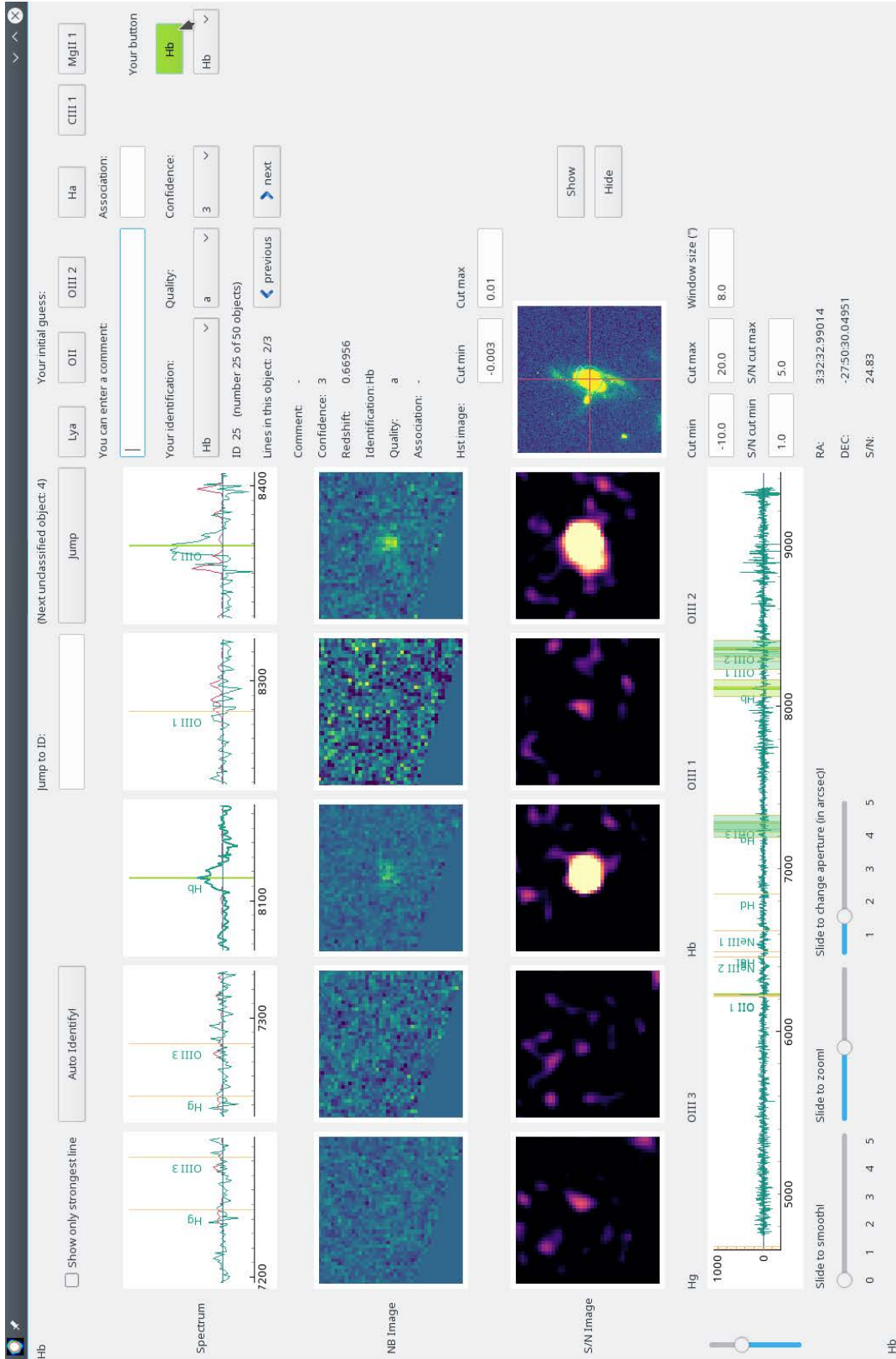


Figure 2.2.: Screenshot of QtClassify displaying the customisable button.

2.3.7. Output catalogue

After classifying all objects and emission lines or if you want to end the classification session, the main window can simply be closed by clicking at cross in the top right corner. Any identifications will be stored automatically in a catalogue that was specified as the output catalogue, which will be overwritten in case it already exists (so the user should be careful). To avoid unpleasant surprises, QtClassify checks if the specified output catalogue name exists when starting the GUI and a pop-up opens prompting the user to either continue by loading the existing catalogue or to choose a new output catalogue name and begin again. In case an existing catalogue is loaded, remember to not press the 'Auto identify!' button (14), as this will overwrite any previous classifications. The output catalogue will contain the same columns and content of the original input catalogue with the following additional columns after classification:

- identification (the classification by the user or the automatic identification)
- short names (a different version of storing the emission line names)
- comment (if input by the user)
- redshift (automatically generated, but can be influenced by the user)
- quality (objective flag, either automatic or input by the user)
- confidence (subjective, should be set by the user)
- association (in case the emission line is part of a different object)

Note that each time you move forward or backward in your catalogue to classify, the output catalogue will automatically be written, so there is no loss of classifications if the programme should crash or if it is closed. This way the classification can be interrupted when needed and picked up at a later time.

2.3.8. Customisation

Apart from the different options to customise the GUI directly in the main window, such as changing the cuts for the displayed images (10) or the scaling range for the spectra (7 and 8), it is also possible to change some aspects of the GUI when starting it, e.g. to avoid having to change the cuts for every classification session since the default values might not suit every purpose. You can also choose between two different colour schemes (old and new) when starting the programme with the switch '-d'. The old design uses cubehelix and jet as colourmaps as well as some basic colours for the lines, while the new design is based on the colourblind-safe, perceptually uniform, sequential colourmaps Viridis and Magma and all displayed lines are taken from those colourmaps. If the user prefers different colourmaps, they can be changed in the script `get_colors.py`. The colourscheme in the example in figure 2.2 is the design using Viridis and Magma.

What is more, you might come across the problem that the options in the identification drop-down menu (5) are not sufficient or your favourite emission line is missing or you want to include He II as an initial guess at the top without having to use the custom button (13). All these problems can be solved by using the two scripts `custom_lines.py` and `custom_options.py` which are specifically created to be easily customised by the user. They store the lines, wavelengths and possible identifications as simple Python lists and dictionaries. To change which lines can be used as initial guesses, go to `custom_options.py` and edit `possible_lines_names` and `dict_other_lines` (there are also explanations on how to do that in the scripts themselves). Got to `custom_lines.py` and follow the instructions therein to change the wavelengths or identifications for the drop down menu under 'Your identification'.

2.4. Possible Improvements

After having used QtClassify for 100 fields, over 1000 LAEs and ~ 1600 other objects in the MUSE-Wide survey, it is safe to say that it is a useful tool for classifying emission lines, but there are still some aspects that could be improved in the future. The programme is quite specific in its science and data application, which is why we decided not to put it in a dedicated paper or letter. To make it more accessible and useful for other applications, it would be helpful to lift some restrictions on the input data that is needed or additionally provide a more detailed tutorial on how to produce the needed input data using for example LSDCat. It has also not been tested for other IFU data or a different format of the data, which might also cause problems for some users.

Improving the automatic redshift determination in order to reduce the number of objects the user has to

examine by hand is another option for future improvement. For this the signal-to-noise information could be used and added up for different guesses for the main line, so that we would also obtain an estimate of the significance of the determined redshift. Then the user would only have to look at objects with a low significance of the pre-determined redshift to classify the objects by hand.

Another aspect that could be overhauled is the design. The placement of the buttons and elements in the GUI could be more intuitive. Some of the sliders are only used rarely (such as the one for the wavelength range that is displayed) and could be moved to a customisable script along with the line lists. The programme was developed over the course of many months and improved over years of usage, making additions as the users requested. After using it extensively and gathering experience it should now be easy to make it more intuitive and streamlined. It has not yet been tested on many different screen settings, though, but since some of the elements are static while other are not, they can be in slightly different places for different computers and screens.

In the future the GUI could also be tested and optimised for different input catalogues from other emission line detection tools or for other input data-sets from different instruments or wavelength ranges. As long as the data that is analysed comes from an integral field spectrograph, the basic method and principles of `QtClassify` still apply.

2.4.1. Gamification

As mentioned in the beginning of this section, an important aspect of the GUI was to keep the user motivated and alert to avoid mistakes. The many options to interact with the data and the additional excitement of discovering new features, such as extended Lyman α halos or blobs, make the process of classification fun and entertaining. However, especially for large catalogues, the process of classification can become monotonous and tiring. To increase the effect of excitement through discovery, keep the users from getting bored and include some aspects of the principles of gamification, here are some possible ideas for additional features:

- Including a health bar, which will be reduced if you mark an object as 'Something Else' or 'Revisit' and will be increased if you return to classify them. This way users would be motivated to classify even difficult emission lines instead of leaving them for later or completely unclassified.
- Gaining points by marking objects as 'b', since such objects are difficult to classify. However, it is advisable to discuss such objects with a second or third user, as an additional incentive to find more lines in the data could lead to an overenthusiastic interpretation of noise peaks.
- Getting a warning after a certain time or a certain number of objects that tells you to take a break. It is not recommended to go through more than 100 emission lines in one sitting, as the concentration will inevitably go down.
- Including a progress bar based on the number of objects you looked at. This would help with longer input catalogues to keep an overview of how much work is still ahead and when is a good time to cut the classification session in half.
- Displaying some statistics that automatically update when you go through the data and tell you e.g. the fraction of LAEs or the number of objects marked with 'a', 'b', or 'c' or confidence '1', '2', or '3'.
- In the same vein, there could be milestones linked to the statistics. After a certain number of classified objects a pop-up could alert the user of the success. This could be combined with a recommendation to take a break after a certain milestone.

2.4.2. Identification of Counterparts

Another useful extension of `QtClassify` would be including the possibility to identify counterparts for the emission line objects in broad-band data such as HST bands or other photometric catalogues. Until now, this is an additional step after the classification of objects and assignment of redshifts based on their emission lines. There is already the option to display an HST image (9) to help with the classification by looking at the broad-band morphology of the object. This information can be taken into account when distinguishing for example between a high redshift LAE and a lower redshift O_{II} emitter, which can look similar in a low signal-to-noise spectrum. If the object is extended in the HST image, this would argue against a high-redshift object.

Thus the broad-band information can already be taken into account in the classification process and could

be extended to multiple bands or photometric catalogues for simultaneously ascribing counterparts to the emission line objects. When using multiple HST bands it would also be possible to get a SED by eye, by looking at the brightness of the object in different bands to look for possible breaks. One or more of the bands would be marked with the positions of objects either already found with MUSE, to detect possible superpositions or emission lines that belong to the same object but were sorted into multiple objects due to extended emission, or objects from existing catalogues (usually based on photometry). Ideally, the user would be able to select the counterpart that most likely corresponds to the MUSE emission line object, based on position, shape and SED. The selected counterpart(s) would then be stored in the output table, possibly with a comment or a confidence, similar to the classification confidence flag. Clicking on the possible counterpart could also display information on that object from the additional catalogue, like photometric or spectroscopic redshift (if available), to give the user an idea of previous classifications. This should only be a guideline, however, since photometric redshifts are usually less secure than spectroscopic ones.

In some cases, especially for the high redshift LAEs, there might not be a counterpart visible, which also should be marked in the output table. Even if there is an object close by, this object is not necessarily the counterpart. Considering only the spatial coincidence would be what an automatic matching algorithm would do, but using a GUI like QtClassify would avoid such mistakes.

Taking all these advantages into consideration, including the option to manually select the broad-band counterpart of the emission line objects in QtClassify would be the next step of improvement.

2.5. Access

QtClassify is hosted on Bitbucket¹ and free to use for everyone interested as long as credit is given. The paper 'A first catalogue of 8831 emission line galaxies in the MUSE-Wide survey' by Herenz et al. (2017) contains a brief description of QtClassify in the Appendix, which can be cited if QtClassify is used in a publication. There is also the possibility of leaving feedback or filing a bug report in Bitbucket. However, as of the conclusion of this PhD thesis, the maintenance of the code can not be guaranteed. It is however created in a way that makes it easy for others to develop and improve on their own, since there is ample documentation, well documented code and a modular setup which allows users to easily customise for example which lines should be used and insert their own wavelengths. The code can be found on Bitbucket in this location: <https://bitbucket.org/Leviosa/qtclassify>.

QtClassify is also submitted to the Astrophysics Source Code Library (ASCL) and can be found here: <http://ascl.net/code/v/1628>. This can also be used to cite it in publications.

2.6. Process

QtClassify is one link in the chain to creating a catalogue of objects for integral field spectroscopic data such as the MUSE instrument provides. In this section I want to sketch a possible work flow to illustrate the process of arriving at such a catalogue and the role QtClassify can play in this by using the example of the MUSE-Wide survey.

After the observations and data reduction, we used LSDCat to find emission lines in the MUSE data and group them into objects.

With the catalogue provided by LSDCat as input for QtClassify it is possible to classify the emission lines. To ensure the best results, we chose to make this a team effort by including several people in the process. For the first 24 fields of the MUSE-Wide survey, each field was classified by three people individually using QtClassify. When this was done, we had consolidation meetings in which the three classifiers met with an additional, unbiased fourth classifier, that had not yet classified the emission lines of that field. Comparing the results of the three classifiers and based on the confidence flags given, as well as on the opinion of the fourth classifier, a decision was made for each object. Although in most cases the decisions were straight forward, some objects needed further discussion. Other reasons for distributing the task of classification and consolidating the results are possible personal biases, especially when it comes to the confidence flag. Some users might consider a classification straight forward, while other have doubts. Therefore it is important to not only consolidate the actual classification, but also the confidence and quality flags. The

¹Bitbucket, which is owned by Atlassian, lets you host your software projects both for commercial and private use in a Git revision control system. This makes it possible for people to work on a project together in different branches which can be merged and lets you track the changes or file bug reports.

confidence flag can later be used for analysis that rely on a correct classification, e.g. by excluding low confidence objects.

After gaining experience from the first 24 fields, for the next 36 fields of MUSE-Wide we decided to divert from our initial strategy slightly, by reducing the number of initial classifiers to two per field, but keeping a third person for the consolidation process with alternating combinations of people for these consolidation meetings. This reduces the workload per person while still reducing personal biases and allowing for discussions of uncertain objects.

Since the absorption line and continuum objects are not covered by LSDCat or QtClassify, in order to obtain a complete catalogue of all objects in the datacube (not just emission line objects), further tools or methods are needed. Comparisons with existing broad-band data can help identify weak continuum or absorption line sources, which have to be analysed separately to obtain reliable redshifts.

2.6.1. Conclusion

In Astronomy, the tools we use are usually created by fellow scientists instead of dedicated developers. This often results in software that is designed for specialised purposes, as is the case for the GUI at hand. Having knowledge of the application of the tool you are developing and working closely with the users who are constantly testing and recommending improvements is extremely useful for creating a valuable and useful tool. The drawback of course is the limited application and apart from the handful of people who worked with QtClassify on MUSE-Wide the tool will probably not reach a wider audience. Nevertheless, it was an extremely educational and fun process and putting the time and effort into developing the GUI definitely improved the quality, speed and enjoyment of the otherwise tedious process of classifying emission lines.



Equivalent Width Distribution of Lyman α Emitters with MUSE

Abstract: In this chapter I use the equivalent width (EW_0) to analyse Lyman α lines of high redshift Lyman α emitters (LAEs) and determine the distribution of EWs with a new method. The aim is to better understand the interplay between the Lyman α emission line and the galaxy by obtaining reliable EW_0 distributions, using the advantages of integral field spectroscopy combined with broad-band Hubble Space Telescope (HST) data.

I analyse the emission lines of LAEs detected in the first 60 fields of the MUSE-Wide survey as well as in the MUSE-Deep survey and find UV continuum counterparts in archival HST data. The spectral resolution of MUSE allows for accurate fitting of the Lyman α line to determine the double peak fraction, asymmetry and line flux while taking into account the extended nature of Lyman α for precise measurements of the EW_0 . The high spatial resolution of the HST data makes it possible to gain morphological information of the rest-UV emission of the LAEs to capture the UV continuum even of faint objects containing of multiple components.

The object with the highest measured equivalent width has $EW_0 = 601 \pm 110 \text{ \AA}$, much above what normal stellar population models could produce. I find a fraction of 9% of LAEs with equivalent widths over 240 \AA in the full sample and almost 20% for the MUSE-Wide LAEs from histograms of EWs. However, this method only reflects on the sample limits, not the actual EW distributions, which is why I construct the equivalent width distribution function (EWDF) as a property of the population of LAEs, similar to the luminosity function. Thus when accounting for the full Lyman α emission by not restricting the aperture and by correcting for the sample limitations, the equivalent width distribution can be measured with higher reliability.

3.1. Introduction

Understanding the formation and evolution of galaxies and the first stars is one of the most prominent endeavors in Astronomy today. The question how stars in extremely metal poor environments form can be tackled by observing high-redshift galaxies in the early universe. Most high-redshift galaxies are not easy to observe, though, and it has been predicted (Partridge & Peebles, 1967) that star-forming galaxies at high-redshift will be visible through their high Lyman α radiation.

Young, massive, hot stars in star-forming regions produce hydrogen ionizing radiation. When the hydrogen recombines, the most likely outcome of the recombination process will be a Lyman α photon. The higher the number of such massive, hot stars, the more ionizing photons are produced, which means the star formation rate and initial mass function (IMF) influence the Lyman α radiation as well. It is often assumed that a low metallicity in the gas forming the stars will have an influence both on the IMF and on the ionizing

photons produced in such stars (Schaerer, 2002, Pallottini et al., 2015, Raiter et al., 2010, Sobral et al., 2015). Thus a high star-formation rate, a top-heavy IMF and low metallicity can result in a high number of Lyman α photons and a high equivalent width (Schaerer, 2003), which is usually used to compare the strength of the Lyman α lines with the rest-frame UV continuum.

On their way to the observer the Lyman α photons pass the interstellar, circumgalactic and intergalactic medium (ISM, CGM and IGM, see the introduction in chapter 1). Since the photons are resonantly scattered by neutral hydrogen both in spatial direction and in frequency space, the strength and shape (Verhamme et al., 2006) of the Lyman α line will be affected by the neutral hydrogen column density (Shibuya et al., 2014) as well as the dust content, the morphology and kinematics of the gas (Neufeld, 1991, Gronke & Dijkstra, 2014, Behrens et al., 2014, Behrens & Braun, 2014, Laursen et al., 2013). A high neutral hydrogen column density will cause the Lyman α photons to scatter, which increases their path lengths and thus also the probability to hit a dust grain and be destroyed or to get scattered out of the line of sight (see Dijkstra, 2014 for a review). The morphology of the CGM can also play an important role in the radiative transfer of Lyman α photons and it has been shown that most LAEs have halos of extended Lyman α emission (Steidel et al., 2011, Momose et al., 2014, Momose et al., 2016, Wisotzki et al., 2016).

The IGM, too, can influence the EW of the Lyman α line, which is why LAEs are often used as probes of the epoch of reionisation (Malhotra & Rhoads, 2004, Malhotra & Rhoads, 2002, Stark et al., 2010, Lake et al., 2015). Although it is vital to account for the effects of the IGM on the Lyman α radiation when deriving the intrinsic equivalent width, there is no consensus yet on the best way to do so. Some studies assume an intrinsically symmetric Lyman α line, the red part of which is attenuated by the IGM, which would lead to an apparent EW of half the intrinsic value (Kashikawa et al., 2012). However, it has been shown that the resulting asymmetric line profile can be explained on the basis of the gas kinematics alone (Verhamme et al., 2006) and the Lyman α line is often shifted with respect to the intrinsic redshift (Hashimoto et al., 2015), making the correct treatment of the IGM attenuation difficult.

From analysing the EWs of LAEs, we can learn about the properties of these galaxies. It has often been quoted that there is an intrinsic upper limit of EW of 240 Å (Laursen et al., 2013, Schaerer, 2003). While a lower EW can still be explained by normal stellar populations, anything above that value implies unusual conditions in the star-forming regions of the galaxy. Therefore it is important to estimate the number of high equivalent width LAEs and to look at the high EW part of the EW distribution, in order to test this upper limit and to gain insight into the possible properties and mechanisms at work in LAEs.

In this chapter, I use data from the MUSE-Wide and -Deep surveys, containing a large sample of LAEs over a wide redshift range ($\sim 3 < z < \sim 6$), observed with the MUSE instrument (Bacon et al., 2010) at the VLT. Typically, LAEs are found using the narrow-band technique (e.g. Hu et al., 1998, Malhotra & Rhoads, 2004, Ouchi et al., 2008), which however needs spectroscopic follow-up observations to rule out the possibility of low redshift interlopers (e.g. Erb et al., 2014). MUSE is an integral field spectrograph, though, which makes it ideal for the detection and observation of LAEs (e.g. Wisotzki et al., 2016, Bacon et al., 2015). Therefore, if there is indeed a fraction of objects with rest-frame EW of over 240 Å indicating extreme conditions in these galaxies, I should be able to see them.

3.2. Data

Measuring equivalent widths can be dissected into two parts: Determining the Lyman α line flux and the rest-frame UV continuum flux density. For both measurements I am using different kinds of data, which makes it possible to combine the power of integral field spectroscopy from MUSE with the depths of the HST broad-band photometry. Therefore, special care has to be taken to properly combine the information gained from both data sets, which is discussed in more detail in section 3.4.

MUSE is an integral field spectrograph (Bacon et al., 2010) and as such has both spatial and spectral dimensions. It has a field-of-view (FoV) of one arcmin² and a spatial resolution of 0.2 arcsec. The spectral range covers 4750 Å to 9350 Å which allows for the detection of LAEs in a redshift range of $2.9 < z < 6.7$, as the Lyman α line at a wavelength of 1215.67 Å conveniently shifts into the optical at this redshift range. Since this makes MUSE ideal to study high-redshift LAEs, there are several projects within the MUSE consortium which are aimed at studying their properties. In this thesis I use data from the MUSE-Wide¹ (Herenz et al., 2017, Urrutia et al., 2019) and -Deep (Bacon et al., 2017) survey. Especially the MUSE-Wide survey is aimed at discovering a large number of bright LAEs blindly, while the MUSE-Deep survey uses

¹The data and data products such as cut-outs, mini-cubes and extracted spectra as well as emission line catalogues can be found at <https://musedeep.aip.de/project/>.

HST detections as basis for most of their objects, but includes detections based on emission lines as well.

The two surveys compliment each other, as they have different depths and sizes. The MUSE-Wide survey provides a more shallow approach, with one hour exposure time for each field, but covers a large FoV (60 fields are used in this study, but the total goal is to include all 100 fields eventually), while the MUSE-Deep survey focuses on the Hubble Ultra Deep Field (HUDF, Beckwith et al., 2006) with nine fields of ten hours exposure time each with a final contiguous area of 9.92 arcmin^2 (Bacon et al., 2017). The third and deepest tier in this construct is the ultra deep field UDF 10, located in the deepest part of the HUDF and reaching a total exposure time of 31 hours. Combined these three depths allow for different approaches for example for the luminosity function of LAEs, probing the bright end (Herenz et al., 2019) as well as the faint end (Drake et al., 2017a) but also the equivalent width distribution for UV faint LAEs (Hashimoto et al., 2017), which is a complimentary study to the one at hand.

3.2.1. MUSE-Wide

The MUSE-Wide data, located in the GOODS-South (Giavalisco et al., 2004) and CANDELS-Deep/CDFS (Koekemoer et al., 2011) as well as COSMOS (Scoville et al., 2007) fields, were taken at the European Southern Observatory (ESO)-VLT between September and December 2014 (first 24 fields, see Herenz et al., 2017) and from August 2015 to May 2016 (next 36 fields, Urrutia et al., 2019) during guaranteed time observations (GTOs) of the MUSE consortium. Each MUSE-Wide field is roughly one square arcmin in size and each field consists of four exposures, rotated by 90 degrees and dithered randomly in order to avoid spatial correlations in the instrument. That way spatial pixels (spaxels) at the edges of the FoV have less exposure time, which is why the fields overlap by $\sim 4''$.

3.2.2. MUSE-Deep

As mentioned above, the MUSE-Deep survey consists of the ten hour deep mosaic fields and the 31 hour ultra deep field (however, in this thesis 'MUSE-Deep' always refers to both the mosaic and ultra deep fields). The MUSE-Deep data was acquired using GTO time as well, between September, October, November and December 2014, August, September, October, and December 2015 and February 2016 (Bacon et al., 2017). Just as for the MUSE-Wide data, the MUSE-Deep fields were also dithered (with an offset of $2''$) and taken in multiple exposures with 90 degree rotations, but the integration time was 25 minutes instead of 15 minutes.

3.2.3. Photometry

The integral field spectroscopic data from MUSE was used to detect the LAEs and measure the line properties such as line flux, FWHM, asymmetry and peak separation (in case of a double peak, which is typical for Lyman α lines). However, even for the MUSE-Deep data, the exposure time is not long enough to reliably measure the rest-frame UV continuum of the LAEs, which is needed to for the determination of the equivalent widths. Therefore, I use deep, broad-band data from HST at wavelengths longer than the redshifted Lyman α line so as not to contaminate the bands. I use the bands Advanced Camera for Surveys (ACS) F606W, ACS F775W, ACS F814W, WFC3 F125W and WFC3 F160W for measuring the UV continuum flux density and the band ACS F814W to determine the UV continuum counterparts. I then fit them using `GalFit` (Peng et al., 2002, Peng et al., 2010) where possible, since many LAEs (especially in the MUSE-Deep survey) do not have visible HST counterparts (see e.g. Marino et al., 2018, Bacon et al., 2017, Maseda et al., 2018). From the `GalFit` fits I obtain the magnitudes as well as morphological properties such as sizes and axis ratios.

3.3. Sample Selection

Integral field spectroscopy is ideal for detecting LAEs and we use the full two-dimensional spatial and combined spectral information of MUSE to detect emission line objects using `LSDCat`² (see Herenz & Wisotzki, 2017) for MUSE-Wide, a powerful tool aimed at finding emission line objects in integral field spectroscopic data using a matched-filtering approach (see Herenz & Wisotzki, 2017 and Herenz et al., 2017 for a more detailed description) to maximise the signal-to-noise (S/N). For the first 24 fields, we used

²LSDCat is available via the Astrophysics Source Code Library: <http://www.ascl.net/1612.002>

an S/N detection threshold of 8 (effectively 6.4 with the new effective noise description, see [Urrutia et al., 2019](#)), for the next 36 fields we lowered the threshold to 5. The emission lines were then grouped together by LSDCat into individual objects where the line with the highest S/N is denominated the 'lead line'.

3.3.1. Classification of Emission Lines

After detecting all emission line object candidates, the next step is to classify the objects, which is done with the help of the custom GUI `QtClassify`, described in detail in this thesis in chapter 2 (see also the appendix of [Herenz et al., 2017](#) and online at <http://ascl.net/code/v/1628>). `QtClassify` enables the user to load a MUSE datacube, a catalogue created by LSDCat and ancillary HST data if needed. In case of an object with multiple emission lines that can be matched with a list of line wavelengths while assuming a redshift, the classification is done automatically and can be verified by the user. Each unclear emission line can be examined spatially and spectrally and through assumptions of the redshift of the object, positions of other possible lines are shown. Clues that the detected emission line is Lyman α are the lack of any other emission line as well as the line shape, which is typically asymmetrical and sometimes has a double peak. In cases of doubt we also use additional broad-band data to see if the object drops out in bands redwards of Lyman α . For a detailed discussion of the procedure see chapter 2 and [Herenz et al. \(2017\)](#) and [Urrutia et al. \(2019\)](#).

For the MUSE-Deep survey, I use the catalogue produced by [Inami et al. \(2017\)](#), which in contrast to the MUSE-Wide approach is not based solely on emission line selection. Instead, several different methods were used to construct a catalogue, including an automated emission line detection software similar to LSDCat based also on a matched filtering approach (ORIGIN, Mary et al. in prep., [Bacon et al., 2017](#)) and MUSELET ([Piqueras et al., 2017](#)) based on SExtractor ([Bertin & Arnouts, 1996](#)) and using narrow-band images created by collapsing 5 wavelength layers and subtracting the continuum ([Inami et al., 2017](#)). To classify the detected emission lines, the software MARZ ([Hinton et al., 2016](#)) was modified and the classification was done semi-automatically.

The total sample of LAEs used in this chapter consists of 1465 LAEs, 668 of which come from the MUSE-Wide data (with 237 of those from the initial 24 fields, [Herenz et al., 2017](#), the other 431 from the next 36 fields³) and 797 from the MUSE-Deep survey⁴. I thus get a number of ~ 11 LAEs per field in the MUSE-Wide survey (and therefore also roughly per square arcsecond or even more, since the fields overlap slightly). I build on the work of [Hashimoto et al. \(2017\)](#), who construct an equivalent width distribution with the MUSE-Deep survey LAEs only and I add the MUSE-Wide detections to get better statistical constraints and probe a larger field-of-view and a larger range of Lyman α and UV luminosities. To keep the method consistent, I derive the UV continuum counterparts and flux for the MUSE-Deep as well as for the MUSE-Wide objects.

3.3.2. Counterparts

After obtaining the LAE sample, I determine the UV continuum counterparts of our LAEs in the HST data using the band ACS F814W for each object where possible (both for MUSE-Wide and -Deep). Any object within a radius of $0''.5$ was taken into consideration as a counterpart (as this distance was found to be the 3σ positional error in [Herenz et al., 2017](#)). This distance criterion is used as a starting point, as it is expected that not all Lyman α emission lines we find in the MUSE data will have a UV continuum counterpart that is bright enough to be visible in the HST data. Therefore using the closest HST object as a counterpart for the found LAE would be insufficient. In case of more than one counterpart candidate within the $0''.5$, additional HST bands were examined to visually determine if the SED matches what is expected from LAEs, namely that there is little to no flux to the blue of the line (since that would be absorbed by the IGM, but see chapter 5 where I search for Lyman continuum) and the spectrum declines towards the red.

The redshift distribution of our sample can be seen in figure 3.1, as well as the number of objects with no counterparts in the HST data, the number of objects with a counterpart that has a simple profile and the number of objects with counterparts that are composed of multiple components. The drop in the number of LAEs at high redshifts is influenced by the increased noise in the red part of the MUSE spectra (see [Herenz et al., 2019](#) for a discussion on the selection function and Lyman α luminosity function). At redshifts above

³This sample already includes 21 COSMOS fields, which were not included in the data release in [Urrutia et al., 2019](#)

⁴[Inami et al., 2017](#) give a number of 692 LAEs with high confidence and an additional 115 LAEs with lower confidence, ten of which I exclude as they have no measurable Lyman α flux

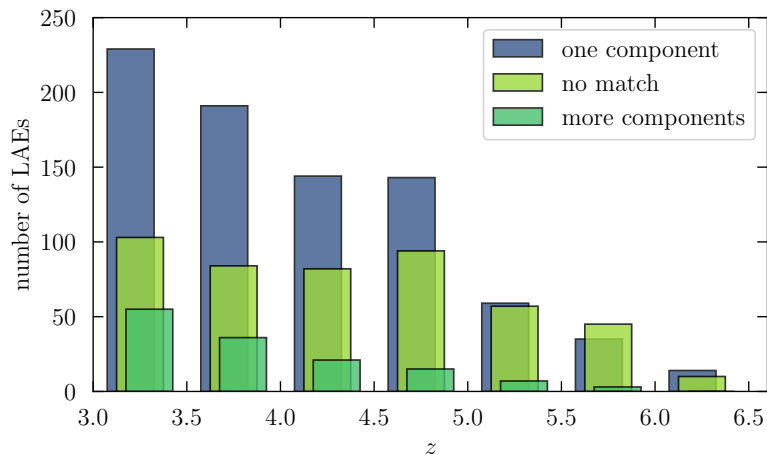


Figure 3.1.: Redshift distribution of the sample of LAEs, divided by their UV continuum counterparts in HST. The blue histogram shows objects with UV counterparts consisting of only one component, meaning the counterpart is either a point source or not resolved into multiple components. The light green histogram shows objects without a visible UV counterpart and the dark green histogram shows objects with UV counterparts consisting of more than one component. In such cases, the HST counterpart could not be fitted with a single Sérsic (Sersic, 1968) profile (see section 4.1), as it was clumpy or asymmetric. Since all counterparts were determined by eye by multiple people, the cases with multiple components are most likely due to mergers or irregular shapes, which are common at such high redshifts. There is still the possibility that some of the LAEs with multiple components are superpositions of different objects, but they will be rare since we check for additional emission lines in the MUSE spectra and for different SEDs in the separate components of the counterparts the HST data.

$z > 5$ the number of objects with and without a UV continuum counterpart is almost the same, while at lower redshifts most objects can be seen in the HST data.

In total, of our sample of 1465 LAEs, 494 do not have a UV continuum counterpart, which is roughly a third of the objects. For those I can only estimate lower limits for the equivalent widths, using the limiting magnitudes in the relevant band as an estimate of the maximum continuum flux density (see Marino et al., 2018, Bacon et al., 2017, Maseda et al., 2018 for studies of 'dark' galaxies not visible in the rest-frame UV continuum). Most objects (826 or 56%) have a counterpart with only one component visible in the HST data, but 145 ($\sim 10\%$) have multiple components. In these cases, I add the continuum flux density of all components for the equivalent widths estimation. The reasons for the appearance of multiple components in the UV continuum counterparts could be either that there are multiple objects in the line of sight at other redshifts which do not belong to the Lyman α line we observe in the MUSE data (but we reject such interlopers as best as possible when determining the counterparts) or the LAEs could be intrinsically clumpy or irregular. This could be caused by 1) a merger that is underway, which could potentially boost the star formation and thus the Lyman α equivalent width, 2) the LAEs could consist of multiple star-forming region and have an irregular shape or 3) there could be multiple objects at close range that all contribute to the Lyman α emission. Galaxies at high redshifts do not necessarily follow the same morphological patterns we see in the nearby universe and even though most of our LAEs are not resolved in HST, it is interesting to see a significant fraction of LAEs at redshifts $3 < z < 6$ being clumpy or possibly merging.

3.4. Determination of Equivalent Widths

To investigate the strength of the Lyman α lines, I measure the equivalent widths. As mentioned above, I need both the Lyman α line flux, the UV continuum flux density at the position of the line and the redshift for the rest-frame equivalent width EW_0 . I measure both fluxes independently, the Lyman α line flux directly in the MUSE data, the UV continuum flux density in the deeper HST broad-band data. This has the advantage that I get good estimates of the continuum even for faint objects that would be too faint to be observable in the continuum of spectra extracted from MUSE.

3.4.1. Line Fluxes

For the MUSE-Wide objects, the line flux was measured using LSDCat (Herenz & Wisotzki, 2017), which measures the line flux in apertures that are multiples of the Kron radius (see Herenz et al., 2017, here I use the flux in three Kron radii) and uses the full three dimensional information from the MUSE cube to include the entire Lyman α line flux. This is important because we know that most (or maybe all) LAEs have extended Lyman α halos (see Wisotzki et al., 2016, Leclercq et al., 2017b, Saust et al. in prep., Wisotzki et al., 2018), on average ten times more extended than the UV continuum counterpart (Wisotzki et al., 2016). This extended emission is often not taken into account when measuring Lyman α equivalent widths if the line flux is measured from slit spectroscopy or narrow-band images where the object is assumed to have the same size in Lyman α as in the complimentary broad-band images. That way, up to 90% of the line flux could be omitted, which shows the strength of using integral field spectroscopy for accurately measuring the equivalent widths of Lyman α lines. Using an aperture of three Kron radii is a good compromise between including a large fraction of the Lyman α line flux and keeping the error in the line flux measurements low. Increasing the aperture would include more line flux, but would also make the measurement more noisy.

For the MUSE-Deep LAEs, I use Lyman α line flux measurements from spectra obtained by summing over the segmentation map from Rafelski et al. (2015) which were convolved with the MUSE PSF (see Inami et al., 2017). For all other line properties, such as asymmetry, full width at half maximum (FWHM), double peak fraction and peak separation I use one dimensional spectra weighted by the PSF so as to maximise the S/N in the Lyman α line. This assumes that there are no spatial variations in the line shape properties and the line has the same properties in the halo as in the central part of the LAE. This simplification is sufficient for the purpose of this thesis, but it should be noted that a spatial variation of Lyman α line properties of our high-redshift LAEs is more than likely (see e.g. Erb et al., 2018 for an example at $z = 2.3$).

3.4.2. Continuum Flux Densities

After the line flux, the next ingredient to the equivalent width is the continuum flux density, which I measure myself instead of falling back on existing catalogues, since some of the UV continuum counterparts we identified were rather faint and not catalogued before or I combine multiple components for one LAE. I also avoid using simple circular apertures so as not to lose any continuum flux. Instead, I fit our galaxies in the HST broad-bands with a Sérsic (Sersic, 1968) model using Galfit (Peng et al., 2002, Peng et al., 2010), which gives me the magnitudes but also basic shape parameters such as effective radius, Sérsic index and axis ratio, which I compare to the equivalent widths in section 4.3.2 in the methods chapter. I start by modelling the rest-frame UV continuum of our LAEs in the ACS F814W band, which is the deepest in the MUSE-Wide survey area. The shape parameters I gain from this fit are used for the other, shallower HST bands as fixed priors, with only the magnitude as a free parameter to ensure that I am not influenced by noise for faint objects and that I am using the same area to measure the magnitudes in each band. This assumes that the morphology of the LAEs does not change much with wavelength but most of the objects are mostly unresolved anyway.

For the equivalent width, I need the continuum flux density at the position of the Lyman α line (see equation 1.5 on page 13 in the introduction to this thesis). Using the same band the Lyman α line falls in to measure the continuum flux density is not ideal, though, as the Lyman α line would have to be subtracted so it does not contaminate the measurement. Instead, I use the band to the left (blue) side of the band where the line falls in. To correct the measured flux at the effective wavelength of the band to the Lyman α line position, I need to know the continuum slope (β -slope) of the spectrum. For this I use bands to the left of the emission line and fit a simple linear relation to the continuum flux density measurements. Since the measured β -slopes scatter significantly and have large errorbars for faint objects, I use the median value of $\beta = -1.92$ of the entire sample as a fixed value for each individual object (for a more in-depth discussion on the measurement of the β -slope see section 4.1.2 on page 70).

As mentioned above, for a third of our sample I could not reliably fit the continuum flux density from the HST broad-band data due to a lack of a visible counterpart. Any object with a continuum flux measurement below 3σ of the limiting flux in the respective band (see section 4.1.1 for the limiting magnitudes in each band) was classified as not detected and the limiting flux for a 3σ detection was used instead (those objects are marked in light green in figures 3.1 and 3.3). These LAEs were excluded from determining the median β -slope as well.

A comparison between the continuum flux density measured from the broad-band data and measured directly from spectra extracted from MUSE can be seen in figure 3.2. I measure the continuum between the rest-frame wavelengths of $\lambda = 1230 - 1280 \text{ \AA}$ (which is close enough to the Lyman α line to neglect

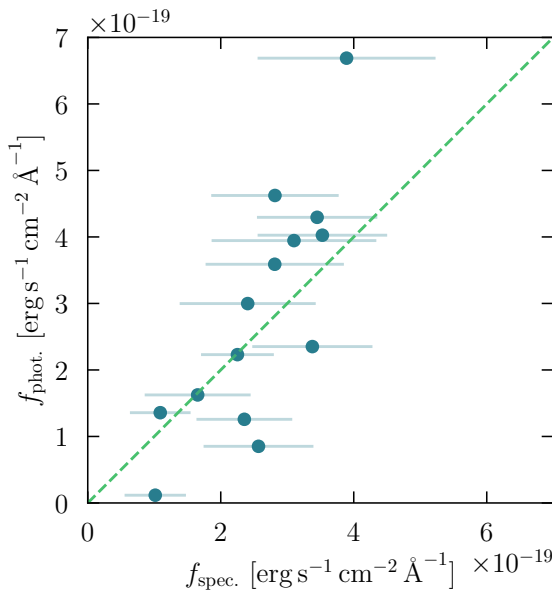


Figure 3.2.: Continuum flux densities at the wavelength of the Lyman α line, measured directly from the LAE spectra (x -axis) compared to the photometric measurements using HST bands (y -axis). The dashed line is the one-to-one relation. The continuum flux density from the spectra is the median value between the rest-frame wavelengths of $\lambda = 1230 - 1280 \text{ \AA}$, detected at more than 2σ .

the β -slope correction) and show any objects where the continuum could be measured with a significance above 2σ . As can be seen, most LAEs have continuum fluxes that are not visible in the MUSE data and only 15 objects were bright enough for this comparison. Most of them are from the MUSE-Deep sample, because of the longer exposure time which increases the S/N in the UV continuum of the spectrum. There is a clear correlation between the two methods of measuring the continuum flux density, which indicates that our method using broad-band data gives the same results as I would get from using the spectra directly (if they were deep enough to detect the continuum in each object).

3.4.3. Distribution of Equivalent Widths

The distribution of equivalent widths is shown in figure 3.3 with an exponential fit to determine the scaling factor. Objects where only a lower limit for the equivalent width could be measured are excluded from the fit but shown in the histogram as the light green bars. Table 3.1 gives an overview of the number of objects in the full sample as well as in the MUSE-Wide and -Deep samples that have equivalent widths above $EW_0 > 100 \text{ \AA}$ and $EW_0 > 240 \text{ \AA}$. The strength of the MUSE-Wide survey in detecting extreme LAEs is evident when looking at the statistics: Including the limiting EW_0 , almost 20% of LAEs in MUSE-Wide have EW_0 higher than predicted by normal stellar population models (which set the limit at 240 \AA), while only $\sim 1\%$ of objects in MUSE-Deep have such high equivalent widths.

It is interesting to look at this discrepancy in the fraction of high equivalent widths between the two surveys more closely. It could be either because of how the different surveys were constructed or because of an intrinsic property of LAEs. MUSE-Wide and -Deep differ in the detection method of LAEs, the size of their footprints and their respective depths. While the MUSE-Wide sample was detected on the Lyman α line, the MUSE-Deep sample was constructed mostly from broad-band photometry catalogues with additional searches for Lyman α emission. However, using LSDCat on the deepest MUSE-Deep field, the UDF 10, and comparing it to the catalogue by Inami et al. (2017) used here showed only minor differences between the detected objects. The two procedures of object detection are therefore comparable and do not explain the difference in the high equivalent widths fractions in the two surveys.

Another difference between the surveys is the larger footprint of the MUSE-Wide survey. Especially for the rare LAEs with very high equivalent widths, low-number statistics could have led to a chance abundance of such objects in the MUSE-Wide fields and a lack in the MUSE-Deep fields. However, as can be seen in figure 3.4, while there are indeed fewer objects with high EWs in MUSE-Deep, there are also more LAEs with low EWs in the deeper survey, which eventually dominates the low fraction of high

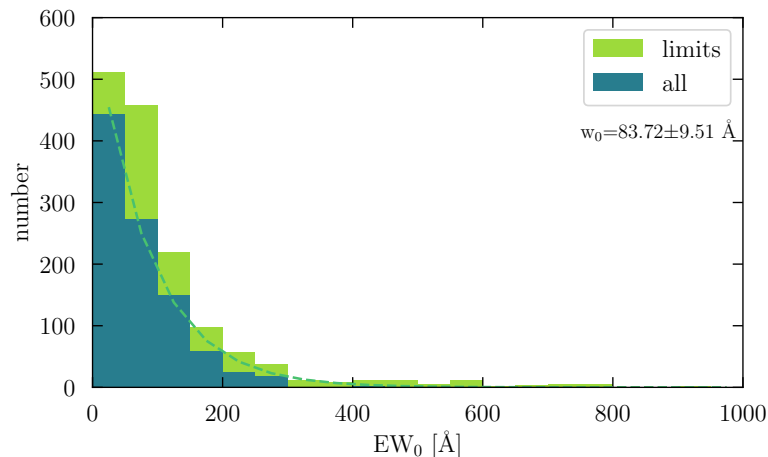


Figure 3.3.: Histogram of the Lyman α equivalent widths of the full sample, including MUSE-Wide and -Deep objects. The blue histogram shows the objects with a secure measurement, the light green shows the objects with limiting EW_0 . However, the latter were not included in the exponential fit $N = N_0 \exp(-EW_0/w_0)$ that is shown as the dashed line. The bin width is 50 \AA and the scale factor w_0 is measured as $83.72 \pm 9.51 \text{ \AA}$.

	total (with limits)	MUSE-Wide (with limits)	MUSE-Deep (with limits)
# of measurements	971 (1465)	490 (668)	481 (797)
$EW_0 > 100 \text{ \AA}$	254 (496)	208 (367)	46 (129)
$EW_0 > 100 \text{ \AA}$ in %	26.2 % (33.9 %)	42.4 % (54.9 %)	9.6 % (16.2 %)
$EW_0 > 240 \text{ \AA}$	25 (132)	23 (125)	2 (7)
$EW_0 > 240 \text{ \AA}$ in %	2.6 % (9.0 %)	4.7 % (18.7 %)	0.4 % (0.9 %)

Table 3.1.: Overview of the numbers of high EW_0 of the different samples. The numbers are given for objects with secure EW_0 measurements, with the total number (including limiting EW_0) given in brackets.

EW objects in MUSE-Deep. This is due to the different observation times of the surveys. The longer the observation time, the lower the observable Lyman α flux limit. While I also go deeper in the UV continuum, this shift is much smaller, which results in a higher number of detected objects with faint Lyman α fluxes and thus small Lyman α equivalent widths. If there is no correction for the flux limits of the sample, any statement about the fraction of LAEs with high EWs only reflects on the survey construction and not the intrinsic properties of LAEs or the distribution of their Lyman α equivalent widths. Nevertheless it is interesting to mention that I do find a number of high EW objects with $EW_0 > 240 \text{ \AA}$, which is often cited as the limit for normal stellar populations based on [Charlot & Fall \(1993\)](#). The limit of 240 \AA can be increased however for decreasing metallicity, as shown e.g. in [Schaerer \(2003\)](#), meaning that such observed high EW_0 could be explained by unusual stellar populations.

When comparing to the literature it is clear that there is no consensus yet on the fraction of high EW_0 values among high redshift LAEs. There are studies that find a larger fraction of $EW_0 > 240 \text{ \AA}$ than shown here, such as [Malhotra & Rhoads \(2002\)](#) (who analyse 150 LAEs and find a fraction of 60% at $z = 4.5$) or [Shimasaku et al. \(2006\)](#) (30 – 40% at $z = 5.7$), but in addition to the different line flux depths, such studies often correct for the Lyman α absorption of the IGM, which can decrease the Lyman α line flux by as much as half. Studies that do not find high EW_0 are [Gawiser et al. \(2006\)](#), [Gronwall et al. \(2007\)](#) and [Guaita et al. \(2011\)](#) at redshifts $z \sim 3.1$ and $z \sim 2.1$. [Gawiser et al. \(2006\)](#) caution that non-detections on broad-band data used for the continuum flux density can lead to extremely large EW_0 for spurious detections in narrow-band images. This mostly applies to narrow-band selected LAE samples and since the MUSE-Wide and -Deep samples were constructed using the spectroscopic information of MUSE where we can confirm the presence of an emission line (and thus classify the objects correctly), this danger is lower in our study. Nevertheless, there is still the possibility of a few false detections in the MUSE data, which is discussed for example in [Urrutia et al., 2019](#).

Another possible explanation for the lack of high EW_0 in some studies is a redshift evolution of the

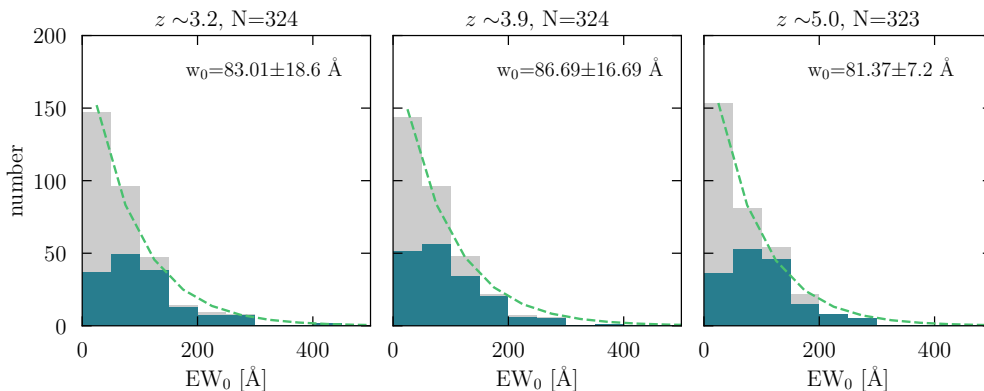


Figure 3.4.: Similar to figure 3.3, but the equivalent width histograms are shown in three different redshift bins. The grey histograms show the full sample of objects (excluding objects with lower limits for the rest-frame EW), the blue histograms show only the MUSE-Wide objects. Each redshift bin is chosen such that the three redshift bins have approximately the same number of objects (given at the top). Each individual histogram is binned with a bin-width of 50 \AA and fitted with an exponential function following the form $N = N_0 \exp(-EW_0/w_0)$ (green dashed line), and the scale factors w_0 are given in each panel.

equivalent widths of Lyman α . Building on the work by Hashimoto et al. (2017), I divide our sample into three redshift bins, with roughly the same number of objects in each redshift bin (see figure 3.4). I thus get histograms of equivalent widths for the redshifts $z \sim 3.2$, $z \sim 3.9$ and $z \sim 5.0$ that are quite similar to each other. In fact, the measured scale factors suggests that there is no evolution with redshift in the distribution of EW_0 (within the errorbars). This result is plausible if we keep in mind that there is also no redshift evolution in the Lyman α luminosity function (shown e.g. for the first 24 MUSE-Wide fields in Herenz et al., 2019 and for the HDFS in Drake et al., 2017b). In combination with the discrepancy in the fraction of high EW objects between MUSE-Wide and -Deep, one should keep in mind that the histograms shown in figure 3.4 are only indicative and useful for comparing with other studies (e.g. Hashimoto et al., 2017), but do not show the real EW distribution, as they are influenced by the survey limits. This is especially obvious when comparing the blue and grey histograms showing the MUSE-Wide and -Deep samples. When looking at the MUSE-Wide sample alone, the lowest EW bins ($0 \text{ \AA} < EW_{0, Ly\alpha} < 50 \text{ \AA}$) have smaller numbers than the next highest EW bins ($50 \text{ \AA} < EW_{0, Ly\alpha} < 100 \text{ \AA}$), clearly indicating the incompleteness in the lowest bin. A more useful equivalent width distribution function that takes into account the selection functions of the surveys is introduced in the next section (3.5).

Notably, the highest securely measured EW_0 I find is at $EW_0 = 600.63 \pm 109.99 \text{ \AA}$, which is a clear indication of an unusual underlying stellar population. Kashikawa et al. (2012) find an even higher equivalent width of $EW_0 = 872_{-298}^{+844} \text{ \AA}$ at a redshift of $z = 6.5$, which could be a hint at a candidate population III galaxy. However, as mentioned above, they apply an IGM correction and assume that half of the Lyman α emission line flux was absorbed by the IGM. Their uncorrected value is at $EW_0 = 436_{-149}^{+422} \text{ \AA}$, even somewhat lower than our highest value. Several studies simulating Lyman α radiative transfer have recently found that the asymmetry of the line can be produced already in the galaxy itself through kinematics and neutral hydrogen density of the ISM (see e.g. Verhamme et al., 2015), shifting a significant fraction of the line flux to the red, main peak, even before the emission reaches the IGM. Therefore I do not correct for IGM absorption, so as not to overestimate the Lyman α line flux and thus the EW_0 .

3.5. The Lyman α Equivalent Width Distribution Function

We have seen above by comparing the MUSE-Wide and -Deep surveys that the standard way of looking at distributions of EWs by constructing histograms mostly reflects the line flux detection limit of the data and not so much the underlying properties of the LAE population itself. An ideal equivalent width distribution should be corrected for the selection function of the survey and it should be possible to extrapolate the equivalent width distribution so as to avoid problems with incomplete bins due to the line flux detection limit. Another benefit of this equivalent width distribution function (EWDF) is that it also offers a new way of estimating the LAE fraction as a function of the equivalent widths. Therefore I show in this section a

new way of deriving the equivalent width distribution with the EWDF, following the same principle as the construction of the luminosity function.

3.5.1. Luminosity Functions

The luminosity function is one of the fundamental properties that describe a population of galaxies and answers the question how many galaxies per co-moving volume (usually in Mpc^{-3}) are in a certain luminosity range.

For the first 24 MUSE-Wide fields, [Herenz et al. \(2019\)](#) construct the Lyman α LF using (among others) the classic $1/V_{\text{max}}$ method in the luminosity range $42.5 \leq \log L_{\text{Ly}\alpha} [\text{erg s}^{-1}] \leq 43.5$. The differential LF giving the volume density $\phi_{1/V_{\text{max}}}$ as a function of the average Lyman α luminosity in a bin $\langle L_{\text{Ly}\alpha} \rangle$ can be written as

$$\phi_{1/V_{\text{max}}}(\langle L_{\text{Ly}\alpha} \rangle) = \frac{1}{\Delta L_{\text{Ly}\alpha}} \sum_k \frac{1}{V_{\text{max},k}} \quad (3.1)$$

in which $\Delta L_{\text{Ly}\alpha}$ is the width of the bin and the sum is running over all sources k in the bin (see [Herenz et al., 2019](#)). For each LAE i in the survey I can define the maximum accessible volume $V_{\text{max},i}$ with the selection function $f_C(L, z)$ which depends on the luminosity and the redshift of the object. The selection function is integrated over the differential cosmological volume element dV/dz and integrated over redshift from the lower (z_{min}) to the upper (z_{max}) redshift limit in the considered range. To get the maximum volume for each LAE, the angular momentum ω has to be multiplied with the resulting integral (for the first 24 MUSE-Wide fields shown in [Herenz et al. \(2019\)](#) this is $\omega = 22.2 \text{ arcmin}^2$).

$$V_{\text{max},i} = \omega \int_{z_{\text{min}}}^{z_{\text{max}}} f_C(L_{\text{Ly}\alpha,i}, z) \frac{dV}{dz} dz \quad (3.2)$$

For the sample of LAEs at hand, the selection function of the next 36 MUSE-Wide fields, as well as for the MUSE-Deep fields UDF 10 and the mosaic fields was derived. The Lyman α luminosity function can then be extended towards fainter LAEs thanks to the inclusion of the MUSE-Deep fields (see figure 3.5). Here I limit the objects to those with a value of the selection function above $f_C > 0.15$, just as in ([Herenz et al., 2019](#)), as the uncertainties of the selection function get unreliable below this value.

