UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE ASTRONOMÍA

A COMPREHENSIVE STUDY OF LYMAN ALPHA EMISSION IN THE $3 < z < 4.6$ GALAXY POPULATION

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN CIENCIAS, MENCIÓN ASTRONOMÍA

GRECCO ÁLVARO OYARZÚN MARTÍNEZ

PROFESOR GUÍA:
GUILLERMO BLANC MENDIBERRI

MIEMBROS DE LA COMISIÓN:
VALENTINO GONZÁLEZ CORVALÁN
RICARDO MUÑOZ VIDAL
NELSON PADILLA

Este trabajo ha sido parcialmente financiado por CONICYT, Beca Magíster Nacional 2014, Folio 22140924

SANTIAGO DE CHILE
2016

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE MAGÍSTER EN CIENCIAS, MENCIÓN ASTRONOMÍA
POR: GRECCO ÁLVARO OYARZÚN MARTÍNEZ
FECHA: 2016
PROF. GUÍA: SR. GUILLERMO BLANC MENDIBERRI

A COMPREHENSIVE STUDY OF LYMAN ALPHA EMISSION IN THE $3 < z < 4.6$ GALAXY POPULATION

La observación de emisión Lyman alpha (Lyα) proveniente de galaxias es una técnica ampliamente utilizada para el estudio del universo a alto redshift (e.g. Shapley et al. 2003, Ouchi et al. 2008, Stark et al. 2010, Blanc et al. 2011). Sin embargo, a pesar de su importancia, los estudios estadísticos de esta emisión no abordan posibles sesgos al elaborar las muestras de objetos, induciendo discrepancias. Diferentes metodologías inducen selecciones en masa estelar (M_*), tasa de formación estelar (SFR) y metalicidad. Los perjuicios asociados a la forma actual de estudiar la emisión Lyα no solamente se restringen a sesgos en los resultados obtenidos, sino que también limitan nuestro entendimiento de este proceso radiativo en la población de galaxias. En consecuencia, nuestra comprensión de la emisión Lyα en dicha población es actualmente altamente especulativo. Todas estas razones justifican la necesidad de llevar a cabo un estudio de emisión Lyα en la población de galaxias, lo cual es el objetivo de esta tesis.

El presente trabajo detalla los resultados de un exhaustivo análisis de emisión Lyα a alto redshift. Para ello, primero diseñamos una muestra de 629 galaxias observadas por CANDELS (Koekemoer et al. 2011, Grogin et al. 2011). Dicha muestra se encuentra especialmente diseñada para abarcar el rango de masa estelar $7.6 < \log(M_*[M_\odot]) < 10.6$ y rango de redshift $3 < z < 4.6$. Tras esta selección, realizamos espectroscopía de los objetos con el Michigan/-Magellan Fiber System (M2FS; Mateo et al. 2012). De esta forma, contamos con mediciones de flujo de Lyα, además de la fotometría CANDELS, la cual permite obtener masa estelar, luminosidad, tasa de formación estelar, y extinción para cada galaxia. Observamos que el Ancho Equivalente y Fracción de escape de Lyα decrecen con la masa estelar, luminosidad en el UV, tasa de formación estelar, y extinción de las galaxias. Además, introducimos aplicaciones de estadística Bayesiana en este campo de la astrofísica. Esto nos permite concluir que un modelo de distribución exponencial es el más apropiado para reproducir nuestras distribuciones de Ancho Equivalente de Lyα, además de caracterizar la completitud y significancia de las correlaciones que observamos. También estudiamos los impactos que tienen diferentes técnicas de selección de galaxias en las estadísticas observadas de Lyα. Lyman Break Galaxies y estudios de Lyα emitters seleccionan preferentemente galaxias de baja masa estelar y poco polvo interestelar. Finalmente, realizamos la primera predicción semi-analítica de la fracción de Lyα emitters hasta redshift 7, la cual se puede utilizar para restringir observacionalmente la época de reionización. Todos estos resultados no solamente contribuyen a comprender de mejor manera este tipo de radiación, sino que también establecen un nuevo marco para el análisis estadístico de este trazador del universo temprano.

A mi abuelo René

Agradecimientos

Este trabajo no hubiese sido posible sin el incondicional apoyo de mi familia. Elementos como su formación moral y educacional han permitido que mis metas tengan como único límite el esfuerzo y la perseverancia. Es también importante destacar el ejemplo que he recibido de cada profesor con el que he trabajado. Muchas gracias a los profesores Leonardo Bronfman, Ricardo Muñoz, Julio Chanamé, Guillermo Blanc, y Valentino González. Mis agradecimientos por el conocimiento que me han entregado, así como también por lo respetuoso y ameno de su trato. Especiales agradecimientos a Guillermo Blanc, mi profesor guía, por darme la oportunidad de liderar un proyecto envidiable. Gracias por permanentemente motivarme a ser un investigador más capaz e independiente cada día. Quisiera también agradecer a los funcionarios del departamento. Su amable presencia permitió que mi estadía fuese cómoda y agradable. Por último, pero no por ello menos importante, quisiera agradecer a mis compañeros por entregarme su ayuda cuando fuese necesario. Su presencia hizo de mi estadía en el departamento una experiencia mucho más entretenida y completa. Muchas gracias.

Table of Contents

List of Tables

List of Figures

Introduction

In hand with observational progress, the understanding of high redshift galaxies has progressed immensely in the last few decades. Almost twenty years ago, the only efficient way of observing these galaxies was by detecting breaks in their spectra with broadband photometry. For example, the two main classes of rest UV selected galaxies are Lyman Break Galaxies (LBGs) and Lyman-alpha Emitters (LAEs). While the former are observed by their Lyman Break at 912 Å, the latter are detected by their Lyα emission at 1216 Å, which is produced in star forming regions and is subject to resonant scattering in the neutral hydrogen of the ISM. After being, at first, used as a galaxy detection method, Lyα emission from galaxies is now widely used to derive a wide variety of information about the physical properties of galaxies and the universe as a whole.

A lot of effort has been dedicated to study the process of Lyα emission within galaxies. Radiative transfer analysis focuses on the dependence of Lyα escape on ISM kinematics (Rivera-Thorsen et al. 2015), clumpiness (Duval et al. 2014), and outflows (Dijkstra & Kramer 2012, Gronke & Dijkstra 2016). These studies yield elementary insights that, when combined with observations of the Lyα escape fraction (f_{esc}; Hayes et al. 2011, Ciardullo et al. 2014), provide a coherent picture for Lyα emission. Still, despite the complex modeling required to explain this particular type of radiation at small scales, Lyα is actively used as a technique for solving cosmological scale paradigms. Among other things, Lyα emission is being used to study the properties of neutral hydrogen in the interstellar, circumgalactic and intergalactic mediums. For instance, the fraction of LAEs as a function of redshift is used as a proxy for the fraction of neutral gas in the IGM (Stark et al. 2011, Ono et al. 2012, Tilvi et al. 2014). If successful, this insight can be used to trace the epoch of re-ionization, which we have been able to constrain using QSO sightlines (Fan et al. 2006) and CMB measurements (Bennett et al. 2013, Planck Collaboration et al. 2015). Even more, with the hope of tracing the sources responsible for re-ionization, Lyα emission is also conveniently used for characterizing luminosity functions at high redshift (Ouchi et al. 2008, Dressler et al. 2015). Nevertheless, the cosmological questions that we are trying to answer by means of this complex emission are not restricted to the re-ionization of the universe. For instance, Lyα emission will also be used for studying dark energy through baryonic acoustic oscillations (BAOs) in the clustering of LAEs (Adams et al. 2011, Blanc et al. 2011).

As evidenced by small scale studies, Lyα emission is highly stochastic and complex. The observed equivalent width ($W_{Ly\alpha}$) of Lyα lines is highly dependent on neutral gas opacity and kinematics, clumpiness of the ISM, dust distribution, and, therefore, line-of-sight towards the observer. As a consequence, if we are to use Lyα emission as a tracer, a careful and thorough

approach is required. However, limitations in survey design difficult such approach. For instance, spectroscopic studies of UV continuum detected galaxies, typically LBGs (e.g. Shapley et el. 2003, Stark et al. 2010, Ono et al. 2012), reject galaxies with no significant Lyman Break and/or red rest-frame UV spectral energy distributions (SEDs), i.e., these studies do not account for passive and heavily-extincted galaxies. Observationally, this technique is limited by the M_{UV} sensitivity of the survey, which also translates into low SFR incompleteness. On the other hand, samples where galaxies are detected directly in Lyα using narrowband imaging (e.g. Gronwall et al. 2007, Ouchi et al. 2008) or blind spectroscopy (e.g. Blanc et al. 2011, Dressler et al. 2011) select on emission line flux and $W_{Ly\alpha}$. Such methodology does not require continuum selection, which allows for line detection in faint objects. However, follow-up rarely allows for more than spectroscopic confirmation of the line, limiting the use of this technique to the measurement of line fluxes, $W_{Ly\alpha}$, and emission profiles.

Comprehensive studies of Lyα emission in the overall galaxy population are, therefore, of significant importance. Correlations between galaxy properties and emission are the statistical manifestation of the radiative escape of $Ly\alpha$ photons. Properties such as stellar mass, star formation rate (SFR), UV slope, and merger rate reveal information into how the escape of Lyα radiation is affected by stellar population ages, gas fraction, dust content and ISM turbulence, respectively. Acknowledgment of such relations is required for drawing conclusions from high-redshift Lyα surveys. Oyarzún et al. (2016) show how the normalization and e-folding scale of the EW distribution of $3 < z < 4.6$ galaxies anti-correlate with M_*. In other words, higher mass objects typically have lower $W_{Ly\alpha}$. Higher mass galaxies typically have a lower gas fraction, but higher gas mass. More neutral gas contributes to increase the scatter of Lyα photons, spreading this radiation throughout the galaxy and toward the circumgalactic medium. Higher-mass galaxies also have more dust extinction, which leads to more Lyα photons absorption. The escape fraction of this radiation is, therefore, severely affected by the stellar mass of galaxies. These results even explain some discrepancies between LBG and narrowband studies of Lyα emission originated in sample selection effects. This example highlights the need for proper assessment of samples studied.

In this work, we present a comprehensive analysis of Lyα emission in the $3 < z < 4.6$ galaxy population. To this end, we further analyze the data introduced by Oyarzún et al. (2016). This sample is composed of 629 galaxies in the M_* range $7.6 < \log(M_*[M_\odot]) < 10.6$ from the 3D-HST/CANDELS survey (Koekemoer et al. 2011, Grogin et al. 2011). Our work is based on spectroscopic observations of the sample using the Michigan/Magellan Fiber System (M2FS; Mateo et al. 2012), allowing us to measure Lyα fluxes and use 3D-HST/CANDELS ancillary data. In particular, we study Lyα emission dependence on the M_*, SFR, UV luminosity and dust extinction of galaxies. To do so, we introduce a Bayesian approach to properly compare $W_{Ly\alpha}$ distribution models, account for incompleteness, and quantify the significance of observed correlations. We also use our results to simulate high-redshift Lyα emission surveys, allowing us to infer on their biases and imprints on the resulting samples. We finally use the correlations we recover to predict the fraction of LAEs as a function of redshift. This semi-analytic approach provides a baseline for comparing observational drops in the fraction of LAEs at high redshift. This work is structured as follows. In Section 1, we describe our sample and data set. In Section 2, we describe our line detection and measurement methodologies. We explain our Bayesian $W_{Ly\alpha}$ distribution analysis in Section 3. We show our results on the $W_{Ly\alpha}$ distribution dependence on different properties and selection techniques

in Sections 4 and 5. We state in Section 6 our inferences on higher redshift Lyα emission. We present the summary and discussion in the Conclusion. Throughout this work, all magnitudes are in the AB system (Oke & Gunn 1983). We also adopt a ΛCDM cosmology with $H_0 = 70$ km s^{-1} Mpc^{-1}, $\Omega_m = 0.3$, and $\Omega_\lambda = 0.7$.

Body Text

1. Dataset

1.1. Sample Selection

Our sample is composed of 629 galaxies in the COSMOS, GOODS-S and UDS fields. Every object is observed under the 3D-HST/CANDELS program, providing HST and Spitzer photometry from 3800 Å to 7.9 μm (44 bands for COSMOS, 40 for GOODS-S, and 18 for UDS). We construct our sample using 3D-HST outputs (Skelton et al. 2014). According to these, our 629 photometric redshifts satisfy $3.25 < z_{3D-HST} < 4.25$ and have a 95 % probability of $2.9 < z < 4.25$. Every galaxy also complies with a photometric redshift reliability parameter $Q_z \leqslant 3$ selection to remove catastrophic outliers (Brammer et al. 2008). In terms of M_*, our galaxies are homogeneously distributed in the range $8 < \log(M_*[M_\odot])_{3D-HST} < 10.4$. These values are obtained assuming exponentially declining star formation histories (SFHs) with a minimum e-folding time of $log_{10}(\tau/yr) = 7$ (Skelton et al. 2014).

1.2. Data

Spectroscopy of the complete sample was conducted at the Magellan Clay 6.5 m telescope during 2014 December and 2015 February. To this end, we used the M2FS, a multi-object fiber-fed spectrograph. This instrument's 1.″2 fibers allow for observation of 256 targets within 30 arcmin in a single exposure. We are then able to observe each field's 210 targets simultaneously, while using 40 fibers for sky apertures and five for calibration stars. We used this spectrograph in LoRes mode, which features an expected resolution of R=2000 and a continuum sensitivity of V=24 with S/N=5 in 2 hours (Mateo et al. 2012). The final data set consists on six exposure hours on each of the three fields with an average seeing of 0.″6.

For data reduction, we develop a custom M2FS pipeline. This routine features standard bias subtraction and dark correction. For wavelength calibration, we use HgArNe lamps observed on each night. The wavelength solution is obtained separately for each fiber, with a typical rms uncertainty of 0.03Å. We further correct each fiber's solution using the sky-lines in the science spectra. For flat-fielding, we use sky-flats obtained during either twilight or dusk. The correction is calculated separately for each fiber, and features illumination, fiber profile, and a correction to account for shifting of fiber spectra on the detector due to thermal

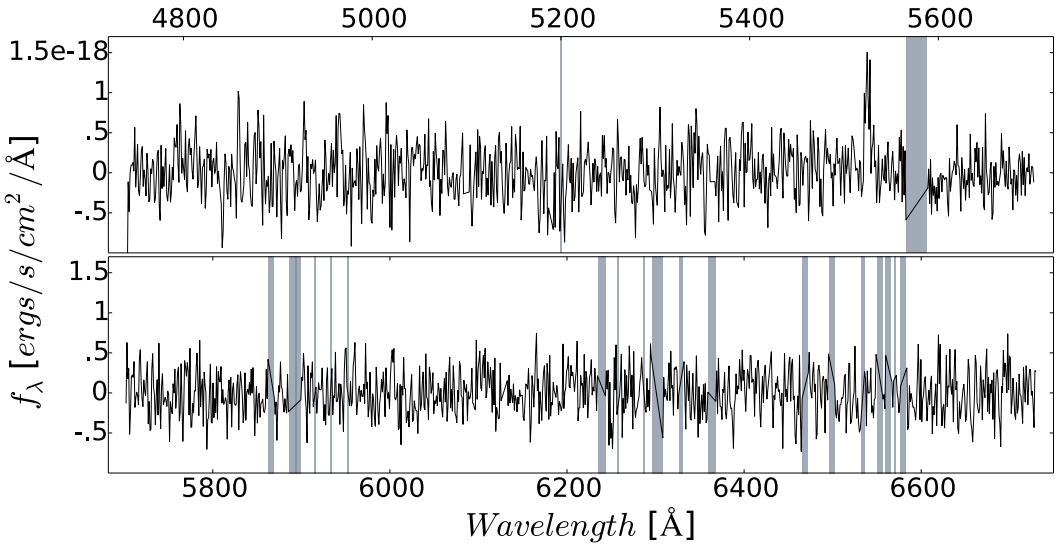

Figure 1 Reduced and continuum subtracted spectrum for target GS39574. Shadows show the sky residuals we mask after skysubtraction is performed. We observe a Lyα line at 5525 Å for this galaxy and associate a S/N$\sim$ 6 to the detection. We measure a Lyα flux of 1.1×10^{-17} erg s^{-1} cm^{-2}, $z_{Ly\alpha} \sim 3.36$, and $W_{Ly\alpha}\sim$ 70 Å for this line. For reference, we estimate log(M$_*$[M$_\odot$]) $\sim$8.2 and log(SFR[M$_\odot/yr$]) $\sim$0.7 for this object.

effects. Despite the fact that we are not using dome-flats, we estimate the uncertainties in our flat-fielding to be $< 5\,\%$.

Sky subtraction is made using the 40 sky fibers homogeneously distributed over the field of view and across the detector, where fibers are grouped in blocks. We find the sky solution to be more dependent on fiber location on the CCD than the sky, leading us to perform sky subtraction separately for each frame fiber block. The solution is computed using a non-parametric spline fit, yielding satisfactory sky subtraction for most sky lines. For the bright sky lines that we are not able to properly subtract, we build emission line masks. For consistency, we use the same mask for all spectra, except for particularly noisy fibers. In those rare cases (11 of 629), we mask broad sections of the spectrum. Due to fiber malfunction, we could not obtain spectra for three of the 629 targets.

Flux calibration is performed using five $M_V = 19 - 22$ calibration stars on each exposure. Due to atmospheric differential refraction (ADR) comparable to the $1\farcs2$ fiber size, we need to recover the intrinsic spectrum of each star. To do so, we first correct each field for Galactic extinction and, then, fit stellar templates from the Pickles library (Pickles 1998) to the continuum normalized instrumental spectra of the calibration stars. We then use the five stars on each exposure to obtain an average sensitivity curve. We estimate the rms uncertainty of our method to be about $\sim 15\,\%$. After this calibration, we correct the fluxes in our spectra for Galactic extinction. A sample sky subtracted spectrum is shown in Figure 1, including the lines we mask.

Considering the three fields, the science area we survey is of ~ 550 arcmin^2. The resulting spectral FWHM line resolution is of ~ 2Å, and we reach a 1σ continuum flux density limit

of $\sim 4 \times 10^{-19}$ erg s^{-1} cm^{-2} Å^{-1} per pixel in our 6 hr of exposure. We estimate a 5σ emission line flux sensitivity of $\sim 8 \times 10^{-18}$ erg s^{-1} cm^{-2} in our final spectra.

1.3. Sample Properties

As stated by Skelton et al. (2014), 3DHST outputs for our galaxies were obtained using FAST (Kriek et al. 2009). These calculations assume exponentially declining star formation histories (SFHs) with a minimum e-folding time of $log_{10}(\tau/yr) = 7$ (Skelton et al. 2014). However, since recent studies suggest that constant SFHs (cSFHs) are more adequate to reproduce high-redshift obervations (González et al. 2014), we perform our own executions of FAST assuming cSFHs. As FAST requires redshifts as input, we first run EAZY (Brammer et al. 2008). The executions of EAZY on CANDELS/IRAC photometry yield a most probable redshift z_{EAZY}. Our 629 objects satisfy $3 < z_{EAZY} < 4.25$ (Figure 2), with a median uncertainty $\sigma_{EAZY} = 0.1$. EAZY outputs also include 2σ constraints, which for our sample are limited to $2.95 < z < 4.5$. From now on, we use our spectroscopic redshifts ($z_{Ly\alpha}$) for detections and z_{EAZY} for non-detections. Our FAST executions yield a mass coverage of $7.6 < \log(\mathrm{M}_*[\mathrm{M}_\odot]) < 10.6$ (Figure 2), with a characteristic uncertainty of $\log(\mathrm{M}_*[\mathrm{M}_\odot]) \sim 0.2$. Our FAST outputs also include SFRs, for which we obtain most values in the range $1 - 100\mathrm{M}_\odot/yr$ (further analysis and plots therein).

1.4. Considerations

The flux calibration procedure performed in our spectra is based on using stars, so therefore corrects for a $\sim 32\,\%$ fiber flux loss, which corresponds to a point-source Lyα surface brightness distribution. In cases of extended Lyα emission halos (Matsuda et al. 2012, Feldmeier et al. 2013, Hayes et al. 2013, Momose et al. 2014, Caminha et al. 2015, Wisotzki et al. 2015, Patrício et al. 2016), the fluxes we derive are mostly associated to galaxies themselves and the inner parts of their Lyα halos. At the cosmic time of our sample, fiber diameter corresponds to a scale of $\sim$ 8 kpc, roughly a factor of 4 larger than the typical effective diameter of galaxies (Bond et al. 2012, Law et al. 2012).

