A Comprehensive Study of Lyα Emission in the High-redshift Galaxy Population

Grecco A. Oyarzúñ1,2, Guillermo A. Blanc1,3,4, Valentino González1,4,5, Mario Mateo6, and John I. Bailey, III6
1 Departamento de Astronomía, Universidad de Chile, Casilla 36-D, Santiago, Chile; goyarzun@ucsc.edu
2 Astronomy Department, University of California, Santa Cruz, CA 95064, USA
3 The Observatories of the Carnegie Institution for Science, 813 Santa Barbara Street, Pasadena, CA 91101, USA
4 Centro de Astrofísica y Tecnologías Átnicas (CATA), Camino del Observatorio 1515, Las Condes, Santiago, Chile
5 Chinese Academy of Sciences South America Center for Astronomy, China-Chile Joint Center for Astronomy, Camino del Observatorio 1515, Las Condes, Santiago, Chile
6 Department of Astronomy, University of Michigan, Ann Arbor, MI 48109, USA

Received 2017 February 8; revised 2017 May 15; accepted 2017 May 24; published 2017 July 13

Abstract

We present an exhaustive census of Lyman alpha (Lyα) emission in the general galaxy population at 3 < z < 4.6. We use the Michigan/Magellan Fiber System (M2FS) spectrograph to study a stellar mass (M*) selected sample of 625 galaxies homogeneously distributed in the range 7.6 < log M*/M⊙ < 10.6. Our sample is selected from the 3D-HST/ CANDELS survey, which provides the complementary data to estimate Lyα equivalent widths (WLyα) and escape fractions (fesc) for our galaxies. We find both quantities to anti-correlate with M*, star formation rate (SFR), UV luminosity, and UV slope (β). We then model the WLyα distribution as a function of MUV and β using a Bayesian approach. Based on our model and matching the properties of typical Lyman break galaxy (LBG) selections, we conclude that the WLyα distribution in such samples is heavily dependent on the limiting MUV of the survey. Regarding narrowband surveys, we find their WLyα selections to bias samples toward low M*, while their line-flux limitations preferentially leave out low-SFR galaxies. We can also use our model to predict the fraction of Lyα-emitting LBGs at 4 < z < 7. We show that reported drops in the Lyα fraction at z > 6, usually attributed to the rapidly increasing neutral gas fraction of the universe, can also be explained by survey MUV incompleteness. This result does not dismiss reionization occurring at z ~ 7, but highlights that current data is not inconsistent with this process taking place at z > 7.

Key words: dark ages, reionization, first stars – galaxies: evolution – galaxies: formation – galaxies: high-redshift – galaxies: statistics – ISM: lines and bands

1. Introduction

In hand with observational progress, the understanding of high-redshift galaxies has progressed immensely in the last few decades. Almost 20 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-frame UV-selected galaxies are Lyman Break Galaxies (LBGs; Steidel et al. 1996, 1999, 2003; Shapley et al. 2003; Stark et al. 2009; Bouwens et al. 2015) and Lyα Emitters (LAEs; Cowie et al. 1996; Cowie & Hu 1998; Hu et al. 1998; Kashikawa et al. 2006; Shimakawa et al. 2006; Gronwall et al. 2007; Ouchi et al. 2008; Blanc et al. 2011; Sobral et al. 2015). While the former are observed by their Lyman break at 912 Å, the latter are detected by their Lyman alpha (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, clumpiness, and outflows (Dijkstra & Kramer 2012; Duval et al. 2014; Rivera-Thorsen et al. 2015; Gronke & Dijkstra 2016; Gronke et al. 2016). These studies yield elementary insights that, when combined with observations of the Lyα escape fraction (fesc; 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 reionization, which we have been able to constrain using QSO sightlines (Fan et al. 2006) and cosmic microwave background (CMB) measurements (Bennett et al. 2013; Planck Collaboration 2016a, 2016b). Moreover, with the hope of tracing the sources responsible for reionization, Lyα emission is also conveniently used for characterizing luminosity functions (LFs) at high redshift (Ouchi et al. 2008; Sobral et al. 2009; Dressler et al. 2015; Santos et al. 2016). Nevertheless, the cosmological questions that we are trying to answer by means of this complex emission are not restricted to the reionization of the universe. For instance, Lyα emission can be used to trace the cosmic web (Gould & Weinberg 1996; Cantalupo et al. 2005, 2012, 2014; Kollmeier et al. 2010) and young, metal-poor stellar populations in the early universe (Sobral et al. 2015). Even more, Lyα emission will also be used to study dark energy through baryonic acoustic oscillations 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 (WLyα) of Lyα lines is highly dependent on neutral gas opacity and kinematics, clumpiness of the ISM, dust distribution, and, therefore, line of sight toward the observer. Furthermore, typical Lyα emitters are extreme objects in terms of their luminosity, metallicity, and emission line diagnostics (Trainor et al. 2016).
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 make such an approach difficult. For instance, spectroscopic studies of UV continuum detected galaxies, typically LBGs (e.g., Shapley et al. 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 translates to incompleteness at low SFR. 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 Lyα equivalent width ($W_{Lyα}$) and emission line flux. Such methodology does not require continuum selection, which allows for line detection in faint objects. However, follow-up 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, $W_{Lyα}$, and emission profiles. Even more, such surveys can have non-negligible amounts of line interlopers (Sobral et al. 2017).

Comprehensive studies of Lyα emission in the overall galaxy population are, therefore, of crucial importance. Correlations between galaxy properties and emission are the statistical manifestation of the radiative escape of Lyα photons. Properties such as stellar mass ($M_*$), star formation rate (SFR), UV slope ($\beta$), and merger rate reveal information of 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 to draw conclusions from high-redshift Lyα surveys. Oyarzún et al. (2016) show how the normalization and e-folding scale of the $W_{Lyα}$ distribution of $3 < z < 4.6$ galaxies anti-correlate with $M_*$. In other words, higher $M_*$ objects typically have lower $W_{Lyα}$. More massive galaxies typically have lower gas fractions, but higher gas mass (Tacconi et al. 2013; Song et al. 2014; Genzel et al. 2015). More neutral gas contributes to the increase in the scatter of Lyα photons, spreading this radiation throughout the galaxy and toward the circumgalactic medium. Higher $M_*$ galaxies also have more dust extinction (Franx et al. 2003; Daddi et al. 2004, 2007; Förster Schreiber et al. 2004; Pérez-González et al. 2008), which leads to more Lyα photon 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 that originated in sample selection effects. This example highlights the need for proper assessment.

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 (Grogan et al. 2011; Koekemoer 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 the 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 $W_{Lyα}$ 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 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 2, we describe our sample and data set. In Section 3, we describe our line detection and measurement methodologies. We explain our Bayesian $W_{Lyα}$ distribution analysis in Section 4. We show our results on the $W_{Lyα}$ distribution dependence on different properties and selection techniques in Sections 5 and 6. We state in Section 7 our inferences on higher redshift Lyα emission. We present our conclusions in Section 8.

Throughout this work, all magnitudes are in the AB system (Oke & Gunn 1983). A CDM cosmology with $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_m = 0.3$, and $\Omega_\Lambda = 0.7$ was assumed whenever needed.

2. Data Set

2.1. Sample Selection

Our sample was initially composed of 629 galaxies in the COSMOS, GOODS-S, and UDS fields. Every object was observed under the 3D-HST/CANDELS program, providing HST and Spitzer photometry from 3800 A to 7.9 $\mu$m (9 bands for COSMOS, 14 for GOODS-S, and 9 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 compiles with a photometric redshift reliability parameter $Q < 3$ 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 3D-HST output values are based on the Bruzual & Charlot (2003) stellar population synthesis model library with a Chabrier (2003) IMF and solar metallicity. Exponentially declining star formation histories (SFHs) with a minimum e-folding time of $t^{\text{MM}} = 7$ and a Calzetti et al. (2000) dust attenuation law were also assumed for the calculation (Skelton et al. 2014). We further emphasize that sample selection was performed based on the values of $z_{3D-HST}$ and $M_*$ detailed in this section, whereas all further analysis in this paper is based on our own estimates (Section 2.3).

2.2. Data

Spectroscopy of the full 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. The 1"2 fibers of this instrument allow the observation of 256 targets within 30 arcmin in a single exposure. We are then able to observe 210 targets in each field 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 hr (Mateo et al. 2012). The final data set consists of six exposure hours on each of the three fields with an average seeing of $0.6^\prime$.

For data reduction, we developed 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 0.03 Å. We further correct the solution for each fiber using the sky lines in the science spectra. For flat-fielding, we use sky-flats obtained during either twilight or dawn. The correction is calculated separately for each fiber, and features illumination, fiber profile, and a correction to account for the shifting of fiber spectra on the detector due to thermal 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 performed 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 in the CCD than in 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), we mask broad sections of the spectrum. Due to fiber malfunction, we could not obtain spectra for 4 of the 629 targets, leaving our final sample in 625 objects.

Flux calibration is performed using five $M_V = 19–22$ calibration stars on each exposure. Due to atmospheric differential refraction (ADR) comparable to the 1/2" fiber size, we need to recover the intrinsic spectrum of each star. To do so, we 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 ~15%. After this calibration, we correct the fluxes in our spectra for Galactic extinction. Samples of sky-subtracted spectra are shown in Figure 1.

Considering the three fields, the science area we survey is ~550 arcmin$^2$. The resulting spectral FWHM line resolution is ~2 Å, and we reach a 1σ continuum flux density limit of ~4 × 10$^{-19}$ erg s$^{-1}$ cm$^{-2}$ Å$^{-1}$ per pixel in our 6 h of exposure. We estimate a 5σ emission line-flux sensitivity of ~8 × 10$^{-18}$ erg s$^{-1}$ cm$^{-2}$ in our final spectra (see Figure 1). We identify 120 Ly$\alpha$ emission lines with S/N ≥ 5.5 in our data (details in Section 3.1). Given that we have 625 observed targets, we are then recovering a Ly$\alpha$-emitting galaxy with S/N ≥ 5.5 every ~5 objects.

2.3. Sample Properties

As stated by Skelton et al. (2014), 3D-HST outputs for these galaxies were obtained using FAST (Kriek et al. 2009). These calculations assume exponentially declining SFHs with a minimum e-folding time of log$_{10}$(τ/yr) = 7 (Skelton et al. 2014). However, since recent studies suggest that it is more adequate to reproduce high-redshift observations with constant SFHs (cSFHs; González et al. 2014), or even rising SFHs (Maraston et al. 2010), we perform our own executions of FAST assuming cSFHs. For a detailed discussion on this topic, we refer the reader to Conroy (2013). Our FAST executions also adopt the Bruzual & Charlot (2003) stellar population synthesis model library, a Chabrier (2003) IMF, and solar metallicity, similarly to Skelton et al. (2014). We do not account for nebular emission lines in our SED fitting, which, in principle, can overestimate our reported $M_*$ by a factor of up to 4 (Atek et al. 2011; Conroy 2013; Stark et al. 2013) or even ~7 for strong emitters (de Barros et al. 2014). However, our galaxies are at $z \sim 4$. At this redshift, the overestimate is within a factor of $\lesssim 1.1$ (Stark et al. 2013; Salmon et al. 2015), since the 4.5 μm Spitzer/IRAC band is unaffected. Our FAST outputs also include extinction-corrected SFRs. The SFRs are derived from the SED, and the extinction correction assumes a Calzetti et al. (2000) attenuation law with $R_V = 4.05$.