3.5.2. Constructing the Equivalent Width Distribution Function

The next step is to construct the EWDF. With the knowledge of the $1/V_{\text{max}}$ values for each LAE combined with the EWs, I can get the number densities of LAEs per EW bin, as shown in figure 3.6. The number density of LAEs with faint Lyman α lines drops at higher redshifts due to observational difficulties, which is why I correct for this. To do so, I compute the sum $s_{V,20}$ of the $1/V_{\text{max}}$ values of the lowest 20th percentile of Lyman α luminosity values and scale the entire equivalent width distribution function by the ratio between the sum of the portion of the sample with the lower redshifts ($2.9 < z < 4.0$) and the sum $s_{V,20}$ for the higher redshifts and for the entire sample. This is done for each of the four different areas (MUSE-Wide 24, MUSE-Wide 36 MUSE-Deep mosaic and MUSE-Deep UDF 10) individually. A simple scaling of the equivalent width distribution is sufficient, since the objects with the low Lyman α luminosity will be spread over the whole equivalent width range, not just the low EW end. After this correction, the equivalent width distribution functions for the three redshift ranges show no evolution, just as the Lyman α LF.

The equivalent width distribution function can be fit in a similar way as the normal luminosity function, with a Schechter function ([Schechter, 1976](#)) where I replace the luminosity with the equivalent width EW_0 . A reasonably good fit is an exponential function, meaning the faint-end slope of the Schechter function is set to zero, $\alpha = 0$, with ϕ^* being the normalisation in Mpc^{-3} and w_0 the characteristic equivalent width or scale factor, similar to the exponential fit I use for the equivalent width histograms in figures 3.3 and 3.4.

$$\phi(\text{EW}_0) d\log \text{EW}_0 = \log(10) \phi^* \left(\frac{\text{EW}_0}{w_0} \right)^{\alpha+1} e^{-\text{EW}/w_0} d\log \text{EW}_0 \quad (3.3)$$

The fitting results for the entire redshift range $2.9 < z < 6.7$ is $\phi^* = 0.115 \pm 0.007$ and $w_0 = 38.0 \pm 3.1 \text{ \AA}$, clearly different from what I measure from the simple equivalent width histograms.

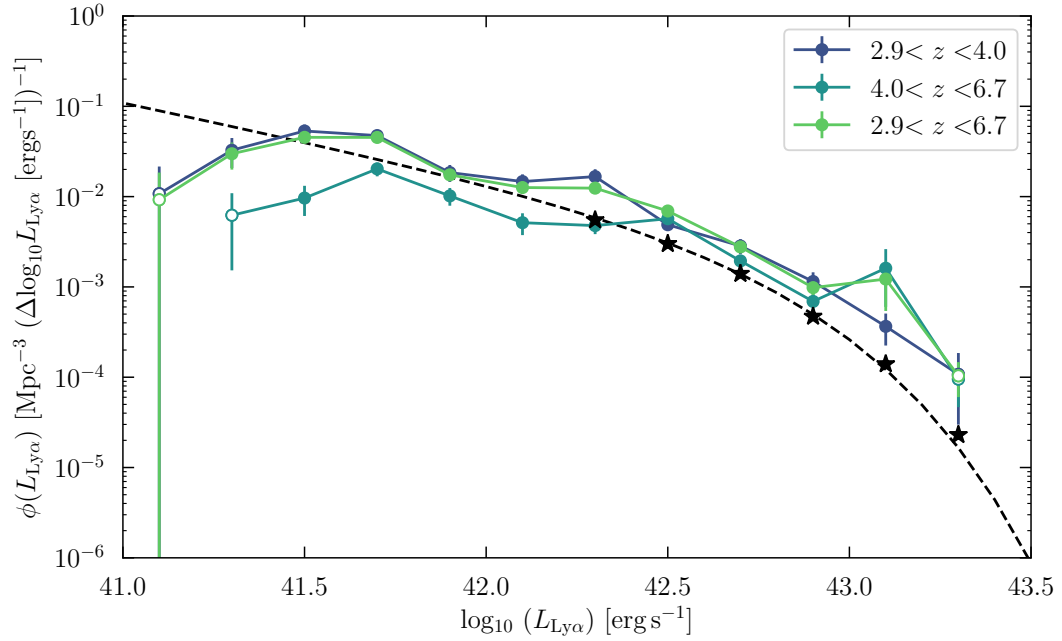


Figure 3.5.: The Lyman α luminosity function for the full sample of LAEs (including the first 60 MUSE-Wide fields as well as the MUSE-Deep mosaic fields and the UDF 10). The different colours show three different redshift ranges, with blue the lower redshift range $2.9 < z < 4.0$, dark green the higher redshift range $4.0 < z < 6.7$ and light green the total sample with $2.9 < z < 6.7$ (see legend). The open symbols indicate bins that are partly empty. The black stars and the dashed line, indicating the Schechter function fit to the data, are taken from [Herenz et al. \(2019\)](#), who show the luminosity function for the first 24 MUSE-Wide fields.

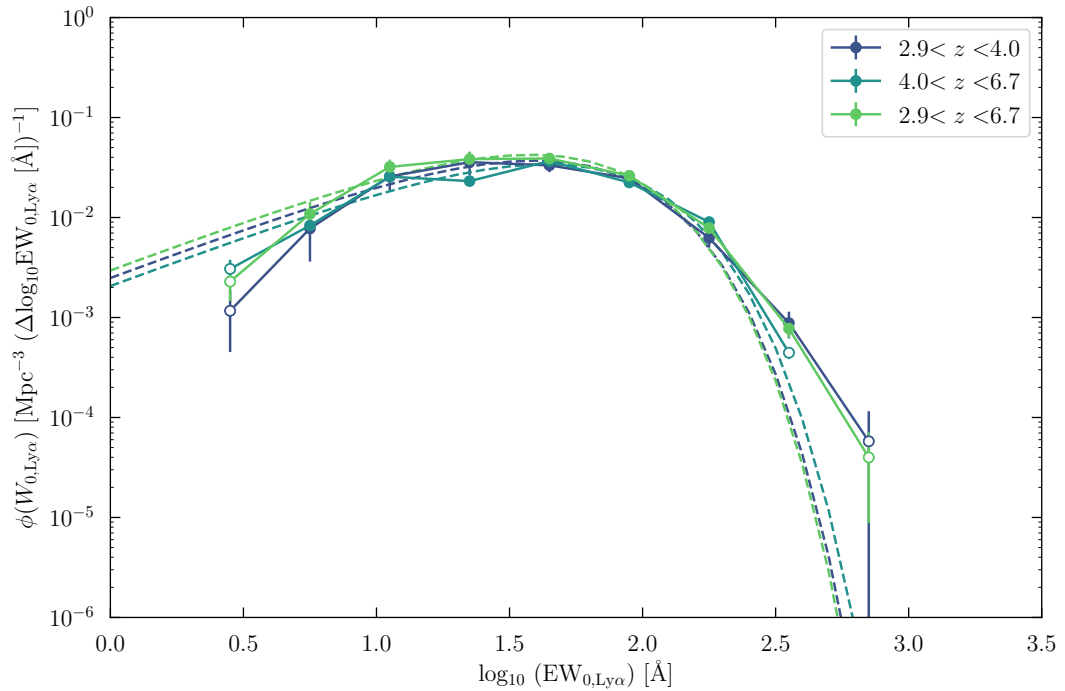


Figure 3.6.: Equivalent width distribution function for the full sample of LAEs (again including the first 60 MUSE-Wide fields as well as the MUSE-Deep mosaic fields and the UDF 10). The different colours show three different redshift ranges (see the legend). The dashed lines show fits to the data.

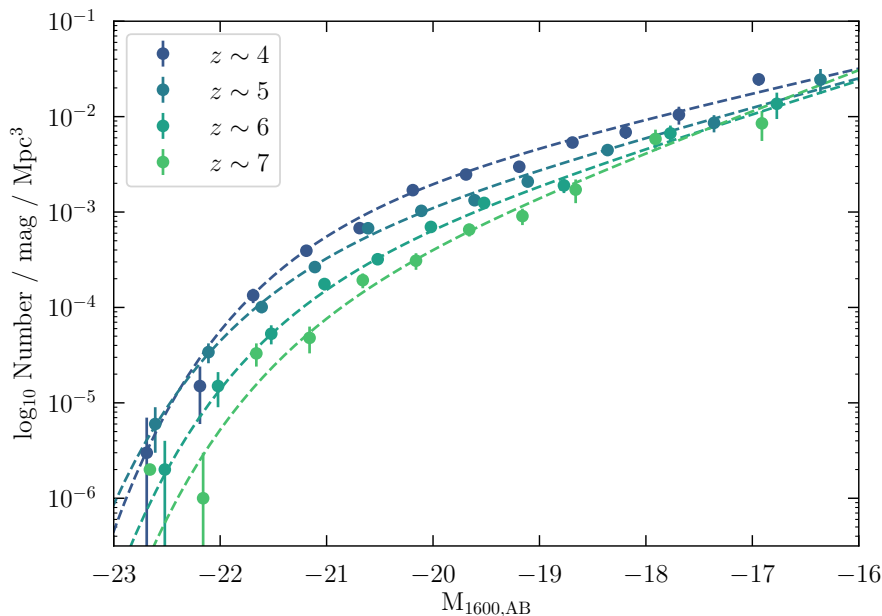


Figure 3.7.: UV luminosity function for redshifts $z \sim 4$, ~ 5 , ~ 6 and ~ 7 , taken from [Bouwens et al. \(2015\)](#). The dashed lines show Schechter fits to the LFs. There is a clear evolution of the number density of galaxies with redshift.

3.5.3. Lyman α Emitter Fractions

With the newly constructed EWDF I can now obtain the LAE fraction (the number of galaxies that emit Lyman α emission compared to the entire galaxy population) by using the UV continuum luminosity function of the whole galaxy population at the same redshifts. [Bouwens et al. \(2015\)](#) have derived the LF (see figure 3.7) at redshifts $z \sim 4$, ~ 5 , ~ 6 , ~ 7 and ~ 8 (the latter being outside the redshift range for LAEs observable with MUSE). In contrast with the Lyman α LF from [Herenz et al. \(2019\)](#) shown in figure 3.5, the UV LF from [Bouwens et al. \(2015\)](#) in figure 3.7 does show a redshift dependent change in the number density with fewer galaxies at higher redshifts.

The EWDF can be integrated up to different equivalent width limits, which in conjunction with the UV luminosity function provides a new method of measuring the LAE fraction at different equivalent width cuts, see figure 3.8. For this figure I integrate the UV luminosity function of [Bouwens et al. \(2015\)](#) (shown in figure 3.7) in the range of $-22 < M_{UV} < -16$ for the redshifts $z \sim 4$, ~ 5 and ~ 6 . The integrated UV LF gives the number density of galaxies at the different redshifts, which is then divided by the integrated EW distribution function. For this I use the EW distribution function of the full redshift range, as it does not change with redshift, as we can see in figure 3.6 (and is a direct result of the invariability of the Lyman α luminosity function with redshift).

With this new way of measuring the LAE fraction as a function of the Lyman α equivalent width I can show that the fraction of LAEs drops with increasing equivalent width, in agreement with e.g. [Stark et al. \(2011\)](#), who looked at $EW_{0,Ly\alpha} < 25 \text{ \AA}$ and $< 55 \text{ \AA}$. The exact values of the LAE fraction depend on the selection function, though, which was adapted from the measured values of the first 24 MUSE-Wide fields and extended to the other fields, where the precise measurement of the selection function is still to be obtained. Therefore the values shown in figure 3.8 are a preliminary proof of concept, but already match well with previous measurements of the LAE fraction with different methods. The lower LAE fractions from [Stark et al. \(2011\)](#) could be due to aperture effects, as the MUSE data captures the full extend of the Lyman α emission which might be missed in slit spectroscopy.

The high LAE fractions for low equivalent widths will be affected by errors in the selection function, but are also a result of the extrapolation of the EW distribution using the EWDF. It should also be noted that the definition of what constitutes a Lyman α emitter also varies. Often a limit of $EW_{0,Ly\alpha} > 20 \text{ \AA}$ or $EW_{0,Ly\alpha} > 25 \text{ \AA}$ is used, which disregards LAEs with lower equivalent widths. However, the high LAE fractions in figure 3.8, especially for the higher redshifts, mainly occur for very low equivalent widths below the common definition of $EW_{0,Ly\alpha} > 25 \text{ \AA}$, which could hint at a possible ubiquity of LAEs among

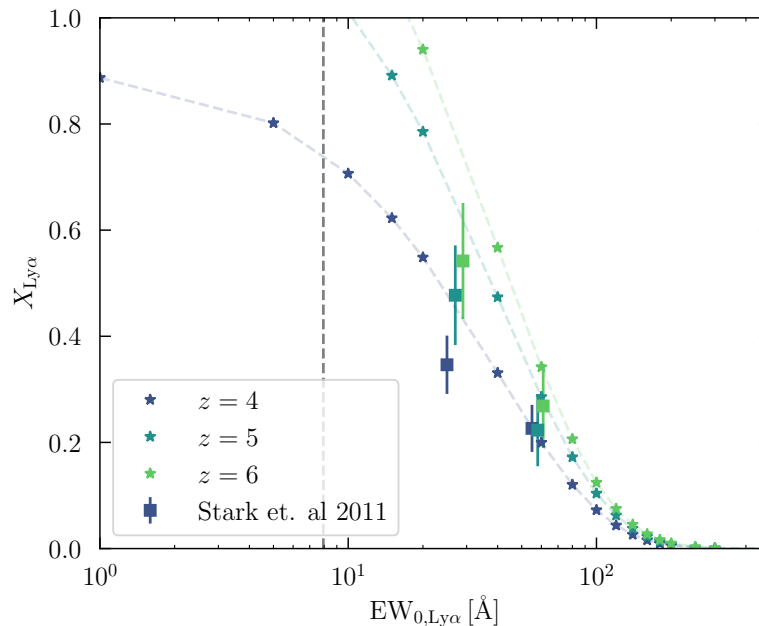


Figure 3.8.: LAE fractions plotted over different $EW_{0, Ly\alpha}$ cuts in \AA . The different colours show different redshifts and the squares show results from Stark et al., 2011 for $EW_{0, Ly\alpha} < 25 \text{\AA}$ and $< 55 \text{\AA}$ and a UV magnitude cut of $-20.25 < M_{UV} < -18.75$ and similar redshifts. The dashed vertical line indicates the limit for extrapolation where the fit to the EWDF diverges from the measured values.

the general galaxy population. The LAE fraction for the first 24 MUSE-Wide fields had already been studied in Caruana et al. (2018) by comparing the LAEs to the HST continuum selected galaxies in these fields. Comparing to this previous study shows that the LAE fractions match well for equivalent width thresholds of $EW_0 > 25 \text{\AA}$, $EW_0 > 50 \text{\AA}$ and $EW_0 > 75 \text{\AA}$, but only go to an LAE fraction of $\sim 60\%$ for $EW_0 > 50 \text{\AA}$ in Caruana et al. (2018). Also, while they find no evolution with redshift, here I show that when using the UV LF for the general galaxy population (which does have a redshift dependence) and the EWDF for the number of LAEs at different redshifts (which does not), I do find a redshift evolution in the fraction of LAEs.

3.6. Summary and Discussion

In this chapter I derive the equivalent widths of 1465 LAEs in the MUSE-Wide and -Deep surveys and construct equivalent width distributions split into multiple wavelength ranges. My main results are

- When matching the Lyman α emission lines to UV continuum counterparts I find that a third of our LAEs do not have counterparts visible in the HST broad-band data. Of the remaining objects, 10% have multiple components in their counterparts, hinting at merging or a clumpy morphology.
- I find a fraction of 9% of the full sample of our LAEs have $EW_0 > 240 \text{\AA}$, which is comparable to other studies in the literature, especially when considering that most Lyman α equivalent width studies apply a large IGM correction, often doubling the line flux and thus the equivalent width measurement. In addition I argue that such fractions are influenced by the Lyman α line flux depth of the survey and should not be taken at face value.
- I find an object with an extremely high equivalent width of $EW_0 = 600.63 \pm 109.99 \text{\AA}$.
- With the equivalent widths and the Lyman α luminosity function I construct for the first time the equivalent width distribution function (EWDF), which is a property of the LAE population independent of the survey construction (such as flux limits).
- I neither find a redshift evolution of the equivalent width distribution function nor of the equivalent width histograms.
- With the addition of the UV luminosity function of Bouwens et al. (2015) I show a new way of measuring the LAE fraction as a function of the equivalent widths.

The large fraction of LAEs with high equivalent widths (over 240 \AA) can be explained not only by the flux limits, but also by the method I used. In the literature, the traditional approach to finding and measuring equivalent widths of LAEs is the narrow-band technique. A narrow-band filter is compared to a broad-band filter and an object that appears only (or has a much stronger signal) in the former is classified as an emission line object. Broad-bands bluewards of the object should contain no signal at the position of the object in order for it to be an LAE. Although this technique is well suited for a large field-of-view and thus a larger sample of LAEs, it has several drawbacks. The spectral shape of the line is not known and it is possible that part of the line flux is outside of the narrow-band wavelength range. This results in lower equivalent widths through lower measured line fluxes. This effect is enhanced if the line falls in the broad-band that is used to measure the UV continuum. Not only will the Lyman α line flux be underestimated, when subtracting it from the continuum band, the continuum flux density will be over-estimated, doubling the effect of underestimating the equivalent width. Integral field spectroscopy does not suffer from this problem while slit spectroscopy suffers from a similar problem in spatial direction.

Another aspect to be considered is the aperture in which the Lyman α line flux and the continuum flux density are measured. Often a simple circular aperture is used for both, which does not take into account the extended Lyman α emission (which can be ten times more extended than the continuum, see [Wisotzki et al., 2016](#)). Apertures with different sizes should be used for the line flux and the continuum flux density in order to avoid aperture losses for the line flux and a low signal-to-noise (due to a large aperture) for the continuum flux density. Both effects add to the reason why in many studies the Lyman α equivalent widths are smaller than shown here.

For future work, it would be interesting to study the SEDs of the objects, since I already measured the magnitudes of the LAEs in several different bands (ACS F606W, ACS F775W, ACS F814W, WFC3 F125W and WFC3 F160W). Modelling the SEDs will give us insights into properties of these LAEs, like stellar masses, ages and metallicity and compare their influence on the equivalent widths distribution function. Since the introduction of the EWDF in this chapter has to be seen as preliminary, there are several ways in which to improve the method. Apart from refining the selection functions for the remaining MUSE-Wide fields and the MUSE-Deep fields, one could also fit a proper Schechter function or even other functions to the data instead of a simple exponential function to further constrain the nature of the EWDF.



Methods: Measuring Equivalent Widths

In chapter 3 I have shown the results of deriving the Lyman α EW distribution function from the MUSE-Wide and -Deep survey in the form of a paper draft. However, there were many steps to the analysis and secondary results that went into this work. In this chapter I want to discuss in detail some of the steps that are only briefly mentioned in chapter 3 and highlight interesting findings that might not be suitable for a paper but went into the analysis.

4.1. Fitting the UV Continuum

As mentioned in chapter 3, for measuring the EWs of the LAEs found in the MUSE data, the rest-frame UV continuum flux density was obtained from the deeper broad-band HST data instead of directly from the extracted MUSE spectra, since they were not deep enough. For this purpose I determine the UV continuum counterparts in the band ACS F814W and fit them with `Galfit`, gaining information not only on the magnitudes, but also the sizes, number of components and axis ratio of the UV continuum counterparts. In this section I will go into more detail on the process of determining the correct counterparts and fitting them with `Galfit`.

4.1.1. Counterparts and `Galfit` Models

To determine the UV continuum counterparts, I look for objects within a radius of 0.5 from the Lyman α position. Each potential counterpart and each component in case there were multiple candidates was examined by three different people (Tanya Urrutia, Kasper Schmidt and myself), using various HST bands (ACS F606W, ACS F775W, ACS F814W, WFC3 F125W and WFC3 F160W). The criteria for counterparts were a reasonably small size (at the redshift of the LAEs, between $z = 2.9$ and $z = 6.7$, we do not expect to be able to resolve e.g. spiral arms), close proximity to the Lyman α emission line coordinates obtained from MUSE, no strong emission in bands bluewards of the Lyman α line and decreasing emission in bands redwards of Lyman α . The emission of the part of the spectrum to the blue side of the Lyman α line should be scattered and absorbed by neutral hydrogen in the IGM once they are redshifted to the Lyman α wavelength and emission at even smaller wavelength, below the Lyman limit at 912 Å, will ionise the neutral hydrogen in the IGM. However, in rare cases there could be a free line-of-sight allowing a fraction of the Lyman continuum to get through (see chapter 5 where I discuss the search for Lyman continuum leaker candidates). Nevertheless, even in these rare cases, the emission is expected to drop significantly in bands to the blue side of the Lyman α line.

Once the counterparts are determined, the objects are fit using `Galfit`, a fitting algorithm created for HST images that is fast and efficient in fitting multiple components at once (see Peng et al., 2002 and Peng et al., 2010). As the LAEs I want to fit are at high redshifts and small even in the HST data, I fit them with

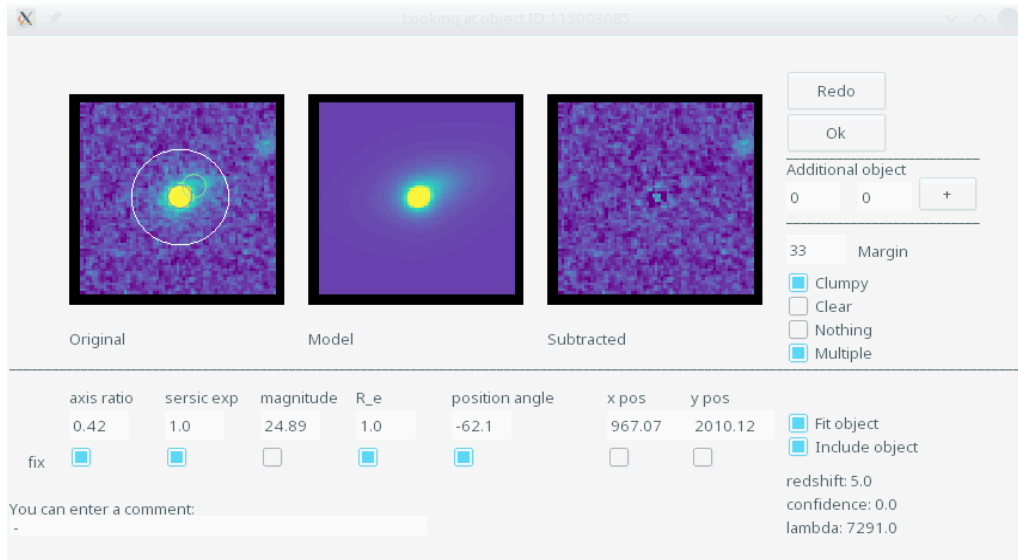


Figure 4.1.: Screenshot of the Galfit wrapper I developed to fit the UV continuum counterparts for the MUSE LAEs. In this example an object is shown in ACS F814W that was fitted with two components (marked as the two smaller circles in the first panel) to better match its shape, which is not perfectly symmetrical but has a bright tail to the top right. Each panel has a size of 1 arcsec^2 (33 pixels in ACS, as shown on the right in the box labeled 'Margin'), the white circle in the first panel has a diameter of $0''.5$. The first panel shows the original data, the second the Galfit model and the third the residual. The fitting parameters axis ratio, Sérsic exponent, magnitude, effective radius (R_e), position angle and coordinates are shown below the panels and can be fixed if needed. If there are multiple components visible or close neighbours that should be subtracted, objects can be added by first clicking on the button on the right labeled '+' and then on the position on the data where the additional object or component is located. For each new object, the user can decide whether to include it in the counterpart of the LAE or just fit it to subtract it from the main fit.

a simple Sérsic (Sersic, 1968) profile (one per component). The equation for the Sérsic profile is:

$$\Sigma(r) = \Sigma_e e^{-\kappa[(R/R_e)^{1/n} - 1]} \quad (4.1)$$

Here, Σ_e is the surface brightness at the effective radius R_e , n is the Sérsic (power-law) index and $\kappa \approx 2n - 0.331$ (Peng et al., 2002) makes sure that half of the total flux will be within one effective radius. The reason the Sérsic profile is often used in modelling galaxies is its versatility. Depending on the Sérsic index n , the profile can become a Gaussian for $n = 0.5$, an exponential profile at $n = 1$ and a de Vaucouleurs profile (de Vaucouleurs, 1948, one of the earliest models for fitting galaxy profiles) with $n = 4$.

To facilitate the fitting, I use a wrapper (see figure 4.1) that automatically generates input files for Galfit, first based on initial priors, then on previous runs, and stores the output data in a coherent way. The procedure is to fit the morphological properties of the objects in the ACS F814W band and then use them as fixed values for the other bands (ACS F606W, ACS F775W, WFC3 F125W and WFC3 F160W) with the magnitude as a free parameter. That way fainter flux at higher wavelength can be captured within the same area as in the deepest band. With the help of the wrapper, close objects that might influence the fit of the LAE can be fit as well, so they do not artificially increase the measured magnitude of the LAE. In the same way, multiple components can be fit and included in the total flux of the UV continuum counterpart. The fit is run again until the model stays stable (meaning the same fitting parameters are recovered each turn). If that is not achieved, I fix some parameters (like the size and axis ratio if the object is too small or the Sérsic index if it has a difficult shape) in order to at least recover the magnitude which is needed for the equivalent widths.

In many cases, there is either no visible counterpart in the UV continuum or only a faint one that was not distinguishable by eye from noise. Since I still fit the magnitude with fixed fitting parameters, I have at least an estimate for each object in each band. To not only rely on determining whether or not there is a counterpart by eye, I use the limiting magnitudes in each band as a cut (see table 4.1 for the limiting magnitudes). Any object not detected at a 3σ level was assigned the limiting flux in the respective band to use as the continuum flux density and this band was excluded from the EW and β -slope estimations.

Band	$\text{mag}_{\text{AB,lim}} (5 \sigma)$	flux (5σ)	flux (3σ)	flux (2σ)
ACS F435W	28.95	0.15	9.22	6.14
ACS F606W	29.35	5.67	3.40	2.27
ACS F775W	28.55	6.99	4.20	2.80
ACS F814W	28.84	4.89	2.94	1.96
ACS F850LP	28.55	5.05	3.03	2.02
WFC 3 F105W	28.45	4.08	2.45	1.63
WFC 3 F125W	28.34	3.22	1.93	1.29
WFC 3 F160W	28.16	2.51	1.51	1.00

Table 4.1.: Limiting magnitudes and fluxes given in $10^{-21} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1}$ for different σ detection values, taken from [Guo et al. \(2013\)](#). For the ACS bands I use GOODS-S V2.0, for the WFC3 I use the CANDELS-deep values. To convert to flux values I use the effective wavelengths given in [Guo et al. \(2013\)](#).

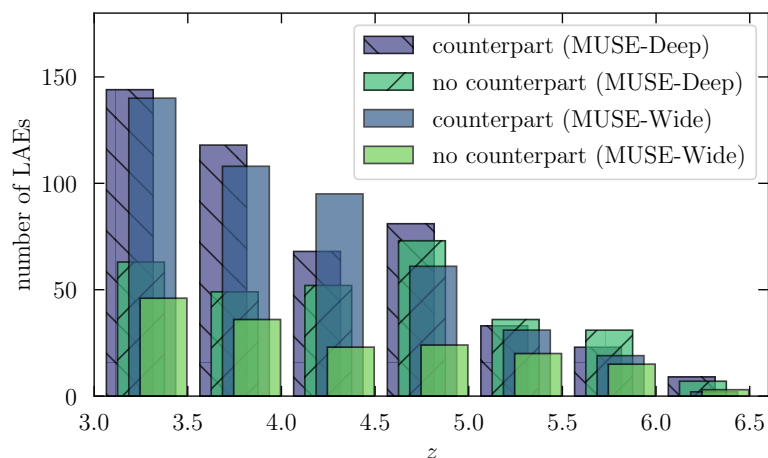


Figure 4.2.: Histogram showing the redshift distribution of the LAEs in the MUSE-Wide and MUSE-Deep surveys (hatched histogram). The samples are divided into objects with UV continuum counterparts (blue histograms) and without counterparts (green histograms).

Objects with no significant detection in any band have lower limits as equivalent widths.

Around 30% of LAEs in the MUSE-Wide and -Deep surveys have no visible UV continuum counterpart. The distribution of objects with and without counterparts in the two surveys MUSE-Wide and -Deep can be seen in figure 4.2. The MUSE-Deep LAEs have fractionally more LAEs without counterparts (compared to LAEs with counterparts) than the MUSE-Wide LAEs, which is caused by the longer exposure time in MUSE-Deep, making it easier to detect objects in Lyman α that are too faint to see in the HST photometry.

Comparison with Catalogues

In order to be consistent with the method I used for the MUSE-Wide equivalent widths, I model the MUSE-Deep objects in the same manner and compare the fluxes to literature values from the Ultra-Violet Hubble Ultra Deep Field (UVUDF) catalogue by [Rafelski et al. \(2015\)](#). `Galfit` gives AB magnitudes m_{AB} , which are converted to flux densities f_{λ} using the effective wavelengths λ_{eff} of the band in question, with $[c] = \text{\AA}/\text{s}$ the speed of light and $[\lambda] = \text{\AA}$.

$$f_{\lambda} = 10^{-\frac{m_{\text{AB}}+48.60}{2.5}} \frac{\text{erg}}{\text{cm}^2} \frac{c}{\lambda_{\text{eff}}^2} \quad (4.2)$$

Figure 4.3 shows a comparison between the flux density values in four different bands for the `Galfit` measurements and the [Rafelski et al. \(2015\)](#) values. At larger flux values, the `Galfit` derived fluxes tend to be slightly smaller than the [Rafelski et al. \(2015\)](#) fluxes. This might be because I fix the shape parameters (and only fit the magnitude) in all bands except ACS F814W, thus possibly losing some of the continuum flux for bright objects. For fainter objects the opposite is the case, as the `Galfit` derived fluxes are slightly

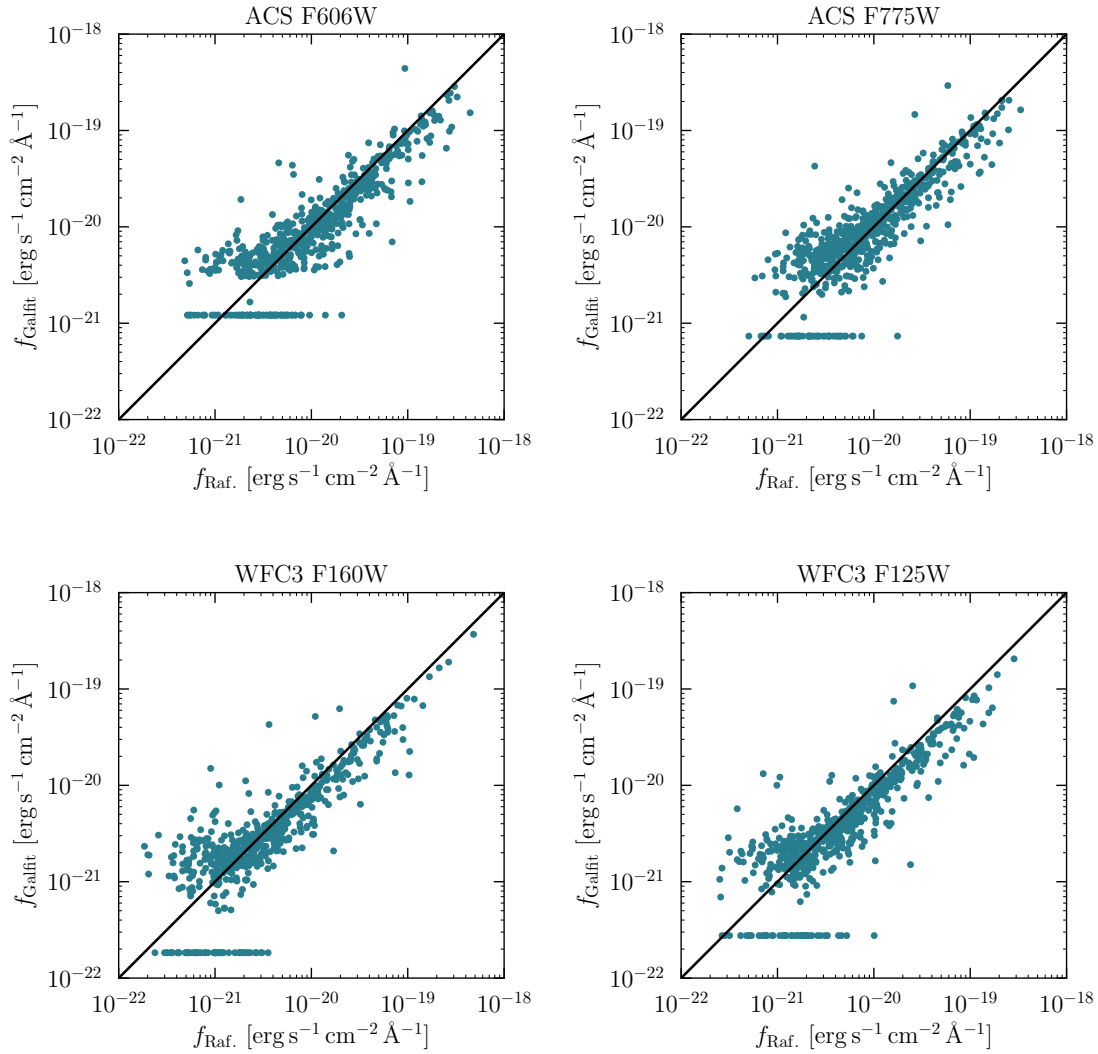


Figure 4.3.: Flux comparison between catalogued UV continuum flux densities (Rafelski et al., 2015) and the measured flux densities for MUSE-Deep obtained from the Galfit fit for four different HST bands. Top left: ACS F606W, Top right: ACS F775W, Bottom left: WFC3 F160W, Bottom right: WFC3 F125W. The flux densities match well for all bands, the line at the bottom of each plot shows the objects where no counterpart could be fit with Galfit.

higher than the Rafelski et al. (2015) ones. Overall the correlation shows that the method using Galfit to derive the continuum flux density works well.

4.1.2. The β -Parameter

From the flux density measurements in the different HST bands I can construct the UV continuum slope (the β -slope) of the objects. This is important for the EWs, for which I need the continuum flux density at the wavelength of the Lyman α emission line, which I cannot measure directly from the spectra (see section 3.4.2). The continuum flux density measured from the HST photometry however is not at the correct wavelength. Therefore I have to extrapolate the continuum to the Lyman α line position. As mentioned in chapter 3 in section 3.4.2 I use a median β -slope of $\beta = -1.92 \pm 0.55$ as a fixed value for all objects. In this section I describe how I arrive at this value and compare it to the β -slopes obtained by Hashimoto et al. (2017).

The β -parameter is the slope of the spectrum, correlating the flux density f at a certain wavelength λ to the wavelength λ .

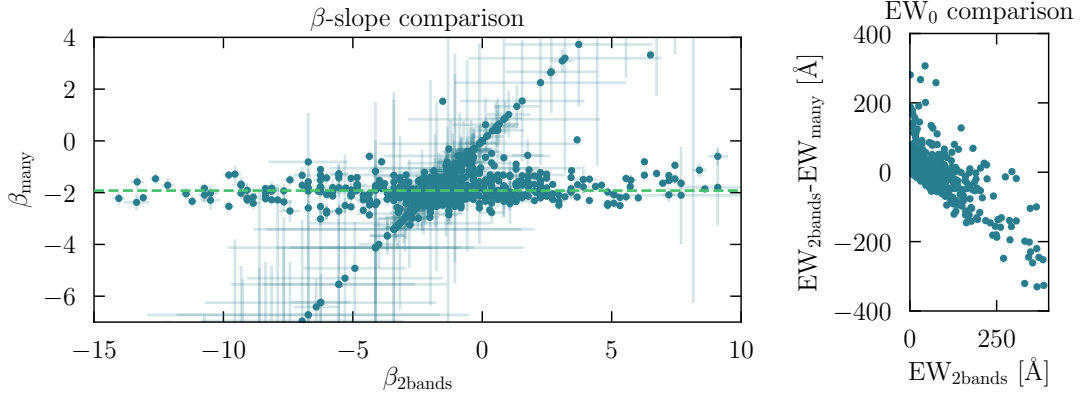


Figure 4.4.: Left panel: Comparison between deriving the β -slope from the two HST bands adjacent to the Lyman α line wavelength (x -axis) and using as many bands as possible (y -axis), usually three or four. The dashed green line shows the median value for the β -parameters obtained using many bands. Dots lying on the vertical line indicate objects where there were only two bands available anyway, which is why the values are on the one-to-one relation. From this plot we can see that using only two bands to determine the β -slope, even if they are close to the Lyman α line, will produce a much larger scatter than using as many bands as I have available. Right panel: Comparison between EWs derived using only the two closest HST bands for the β -slope and using three or four bands. The y -axis shows the difference between both, the x -axis shows the EW using only two bands. Especially at higher EWs, the ones derived from β -slopes using only two bands give EWs that are too small, since the β -slope is usually redder in this case.

$$f_{\lambda} \propto \lambda^{\beta} \quad (4.3)$$

If I know the flux density f at two or more different wavelength λ_1 and λ_2 , I can derive the β -slope.

$$\frac{f_{\lambda_1}}{f_{\lambda_2}} = \left(\frac{\lambda_1}{\lambda_2} \right)^{\beta} \quad (4.4)$$

$$\beta = \frac{\log_{10} \left(\frac{f_{\lambda_1}}{f_{\lambda_2}} \right)}{\log_{10} \left(\frac{\lambda_1}{\lambda_2} \right)} \quad (4.5)$$

If I have two or more bands, I can fit a linear relation to the logarithm of the measured flux density values. With the β -slope, I can now derive the flux density at a different position in the spectrum. If I want to get the EWs, I use the flux density at the position of the Lyman α line. For comparing the rest-UV continuum to the Lyman continuum I also need the β -slope as I use the flux density at 1500 \AA for the UV continuum in this case (see chapters 5 and 6 where the escape fraction of Lyman continuum is discussed).

Using as many bands as possible to constrain the β -slope can decrease the error, but bands at larger wavelength will have a fainter signal, as the spectrum goes down (see stellar population models in section 6.3). If only the two adjacent HST bands are used, the error might be larger, but the slopes could still be more reliable. To test this I derive β -slopes from two bands and from all suitable bands and compare the results in figure 4.4. As can be seen, the spread of values in the β -slopes using only two bands is larger than the β -slopes derived from fitting a linear relation to multiple bands. Values range from $\beta \lesssim -15$ to $\beta \gtrsim 10$, while the spread in values for the β -slopes for multiple bands is only caused by cases where only two bands are available anyway. When comparing to the literature, most literature studies find β -slopes with values around $\beta \lesssim -2$ (Bouwens et al., 2009 and Castellano et al., 2012 for LBGs, but Karman et al., 2017 find even smaller values for faint LAEs at similar redshifts). A larger β -slope indicates a more red spectrum, which indicates dust absorption in the galaxy. In figure 4.4, the median of the β -slope derived from multiple bands is indicated as a dashed line at $\beta = -1.92$ (only including objects where more than two bands were available), which matches well with results found in the literature. To compare with Hashimoto et al. (2017), who derive the β -slopes for the MUSE-Deep survey, I divide my sample in the same three redshift bins (see table 4.2). The values match well within the errorbars, however our β -slopes are on average a little more red (and thus dusty) than the Hashimoto et al. (2017) values.

z range	β_{H2017} (mean)	β_{H2017} (median)	β_{here} (mean)	β_{here} (median)
$2.90 < z < 4.44$	-1.62 ± 0.72	-1.73	-1.85 ± 0.55	-1.91
$4.44 < z < 5.58$	-2.17 ± 1.57	-2.22	-1.93 ± 0.52	-1.99
$5.58 < z < 6.66$	-2.10 ± 1.05	-2.321	(-1.8 ± 1.4)	(-1.82)

Table 4.2.: Comparison between β -slopes from Hashimoto et al. (2017) (H2017) for the MUSE-Deep survey (second and third column) and my own measurements (fourth and fifth column) divided into three redshift bins. For this table I only include β -slope measurements with more than two available HST bands, except for the last redshift bin, where only WFC3 F125W and WFC3 F160W were used (values given in brackets). If I include measurements with only two bands in the other redshift bins, the mean and median slopes become redder and the errorbars more than double. Therefore I exclude them for this comparison.

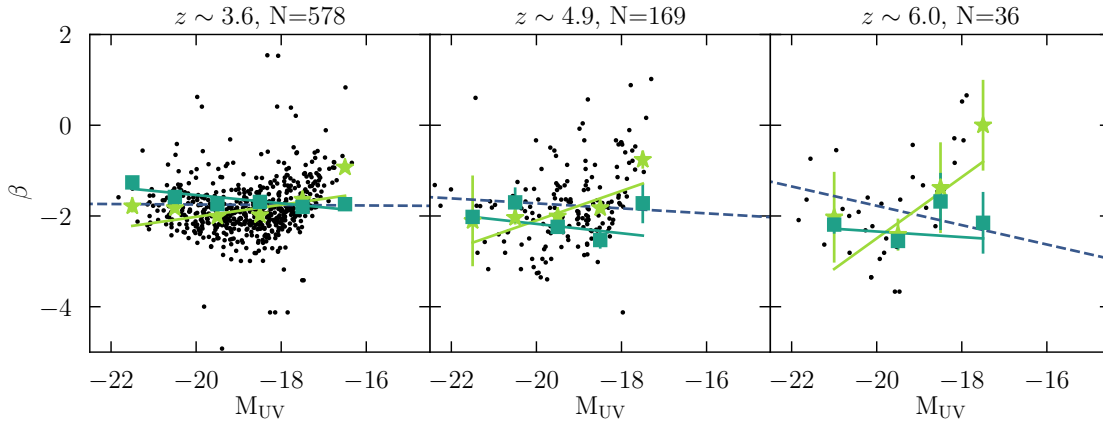


Figure 4.5.: Similar to figure 2 shown in Hashimoto et al. (2017) who use the same three redshift bins (shown in each panel). The black dots show the β -parameters over the absolute magnitude M_{UV} at 1500 \AA for each LAE in MUSE-Wide and -Deep (given that at least two bands to the red of the Lyman α line could be fitted with Galfit). The blue dashed line shows a linear fit to all β -slopes, taking the error for each β value into account. The light green stars show biweight means for bins in M_{UV} , similar to Hashimoto et al. (2017), with errorbars derived from 1000 bootstrapping resamplings where possible. The green squares are values taken from Hashimoto et al. (2017). The redshift bins are the same as in Hashimoto et al. (2017), the number of objects in each plot is written at the top.

In Hashimoto et al. (2017), the β -slope is compared to the absolute magnitude M_{1500} at $\lambda = 1500 \text{ \AA}$ of the LAEs. In order to compare to their results, I also derive M_{1500} through the apparent magnitude m_{1500} and the flux density at this position $f_{\lambda_{1500}}$.

$$f_{\nu_{1500}} = f_{\lambda_{1500}} \frac{(1500 \text{ \AA}(1+z))^2}{c} \quad (4.6)$$

$$m_{1500} = \frac{\log_{10}(f_{\nu_{1500}})}{-0.4} - 48.6 = \log_{10}\left(f_{\lambda_{1500}} \frac{(1500 \text{ \AA}(1+z))^2}{c}\right) / (-0.4) - 48.6 \quad (4.7)$$

$$M_{1500} = m_{1500} - 5\log_{10}(d_L/10\text{pc}) + 2.5\log_{10}(1+z) \quad (4.8)$$

Here, z is the redshift and d_L is the luminosity distance. Using these conversions, I compare the β -slopes to those measured in Hashimoto et al. (2017), see figure 4.5. They use a similar sample of objects as I do, but they are limited to the MUSE-Deep survey, while I am using both the MUSE-Wide and -Deep LAEs. In order to exclude large outliers, I calculate the median in each redshift bin as well as the standard deviation and exclude objects with β values larger (smaller) than the median plus (minus) two times the standard deviation.

A trend has been observed (e.g. by Bouwens et al., 2009, Bouwens et al., 2014 and Castellano et al., 2012) that galaxies with lower UV luminosities tend to have bluer spectra, which is also seen in Hashimoto et al., 2017. In contrast, the β -slope values in figure 4.5 go up again for faint objects. While Bouwens

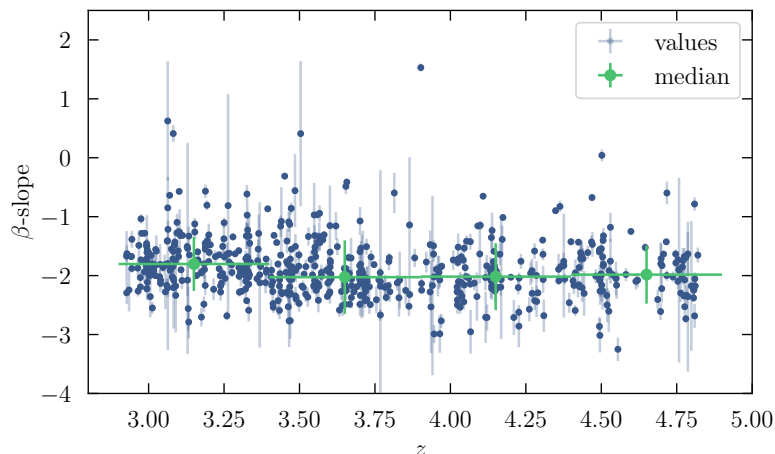


Figure 4.6.: UV continuum slope (β -slope) over redshift (only for values with more than two available HST bands to measure the β -slope). The green dots show the median values in redshift bins with a width of 0.5.

et al. (2014) find that galaxies at magnitudes fainter than $M_{UV} \sim -19$ show a weaker β -slope correlation, the inversion might also be due to a selection effect, as the faintest objects seem to have more flat spectra when fitted with `Galfit`. The infrared bands (WFC3 F125W and WFC3 F160W) used for measuring the β -slope are less deep than the optical bands. Therefore, Hashimoto et al. (2017) only use them for the highest redshift objects. Since most objects have a negative β -slope, faint objects will be even fainter in the infrared bands. Due to noise, faint objects have a measurement error, that could either make them seem more bright than they really are or more faint. If the first, they would appear to have a more red slope, as the signal in the infrared band would increase. If the second, they would not be visible and the infrared band would not be used to determine the β -slope at all, thus the slope would be more blue or the β -slope might not be measurable. This would explain the lack of faint objects with a very negative slope. With these considerations, one would assume that the spread of the distribution towards fainter objects would be asymmetrical, which is clearly the case. The asymmetric faint range is more pronounced in my own measurements (causing the slope of the β -parameter to go up again at faint magnitudes), probably because Hashimoto et al. (2017) only use the infrared bands for the higher redshift bins. While fitting the biweight mean values in the two higher redshift bins do not give the same result as seen in Hashimoto et al. (2017) for the reasons just described, the linear fit to the full data match the expected trend of steeper slopes for fainter UV magnitudes. The trend found in e.g. Bouwens et al. (2009) and Bouwens et al. (2014) that higher redshift objects have steeper spectra (and thus smaller β -slopes) could not be observed, see figure 4.6.

After comparisons with the literature and comparisons between EWs derived from β -slope values using different methods (see figure 4.4) it is clear that the individually measured values are not secure enough. The scatter is mostly due to insecure measurements and not due to intrinsically different spectral slopes. Therefore, I use as the β -slope the median of my measurements (from those objects with more than two available bands) for all LAEs, which is $\beta = -1.92 \pm 0.55$. This also helps with consistency, as this can be used for objects with only one UV continuum band that has a detection and for objects with only upper limits as continuum flux density.

4.2. Morphology of Lyman α Emitters

Since I fit the LAEs with a Sérsic profile using `Galfit` (as explained above in section 4.1), I not only obtain the UV magnitudes of the LAEs but also morphological parameters such as axis ratio, effective radius, position angle and Sérsic index. While it has been shown that properties such as concentration and axis ratio are of limited use when determining the state of the galaxy, e.g. whether it is merging (Abruzzo et al., 2018), it is interesting to see if the internal properties of the LAE that influence the strength of the Lyman α line are connected to its morphology. Figure 4.7 shows a comparison between the rest-frame EW_0 values and the effective Radius R_e , the Sérsic index and the β -slope.

The top left panel shows the size of the UV counterpart, the top right shows the Sérsic index, the bottom left the axis ratio and the bottom right the UV continuum slope (β -slope), each over the rest-frame EW. The

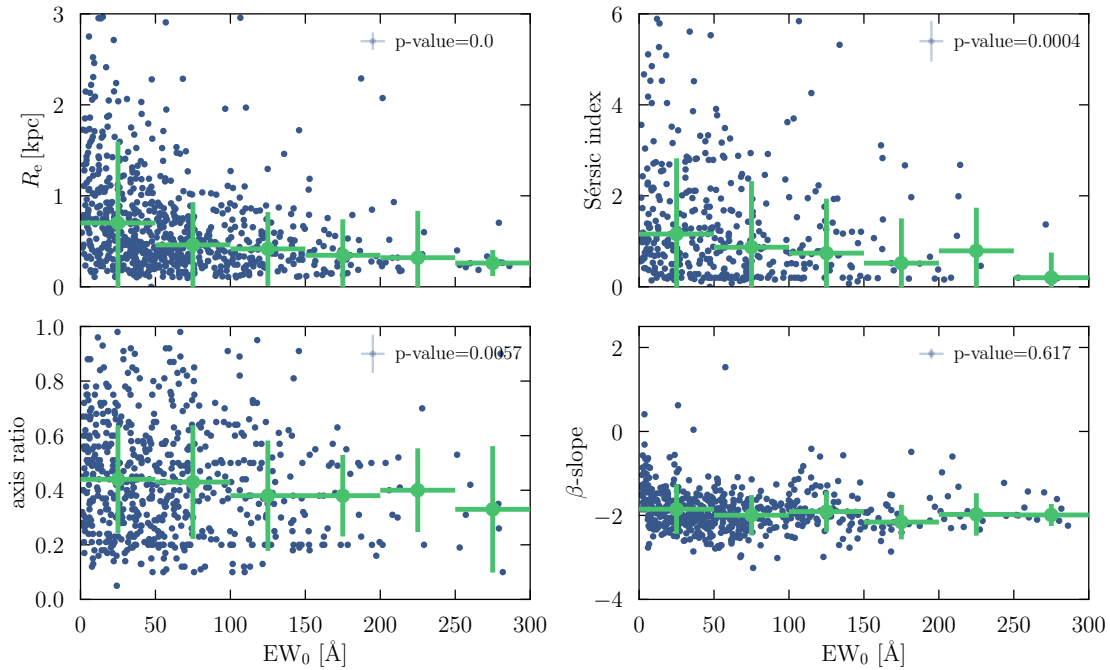


Figure 4.7.: Four different measures of properties of the UV continuum counterparts of the LAEs (measured in the HST photometry) compared to the rest-frame EW of the Lyman α line (measured from the MUSE data). The plots include both the objects from the MUSE-Wide and -Deep surveys. The green dots show the median values in bins of $EW_0 = 50 \text{ \AA}$ (indicated by the widths of their errorbars in x -direction with the standard deviation indicated by the y -axis errorbar). The median error sizes of the measured individual values are shown in grey at the top right corner of each panel (to avoid crowding the plots with errorbars), as well as the p -value of the correlation between the UV counterpart property and EW_0 . In cases of objects with multiple components in their UV continuum counterpart, I use the morphological parameters of the brightest component.

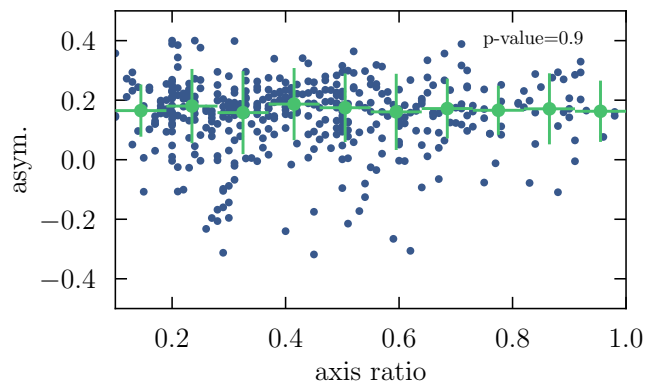


Figure 4.8.: Asymmetry of the Lyman α line (see section 4.3 is compared to the axis ratio of UV counterpart. There is no correlation between the two.