We find a median redshift offset of $\Delta z = z_{Ly\alpha} - z_{EAZY} = 0.24$ for our detections. This offset correlates with $W_{Ly\alpha}$, hinting at EAZY template fitting biases when characterizing the Lyman Break. We show in Oyarzún et al. (2016) that this offset does not have any noticeable effects on our sample dependence on M_*, which is the primary selection criterion for our galaxies. We will present a detailed study of the causes behind this offset in a future study.

We stress 3D-HST/CANDELS mass incompleteness is restricted to $\log(\mathrm{M}_*[\mathrm{M}_\odot]) < 8.5$ at $z \sim 4$ (Duncan et al. 2014), which corresponds to about one quarter of the sample. Our FAST measured SFRs are obtained using UV photometry and slopes, which has associated timescales of 100 Myr (Kennicutt 1998). The uncertainties in this methodology are well documented (Kennicutt & Evans 2012, Boquien et al. 2014), and must be considered when comparing to Hα SFRs (timescales of 10 Myr; Boquien et al. 2014). Finally, we want to stress that this is a homogeneously M_* selected sample, designed to study Lyα emission statistics

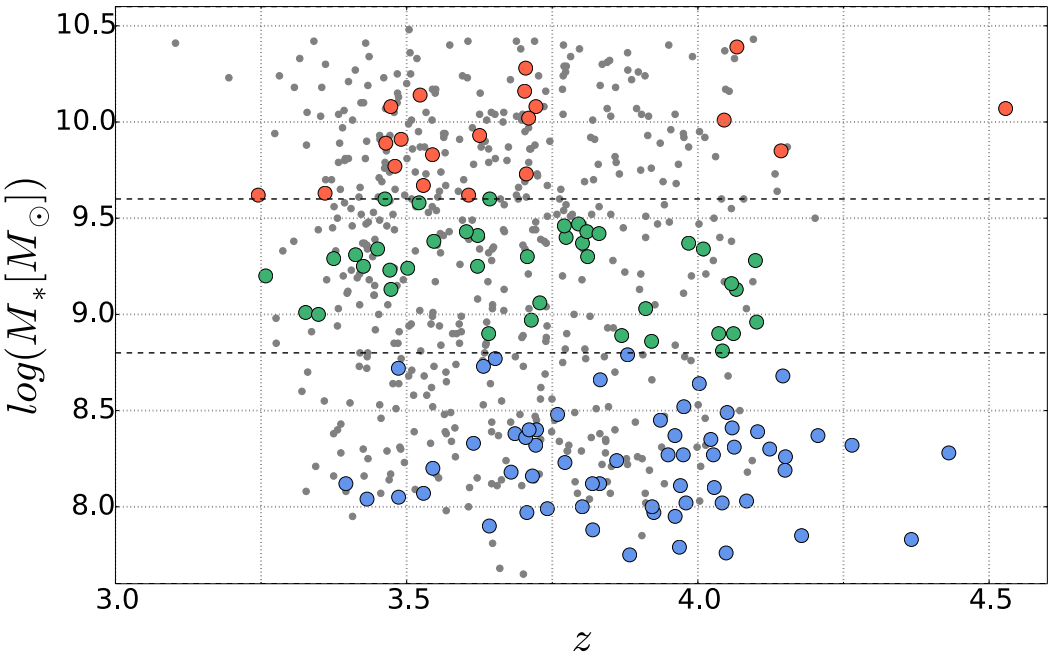

Figure 2 Distribution of our sample in the redshift-M_* plane. The grey points show the location of the 629 objects using photometric redshifts (z_{EAZY}, $M_*(z_{EAZY})$). In dashed lines are plotted the boundaries -log($M_*[M_\odot]$)=8.8, 9.6- used to define the low, medium, and high mass subsamples. We plot the detections ($z_{Ly\alpha}$, $M_*(z_{Ly\alpha})$) in the respective bins as blue, green, and red datapoints. We have a total of 120 detections down to S/N=5.5. Note that detections are plotted twice (using z_{EAZY} and $z_{Ly\alpha}$).

dependence on galaxy properties. As a consequence, it is by no means representative of the M_* or Lyα luminosity functions at $3 < z < 4.6$. This must be taken into account when comparing this sample to analogues directly drawn from the galaxy population (i.e. LBGs or narrowband samples).

2. Lyα Measurements

2.1. Line Detection

For line detection, we use an automated maximum likelihood fitting routine after continuum subtraction. We assume intrinsic gaussian profiles of the form:

$$f_{rest}(\lambda) = \frac{F}{\sqrt{2\pi\sigma_\lambda^2}} \exp\left[-\frac{(\lambda-\lambda_0)^2}{2\sigma_\lambda^2}\right] \quad (1)$$

where F, λ_0, and σ_λ compose the parameter space explored by the maximum likelihood.

In order to account for false positives, we run the line detection routine on the 115 sky fibers. The results are shown in Figure 3. We detect four lines above 4σ and none above 5σ. Therefore, down to 5σ, we are confident of having fewer than 5 false detections in our

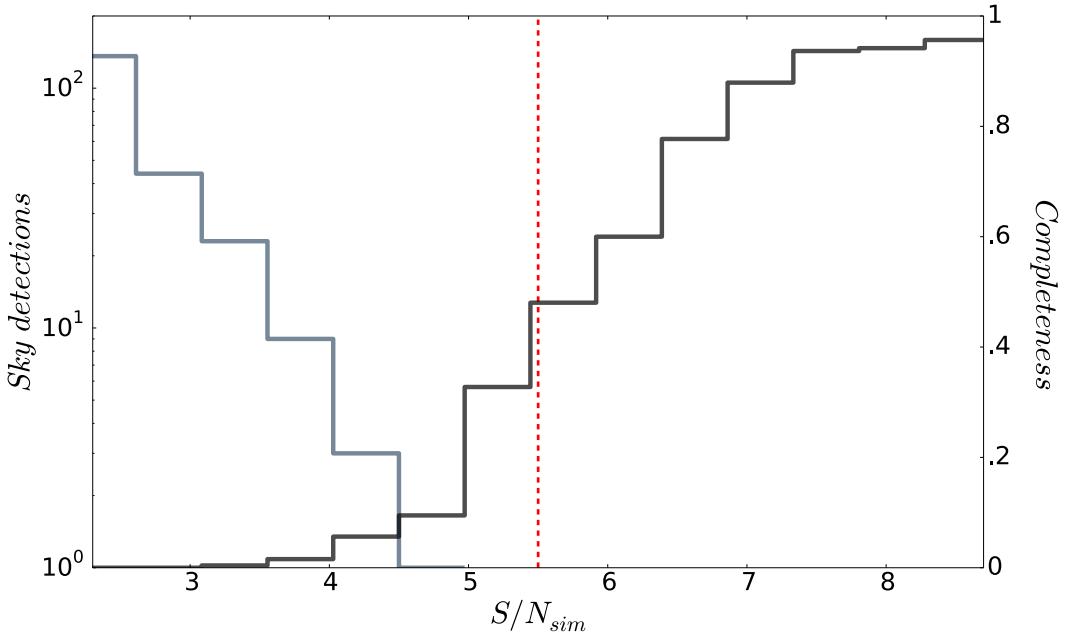

Figure 3 Completeness (black) and false positives (grey) yielded by our line detection method. The completeness histogram is recovered simulating $\sim 10^3$ gaussian emission lines on the 115 reduced skyspectra. Hence, for this curve, the x-axis corresponds to S/N_{sim}. The false positives histogram is obtained running our line detection code on the 115 skyspectra. The dashed red line shows our line detection threshold S/N=5.5.

629 targets. This translates into $\lesssim 5\,\%$ contamination using signal-to-noise $S/N^* = 5.5$ as our threshold, considering we have 120 detections with $S/N \geqslant 5.5$. We also characterize our detection completeness (see Figure 3). To obtain it, we use $p(S/N_i > S/N^* | S/N)$, with S/N_i the measured signal-to-noise and S/N^* the imposed detection threshold. We define the simulated signal-to-noise as $S/N = \sqrt{\sum d\lambda^2 \sigma_k^2}$, with $d\lambda$ the wavelength dispersion in the spectrum and σ_k the flux uncertainty for pixel k. We find the most accurate representation summing over an interval of 20Å centered at 5500 Å. To recover the actual completeness shown in Figure 3, we simulate $\sim 10^3$ lines on the 115 sky-spectra sampling fluxes of $10^{-19} - 10^{-17}$ ergs s^{-1} cm^{-2}, FWHMs between $5 - 13$ Å, and wavelengths of 4800-6700 Å. The fraction for which we measure $S/N_i > S/N^*$ is our completeness. Now that we use $S/N^* = 5.5$ as our detection threshold, our imposed line flux sensitivity in the final spectra is, on average, $\sim 1 \times 10^{-17}$ erg s^{-1} cm^{-2}. We do not measure any object to have more than one emission line with $S/N_i > S/N^*$, dismissing the existence of any evident AGN in our spectra.

2.2. Line Profiles

The radiative transfer and escape of Lyα radiation from galaxies can be highly complicated. As a matter of fact, the resonant nature of this line has led to thorough modeling of its radiative escape (e.g. Verhamme et al. 2006, Dijkstra & Kramer 2012). Such complications imply that the flux profile of a Lyα line is not always well reproduced by the usual gaussian profile. Hence, for flux measurements, we adopt a more sophisticated model. Similarly to

McLinden et al. (2011) and Chonis et al. (2013), we fit double-peaked gaussian profiles of the form

$$f_{rest}(\lambda) = f_{blue}(\lambda) + f_{\rm red}(\lambda) \tag{2}$$

where $f_{blue}(\lambda)$ represents the blue emission component and $f_{\rm red}(\lambda)$ the blue component. We assume each of these components to be asymmetric, i.e., they follow Equation (1), with σ defined as:

$$\begin{aligned} \sigma_\lambda &\equiv \sigma_b \text{ if } \lambda < \lambda_0 \\ \sigma_\lambda &\equiv \sigma_r \text{ if } \lambda > \lambda_0 \end{aligned} \tag{3}$$

Before fitting, we convolve the profiles given by (2) and (3) with the spectral resolution. This allows us to properly characterize the errors in our measurements and methodology. In case that there is some sky contamination (which happens for 28 of the 120 detections), we only fit single-peaked profiles.

2.3. Lyα Equivalent Width and Escape Fraction

There are typically two diagnostics used to characterize the prominence of Lyα emission in galaxies: the equivalent width ($W_{Ly\alpha}$) and the Lyα escape fraction (f_{esc}). The rest $W_{Ly\alpha}$ is defined as the fraction between line flux and UV continuum flux in the rest frame of the galaxy. Explicitly:

$$W_{Ly\alpha} = \frac{F}{f_\lambda} \frac{1}{(1 + z_{Ly\alpha})} \tag{4}$$

with F the Lyα flux we measure in the spectra and f_λ the observed flux at rest 1700Å from 3DHST rest-frame colors (Skelton et al. 2014).

On the other hand, f_{esc} is the fraction between the number of Lyα photons that escape the galaxy and the number produced. This diagnostic is typically indirectly recovered using SFRs derived from the Lyα line and intrinsic SFRs. The latter is typically calculated from UV continuum measurements or Hα fluxes, subject to extinction correction. In our case, we use the Lyα fluxes we measure and the SFRs from FAST (which are intrinsic, i.e., extinction corrected) to calculate f_{esc} for our objects. We use the definition from Blanc et al. (2011):

$$\begin{aligned} f_{esc}(Ly\alpha) = \frac{L(Ly\alpha)_{obs}}{L(Ly\alpha)_{intrinsic}} = \\ \frac{\mathrm{SFR}(Ly\alpha)}{\mathrm{SFR(UV)}_{corr}} = \frac{5.4 \times 10^{-42} L_{Ly\alpha}/8.7}{\mathrm{SFR(UV)}_{corr}} \end{aligned} \tag{5}$$

with $L_{Ly\alpha}$ the luminosity associated to the Lyα line and $\mathrm{SFR(UV)}_{corr}$ the extinction corrected SFRs. This assumes a Lyα to Hα ratio of 8.7 and then uses the Hα SFR calibration from Kennicutt (1998) for a Chabrier (2003) IMF (Kennicutt & Evans 2012). In case of non-detections, we use the corresponding $\mathrm{S/N}^* = 5.5$ line fluxes to derive upper limits on $W_{Ly\alpha}$ and f_{esc} for every object in our sample. It must be noted that our values for SFR(Lyα) and $\mathrm{SFR(UV)}_{corr}$ are tied to the assumption of a Chabrier (2003) IMF. Our calculations do not

account for UV excess associated to binaries, which induce uncertainties on f_{esc} that can be very difficult to account for (Stanway et al. 2016).

We use our f_{esc} measurements to look for possible AGNs. We perform cross-matching with NASA/IPAC Extragalactic Database (NED[1]) for all objects with $f_{esc} > 1$ and find none of them to be a reported X-ray source.

3. The Lyα Equivalent Width Distribution

3.1. Bayesian Inference

Measurement of $W_{Ly\alpha}$ for a galaxy sample yields the $W_{Ly\alpha}$ distribution. Since $W_{Ly\alpha}$ is directly measured from the data, characterization of the $W_{Ly\alpha}$ distribution can be naively considered straightforward. However, careful consideration of uncertainties and completeness can yield important insights on underlying information. Therefore, for proper characterization of uncertainties, significance, and trends in our results, we use Bayesian statistics. In this section, we explain how to recover the $W_{Ly\alpha}$ distribution within this framework, complementary to the one introduced in Treu et al. (2012).

Different probability distribution models can be adopted for reproducing $W_{Ly\alpha}$ distribution measurements. For instance, studies use gaussian (e.g. Guaita et al. 2010), exponential (e.g. Zheng et al. 2014), and lognormal distributions. For our analysis, we define the probability distribution as $p(W_{Ly\alpha}|W_n)$. Let W_n be the parameter space associated to the model. From now on, we describe the Bayesian approach to recover the posterior distribution of W_n. By means of this approach, we include the uncertainties in sample size, flux measurements, and photometry in the estimation of the posterior. The description we provide is not limited to this particular work, allowing for further application in similar datasets. We only present here the fundamental equations, as this procedure is already described in detail in Oyarzún et al. (2016).

Our Bayesian analysis is based on Lyα line flux F instead of $W_{Ly\alpha}$, as introduced in equation (4). This approach simplifies the equations, since we can assume F and f_λ to be normally distributed, which cannot be done for $W_{Ly\alpha}$. According to Bayes' Theorem, the posterior distribution $p(W_n|\{F\})$, i.e., the parameter space probability distribution given our data set $\{F\}$, is:

$$p(W_n|\{F\}) = \frac{p(\{F\}|W_n)p(W_n)}{p(\{F\})} \tag{6}$$

The likelihood is just the product of the individual likelihood for every galaxy, i.e., $p(\{F\}|W_n) = \prod p(F_{\rm i}|W_n)$. For a detection, it is given by:

$$p(F_{\rm i}|W_n) = \int_0^\infty p(F_{\rm i}|F)p(F|W_n){\rm d}F \tag{7}$$

[1]The NASA/IPAC Extragalactic Database (NED) is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

Table 1. Lyα Equivalent Width distribution model comparison

Likelihood Output	Low mass			Medium mass			High mass			Complete Sample		
	Exp	Gaussian	Logn	Exp	Gaussian	Logn	Exp	Gaussian	Logn	Exp	Gaussian	Logn
Model odds[a]	0.93	0.06	0.01	0.97	0.01	0.02	1	10^{-3}	10^{-5}	0.98	10^{-6}	0.02
Peak odds[b]	0.25	0.02	0.73	0.23	10^{-3}	0.77	0.97	10^{-3}	0.03	0.19	10^{-7}	0.81
A[c]	0.75	0.55	0.55	0.55	0.4	0.5	0.35	0.2	0.2	0.4	0.3	0.45
W_0[c]	46	74	0.7	26	46	0.85	14	26	0.75	38	64	1.05
μ[c]	-	-	3.85	-	-	3.05	-	-	3	-	-	3.05

[a]Obtained integrating the likelihood over the whole parameter space.

[b]Calculated with the maximum of the likelihood.

[c]Model parameters for corresponding likelihood maximum. Note that these values are different from the ones in Oyarzún et al. (2016), since here we are just working with the likelihood.

where $p(F_{\rm i}|F)$ is the line flux probability distribution for the corresponding galaxy valued at a flux F, which we consider to distribute normally. On the other hand, the term $p(F|W_n)$ is just the probability distribution for F given W_n. Using the definition of $W_{Ly\alpha}$ in Equation (4), the term translates to the product distribution $p(F|W_n) = p(W_{Ly\alpha}|W_n)p(f_\lambda)$. At this point, we include the probability distribution for the continuum, which we also assume to be gaussian.

The limiting line flux $F_{\rm i}^*$ for discerning detections from noise is given by our S/N threshold, ie., $F_{\rm i}^* = {\rm S/N}^*\sigma_{\rm i}$. For galaxies with no detections above $F_{\rm i}^*$, we adopt the following value for the likelihood:

$$p(F_{\rm i} < F_{\rm i}^*|W_n) = \int_0^\infty \left(1 - p(F_{\rm i} > F_{\rm i}^*|F)\right) p(F|W_n){\rm d}F \tag{8}$$

with $p(F_{\rm i} > F_{\rm i}^*|F)$ the detection completeness at a line flux F (see Section 2.1).

Using the expressions for detections and non-detections, the posterior distribution takes the final form:

$$p(W_n|\{F\}) = \frac{p(W_n)}{p(\{F\})} \prod_D p(F_{\rm i}|W_n) \prod_{ND} p(F_{\rm i} < F_{\rm i}^*|W_n) \tag{9}$$

With $p(W_n)$ the priors of the model parameters and $p(\{F\})$ a normalization constant reflecting the likelihood of the model. The form of Equation (9) is general and will be used as starting point for multiple analysis throughout.

3.2. Model Comparison

As we introduced in the previous section, multiple probability distributions can be adopted for representation of the $W_{Ly\alpha}$ distribution. In order to perform model selection, several elements are taken into account, such as model complexity and number of parameters. In this section, we describe our methodology to perform such selection from a quantitative standpoint. By means of a Bayesian approach, we recover probability ratios for the different models, providing insight into how we perform the selection given our measurements. The

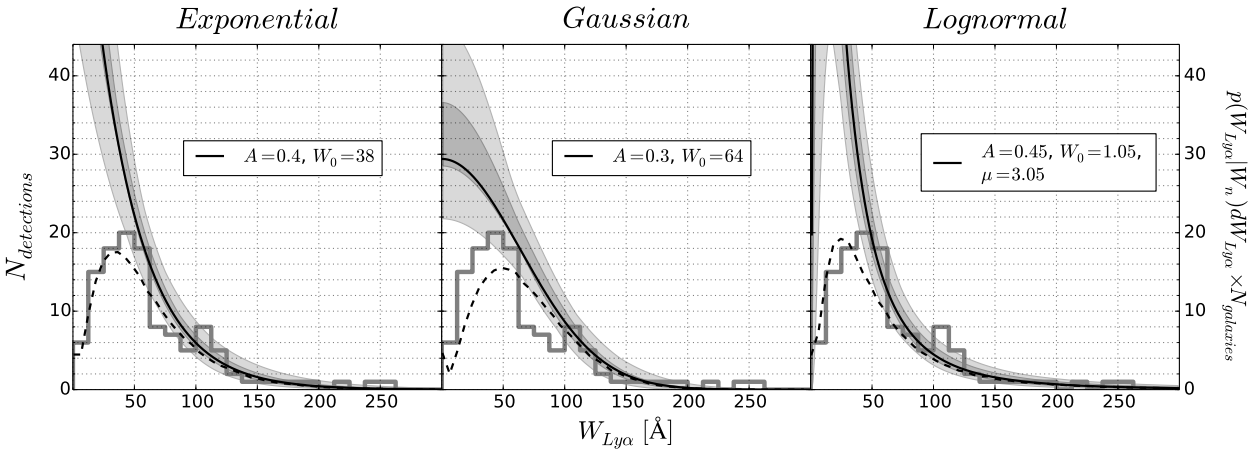

Figure 4 We show in histograms the observed rest frame $W_{Ly\alpha}$ distribution for the whole sample. These histograms account for the 120 measured $W_{Ly\alpha}$ with S/N$\geqslant$5.5. From left to right, we overplot exponential, gaussian, and lognormal model constraints, respectively. The solid black lines correspond to the peak of the likelihood, i.e., the most probable solutions. The dashed lines represent these solutions corrected for our completeness. The shaded regions show the 1σ and 3σ confidence levels yielded by the likelihood.

analysis presented here is a quantitative implementation of Occam's Razor and is not unique to our dataset, i.e., it can be applied to any dataset modeling.