We are also required to run EAZY (Brammer et al. 2008) on the photometry from CANDELS/IRAC, since FAST requires redshifts as input. The executions of EAZY yield a most probable redshift $z_{\text{EAZY}}$. Our 625 objects satisfy 3 < $z_{\text{EAZY}}$ < 4.25, with a median uncertainty $\sigma_{\text{EAZY}} = 0.1$ (Figure 2). EAZY outputs also include 2σ constraints, which is limited to 2.95 < $z < 4.5$. Our FAST executions yield a mass coverage of 7.6 < log$(M_\ast[M_\odot]) < 10.6$ (Figure 3), with a characteristic uncertainty of log$(M_\ast[M_\odot])$ ~ 0.3. Most of our SFR values are in the range 1–100 $M_\odot$ yr$^{-1}$ (further analysis and plots below). From now on, we use the FAST outputs based on the spectroscopic redshifts ($z_{\text{Ly}\alpha}$) for the 120 detections (S/N ≥ 5.5; Section 3.1) and the outputs based on $z_{\text{EAZY}}$ for the 505 non-detections.

2.4. Considerations

The flux calibration procedure performed in our spectra is based on using stars; therefore, it corrects for a ~32% fiber flux loss, which corresponds to a point-source Ly$\alpha$ surface brightness distribution. In cases of extended Ly$\alpha$ emission halos (Steidel et al. 2011; Matsuda et al. 2012; Feldmeier et al. 2013; Hayes et al. 2013; Momose et al. 2014; Caminha et al. 2016; Patricio et al. 2016; Wisotzki et al. 2016), the fluxes we derive are mostly associated with the galaxies themselves and the inner parts of their Ly$\alpha$ halos. Furthermore, Ly$\alpha$ emission from a galaxy can show significant misalignment with the UV continuum (Rauch et al. 2011), although such cases are not the norm (Finkelstein et al. 2011; Jiang et al. 2013b). At the cosmic time of our sample, the fiber diameter corresponds to a scale of ~8 kpc, roughly a factor of four larger than the typical effective diameter of galaxies (Bond et al. 2012; Law et al. 2012).

We find a median redshift offset of $\Delta z = z_{\text{Ly}\alpha} - z_{\text{EAZY}} = 0.24$ in our detections (see Figure 2). We show in Oyarzún et al. (2016) that this offset does not have any noticeable effects on our sample dependence on $M_\ast$, which is the primary selection criterion for our galaxies. A very similar offset has also been found in the MUSE-Wide Survey (Herenz et al. 2017). We find the offset to correlate with $W_{\text{Ly}\alpha}$ (see Figure 2), hinting at biases in photometric redshift fitting when a strong Ly$\alpha$ emission line is present. A plausible scenario is one in which the Lyman break from the template is fitted slightly blueshifted to account for the flux excess in the redder band. Thorough simulations beyond the scope of this work are required to explore the causes behind this bias.

We stress that the 3D-HST/CANDELS mass incompleteness is restricted to log$(M_\ast[M_\odot]) < 8.5$ at $z \sim 4$ (at least for GOODS-S; Duncan et al. 2014), which corresponds to about $^7$ For the redshift range of our galaxies, the H$\beta$, O II, and O III emission lines fall between the $H_{\text{160}}$ and 3.6 μm bands. H$\alpha$ only contributes to the 3.6 μm Spitzer/IRAC band for $z > 3.75$.
one-quarter of the sample. We want to stress that this is a homogeneously selected sample, designed to study Ly\(\alpha\) emission statistics dependence on galaxy properties. As a consequence, it is by no means representative of the \(M_*\) or Ly\(\alpha\) LFs 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).

3. Ly\(\alpha\) Measurements

3.1. Line Detection

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

\[
f_{\text{res}}(\lambda) = \frac{F}{\sqrt{2\pi\sigma_\lambda^2}} \exp\left[-\frac{(\lambda - \lambda_0)^2}{2\sigma_\lambda^2}\right],
\]

3.2. Line Detection

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

\[
f_{\text{res}}(\lambda) = \frac{F}{\sqrt{2\pi\sigma_\lambda^2}} \exp\left[-\frac{(\lambda - \lambda_0)^2}{2\sigma_\lambda^2}\right],
\]

3.3. Line Detection

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

\[
f_{\text{res}}(\lambda) = \frac{F}{\sqrt{2\pi\sigma_\lambda^2}} \exp\left[-\frac{(\lambda - \lambda_0)^2}{2\sigma_\lambda^2}\right],
\]

Figure 1. Reduced and continuum subtracted spectra of six Ly\(\alpha\) lines (black) from our study. These examples show the wide range of profiles and galaxy properties encompassed by our 120 detections (S/N \(\geq 5.5\)). We also include in these plots the line profiles we fit to each case (red, dashed). The purpose of profile fitting in this work is limited to line-flux measurements.

Figure 2. Measured spectroscopic Ly\(\alpha\) redshifts (\(z_{\text{Ly}\alpha}\)) as a function of the corresponding photometric counterparts (\(z_{\text{EAZY}}\)) for the 120 Ly\(\alpha\) detections (S/N \(\geq 5.5\)). Low-, medium-, and high-\(W_{\text{Ly}\alpha}\) galaxies are shown as cyan circles, blue diamonds, and magenta pentagons, respectively. We find a median deviation of \(\Delta z_{\text{Ly}\alpha} = z_{\text{Ly}\alpha} - z_{\text{EAZY}} = 0.24\) from the 1:1 relation (dashed). As revealed by this figure, we find the deviation to correlate with \(W_{\text{Ly}\alpha}\). Hence, we associate this deviation with photometric redshift fitting biases when a strong Ly\(\alpha\) emission line is present.

Figure 3. Redshift–\(M_*\) distribution of the galaxies composing our sample, according to our executions of EAZY and FAST (Section 2.3). We divide our sample into three \(M_*\) bins using as boundaries \(\log(M_*/M_\odot) = 8.8, 9.6\). We find 120 Ly\(\alpha\) detections (S/N \(\geq 5.5\)), of which 60 are low \(M_*\) (blue circles), 40 are medium \(M_*\) (green diamonds), and 20 are high \(M_*\) (red pentagons). Non-detections are shown as gray points. For detections, we plot \(z_{\text{Ly}\alpha}, M_*(z_{\text{Ly}\alpha})\), whereas for non-detections we use \(z_{\text{EAZY}}, M_*(z_{\text{EAZY}})\).
where $F$, $\lambda_0$, and $\sigma_\ell$ 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 4. We detect four lines above $4\sigma$ and none above $5\sigma$. Therefore, down to $5\sigma$, we expect at most two false detections in our 120 lines, translating into $\lesssim 2\%$ contamination using signal-to-noise $S/N^* = 5.5$ as our threshold. We also characterize our detection method (see Figure 4). To obtain it, we use $p(S/N_\ell > S/N * S/N)$, with $S/N_\ell$ the measured signal-to-noise ratio and $S/N^*$ the imposed detection threshold. We define the simulated signal to noise as $S/N_\text{sim} = F / \sqrt{\sum_k \sigma_\ell^2}$, with $\lambda$ the wavelength dispersion in the spectrum and $\sigma_\ell$ the flux uncertainty for pixel $k$. We find the most accurate representation by summing over an interval of 20 Å centered at 5500 Å. To recover the actual completeness shown in Figure 4, we simulate $\sim 10^5$ lines on the 115 sky-spectra sampling fluxes of $10^{-19}$–$10^{-17}$ erg s$^{-1}$ cm$^{-2}$, FWHMs between 5–13 Å, and wavelengths of 4800–6700 Å. The fraction for which we measure $S/N_\ell > S/N^*$ is our completeness. We use $S/N^* = 5.5$ as our detection threshold, which corresponds to a line-flux sensitivity in the final spectra of $\sim 1 \times 10^{-17}$ erg s$^{-1}$ cm$^{-2}$ (see Figure 1).

### 3.2. Line Profiles

The radiative transfer and escape of Ly$\alpha$ 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 Ly$\alpha$ line is not always well reproduced by the usual Gaussian profile (Chonis et al. 2013; Trainor et al. 2015). Hence, for flux measurements, we adopt a more sophisticated model. Similarly to McLinden et al. (2011) and Chonis et al. (2013), we fit double-peak Gaussian profiles of the form

$$f_\text{rest} (\lambda) = f_\text{blue} (\lambda) + f_\text{red} (\lambda),$$

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

$$\sigma_\lambda = \sigma_0 \text{ if } \lambda < \lambda_0,$$

$$\sigma_\lambda = \sigma_0 \text{ if } \lambda > \lambda_0.$$  

(3)

Before fitting, we convolve the profiles given by Equations (2) and (3) with the spectral resolution. This allows us to properly characterize the errors in our measurements and methodology. In case there is some sky contamination (which happens for 28 of the 120 detections), we only fit single-peaked profiles. The resulting fits to six of our emission lines are shown in Figure 1.

### 3.3. Ly$\alpha$ Equivalent Width and Escape Fraction

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

$$W_{Ly\alpha} = \frac{1}{f_\ell (1 + z_{Ly\alpha})},$$

(4)

with $F$ the Ly$\alpha$ flux we measure in the spectra and $f_\ell$ the observed flux at rest frame 1700 Å from 3D-HST rest-frame colors (Skelton et al. 2014). It must be noted that the 3D-HST rest-frame colors come from the best-fit template to the photometry, i.e., they are not direct measurements. Still, this allows us to use, in principle, the same rest-frame wavelength for every object, regardless of its actual redshift. In addition, we can also derive a value of $f_\ell$ even if the rest-frame 1700 Å photometry is missing. For reference, we find that the templates match the photometry at $\sim 1700$ Å of every object to a typical agreement of $\sim 90\%$. For the uncertainties on $f_\ell$, however, we use the errors on the photometry. We do not use any $f_\ell$ directly measured from our data, since every galaxy has a continuum fainter or comparable to the $1\sigma$ errors in the spectra.

On the other hand, $f_{esc}$ is the fraction of the number of Ly$\alpha$ photons that escape the galaxy from the number produced. This diagnostic is typically indirectly recovered using SFRs derived from the Ly$\alpha$ line and intrinsic SFRs. The latter are typically calculated from UV continuum measurements or H$\alpha$ fluxes, subject to extinction correction. In our case, we use the intrinsic (i.e., extinction-corrected) SFRs from FAST. The observed SFRs are derived from the SED, whereas the extinction corrections assume a Calzetti et al. (2000) attenuation law with $R_V = 4.05$. The explicit definition is then (e.g., Blanc et al. 2011)

$$f_{esc}(Ly\alpha) = \frac{L(Ly\alpha)_{obs}}{L(Ly\alpha)_{intrinsic}} = \frac{\text{SFR}(Ly\alpha)}{\text{SFR}(UV)_{corr}} = 4.4 \times 10^{-42} L_{Ly\alpha}/8.7,$$

(5)

with $L_{Ly\alpha}$ the luminosity associated with the Ly$\alpha$ line and SFR (UV)$_{corr}$ the extinction-corrected SFRs. This assumes a $L_{Ly\alpha}$ to H$\alpha$ ratio of 8.7 and the H$\alpha$ SFR calibration for a Chabrier (2003) IMF of $4.4 \times 10^{-42}$ (Matthee et al. 2016). In case of non-detections, we use the corresponding $S/N^* = 5.5$ line fluxes to derive upper limits on $W_{Ly\alpha}$ and $f_{esc}$. Our calculations do not account for the UV excess associated with binaries,
which induce uncertainties on \( f_{\text{esc}} \) that can be very difficult to account for (Stanway et al. 2016).