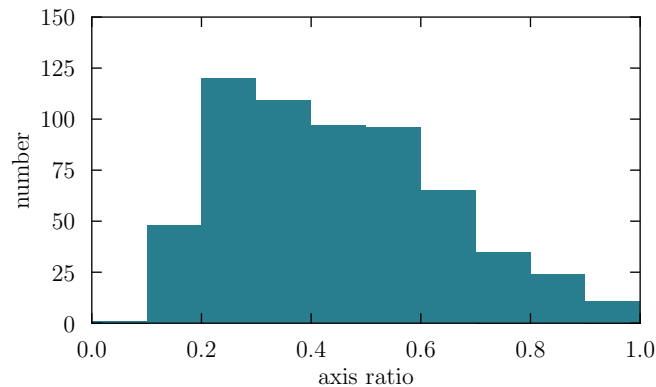


Figure 4.9.: Distribution of axis ratio values. The median of the axis ratio values is at $b/a = 0.42 \pm 0.2$ and the distribution is skewed towards more elongated shapes.

properties are derived from fitting the galaxies with `Galfit` in the HST band ACS F814W (except for the β -slope). The β -slope was measured using the magnitudes bands (mostly ACS F775W, ACS F814W, WFC3 F125W and WFC3 F160W, also fitted with `Galfit`) to the right of the Lyman α wavelength that did not contain the line itself, so as to avoid contamination (as discussed above, see section 4.1.2). However, for the EW, the median β -slope (-1.92) was used instead of the individual measurements, as their scatter is likely caused by noise. There is no correlation between EW_0 and the β -slope, indicating that the dust content of the galaxy might not have an influence on the Lyman α EW.

For the other three panels, there seems to be a slight anti-correlation between the size, the Sérsic index and the axis ratio and EW_0 (albeit with a low significance for the Sérsic index and the axis ratio). [Bond et al. \(2012\)](#) look at LAEs at redshifts $z = 2.1 - 3.1$ and find not correlation between the EW and the half-light radius. The trend we see here could be caused by a bias, as faint UV counterparts are more difficult to fit, but will generally have higher EW_0 (which in itself is probably a selection bias, as can be seen in figure 4.18, discussed in section 4.4.2). For very faint counterparts, the size, Sérsic index and axis ratio were set to a fixed value to try to at least measure the magnitude. Those counterparts are not included in figure 4.7, though.

If I include a cut in magnitude of $M_{UV} < -18$ to minimise the influence of UV faint objects, the correlations persist. Still, fits of fainter UV continuum objects are more easily influenced by noise, making their property measurements more unreliable. Keeping these cautions in mind, the trends could be interpreted in a way that smaller, more compact objects have higher EW_0 . The axis ratio can be used as a proxy of the viewing angle, with more round objects (higher axis ratio) being viewed face-on. It has been proposed that face-on galaxies should have a higher Lyman α escape and thus higher EW_0 ([Verhamme et al., 2012](#), [Behrens & Braun, 2014](#)), as the Lyman α photons do not have to travel through the dusty disc and can escape more easily. This would contradict the observation made here, where the trend points towards the opposite interpretation. It could be that the Lyman α photons are indeed escaping primarily perpendicular to the disc's plane of the LAE, but are scattered back into the line of sight to make them visible from an edge-on view as well. Another interpretation would be that LAEs do not typically show the same morphologies as low-redshift spiral galaxies. They are known to be clumpy, often merging systems (even at low redshifts, as can be seen in the LARS sample, e.g. [Östlin et al., 2014](#), [Hayes et al., 2014](#), [Messa et al., 2019](#)) and the Lyman α morphology does not necessarily follow the UV morphology ([Östlin et al., 2009](#), [Hayes et al., 2014](#)). In figure 4.8 the axis ratio (or viewing angle) of the LAEs is not correlated with the Lyman α line asymmetry, which is itself an indicator of a complex radiative transfer and higher neutral hydrogen column density. [U et al., 2015](#) have found in their analysis that for more symmetric UV morphologies, the Lyman α skewness (similar to the asymmetry shown in this work) has a larger scatter, which is not observed in our data. However, since there is no correlation between the morphology and the line shape, the interpretation of the axis ratio as a viewing angle might not be accurate at these high redshifts, where the galaxies are not well resolved. As mentioned above, high redshift galaxies might not follow the same UV morphology as low redshift galaxies and the axis ratio I measure could be due to irregularities more than the tilt of the LAEs' galaxy plane.

Looking at the values of the morphological parameters themselves, there is a large variation in the Sérsic index values, indicating a variety of morphologies in high redshift LAEs, but the median Sérsic index is at

	total	MUSE-Wide	MUSE-Deep
percentage of objects with blue bump	29.8%	26.5%	39.5%
median EW_0 (no blue bump)	67.6 Å	110.0 Å	50.0 Å
median EW_0 (with blue bump)	74.0 Å	113.5 Å	60.7 Å

Table 4.3.: Overview of the percentages of blue bumps in Lyman α lines and the median EW_0 for MUSE-Wide and -Deep. In the total sample, the fraction of blue bumps is around $\sim 30\%$ and lines with a blue bump have generally a slightly higher EW_0 . The blue bump fraction is even higher in the MUSE-Deep survey, possibly because the longer exposure time makes it possible to more accurately determine the presence of a blue bump. As can be seen in figure 4.10, the blue bump fraction goes up with increasing S/N of the line, which means that blue bumps are often simply not detected.

$n = 1.01 \pm 1.54$, meaning most LAEs are consistent with an exponential profile (as found by e.g. Gronwall et al., 2011 for LAEs at $z \sim 3.1$). The median axis ratio is $b/a = 0.42 \pm 0.2$, consistent with the value of $b/a = 0.45$ found by Gronwall et al. (2011). Another noteworthy result is the distribution of the axis ratios (see figure 4.9), which has a clear maximum and is not flat as seen for local spiral galaxies (e.g. Lambas et al., 1992, Padilla & Strauss, 2008). One should keep in mind, though, that for faint objects the axis ratio derived from `GalFit` is mostly influenced by noise (see Gronwall et al., 2011 for a discussion).

For a future analysis it might be interesting to compare the morphological parameters of LAEs in the UV continuum to that of the Lyman α line itself, since most LAEs have extended Lyman α halos. Not only would it be interesting to see if compact galaxies have a larger Lyman α halo (since they seem to have larger Lyman α EWs as well), but one could also use the axis ratio in combination with the rotation angle of both the UV continuum and the Lyman α haloe to study whether the galaxy plane aligns with the Lyman α emission. This would only be possible for LAEs with large enough halos to measure their axis ratios and also relies on the assumption that LAEs have a rather regular shape both in the UV as well as in the Lyman α emission, which as we have seen, might not necessarily hold.

4.3. Fitting the Lyman α line

As explained in section 1.3.3, the Lyman α emission line profile can be influenced by the ISM to produce a characteristic double peak with a blue bump and a red main line, both usually asymmetric. Several studies (e.g. Verhamme et al., 2015, Verhamme et al., 2017) have predicted and shown that the line shape properties are connected with the Lyman α escape fraction and EW of the LAEs. In addition to this, as we have learned in the introduction to this thesis in section 1.3.3, there is a correlation between the peak separation and the shift of the red peak as well as between the FWHM of the line and the shift of the red peak (Verhamme et al., 2018). Thus both the peak separation and the FWHM can be used to more accurately estimate the intrinsic redshift of the LAE if the Lyman α line is the only one available to measure the redshift from. To access this information, I fit the Lyman α emission lines of our sample with an asymmetric Gaussian and determine which lines have a blue bump. For this I use spectra extracted from the MUSE datacubes that were weighted with the MUSE PSF, to increase the signal in the Lyman α line, but not necessarily recover the full flux information. Again I assume that the line shape is not influenced by the aperture size that is used to extract the spectrum, which might not be the case for all objects, but is a reasonable simplification.

To fit the Lyman α emission line, I use the asymmetric Gaussian function described in Shibuya et al. (2014). They show that when fitting the line with a symmetric Gaussian, the peak wavelength of the line is often not accurately recovered, which makes it difficult to apply the correction for the redshift of the LAEs suggested by Verhamme et al. (2018).

$$f(\lambda) = A \exp\left(-\frac{(\lambda - \lambda_0^{\text{asym}})^2}{2\sigma_{\text{asym}}^2}\right) + f_0 \quad (4.9)$$

Here, f_0 is the continuum level, A is the amplitude and $\lambda - \lambda_0^{\text{asym}}$ is the peak wavelength of the line. The latter was taken from the LSDCat measurements as a first guess. The asymmetric dispersion is σ_{asym} , consisting of $\sigma_{\text{asym}} = a_{\text{asym}}(\lambda - \lambda_0^{\text{asym}}) + d$. Here, d is the typical width of the emission line and a_{asym} is the asymmetric parameter. A positive asymmetry value suggests a line with a red wing, which is the case for most of the (main) Lyman α lines, a negative value means the line has a blue wing. After fitting the line, I derive the FWHM value manually from the asymmetric Gaussian fit and corrected for the spectral

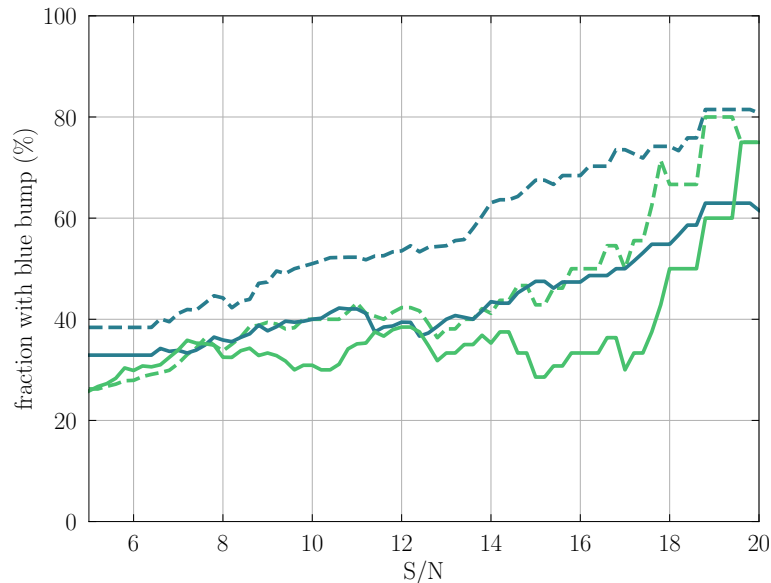


Figure 4.10.: Fraction of objects with a blue bump over S/N. The different dashed and solid lines indicate two different inspectors, one of which was clearly more enthusiastic when looking for double peaks than the other. The different coloured lines show the results for the first 24 MUSE-Wide fields (blue) and for the next 36 (green).

line spread function (LSF) of MUSE. For this I use the value for the MUSE-Deep Mosaic fields (with 10 hours exposure time) given in Bacon et al. (2017), which has a wavelength dependency of the FWHM that follows $F_{\text{mosaic}}(\lambda) = 5.835 \cdot 10^{-8} \lambda^2 - 9.080 \cdot 10^{-4} \lambda + 5.983$ and can be approximated by a Gaussian.

4.3.1. Blue Bump Fraction

The first step to fitting the Lyman α line is determining whether the line is split into a double peak with a visible blue bump. This is done by looking at the spectra and fitting them interactively (with the help of a GUI) with one or two asymmetric Gaussian profiles (see above). Although an automated detection of blue bumps would reduce the human error, it only makes sense when additional information is used (e.g. from models) in order to extract more information from the data than visually possible. An automated determination of the presence of a blue bump in a spectrum without modelling the line shape would thus only use artificial criteria from visual inspection anyways.

The result of the visual inspection is summarised in table 4.3. In total, around 30% of our objects have a blue bump, even more ($\sim 40\%$) if I only look at the MUSE-Deep sample alone. The difference between the two samples could either be the longer exposure time in MUSE-Deep, which could increase the visibility of the blue bumps, or the differences in the S/N calculation of the emission lines could have led to MUSE-Deep objects having more high S/N lines (as can be seen below and in figure 4.10, the S/N influences the blue bump fraction). Objects with a blue bump tend to have a slightly higher EW (see table 4.3). If I assume that most emission lines (more than the 30% I find) intrinsically have a blue bump which is mostly eaten away by the intervening IGM, this could explain why I see higher EW_0 in cases where the blue bump was not absorbed by the IGM. Therefore I cannot rule out the possibility that the blue bump is not connected to a stronger Lyman α emission.

The high fraction of objects of blue bumps matches well with literature results. Kulas et al. (2012) find the same result of 30% for their LAEs at redshift $z \sim 2 - 3$, while Yamada et al. (2012) even find a fraction as high as 50% (for LAEs at $z = 3.1$).

In order to gauge the reliability of the visual inspection performed to determine the blue bump fraction, I use a Monte-Carlo-like simulation by creating 1000 randomised spectra with varying S/N values. The basis for this are ten spectra with intrinsically high S/N ratios, five of them with clear blue bumps and 5 without blue bumps. These spectra are then artificially degraded to lower S/N values. That way, it could be tested to which S/N ratio it is possible to find a blue bump reliably by eye (see figure 4.11). For this test I do not separate into false positives (classifying a single line as a double peak) and false negatives

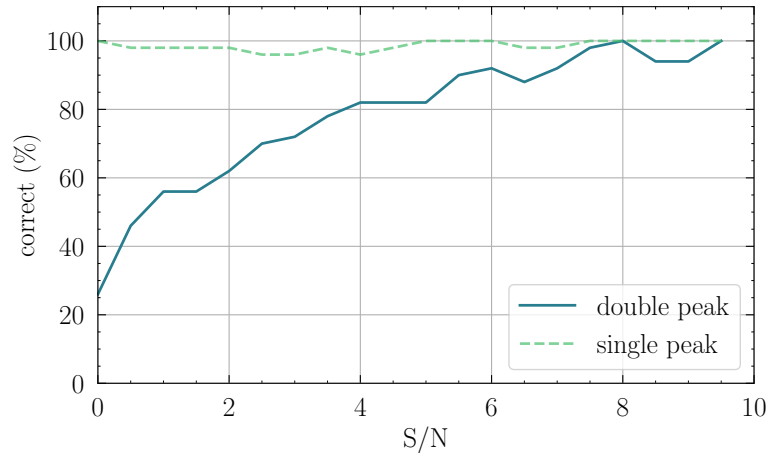


Figure 4.11.: Percentage of correctly identified double (single) peaks shown in dark solid (light dashed) green lines. For lines with a S/N of four and above, 80% of the double peaks can be identified correctly.

(overlooking a double peak), but I only consider whether the line was correctly identified. Even at low S/N values, single peaks can be found with almost perfect accuracy, while double peaks need a S/N of over three for an accuracy of at least 75%. At the S/N values used as a detection limit in MUSE-Wide (6.4 and 5 respectively for the first 24 fields and the next 36) I already reach an accuracy of over 80%.

In order to further test whether the blue bump fraction depends on a potential bias of an overenthusiastic inspector, I compare the results of two independent inspectors¹ using the same scripts to detect blue bumps (see figure 4.10). There is indeed a difference in the fraction of blue bumps found by the two different inspectors of up to 20% for S/N values as high as 14. However, at the lowest and at the highest S/N values the two agree again. There is also a curious and consistent (for both inspectors) difference between the first 24 MUSE-Wide fields and the next 36 fields (indicated by the differently coloured lines). The different S/N detection cuts used for LSDCat should not be responsible for this discrepancy, as this only affects the starting point of this graph (with the blue lines for the first 24 fields starting at S/N = 6.4). We also used a different method to determine the effective noise in Herenz et al. (2017) for the first 24 fields and in Urrutia et al. (2019) for the next 36 fields², but this was also corrected for the first 24 fields (but explains the different detection cuts of 6.4 and 5). Another possible reason could be the different observing conditions of the two sets of fields. The first 24 MUSE-Wide fields in the Chandra Deep Field-South (CDF-S) had a median FWHM of the PSF (fitted with a Gaussian) of $\text{FWHM}_{\text{PSF,G}} = 0.847$, while the next 20 MUSE-Wide fields in the CDF-S had on average a higher seeing of $\text{FWHM}_{\text{PSF,G}} = 1.016$ (computed from table 2 in Urrutia et al., 2019). In the sample in figure 4.10 I also use the 16 additional COSMOS fields. However, the difference in the seeing conditions between the two samples should influence the spatial, but not necessarily the spectral properties and thus might not explain the different observed fractions of blue bumps.

As we have seen, the observed fraction of blue bumps in LAE spectra depends on the S/N of the line and can be as high as 80% for the highest S/N lines. I thus assume that most Lyman α spectra have a blue bump that is usually hidden in the noise or absorbed in the IGM. I use the fits to the two parts of the Lyman α line to correct the line fluxes to include the blue bumps. This is because for the MUSE-Wide sample, the line flux is taken from LSDCat in a certain wavelength window, which not always includes the full blue bump, especially for high peak separations. I therefore re-scale the measured line flux with the ratio between blue bump and main line to include the full line flux in the EW estimation. I do the same for the MUSE-Deep objects' line flux measurements.

In figure 4.12 I compare the ratio between blue bump and total line flux to the rest-frame EW and find that stronger Lyman α lines with higher EWs have a more pronounced blue bump, which matches well with Erb et al. (2014) for LAEs at $z \sim 2 - 3$ (who use a slightly different definition though, as they do not divide the lines into blue and red peaks, but blue- and redshifted emission along the systemic redshift and thus find an even stronger correlation). A strong blue bump can indicate the presence of holes or low neutral hydrogen column densities in an outflowing gas cloud. Both could also facilitate the escape of Lyman continuum

¹Sincere thanks goes to Kasper Borello Schmidt for clicking through over 600 spectra.

²Work is underway to use a fully consistent procedure of data reduction, emission line detection and classification of all 100 MUSE-Wide fields, which should hopefully be completed in 2020.

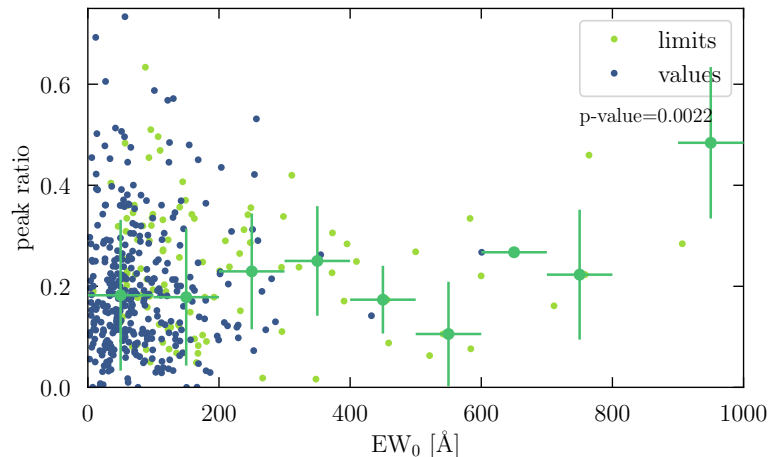


Figure 4.12.: The ratio of the blue bump compared to the total line flux is shown over EW_0 . The blue dots show the measured EW_0 values, while the light green dots show lower limits for EW_0 . The dark green dots and errorbars show the median peak ratio in bins of $EW_0 = 100 \text{ \AA}$. There is a weak correlation between the ratio of the blue bump to the total line flux and EW_0 with higher EW_0 having a more prominent blue bump.

	all			secure			limits		
	mean	median	std	mean	median	std	mean	median	std
peak ratio f_{bb}/f_t	0.21	0.18	0.14	0.2	0.17	0.14	0.25	0.23	0.16
peak sep. [km s^{-1}]	571.6	499.4	348.5	600.7	518.7	366.1	482.8	435.4	269.6
FWHM [km s^{-1}]	217.8	205.9	105.2	234.8	224.9	105.6	184.2	171.6	95.9
asymmetry	0.14	0.16	0.13	0.15	0.16	0.12	0.13	0.15	0.14
EW_0 [\AA]	111.9	69.3	166.7	73.1	55.9	63.3	187.9	98.6	256.4

Table 4.4.: Overview of the mean, median and standard deviation values (to gauge the spread in the distribution) of the measured emission line properties. The first columns give the values for the full sample, the second only for objects with secure EW_0 measurements (objects with a UV continuum counterpart) and the last columns for the objects with limiting EW_0 (without UV continuum counterparts). The first line gives the flux ratio between the blue bump (f_{bb}) and the total flux (f_t including both the red main peak and the blue bump). The second line contains the peak separation in km s^{-1} , the third line the FWHM, also in km s^{-1} , the fourth line gives the asymmetry and the last line the rest-frame equivalent width in \AA .

emission, which would otherwise get absorbed in the neutral hydrogen. Therefore, the strength of the blue bump or the fraction of Lyman α emission that is blueshifted can be used as a proxy for Lyman continuum leakage (see also section 1.4.2 in the introduction).

The median peak ratio (blue bump to total) is $f_{bb}/f_t = 0.19 \pm 0.14$, which means on average the blue bump makes up 20% of the total flux of the Lyman α line. Notably, there are also 17 of 447 lines where the blue bump is stronger than the red peak, which is a sign of infalling material or an outflow with a low neutral hydrogen column density, which enables the Lyman α photons to escape from the near side of the galaxy (Verhamme et al., 2017, Alexandroff et al., 2015, Erb et al., 2010).

4.3.2. The Lyman α Line Profile

In this section I discuss the line shape properties measured from the Lyman α line profile, which are the peak separation, the asymmetry of the line and the FWHM. The measurements of the peak separation and FWHM are used to get more accurate estimates of the redshift of the LAEs (see Verhamme et al., 2018), as discussed in the introduction in section 1.3.3. Since both the peak separation and the FWHM are tracers of the neutral hydrogen column density in the ISM of the galaxy, they show a correlation, as can be seen in figure 4.13.

The mean peak separation of our objects with a visible double peak is $\sim 571.6 \pm 348.5 \text{ km s}^{-1}$ (see table 4.4 for an overview of the mean values for the different line properties), which can be translated approximately

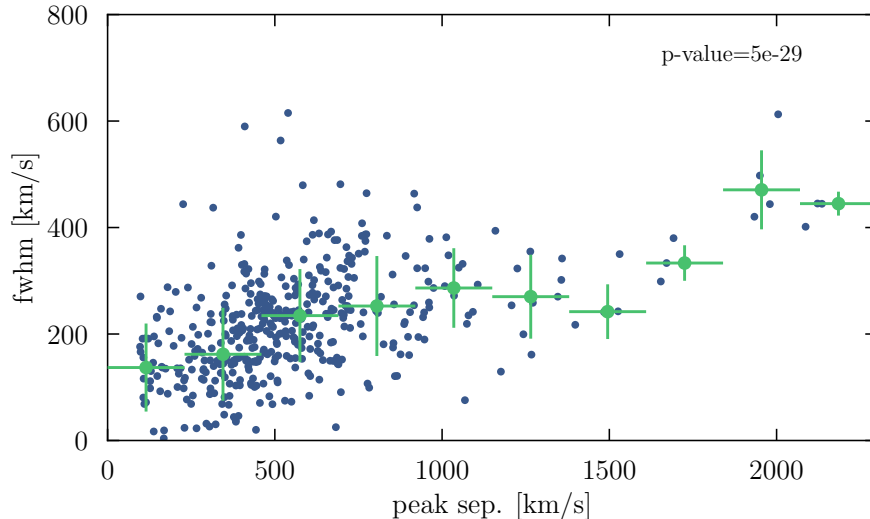


Figure 4.13.: The FWHM of the main (red) peak of the Lyman α line is plotted over the peak separation, both in km s^{-1} . The green dots with errorbars show binned median values. There is a clear correlation between the two values, with lines having higher FWHM have a higher peak separation. This is expected from theoretical models as well (see Verhamme et al., 2015 and observed in Verhamme et al., 2018). The peak separation is a measure of the neutral hydrogen column density, which will scatter the Lyman α photons spatially and in frequency, which in turn results in a broadening of the line.

to a shift of the red peak with respect to the systemic redshift of $\Delta v \sim 285.8 \text{ km s}^{-1}$. In comparison, Shapley et al. (2003) find a shift of $\Delta v \sim 650 \text{ km s}^{-1}$ for a sample of LBGs, Kulas et al. (2012) find values with a mean of $\Delta v \sim 370 \text{ km s}^{-1}$ (for half the peak separation), which is closer to our value. Hashimoto et al. (2013) and Hashimoto et al. (2015) look at LAEs at $z \sim 2$ and find an even smaller offset between Lyman α and nebular lines of $\Delta v \sim 175 \text{ km s}^{-1}$, similar to $\Delta v \sim 180 \text{ km s}^{-1}$ from Song et al. (2014). They find an anti-correlation between the Lyman α EW_0 and the velocity offset, which I do not observe between the peak separation and EW_0 (see figure 4.14). For local LAE analogues, the velocity shift is as small as $\Delta v \sim 150 \text{ km s}^{-1}$ on average (Verhamme et al., 2017). However, I see an anti-correlation between the contribution of the blue bump to the total line flux and the peak separation in figure 4.15. A small peak separation indicates that the Lyman α emission had to scatter less to escape the galaxy, which means the neutral hydrogen column densities in the ISM are low. There is a similar anti-correlation between the asymmetry of the line and the peak separation (see figure 4.16) with larger peak separations showing more symmetric emission line profiles, which matches theoretical models (e.g. Verhamme et al., 2015). However, since the peak separation is connected to the neutral hydrogen column density, one would expect an anti-correlation between the peak separation and the equivalent widths of the lines (see figure 4.14), which is not the case here. Although there is no clear anti-correlation, it should be noted, that the highest peak separations only occur for objects with small EW_0 , while most of the limiting EW_0 are in a regime of lower peak separations.

When it comes to the width of the line, the FWHM, the mean value I find is $\text{FWHM} = 217.8 \pm 105.2 \text{ km s}^{-1}$, which in turn is smaller than what is found by other studies, e.g. the average FWHM of Lyman α lines of galaxies at $z = 2 - 3$ is found to be $\sim 650 \text{ km s}^{-1}$ by Steidel et al. (2010), and Erb et al. (2010) find an LAE with a FWHM of $\sim 850 \text{ km s}^{-1}$. The latter study is analysing a single object, though and among our sample of LAEs, the widest line has a FWHM that is similar ($\sim 844 \text{ km s}^{-1}$). As mentioned, from e.g. Verhamme et al. (2008) we learn that the FWHM reflects the neutral hydrogen column density of the ISM in the LAEs and for expanding shell models, line widths of $\text{FWHM} > 500 \text{ km s}^{-1}$ need neutral hydrogen column densities of over $N > 10^{20} \text{ cm}^{-2}$ and vice versa. This suggests, that most of our objects have small neutral hydrogen column densities. In conclusion, I find that our sample of LAEs have large peak separations on average, making them more similar to values found for LBGs in the literature. Since the distinction between LBGs and LAEs is often an arbitrary cut in EW_0 of $\sim 20 \text{ \AA}$ which I do not make, this result could be caused by the ability to find even faint emission lines in our MUSE data. The width of our lines is rather smaller than found in other studies, which hints at smaller neutral hydrogen column densities. However, the spread of values is large and I do find extreme $\text{FWHM} \sim 800 \text{ km s}^{-1}$.

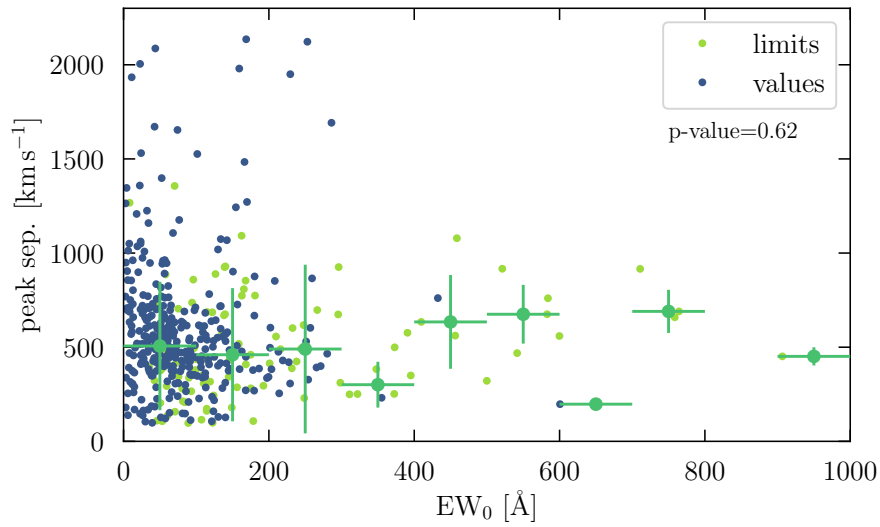


Figure 4.14.: The peak separation (in km s^{-1}) is plotted over the rest-frame equivalent widths EW_0 in \AA . The blue dots show the securely measured EW_0 values, while the light green dots show objects that only have lower limits for EW_0 (with no significantly visible UV continuum counterpart). The dark green dots with errorbars show median peak separation values for EW_0 bins of 100\AA . There is no correlation between the two values, although it would have been expected that objects with larger peak separations would have lower EW_0 (see Verhamme et al., 2017).

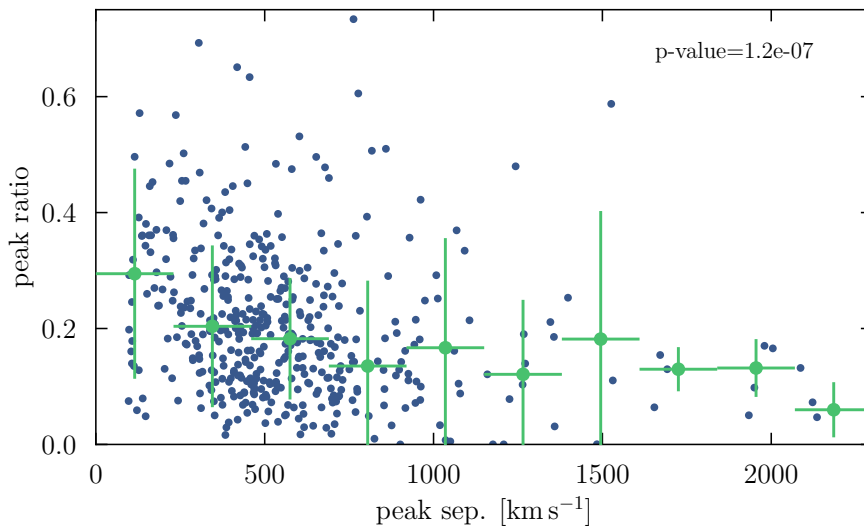


Figure 4.15.: The flux ratio of the two peaks is plotted over the peak separation. The ratio of the peaks is defined as the fraction of the line flux of the blue bump compared to the total line flux (blue bump and red main peak). There is an anti-correlation between the two values, as the peak separation goes down with increasing contribution by the blue bump to the line flux. One reason for this effect could be the IGM absorption of the blue bump, which would decrease its line flux, while the red main peak is shifted out of the resonance frequency. The farther away the blue bump is from the resonance frequency (thus the larger the peak separation), the more likely it is to be reduced by the IGM.

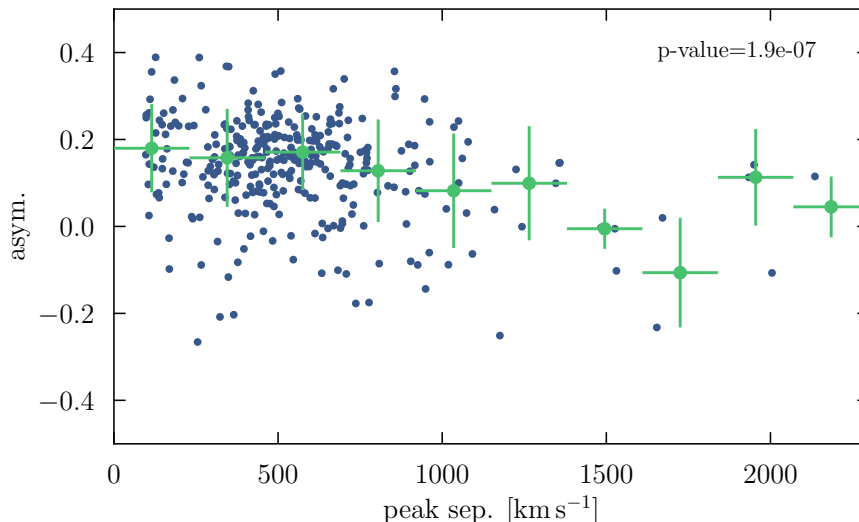


Figure 4.16.: The asymmetry of the line is compared to the peak separation (given in km s^{-1}). The green dots and errorbars again show the median asymmetry in peak separation bins. The asymmetry is defined such that positive values mean the line has a red wing, while negative values mean the line has a blue wing (and 0 means no asymmetry). There is a slight trend that can be seen from the plot, as lines with small peak separations tend to have a higher asymmetry (with a blue wing, which is the most typical line shape for Lyman α), while lines with larger peak separations have smaller asymmetries. This might be caused by the lines also getting broader, which could decrease the asymmetry but also make it harder to measure, since the line shape could be more susceptible to noise.

4.4. Equivalent Widths

The description on how I derive the EW and EW distribution (function) can be found in the previous chapter in section 3.4. Here I want to dig deeper into the dependence of the scale factor on Lyman α luminosity and UV magnitude cuts and compare to values from the literature. I also briefly discuss the occurrence and possible explanation of the Ando effect (see introduction section 1.3.2) in our sample.

4.4.1. Distribution of Equivalent Widths

I measure the EWs from both the MUSE-Wide and -Deep surveys, which have different properties. In figure 4.17 I compare the Lyman α line fluxes between the two surveys. It can be seen that the line fluxes are distributed more widely for MUSE-Wide, while there are more MUSE-Deep objects with smaller line fluxes. They also cluster in some places due to the differences in the detection methods that were used for different ID numbers. There is a slight trend with redshift, as higher redshift objects tend to have smaller line fluxes, which is seen in both surveys. This trend vanishes when looking at the relation between the EW_0 and redshift. There is also a small skew in the distribution of line fluxes in the MUSE-Wide sample in figure 4.17, as higher IDs have smaller line fluxes. This is caused by the different detection cuts in S/N between the first 24 fields (corrected $S/N_{\text{cut}} = 6.4$) and the following fields ($S/N_{\text{cut}} = 5$).

Similar to Hashimoto et al. (2017), I construct the EW histograms (as shown in figure 3.3 on page 60) for different cuts in redshift, UV magnitude ($M_{\text{UV,limit}}$) and Lyman α luminosity ($\log(L_{\text{Ly}\alpha,\text{limit}})$) to compare to other studies. The results can be found in table 4.5. For this I leave out the first EW_0 bin from the plot ($< 50 \text{ \AA}$), as this is usually incomplete and destabilised the fit by increasing the errors. On average, the scale factors from our distributions are smaller than what is found in the literature. Keep in mind however, that as I discuss in section 3.5, the scale factors derived from the histograms are influenced more by the line flux limit of the survey than underlying properties of the population of LAEs.

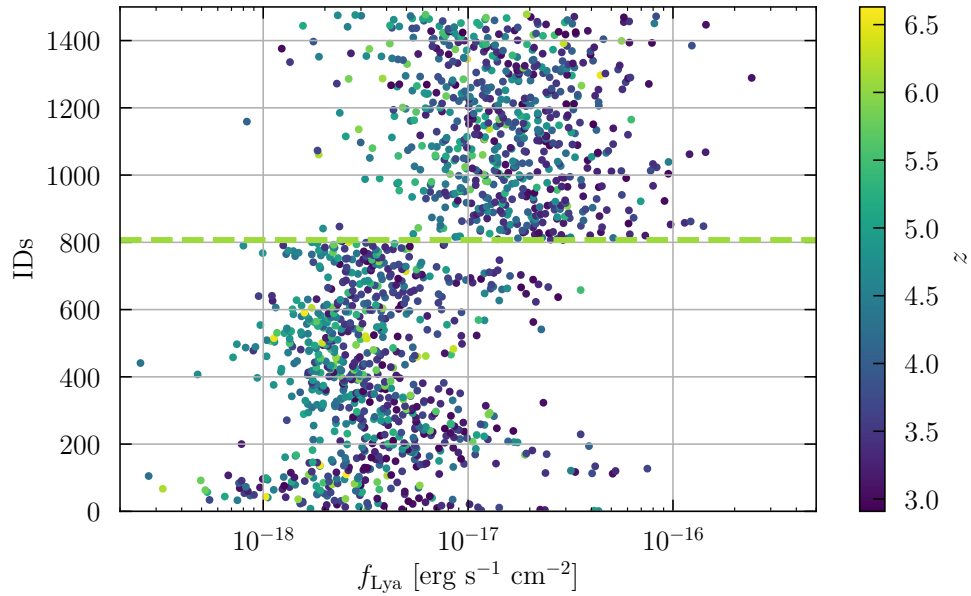


Figure 4.17.: For this plot, the objects were sorted by ID and then plotted with a running number, which means the lower part (below the green dashed line) shows the the MUSE-Deep objects and the upper part (above the green dashed line) the MUSE-Wide objects. Note that not the actual IDs were plotted over the Lyman α line flux, but just the index (since the naming convention in MUSE-Deep and MUSE-Wide are very different). The colour indicates the redshift with dark purple being the lowest redshift ($z \sim 3$) and bright yellow being the highest redshift objects ($z \sim 6.7$).

Study	Redshift	$M_{UV,limit}$ [mag _{AB}]	$\log(L_{Ly\alpha,limit})$ [erg s ⁻¹]	w_0 (study) [Å]	w_0 (here) [Å]
Gronwall et al. (2007)	3.1	-18.0	42.0	75 ± 6	48 ± 4
Ciardullo et al. (2012)	3.1	-18.6	42.0	64 ± 9	35 ± 2
Hashimoto et al. (2017)	~ 3.6	-18.0	42.0	74 ± 19	59 ± 1
Hashimoto et al. (2017)	~ 3.6	-18.6	42.0	60 ± 20	40 ± 2
Zheng et al. (2014)	4.5	-17.0	42.4	50 ± 11	92 ± 23
Hashimoto et al. (2017)	4.9	-17.0	42.4	143 ± 64	91 ± 58
Kashikawa et al. (2011)	5.7	-18.0	42.0	108 ± 20^a	59 ± 4
Kashikawa et al. (2011)	6.6	-18.0	42.0	79 ± 19^a	78 ± 28
Hashimoto et al. (2017)	~ 6.0	-18.0	42.0	157 ± 110	78 ± 28

Table 4.5.: Comparison with scale factors from the literature for different cuts in redshift, absolute magnitude and Lyman α luminosity, similar to table 6 in Hashimoto et al. (2017). The values for Kashikawa et al. (2011) were taken from Zheng et al. (2014).

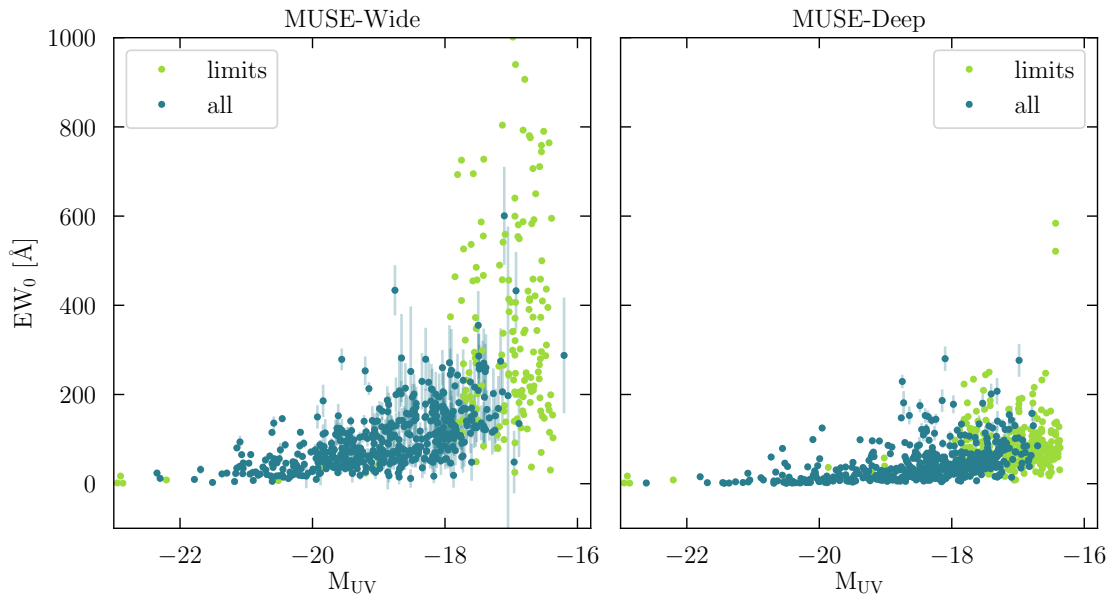


Figure 4.18.: Both panels show the rest-frame EWs over the UV magnitudes. For objects where the UV magnitude could only be given as an upper limit (since there was no UV counterpart visible in the HST data), limiting EWs are shown in green. The left panel shows the first 60 MUSE-Wide fields, the right panel the MUSE-Deep fields. In the latter sample, the EWs are on average smaller and there are more limiting EWs, since the MUSE observations were deeper and weaker Lyman α lines could be detected.

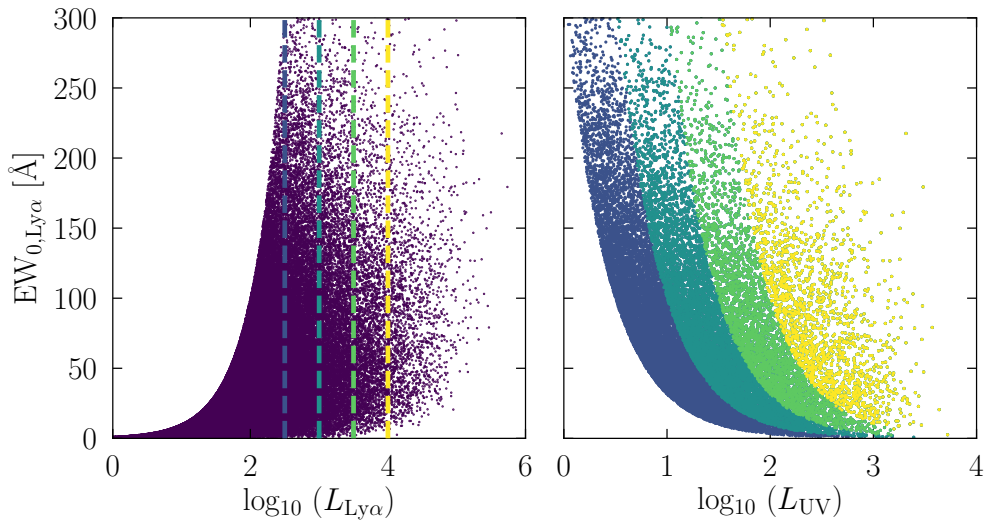


Figure 4.19.: Left panel: Lyman α equivalent widths over Lyman α line luminosity for 10 000 simulated objects (purple dots). The dashed lines indicate different luminosity cuts in the Lyman α line. Right panel: Lyman α equivalent widths over the UV continuum luminosity. The different colours correspond to the different line luminosity cuts in the left panel. This clearly shows that the lower cut-off is caused by the line luminosity limit. Keep in mind that when comparing to figure 4.18, the x -axis here shows brighter objects with increasing values, while for figure 4.18 UV magnitudes were used, which means brighter objects are at smaller values.

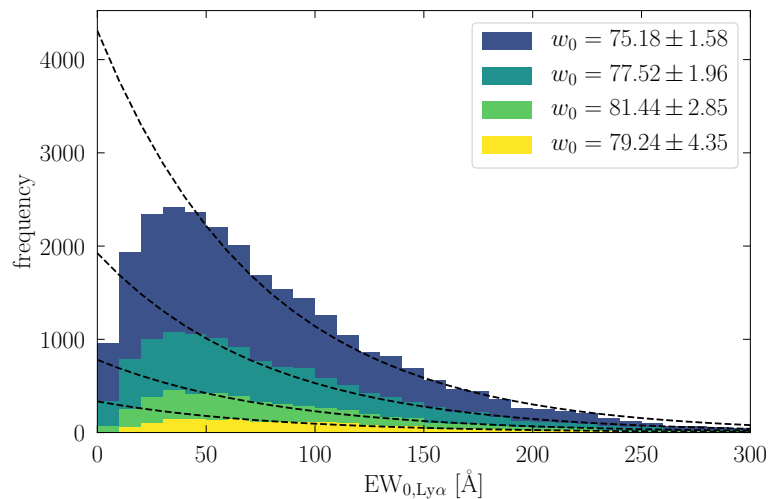


Figure 4.20.: Distributions of equivalent widths corresponding to the different line flux cuts in figure 4.19. The histograms are fit with an exponential function starting at the highest bin, the scale lengths are shown in the legend with their respective errors.

4.4.2. Ando Effect

As described in the introduction in section 1.3.2, many studies (e.g. Ando et al., 2006, Shimasaku et al., 2006, Ouchi et al., 2008, Stark et al., 2010, Furusawa et al., 2016 and others) have found an apparent correlation between the rest-frame EWs and the absolute UV magnitudes of LAEs. As can be seen in figure 4.18, there is a clear lack of objects with large EW_0 and bright absolute UV magnitudes. If the effect is real, this could be correlated with the mass of the galaxy as less massive galaxies are fainter than more massive ones. There has also been the claim that the observed correlation is only a selection effect. If bright galaxies are less common than fainter galaxies and strong Lyman α emission (producing high equivalent widths) is less common than weak Lyman α emission, both effects combined could account for the apparent absence of bright objects with high EWs, simply because they are more rare and thus less likely to be observed, not because this combination of properties is intrinsically prohibited.

Another effect that is visible in figure 4.18 is the lack of LAEs with small EWs and bright UV continua (stronger for the MUSE-Wide sample). To investigate this lower cut-off in the plot as well as the missing UV bright objects with high EWs (the Ando effect), I can make some statistical considerations, similar to Hashimoto et al. (2017) (see figure 1.12 in the introduction). For the distribution of UV continuum values I use a Schechter function with $\alpha = -1.6$ (similar to the LF in Bouwens et al., 2015), for the distribution of EWs I use an exponential function with a scale factor of $w_0 = 50$. I then create 10 000 objects with UV continuum and EW values drawn randomly and independently from both distributions. The Lyman α line flux is derived from the equivalent width and the UV continuum luminosity. The result can be seen in the left panel of figure 4.19 where I show the Lyman α equivalent width $EW_{0,Ly\alpha}$ over the logarithmic line luminosity $L_{Ly\alpha}$ in arbitrary units. If I now introduce a lower limit in the line flux luminosity (simulating observation limits) and look at the EW values over the UV continuum luminosity (see the right panel in figure 4.19), both the Ando effect and the lower cut-off are clearly visible. The cut-off is determined by the line flux limit of the survey and the Ando effect can be indeed explained by the rarity of both high EW objects and UV bright objects, since I do not input any other correlation between the equivalent width and the UV luminosity in the statistical simulation.

Another interesting statistical effect is the lack of small equivalent widths when introducing a line flux limit. This can be seen in the EW histograms in figure 4.20. While I initially used an exponential function for the distribution of equivalent widths (with a scale factor of $w_0 = 50$), the resulting histograms look very different. The numbers drop towards low equivalent widths and the shape could not be reproduced, as I get scale factors that are much larger than what was put in (around $w_0 \sim 80$). This goes to show again that without a proper handling of the survey observational limits, such EW histograms do not reflect the underlying properties of the LAE population.

Lyman Continuum Emission from MUSE Lyman α Emitters

Abstract: During and after the epoch of reionisation, star-forming galaxies contribute Lyman continuum (LyC) radiation to the total ionising photon budget of the IGM. Such galaxies can be efficiently selected through their Lyman α emission line, which is connected to the escape fraction of their emitted LyC emission. I explore this connection in order to constrain the LyC emission from Lyman α Emitters (LAE) by investigating correlations between the Lyman α line profiles and LyC escape. I search for LyC emission in broad-band imaging from the HDUV legacy survey, both in individual objects as well as in stacks of LAEs detected by their Lyman α lines as part of the MUSE Guaranteed Time Observation programs MUSE-Wide and -Deep. I mask contamination from bright neighbours, constrain the redshift to $3 < z < 4$ and require the objects to be detected in at least one Hubble Space Telescope (HST) band for stacking, leaving us with 323 objects. To further select only objects which can add a significant signal to the stack, I sort them by their ultra-violet (UV) continuum flux and am left with a sample of 60 objects. Additionally, I group these objects by Lyman α line properties measured through spectral fitting of the MUSE data and stack the subgroups. I identify six individual LyC leaker candidates. I detect LyC emission in the stack of all 60 LAEs with a signal-to-noise ratio of $S/N = 5.5$, which remains at 4.8 even when the individual LyC candidates are removed from the stack. When stacking subsets grouped by Lyman α line properties, there are detections in all subsets. I find that the most significant signal can be achieved when stacking LAEs with Lyman α emission lines that are asymmetric, have rather smaller equivalent widths (EWs), a small full width at half maximum (FWHM) and have no double peak. This creates a possible tension between our results and models and observations at low-redshift, which find higher LyC escape for narrow lines and large equivalent widths. Using the large number of LAEs found in MUSE data combined with the high spatial resolution HDUV data we detect LyC emission in the stacked photometric data of LAEs, as well as significantly increase the current sample of individual high-redshift LyC leaker candidates.

5.1. Introduction

LyC leakers may be the key to understanding the origins of the photons needed to reionise the Universe and to keep it ionised after the epoch of reionisation (EoR), the time after the cosmic dark ages of the Universe. This important phase, when the first galaxies formed and the intergalactic medium (IGM) in the Universe went from a mostly neutral to a mostly ionised state, has been studied in multiple ways in order to understand its duration, process and the sources powering it.

Through Gunn-Peterson absorption observations of quasar spectra, the end of the EoR is constrained to $z \sim 6-8$ (Fan et al., 2006a), with a starting point at $z \sim 7.8-8.8$, derived from observations of the Thomson optical depth in polarization maps of the cosmic microwave background (CMB, Planck Collaboration et al., 2016). While it is possible to constrain the time-frame of the EoR and even the patchy process, there is less

consensus yet as to what provided the necessary ionising photons.

The current contestants are either young, massive stars in star-forming galaxies, or quasars powered by early black holes. Thus, either LAEs and Lyman break galaxies (LBGs) or active galactic nuclei (AGN) are leaking LyC radiation into the surrounding circumgalactic and intergalactic medium. However, the number density of bright AGN decreases rapidly at $z > 3$ (e.g. Masters et al., 2012) and the escape fraction of ionising photons from AGN seems to be not as high as needed (e.g. Micheva et al., 2017a). If the bright AGN are not numerous enough, a large number of low luminosity AGN could contribute significantly to the reionisation budget, as has been claimed e.g. by Giallongo et al. (2015) and Grazian et al. (2018). Their results have been put into doubt however by Parsa et al. (2018), who found significantly fewer faint AGN at redshifts $z > 4$, as do the recent results by Subaru (Onoue et al., 2017) at redshifts $z \sim 6$, again pointing to a less significant role of AGN for reionisation and towards LAEs as the main culprits. Since we know from the Lyman α emission line that LAEs produce hydrogen ionising radiation, the question is how much of that radiation can escape the galaxies before being converted to Lyman α photons or absorbed by dust. Models predict that an average escape fraction of LyC emission of $\sim 10\%$ (e.g. Pawlik et al., 2009, Mitra et al., 2015, Giallongo et al., 2015, Madau & Haardt, 2015, Feng et al., 2016) is needed to ionise the IGM and keep the Universe ionised.

Since the IGM becomes increasingly opaque to LyC photons at higher redshifts, it is not possible to observe LyC leakers directly at the EoR. Instead, there have been efforts to observe LyC leakage from LAEs and LBGs after the EoR, assuming there is no strong evolution of their properties (Ouchi et al., 2008, Hu et al., 1998, Blanc et al., 2011). This assumption is justified by a lack of evolution in most properties of LAEs up to redshifts of $z \sim 6$ where the number density drops, again indicating the endpoint of reionisation (e.g. Caruana et al., 2012, Ouchi et al., 2008, Kashikawa et al., 2011). That way, understanding the connection between LyC leakage and galaxy properties that are easier to observe after the EoR, may help predict the LyC leakage at the EoR, via e.g. extrapolation or correlation with other properties.

But even so, there are only a few confirmed LyC leakers at intermediate redshifts of $z \sim 2 - 5$ (Vanzella et al., 2016, de Barros et al., 2016, Shapley et al., 2016, Vanzella et al., 2018, Bian et al., 2017) and only one with a high escape fraction of $f_{\text{esc}}^{\text{LyC}} = 50\%$ (de Barros et al., 2016). Line-of-sight (LoS) variations in the neutral fraction of the IGM, which increases with redshift and absorbs LyC radiation before it reaches us, makes the detection of individual high-redshift LyC leakers rare. This is why low-redshift analogues, such as Green Peas (Cardamone et al., 2009) or Lyman break analogues (LBAs, Overzier et al., 2009), are used to study LyC leakage. Green Peas are highly star-forming with prominent [O III] emission lines and redshifts of $0.112 \lesssim z \lesssim 0.360$, making them appear green in the SDSS images where they were first discovered (Cardamone et al., 2009). The escape fractions for some such objects have recently been found to be as high as $f_{\text{esc}}^{\text{LyC}} = 73\%$ for the full sample of 11 objects known so far (Izotov et al., 2016a, Izotov et al., 2016b, Izotov et al., 2018a, Izotov et al., 2018b). Apart from Green Peas, there are other galaxies in the low-redshift Universe that are leaking LyC emission, such as Haro11 (Hayes et al., 2007 with $f_{\text{esc}}^{\text{LyC}} = 9\%$), Tol 1247 with $f_{\text{esc}}^{\text{LyC}} = 2.4\%$ (Leitet et al., 2013), Mrk54 with $f_{\text{esc}}^{\text{LyC}} = 6.2\%$ (Deharveng et al., 2001) and J0921 with $f_{\text{esc}}^{\text{LyC}} \sim 1\%$ (Borthakur et al., 2014).

