



# The Universe Is Reionizing at $z \sim 7$ : Bayesian Inference of the IGM Neutral Fraction Using Ly $\alpha$ Emission from Galaxies

Charlotte A. Mason<sup>1</sup> , Tommaso Treu<sup>1</sup> , Mark Dijkstra<sup>2</sup>, Andrei Mesinger<sup>3</sup>, Michele Trenti<sup>4,5</sup> , Laura Pentericci<sup>6</sup> ,  
Stephane de Barros<sup>7,8</sup> , and Eros Vanzella<sup>8</sup>

<sup>1</sup> Department of Physics and Astronomy, UCLA, Los Angeles, CA, 90095-1547, USA; [cmason@astro.ucla.edu](mailto:cmason@astro.ucla.edu)

<sup>2</sup> Institute of Theoretical Astrophysics, University of Oslo, P.O. Box 1029, NO-0315 Oslo, Norway

<sup>3</sup> Scuola Normale Superiore, Piazza dei Cavalieri 7, I-56126 Pisa, Italy

<sup>4</sup> School of Physics, University of Melbourne, Parkville, Victoria, Australia

<sup>5</sup> ARC Centre of Excellence for All Sky Astrophysics in Three Dimensions (ASTRO 3D), Australia

<sup>6</sup> INAF Osservatorio Astronomico di Roma, Via Frascati 33, I-00040 Monteporzio (RM), Italy

<sup>7</sup> Observatoire de Genève, Université de Genève, 51 Ch. des Maillettes, 1290 Versoix, Switzerland

<sup>8</sup> INAF Osservatorio Astronomico di Bologna, via Ranzani 1, I-40127 Bologna, Italy

Received 2017 September 14; revised 2018 February 13; accepted 2018 February 15; published 2018 March 20

## Abstract

We present a new flexible Bayesian framework for directly inferring the fraction of neutral hydrogen in the intergalactic medium (IGM) during the Epoch of Reionization (EoR,  $z \sim 6$ – $10$ ) from detections and non-detections of Lyman Alpha (Ly $\alpha$ ) emission from Lyman Break galaxies (LBGs). Our framework combines sophisticated reionization simulations with empirical models of the interstellar medium (ISM) radiative transfer effects on Ly $\alpha$ . We assert that the Ly $\alpha$  line profile emerging from the ISM has an important impact on the resulting transmission of photons through the IGM, and that these line profiles depend on galaxy properties. We model this effect by considering the peak velocity offset of Ly $\alpha$  lines from host galaxies' systemic redshifts, which are empirically correlated with UV luminosity and redshift (or halo mass at fixed redshift). We use our framework on the sample of LBGs presented in Pentericci et al. and infer a global neutral fraction at  $z \sim 7$  of  $\bar{x}_{\text{HI}} = 0.59_{-0.15}^{+0.11}$ , consistent with other robust probes of the EoR and confirming that reionization is ongoing  $\sim 700$  Myr after the Big Bang. We show that using the full distribution of Ly $\alpha$  equivalent width detections and upper limits from LBGs places tighter constraints on the evolving IGM than the standard Ly $\alpha$  emitter fraction, and that larger samples are within reach of deep spectroscopic surveys of gravitationally lensed fields and *James Webb Space Telescope* NIRSpec.

*Key words:* dark ages, reionization, first stars – galaxies: evolution – galaxies: high redshift – intergalactic medium

## 1. Introduction

In the first billion years of the universe's history, intergalactic hydrogen atoms, formed at Recombination, were ionized (e.g., Robertson et al. 2015; Mesinger 2016; Planck Collaboration et al. 2016). This reionization of the intergalactic medium (IGM) was driven by the first sources of light: stars, and accretion disks around black holes, in galaxies. By understanding the process and timeline of reionization, we can learn about the nature of these nascent populations of galaxies.

Ground-breaking observations within the last decade have provided significant information about this Epoch of Reionization (EoR,  $z \sim 6$ – $10$ ). With the largest near-IR instruments in space and on the ground, we have now discovered large populations of galaxies at  $z \gtrsim 6$  (e.g., McLure et al. 2010; Trenti et al. 2011; Bradley et al. 2012; Illingworth et al. 2013; Schenker et al. 2013b; Schmidt et al. 2014; Yue et al. 2014; Bouwens et al. 2015b; Finkelstein et al. 2015a; Calvi et al. 2016). Young stars in these galaxies are likely the primary sources of reionizing photons (e.g., Bouwens et al. 2003; Lehnert & Bremer 2003; Bunker et al. 2004; Yan & Windhorst 2004; Finkelstein et al. 2012a; Robertson et al. 2013; Schmidt et al. 2014), though a contribution from AGN cannot be excluded (Giallongo et al. 2015; Madau & Haardt 2015; Onoue et al. 2017): we do not know if sufficient hard ionizing photons escape from galaxies, as we do not fully understand the interactions between these early galaxies and their surrounding media.

