



Predicting Ly α Emission from Galaxies via Empirical Markers of Production and Escape in the KBSS*

Ryan F. Trainor¹ , Allison L. Strom² , Charles C. Steidel³ , Gwen C. Rudie² , Yuguang Chen³, and Rachel L. Theios^{3,4} ¹Department of Physics & Astronomy, Franklin & Marshall College, 415 Harrisburg Pike, Lancaster, PA 17603, USA; ryan.trainor@fandm.edu²Carnegie Observatories, 813 Santa Barbara St., Pasadena, CA 91101, USA³Cahill Center for Astrophysics, MC 249-17, 1200 E. California Blvd., Pasadena, CA 91125, USA

Received 2019 August 13; revised 2019 September 24; accepted 2019 September 29; published 2019 December 12

Abstract

Ly α emission is widely used to detect and confirm high-redshift galaxies and characterize the evolution of the intergalactic medium (IGM). However, many galaxies do not display Ly α emission in typical spectroscopic observations, and intrinsic Ly α emitters represent a potentially biased set of high-redshift galaxies. In this work, we analyze a set of 703 galaxies at $2 \lesssim z \lesssim 3$ with both Ly α spectroscopy and measurements of other rest-frame ultraviolet and optical properties in order to develop an empirical model for Ly α emission from galaxies and understand how the probability of Ly α emission depends on other observables. We consider several empirical proxies for the efficiency of Ly α photon production, as well as the subsequent escape of these photons through their local interstellar medium. We find that the equivalent width of metal-line absorption and the O3 ratio of rest-frame optical nebular lines are advantageous empirical proxies for Ly α escape and production, respectively. We develop a new quantity, $X_{\text{LIS}}^{\text{O3}}$, that combines these two properties into a single predictor of net Ly α emission, which we find describes $\sim 90\%$ of the observed variance in Ly α equivalent width when accounting for our observational uncertainties. We also construct conditional probability distributions demonstrating that galaxy selection based on measurements of galaxy properties yield samples of galaxies with widely varying probabilities of net Ly α emission. The application of the empirical models and probability distributions described here may be used to infer the selection biases of current galaxy surveys and evaluate the significance of high-redshift Ly α (non)detections in studies of reionization and the IGM.

Unified Astronomy Thesaurus concepts: Lyman-alpha galaxies (978); Galaxies (573); High-redshift galaxies (734)

1. Introduction

The Ly α line of hydrogen is a powerful tool for detecting and characterizing high-redshift galaxies. Large samples of galaxies have been selected through Ly α emission via narrowband surveys (e.g., Cowie & Hu 1998; Rhoads et al. 2000; Steidel et al. 2000; Trainor et al. 2015; Ouchi et al. 2018) or IFU spectroscopy (e.g., Bacon et al. 2015). Likewise, its strength in emission or absorption makes Ly α extremely efficient for spectroscopically confirming the redshifts of galaxies selected by other means, including broadband imaging surveys.

Ly α also provides valuable information about the properties of galaxies and their surrounding gas, in both emission and absorption (e.g., Rakic et al. 2012; Rudie et al. 2012). Ly α emission clearly signifies the presence of embedded star formation and/or active galactic nucleus (AGN) activity⁵ in a galaxy, and resonant scattering of Ly α photons can cause this light to trace the gas

distribution on scales comparable to the virial radius of the galaxy halo (e.g., Steidel et al. 2011; Momose et al. 2014; Wisotzki et al. 2016), or even beyond the halo radius for very luminous quasars (e.g., Cantalupo et al. 2014; Martin et al. 2015). In addition, the apparent Ly α emission from galaxies at the highest redshifts is a useful diagnostic of the intergalactic medium (IGM): the apparent drop-off in the fraction of galaxies exhibiting strong Ly α emission at $z \gtrsim 6-7$ (e.g., Pentericci et al. 2011; Schenker et al. 2012) likely points to the increasing neutral fraction at this epoch, and this evolution thus constrains the tail end of cosmic reionization (Robertson et al. 2015).

However, the same physical processes of emission and scattering that make Ly α such a promising tool also introduce significant challenges to its utility. Because strong Ly α emission facilitates efficient galaxy detection and redshift confirmation—but not all galaxies exhibit strong Ly α emission—there are potential selection biases both in Ly α -selected galaxy samples and in samples of broadband-selected galaxies that are vetted through rest-UV spectra. In particular, star formation is a necessary but insufficient condition for detectable Ly α emission; only $\sim 50\%$ of L_* galaxies at $z \sim 3$ show Ly α in net emission in slit spectroscopy (Shapley et al. 2003; Steidel et al. 2011), although this fraction appears to increase toward lower galaxy masses and continuum luminosities (e.g., Stark et al. 2013; Oyarzún et al. 2016). As such, Ly α -selected (or Ly α -confirmed) galaxy samples will lack non-star-forming galaxies, as well as a large fraction of star-forming galaxies. Resonant scattering, as well as the “fluorescent” generation of Ly α recombination emission, can be used to map out extended H I illuminated by an external or internal engine, but this same scattering often serves to impede the

* Based on data obtained at the W. M. Keck Observatory, which is operated as a scientific partnership among the California Institute of Technology, the University of California, and NASA, and was made possible by the generous financial support of the W. M. Keck Foundation.

⁴ Carnegie Fellow.

⁵ While an external ionizing field can in principle illuminate “dark galaxies” unpolluted by star formation, current detection limits prohibit the detection of these pristine halos in all but the most extreme environments (see, e.g., Cantalupo et al. 2005; Kollmeier et al. 2010; Cantalupo et al. 2012; Trainor & Steidel 2013). In general, Ly α is therefore an effective tracer of *local* ionizing photon production.

identification of the size, location, and intrinsic properties of the energizing source.

Finally, our incomplete knowledge of the galaxy-scale determinants of strong Ly α emission impedes our ability to use it as an IGM tracer. Given that many star-forming galaxies do not show strong Ly α emission at $z \sim 3$, where the neutral fraction of the IGM is minimal, it is not always clear which galaxies at $z \gtrsim 6-7$ are intrinsic emitters of Ly α photons and whether their lack of apparent emission indicates suppression by the IGM. Recent studies of Ly α as a tracer of IGM opacity have been careful to compare galaxies at similar redshift epochs, where intrinsic galaxy evolution is likely to be minimal (e.g., from $z \sim 7$ to $z \sim 6$; Mason et al. 2018; Pentericci et al. 2018; Hoag et al. 2019). However, both the overall trend in increasing intrinsic Ly α emission as a function of redshift (e.g., Stark et al. 2010) and individual measurements of strong Ly α emission even at $z > 7$ (Roberts-Borsani et al. 2016; Stark et al. 2017) indicate that the variation of intrinsic Ly α emission among galaxies is important to understand for a full accounting of the reionization process. Similarly, the uncertain connections between Ly α emission and physical properties of galaxies prevent the clear identification of the selection biases intrinsic to Ly α selection and redshift confirmation, even as we expect some such biases to be present.

Advancing the utility of Ly α emission as a tool for detecting and characterizing galaxies therefore requires a model for understanding—and potentially predicting—when this emission is expected based on a galaxy’s other properties. Any such model of galaxy-scale Ly α emission must include two disparate sets of processes: (1) the production of Ly α photons in H II regions, and (2) the subsequent transmission (or absorption) of these photons through the surrounding HI gas.

The latter set of processes—those pertaining to Ly α scattering, transmission, and escape from galaxies—have been subject to detailed study for two decades, eased in part by the availability of interstellar medium (ISM) diagnostics near Ly α in the rest-UV spectra of galaxies. In particular, much work has demonstrated that Ly α emission (parameterized by the Ly α equivalent width, EW_{Ly α}) correlates strongly with the optical depth or covering fraction of Lyman-series lines or low-ionization metal lines (hereafter LIS lines, parameterized by EW_{LIS}), in the sense that galaxies with stronger Ly α emission exhibit weaker Lyman-series or LIS absorption (Kunth et al. 1998; Shapley et al. 2003; Steidel et al. 2010; Jones et al. 2012; Trainor et al. 2015; Du et al. 2018). Weak absorption lines are likely an indicator that the covering fraction or optical depth of HI gas is relatively small, and in some cases the covering fraction of LIS and/or Lyman-series lines have been shown to be much less than unity, particularly for Ly α -emitting or Lyman-continuum-emitting galaxies (Trainor et al. 2015; Steidel et al. 2018).⁶

In a related phenomenon, strong Ly α emission lines are found to be narrow in velocity space and close to the systemic redshift of the emitting galaxy (e.g., Erb et al. 2014; Trainor et al. 2015), as well as spatially compact (e.g., Steidel et al. 2011; Momose et al. 2016; but see Wisotzki et al. 2016). These relationships suggest that net Ly α emission is maximized when these photons can escape their parent galaxy with minimal scattering both in velocity and in physical space.

Common among each of the above observables—LIS and Lyman-series absorption, spatial and spectral scattering—is

that they are associated with modulation of the Ly α photons that occurs after these photons have left their original star-forming regions. Conversely, more recent studies have identified trends linking EW_{Ly α} to signatures of the star-forming regions themselves. Much of this recent work has been enabled by the development of efficient near-infrared spectrometers such as MOSFIRE (McLean et al. 2012) and the *Hubble Space Telescope* (HST)/WFC3-IR grisms (e.g., Atek et al. 2010; Brammer et al. 2012) that can detect the faint rest-frame optical emission lines used to characterize the gas around young stars in high- z galaxies.

McLinden et al. (2011), Finkelstein et al. (2011), and Nakajima et al. (2013) found strong [O III] lines in a total of eight Ly α -selected galaxies at $z \sim 2-3$, while Song et al. (2014) localized 10 Ly α -selected galaxies in the N2-BPT⁷ plane, suggesting that high-redshift galaxies with strong Ly α emission have low metallicities and high nebular excitation. Trainor et al. (2016) found similarly extreme N2-BPT line ratios (i.e., high O3 and low N2) for a stacked sample of 60 $L \sim 0.1L_*$ galaxies with strong Ly α emission at $z \sim 2.5$, while Erb et al. (2016) and Hagen et al. (2016) demonstrated that galaxies selected for strong [O III] emission and/or weak N2 ratios are strong Ly α emitters and share many physical properties with Ly α -selected galaxies.

In addition to presenting results for faint Ly α -emitting galaxies, Trainor et al. (2016) also show that galaxies ranging from strong Ly α emitters to Ly α absorbers can be described as a sequence in the N2-BPT plane, a phenomenon that appears to be primarily linked to the variation in the excitation state of gas in star-forming H II regions, rather than to extreme variation in gas-phase metallicity. In that paper, we also argue that Ly α emission and high nebular excitation are linked by their association with strong sources of ionizing emission within galaxies, including massive stars with low Fe abundances as discussed at length by Steidel et al. (2016) and Strom et al. (2017). Taken together, the results of these rest-frame optical studies are consistent with the expectation that Ly α production is accompanied by numerous other forms of recombination emission and collisionally excited emission that originate in the same star-forming regions as Ly α , although the subsequent transmission of these non-Ly α photons is much less sensitive to the surrounding HI distribution.

It is therefore clear that the net Ly α emission on galaxy scales depends on both the properties of star-forming regions (the sites of Ly α production) and the distribution of the surrounding HI gas that modulates Ly α escape. Here we propose a holistic, empirical framework for accounting for both of these processes. Using the largest sample of galaxies with simultaneous spectroscopic measurements of Ly α , the rest-UV continuum, and a series of rest-frame optical transitions, we identify empirical discriminants of Ly α production and escape, and we demonstrate that the combination of these observable markers can predict the net Ly α emission of galaxies more reliably than individual galaxy properties.

The paper is organized as follows: Section 2 presents details of our galaxy observations and the assembly of our sample, Section 3 describes the methods used to quantify our empirical markers of the efficiency of Ly α production and escape in a given galaxy and their correlations with EW_{Ly α} , Section 4 presents our combined model for predicting Ly α based on multiple markers,

⁶ Note that the HI covering fraction inferred from Lyman-series lines may not be the same as the LIS covering fraction (see, e.g., Henry et al. 2015), but both are found to decrease on average with increasing EW_{Ly α} .

⁷ The N2-BPT compares $\log([\text{N II}] \lambda 6583/\text{H}\alpha)$ [N2] vs. $\log([\text{O III}] \lambda 5007/\text{H}\beta)$ [O3], an emission-line diagnostic similar to those presented by Baldwin et al. (1981), but introduced by Veilleux & Osterbrock (1987) in its modern form.

Section 5 presents the conditional probability distribution of detecting net Ly α emission as a function of other galaxy properties, Section 6 provides discussion comparing our results to previous work, and Section 7 summarizes our conclusions.

2. Observations

2.1. Galaxy Sample

The galaxies presented here are selected from the Keck Baryonic Structure Survey (KBSS; Rudie et al. 2012) and KBSS-MOSFIRE (Steidel et al. 2014; Strom et al. 2017), which together comprise a set of rest-UV and rest-optical spectra of more than 1000 galaxies across 15 fields at $1.5 \lesssim z \lesssim 3.5$. The galaxies are selected using optical colors in the U_n , G , and \mathcal{R} bands to identify Lyman break galaxy analogs over the target range of redshifts; a more detailed description of the photometric selection is given by Steidel et al. (2004). The full distributions of KBSS-MOSFIRE galaxies' redshifts, masses, and star formation rates (SFRs) are given in Strom et al. (2017), but they occupy the ranges $10^9 \lesssim M_*/M_\odot \lesssim 10^{11}$ and $3 \lesssim \text{SFR}/(M_\odot \text{ yr}^{-1}) \lesssim 300$, as inferred from reddened stellar population synthesis models⁸ fit to Keck LRIS and MOSFIRE broadband photometry. The galaxies have typical dark matter halo masses $M_h \approx 8 \times 10^{11} M_\odot$ estimated through galaxy-galaxy clustering measurements of the spectroscopic KBSS sample (Trainor & Steidel 2012). The median apparent magnitude of the KBSS-MOSFIRE sample is $\langle \mathcal{R} \rangle = 24.4$, which corresponds to an absolute magnitude $(M_{\text{UV}}) = -20.7$ ($\langle L \rangle \approx L_*$; Reddy et al. 2008) at the median redshift of the sample, $\langle z \rangle = 2.3$.

For this work, we select the subset of the full KBSS-MOSFIRE sample that have rest-UV spectroscopic coverage of the Ly α line and surrounding region ($1208 \text{ \AA} < \lambda_{\text{rest}} < 1227 \text{ \AA}$) from Keck/LRIS (Oke et al. 1995; Steidel et al. 2004), as well as a galaxy redshift measured from rest-optical MOSFIRE (McLean et al. 2012) spectroscopy (typically from the [O III] $\lambda\lambda 4959, 5008$ doublet and/or H α). Details of the rest-UV and rest-optical spectroscopic data are given in the sections below. The total subset of the KBSS-MOSFIRE sample used in this paper comprises 703 galaxies, although most of the individual empirical parameters described in Section 3 are not measured robustly for every galaxy. The requirements for making each measurement and the number of galaxies for which each is made are given explicitly in Section 3, with numbers also given in Table 1.

2.2. LRIS Observations

The rest-UV galaxy spectra were obtained with Keck I/LRIS-B (Steidel et al. 2004) in multislit mode over a series of observing runs between 2002 and 2016 August. Approximately 2/3 of spectra were taken using the 4000/3400 grism, which produces a resolution $R \sim 800$ in the typical seeing conditions of $0''.6\text{--}0''.8$.⁹ The remaining spectra were taken using the 600/4000 grism, which produces a resolution $R \sim 1300$ in the same seeing conditions. The blue edge of the LRIS-B observations is typically determined by the atmospheric limit at $\lambda \sim 3200 \text{ \AA}$,

⁸ The SED fitting is performed using Bruzual & Charlot (2003) solar-metallicity models, a Chabrier (2003) initial mass function, and a Calzetti et al. (2000) attenuation curve. For a full description of the fitting methodology, see Reddy et al. (2012).

⁹ LRIS slit masks for the observations were made with $1''.2$ slits, so the spectral resolution of the (typically spatially unresolved) galaxies in our spectroscopic sample is limited by the smaller seeing disk diameter.