Still, observations of low-redshift galaxies that are leaking Lyman continuum photons are rare and have to be observed from space. At redshifts $z > 3$ the Lyman α line moves into the optical wavelength range, making it possible to find suitable Lyman continuum candidates with ground-based telescopes. The increased distance makes it difficult to observe the Lyman continuum in individual galaxies, however.

To overcome the observational difficulties of single LyC leakers at high redshift, several studies use stacks of LBGs and LAEs to find LyC emission (e.g. Naidu et al., 2018, Marchi et al., 2018, Japelj et al., 2017, Marchi et al., 2018, Grazian et al., 2016, Grazian et al., 2017 and Rutkowski et al., 2016, Rutkowski et al., 2017 at lower redshifts), none of which observe large enough escape fractions. Recently, Steidel et al. (2018) performed a stacking analysis of spectra of Lyman break selected galaxies at $z \sim 3$, finding $f_{\text{esc}}^{\text{LyC}} = 9\%$ for the full sample, while Fletcher et al. (2019) find 30% individual LyC leakers among their sample of 54 LAEs, but in their stacks without these 18 candidates, they find an upper limit in the LyC escape fraction of 0.3%.

To facilitate the search for individual LyC leakers as well as the selection of appropriate sources for stacking, studies have suggested links between LyC radiation and other properties of the LyC candidates (Verhamme et al., 2015, Verhamme et al., 2017, Marchi et al., 2018, Izotov et al., 2018b, Dijkstra, 2014). Assuming they are caused by scattered Lyman α photons, small Lyman α halos can indicate that there is only a small amount of neutral hydrogen in the circumgalactic medium (CGM) of the galaxy, making LyC escape more easily (Marchi et al., 2018 find that UV compact galaxies also leak more LyC photons). Studying the Lyman α halo sizes of individual objects is challenging, but recent studies have shown that

most LAEs at high redshifts have Lyman α halos (Wisotzki et al., 2016, Leclercq et al., 2017b, Saust et al. in prep., Wisotzki et al., 2018), making it possible to predict LyC emission from the halo size.

Other possible signs of LyC leakage can be seen in the rest-frame optical spectra of galaxies. A high [O III]/[O II] ratio has been proposed as an indicator for LyC leakage, as it is a sign of density-bounded H II regions (first proposed by Jaskot & Oey, 2013 and Nakajima & Ouchi, 2014, see also Nakajima et al., 2016, Faisst, 2016). Although some individual LyC candidates have shown high [O III]/[O II] ratios, it does not by itself seem to guarantee high LyC escape fractions (see Rutkowski et al., 2017 at lower and Naidu et al., 2018 at high redshift) and the possible correlation includes a lot of scatter (as recently shown by Izotov et al., 2018b).

Other suggested indicators of LyC leakage are Lyman α spectral line shape properties, such as a small velocity offset of the Lyman α line with respect to the systemic redshift (and thus a small peak separation in case of a blue bump), large equivalent widths and a narrow line (predicted through radiative transfer modelling by Verhamme et al., 2015 and observationally supported by Verhamme et al., 2017, Marchi et al., 2018, Izotov et al., 2018b, Dijkstra, 2014). These Lyman α line shape properties all indicate an easy escape of Lyman α photons without much scattering in neutral hydrogen, thus pointing towards a low neutral hydrogen column density. As LyC emission is easily absorbed by neutral hydrogen, a low neutral hydrogen column density increases the LyC escape fraction. Therefore, I investigate the connection between LyC escape and Lyman α line shape properties by stacking groups of objects based on these properties. Using high-redshift LAEs is advantageous as it makes it possible to detect them by their Lyman α emission line from the ground, where IFUs such as MUSE are available. This eliminates the need for narrow-band imaging with spectroscopic follow-up observations, but makes it necessary to either observe very UV bright LAEs or stack them in order to increase the possible Lyman continuum signal, as done in this work.

For the analysis in this chapter I use both the integral field spectroscopy data from MUSE and HST photometry. In section 5.3 I highlight the discovery of individual LyC leaker candidates, and then show the stacks for the core sample of LAEs and describe the masking and selection of objects. I then show the results for stacking different groups of objects based on their Lyman α line shape properties and describe how to avoid low-redshift interlopers and discuss error estimations and in section 5.8 I interpret the results and give a brief summary.

5.2. Data

We use a combination of data from MUSE (Bacon et al., 2010) and HST imaging. MUSE is an integral-field spectrograph of the European Southern Observatory (ESO) Very Large Telescope (VLT), with one of its primary goals being the exploration of the formation and evolution of galaxies. MUSE covers the optical wavelength range of 4750 Å to 9350 Å, corresponding to redshifted Lyman α at redshifts $2.9 < z < 6.7$ at an average spectral resolution of $R \sim 3000$.

The field-of-view (FoV) covers roughly one square arcminute. Integral field spectroscopy is ideally suited for the search for LAEs that might be leaking LyC photons, compared to the traditional drop-out or Lyman break techniques, which compare several narrow- and/or broad-bands to find the Lyman break. This can be a bias against LyC leakers with a high escape fraction since the narrow-band technique relies on a drop in the spectrum which might be less pronounced in LyC leakers (e.g. Cooke et al., 2014, Vanzella et al., 2016, but see also Steidel et al., 2018 who do find LyC emission in LBGs). Instead of focusing on the Lyman break, the MUSE data allows us to find the Lyman α lines directly, independent of the ratio between the rest-UV continuum and the LyC. Additionally there is no need for spectroscopic follow-up observations as would be the case for photometric searches for LAEs and we get a redshift estimate from the Lyman α line, which was not the case for previous studies (e.g. Naidu et al., 2017).

5.2.1. MUSE Data

I use two MUSE datasets obtained during Guaranteed Time Observations (GTO) of the MUSE consortium. These observations were aimed at finding and analysing LAEs and three different observational depths were used: The MUSE-Wide survey being the least deep with one hour observation time per pointing in the CDF-S and COSMOS regions. It was observed from September to December 2014 (first 24 fields, see Herenz et al., 2017) and from August 2015 to May 2016 (next 36 fields, Urrutia et al., 2019). In this study, only the CDF-S data is used, as this is covered by the HDUV (Oesch et al., 2015, Oesch et al., 2018, see

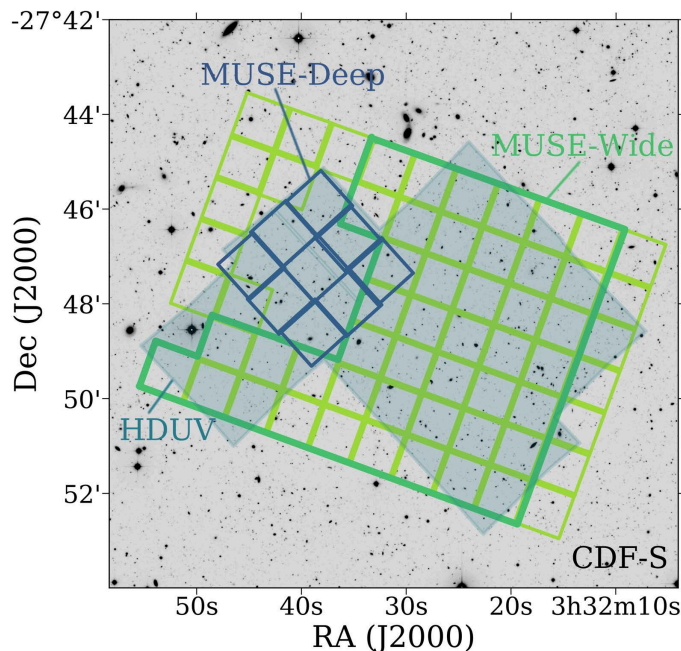


Figure 5.1.: Footprint of MUSE pointings in the CDF-S region. The 60 individual MUSE-Wide pointings are shown in bright green, the 44 fields used in this chapter are circled with a thick green line, the nine MUSE-Deep fields (that also cover the UDF 10) are marked blue. The HDUV footprint is shown as the shaded area and covers most of the MUSE-Wide fields as well as MUSE-Deep.

figure 5.1).

The MUSE-Deep survey encompasses the medium- and ultra-deep observations in the HUDF with observations of ten hours in the nine medium-deep pointings and 31 hours in the ultra deep UDF 10, both merged into a single catalogue and used in this chapter (see Bacon et al., 2017 and Inami et al., 2017 for detailed information).

The MUSE data gives us not only the necessary positional information for the LAEs to search for LyC emission in (stacks of) the complimentary HST data, but it also provides information on the spectral shape of the Lyman α line. In total, there are 1162 objects in both the MUSE-Wide and -Deep catalogues that are covered by the HDUV legacy survey (see figure 5.1). This is slightly less than I used in chapter 3 as I exclude the MUSE-Wide COSMOS objects for this study.

MUSE-Wide

To find emission line objects in the MUSE-Wide survey, and thus LAEs, we used the Line Source Detection and Cataloguing Tool (LSDCat¹, described in Herenz & Wisotzki, 2017), which uses a 3D matched-filtering approach. This way we secure selection completeness of the emission line sources e.g. for faint LAEs that do not have a counterpart in the ancillary HST data. Therefore we do not use the HST data as a prior to find emission line objects in the MUSE data for MUSE-Wide. Catalogues for the first 24 fields can be found in Herenz et al. (2017) and the survey description and first data release in Urrutia et al. (2019). In this study I investigate the 39 MUSE-Wide pointings that overlap the HDUV imaging (see figure 5.1), where LyC can be detected.

We classified the emission line objects detected by LSDCat using the graphical user interface QtClassify² (see Appendix A in Herenz et al., 2017 and also chapter 2 in this thesis), which helps to identify the redshifts of the objects by looking at the full 3D information that MUSE provides. In addition to the actual classification, we also attribute a confidence (a value between 1 for low and 3 for high confidence) to the classification, which is a subjective measure but used to gauge how reliable the classification is. Mostly a low confidence is driven by a low signal-to-noise-ratio or a weak line flux, but sometimes the line is not uniquely identifiable, which is important in later steps when excluding possible interlopers (see section 5.7).

¹LSDCat is available via the Astrophysics Source Code Library: <http://www.ascl.net/1612.002>

²QtClassify is available via the Astrophysics Source Code Library: <http://ascl.net/1703.011>

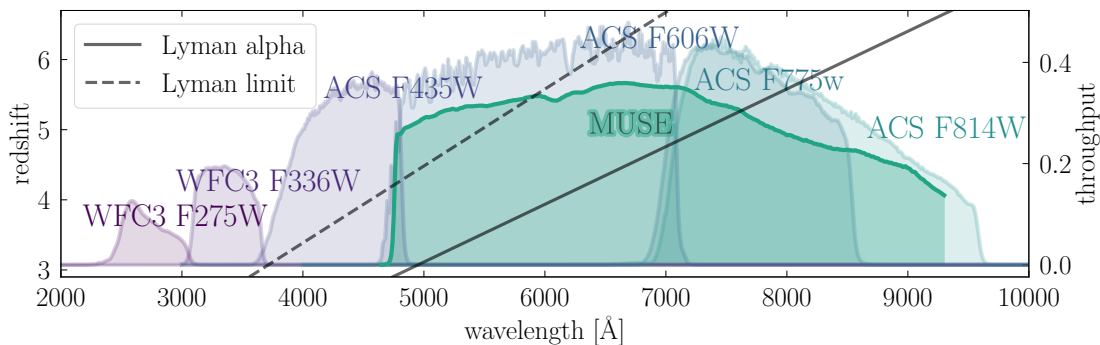


Figure 5.2.: HST bands used in this study. The x -axis shows the wavelength in \AA , the left y -axis shows the redshift and corresponds to the solid and dashed lines, the right y -axis shows the filter throughput and corresponds to the coloured filter curves. The solid line shows where the Lyman α line would fall depending on redshift, the dashed line shows the Lyman limit redwards of the LyC. Only the redshift range where Lyman α is visible for MUSE is shown. The two bands with the lowest wavelengths (WFC3 F275W and WFC3 F336W) are from the HDUV survey and are used in this study to find LyC emission.

Each object was classified by at least two investigators and then consolidated with a third. This reduces the risk of wrong classifications and thus the introduction of interlopers in the LAE sample, which is crucial for confirming LyC leakers and stacking the LAEs.

MUSE-Deep

While the MUSE-Wide sample is purely emission line selected, the MUSE-Deep catalogue (Bacon et al., 2017) got constructed using various methods such as an automated emission line detection software based on a matched filter approach (ORIGIN, Mary et al. in prep., Bacon et al., 2017), with a prior removal of continuum emission through a Principal Component Analysis (PCA). In addition to ORIGIN, the software MUSELET was used (Piqueras et al., 2017), which is based on SExtractor (Bertin & Arnouts, 1996) and works with continuum subtracted narrow-band images created from MUSE data by collapsing the five closest wavelength layers (Inami et al., 2017). The emission lines were then classified semi-automatically using a slightly modified version of MARZ (Hinton et al., 2016). At least two investigators classified each of the emission line objects, again with a third person checking the final results.

Both for the MUSE-Wide and the MUSE-Deep LAEs, I use extracted one dimensional spectra weighted by the point spread function (PSF) to measure all line properties except for the line flux, where I use a spectrum summed over the segmentation map from Rafelski et al. (2015) convolved with the MUSE PSF for MUSE-Deep (Inami et al., 2017), and measurements from LSDCat within three Kron radii for MUSE-Wide.

5.2.2. Photometry

While the MUSE data is ideal for finding LAEs, it is not deep enough to accurately constrain the (Lyman and rest-UV) continuum of the LAEs needed not only for the detection of LyC leakers, but also for the equivalent widths. Therefore, a number of additional HST bands is used (as illustrated in figure 5.2). The high spatial resolution of the HST data also makes it possible to clearly identify the LAE counterparts (see section 3.3.2) and avoid contamination by foreground objects as much as possible (except for chance alignments, see section 5.7).

Hubble Deep UV (HDUV) legacy survey

The redshift range MUSE covers for LAEs corresponds to a wavelength range for the Lyman limit (and thus the LyC bluewards of the Lyman limit at 912 \AA) of $3563 - 7014 \text{ \AA}$. Therefore the two bands used in this work for stacking LAEs in order to find a possible LyC signal are WFC3 F275W and WFC3 F336W (see figure 5.2), taken from the HST WFC3/UVIS program, the HDUV legacy survey (Oesch et al., 2015, Oesch et al., 2018), which includes the UVUDF (Teplitz et al., 2013, Rafelski et al., 2015). The HDUV

MUSE-ID	Guo-ID	Skelton-ID	Rafelski-ID	Coe-ID	Survey	redshift
7193			6470	4093	Deep	3.09
135005170	13707	26130	4873	5742	Wide	3.16
109004028	7570	15601			Wide	3.27
1087			3506	3416	Deep	3.46
1228			2336	2245	Deep	3.51
141011083		27114			Wide	3.66

Table 5.1.: Individual LyC leaker candidates and their IDs in different catalogues, sorted by redshift.

filters F275W and F336W go down to a magnitude of 27.5 to 28 mag_{AB} at 5σ (Naidu et al., 2017) and cover a wavelength range from $\sim 2435 \text{ \AA}$ to $\sim 3032 \text{ \AA}$ and $\sim 3096 \text{ \AA}$ to $\sim 3639 \text{ \AA}$, respectively. This data was already used to identify LyC emitter candidates at a lower redshift ($z \sim 2$, see Naidu et al., 2017) using several HST bands to construct spectral energy distributions (SEDs), finding an absolute ionising photon escape fraction of $> 60\%$ for their candidates.

Other HST Bands

In order to measure the rest-UV continuum for the rest-frame EWs of the LAEs in MUSE-Wide, I use filters at longer wavelengths than the Lyman α emission line, to exclude flux contamination from the line itself. Since not all LAEs were already discovered in the HST data, I use `GalFit` (Peng et al., 2002, Peng et al., 2010) to fit the UV counterparts (if possible) to get the magnitudes in each band (see section 4.1 in this thesis for more information on the `GalFit` fitting process). Around $\sim 30\%$ of our LAEs are too faint to have visible UV counterparts, even in the deep HST data (Marino et al., 2018, Bacon et al., 2017, Maseda et al., 2018) and are not considered here.

5.3. Individual Lyman Continuum Leaker Candidates

After visually checking our LAE sample for interlopers and removing contaminants (see section 5.5.1 and 5.7), I find that among our sample there are six objects for which LyC emission is detected individually in the HST WFC3 F336W band (see table 5.1). To get the signal-to-noise ratio at the position where I expect the LyC emission, account for the HST point spread function (PSF) and get rid of pixel-to-pixel correlations, I convolve the cut-outs with a Gaussian function with $\sigma_G = 0''.1$. I measure the flux (f) in the WFC3 F336W band inside an aperture with a diameter of $0''.5$ centered on the UV continuum position in the HST data (the centre of the cut-out) after having subtracted the background B in each band (for a description see section 6.2 of this thesis). The standard deviation in the background (σ_B) of the filtered image is used to estimate the signal-to-noise of the object (see equation 5.1), where \sqrt{N} (with N being the Number of pixels in the aperture) comes into play because we need the standard deviation in the aperture, not in a single pixel. I correct for convolving with a Gaussian of $\sigma_G = 0''.1$ and get the signal-to-noise ratio:

$$S/N = f / (\sigma_B \sigma_G 2 \sqrt{\pi} \sqrt{N}) \quad (5.1)$$

In total, I find six objects with a S/N value above 3 in the HST WFC3 F336W band at a redshift of $z < 4$ (see section 5.4 and figure 5.5 below for a discussion of the redshift cut). All of the candidates (sorted by redshift) are listed in the table 5.2. An overview of the cut-outs in different bands can be found in figure 5.3 and the Lyman α emission lines are shown in figure 5.4.

5.3.1. Evaluation of Candidates

I divide the list of LyC leaker candidates into 'likely' and 'uncertain'. According to my selection procedure (section 5.5.1), the candidates are in the redshift range $3 < z < 4$, do not have bright neighbours, are fully in the field-of-view (FoV) of both MUSE and the HDUV and have detections in at least one HST band. I discuss for each candidate the likelihood of it being a real LyC leaker based on several criteria, each of which individually might not be sufficient to rule out the possibility of the LyC emitter candidate to be real, but if several criteria apply the candidate is not reliable. The criteria are:

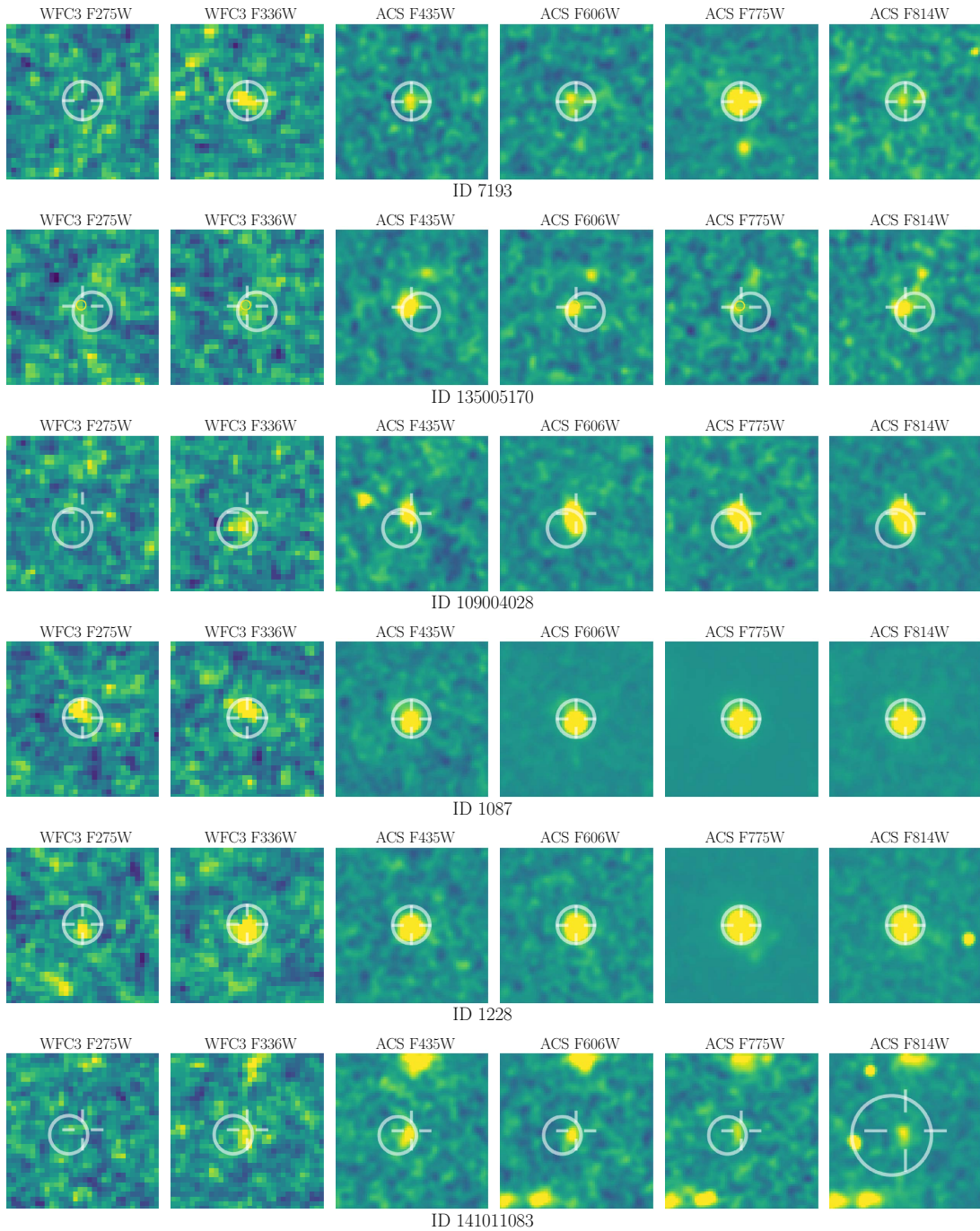


Figure 5.3.: Cut-outs (with S/N values ranging from -2 in dark blue to $+2$ in bright yellow) of all LyC leaker candidates. The IDs are (from top to bottom): 7193, 135005170, 109004028, 1087, 1228, 141011083. The shown HST bands are (from left to right): WFC3 F275W, WFC3 F336W, ACS F435W, ACS F606W, ACS F775W and ACS F814W. The white cross is centered on the UV continuum counterpart position (from the ACS F814W band) and the white circle (with a diameter of $0''.5$) is centered on the Lyman α emission line position from MUSE. Each cut-out has a size of $2'' \times 2''$.

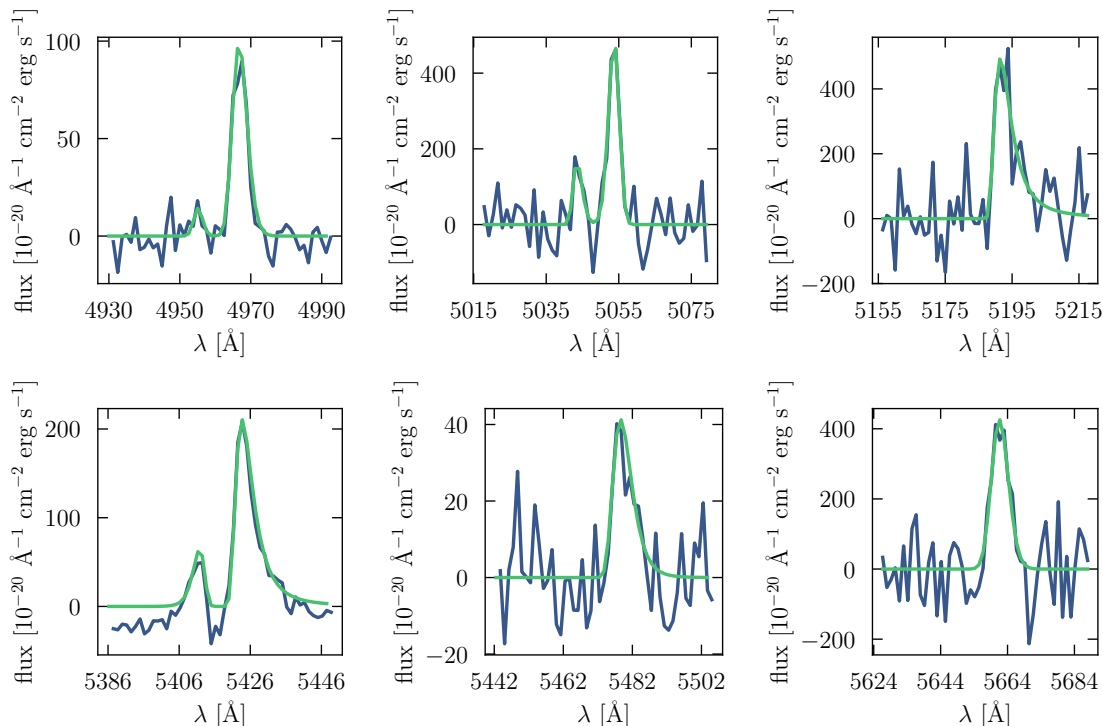


Figure 5.4.: Lyman α emission lines of all LyC leaker candidates, from left to right, top to bottom: IDs 7193, 135005170, 109004028, 1087, 1228, 141011083. The blue line is the median subtracted spectrum, the green line is the asymmetric Gaussian fit.

1) The morphology of the rest-UV continuum and the LyC in the HST bands does not have to be exactly the same, but there should not be a significant offset between the LyC and the rest-UV continuum. LyC can escape through channels where the neutral hydrogen column density is lower than in the rest of the galaxy. However, the Lyman α emission may not follow the same path as the LyC emission, since although both are affected by neutral hydrogen in their path, the Lyman α emission is resonantly scattered which increases its extent and there are several studies that found offsets between the Lyman α emission and the LyC emission (Mostardi et al., 2015, Micheva et al., 2017b, Iwata et al., 2009). Thus the LyC emission can have a different morphology than the rest-UV continuum emission or the Lyman α emission, but should not be more extended than the rest-UV continuum.

2) The two bands containing the LyC emission, WFC3 F275W and WFC3 F336W should have consistent morphologies and positions of the LyC emission. Whether the WFC3 F275W band is expected to have LyC emission at all (since it is at a lower wavelength) depends on the IGM transmission and the intrinsic shape of the LyC part of the spectrum. Thus, a detection in the WFC3 F275W band is not necessary for the LyC emission in the WFC3 F336W band to be considered reliable. On the contrary, more LyC flux in WFC3 F275W than in WFC3 F336W is extremely unlikely due to the absorption in the IGM, which is why the flux ratio between F275W and F336W should be considered for the reliability of the LyC leaker candidates.

3) A low Lyman α classification confidence can also indicate a possible interloper or even a wrong redshift. A solution would be to verify the obtained redshift from the potential Lyman α emission line by comparing to photometric redshifts from the literature. This is not fully conclusive either, though, since a correctly identified line will always give the more reliable redshift.

The two most likely LyC leaker candidates are ID 109004028 and ID 7193, both of which have an emission line classification confidence of 2, their LyC emission coincides with the UV continuum and they have no strong detection in WFC3 F275W (see figure 5.3).

The other four candidates are classified as uncertain. ID 135005170 has detections in both WFC3 F336W and WFC3 F275W, which are slightly more extended than the rest-UV continuum. ID 1087 has a typical Lyman α line with a double peak and a strong asymmetry. However, the possible LyC emission is slightly offset from the rest-UV emission and strong in the WFC3 F275W band, with a F275W/F336W flux ratio of $f_{\lambda, F275W}/f_{\lambda, F336W} = 1.97$. For ID 1228 the possible LyC emission and the rest-UV emission coincide and but there is a (weak) signal in the WFC3 F275W as well. However, the F275W/F336W flux ratio

MUSE-ID	f_{900}	f_{1500}	$(f_{900}/f_{1500})_{\text{obs}}$	$\text{EW}_{0,\text{Ly}\alpha}$	S/N_{336}
7193	6.775 ± 1.361	1.786 ± 0.109	3.794 ± 0.797	28.6 ± 3.6	4.89
135005170	3.049 ± 0.097	2.190 ± 0.250	1.392 ± 0.469	52.0 ± 17.6	3.01
109004028	4.804 ± 1.362	8.407 ± 0.272	0.572 ± 0.163	34.6 ± 7.2	3.59
1087	6.261 ± 1.424	23.882 ± 0.369	0.262 ± 0.060	13.3 ± 0.2	4.85
1228	1.429 ± 1.727	11.837 ± 0.359	1.207 ± 0.150	3.6 ± 0.4	9.25
141011083	5.200 ± 1.360	1.727 ± 0.2541	3.012 ± 0.904	181.6 ± 47.6	4.50

Table 5.2.: Overview of fluxes and ratios for the LyC leaker candidates (also shown in table 5.1), giving the LyC and UV continuum flux densities units of 10^{-20} erg s $^{-1}$ cm 2 as well as their ratios, Lyman α equivalent widths $\text{EW}_{0,\text{Ly}\alpha}$ and the signal-to-noise in the LyC detection band WFC3 F336W S/N_{336} .

is $f_{\lambda,\text{F275W}}/f_{\lambda,\text{F336W}} = 0.7$, which is reasonable. The LyC emission of ID 141011083 follows the same morphology as the rest-UV continuum and is not visible in the WFC3 F275W band. However, the emission line confidence is rather low and the photometric redshift puts it at $z \sim 1.4$ (e.g. Rafelski et al., 2015).

5.4. Lyman Continuum Escape

One possibility to estimate the amount of LyC leakage is by deriving the LyC escape fraction for each object. This relies on assumptions both on the intrinsic luminosity ratio $(L_{\text{UV}}/L_{\text{LyC}})_{\text{int}}$ between the UV continuum and the Lyman continuum and on IGM transmission T_{IGM} , which strongly depends on the line-of-sight and cannot be measured individually but only estimated statistically. The fact that I see LyC emission in individual objects already argues for a line-of-sight with a high IGM transmission, which is why it is not reasonable to use a mean IGM transmission to correct the LyC emission for individual objects (see for example Steidel et al., 2018 for a detailed discussion of this issue). Therefore I focus on the flux ratio between the LyC and the UV continuum flux densities $(f_{\text{LyC}}/f_{\text{UV}})_{\text{obs}}$ which I can measure directly.

5.4.1. Lyman Continuum to UV Continuum Flux Ratios

I use the WFC3 F336W band to get the LyC flux density f_{LyC} , which covers the rest-frame wavelength range between 670 Å to 840 Å. For the UV continuum flux density f_{UV} I use the ACS F814W band for MUSE-Wide and the ACS F775W band for MUSE-Deep, since they are the deepest in the respective area. Usually, the LyC and UV continuum are compared at the wavelengths 900 Å and 1500 Å, respectively, which is why I correct the flux density measurements using a β -slope of -1.92 (see section 4.1.2 and section 5.6.1) for the UV continuum. For the LyC I use a correction based on stellar population synthesis models from Starburst99 (Leitherer et al., 1999, with improvements by Vázquez & Leitherer, 2005, Leitherer et al., 2010 and Leitherer et al., 2014). There is a detailed discussion on the stellar population synthesis models used here in section 6.3 of this thesis. I then measure the flux ratio in the LyC and UV continuum bands (which can be seen table 5.2).

5.4.2. Lyman Continuum Escape Fractions

The flux ratio between LyC and UV continuum $(f_{\text{LyC}}/f_{\text{UV}})_{\text{obs}}$ is directly measurable and does not depend on assumptions on the IGM or the galaxy's stellar populations (and thus the intrinsic ratio $(L_{\text{UV}}/L_{\text{LyC}})_{\text{int}}$). However, most studies on LyC emission estimate the escape fraction to judge the contribution to the ionizing background of the IGM.

The absolute escape fraction $f_{\text{esc}}^{\text{abs}}$ is the ratio of LyC emission that escapes into the IGM to the intrinsic LyC emission. Since this relies on an accurate knowledge of the H I content of the galaxy, the photon production rate, the CGM and the dust attenuation in the galaxy ($\tau_{\text{UV},\text{dust}}$), Steidel et al. (2001) introduced the relative escape fraction (see also section 1.4.2 in the introduction of this thesis for a detailed discussion of the escape fraction).

$$f_{\text{esc}}^{\text{abs}} = f_{\text{esc}}^{\text{rel}} \exp(-\tau_{\text{UV},\text{dust}}) \quad (5.2)$$

The relative escape fraction³ is simply the ratio of the intrinsic UV luminosity L_{UV} and LyC luminosity L_{LyC} divided by the observed flux density ratio $(f_{UV}/f_{LyC})_{obs}$. The observed ratio has to be corrected by the IGM optical depth of LyC photons τ_{IGM} .

$$f_{esc}^{rel} = \frac{(L_{UV}/L_{LyC})_{int}}{(f_{UV}/f_{LyC})_{obs}} \exp(\tau_{IGM}) \quad (5.3)$$

Note that [Steidel et al. \(2018\)](#) discuss in detail the absorption of LyC photons in the IGM as well as in the CGM. If the CGM should be included in the consideration depends on the definition of the escape fraction: Do we mean by escape that the photon leaves the galaxy or also its CGM? In this case I choose to interpret the CGM as part of the galaxy and only account for the absorption in the IGM.

As explained in section 5.4.1, I use the ACS F814W band for MUSE-Wide and the ACS F775W band for MUSE-Deep to measure the UV continuum flux density f_{UV} in the rest-frame wavelength range between 1200 Å and 1500 Å (using the effective wavelength, for redshifts between 3 and 4) and the WFC3 F336W band for the LyC flux density f_{LyC} in the rest-frame wavelength range between 670 Å and 840 Å. Thus, I can write (see e.g. [Siana et al., 2007](#)):

$$f_{esc}^{rel} = \frac{(L_{1500}/L_{900})_{int}}{(f_{1500}/f_{900})_{obs}} \exp(\tau_{IGM}) \quad (5.4)$$

The intrinsic break amplitude $(L_{1500}/L_{900})_{int}$ is often estimated from spectral synthesis models (e.g. [Bruzual & Charlot, 2003](#), [Leitherer et al., 1999](#)), since it is difficult to infer from observations. It is usually set to values between three (see e.g. [Steidel et al., 2001](#), [Grazian et al., 2016](#), [Marchi et al., 2018](#), [Grazian et al., 2017](#), [Japelj et al., 2017](#)), five (see e.g. [Naidu et al., 2018](#)) and 6 – 8 (see e.g. [Siana et al., 2007](#)), depending on the input parameters: star formation history, metallicity, initial mass function (IMF) and age.

I use Starburst99 ([Leitherer et al., 1999](#), with improvements by [Vázquez & Leitherer, 2005](#), [Leitherer et al., 2010](#) and [Leitherer et al., 2014](#)) models with a metallicity of 0.002 and a star formation rate of 1 M_{\odot} per year and derive an intrinsic ratio of $L_{1500}/L_{900} = 1.566$ for 10⁸ Myrs (using the Geneva evolutionary tracks with rotation, see section 6.3). This value is given for luminosities in wavelength, while the papers cited above use frequency space, where my ratio translates to $L_{LyC,\nu}/L_{UV,\nu} = 4.35$ and is thus well in the range of literature values.

Taking the IGM transmission of LyC photons $T_{IGM} = \exp(-\tau_{IGM})$ into account when computing the escape fraction can be done in different ways. There are numerous measurements of T_{IGM} , e.g. using quasar sight lines, but it is not possible to know the exact IGM transmission at all lines-of-sight. Thus, I can either resort to the mean IGM transmission measured from quasar absorption spectra or I can treat the IGM in a statistical way. The former approach was used e.g. in [Grazian et al. \(2016\)](#), but using the mean transmission disregards the stochastic nature of the IGM. The transmission is mainly driven by the presence of Lyman limit systems (LLSs) in the line-of-sight (see section 1.5.1 in the introduction to this thesis for a discussion of LLSs), meaning there is a possibility even at redshifts above $z \sim 4$ of a clear line-of-sight with low H I column densities (see figure 5.5). However, the probability for such clear lines-of-sight goes down with redshift and the LyC IGM transmission tends to 0, which is why I restrict the redshift range in this study to $3 < z < 4$.

For individual LyC leaker candidates it is unreasonable to estimate the relative escape fraction from the mean of the IGM transmission at the given redshift, as it can be assumed that the IGM transmission in the line-of-sight was high if we can see Lyman continuum leakage. Nevertheless, estimating the IGM transmission for a stack of objects is possible. I use transmission curves for 10 000 lines-of-sight (from [Inoue & Iwata, 2008](#) and [Inoue et al., 2014](#)) of the IGM for different redshifts and convolve them with the transmission curve of the WFC3 F336W filter band. For each object in the full stack I draw random lines-of-sight of the same redshift 10 000 times to get an estimate of the mean IGM transmission and the standard deviation.

³Note that I am using a slightly different form of the relative escape fraction here than in the introduction in section 1.4.2. In the introduction I tried to derive the escape fraction in an intuitive way, but most studies in the literature use the form given in this chapter. The difference is that here the ratio of the intrinsic UV to LyC is divided by the observed UV to LyC ratio, while in the introduction the entire fraction is reversed and I divide the observed Lyman to UV continuum ratio by the intrinsic one. Both fractions are equivalent and can be converted into each other. Especially the intrinsic ratio between UV and Lyman continuum is usually given as $(L_{UV}/L_{LyC})_{int}$ instead of $(L_{LyC}/L_{UV})_{int}$, so to be able to compare it to the literature, I use the first version.

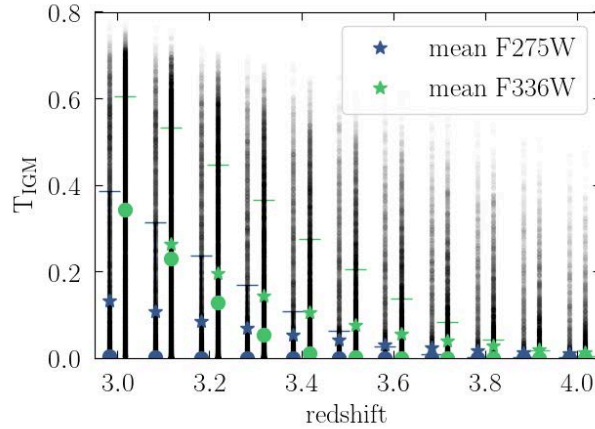


Figure 5.5.: Transmission of the IGM as a function of redshift (Inoue & Iwata, 2008 and Inoue et al., 2014). The grey dots show 10 000 individual lines-of-sight, the blue and green dots show the median values for the WFC3 F275W and WFC3 F336W HST bands respectively (shifted on the x -axis for clarity). The blue and green stars show the mean IGM transmission, the errorbars show the 84th percentile (corresponding to 1σ for a normal distribution). The transmission is already convolved with the filter curve of the bands. While the mean (median) goes down to almost 0 at a redshift of $z \sim 4$ ($z \sim 3.4$) for the WFC3 F336W band, there are still individual lines-of-sight with a transmission of LyC photons of 40% up to a redshift of $z \sim 4$.

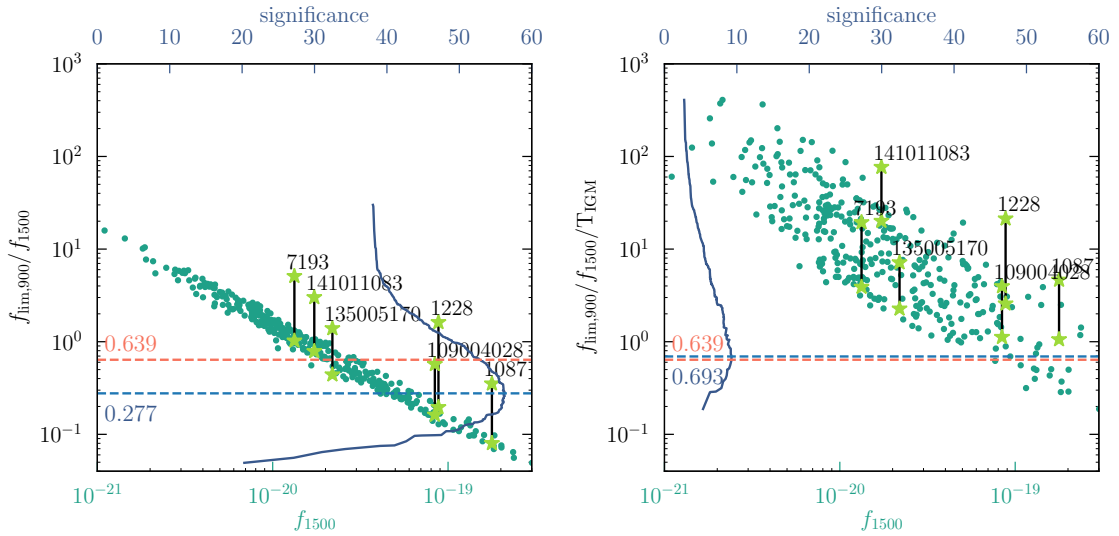


Figure 5.6.: Flux ratio $f_{\text{lim},900}/f_{1500}$ over UV continuum flux f_{1500} . The green dots show all objects which have not been excluded, the light green stars show the individual LyC leaker candidates both at their $f_{\text{lim},900}/f_{1500}$ values as well as their real measured LyC continuum to UV continuum ratio f_{900}/f_{1500} , connected with a black line and annotated with their IDs. The blue solid line corresponds to the x -axis at the top of the plot, showing the significance when stacking all objects below the corresponding y -value. The blue dashed line indicates the maximum of the significance, which would be the ideal stacking sample, and the orange dashed line indicates the theoretical intrinsic value for the flux ratio f_{900}/f_{1500} . Values above this line are not expected from stellar population synthesis models used in this thesis. Left panel: the flux ratio $f_{\text{lim},900}/f_{1500}$ over the UV continuum flux f_{1500} . Right panel: The flux ratio $f_{\text{lim},900}/f_{1500}$ is divided by the mean IGM transmission T_{IGM} of each object.

5.5. Stacking the Selected Sample

For most of the 1162 LAEs I do not detect LyC emission individually, but there is a possibility that some of the LAEs emit LyC which is too faint to be observed in individual objects. Since this relies on low IGM absorption in the line-of-sight as well as on an intrinsic UV brightness of the galaxies, I select a sub-sample of promising LAEs in order to create stacks to increase the signal-to-noise and measure the mean LyC emission for this sample.

5.5.1. Selecting the Lyman α Emitters

For the stacks, the objects need to have a low enough redshift for the IGM to be sufficiently transparent. While there might be a few individual clear lines-of-sight even at redshifts above $z \sim 4$, the mean transmission will be too low to allow much LyC emission to get through (see figure 5.5 above). Thus, I apply a cut in redshift of $3 < z < 4$. Of the initial 1162 objects, 633 were excluded for their redshifts, leaving us with 529 objects.

To only include sources in the stacking analysis that can contribute signal, I need to consider several selection criteria for the sample. The objects need to be securely identified as LAEs, their UV continuum needs to be visible in the HST bands (Marino et al., 2018, Maseda et al., 2018) and also bright enough to make it realistic to expect detectable LyC emission. If the UV continuum is not detected, the LAEs will only contribute noise to the stacks. Therefore I only use objects which have detections in at least one HST band (excluding the one containing the Lyman α line itself). Of the remaining 529 objects, 143 objects were excluded for not being detected in one or more HST bands.

In addition to these criteria, I only want to stack objects that add signal to the stack. For this I measure the standard deviation in the LyC and use it as a 1σ limiting flux at 900 \AA ($f_{\text{lim},900}$). This is divided by the measured flux at 1500 \AA (see left panel of figure 5.6), which gives us the LyC to UV continuum flux ratio assuming each object is detected with a S/N of 1. I then compute the expected significance of stacks with an increasing number of objects in the order given in the left panel of figure 5.6, starting at the bottom where I can find the most promising objects: those with the highest UV continuum. In the right panel in figure 5.6 I additionally divide by the IGM transmission, which changes the order of the objects to include the influence of the redshift (since higher redshifts are less likely to contribute to the stack due to the increased IGM absorption).

I calculate the 'significance' S of the stacks using:

$$S = \frac{\sum_i f_{1500,i} T_{\text{IGM}}(z_i)}{f_{900,\text{lim}} \sqrt{n}} \quad (5.5)$$

with $f_{1500,i}$ being the UV continuum at 1500 \AA for each object, $f_{900,\text{lim}}$ the mean limiting Lyman continuum flux at 900 \AA , n the number of objects and $T_{\text{IGM}}(z_i)$ the IGM transmission at the redshift of the objects (only for the right panel of figure 5.6). The maximum of this significance gives us the ideal sample of objects to stack, which are all objects below the blue dashed line in the left panel of figure 5.6.

Using stellar population synthesis models from Starburst99 (Leitherer et al., 1999, see section 5.4 and section 6.3), I get an intrinsic UV to Lyman continuum ratio of $(L_{\text{UV}}/L_{\text{LyC}})_{\text{int}} = 1.566$ or $(L_{\text{LyC}}/L_{\text{UV}})_{\text{int}} = 0.639$ for LyC over UV continuum, which is marked in figure 5.6. This limit offers another way to estimate the reliability of the individual LyC leaker candidates. Any candidates with measured values (the top green stars) above this limit (dashed orange line) are likely to be uncertain candidates. The two that remain are IDs 19004028 and 1087, the latter of which had already been judged uncertain due to the spatial offset between the LyC emission and UV continuum.

All objects that are above the blue dashed line in the left panel of figure 5.6 and thus do not increase the significance for the stacks, are excluded from the final sub-sample, as well as objects with close, bright or even overlapping neighbours that could not easily be masked after visual inspection. Of the remaining 386 objects, another 43 were excluded based on figure 5.6, contamination with bright neighbours and possible overlap or superposition with other objects. This leaves us with a total of 323 LAEs for stacking, including the 6 individual LyC leaker candidates.

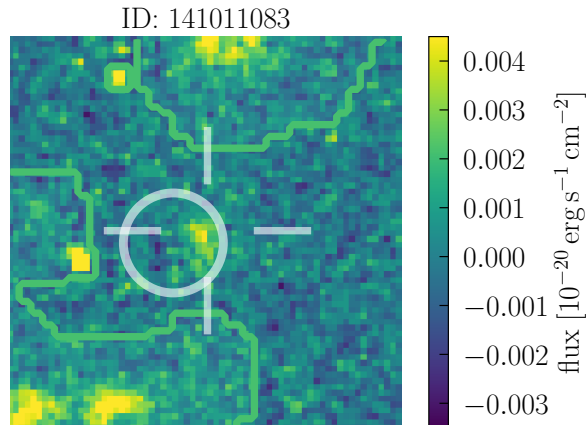


Figure 5.7.: Example of a cut-out with a size of $2'' \times 2''$ for stacking. The mask is indicated as the lighter area inside the green contours. This example shows a cut-out of the HST ACS F814W band (with two masked cosmic rays). The white cross is centered on the UV continuum position (for MUSE-Wide this position is taken from the GALFIT fits) and the white circle is centered on the Lyman α emission obtained with LSDCat which has a diameter of $0''.5$.

5.5.2. Masking

In order to stack the objects, I create cut-outs for each band. Before stacking, bright objects as well as contaminants, cosmic rays and hot/cold pixels are masked. The masks are important since incorrect masking (such as masking only bright objects and cosmic rays with a certain distance from the object position) can lead to an artificial signal. Therefore, I take several steps to creating accurate and robust masks.

Bright objects are masked in the cut-outs by using the segmentation map (from the 3D-HST catalogues by Skelton et al., 2014). To ensure that all objects are completely masked in all bands, I grow the mask by four pixels in the case of ACS bands and two pixels in the case of WFC3 bands (due to the different resolutions). The edge of the FoV of the respective pointing is also masked.

The masked image is then used to get the standard deviation for the remaining pixels in order to perform σ -clipping to get rid of hot/cold pixels and cosmic rays (which are especially numerous in the ACS F814W filter). Individual pixels with values greater than $\sigma > 5$ are masked and this mask is again grown by one pixel for all filters (but only for pixels with more than one connecting masked pixel, so as to only grow cosmic rays, but not hot/cold pixels). In order to not mask the central object, it is checked whether the pixels in the centre of the cut-out (within $0''.2$ from the position) are masked and if the regions of connected masked pixels are larger than what I would expect from a cosmic ray. Therefore, regions of more than 12 connected pixels in the middle of the image are not masked. The radius of $0''.2$ for this ensures that small uncertainties in the position will not cause the main object to be masked and the minimum size of 12 pixels ensures that objects with less than two connected pixels (and then grown to 12) are still masked as they are likely to be cosmic rays. At this stage, the median in the masked cut-out (including the possible central object) is subtracted, so as to remove any local background.

Additionally, all but the central object from the segmentation map are masked (see figure 5.7 for an example). For each filter, the same segmentation map, scaled to the respective pixel size of the bands ($0''.03$ per pixel for ACS bands and $0''.06$ per pixel for WFC3 bands), is used as a basis for the mask, which is then grown. The σ -clipping is done for each individual filter, which is why the masks used for stacking different bands vary slightly.

5.5.3. Stacking

After selecting LAEs according to the criteria described in section 5.5.1 and masking bright neighbours and cosmic rays/hot pixels, the first step is to stack the WFC3 F336W HST cut-outs of all objects in the selected sub-sample. I centered the cut-outs on the UV continuum (HST band ACS F814W) positions of the counterparts, not the Lyman α positions. This is because the Lyman α position is based on the lower

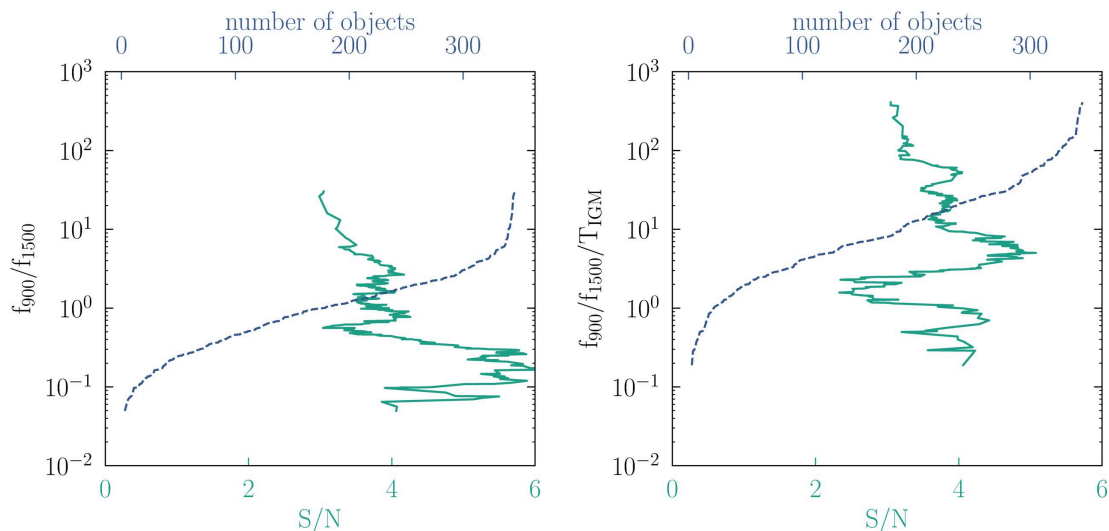


Figure 5.8.: Like the left panel of figure 5.6, the y-axis shows the flux ratio $f_{\text{lim},900}/f_{1500}$, but over the S/N of the actual stacks, which use objects sorted by the order in the left panel of figure 5.6 (without the individual LyC leaker candidates). The top x-axis corresponds to the dashed blue line and shows the number of objects used for each stack. It can be seen that the maximum of the S/N ratio is at the same position as expected from figure 5.6. The left panel shows the order of objects without the IGM correction, the right panel with the IGM correction.

spatial resolution of MUSE while the UV continuum position is measured in the high resolution HST data and also because the Lyman α halo might be asymmetrically extended, but LyC emission would escape from the same position as the UV continuum as it is not resonantly scattered in neutral hydrogen as Lyman α is.

I stack the objects in cut-outs of $5'' \times 5''$ by taking the mean in each pixel position while also taking the masking into account. Using the mean instead of the median, which is influenced far less by outliers, is necessary in this case, since I expect that most objects will not have clear lines-of-sight and the LyC photons of most objects will be absorbed by the IGM. Therefore the median would come out as 0, even if there are several leaking objects.

As discussed in section 5.5.1, I create stacks of an increasing number of objects in the order given in figure 5.6. From this figure I expect to reach a maximum S/N at around 61 objects, which coincides sufficiently well with the maximum seen in figure 5.8. From these considerations it can be seen that stacking all available objects without selecting the most promising ones is not useful, as the S/N will decrease again when adding more UV faint objects to the stack.

Since the S/N ratio is not the value I am interested in, but rather the flux ratio between Lyman and UV continuum or the escape fraction of the LyC, I now want to understand how high the LyC escape fraction in our objects has to be to produce the shown S/N curve in figure 5.8. For this I make statistical calculations using the background noise in WFC3 F336W, the UV flux of the objects and assumptions on the escape fraction of LyC photons (see section 6.4 for a more detailed explanation of this process). I produce 1000 artificial apertures per object with random noise distributed according to the standard deviation in the WFC3 F336W band of $\sigma_{\text{bgr},\text{F336W}} = 1.532 \times 10^{-20} \text{ erg s}^{-1} \text{ cm}^{-2}$ for apertures with $0''.5$, and stack them with an increasing number of objects, just as shown in figure 5.8. This is repeated with escape fractions from $f_{\text{esc}} = 0$ to 1 in steps of 0.01 by adding flux to the artificial WFC3 F336W apertures that is a fraction of the measured UV continuum of the corresponding real objects, such that the desired escape fraction is mimicked. This way, I get a prediction of the expected S/N curves for different escape fraction values. Since the individual curves are noisy themselves and the S/N goes down after around 30 objects, I focus on the range of objects with 90% or higher of the maximum S/N, which is from 20 to 60 objects. This is a little shifted with respect to the maximum of the significance in figure 5.6, which peaks at 61 objects. Therefore I extend the range to 20 to 70 objects. I compute the mean S/N in this range, both in the real stacks and in the modelled stacks with varying LyC escape fractions. The result can be seen in figure 5.9, which can be read as a probability of having an underlying escape fraction at a specific S/N ratio in the stacks.