### 3.4. AGN Contamination

We do not measure any objects to have more than one emission line with \( S/N > S/N^* \), dismissing the existence of evident active galactic nuclei (AGNs) in our sample. We do not find any of the 120 Ly\( \alpha \)-emitting galaxies to have \( S/N > 1 \) emission potentially associated with the CIV \( \lambda 1549 \) line. Moreover, we find no broad-line Ly\( \alpha \) emission (\( \Delta v > 1000 \text{ km s}^{-1} \)), consistent with the rarity of Type I AGNs at high redshift (Dawson et al. 2004), and in strong contrast with the AGN fractions at low redshift (\( \sim 0.4 \); Finkelstein et al. 2009). This is consistent with the fact that bright AGNs are found in \( \lesssim 1\% \) of z > 3 LAEs (Malhotra et al. 2003; Wang et al. 2004; Gawiser et al. 2007; Ouchi et al. 2008; Zheng et al. 2010). Low-luminosity AGN contamination in LAE samples is more difficult to constrain, but Wang et al. (2004) estimate the total AGN fraction to be \( < 17\% \). We use our \( f_{\text{esc}} \) measurements to further dismiss the existence of evident AGNs. We perform cross-matching with the NASA/IPAC Extragalactic Database (NED\(^8\)) for all objects with \( f_{\text{esc}} > 1 \), and we find none of them to be a reported X-ray source.

### 4. The Ly\( \alpha \) Equivalent Width Distribution

#### 4.1. Bayesian Inference

Measurement of \( W_{\text{Ly} \alpha} \) for a galaxy sample yields the \( W_{\text{Ly} \alpha} \) distribution. Since \( W_{\text{Ly} \alpha} \) is directly measured from the data, characterization of the \( W_{\text{Ly} \alpha} \) distribution can be naively considered straightforward. However, careful consideration of uncertainties and completeness can yield important insights into the 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 \( W_{\text{Ly} \alpha} \) distribution within this framework, complementary to the one introduced in Treu et al. (2012).

Different probability distribution models can be adopted to reproduce \( W_{\text{Ly} \alpha} \) distribution measurements. For instance, studies use Gaussian (e.g., Guaita et al. 2010), exponential (e.g., Jiang et al. 2013b; Zheng et al. 2014), and log-normal distributions. For our analysis, we define the probability distribution as \( p(W_{\text{Ly} \alpha}|W) \). Let \( W \) be the parameter space associated with the model. From now on, we describe the Bayesian approach to recover the posterior distribution of \( W \). 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 data sets. 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 the Ly\( \alpha \) line flux \( F \) instead of \( W_{\text{Ly} \alpha} \), as introduced in Equation (4). This approach simplifies the equations, since we can assume \( F \) and \( f_\lambda \) to be normally distributed, which cannot be done for \( W_{\text{Ly} \alpha} \). According to Bayes’ theorem, the posterior distribution \( p(W|F) \), i.e., the parameter space probability distribution

\[
p(W_{\text{Ly} \alpha}|W) = \frac{p(F|W_{\text{Ly} \alpha})p(W_{\text{Ly} \alpha})}{p(F)}
\]  

(6)

The likelihood is just the product of the individual likelihood for every galaxy, i.e., \( p(F|W_{\text{Ly} \alpha}) = \prod p(F|W_{\text{Ly} \alpha}) \). For a detection, it is given by

\[
p(F|W_{\text{Ly} \alpha}) = \int_{0}^{\infty} p(F|F) p(F|W_{\text{Ly} \alpha}) dF,
\]  

(7)

where \( p(F|F) \) 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 \( p(F|W_{\text{Ly} \alpha}) \) is just the probability distribution of \( F \) given \( W_{\text{Ly} \alpha} \). Using the definition of \( W_{\text{Ly} \alpha} \) from Equation (4), the term translates to the product distribution \( p(F|W_{\text{Ly} \alpha}) = p(W_{\text{Ly} \alpha}|W_{\text{Ly} \alpha})p(f_{\lambda}) \). At this point, we include the probability distribution for the continuum, which we also assume to be Gaussian.

The limiting line flux \( F^\text{lim}_1 \) for discerning detections from noise is given by our S/N threshold, i.e., \( F^\text{lim} = S/N^* \). For galaxies with no detections above \( F^\text{lim} \), we adopt the following value for the likelihood:

\[
p(F_l < F^\text{lim}_1|W_{\text{Ly} \alpha}) = \int_{0}^{\infty} (1 - p(F_l > F^\text{lim}_1|F)) p(F|W_{\text{Ly} \alpha}) dF,
\]  

(8)

with \( p(F_l > F^\text{lim}_1|F) \) the detection completeness at a line flux \( F \) (see Section 3.1).

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

\[
p(W_{\text{Ly} \alpha}|F) = \frac{p(W_{\text{Ly} \alpha})}{p(F)} \prod_{i} p(F_i|W_{\text{Ly} \alpha}) \prod_{j} p(F_j < F^\text{lim}_1|W_{\text{Ly} \alpha}),
\]  

(9)

with \( p(W_{\text{Ly} \alpha}) \) 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 a starting point for multiple analysis throughout.

#### 4.2. Model Comparison

As we introduced in the previous section, multiple probability distributions can be adopted for the representation of the \( W_{\text{Ly} \alpha} \) 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 a 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 analysis presented here is a quantitative implementation of Occam’s Razor and is not unique to our data set, i.e., it can be applied to any data set 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 \( p_0(W_{\text{Ly} \alpha}|W_{\text{Ly} \alpha}) \), the modified counterparts we consider are given by

\[
p(W_{\text{Ly} \alpha}|W_{\text{Ly} \alpha}) = A \times H(W_{\text{Ly} \alpha}) \times p_0(W_{\text{Ly} \alpha}|W_{\text{Ly} \alpha})
\]  

\[
+ (1 - A) \times \delta(W_{\text{Ly} \alpha}),
\]  

(10)

where the first term is the scaled probability distribution. It is multiplied by the Heaviside \( H(W_{\text{Ly} \alpha}) \) to ensure it only

---

\(^8\) 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.
For an analytically correct model comparison, analysis of The absolute probability of each model with $p_{M_i}$ for a set of measurements, Bayes theorem gives the probability of model $M_i$ in the model space $\{M\}$:

$$p(M_i|\{F\}) = \frac{p(\{F\}|M_i)p(M_i)}{p(\{F\})},$$

with $p(M)$ the prior for model $M$, 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.,

$$p(M_i|\{F\}) \propto p(\{F\}|M_i) = \int p(W_n)p(\{F\}|W_n)dW_n.$$

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

$$p(M_i|\{F\}) = \frac{\int p(W_n)p(\{F\}|W_n)dW_n}{\sum p(M_i|\{F\})}. $$

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

$$p(M_i|\{F\}) = \frac{\int p(\{F\}|W_n)dW_n}{\sum p(M_i|\{F\})}.$$  

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 $W_{\text{Ly} \alpha}$ distribution. For the rest of this section, we divide our sample into three $M_*$ bins, as shown in Figure 3. We use the complete sample and these three subsamples to contrast the models. The outcomes are presented in Table 1 and Figure 5. Table 1 gives evidence that the best likelihood is obtained with the log-normal model for three of the four distributions. This can be verified in our distribution simulations for the complete sample in Figure 5, especially toward the high $W_{\text{Ly} \alpha}$ tail. Still, when integrating the likelihoods, the log-normal distribution is the least probable. This is a consequence of the extra parameter needed by the model, which penalizes the likelihood 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 $W_{\text{Ly} \alpha}$ measurements fairly well, despite depending on only two parameters. In addition, the uncertainties in Figure 5, especially for low $W_{\text{Ly} \alpha}$, confirm this model is the most adequate to reproduce our measurements. We remark that the procedure for model selection described here considers the lower $W_{\text{Ly} \alpha}$ end of the distribution, which includes our completeness and non-detections. Nonetheless, further $W_{\text{Ly} \alpha}$ distribution analysis in this paper is mostly focused on the higher $W_{\text{Ly} \alpha}$ end and is not strongly dependent on model preference.

From now on, we perform our $W_{\text{Ly} \alpha}$ distribution analysis using the exponential model of Equation (11). We stress that this expression is dependent on the parameters $\lambda$ and $W_0$, with the first being the fraction of galaxies showing emission and the second the e-folding scale of the distribution. We advise caution when using $\lambda$ as a proxy for the fraction of line emitters in the parent population, since it is tied to the adequacy of an exponential profile.

5. $\text{Ly} \alpha$ Emission Dependence on Galaxy Properties

5.1. Stellar Mass

Evidence suggests $\text{Ly} \alpha$ emission is strongly dependent on the $M_*$ of galaxies. Galaxies with higher $M_*$ have been forming stars for longer, leading to greater ISM dust that presumably forms in supernovae and AGB stars (Silva et al. 1998). A greater dust content leads to more $\text{Ly} \alpha$ photon absorption, decreasing $W_{\text{Ly} \alpha}$. This effect has already been observed, at least for high $W_{\text{Ly} \alpha}$, 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 $\text{Ly} \alpha$ photon budget of galaxies. As a matter of fact, $\text{Ly} \alpha$ emission decreases steadily with the age of stellar populations, as seen in Charlot & Fall (1993) and Schaerer (2003). $\text{Ly} \alpha$ 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 star-forming galaxies are bound to have higher gas mass (e.g., Kereš et al. 2005; Finlator et al. 2007), $\text{Ly} \alpha$ photons should be subject to more resonant scattering, therefore decreasing their $W_{\text{Ly} \alpha}$. 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 Ly\(\alpha\) emission dependence on \(M_\alpha\).

Our sample is especially designed to study the dependence of Ly\(\alpha\) emission in \(M_\alpha\). As shown in Figure 3, our objects are selected in redshift and \(M_\alpha\), homogeneously covering the range \(7.6 < \log(M_\alpha[M_\odot]) < 10.6\). For further clarity, we plot in Figure 6 the \(M_\alpha\) and SFRs of our complete sample and detections. Comparison of the \(M_\alpha\) histograms slightly hints at an anti-correlation between the LAE fraction and \(M_\alpha\), at least down to our detection limit. However, since our completeness depends on \(M_\alpha\), a more thorough analysis is required (see below). The existence of any Ly\(\alpha\) emission dependence on \(M_\alpha\) becomes clearer in Figure 7, where we plot \(W_{\text{Ly}\alpha}\) and \(f_{\text{esc}}\) as a function of \(M_\alpha\). We also plot in Figure 7 our upper limits for non-detections. The regions sampled by our non-detections reveal how our completeness is not independent of \(M_\alpha\). Our detections are flux limited, so we achieve lower \(W_{\text{Ly}\alpha}\) for galaxies with brighter UV continuum. Therefore, even though we observe lower \(M_\alpha\) galaxies to have higher \(W_{\text{Ly}\alpha}\) and \(f_{\text{esc}}\), any qualitative conclusions we can draw involving the fraction of LAEs as a function of \(M_\alpha\) 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. For \(M_\alpha < 10^{8.5}\), there is a clear anti-correlation between \(W_{\text{Ly}\alpha}\) and \(M_\alpha\). For more massive galaxies, however, the trend is mostly flat, except for the presence of a few interlopers. Still, our qualitative result is that both Ly\(\alpha\) emission diagnostics show an anti-correlation with \(M_\alpha\). For the rest of this section, we focus on how to thoroughly quantify the effect of \(M_\alpha\) on \(W_{\text{Ly}\alpha}\).