Table 1
Correlations

Quantities		N_{gal} ^a	r_{Sp} ^b	$\log_{10}(p)$
<i>Escape-related Quantities</i>				
EW _{Lyα}	versus $v_{\text{Ly}\alpha}$	496	-0.56	-41.8
	versus EW _{LRIS}	669	0.35	-20.5
	versus $E(B - V)_{\text{SED}}$	637	-0.23	-8.2
	versus $E(B - V)_{\text{neb}}$	208	-0.14	-1.3
EW _{LRIS}	versus $v_{\text{Ly}\alpha}$	479	0.28	-9.5
	versus $f_{\text{esc,rel}}$	368	-0.42	-16.5
	versus $f_{\text{esc,abs}}$	188	-0.50	-12.4
<i>Production-related Quantities</i>				
EW _{Lyα}	versus $M_{*,\text{SED}}$	637	-0.15	-3.9
	versus SFR _{SED}	637	-0.17	-5.0
	versus SFR _{Hα}	208	-0.05	-0.4
	versus sSFR _{Hα}	199	0.23	-3.0
	versus M_{UV}	637	-0.01	-0.1
	versus O32 _{raw}	316	0.43	-15.0
	versus O32 _{corr}	174	0.47	-10.3
	versus O3	395	0.40	-15.3
<i>Production + Escape</i>				
EW _{LRIS}	versus O3	377	0.21	-4.9
	versus O32 _{corr}	174	0.47	-10.3
EW _{Lyα}	versus $X_{\text{LRIS}}^{\text{O3c}}$	377	0.49	-23.8

Notes.

^a Number of galaxies for which correlation is calculated.

^b Spearman correlation coefficient.

^c $X_{\text{LRIS}}^{\text{O3c}}$ is defined by Equation (12).

while the red edge is determined by a dichroic that splits the light between LRIS-B and LRIS-R; dichroics with transition wavelengths $\lambda_{\text{dich}} = 5000, 5600, \text{ or } 6800 \text{ \AA}$ were used to collect the spectra in this sample. These constraints, combined with the redshift distribution of our sample, allow us to sample the rest-frame spectrum of most objects over the range $1000 \text{ \AA} \lesssim \lambda_{\text{rest}} \lesssim 1700 \text{ \AA}$. Each object was typically observed for 1.5 hr of 1800 s integrations, but a subsample of ~ 150 galaxies included in the KBSS-LM1 project (Steidel et al. 2016) were each observed for $\sim 10\text{--}14$ hr.

LRIS-B spectra were reduced using a custom suite of IDL and IRAF routines. Raw 2D spectrograms were rectified using a slit-edge tracing algorithm, and the resulting rectilinear spectrograms were flat-fielded, background-subtracted, and subjected to cosmic-ray rejection. Individual 2D spectrograms were stacked after accounting for shifts in the spatial and spectral directions due to instrument flexure between exposures.¹⁰ A 1D spectrogram was extracted for each object, and wavelength calibration was performed using arc-line spectra observed through the slit mask during daytime telescope operations, which were then individually shifted to match the wavelength of the 5577 \AA sky line in each science spectrum in order to account for instrument flexure. Finally, spectra are corrected to vacuum, heliocentric wavelengths and rebinned to a common wavelength scale. Further details regarding the software routines used in the KBSS data reduction process are given by Steidel et al. (2003, 2010).

¹⁰ Note that MOSFIRE corrects the optical path internally to account for instrument flexure, so analogous shifts are not necessary for our MOSFIRE spectra (Section 2.3).

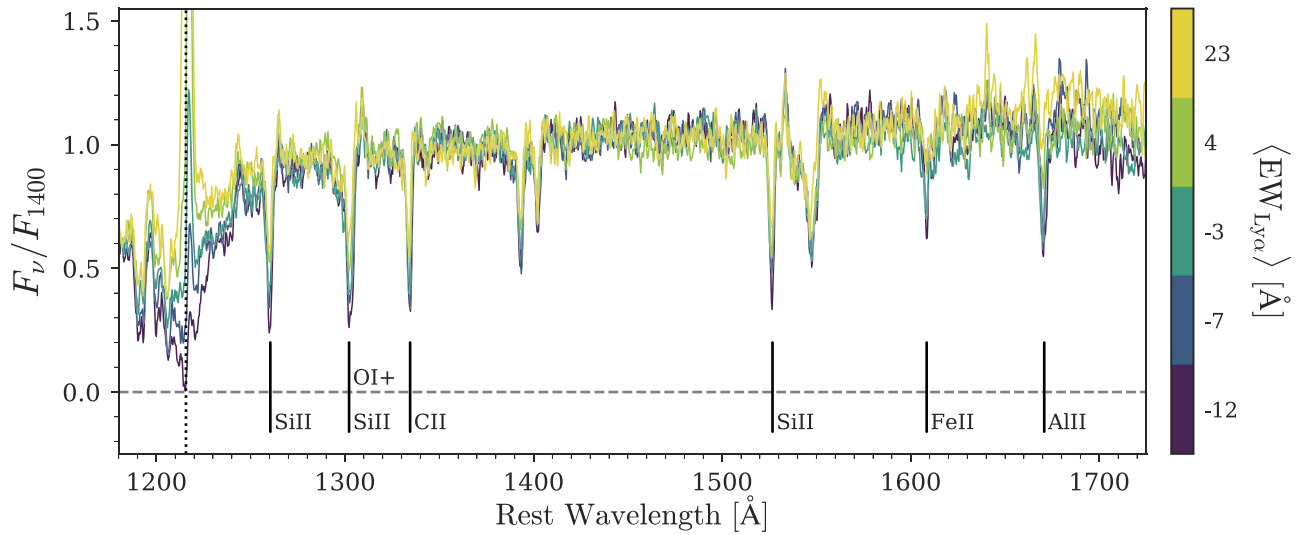


Figure 1. Stacked spectra of 703 galaxies sorted in quintiles of $EW_{Ly\alpha}$; the median $EW_{Ly\alpha}$ for each sample is given by the color bar on the right. The dashed vertical line indicates $Ly\alpha$, while the solid vertical lines indicate the LIS transitions described in Table 2 and Section 3.2.2. The LIS line strengths vary similarly across all six distinct absorption features, such that strong LIS absorption is associated with weak $Ly\alpha$ emission and vice versa.

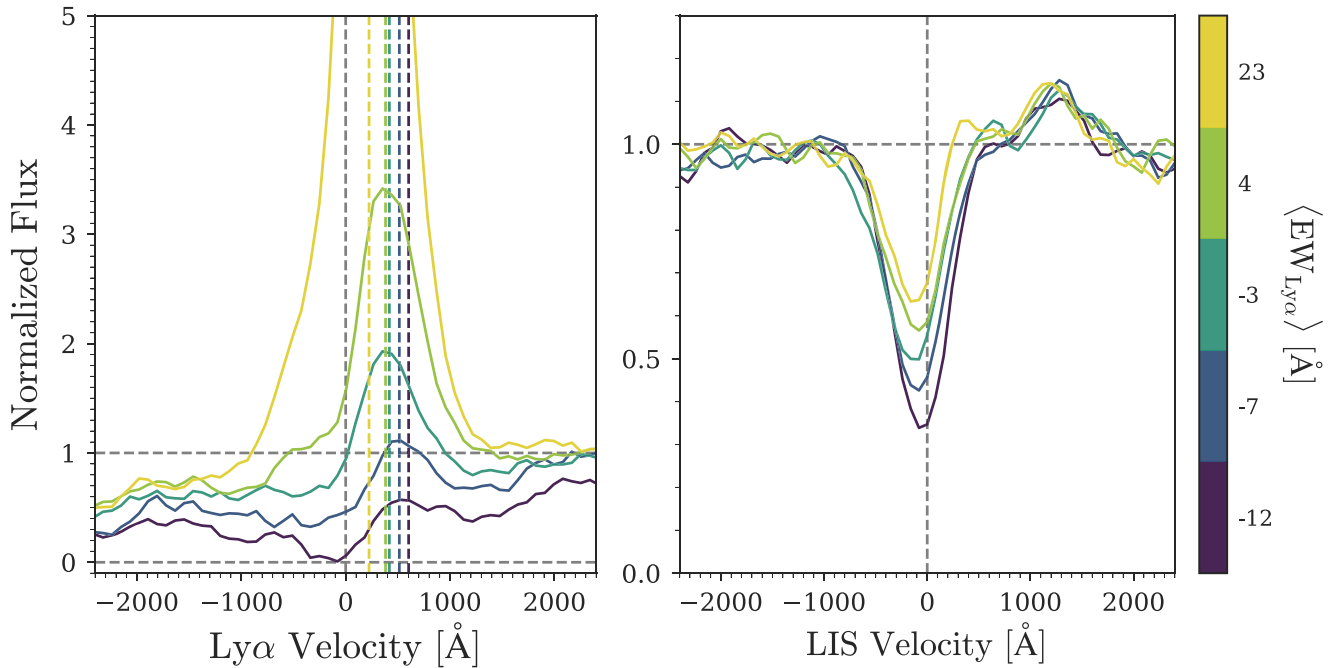


Figure 2. Stacked spectra in quintiles of $EW_{Ly\alpha}$, as in Figure 1. The left panel is zoomed in on the $Ly\alpha$ line, and the dashed colored lines indicate the velocity of the $Ly\alpha$ centroid for each stacked subsample. The continuum for each stack is normalized at $\lambda_{rest} = 1400 \text{ \AA}$ as in Figure 1. The right panel shows the combined LIS absorption profile for each stack, constructed by averaging the profiles of each of the LIS absorption features labeled in Figure 1 according to the process described in Section 3.2.2 including normalization to the local continuum. Note that the stacked profiles show the same trends described in the text for individual objects: increased $EW_{Ly\alpha}$ is strongly associated with decreased (i.e., less redshifted) $v_{Ly\alpha}$ and increased (i.e., weaker and less negative) EW_{LIS} .

Because many objects were observed multiple times over the period since observations of the KBSS galaxy sample began, all extant spectra for a given galaxy were averaged—weighted according to their exposure time and a visual inspection of each spectrum—to produce a final spectrum for each galaxy in our sample. Stacked LRIS spectra are displayed in Figures 1–2.

2.3. MOSFIRE Observations

The MOSFIRE observations for the KBSS-MOSFIRE survey are described in detail elsewhere (Steidel et al. 2014;

Strom et al. 2017). Briefly, galaxies are observed in the J , H , and/or K bands using 0.7 slits, and the resulting 2D spectrograms are reduced using the MOSFIRE data-reduction pipeline¹¹ provided by the instrument team. Wavelength calibration is performed by identifying OH sky lines in all spectral regions except in the red end of the K band ($\lambda_{obs} \gtrsim 2 \mu\text{m}$), where arc lamp spectra are used owing to the paucity of sky lines. The absolute flux scaling, slit-loss corrections, and cross-band calibration are performed through

¹¹ <https://github.com/Keck-DataReductionPipelines/MosfireDRP>

observations of a slit star on each mask, as discussed in detail by Strom et al. (2017). As described in that paper, the spatial extent of the KBSS galaxies (which are marginally resolved in typical atmospheric conditions) causes the slit losses for KBSS galaxies to exceed that measured directly from the slit star.

Line measurements are performed using the IDL program MOSPEC (Strom et al. 2017). Typically, galaxy redshifts and line fluxes are fit simultaneously in a single band (J , H , or K), with all nebular lines in the band constrained to have the same redshift and velocity width. For the [O III] $\lambda\lambda 4960, 5008$ doublet, the known line flux ratio $f_{5008}/f_{4960} = 3$ is also enforced. Line widths and redshifts are not forced to match between bands (e.g., for [O III] in the H band and $H\alpha$ in the K band, as would be observed at $z \approx 2-2.6$), but galaxies with redshift measurements in multiple bands are checked for consistency. Each galaxy is then assigned a nebular redshift z_{neb} based on the rest-frame optical redshift with the smallest uncertainty. For galaxies with redshift measurements in multiple bands, the typical agreement is less than $\Delta z = 0.0002$.

2.4. Spectral Energy Distribution (SED) Models and SFRs

Photometry of the KBSS fields and SED modeling of the KBSS galaxy sample are described by Steidel et al. (2014) and Strom et al. (2017) using SED fitting methodology described by Reddy et al. (2012). Models are from the Bruzual & Charlot (2003) library and assume solar metallicity, a Chabrier (2003) initial mass function (IMF), a Calzetti et al. (2000) attenuation relation, and a constant star formation history with a minimum age of 50 Myr. Stellar masses (M_*), SFRs (SFR_{SED}), and continuum-based reddening ($E(B-V)_{\text{SED}}$) estimates are obtained from the SED fitting.

In addition to SED-based SFRs, $H\alpha$ SFRs are calculated for the majority of the KBSS-MOSFIRE sample as described by Strom et al. (2017) using the MOSFIRE measurements described above. These SFRs and sSFRs ($\text{sSFR} = \text{SFR}_{H\alpha}/M_*$) are calculated assuming a Kennicutt (1998) $H\alpha$ -SFR relation adjusted for a Chabrier (2003) IMF. Dust corrections for $H\alpha$ SFRs are calculated as described in Section 3.2.3.

3. Empirical Galaxy Measurements

As described in Section 1, the primary goals of this paper are (1) to characterize the empirical relationships between $\text{Ly}\alpha$ emission and various other galaxy properties and (2) to interpret these relationships in terms of the production and escape of $\text{Ly}\alpha$ photons. Below, we define the various observables used throughout this paper, and we categorize them in terms of whether they are likely to primarily relate to the production or escape of $\text{Ly}\alpha$ photons.

3.1. $\text{EW}_{\text{Ly}\alpha}$

First, we must define a metric of the *efficiency* of $\text{Ly}\alpha$ emission: the $\text{Ly}\alpha$ equivalent width ($\text{EW}_{\text{Ly}\alpha}$). We focus on $\text{EW}_{\text{Ly}\alpha}$ rather than the total $\text{Ly}\alpha$ luminosity for multiple reasons. First, we find that the $\text{Ly}\alpha$ luminosity is primarily correlated with other descriptors of the total luminosity of a galaxy, so measuring the $\text{Ly}\alpha$ luminosity per unit UV continuum luminosity is a more interesting descriptor of the (in)ability of $\text{Ly}\alpha$ photons to escape relative to non- $\text{Ly}\alpha$ photons at similar wavelengths. Second, characterization of $\text{EW}_{\text{Ly}\alpha}$ is possible in slit spectra even without precise flux calibration, so our predictions for $\text{EW}_{\text{Ly}\alpha}$ can be more easily

applied to actual observations of galaxies with uncertain slit losses.¹²

The value of $\text{EW}_{\text{Ly}\alpha}$ is measured for each object in our sample directly from the 1D object spectrum in a manner similar to that described in Trainor et al. (2015, 2016). Each rest-UV spectrum is first shifted to the rest frame based on its nebular redshift z_{neb} . In order to account for the wide variety of $\text{Ly}\alpha$ line profiles in a systematic manner, the $\text{Ly}\alpha$ line flux, $F_{\text{Ly}\alpha}$, was directly integrated by summing the continuum-subtracted line flux over the range $1208 \text{ \AA} < \lambda_{\text{rest}} < 1227 \text{ \AA}$, roughly the maximum range of wavelengths found to encompass the $\text{Ly}\alpha$ line in our spectra.¹³ In velocity space, this corresponds to a range $-1900 \text{ km s}^{-1} < v < 2800 \text{ km s}^{-1}$ with respect to $\text{Ly}\alpha$. The local continuum flux $f_{\lambda, \text{cont}}$ used for subtraction is estimated as the median flux in the range $1225 \text{ \AA} < \lambda_{\text{rest}} < 1250 \text{ \AA}$. The $\text{Ly}\alpha$ line flux $f_{\text{Ly}\alpha}$ is therefore positive for net $\text{Ly}\alpha$ emission lines and negative for net $\text{Ly}\alpha$ absorption. The line flux and equivalent width are thus defined by the following expressions:

$$F_{\text{Ly}\alpha} = \int_{1208 \text{ \AA}}^{1227 \text{ \AA}} (f_{\lambda} - f_{\lambda, \text{cont}}) d\lambda \quad (1)$$

$$\text{EW}_{\text{Ly}\alpha} = \frac{F_{\text{Ly}\alpha}}{f_{\lambda, \text{cont}}} \frac{1}{1 + z_{\text{neb}}}. \quad (2)$$

Again, this procedure causes galaxies with net $\text{Ly}\alpha$ emission in their 1D slit spectra to have $\text{EW}_{\text{Ly}\alpha} > 0$ and galaxies exhibiting net $\text{Ly}\alpha$ absorption to be assigned $\text{EW}_{\text{Ly}\alpha} < 0$. Note that, in each case, the assigned value of $\text{EW}_{\text{Ly}\alpha}$ is likely to be an underestimate of the intrinsic value owing to the spatial scattering of $\text{Ly}\alpha$ photons, which preferentially lowers the observed ratio of $\text{Ly}\alpha$ photons to continuum photons in a centralized aperture. Typical relative slit losses of $\text{Ly}\alpha$ with respect to the nearby continuum in similar samples are found to be $2-3\times$ (see, e.g., Steidel et al. 2011; Trainor et al. 2015). However, we have no way to determine the relative $\text{Ly}\alpha$ slit loss for the majority of individual objects in our sample, so we do not attempt to do so here. Furthermore, in this work we are primarily concerned with the factors that determine the net $\text{Ly}\alpha$ emission on galaxy scales; variation in the total $\text{Ly}\alpha$ emission of the galaxy-plus-halo system with physical and environmental properties of the galaxies will be discussed in future work.