Absorption features in quasar spectra suggest reionization was largely complete by  $z \sim 6$  ( $< 1$  Gyr after the Big Bang; e.g., Fan et al. 2006; Schroeder et al. 2013; McGreer et al. 2014; Venemans et al. 2015), while the electron scattering optical depth to the CMB (Planck Collaboration et al. 2015, 2016; Greig & Mesinger 2017b) indicates significant reionization was occurring at  $z \sim 7.8$ – $8.8$ . A robust constraint, albeit from a single sightline, on ongoing reionization comes from the absorption spectrum of the  $z = 7.1$  quasar ULASJ1120+0641, where Greig et al. (2017) recently inferred a neutral fraction of  $\bar{x}_{\text{HI}} = 0.40_{-0.19}^{+0.21}$ .

To produce a timeline of reionization consistent with the evolution suggested by observations generally requires optimistic assumptions about the numbers of as yet undetected ultra-faint galaxies (Shull et al. 2012; Robertson et al. 2013; Mason et al. 2015), which are likely the hosts of high redshift gamma-ray bursts (Kistler et al. 2009; Trenti et al. 2012), and/or the production efficiency and escape fraction of hard ionizing photons (Bouwens et al. 2015a; Ma et al. 2015; Grazian et al. 2017; Vanzella et al. 2017). However, the timeline of reionization is not well-constrained, especially beyond  $z \gtrsim 6$ , where quasars become extremely rare (Fan et al. 2001; Manti et al. 2016).

Into the EoR, a powerful probe of the IGM is the Lyman alpha (Ly $\alpha$ , 1216 Å) emission line from galaxies, which is strongly attenuated by neutral hydrogen (Haiman & Spaans 1999;

Malhotra & Rhoads 2004; Santos 2004; Verhamme et al. 2006; McQuinn et al. 2007a; Dijkstra 2014). Observing Ly $\alpha$  at high redshift gives us key insights into both the IGM ionization state and galaxy properties, and while quasars probably live in the densest regions of the early universe (Mesinger 2010), observing galaxies enables us to trace reionization in cosmic volumes in a less biased way.

Dedicated spectroscopic follow-up of young star-forming galaxies at high redshift, identified as photometric dropouts (Lyman Break galaxies, hereafter LBGs), combined with low redshift comparison samples (Hayes et al. 2013; Yang et al. 2016) show that the fraction of LBGs emitting Ly $\alpha$  increases with redshift (Stark et al. 2010; Hayes et al. 2011; Curtis-Lake et al. 2012; Cassata et al. 2015), likely because the dust fraction in galaxies decreases (e.g., Finkelstein et al. 2012b; Bouwens et al. 2014), which reduces the absorption of Ly $\alpha$  (Hayes et al. 2011). However, there is a potential smoking gun signature of reionization at  $z > 6$ : recent observations show a declining fraction of Ly $\alpha$  emitters in the LBG population with redshift (e.g., Fontana et al. 2010; Stark et al. 2010; Caruana et al. 2012, 2014; Treu et al. 2013; Faisst et al. 2014; Pentericci et al. 2014; Schenker et al. 2014; Tilvi et al. 2014), as well as an evolving Ly $\alpha$  luminosity function (e.g., Ouchi et al. 2010; Konno et al. 2017; Ota et al. 2017; Zheng et al. 2017), suggesting an increasingly neutral, but inhomogeneous, IGM (Dijkstra et al. 2013; Mesinger et al. 2015).

Robust conversions from observations to the IGM state are challenging, however, and current constraints from Ly $\alpha$  emission measurements show some tension. The sudden drop in Ly $\alpha$  emission from LBGs suggests a high neutral fraction at  $z \sim 7$ ,  $\bar{x}_{\text{H I}} \gtrsim 0.5$  (Dijkstra et al. 2013; Choudhury et al. 2015; Mesinger et al. 2015), whereas measurements from clustering of Ly $\alpha$  emitters at  $z = 6.6$  imply a lower neutral fraction ( $\bar{x}_{\text{H I}} \lesssim 0.5$ , Ouchi et al. 2010; Sobacchi & Mesinger 2015). These constraints are consistent within  $1\sigma$ , but the qualitative tension motivates a more thorough treatment of the properties of Ly $\alpha$  emitters during reionization. Given this, and that tight constraints on the reionization history can constrain properties of the sources of reionization (e.g., the minimum mass/luminosity of galaxies, and the escape fraction of ionizing photons; Mitra et al. 2015; Bouwens et al. 2016b; Greig & Mesinger 2017a, 2017b), we aim to develop a robust framework for inferring the ionization state of the IGM from observations of Ly $\alpha$  emission from galaxies.