Every model we discuss here is composed of a scaled probability distribution and a Dirac Delta, as proposed in Treu et al. (2012). If we define the standard probability distributions as $p_0(W_{Ly\alpha}|W_n)$, the modified counterparts we consider are given by:

$$p(W_{Ly\alpha}|W_n) = A \times H(W_{Ly\alpha}) \times p_0(W_{Ly\alpha}|W_n) \\ +(1-A) \times \delta(W_{Ly\alpha}) \tag{10}$$

where the first term is the scaled probability distribution. It is multiplied by the Heaviside $H(W_{Ly\alpha})$ to ensure it only represents positive $W_{Ly\alpha}$ values. Hence, this scaled term integrates A, i.e., A can only adopt values between 0 and 1. Note that this term is different to the fraction of detections in our sample, as we consider upper limits for our non-detections. The second term groups the fraction of galaxies that do not emit in Lyα (i.e., no line and/or absorption). As our data is restricted to emission lines, we represent this term using the Dirac Delta.

We explore here exponential, gaussian, and lognormal based distributions. The first two are 2-parameter models, while the lognormal is 3-parameter dependent. Hence, we generalize our parameter space as $W_n = (A, W_0, \mu)$. Then, the expressions for our exponential, gaussian, and lognormal models are, respectively:

$$p(W_{Ly\alpha}|\mathrm{exp}) = \frac{A}{W_0} \mathrm{e}^{-W_{Ly\alpha}/W_0} H(W_{Ly\alpha}) + (1-A)\delta(W_{Ly\alpha}) \tag{11}$$

$$p(W_{Ly\alpha}|\mathrm{gauss}) = \frac{2A}{\sqrt{2\pi W_0^2}}\mathrm{e}^{-W_{Ly\alpha}^2/2W_0^2}H(W_{Ly\alpha}) + (1-A)\delta(W_{Ly\alpha}) \quad (12)$$

$$p(W_{Ly\alpha}|\mathrm{logn}) = \frac{A}{\sqrt{2\pi W_0^2 W_{Ly\alpha}^2}}\mathrm{e}^{-(ln(W_{Ly\alpha})-\mu)^2/2W_0^2} + (1-A)\delta(W_{Ly\alpha}) \quad (13)$$

We now describe our approach to compare the three models. For a set of measurements, Bayes' Theorem gives the probability of model M_i in the model space $\{M\}$

$$p(M_\mathrm{i}|\{F\}) = \frac{p(\{F\}|M_\mathrm{i})p(M_\mathrm{i})}{p(\{F\})} \quad (14)$$

with $p(M_\mathrm{i})$ the prior for model M_i in the set $\{M\}$, which we assume to be equal for the three models. Once again, $p(\{F\})$ is a normalization constant. Therefore, the probability for model M_i is proportional to the likelihood of the model, i.e.:

$$p(M_\mathrm{i}|\{F\}) \propto p(\{F\}|M_\mathrm{i}) = \int_n p(W_n)p(\{F\}|W_n)\mathrm{d}W_n \quad (15)$$

The absolute probability of each model M_i within the set $\{M\}$ is obtained imposing that the models explored cover all possible choices, i.e., $\sum p(M_\mathrm{i}|\{F\}) = 1$. Hence, the probability of M_i given our dataset is

$$p(M_\mathrm{i}|\{F\}) = \frac{\int_n p(W_n)p(\{F\}|W_n)\mathrm{d}W_n}{\sum p(M_\mathrm{i}|\{F\})} \quad (16)$$

For an analytically correct model comparison, analysis of Equation (16) is required. Still, the term in Equation (15) is strongly dependent on the priors assumed for the parameter space W_n of every model. Therefore, different prior selections can have significant effects on the odds for each model. As a workaround, we rewrite the model probabilities as:

$$p(M_\mathrm{i}|\{F\}) = \frac{\int_n p(\{F\}|W_n)\mathrm{d}W_n}{\sum p(M_\mathrm{i}|\{F\})} \quad (17)$$

Effectively, this simplification conveniently limits our analysis to a pure likelihood comparison, i.e., we adopt constant, uninformative priors. This is equivalent to assuming ignorance in linear scales for every parameter. Since our distributions are smooth and single peaked, we are confident in this assumption. The dimensions of our parameter spaces do not go beyond three, and the uncertainties in our parameters are of the order of the most probable values, providing further assurance to our assumptions.

In Oyarzún et al. (2016), we show that a galaxy sample with a broad M_* range yields a composite $W_{Ly\alpha}$ distribution. For the rest of this section, we divide our sample in three M_* bins, as shown in Figure 2. We use the complete sample and these three subsamples to contrast the models. The outcomes are presented in Table 1 and Figure 4. Table 1 evidences that the best likelihood is obtained with the lognormal model for three of the four distributions.

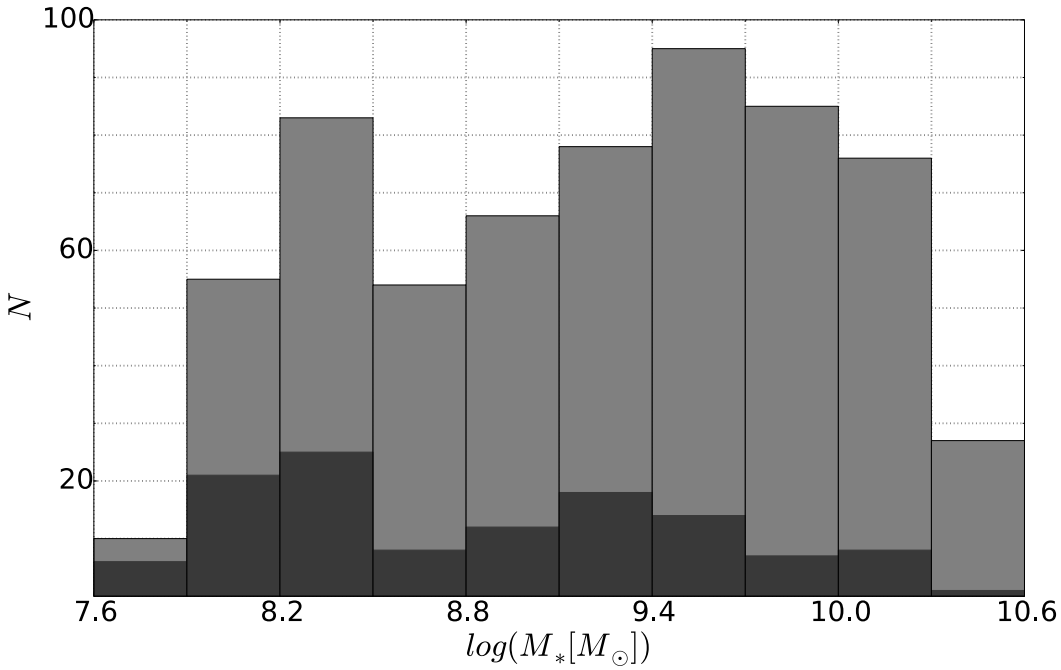

Figure 5 M_* histogram of the 629 galaxies in our sample (grey) and 120 detections (black). The values of M_* we show are obtained using FAST and assuming constant SFHs. For detections, M_* are calculated using spectroscopic redshifts ($z_{Ly\alpha}$). For non-detections we use the peak probability from EAZY (z_{EAZY}). Note that any inferences on the fraction of detections from this histogram must take into account M_* dependent incompleteness.

This can be verified in our distribution simulations for the complete sample in Figure 4, especially towards the high $W_{Ly\alpha}$ tail. Still, when integrating the likelihoods, the lognormal distribution is the least probable. This is a consequence of the extra parameter needed by the model, which penalizes when integrating over the parameter space. Then, according to our analysis, the preferred model is the exponential. While it models the distribution better than the gaussian, it also reproduces our $W_{Ly\alpha}$ measurements fairly well, despite depending on only 2 parameters. The uncertainties in Figure 4, especially for low $W_{Ly\alpha}$, also confirm this model is the most adequate to reproduce our measurements. We remark that the procedure for model selection described here considers the lower $W_{Ly\alpha}$ end of the distribution, which includes our completeness and non-detections. Still, further $W_{Ly\alpha}$ distribution analysis in this paper is mostly focused on the higher $W_{Ly\alpha}$ end, and is not strongly dependent on model preference.

From now on, we perform our $W_{Ly\alpha}$ distribution analysis using the exponential model of Equation (11). We stress this expression is dependent on parameters A and W_0, with the first being the fraction of galaxies showing emission and W_0 the e-folding scale of the distribution. It should be kept in consideration that A is not necessarily the fraction of emitters in the population, but the best constant for reproducing this fraction by an exponential profile.

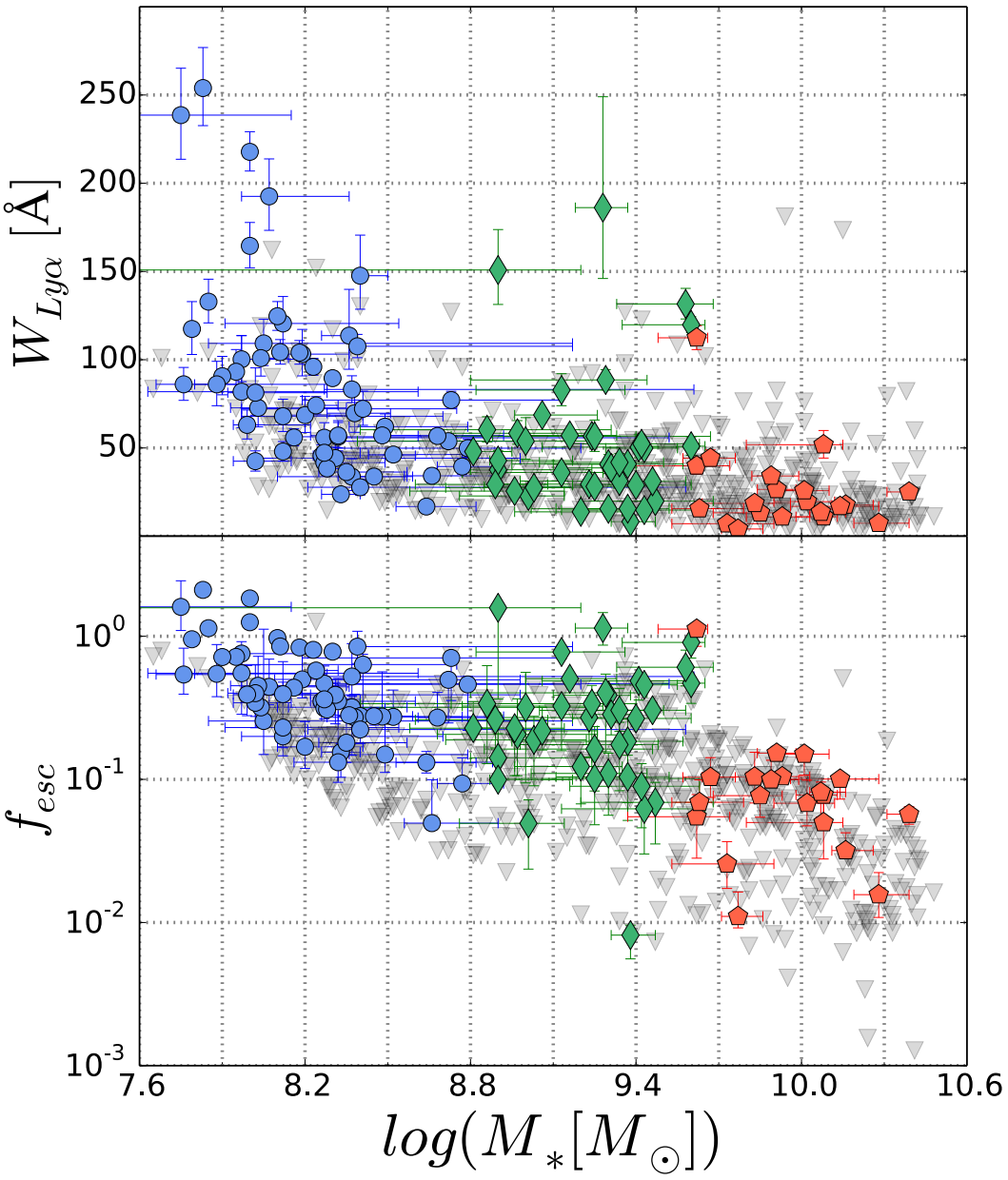

Figure 6 Top: Plot of our $W_{Ly\alpha}$ as a function of M$_*$. Bottom: Plot of our Lyα escape fractions as a function of M$_*$. The detections corresponding to the low-, medium-, and high-mass subsamples are shown as blue circles, green diamonds, and red pentagons, respectively. We observe both Lyα diagnostics, $W_{Ly\alpha}$ and f_{esc}, to anti-correlate with M$_*$. The upper limits associated to non-detections are plotted as grey triangles. Note how our completeness is not independent of M$_*$, which has to be considered when making inferences on the M$_*$ distribution and characteristic $W_{Ly\alpha}/f_{esc}$ of LAEs.

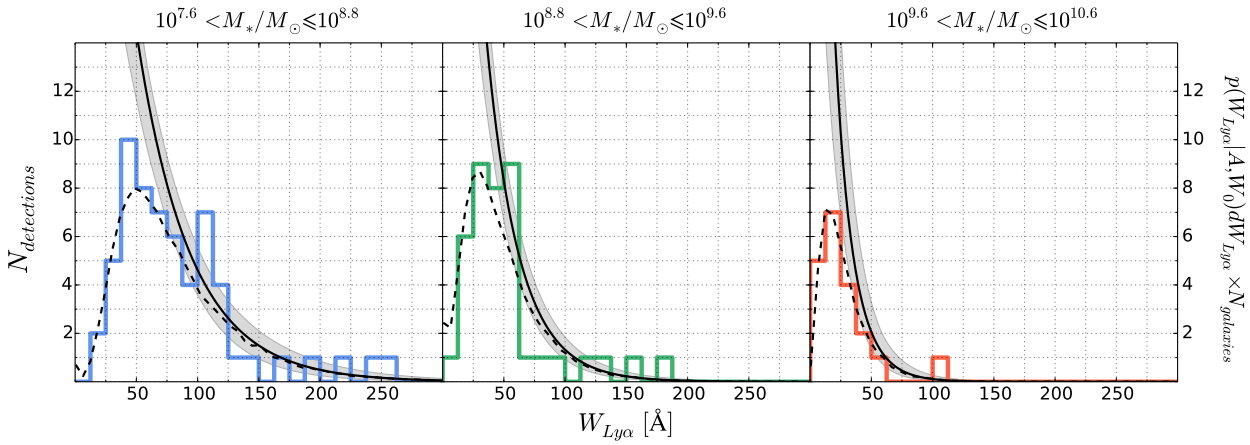

Figure 7 Rest frame $W_{Ly\alpha}$ distributions of the low-, medium-, and high-mass bins, from left to right. We use Monte Carlo simulations to characterize the exponential $W_{Ly\alpha}$ distribution dependence on M_* following Equations (18) and (19). We can then simulate $W_{Ly\alpha}$ distributions for every subsample and obtain 1σ constraints (shaded contours). Keep in mind that these constraints are given by collections of curves resulting from our Monte Carlo simulations. Our best results correspond to the avergage distribution we expect considering every object in the bin. We also plot as dotted lines the completeness corrected counterparts of the average distributions.

4. Lyα Emission Dependence on Galaxy Properties

4.1. Stellar Mass

Evidence suggests Lyα emission is strongly dependent on the M_* of galaxies. Galaxies with higher M_* have been forming stars for longer, leading to higher ISM dust that presumably forms in supernovae and AGB stars (Silva et al. 1998). A greater dust covering fraction leads to more Lyα photons absorption, decreasing $W_{Ly\alpha}$. This effect has already been observed, at least for high $W_{Ly\alpha}$, in Blanc et al. (2011) and Hagen et al. (2014). Similarly, the bulk of M_* is dominated by older stars, which do not contribute significantly to the Lyα photons budget of galaxies. As a matter of fact, Lyα emission decreases steadily with the age of stellar populations, as seen in Charlot & Fall (1993) and Schaerer (2003). Lyα radiative transfer is also severely affected by the neutral gas structure and kinematics of the ISM and circumgalactic medium (Verhamme et al. 2006). Since more massive galaxies are bound to have higher gass mass, Lyα photons should be subject to more resonant scattering. As a consequence, Lyα photons have higher probabilities of being absorbed, decreasing their $W_{Ly\alpha}$. The trends we find in Oyarzún et al. (2016) confirm this qualitative scheme at $z \sim 4$. In this section, we perform a more detailed and robust characterization of Lyα emission dependence on M_*.

Our sample is especially designed to study the dependence of Lyα emission in M_*. As shown in Figure 2, our objects are selected in redshift and M_*, homogeneously covering the range $8 < log(M_*[M_\odot]) < 10.5$. For further clarity, we plot in Figure 5 our sample M_* distribution. We also include in this figure the M_* distribution of our detections. Comparison

of both histograms hints at an anti-correlation between LAEs fraction and M_*, at least down to our detection limit. The existence of a Lyα emission dependence on M_* becomes clear in Figure 6, where we plot $W_{Ly\alpha}$ and f_{esc} as a function of M_*. We also plot in Figure 6 our upper limits for non-detections. The regions sampled by our non-detections reveal how our completeness is not independent of M_*. Our detections are flux limited, so we achieve lower $W_{Ly\alpha}$ for galaxies with brighter UV continuum. Therefore, even though we observe lower M_* galaxies to have higher $W_{Ly\alpha}$ and f_{esc}, any qualitative conclusions we can draw involving the fraction of LAEs as a function of M_* are affected by our completeness. Still, there is a clear upper envelope to the distribution of galaxies in this plot, where we are not affected by incompleteness. Hence, our main result is that both Lyα emission diagnostics show a clear anti-correlation with M_* .

The dependence of Lyα emission on M_* implies that $W_{Ly\alpha}$ distributions in the literature (e.g., Gronwall et al. 2007, Zheng et al. 2014) are influenced by the M_* distribution of the sample. In comparison to deeper M_{UV} surveys, shallower samples are bound to observe lower $W_{Ly\alpha}$. This can lead to incorrect contrast of surveys and misinterpretation of trends. To verify these claims, we divide our sample in three M_* bins (Figure 2) and plot the resulting $W_{Ly\alpha}$ distributions in Figure 7. As expected, there is an apparent anti-correlation between M_* and both, the tail of the $W_{Ly\alpha}$ distribution and the normalization. We perform our first quantitative characterization of the $W_{Ly\alpha}$ distribution dependence on M_* in Oyarzún et al. (2016). In that study, we divide our sample in three M_* bins and obtain the posterior distribution for the exponential parameters separately. This procedure allowed us to fit a linear relation to the final parameters, recovering $A(M_*)$ and $W_0(M_*)$ using expressions of the form:

$$A(M_*) = A_{M_*} \log(M_*[M_\odot]) + A_C \tag{18}$$

$$W_0(M_*) = W_{M_*} \log(M_*[M_\odot]) + W_C \tag{19}$$

In this section, we recover these linear relations directly from the complete sample, i.e., we recover the posterior distribution for the 4-parameter space composed by the linear coefficients in Equations (18) and (19). We adopt here a more robust methodology that does not relies on binning, while also allowing us to constrain the errors on the coefficients directly from the model and measurements. We once again start from Equation (9). As mentioned, our parameter space is now $W_n = (A_{M_*}, A_C, W_{M_*}, W_C)$. As these linear coefficients represent the exponential parameters of Equation (11), a purely analytical approach to obtain the priors is highly complicated. Therefore, in order to determine our priors, we only consider linear scales ignorance. Therefore, the posterior translates to:

$$p(A_{M_*}, A_C, W_{M_*}, W_C|\{F\}) = C\times \tag{20}$$
$$\prod_D p(F_i|A(M_*), W_0(M_*)) \prod_{ND} p(F_i < F_i^*|A(M_*), W_0(M_*))$$

with C a normalization constant.