The overall dependence of Ly\(\alpha\) emission on \(M_\alpha\) implies that \(W_{\text{Ly}\alpha}\) distributions in the literature (e.g., Gronwall et al. 2007; Zheng et al. 2014) are influenced by the \(M_\alpha\) distribution of the sample. In comparison to deeper \(M_{\text{UV}}\) surveys, shallower samples are bound to observe lower \(W_{\text{Ly}\alpha}\). This can lead to incorrect contrast of surveys and misinterpretation of trends. To verify these claims, we divide our sample into three \(M_\alpha\) bins (see Figure 3) and plot the resulting \(W_{\text{Ly}\alpha}\) distributions in Figure 8. As expected, there is an apparent anti-correlation between \(M_\alpha\) and both the tail of the \(W_{\text{Ly}\alpha}\) distribution and the normalization. We perform our first quantitative characterization of the \(W_{\text{Ly}\alpha}\) distribution dependence on \(M_\alpha\) in Oyarzún et al. (2016). In that study, we divide our sample into three \(M_\alpha\) bins and obtain the posterior distribution for the exponential parameters separately. This procedure allowed us to fit a linear relationship to the final parameters, recovering \(A(M_\alpha)\) and \(W_0(M_\alpha)\) using expressions of the form

\[
A(M_\alpha) = A_{MC}\log(M_\alpha[M_\odot]) + A_C, \tag{18}
\]

\[
W_0(M_\alpha) = W_{MC}\log(M_\alpha[M_\odot]) + W_C. \tag{19}
\]

In this section, we recover these linear relations directly from the complete sample, i.e., we recover the posterior distribution for the four-parameter space composed of the linear coefficients in Equations (18) and (19). This more robust methodology does not rely 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 \(W_0 = (A_{MC}, A_C, W_{MC}, W_C)\). As these linear coefficients represent the exponential parameters of Equation (11), deriving non-informative priors is highly complicated. Therefore, in order to determine our priors, we only consider linear scales ignorance. Therefore, the posterior translates to

\[
p(A_{MC}, A_C, W_{MC}, W_C | [F]) = C \times \prod_D p(F_D | A(M_\alpha), W_0(M_\alpha)) \times \prod_{ND} p(F_{D} < F_D | A(M_\alpha), W_0(M_\alpha)), \tag{20}
\]

with \(C\) a normalization constant.

We use MCMC simulations to characterize this four-parameter posterior. Its maximum gives the best solution for \(A(M_\alpha)\) and \(W_0(M_\alpha)\), while the collapsed posteriors yield the uncertainties on the parameters. We can then write Equations (18) and (19) as

\[
A(M_\alpha) = -0.28^{+0.12}_{-0.02}\log(M_\alpha[M_\odot]) + 3.1^{+0.15}_{-1.1}, \tag{21}
\]

\[
W_0(M_\alpha) = -17.7^{+2.6}_{-2.0}\log(M_\alpha[M_\odot]) + 190^{+20}_{-25}. \tag{22}
\]

In the framework of an exponential profile, these relations we recover yield the \(W_{\text{Ly}\alpha}\) probability distribution for an object with known \(M_\alpha\). Hence, we can simulate the expected \(W_{\text{Ly}\alpha}\) distribution for each of the three \(M_\alpha\) subsamples and compare with our direct measurements. The results are presented in Figure 8. Our constraints are consistent with the observed \(W_{\text{Ly}\alpha}\) distributions.

Our results regarding the dependence of Ly\(\alpha\) on \(M_\alpha\) are conclusive. In the range \(10^{8}-10^{10.5}\) \(M_\odot\), the \(W_{\text{Ly}\alpha}\) probability distribution extends to higher \(W_{\text{Ly}\alpha}\) for lower \(M_\alpha\) galaxies. In

| Likelihood | Low Mass | Medium Mass | High Mass | Complete Sample |
|------------|----------|-------------|----------|----------------|
| Output     | Exp  | Gaussian | Log \(n\) | Exp  | Gaussian | Log \(n\) | Exp  | Gaussian | Log \(n\) | Exp  | Gaussian | Log \(n\) |
| Model odds\(^a\) | 0.93 | 0.06 | 0.01 | 0.97 | 0.01 | 0.02 | 1 | \(10^{-3}\) | \(10^{-3}\) | 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 |
| \(A_\alpha\) | 0.75 | 0.55 | 0.55 | 0.55 | 0.4 | 0.5 | 0.35 | 0.2 | 0.2 | 0.4 | 0.3 | 0.45 |
| \(W_{\text{Ly}\alpha}\) | 46 | 74 | 0.7 | 26 | 46 | 0.85 | 14 | 26 | 0.75 | 38 | 64 | 1.05 |
| \(\mu\) | ... | ... | 3.85 | ... | ... | 3.05 | ... | ... | 3 |

Notes.
\(^a\) Obtained by integrating the likelihood over the whole parameter space.
\(^b\) Calculated with the maximum of the likelihood.
other words, more massive galaxies tend to have lower $W_{\lambda_{21}}$. A similar trend is observed for $f_{\text{esc}}$, further highlighting the role of dust and gas mass in the escape of $Ly\alpha$ photons. At $z \sim 2.2$, Matthee et al. (2016) also observe the $f_{\text{esc}}$ anti-correlation from Figure 7, although only when stacking galaxies. Their massive objects showing high $f_{\text{esc}}$, which they associate with dusty gas outflows, seem to lie below the $M_\ast$-SFR sequence at $z \sim 2$. These inferences, combined with the significantly higher $f_{\text{esc}}$ they measure for larger apertures, make sense in a Ly$\alpha$ diffuse halo scheme, which we do not observe due to our aperture size. Studies on the dependence of $f_{\text{esc}}$ on $M_\ast$ also explore $1.9 < z < 3.6$ (Hagen et al. 2014). Their $W_{\lambda_{21}}$-selected LAEs follow a trend similar to the one we find at $z \sim 4$. In summary, the evidence for an anti-correlation between $W_{\lambda_{21}}$ or $f_{\text{esc}}$ and $M_\ast$ is significant, but the scatter seems to depend on measurement methodology and sample selection.

Inferences on the $M_\ast$ distribution of LAEs are not as evident. Hagen et al. (2014) do not find their $1.9 < z < 3.6$ Ly$\alpha$ luminosity-selected LAE number distribution to depend on $M_\ast$. Their results agree with the $z \sim 3.1$ narrowband-selected survey of McLinden et al. (2014). However, we show in Figure 7 that spectroscopic completeness is not independent of $M_\ast$. Therefore, most LAEs survey follow-ups could have higher incompleteness toward lower $M_\ast$. Since our Bayesian analysis takes into account our completeness for every object, we can test the significance of this claim. The coefficient $A_{M_\ast}$ in Equation (18) represents the exponential fraction of LAE dependence on $M_\ast$. As evidenced by Equation (21), our measurements are more than $2 \sigma$ consistent with a decrease in the fraction going to higher $M_\ast$. In the scheme of an exponential model, this translates to an LAE distribution dominated by lower $M_\ast$ galaxies. This result complements the much more significant anti-correlation between $W_\alpha$ and $M_\ast$ we find (see Equation (22)).

### 5.2. SFR

High-redshift galaxies have been observed to follow a correlation between SFR and $M_\ast$, 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; see our Figure 6). In terms of the underlying physics, more massive objects dominate gas accretion in their neighborhood, feeding and triggering 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 $M_\ast$ from the previous section, we expect similar trends between $W_{\lambda_{21}}$ ($f_{\text{esc}}$) and SFR (Figure 6). Even more, star-forming galaxies have a higher neutral gas mass, which can hamper the escape of Ly$\alpha$ photons from galaxies (Verhamme et al. 2006). In fact, it has also been suggested that photoelectric absorption rules Ly$\alpha$ depletion, even over dust attenuation (Reddy et al. 2016). In this section, we explore any Ly$\alpha$ dependence on SFR within our data set. We remark that our SFRs come from SED fitting of 3D-HST photometry using FAST (see Section 2.3), i.e., they have typical associated timescales of 100 Myr (Kennicutt 1998). We stress that our derived SFRs differ from 3D-HST SFRs, since our calculation assumes cSFHs instead of exponentially declining SFHs (Skelton et al. 2014).

We show the $W_{\lambda_{21}}$ and $f_{\text{esc}}$ dependence on SFR in Figures 6 and 9. In the latter, we include upper limits for our non-detections to give an insight into how our incompleteness depends on SFR. A clear anti-correlation between $W_{\lambda_{21}}$ ($f_{\text{esc}}$) and SFR is observed. These results come as no surprise, as they have been previously reported. Most studies of $W_{\lambda_{21}}$ dependence on SFR involve uncorrected SFRs (Pettini et al. 2002; Shapley et al. 2003; Yamada et al. 2005; Gronwall et al. 2007; Tapken et al. 2007; Ouchi et al. 2008; all compiled in Verhamme et al. 2008). Even without dust correction, the anti-correlation is still present in these studies (refer to Figure 19 in Verhamme et al. 2008 and Section 5.4 of this work). Based on a $z \sim 2$ $H_\alpha$ emitters sample, Matthee et al. (2016) also observe a clear anti-correlation between Ly$\alpha$ $f_{\text{esc}}$ and SFR. Interestingly, they do not only observe such a trend in their individual objects, but likewise on their stacks when using different apertures (galaxy diameters of 12 and 24 kpc). As their data set includes $H_\alpha$ fluxes, they can recover SFRs and $f_{\text{esc}}$ using $H_\alpha$...
luminosities. The fact that they observe similar trends with such a different sample suggests that the anti-correlation between $W_{\text{Ly}\alpha}$ ($f_{\text{esc}}$) and SFR is not only independent of redshift, but also observational constraints like aperture and methodology for recovering SFRs. Their comparison, however, is restricted to SFRs higher than $\sim 5 \, M_\odot \, \text{yr}^{-1}$. Most of our low-mass objects have SFRs lower than $\sim 5 \, M_\odot \, \text{yr}^{-1}$, but they seem to follow the same regime as the rest of our sample. Even though $f_{\text{esc}}$ uncertainties and incompleteness increase toward lower-SFR, UV-fainter galaxies (Figure 9), our results suggest that $f_{\text{esc}}$ reaches values of 100% toward SFR $\sim 1–3 \, M_\odot \, \text{yr}^{-1}$. These numbers are consistent with the analysis by Atek et al. (2014). They compare their $z < 0.5 \, \text{SFR(Ly}\alpha)/\text{SFR(UV)}$ measurements with the literature at $z > 2$ (Taniguchi et al. 2005; Gronwall et al. 2007; Guaita et al. 2010; Curtis-Lake et al. 2012; Jiang et al. 2013b). Our finding that $f_{\text{esc}}$ reaches values of 100% toward SFR $\sim 1–3 \, M_\odot \, \text{yr}^{-1}$ overestimates SFR(Ly$\alpha$)/SFR(UV) for $z < 0.5$, but is consistent with higher redshift. As pointed by Atek et al. (2014), it seems that $f_{\text{esc}}$ at fixed SFR(UV) increases with redshift. Then again, it must be kept in mind that these literature results consider uncorrected SFRs.

5.3. UV Luminosity

In this section, we analyze the $M_{\text{UV}}$ distribution of our sample, while also exploring any correlations between $W_{\text{Ly}\alpha}$ 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 toward 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 $M_{\text{UV}}$ galaxies than a population-representative subsample.