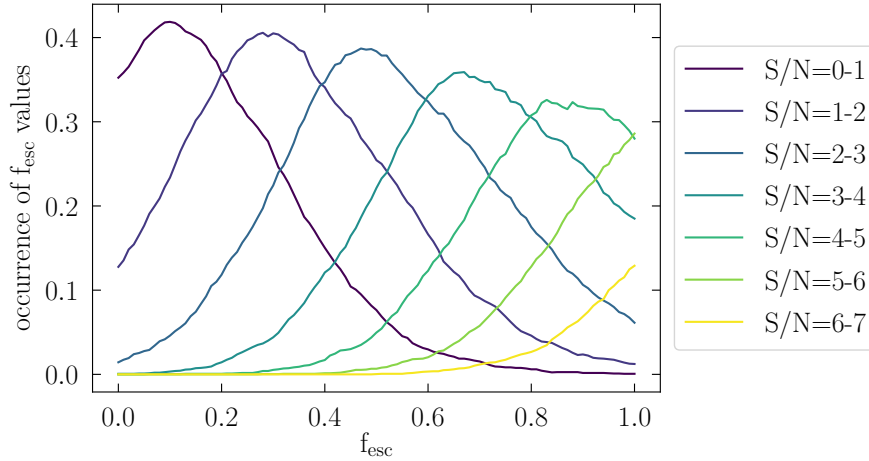


Figure 5.9.: Prediction of the underlying escape fraction for different S/N values in stacks, based on measuring the S/N of stacks of artificial LAEs. For this I used the measured UV continuum for each object in the selected sub-sample to derive a Lyman continuum flux using varying LyC escape fractions. The resulting LyC emission was added to an aperture with random noise (based on the background noise in the band WFC3 F336W) and the artificial objects were stacked. This was done 1000 times, each time measuring the S/N in the stack of artificial objects for which I know the underlying escape fraction in order to understand what S/N values can be produced from objects with different escape fractions. I then compared all stacks with the same S/N values and created distributions of the underlying escape fractions, which is shown in this plot. The x -axis shows the escape fractions in steps of $\Delta f_{\text{esc}} = 0.01$, the y -axis shows the percentage of realisations (out of 1000) that produced a certain mean S/N ratio in the LyC band WFC3 F337W (colour coded, see the legend to the right) in the range between 20 and 70 objects at the given escape fraction on the x -axis. The 1000 different realisations used the UV continuum of the real objects as a basis to compute the Lyman continuum emission, given a certain escape fraction and IGM absorption.

This gives me a method to estimate the escape fraction in the real stacks, which have a mean S/N in the range between 20 and 70 objects, sorted by the order without IGM correction without (with) candidates of $S/N=4.8$ ($S/N=5.5$). The most likely underlying escape fraction is thus $f_{\text{esc}} = 80\%$ according to figure 5.9, but I can rule out an escape fraction of $f_{\text{esc}} = 0\%$ since the probability goes down to 0 and an escape fraction of $f_{\text{esc}} < 50\%$ is ruled out roughly by a 1σ margin.

In the following, I use the 60 objects (plus the two individual LyC leaker candidates not excluded by figure 5.6) with the highest contribution to the S/N ratio for the stacks, since adding more objects only decreases the S/N and makes the results more unreliable as I have discussed above. The result of the stacking can be seen in figure 5.10, where I show the stack in the Lyman continuum on the left and the stack in the UV continuum on the right. The signal in the Lyman continuum stack has a S/N of over 5, but it is offset by ~ 0.2 . This raises doubt about the genuineness of the signal, since the offset is larger than e.g. a pixel shift caused by an indexing problem. There could be several causes, including a problem with the positions from `GalFit`, an error in translating the coordinates to the WFC3 filterband, an inconsistency with the world coordinate system of the WFC3 band, a flaw in the masking process or a few bright, insufficiently masked objects. I discuss all these possible trapdoors in section 6.5 of this thesis and rule out each of them through a number of tests with other bands, other kinds of objects and bootstrapping experiments. None of the possible explanations seem to apply, though and for the time being, the origin of this mysterious offset is not clear. If it the offset is caused by an error in positioning or in the method described here but the signal is still real, the LyC escape fraction is likely very high and on the order of $f_{\text{esc}} = 80\%$. If however the the signal itself was caused by a chance alignment of noise in the data, creating an unlikely, but not impossible high signal-to-noise peak, this would also explain why the derived escape fraction has such a high value, much above most literature escape fraction. Whatever the reason for this offset, in the following I treat the signal as real and explain its implications under this assumption.

I proceed to group the LAEs into two redshift bins (half of the objects with a lower redshift and the other half with higher redshifts) to see if there is a trend (see figure 5.11). On the one hand, at higher redshifts the IGM gets less transparent for LyC photons but on the other hand, the higher the redshift of the object, the lower the wavelength range covered in the WFC3 F336W band. Since the LyC initially increases towards

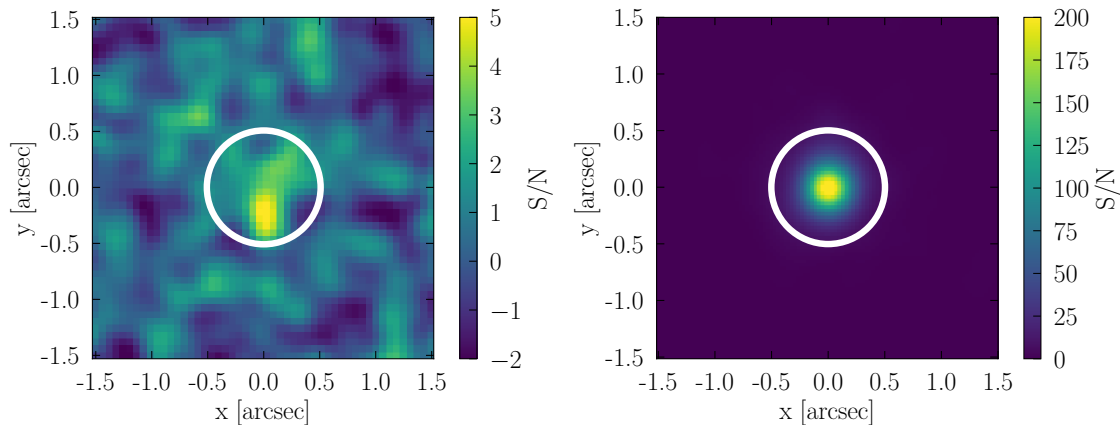


Figure 5.10.: Left panel: Mean stack of 62 objects, including the individual LyC candidates in the WFC3 F336W band. This cut-out has a size of $3'' \times 3''$, the white circle marks the centre (based on the rest-UV positions of the objects) and has a radius of $0''.5$. There is a clear offset in the signal in y -direction. Right panel: The same objects, but in the band ACS F775W for the MUSE-Wide objects and ACS F814W for the MUSE-Deep objects. Note the different S/N scales in the colourbars to the right of the panels.

the blue, this would potentially increase the S/N in the stacks and could explain why we see a detection not only in the lower redshift bin, but also in the higher one.

5.6. Stacking Lyman α Emitter Subsets Based on the Lyman α Line

Similar to what was done in figure 5.11 I now create subsets of objects based on the properties of the Lyman α line. The aim of grouping the LAEs in this way is to boost the LyC signal by selecting objects that have a higher predicted chance of leaking LyC emission, as well as to find the Lyman α line properties that best predict which LAEs can be LyC leaking. Therefore, I stack different groups of objects based on their line properties. This ties in with observations of Verhamme et al. (2017), Fletcher et al. (2019), Marchi et al. (2018) and Izotov et al. (2018b) and predictions by Verhamme et al. (2015). Since I pre-select objects that could contribute signal to the Lyman continuum stack, as explained in section 5.5.1, I again divide the sample of 60 objects (plus two candidates) into two groups for each line property: The half of the selected sample with the higher value of the property and the half with the lower value (for equivalent widths, FWHM and asymmetry). In the case of the peak separation, I stack the part of the selected sample with a double peak and the part without. The result is shown in figure 5.12 and 5.13 and in this section I describe the measurement of the different line properties.

5.6.1. Spectral Properties

I fit an asymmetric function (see equation 5.6) to the PSF-weighted one dimensional spectra of all LAEs and derive the Lyman α redshift, the FWHM of the line and the asymmetry (see section 4.3 for a detailed discussion on the process of fitting the line). In case of a double peak (blue bump), I use a combination of two asymmetric functions to derive the peak separation. From the line flux and the continuum flux density (measured in the HST photometry) I can derive the equivalent widths.

For MUSE-Wide the PSF-weighted extracted spectra are not flux conserving and do not fully include any possible extended Lyman α emission. Increasing the aperture size would also increase the noise in the spectrum, though. If certain line shape features, such as the blue bumps, are different in the outer Lyman α halos than in the central parts, this is not recovered by this method. However, since the LyC emission can only escape through either a low neutral hydrogen column density or holes in the ISM of the LAE and thus from the central regions, it is reasonable to exclude contributions from the Lyman α halo. These spectra are therefore only used for the line shape but not for the line flux. For MUSE-Wide, the line flux is measured by LSDCat using the three-dimensional information of the source in an aperture of three Kron radii, thus including most of the flux in the Lyman α halo (Herenz et al., 2017). The line flux for the MUSE-Deep objects was measured from spectra based on the segmentation maps from Rafelski et al. (2015) which were

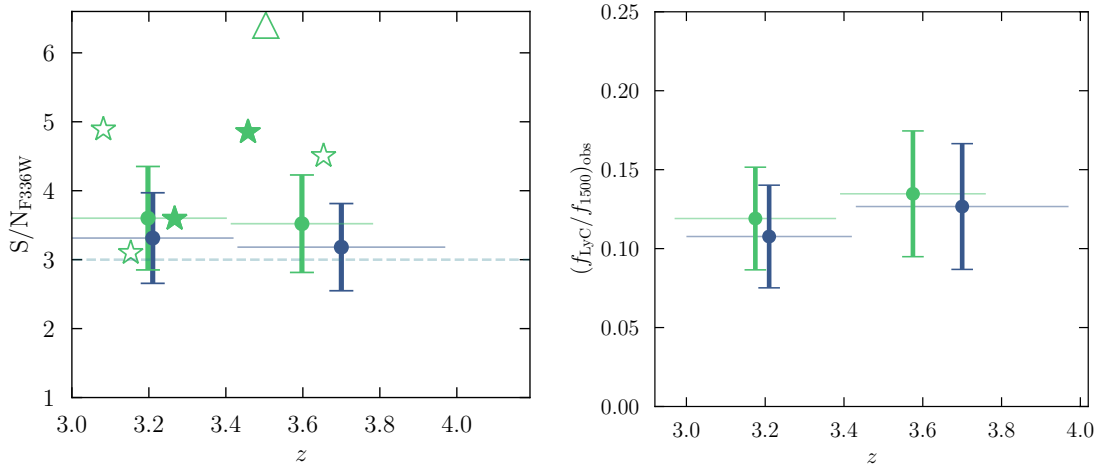


Figure 5.11.: S/N (left panel) and LyC flux to UV continuum flux at 1500 \AA (right panel) of stacks for two different redshift bins. The ratio f_{LyC}/f_{1500} is directly proportional to the escape fraction (see equation 5.4), but does not rely on assumptions about the intrinsic ratio or the IGM absorption and is thus directly measurable. The green dots show the stacks including the LyC leaker candidates, the blue dots the stacks without the candidates (shifted in x -direction for clarity). Each bin contains half of the stacked objects (30). There is no significant difference (within the errorbars) between the stacks in the two redshift bins. The stars in the left panel show individual LyC leaker candidates (the likely ones with filled stars) and the dashed line indicates a S/N ratio of 3.

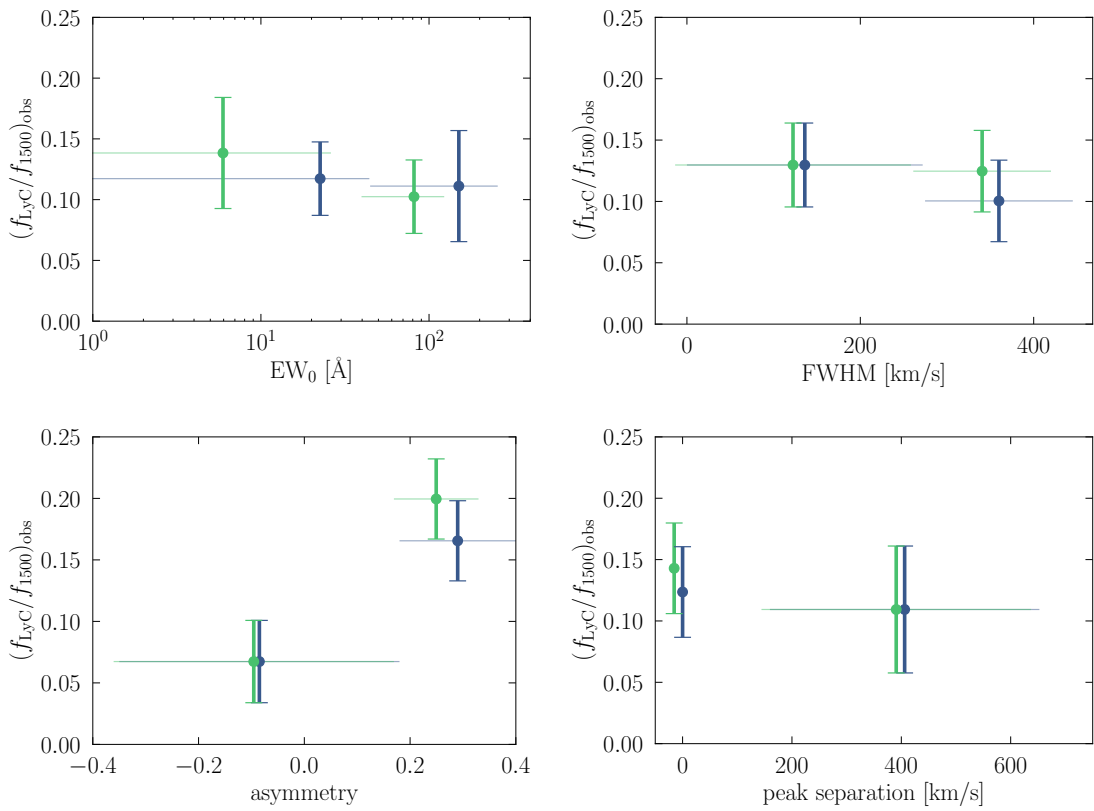


Figure 5.12.: LyC flux to UV continuum flux at 1500 \AA of stacks for different parameters (similar to figure 5.11). Top left panel: EW_0 , top right panel: FWHM, bottom left panel: peak separation and bottom right panel: asymmetry. The symbols and colours are the same as in figure 5.11.

convolved with the MUSE PSF (Inami et al., 2017 and see section 3.4.1 in this thesis).

I fit the Lyman α line with an asymmetric Gaussian from Shibuya et al. (2014):

$$f(\lambda) = A \exp\left(-\frac{(\lambda - \lambda_0)^2}{2\sigma_{\text{asym}}^2}\right) + f_0 \quad (5.6)$$

Similar to a normal Gaussian function, A is the amplitude and f_0 is the continuum level. The latter is not fitted, since we are using median filter subtracted spectra. A first guess for the central wavelength of the line, λ_0 , is taken from the LSDCat catalogue. The asymmetric dispersion σ_{asym} expresses the asymmetry of the line:

$$\sigma_{\text{asym}} = a_{\text{asym}}(\lambda - \lambda_0) + d \quad (5.7)$$

Here, d is the typical width of the emission line and a_{asym} is the asymmetry parameter. As a first guess for the amplitude A of the line, I use twice the value in the spectrum at the line position from LSDCat (since if the line position is not found perfectly by LSDCat, the value at that position will be lower than the actual line peak value). Just as in Shibuya et al. (2014), the first guess for the asymmetry parameter a_{asym} is set to 4.0 and for the line width d to 0.8 for each line. For lines where a double peak is identified (see section 5.6.1 and 4.3.1), the first guess for a blue bump is 7 Å to the blue of the main peak, with an amplitude half of that of the main peak, with the same input parameters for width and asymmetry (but with a blue wing). These first guess input parameters are tested and modified manually where the fit failed.

I look at the possible connection between the flux density ratio $(f_{\text{LyC}}/f_{\text{UVc}})_{\text{obs}}$ and the asymmetry of the line in the bottom right panel of figure 5.12. The flux density ratio is significantly higher in the half of the objects with a high asymmetry, meaning a pronounced red wing of the main peak, which would indicate that an asymmetric line might be favourable for LyC escape.

Estimating the Double Peak Fraction

The gas kinematics in the ISM of the galaxy can influence the shape of the Lyman α line and produce a double peaked line. Several studies have used radiative transfer codes to simulate Lyman α lines and explain the different observed line shapes (e.g. Neufeld, 1991, Dijkstra et al., 2006a, Dijkstra et al., 2006b, Verhamme et al., 2006, Verhamme et al., 2017, Gronke et al., 2015). A simple model using either an expanding (or an infalling) gas shell can explain a stronger red (or blue) peak (see e.g. Gronke, 2017 for models of MUSE-Wide LAEs from Herenz et al., 2017). Thus, the ratio between the line peaks is an indication of the underlying gas kinematics⁴. The peak separation is determined by the neutral hydrogen column density of the gas surrounding the star-forming region in which the ionising photons are produced. The higher the neutral hydrogen column density, the more often the photons have to be scattered to shift enough in frequency to be out of resonance and escape, which is why the line separation gets broader. To find blue bumps in the spectra and to determine the fraction of Lyman α lines that are double peaked compared to the single peaked ones (the double peak fraction), each LAE spectrum is visually inspected (see section 4.3.1).

In cases where I detect a blue bump, I use the wavelength midway between the two peaks for estimating the redshift (according to Verhamme et al., 2018), as this is a good approximation for the systemic redshift. Where there is no double peak, I use the FWHM of the line to estimate the redshift (see section 1.3.3 in the introduction of this thesis).

Peak Separation and Full Width at Half Maximum

The resolution of MUSE sets a lower limit on the measurable peak separation, as does the LSF for the measurable FWHM. One MUSE spectral bin is 1.25 Å, which translates to a peak separation of 78 km s⁻¹ at $z \sim 2.91$ and of 40 km s⁻¹ at $z \sim 6.65$, the redshift range of Lyman α covered by the wavelength range of MUSE (4750 Å to 9300 Å). The LSF is given in Bacon et al. (2017) and corresponds to a spectral resolution of $R = 1617$ at 4800 Å and $R = 3597$ at 9300 Å, I get a minimum resolvable FWHM of 184 km s⁻¹ and 83 km s⁻¹ respectively (see also section 4.3).

To correctly measure the peak separation, which is the velocity or wavelength distance between the red and the blue peak, I would need the systemic redshift, which I estimate using the middle of the double peaks (if available) or the FWHM as described above. The error made by assuming a wrong line position

⁴Note that Gronke & Dijkstra (2016) show that more complicated multiphase models of the ISM also produce blue and red peaks and the parameters derived from the shell models do not match the parameters of the multiphase models well.

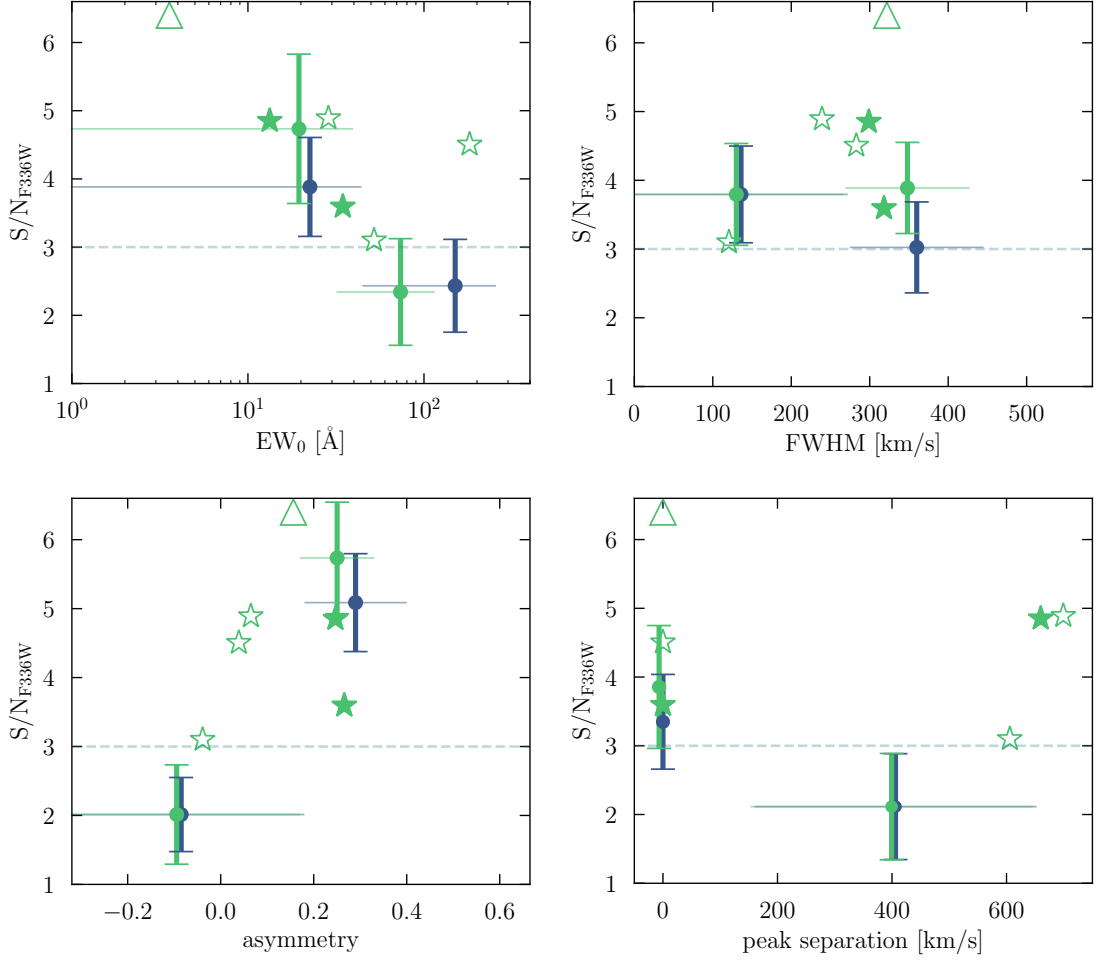


Figure 5.13.: Same as figure 5.12 and the left panel in figure 5.11 but showing the S/N for the stacks in two bins of different Lyman α line properties. There are significant detections in stacks with low EWs, for both FWHM stacks, for asymmetric lines and for Lyman α lines without double peaks.

is small compared to the influence of the peak separation on the final velocity separation of the two lines (which can be around 40 km s^{-1} for a measured peak separation that is off by one MUSE spectral bin of 1.25 \AA at 9300 \AA). I measure the FWHM of the line (taking the LSF into account) directly from the fit of the spectrum and measuring the width at the wavelength where the flux has dropped to half its maximum value, since the parameter d in equation 5.6 cannot be easily translated into the FWHM.

There seems to be a slight trend with higher LyC emission with respect to UV continuum in the group with smaller FWHM of the emission lines (top right panel of figure 5.12), which is not significant within the errorbars however. From predictions we would indeed expect more LyC leakage for narrower lines, as this indicates easy escape of Lyman α photons and little scattering by neutral hydrogen (Verhamme et al., 2015). The difference between the two stacks is not as clear as predictions would have suggested, though.

For the peak separation, I find a slightly stronger signal in the group corresponding to single lines (the bottom left panel in figure 5.12), as the first group contains all objects with no double peak. The group with the doubled peaked lines shows a signal as well.

Equivalent Widths

For the rest-frame equivalent widths (EW_0) I use the line flux $F_{\lambda, \text{Ly}\alpha}^{\text{line}}$ from the MUSE data and the continuum flux density $f_{\lambda, \text{Ly}\alpha}^{\text{cont}}$ from the HST broad-band data. Then the rest-frame equivalent width EW_0 is:

$$EW_0 = \frac{F_{\lambda, \text{Ly}\alpha}^{\text{line}}}{f_{\lambda, \text{Ly}\alpha}^{\text{cont}}} \frac{1}{(z+1)} = \frac{EW}{(z+1)} \quad (5.8)$$

For the continuum flux density $f_{\lambda, \text{Ly}\alpha}^{\text{cont}}$, the MUSE data is not deep enough, therefore I measure it in several HST bands, to also obtain the UV slope (β -parameter) with which I can correct the flux density (measured in the HST bands) to the Lyman α wavelength (section 4.1.2 contains a discussion of the β -parameter).

$$\beta = \frac{\log_{10}\left(\frac{f_{\lambda_1}}{f_{\lambda_2}}\right)}{\log_{10}\left(\frac{\lambda_1}{\lambda_2}\right)} \quad (5.9)$$

Since not all objects had already been found and measured in previous surveys, I fit each object using a custom `GalFit`-wrapper with Sérsic profiles (see section 4.1). I initially fit each object in the ACS F814W band to get the shape parameters and use them in the other bands as fixed priors and only leave the magnitudes as free parameters. Due to the different depths and continuum slopes, not all objects have reliable measurements in all bands. For the β -slope, I use all bands with reliable continuum measurements to the red side of the band containing the Lyman α emission, so as to avoid contamination by the line itself. While I could subtract the Lyman α emission line flux from the band containing the line, this introduces uncertainties due to the extended Lyman α halos, which would have to be taken into account as well. I then fit a linear relation to the logarithm of the individual flux measurements in those bands. The scatter in the β -slopes measured that way increases with increasing absolute magnitude (see for example figure 2 in Hashimoto et al., 2017, who analyse the LAE equivalent width distribution in the MUSE-Deep LAEs). This scatter is likely mostly due to measurement errors rather than intrinsic differences in the spectral slopes, as the standard deviation of the β -slope measurements is $\beta_{\text{med, std}} = 0.55$. I therefore decide to adopt the same β -slope for all EW measurements, using the median of the sigma-clipped β -slope distribution $\beta_{\text{med}} = -1.92$, which is close to the canonical value of -2 (see section 4.1.2 for a more detailed discussion).

The equivalent width distribution has been shown and discussed already in the previous part of this thesis, in chapter 3 and the corresponding methods chapter 4, where I show that 9% of LAEs have equivalent widths over 240 Å in the full sample used here, which is usually quoted as the upper limit for a normal stellar population (e.g. Raiter et al., 2010). Higher values would indicate special conditions such as low metallicity, a top-heavy initial mass function (IMF) or young ages. Of the 60 selected objects for stacking, only one has an EW_0 over 240 Å. This can be explained both by our redshift cut and by the selection procedure, which favours objects with a bright UV continuum.

Also, the conditions for the escape of Lyman α photons have to be ideal to produce such high EW_0 . While Verhamme et al. (2017) have found a correlation between Lyman α and LyC escape fraction (and thus EW) for local LyC leaking galaxies up to EW_0 s around 100 Å, recently Steidel et al. (2018) have discussed that in their study this is only the case for lower EW_0 of up to 50 Å. A high Lyman α EW_0 means that more LyC photons were converted to Lyman α in neutral hydrogen clouds and could not escape the galaxy (see also Raiter et al., 2010).

Looking at the ratio between LyC and UV continuum in the two stacks with different EW_0 shows no clear trend (see top left panel in figure 5.12), contrary to other studies (see Marchi et al., 2018, Verhamme et al., 2017, Verhamme et al., 2015, Steidel et al., 2018, Fletcher et al., 2019). There is a signal in both groups and the group with lower EW_0 has a slightly higher LyC over UV continuum flux fraction. However, only one of our six individual LyC candidates has an $\text{EW}_0 > 50$ Å. The LAEs with the potentially highest EW_0 s but without counterpart in the rest-frame UV continuum HST bands are not included in the plot as explained in 5.5.1.

5.7. Interlopers, Superpositions and Error Estimation

As the offset in the detected LyC signal gives reason to doubt whether the offset signal is real, it is important to understand how interlopers and superpositions can influence the stacking process. There are three main causes for erroneously including objects in the stacks that might distort the results: (i) including noise that was misidentified as an LAE (spurious detections), (ii) including objects that were classified as LAEs but are actually low-redshift interlopers and (iii) including LAEs that are superposed with another, low-redshift object whose continuum can mimic the LyC of the LAE.

It is especially important to exclude low-redshift interlopers and superpositions from our sample of LAEs, as they can display the expected SED of a LyC leaker. As mentioned in section 5.2.1 and also in chapter 2, we classified the emission line objects visually. Signs of LAEs are an asymmetric line shape due to either IGM absorption of the blue part or a high neutral hydrogen fraction of the ISM or CGM of the host galaxy, a double peaked line, again with possible asymmetry, no other lines in the spectrum visible, a compact source

(as seen in the HST data) and an extended Lyman α halo, which is common for LAEs (see e.g. Wisotzki et al., 2016, Leclercq et al., 2017b).

If I add the ancillary HST data, I can make use of its high spatial resolution as well to determine the nature of a galaxy. Any object in the HST data within $0''.5$ of the detected emission line is considered to be a possible counterpart, since 83% of all emission line objects in the MUSE-Wide survey have a counterpart in the Skelton et al. (2014) catalogue within $0''.5$ (Herenz et al., 2017). This way, I am able to visually check the SEDs by looking at multiple HST bands, since high-redshift LAEs are expected to show a drop in the spectrum at the Lyman limit (although Green Peas do not show this behaviour).

With the prior knowledge of the position of the LAE and the high resolution of the HST images, one would be quick to rule out the possibility of an interloper based on the concurrence of the possible LyC detection and the Lyman α emission in HST. However, in one case the LAE lies directly on top of an [O III] $\lambda 5007$ emitter, only visible through the [O III] $\lambda 5007$ emission line in the MUSE spectrum. Without the additional information from MUSE, such an object would have been classified as a clear LyC detection. In cases where objects without additional emission lines are exactly in the line-of-sight to the LAE in question, not even the MUSE data help to rule out such interlopers. In order to estimate how often cases like these might occur, I compare our shallower MUSE-Wide data to the MUSE-Deep data (see section 5.7.1) and also do a test with random sky positions (see section 5.7.2).

Thus, our fixed criteria for ruling out interlopers were the proximity between the signal in the WFC3 F336W band and the MUSE detection and the absence of any other emission lines in the MUSE spectrum. A drop in flux in the band bluewards of Lyman α is also an indicator that the object is indeed an LAE.

5.7.1. Estimating the Reliability of Lyman α Emitter Classifications

To assess the statistical reliability of our classifications in the MUSE-Wide data, we produced shallow versions (1.6 hours integration time) of the MUSE-Deep fields (that have 10 hour integration time) and compare the MUSE-Wide classifications to the MUSE-Deep classifications (see Urrutia et al., 2019 for a detailed discussion). There are a total of 676 objects in the overlapping region, 283 of which are classified as LAEs. Comparing the catalogues shows that 30 objects were classified as LAEs in the MUSE-Wide catalogue that are not in the MUSE-Deep data, which suggests that they are spurious sources in the MUSE-Wide catalogue. This amounts to 10.6% of the LAEs in the MUSE-Wide data. Most of these sources have a signal-to-noise ratio of close to five, which is the detection limit we set. Since they do not have an HST counterpart, such spurious sources do not contaminate the sample of LyC candidates and would not enter in the stacking process.

There are also real objects that were misclassified in the MUSE-Wide data compared to the MUSE-Deep catalogue. In total, 30 objects were given different classifications, which is 4.4%. In 12 cases, an LAE was wrongly classified as either an [O III] or [O II] emitter and in two cases a C III emitter was classified as an LAE. Since the MUSE-Deep classifications are based on data with a longer exposure time and thus a higher S/N in the lines, we assume them to be correct.

From these analysis it seems that in general it is more likely to misclassify an LAE as something else (thus missing it and excluding it from the stacks) than the other way around, which only happens in 0.7% of cases. Thus, I can conclude that it is highly unlikely that any of the individual LyC candidates is such a misclassification, while for the stacking sample, at most one object might be such an interloper.

5.7.2. Random Positions

To estimate the possible contamination with interlopers or false detection of our sample, I use random positions in the sky. At some of these random positions, there will likely be real objects that are indistinguishable from LAEs. I can assume that the number of real objects at the random positions is the same as the number of chance superpositions of real objects with real LAEs. These objects would contaminate the WFC3 F336W band and would be wrongly identified as LyC emission. Estimating the number of real objects I detect in the random positions thus helps to estimate the number of contaminating superpositions. In addition to understanding how many of the LyC leaker candidates might be such superpositions, I also use the cut-outs at the random positions to stack in the same way as for the real objects, to see if I detect any signal in any of the groups at supposedly empty positions. This could be caused by insufficient masking of neighbouring objects or cosmic rays.

For each real object, I use a random position in the same MUSE field, only excluding positions at the edge of the FoV. I then treat it the same way as I would a real object, going through the same process of

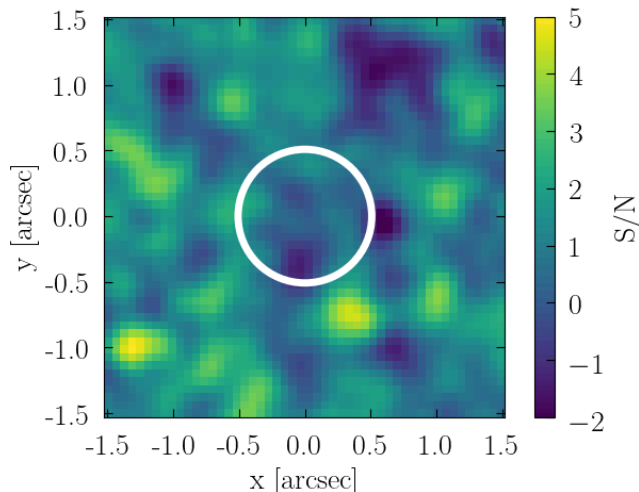


Figure 5.14.: Same as figure 5.10, but with random positions. There is no significant signal.

rejecting possible interlopers and masking bright neighbours as described above (see section 5.7). This leaves us with two detections out of 323 positions with a signal-to-noise ratio over three in the LyC band WFC3 F336W. Thus I can expect that at most two of our LyC leaker candidates might be contaminated by interlopers because they are detected at a similar S/N ratio level. The results of the stacking can be seen in figure 5.14, with a signal-to-noise ratio of 0.35. In the stack with the random positions there is no detection and no offset and also no noise peak in the cut-out that has the same strength as the offset signal I see in the stack of the real objects (in figure 5.10).

5.7.3. Ruling out AGN Contamination

As mentioned in the introduction, some studies (e.g. Giallongo et al. (2015) and Grazian et al. (2018)) have suggested a large population of faint AGN as contributors for the ionising photons at the EoR. To exclude the possibility of a contamination of our sample by AGN, there are several possible examinations.

First, I inspected the Lyman α line profiles to look for broad Lyman α lines that are typical for type I AGN (Osterbrock & Ferland, 2006) often with FWHMs of several thousand km s^{-1} . For the MUSE-Wide sample, there are two known quasars: ID 104014050 and ID 115003085 (which is at a redshift of $z = 5$ and therefore excluded anyway) with a FWHM of 615 km s^{-1} and 1025 km s^{-1} , respectively (see Norman et al., 2002 and Mainieri et al., 2005 who already discovered them). Both are excluded from the stacks.

Second, we cross-matched the MUSE-Wide objects with an X-ray catalogue from the CDF-S 7 Ms Chandra observations (Luo et al., 2017) but do not find any new AGN at redshifts $z > 3$, apart from the two AGN mentioned above (for a detailed discussion, see Urrutia et al., 2019). In order to find a possible contribution of faint AGN, we stacked the $z > 3$ objects from MUSE-Wide in X-rays using the CDF-S X-ray data, but again find no detection (see section 4.5.1, figure 13 in Urrutia et al., 2019). This means the LAEs are not dominated by an AGN population, but it does not exclude the presence of a few, faint AGN.

Third, I stack the MUSE spectra of our LAE sample and search for C IV, He II, N v and C III] emission, which would hint at photons with energies above 47.9 eV, 24.6 eV, 77.5 eV and 24.4 eV respectively⁵, as they can be used for AGN diagnostics (e.g. Feltre et al., 2016). Although these emission lines could also be produced by normal star-forming galaxies without an AGN, the line ratios and FWHM can rule out significant AGN contamination. As a redshift estimation, I cannot use the peak of the Lyman α line, as it can be shifted with respect to the systemic redshift. Instead I use the corrected redshifts based on the prescriptions by Verhamme et al. (2018) (see introduction section 1.3.3) using the correlation between the systemic redshift and the mean wavelength of the two peaks (in case there is a blue bump) and between the systemic redshift and the FWHM. In order to minimise the influence of noisy parts of the spectrum (caused e.g. by sky lines), I mask any wavelength bin with a standard deviation of more than five in the corresponding error spectrum. As can be seen in figure 5.15, there is a clear detection of both the C IV doublet and the C III] doublet, but no significant He II or N v lines. Table 5.3 gives an overview of the

⁵Ionisation energies taken from <https://physics.nist.gov/PhysRefData/ASD/ionEnergy.html>

Lines	S/N	Ratio	FWHM
C IV ₁₅₄₈ , C IV ₁₅₅₁	9.6	2.0	228 km s ⁻¹
C III] ₁₉₀₇ , C III] ₁₉₀₉	15.8	1.1	185 km s ⁻¹
C IV ₁₅₅₀ , C III] ₁₉₀₈		~ 0.48	

Table 5.3.: Line ratios of C III] and C IV in the stacked spectra.

different line ratios and FWHM of the C IV and C III] lines. The narrowness of the lines is a clear indication against a dominant AGN contamination of our sample.

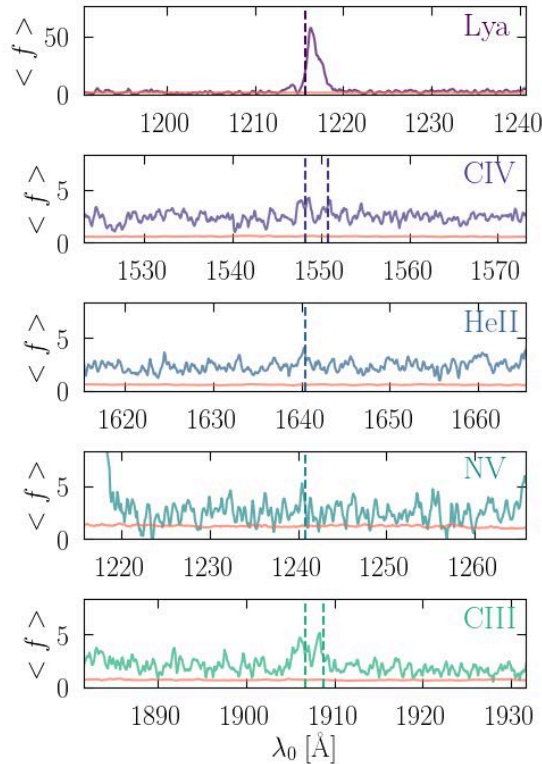


Figure 5.15.: Mean flux from stacks of the sample of 323 LAEs (including the individual LyC leaker candidates) over rest-frame wavelength in \AA . The top panel shows the mean Lyman α line, the second panel the wavelength range around the C IV line, the third panel the wavelength range around the He II line, the fourth panel the NV line and the bottom panel the C III] lines. The positions of the emission lines are marked with dashed lines. Note the different scaling on the y-axis for the Lyman α line, which is much stronger. The red line shows the noise spectrum. I show the sample of 323 LAE (which have been selected to have redshifts between $3 < z < 4$ and detections in at least two HST bands) instead of the selected 60 objects, because for this plot the potential for LyC signal is not important and faint spectra will contribute less in the mean flux anyway.

5.8. Discussion

I find six individual objects with possible LyC emission. Given that the sample consists of 323 LAEs, this means I find LyC emission in only $\sim 2\%$ of the objects. This is in contrast to a recent study by [Fletcher et al. \(2019\)](#), who find LyC emission in $\sim 30\%$ of their sample of narrow-band selected LAEs at a redshift of $z \sim 3.1$. Apart from a different sample selection, [Fletcher et al. \(2019\)](#) also allow slightly larger spatial offsets between the Lyman α emission and the UV or Lyman continuum (up to $0''.8$), while I set the limit to $0''.5$. However, as can be seen in their figure three, most of their objects have smaller spatial separations. The addition of the WFC3 F275W band for this study allows me to exclude interlopers based on their bright

emission in this band, which is not done in [Fletcher et al. \(2019\)](#) and could contribute to the difference in the number of LyC leaker candidates.

In addition to identifying individual LyC leaker candidates, I have grouped the LAEs based on their Lyman α line properties and stacked the groups in order to find possible correlations between the line properties and LyC leakage (see figure 5.12).

Based on the tentative connections I find between the Lyman α line shape and LyC emission, the ideal LyC leaker candidate has a Lyman α line which is asymmetric while the other line shape properties seem to have little influence on the strength of the LyC emission. A possible explanation for this result, which is in slight tension with theoretical predictions and other studies (see [Marchi et al., 2018](#), [Verhamme et al., 2017](#), [Verhamme et al., 2015](#), [Steidel et al., 2018](#), [Fletcher et al., 2019](#)), is that lower equivalent width objects are bright in the UV continuum and thus also in the LyC. By making a pre-selection in UV continuum flux, I might automatically select objects with lower equivalent widths. Similar studies, such as [Steidel et al. \(2018\)](#) who stack spectra of LBGs, find the expected correlation between Lyman α EW and LyC escape fraction. However, the group containing the smaller equivalent width, contains objects with $0 < EW_0 < 40 \text{ \AA}$, while their sample only reaches $EW_0 = 50 \text{ \AA}$, and also includes Lyman α absorbers. They argue that there might be a turnover at around 50 \AA in the correlation between LyC escape fraction and Lyman α EW. [Fletcher et al. \(2019\)](#) also find a tentative correlation between Lyman α EW and LyC emission, although with a large scatter. Their sample reaches equivalent widths up to 150 \AA . In conclusion, it seems difficult to infer possible LyC leakage directly from properties of the Lyman α emission line. Especially the correlation between Lyman α rest-frame equivalent width and LyC leakage up to high equivalent widths as seen in low-redshift Lyman continuum leakers ([Verhamme et al., 2017](#)) could not be seen as clearly here. Galaxies with high Lyman α EWs have typically large Lyman α fluxes and/or faint UV continua. If the latter is the case, the LyC might be too weak and not be observable in the data at hand. If the former is the case, this might be due to LyC photons having been converted to Lyman α photons, increasing the Lyman α flux and subsequently reducing the LyC escape.

5.9. Summary

In this chapter I explore the possibility of stacking HDUV images of LAEs found in the MUSE-Wide and -Deep GTO surveys to estimate the LyC escape fraction. In addition to stacking a selected sample (with restrictions on the redshift and UV continuum magnitude), I also stack subgroups of objects based on their Lyman α line properties. The main results are as follows:

- I find six individual LyC leaking objects, two of which are promising candidates.
- There is a significant detection of LyC for the mean stack of the selected sample of promising objects with $S/N = 5.5$. Even if I exclude the individual candidates, the S/N ratio reaches 4.8 and is a significant detection as well. This points towards LyC leakage in more objects than the individual candidates.
- The detected signal in LyC has a significant offset of $\sim 0''.2$, raising the question whether the signal could possibly be dominated by a chance alignment of noise.
- If I trust the found signal, the LyC leakage is strongest for objects whose Lyman α line is asymmetric, but there is no correlation between the LyC leakage and other properties of the Lyman α line.

A further step towards understanding the connection between LAEs and LyC emission would be to look at the Lyman α halo sizes of the individual candidates as well as the stack of the selected sample. If the Lyman α photons are not significantly scattered in the neutral hydrogen in the CGM of the galaxy, thus producing only a small Lyman α halo, this could also facilitate the escape of LyC emission and we would see a higher LyC escape fraction for more compact Lyman α halos.

Methods: Measuring Lyman Continuum Emission

In chapter 5 I have searched for Lyman continuum in Hubble Deep UV (HDUV) data from Lyman α Emitters (LAEs) found with Multi-Unit Spectroscopic Explorer (MUSE), both in stacks of objects and in individual objects. In order to perform this task, the objects were selected carefully, flux measurements and positions of objects were compared to the HDUV data catalogue, aperture sizes were chosen, the Hubble Space Telescope (HST) data was corrected for any background offsets, the noise was modelled to predict the expected signal from Lyman continuum leakage with respect to different escape fractions, stellar population models were used to determine the intrinsic ratio between ultra-violet (UV) continuum and Lyman continuum and the final object selection was screened for active galactic nucleus (AGN) contamination. All these additional tests and preparations were not included in detail in chapter 5, as it shows only the main analysis, but are nevertheless an important part of the process, which is why they are described in this methods chapter.

6.1. Comparison with the HDUV Catalogue

As a first step, I compare the HDUV and MUSE catalogues. Eventually, I am using the positional measurements done with `Galfit` on the ACS F814W band (see chapter 4, section 4.1 for a description of the `Galfit` fitting process) and circular apertures for flux measurements in the relevant bands (mostly WFC3 F336W for the Lyman continuum from the HDUV survey and ACS F775W and ACS F814W for the UV continuum), but to make sure the measurements make sense and the objects are correctly matched, the comparison between the two catalogues is useful.

6.1.1. Matching Catalogues

To check that the flux measurements in the HST bands (both for the Lyman continuum and UV continuum) are reasonable, I compare them to the measurements of the HDUV survey (Oesch et al., 2018)¹. For my own measurements I use circular apertures centered on the ACS F814W positions of the counterparts of the LAEs found in MUSE. This assumes that the Lyman continuum will follow a similar path through the galaxy as the UV continuum, as compared to the Lyman α photons, which will have a more complex radiative transfer through the interstellar medium (ISM) and circumgalactic medium (CGM) and thus might not perfectly align with the UV and Lyman continua (see introduction section 1.4.1). The circular apertures also ensure that the same area will contribute to the flux measurements in all bands, to be able to use the measurements for the flux ratios, comparing Lyman and UV continuum flux.

¹The catalogue I am comparing to can be found at <https://archive.stsci.edu/prepds/hduv/index.html>, under 'Source Catalogue'. I am using the one for the GOODS-S region

The HDUV data provide the HST bands WFC3 F275W and WFC3 F336W for probing the Lyman continuum region of the spectral energy distributions (SEDs) of the MUSE LAEs. They have a 5σ depth of 27.5 and 27.9 magnitudes respectively, measured with 10 000 apertures ($0''.4$ diameter) in empty sky regions per field (Oesch et al., 2018).

To match the catalogues, I use a maximum distance of $0''.5$ and assign the counterpart with the smallest distance within this radius to the MUSE ID. Of the 1288 LAEs in the MUSE-Wide and -Deep sample I use, there are no counterparts for 669 objects (also counting LAEs that are not covered by the area of the HDUV survey). Figure 6.1 shows the spatial offsets between the HDUV and MUSE positions (the latter are measured from the ACS F814W images using Galfit). They are the same positions that are used to measure the flux in circular apertures (see below).

6.1.2. Determining Aperture Sizes

When determining the aperture size for measuring fluxes, I need to make sure to use apertures large enough to account for the PSF but also for the possible positional shift between the ACS F814W and WFC3 F336W bands, which could be caused by differences in astrometry corrections. As can be seen in figure 6.1, the positional offsets are small. The FWHM of the PSF of both WFC3 F275W and WFC3 F336W is around $\sim 0''.1$ (Oesch et al., 2018) which is consistent with the mean (median) difference between the positions from HDUV and the ACS F814W Galfit positions of $0''.138$ ($0''.073$). It should also be noted that one pixel in the WFC3 bands has a size of $0''.06$.

Another factor to be taken into account when deciding on an aperture size is not only the positional precision, but also the sizes of the objects themselves. From modelling the objects with Galfit (Peng et al., 2002) in the ACS F814W band I know that the mean (median) effective radius of LAEs that could be fitted with Galfit² is $0''.105 \pm 0''.093$ ($0''.0741$), only marginally larger than the FWHM of the PSF in WFC3. However, combining the median effective radius r_e of the objects and the FWHM of the PSF FWHM_{PSF} I get a larger expected size in WFC3 F336W. As a simple estimation of the expected object sizes, I convert both to standard deviations with $r_e = 0.67 \sigma_r$ and $\text{FWHM}_{\text{PSF}} = 2 \sqrt{2 \ln 2} \sigma_{\text{PSF}}$ by assuming that the object and PSF are two Gaussians, thus I can estimate the mean expected apparent object size $\sigma_{\text{obj}, \text{WFC3}}$ as a convolution of two Gaussians:

$$\sigma_{\text{obj}} = \sqrt{\sigma_r^2 + \sigma_{\text{PSF}}^2} = \sqrt{(0.105''/0.67)^2 + (0.1''/(2 \sqrt{2 \ln 2}))^2} = 0.16'' \quad (6.1)$$

This gives us $\text{FWHM}_{\text{obj}, \text{WFC3}} = 2 \sqrt{2 \ln 2} 0.16'' = 0.38''$ as a mean expected FWHM of the objects in WFC3 F336W, which can be interpreted as a minimum aperture diameter for measuring fluxes in WFC3 F336W. Thus, for flux measurements in UV continuum and Lyman continuum bands I use an aperture of $0''.5$ (diameter), which encompasses the central positions of 80% of objects, as can be seen in figure 6.1. This aperture is a good compromise between being large enough to recover the full flux for the majority of objects and small enough to not include too much empty sky (which increases the noise).

6.1.3. Flux Comparison

After confirming that the positional offsets are small, I now compare the measured flux values. In figure 6.2 and 6.3 it can be seen that our aperture measurements tend to be smaller than the HDUV measured flux values. This can be explained by the fact that brighter objects will lose flux in aperture photometry, that can be more easily recovered with the HDUV method. For their catalogues, they use the same approach as Skelton et al. (2014) in the 3D-HST survey: Using SExtractor (Bertin & Arnouts, 1996), the sources are detected in a combination of the WFC3 infrared bands F125W, F140W and F160W. The IDs are matched to the 3D-HST catalogue, but only for the relevant area (since the HDUV footprint is smaller). To get the photometry of their objects, they use circular apertures of $0''.7$ (diameter), which is slightly larger than the $0.5''$ apertures we use. Additionally, they correct the aperture measurements to total fluxes by using a correction factor (the same as in the 3D-HST catalogue, see Skelton et al., 2014) based on the WFC3 F160W measurements, where they get the total flux from growth-curve measurements.

Since the flux measurements are consistent with the HDUV catalogue, I now analyse the standard deviation. As described in section 5.3, I measure the standard deviation from convolving with a Gaussian of

²I am excluding objects with an effective radius that was either fixed due to problems with the model, or had an error of the same size or larger than the effective radius itself.