We use MC simulations to characterize this 4-parameter posterior. Its maximum gives the best solution for $A(M_*)$ and $W_0(M_*)$, while the collapsed posteriors yield the uncertainties

on the parameters. We can then write Equations (18) and (19) as:

$$A(\mathrm{M}_*) = -0.28^{+.12}_{-.02} \log\left(\mathrm{M}_*[\mathrm{M}_\odot]\right) + 3.1^{+0.15}_{-1.1} \tag{21}$$

$$W_0(\mathrm{M}_*) = -17.7^{+2.6}_{-2.0} \log\left(\mathrm{M}_*[\mathrm{M}_\odot]\right) + 190^{+20}_{-25} \tag{22}$$

In the framework of an exponential profile, these relations we recover yield the $W_{Ly\alpha}$ probability distribution for an object with known M_*. Hence, we can simulate the expected $W_{Ly\alpha}$ distribution for each of the three M_* subsamples and compare with our direct measurements. The results are presented in Figure 7. Indeed, our constraints are consistent with the observed $W_{Ly\alpha}$ distributions, further confirming our Bayesian outcome within our dataset.

Our results regarding the dependence of Lyα on M_* are conclusive. In the range $10^8 - 10^{10.5}\mathrm{M}_\odot$, the $W_{Ly\alpha}$ probability distribution extends to higher $W_{Ly\alpha}$ for lower M_* galaxies. In other words, more massive galaxies tend to have lower $W_{Ly\alpha}$. A similar trend is observed for f_{esc}, further highlighting the role of dust and gas mass in the escape of Lyα photons. At $z \sim 2.2$, Matthee et al. (2016) also observe the f_{esc} anti-correlation from Figure 6, although only when using stacks. Their massive objects showing high f_{esc} seem to lie below the M_*-SFR sequence at $z \sim 2$, which they associate to dusty gas outflows. These inferences, combined with the significantly higher f_{esc} they measure for larger apertures, make sense in a Lyα diffuse halo scheme we do not observe due to our aperture size. There are also $1.9 < z < 3.8$ studies on the dependence of f_{esc} on M_* using $W_{Ly\alpha}$ selected LAEs (Hagen et al. 2014). They find a trend similar to the one we find at $z \sim 4$. Then, we can say that the evidence for an anti-correlation between $W_{Ly\alpha}$ (or f_{esc}) and M_* is significant, but the scatter seems to depend on measurement methodology and sample selection. We discuss in detail the potential selection effects induced by Lyα surveys in Section 5.

Inferences on the M_* distribution of LAEs are not as evident. Hagen et al. (2014) do not find their $1.9 < z < 3.6$ Lyα luminosity selected LAEs distribution to depend on M_*. Their results agree with the $z \sim 3.1$ narrowband selected survey of McLinden et al. (2014). However, in Figure 6 we reveal that survey completeness is not independent of M_*. Therefore, most LAEs surveys could have higher incompleteness towards lower M_*. Since our Bayesian analysis takes into account our completeness for every object, we can test the significance of this claim. The coefficient A_{M_*} in Equation (18) represents the exponential fraction of LAEs dependence on M_*. As evidenced by Equation (21), our measurements are more than 1σ consistent with a decrease in the fraction going to higher M_*. In the scheme of an exponential model, this translates into a LAEs distribution dominated by lower M_* galaxies. This result comes to complement the much more significant anti-correlation between W_0 and M_* we find (see Equation (22)).

4.2. SFR

High-redshift galaxies have been observed to follow a correlation between SFR and M_*, known as the star-forming main sequence (e.g., Kereš et al. 2005, Finlator et al. 2007, Stark et al. 2009, González et al. 2011, Whitaker et al. 2012). In terms of the underlying physics, more massive objects dominate gas accretion in their neighborhood, feeding and triggering

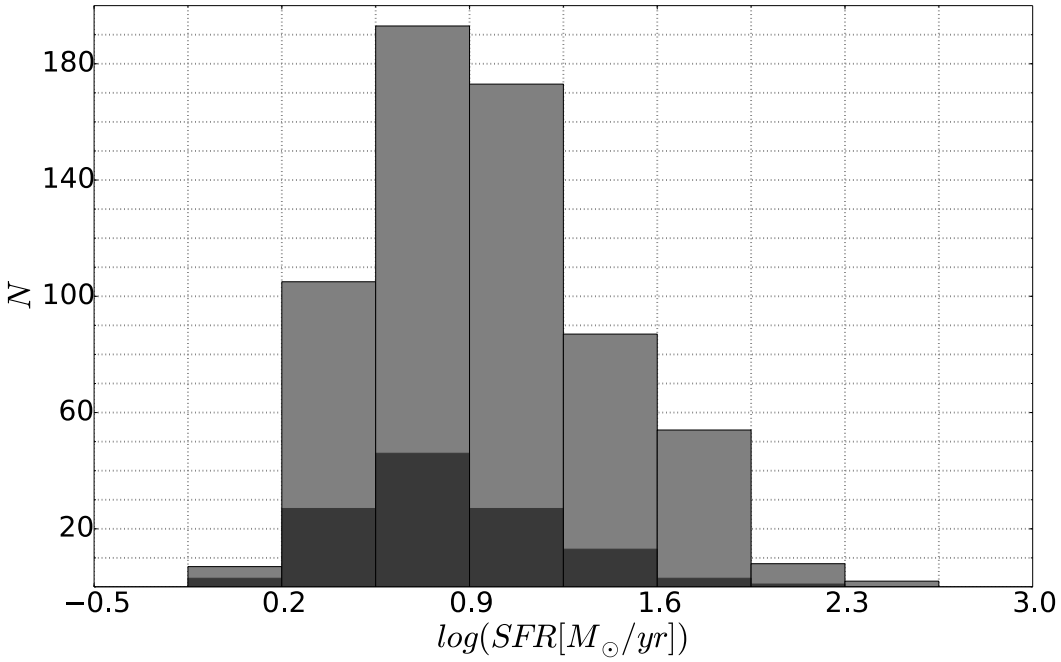

Figure 8 Distribution in SFR for our sample (grey) and Lyα detections (black). Our extinction-corrected SFRs are derived from SED fitting to CANDELS photometry using FAST. We remark that this plot is shown only for sample description purposes. Since our targets are M_* selected, we do not expect them to replicate the SFR distribution of the galaxy population and/or Lyα emitting galaxies. Keep in mind that our detection completeness is strongly dependent on SFR (see Figure 9).

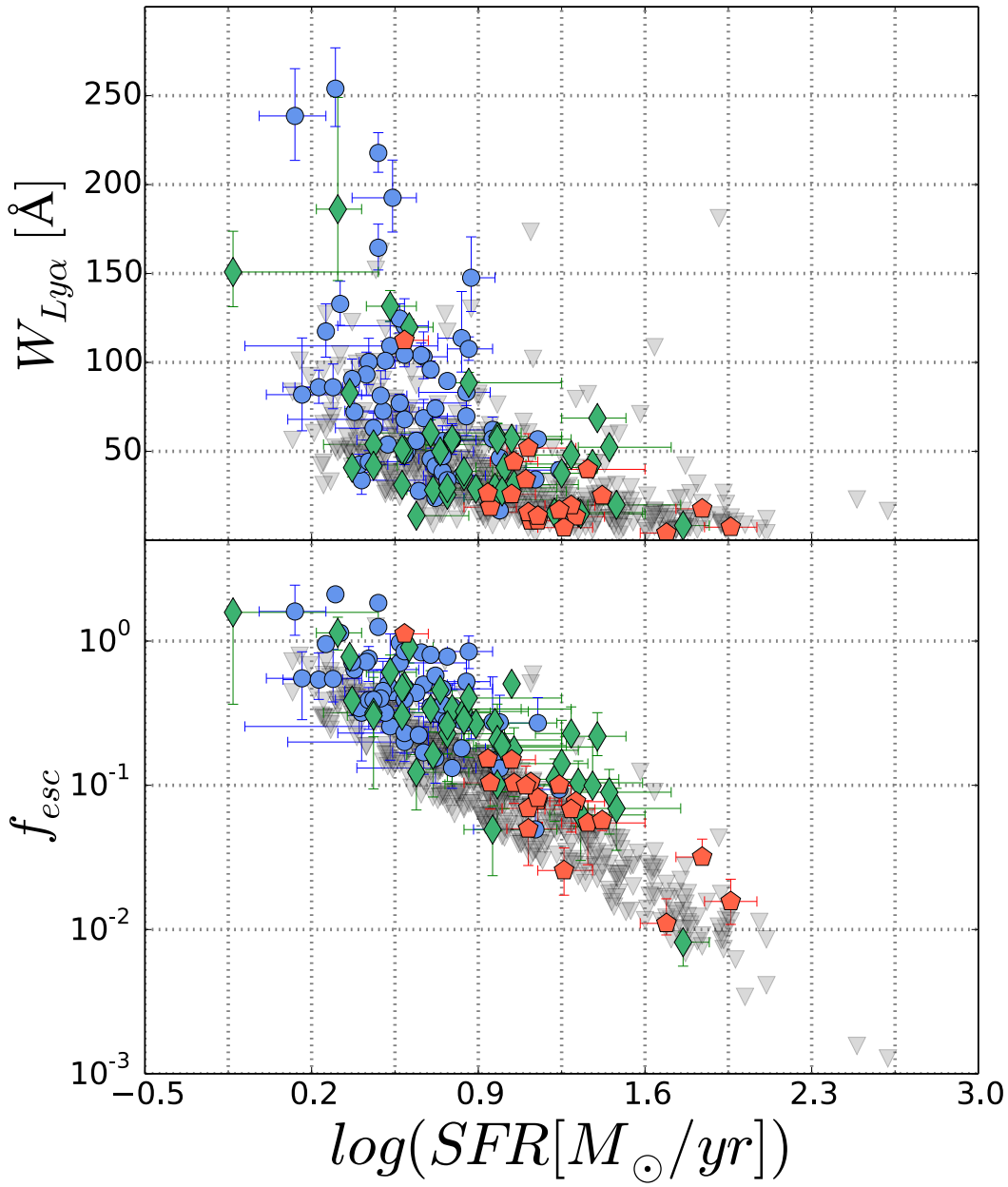

Figure 9 Plot of our $W_{Ly\alpha}$ (top panel) and f_{esc} (bottom panel) as a function of SFR. Detections corresponding to our low-, medium-, and high-mass bins are shown as blue circles, green diamonds, and red pentagons, respectively. We find both measurements to anti-correlate with the SFR we measure from SED fitting. For the plot in the bottom panel, we use the same value for SFRs when determining the escape fraction and galaxy SFR. For reference, we show our non-detections upper limits as grey triangles. The region that they trace clearly shows that our detection completeness is strongly dependent on SFR.

star-formation. Such gas infall seems to dominate over galaxy growth at high-redshift (Kereš et al. 2005, Finlator et al. 2007). This scheme implies that more massive objects form stars at higher rates, at least down to our observational limitations and modeling of high-redshift ISM. Given our results on M_* from the previous section, we expect similar trends between $W_{Ly\alpha}/f_{esc}$ and SFR. Even more, star-forming galaxies have a higher neutral gas budget available, which is an important factor in ruling the escape of Lyα photons from galaxies (Verhamme et al. 2006). It has also been suggested that photoelectric absorption rules Lyα depletion, even over dust attenuation (Reddy et al. 2016). In this section, we explore any Lyα dependence on SFR. We remark that our SFRs come from SED fitting of CANDELS photometry using FAST (see Section 1.3), i.e., they have typical associated timescales of 100 Myr (Kennicutt 1998). We stress that our derived SFRs differ from 3DHST SFRs, since our calculation assumes cSFHs instead of exponentially declining SFHs (Skelton et al. 2014).

We plot our sample and detections distribution as a function of SFR in Figure 8. We also show $W_{Ly\alpha}$ and f_{esc} dependence on SFR in Figure 9 and include upper limits for our non-detections to give an insight into how our incompleteness depends on SFR. A clear anti-correlation between $W_{Ly\alpha}/f_{esc}$ and SFR is observed. These results come as no surprise, as previously observed in the literature. Most studies of $W_{Ly\alpha}$ dependence on SFR involve uncorrected SFRs (Tapken et al. 2007, Pettini et al. 2002, Shapley et al. 2003, Yamada et al. 2005, Gronwall et al. 2007, and Ouchi et al. 2008 ; all compiled in Verhamme et al. 2008). Even without dust correction, the anti-correlation we observe is still present in these studies (see Figure 19 in Verhamme et al. 2008). Still, their scatter is much more significant than in our plot, to the point of no trend existence at all (Atek et al. 2014). We argue that the trend is more significant when using extinction corrected SFRs, reducing the scatter. We verify this claim in our analysis of Lyα dependence on M_{UV} in Section 4.3.

Based on a $z \sim 2$ Hα emitters sample, Matthee et al. (2016) also observe a clear anti-correlation between Lyα f_{esc} and SFR. Interestingly, they do not only observe such trend in their individual objects, but also on their stacks using different apertures (galaxy diameters of 12 kpc and 24 kpc). As their dataset includes Hα fluxes, they can recover SFRs and f_{esc} using Hα luminosities. The fact that they observe similar trends with such a different sample suggests that this anti-correlation between Lyα/f_{esc} and SFR is not only independent of redshift, but also observational constraints like aperture and methodology for recovering SFRs. This comparison, however, is restricted to SFRs higher than $\sim 5 M_\odot$/yr, which is the lower SFR measured by Matthee et al. (2016). Most of our low-mass objects have SFRs lower than $\sim 5 M_\odot$/yr, but they seem to follow the same regime as the rest of our sample. Naturally, the uncertainties in f_{esc} increase towards lower SFRs, UV fainer galaxies (Figure 9), but our results suggest that f_{esc} reaches values of 100 % towards SFR$\sim 1 - 3 M_\odot$/yr. It is worth mentioning, however, that our results are biased to measure only the highest f_{esc} objects at a given SFR. This is a result of flux limitations in Lyα surveys, which are unavoidable regardless of the methodology. Therefore, even though it seems likely that most low SFR Lyα emitters have high f_{esc}, deeper surveys are needed to dismiss the existence of low f_{esc} LAEs.

4.3. UV Luminosity

In this section, we analyze the M_{UV} distribution of our sample, while also exploring any correlations between $W_{Ly\alpha}$ and UV luminosities. It must be noted, though, that our sample is not representative of the galaxy population at $3 < z < 4$. First, our galaxies are homogeneously distributed in M_*, i.e., they are not a random sample from $3 < z < 4$ CANDELS objects. Second, we are affected by CANDELS completeness, which decreases towards lower M_* galaxies. Since more massive galaxies tend to have higher UV luminosities (Stark et al. 2009, González et al. 2014), our sample has a higher contribution of bright M_{UV} galaxies than a population representative subsample.

We show in Figure 10 the M_{UV} distribution for the complete sample and detections. We also present the dependence of $W_{Ly\alpha}$ on M_{UV} in Figure 11. As expected from the SFR-$W_{Ly\alpha}$ anti-correlations we recover in Section 4.2, a similar trend is observed for UV luminosities. This anti-correlation comes as no surprise, since brighter UV galaxies tend to have higher M_* at the cosmic time of our sample. Galaxies brighter in the UV have been subject to more intensive star-formation events in 100 Myr timescales. Typically, the higher neutral gas and turbulence associated with higher SFRs boosts the scatter of Lyα photons. In combination, higher dust extinction, older age of stellar populations, and greater neutral gas mass in UV brighter galaxies seem to dominate over the increased Lyα and UV continuum output from higher SFRs. This result is consistent with a scenario where Lyα is more prominent in galaxies with higher specific SFR (sSFR).

Most analysis of Lyα emission dependence on M_{UV} has been performed using un-corrected SFRs. The literature compilation shown in Verhamme et al. (2008) reveals how observed UV SFRs anti-correlate with $W_{Ly\alpha}$ from z$\sim$ 2.7 to 5.7 (Tapken et al. 2007, Pettini et al. 2002, Shapley et al. 2003, Yamada et al. 2005, Gronwall et al. 2007, and Ouchi et al. 2008). Analysis using explicitly M_{UV} have been also performed in the surveys of Shimasaku et al. (2006), Ouchi et al. (2008), Vanzella et al. (2009), Balestra et al. (2010), Stark et al. (2010), Schaerer et al. (2011), and Cassata et al. (2011), yielding similar trends. Simulations also predict such correlations (Shimizu et al. 2011). However, regardless of methodology, the scatter in these correlations is non-negligible. Even more, Atek et al. (2014) question the existence of the correlation in their sample. We argue that the scatter observed in the $W_{Ly\alpha}$ dependence on M_{UV}/SFR_{obs} is a consequence of dust degeneracy. As we find in our analysis on dust-corrected SFRs from Section 4.2, the trend is better defined after correcting for dust. Galaxies brighter in the UV naturally have a greater Lyα photon production, but the effect of M_* and/or extinction affects the escape fraction. In this scenario, Lyα photon escape is a complex process simultaneously ruled by different properties of high-redshift galaxies. We explore such property space in our analysis of the $W_{Ly\alpha}$ distribution dependence on the M_*-SFR sequence in Section 4.4.

Implications involving the fraction of detections as a function of M_{UV} are not straightforward. In principle, this fraction seems to correlate with the UV luminosities of our galaxies, as opposed to what is observed for characteristic $W_{Ly\alpha}$. Nevertheless, our Lyα measurements are flux-limited. Therefore, our detection completeness in $W_{Ly\alpha}$ is higher for brighter objects, leading to biases difficult to account for, as noted in Nilsson et al. (2009). Under these circumstances, the ideal approach is to consider both detections and non-detections, while

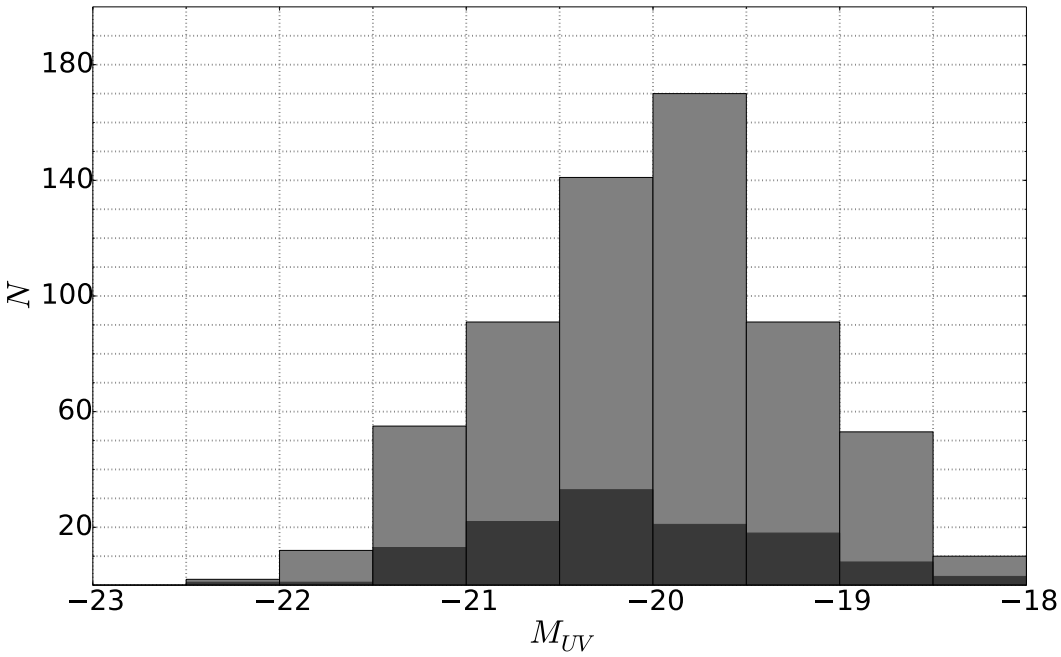

Figure 10 Histograms of our sample (grey) and detections (black) as a function of UV absolute magnitude. The distribution of our sample and detections is roughly similar, although our incompleteness is higher towards fainter objects (Figure 11). Therefore, our analysis suggests the fraction of LAEs decreases with M_{UV}.

taking into account the uncertainties for line and continuum fluxes (F and f_λ, respectively). Hence, we encourage further interpretations of these results to focus on the analysis performed in the M_*-SFR plane (Section 4.4).