In order to determine $M_{\text{UV}}$ for our objects, we use the CANDELS $i_{775}$ band. We show in Figure 10 the corresponding distribution of the complete sample and detections. We also present the dependence of $W_{\text{Ly}\alpha}$ on $M_{\text{UV}}$ in this figure. As expected from the SFR–$W_{\text{Ly}\alpha}$ anti-correlations we recover in Section 5.2, 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 (Stark et al. 2009; González et al. 2014). Galaxies brighter in the UV have been subject to more intensive star formation events in 100 Myr timescales. Typically, higher neutral gas, turbulence, and bulk gas motions are associated with higher SFRs, boosting the scatter of Ly$\alpha$ photons. In combination, higher dust extinction, older age of stellar populations, and greater neutral gas mass in UV brighter galaxies seem to rule Ly$\alpha$ statistics dependence on SFR.

Most analyses of Ly$\alpha$ emission dependence on $M_{\text{UV}}$ have been performed using uncorrected SFRs. The literature compilation shown in Verhamme et al. (2008) reveals how observed UV SFRs anti-correlate with $W_{\text{Ly}\alpha}$ from $z \sim 2.7$ to 5.7 (Pettini et al. 2002; Shapley et al. 2003; Yamada et al. 2005; Gronwall et al. 2007; Tapken et al. 2007; Ouchi et al. 2008). Analyses explicitly using $M_{\text{UV}}$ have also been 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 up to $z \sim 6$. More recent studies confirm these trends (Jiang et al. 2013b, 2016; Zheng et al. 2014). Simulations likewise predict such correlations (see Shimizu et al. 2011). However, regardless of the methodology, the scatter in these correlations is non-negligible. Moreover, Atek et al. (2014) question the existence of the correlation in their sample. We argue that the scatter observed in the $W_{\text{Ly}\alpha}$ dependence on $M_{\text{UV}}/\text{SFR}_{\text{obs}}$ can be a consequence of...
of the role played by dust. Galaxies brighter in the UV naturally have a greater Lyα photon production, but the anti-correlation with $M_\ast$ affects the escape fraction (probably through increased dust extinction). In this scenario, Lyα photon escape is a complex process simultaneously ruled by different properties of high-redshift galaxies. We explore such a property space in our analysis of the $W_{\text{Ly}\alpha}$ distribution dependence on the $M_{1500-3600}$ relationship in Section 5.5.

Implications involving the fraction of detections as a function of $M_{1500-3600}$ 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 $W_{\text{Ly}\alpha}$. Nevertheless, our Lyα measurements are flux limited. Therefore, our detection completeness in $W_{\text{Ly}\alpha}$ 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 taking into account the uncertainties for line and continuum fluxes ($F$ and $f_{\text{esc}}$, respectively). Hence, we encourage further interpretations of these results to focus on the analysis performed on the $M_{1500-3600} - \beta$ plane (Section 5.5).

5.4. UV Slope

Measurement of the UV slopes of high-redshift galaxies 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 Lyα 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-frame 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 our method to estimate the UV slopes for our objects and show our results on the 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 on the photometry between rest frame 1400 and 3500 Å. These calculations correspond, in principle, to 8–15 bands between 1500 and 3600 Å. However, since we only use fluxes with $S/N > 3$, our median number of bands is eight. Naturally, we further require at least two bands to associate a slope to our targets, which means we can measure the UV slope for 611 of the 625 observed objects composing our sample.

Our FAST outputs include dust extinction in the V-band, $A_V$, for every object. However, if we want to associate a reddening $E(B-V)$ with every galaxy, we need to use our derived UV slopes. We use the relation

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

(23)

where we assume a pristine slope of $\beta = -2.23$ (Meurer et al. 1999). We also require the adoption of an attenuation law $k_\lambda$ (e.g., Calzetti et al. 2000). Reddy et al. (2015) recover a more appropriate attenuation law than Calzetti et al. (2000) for high-redshift galaxies, so we use the $k_\lambda$ of the former. It is worth noting, nevertheless, that both attenuation laws yield almost identical results.

Since our targets are $M_\ast$ selected, we have a higher contribution of massive objects in comparison to the $M_\ast$ 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 a lower contribution from such galaxies. In any case, we show in Figure 11 the UV slope histogram of our sample and our Lyα emitters. Just by comparing both distributions, it is clear that the fraction of emitters increases toward bluer galaxies. We present a quantitative analysis of this claim in Section 5.5.

We also show in Figure 11 our results on $W_{\text{Ly}\alpha}$ and $f_{\text{esc}}$ as a function of $\beta$ and $E(B-V)$. Since lower mass galaxies have bluer UV slopes than more massive ones, the trends we find are complementary to our previous results. There is a correlation between the steepness of the UV spectrum and $W_{\text{Ly}\alpha}/f_{\text{esc}}$, although with significant scatter. As extinction seems to play a major role in Lyα photon escape from galaxies, mainly through scattering and absorption (Blanc et al. 2011; Hagen et al. 2014), these correlations come 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). Regarding the scatter we find at fixed $\beta$ (see also Blanc et al. 2011), it is consistent with a scenario where the observed Lyα flux is mostly affected by the dependence of the
dust-covering fraction on the line of sight. We discuss this picture in detail in Section 5.5.

When comparing the more dusty Ly$\alpha$ emitters in our sample with results from the literature, however, some differences show up. The $1.9 < z < 3.8$ survey from Blanc et al. (2011) and the $z < 0.5$ study from Atek et al. (2014) find Ly$\alpha$ emitters up to $E(B-V) \sim 1$. Similarly, Matthee et al. (2016) find a population of dusty LAEs at $z \sim 2.23$, and they speculate on how dusty gas outflows might be the feature driving the escape of Ly$\alpha$ radiation. Ly$\alpha$ sources up to $z \sim 3$ are Herschel (Oteo et al. 2011, 2012; Casey et al. 2012; Sandberg et al. 2015) and SCUBA (Geach et al. 2005; Hine et al. 2016) detected. Moreover, Ly$\alpha$ emission has also been measured in submillimeter galaxies (e.g., Chapman et al. 2005). On the other hand, our 625 object sample features $\sim 5$ objects with $E(B-V) > 0.4$, but none of them qualifies as a detection. We cannot state whether the absence of such LAEs in our sample is representative of $z \sim 4$. True enough, the fraction of dusty galaxies at $z \sim 4$ is expected to be lower than at $z < 4$. Along such lines, cosmic evolution in the fraction of dusty LAEs has 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 the selection effects associated with LBG surveys. We conclude that even though Ly$\alpha$ emitting galaxies are mostly low-dust objects (e.g., Song et al. 2014), there is also a population of dustier, low-W$_{Ly\alpha}$ LAEs. Still, the significance of their numbers at $z \geq 4$ is still an open question.

Our results also confirm at $z \sim 4$ a trend in $f_{esc}$ that has 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 to the one we show in Figure 11. Hayes et al. (2011) at $z \sim 2.2$, Blanc et al. (2011) at $1.9 < z < 3.8$, Song et al. (2014) at $z \sim 2.1-2.5$, 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 11. Considering that our results are dominated by upper limits, the two higher redshift relations are roughly consistent with our measurements. Still, of particular interest might be 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 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 $E(B-V) > 0.2$.

5.5. M$_{UV}$–$\beta$ Sequence

We have characterized in this paper the dependence of Ly$\alpha$ emission on M$_{a}$, SFR, UV luminosity, and UV slope. We find W$_{Ly\alpha}$ and $f_{esc}$ to anti-correlate with these four properties. However, these results might not be independent, since these properties are correlated (see Figure 6). For instance, M$_{a}$ 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). Similarly, a relation between M$_{UV}$ and $\beta$ has been studied in LBGs (Bouwens et al. 2009, 2014). As we discuss throughout this work, Ly$\alpha$ escape from galaxies is likely to be a process ruled by many parameters, such as age of the population, gas column density, extinction, and SFR. Therefore, it is interesting to explore Ly$\alpha$ emission dependence on multi-dimensional spaces. In this section, we study W$_{Ly\alpha}$ dependence on M$_{UV}$ and $\beta$. There are two reasons to justify exploring this parameter space in particular. First, both properties are observables, i.e., the amount of assumptions involved in their calculation is kept to a minimum (unlike, for example, M$_{a}$ and SFR). Second, these
two properties are directly related to elements that rule Lyα escape. The UV luminosity of a galaxy traces its ionizing output and gas content, whereas the UV slope is a proxy for dust. Hence, in this section, we focus on characterizing Lyα emission in the $M_{UV}$–$\beta$ plane. Similarly to previous analyses in this paper, we take advantage of a Bayesian approach to obtain our results.

We first present in Figure 12 the location of our detections and non-detections in this plane. In qualitative terms, 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 $W_{Ly\alpha}$ and $M_{UV}$ is still clearly observed when comparing the low $M_{UV}$ with the high $M_{UV}$ subsamples, independent of $\beta$. For the UV slope, however, the trend between $W_{Ly\alpha}$ and $\beta$ does not seem that clear anymore. The normalization of the distribution, however, looks to be highly dependent on the UV slope. To characterize the significance of these insights, we simultaneously model the exponential profile parameters of Equation (11) as a function of $M_{UV}$ and $\beta$. We start again with linear expressions of the form

$$A(M_{UV}, \beta) = A_{M_{UV}} M_{UV} + A_\beta \beta + A_C,$$

$$W_0(M_{UV}, \beta) = W_{0_{M_{UV}}} M_{UV} + W_{0_\beta} \beta + W_C.$$ (24)

Once the parameterization and priors are set, we can obtain the posterior distribution of $[W_i] = (A_{M_{UV}}, A_\beta, A_C, W_{0_{M_{UV}}}, W_{0_\beta}, W_C)$ using Equation (9). We assume the priors to be independent, which translates to $p([W_i]) \propto \prod p(W_i)$. 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 ($A_{M_{UV}}, A_C$ and $W_\beta$). Therefore, our

![Figure 9](image1.png)

**Figure 9.** Plot of our $W_{Ly\alpha}$ (top panel) and $f_{esc}$ (bottom panel) as a function of intrinsic SFR. Detections corresponding to our low-, medium-, and high-mass bins are shown as the 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 fractions and galaxy SFRs. For reference, we show our non-detection upper limits as gray triangles. We use these non-detections to give a rough estimate of our 40% completeness (dashed), confirming how strongly our sensitivity depends on SFR.

![Figure 10](image2.png)

**Figure 10.** Top: histograms of sample (gray) and detections (black) as a function of UV absolute magnitude. We use the CANDELS $i_{775}$ band for the derivation of $M_{UV}$. For better comparison, we divide the histogram values for the full sample by five. The photometric measurement on this particular band is available for 621 of the 625 galaxies. Center: dependence of $W_{Ly\alpha}$ on UV absolute magnitude. Bottom: dependence of $f_{esc}$ on $M_{UV}$. Blue circles, green diamonds, and red pentagons correspond to low-, medium-, and high-mass subsample detections, respectively. Non-detection upper limits are shown as gray triangles, and we use them to give a rough estimate of our 40% completeness (dashed). 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 narrowband surveys (Ouchi et al. 2008).
W_0 < 0, we impose A = 0, i.e., the W_{Lyα} distribution is a Dirac delta (i.e., no Lyα emission and/or absorption).