The conversion from the evolving transmission of Ly $\alpha$  emission from galaxies to a constraint on the IGM ionization state is non-trivial and involves physics from pc to Gpc scales. Multiple observations (e.g., Treu et al. 2013; Pentericci et al. 2014; Becker et al. 2015) and simulations (Furlanetto et al. 2006; McQuinn et al. 2007b) suggest reionization of the IGM is likely a “patchy” process, with large ionized bubbles growing faster in overdense regions filled with star-forming galaxies. An accurate model of reionization must include realistic large-scale IGM structure (Trac et al. 2008; Iliev et al. 2014; Sobacchi & Mesinger 2014).

Irrespective of reionization, as a highly resonant line, Ly $\alpha$  photons experience significant scattering within the interstellar medium (ISM) of their host galaxies, and absorption within the circumgalactic medium (CGM) which affects the visibility of emission (Verhamme et al. 2006, 2008; Dijkstra et al. 2007; Laursen et al. 2011). ISM effects on Ly $\alpha$  are likely to correlate

with galaxy mass and star formation rate (SFR) via dust absorption, neutral hydrogen column density and covering fraction, and outflows (Erb et al. 2014; Erb 2015; Oyarzún et al. 2016; Hayward & Hopkins 2017; Yang et al. 2017).

UV faint galaxies ( $M_{\text{UV}} > M^* \sim -20$ ) tend to be the strongest Ly $\alpha$  emitters at all redshifts due to lower dust masses and neutral hydrogen column densities (Yang et al. 2016, 2017). However, a small sample of UV bright galaxies at  $z > 7.5$  with strong *Spitzer*/IRAC excesses have recently been observed with Ly $\alpha$  (Finkelstein et al. 2013; Oesch et al. 2015; Zitrin et al. 2015; Roberts-Borsani et al. 2016; Stark et al. 2017), at a redshift when the IGM is expected to be significantly neutral (Planck Collaboration et al. 2016; Greig & Mesinger 2017b). Are these objects a new class of highly ionizing galaxies (Stark et al. 2017), emitting Ly $\alpha$  with very high EW, so some flux is still observable even after attenuation in the IGM? Do they inhabit large ionized bubbles in the IGM at high redshift? How do different halo environments and ISM properties affect the impact on reionization on galaxies?

Dijkstra et al. (2011) first considered the effects of the ISM on Ly $\alpha$  photons during reionization, using shell models (e.g., Verhamme et al. 2006; Gronke et al. 2015a) to mimic the ISM radiative transfer, and showed ISM effects had a large impact on the transmission of Ly $\alpha$  photons through the reionizing IGM. As described above, the Ly $\alpha$  photons’ journey through the ISM depends on galaxy properties. However, previous constraints on the evolving transmission of Ly $\alpha$  emission at  $z \gtrsim 6$  have limited treatment of this effect: Dijkstra et al. (2011) and Mesinger et al. (2015) parametrically accounted for the ISM but assumed the LBG galaxy population is homogeneous; Jensen et al. (2013) obtained similar results combining cosmological hydro-simulations of reionization with a different sub-grid prescription for Ly $\alpha$  radiative transfer in the ISM; simpler models do not treat the ISM but consider two bins of UV bright and faint galaxies (e.g., Treu et al. 2012).

In this paper we introduce a flexible modeling framework to enable Bayesian inference of the IGM neutral fraction from detections and non-detections of Ly $\alpha$  from LBGs. Our framework includes realistic cosmological IGM simulations that contain the large-scale structure of the reionizing IGM. We generate thousands of sightlines through these simulations to halos born from the same density field as the IGM, and populate these halos with simple, but realistic, ISM properties drawn from empirical models, which, for the first time in a reionization model, are linked to observable galaxy properties.

Our model asserts the impact of the ISM on the Ly $\alpha$  line profile is the most important galaxy property to consider when trying to make accurate inferences about reionization. In our model we include this effect via the peak velocity offset of the Ly $\alpha$  line profile from systemic ( $\Delta v$ ), which correlates with galaxy mass (or UV magnitude at fixed redshift), for which there are a handful of measurements at  $z \gtrsim 6$  (Pentericci et al. 2016; Bradač et al. 2017; Mainali et al. 2017; Stark et al. 2017). Galaxies with high Ly $\alpha$  velocity offsets have higher probabilities of transmitting Ly $\alpha$  photons through the IGM. This effect is robustly accounted for in our model as a nuisance parameter in our inference.