Uncertainty in $\text{EW}_{\text{Ly}\alpha}$ is determined based on the uncertainty in the $\text{Ly}\alpha$ flux (usually a small factor), as well as the uncertainty in the local continuum (usually the dominant factor, particularly for high- $\text{EW}_{\text{Ly}\alpha}$ sources). In some cases, correlated noise in the continuum spectrum causes the formal uncertainty in the local continuum to be unrealistically low based on a visual inspection of the spectrum. To account for this fact, a separate estimate of the continuum flux $f_{\lambda, \text{cont}}$ is measured for each galaxy over the range $1220 \text{ \AA} < \lambda_{\text{rest}} < 1300 \text{ \AA}$ (i.e., over a $\sim 3\times$ larger range of wavelengths than the first estimate), and this second continuum estimate is used to recalculate $F_{\text{Ly}\alpha}$ and $\text{EW}_{\text{Ly}\alpha}$. If the new $\text{EW}_{\text{Ly}\alpha}$ value differs from the original value by more than the uncertainty on $\text{EW}_{\text{Ly}\alpha}$ calculated originally, then the uncertainty on $\text{EW}_{\text{Ly}\alpha}$ is replaced by the absolute

¹² However, note that the spatial scattering of $\text{Ly}\alpha$ photons (as described in Section 1) causes $\text{EW}_{\text{Ly}\alpha}$ to be sensitive to the differential slit losses in $\text{Ly}\alpha$ vs. the continuum (although $\text{EW}_{\text{Ly}\alpha}$ is still less sensitive to slit losses than the total $\text{Ly}\alpha$ luminosity).

¹³ This velocity range is also chosen to minimize contaminating absorption due to the Si III $\lambda 1206$ transition.

value of the difference between the two estimates of $EW_{Ly\alpha}$. This procedure increases the uncertainty on $EW_{Ly\alpha}$ for 59 objects (8% of the total sample), and the total variance in $EW_{Ly\alpha}$ associated with our estimated measurement error among our entire sample increases by 5%.

Furthermore, if the measured continuum value is smaller than $2\times$ the estimated uncertainty on the continuum, then $EW_{Ly\alpha}$ is defined to be a lower limit:

$$EW_{Ly\alpha} > \frac{F_{Ly\alpha}}{2\sigma_{\lambda,cont}} \frac{1}{1 + z_{neb}}, \quad (3)$$

where $\sigma_{\lambda,cont}$ is the formal uncertainty in the local $Ly\alpha$ continuum. This correction applies to 13 objects (2% of our total sample), and $EW_{Ly\alpha}$ for these objects is assumed to take the value of their 2σ error in the analysis that follows. In this manner, a value and uncertainty for $EW_{Ly\alpha}$ are estimated for each of the 703 galaxies in our sample.

Spearman rank correlation statistics between $EW_{Ly\alpha}$ and a series of other empirical quantities measured among the galaxies in our sample are given in Table 1 below. Definitions and measurement methodologies for each of these quantities are given in the sections that follow.

3.2. Proxies for $Ly\alpha$ Escape

3.2.1. $v_{Ly\alpha}$

As discussed in Section 1 above, increasing shifts of the $Ly\alpha$ emission line with respect to the systemic redshift are associated with decreasing $EW_{Ly\alpha}$, and this trend likely corresponds to the fact that $Ly\alpha$ photons must scatter significantly in redshift and/or physical space to escape regions of high optical depth. For this reason, the difference of the $Ly\alpha$ redshift ($z_{Ly\alpha}$) from systemic (z_{neb}) may be regarded as a proxy for the optical depth experienced by $Ly\alpha$ photons transiting the galaxy ISM and CGM and is thus related to the probability of $Ly\alpha$ photon escape.

We define the $Ly\alpha$ redshift and offset velocity based on the centroid of the $Ly\alpha$ line flux:

$$\lambda_{Ly\alpha}^{obs} = \frac{\int \lambda f_{\lambda} d\lambda}{\int f_{\lambda} d\lambda} \quad (4)$$

$$z_{Ly\alpha} = \frac{\lambda_{Ly\alpha}^{obs} - \lambda_{Ly\alpha}^{rest}}{\lambda_{Ly\alpha}^{rest}} \quad (5)$$

$$v_{Ly\alpha} = \left(\frac{z_{Ly\alpha} - z_{neb}}{1 + z_{neb}} \right) c, \quad (6)$$

where c is the speed of light and the integrals in Equation (4) are evaluated over the range $1208 \text{ \AA} < \lambda_{rest} < 1227 \text{ \AA}$, as is done to estimate the total $Ly\alpha$ flux (Equation (1)). Note, however, that in this case we integrate the raw flux over the $Ly\alpha$ region rather than integrating the continuum-subtracted flux. This choice does not appreciably change the assigned value of $v_{Ly\alpha}$ when $EW_{Ly\alpha}$ is large, but it significantly reduces the noise on $v_{Ly\alpha}$ when $EW_{Ly\alpha} \rightarrow 0$, in which case the denominator would approach zero for an analogous equation to Equation (4) weighted by continuum-subtracted flux.

Note also that the flux in the $Ly\alpha$ transition need not exceed $f_{\lambda,cont}$ in order to measure an $Ly\alpha$ velocity; if a galaxy exhibits $Ly\alpha$ absorption that is preferentially blueshifted, the assigned $Ly\alpha$ velocity will be positive and thus similar to a galaxy with redshifted $Ly\alpha$ emission.

Velocity uncertainties are determined by a Monte Carlo analysis in which a randomly generated error array consistent with the per-pixel uncertainty is added to the $Ly\alpha$ region of the spectrum and the velocity is measured as above. This process is repeated 1000 times per spectrum, and the estimated velocity uncertainty is $1.5\times$ the median absolute deviation¹⁴ of the Monte Carlo velocity values. While velocity measurements can be made in this way for every spectrum in our sample, we include in our analysis of $v_{Ly\alpha}$ only those spectra with an $Ly\alpha$ velocity uncertainty $\sigma_{v_{Ly\alpha}} < 750 \text{ km s}^{-1}$.¹⁵ Because the $Ly\alpha$ velocity also depends on the accuracy of z_{neb} , we also require that $H\alpha$ and/or $[O III] \lambda 5008$ are detected with at least 5σ significance. When these cuts are made, 496 galaxies in our sample have reliably measured values of $v_{Ly\alpha}$. The Spearman rank correlation statistic between $v_{Ly\alpha}$ and $EW_{Ly\alpha}$ is $r_{Sp} = -0.56$ ($p = 1.6 \times 10^{-42}$), indicating a strong, highly significant correlation. This relationship is consistent with previous work (e.g., Erb et al. 2014), as well as our results from stacked spectra shown in Figure 2 (left panel).

As described in Section 1, our eventual goal is to predict the value of $EW_{Ly\alpha}$ for a galaxy in the absence of its direct measurement (since the $Ly\alpha$ flux is not always directly observable). Unfortunately, $v_{Ly\alpha}$ may be ineffective as such a predictor for two reasons. First, $v_{Ly\alpha}$ cannot be measured in cases where $Ly\alpha$ is not directly measurable (e.g., when the transition is censored by the IGM or contaminated by other emission). Second, even in cases where $v_{Ly\alpha}$ is measurable (e.g., among some galaxies at high redshift), any intergalactic absorption that suppresses $EW_{Ly\alpha}$ is also likely to change the observed value of $v_{Ly\alpha}$. Because scattering of $Ly\alpha$ by both the ISM and the surrounding IGM will produce degenerate shifts on $v_{Ly\alpha}$, $v_{Ly\alpha}$ itself cannot be expected to separate between these two effects. With this in mind, we caution against the use of $v_{Ly\alpha}$ to predict the intrinsic value of $EW_{Ly\alpha}$.

3.2.2. EW_{LIS}

The second proxy we use for the ease of $Ly\alpha$ photon escape is the strength of absorption lines corresponding to low-ionization interstellar gas. Unfortunately, even the strongest interstellar absorption features are difficult to measure reliably in individual spectra. In order to increase the significance of these detections, we construct a “mean” LIS absorption profile for each galaxy spectrum as follows.

Seven LIS transitions covered by the majority of our rest-UV spectra are identified in Table 2. O I $\lambda 1302$ and Si II $\lambda 1304$ are blended at the typical spectral resolution of our observations, so six distinct absorption features can be individually measured. For each transition, the spectral region within $\pm 5000 \text{ km s}^{-1}$ of the rest-frame wavelength is interpolated onto a grid in velocity space and normalized to its local continuum (defined as the median flux in the region $> 1000 \text{ km s}^{-1}$ and $< 5000 \text{ km s}^{-1}$

¹⁴ Note that $\sigma \approx 1.5 \times \text{MAD}$ is a simple estimator of scale that is insensitive to outliers and recovers the usual standard deviation when applied to a Gaussian distribution (see, e.g., Rousseeuw & Croux 1993).

¹⁵ Objects with larger velocity uncertainties typically have relatively low signal-to-noise ratios in the UV continua, as well as minimal $Ly\alpha$ absorption and emission, making the “velocity” of the $Ly\alpha$ line an ill-defined quantity.

Table 2
LIS Transitions

Ion	λ_{vac}^a (Å)	f_{osc}^b	EW_{ion}^c (Å)
Si II	1260.418	1.22	1.74 ± 0.06
O I	1302.169	0.0520	2.37 ± 0.08^d
Si II	1304.370	0.0928	2.37 ± 0.08^d
C II	1334.532	0.129	1.54 ± 0.08
Si II	1526.707	0.133	1.40 ± 0.10
Fe II	1608.451	0.0591	1.11 ± 0.14
Al II	1670.787	1.77	1.13 ± 0.21

Notes.^a Vacuum wavelength of transition.^b Oscillator strength from the NIST Atomic Spectra Database (www.nist.gov/pml/data/asd.cfm).^c Equivalent width of absorption in a stacked spectrum of all 703 galaxies in the sample (Figure 1).^d The O I $\lambda 1302$ and Si II $\lambda 1304$ absorption lines are blended, so they are measured as a single absorption feature with the given (combined) EW.

from the transition wavelength in either direction). The six LIS absorption profiles are then averaged with equal weighting, and an effective rest-frame equivalent width in absorption is measured for the stacked profile via direct integration according to the following expression, which we define as EW_{LIS} :

$$\text{EW}_{\text{LIS}} = \int_{\lambda_1}^{\lambda_2} \left(1 - \frac{f_{\lambda}}{f_{\text{cont}}} \right) d\lambda, \quad (7)$$

where λ_1 and λ_2 correspond to $\pm 1000 \text{ km s}^{-1}$ from the rest-frame line center. The uncertainty on EW_{LIS} is defined to be the standard deviation of absorption equivalent widths calculated as above for random 2000 km s^{-1} intervals in nearby regions of the rest-UV spectrum. Because the reliability of our EW_{LIS} measurement is extremely sensitive to the strength of the far-UV continuum, we only consider measurements of EW_{LIS} for which the local continuum is detected with signal-to-noise ratio $(S/N) > 20$ in the stacked profile; this sample includes 625 objects.

For 162 objects in this sample, one or more LIS transitions fall above the red edge of the LRIS-B spectrum. For 19 objects, one or more LIS transitions are flagged as discrepant: they either correspond to an equivalent width more than 15σ away from the median equivalent width of the other transitions, or they otherwise lie in a region of the spectrum that appears significantly noisier than average based on a visual inspection. In any of these cases, the missing or flagged transitions are omitted, and the mean EW_{LIS} value and uncertainty are calculated from the remaining transitions. In total, the number of objects for which 6 (5, 4, 3, 2, 1) transitions contribute to EW_{LIS} is 456 (75, 56, 42, 8, 0).

As described above, the vast majority of cases where one or more transitions are omitted occur because of a lack of red spectral coverage, such that the redder transitions in Table 2 are preferentially omitted. Given that the two strongest transitions are also the two bluest (and thus least likely to be omitted), there is potential for our spectral coverage to introduce a trend between EW_{LIS} and the number of included transitions. Separating galaxies by the number of included LIS transitions (N_{LIS}), the median value of EW_{LIS} for each subset is (1.43, 1.49, 1.27, 1.40, 1.55 Å) for $N_{\text{LIS}} = (6, 5, 4, 3, 2)$. The lack of

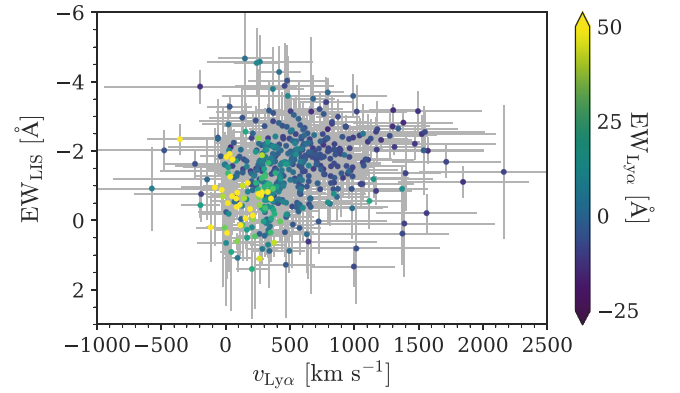


Figure 3. Centroid $\text{Ly}\alpha$ velocity vs. rest-frame equivalent width of LIS absorption for 452 galaxies, with colors denoting $\text{Ly}\alpha$ equivalent width. Two clear trends are visible: (1) stronger LIS absorption ($\text{EW}_{\text{LIS}} < 0$) is associated with increasing $\text{Ly}\alpha$ redshift ($v_{\text{Ly}\alpha} > 0$), and (2) strong $\text{Ly}\alpha$ emission ($\text{EW}_{\text{Ly}\alpha} > 0$, yellow points) is associated with both weak LIS absorption and small $\text{Ly}\alpha$ redshifts.

a systematic trend between EW_{LIS} and N_{LIS} suggests that our EW_{LIS} values are not particularly sensitive to the precise subset of LIS transitions included.

The distribution of EW_{LIS} versus $v_{\text{Ly}\alpha}$ is shown in Figure 3 for the 479 objects for which both quantities are measured robustly according to the criteria described above. The Spearman correlation coefficient for these two parameters is $r_{\text{Sp}} = 0.28$ ($p = 3.2 \times 10^{-10}$; see Table 1), indicating a moderate (although highly statistically significant) correlation between these two proxies for ISM optical depth (or porosity) and the likely ease of $\text{Ly}\alpha$ photon escape. The color-coding by $\text{EW}_{\text{Ly}\alpha}$ in Figure 3 demonstrates that strong $\text{Ly}\alpha$ emission is associated with weak LIS absorption and a small shift of the $\text{Ly}\alpha$ line with respect to systemic, in agreement with the expectations outlined above. The correlation between EW_{LIS} and $\text{EW}_{\text{Ly}\alpha}$ is moderately strong and highly significant ($r_{\text{Sp}} = -0.35$, $p = 3 \times 10^{-21}$).

We note that EW_{LIS} has multiple practical advantages over $v_{\text{Ly}\alpha}$ as a predictor of $\text{EW}_{\text{Ly}\alpha}$. Unlike $v_{\text{Ly}\alpha}$, EW_{LIS} is likely to be unaffected by IGM absorption, since the metallicity of intergalactic gas will be negligible compared to that of the enriched galactic outflows traced by metal-line absorption. In addition, in the event that the $\text{Ly}\alpha$ transition is censored by the IGM, local H I, or contaminating emission, the longer-wavelength LIS transitions may still be measurable in many realistic cases at both low and high redshifts. For these reasons, our analysis that follows utilizes EW_{LIS} as our primary proxy for $\text{Ly}\alpha$ escape.