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

\[
A(M_{UV}, \beta) = 0.08^{+0.1+5}_{-0.06} M_{UV} - 0.6^{+0.1+5}_{-1.1} \beta + 1.1^{+5}_{-1.4},
\]

\[
W_0(M_{UV}, \beta) = 11^{+2.0}_{-1.8} M_{UV} - 7.8^{+5.2}_{-5.5} \beta + 235^{+45}_{-42}.
\]

We show with more detail the results for these coefficients in Figure 13. The early analysis based on the sample binning from Figure 12 is verified in Figure 13. The normalization factor A anti-correlates with \beta with a significance \geq 2\sigma. There also seems to be an anti-correlation between A and M_{UV}, but to a level of \sim 1\sigma. For W_0, the e-folding scale of the distribution, the behavior is the opposite. The extent of the distribution correlates with M_{UV} and shows a weak anti-correlation with \beta.

We now focus on the interpretation of these results. The fact that beta correlates mostly with A suggests that the UV slope is somehow related to a stochastic probability of being able to observe or not a particular object in Lyα. On the other hand, the stronger correlation of M_{UV} with W_{Lyα}, suggests that the UV luminosity is associated with physical processes that determine the magnitude of the resulting W_{Lyα}. A possible scenario is one in which M_{UV} traces SFR and, therefore, the total cold gas mass of galaxies. Hence, higher UV luminosities translate into increased levels of Lyα photon scattering into an extended halo, smoothly reducing the W_{Lyα} of the central source. At the same time, in a significantly clumpy ISM in which dust is well mixed with the gas, a large contrast in gas density/column will imply that dust can effectively block the bulk of both UV and Lyα radiation in some parts of the disk. Such impact will be significant toward certain lines of sight, with little to no effect on others. Such a “covering factor” scenario could explain that \beta correlates more significantly with the normalization (A) than with the shape of the distribution (W_0). This is consistent with the finding that Lyα and UV suffer from similar levels of extinction by dust in LAEs (e.g., Blanc et al. 2011). Moreover, the scatter in the f_{esc} of Lyα photons at fixed reddening (see Figure 11 and Blanc et al. 2011) is also consistent with this scenario where wide ranges of dust absorption lines of sight and photon scattering halos rule the escape of Lyα radiation.

Naturally, this interpretation is a simplified description of Lyα escape from galaxies. Going no further, Figure 13 reveals how both distribution parameters do not solely depend on one property. Ideally, studies of W_{Lyα} and f_{esc} in three-parameter spaces (e.g., M_s, M_{UV}, \beta) can yield predictive and more accurate parameterization of Lyα emission. However, such analysis must be performed in much larger data sets than ours, at least if significant enough results are to be obtained. Furthermore, hydrodynamical simulations can also give insight into how Lyα escape depends on the line of sight and the distribution of gas, dust, and star-forming regions (e.g., Verhamme et al. 2012), especially if performed in a statistically significant sample.

6. Lyα Dependence on Sample Selection

6.1. LBG Samples

The Lyman break selection technique has proven to be a very efficient method for detecting high-redshift galaxies (e.g., Steidel et al. 1996; Shapley et al. 2003; Stark et al. 2010;
Ono et al. 2012). The fact that the Lyman break is in the optical region of the observed spectrum for galaxies at redshift $z \gtrsim 3$ allows for efficient detection from ground telescopes. By only requiring the use of a few broadband filters, several galaxies can be detected in a deep single exposure. Still, to avoid aliasing with the Balmer break, there are unavoidable biases associated with this technique. First, galaxies with no prominent Lyman break are, by construction, not detected. As a consequence, either extremely young or passive, heavily extincted galaxies are underrepresented in LBG surveys. As a second, these surveys also impose color restrictions on the UV sensitivity of the survey, creating incompleteness at low SFR. In the case of Lyα, missing the highest $\beta$ galaxies in the universe. Indeed, the $M_{UV}$ dependence on the Lyα emission. This observational limit requires the use of a few broadband filters, several galaxies can be detected in a deep single exposure. Still, to avoid aliasing with the Balmer break, there are unavoidable biases associated with this technique. First, galaxies with no prominent Lyman break are, by construction, not detected. As a consequence, either extremely young or passive, heavily extincted galaxies are underrepresented in LBG surveys. Second, these surveys also impose color restrictions on the slope of galaxies, further increasing the selection toward, in principle, bluer UV objects. Third, this technique is limited by the $M_{UV}$ sensitivity of the survey, creating incompleteness at low SFR. In the case of Lyα emission, this observational limit can have a significant downside. As shown in Section 5.2, there is a clear anti-correlation between UV luminosity 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. Indeed, the deepest surveys nowadays reach $\sim$50% completeness at observed $m_{UV} \sim 29.5$ (e.g., Bouwens et al. 2015; Bowler et al. 2015, 2017; Ishigaki et al. 2017). These limits translate to higher redshift $\sim$50% completeness at $M_{UV} \sim -18.3$, $-19$, $-19.5$, and $-19.7$ ($z \sim 4$, 5, 6, and 7, respectively). These limitations become problematic when comparing the results between LBG and narrowband samples. Presumably due to not requiring a continuum detection, the $W_{Ly\alpha}$ observed in narrowband surveys are systematically higher (e.g., Zheng et al. 2014).

Our data set and results can be useful for characterizing the effects that selection techniques can have on Lyα emission. Even though it is possible to identify these selections directly in our sample, any comprehensive analysis must consider our galaxy selection procedure. Since our sample follows neither the $M_*$ nor $M_{UV}$ functions of the galaxy population, correcting requires assumption of $M_*$ and/or $M_{UV}$ distributions. Nevertheless, we can still simulate Lyα emission samples on CANDELS galaxies using our results of $W_{Ly\alpha}$ dependence on the $M_{UV}–\beta$ plane given by Equations (27) and (28). In this section, we describe our simulation of a $z \sim 4$ LBG sample from the 3D-HST catalogs. We then compare the properties of LBGs with the parent distribution, focusing on $M_*$, SFR, $M_{UV}$, and $\beta$. We finally simulate Lyα fractions for each sample and conclude on the affects of LBG selection at $z \sim 4$.

LBGs at $z \sim 4$ are typically selected using B-drops and imposing color selections on redder filters. As we are simulating this survey in CANDELS, our detection limits are given by the depth of their $H_{160}$ images. For the $z \sim 4$ LBG selection, we adopt the same methodology applied in Bouwens et al. (2012):

\[ (B_{435} - V_{606} > 1.1) \land (B_{435} - V_{606} > (V_{606} - z_{850}) + 1.1) \land (V_{606} - z_{850} < 1.6). \]  

After color selection, we impose $5\sigma$ detections in the $V_{606}$ filter, to which we now refer to as the detection band. As commonly applied in LBG surveys, we also require all candidates to have $<1.5\sigma$ measurements in the $U$ band, since the Lyman break must be redshifted out of this filter. Before performing any analysis on the CANDELS LBG sample, we run EAZY and FAST according to our prescriptions (see Section 2.3). Of the $\sim$3200 CANDELS galaxies that comply with our selection, $\sim$800 classify as low-redshift interlopers ($z_{\text{EAZY}} < 3$), according to our outputs. We check EAZY $3\sigma$ constraints for these interlopers and find most to have $z_{\text{EAZY}} < 4$. Hence, we find a low-redshift contamination in $z \sim 4$ CANDELS LBGs that is higher than the typically reported number of $\sim$10% (Bouwens et al. 2007). From now on, we remove these presumed $z_{\text{EAZY}} < 3$ contaminants and work with a $z \sim 4$ LBG sample of nearly 2300 objects.

Using our outputs, we plot relevant properties of CANDELS LBGs in Figure 14. To ensure that we do not venture way beyond CANDELS completeness, we restrict the upcoming analysis to galaxies with $M_* > 10^{8.5} M_\odot$. We construct two $3 < z < 4.2$ CANDELS samples for comparison: all galaxies with $M_* > 10^{8.5} M_\odot$ and all V-band detected $M_* > 10^{8.5} M_\odot$. 

![Figure 12](image_url)
objects. As seen in the central panel, the major LBG selection effect is associated with the detection band threshold. These left-out objects are either red, heavily extincted galaxies, or blue, intrinsically faint objects (Quadri et al. 2007). Indeed, as revealed by the comparison between LBGs and the detection-band-selected sample, any selection imprints associated with color requirements are minor. It is still worth noting, however, that these color criteria seem to neglect blue rather than red galaxies at \( z \sim 4 \) (center of Figure 14). We find the \((B - V > 1.1)\) Lyman break cut to be the driving criterion behind this selection effect (see Equation (29)). We remark that these insights might not be valid at different redshifts or lower \( M_{\alpha} \), however. Some studies have indeed explored differences at higher redshift (e.g., Jiang et al. 2013a, 2013b, 2016). In analyses that explore Ly\( \alpha \) emission strength, UV continuum properties, morphologies, and sizes, they find no significant differences between LBGs and LAEs at \( z \gtrsim 6 \).

The insights we present can be further tested by simulating the fraction of LAEs above an \( W_{\text{Ly} \alpha} \) threshold. We use our results from Section 5.5 to estimate the Ly\( \alpha \) fraction for each sample and plot our results in the right panel of Figure 14. Indeed, depending on the depth of the detection band, the fraction of low UV luminosity galaxies composing the sample can vary significantly. As a consequence, deeper surveys can potentially recover a higher fraction of line emitters at high \( W_{\text{Ly} \alpha} \). To deal with this selection, it has been proposed to compare the Ly\( \alpha \) fluxes over narrow UV luminosity samples (Stark et al. 2011; Mallery et al. 2012; Ono et al. 2012; Schenker et al. 2012). Our results in this paper further emphasize the need to compare galaxies of the same luminosity and properly characterize completeness when studying the evolution of the Ly\( \alpha \) fraction with cosmic time. Regarding color selections, we find the Ly\( \alpha \) fraction to be slightly lower for LBGs than for the \( V \)-band-detected sample. As we suggested, this is a consequence of LBG color cuts leaving out some of the bluest galaxies.

6.2. Narrowband Samples

Ly\( \alpha \)-emitting galaxies can also be selected using narrowband imaging or blind spectroscopy (e.g., Gronwall et al. 2007; Ouchi et al. 2008; Adams et al. 2011). By establishing a \( W_{\text{Ly} \alpha} \) detection threshold between the narrow- and broadband flux measurement, this technique allows for efficient line-emitter 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 narrowband measurements of LAEs are followed up by spectroscopy, any sample selection effects induced by the narrowband technique are already present in the sample. The \( W_{\text{Ly} \alpha} \) 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 \( W_{\text{Ly} \alpha} \) cuts. For instance, Vargas et al. (2014) select sources with \( W_{\text{Ly} \alpha} > 20 \text{ Å} \) at \( z \sim 2.1 \), Zheng et al. (2014) select \( W_{\text{Ly} \alpha} > 9 \text{ Å} \) at \( z \sim 4.5 \), and Sobral et al. (2017) go as low as \( W_{\text{Ly} \alpha} > 5 \text{ Å} \) at \( z \sim 2.23 \). If \( W_{\text{Ly} \alpha} \) selections induce important biases on galaxy samples, the comparison of different surveys is not straightforward.

In this section, we explore the effects that narrowband selections have on the population of LAEs, focusing on the \( M_{\alpha} \)-SFR plane. The insights we present here are based on Equations (27) and (28), i.e., our \( M_{\alpha} - \beta \) model. We show in Figure 15 the outcome of \( W_{\text{Ly} \alpha} \) (top) and line-flux selections (bottom) on the \( M_{\alpha} \)-SFR sequence. The red contours show \( 3.5 < z < 4.5 \) CANDELS galaxies with \( M > 10^{8.5} M_\odot \).
clarify, however, we concentrate our analysis on the main sequence, removing objects with young and old ages (red dots). This approach allows our results to be dominated by objects optimally fitted by our FAST executions.