4.4. Extinction

Measurement of high-redshift galaxies UV slopes is a direct way of tracing the amount of dust inside galaxies, given the assumption of an extinction law and an intrinsic spectral shape. This is of particular interest for LAE surveys, since simulations (e.g., Verhamme et al. 2008) and observations at low-redshift (Hayes et al. 2011, Atek et al. 2014) suggest that dust plays an important role in the escape of Lyα photons. Since we have rest UV photometry from CANDELS for all our objects, we can determine their UV slopes and study their effect on Lyα emission at $z \sim 4$. In this section, we detail or method to estimate the UV slopes for our objects and show our results on Lyα dependence on this galaxy property.

To determine UV slopes, we fit a power law $f_\lambda = f_0 \lambda^\beta$ (Calzetti et al. 1994) to the photometry of each object. For fitting, we just use a standard least-squares routine. To do so, we use the best-fit SED between rest 1200Å and 3600Å, which come from 3D-HST rest-frame colors. For the uncertainties, however, we use the observed SED. These calculations effectively correspond to an average of 8 bands between 1400Å and 3500Å. We use the slopes to also account for extinction $E(B-V)$. Similarly to Blanc et al. (2011), we assume an pristine slope of $\beta = -2.23$ (Meurer et al. 1999) and a Calzetti et al. (2000) extinction law,

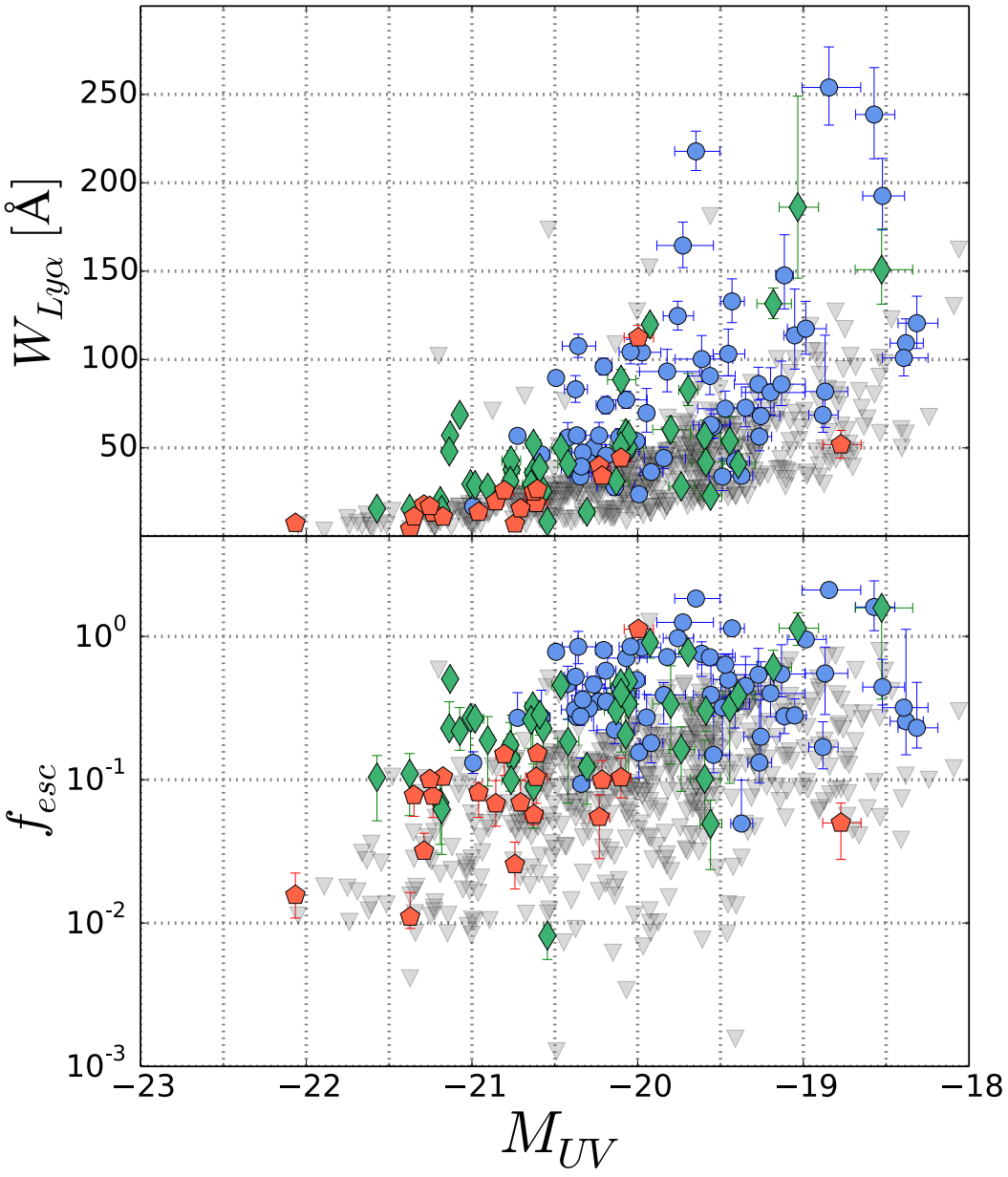

Figure 11 We show in this plot the dependence of $W_{Ly\alpha}$ (top) and f_{esc} (bottom) on UV absolute magnitude. The blue circles, green diamonds, and red pentagons correspond to detections on the low-, medium-, and high-mass subsamples, respectively. Non-detections upper limits are shown as grey triangles. Our results suggest that $W_{Ly\alpha}$ and f_{esc} anti-correlate with UV luminosity. These trends are consistent with studies performed on LBG samples (Stark et al. 2010, Schaerer et al. 2011) and narrow-band surveys (Ouchi et al. 2008).

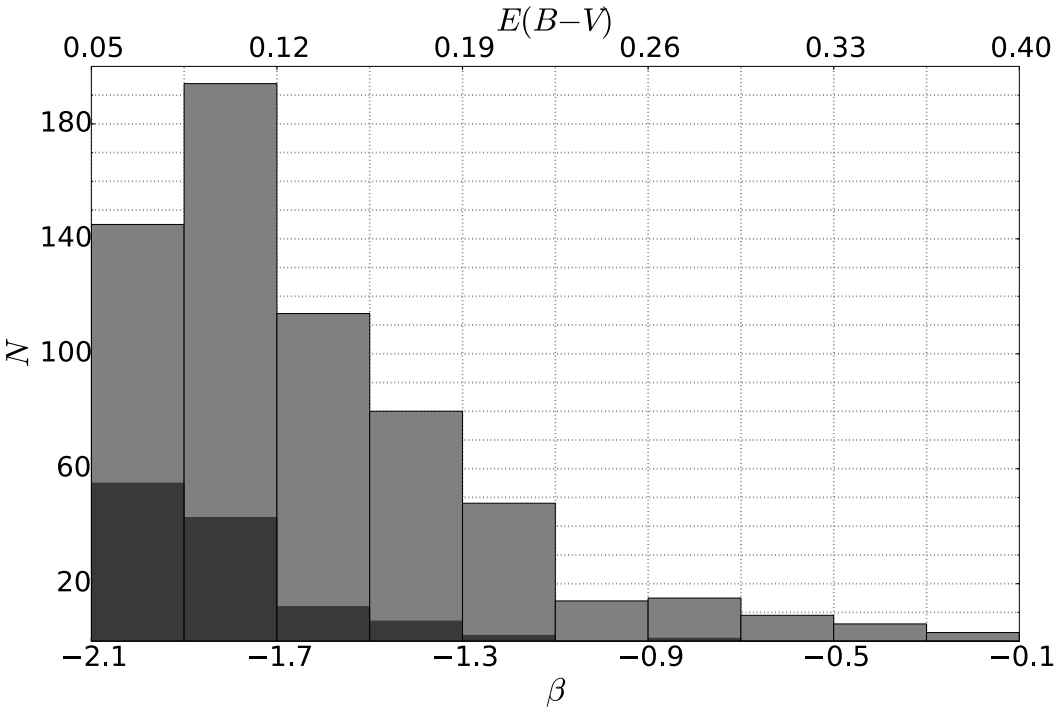

Figure 12 We show as grey histograms our sample distribution as a function of UV slope (β) and/or E(B-V). We obtain E(B-V) assuming a Calzetti et al. (2000) extinction law with a pristine slope of $\beta = -2.23$ (Meurer et al. 1999). We include in this plot the distribution for our 120 detections as black histograms. At ~ 4, our M_* selected sample is dominated by blue slopes, with this also the case for our detections.

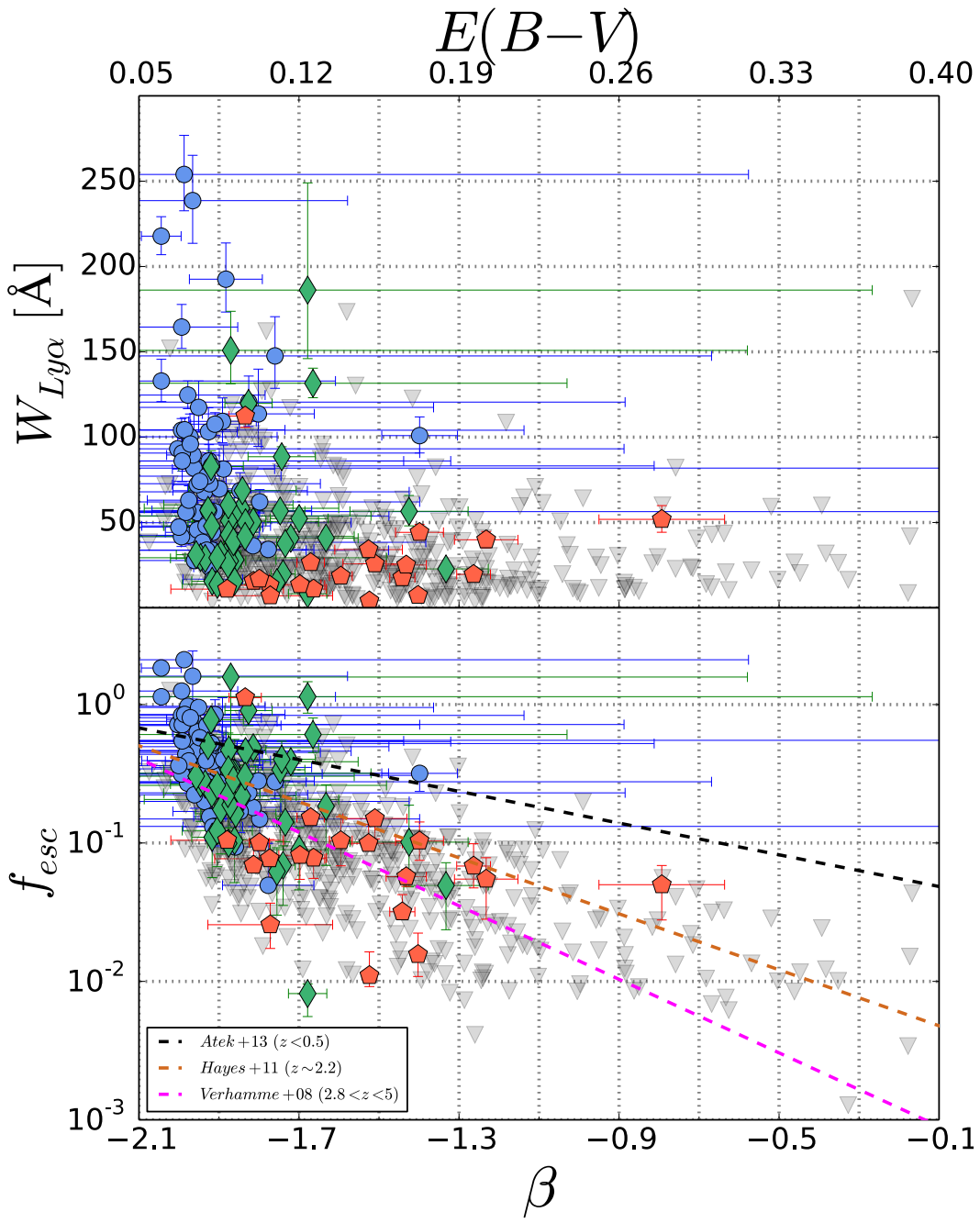

Figure 13 We show in this plot $W_{Ly\alpha}$ and f_{esc} dependence on UV slope (β) and E(B-V). We obtain E(B-V) assuming a Calzetti et al. (2000) extinction law with a pristine slope of $\beta = -2.23$ (Meurer et al. 1999). Blue circles, green diamonds, and red pentagons differentiate our low-mass, medium-mass, and high-mass detections, respectively. Grey triangles represent our upper limits from non-detections. Our results clearly show that high $W_{Ly\alpha}/f_{esc}$ objects are associated to extremely blue slopes, consistent with the notion that dust boosts the absorption of Lyα photons. We also include several trends from the literature for comparison. Discrepancies towards low $E(B-V)$ between the literature and our results are most likely a consequence of our incompleteness.

leading to:

$$f_\lambda = f_0 \lambda^\beta \propto \lambda^{-2.23} 10^{0.4 k_\lambda E(B-V)} \tag{23}$$

Figure 12 shows the $E(B-V)$ distribution for our sample and detections. Since our targets are M_* selected, we have a higher contribution of massive objects in comparison to the M_* distribution of the galaxy population. As more massive objects tend to have higher $E(B-V)$ and redder UV slopes, we expect $z \sim 4$ samples representative of the galaxy population to have lower contribution from such galaxies. We show in Figure 13 our results on $W_{Ly\alpha}$ and f_{esc} as a function of β and $E(B-V)$. Since lower mass galaxies have bluer UV slopes than more massive ones, the trends we find are complimentary to our previous results. There is a correlation between steepness of UV spectrum and $W_{Ly\alpha}$, although with significant scatter. As extinction seems to play a major role in Lyα photons escape from galaxies, mainly through scattering and absorption (Blanc et al. 2011, Hagen et al. 2014), this correlation comes as no surprise.

Qualitatively, our results agree with the measurements from Shapley et al. (2003), Pentericci et al. (2009), Blanc et al. (2011), and Atek et al. (2014). Galaxies with $W_{Ly\alpha} > 100$Å are mostly associated to $E(B-V) < 0.15$. Some differences show up, though, when comparing the more dusty Lyα emitters from each sample. We have only one object with $E(B-V) > 0.25$, while the $1.9 < z < 3.8$ survey from Blanc et al. (2011) and $z < 0.5$ study from Atek et al. (2014) find Lyα emitters up to $E(B-V) = 0.4$. Similarly, Mathee et al. (2016) find a population of dusty LAEs at $z \sim 2.23$ and speculate on how dusty gas outflows might be the feature driving the escape of Lyα. We cannot state whether the absence of such LAEs in our sample is representative of $z \sim 4$ LAEs. Truly enough, the fraction of dusty galaxies at $z \sim 4$ is expected to be lower than at $z < 4$. Our original 629-object sample does not feature any objects with $E(B-V) > 0.4$, further backing this statement. Even though Bouwens et al. (2012) measure galaxies with $E(B-V) > 0.4$ at $z \sim 4$, the actual fraction in the LBG population seems to be fairly low.

Evolutions in the fraction of dusty LAEs with redshift have already been discussed in the literature. Blanc et al. (2011) study any dust evolution in their $1.9 < z < 3.8$ LAEs sample and find no significant trend. Hagen et al. (2014) study the same sample, and show that there is little anti-correlation, if any, between $E(B-V)$ and redshift. Shapley et al. (2003) do observe such evolution in the range $2 < z < 3.5$, but question the validity given selection effects associated to LBG surveys. We conclude that, even though Lyα emitting galaxies are mostly low-dust objects, there is also a fraction of dustier, low-$W_{Ly\alpha}$ LAEs. It is unclear if we do not observe them as a consequence of our selection, flux limitations, or just because they are very uncommon at $z \sim 4$.

We also recover in Figure 13 Lyα f_{esc} dependence on $E(B-V)$. Similarly to $W_{Ly\alpha}$, f_{esc} anti-correlates with the dust content of galaxies. Our result confirms at $z \sim 4$ a trend already been observed at lower redshift. Atek et al. (2014) perform a f_{esc} study at $z < 0.5$, and recover a similar anti-correlation we show in Figure 13. Hayes et al. (2011) at $z \sim 2.2$, Blanc et al. (2011) at $1.9 < z < 3.8$, and Matthee et al. (2016) at $z \sim 2.23$ also observe the same trends. Verhamme et al. (2008) replicate such trends using radiative simulations of galaxies in the range $2.8 < z < 5$, confirming that qualitative explanations for these observational relations are well supported by theory. We show several best fit relations from the literature

in our lower plot of Figure 13. Considering that our results are dominated by upper limits, all relations are roughly consistent with our measurements. We speculate that the discrepancy towards lower extinction is due to our incompleteness, which only allows us to measure the objects with high escape fractions. Nevertheless, more interesting is the potential redshift evolution suggested by these relations. There is a clear decrease in f_{esc} at high dust contents when going from low to high redshift. If real, this trend could back our previous analysis on the fraction of more dusty LAEs and their evolution as a function of cosmic time. At higher redshift (Verhamme et al. 2008, this work), low-$W_{Ly\alpha}$, very dusty LAEs do not seem to be common, driving the $E(B-V) - f_{esc}$ relation down significantly for $E(B-V) > 0.2$. However, at lower redshift, such objects are actually observed, driving the relation up for $E(B-V) > 0.2$. We go back to the impact that dust may have on Lyα emission studies in Section 6.

4.5. M_*-SFR Sequence

We characterize in Section 4.1 the dependence of Lyα emission on M_*, and in Section 4.2 the dependence on SFR. We find $W_{Ly\alpha}$ and f_{esc} to anti-correlate with both properties. However, these results might not be independent, since M_* and SFR are known to follow a relation at high redshift referred to as the main sequence (Kereš et al. 2005, Finlator et al. 2007, Noeske et al. 2007, Daddi et al. 2007). As we discuss throughout this work, Lyα escape from galaxies is likely to be a process ruled by many elements, such as age of the population, gas column density, extinction, and SFR. Therefore, it is interesting to explore Lyα emission dependence on a 2-parameter space. In this section in particular, we study $W_{Ly\alpha}$ dependence on M_* and SFR. By means of a Bayesian approach, we characterize to which extent $W_{Ly\alpha}$ dependence on both properties is independent.

First, we present in Figure 14 our detections and non-detections location in this plane. Roughly speaking, our detections sample most of the sequence initially covered by our targets. For further insight into how our observations depend on this sequence, we construct four subsamples for visualization. The anti-correlation we observe between characteristic $W_{Ly\alpha}$ and M_* is still clearly observed when comparing the low M_* with the high M_* subsample. Of special interest is the apparent SFR dependence in the range $10^{8.8} - 10^{9.6} M_\odot$, hinting at a non-negligible SFR dependence at fixed M_*.