The paper is structured as follows: in Section 2 we explain the ISM, CGM, and IGM radiative transfer modeling components of our model; in Section 3 we describe our flexible Bayesian framework for inferring the neutral fraction

$\bar{x}_{\text{HI}}$ ; in Section 4 we give our results, including key insights from the model, the inferred value of  $\bar{x}_{\text{HI}}$  from current observations, and forecasts for spectroscopic surveys with the *James Webb Space Telescope (JWST)*; we discuss our results in Section 5 and present a summary and conclusions in Section 6.

We use the Planck Collaboration et al. (2015) cosmology where  $(\Omega_{\Lambda}, \Omega_{\text{m}}, \Omega_{\text{b}}, n, \sigma_8, H_0) = (0.69, 0.31, 0.048, 0.97, 0.81, 68 \text{ km s}^{-1} \text{ Mpc}^{-1})$ , and all magnitudes are given in the AB system.

## 2. ISM, CGM, and IGM Radiative Transfer Modeling

$\text{Ly}\alpha$  photons are significantly affected by the neutral hydrogen they encounter within the ISM of their source galaxies, their local CGM, and the IGM through which they travel to our telescopes. To make constraints in the Epoch of Reionization, we must model  $\text{Ly}\alpha$  radiative transfer in all three media. Here we describe the combination of empirical formalisms and numerical simulations to model the effect of the ISM (Section 2.1) and the CGM and IGM (Section 2.2) on  $\text{Ly}\alpha$  transmission.

### 2.1. ISM $\text{Ly}\alpha$ Radiative Transfer

$\text{Ly}\alpha$  photons are produced predominantly via recombination in H II regions around young stars and have a high cross-section for resonant scattering (for a detailed review, see Dijkstra 2014). As the ISM of individual galaxies contains a large amount of neutral hydrogen gas to escape the ISM,  $\text{Ly}\alpha$  photons must diffuse both spatially and spectrally (e.g., Shapley et al. 2003; McLinden et al. 2011; Chonis et al. 2013; Mostardi et al. 2013; Song et al. 2014). This produces the fiducial double-peaked  $\text{Ly}\alpha$  lineshape, for which the red (blue) peak is enhanced for outflows (inflows; Zheng & Miralda-Escude 2002; Verhamme et al. 2006).

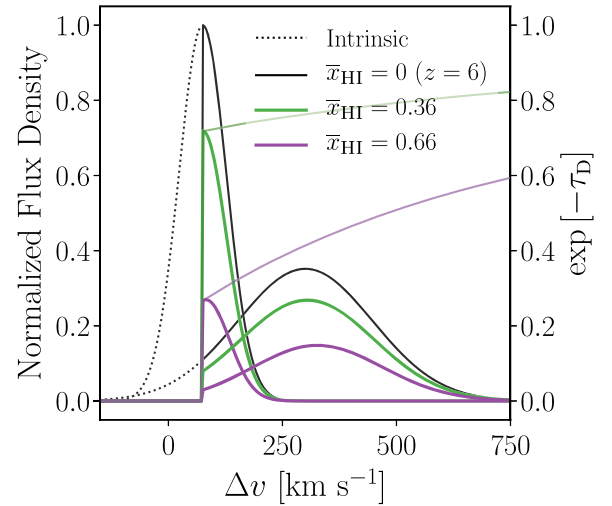
In this work, we model the  $\text{Ly}\alpha$  lineshape after transmission through the ISM as a Gaussian, centered at a velocity offset  $\Delta v$  from the systemic redshift of the galaxy (due to scattering through the ISM, described in Section 2.1.1) with a velocity dispersion  $\sigma_{\alpha}$  (due to scattering and thermal broadening in the ISM, described in Section 2.1.2). We refer to this lineshape as “intrinsic”; examples are shown as dotted black lines in Figure 1. As described in Section 2.2, even after reionization residual neutral gas in the IGM and CGM will absorb all blue flux at  $z \gtrsim 6$ .

#### 2.1.1. Modeling $\text{Ly}\alpha$ Velocity Offsets

Numerous studies of star-forming galaxies at  $z \lesssim 4$  have identified the column density of neutral hydrogen ( $N_{\text{HI}}$ ) within the ISM as a key mediator of  $\text{Ly}\alpha$  radiative transfer.  $\text{Ly}\alpha$  photons traveling through highly dense neutral ISM scatter more frequently and emerge with larger velocity offsets than galaxies with lower  $N_{\text{HI}}$  (Shibuya et al. 2014; Hashimoto et al. 2015; Yang et al. 2016, 2017; Guaita et al. 2017).