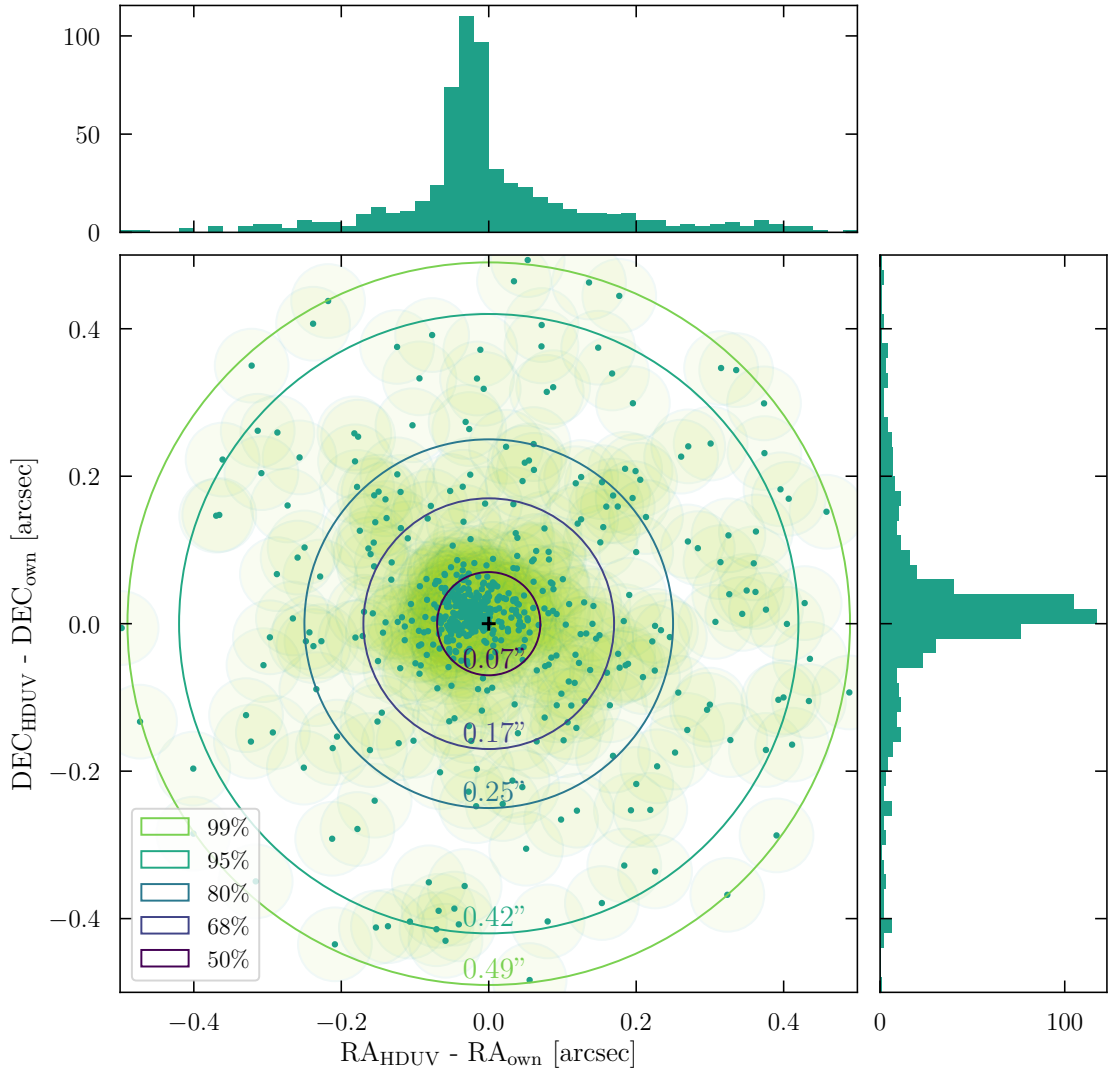


Figure 6.1.: Comparison between RA and DEC positions from the HDUV catalogue and my own `Galfit` positions in ACS F814W of the LAE counterparts found in MUSE (for a description on how the counterparts were determined, see section 3.3.2). The central panel shows the differences in positions, the histograms in the top and right panel show the distributions in RA and DEC offsets with a bin size of $0''.02$. The transparent light green circles in the main panel indicate the size of then FWHM of the PSF, which is around $\sim 0.1''$ (Oesch et al., 2018) in both the WFC3 F275W and WFC3 F336W images. This would be the sizes of the objects assuming point sources. The central panel also shows different aperture sizes (given as radius at the lowest points of the apertures in the middle panel) corresponding to different percentiles of objects within these apertures, as given in the legend. The outermost circle shows 99% of objects are within a distance of $0''.49$, but since objects are only counted as counterparts at distances of $0''.5$ or less, this is not surprising. The 80th percentile circle has a radius of $0''.25$ and 50% of the objects have counterparts closer than $0''.07$. As can be seen by the black cross in the middle of the panel, marking a distance of 0, and by the histograms which peak with a slight offset to 0, there seems to be a systematic difference between positions found in the HDUV survey and the `Galfit` positions. This offset has a mean positional shift of $0''.014$ between HDUV and MUSE positions. However, since the FWHM of the PSF is $\sim 0.1''$ and one pixel is $0''.06$, this positional shift can be safely ignored. It is not enough to explain the spatial offset found in the Lyman continuum stack, see section 6.5.

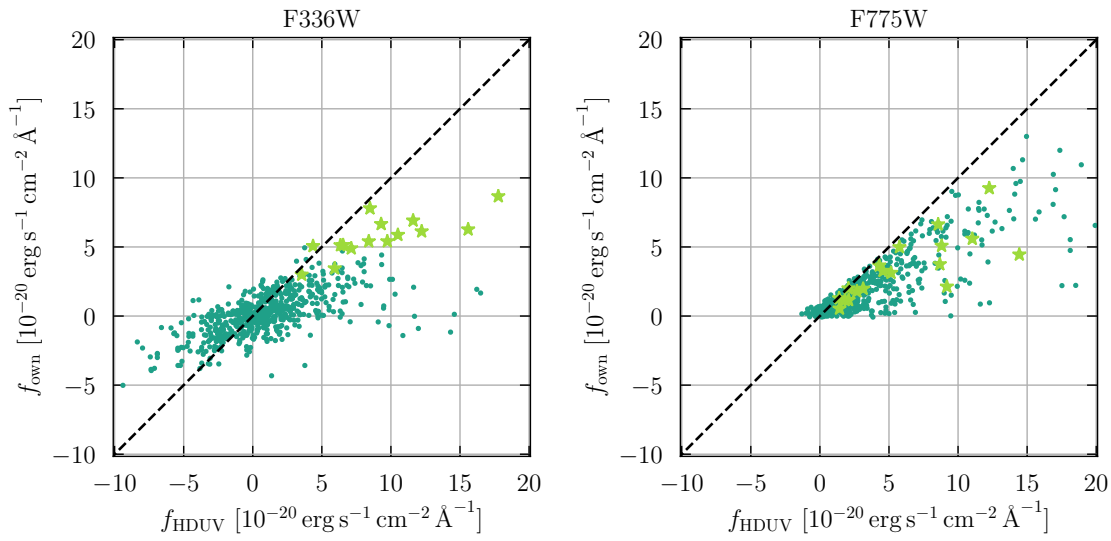


Figure 6.2.: Comparison between flux measurements from the HDUV survey and my own aperture measurements given in $f_{\lambda} = 10^{-20} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1}$. The bright green stars indicate the objects with a S/N ratio over three in the WFC3 F336W band (which includes the individually detected LyC leaker candidates). The left panel shows the flux values for the WFC3 F336W, the right panel for ACS F775W.

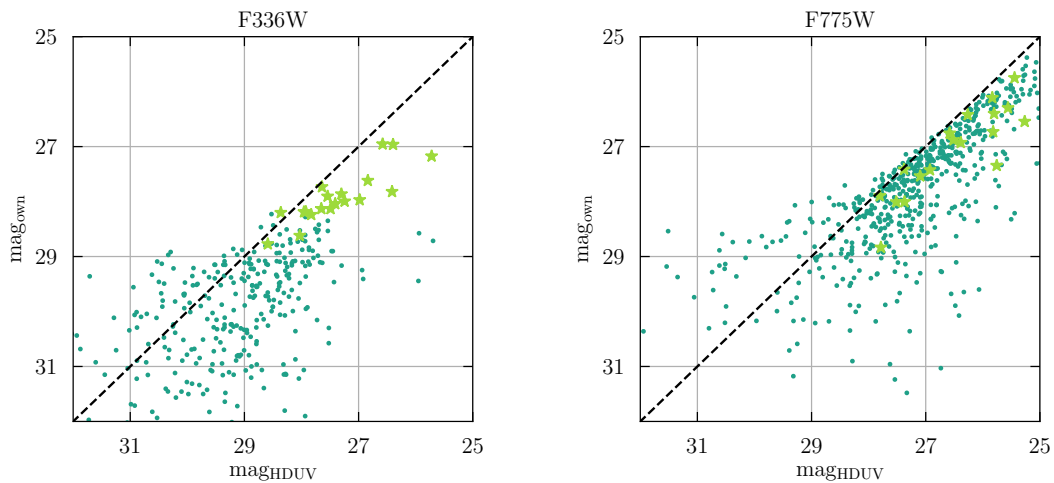


Figure 6.3.: Comparison between magnitude measurements from the HDUV survey and my own aperture measurements (similar to figure 6.2). The x -axis shows the magnitudes from the HDUV catalogue and the y -axis shows the magnitudes I measured myself. The left panel shows the flux values for the WFC3 F336W, the right panel for ACS F775W.

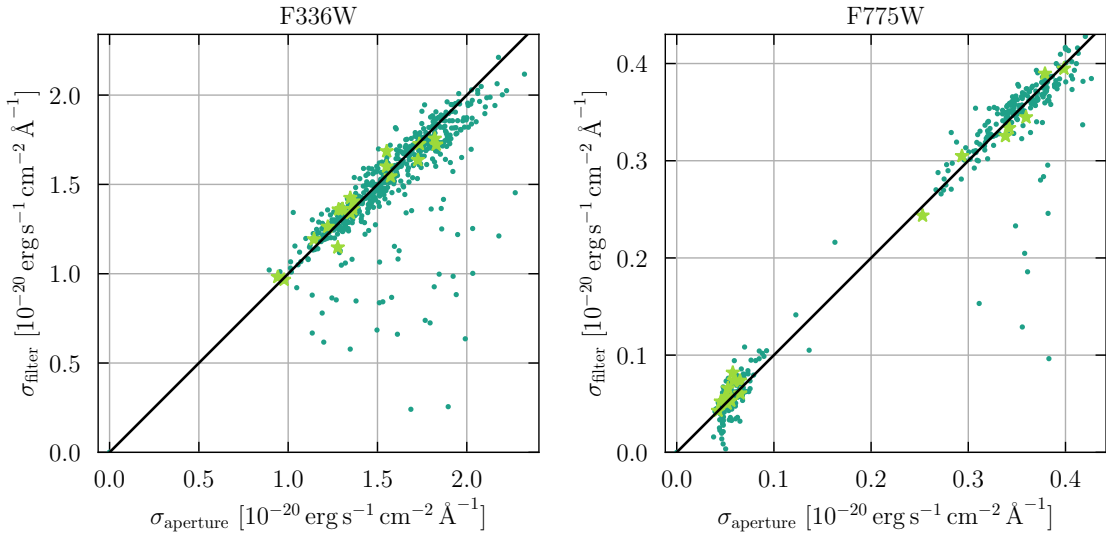


Figure 6.4.: Comparison between the standard deviation σ derived from apertures (x-axis) and filtering (y-axis) the data. Both methods to measure the standard deviation give consistent results. The smaller σ values in the ACS F775W band are in the UDF area where there is a longer observation time available in HST.

$\sigma_G = 0.1$ to get the S/N values. Another option is to measure the background noise in a number of random apertures of the same size as used for measuring the flux values. I compare both approaches in figure 6.4 which shows that for most objects, both methods give similar results.

In figure 6.5 the measured standard deviations are compared between the HDUV survey and my own aperture measurements for two different bands. The standard deviation of the aperture flux measurements in the WFC3 F336W band (excluding objects with a formal detection of $3\sigma \leq f_{\text{ap}}$) is $\sigma_{f,\text{ap}} = 1.58$, while the median of the standard deviations for the aperture flux measurements in the same band is $\langle \sigma_{f,\text{ap}} \rangle = 1.47$, which matches reasonably well, but cautions that on average the derived standard deviation for each flux measurement might be slightly underestimated. This is true if I assume that there is no signal in any of the aperture flux measurements in the WFC3 F336W band and what I measure is purely randomly distributed. Thus another explanation for the slightly higher standard deviation of the flux values than the median measured standard deviation for individual objects could be that there is indeed some real flux in some apertures (which matches the observations in the stack of the full sample, where I do see a detection). However, if I look at the WFC3 F275W band, the median of the standard deviations (excluding detections) is also slightly smaller at $\langle \sigma_{f,\text{ap}} \rangle = 3.57$ than the standard deviations of the measured flux values $\sigma_{f,\text{ap}} = 3.78$ and due to the increased opacity of the IGM I do not expect a significant signal in this band.

In figure 6.6 I compare directly my standard deviation measurements to the HDUV measurements. They do not seem to follow a correlation, with my error estimates typically smaller than the HDUV values. Especially in ACS F775W, where I use two different depths of data (deeper for MUSE-Deep), the error measurements are much below those by the HDUV survey. The fact that the errors are uncorrelated either means there is an inconsistency in the comparison or that the noise values are determined by random processes and not inherently different for each object (except for the two different depths of the data in ACS F775W).

I now want to determine whether the measured flux errors are reliable or only scatter randomly according to a normal distribution. If the former, I can use them for individual objects in the sample selection, if the latter, I can safely assume that the error for each measurement can be approximated by the median error of all objects. This would be more save, as it avoids the issue of the error measurements having errors themselves, thus not representing the true signal-to-noise of an object. Also, it is reasonable to assume that the background noise is uniform across the field of view.

In figure 6.7 I look at the distribution of flux and σ measurements as histograms (the values are the same as in 6.2 and 6.6). It can be seen that the flux values measured in the HDUV catalogue have a larger standard variation, in agreement with the right panel in figure 6.7. The distributions are not purely normal however, as they have many outliers with extreme values. If I σ -clip the distributions by only including objects with $\bar{f} - 3\sigma < f < \bar{f} + 3\sigma$, where \bar{f} is the median, the flux and σ distributions for the MUSE values become

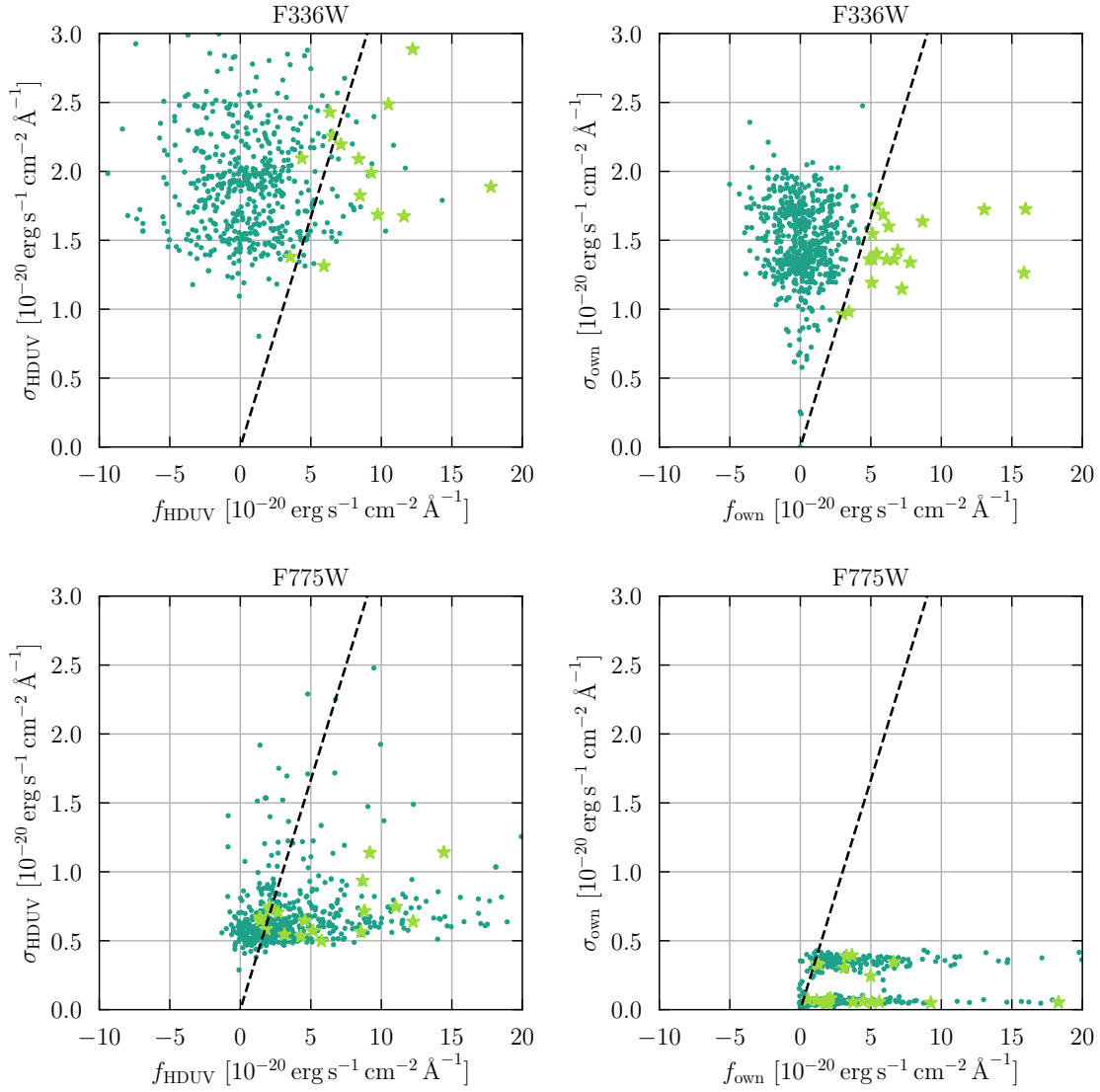


Figure 6.5.: Comparison between the standard deviation σ (y-axis) and the flux measurements (x -axis) for the HDUV survey (left panels) and my aperture measurements (right panels) for the bands WFC3 F336W (top panels) and ACS F775W (bottom panels). The bright green stars again indicate objects with a S/N ratio over three in the WFC3 F336W band. The dashed line indicates the 3σ margin, everything to the right of the line would be considered a detection in the respective band (which is why there are only green stars to the right of the dashed line in the top right panel). The two separate distributions in the panel corresponding to the ACS F775W band and the aperture measurements show the two different depths of data that was used. For the MUSE-Deep objects, deeper HST data was available than for the MUSE-Wide sample.

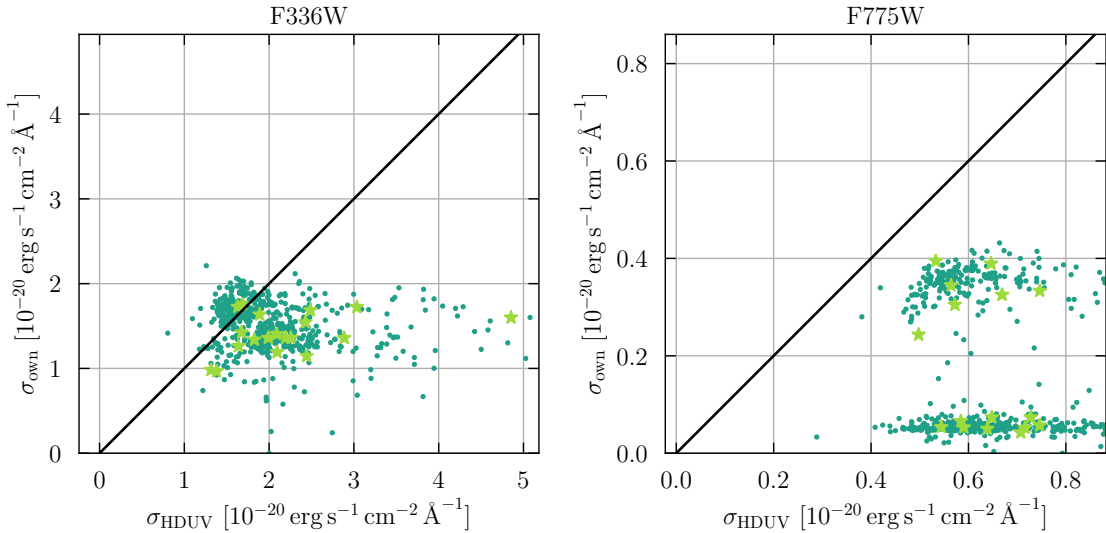


Figure 6.6.: Comparison between the standard deviation σ from MUSE and HDUV. The left panel shows the band WFC3 F336W and the right panel shows the band ACS F775W. The two distinct distributions visible in ACS F775W are due to the two different depths of the data.

normal (albeit at a low probability, with a p-value of $p = 0.109$ and $p = 0.103$ respectively). This deviation from normal distributions (at least if I don't exclude outliers) could be explained by some interlopers that do have measurable flux in the WFC3 F336W band. This is supported by the fact that both distributions are not symmetrical around 0, but lean more to the positive side. The tail of lower values in my errors (right panel of figure 6.7) could be caused by the areas where multiple observations overlap, thus increasing the depth of the data. The trail of higher error values in the HDUV errors could be at the edge of the field of view or close to neighbouring objects, which I mask.

6.2. Correcting Background Offsets in Flux Measurements

To get an accurate measurement of the flux and also the signal-to-noise ratio of all objects in the various bands, I need to examine if the HST bands have a local background that needs to be subtracted (note that the background subtraction discussed in this section was already taken into account for the comparison with the HDUV catalogue in section 6.1.3). The assumption is that for a perfect data reduction, the background will be normally distributed around 0. If that is not the case, the measured flux value has to be corrected for the background contribution. If the local background is normally distributed around a different value than zero, I can use the mean or median of the background measurements to correct the flux values. If the background values deviate too much from a normal distribution, that would indicate local background variations that need to be treated individually instead of correcting for the background with the same value in each object. However, in case of a normal distribution, the error made when correcting each object individually for the local background would be higher than correcting for the median background value in the image.

To investigate whether there are (local) background variations, I measure the background values in the cut-outs of $5 \times 5''$ for each object by using the median flux in all pixels in the cut-out that are not masked and are not within the central aperture (the masking process is described in section 5.5.2). Figure 6.8 shows histograms of the background measurements, together with the p-value assuming a normal distribution, the median m and the standard deviation σ of the distribution. The p-value is a 2-sided Chi squared probability based on the D'Agostino and Pearson's test³. This means, the lower the p-value, the more unlikely it is that the underlying distribution is a normal distribution. This should be seen as a rough guideline, as there is no fixed limit up to which a p-value can be assumed to rule out the null hypothesis with certainty.

For the histograms and statistical measurements, the background measurements are σ -clipped first (values with $f_{\text{bgr}} < 3\sigma$ are excluded), to make sure the estimate of the distribution is not influenced by regions where a close, insufficiently masked object drives the background up. To make it easy to compare to the actual

³I use the python module `scipy.stats` with `scipy.stats.normaltest` for this test.

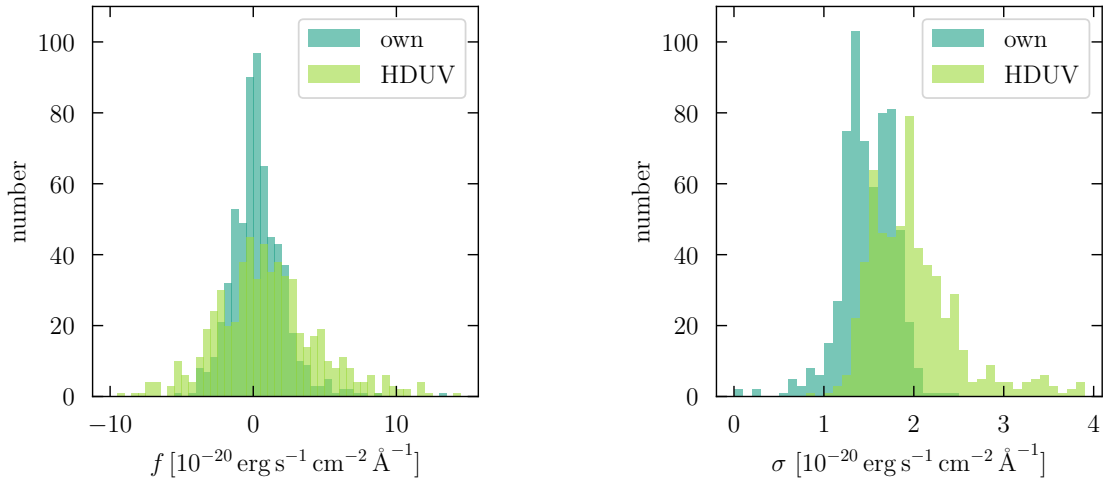


Figure 6.7.: Distributions of measured fluxes f (left panel) and standard deviations σ (right panel) in the WFC3 F336W band for my own aperture photometry (blue) and the flux values in the HDUV survey catalogue (green)

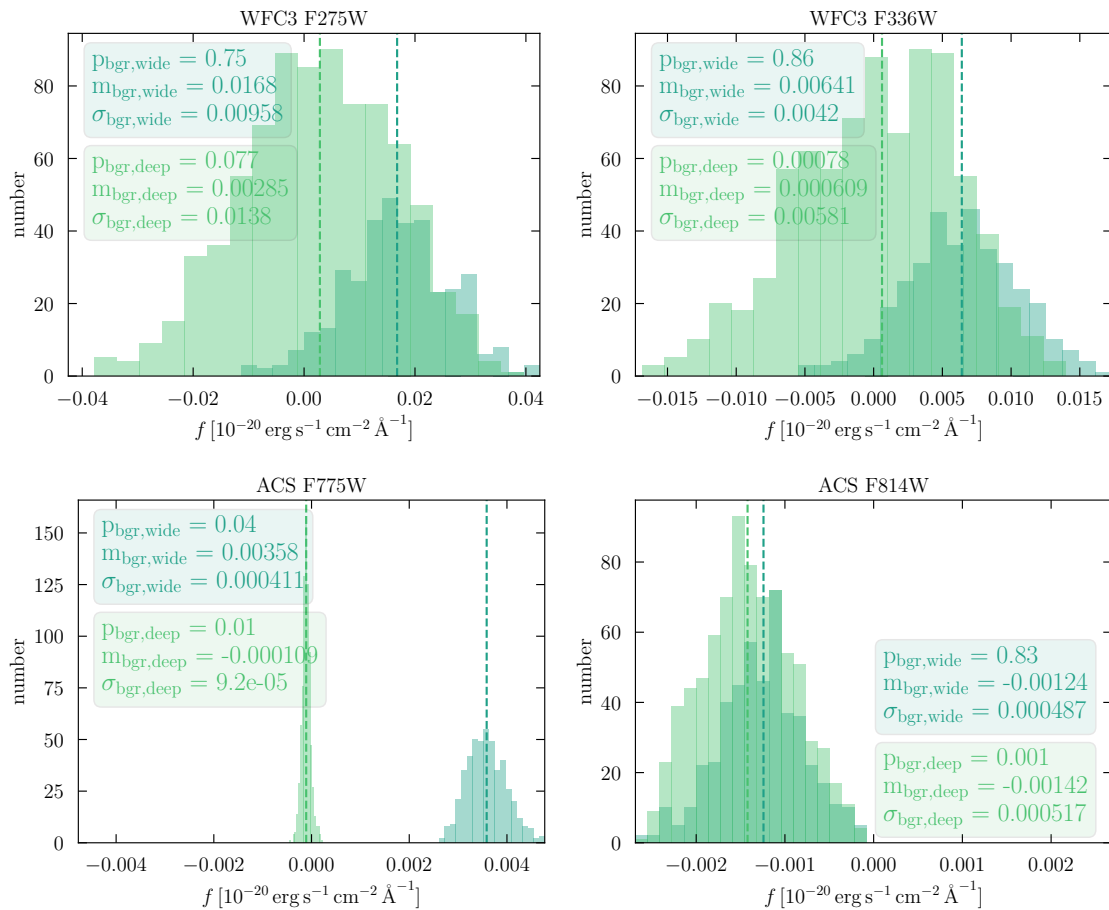


Figure 6.8.: Distribution of background measurements for apertures of $0''.5$ diameter per pixel. The text-boxes give the p-value, the median m (indicated by the dashed lines) and the standard deviation σ of the distribution of background values in units of $10^{-20} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1}$. The figures are centered on 0 to show whether the background is more negative or more positive. The sample is split into MUSE-Wide (blue) and -Deep (green).

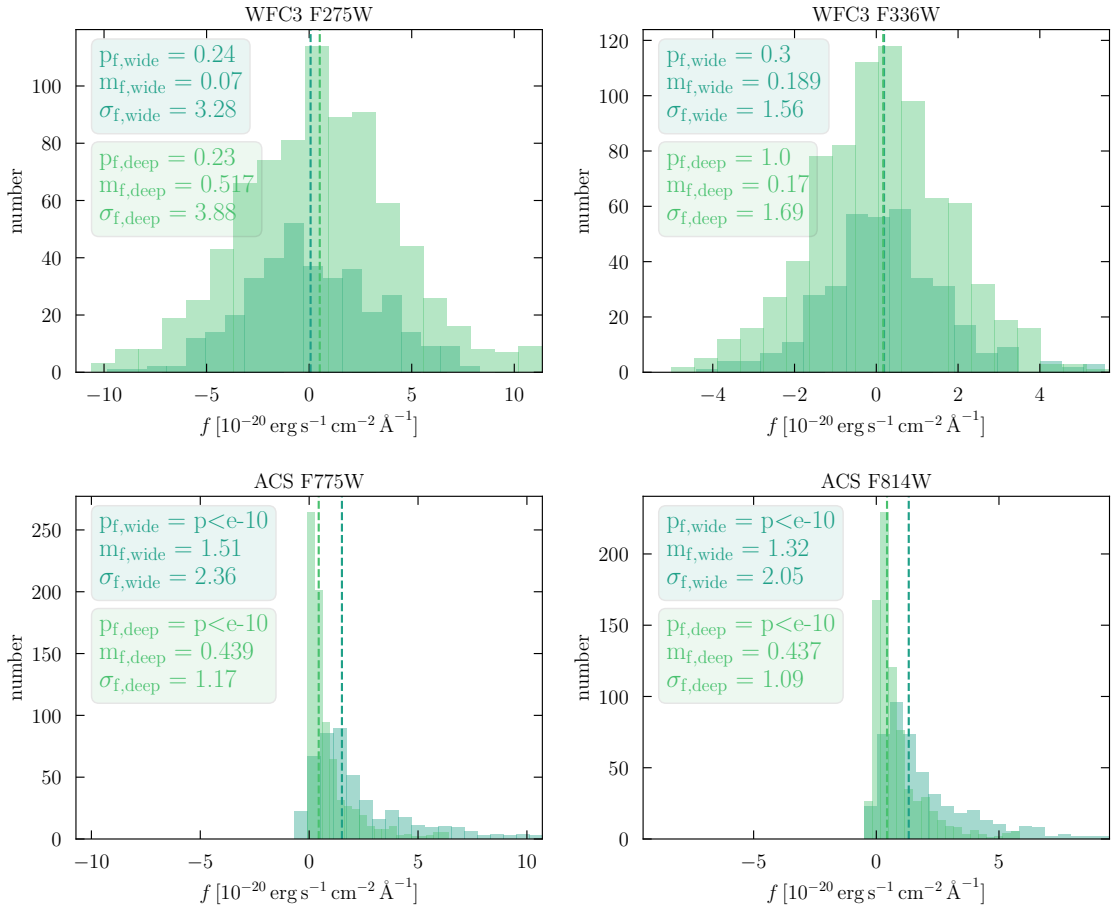


Figure 6.9.: Similar to figure 6.8, but showing the background corrected flux measurements instead of the background, separated again into MUSE-Wide (blue) and -Deep (green).

measured flux values (see figure 6.9), the background is measured per pixel and then multiplied by the number of pixels in an aperture with a diameter of $0''.5$ (which is also used to measure the flux values).

The first plot (6.8) shows the distributions for the HST bands WFC3 F275W, WFC3 F336W, ACS F775W and ACS F814W, which are the relevant bands for this study. The sample is split into MUSE-Wide (blue) and MUSE-Deep (green), since except for the ACS F814W band, both samples have different background values. However, the ACS F814W band is also the only one where both the MUSE-Wide and -Deep samples have background value distributions that are not compatible with a scatter around 0. Therefore for this band I use the background level that is the mean of both distributions, $m_{bgr} = -0.00133 \cdot 10^{-20} \text{ erg s}^{-1} \text{ cm}^{-2}$. In the bands WFC3 F275W, WFC3 F336W and ACS F775W, the MUSE-Deep objects are in parts of the images (the UDF) where the background level is compatible with 0, which is why no background subtraction is performed here. The WFC3 F336W and F275W data taken from the HDUV survey contain the version 2 image of the UVUDF (Rafelski et al., 2015 and Teplitz et al., 2013) in the UDF, which is the same area as the MUSE-Deep data.

When splitting the sample into MUSE-Wide and -Deep, I thus find that all distributions of background noise are roughly consistent with normal distributions, which means their scatter can be interpreted as random. Therefore I subtract (or add) the median background level to each individual object's cut-out before measuring the flux and before stacking (after stacking, I do not subtract any possible background anymore). I only correct the background for the MUSE-Wide objects (except for band ACS F814W), as the MUSE-Deep areas do not show a significant background level in WFC3 F275W, WFC3 F336W and ACS F775W.

To show that the background would offset the flux measurements significantly, figure 6.9 shows the background corrected flux measurements of the actual objects in each band. For bands WFC3 F275W and WFC3 F336W, the scatter in the flux measurements are of the same order as the scatter in the background, which means that most objects do not have any detection in those bands (which is to be expected, as they

HST band	$f_{\text{bgr}} [10^{-20} \text{ erg s}^{-1} \text{ cm}^{-2}]$			
	MUSE-Wide		MUSE-Deep	
	/pixel	/aperture	/pixel	/aperture
WFC3 F275W	0.0168	0.923	/	/
WFC3 F336W	0.00641	0.353	/	/
ACS F775W	0.00358	0.782	/	/
ACS F814W	-0.00133	-0.29	-0.00133	-0.29

Table 6.1.: Background correction values f_{bgr} for the four relevant fields WFC3 F275W, WFC3 F336W, ACS F775W and ACS F814W, taken from the median of the measured background value distributions (figure 6.8). The first number gives the background per pixel, the second per aperture with a diameter of $0''.5$ (the same aperture as for the flux measurements). The median value of the background for ACS F814W is the same for both MUSE-Wide and -Deep, as there is no significant difference between the two sub-samples in this band. For the other bands, no background correction is needed for objects in the MUSE-Deep area.

probe the Lyman continuum). Ignoring the background level for the MUSE-Wide objects would however artificially increase the measured flux in a stack of WFC3 F336W when looking for Lyman continuum emission, possibly leading to a spurious detection or overestimation of the flux. For ACS F775W and ACS F814W the correct handling of the background level is equally important, as the flux measurements in ACS F775W (ACS F814W) would otherwise be overestimated (underestimated). Table 6.1 shows an overview of the different values used to correct the background for each band.

6.3. Stellar Population Models

When measuring flux in HST bands in order to determine the Lyman continuum or UV continuum flux values and compare both, the flux is measured not at a specific wavelength, but over the whole width of the filter band. The rest-frame wavelength range covered by this filter band depends on the redshift of the object. Often the measured flux is interpreted as the flux at the effective wavelength of the filter band, but this ignores the shape of the spectrum of the object. Additionally, to be comparable, the LyC and UV continuum are usually defined at the positions of 900 \AA and at 1500 \AA rest-frame respectively. Therefore, I use stellar population models to determine the intrinsic LyC to UV continuum ratio $(L_{\text{UV}}/L_{\text{LyC}})_{\text{int}}$ and to correct the flux measurements from photometry to 900 \AA and at 1500 \AA .

For the UV continuum part of the spectrum, the slope can in principle be measured (see section 4.1.2) but the measurements scatter significantly. This is most likely due to uncertainties rather than intrinsic differences in the β -slope. Therefore I can either use a median value of a large sample of objects, as described in section 4.1.2 or the slope of a theoretical spectrum. For the Lyman continuum part of the spectrum however, I do not have any measurements of the slope at all, therefore using a model spectrum is the only option.

One of the most used evolutionary codes for stellar population synthesis models is Starburst99 (Leitherer et al., 1999, with improvements by Vázquez & Leitherer, 2005, Leitherer et al., 2010 and Leitherer et al., 2014). Using a wide range of input parameters and taking both stellar and nebular emission into account, it produces spectra for star-forming galaxies, as well as other spectrophotometric properties. I use the stellar population models⁴ for two purposes: 1) To get an estimate of the expected flux ratio between the UV continuum and the Lyman continuum needed for measuring the LyC escape fraction and 2) to correct the flux measurements from HST to the wavelength positions of the UV and Lyman continua at 900 \AA and 1500 \AA . The input parameters relevant here are the star formation rate, the stellar evolutionary tracks, the metallicity and the age of the stellar population. All other parameters were left at the default values and the computation of the number of ionizing photons was activated to get more accurate results below the Lyman limit. An overview of the different models used here can be found in table 6.2.

Model name	Parameter settings
Gv00	Geneva evolutionary tracks without rotation.
Gv40	Geneva evolutionary tracks with rotation.
Gvstd	Geneva evolutionary tracks from 1994 with standard mass loss.
high sf	Star formation rate of $10 M_{\odot}$ per year.
002, 014	Metallicity settings for the 2012 Geneva evolutionary tracks.

Table 6.2.: Overview of the different models that are used in this thesis. The standard Geneva evolutionary tracks marked with 'Gvstd' are described in [Meynet et al. \(1994\)](#) and the Geneva evolutionary tracks with and without rotation are described in [Ekström et al. \(2012\)](#) and marked with 'Gv00' and 'Gv40' respectively.

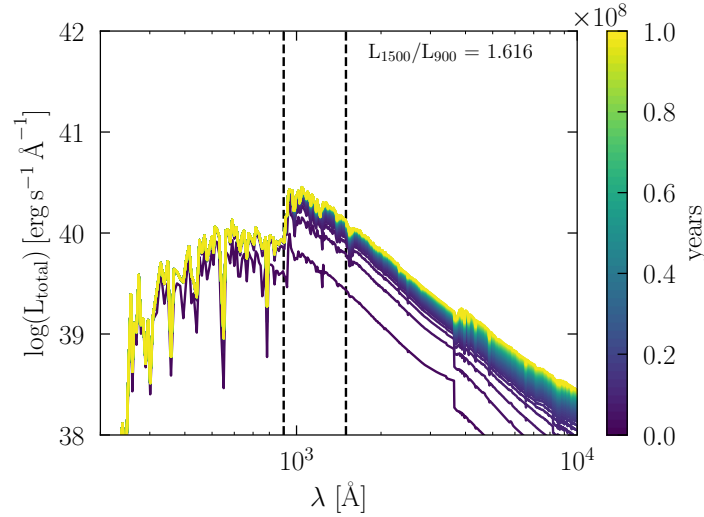


Figure 6.10.: Stellar population model spectra from Starburst99, using the default parameters with a continuous star formation rate of $1 M_{\odot}$ per year and a Kroupa initial mass function (IMF) with Geneva evolutionary tracks ([Ekström et al., 2012](#)) without rotation and a solar metallicity of 0.014. The logarithmic luminosity is plotted over the logarithmic wavelength in Ångstrom. The colours indicate the ages of the stellar populations, with dark purple being the youngest at 2 Myrs and bright yellow being the oldest at 100 Myrs (see the colourbar). The dashed lines indicate 900 Å and 1500 Å and the ratio L_{1500}/L_{900} given in the top right corner corresponds to the yellow (oldest) curve.

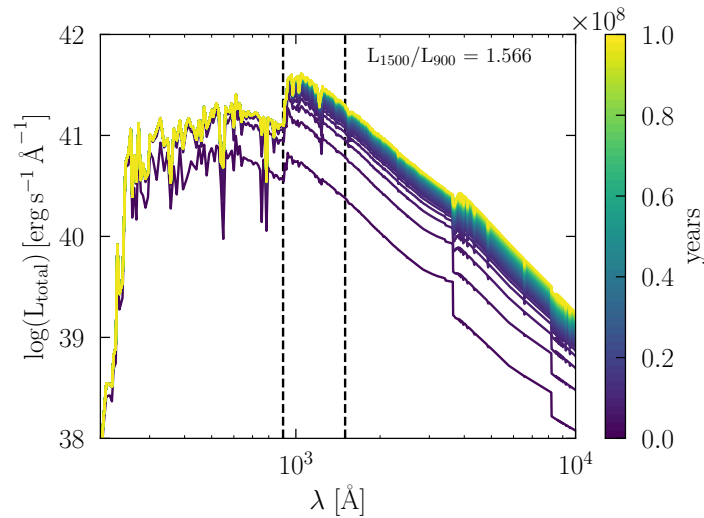


Figure 6.11.: Same as figure 6.10, but with different input parameters for Starburst99 that resemble more a typical LAE at high redshift. The constant star formation rate is set to $10 M_{\odot}$ per year, the evolutionary tracks are Geneva evolutionary tracks with rotation ([Ekström et al., 2012](#)) and a lower metallicity of 0.002.

Model name	evolutionary tracks	metallicity Z	SFR	L_{1500}/L_{900}
Gv00 002	Geneva 2012, no rotation	$Z=0.002$	$1 M_{\odot}/\text{yr}$	1.612
Gv00 014	Geneva 2012, no rotation	$Z=0.014$	$1 M_{\odot}/\text{yr}$	1.616
Gv40 002	Geneva 2012, with rotation	$Z=0.002$	$1 M_{\odot}/\text{yr}$	1.566
Gv40 014	Geneva 2012, with rotation	$Z=0.014$	$1 M_{\odot}/\text{yr}$	1.172
Gv40 002 high sf	Geneva 2012, with rotation	$Z=0.002$	$10 M_{\odot}/\text{yr}$	1.566
Gv40 014 high sf	Geneva 2012, with rotation	$Z=0.014$	$10 M_{\odot}/\text{yr}$	1.171
Gvstd 001	Geneva 1994, standard mass loss	$Z=0.001$	$1 M_{\odot}/\text{yr}$	1.745
Gvstd 004	Geneva 1994, standard mass loss	$Z=0.004$	$1 M_{\odot}/\text{yr}$	1.708
Gvstd 008	Geneva 1994, standard mass loss	$Z=0.008$	$1 M_{\odot}/\text{yr}$	1.730
Gvstd 02	Geneva 1994, standard mass loss	$Z=0.020$	$1 M_{\odot}/\text{yr}$	1.698
Gvstd 04	Geneva 1994, standard mass loss	$Z=0.040$	$1 M_{\odot}/\text{yr}$	1.734

Table 6.3.: Overview of the different L_{1500}/L_{900} ratios at 10^8 Myrs depending on the choice of evolutionary tracks, metallicity Z and star formation rate (SFR). Any parameters not mentioned are left at the default values (except for the star formation, which is always set to constant). The smallest ratio of 1.171 is produced by the model using the Geneva 2012 evolutionary tracks with a metallicity of $Z = 0.014$ and the larger SFR of $10 M_{\odot}/\text{yr}$, while the highest ratio of 1.745 is produced by the model using the Geneva 1994 evolutionary tracks with a metallicity of $Z = 0.001$. The mean ratio of the models analysed here is $L_{1500}/L_{900} = 1.574$.

6.3.1. The Intrinsic UV to Lyman Continuum Flux Ratio

For a realistic estimate of the LyC escape fraction, it is important to know the intrinsic flux ratio between the LyC and UV continuum $(L_{UV}/L_{LyC})_{\text{int}}$. As described above, this is not directly observable due to IGM attenuation at higher redshifts, which is why I use stellar population synthesis models from Starburst99 with parameters that match high-redshift LAEs as closely as possible (since these are the objects I am analysing in chapter 5).

Figure 6.10 shows a Starburst99 model spectrum with default input parameters, except for the star formation rate, which is set to constant. Since the LyC part of the stellar population spectrum is dominated by massive, hot stars that have a short lifetime, in a star burst the LyC will be low, as those stars die quickly. With a constant star formation rate, the high mass population is replenished fast enough to produce a noticeable LyC which does not vary with time as much as the UV continuum (see figure 6.10). With increasing age, the UV continuum increases, causing the ratio $(L_{UV}/L_{LyC})_{\text{int}}$ to increase as well. At an age of 10^8 years the ratio is $L_{1500}/L_{900} = 1.616$.

As discussed in the introduction in section 1.3.2, various studies (e.g. McLinden et al., 2014, Acquaviva et al., 2011, Karman et al., 2017) have examined the stellar population properties of LAEs at high redshifts. Typically LAEs have a low dust content, a subsolar metallicity (e.g. Acquaviva et al., 2011, McLinden et al., 2014, Nilsson et al., 2007), a high star-formation rate and ages that range from 1 Myr for faint, lensed LAEs (Karman et al., 2017) over 10 Myrs (Gawiser et al., 2007, Ono et al., 2010) to 1.81 Gyrs (Acquaviva et al., 2011, Finkelstein et al., 2015). Star formation rates can also vary between $\sim 1 - 100 M_{\odot}/\text{yr}$ for LAEs at $z = 3 - 4$ (e.g. Ono et al., 2010, Gawiser et al., 2007). Figure 6.11 shows a more realistic spectrum based on average values found in the literature for an LAE with a higher (compared to the default values) star formation rate of $10 M_{\odot}$ per year, a lower metallicity of 0.002 and evolutionary tracks with rotation. While the higher star-formation rate shifts the spectrum to higher luminosities, the changed metallicity and evolutionary tracks change the shape of the spectrum, but influence the ratio L_{1500}/L_{900} only slightly, which has a value of 1.566 for this model spectrum.

As can already be seen in figures 6.10 and 6.11, the UV continuum increases over time, causing the ratio L_{1500}/L_{900} to increase as well. In figure 6.12, four different model spectra are compared with respect to the evolution of the UV over Lyman continuum ratio L_{1500}/L_{900} with time. As can be seen, the star formation rate does not have a big influence on the ratio L_{1500}/L_{900} as long as the star formation is constant, while the metallicity has a significant influence (at least for the Geneva evolutionary tracks with rotation). The two metallicities shown are the only two available at the moment for the Geneva evolutionary tracks 2012 with rotation. Figures 6.13 and 6.14 show two different sets of models and their influence of metallicity

⁴I use the online interface for running the stellar population models of Starburst99, found at <http://www.stsci.edu/science/starburst99/docs/parameters.html>.

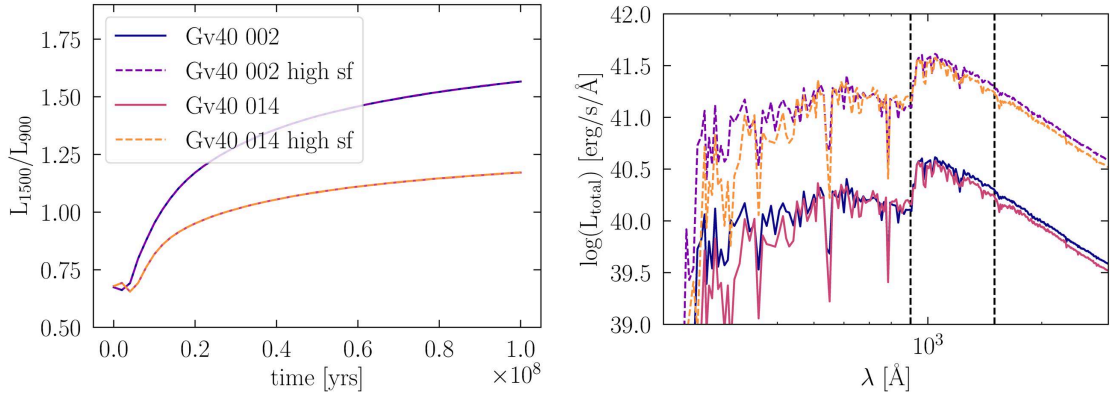


Figure 6.12.: Left panel: Ratio L_{1500}/L_{900} over time for four different sets of input parameters, using Geneva evolutionary tracks with rotation (Ekström et al., 2012). Two of the models (red and orange) have solar metallicities of 0.014, the other two (purple and pink) have a lower metallicity of 0.002. Two models (dashed lines) have a higher star formation rate of $10 M_{\odot}$ per year, the other two have a lower star formation rate of $10 M_{\odot}$ per year. Right panel: Spectra of the models in the left panel. The higher star formation rate only shifts the spectra to higher luminosities, but does not change their shapes.

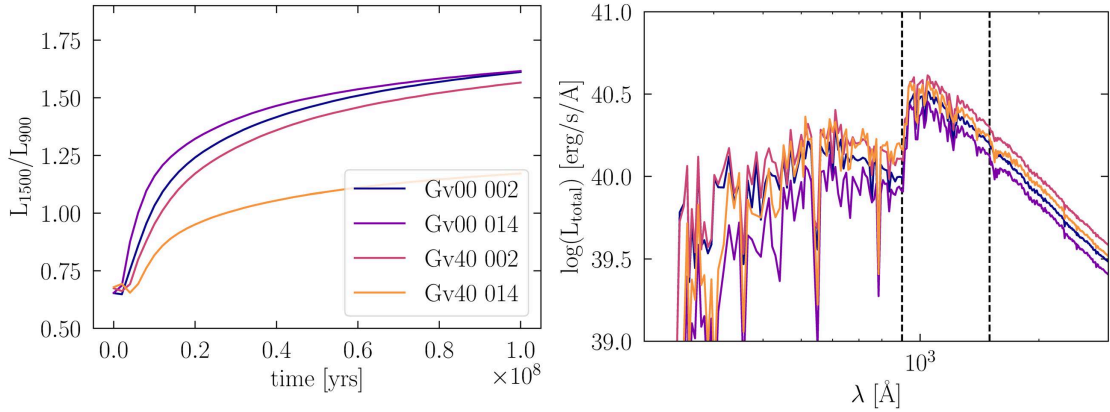


Figure 6.13.: Same as figure 6.12, but with different models showing two different evolutionary tracks (Geneva 2012), with (Gv40) and without (Gv00) rotation and two different metallicities (solar and $Z = 0.002$). The evolutionary tracks with rotation are influenced by the change in metallicity much more than the ones without.

and evolutionary tracks on the ratio L_{1500}/L_{900} . While a higher metallicity decreases the ratio between UV continuum and Lyman continuum for the Geneva 2012 evolutionary tracks with rotation, for the ones without rotation and the Geneva 1994 tracks the influence of the metallicity is smaller and has the opposite effect: The ratio increases slightly with increasing metallicity.

Thus the choice of the stellar population model can change the assumed intrinsic ratio L_{1500}/L_{900} by a factor of ~ 1.5 , assuming an age of 10^8 Myrs. An overview of different evolutionary models and the resulting L_{1500}/L_{900} ratios is given in table 6.3. Since the 2012 Geneva evolutionary tracks with rotation are the latest available ones and have been shown to be successful at producing realistic stellar properties (see Ekström et al., 2012) which are similar to the LAEs in our sample, I use them in the further analysis (and the following plots in this section), together with a metallicity of 0.002 and a star formation rate of $1 M_{\odot}$ per year, which gives a ratio of $L_{1500}/L_{900} = 1.566$ for 10^8 Myrs. This is also close to the mean ratio of $L_{1500}/L_{900} = 1.574$ of all models analysed here.

6.3.2. Correcting Flux Measurements Using Stellar Population Models

Knowing the intrinsic ratio L_{1500}/L_{900} now, I also need to know the flux at the two wavelength 900 and 1500 \AA . The flux at the Lyman and UV continuum was measured using HST filter bands: The WFC3

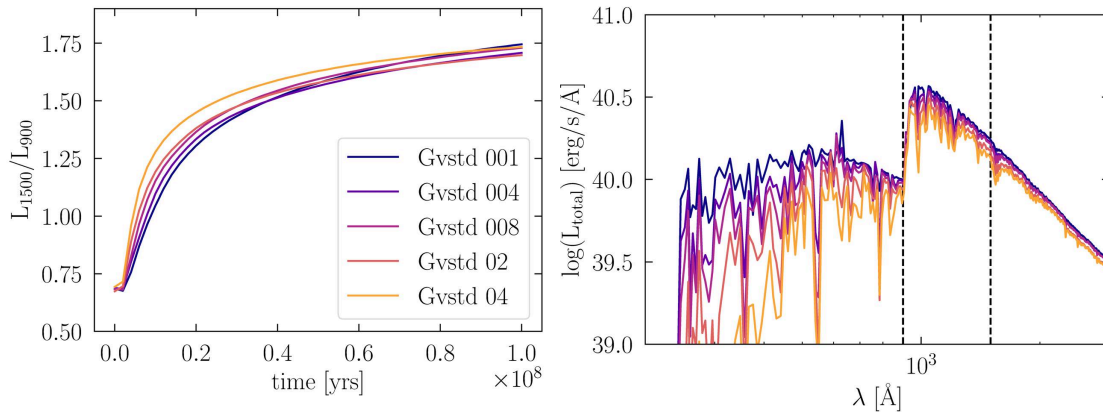


Figure 6.14.: Same as figure 6.12, but with Geneva 1994 evolutionary tracks with standard mass loss and five different metallicities. For these tracks, the metallicity does not significantly influence the ratio.

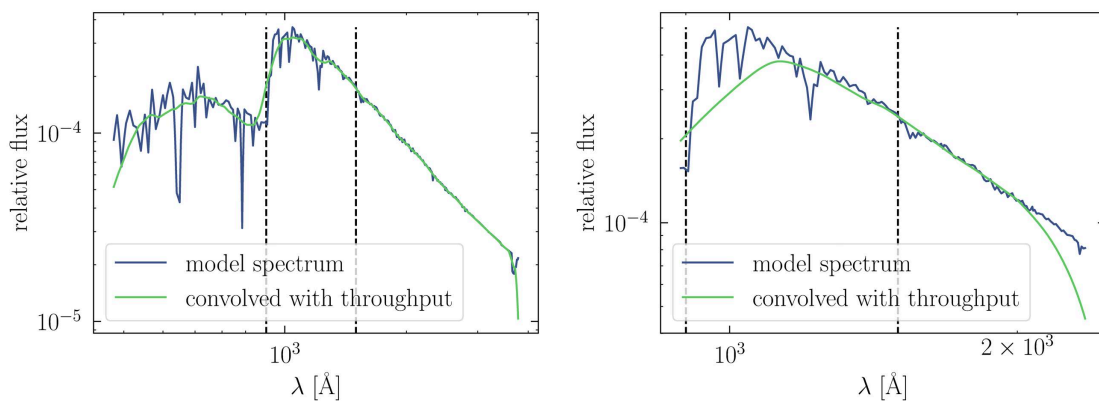


Figure 6.15.: The model spectrum (Gv40 002 at $10^8 Myrs$) from Starburst99 is shown in blue, the model spectrum with the filter throughput of WFC3 F336W (left panel) and ACS F814W (right panel) is shown in green. This means at each wavelength, the green spectrum shows the sum of the model spectrum multiplied by the filter throughput. The effect is similar to smoothing the spectrum, which softens the Lyman continuum break. Both spectra are normalised to a total flux of one. The dashed lines indicate 900 and 1500 Å.

F336W filter was used for the Lyman continuum and the ACS F814W (for MUSE-Wide) and ACS F775W (for MUSE-Deep) filter for the UV continuum. To get estimates of the flux at the wavelengths 900 and 1500 Å, I use the model spectra from Starburst99 combined with the filter curves of the bands (see figure 6.15 for an example for the LyC band WFC3 F336W). Since I want to know the flux value at a specific wavelength, I cannot use the model spectrum folded by the filter band throughput, since that would give me only a flux correction for measuring the flux in the same band, but at a different position. This would underestimate the Lyman continuum break at 912 Å, as can be seen in figure 6.15 by the smoothing of the break in the green line and the difference between the two lines at that position.