3.2.3. $E(B - V)$

Because the escape of $\text{Ly}\alpha$ photons depends on the distribution of dust in galaxies as well as H I, we also consider the relationship between $\text{EW}_{\text{Ly}\alpha}$ and $E(B - V)$ (see also discussion in Trainor et al. 2016 and Theios et al. 2019).

$E(B - V)_{\text{SED}}$ is measured via SED fitting as described in Section 2.4 for 637 galaxies. Comparing these values to $\text{EW}_{\text{Ly}\alpha}$ yields a moderate, highly significant correlation ($r_{\text{Sp}} = -0.23$, $p = 3 \times 10^{-8}$).

We also measure $E(B - V)_{\text{neb}}$ based on the Balmer decrement ($\text{H}\alpha/\text{H}\beta$) as described by Strom et al. (2017). Briefly, the slit-loss-corrected $\text{H}\alpha$ and $\text{H}\beta$ fluxes are compared to the canonical ratio $\text{H}\alpha/\text{H}\beta = 2.86$ for case B recombination

at $T = 10^4$ K (Osterbrock 1989). Galaxies with $H\alpha/H\beta < 2.86$ are assigned $E(B - V)_{\text{neb}} = 0$, while galaxies with $H\alpha/H\beta > 2.86$ are assigned a value of $E(B - V)_{\text{neb}}$ based on a Cardelli et al. (1989) Galactic attenuation relation. The median value of $E(B - V)_{\text{neb}}$ for KBSS galaxies is 0.26, and the interquartile range is 0.06–0.47 (Strom et al. 2017).

As discussed by Trainor et al. (2016) and Strom et al. (2017), the $H\beta$ and $H\alpha$ emission lines are measured in separate exposures in KBSS-MOSFIRE observations; at typical redshifts $2.0 < z < 2.6$, the lines fall in the H and K near-IR atmospheric bands, respectively. For this reason, we present values of $E(B - V)_{\text{neb}}$ only for those galaxies with $>5\sigma$ measurements of $H\alpha/H\beta$, including the uncertainties in the individual line fluxes as well as the cross-band calibration. This cut limits our sample of secure $E(B - V)_{\text{neb}}$ measurements to 208 galaxies, which display a weak correlation with $\text{EW}_{\text{Ly}\alpha}$ ($r_{\text{Sp}} = -0.14$, $p = 0.05$).

3.2.4. f_{esc}

The most direct measure of the efficiency of $\text{Ly}\alpha$ photon escape is the actual escape fraction of $\text{Ly}\alpha$, hereafter f_{esc} .¹⁶ Any determination of this escape fraction relies on an estimation of the true number of $\text{Ly}\alpha$ photons produced in galaxies, which can then be compared to the observed $\text{Ly}\alpha$ flux. In practice, the observed $\text{Ly}\alpha$ flux is compared to the observed $H\alpha$ flux, with the latter value scaled by the expected intrinsic flux ratio ($F_{\text{Ly}\alpha}/F_{\text{H}\alpha})_{\text{int}} \approx 8.7$ for case B recombination.¹⁷

$H\alpha$ provides an effective proxy for the intrinsic $\text{Ly}\alpha$ luminosity because the former is not significantly scattered by H I ; however, it nonetheless suffers extinction by interstellar dust. The intrinsic $H\alpha$ flux (and thus the intrinsic $\text{Ly}\alpha$ flux) can therefore only be determined using the absolute attenuation $A_{\text{H}\alpha}$, which is typically estimated from the inferred nebular reddening (i.e., $E(B - V)$) as defined in Section 3.2.3 and the application of an attenuation relation that is appropriate to the galaxy at hand. As discussed by Theios et al. (2019), no single attenuation relation is able to self-consistently describe the host of photometric and spectroscopic properties inferred for KBSS galaxies. Given these ambiguities in the attenuation correction, we present both the relative (i.e., dust-uncorrected) and absolute (i.e., dust-corrected) $\text{Ly}\alpha$ escape fractions based on the following definitions:

$$f_{\text{esc,rel}} = \frac{F_{\text{Ly}\alpha}}{8.7 \times F_{\text{H}\alpha,\text{obs}}} \quad (8)$$

$$f_{\text{esc,abs}} = \frac{F_{\text{Ly}\alpha}}{8.7 \times F_{\text{H}\alpha,\text{corr}}}, \quad (9)$$

where $F_{\text{H}\alpha,\text{obs}}$ is the observed $H\alpha$ flux and $F_{\text{H}\alpha,\text{corr}}$ is that corrected based on the observed $E(B - V)_{\text{neb}}$ and the application of a Cardelli et al. (1989) attenuation relation. In both

¹⁶ f_{esc} here should not be confused with the escape fraction of Lyman continuum (i.e., ionizing) photons. The f_{esc} defined here is described elsewhere in the literature as $f_{\text{esc,Ly}\alpha}$, but we will omit the $\text{Ly}\alpha$ subscript for simplicity in this paper.

¹⁷ Note that various values are assumed for $(F_{\text{Ly}\alpha}/F_{\text{H}\alpha})_{\text{int}}$ in the literature, but the uncertainty on the aperture correction for $\text{Ly}\alpha$ in our data dwarfs the uncertainty on the intrinsic flux ratio, and our measured trends between f_{esc} and other parameters are insensitive to the chosen value regardless. The value 8.7 is motivated by the calculations of Dopita & Sutherland (2003) and is consistent with previous studies (Atek et al. 2009; Hayes et al. 2010; Henry et al. 2015; Trainor et al. 2015).

cases, $F_{\text{Ly}\alpha}$ is the *observed* $\text{Ly}\alpha$ flux as defined in Equation (1), which is not dust-corrected.

We calculate $f_{\text{esc,rel}}$ via Equation (8) for 368 galaxies for which we have a 5σ detection of $H\alpha$ (we take $f_{\text{esc,rel}} = 0$ where $F_{\text{Ly}\alpha} \leq 0$), as well as a measurement of $\text{EW}_{\text{Ly}\alpha}$. As described in Section 3.2.3, only 208 galaxies have a robust measurement of $E(B - V)_{\text{neb}}$; when combined with the S/N cuts on $H\alpha$ and $\text{EW}_{\text{Ly}\alpha}$, this leaves 188 galaxies for which $f_{\text{esc,abs}}$ may be calculated with confidence via Equation (9) (although subject to the remaining uncertainty in the attenuation law, as well as differential slit losses in $\text{Ly}\alpha$ versus $H\alpha$).

Figure 4 displays the relationship between $\text{EW}_{\text{Ly}\alpha}$ and f_{esc} in both its absolute and relative forms. In either case, the two quantities have a relatively strong correlation given the measurement uncertainties ($r_{\text{Sp}} = -0.50$, $p = 6 \times 10^{-13}$ for 188 galaxies for $f_{\text{esc,abs}}$; Table 1). Again, our analysis that follows is restricted to using $\text{EW}_{\text{Ly}\alpha}$ as a proxy for $\text{Ly}\alpha$ escape because f_{esc} , like $v_{\text{Ly}\alpha}$, is not typically measurable in cases where we would like to predict $\text{EW}_{\text{Ly}\alpha}$.

3.3. Proxies for $\text{Ly}\alpha$ Production

3.3.1. SFR, Mass, and Luminosity

We now consider parameters that may be associated with $\text{Ly}\alpha$ production. As noted above, $H\alpha$ luminosity should be a fairly direct proxy for the intrinsic luminosity of a galaxy in $\text{Ly}\alpha$. However, it is less related to the *efficiency* of $\text{Ly}\alpha$ production as described by $\text{EW}_{\text{Ly}\alpha}$: dust-corrected $L_{\text{H}\alpha}$ is uncorrelated with $\text{EW}_{\text{Ly}\alpha}$ in our sample ($r_{\text{Sp}} = -0.05$, $p = 0.4$) for the 208 galaxies with robust estimations of $E(B - V)$ and $A_{\text{H}\alpha}$ (see Section 3.2.3); the same correlation holds between $\text{EW}_{\text{Ly}\alpha}$ and $\text{SFR}_{\text{H}\alpha}$ since $\text{SFR}_{\text{H}\alpha}$ is linearly related to $L_{\text{H}\alpha}$.¹⁸

Our photometry-based estimates of SFR_{SED} display slightly stronger relationships with $\text{EW}_{\text{Ly}\alpha}$ ($r_{\text{Sp}} = -0.17$, $p = 10^{-5}$), and the correlation for stellar mass is very similar ($r_{\text{Sp}} = -0.15$, $p = 10^{-4}$) for the 637 galaxies with SED fits and $\text{EW}_{\text{Ly}\alpha}$ measurements. sSFR ($\text{SFR}_{\text{H}\alpha}/M_*$) displays a slightly stronger correlation with $\text{EW}_{\text{Ly}\alpha}$ ($r_{\text{Sp}} = 0.23$, $p = 10^{-3}$), with lower significance due to the smaller sample size of objects with the necessary measurements of both $\text{SFR}_{\text{H}\alpha}$ and M_* (199 galaxies).

Rest-UV absolute magnitudes M_{UV} are measured from the G - and \mathcal{R} -band magnitudes. The apparent magnitude corresponding to a rest-frame wavelength $\lambda_{\text{rest}} = 1450 \text{ \AA}$ is estimated by taking a weighted average of the G and \mathcal{R} based on the redshift of the galaxy.¹⁹ This apparent magnitude m_{UV} is then converted to the absolute magnitude M_{UV} based on the redshift of the source and the luminosity distance calculated assuming a ΛCDM cosmological model with $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_m = 0.3$, and $\Omega_\Lambda = 0.7$. In this manner, M_{UV} is measured for each of the 637 galaxies in our SED-fit sample. This parameter shows the weakest relationship with $\text{EW}_{\text{Ly}\alpha}$ of any quantity we measure, with $r_{\text{Sp}} = -1.5 \times 10^{-2}$ and $p = 0.71$. This lack of association between UV luminosity and $\text{EW}_{\text{Ly}\alpha}$ is

¹⁸ This calculation includes 208 galaxies with robust dust corrections as described in Section 3.2.3; the correlation with dust-uncorrected $L_{\text{H}\alpha}$ is similarly weak despite the much larger sample.

¹⁹ Note that the $\text{Ly}\alpha$ line falls within the G band for $z \gtrsim 2.45$, which includes roughly one-quarter of our galaxy sample. For the galaxies in this redshift interval, we correct the inferred value of M_{UV} based on the spectroscopic measurement of $\text{EW}_{\text{Ly}\alpha}$. This correction produces a median change $\Delta M_{\text{UV}} \approx 0.03$, although $\Delta M_{\text{UV}} \approx 0.5$ for the few most extreme $\text{Ly}\alpha$ emitters in our sample ($\text{EW}_{\text{Ly}\alpha} \gtrsim 100 \text{ \AA}$).

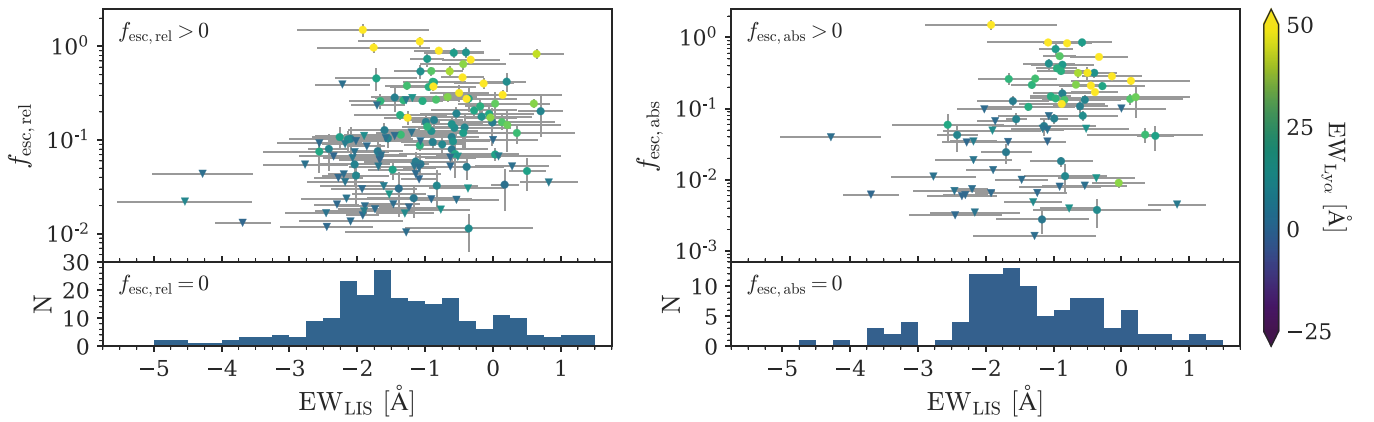


Figure 4. Ly α photon escape fraction vs. rest-frame equivalent width of LIS absorption, with color-coding by Ly α equivalent width. Circles indicate formal detections, while triangles indicate 2σ upper limits on f_{esc} . The left panel gives the escape fraction of Ly α photons relative to H α (Equation (8)) for 368 galaxies with detected Ly α and H α emission. The right panel gives the absolute escape fraction of Ly α photons (Equation (9)) for 188 galaxies that also have robust estimates of the Balmer decrement (used to dust-correct the H α flux). Bottom panels show the EW $_{\text{LIS}}$ distribution for galaxies with $F_{\text{Ly}\alpha} < 0$.

remarkable given the high EW $_{\text{Ly}\alpha}$ values associated with faint, high- z galaxies in other recent work; these trends are discussed further in Section 6.

3.3.2. O32

By definition, the intrinsic EW $_{\text{Ly}\alpha}$ of a galaxy is the ratio of Ly α photons to UV continuum photons, where the latter are generated directly by OB stars and the former are generated by the gas excited and ionized by these same stars. It is therefore sensible that EW $_{\text{Ly}\alpha}$ would be strongly associated with the excitation and ionization states of the gas in star-forming regions.

The O32 line ratio is one commonly used indicator of nebular ionization (Sanders et al. 2016; Steidel et al. 2016; Strom et al. 2017, 2018):

$$\text{O32} \equiv \log \left(\frac{[\text{O III}] \lambda\lambda 4960, 5008}{[\text{O II}] \lambda\lambda 3727, 3729} \right). \quad (10)$$

For the ionization-bounded H II regions typically assumed in photoionization models of star-forming galaxies, O32 is approximately proportional to $\log(U)$, where U denotes the “ionization parameter,” the local number of hydrogen-ionizing photons per hydrogen atom (see discussion by Steidel et al. 2016). Furthermore, O32 has previously been found to correlate strongly with Ly α emission (e.g., Nakajima et al. 2016; Trainor et al. 2016).

Notably, however, recent work has suggested that elevated O32 values may correspond in some cases to density-bounded H II regions, in which the ionized region is not entirely surrounded by neutral gas²⁰ (Nakajima et al. 2013; Izotov et al. 2016; Trainor et al. 2016). In particular, several recent detections of escaping Ly-continuum (rest-frame H ionizing) photons from galaxies at low and high redshift have been accompanied by elevated O32 ratios (e.g., de Barros et al. 2016; Izotov et al. 2016, 2018; Fletcher et al. 2019; but see Borthakur et al. 2014; Shapley et al. 2016, who find Ly-continuum escape in the absence of extreme O32). In these scenarios, an association of large O32 with high EW $_{\text{Ly}\alpha}$ may

reflect a combination of both increased ionizing photon production *and* increased probability of photon escape due to the lack of surrounding neutral gas.