Low mass galaxies, especially at high redshifts, are less likely to contain significant fractions of neutral gas due to enhanced photoionization feedback. Additionally, strong star formation feedback may drive outflows and/or reduce the covering fraction of neutral gas in the ISM, which can facilitate  $\text{Ly}\alpha$  escape (Jones et al. 2013; Trainor et al. 2015; Leethochawalit et al. 2016).

Recently, a correlation has been suggested between UV magnitude and  $\text{Ly}\alpha$  velocity offset (Schenker et al. 2013a; Erb et al. 2014; Song et al. 2014; Stark et al. 2015, 2017;



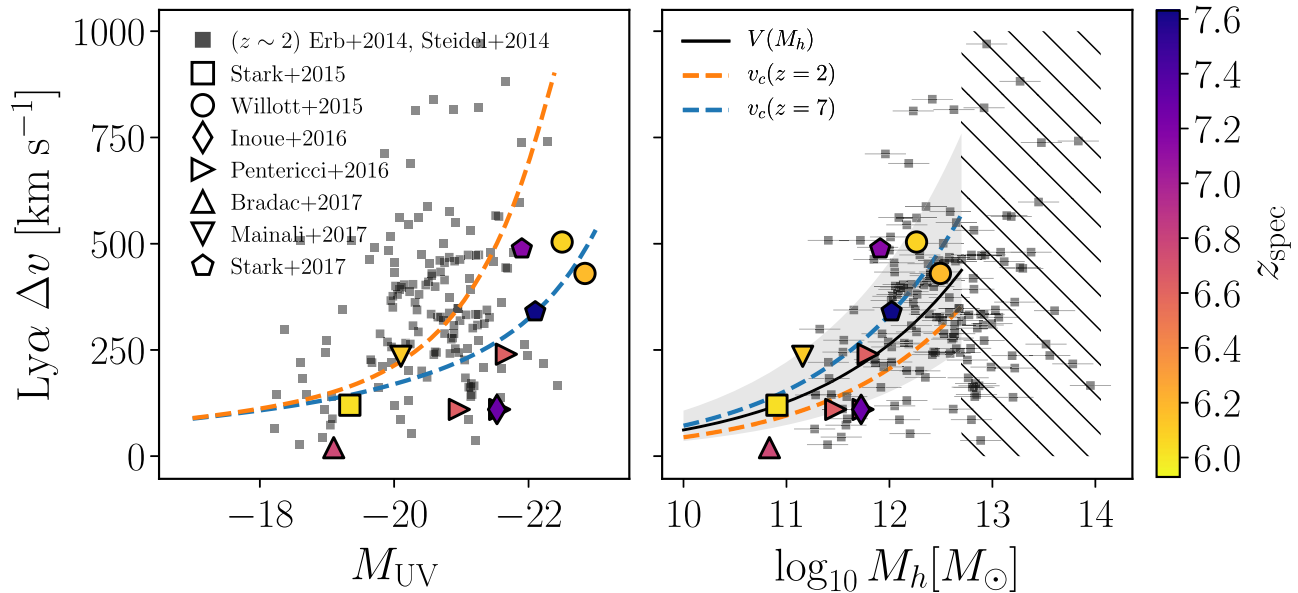
**Figure 1.** Effect of the IGM on simulated line profiles. We show two example intrinsic line profiles (black dotted lines), with peak velocity offsets of 75 and 300  $\text{km s}^{-1}$ , with flux densities normalized to that of the line at 75  $\text{km s}^{-1}$ . This is the line after transmission through the ISM. The solid black shows the line shape in an ionized universe at  $z \sim 6$ , where all flux bluer than the halo’s circular velocity is resonantly absorbed by neutral hydrogen in the local CGM/IGM (i.e., they experience only  $\tau_{\text{HII}}$ ). The colored lines show the emission lines after transmission through a reionizing IGM with damping wing optical depths  $\tau_{\text{D}}$ , where the median IGM attenuation is also plotted (lighter line, corresponds to right axis). Lines emitted with high velocity offsets are less attenuated by the IGM: for the line with  $\Delta v = 75 \text{ km s}^{-1}$ ,  $\sim 70\%$  of the emitted flux is observed for  $\bar{x}_{\text{HI}} = 0.36$  (green); for the line with  $\Delta v = 300 \text{ km s}^{-1}$ , this fraction rises to  $\sim 75\%$ . For  $\bar{x}_{\text{HI}} = 0.66$  (purple),  $\sim 30\%$  of the total flux is transmitted from the line with  $\Delta v = 75 \text{ km s}^{-1}$ , while  $\sim 40\%$  is emitted for the line at  $\Delta v = 300 \text{ km s}^{-1}$ .