Instead, I calculate the correction factor by dividing the value of the model spectrum at the desired wavelength by the model spectrum folded by the filter band throughput, both normalised to one. That way I get the factor by which I have to multiply a measured flux value in order to get the flux value at 900 or 1500 Å. This is illustrated in figure 6.16 and shows that for most redshifts in the relevant range, the correction for the LyC is rather small. This can already be seen in the model spectra e.g. in figure 6.10, as the spectrum in the LyC range shows a curve that gets smoothed over by taking the throughput into account. The correction for the UV continuum is larger at smaller redshifts and decreases almost linearly. This correction factor is used to correct flux measurements in the LyC (also in chapter 5).

Note that I also show the β -slope of the UV continuum at different wavelength positions in the right panel of figure 6.16 and get a slightly lower value than typically assumed in the literature (see discussion in chapter 4 in section 4.1.2).

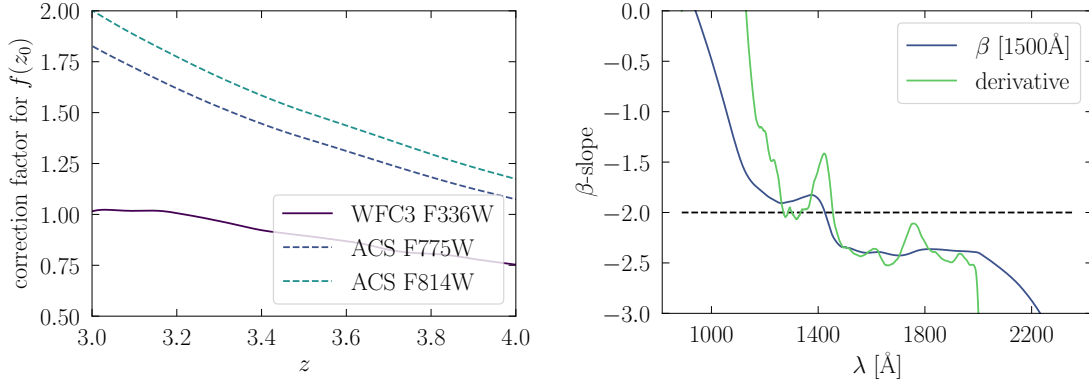


Figure 6.16.: Left panel: Correction factors for measuring the LyC in the WFC3 F336W filter band (solid line) and for measuring the UV continuum in the ACS F775W and ACS F814W band (dashed lines). For simplicity, the x -axis shows the redshift of the object instead of the rest-frame wavelength. Right panel: The β -parameter, meaning the UV continuum slope of the spectrum, is computed from the model shown in the right panel in figure 6.15. For the blue curve, at each wavelength, the reference wavelength was 1500 Å. For the green curve, the derivative at each wavelength is shown. The value for β in the reliable range (1200 Å to 1900 Å) in this model is lower than usually assumed ($\beta = -2$, indicated by the black dashed line) since the models do not account for dust reddening.

6.4. Noise Models

With knowledge of the background noise in WFC3 F336W and the UV continuum flux of each object (as well as the IGM transmission) I can estimate what S/N to expect in our stacks given a certain escape fraction (which is expressed as flux in the band WFC3 F336W that is a certain fraction of the flux in the UV continuum). I already explained this process in the previous chapter in section 5.5.1 but I will now go into more detail.

I measure the background noise in WFC3 F336W in apertures of 0".5, which is

$$\sigma_{\text{bgr},\text{F336W}} = 1.532 \cdot 10^{-20} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1} \quad (6.2)$$

This way I can create apertures with random noise which follows a normal distribution with a standard deviation of $\sigma_{\text{bgr},\text{F336W}}$ and add flux that is a fraction of the UV continuum to simulate an escape fraction. For the UV continuum flux f_{UV} I use the flux measured in either the band ACS F775W (for MUSE-Deep objects) or ACS F814W (for MUSE-Wide objects) and correct it to the wavelength $\lambda = 1500 \text{ \AA}$ using a β -slope of $\beta = -1.92$, as is described in section 4.1.2. This flux can be used to derive the expected Lyman continuum flux f_{LyC} as

$$f_{\text{LyC}} = \frac{f_{\text{UV},1500 \text{ \AA}}}{1.566} f_{\text{esc}} T_{\text{IGM}} \quad (6.3)$$

where 1.566 is the theoretical ratio $L_{\text{UV}}/L_{\text{LyC}}$ of UV to Lyman continuum, derived using stellar population models (see section 6.3), f_{esc} is the escape fraction and T_{IGM} is the mean IGM transmission at the redshift of the object (derived from 10 000 theoretical lines-of-sight, see section 5.4). Now I can add this LyC flux to the random noise generated from $\sigma_{\text{bgr},\text{F336W}}$ and get a S/N ratio with $f_{\text{LyC}}/\sigma_{\text{bgr},\text{F336W}}$.

I use the same order of objects I get from section 5.5.1 and stack an increasing number of objects. The result can be seen in figure 6.17, where I show four examples of random apertures as well as the real measured S/N over the number of objects. The four different randomly drawn examples show that due to the noise level, the S/N curve can vary wildly with different individual realisations of such a random experiment. Therefore I perform 1000 such stacking procedures with random noise and different escape fractions to obtain a distribution.

Figure 6.18 shows the S/N ratio distributions for different numbers of objects, sorted as described above in section 5.5.1 and shown in figure 5.8. For any number of objects per stack (number on the x -axis), there is a distribution of possible S/N values, with a width determined by the background noise. Even with $f_{\text{esc}} = 0\%$, there is still a chance of around 5% to see a S/N ratio of almost two in a stack with any number of objects.

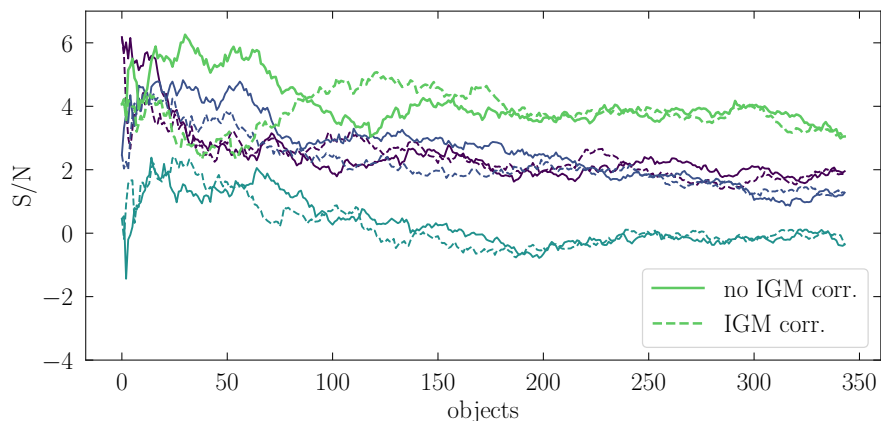


Figure 6.17.: Example of four different realisations of stacking random apertures, all with an escape fraction of $f_{\text{esc}} = 50\%$. The solid lines show the objects sorted without IGM correction (see left panel of figure 5.8), the dashed lines show the objects sorted with the IGM correction (right panel in figure 5.8). The x -axis shows the number of objects and the y -axis the S/N ratio.

For $f_{\text{esc}} > 0\%$, the S/N reaches a maximum at a low number of objects (usually at around 30) and decreases even if more objects are used for the stack. For the stacking order without IGM correction, the median of the maximum S/N is at 29 objects with a range of 20 to 58 objects at 90% of the maximum S/N or higher. For a stacking order with IGM correction, the maximum is at 16 objects with a range of 7 to 61 objects.

Now I know the ideal selection and number of objects to stack (around 60) to increase the S/N ratio, I can compare the results of the stacks of the real objects to the simulations with random noise. Since the S/N ratio is itself noisy, I use the mean of stacks of an increasing number of objects in the range of 20 to 70 objects. This is a little larger than the ideal range found with figure 6.18, but since the maximum of the significance in figure 5.6 was at 61, I increase the range to 70. The result can be seen in figure 5.9 on page 101, which shows how often a certain mean S/N ratio in this range occurs at different escape fractions. Another way of getting the LyC escape fraction from the measured S/N ratio in the real stack is shown in figure 6.19, which is basically the inverse of figure 5.9. Different curves with occurrences of S/N values are shown, each corresponding to a different escape fraction. Even for a LyC escape fraction of 100% I could find S/N ratios below 3 and for an escape fraction of 0% there are a few cases with S/N values above 3. This shows that the signal I find can merely give an indication with a certain probability of the underlying escape fraction. The most likely escape fraction in our data is $f_{\text{esc}} \sim 80\%$, but since the detected signal has an offset (see next section below), the real underlying escape fraction might be much lower.

6.5. Spatial Offset

When stacking the selected LAEs in the LyC band WFC3 F336W, the detected signal has an offset from the expected centre and looks elongated, both along the y -axis (the north-south axis, see section 5.5.2 in the previous chapter). In this section I describe several tests I make to narrow down the cause of this offset, which can be assumed not to be a random effect due to its orientation and size. In this section, the individual LyC leaker candidates are excluded to make sure they do not dominate the offset. The following possible causes for the offset were explored and excluded:

- There could be a systematic problem with masking bright pixels or close neighbours. However, when not masking anything, the offset is still seen in the stack, along with some bright objects that are normally masked (see figure 6.20).
- A problem with the positions of the LAEs (determined from the ACS F814W band using *Galfit*) would result in offsets in other bands than WFC3 F336W as well, which is not seen (see below).
- The WFC3 F336W cut-out of each object included in the sample was visually inspected by two inspectors. While some objects do show a noise peak or a faint signal at the position of the offset, it is not obvious which objects exactly could have caused the problem.
- Still, the offset signal could be caused by a single interloper or unmasked contaminant, which would mean it would not originate from Lyman continuum emission. To test this, the stacks for determining

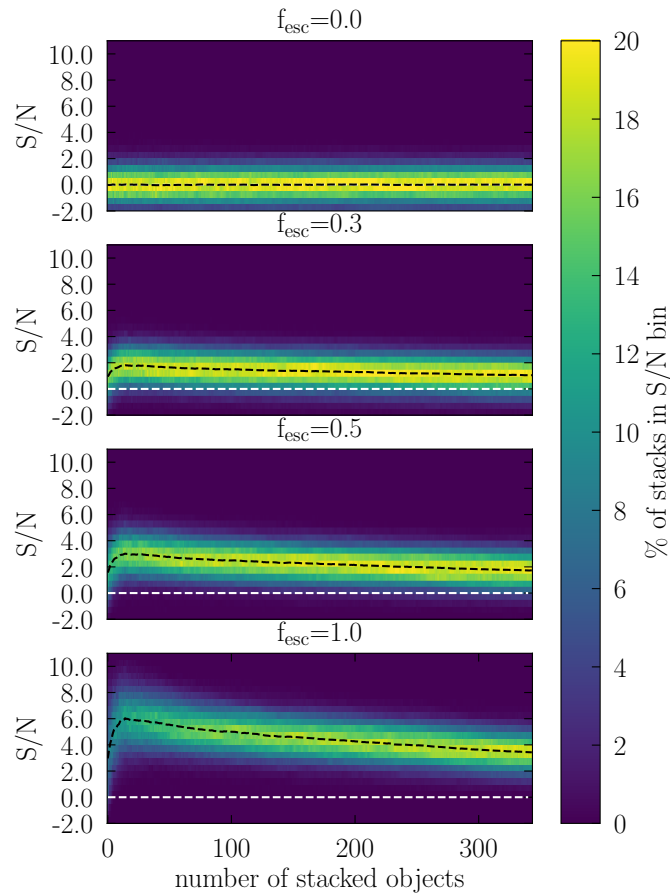


Figure 6.18.: The plots show the S/N over the number of stacked objects for the stacking order without the IGM correction. the x -axis gives the number of objects per stack. A total of 1000 stacks were produced per x -axis value and the distribution of S/N values in bins of 0.5 is shown as the colour coding. The four different panels show different escape fractions, with the top panel corresponding to an escape fraction of $f_{\text{esc}} = 0\%$, the second to $f_{\text{esc}} = 30\%$, the third to $f_{\text{esc}} = 50\%$ and the bottom panel to $f_{\text{esc}} = 100\%$. The white dashed line marks a S/N value of 0, the black dashed line is the median S/N value of all 1000 realisations.

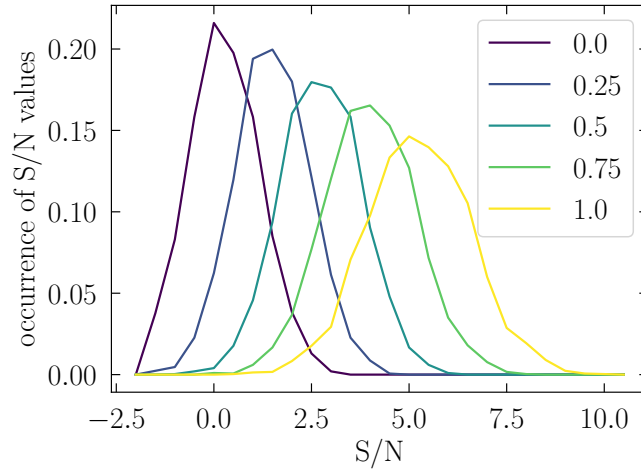


Figure 6.19.: Same as figure 5.9, but instead of the escape fraction, the S/N ratio is shown on the x -axis and the escape fraction is colour coded (see legend). From this plot I can estimate the likelihood that a measured S/N value has been caused by a certain escape fraction in the stacked objects.

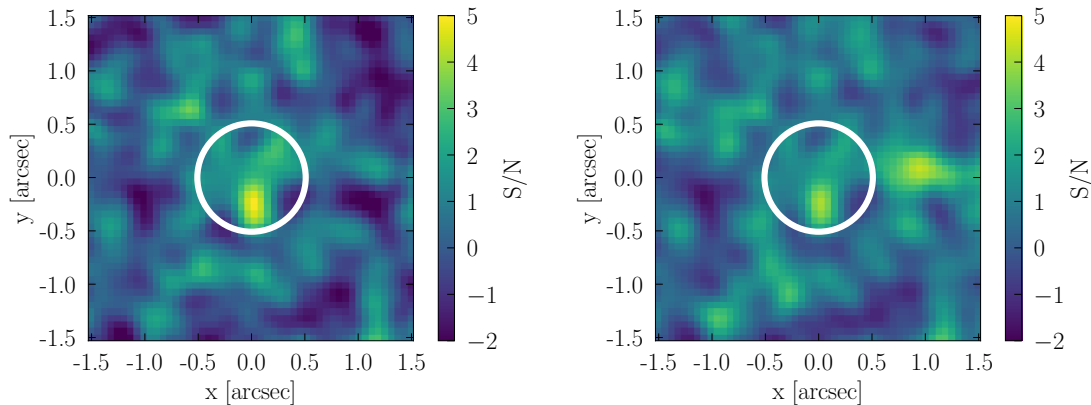
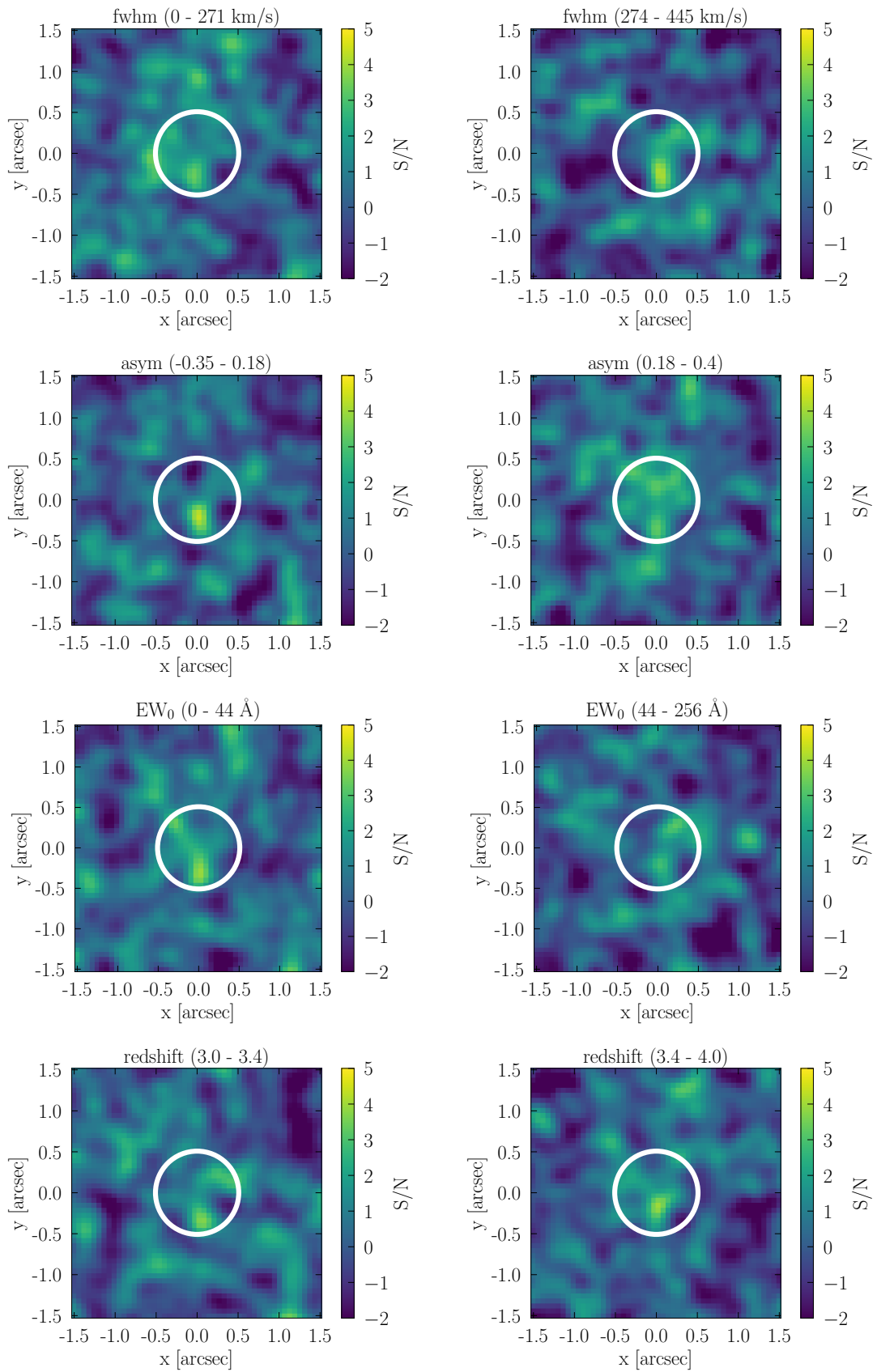


Figure 6.20.: Left panel: Stack of the 60 most promising LAEs in the WFC3 F336W band. Right panel: The same stack without masking any bright neighbours or pixels. The offset is visible in both images, although the signal is stronger in the stack with the masking.

the influence of Lyman α line properties are examined (discussed in section 5.6). For these plots, the sample was split into two equal parts, one with high and one with low equivalent widths (the same for high and low FWHM and asymmetry), but the offset was detected in all stacks at least marginally (see figure 6.21). This excludes the possibility of a single interloper (or imperfectly masked bright neighbour) causing the offset signal, as in this case only one of the stacks containing half of the objects would display the signal. There could still be multiple objects (instead of just one) with random noise peaks at the position of the offset, as found by the visual inspection. To test whether this offset signal is consistent, I perform bootstrapping experiments (see below).

6.5.1. Shape and Offset of Stacks in Different Bands

Another possibility is a problem with the translation between pixel coordinates to the world coordinate system (WCS), either by a shift or even a distortion. This could be present in all bands to varying degrees or only in the WFC3 data or even only in the WFC3 F336W band and thus not be detected in any other band. I test this by fitting the stacked LAEs with Gaussians in the different HST bands, which can be seen in figure 6.22 for the two dimensional data and in figure 6.23 for the data collapsed along the x - and y -axes. The top left panel shows the WFC3 F336W band in which the offset was noticed. All bands are oriented with north facing up along the y -axis (declination) and east facing left along the x -axis (right ascension).



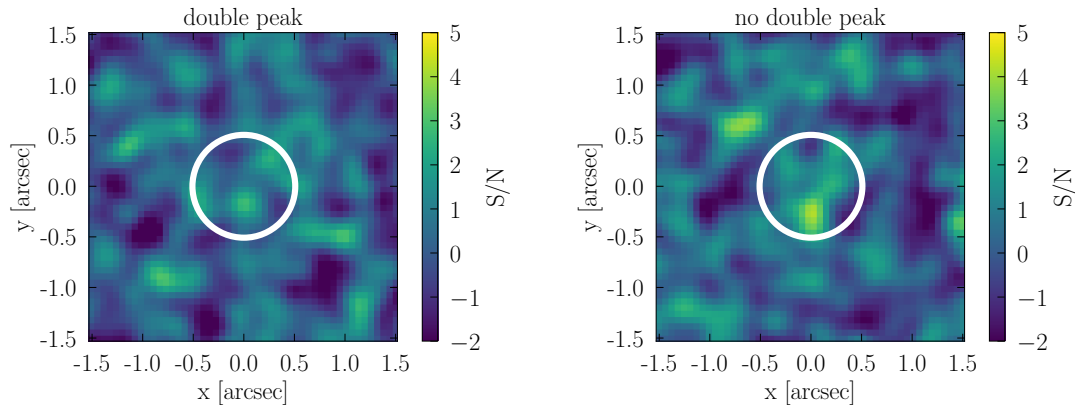


Figure 6.21.: Figure continued from previous page: Stacks of objects based on their (Lyman α line) properties. The 60 most promising objects were separated into two samples: the left panels contain the low values of the respective properties, the right panels the higher values, except for the peak separation, where the sample was separated into objects with and without a double peak. As can be seen, the offset is present in all stacks (with a possible exception of the highly asymmetric line profiles), indicating that the signal is not caused by a single or a few objects alone.

When fitting a simple two dimensional Gaussian to the stack in the WFC3 F336W band, the offset has a size in the x -direction of $0''.029$ and in y -direction of $-0''.154$. However, the shift in x -direction is likely not significant and only influenced by the brighter area to the top right of the centre, which looks like a random noise peak. To minimise the effect of noise peaks that are farther away from the center and can thus artificially elongate the fitted 2D Gaussian, I weigh the fit to the WFC3 F336W stack with a 2D Gaussian with standard deviations of $0''.25$ centered on the middle. Still, the offset in y -direction is bigger than measured from the fitting and is also influenced by the bright spot, just as the tilt of the Gaussian, which is clearly elongated in the y -direction. None of the other probed bands show a similarly strong spatial offset or a similar shape of the fitted Gaussian. All bands show an insignificant offset to the left and bottom of the size of less than one HST pixel in the respective bands ($0''.03$ in the ACS bands and $0''.06$ in the WFC3 bands, which is nevertheless twice as large than the offset found in section 6.1.1 between the Galfit positions and those from the HDUV catalogue). There is a small elongation of the fitted Gaussian in the band ACS F435W, but not as strong as in WFC3 F336W. The other bands also show a difference between the standard deviation of the fitted Gaussian in x and y direction, but none is as significant as in ACS F435W or even WFC3 F336W. I can therefore conclude that the offset and elongation are not systematic in all bands and the effect has to be caused by either a different treatment of the WFC3 F336W band or another effect only affecting this band.

In this case the offset should be visible not only for the potential LyC emission of the LAEs but also in other objects. Therefore I instead stack objects at lower redshifts that are not LAEs, which can be seen in the top right panel in figure 6.22. I set the redshift range for these test objects to $0.8 < z < 2.9$ to make sure no large, low-redshift galaxies that might be elongated influence the stack. Since the signal in the LyC of the LAEs is weak, I select objects that are faint (in ACS F814W) with magnitudes between $27.5 < m < 33$. I take these objects from the HDUV catalogue (Oesch et al., 2018) that I already used in section 6.1. After the selection criteria, I am left with 2084 low redshift objects that I then spatially match to my real LAEs to probe similar parts of the HDUV data. I assign each LAE the low redshift counterpart that is closest, as long as it is within $10''$. This leaves me with 1025 counterparts to my 1479 LAEs. Unfortunately, of the 60 most promising objects that I stack, only 48 obtained a low redshift counterpart this way. Nevertheless, as can be seen in the top right panel of figure 6.22, there is neither an offset nor an elongation for the lower redshift objects. Despite requiring low magnitudes, the signal is still much stronger in the low redshift objects than it is in the LyC of the LAEs, but if I further restrict the magnitude to get fainter objects, I lose in the number of available low redshift counterparts. This experiment shows that I can exclude a problem in the WFC3 F336W band as a reason for the offset in the LyC emission.

There is also the possibility that the charge-coupled device (CCD) readout causes the observed offset in the LyC signal, since the charge transfer efficiency is known to be low for faint signals. The usually observed effect of this is a 'trailing' of faint sources in readout direction (e.g. Noeske et al., 2012, Anderson

& Bedin, 2010⁵). If that were the case here, I would still expect most of the signal to be in the centre, with a faint trail in y -direction, which is not the case. Instead, the signal seems to be shifted entirely, which argues against a problem with the charge transfer efficiency.

6.5.2. Bootstrapping Analysis of Offset

From looking at figure 6.21, where the sample of the 60 most promising objects is split into two based on their properties, I know that the offset is not caused by a single object, since it is visible in all stacks. It could still be possible that the offset and the signal as a whole is dominated by only a few objects, though. To make sure this is not the case, I perform 10 000 bootstrapping experiments by randomly selecting from the sample of 60 objects (random sampling with putting the picks back). The result of this test can be seen in figure 6.24. Note that in order to stabilise the fitting procedure, I convolved the resulting stacks with a two dimensional Gaussian, so as to reduce the influence of noise peaks or insufficiently masked objects in the vicinity of the center (the same as in section 6.5.1). Otherwise if an object with a noise peak is picked multiple times, the fit will be skewed and elongated unreasonably.

From the bootstrapping experiment I can deduce that the offset and signal are not produced by only a few objects. If that were the case, the distribution of offsets in y -direction (green histogram in figure 6.24) would peak at 0, as does the distribution of the x offsets, with only a secondary peak at the position of the offset. However, the distribution is broader than for the x positions with a peak at a negative offset. The situation is less clear for the elongation of the Gaussian (right panel in figure 6.24), as the x and y sigma distributions look similar. The width of the distributions can be explained by noise peaks influencing the fits and creating asymmetric Gaussians.

In conclusion, the bootstrapping showed that the offset is not influenced by only a few objects but the elongation might not be real.

⁵See also http://www.stsci.edu/hst/wfc3/ins_performance/CTE/

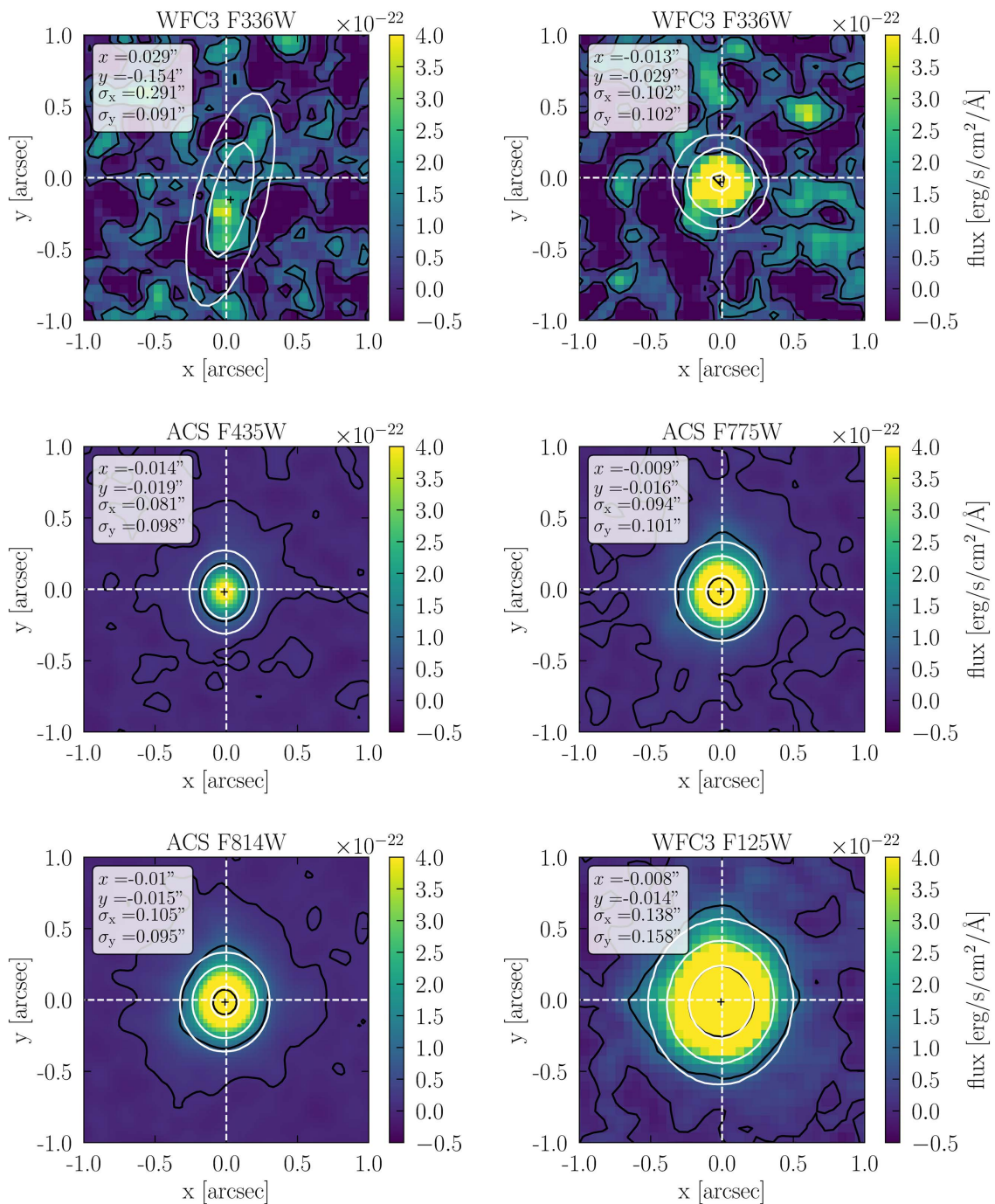


Figure 6.22.: Stacks of the most promising 60 objects in different bands. The cut-outs have a size of $2'' \times 2''$, are smoothed with a Gaussian of $0''.05$ and centered on the `GalFit` positions (indicated by the white dashed lines). The unsmoothed data was fitted with a two dimensional Gaussian (white circles), with the parameters of the fit given in the white boxes in arcseconds. The contour levels (black for the real data and white for the fit) are at 10^{-23} , 10^{-22} and 10^{-21} $\text{erg s}^{-1} \text{cm}^{-2} \text{Å}^{-1}$.

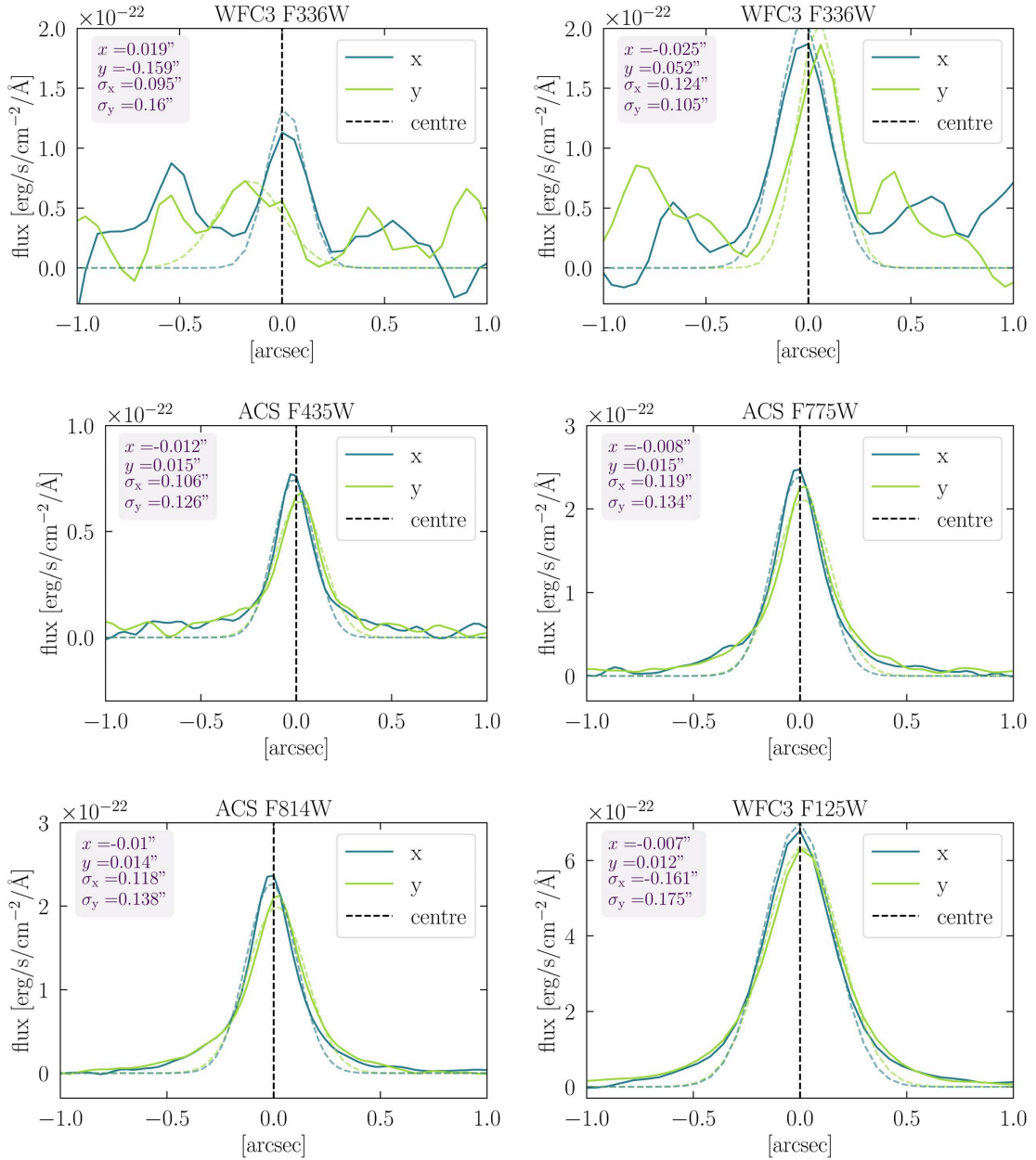


Figure 6.23.: Since the offset is seen mostly along the y-axis I disentangle the axes by collapsing the data (taking the mean) both to the x-axis (dark green) and the y-axis (light green), for the same bands as shown in figure 6.22. The dashed lines show one dimensional Gaussian fits to the data with the fitting parameters given in the purple box. The centre is indicated by a black dashed line.

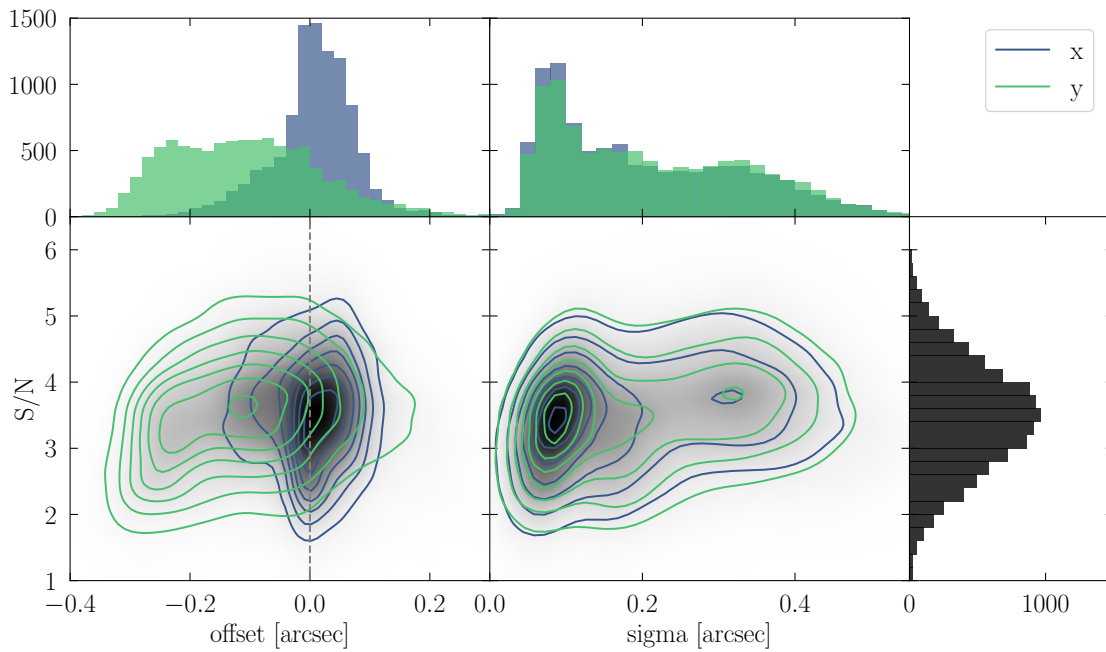


Figure 6.24.: Results of 10 000 bootstrap realisations of stacking the 60 most promising objects and measuring the resulting offsets and widths (sigmas) of the fitted Gaussians (see figure 6.22 for the fitting). The left panel shows the S/N ratios over the offsets as a density plot with a histogram at the top, divided into x positions (blue) and y positions (green). The right panel shows the distribution of sigmas instead of offsets. At the right side, the S/N distribution is shown as a histogram. While the spread in offsets is large and can be explained by the varying signal-to-noise in the stacks, the offset in y -direction stays significant in the bootstrapping experiment. This is not the case for the sigma of the Gaussian, where both the elongation in x and in y direction have the same distribution. This could be caused by the influence of nearby noise peaks that can distort the fitted Gaussian easily, but argues against a systematic elongation in either direction.

Conclusion and Outlook

This thesis gave insights into the properties of two linked categories of high redshift objects: Lyman α emitters and Lyman continuum leakers. The goal was to understand their properties, especially with respect to the Lyman α emission line and the connection between the Lyman α line and the Lyman continuum leakage. At high redshifts, the Lyman α line is one of the most useful tools to study star-forming galaxies that might be the progenitors to our own Milky Way, possibly containing some of the first stars created in the Universe. Additionally, LAEs are candidates for providing the Lyman continuum photons for the ionising background of the IGM, both at intermediate redshifts and at the epoch of reionisation.

7.1. Summary

Before moving on to scientific results, I described in detail in chapter 2 the graphical user interface (GUI) I developed together with the MUSE-Wide team to facilitate the classification of emission line objects found in integral field spectroscopic data, such as Multi-Unit Spectroscopic Explorer (MUSE) provides. Since the use of integral field spectroscopy in the optical range is a relatively new development, there are not many tools yet to help visualise the two spatial and one spectral dimension and to use the data to their full potential. QtClassify offers various possibilities to access the data and make the process of classifying objects efficient and fun. A simple routine for the automated matching of emission lines to redshifts also speeds the process along and provides a good foundation for building a redshift catalogue. The GUI has been tested on 100 MUSE-Wide fields and used to classify over 1000 LAEs and ~ 1600 other objects in the MUSE-Wide survey.

With the classified Lyman α emitters and their redshifts, the next step was to analyse the emission lines and derive the escape fraction distribution in chapter 3. I included both the MUSE-Wide and -Deep surveys and measured the UV continuum from counterparts in the ancillary Hubble Space Telescope (HST) broadband data. A third of the LAEs detected in MUSE were found to have no visible counterpart. Interestingly, I found that 9% of our LAEs have equivalent widths above 240 \AA , the often quoted theoretical value for normal stellar populations. Those objects could contain low-metallicity stars or unusually high star formation rates, also exemplified by an object with a measured equivalent width of $EW_0 \sim 600 \text{ \AA}$. I first showed equivalent width histograms (also for different redshifts, where I found no significant evolution) and point out the weakness of deriving a distribution of equivalent widths from histograms alone, as they reflect the survey limitations in line flux more than the properties of the population of LAEs. To remedy this problem, I constructed for the first time the equivalent width distribution function (EWDF) based on the selection functions of the surveys used. This can be fit with an exponential function to obtain a more reliable scale factor and the EWDF can also be used in combination with the UV luminosity function to get the LAE fraction among high-redshift galaxies (at varying equivalent widths).

In a chapter on the methods used to derive the equivalent widths and the distribution function, I also discussed some additional results and correlations that were found in the data (chapter 4). The GalFit models of the UV continuum counterparts were described and the UV continuum slope was measured,

where I found a median value of $\beta = -1.92$, slightly more dusty than in similar studies in the literature. Among the measured UV continuum counterpart properties were the magnitudes, the axis ratio, Sérsic index and size, which anti-correlates with the Lyman α equivalent width, showing that compact objects have a higher escape of Lyman α emission. I also discussed the line shape of the Lyman α emission line, showing that $\sim 30\%$ of our lines have a blue bump, which is an indicator of a possible outflow in the ISM. The fraction could be even higher, though, if the effect of signal-to-noise in the spectrum is accounted for.

Moving from Lyman α emission to Lyman continuum emission, I showed the approach of stacking Hubble Space Telescope (HST) data in the band WFC3 F336W based on the LAEs found in the MUSE-Wide and -Deep surveys in chapter 5. This band probes the Lyman continuum bluewards of the Lyman limit. I found six possible individual Lyman continuum leaker candidates, two of which are good candidates, four are unreliable candidates. There is a detailed discussion in this chapter on the different caveats and pitfalls one needs to take into account when identifying Lyman continuum leaker candidates, like contamination from low redshift interlopers or bright, close neighbours. Next I selected the most useful LAEs for creating a stacking sample of 60 objects (plus the two promising candidates) and find a signal in the stack (including the candidates) with a signal-to-noise of 5.5. After statistical considerations concerning the noise, I concluded that such a high signal would only be caused by an escape fraction of $f_{\text{esc}} > 80\%$. Separating the 60 objects by their Lyman α emission line properties yielded no clear results with respect to the LyC over UV continuum flux ratio, although asymmetric lines show a slightly stronger signal and thus higher Lyman continuum escape fraction.

When stacking the full sample of selected objects in the WFC3 F336W band to search for Lyman continuum emission, the signal that I found is offset by $\sim 0''.2$. This offset was discussed in detail in the chapter on the methods of detecting Lyman continuum emission (chapter 6). I performed various tests to find the cause of the offset and I ruled out problems with the coordinates, a possible positional inaccuracy in the WFC3 F336W band, insufficiently masked objects and low redshift interlopers. Bootstrapping analysis additionally ruled out the possibility of single objects causing the signal, stacking in different bands showed no offset and even stacking different objects in the same band had a clear, unshifted signal in the centre. The origin of the offset could thus not be fully determined, but it is narrowed down to either a real signal that predominantly appears in a preferred direction, a real signal at the centre but overlaid with a contamination that is very close nearby or a chance alignment of random noise, rather than a Lyman continuum signal with an underlying escape fraction of $f_{\text{esc}} > 80\%$.

These difficulties highlight the challenges of finding Lyman continuum emission at high redshift, especially from faint star-forming galaxies such as LAEs. While I could not answer the question if LAEs are responsible for the bulk of ionising emission that forms the metagalactic ionising background of the Universe and keeps the hydrogen in the IGM ionised, the issues raised here should caution future studies and give some ideas for useful tests, as the results on the escape fraction values of Lyman continuum emission remain inconclusive.

7.2. Outlook

In this section I will give an overview of future improvements of the processes and analysis shown in this thesis. For the purpose of classifying emission lines found in integral field spectroscopic data I developed the GUI `QtClassify`, which also contains a simple algorithm for automatic classification and redshift determination. This algorithm could be improved in the future by including more sophisticated redshift matching techniques by e.g. comparing to template spectra. Another improvement would be the inclusion of a process to determine counterparts in photometric data to the found objects in MUSE, adding this step of the process to the GUI as well. In the future, the GUI could also be tested data from other integral field spectrographs or with input catalogues that are not obtained by LSDCat, as it should be applicable to other data and science cases as well.

Improvements of the Equivalent Width Distribution Function

Moving on the analysis of the found and classified LAEs, I have shown in this thesis that creating histograms of equivalent widths is influenced by the properties of the survey more than the properties of the LAE population, which is why the construction of the EWDF is much more significant. I constructed a preliminary EWDF from LAEs in the MUSE-Wide and -Deep surveys, but as I mentioned the full construction of the selection function is still to be obtained. In addition, the analysis done in this thesis was not performed using the full sample of MUSE-Wide LAEs available (by now the full 100 fields are available, while I only used the first 60 fields so far). This will constrain the EWDF further and make it possible to get more precise estimations of the fraction of LAEs among the general galaxy population.

Alignment of UV Continuum and Lyman α Emission

With the measured properties of both the morphology of the UV continuum counterpart of the LAEs as well as the properties of the emission line, it is also possible to further study their connection. Since I showed in this thesis that the axis ratio (as a proxy for the viewing angle) of the LAE is not necessarily connected to the Lyman α equivalent widths, the question remains if the Lyman α emission indeed escapes the galaxy mostly face-on or if there are other mechanisms at play. Another way to test to alignment of Lyman α emission and UV continuum emission would be to use the spatially extended Lyman α halos to measure the angle in which Lyman α escapes. The Lyman α halos could follow the same morphology as the UV continuum or they could be elongated perpendicular to the plane of the galaxy (estimated from the axis ratio). This analysis is only possible for some of the (Lyman α) brightest objects yet, that allow us to study morphology of the Lyman α emission. One could nevertheless use the MUSE-Wide LAEs and create spatial cut-outs of the spectral range of their Lyman α line and rotate them to align in the UV continuum. If the Lyman α emission now also shows an asymmetric or elongated signal, that would be a clear indication of a correlation between the UV continuum and Lyman α halo morphologies. Such a result would then argue against a mostly irregular morphology of high-redshift galaxies. The stacks could also be divided into groups of different equivalent widths or compactness of the objects.

Investigating the Blue Bump

The Lyman α halo was assumed to be uniform in its line shape properties in this thesis, but this is not necessarily the case. Different parts of the LAE or of its halo could produce Lyman α lines with different properties due to differences in neutral hydrogen column density, star formation or kinematics. A radial trend could again be investigated by stacking LAE halos, but this way we would lose the spatial information. Instead, the MUSE-Deep data (especially the UDF 10) with its longer exposure time could be used to study Lyman α halos individually with respect to different line shape properties. Specifically the existence of the blue bump is interesting in this respect, as it makes it possible to study the neutral hydrogen column density from the separation between the two peaks and gives us information on the kinematics as well. It could also be a potential indicator of Lyman continuum leakage, which has not been studied in this thesis as the individual LAE halo sizes for the MUSE-Wide sample are not yet measured. The blue bump itself is interesting to study as well. I have shown already that its appearance depends on the signal-to-noise in the data, which could mean that most (if not all) LAEs intrinsically have blue bumps. This analysis was only done with the MUSE-Wide objects yet, but it would be ideal to include the MUSE-Deep spectra as well.

Improving the Lyman Continuum Detection

When it comes to the detection of Lyman continuum emission from high-redshift LAEs found in MUSE, a good improvement would be to obtain deeper data with the WFC3 F336W HST band to increase the

signal-to-noise and more securely rule out the possibility of interlopers and confirm the presence of Lyman continuum leakage. Increasing the HDUV footprint to include more MUSE-Wide fields would also be helpful, especially for finding the source of the mysterious offset. Detecting Lyman continuum emission is not only possible in photometry, of course, but also in the MUSE spectra directly. As I have cautioned, the MUSE data is not as deep in the Lyman continuum as the HST data, but again we could use stacking to try and constrain at least an upper limit of Lyman continuum leakage. An added complication is the redshift coverage of MUSE, which only allows us to study Lyman continuum emission in the MUSE spectra directly at redshifts of $z \sim 4$ and above. At these redshifts, the IGM transmission of Lyman continuum is already very low, making it unlikely although not impossible that Lyman continuum could be detected with a high significance. However, especially the MUSE-Deep objects are promising candidates for such a test, as the MUSE-Deep data has a longer exposure time than MUSE-Wide. Ideally, though, one would need spectra that extend more towards the blue in order to detect Lyman continuum emission also from LAEs at lower redshifts with a higher IGM transmission. This could be achieved with the planned instrument BlueMUSE (Richard, 2019), that is proposed to go down to a wavelength of $\lambda \sim 3500\text{\AA}$, allowing the study of Lyman continuum emission from LAEs down to redshifts of $z \sim 3$. BlueMUSE is currently in its first stages, but is a promising upcoming tool for studying LAEs and their Lyman continuum leakage. As I have shown in this thesis, integral field spectroscopy is ideal to find LAEs without any photometric prior (from narrow-band imaging) and without the need of follow-up spectroscopy, which is why sticking to the concept of MUSE for future instruments is useful. The lower wavelength range of BlueMUSE would make it possible to observe Lyman continuum emission from LAEs at a redshift where the IGM is transparent enough to make a detection possible, but the drawback would be that the observed redshifts are further away from the Epoch of Reionisation (EoR). The smaller redshifts allow for a better spatial resolution of the LAEs, though, which makes it possible to probe the CGM through their Lyman α halos in lower-redshift analogues of the LAEs likely responsible for the EoR. Another benefit of BlueMUSE would be the planned increased spectral resolution, which would allow a more detailed study of the Lyman α line structure, possibly detecting even more than two peaks in more complex kinematic configurations or studying the variations of Lyman α line shape properties within their halos.

Open questions

In this thesis I have addressed questions concerning the nature of high-redshift LAEs and their emitted Lyman α and Lyman continuum emission. The distribution of equivalent widths among LAEs and the occurrence of unusually high equivalent widths can give an insight into the underlying stellar populations of the LAEs but also the neutral hydrogen column density. The specific internal properties of high-redshift LAEs remain elusive however, such as the source of the ionising emission that produces such high equivalent widths. The most common assumption is that it comes from star formation at low metallicities, but other sources could be X-Ray binaries or Wolf-Rayet stars, which produce a large number of Lyman α photons themselves. Since LAEs are typically not very massive, even a low number of Wolf-Rayet stars could dominate the spectrum, which is usually not accounted for in models of LAEs.

Not only the sources of ionising emission, but also their radiative transfer through the galaxies is still a topic of current research. The models are getting more and more sophisticated when it comes to complexity of the kinematics and composition of the ISM that the Lyman α photons have to pass on their way. However, since the morphologies of high-redshift LAEs are poorly understood, it is not clear if Lyman α emission escapes in the same regions as it is produced, this following the UV continuum morphology of the galaxy or if it is scattered until the Lyman α halo has little resemblance with the UV morphology. Related to this question is probably the most important topic for current upcoming studies, the connection between the emission from the LAEs and the IGM, which is the circumgalactic medium (CGM) that the Lyman α and Lyman continuum emission has to pass through. How does it influence the escape fractions and can we understand its properties by looking at the imprint it leaves on the Lyman α line? We now know that most or all LAEs have Lyman α halos, but their properties and connections to the galaxy and the IGM need to be studied. Especially their morphology is interesting, as it could be possible that the Lyman α emission predominantly escapes in the halo for some galaxies instead of in the same region as the UV continuum, which could lead to missing them in surveys. Due to this, the line between LAEs and Lyman break galaxies (LBGs) or even between LAEs and the rest of the galaxy population is becoming thinner, but also because as we have seen here, the LAE fraction at low equivalent widths is close to unity. If Lyman α emission can be found in most galaxies, this would make LAEs the best candidates to reionise the Universe. By now, the consensus leans towards this explanation (as opposed to AGN) but the numbers still do not match perfectly. Eventually, more research is needed to fully understand the epoch of reionisation and the role that LAEs can play in it.