O32 measurements for the KBSS sample are calculated as described in Strom et al. (2017). Briefly, the [O III] $\lambda\lambda 4960, 5008$ and [O II] $\lambda\lambda 3727, 3729$ line fluxes are measured as described in Section 2.3. We require that both [O III] and [O II] be detected with $S/N > 3$, resulting in a sample of 316 galaxies with a measured raw O32 value. Dust-corrected O32 values are measured for a smaller sample of 174 objects that meet both the requirements described above and the cut on the S/N of the Balmer decrement described in Section 3.2.4. For these measurements, each of the [O III] and [O II] emission lines is corrected for extinction using the measured Balmer decrement and a Cardelli et al. (1989) attenuation curve before calculating the line ratio. These two O32 estimators have the highest individual correlations with EW $_{\text{Ly}\alpha}$ of any “production”-related parameter: $r_{\text{Sp}} = 0.43$ ($p = 10^{-15}$) for the raw O32 measurements with a slightly higher correlation strength and lower significance for the smaller sample of dust-corrected O32 values (Table 1). However, O32 also displays a strong correlation with EW $_{\text{LIS}}$ ($r_{\text{Sp}} = 0.47$), perhaps reinforcing the idea that O32 is not wholly a measure of Ly α production.

3.3.3. O3

The O3 ratio is another indicator of nebular ionization and excitation:

$$\text{O3} \equiv \log \left(\frac{[\text{O III}] \lambda 5008}{\text{H}\beta} \right). \quad (11)$$

As discussed by Trainor et al. (2016), the O3 ratio is strongly associated with O32 for the high-excitation galaxies typical at $z \gtrsim 2$: the two quantities are correlated with $r_{\text{Sp}} = 0.74$ in the KBSS sample. Likewise, Strom et al. (2018) demonstrate that O3 is an effective indicator of $\log(U)$ through extensive photoionization modeling of the KBSS galaxy sample. O3 therefore has many of the same advantages as O32 for indicating Ly α production.

However, O3 has two significant advantages over O32. First, O3 relies on two emission lines at similar wavelengths, which makes the ratio insensitive to both dust extinction and cross-band calibration. Second, O3 is insensitive to the differences between density-bounded and ionization-bounded H II regions,

²⁰ Essentially, the local ratio of O II to O III increases toward the edge of the Strömgen sphere for an ionization-bounded nebula. For a nebula that is optically thin to ionizing photons, this ionization front (and its associated region of stronger O II emission) is not present. See, e.g., Pellegrini et al. (2012).

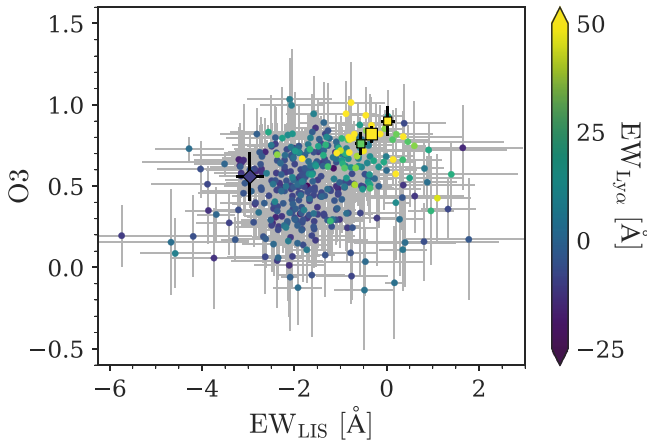


Figure 5. O3 ratio ($\equiv \log([\text{O III}]/\text{H}\beta)$) vs. rest-frame equivalent width of LIS absorption for 377 galaxies, with colors denoting $\text{Ly}\alpha$ equivalent width. Note that O3 and EW_{LIS} are not strongly correlated with each other, but $\text{EW}_{\text{Ly}\alpha} > 0$ is strongly associated with both weak LIS absorption ($\text{EW}_{\text{LIS}} \approx 0$) and strong [O III] emission ($\text{O3} \gtrsim 0.5$). Square boxes with black borders correspond to stacked measurements of faint $L \sim 0.1L_*$ $\text{Ly}\alpha$ -selected galaxies from Trainor et al. (2015, 2016). The large square denotes the full sample, and the smaller squares denote measurements based on splitting about the median value of $\text{EW}_{\text{Ly}\alpha}$. The diamond with black border represents MS 1512-cB58 based on measurements from Pettini et al. (2002) and Teplitz et al. (2000).

so it may indicate nebular excitation (and $\text{Ly}\alpha$ production) in a manner more decoupled from the physics of $\text{Ly}\alpha$ escape. Based on these advantages, we rely on O3 as our primary metric of $\text{Ly}\alpha$ production efficiency for the remainder of this work.

Using the same line-fitting process described above to estimate the line fluxes and uncertainties, we calculate the O3 ratio for every galaxy that has $\text{S/N} > 3$ for both [O III] and $\text{H}\beta$, a total of 395 objects. O3 has a correlation with $\text{EW}_{\text{Ly}\alpha}$ that is only marginally weaker than the corresponding correlation for O32 ($r_{\text{Sp}} = 0.40$, $p = 5 \times 10^{-16}$).

3.4. Summary of Correlations with $\text{Ly}\alpha$

Again, Spearman rank correlation statistics for $\text{EW}_{\text{Ly}\alpha}$ and EW_{LIS} with other measured quantities are presented in Table 1. Each quantity is calculated for a different number of objects according to the cuts described above, and the p values of every measured correlation (which depend on both the measured r and the number of objects in the sample) are highly significant ($p \ll 1$ in nearly all cases). However, there is a wide range of r values, indicating that certain parameters explain only a small fraction in the total variation in $\text{EW}_{\text{Ly}\alpha}$ despite the statistical significance of their correlation.

Note that EW_{LIS} is strongly correlated with the escape fraction of $\text{Ly}\alpha$ as expected based on the arguments that both of these quantities are related to the ability of $\text{Ly}\alpha$ photons to escape galaxies (see Section 3.2.2 and Figure 4). Conversely, EW_{LIS} has a much weaker²¹ correlation with O3 despite the fact that both quantities show relatively strong correlations with $\text{EW}_{\text{Ly}\alpha}$. We interpret this relationship in the sections that follow, but it is suggestive of the fact that these two quantities capture different processes (escape and production) related to the observed $\text{EW}_{\text{Ly}\alpha}$.

Figure 5 visually demonstrates this same result. While O3 and EW_{LIS} are themselves not closely correlated, there is a

²¹ While the correlation is highly significant at $p < 10^{-6}$, the low rank correlation coefficient $r_{\text{Sp}} = 0.21$ indicates that most of the variation in EW_{LIS} is not associated with variation in O3.

clear trend toward high $\text{EW}_{\text{Ly}\alpha}$ in the upper right corner of Figure 5 (i.e., the region of high O3 and/or $\text{EW}_{\text{LIS}} \sim 0$) and low $\text{EW}_{\text{Ly}\alpha}$ in the lower left corner (i.e., the region of low O3 and/or strongly negative EW_{LIS}).

Figure 5 also includes measurements from stacked spectra of faint $L \sim 0.1L_*$ $\text{Ly}\alpha$ -selected galaxies from the KBSS $\text{Ly}\alpha$ survey; these measurements are shown as boxed points. The EW_{LIS} measurements are described by Trainor et al. (2015), while the O3 measurements are described by Trainor et al. (2016). The faint galaxy measurements follow the same trend as the individual measurements from the brighter KBSS galaxies, in that high $\text{EW}_{\text{Ly}\alpha}$ is associated with the upper right corner of the parameter space.

For comparison, we also include measurements of MS 1523-cB58, a gravitationally lensed galaxy with $M_* \approx 10^9 M_\odot$, $\text{SFR} \approx 50\text{--}100 M_\odot \text{ yr}^{-1}$, and a young age ~ 9 Myr (Siana et al. 2008). The reported O3 value is based on spectroscopic measurements reported by Teplitz et al. (2000), while the $\text{Ly}\alpha$ and LIS equivalent widths are new measurements from the Keck/ESI spectrum presented by Pettini et al. (2002). Despite its young age and large SFR—both of which would predict a high rate of $\text{Ly}\alpha$ production—the spectrum of MS 1512-cB58 (hereafter cB58) displays net $\text{Ly}\alpha$ absorption,²² consistent with its deep LIS absorption lines ($\text{EW}_{\text{LIS}} \approx -3 \text{ \AA}$). The galaxy cB58 thus obeys the same association as the KBSS galaxies between $\text{Ly}\alpha$ emission and position in the O3– EW_{LIS} parameter space.

The structure of Figure 5 therefore suggests that a linear combination of O3 and EW_{LIS} would better predict $\text{EW}_{\text{Ly}\alpha}$ than either quantity alone. That is, we could in principle define a single parameter that is maximized when both O3 and EW_{LIS} predict strong $\text{EW}_{\text{Ly}\alpha}$, is minimized when both O3 and EW_{LIS} predict weak $\text{EW}_{\text{Ly}\alpha}$, and takes intermediate values when O3 and $\text{EW}_{\text{Ly}\alpha}$ have contradictory implications for the value of $\text{EW}_{\text{Ly}\alpha}$. We develop such a model in Section 4 below.

4. Combined Model

Motivated by the arguments above, we construct a new parameter $X_{\text{LIS}}^{\text{O3}}$ with the following definition:

$$X_{\text{LIS}}^{\text{O3}} = \alpha(\text{EW}_{\text{LIS}}/\text{\AA}) + (1 - \alpha)\text{O3}. \quad (12)$$

This parameter has the behavior described at the end of Section 3.4: $X_{\text{LIS}}^{\text{O3}}$ is maximized when both O3 and EW_{LIS} are maximized (i.e., when our proxies for both $\text{Ly}\alpha$ production and $\text{Ly}\alpha$ escape suggest that $\text{EW}_{\text{Ly}\alpha}$ should be strong). Likewise, $X_{\text{LIS}}^{\text{O3}}$ will take smaller values when either or both of O3 and EW_{LIS} are small (i.e., when $\text{EW}_{\text{Ly}\alpha}$ is expected to be small according to Figure 5). We therefore may expect any equation with the form of Equation (12) to predict a strong, monotonically increasing relationship between $\text{EW}_{\text{Ly}\alpha}$ and $X_{\text{LIS}}^{\text{O3}}$.

We then tune α ²³ to maximize the predictive power of this relationship. Specifically, we choose the value of α that maximizes the rank correlation coefficient between $X_{\text{LIS}}^{\text{O3}}$ and $\text{EW}_{\text{Ly}\alpha}$, yielding

²² Note that the detailed cB58 $\text{Ly}\alpha$ profile displays weak $\text{Ly}\alpha$ emission superimposed on a much stronger damped $\text{Ly}\alpha$ absorption profile, as described by Pettini et al. (2000, 2002). Due to the lower S/N of our KBSS $\text{Ly}\alpha$ measurements, we simply describe each galaxy as a net absorber or emitter for the purposes of this paper.

²³ Note that α is a dimensionless number that sets the weighting of EW_{LIS} (measured in \AA) relative to O3 (measured in dex); this arbitrary parameterization was chosen so that typical values of $X_{\text{LIS}}^{\text{O3}}$ would be of order unity for the galaxies in our sample.

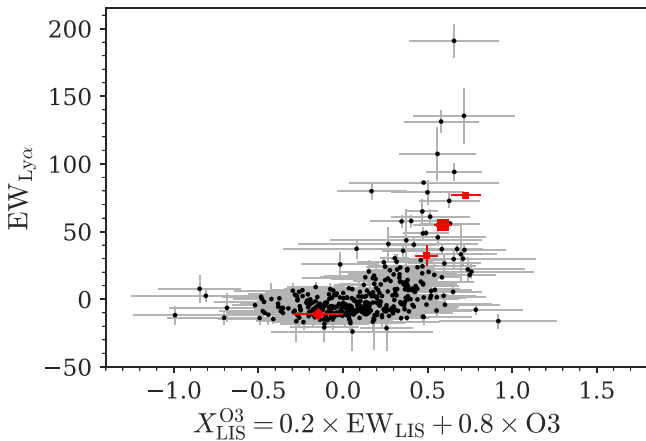


Figure 6. Ly α equivalent width vs. $X_{\text{LIS}}^{\text{O3}}$, a linear combination of O3 and EW_{LIS} that maximizes the Spearman rank correlation with $\text{EW}_{\text{Ly}\alpha}$: $r_{\text{Sp}} = 0.49$, $p = 1.5 \times 10^{-24}$. Approximately half of the ordering in observed $\text{EW}_{\text{Ly}\alpha}$ is explained by these two variables alone, and 90% of the variance in $\text{EW}_{\text{Ly}\alpha}$ is accounted for by the combination of an exponential model and the 2D measurement uncertainties (Section 4.1). As in Figure 5, red squares correspond to stacked measurements of faint $L \sim 0.1L_*$ Ly α -selected galaxies from Trainor et al. (2015, 2016), and the red diamond corresponds to measurements of MS 1512-cB58 from Pettini et al. (2002) and Teplitz et al. (2000).

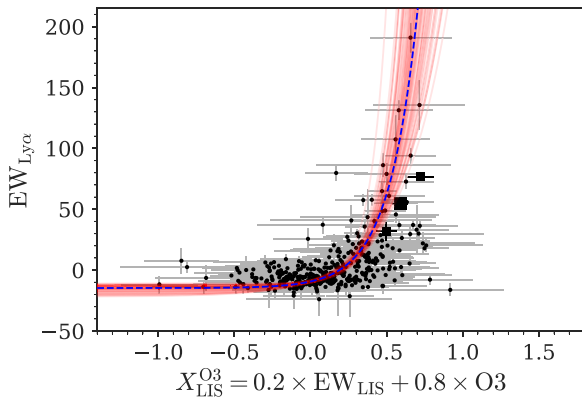


Figure 7. Data are the same as in Figure 6, with a best-fit exponential relationship (blue dashed curve) and 100 fits to bootstrap realizations of the data (faint red curves). The faint galaxy stacks and cB58 spectrum (square and diamond symbols) are not used in calculating the best-fit model parameters, but their positions are well described by our exponential model.

a maximum correlation of $r_{\text{Sp}} = 0.49$ for $\alpha \approx 0.2$.²⁴ Repeating this procedure on 1000 bootstrap samples, we find that the optimum value of α is constrained to 0.19 ± 0.06 , and the bootstrap samples are correlated with $\text{EW}_{\text{Ly}\alpha}$ with $r_{\text{Sp}} = 0.49 \pm 0.04$ for fixed $\alpha = 0.2$. The relationship between $\text{EW}_{\text{Ly}\alpha}$ and $X_{\text{LIS}}^{\text{O3}}$ is displayed in Figure 6. As expected from Figure 5, strong Ly α emission is closely associated with large $X_{\text{LIS}}^{\text{O3}}$.

4.1. Exponential Model and Variance

Motivated by the distribution of points in Figure 6, we fit an exponential model to the data. Because there are large uncertainties in both axes, we choose model parameters to minimize the 2D distance between the model and data, scaled

by the corresponding uncertainties in each dimension. In effect, we define a 2D analog of the traditional χ^2 parameter:

$$\chi_{2\text{D}}^2 = \sum_i \left(\left(\frac{x_i - x_{c,i}}{\sigma_{x,i}} \right)^2 + \left(\frac{y_i - y_{c,i}}{\sigma_{y,i}} \right)^2 \right), \quad (13)$$

which our model fitting seeks to minimize. In the above equation, x_i and y_i represent the values of $\text{EW}_{\text{Ly}\alpha}$ and $X_{\text{LIS}}^{\text{O3}}$ for the i th object in our sample, $\sigma_{x,i}$ and $\sigma_{y,i}$ are the associated uncertainties for that object, and $(x_{c,i}, y_{c,i})$ is the closest point on the model curve to the observed values x_i and y_i , scaled by their corresponding uncertainties. Our model takes the following form:

$$\text{EW}_{\text{Ly}\alpha} = \text{EW}_0 + A e^{X_{\text{LIS}}^{\text{O3}}/\beta}, \quad (14)$$

where the best-fit coefficients and 1D marginalized uncertainties are found to be

$$\text{EW}_0 = -15 \pm 2 \quad (15)$$

$$A = 5 \pm 2 \quad (16)$$

$$\beta = 0.19 \pm 0.04. \quad (17)$$

The uncertainties in the model parameters are calculated by repeating the fit on 500 bootstrap samples of the data, where each bootstrap sample contains 377 points and their corresponding uncertainties selected at random (with replacement) from the true set of 377 points in the sample. Figure 7 displays the best-fit curve along with 100 representative fits to the bootstrap samples to demonstrate the uncertainty in the fit model.

The best-fit curve corresponds to $\chi_{2\text{D}}^2 = 515$ for 377 observations, or $\chi_{2\text{D}}^2/N_{\text{dof}} = 1.36$. Being above unity, this value indicates that the typical observations differ from the best-fit model by more than their formal uncertainties, such that there is intrinsic scatter in the relationship between $\text{EW}_{\text{Ly}\alpha}$ and $X_{\text{LIS}}^{\text{O3}}$ that is described neither by our model nor by our estimated observational uncertainties.