Mainali et al. 2017), again indicating that galaxy mass and/or SFR strongly affects  $\text{Ly}\alpha$  escape. However, galaxies with the same UV magnitudes at different redshifts likely have very different masses because of increasing SFR at high redshift (e.g., Behroozi et al. 2013b; Barone-Nugent et al. 2014; Finkelstein et al. 2015b; Mason et al. 2015; Harikane et al. 2016, 2017), so one should be cautious of comparing galaxies with the same UV magnitudes at different redshifts. We plot a compilation of  $M_{\text{UV}}-\Delta v$  measurements from the literature (Erb et al. 2014; Steidel et al. 2014; Stark et al. 2015, 2017; Willott et al. 2015; Inoue et al. 2016; Pentericci et al. 2016; Bradač et al. 2017; Mainali et al. 2017) in Figure 2 (left), where it is clear the high redshift galaxies have lower  $\Delta v$  at given  $M_{\text{UV}}$  compared to the low redshift galaxies, probably because they have lower mass.

To model the effect of the ISM on  $\text{Ly}\alpha$  escape, we assume the column density of neutral hydrogen within the ISM is determined by halo mass and is the most important quantity for understanding the emerging  $\text{Ly}\alpha$  line profile. This is likely an oversimplification—for example, “shell” models take  $\sim 6$  parameters to model  $\text{Ly}\alpha$  lines (Verhamme et al. 2006, 2008; Gronke & Dijkstra 2016)—but is an efficient first-order approach. With this in mind, we assume a correlation between  $\Delta v$  and halo mass of the form  $\Delta v \sim (M_h)^m$ , where we determine  $m$  empirically from observations, as described below.

We take a sample of 158  $z \sim 2-3$  galaxies with both UV magnitudes and  $\text{Ly}\alpha$  velocity offsets from Erb et al. (2014) and Steidel et al. (2014). The Steidel et al. (2014) sample (from the KBSS-MOSFIRE survey) is effectively complete at  $M_{\text{UV}} \lesssim -20$ , where  $\sim 90\%$  of their photometrically selected LBGs have





**Figure 2.** Ly $\alpha$  velocity offset as a function of UV absolute magnitude (left) and halo mass (right, derived from the Mason et al. 2015 UV magnitude–halo mass relation) for a collection of data from the literature (Erb et al. 2014; Steidel et al. 2014; Willott et al. 2015; Stark et al. 2015, 2017; Inoue et al. 2016; Pentericci et al. 2016; Mainali et al. 2017; Bradač et al. 2017). The gray squares show data from a  $z \sim 2$  sample, while the colored points are at  $z > 6$ . We take the  $z \sim 2$  distribution as complete and intrinsic, and fit a log-normal distribution to the  $\Delta v - M_h$  points, as shown in Equation (1). The median  $\Delta v - M_h$  fit is shown as a black solid line, and the gray shaded region shows the  $\sigma_v$  scatter. We add a 0.2 mag uncertainty to the UV magnitude measurements to account for scatter in the UV magnitude–halo mass relation and use the propagated uncertainties in halo mass in the  $\Delta v - M_h$ . The hashed region in the right panel indicates the galaxies with  $M_{UV} < -21$ , which are discarded from fitting due to large uncertainties in assigning their halo masses. We plot the circular velocities,  $v_c$ , of halos at  $z \sim 2$  (dashed orange) and  $z \sim 7$  (dashed blue) for comparison. The  $\Delta v - M_h$  relation closely traces the circular velocities, suggesting galaxy mass is a key mediator of Ly $\alpha$  radiative transfer.

rest-frame optical emission lines detected in deep near-IR spectroscopy with *Keck*/MOSFIRE (McLean et al. 2012). The Erb et al. (2014) sample comprises 36 galaxies selected as Ly $\alpha$  emitters in narrow-band photometry with  $-21 \lesssim M_{UV} \lesssim -18$ ; all these objects had rest-frame optical lines detected in MOSFIRE observations. We note the Erb et al. (2014) sample does not include faint Ly $\alpha$  emitters ( $W \lesssim 25 \text{ \AA}$ ), which may have higher velocity offsets, given observed anti-correlations between Ly $\alpha$  EW and  $\Delta v$  (Hashimoto et al. 2013; Shibuya et al. 2014; Erb et al. 2014). While these samples are the largest available to measure an  $M_{UV} - \Delta v$  correlation, future rest-frame optical follow-up of large samples of galaxies with detected Ly $\alpha$  emission (e.g., from the HETDEX and MUSE-Wide spectroscopic surveys; Song et al. 2014; Herenz et al. 2017) will provide more complete information about the relationships between Ly $\alpha$  radiative transfer in the ISM and galaxy properties.