Bibliography

- Abel, T., Bryan, G. L., & Norman, M. L. 2002, *Science*, 295, 93
- Abruzzo, M. W., Narayanan, D., Davé, R., & Thompson, R. 2018, arXiv e-prints, arXiv:1803.02374
- Acquaviva, V., Gawiser, E., & Guaita, L. 2011, *Astrophysical Journal*, 737, 47
- Acquaviva, V., Vargas, C., Gawiser, E., & Guaita, L. 2012, *Astrophysical Journal, Letters*, 751, L26
- Adams, J. J., Blanc, G. A., Hill, G. J., et al. 2011, *Astrophysical Journal, Supplement*, 192, 5
- Adelberger, K. L., Steidel, C. C., Shapley, A. E., & Pettini, M. 2003, *Astrophysical Journal*, 584, 45
- Alexandroff, R. M., Heckman, T. M., Borthakur, S., Overzier, R., & Leitherer, C. 2015, *Astrophysical Journal*, 810, 104
- Anderson, J. & Bedin, L. R. 2010, *Publications of the ASP*, 122, 1035
- Ando, M., Ohta, K., Iwata, I., et al. 2006, *Astrophysical Journal, Letters*, 645, L9
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *Astronomy and Astrophysics*, 558, A33
- Bacon, R., Accardo, M., Adjali, L., et al. 2010, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 7735, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 8
- Bacon, R., Brinchmann, J., Richard, J., et al. 2015, *Astronomy and Astrophysics*, 575, A75
- Bacon, R., Conseil, S., Mary, D., et al. 2017, *Astronomy and Astrophysics*, 608, A1
- Baker, J. G. & Menzel, D. H. 1938, *Astrophysical Journal*, 88, 52
- Bassett, R., Ryan-Weber, E. V., Cooke, J., et al. 2019, *Monthly Notices of the RAS*, 483, 5223
- Beckwith, S. V. W., Stiavelli, M., Koekemoer, A. M., et al. 2006, *Astronomical Journal*, 132, 1729
- Behrens, C. & Braun, H. 2014, *Astronomy and Astrophysics*, 572, A74
- Behrens, C., Dijkstra, M., & Niemeyer, J. C. 2014, *Astronomy and Astrophysics*, 563, A77
- Berg, D. A., Chisholm, J., Erb, D. K., et al. 2019, *Astrophysical Journal, Letters*, 878, L3
- Bergvall, N., Zackrisson, E., Andersson, B. G., et al. 2006, *Astronomy and Astrophysics*, 448, 513
- Bertin, E. & Arnouts, S. 1996, *Astronomy and Astrophysics, Supplement*, 117, 393
- Bian, F., Fan, X., McGreer, I., Cai, Z., & Jiang, L. 2017, *Astrophysical Journal, Letters*, 837, L12
- Blanc, G. A., Adams, J. J., Gebhardt, K., et al. 2011, *Astrophysical Journal*, 736, 31
- Bond, N. A., Gawiser, E., Guaita, L., et al. 2012, *Astrophysical Journal*, 753, 95
- Borthakur, S., Heckman, T. M., Leitherer, C., & Overzier, R. A. 2014, *Science*, 346, 216
- Boutsia, K., Grazian, A., Giallongo, E., et al. 2011, *Astrophysical Journal*, 736, 41
- Bouwens, R. J., Illingworth, G. D., Franx, M., et al. 2009, *Astrophysical Journal*, 705, 936
- Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2014, *Astrophysical Journal*, 793, 115
- Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2015, *Astrophysical Journal*, 803, 34
- Bouwens, R. J., Smit, R., Labbé, I., et al. 2016, *Astrophysical Journal*, 831, 176
- Bowler, R. A. A., McLure, R. J., Dunlop, J. S., et al. 2017, *Monthly Notices of the RAS*, 469, 448
- Bowman, J. D., Rogers, A. E. E., Monsalve, R. A., Mozdzen, T. J., & Mahesh, N. 2018, *Nature*, 555, 67
- Bridge, J. S., Hayes, M., Melinder, J., et al. 2018, *Astrophysical Journal*, 852, 9
- Brinchmann, J., Inami, H., Bacon, R., et al. 2017, *Astronomy and Astrophysics*, 608, A3
- Bruzual, G. & Charlot, S. 2003, *Monthly Notices of the RAS*, 344, 1000

- Calzetti, D., Kinney, A. L., & Storchi-Bergmann, T. 1994, *Astrophysical Journal*, 429, 582
- Cantalupo, S., Porciani, C., & Lilly, S. J. 2008, *Astrophysical Journal*, 672, 48
- Cardamone, C., Schawinski, K., Sarzi, M., et al. 2009, *Monthly Notices of the RAS*, 399, 1191
- Caruana, J., Bunker, A. J., Wilkins, S. M., et al. 2012, *Monthly Notices of the RAS*, 427, 3055
- Caruana, J., Bunker, A. J., Wilkins, S. M., et al. 2014, *Monthly Notices of the RAS*, 443, 2831
- Caruana, J., Wisotzki, L., Herenz, E. C., et al. 2018, *Monthly Notices of the RAS*, 473, 30
- Cassata, P., Tasca, L. A. M., Le Fèvre, O., et al. 2015, *Astronomy and Astrophysics*, 573, A24
- Castellano, M., Fontana, A., Grazian, A., et al. 2012, *Astronomy and Astrophysics*, 540, A39
- Charlot, S. & Fall, S. M. 1993, *Astrophysical Journal*, 415, 580
- Chisholm, J., Orlitová, I., Schaerer, D., et al. 2017, *Astronomy and Astrophysics*, 605, A67
- Chonis, T. S., Blanc, G. A., Hill, G. J., et al. 2013, *Astrophysical Journal*, 775, 99
- Ciardullo, R., Gronwall, C., Wolf, C., et al. 2012, *Astrophysical Journal*, 744, 110
- Ciardullo, R., Zeimann, G. R., Gronwall, C., et al. 2014, *Astrophysical Journal*, 796, 64
- Cicone, C., De Breuck, C., Chen, C.-C., et al. 2019, in *Bulletin of the AAS*, Vol. 51, 82
- Clarke, C. & Oey, M. S. 2002, *Monthly Notices of the RAS*, 337, 1299
- Cooke, J., Ryan-Weber, E. V., Garel, T., & Díaz, C. G. 2014, *Monthly Notices of the RAS*, 441, 837
- Cowie, L. L., Barger, A. J., & Hu, E. M. 2011, *Astrophysical Journal*, 738, 136
- Cowie, L. L., Barger, A. J., & Trouille, L. 2009, *Astrophysical Journal*, 692, 1476
- Crighton, N. H. M., Prochaska, J. X., Murphy, M. T., et al. 2019, *Monthly Notices of the RAS*, 482, 1456
- Cybert, R. H., Fields, B. D., Olive, K. A., & Yeh, T.-H. 2016, *Reviews of Modern Physics*, 88, 015004
- de Barros, S., Vanzella, E., Amorín, R., et al. 2016, *Astronomy and Astrophysics*, 585, A51
- de Vaucouleurs, G. 1948, *Annales d'Astrophysique*, 11, 247
- Deharveng, J.-M., Buat, V., Le Brun, V., et al. 2001, *Astronomy and Astrophysics*, 375, 805
- Diener, C., Wisotzki, L., Schmidt, K. B., et al. 2017, *Monthly Notices of the RAS*, 471, 3186
- Dijkstra, M. 2014, *Publications of the Astron. Soc. of Australia*, 31, e040
- Dijkstra, M. 2017, arXiv e-prints, arXiv:1704.03416
- Dijkstra, M., Haiman, Z., & Spaans, M. 2006a, *Astrophysical Journal*, 649, 14
- Dijkstra, M., Haiman, Z., & Spaans, M. 2006b, *Astrophysical Journal*, 649, 37
- Drake, A. B., Garel, T., Wisotzki, L., et al. 2017a, *Astronomy and Astrophysics*, 608, A6
- Drake, A. B., Guiderdoni, B., Blaizot, J., et al. 2017b, *Monthly Notices of the RAS*, 471, 267
- Drozdovsky, I., Yan, L., Chen, H.-W., et al. 2005, *Astronomical Journal*, 130, 1324
- Duval, F., Östlin, G., Hayes, M., et al. 2016, *Astronomy and Astrophysics*, 587, A77
- Duval, F., Schaerer, D., Östlin, G., & Laursen, P. 2014, *Astronomy and Astrophysics*, 562, A52
- Ekström, S., Georgy, C., Eggenberger, P., et al. 2012, *Astronomy and Astrophysics*, 537, A146
- Ellis, R. S. 2014, arXiv e-prints, arXiv:1411.3330
- Erb, D. K., Pettini, M., Shapley, A. E., et al. 2010, *Astrophysical Journal*, 719, 1168
- Erb, D. K., Steidel, C. C., & Chen, Y. 2018, *Astrophysical Journal*, 862, L10
- Erb, D. K., Steidel, C. C., Trainor, R. F., et al. 2014, *Astrophysical Journal*, 795, 33
- Faisst, A. L. 2016, *Astrophysical Journal*, 829, 99
- Fan, X., Carilli, C. L., & Keating, B. 2006a, *Annual Review of Astron and Astrophys*, 44, 415
- Fan, X., Strauss, M. A., Becker, R. H., et al. 2006b, *Astronomical Journal*, 132, 117
- Feldmeier, J. J., Hagen, A., Ciardullo, R., et al. 2013, *Astrophysical Journal*, 776, 75
- Feltre, A., Charlot, S., & Gutkin, J. 2016, *Monthly Notices of the RAS*, 456, 3354
- Feng, Y., Di-Matteo, T., Croft, R. A., et al. 2016, *Monthly Notices of the RAS*, 455, 2778
- Finkelstein, K. D., Finkelstein, S. L., Tilvi, V., et al. 2015, *Astrophysical Journal*, 813, 78
- Finkelstein, S. L., D'Aloisio, A., Paardekooper, J.-P., et al. 2019, *Astrophysical Journal*, 879, 36
- Finkelstein, S. L., Papovich, C., Dickinson, M., et al. 2013, *Nature*, 502, 524
- Finkelstein, S. L., Rhoads, J. E., Malhotra, S., Pirzkal, N., & Wang, J. 2007, *Astrophysical Journal*, 660, 1023
- Fixsen, D. J. 2009, *Astrophysical Journal*, 707, 916
- Fletcher, T. J., Tang, M., Robertson, B. E., et al. 2019, *Astrophysical Journal*, 878, 87
- Forero-Romero, J. E., Yepes, G., Gottlöber, S., & Prada, F. 2012, *Monthly Notices of the RAS*, 419, 952
- Fujita, A., Martin, C. L., Mac Low, M.-M., & Abel, T. 2003, *Astrophysical Journal*, 599, 50
- Fumagalli, M., O'Meara, J. M., & Prochaska, J. X. 2016, *Monthly Notices of the RAS*, 455, 4100
- Furlanetto, S. R. & Briggs, F. H. 2004, *New Astronomy Review*, 48, 1039
- Furlanetto, S. R., Schaye, J., Springel, V., & Hernquist, L. 2005, *Astrophysical Journal*, 622, 7
- Furusawa, H., Kashikawa, N., Kobayashi, M. A. R., et al. 2016, *Astrophysical Journal*, 822, 46
- Garel, T., Blaizot, J., Guiderdoni, B., et al. 2012, *Monthly Notices of the RAS*, 422, 310

- Garilli, B., Fumana, M., Franzetti, P., et al. 2010, *Publications of the ASP*, 122, 827
- Gawiser, E., Francke, H., Lai, K., et al. 2007, *Astrophysical Journal*, 671, 278
- Gawiser, E., van Dokkum, P. G., Gronwall, C., et al. 2006, *Astrophysical Journal, Letters*, 642, L13
- George, E. M., Reichardt, C. L., Aird, K. A., et al. 2015, *Astrophysical Journal*, 799, 177
- Giallongo, E., Grazian, A., Fiore, F., et al. 2015, *Astronomy and Astrophysics*, 578, A83
- Giavalisco, M., Ferguson, H. C., Koekemoer, A. M., et al. 2004, *Astrophysical Journal*, 600, L93
- Giri, S. K., Zackrisson, E., Binggeli, C., Pelckmans, K., & Cubo, R. 2019, arXiv e-prints, arXiv:1903.01483
- Grazian, A., Giallongo, E., Boutsia, K., et al. 2018, *Astronomy and Astrophysics*, 613, A44
- Grazian, A., Giallongo, E., Gerbasi, R., et al. 2016, *Astronomy and Astrophysics*, 585, A48
- Grazian, A., Giallongo, E., Paris, D., et al. 2017, *Astronomy and Astrophysics*, 602, A18
- Gronke, M. 2017, *Astronomy and Astrophysics*, 608, A139
- Gronke, M., Bull, P., & Dijkstra, M. 2015, *Astrophysical Journal*, 812, 123
- Gronke, M. & Dijkstra, M. 2014, *Monthly Notices of the RAS*, 444, 1095
- Gronke, M. & Dijkstra, M. 2016, *Astrophysical Journal*, 826, 14
- Gronwall, C., Bond, N. A., Ciardullo, R., et al. 2011, *Astrophysical Journal*, 743, 9
- Gronwall, C., Ciardullo, R., Hickey, T., et al. 2007, *Astrophysical Journal*, 667, 79
- Guaita, L., Acquaviva, V., Padilla, N., et al. 2011, *Astrophysical Journal*, 733, 114
- Guaita, L., Melinder, J., Hayes, M., et al. 2015, *Astronomy and Astrophysics*, 576, A51
- Gunn, J. E. & Peterson, B. A. 1965, *Astrophysical Journal*, 142, 1633
- Guo, Y., Ferguson, H. C., Giavalisco, M., et al. 2013, *Astrophysical Journal, Supplement*, 207, 24
- Haiman, Z., Spaans, M., & Quataert, E. 2000, *Astrophysical Journal, Letters*, 537, L5
- Hashimoto, T., Garel, T., Guiderdoni, B., et al. 2017, *Astronomy and Astrophysics*, 608, A10
- Hashimoto, T., Ouchi, M., Shimasaku, K., et al. 2013, *Astrophysical Journal*, 765, 70
- Hashimoto, T., Verhamme, A., Ouchi, M., et al. 2015, *Astrophysical Journal*, 812, 157
- Hayes, M., Östlin, G., Atek, H., et al. 2007, *Monthly Notices of the RAS*, 382, 1465
- Hayes, M., Östlin, G., Duval, F., et al. 2014, *Astrophysical Journal*, 782, 6
- Hayes, M., Östlin, G., Schaerer, D., et al. 2013, *Astrophysical Journal*, 765, L27
- Henry, A., Berg, D. A., Scarlata, C., Verhamme, A., & Erb, D. 2018, *Astrophysical Journal*, 855, 96
- Henry, A., Scarlata, C., Martin, C. L., & Erb, D. 2015, *Astrophysical Journal*, 809, 19
- Herenz, E. C., Gruyters, P., Orlitova, I., et al. 2016, *Astronomy and Astrophysics*, 587, A78
- Herenz, E. C., Urrutia, T., Wisotzki, L., et al. 2017, *Astronomy and Astrophysics*, 606, A12
- Herenz, E. C. & Wisotzki, L. 2017, *Astronomy and Astrophysics*, 602, A111
- Herenz, E. C., Wisotzki, L., Saust, R., et al. 2019, *Astronomy and Astrophysics*, 621, A107
- Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, *Astrophysical Journal, Supplement*, 208, 19
- Hinton, S. R., Davis, T. M., Lidman, C., Glazebrook, K., & Lewis, G. F. 2016, *Astronomy and Computing*, 15, 61
- Hu, E. M., Cowie, L. L., & McMahon, R. G. 1998, *Astrophysical Journal, Letters*, 502, L99
- Hubble, E. 1929, *Proceedings of the National Academy of Science*, 15, 168
- Hunter, J. D. 2007, *Computing In Science & Engineering*, 9, 90
- Inami, H., Bacon, R., Brinchmann, J., et al. 2017, *Astronomy and Astrophysics*, 608, A2
- Inoue, A. K. & Iwata, I. 2008, *Monthly Notices of the RAS*, 387, 1681
- Inoue, A. K., Shimizu, I., Iwata, I., & Tanaka, M. 2014, *Monthly Notices of the RAS*, 442, 1805
- Iwata, I., Inoue, A. K., Matsuda, Y., et al. 2009, *Astrophysical Journal*, 692, 1287
- Izotov, Y. I., Orlitová, I., Schaerer, D., et al. 2016a, *Nature*, 529, 178
- Izotov, Y. I., Schaerer, D., Thuan, T. X., et al. 2016b, *Monthly Notices of the RAS*, 461, 3683
- Izotov, Y. I., Schaerer, D., Worseck, G., et al. 2018a, *Monthly Notices of the RAS*, 474, 4514
- Izotov, Y. I., Worseck, G., Schaerer, D., et al. 2018b, *Monthly Notices of the RAS*, 478, 4851
- Japelj, J., Vanzella, E., Fontanot, F., et al. 2017, *Monthly Notices of the RAS*, 468, 389
- Jaskot, A. E. & Oey, M. S. 2013, *Astrophysical Journal*, 766, 91
- Jaskot, A. E. & Ravindranath, S. 2016, *Astrophysical Journal*, 833, 136
- Jiang, L., Egami, E., Fan, X., et al. 2013, *Astrophysical Journal*, 773, 153
- Jung, I., Finkelstein, S. L., Livermore, R. C., et al. 2018, *Astrophysical Journal*, 864, 103
- Karman, W., Caputi, K. I., Caminha, G. B., et al. 2017, *Astronomy and Astrophysics*, 599, A28
- Kashikawa, N., Nagao, T., Toshikawa, J., et al. 2012, *Astrophysical Journal*, 761, 85
- Kashikawa, N., Shimasaku, K., Matsuda, Y., et al. 2011, *Astrophysical Journal*, 734, 119
- Keenan, R. P., Oey, M. S., Jaskot, A. E., & James, B. L. 2017, *Astrophysical Journal*, 848, 12
- Kerutt, J. 2017, *QtClassify: IFS data emission line candidates classifier*
- Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, *Astrophysical Journal, Supplement*, 197, 36

- Kogut, A., Spergel, D. N., Barnes, C., et al. 2003, *Astrophysical Journal*, Supplement, 148, 161
- Kornei, K. A., Shapley, A. E., Erb, D. K., et al. 2010, *Astrophysical Journal*, 711, 693
- Kulas, K. R., Shapley, A. E., Kollmeier, J. A., et al. 2012, *Astrophysical Journal*, 745, 33
- Kunth, D., Mas-Hesse, J. M., Terlevich, E., et al. 1998, *Astronomy and Astrophysics*, 334, 11
- Lake, E. A., Zheng, Z., Cen, R., et al. 2015, *IAU General Assembly*, 22, 54115
- Lambas, D. G., Maddox, S. J., & Loveday, J. 1992, *Monthly Notices of the RAS*, 258, 404
- Laporte, N., Ellis, R. S., Boone, F., et al. 2017, *Astrophysical Journal*, 837, L21
- Larson, R. L., Finkelstein, S. L., Pirzkal, N., et al. 2018, *Astrophysical Journal*, 858, 94
- Laursen, P., Duval, F., & Östlin, G. 2013, *Astrophysical Journal*, 766, 124
- Le Fèvre, O., Tasca, L. A. M., Cassata, P., et al. 2015, *Astronomy and Astrophysics*, 576, A79
- Leavitt, H. S. & Pickering, E. C. 1912, *Harvard College Observatory Circular*, 173, 1
- Leclercq, F., Bacon, R., Wisotzki, L., et al. 2017a, *Astronomy and Astrophysics*, 608, A8
- Leclercq, F., Bacon, R., Wisotzki, L., et al. 2017b, *Astronomy and Astrophysics*, 608, A8
- Lee, H.-W. 2013, *Astrophysical Journal*, 772, 123
- Leitet, E., Bergvall, N., Hayes, M., Linné, S., & Zackrisson, E. 2013, *Astronomy and Astrophysics*, 553, A106
- Leitet, E., Bergvall, N., Piskunov, N., & Andersson, B.-G. 2011, *Astronomy and Astrophysics*, 532, A107
- Leitherer, C., Ekström, S., Meynet, G., et al. 2014, *Astrophysical Journal*, Supplement, 212, 14
- Leitherer, C., Hernandez, S., Lee, J. C., & Oey, M. S. 2016, *Astrophysical Journal*, 823, 64
- Leitherer, C., Ortiz Otálvaro, P. A., Bresolin, F., et al. 2010, *Astrophysical Journal*, Supplement, 189, 309
- Leitherer, C., Schaerer, D., Goldader, J. D., et al. 1999, *Astrophysical Journal*, Supplement, 123, 3
- Lemaître, G. 1927, *Annales de la Société Scientifique de Bruxelles*, 47, 49
- Levesque, E. M. & Richardson, M. L. A. 2014, *Astrophysical Journal*, 780, 100
- Lintott, C. J., Schawinski, K., Slosar, A., et al. 2008, *Monthly Notices of the RAS*, 389, 1179
- Luo, B., Brandt, W. N., Xue, Y. Q., et al. 2017, *Astrophysical Journal*, Supplement, 228, 2
- Lyman, T. 1906, *Astrophysical Journal*, 23, 181
- Madau, P. & Dickinson, M. 2014, *Annual Review of Astron and Astrophys*, 52, 415
- Madau, P. & Haardt, F. 2015, *Astrophysical Journal*, Letters, 813, L8
- Madau, P., Haardt, F., & Rees, M. J. 1999, *Astrophysical Journal*, 514, 648
- Madau, P. & Shull, J. M. 1996, *Astrophysical Journal*, 457, 551
- Mainieri, V., Rigopoulou, D., Lehmann, I., et al. 2005, *Monthly Notices of the RAS*, 356, 1571
- Maiolino, R., Schneider, R., Oliva, E., et al. 2004, *Nature*, 431, 533
- Malhotra, S. & Rhoads, J. E. 2002, *Astrophysical Journal*, Letters, 565, L71
- Malhotra, S. & Rhoads, J. E. 2004, *Astrophysical Journal*, Letters, 617, L5
- Marchi, F., Pentericci, L., Guaita, L., et al. 2018, *Astronomy and Astrophysics*, 614, A11
- Marchi, F., Pentericci, L., Guaita, L., et al. 2019, *arXiv e-prints*, arXiv:1903.08593
- Marino, R. A., Cantalupo, S., Lilly, S. J., et al. 2018, *Astrophysical Journal*, 859, 53
- Mas-Hesse, J. M., Kunth, D., Tenorio-Tagle, G., et al. 2003, *Astrophysical Journal*, 598, 858
- Maseda, M. V., Bacon, R., Franx, M., et al. 2018, *Astrophysical Journal*, Letters, 865, L1
- Maseda, M. V., Brinchmann, J., Franx, M., et al. 2017, *Astronomy and Astrophysics*, 608, A4
- Masters, D. & Capak, P. 2011, *Publications of the ASP*, 123, 638
- Masters, D., Capak, P., Salvato, M., et al. 2012, *Astrophysical Journal*, 755, 169
- Matsuda, Y., Yamada, T., Hayashino, T., et al. 2012, *Monthly Notices of the RAS*, 425, 878
- McCandliss, S., Calzetti, D., Ferguson, H. C., et al. 2019, in *Bulletin of the AAS*, Vol. 51, 535
- McLinden, E. M., Finkelstein, S. L., Rhoads, J. E., et al. 2011, *Astrophysical Journal*, 730, 136
- McLinden, E. M., Rhoads, J. E., Malhotra, S., et al. 2014, *Monthly Notices of the RAS*, 439, 446
- McQuinn, M. 2016, *Annual Review of Astron and Astrophys*, 54, 313
- Meiksin, A. A. 2009, *Reviews of Modern Physics*, 81, 1405
- Messa, M., Adamo, A., Östlin, G., et al. 2019, *Monthly Notices of the RAS*, 1480
- Meurer, G. R., Heckman, T. M., & Calzetti, D. 1999, *Astrophysical Journal*, 521, 64
- Meynet, G., Maeder, A., Schaller, G., Schaerer, D., & Charbonnel, C. 1994, *Astronomy and Astrophysics*, Supplement, 103, 97
- Micheva, G., Iwata, I., & Inoue, A. K. 2017a, *Monthly Notices of the RAS*, 465, 302
- Micheva, G., Iwata, I., Inoue, A. K., et al. 2017b, *Monthly Notices of the RAS*, 465, 316
- Micheva, G., Oey, M. S., Keenan, R. P., Jaskot, A. E., & James, B. L. 2018a, *Astrophysical Journal*, 867, 2
- Micheva, G., Östlin, G., Zackrisson, E., et al. 2018b, *Astronomy and Astrophysics*, 615, A46
- Mitra, S., Choudhury, T. R., & Ferrara, A. 2015, *Monthly Notices of the RAS*, 454, L76
- Mitra, S., Ferrara, A., & Choudhury, T. R. 2013, *Monthly Notices of the RAS*, 428, L1

- Momcheva, I. G., Brammer, G. B., van Dokkum, P. G., et al. 2016, *Astrophysical Journal*, Supplement, 225, 27
- Momose, R., Ouchi, M., Nakajima, K., et al. 2014, *Monthly Notices of the RAS*, 442, 110
- Momose, R., Ouchi, M., Nakajima, K., et al. 2016, *Monthly Notices of the RAS*, 457, 2318
- Mostardi, R. E., Shapley, A. E., Nestor, D. B., et al. 2013, *Astrophysical Journal*, 779, 65
- Mostardi, R. E., Shapley, A. E., Steidel, C. C., et al. 2015, *Astrophysical Journal*, 810, 107
- Naidu, R. P., Forrest, B., Oesch, P. A., Tran, K.-V. H., & Holden, B. P. 2018, *Monthly Notices of the RAS*, 478, 791
- Naidu, R. P., Oesch, P. A., Reddy, N., et al. 2017, *Astrophysical Journal*, 847, 12
- Nakajima, K., Ellis, R. S., Iwata, I., et al. 2016, *Astrophysical Journal*, Letters, 831, L9
- Nakajima, K. & Ouchi, M. 2014, *Monthly Notices of the RAS*, 442, 900
- Nestor, D. B., Shapley, A. E., Kornei, K. A., Steidel, C. C., & Siana, B. 2013, *Astrophysical Journal*, 765, 47
- Nestor, D. B., Shapley, A. E., Steidel, C. C., & Siana, B. 2011, *Astrophysical Journal*, 736, 18
- Neufeld, D. A. 1990, *Astrophysical Journal*, 350, 216
- Neufeld, D. A. 1991, *Astrophysical Journal*, Letters, 370, L85
- Nilsson, K. K., Møller, P., Møller, O., et al. 2007, *Astronomy and Astrophysics*, 471, 71
- Nilsson, K. K., Møller-Nilsson, O., Møller, P., Fynbo, J. P. U., & Shapley, A. E. 2009, *Monthly Notices of the RAS*, 400, 232
- Noeske, K., Baggett, S., Bushouse, H., et al. 2012, WFC3 UVIS Charge Transfer Efficiency October 2009 to October 2011, Tech. rep.
- Norman, C., Hasinger, G., Giacconi, R., et al. 2002, *Astrophysical Journal*, 571, 218
- Oesch, P., Montes, M., & HDUV Survey Team. 2015, IAU General Assembly, 22, 2252813
- Oesch, P. A., Montes, M., Reddy, N., et al. 2018, *Astrophysical Journal*, Supplement, 237, 12
- Ono, Y., Ouchi, M., Shimasaku, K., et al. 2010, *Monthly Notices of the RAS*, 402, 1580
- Onoue, M., Kashikawa, N., Willott, C. J., et al. 2017, *Astrophysical Journal*, Letters, 847, L15
- Osterbrock, D. E. 1989, *Astrophysics of gaseous nebulae and active galactic nuclei*
- Osterbrock, D. E. & Ferland, G. J. 2006, *Astrophysics of Gaseous Nebulae and Active Galactic Nuclei* (University Science Books)
- Östlin, G., Hayes, M., Duval, F., et al. 2014, *Astrophysical Journal*, 797, 11
- Östlin, G., Hayes, M., Kunth, D., et al. 2009, *Astronomical Journal*, 138, 923
- Ota, K., Iye, M., Kashikawa, N., et al. 2017, *Astrophysical Journal*, 844, 85
- Oteo, I., Bongiovanni, A., Pérez García, A. M., et al. 2012, *Astronomy and Astrophysics*, 541, A65
- Ouchi, M., Mobasher, B., Shimasaku, K., et al. 2009, *Astrophysical Journal*, 706, 1136
- Ouchi, M., Shimasaku, K., Akiyama, M., et al. 2008, *Astrophysical Journal*, Supplement, 176, 301
- Ouchi, M., Shimasaku, K., Furusawa, H., et al. 2010, *Astrophysical Journal*, 723, 869
- Overzier, R. A., Heckman, T. M., Tremonti, C., et al. 2009, *Astrophysical Journal*, 706, 203
- Oyarzún, G. A., Blanc, G. A., González, V., et al. 2016, *Astrophysical Journal*, Letters, 821, L14
- Paalvast, M., Verhamme, A., Straka, L. A., et al. 2018, *Astronomy and Astrophysics*, 618, A40
- Padilla, N. D. & Strauss, M. A. 2008, *Monthly Notices of the RAS*, 388, 1321
- Pallottini, A., Ferrara, A., Pacucci, F., et al. 2015, *Monthly Notices of the RAS*, 453, 2465
- Papovich, C., Finkelstein, S. L., Ferguson, H. C., Lotz, J. M., & Giavalisco, M. 2011, *Monthly Notices of the RAS*, 412, 1123
- Parsa, S., Dunlop, J. S., & McLure, R. J. 2018, *Monthly Notices of the RAS*, 474, 2904
- Partridge, R. B. & Peebles, P. J. E. 1967, *Astrophysical Journal*, 147, 868
- Pawlik, A. H., Schaye, J., & van Scherpenzeel, E. 2009, *Monthly Notices of the RAS*, 394, 1812
- Payne, C. H. 1925, PhD thesis, RADCLIFFE COLLEGE.
- Peebles, P. J. E. 1968, *Astrophysical Journal*, 153, 1
- Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, *Astronomical Journal*, 124, 266
- Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2010, *Astronomical Journal*, 139, 2097
- Piqueras, L., Conseil, S., Shepherd, M., et al. 2017, arXiv e-prints, arXiv:1710.03554
- Planck Collaboration, Adam, R., Aghanim, N., et al. 2016, *Astronomy and Astrophysics*, 596, A108
- Planck Collaboration, Akrami, Y., Arroja, F., et al. 2018, arXiv e-prints, arXiv:1807.06205
- Prochaska, J. X., O'Meara, J. M., & Worseck, G. 2010, *Astrophysical Journal*, 718, 392
- Prochaska, J. X. & Wolfe, A. M. 1999, *Astrophysical Journal*, Supplement, 121, 369
- Rafelski, M., Teplitz, H. I., Gardner, J. P., et al. 2015, *Astronomical Journal*, 150, 31
- Raiter, A., Schaerer, D., & Fosbury, R. A. E. 2010, *Astronomy and Astrophysics*, 523, A64
- Rakic, O., Schaye, J., Steidel, C. C., & Rudie, G. C. 2011, *Monthly Notices of the RAS*, 414, 3265

- Rees, M. J. 1998, *Proceedings of the National Academy of Science*, 95, 47
- Reichardt, C. L. 2016, in *Astrophysics and Space Science Library*, Vol. 423, *Understanding the Epoch of Cosmic Reionization: Challenges and Progress*, ed. A. Mesinger, 227
- Remolina-Gutiérrez, M. C. & Forero-Romero, J. E. 2019, *Monthly Notices of the RAS*, 482, 4553
- Rhoads, J. E., Malhotra, S., Dey, A., et al. 2000, *Astrophysical Journal, Letters*, 545, L85
- Richard, J. 2019, in *The Very Large Telescope in 2030*, 24
- Rivera-Thorsen, T. E., Hayes, M., Östlin, G., et al. 2015, *Astrophysical Journal*, 805, 14
- Robertson, B. E., Furlanetto, S. R., Schneider, E., et al. 2013, *Astrophysical Journal*, 768, 71
- Rutkowski, M. J., Scarlata, C., Haardt, F., et al. 2016, *Astrophysical Journal*, 819, 81
- Rutkowski, M. J., Scarlata, C., Henry, A., et al. 2017, *Astrophysical Journal, Letters*, 841, L27
- Scarlata, C., Colbert, J., Teplitz, H. I., et al. 2009, *Astrophysical Journal*, 706, 1241
- Schaerer, D. 2002, *Astronomy and Astrophysics*, 382, 28
- Schaerer, D. 2003, *Astronomy and Astrophysics*, 397, 527
- Schaerer, D., de Barros, S., & Stark, D. P. 2011a, *Astronomy and Astrophysics*, 536, A72
- Schaerer, D., Hayes, M., Verhamme, A., & Teyssier, R. 2011b, *Astronomy and Astrophysics*, 531, A12
- Schechter, P. 1976, *Astrophysical Journal*, 203, 297
- Schmidt, K. B., Wisotzki, L., Urrutia, T., et al. 2019, *Astronomy and Astrophysics*, 628, A91
- Schroeder, J., Mesinger, A., & Haiman, Z. 2013, *Monthly Notices of the RAS*, 428, 3058
- Scoville, N., Aussel, H., Brusa, M., et al. 2007, *Astrophysical Journal, Supplement*, 172, 1
- Sersic, J. L. 1968, *Atlas de Galaxias Australes*
- Shapley, A. E., Steidel, C. C., Pettini, M., & Adelberger, K. L. 2003, *Astrophysical Journal*, 588, 65
- Shapley, A. E., Steidel, C. C., Pettini, M., Adelberger, K. L., & Erb, D. K. 2006, *Astrophysical Journal*, 651, 688
- Shapley, A. E., Steidel, C. C., Strom, A. L., et al. 2016, *Astrophysical Journal, Letters*, 826, L24
- Shibuya, T., Kashikawa, N., Ota, K., et al. 2012, *Astrophysical Journal*, 752, 114
- Shibuya, T., Ouchi, M., Nakajima, K., et al. 2014, *Astrophysical Journal*, 788, 74
- Shimasaku, K., Kashikawa, N., Doi, M., et al. 2006, *Publications of the ASJ*, 58, 313
- Shukla, H., Mellema, G., Iliev, I. T., & Shapiro, P. R. 2016, *Monthly Notices of the RAS*, 458, 135
- Siana, B., Teplitz, H. I., Colbert, J., et al. 2007, *Astrophysical Journal*, 668, 62
- Skelton, R. E., Whitaker, K. E., Momcheva, I. G., et al. 2014, *Astrophysical Journal, Supplement*, 214, 24
- Sobral, D., Matthee, J., Darvish, B., et al. 2015, *Astrophysical Journal*, 808, 139
- Song, M., Finkelstein, S. L., Gebhardt, K., et al. 2014, *Astrophysical Journal*, 791, 3
- Stark, D. P., Ellis, R. S., Chiu, K., Ouchi, M., & Bunker, A. 2010, *Monthly Notices of the RAS*, 408, 1628
- Stark, D. P., Ellis, R. S., & Ouchi, M. 2011, *Astrophysical Journal, Letters*, 728, L2
- Stasińska, G., Izotov, Y., Morisset, C., & Guseva, N. 2015, *Astronomy and Astrophysics*, 576, A83
- Steidel, C. C., Adelberger, K. L., Shapley, A. E., et al. 2000, *Astrophysical Journal*, 532, 170
- Steidel, C. C., Bogosavljević, M., Shapley, A. E., et al. 2011, *Astrophysical Journal*, 736, 160
- Steidel, C. C., Bogosavljević, M., Shapley, A. E., et al. 2018, *Astrophysical Journal*, 869, 123
- Steidel, C. C., Erb, D. K., Shapley, A. E., et al. 2010, *Astrophysical Journal*, 717, 289
- Steidel, C. C., Giavalisco, M., Dickinson, M., & Adelberger, K. L. 1996, *Astronomical Journal*, 112, 352
- Steidel, C. C., Pettini, M., & Adelberger, K. L. 2001, *Astrophysical Journal*, 546, 665
- Sunyaev, R. A. & Zeldovich, Y. B. 1972, *Comments on Astrophysics and Space Physics*, 4, 173
- Tapken, C., Appenzeller, I., Mehlert, D., Noll, S., & Richling, S. 2004, *Astronomy and Astrophysics*, 416, L1
- Tasca, L. A. M., Le Fèvre, O., Ribeiro, B., et al. 2017, *Astronomy and Astrophysics*, 600, A110
- Teplitz, H. I., Rafelski, M., Kurczynski, P., et al. 2013, *Astronomical Journal*, 146, 159
- Tilvi, V., Papovich, C., Finkelstein, S. L., et al. 2014, *Astrophysical Journal*, 794, 5
- Tumlinson, J. & Shull, J. M. 2000, *Astrophysical Journal, Letters*, 528, L65
- Tumlinson, J., Shull, J. M., & Venkatesan, A. 2003, *Astrophysical Journal*, 584, 608
- Tumlinson, J., Werk, J. K., Thom, C., et al. 2011, *Astrophysical Journal*, 733, 111
- U, V., Hemmati, S., Darvish, B., et al. 2015, *Astrophysical Journal*, 815, 57
- Urrutia, T., Wisotzki, L., Kerutt, J., et al. 2019, *Astronomy and Astrophysics*, 624, A141
- Vanzella, E., de Barros, S., Castellano, M., et al. 2015, *Astronomy and Astrophysics*, 576, A116
- Vanzella, E., de Barros, S., Vasei, K., et al. 2016, *Astrophysical Journal*, 825, 41
- Vanzella, E., Giavalisco, M., Inoue, A. K., et al. 2010a, *Astrophysical Journal*, 725, 1011
- Vanzella, E., Nonino, M., Cupani, G., et al. 2018, *Monthly Notices of the RAS*, 476, L15
- Vanzella, E., Siana, B., Cristiani, S., & Nonino, M. 2010b, *Monthly Notices of the RAS*, 404, 1672
- Vargas, C. J., Bish, H., Acquaviva, V., et al. 2014, *Astrophysical Journal*, 783, 26

- Vázquez, G. A. & Leitherer, C. 2005, *Astrophysical Journal*, 621, 695
- Venemans, B. P., Röttgering, H. J. A., Miley, G. K., et al. 2005, *Astronomy and Astrophysics*, 431, 793
- Verhamme, A., Dubois, Y., Blaizot, J., et al. 2012, *Astronomy and Astrophysics*, 546, A111
- Verhamme, A., Garel, T., Ventou, E., et al. 2018, *Monthly Notices of the RAS*, 478, L60
- Verhamme, A., Orlitová, I., Schaerer, D., & Hayes, M. 2015, *Astronomy and Astrophysics*, 578, A7
- Verhamme, A., Orlitová, I., Schaerer, D., et al. 2017, *Astronomy and Astrophysics*, 597, A13
- Verhamme, A., Schaerer, D., Atek, H., & Tapken, C. 2008, *Astronomy and Astrophysics*, 491, 89
- Verhamme, A., Schaerer, D., & Maselli, A. 2006, *Astronomy and Astrophysics*, 460, 397
- Weilbacher, P. M., Monreal-Ibero, A., Verhamme, A., et al. 2018, *Astronomy and Astrophysics*, 611, A95
- Wisotzki, L., Bacon, R., Blaizot, J., et al. 2016, *Astronomy and Astrophysics*, 587, A98
- Wisotzki, L., Bacon, R., Brinchmann, J., et al. 2018, *Nature*, 562, 229
- Wolfe, A. M., Gawiser, E., & Prochaska, J. X. 2005, *Annual Review of Astron and Astrophys*, 43, 861
- Wuyts, S., Labbé, I., Förster Schreiber, N. M., et al. 2008, *Astrophysical Journal*, 682, 985
- Yamada, T., Matsuda, Y., Kousai, K., et al. 2012, *Astrophysical Journal*, 751, 29
- Yang, H., Malhotra, S., Gronke, M., et al. 2016, *Astrophysical Journal*, 820, 130
- Yang, H., Malhotra, S., Rhoads, J. E., & Wang, J. 2017, *Astrophysical Journal*, 847, 38
- Yoshida, N., Omukai, K., Hernquist, L., & Abel, T. 2006, *Astrophysical Journal*, 652, 6
- Yuma, S., Ohta, K., Yabe, K., et al. 2010, *Astrophysical Journal*, 720, 1016
- Zackrisson, E., Inoue, A. K., & Jensen, H. 2013, *Astrophysical Journal*, 777, 39
- Zheng, Z. & Wallace, J. 2014, *Astrophysical Journal*, 794, 116
- Zheng, Z.-Y., Wang, J.-X., Malhotra, S., et al. 2014, *Monthly Notices of the RAS*, 439, 1101
- Zitrin, A., Labbé, I., Belli, S., et al. 2015, *Astrophysical Journal*, 810, L12

Miscellaneous

B.1. List of Publications

The following is a list of publications I have been involved in as a co-author (in chronological order), mostly as part of the MUSE consortium and the MUSE-Wide team at the AIP.

- **Bacon et al., 2015:** *The MUSE 3D view of the Hubble Deep Field South*
R. Bacon, J. Brinchmann, J. Richard, T. Contini, A. Drake, M. Franx, S. Tacchella, J. Vernet, L. Wisotzki, J. Blaizot, N. Bouché, R. Bouwens, S. Cantalupo, C. M. Carollo, D. Carton, J. Caruana, B. Clément, S. Dreizler, B. Epinat, B. Guiderdoni, C. Herenz, T.-O. Husser, S. Kamann, **J. Kerutt**, W. Kollatschny, D. Krajinovic, S. Lilly, T. Martinsson, L. Michel-Dansac, V. Patricio, J. Schaye, M. Shirazi, K. Soto, G. Soucail, M. Steinmetz, T. Urrutia, P. Weilbacher, and T. de Zeeuw
Astronomy & Astrophysics (2015), 575, A75, doi: [10.1051/0004-6361/201425419](https://doi.org/10.1051/0004-6361/201425419)
- **Wisotzki et al., 2016:** *Extended Lyman α haloes around individual high-redshift galaxies revealed by MUSE*
L. Wisotzki, R. Bacon, J. Blaizot, J. Brinchmann, E. C. Herenz, J. Schaye, N. Bouché, S. Cantalupo, T. Contini, C. M. Carollo, J. Caruana, J.-B. Courbot, E. Emsellem, S. Kamann, **J. Kerutt**, F. Leclercq, S. J. Lilly, V. Patrício, C. Sandin, M. Steinmetz, L. A. Straka, T. Urrutia, A. Verhamme, P. M. Weilbacher, and M. Wendt
Astronomy & Astrophysics (2016), 587, A98, doi: [10.1051/0004-6361/201527384](https://doi.org/10.1051/0004-6361/201527384)
- **Herenz et al., 2017:** *The MUSE-Wide survey: A first catalogue of 831 emission line galaxies*
E. C. Herenz, T. Urrutia, L. Wisotzki, **J. Kerutt**, R. Saust, M. Werhahn, K. B. Schmidt, J. Caruana, C. Diener, R. Bacon, J. Brinchmann, J. Schaye, M. Maseda, and P. M. Weilbacher
Astronomy & Astrophysics (2017), 606, A12, doi: [10.1051/0004-6361/201731055](https://doi.org/10.1051/0004-6361/201731055)
- **Hashimoto et al., 2017:** *The MUSE Hubble Ultra Deep Field Survey. X. Ly α equivalent widths at $2.9 < z < 6.6$*
T. Hashimoto, T. Garel, B. Guiderdoni, A. B. Drake, R. Bacon, J. Blaizot, J. Richard, F. Leclercq, H. Inami, A. Verhamme, R. Bouwens, J. Brinchmann, S. Cantalupo, M. Carollo, J. Caruana, E. C. Herenz, **J. Kerutt**, R. A. Marino, P. Mitchell, and J. Schaye
Astronomy & Astrophysics (2017), 608, A10, doi: [10.1051/0004-6361/201731579](https://doi.org/10.1051/0004-6361/201731579)

- **Diener et al., 2017:** *The MUSE-Wide survey: detection of a clustering signal from Lyman α emitters in the range $3 < z < 6$*
C. Diener, L. Wisotzki, K. B. Schmidt, E. C. Herenz, T. Urrutia, T. Garel, **J. Kerutt**, R. L. Saust, R. Bacon, and S. Cantalupo
Monthly Notices of the RAS (2017), 471(3):3186–3192, doi: [10.1093/mnras/stx1677](https://doi.org/10.1093/mnras/stx1677)
- **Caruana et al., 2018:** *The MUSE-Wide survey: a measurement of the Ly α emitting fraction among $z > 3$ galaxies*
J. Caruana, L. Wisotzki, E. C. Herenz, **J. Kerutt**, T. Urrutia, K. B. Schmidt, R. Bouwens, J. Brinchmann, S. Cantalupo, M. Carollo, C. Diener, A. Drake, T. Garel, R. A. Marino, J. Richard, R. Saust, J. Schaye, and A. Verhamme
Monthly Notices of the RAS (2018), 473:30–37, doi: [10.1093/mnras/stx2307](https://doi.org/10.1093/mnras/stx2307)
- **Verhamme et al., 2018:** *Recovering the systemic redshift of galaxies from their Lyman alpha line profile.*
A. Verhamme, T. Garel, E. Ventou, T. Contini, N. Bouché, E. Herenz, J. Richard, R. Bacon, K. Schmidt, M. Maseda, R. Marino, J. Brinchmann, S. Cantalupo, J. Caruana, B. Clément, C. Diener, A. Drake, T. Hashimoto, H. Inami, **J. Kerutt**, W. Kollatschny, F. Leclercq, V. Patrício, J. Schaye, L. Wisotzki, and J. Zabl
Monthly Notices of the RAS (2018), 478:L60–L65, doi: [10.1093/mnrasl/sly058](https://doi.org/10.1093/mnrasl/sly058)
- **Wisotzki et al., 2018:** *Nearly all the sky is covered by Lyman- α emission around high-redshift galaxies*
L. Wisotzki, R. Bacon, J. Brinchmann, S. Cantalupo, P. Richter, J. Schaye, K. B. Schmidt, T. Urrutia, P. M. Weilbacher, M. Akhlaghi, N. Bouché, T. Contini, B. Guiderdoni, E. C. Herenz, H. Inami, **J. Kerutt**, F. Leclercq, R. A. Marino, M. Maseda, A. Monreal-Ibero, T. Nanayakkara, J. Richard, R. Saust, M. Steinmetz, and M. Wendt
Nature (2018), 562:229–232, doi: [10.1038/s41586-018-0564-6](https://doi.org/10.1038/s41586-018-0564-6)
- **Herenz et al., 2019:** *The MUSE-Wide Survey: A determination of the Lyman α emitter luminosity function at $3 < z < 6$*
E. C. Herenz, L. Wisotzki, R. Saust, **J. Kerutt**, T. Urrutia, C. Diener, K. B. Schmidt, R. A. Marino, G. de la Vieuville, L. Boogaard, J. Schaye, B. Guiderdoni, J. Richard, and R. Bacon.
Astronomy & Astrophysics (2019), 621:A107, doi: [10.1051/0004-6361/201834164](https://doi.org/10.1051/0004-6361/201834164)
- **Urrutia et al., 2019:** *The MUSE-Wide Survey: survey description and first data release*
T. Urrutia, L. Wisotzki, **J. Kerutt**, K. B. Schmidt, E. C. Herenz, J. Klar, R. Saust, M. Werhahn, C. Diener, and J. Caruana
Astronomy & Astrophysics (2019), 624:A141, doi: [10.1051/0004-6361/201834656](https://doi.org/10.1051/0004-6361/201834656)
- **Schmidt et al., 2019:** *Three Dimensional Optimal Spectral Extraction (TDOSE) from Integral Field Spectroscopy*
K. B. Schmidt, L. Wisotzki, T. Urrutia, **J. Kerutt**, D. Krajnovic, E. C. Herenz, R. Saust, T. Contini, B. Epinat, and H. Inami
Astronomy & Astrophysics (2019), 628:A91, doi: [10.1051/0004-6361/201935857](https://doi.org/10.1051/0004-6361/201935857)

B.2. Software and Packages

While working on this thesis I made use of many software packages, scripts and python packages created by people who rarely get credit for their valuable work. Without programmers solving problems for us so we do not have to reinvent the wheel, science would not have advanced at the current pace. Here is a non-exhaustive list of the most important tools that helped me in my work.

- This thesis was set with LaTeX, <https://www.latex-project.org/>.
- python 3.4.6 (mostly), <https://www.python.org/>
- astropy: Astropy Collaboration et al., 2013
- matplotlib: Hunter, 2007
- numpy and scipy
- LSDCat: Herenz & Wisotzki, 2017, available download at <http://muse-vlt.eu/science/tools> and via the Astrophysics Source Code Library at <http://ascl.net/1612.002>
- Galfit: Peng et al., 2002, Peng et al., 2010 at <https://users.obs.carnegiescience.edu/peng/work/galfit/galfit.html>
- pyqtgraph: <http://pyqtgraph.org/>, QtGui, QtCore
- QtClassify: Kerutt, 2017, Herenz et al., 2017 at <https://bitbucket.org/Leviosa/qtclassify> and publicly available via the Astrophysics Source Code Library at <http://ascl.net/1703.011>
- TDOSE: Schmidt et al., 2019

B.3. Abbreviations and Acronyms

ACS	Advanced Camera for Surveys
AGB	asymptotic giant branch
AGN	active galactic nucleus
BCG	brightest cluster galaxy
CCD	charge-coupled device
CDF-S	Chandra Deep Field-South
CGM	circumgalactic medium
CMB	cosmic microwave background
DEC	declination
DLA	damped Lyman α system
EoR	epoch of reionisation
ESA	European Space Agency
ESO	European Southern Observatory
EW	equivalent width
EW₀	rest-frame equivalent width
EWDF	equivalent width distribution function
FoV	field-of-view
FWHM	full width at half maximum
GP	Green Peas
GRB	gamma-ray burst
GTO	guaranteed time observation
GUI	graphical user interface
HDFS	Hubble Deep Field South
HDUV	Hubble Deep UV
HETDEX	Hobby-Eberly Telescope Dark Energy Experiment
HST	Hubble Space Telescope
HUDF	Hubble Ultra Deep Field
IFU	integral field unit

IGM	intergalactic medium
IMF	initial mass function
int	intrinsic
ISM	interstellar medium
LAE	Lyman α Emitter
LAH	Lyman α halo
LARS	Lyman α Reference Sample
LBA	Lyman break analogue
LBG	Lyman break galaxy
LF	luminosity function
LLS	Lyman limit system
LoS	line-of-sight
LSDCat	Line Source Detection and Cataloguing Tool
LSF	line spread function
LyC	Lyman continuum
MUSE	Multi-Unit Spectroscopic Explorer
NASA	National Aeronautics and Space Administration
PCA	Principal Component Analysis
PSF	point spread function
QSO	quasi-stellar object
RA	right ascension
SDSS	Sloan Digital Sky Survey
SED	spectral energy distribution
SFR	star formation rate
S/N	signal-to-noise
UDF	Ultra Deep Field
UV	ultra-violet
UVC	UV continuum
UVUDF	Ultra-Violet Hubble Ultra Deep Field
VLT	Very Large Telescope
WCS	world coordinate system
WFC3	Wide Field Camera 3
yrs	years
z	redshift
ZAMS	zero age main sequence

B.4. Selbstständigkeitserklärung

Hiermit erkläre ich, dass diese Arbeit an keiner anderen Hochschule eingereicht sowie selbständig von mir und nur mit den angegebenen Mitteln angefertigt wurde.

Potsdam, October 29, 2019

B.5. Acknowledgements

This work would not have been possible without the help of my friends, colleagues and family. First I want to thank my supervisor Lutz Wisotzki for giving me the opportunity to do a PhD and for guiding me through the process with honesty, enthusiasm, knowledge and trust. He once told me that maybe, if I work hard and everything goes well, there could be a chance for me to actually get my PhD. Coming from him, this really gave me hope! And what do you know, here we are. I also want to thank my second supervisor Philipp Richter and my external reviewer Dawn Erb as well as the rest of my committee.

Thank you to...

- Sebastian Raabe: Each time I wanted to quit I would ask "Can I please study something else now?" and you would always reply with "Of course, I'll stand by you no matter what".
- My parents Angelika Kerutt and Diedrich Bühler: It has been quite an extraordinary voyage you sent me on and I will always be grateful for your support and for lighting the flame.
- My friends: You stuck with me, listened to my whining with superhuman patience and believed in me when I couldn't. I am sorry and I promise I am better now, because of you.
- My colleagues: It is always nice to have company on a long and winding road, and I had both people going the way with me as well as people who have been there before.
- Fermat: All the times I waited, heart beating, for you to reboot after I messed up a loop and drowned you in processes. I was not always kind, but you never let me down and always came back to life.
- MUSE: You beautiful medusa, it was a pleasure to visit you, listen to your clicking and extract science from the ones and zeros you converted the light in.
- The park: I fled to you, I hid in you, I found peace and quiet and serenity and mushrooms and hornets and moths and Rapunzel's tower. And when I returned I was rested and ready to face my dragons.

I always tell people who ask how someone can study Physics that I would not have survived a single semester without the friends that I made along the way. We were in this together and we helped each other and I had the privilege to get to know some of the most amazing humans: Thank you so much to Lisa for your passion, to Antje for inspiration, to Nadin for kindness, to Levke for strength, to Mario for patience. Thank you to Mirna, Lucy, Fabian, Flo, Steffi, Markus and Saskia for being there for me through thick and thin. And thank you to Jan for reading this monster and finding all the missing half-spaces.

When I started the PhD journey, I already knew the AIP, I knew the group and my colleagues and I specifically chose to do my PhD here because I knew that if you are surrounded by people you like, it will make things much easier. I want to thank Tanya for reducing the data and being the driving force behind MUSE-Wide, Mirko for the pep-talks on (im)possible time schedules, Christian for a special cat and inspiration for QtClassify, Mark for patience and selective German comprehension, Kasper for an open ear, for looking at blue bumps and the introduction, Gab for discussions on other worlds, Hakan for visiting other worlds, Mohamad for accepting I am from another world, Karl-Heinz for saving fermat and being supportive and Davor for taking my mind off science during lunch. I never felt alone, neither with my struggles nor with my joys, as my fellow PhD students kept me company: Thanks to Rikke, Sabine, Justus, Khadiga, Sanja, Yohana, Jenn, Kris and all the others for crying and laughing together. And Anika, you joined the group with an inspiring determination and helped me through many fits of self doubt. Thanks for being an inspiration! Thank you also to the rest of the Galaxies and Quasars group and all my colleagues at the AIP and to the Python Club. Science is not only about gaining knowledge, but also about sharing and discussing and I was fortunate to be part of the amazing MUSE consortium. The Busy Week meetings were some of the highlights of my PhD and I want to thank Roland for creating MUSE and making this work possible. Many thanks also to Anne, who invited me to work with her and showed me how exciting Lyman continuum emission can be.

Some last thanks go to the more abstract help that I had during this exciting, but trying time: Thank you to Susi and Clair for the moonshine, thank you to Sally for the dream, to Gardens for the fire, to Wolfgang Schäfer for showing me the relevance of Physics, to Seven of Nine (my laptop) for staying alive and to Seven of Nine herself. Yes, the Star Trek character. Often when people are asked how they got into Astronomy they will start telling you about the first time they looked into a telescope, when they were five years old, in the backyard of their grandparents' house and saw Saturn. Well I have a different story, I owe my passion for Astronomy partly to an imagined experience, but that makes it no less real. And yet, I still want to thank my grandparents, especially Franz Kerutt. A card you had written for me years ago fell into my lap this week. It said 'Good luck for the future.' I had and you would be proud, you always were. And finally, a special thanks goes to everyone I forgot to mention and especially to you, for reading this. LLAP!

"Survival is insufficient."
– Seven of Nine