In order to determine the fraction of the total variance in $\text{EW}_{\text{Ly}\alpha}$ captured by the combination of our model and our measurement uncertainties, we perform the following exercise. We begin by assigning each object a fiducial value of $\text{EW}_{\text{Ly}\alpha}$ and $X_{\text{LIS}}^{\text{O3}}$ according to the nearest point to the observed values on the best-fit curve (where proximity to the curve is calculated in 2D, weighted by the uncertainty in each dimension). Each point is then perturbed in both dimensions, with the perturbation drawn from a Gaussian distribution $\phi(\mu, \sigma)$, with $\mu = 0$ and σ equal to the estimated uncertainty in $\text{EW}_{\text{Ly}\alpha}$ or $X_{\text{LIS}}^{\text{O3}}$ for that object. The resulting simulated data thus represent a hypothetical sample consistent with the model, with scatter given only by the observational uncertainties.

A new fit to the resulting simulated data set is calculated (i.e., new coefficients are calculated for Equations (15)–(17)), and the following statistics are calculated to assess the variance in the simulated data: (1) the Spearman rank correlation coefficient r_s of the simulated data, and (2) the $\chi_{2\text{D}}^2$ coefficient assessing the goodness of fit of the simulated data to its own best-fit model. Repeating this process 100 times, we find that the resulting distribution of statistics have $\langle r_s \rangle = 0.59 \pm 0.02$ and $\langle \chi_{2\text{D}}^2/N_{\text{dof}} \rangle = 0.98 \pm 0.04$. Comparing these values to the same statistics for the real data ($r_s = 0.49$, $\chi_{2\text{D}}^2/N_{\text{dof}} = 1.36$), we see that the simulated data have a tighter correlation and

²⁴ This calculation is performed for the 377 galaxies with robust measurements of $\text{EW}_{\text{Ly}\alpha}$, EW_{LIS} , and O3; see Table 1.

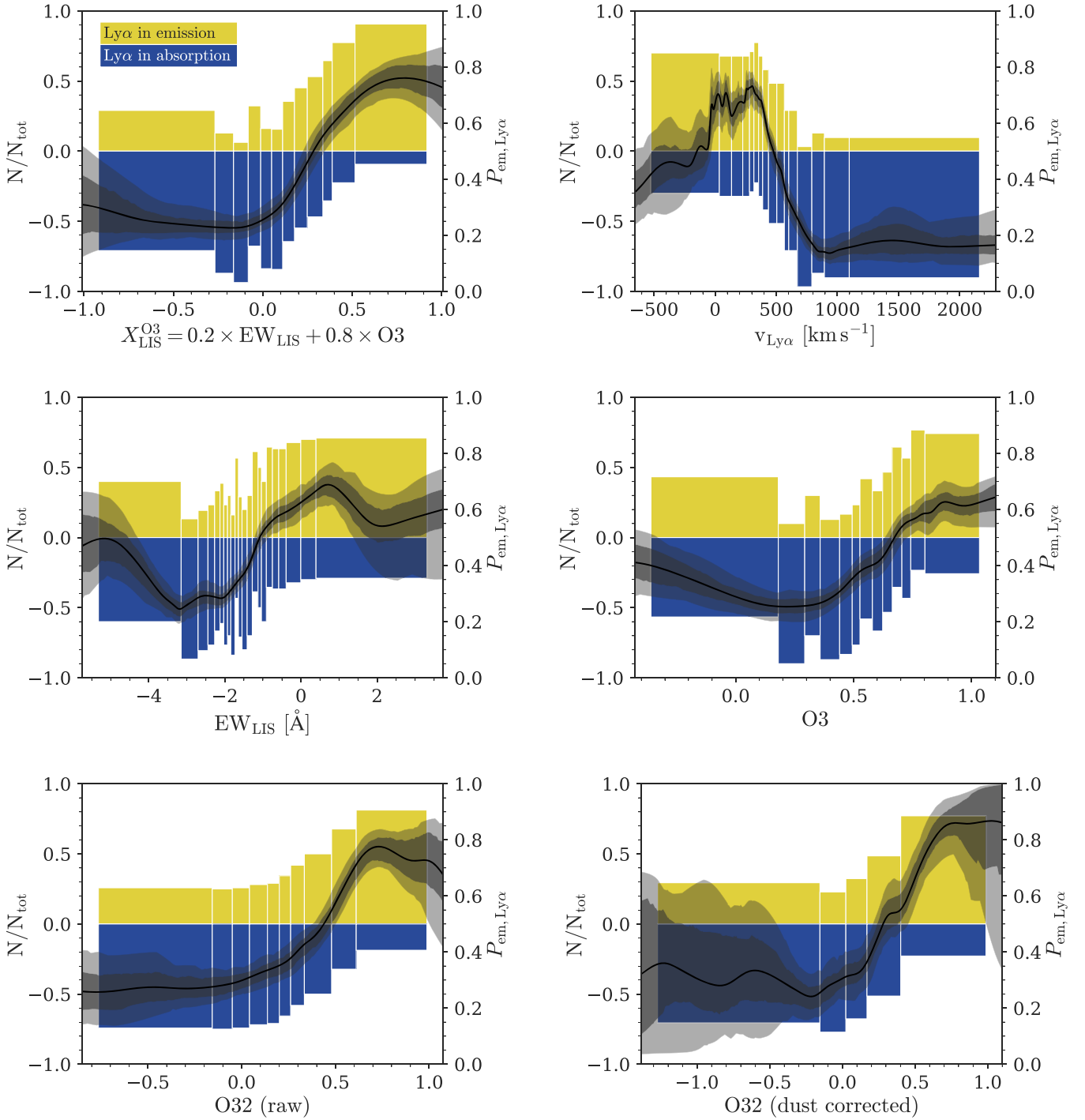


Figure 8. Conditional probability distributions for detecting Ly α in net emission (i.e., with $\text{EW}_{\text{Ly}\alpha} > 0$) as a function of other galaxy parameters. Yellow (blue) bars represent the fraction of Ly α emitters (absorbers) found within a given bin as shown by the left-hand vertical axis. Bin sizes and boundaries are determined in order to ensure that at least 30 objects are included per bin while allowing the number of bins to vary. Black lines indicate the conditional probability of Ly α detection (according to the right-hand vertical axis) at a given value of the parameter shown on the horizontal axis. Dark-gray (light-gray) shaded regions indicate the 68% (95%) confidence intervals on this conditional probability. Curves are determined based on the unbinned, nonparametric model described in Section 5.1, which depends on the measured parameter values and their uncertainties. $P_{\text{em}}^{\text{Ly}\alpha}$ represents $P(\text{EW}_{\text{Ly}\alpha} > 0 | X)$, where X is the parameter given on the horizontal axis. The parameters displayed here are all relatively effective predictors of Ly α emission, with $X_{\text{LIS}}^{\text{O3}}$ and O32 being particularly effective.

closer agreement with the fitted relation than do our real data; again, this indicates that the real data have additional sources of intrinsic scatter not described by our model or our estimated measurement errors.

We therefore model the intrinsic scatter in the relationship of $\text{EW}_{\text{Ly}\alpha}$ and $X_{\text{LIS}}^{\text{O3}}$ by assuming an underlying equation of the

following form:

$$\text{EW}_{\text{Ly}\alpha} = f(X_{\text{LIS}}^{\text{O3}}) + \phi(0, \sigma_{\text{int}}), \quad (18)$$

where $\phi(\mu, \sigma)$ is a number drawn from the Gaussian distribution with $\mu = 0$, and σ_{int} describes the intrinsic scatter

in $\text{EW}_{\text{Ly}\alpha}$ at fixed $X_{\text{LIS}}^{\text{O3}}$.²⁵ In this manner, we can simulate values of $\text{EW}_{\text{Ly}\alpha}$ and $X_{\text{LIS}}^{\text{O3}}$ drawn from the model distribution (including random perturbations corresponding to the estimated measurement uncertainties, as above), but with an additional term corresponding to the assumed intrinsic scatter that can be increased until the simulated data have similar total scatter (parameterized by r_s and $\chi_{2\text{D}}^2$) to the observed data.

In practice, we find that a value $\sigma_{\text{int}} = 7 \pm 1 \text{ \AA}$ produces the best match to the statistical properties of the observed data, with $r_s = 0.50 \pm 0.03$ and $\chi_{2\text{D}}^2/N_{\text{dof}} = 1.33 \pm 0.06$. Adopting this value for σ_{int} implies that the intrinsic variance in $\text{EW}_{\text{Ly}\alpha}$ not described by our model is $\sigma_{\text{int}}^2 = 50 \text{ \AA}^2$, while the total variance in $\text{EW}_{\text{Ly}\alpha}$ in our data set is $\sigma_{\text{obs}}^2 = 512 \text{ \AA}^2$. Assuming that we have $\sigma_{\text{obs}}^2 = \sigma_{\text{mod}}^2 + \sigma_{\text{int}}^2 + \sigma_{\text{meas}}^2$, we find that $\sim 90\%$ of the total variance in $\text{EW}_{\text{Ly}\alpha}$ is accounted for by our exponential model and the estimated measurement errors.

The apparent success of our two-parameter model for predicting $\text{EW}_{\text{Ly}\alpha}$ deserves some inspection, particularly in light of the well-known tendency (described in Section 1 and below) for Ly α emission to display substantial scatter with respect to galaxy properties. We address this topic in Section 6.

5. Conditional Probabilities for Ly α Detection

Despite the apparent success of the model above in self-consistently describing the behavior of $\text{EW}_{\text{Ly}\alpha}$, it has several shortcomings. Specifically, the model described above allows us to predict the net Ly α emission of a given galaxy based on measurements of EW_{LIS} and O3, but Figures 6–7 reveal substantial observational scatter in this relation that is not described by our exponential model. Furthermore, it is not obvious that an exponential model is physically meaningful for describing the dependence of Ly α emission on these properties.

An alternative method of describing the dependence of Ly α emission on galaxy properties would be to relinquish analytical functions for $\text{EW}_{\text{Ly}\alpha}$ in favor of a nonparametric model for the conditional probability of detecting Ly α , given a value for one or more other galaxy parameters. While this method does not provide an expected numerical value for $\text{EW}_{\text{Ly}\alpha}$, it allows us to explicitly describe how the detection fraction (as well as the stochasticity in observed Ly α emission) varies as a function of galaxy properties.

5.1. Methodology

For the majority²⁶ of the empirical parameters listed in Table 1, we construct conditional probability functions in two ways. First, we bin the full set of galaxies for which each indicator is measured into bins with widths that are allowed to vary in order to contain at least 30 galaxies per bin.²⁷ Within each bin, the fraction of galaxies with detected Ly α in net emission ($\text{EW}_{\text{Ly}\alpha} > 0$) is plotted as a yellow bar in the corresponding panel of Figures 8–9; the fraction of galaxies that are net Ly α

absorbers ($\text{EW}_{\text{Ly}\alpha} \leq 0$) is plotted as a blue bar in the negative direction. This empirical Ly α emitter fraction as a function of an observed parameter X can be interpreted as the conditional probability distribution $P(\text{EW}_{\text{Ly}\alpha} > 0 | X)$, hereafter $P_{\text{em}}^{\text{Ly}\alpha}(X)$.

Displaying the empirical Ly α emitter fraction in this way has the useful property that every galaxy contributes to the number of absorbers or emitters for a single bin, which means that each bin is independent. However, assigning each galaxy to a specific bin based on the observed value of a given Ly α -predicting parameter neglects the fact that the parameter values that define the horizontal axes of Figures 8–9 have their own observational uncertainties, which inhibits their assignment to a single specific bin. Likewise, the observational uncertainty on $\text{EW}_{\text{Ly}\alpha}$ prevents a clean separation between observed Ly α emitters and absorbers. For this reason, we construct a second, unbinned estimator of $P_{\text{em}}^{\text{Ly}\alpha}(X)$ that explicitly incorporates both of these uncertainties.

Our unbinned estimator of $P_{\text{em}}^{\text{Ly}\alpha}(X)$ represents each galaxy observation as a pair of 1D Gaussian probability distributions of the form $\phi(X | \mu = X_i, \sigma = \sigma_{X,i})$, where X_i and $\sigma_{X,i}$ are the observed value and observational uncertainty, respectively, on parameter X for galaxy i . Two distributions of this form are generated for each galaxy, with one distribution being normalized by the probability of the observed galaxy being an Ly α emitter and the second being normalized by the probability of being an absorber. Formally, we calculate the probability of an observed galaxy being an Ly α emitter from the measured value of $\text{EW}_{\text{Ly}\alpha}$ and its corresponding uncertainty $\sigma_{\text{EW,Ly}\alpha}$:

$$P_{\text{em},i} = \frac{1}{2} \operatorname{erf} \left(\frac{\text{EW}_{\text{Ly}\alpha,i}}{\sqrt{2} \sigma_{\text{EW,Ly}\alpha,i}} \right) + \frac{1}{2} \quad (19)$$

$$P_{\text{abs},i} = 1 - P_{\text{em},i}. \quad (20)$$

Thus, a galaxy with $\text{EW}_{\text{Ly}\alpha} > 0$ and $\text{EW}_{\text{Ly}\alpha} \gg \sigma_{\text{EW,Ly}\alpha}$ will have $P_{\text{em}} \approx 1$ and $P_{\text{abs}} \approx 0$; the “emitter” and “absorber” Gaussian probability distributions are then normalized such that their integrals equal P_{em} and P_{abs} , respectively. The inferred incidence η of Ly α emitters (absorbers) in our sample is then the sum of all the emitter (absorber) distributions:

$$\eta_{\text{em}}(X) = \sum_i P_{\text{em},i} \phi(X | X_i, \sigma_{X,i}) \quad (21)$$

$$\eta_{\text{abs}}(X) = \sum_i P_{\text{abs},i} \phi(X | X_i, \sigma_{X,i}). \quad (22)$$

A galaxy that is a clear Ly α emitter ($P_{\text{em}} \approx 1$) will therefore increase the integral of η_{em} by one over the distribution of X . This contribution to the incidence will be localized to a specific value²⁸ of X if $\sigma_{X,i}$ is small, whereas the contribution to the total incidence will be spread across a large fraction of the distribution if $\sigma_{X,i}$ is large. Finally, the inferred probability of Ly α emission given the observation of a galaxy with measured parameter value X is taken to be the inferred incidence of Ly α emitters divided by the total incidence of emitters:

$$P_{\text{em}}^{\text{Ly}\alpha}(X) = \frac{\eta_{\text{em}}(X)}{\eta_{\text{em}}(X) + \eta_{\text{abs}}(X)}. \quad (23)$$

²⁵ Note that σ_{int} is assumed not to vary with $X_{\text{LIS}}^{\text{O3}}$ for simplicity. While the observed distribution of $\text{EW}_{\text{Ly}\alpha}$ shows significantly more scatter at large $X_{\text{LIS}}^{\text{O3}}$ than at smaller values, we find that this effect is entirely consistent with the trend of increasing measurement uncertainties on both axes as $X_{\text{LIS}}^{\text{O3}}$ and $\text{EW}_{\text{Ly}\alpha}$ increase.

²⁶ We do not present conditional probability functions for $E(B - V)_{\text{neb}}$; the conditional probability density function is similar to that of $E(B - V)_{\text{SED}}$ but weaker.

²⁷ Note that the number of bins therefore depends on the total number of galaxies for which a given parameter is measured; see Table 1.

²⁸ To reduce the sample variance of this estimator, we replace $\sigma_{X,i}$ with the distance ΔX_i to its nearest neighbor in the observed distribution of X in cases where $\Delta X_i > \sigma_{X,i}$. This occurs for $\sim 2\%$ of objects, typically in the extrema of the distribution.