As described above, it is difficult to directly compare galaxies at fixed UV magnitude across cosmic time, so we map UV magnitudes to halo mass. To first order, the depth of a halo’s gravitational potential well is the dominant influence on galaxy properties independent of redshift (Behroozi et al. 2013a; Mason et al. 2015; Moster et al. 2017). We assume no redshift evolution between halo mass and velocity offset. We convert UV magnitude to halo mass using the successful model derived by Mason et al. (2015), which assumes the SFR is proportional to the halo mass assembly rate at a given halo mass and redshift, and is consistent with  $M_{UV} - M_h$  measurements from clustering at  $z \sim 7$  (Barone-Nugent et al. 2014; Harikane et al. 2016, 2017). We add a 0.2 mag uncertainty to the UV magnitude measurements to account for scatter in the UV magnitude–halo mass relation (e.g., Finkelstein et al. 2015b) and use the propagated uncertainties in halo mass in  $\Delta v - M_h$ .

In the right panel of Figure 2 we plot the literature  $\Delta v$  measurements as a function of the estimated halo masses. Due to the uncertainties in mapping from UV magnitude to halo mass for very bright galaxies at  $z \lesssim 4$ , which may be significantly more starbursty than average, we discard the  $z \sim 2$  galaxies with  $M_{UV} < -21$  from further analysis. Likewise, we exclude from this inference the galaxies at  $z \sim 7$  with  $M_{UV} < -22$ , deferring their analysis to a later paper (Mason et al. 2018).

When we transform to halo mass, the high redshift literature points clearly lie within the low redshift data space. This suggests halo mass is a useful approximately redshift independent indicator of Ly $\alpha$  escape routes. We note gravitationally lensed objects at intermediate redshifts suggest these trends hold at low mass/luminosity (e.g., a lensed  $M_{UV} = -17$  galaxy at  $z \sim 3$  was recently observed with a Ly $\alpha$  velocity offset of 51 km s $^{-1}$ ; Vanzella et al. 2016). Further studies, using NIRSpec on *JWST*, will be able to investigate these trends at high redshifts.

The distribution is well-described by a log-normal distribution with a peak that increases with increasing luminosity, and approximately constant variance:

$$p(\Delta v | M_h) = \frac{\exp\left[-\frac{(\log_{10} \Delta v - V(M_h))^2}{2\sigma_v^2}\right]}{\Delta v \ln 10 \sqrt{2\pi} \sigma_v}, \quad (1)$$

where  $V$  is a linear relation corresponding to the most likely  $\log_{10}(\Delta v)$  at a given halo mass,

$$V(M_h) = m \log_{10}\left(\frac{M_h}{1.55 \times 10^{12} M_\odot}\right) + c. \quad (2)$$

To find the parameters  $m$ ,  $c$ , and  $\sigma_v$ , we take Equation (1) as the likelihood function and perform a Bayesian inference on the

$z \sim 2$  galaxies with  $M_{\text{UV}} > -21$ , with uniform priors on the parameters. The inferred parameters are  $m = 0.32 \pm 0.07$ ,  $c = 2.48 \pm 0.03$ , and  $\sigma_v = 0.24 \pm 0.02$ . We show this relation in Figure 2.

We can obtain an approximate relation between velocity offset, UV magnitude, and redshift by approximating the Mason et al. (2015) UV magnitude–halo mass relation as broken linear:  $\log_{10} M_h [M_\odot] \approx \gamma(M_{\text{UV}} + 20.0 + 0.26z) + 11.75$ , where  $\gamma = -0.3$  for  $M_{\text{UV}} \geq -20.0 - 0.26z$ , and  $\gamma = -0.7$  otherwise. The mean velocity offset in  $\text{km s}^{-1}$  can then be approximated as

$$\log_{10} \Delta v(M_{\text{UV}}, z) \approx 0.32\gamma(M_{\text{UV}} + 20.0 + 0.26z) + 2.34. \quad (3)$$

In this work we sample directly from the distribution in Equation (2) to calculate velocity offsets directly for simulated halos, including scatter.

In Figure 2 we also plot the circular velocity ( $v_c = [10GM_h H(z)]^{1/3}$ ) at  $z = 2$  and  $z = 7$  for comparison with the observed data. The circular velocities are higher at low redshifts, as halos are less dense and more extended, but there is a clear similarity in our derived trend  $\Delta v \sim M_h^{0.32}$  and the circular velocity  $v_c \sim M_h^{1/3}$ . Investigating these trends with larger samples at low redshifts with dynamical mass measurements (e.g., Erb et al. 2014; Trainor et al. 2015) could determine to what extent  $\text{Ly}\alpha$  radiative transfer depends on the gravitational potential of the halo.