To explore the effect of \( W_{\text{Ly}\alpha} \) selections (top panel of Figure 15), we base our analysis on the \( z \sim 2-2.3 \) narrowband survey of Sobral et al. (2017), whose galaxies are limited to \( W_{\text{Ly}\alpha} > 5 \text{ Å} \) and fluxes \( > 2 \times 10^{-17} \text{ erg cm}^{-2} \text{s}^{-1} \) (5σ). We simulate a flux-limited-only Ly\( \alpha \) sample and use it as baseline (black). We then show the region where \( W_{\text{Ly}\alpha} > 5 \text{ Å} \) (1σ; blue) and \( W_{\text{Ly}\alpha} > 20 \text{ Å} \) (1σ; green) objects lie. The \( W_{\text{Ly}\alpha} > 5 \text{ Å} \) contours are intended to reproduce Sobral et al. (2017) selections, whereas the more restrictive selection \( W_{\text{Ly}\alpha} > 20 \text{ Å} \) is representative of multiple surveys (Gronwall et al. 2007; Hagen et al. 2014; Vargas et al. 2014). As confirmed by this plot, galaxies with higher characteristic \( W_{\text{Ly}\alpha} \) (i.e., low \( M_{*} \), low SFR) are more likely to be selected by narrowband samples, even though the effect is minor for the typical narrowband cuts of \( W_{\text{Ly}\alpha} \leq 20 \text{ Å} \). Still, selections based solely on \( W_{\text{Ly}\alpha} \) systematically fail to remove AGNs and neglect bright Ly\( \alpha \) emitters (Sobral et al. 2017). All these insights remark on the importance of using low \( W_{\text{Ly}\alpha} \) cuts and thorough interloper controls.

After narrowband outcomes are used for sample selection, spectroscopic observations follow. However, these follow-ups of Ly\( \alpha \) emitting galaxies are flux limited, just like the ones presented in this work. Still, as we consider completeness in our modeling and simulations of Section 5.5, 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 \( f_{\text{UV}} \) for every object, we obtain the probability of \( L_{\text{Ly}\alpha} > L_{\text{Ly}\alpha}^{*} \) for every galaxy. The outcome (bottom panel of Figure 15) reveals that flux selections bias samples toward high-SFR LAEs.

Discrepancies in the location of LAEs in the \( M_{*}\text{-SFR} \) plane have already been observed in the literature. In Figure 10 of their paper, Hagen et al. (2014) plot the \( M_{*}\text{-SFR} \) relation of their \( z \sim 2-3 \) LAEs alongside the \( z \sim 2.1 \) counterparts of Vargas et al. (2014). The Hagen et al. (2014) objects are part of the HETDEX Pilot Survey (Adams et al. 2011), for which detections are constrained to line fluxes \( > 7-10 \times 10^{-17} \text{ erg cm}^{-2} \text{s}^{-1} \) (5σ) and \( W_{\text{Ly}\alpha} > 20 \text{ Å} \). In contrast, the Vargas et al. (2014) flux depth is higher than that of HETDEX, reaching \( 2 \times 10^{-17} \text{ erg cm}^{-2} \text{s}^{-1} \) (5σ), while also selecting sources with \( W_{\text{Ly}\alpha} > 20 \text{ Å} \). Therefore, Hagen et al. (2014) are comparing two \( W_{\text{Ly}\alpha} \)-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 (see \( z \approx 10^{9} M_{\odot} \) galaxies in Figure 15). 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 5.5. This explanation is backed by the fact that these discrepancies are not caused by inconsistent \( M_{*} \) or SFR derivations. Both studies assume a Salpeter (1955) IMF, cSFHs, and a fixed metallicity of \( Z = 0.2 Z_{\odot} \).

Vargas et al. (2014) and Hagen et al. (2014) find \( M_{*} < 10^{9} M_{\odot} \) LAEs to lie above the \( M_{*}\text{-SFR} \) relation. A similar trend is reported in the work of Karman et al. (2017), who study Ly\( \alpha \) emitters down to \( 10^{6} M_{\odot} \). Karman et al. (2017) point out that this offset can be a consequence of how uncertain SFRs are for low-\( M_{*} \) starburst galaxies. This explanation is consistent with the fact that most of our \( M_{*} < 10^{9} M_{\odot} \) Ly\( \alpha \) emitters are constrained to the lowest age locus of the \( M_{*}\text{-SFR} \) plane (see Figure 6). Moreover, Finkelstein et al. (2015) observe the same lowest age, low-\( M_{*} \) Ly\( \alpha \) emitters at \( z \approx 4.5 \). To answer whether low-\( M_{*} \) LAEs lie above star-forming galaxies or the slope of the sequence changes toward \( M_{*} < 10^{9} M_{\odot} \), different approaches are required. For instance, no discrepancies are found between \( z \approx 2 \) LAEs and H\( \alpha \) sources (see Matthee et al. 2016), although most of their galaxies have \( M_{*} > 10^{9} M_{\odot} \). In summary, evidence suggests that \( z > 2, M_{*} < 10^{9} M_{\odot} \) LAEs lie above extrapolations of the \( M_{*}\text{-SFR} \) relation. It is unclear whether this offset is real or the position of the relation at low \( M_{*} \) is far from certain.
The Astrophysical Journal, 843:133 (23pp), 2017 July 10

Oyarzún et al.

7. Inferences on the $4 \leq z \leq 7$ Lyα 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, 2011; Ono et al. 2012; Schenker et al. 2012; Tilvi et al. 2014; Cassata et al. 2015), bringing us closer to the goal of constraining the epoch of reionization. The Lyα fraction 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; Treu et al. 2013; Tilvi et al. 2014; Furusawa et al. 2016). Along these lines, our characterization of $W_{\text{Ly\alpha}}$ on the $M_{\text{UV}}$–$\beta$ plane can be used to simulate $W_{\text{Ly\alpha}}$ distributions at higher redshift. In this section, we apply the observed $M_{\text{UV}}$–$\beta$ relations from Bouwens et al. (2014) and LFs from Bouwens et al. (2015) to simulate high-redshift $W_{\text{Ly\alpha}}$ distributions. By means of these simulations, we can predict the Lyα fraction in galaxies up to $z \sim 7$, providing the first semi-analytical constraint to this tracer toward the reionization epoch. Furthermore, we also simulate dropouts from CANDELS photometry to explore the effects of observational limitations on the inferred Lyα fractions at high redshift.

Our results from this section are based on assuming the same $z \sim 4$ $W_{\text{Ly\alpha}}$ dependence on $M_{\text{UV}}$–$\beta$ at every redshift. We remark that testing this assumption with currently available data sets is challenging, since different sample selections and incompleteness levels can affect the observed $W_{\text{Ly\alpha}}$ dependence on the $M_{\text{UV}}$–$\beta$ plane. It is also worth noting that the analysis described in this section does not account for any effects related to changes in the merger fraction, IGM opacity (e.g., Gunn & Peterson 1965; Becker et al. 2001; Fan et al. 2006), and/or conditions of the ISM with cosmic time (e.g., Carilli & Walter 2013). As the number of complete, unbiased Lyα surveys grows (e.g., Cassata et al. 2015; Hathi et al. 2016; this work), we will be able to test these assumptions and explore their dependence on redshift.

We now detail our simulation of high-redshift LBG samples. For the LFs, we use Bouwens et al.’s (2015) best Schechter parameters (excluding CANDELS-EGS). We start by drawing objects following the LF and then associate a UV slope according to the best-fit relations from Table 3 of Bouwens et al. (2014). We finish by adding an intrinsic scatter of 0.35 to the slopes at every redshift (Bouwens et al. 2012, 2014; Castellano et al. 2012). Similarly, we perform an analogue procedure following the $M_{\text{UV}}$ distribution of CANDELS LBGs, which allows us to assess the effect of magnitude incompleteness. In order to do so, we make use of the optical and IR photometry publicly available from the 3D-HST catalogs (Skelton et al. 2014). For every redshift, we use the selections from Bouwens et al. (2015), since they are based on CANDELS photometry:

$$z \sim 4: (B_{435} - V_{606} > 1) \land (i_{775} - J_{125} < 1)$$

$$\land (B_{435} - V_{606} > 1.6(i_{775} - J_{125} + 1))$$

$$z \sim 5: (V_{606} - i_{814} > 1.3) \land (i_{814} - H_{160} < 1.25)$$

$$\land (V_{606} - i_{814} > 0.72(i_{814} - H_{160}) + 1.3)$$

$$z \sim 6: (i_{814} - J_{125} > 0.8) \land (J_{125} - H_{160} < 0.4)$$

$$\land (i_{814} - J_{125} > 2(J_{125} - H_{160}) + 0.8)$$

$$z \sim 7: (i_{814} - J_{125} > 2.2) \land (J_{125} - H_{160} < 0.4)$$

$$\land (i_{814} - J_{125} > 2(J_{125} - H_{160}) + 2.2).$$

Similarly to our Section 6.1 selection, we impose S/N cuts in the detection bands. For $z \sim 4$ and 5 dropouts, we require 3σ detections in the $V_{606}$, $z_{850}$ bands. For $z \sim 6$ and 7 dropouts, we impose at least 3σ detections in the $Y_{102}$ and $J_{125}$ bands, respectively. To associate a dropout redshift with every galaxy, however, we use the photometric redshifts from 3D-HST instead of the actual dropout from where it was selected. By doing so, we are in agreement with the methodology of Bouwens et al. (2015). We finally complete these CANDELS LBG samples by associating UV slopes to every galaxy. Just as for the complete sample, we associate the slopes following the $M_{\text{UV}}$–$\beta$ relations from Bouwens et al. (2014).
distribution negligible. According to our Equations (27) and (28), this implies that the probability of observing $W_{\text{Ly}\alpha}$ in emission at fixed $M_{\text{UV}}$ should increase with redshift. The right panel of Figure 16 shows the absolute magnitude distribution of $4 \leq z \leq 7$ LBGs as found by Bouwens et al. (2015; solid lines). Note the growth of the LF as a function of redshift. Since the LF becomes steeper as the redshift increases, the probability of observing a high-$W_{\text{Ly}\alpha}$ galaxy should also increase. One more important point regarding Figure 16 must be noted. As initially suggested, CANDELS incompleteness affects the observed $M_{\text{UV}}$ distribution of LBGs (shown as crosses). The fact that there is more contribution from brighter galaxies could, in principle, bias samples toward lower $M_{\text{UV}}$ when compared to the LFs. We now focus on quantifying all these effects within the framework of our model.

We simulate the $W_{\text{Ly}\alpha}$ distributions for the two sample sets we constructed. To do so, we use the $M_{\text{UV}}$ and $\beta$ of each object to draw $W_{\text{Ly}\alpha}$ distributions following our sampling of Equations (27) and (28). This methodology allows us to simulate the $W_{\text{Ly}\alpha}$ distribution for the number of galaxies in every redshift subsample, properly taking into account the uncertainties on each $M_{\text{UV}}$ domain and the errors in our modeling. We present the corresponding $4 \leq z \leq 7$ $\text{Ly}\alpha$ fractions for such galaxies in Figure 17. To this end, we select using the thresholds $W_{\text{Ly}\alpha}^*=25$ Å and $W_{\text{Ly}\alpha}^*=55$ Å (Stark et al. 2011). Still, as we note in Section 5.3, LBGs with $M_{\text{UV}}$ distributions dominated by fainter objects can yield higher fractions of $\text{Ly}\alpha$. In recent works, characterization has been performed separately in galaxies with $M_{\text{UV}} < -20.25$ and $M_{\text{UV}} > -20.25$ (Stark et al. 2011; Ono et al. 2012; Schenker et al. 2012; Tilvi et al. 2014; Cassata et al. 2015). Following the same approach, we now focus our analysis on samples in the ranges $-21.75 < M_{\text{UV}} < -20.25$ and $-20.25 < M_{\text{UV}} < -18.75$ (Figure 16), which are well covered by our data set (see Figure 12).