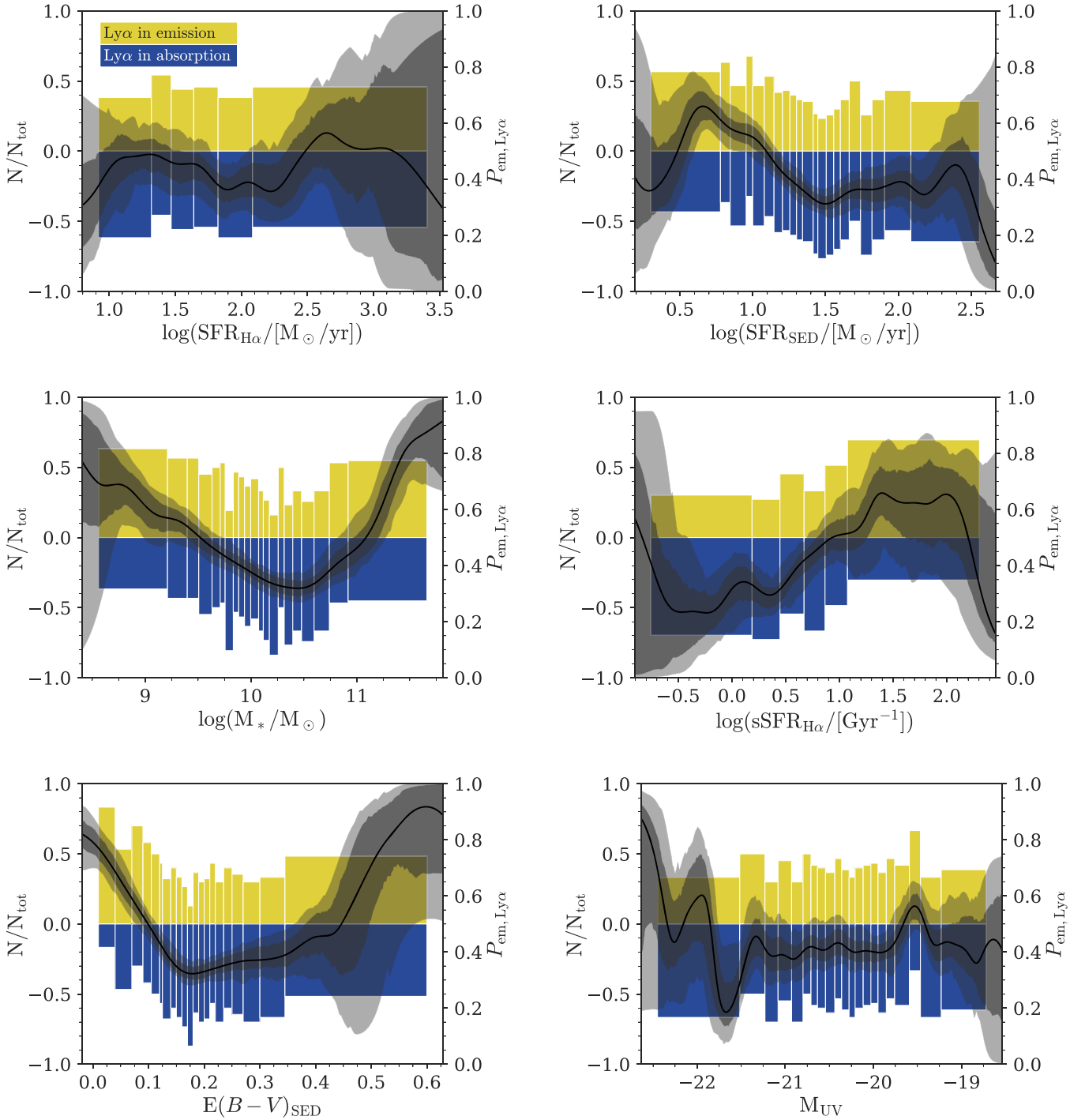


Figure 9. Continuation of Figure 8. Conditional probability distributions for the probability of Ly α detection as a function of SFR, SED-fit parameters, and other photometric galaxy parameters. The parameters in this figure are all relatively weak predictors of $EW_{Ly\alpha}$ compared to the parameters displayed in Figure 8; measuring the value of one of these parameters does not generally provide a strong prior on the probability of detecting Ly α in emission.

This conditional probability is displayed as a black curve in each panel of Figures 8–9. Confidence intervals for this curve are calculated by repeating the process above for 100 bootstrap realizations of the data and identifying the central 68% and 95% intervals; these are shown as gray shaded regions in Figures 8–9.

5.2. Analysis of Conditional Probability Distributions

The conditional probability distributions displayed in Figures 8–9 reinforce the relationships between $EW_{Ly\alpha}$ and

other galaxy properties shown in Table 1. Net Ly α emission is strongly associated with weak LIS absorption ($EW_{LIS} \gtrsim -1$) and high ionization/excitation ($O3 \gtrsim 0.7$; $O32 \gtrsim 0.5$). In particular, our X_{LIS}^{O3} metric is able to more clearly discriminate between Ly α emitters and absorbers than either EW_{LIS} or $O3$ alone; even the most extreme values of the latter metrics only predict $EW_{Ly\alpha} > 0$ with $\sim 60\%$ probability, whereas the highest values of X_{LIS}^{O3} predict Ly α in net emission in $\sim 80\%$ of cases. Likewise, $\lesssim 25\%$ of galaxies with $X_{LIS}^{O3} < 0$ display Ly α in net emission.

The distribution for $v_{\text{Ly}\alpha}$ is also shown in Figure 8, which demonstrates the strong dependence of $P_{\text{em}}^{\text{Ly}\alpha}$ on $v_{\text{Ly}\alpha}$ over the range $0 \lesssim v_{\text{Ly}\alpha} \lesssim 800 \text{ km s}^{-1}$ and very low probability of Ly α emission for galaxies with $v_{\text{Ly}\alpha} \gtrsim 800 \text{ km s}^{-1}$ ($P_{\text{em}}^{\text{Ly}\alpha} \approx 0.1\text{--}0.2$). $P_{\text{em}}^{\text{Ly}\alpha}$ peaks for $v_{\text{Ly}\alpha} \approx 0$, consistent with the model described in Sections 1 and 3.2.1.

Other than $v_{\text{Ly}\alpha}$, O32 is again the most direct predictor of net Ly α emission or absorption: even the dust-uncorrected (i.e., raw) values are approximately as effective at predicting $\text{EW}_{\text{Ly}\alpha} > 0$ as $X_{\text{LIS}}^{\text{O3}}$, and the dust-corrected measurements predict Ly α emission with $>80\%$ probability at the highest O32 values and $\lesssim 20\%$ at the lowest values. As discussed in Sections 3.3.2–3.3.3, this strong correlation may be due in part to the fact that dust-corrected O32 is a quite direct measure of the ionization parameter in ionization-bounded nebulae, but it may also be due to the fact that elevated O32 may indicate density-bounded nebulae that facilitate Ly α escape as well as production. Nonetheless, an observer who merely wishes to predict the net Ly α emission of a galaxy (remaining agnostic to the circumstances that facilitate this emission) will find O32 to be an effective indicator of this emission. However, the caveat to the effectiveness of this indicator remains the observational difficulty of obtaining high-S/N measurements of the [O III] and [O II] emission lines (the latter of which can be extremely faint in the high-ionization galaxies typical at high z), as well as the Balmer lines necessary to correct for differential attenuation by dust. This effect is seen in the small number of bins (which must each contain 30+ galaxies) in each of the O32 plots, as well as in the broad confidence intervals for the corresponding unbinned relations.

Note that in many of the panels of Figures 8–9, the trends apparently reverse toward the extrema of a given parameter value. Given that these regions of parameter space are sparsely populated (as can be seen based on the width of the blue and yellow bars) and the confidence intervals on $P_{\text{em}}^{\text{Ly}\alpha}$ diverge, we interpret the majority of this apparent aberrant behavior as a regression toward the overall average rate of Ly α emission ($P_{\text{em}}^{\text{Ly}\alpha}(X) \approx \langle P_{\text{em}}^{\text{Ly}\alpha} \rangle \approx 0.5$) when the parameter uncertainties are large. This effect is particularly noticeable in the panel for EW_{LIS} , where the highest and lowest bins appear particularly dominated by observational error—in this case, we expect that the trend between EW_{LIS} and $\text{EW}_{\text{Ly}\alpha}$ is monotonically positive over the range in which both parameters are well measured. Conversely, the relatively tight confidence intervals for $X_{\text{LIS}}^{\text{O3}}$ and O32 appear to indicate real flattening in their relationships with $P_{\text{em}}^{\text{Ly}\alpha}$; some non-negligible stochasticity in $\text{EW}_{\text{Ly}\alpha}$ appears to be present that is not accounted for by these factors even when they are well measured, as demonstrated by the $\sim 20\%$ of Ly α emitters (absorbers) that are present even at the lowest (highest) values of these parameters.

Figure 9 displays the conditional probability distributions for several parameters that are only weakly correlated with $\text{EW}_{\text{Ly}\alpha}$, again reinforcing the results shown in Table 1. In particular, $\text{SFR}_{\text{H}\alpha}$ and M_{UV} display negligible predictive power related to Ly α emission over the parameter space sampled by our galaxies. $E(B - V)_{\text{SED}}$ is an effective predictor of $P_{\text{em}}^{\text{Ly}\alpha}$ at the lowest reddenings ($E(B - V)_{\text{SED}} \lesssim 0.2$),²⁹ but Ly α emitters and absorbers appear almost equally common among galaxies with higher reddening values. Surprisingly, the incidence of Ly α emitters appears to grow slightly with increasing

$E(B - V)_{\text{SED}}$ for $E(B - V)_{\text{SED}} > 0.2$. This effect may be dominated by the relatively uncertain values of $E(B - V)_{\text{SED}}$, which is somewhat degenerate with stellar population age, causing a regression toward the population mean. $P_{\text{em}}^{\text{Ly}\alpha}$ shows a moderate increase as SFR_{SED} decreases or sSFR increases.

Stellar mass (M_*) displays perhaps the most interesting relationship with $\text{EW}_{\text{Ly}\alpha}$ that is not obvious from the Spearman correlation alone: Ly α emission probability increases significantly among both the lowest- and highest-mass galaxies. The low-mass relationship (along with the trends in SFR and sSFR) may reflect the tendency for low-mass, low-SFR galaxies to have relatively porous ISMs and high nebular excitation (see, e.g., Trainor et al. 2015, 2016). Although galaxies with clear signatures of AGN activity were removed from our sample, the excess of Ly α emitters at high galaxy masses may reflect residual AGNs in our sample. It is also possible that some galaxy masses in our sample are overestimated owing to strong line emission that contaminates the rest-frame optical photometry; SED-fit masses are especially sensitive to this effect at high redshift (see, e.g., Schenker et al. 2013). In this case, the increase in Ly α emission among galaxies with large inferred stellar masses could be caused by the underlying association between Ly α and optical emission line strength (i.e., nebular excitation). The Ly α -emitting behavior of high-mass galaxies will be investigated in future work.

In general, these conditional probability distributions may be used to inform analyses of the Ly α detection fraction of galaxies at the highest redshifts, where the opacity of the neutral IGM may suppress the observed Ly α emission from intrinsic Ly α emitters. By using other observed properties of galaxies as priors input to the distributions above, it will be possible to more accurately characterize the degree to which evolution in both IGM and galaxy properties shapes the distribution of observed Ly α emission.

6. Comparison to Recent Work

A few recent studies in the low-redshift universe have measured correlations between Ly α emission and other galaxy properties with the goal of predicting the Ly α emission. Hayes et al. (2014) present data from the Lyman-Alpha Reference Sample (LARS; Östlin et al. 2014), comparing Ly α emission with 12 different global galaxy properties derived from imaging and spectroscopy. They find significant correlations in which normalized³⁰ Ly α emission is highest among galaxies with low SFR, low dust content (inferred by nebular line ratios or the UV slope), low mass, and nebular properties indicative of high excitation and low metallicity. The Hayes et al. (2014) study differs from the work presented here in that their individual measurements have much higher S/Ns but are much fewer in number (12 galaxies in LARS vs. the 703 galaxies in this paper). Furthermore, the original LARS sample did not include rest-UV continuum spectroscopy covering the interstellar absorption lines; while these data were later collected via *HST*/COS spectroscopy and presented by Rivera-Thorsen et al. (2015), there is currently no simultaneous analysis of the predictive power of combined rest-UV and rest-optical spectroscopic diagnostics of Ly α emission in LARS.

²⁹ See Section 6 for a description of other recent studies of $E(B - V)$ versus $\text{EW}_{\text{Ly}\alpha}$.

³⁰ Hayes et al. (2014) consider several metrics for Ly α emission, including the total Ly α luminosity ($L_{\text{Ly}\alpha}$), $\text{EW}_{\text{Ly}\alpha}$, $L_{\text{Ly}\alpha}/L_{\text{H}\alpha}$, and $f_{\text{esc,abs}}$; only the latter three quantities show strong correlations with galaxy properties.

Recent results by Yang et al. (2017) are more directly comparable to those presented here: Yang et al. (2017) analyze *HST*/COS spectra of 43 “green pea” galaxies at $z \sim 0.1$ – 0.3 with Sloan Digital Sky Survey optical spectroscopy. The authors find that the escape fraction of Ly α is anticorrelated with the velocity width of the Ly α line profile, the nebular dust extinction, and the stellar mass, as well as positively correlated with the O32 ratio. Each of these relationships has Spearman rank correlation coefficients of $r \sim 0.5$ – 0.6 ; while the contributions of observational uncertainties to this scatter are not explicitly calculated, the quoted uncertainties suggest that these contributions are negligible. Furthermore, Yang et al. (2017) fit a linear combination of the nebular extinction $E(B - V)$ and the velocity offset of the red Ly α peak, finding that the resulting relation fits the observed Ly α escape fraction with a 1σ scatter of 0.3 dex. Notably, this multiparameter relationship is similar to our own work in Section 4, but it differs in that both included parameters (the Ly α velocity offset and inferred dust extinction) fall into the category of empirical parameters we have associated with Ly α escape (Section 3.2), rather than including a proxy for the efficiency of Ly α production (Section 3.3). As with the Hayes et al. (2014) study, Yang et al. (2017) have the advantage over our own work of measuring individual spectroscopic parameters at high S/N, but they include a much smaller sample size. The Yang et al. (2017) sample also only includes galaxies selected to be spatially compact, low mass, and high excitation, whereas the KBSS sample includes $15\times$ more galaxies over a much broader range of properties. Nonetheless, the different selection biases and relative advantages of low- and high-redshift galaxy samples make these $z \sim 0$ – 0.3 surveys (including continued *HST*/COS spectroscopy) an effective complement to the work presented here.

While not directly focused on predicting Ly α emission, recent work from the HiZELS survey (Geach et al. 2008; Sobral et al. 2009) has pointed to the complex relationships between Ly α and other recombination line emission. In particular, Oteo et al. (2015) demonstrate that galaxies selected on the basis of H α emission display only weak Ly α emission on average. This result is consistent with the first panel of Figure 9 and the discussion in Section 3.3.1 of this paper, as we also find $L_{\text{H}\alpha}$ to be a poor predictor of Ly α emission. Although the HiZELS galaxies do not have deep rest-UV spectra, we expect that their low Ly α escape fractions would be associated with strong LIS absorption, such that they may lie near cB58 in the O3–EW_{LIS} plane displayed in Figure 5 (i.e., the parameter space associated with high rates of intrinsic Ly α production but low rates of Ly α escape).

In more recent work, Du et al. (2018) present a systematic study of the redshift evolution of rest-UV spectroscopic properties of galaxies over $z \approx 2$ – 4 , including the variation of Ly α with other galaxy properties in this epoch. The Du et al. (2018) spectroscopic sample is also drawn from the KBSS and has substantial overlap with the galaxies presented here. Broadly, the authors find that the relationship between EW_{Ly α} and EW_{LIS} is invariant with redshift for $2 \lesssim z \lesssim 4$, with a similarly invariant (but weaker) relationship between EW_{Ly α} and $E(B - V)_{\text{SED}}$. In addition to these indicators of Ly α escape, Du et al. (2018) argue that the association between the equivalent width of C III] and EW_{Ly α} (which they find to be similar at $z \sim 2$ and $z \sim 3$) represents the dependence of EW_{Ly α} on the intrinsic production rate of Ly α emission. One

interesting point of comparison between their results and those presented here regards the relationship between EW_{Ly α} and M_{UV} ; in their $z \sim 2$ galaxy bin (which is most similar to the galaxies presented here), Du et al. (2018) find no relationship between EW_{Ly α} and M_{UV} , but they find an increasingly strong relationship with increasing redshift. Likewise, Oyarzún et al. (2017) find a positive correlation between M_{UV} and EW_{Ly α} , but the relationship appears to rely on the inclusion of lower-luminosity galaxies than are included in this sample (although similar to the galaxies described in Trainor et al. 2015, 2016; see Figures 5 and 6 here) and the extension to higher redshifts. This variation may help explain the high EW_{Ly α} values seen generically in the lowest-luminosity galaxies at $z \sim 2$ (e.g., Stark et al. 2013), which are perhaps better analogs for typical galaxies at the highest redshifts than the more luminous $z \sim 2$ galaxies described in this paper.