We now focus on characterizing the significance of these observed trends. To do so, we simultaneously model the exponential profile parameters of Equation (11) as a function of M_* and SFR. We start again with linear expressions of the form

$$A(\mathrm{M_*},\, \mathrm{SFR}) = A_{\mathrm{M_*}} log(\mathrm{M_*}[\mathrm{M_\odot}]) + A_{\mathrm{SFR}} log(\mathrm{SFR}[\mathrm{M_\odot}/yr]) + A_C \quad (24)$$

$$W_0(\mathrm{M_*},\, \mathrm{SFR}) = W_{\mathrm{M_*}} log(\mathrm{M_*}[\mathrm{M_\odot}]) + W_{\mathrm{SFR}} log(\mathrm{SFR}[\mathrm{M_\odot}/yr]) + W_C \quad (25)$$

Once the parametrization and priors are set, we can obtain the posterior distribution of $\{W_n\} = (A_{\mathrm{M_*}}, A_{\mathrm{SFR}}, A_C, W_{\mathrm{M_*}}, W_{\mathrm{SFR}}, W_C)$ using Equation (9). We assume the priors to be

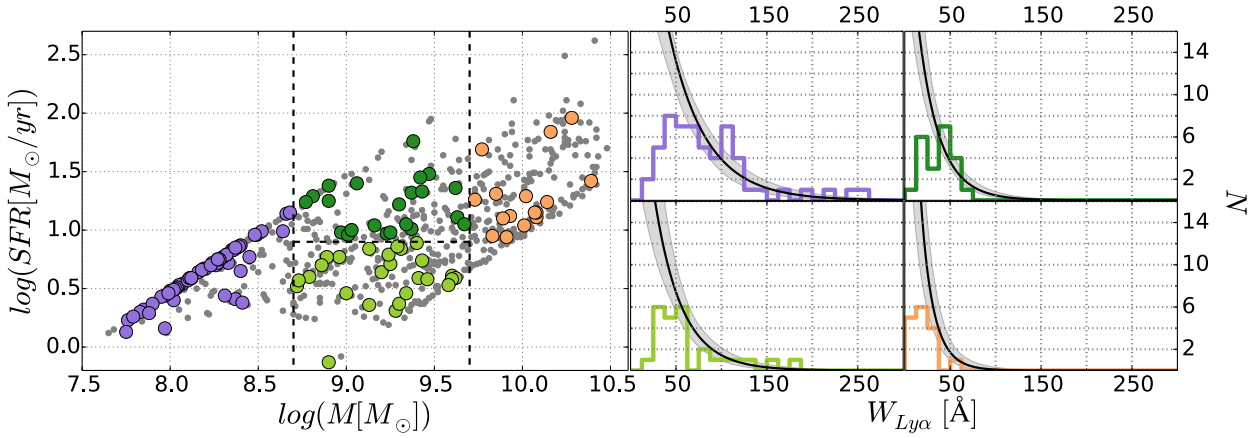

Figure 14 Left: Location of our objects in the M$_*$-SFR plane. Galaxies with detections are plotted with colored circles, and their properties are obtained using spectroscopic redshifts, i.e., M$_*$=M$_*$($z_{Ly\alpha}$) and SFR=SFR($z_{Ly\alpha}$). The grey points correspond to non-detections, for which M$_*$ and SFR are recovered using z_{EAZY}. For visualization, we divide our sample in four regions delimited by dashed black lines. The different detection colors are just for discerning the four subsamples. Right: observed rest $W_{Ly\alpha}$ distributions for the respective M$_*$-SFR subsamples. Equations (27) and (28) state our best solution, which we plot as black lines. The shaded contours represent the 1σ constraints to each subsample according to our Monte Carlo simulations of Equations (24) and (25).

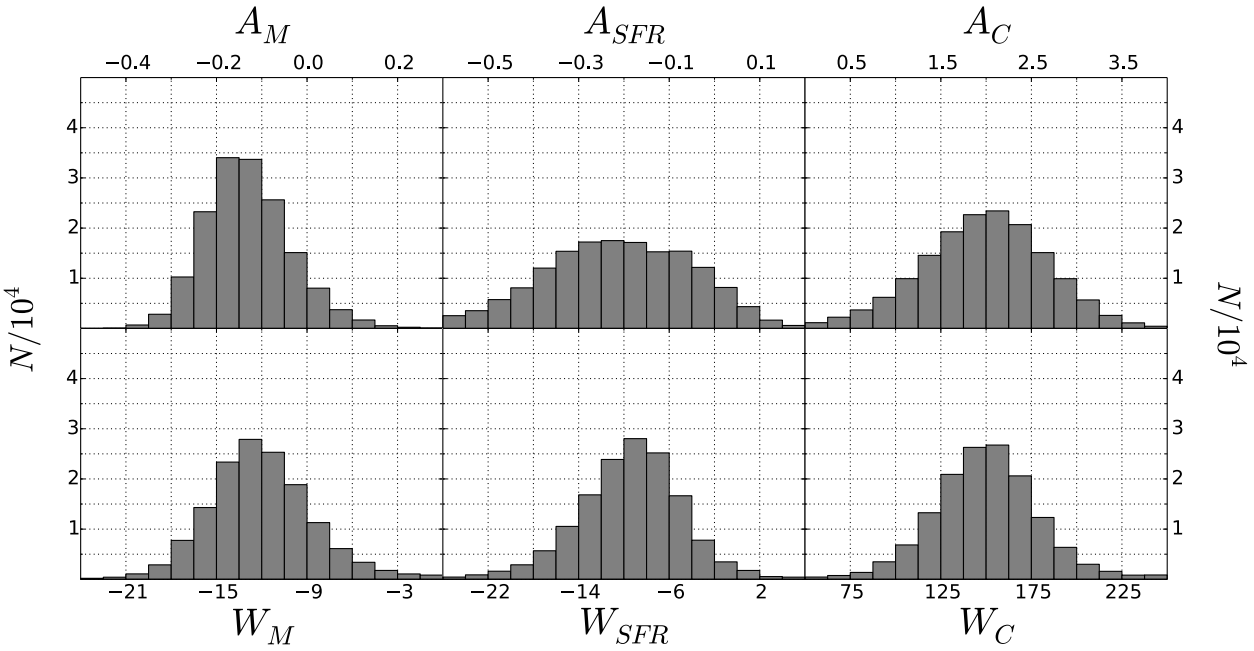

Figure 15 Results for our Monte Carlo simulation on the dependence of Lyα emission on the M$_*$-SFR plane. The six parameters correspond to the coefficients in Equations (24) and (25). The histograms are obtained after collapsing the posterior distribution for every parameter. Left: A, W_0 dependence on M$_*$. Center: A, W_0 dependence on SFR. Right: Additive constants for A and W_0.

independent, which translates into $p(\{W_n\}) \propto \prod p(W_{\rm i})$. Once again, we impose ignorance on the parameters, i.e., we adopt non-informative priors. We study the posterior distribution assuming parameter ignorance in both linear and logarithmic scales. We decide for the first, since logarithmic priors diverge for parameters than can adopt values close to zero ($A_{\rm M_*}$, $A_{\rm SFR}$ and $W_{\rm SFR}$). Hence, our prior is simply a constant $p(\{W_n\})$. Considering that $p(\{F\})$ is just a normalization factor in Equation (9), the posterior distribution of the 6-parameter space can be obtained:

$$p(W_n|\{F\}) \propto \prod_D p(F_{\rm i}|W_n) \prod_{ND} p(F_{\rm i} < F_{\rm i}^*|W_n) \tag{26}$$

The value for parameters A, $\rm W_0$ is constrained in the intervals $A = [0, 1]$ and $W_0 = (0, \infty)$. Still, in the case of extreme $\rm M_*$ and/or SFRs, our linear parameterizations can yield values outside these intervals. As a solution, we just impose A, $\rm W_0$ to saturate outside their corresponding ranges. This condition can be considered simply as an indirect prior on the parameters. For the particular case of $W_0 < 0$, we impose $A = 0$, i.e., the $W_{Ly\alpha}$ distribution is a Dirac Delta.

We recover our best solution from the Monte Carlo simulation and obtain the uncertainties on our parameters using the collapsed distributions. The results are the following:

$$A(\rm M_*,\ SFR) = \\ -0.14^{+.0.1}_{-0.1} log(\rm M_*[M_\odot]) - 0.31^{+0.3}_{-0.1} log(\rm SFR[M_\odot/\mathit{yr}]) + 2.2^{+0.5}_{-1.0} \tag{27}$$

$$W_0(\rm M_*,\ SFR) = \\ -13^{+5.0}_{-2.0} log(\rm M_*[M_\odot]) - 9^{+5.0}_{-5.0} log(\rm SFR[M_\odot/\mathit{yr}]) + 155^{+30}_{-30} \tag{28}$$

We show with more detail in the results for these coefficients in Figure 15. Both A and W_0 decrease with $\rm M_*$ with a significance $\gtrsim 2\sigma$. For SFR, however, the trends are not as eloquent. The fraction A and e-folding scale W_0 seem to decrease with SFR, but only with a significance $\gtrsim 1\sigma$. This is a potentially interesting result, if these trends are real. We observe that, at fixed $\rm M_*$, galaxies with lower neutral gas fractions show higher $W_{Ly\alpha}$. In the scenario of resonant Lyα scattering, higher gas fractions imply a more extended Lyα halo (Verhamme et al. 2006), giving plausible expectations for this correlation. However, at fixed $\rm M_*$, higher SFR objects are older. Hence, at fixed $\rm M_*$, we are observing a regime where age of the population opposes to neutral gas column density, and our trends suggest the effect of the latter is more dominant. We insist, nevertheless, that this trend can be a consequence of statistical fluctuations caused by our limited sample size.

Previous studies of Lyα emission in the $\rm M_*$-SFR sequence typically observe LAEs to sample the high SFR regions of the plane. For example, Hagen et al. (2014) and Vargas et al. (2014) find their LAEs at $z \sim 2-3$ to have ten times greater SFRs than the extrapolation of the main sequence measured for more massive galaxies. At $z \sim 4.5$, Finkelstein et al. (2015) find most of their LAEs to occupy the lowest age locus in the sequence, in partial agreement with the other results at lower redshift. However, Finkelstein et al. (2015) sample is narrow-band selected, similarly to Hagen et al. (2014) and Vargas et al. (2014) samples, which could imprint selection biases on the $\rm M_*$-SFR plane. Interestingly, Finkelstein et al. (2015) also

find a population of older, more evolved LAEs in the main sequence with masses of $10^{10}M_{\odot}$. In contrast, we observe our objects not only to sample this evolved region of the sequence, but the whole plane. This hints that sample selection may play a part in this discrepancy. While our sample is M_* selected, Finkelstein et al. (2015) targets are narrow-band selected, which could lead to characterization of different types of galaxies. We discuss in Section 5 multiple sample selection techniques and their effects on the M_*-SFR sequence, along with implications on the resulting $W_{Ly\alpha}$ distributions.

Ideally, studies of $W_{Ly\alpha}$ and f_{esc} in a 3-parameter space of M_*, SFR, and $E(B-V)$ can yield predictive parameterization of Lyα emission. However, such analysis must be performed on much larger datasets than ours, at least if significant enough results are to be obtained. In future work, our models will include uncertainties in M_*-SFR and explore the M_*-β plane. This will not only allow us characterize the M_*-SFR-β space in more detail, but also better assess the existence of this SFR dependence at fixed M_*.

5. On Lyα Dependence on Sample Selection

5.1. LBG Samples

The Lyman Break selection technique has proven to be a very efficient method for detecting high-redshift galaxies (e.g. Steidel 1996, Shapley et el. 2003, Stark et al. 2010, Ono et al. 2012). The fact that the Lyman Break is on the optical region of the observed spectrum for galaxies at redshift $z \geqslant 3$ allows for efficient detection from ground telescopes. By only requiring the use of, in principle, two broadband filters, the amount of galaxies that can be detected in a deep single exposure is significant. Still, for avoiding aliasing with the Balmer break, there are unavoidable biases associated to this technique. First, galaxies with no prominent Lyman break are, by construction, not detected. As a consequence, passive and heavily-extincted galaxies are underrepresented in LBGs surveys. Second, these surveys also impose color restrictions to the slope of galaxies, further increasing the selection towards bluer UV objects. Third, this technique is limited by the M_{UV} sensitivity of the survey, creating low SFR incompleteness. In the case of Lyα emission, this observational limit can have a significant downside. As shown in Section 4.2, there is a clear anti-correlation between SFR and $W_{Ly\alpha}$, leading to the possibility that Lyα studies in LBG samples are missing the highest $W_{Ly\alpha}$ galaxies in the universe. This possibility becomes increasingly likely when comparing LBGs and narrow-band samples results. The $W_{Ly\alpha}$ observed in narrowband surveys are systematically higher (e.g. Zheng et al. 2014), presumably due to not requiring a continuum detection.

Our dataset and results can be useful for characterizing the effects selection techniques can have on Lyα emission. Even though it is possible to identify these selections in our sample, any comprehensive analysis must consider our galaxy selection procedure. Since our sample follows neither the M_* or M_{UV} functions of the galaxy population, correcting requires assumtion of M_* and/or M_{UV} distributions. However, we can still simulate Lyα emission samples on CANDELS galaxies using our results on $W_{Ly\alpha}$ dependence on the M_*-SFR plane

Table 2. CANDELS bands used for LBG selection

Band	AEGIS	COSMOS	GOODS-N	GOODS-S	UDS
U	U_{383}	U_{383}	U_{359}	U_{375}	u_{383}
B	G_{487}	B_{445}	F435W	F435W	B_{441}
V	F606W	F606W	F606W	F606W	F606W
i	I_{768}	I_{768}	F775W	F775W	i_{766}
z	Z_{887}	Z_{887}	F850LP	F850LP	z_{906}
Y	J_{104}	Y_{102}	-	-	-
J	F125W	F125W	F125W	F125W	F125W

given by Equations (27) and (28). In this section, we describe our simulation of a $z \sim 4$ LBG sample from 3DHST catalogs. We then compare the properties of LBGs with the parent distribution, focusing on the M_*-SFR plane. We finally derive simulated $W_{Ly\alpha}$ distributions for the samples and conclude on the effects of LBG selection at $z \sim 4$.

LBGs at $z \sim 4$ are typically selected using B-dropouts and imposing color selections on redder filters. We summarize the filters we use in Table 2. As we are simulating this survey in CANDELS, our detection limits are given by the depth of their images. We detail the limiting magnitudes for each field in Table 3. For the $z \sim 4$ LBG selection, we adopt the same methodology applied in Bouwens et al. (2012):

$$\begin{aligned} &\text{B-dropouts: } (B - V > 1.1) \wedge \\ &(B - V > (V - z) + 1.1) \wedge (V - z < 1.6) \end{aligned} \tag{29}$$

After color selection, we impose 4.5σ detections in the i filter, to which we now refer to as the detection band. Finally, we also require non-detections in bands bluer than the Lyman Break, which we impose by requiring our targets to have $< 1\sigma$ significance in the U band. Before performing any analysis on CANDELS LBG sample, we run EAZY and FAST according to our prescriptions (see Section 1.3). Of the total 1232 CANDELS galaxies that comply with our selection, 324 classify as low redshift interlopers ($z_{EAZY} < 3$), according to our outputs. We check EAZY 3σ constraints for these interlopers and find most to have $z_{99} < 4$. Hence, we find a low redshift contamination in $z \sim 4$ CANDELS LBGs higher than the typically reported number of $\sim 10\,\%$ (Bouwens et al. 2007). From now on, we remove these presumed contaminants and work with a 908 $z \sim 4$ LBG sample.

Using our outputs, we plot the SFR-M_* plane for CANDELS LBGs in Figure 16. For contrast, we construct two $3.5 < z < 4.5$ CANDELS samples: all galaxies with $M_* > 10^{7.5}$ and all i-band detected $M_* > 10^{7.5}$ objects. As seen in Figure 16, galaxies with high M_* and high SFRs are not classified as LBGs, as a consequence of color selection. These massive galaxies with weak Lyman Break and/or redder UV slopes do not comply with the LBG selection criteria and are, therefore, dismissed. These effects are normally related to prominent UV extinction consequence of high dust presence within galaxies (Quadri et al. 2007).

The results we present in this work, especially Lyα emission dependence on the M_*-SFR sequence (Section 4.5), suggest this selection has no relevant consequences on the shape of high-$W_{Ly\alpha}$ end of the distribution. However, even if the shape does not change, the fraction of LAEs above an $W_{Ly\alpha}$ threshold might. To verify this, we use our results from Section 4.5 to simulate $W_{Ly\alpha}$ distributions for each sample and plot our results in Figure 16. Indeed, when

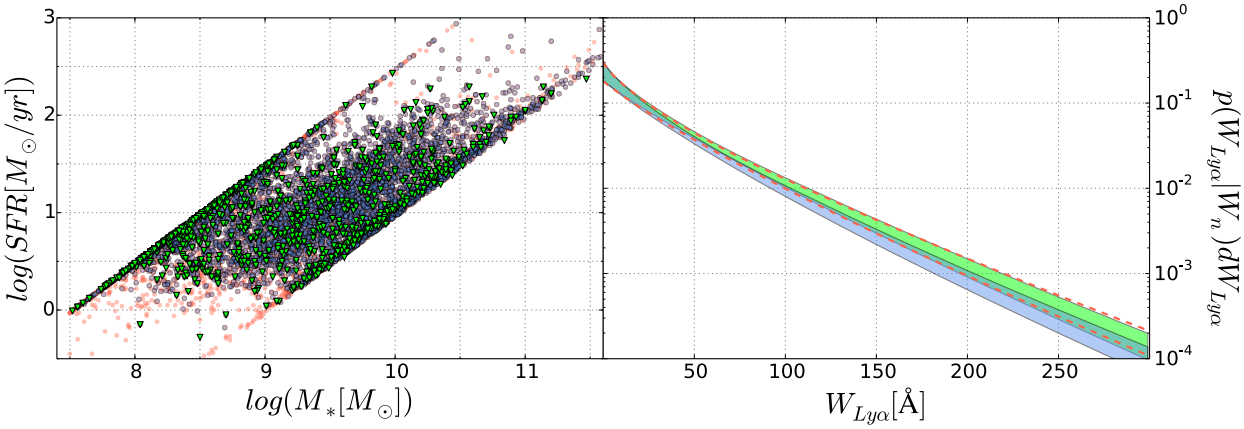

Figure 16 Left: M_*-SFR sequence for $3.5 < z < 4.5$ CANDELS galaxies with $M_* > 10^{7.5} M_\odot$ (red). Galaxies that satisfy i $> 4.5\sigma$ are shown in blue. Our LBG selected galaxies from CANDELS are plotted as triangles (green). The detection band requirement in LBG selection removes low M_* and low SFR objects. The color selection removes high M_* and high SFR galaxies. Right: Simulated $W_{Ly\alpha}$ distributions for the three samples according to our Lyα modeling on the M_*-SFR plane. The distributions shown correspond to CANDELS $3.5 < z < 4.5$ objects (red), LBGs (green), and detection band selected objects (blue). The curves enclose 1σ uncertainties on the probabilities, which are a result of our simulation and sample distributions. We conclude that color selections on LBGs increase the probability of high $W_{Ly\alpha}$ in comparison to a purely i-band selected sample. However, detection band requirements lead to the exclusion of very low mass objects, bringing LBGs $W_{Ly\alpha}$ distribution closer to the parent CANDELS sample. We remark that this insight is valid down to the M_* range we are capable of probing with CANDELS.