We plot in Figure 17 the resulting $\text{Ly}\alpha$ fractions. We find the complete fractions (i.e., derived from the LFs; black circles) to be consistent with a steady increase up to $z \sim 7$, which has been previously assumed as baseline for measuring drops in the $\text{Ly}\alpha$ fraction (Stark et al. 2010, 2011; Ono et al. 2012). Apart from backing this assumption, our result provides a more robust baseline for comparison. We reiterate that a closer look at Figure 16 reveals the explanation for such a trend, at least within our approach. First, the LF of LBGs becomes steeper as redshift increases (Bouwens et al. 2015). Second, the $M_{\text{UV}}-\beta$ relation favors bluer slopes as redshift increases. The combination of both relationships explain an increase in the $\text{Ly}\alpha$ fraction as a function of redshift for LBGs. It has already been discussed (Blanc et al. 2011; Cassata et al. 2015) how the growth in the $\text{Ly}\alpha$ fraction is likely tied to changes in the $f_{\text{esc}}$ of $\text{Ly}\alpha$ photons, at least for $3 < z < 5$. This change in $f_{\text{esc}}$ is likely tied to evolution in the dust content and ISM of galaxies, which we are empirically accounting for in this model. Under this picture, Lyman continuum (LyC) photon escape should follow a similar trend. Indeed, Faisst (2016) empirically predicts that LyC photon escape fraction increases with redshift at fixed $M_*$.
First, our characterization is based on the whole distribution, which may misrepresent the actual fraction of galaxies with haloes. Indeed, our fractions for populations with significant incompleteness artificially lowers the Lyα fraction. This effect must be accounted for when associating Lyα drops with reionization. Keep in mind that our model assumes an exponential Lyα distribution, which might underestimate the fraction for the tail of the distribution. Therefore, Lyα studies of even the deepest LBG surveys are sufficiently incomplete to bias the reported Lyα fractions (see Figures 16 and 17). Taking into account this effect is essential if we are to use Lyα to further constrain the reionization epoch during the next decade.

It is important to clarify the limitations of our modeling. First, our characterization is based on the whole Lyα distribution, which may misrepresent the actual fraction of high-Lyα galaxies. In fact, the analysis we perform in Section 4 reveals that a log-normal profile might be better for reproducing the tail of the distribution (Figure 5). For example, Zheng et al. (2014) find an exponential profile to only adequately represent the low-Lyα end of the distribution, leading them to model the high-Lyα tail separately. As a consequence, we encourage the reader to focus most of their analysis on our \( W_{\text{Ly}\alpha}=25 \) Å results (top panel in Figure 17). Along those lines, we expect our modeling to underestimate the fraction for populations with significantly higher-Lyα tails. Indeed, our fractions for galaxies with \(-20.25 < M_{\text{UV}} < -18.75\) are systematically lower than Stark et al. (2011) measurements (see Figure 17).

In summary, our results suggest that the Lyα fraction steadily increases up to \( z \sim 7 \) in an ionized universe. However, drops in this fraction are not necessarily related to changes in the opacity of the IGM, since we find that sample \( M_{\text{UV}} \) incompleteness can alternatively explain them. In fact, the uncertainties in both literature measurements and our predictions are still large enough to set assertive constraints on the Lyα trends toward \( z \sim 7 \) (see Robertson et al. 2015). Our results, therefore, further highlight that Lyα constraints on \( z_{\text{re}} \) (the redshift at which the fraction of ionized hydrogen is 0.5) are still inconclusive.

It is worth discussing whether \( z_{\text{re}} > 7 \) is in agreement with observational reionization constraints besides Lyα. Recently, the Planck Collaboration et al. (2016a) analysis of CMB anisotropies has yielded \( z_{\text{re}} \sim 8.2_{-1.2}^{+1.0} \). Systematic detection of Lyα-emitting galaxies in the range \( z \sim 7-9 \) (Zitrin et al. 2015; Furusawa et al. 2016; Stark et al. 2017) still leads to the question whether reionization is already in place at \( z \sim 7 \). Several studies have also explored Lyα LF behavior at \( z \sim 9 \), although observational limitations require deeper/wider surveys to obtain more meaningful conclusions (Sobral et al. 2009). Therefore, and since reionization is predicted to happen so rapidly (Finlator et al. 2009), claims of \( z_{\text{re}} > 7 \) are not in tension with measurements of the Lyα fraction at \( z \sim 7 \).

Interestingly enough, it is not even clear whether reionization happens first around fainter, lower density environments (“outside-in,” Kashikawa et al. 2006), or brighter, higher density regions (“inside-out,” Santos et al. 2016). This overall picture suggests that the role of \( M_{\text{UV}} \) in reionization studies will no longer restrict itself to sample control, but it will also extend to parameterizing property. Therefore, the natural approach for future studies would be to constrain reionization as a smooth
8. Conclusions

In this work, we present an exhaustive analysis of Ly$\alpha$ emission at $3 < z < 4.6$. To this end, we $M^*_a$-select 625 galaxies from the CANDELS survey, allowing us to study Ly$\alpha$ 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 6.5 m telescope. We then use a Bayesian approach for proper statistical handling of our results. By means of this framework, we are capable of characterizing Ly$\alpha$ emission in the high-redshift galaxy population. In summary, our conclusions are the following.

1. We present 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 combine this approach with Monte Carlo simulations for robust modeling of observed trends. Combining all of these features allows us to properly state the significance and uncertainties in our results.

2. We take advantage of the Bayesian framework to give an insight into how to compare multiple $W_{Ly\alpha}$ distribution models from a quantitative standpoint. In this paper, we explore the extent to which exponential, Gaussian, and log-normal probability distributions can reproduce $W_{Ly\alpha}$ measurements. We conclude that, in our case, an exponential profile is the most adequate. However, as we find in Section 5, this profile can struggle to reproduce the tail of high-$W_{Ly\alpha}$ populations (e.g., samples with low $M^*_a$, SFR, UV luminosity, or $\beta$).

3. Our measured $W_{Ly\alpha}$ and $f_{esc}$ strongly anti-correlate with $M^*_a$. We associate these trends with the higher dust fraction and gas mass in more massive galaxies, boosting the scattering and absorption of Ly$\alpha$ photons. We model both exponential $W_{Ly\alpha}$ distribution parameters ($A$, $W_0$) using linear relations dependent on $M^*_a$, and find them both to anti-correlate with $M^*_a$. Our modeling is also capable of reproducing our observed $W_{Ly\alpha}$ distributions when binning the sample (Figure 8).

4. We also explore the dependence of Ly$\alpha$ emission on $M_{UV}$. We find $W_{Ly\alpha}$ and $f_{esc}$ to be typically higher for UV faint objects, which have been previously observed in the literature (Stark et al. 2010; Schaerer et al. 2011). Since $z \approx 4$ galaxies seem to follow an $M^*_a - M_{UV}$ sequence (González et al. 2014), this result is consistent with our $M^*_a$ trends. This confirms that the role played by higher dust fraction and gas mass rules $W_{Ly\alpha}$ dependence on UV luminosity. We also observe $W_{Ly\alpha}$ and $f_{esc}$ to anti-correlate with the SFR and UV slope, confirming the qualitative Ly$\alpha$ escape scheme presented.

5. The relatively uniform location of our targets in the $M_{UV} - \beta$ 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 $W_{Ly\alpha}$ distribution parameters. Our results suggest that the probability of observing Ly$\alpha$ emission from a galaxy is determined by its dust content, whereas the actual magnitude of $W_{Ly\alpha}$ is mostly ruled by the UV luminosity of the object. This characterization also allows us to simulate $W_{Ly\alpha}$ distributions from $M_{Ly\alpha}$- galaxy samples (see below).

6. Using CANDELS, we mimic a $z \approx 4$ LBG survey based on B-dropsouts. We simulate $W_{Ly\alpha}$ distributions from the LBG sample and verify that the lower probability of measuring high $W_{Ly\alpha}$ when compared to the population is mostly a consequence of detection band limitations. In other words, shallower LBG surveys are bound to include fewer low-$M^*_a$ galaxies in their samples, decreasing the probability of including high-$W_{Ly\alpha}$ objects. We find color constraints imposed by the LBG technique to also lower the Ly$\alpha$ fraction, although the effect is minor. These results highlight the importance of comparing surveys with similar sensitivities and $M_{UV}$ distributions.

7. Our results on the $M_{UV} - \beta$ plane also allow us to explore the effects $W_{Ly\alpha}$ and line-flux limitations induce on Ly$\alpha$ narrowband surveys. We find $W_{Ly\alpha}$ cuts bias samples toward low-$M^*_a$ objects, whereas flux limitations seem to preferentially leave out low-SFR galaxies. These insights contribute to explain the location of LAEs within the main sequence (e.g., some of the differences between Vargas et al. 2014 and Hagen et al. 2014).

8. We generate $4 \leq z \leq 7$ LBG samples following reported $M_{UV} - \beta$ relations and LFs from the literature (Bouwens et al. 2014, 2015). Assuming the same $W_{Ly\alpha}$ dependence on $M_{UV} - \beta$ that we find at $z \approx 4$, we estimate the fraction of Ly$\alpha$-emitting galaxies in LBG samples above an $W_{Ly\alpha}$ threshold. Our findings are consistent with observational measurements in the literature, that suggest an increase in the fraction up to $z \approx 7$ (Stark et al. 2010, 2011; Cassata et al. 2015). This result constitutes the first semi-analytical constraint to the Ly$\alpha$ fraction at $z \approx 7$, replacing extrapolations of lower redshift regimes.

9. We simulate the $4 \leq z \leq 7$ Ly$\alpha$ fraction in LBG samples drawing from the CANDELS $M_{UV}$ distribution. We conclude that $M_{UV}$ incompleteness can lower the Ly$\alpha$ fraction, reproducing drops at $z \approx 7$. Claims of drops at this redshift need to thoroughly account for this effect if they are to be used to constrain the reionization epoch. This result highlights that reported drops in the Ly$\alpha$ fraction are not inconsistent with $z_{re} > 7$. Given the uncertainties in both predictions and measurements, more work is needed to conclude on the actual Ly$\alpha$ trends toward $z \approx 7$ and beyond.

We thank the referee for useful feedback that improved the quality of this paper. G.O. was supported by CONICYT, Beca Magister Nacional 2014, Folio 22140924. G.B. was supported by CONICYT/FONDECYT, Programa de Iniciación, Folio 11150220. V.G. was supported by CONICYT/FONDECYT, Programa de Iniciación, Folio 11160832. We also acknowledge the NSF/MRI grant for the M2FS, number AST 0923160. 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.

References

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
Atek, H., Siana, B., Scarlata, C., et al. 2011, ApJ, 743, 121
Balestra, I., Mainieri, V., Poppesso, P., et al. 2010, A&A, 512, A12