The results of Du et al. (2018) are in general agreement with those presented here, with the exception that Du et al. (2018) limit their analysis to composite rest-UV spectra (i.e., they include no rest-optical data, nor do they analyze the spectra of individual galaxies), and they consider the redshift evolution of these trends. The Du et al. composite spectra achieve higher-S/N measurements of individual features than the measurements we consider here, but they also smooth over the intrinsic object-to-object variation among galaxies—variation that we highlight in this paper, particularly in Section 5. Together, therefore, these two studies provide a comprehensive view of the average trends among net Ly α emission and the processes of production and escape, while also demonstrating the substantial stochasticity that accompanies these broader trends.

Another complementary aspect of these works is that Du et al. (2018) demonstrate that individual parameters related to Ly α production and escape (i.e., EW_{LIS}, $E(B - V)$), and proxies for nebular excitation) show similar relationships with EW_{Ly α} across $2 < z < 4$, despite the fact that the ubiquity of EW_{Ly α} emission itself (and its dependence on M_{UV} , SFR, and M_*) evolves significantly over this period. This invariance suggests that the models for Ly α emission developed here—particularly those shown in Figures 6 and 8—may be expected to remain useful even at higher redshifts where the intrinsic Ly α emission of galaxies is more difficult to directly measure.

7. Conclusions

We have presented an empirical analysis of factors affecting Ly α production and escape in a sample of 703 star-forming galaxies from the KBSS at $z \approx 1.5$ – 3.5 . Our primary indicators of Ly α escape efficiency include the velocity offset of the Ly α line and the equivalent width in absorption of low-ionization interstellar lines, EW_{LIS}, and we find that these proxies for Ly α escape are strongly associated with the directly measured Ly α escape fraction, f_{esc} . Our indicators of Ly α production include the O3 and O32 ratios, which we argue are effective diagnostics of the ionization conditions within the H II regions from which Ly α photons originate. Several other galaxy parameters, including stellar mass, SFR, luminosity, and reddening, are shown to have much weaker relationships with observed Ly α emission.

We propose that EW_{LIS} and O3 are the most useful predictors of EW_{Ly α} because of their potential for observability in cases when the Ly α line is not directly detectable (or may be strongly affected by IGM absorption) and because of their strong individual correlations with EW_{Ly α} and lack of

correlation with each other. We then construct a new quantity, $X_{\text{LIS}}^{\text{O3}}$, which is a linear combination of EW_{LIS} and O3 with coefficients chosen to maximize the association between $X_{\text{LIS}}^{\text{O3}}$ and $\text{EW}_{\text{Ly}\alpha}$.

We find that the combination of O3 and EW_{LIS} predicts net $\text{EW}_{\text{Ly}\alpha}$ with less scatter than any single variable captured by our survey that does not require measurement of the $\text{Ly}\alpha$ line; $\sim 50\%$ of the ordering in observed $\text{EW}_{\text{Ly}\alpha}$ is captured by $X_{\text{LIS}}^{\text{O3}}$. After accounting for measurement uncertainties and fitting an exponential model for $\text{EW}_{\text{Ly}\alpha}$ as a function of $X_{\text{LIS}}^{\text{O3}}$, we estimate that the combination of our model and observational error account for 90% of the total variance in $\text{EW}_{\text{Ly}\alpha}$ at fixed $X_{\text{LIS}}^{\text{O3}}$.

We also estimate the conditional probability of detecting net $\text{Ly}\alpha$ emission or absorption in slit spectroscopy of a galaxy as a function of various galaxy parameters. We find that galaxies with $X_{\text{LIS}}^{\text{O3}} > 0.6$ have an 80% probability of being net $\text{Ly}\alpha$ emitters, while those with $X_{\text{LIS}}^{\text{O3}} < 0$ have less than a 25% probability of exhibiting net emission. Similarly strong variation in the probability of net $\text{Ly}\alpha$ emission is seen when adopting a prior based on O32, while constraints on photometric or SED-fit parameters or $\text{H}\alpha$ -based SFR have negligible utility as priors over the parameter space probed by our sample.

Given the many factors affecting net $\text{Ly}\alpha$ emission, our two-parameter model for $X_{\text{LIS}}^{\text{O3}}$ is remarkably successful at describing the variation in $\text{Ly}\alpha$ emission across a large, heterogeneous set of star-forming galaxies. We suggest that this success indicates that the wide variety of processes affecting $\text{Ly}\alpha$ emission can be broadly categorized as relating to $\text{Ly}\alpha$ production or escape, and capturing these two different “meta-parameters” is an essential component of any model for $\text{Ly}\alpha$ emission from galaxies.

We are indebted to the staff of the W. M. Keck Observatory who keep the instruments and telescopes running effectively. We also wish to extend thanks to those of Hawaiian ancestry on whose sacred mountain we are privileged to be guests. This work has been supported in part by the US National Science Foundation through grants AST-0908805 and AST-1313472. We also acknowledge support from the JPL/Caltech President’s and Director’s Program.

Software: Astropy (Astropy Collaboration et al. 2013), Matplotlib (Hunter 2007).

ORCID iDs

Ryan F. Trainor  <https://orcid.org/0000-0002-6967-7322>
 Allison L. Strom  <https://orcid.org/0000-0001-6369-1636>
 Charles C. Steidel  <https://orcid.org/0000-0002-4834-7260>
 Gwen C. Rudie  <https://orcid.org/0000-0002-8459-5413>
 Rachel L. Theios  <https://orcid.org/0000-0002-4236-1037>

References

- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, 558, A33
- Atek, H., Kunth, D., Schaerer, D., et al. 2009, *A&A*, 506, L1
- Atek, H., Malkan, M., McCarthy, P., et al. 2010, *ApJ*, 723, 104
- Bacon, R., Brinchmann, J., Richard, J., et al. 2015, *A&A*, 575, A75
- Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, *PASP*, 93, 5
- Borthakur, S., Heckman, T. M., Leitherer, C., & Overzier, R. A. 2014, *Sci*, 346, 216
- Brammer, G. B., van Dokkum, P. G., Franx, M., et al. 2012, *ApJS*, 200, 13
- Bruzual, G., & Charlot, S. 2003, *MNRAS*, 344, 1000
- Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, *ApJ*, 533, 682
- Cantalupo, S., Arrigoni-Battaia, F., Prochaska, J. X., Hennawi, J. F., & Madau, P. 2014, *Natur*, 506, 63
- Cantalupo, S., Lilly, S. J., & Haehnelt, M. G. 2012, *MNRAS*, 425, 1992
- Cantalupo, S., Porciani, C., Lilly, S. J., & Miniati, F. 2005, *ApJ*, 628, 61
- Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, *ApJ*, 345, 245
- Chabrier, G. 2003, *PASP*, 115, 763
- Cowie, L. L., & Hu, E. M. 1998, *AJ*, 115, 1319
- de Barros, S., Vanzella, E., Amorín, R., et al. 2016, *A&A*, 585, A51
- Dopita, M. A., & Sutherland, R. S. 2003, *Astrophysics of the Diffuse Universe* (Berlin, New York: Springer)
- Du, X., Shapley, A. E., Reddy, N. A., et al. 2018, *ApJ*, 860, 75
- Erb, D. K., Pettini, M., Steidel, C. C., et al. 2016, *ApJ*, 830, 52
- Erb, D. K., Steidel, C. C., Trainor, R. F., et al. 2014, *ApJ*, 795, 33
- Finkelstein, S. L., Hill, G. J., Gebhardt, K., et al. 2011, *ApJ*, 729, 140
- Fletcher, T. J., Robertson, B. E., Nakajima, K., et al. 2019, *ApJ*, 878, 87
- Geach, J. E., Smail, I., Best, P. N., et al. 2008, *MNRAS*, 388, 1473
- Hagen, A., Zeimann, G. R., Behrens, C., et al. 2016, *ApJ*, 817, 79
- Hayes, M., Östlin, G., Duval, F., et al. 2014, *ApJ*, 782, 6
- Hayes, M., Östlin, G., Schaerer, D., et al. 2010, *Natur*, 464, 562
- Henry, A., Scarlata, C., Martin, C. L., & Erb, D. 2015, *ApJ*, 809, 19
- Hoag, A., Bradač, M., Huang, K., et al. 2019, *ApJ*, 878, 12
- Hunter, J. D. 2007, *CSE*, 9, 90
- Izotov, Y. I., Schaerer, D., Thuan, T. X., et al. 2016, *MNRAS*, 461, 3683
- Izotov, Y. I., Worseck, G., Schaerer, D., et al. 2018, *MNRAS*, 478, 4851
- Jones, T., Stark, D. P., & Ellis, R. S. 2012, *ApJ*, 751, 51
- Kennicutt, R. C., Jr. 1998, *ARA&A*, 36, 189
- Kollmeier, J. A., Zheng, Z., Davé, R., et al. 2010, *ApJ*, 708, 1048
- Kunth, D., Mas-Hesse, J. M., Terlevich, E., et al. 1998, *A&A*, 334, 11
- Martin, C. L., Dijkstra, M., Henry, A., et al. 2015, *ApJ*, 803, 6
- Mason, C. A., Treu, T., Dijkstra, M., et al. 2018, *ApJ*, 856, 2
- McLean, I. S., Steidel, C. C., Epps, H. W., et al. 2012, *Proc. SPIE*, 8446, 84460J
- McLinden, E. M., Finkelstein, S. L., Rhoads, J. E., et al. 2011, *ApJ*, 730, 136
- Momose, R., Ouchi, M., Nakajima, K., et al. 2014, *MNRAS*, 442, 110
- Momose, R., Ouchi, M., Nakajima, K., et al. 2016, *MNRAS*, 457, 2318
- Nakajima, K., Ellis, R. S., Iwata, I., et al. 2016, *ApJL*, 831, L9
- Nakajima, K., Ouchi, M., Shimasaku, K., et al. 2013, *ApJ*, 769, 3
- Oke, J. B., Cohen, J. G., Carr, M., et al. 1995, *PASP*, 107, 375
- Osterbrock, D. E. 1989, *Astrophysics of Gaseous Nebulae and Active Galactic Nuclei* (Mill Valley, CA: Univ. Science Books)
- Östlin, G., Hayes, M., Duval, F., et al. 2014, *ApJ*, 797, 11
- Oteo, I., Sobral, D., Ivison, R. J., et al. 2015, *MNRAS*, 452, 2018
- Ouchi, M., Harikane, Y., Shibuya, T., et al. 2018, *PASJ*, 70, S13
- Oyarzún, G. A., Blanc, G. A., González, V., Mateo, M., & Bailey, J. I. 2017, *ApJ*, 843, 133
- Oyarzún, G. A., Blanc, G. A., González, V., et al. 2016, *ApJL*, 821, L14
- Pellegrini, E. W., Oey, M. S., Winkler, P. F., et al. 2012, *ApJ*, 755, 40
- Pentericci, L., Fontana, A., Vanzella, E., et al. 2011, *ApJ*, 743, 132
- Pentericci, L., Vanzella, E., Castellano, M., et al. 2018, *A&A*, 619, A147
- Pettini, M., Rix, S. A., Steidel, C. C., et al. 2002, *ApJ*, 569, 742
- Pettini, M., Steidel, C. C., Adelberger, K. L., Dickinson, M., & Giavalisco, M. 2000, *ApJ*, 528, 96
- Rakic, O., Schaye, J., Steidel, C. C., & Rudie, G. C. 2012, *ApJ*, 751, 94
- Reddy, N. A., Pettini, M., Steidel, C. C., et al. 2012, *ApJ*, 754, 25
- Reddy, N. A., Steidel, C. C., Pettini, M., et al. 2008, *ApJS*, 175, 48
- Rhoads, J. E., Malhotra, S., Dey, A., et al. 2000, *ApJL*, 545, L85
- Rivera-Thorsen, T. E., Hayes, M., Östlin, G., et al. 2015, *ApJ*, 805, 14
- Roberts-Borsani, G. W., Bouwens, R. J., Oesch, P. A., et al. 2016, *ApJ*, 823, 143
- Robertson, B. E., Ellis, R. S., Furlanetto, S. R., & Dunlop, J. S. 2015, *ApJL*, 802, L19
- Rousseuw, P. J., & Croux, C. 1993, *J. Am. Stat. Assoc.*, 88, 1273
- Rudie, G. C., Steidel, C. C., Trainor, R. F., et al. 2012, *ApJ*, 750, 67
- Sanders, R. L., Shapley, A. E., Kriek, M., et al. 2016, *ApJ*, 816, 23
- Schenker, M. A., Ellis, R. S., Konidaris, N. P., & Stark, D. P. 2013, *ApJ*, 777, 67
- Schenker, M. A., Stark, D. P., Ellis, R. S., et al. 2012, *ApJ*, 744, 179
- Shapley, A. E., Steidel, C. C., Pettini, M., & Adelberger, K. L. 2003, *ApJ*, 588, 65
- Shapley, A. E., Steidel, C. C., Strom, A. L., et al. 2016, *ApJL*, 826, L24
- Siana, B., Teplitz, H. I., Chary, R.-R., Colbert, J., & Frayer, D. T. 2008, *ApJ*, 689, 59
- Sobral, D., Best, P. N., Geach, J. E., et al. 2009, *MNRAS*, 398, 75
- Song, M., Finkelstein, S. L., Gebhardt, K., et al. 2014, *ApJ*, 791, 3
- Stark, D. P., Ellis, R. S., Charlot, S., et al. 2017, *MNRAS*, 464, 469

- Stark, D. P., Ellis, R. S., Chiu, K., Ouchi, M., & Bunker, A. 2010, *MNRAS*, **408**, 1628
- Stark, D. P., Schenker, M. A., Ellis, R., et al. 2013, *ApJ*, **763**, 129
- Steidel, C. C., Adelberger, K. L., Shapley, A. E., et al. 2000, *ApJ*, **532**, 170
- Steidel, C. C., Adelberger, K. L., Shapley, A. E., et al. 2003, *ApJ*, **592**, 728
- Steidel, C. C., Bogosavljević, M., Shapley, A. E., et al. 2011, *ApJ*, **736**, 160
- Steidel, C. C., Bogosavljević, M., Shapley, A. E., et al. 2018, *ApJ*, **869**, 123
- Steidel, C. C., Erb, D. K., Shapley, A. E., et al. 2010, *ApJ*, **717**, 289
- Steidel, C. C., Rudie, G. C., Strom, A. L., et al. 2014, *ApJ*, **795**, 165
- Steidel, C. C., Shapley, A. E., Pettini, M., et al. 2004, *ApJ*, **604**, 534
- Steidel, C. C., Strom, A. L., Pettini, M., et al. 2016, *ApJ*, **826**, 159
- Strom, A. L., Steidel, C. C., Rudie, G. C., Trainor, R. F., & Pettini, M. 2018, *ApJ*, **868**, 117
- Strom, A. L., Steidel, C. C., Rudie, G. C., et al. 2017, *ApJ*, **836**, 164
- Teplitz, H. I., McLean, I. S., Becklin, E. E., et al. 2000, *ApJL*, **533**, L65
- Theios, R. L., Steidel, C. C., Strom, A. L., et al. 2019, *ApJ*, **871**, 128
- Trainor, R., & Steidel, C. C. 2013, *ApJL*, **775**, L3
- Trainor, R. F., & Steidel, C. C. 2012, *ApJ*, **752**, 39
- Trainor, R. F., Steidel, C. C., Strom, A. L., & Rudie, G. C. 2015, *ApJ*, **809**, 89
- Trainor, R. F., Strom, A. L., Steidel, C. C., & Rudie, G. C. 2016, *ApJ*, **832**, 171
- Veilleux, S., & Osterbrock, D. E. 1987, *ApJS*, **63**, 295
- Wisotzki, L., Bacon, R., Blaizot, J., et al. 2016, *A&A*, **587**, A98
- Yang, H., Malhotra, S., Gronke, M., et al. 2017, *ApJ*, **844**, 171