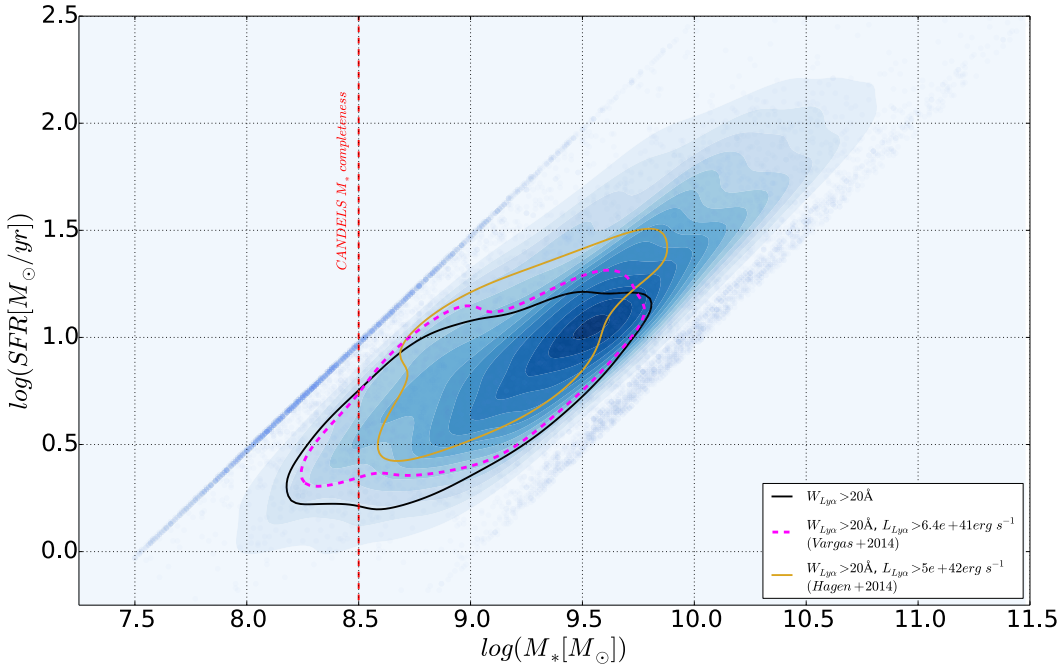

Figure 17 We show in this plot the selection effects produced by narrow-band selected samples of LAEs. As parent sample, we show as blue contours the M_*-SFR sequence for $3.5 < z < 4.5$ CANDELS galaxies with $M_* > 10^{7.5} M_\odot$. For clarity, we remove galaxies in the youngest and oldest age tracks (blue datapoints). We simulate the intrinsic Lyα population at $z \sim 4$, and show as black lines 1σ contours for the resulting sample after performing an $W_{Ly\alpha} > 20$Å selection. Selections in $W_{Ly\alpha}$ are more likely to remove high M_* and high SFR LAEs. We also find luminosity selections to remove mostly low SFR objects. We combine both observational constraints, and show the region sampled by the surveys described in Vargas et al. (2014; magenta) and Hagen et al. (2014; brown), assuming the same rest $W_{Ly\alpha}$ and Lyα luminosity selections at our redshift. We predict the deeper observations from Vargas et al. (2014) to measure lower SFR LAEs than Hagen et al. (2014). The red line shows CANDELS M_* completeness at $z \sim 4$ (Duncan et al. 2014).

comparing LBGs to the CANDELS detection band limited sample, the fraction of emitters towards high $W_{Ly\alpha}$ increases. Since color selection for LBGs at this redshift do not heavily depend on the survey (Stark et al. 2009, González et al. 2011, Bouwens et al. 2014), this selection bias is not likely to dominate results on the fraction of LAEs at $z \sim 4$. However, when tracing evolution in the fraction of LAEs in LBG samples as a function of redshift, dismissing this effect is not straightforward. For example, color selections at $z \sim 5, 6$ are not as well standardized (Stark et al. 2009, Bouwens et al. 2014), which can induce discrepancies in the fraction of LAEs from survey to survey. One more potential selection bias has to be considered, however. Depending on the depth in the detection band, the fraction of low M_* galaxies composing the sample can vary significantly. As a consequence, deeper surveys can potentially recover a higher fraction of LAEs at high $W_{Ly\alpha}$ (Shapley et al. 2003, Stark et al. 2010). To deal with this selection, it has been proposed to compare the fraction of LAEs at a given UV luminosity (Stark et al. 2011, Mallery et al. 2012, Schenker et al. 2012, Ono et al. 2012). Our results in this paper further emphasize the need for comparing galaxies of the same luminosity and properly characterize completeness when studying LAEs fraction evolution in cosmic time.

5.2. Narrowband Samples

Lyα emitting galaxies can be also selected using narrow-band imaging or blind spectroscopy (e.g. Gronwall et al. 2007, Ouchi et al. 2008, Adams et al. 2011). By establishing an $W_{Ly\alpha}$ detection threshold between the narrow- and broad-band flux measurement, this technique allows for efficient line emitters selection. Since only line detection is required, such surveys can trace fainter objects than the LBG technique, possibly leading to selection of the youngest and faintest galaxies at high-redshift. However, even though most narrow-band measurements of LAEs are followed up by spectroscopy, any sample selection effects induced by the narrow-band technique are already present in the sample. The $W_{Ly\alpha}$ threshold used for detection, which is determined by the ability to separate low redshift interlopers from high-redshift LAEs, can adopt a wide range of values. Depending on redshift, observer frame thresholds translate into different rest frame $W_{Ly\alpha}$ cuts. For instance, Vargas et al. (2014) select sources with $W_{Ly\alpha} > 20$Å at $z \sim 2.1$, while Zheng et al. (2014) select $W_{Ly\alpha} > 9$Å at $z \sim 4.5$. If $W_{Ly\alpha}$ selections induce important biases on galaxy samples, comparison of different surveys is not straightforward. In this section, we explore the effects such selections have in the population of LAEs, focusing on the M_*-SFR plane. We remark that the insights we present here are based on Equations (27) and (28), as any analysis performed directly on our detections is affected by our M_*-SFR dependent completeness.

We show in Figure 17 the outcome of $W_{Ly\alpha}$ and line flux selections in the M_*-SFR sequence. The blue datapoints show $3.5 < z < 4.5$ CANDELS galaxies with M$> 10^{7.5}$M$_\odot$. For clarity, however, we concentrate our analysis on the main sequence (blue contours), removing objects with the youngest and oldest ages. This approach allows our results to be dominated by objects optimally fitted by our FAST executions. We simulate a Lyα sample, and show the region where $W_{Ly\alpha} < 20$Å objects lie (1σ contours) in black. As confirmed by this plot, galaxies with higher characteristic $W_{Ly\alpha}$ (i.e., lower M_*) are more likely to be selected by narrow-band samples. Hence, $W_{Ly\alpha}$ cuts imprint a bias towards low M_*, SFRs galaxies. After

narrow-band outcomes are used for sample selection, spectroscopic observations follow. However, these follow-ups of Lyα emitting galaxies are flux limited, just like the ones presented in this work. However, as we consider completeness in our modeling and simulations of Section 4.5, we can still make inferences on flux limited studies. We use our Monte Carlo simulation outputs to also assess the effects of line luminosity selections. Using the corresponding M_*, SFR, and f_{UV} for every object, we obtain the probability of $L_{Ly\alpha} > L^*_{Ly\alpha}$ for every galaxy. We find flux selections to primarily select high-SFR objects.

As we see in Figure 17, flux limitations and $W_{Ly\alpha}$ cuts can produce selection biases in the M_*-SFR sequence. The results we obtain suggest narrow-band surveys are biased to mostly select low M_*, high SFR targets. This trend has already been observed in the $z \sim 2.1$ narrow-band survey of Guaita et al. (2010). In Vargas et al. (2014), they analyze the properties of their LAEs and find them to lie above the M_*-SFR relation. In an interesting comparison, Hagen et al. (2014) plot their $z \sim 2-3$ LAEs in the M_*-SFR relation along Vargas et al. (2014) results in Figure 10 of their paper. They find both survey LAEs to lie above the main sequence. Hagen et al. (2014) objects are part of the HETDEX Pilot Survey (Adams et al. 2011), which is restricted to fluxes $> 7 - 10 \times 10^{-17}$ erg cm^{-2} s^{-1} and $W_{Ly\alpha} > 20$Å. In contrast, Guaita et al. (2010) flux depth is higher than HETDEX, reaching 2×10^{-17} erg cm^{-2} s^{-1}, while also selecting sources with $W_{Ly\alpha} > 20$ Å. Therefore, Hagen et al. (2014) are comparing two $W_{Ly\alpha}$ selected surveys with different line flux depths. Our results in this section would, then, predict Vargas et al. (2014) LAEs to sample lower M_* and SFRs because they go deeper (Figure 17). This is in fact the pattern observed in Figure 10 of Hagen et al. (2014), i.e., we can qualitatively reproduce their comparison with our simulations from the results of Section 4.5.

Our findings in this section suggest that narrow-band selected samples are biased to study low M_*, high SFR galaxies, with the effect depending on the $W_{Ly\alpha}$ cut and flux limitations of spectroscopic follow-up. This gives an explanation for previous observations of this trend at $z \sim 2-3$ (Vargas et al. 2014, Hagen et al. 2014). However, it is not enough to explain claims of LAEs lying above the M_*-SFR sequence. We cannot claim whether this offset is real or the position of the M_*-SFR relation at $z \sim 2-3$ is not well understood. However, recent studies have found not such offset at $z \sim 2.23$ (see Matthee et al. 2016). They use an Hα selected sample for their analysis, hinting at selection effects or SED SFR measurements playing a role in this offset. Furthermore, such trend has not been observed at higher redshift (Finkelstein et al. 2015 and this work).

6. Inferences on $4 < z < 7$ LAEs fraction

Probably the most important use for Lyα emission is tracing the neutral hydrogen fraction in the IGM. Several studies over the last decade have constrained the fraction of Lyα emitting galaxies as a function of redshift (Stark et al. 2010 and 2011, Schenker et al. 2012, Ono et al. 2012, Tilvi et al. 2014), bringing us closer to the goal of constraining the epoch of reionization. The fraction of LAEs is fairly well understood at $z < 5$ (Stark et al. 2011, Cassata et al. 2015), with most efforts nowadays focusing on $z \sim 7, 8$ (Ono et al. 2012, Tilvi et al. 2014, Furusawa et al. 2016). From this standpoint, there is not much our $z \sim 4$ measurements can

Table 3. Summary of B-, V-, i-, and z-dropout samples

Field	Science Area[a] (arcmin^2)	B-dropouts N[d]	B-dropouts Limit[b]	V-dropouts N[d]	V-dropouts Limit[b]	i-dropouts N[d]	i-dropouts Limit[b]	z-dropouts N[d]	z-dropouts Limit[b]	CANDELS[c] Limit[b]
AEGIS	192.4	83	i < 27.0	112	$z < 26.2$	18	$z < 26.2$	8	$J < 24.8$	$H < 24.3$
COSMOS	183.9	91	i < 27.0	53	$z < 26.0$	16	$z < 26.0$	10	$J < 25.7$	$H < 24.0$
GOODSS-N	157.8	315	i < 26.9	137	$z < 26.7$	29	$z < 26.7$	-	-	$H < 24.3$
GOODSS-S	171.0	264	i < 26.9	93	$z < 26.5$	45	$z < 26.5$	-	-	$H < 24.5$
UDS	191.2	155	i < 26.7	72	$z < 25.9$	10	$z < 25.9$	-	-	$H < 24.3$
Total	896.3	908	-	467	-	118	-	18	-	-

[a]Corresponding science area from 3DHST photometric catalogs.

[b]3DHST median 5σ limit from galaxy catalogs.

[c]Detection requirements for CANDELS objects.

[d]Numbers include only galaxies for which our FAST outputs associate values for M_* and SFR.

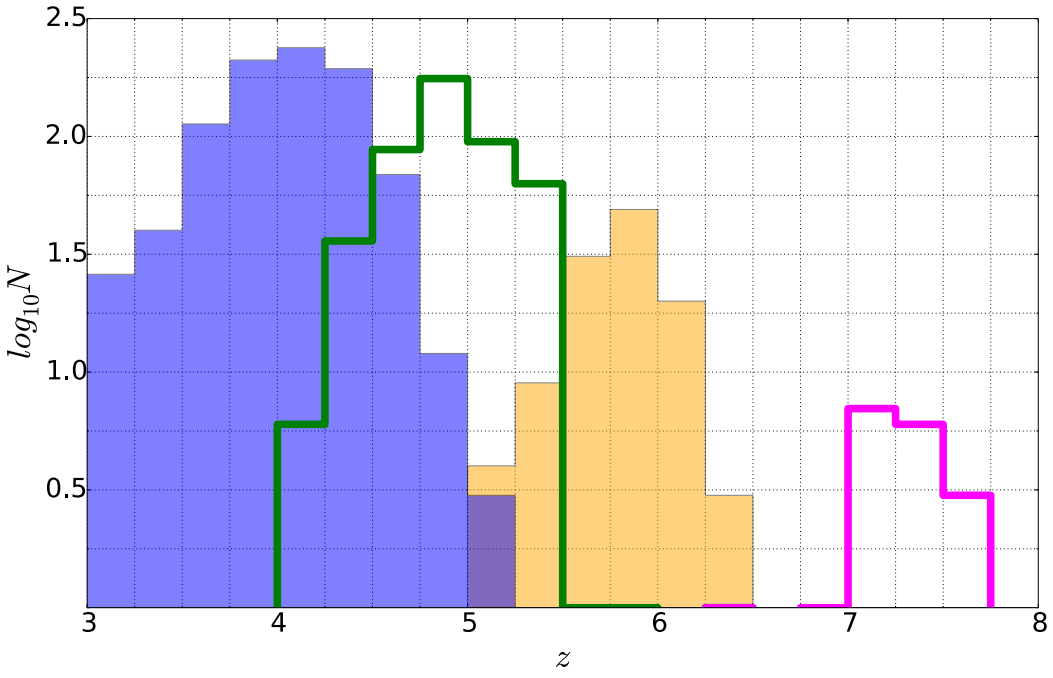

Figure 18 Photometric redshift histograms for our LBG samples at $z \sim 4, 5, 6$ and 7 (blue, green, yellow, and magenta, respectively). The numbers shown are obtained after performing color cuts, detection band S/N selections, and correction for presumed low- and high-redshift contaminants ($\lesssim 15\,\%$). We also impose LBGs to have low S/N on filters bluer than the corresponding Lyman break.

contribute to such campaign. However, our characterization of $W_{Ly\alpha}$ in the M_*-SFR plane can be used to simulate $W_{Ly\alpha}$ distributions at higher redshift. In this section, we use V-, i-, and z-dropouts from CANDELS to simulate the $W_{Ly\alpha}$ distribution for $z \sim 5, 6$, and 7 LBGs. This allows us to simulate the fraction of LAEs up to $z \sim 7$, providing the first semi-analytical constraint to the expected fraction of Lyα emitting galaxies towards the reionization epoch. We stress that our result is based on assuming the same $z \sim 4$ $W_{Ly\alpha}$ dependence on M_* and SFR at $z \sim 5, 6$ and 7. Hence, this analysis does not account for any effects related to merger fraction, IGM opacity, and/or changes in the condition of the ISM as a function of cosmic time.

We use CANDELS photometry for the selection of dropouts. For $z \sim 4$, we use the same LBG sample described in Section 5.1. For V-, i-, and z-dropouts, we use the criteria from Bouwens et al. (2012):

$$\text{V-dropouts: } [(V - \text{i} > 0.9(\text{i} - z) + 1.5) \vee (V - \text{i} > 2)] \wedge (V - \text{i} > 1.2) \wedge (\text{i} - z < 0.8) \tag{30}$$

$$\text{i-dropouts: } (\text{i} - z > 1.3) \wedge (z - J < 0.9) \tag{31}$$

$$\text{z-dropouts: } (z - Y > 0.7) \wedge (Y - J < 0.8) \wedge (z - Y > 1.4(Y - J) + 0.42) \tag{32}$$

Similarly to our B-dropouts selection, we impose significant S/N in the detection band. For $z \sim 5$ and 6 dropouts, we require 3σ detection in the z band. For $z \sim 7$ dropouts, we impose at least a 3σ detection in the J band. We also require the candidates to have $< 1.5\sigma$ significance on every band bluer than the corresponding Lyman Break. Lastly, we remove from our dropouts all galaxies with photometric redshifts lower than 4, 5, and 6, respectively. We also had to remove four presumed $z \sim 7$ interlopers from our $z \sim 6$ LBG sample. The actual bands we use for each field are summarized in Table 2. We show in Table 3 the CANDELS limiting magnitudes and number counts for our LBG selection, and we present in Figure 18 the photometric redshift distribution for the final dropouts. Once again, we run EAZY and FAST according to our prescriptions (see Section 1.3) for the results presented in this section. We then simulate the $W_{Ly\alpha}$ distributions for the four LBG samples. To do so, we use the M_* and SFR of each object to draw $W_{Ly\alpha}$ distributions following Equations (27) and (28). This methodology allows us to simulate the $W_{Ly\alpha}$ distribution for the number of galaxies in every dropout subsample, properly taking into account the uncertainties on each M_*-SFR domain.

It is essential to mention the role played by the M_{UV} distribution of samples when recovering the fraction of Lyα emitting objects. LBGs with M_{UV} distributions dominated by fainter objects can yield higher fractions of Lyα, as we note in Section 4.3. For instance, Stark et al. (2010) find an increase in the fraction of Lyα from $z \sim 4$ to 6 (Figure 15 from their paper). However, the M_{UV} distribution for each of their dropout samples (refer to their Figure 6) can mimic this evolution. In Stark et al. (2011), Lyα fraction comparison is performed using narrower M_{UV} ranges. In fact, the existence of this increase in the fraction is quite dependent on this range and the $W_{Ly\alpha}$ threshold imposed. Furthermore, we show in this paper

how the tail of the $W_{Ly\alpha}$ distribution can depend heavily on galaxy properties and selection. As a consequence, we remark the importance of sample characterization and completeness when studying Lyα fraction evolution. In recent work, characterization has been performed for LBGs with $M_{UV} < -20.25$ and $M_{UV} > -20.25$ (Stark et al. 2011, Schenker et al. 2012, Ono et al. 2012). Other studies cover the same luminosity ranges without necessarily using LBGs (Tilvi et al. 2014, Cassata et al. 2015). We find $z \sim 6, 7$ samples from CANDELS to have questionable completeness even for $M_{UV} < -20.25$. As a consequence, we question any inferences based on LBG samples fainter than $M_{UV} = -20.25$, as the uncertainties associated are too large.

We now focus on our analysis of LBG samples in the range $-21.75 < M_{UV} < -20.25$. We present the $4 < z < 7$ Lyα fraction for such galaxies in Figure 19. To this end, we select using the thresholds $W^*_{Ly\alpha}$= 25 Å and $W^*_{Ly\alpha}$=55 Å (Stark et al. 2011). Aside from Stark et al. (2011), we also include the results from Mallery et al. (2012), Ono et al. (2012), Tilvi et al. (2014), and Cassata et al. (2015) for the same M_{UV} range (Figure 19). Our results are in perfect agreement with reported trends in the literature up to $z \sim 6$, and within an agreement of 1σ at $z \sim 7$. The overall trend is an increase in Lyα fraction up to $z \sim 5$, which flattens at $z \sim 6$ and 7. Blanc et al. (2011) and Cassata et al. (2015) hint that changes in f_{esc} as a function of redshift are responsible for this growth in Lyα fraction. Such changes in f_{esc} are possibly related to evolution in the amount of dust within galaxies as a function of redshift (see our Section 4.4).

Interestingly, we observe a plateau in the fraction from redshift $z \sim 5$ to $z \sim 7$. Careful analysis of our dropouts reveals the reason behind this flat regime. Galaxies with $-21.75 < M_{UV} < -20.25$ that classify as $z \sim 5$ LBGs are mostly $10^8 - 10^{10}$M$_\odot$ objects with a wide range of SFRs. Similarly, $z \sim 6, 7$ counterparts are primarily $10^{9.5}$M$_\odot$ galaxies, which is just in the middle of the previous range mentioned. This is consistent with the fraction remaining constant. These results suggest that extrapolation of an increase in the fraction up to $z \sim 7$ (Stark et al. 2011, Ono et al. 2012) is not correct. Hence, even when the robustness of this plateau is questionable, it raises some concerns about the significance of previously reported $z = 7$ drops in the fraction of LAEs (Stark et al. 2010, Ono et al. 2012).

Later results from Tilvi et al. (2014) estimate reionization to take place at $z > 7$. Their measurements seem to lie below our trend, but without any constraints we can set beyond $z \sim 7$, we cannot strictly compare. However, their results are not based on LBGs, but on a photometric redshift selected sample. Therefore, according to our results from Section 5.1, they include more high M$_*$ objects than a standard LBG sample. Hence, their fractions are bound to be lower than LBG derived constraints. In future work, we will estimate such fractions using photometric redshift selected samples. In any case, since reionization is predicted to happen so rapidly (Finlator et al. 2009), claims of its occurrence at $z \sim 8$ are by no means contradictory to our results. Even more, systematic detection of Lyα emitting galaxies in the range $z \sim 7 - 9$ (Zitrin et al. 2015, Furusawa et al. 2016, Stark et al. 2016) still question whether reionization is in place at $z \sim 8$.

To explain the evolution of Lyα fraction as a function of cosmic time, we speculate on a possible qualitative scheme. At high-redshift, Lyα emission escape from galaxies is likely to be mostly ruled by neutral gas scattering, with negligible dust absorption. Once this emission

reaches the outskirts of galaxies, however, further photon escape is tied to the transparency of the IGM. Hence, we argue that observed Lyα fractions at $z > 6$ are consequence of both, ionization of the universe and Lyα escape fraction in minimal presence of dust. At $z < 6$, however, Lyα escape from galaxies is strongly dependent on dust fractions. The build up of Lyα absorbing particles at lower redshift defines a new regime, dominated by a decrease in Lyα fraction as the universe becomes older (Cassata et al. 2015). The results we find in the dependence of Lyα escape fraction on $E(B-V)$ and redshift (Figure 13) are also consistent with this picture. Furthermore, the existence of bright, dusty, low-$W_{Ly\alpha}$ LAEs seemingly becomes significant just for $z < 4$ (Section 4.4). In summary, such a qualitative scheme we propose adds some complexity to the determination of drops in LAEs fraction, especially at $z > 5$, since we can no longer work on a baseline set by the extrapolation of a lower redshift regime.

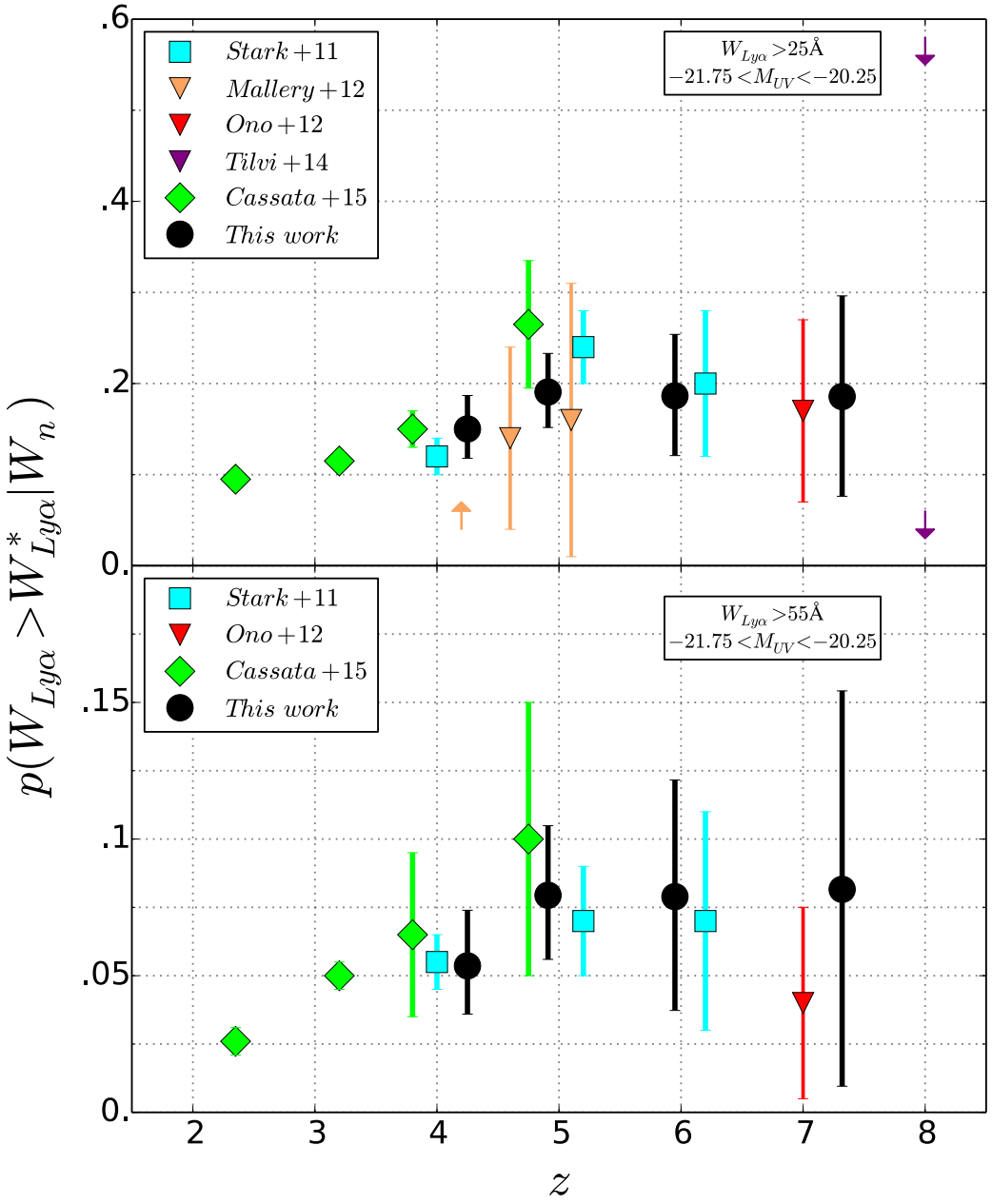

Figure 19 Fraction of Lyα emitting galaxies in $-21.75 < M_{UV} < -20.25$ LBGs as a function of redshift. We show the fractions for $W_{Ly\alpha} > 25$Å(top) and $W_{Ly\alpha} > 55$Å(bottom). Plotted datapoints include observational constraints from Stark et al. (2011; cyan), Mallery et al. (2012; brown), Cassata et al. (2015; green), Ono et al. (2012; red), and Tilvi et al. (2014; orange). The datapoint from Ono et al. (2012) is a compilation of results from Fontana et al. (2010), Pentericci et al. (2011), Vanzella et al. (2011), Schenker et al. (2012), and Ono et al. (2012). The two datapoints from Tilvi et al. (2014) come from completeness corrected measurements (higher fraction) and framework from Treu et al. (2013; lower fraction). Our constraints (black) at $z \sim 4$ come from the Lyα trends we observe in the M$_*$-SFR plane. Our datapoints at $z \sim 5, 6$ and 7 are obtained from simulations assuming the same $z \sim 4$ relation. Some datapoints are slightly offset in redshift for visualization. The results from Tilvi et al. (2014) and Cassata et al. (2015) are obtained from photometric redshift selected samples, i.e., they are complimentary to LBGs.

Conclusion

In this work, we introduce an exhaustive analysis of Lyα emission at $z \sim 4$. To this end, we M_* select 629 galaxies from the CANDELS survey, allowing us to study Lyα emission over a diverse and heterogeneous galaxy sample. We conduct spectroscopic observations of our targets with the M2FS, a multi-fiber spectrograph at the Clay 6.5m telescope. We then use a Bayesian approach for proper statistical handling of our results. By means of this framework, we are capable of characterizing Lyα emission in the high-redshift galaxy population. In summary, our conclusions are the following.

1. We introduce a Bayesian methodology to measure the $W_{Ly\alpha}$ distribution considering the completeness in fluxes measured and the uncertainties in both, spectroscopy and photometry. We also use this framework to compare multiple $W_{Ly\alpha}$ distribution models, and we conclude that the exponential profile is the most adequate to represent our measurements. We also combine this approach with Monte Carlo simulations for robust modeling of observed trends. Combination of all these features allows us to properly state the significance and uncertainties in our results.

2. Our measured $W_{Ly\alpha}$ and f_{esc} strongly anti-correlate with M_*. We associate these trends to the higher dust fraction and gas mass in more massive galaxies, which boost the scattering and absorption of Lyα photons. We model both exponential $W_{Ly\alpha}$ distribution parameters (A, W_0) using linear relations dependent on M_*, and find them both to anti-correlate with M_*. Our modeling is also capable of reproducing our observed $W_{Ly\alpha}$ distributions when binning the sample (Figure 7).

3. We also explore the dependence of Lyα emission on M_{UV}. We find $W_{Ly\alpha}$ and f_{esc} to be typically higher for UV fainter objects, which has been previously observed in the literature (Stark et al. 2010, Schaerer et al. 2011). Since $z \sim 4$ galaxies seem to follow a $M_* - M_{UV}$ sequence, this result is consistent with our M_* trends. This confirms that the role played by higher dust fraction and gas mass overcomes the greater Lyα photon output from brighter UV objects. We also observe $W_{Ly\alpha}$ and f_{esc} to anti-correlate with SFR and $E(B-V)$, confirming the qualitative Lyα escape scheme presented.

4. The relatively uniform location of our targets in the M_*-SFR plane allows us to characterize the $W_{Ly\alpha}$ distribution in this space. Once more, we use a Bayesian framework and Monte Carlo simulations to obtain linear representations of exponential $W_{Ly\alpha}$ parameters. Even when our sample size restricts the significance of our results, we still find a clear anti-correlation between M_* and $W_{Ly\alpha}$. We also recover an anti-correlation between SFR and $W_{Ly\alpha}$ at fixed M_*, although the significance of this result is $< 2\sigma$ (Figure 15). This characterization

is not only limited to the recovery of correlations involving M_* and SFR, however. We also use these results to simulate $W_{Ly\alpha}$ distributions for M_*, SFR galaxy samples (see below).

5. Using CANDELS, we mimic a $z \sim 4$ LBG survey based on B-dropouts. We find the color selection imposed by the LBG technique to leave out high M_*, high SFR galaxies. We simulate $W_{Ly\alpha}$ distributions for the LBG sample and verify that LBGs $W_{Ly\alpha}$ distribution features less low $W_{Ly\alpha}$ objects than a purely detection band limited sample. However, more worrying is the potential bias induced by depth in the detection band. Deeper LBG surveys are bound to include lower M_* galaxies in their sample, increasing the probability of measuring high $W_{Ly\alpha}$. This result highlights the importance of comparing surveys with similar sensitivities and M_{UV} distributions.

6. Our results on the M_*-SFR plane also allow us to explore the effects flux and $W_{Ly\alpha}$ limits induce on Lyα narrow-band surveys. We find $W_{Ly\alpha}$ cuts to bias samples towards low M_* objects, while flux limitations seem to leave out low SFR galaxies. If real, these trends can contribute to explain reports of LAEs location in the main sequence (Vargas et al. 2014, Hagen et al. 2014, Finkelstein et al. 2015).

7. We generate LBG samples at $z \sim 4, 5, 6$ and 7 from CANDELS to simulate $W_{Ly\alpha}$ distributions at higher redshift. Assuming the same M_*-SFR dependence we find at $z \sim 4$, we estimate the fraction of Lyα emitting galaxies in LBG samples above an $W_{Ly\alpha}$ threshold. Our findings are consistent with observational measurements in the literature suggesting an increase in the fraction up to $z \sim 5$ (Stark et al. 2011, Cassata et al. 2015), while hinting at a plain regime towards higher redshift. This result constitutes the first semi-analytical constraint to Lyα fractions at $z \sim 7$, replacing extrapolations of lower redshift regimes. Since our findings are consistent with fractions measured at $z \sim 7$, claims of drops at this redshift need to be more thorough.

G.O. was supported by CONICYT, Beca Magíster Nacional 2014, Folio 22140924. G.B. is supported by CONICYT/FONDECYT, Programa de Iniciación, Folio 11150220. This work is based on observations taken by the 3D-HST Treasury Program (GO 12177 and 12328) with the NASA/ESA HST, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555. This research has made use of the NASA/IPAC Extragalactic Database (NED), which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. This paper includes data gathered with the 6.5 meter Magellan Telescopes located at Las Campanas Observatory, Chile.

Bibliography

Adams, J. J., Blanc, G. A., Hill, G. J., et al. 2011, ApJS, 192, 5

Atek, H., Kunth, D., Schaerer, D., et al. 2014, A&A, 561, A89

Balestra, I., Mainieri, V., Popesso, P., et al. 2010, A&A, 512, A12

Bennett, C. L., Larson, D., Weiland, J. L., et al. 2013, ApJS, 208, 20

Blanc, G. A., Adams, J. J., Gebhardt, K., et al. 2011, ApJ, 736, 31

Bond, N. A., Gawiser, E., Guaita, L., et al. 2012, ApJ, 753, 95

Boquien, M., Buat, V., & Perret, V. 2014, A&A, 571, A72

Bouwens, R. J., Illingworth, G. D., Franx, M., & Ford, H. 2007, ApJ, 670, 928

Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2012, ApJ, 754, 83

Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2014, ApJ, 793, 115

Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, ApJ, 686, 1503-1513

Calzetti, D., Kinney, A. L., & Storchi-Bergmann, T. 1994, ApJ, 429, 582

Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682

Caminha, G. B., Karman, W., Rosati, P., et al. 2015, arXiv:1512.05655

Cassata, P., Le Fèvre, O., Garilli, B., et al. 2011, A&A, 525, A143

Cassata, P., Tasca, L. A. M., Le Fèvre, O., et al. 2015, A&A, 573, A24

Chabrier, G. 2003, PASP, 115, 763

Charlot, S., & Fall, S. M. 1993, ApJ, 415, 580

Chonis, T. S., Blanc, G. A., Hill, G. J., et al. 2013, ApJ, 775, 99

Ciardullo, R., Zeimann, G. R., Gronwall, C., et al. 2014, ApJ, 796, 64

Daddi, E., Dickinson, M., Morrison, G., et al. 2007, ApJ, 670, 156

Dijkstra, M., & Kramer, R. 2012, MNRAS, 424, 1672

Dressler, A., Martin, C. L., Henry, A., Sawicki, M., & McCarthy, P. 2011, ApJ, 740, 71

Dressler, A., Henry, A., Martin, C. L., et al. 2015, ApJ, 806, 19

Duncan, K., Conselice, C. J., Mortlock, A., et al. 2014, MNRAS, 444, 2960

Duval, F., Schaerer, D., Östlin, G., & Laursen, P. 2014, A&A, 562, A52

Fan, X., Strauss, M. A., Becker, R. H., et al. 2006, AJ, 132, 117

Feldmeier, J. J., Hagen, A., Ciardullo, R., et al. 2013, ApJ, 776, 75

Finkelstein, K. D., Finkelstein, S. L., Tilvi, V., et al. 2015, ApJ, 813, 78

Finlator, K., Davé, R., & Oppenheimer, B. D. 2007, MNRAS, 376, 1861

Finlator, K., Özel, F., Davé, R., & Oppenheimer, B. D. 2009, MNRAS, 400, 1049

Furusawa, H., Kashikawa, N., Kobayashi, M. A. R., et al. 2016, ApJ, 822, 46

González, V., Labbé, I., Bouwens, R. J., et al. 2011, ApJL, 735, L34

González, V., Bouwens, R., Illingworth, G., et al. 2014, ApJ, 781, 34

Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35

Gronke, M., & Dijkstra, M. 2016, arXiv:1604.06805

Gronwall, C., Ciardullo, R., Hickey, T., et al. 2007, ApJ, 667, 79

Guaita, L., Gawiser, E., Padilla, N., et al. 2010, ApJ, 714, 255

Hagen, A., Ciardullo, R., Gronwall, C., et al. 2014, ApJ, 786, 59

Hayes, M., Schaerer, D., Östlin, G., et al. 2011, ApJ, 730, 8

Hayes, M., Östlin, G., Schaerer, D., et al. 2013, ApJL, 765, L27

Kennicutt, R. C., Jr. 1998, ARA&A, 36, 189

Kennicutt, R. C., & Evans, N. J. 2012, ARA&A, 50, 531

Kereš, D., Katz, N., Weinberg, D. H., & Davé, R. 2005, MNRAS, 363, 2

Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36

Kriek, M., van Dokkum, P. G., Labbé, I., et al. 2009, ApJ, 700, 221

Law, D. R., Steidel, C. C., Shapley, A. E., et al. 2012, ApJ, 759, 29

Mallery, R. P., Mobasher, B., Capak, P., et al. 2012, ApJ, 760, 128

Mateo, M., Bailey, J. I., Crane, J., et al. 2012, Proc. SPIE, 8446, 84464Y

Matthee, J., Sobral, D., Oteo, I., et al. 2016, MNRAS, 458, 449

Matsuda, Y., Yamada, T., Hayashino, T., et al. 2012, MNRAS, 425, 878

McLinden, E. M., Finkelstein, S. L., Rhoads, J. E., et al. 2011, ApJ, 730, 136

McLinden, E. M., Rhoads, J. E., Malhotra, S., et al. 2014, MNRAS, 439, 446

Meurer, G. R., Heckman, T. M., & Calzetti, D. 1999, ApJ, 521, 64

Momose, R., Ouchi, M., Nakajima, K., et al. 2014, MNRAS, 442, 110

Nilsson, K. K., Möller-Nilsson, O., Møller, P., Fynbo, J. P. U., & Shapley, A. E. 2009, MNRAS, 400, 232

Noeske, K. G., Weiner, B. J., Faber, S. M., et al. 2007, ApJL, 660, L43

Oke, J. B., & Gunn, J. E. 1983, ApJ, 266, 713

Ono, Y., Ouchi, M., Mobasher, B., et al. 2012, ApJ, 744, 83

Ouchi, M., Shimasaku, K., Akiyama, M., et al. 2008, ApJS, 176, 301-330

Oyarzún, G. A., Blanc, G. A., González, V., et al. 2016, ApJL, 821, L14

Patrício, V., Richard, J., Verhamme, A., et al. 2016, MNRAS, 456, 4191

Pentericci, L., Grazian, A., Fontana, A., et al. 2009, A&A, 494, 553

Pettini, M., Rix, S. A., Steidel, C. C., et al. 2002, ApJ, 569, 742

Pickles, A. J. 1998, PASP, 110, 863

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2015, arXiv:1502.01589

Quadri, R., Marchesini, D., van Dokkum, P., et al. 2007, AJ, 134, 1103

Reddy, N. A., Steidel, C. C., Pettini, M., Bogosavljevic, M., & Shapley, A. 2016, arXiv:1606.03452

Rivera-Thorsen, T. E., Hayes, M., Östlin, G., et al. 2015, ApJ, 805, 14

Schaerer, D. 2003, A&A, 397, 527

Schaerer, D., de Barros, S., & Stark, D. P. 2011, A&A, 536, A72

Schenker, M. A., Stark, D. P., Ellis, R. S., et al. 2012, ApJ, 744, 179

Shapley, A. E., Steidel, C. C., Pettini, M., & Adelberger, K. L. 2003, ApJ, 588, 65

Shimasaku, K., Kashikawa, N., Doi, M., et al. 2006, PASJ, 58, 313

Shimizu, I., Yoshida, N., & Okamoto, T. 2011, MNRAS, 418, 2273

Silva, L., Granato, G. L., Bressan, A., & Danese, L. 1998, ApJ, 509, 103

Skelton, R. E., Whitaker, K. E., Momcheva, I. G., et al. 2014, ApJS, 214, 24

Stanway, E. R., Eldridge, J. J., & Becker, G. D. 2016, MNRAS, 456, 485

Stark, D. P., Ellis, R. S., Bunker, A., et al. 2009, ApJ, 697, 1493

Stark, D. P., Ellis, R. S., Chiu, K., Ouchi, M., & Bunker, A. 2010, MNRAS, 408, 1628

Stark, D. P., Ellis, R. S., & Ouchi, M. 2011, ApJL, 728, L2

Steidel, C. C., Giavalisco, M., Pettini, M., Dickinson, M., & Adelberger, K. L. 1996, ApJL, 462, L17

Tapken, C., Appenzeller, I., Noll, S., et al. 2007, A&A, 467, 63

Tilvi, V., Papovich, C., Finkelstein, S. L., et al. 2014, ApJ, 794, 5

Treu, T., Trenti, M., Stiavelli, M., Auger, M. W., & Bradley, L. D. 2012, ApJ, 747, 27

Vanzella, E., Giavalisco, M., Dickinson, M., et al. 2009, ApJ, 695, 1163

Vargas, C. J., Bish, H., Acquaviva, V., et al. 2014, ApJ, 783, 26

Verhamme, A., Schaerer, D., & Maselli, A. 2006, A&A, 460, 397

Verhamme, A., Schaerer, D., Atek, H., & Tapken, C. 2008, A&A, 491, 89

Whitaker, K. E., van Dokkum, P. G., Brammer, G., & Franx, M. 2012, ApJL, 754, L29

Wisotzki, L., Bacon, R., Blaizot, J., et al. 2016, A&A, 587, A98

Yamada, S. F., Sasaki, S. S., Sumiya, R., et al. 2005, PASJ, 57, 881

Zheng, Z.-Y., Wang, J.-X., Malhotra, S., et al. 2014, MNRAS, 439, 1101

Zitrin, A., Labbé, I., Belli, S., et al. 2015, ApJL, 810, L12