### 2.1.2. Modeling $\text{Ly}\alpha$ Line Widths

The widths of  $\text{Ly}\alpha$  lines are also likely dominated by radiative transfer effects that both shift and broaden the line (Verhamme et al. 2006, 2008; Gronke et al. 2016).  $\text{Ly}\alpha$  velocity dispersions are also observed to be systematically higher than in nebular emission lines, which are not resonantly scattered (Trainor et al. 2015).

For simplicity we model the FWHM of the  $\text{Ly}\alpha$  lines as equal to the velocity offset of the line, which accounts for the broadening of the lines through scattering and is a good approximation for the observed correlation between  $\text{Ly}\alpha$  FWHM and velocity offset (Yang et al. 2016, 2017; A. Verhamme et al. 2018, in preparation).

### 2.1.3. EW Distribution in an Ionized Universe

The key observable of  $\text{Ly}\alpha$  emission lines at high redshift is their equivalent width (EW or  $W$ ), a measure of the brightness of the emission line relative to the UV continuum. As  $\text{Ly}\alpha$  photons from high redshift galaxies are attenuated by neutral gas in the intervening CGM and IGM, we observed only a fraction,  $\mathcal{T}_{\text{IGM}}$  (the  $\text{Ly}\alpha$  transmission fraction), of the emitted EW (i.e.,  $W_{\text{obs}} = W_{\text{em}} \times \mathcal{T}_{\text{IGM}}$ ), where  $W_{\text{em}}$  is the emitted equivalent width without any damping due to reionization.

In this work we consider the differential evolution of  $\text{Ly}\alpha$  equivalent widths between  $z \sim 6$  and  $z \sim 7$ , and assume the distribution of equivalent widths changes only because of the increasing neutrality of the IGM due to reionization. This is likely a simplification, as trends at lower redshifts show increasing EW with redshift as dust decreases in galaxies (Hayes et al. 2011), but the time between  $z \sim 6$  and  $z \sim 7$  is short ( $< 200$  Myr). If the underlying EW distribution does evolve significantly during that time, it will likely be to increase the emitted EW (due to decreasing dust; Hayes et al. 2011); thus the reduction due to reionization would need to be

greater to match the observed EW distribution at  $z \sim 7$  (Dijkstra et al. 2011). Papovich et al. (2011) suggests there may be an increase in gas reservoirs with increasing redshifts due to rapid accretion rates, which could also reduce the emission of strong  $\text{Ly}\alpha$ .

Thus observed equivalent widths at  $z \sim 7$  are  $W_7 = W_6 \times \mathcal{T}_{\text{IGM},7}/\mathcal{T}_{\text{IGM},6}$ , where  $\mathcal{T}_{\text{IGM},z}$  is the transmission fraction of  $\text{Ly}\alpha$  emission for a single object at redshift  $z$ . In Section 2.2 we describe the calculation of transmission fractions along thousands of lines of sight using state-of-the-art cosmological reionization simulations.

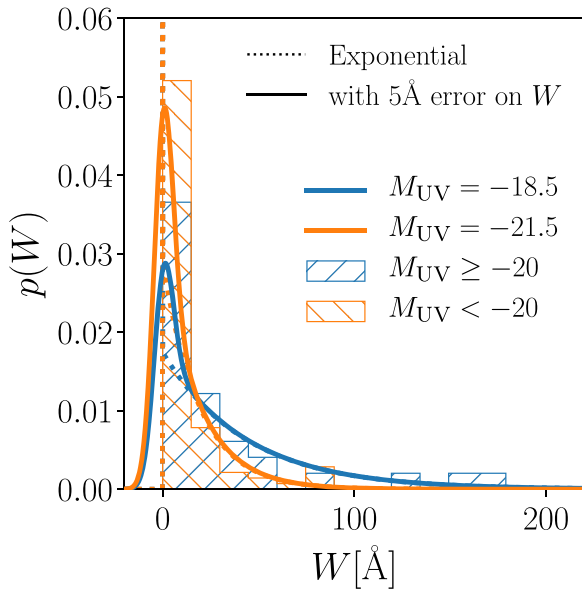
A key input to the model, then, is the  $z \sim 6$  distribution of EW as a function of galaxy properties.  $\text{Ly}\alpha$  EWs for UV continuum-selected galaxies have an observed equivalent width distribution with a peak at zero and some tail to high EW—which is usually parameterized as an exponential function (Dijkstra & Wyithe 2012), log-normal (Schenker et al. 2014), or truncated normal distribution plus a delta function (Treu et al. 2012). At  $z \lesssim 2$ , where large samples exist, including the local “Green Peas,”  $\text{Ly}\alpha$  EW is observed to anti-correlate strongly with UV luminosity (Shapley et al. 2003; Stark et al. 2011; Hashimoto et al. 2013) SFR (Verhamme et al. 2008), HI covering fraction (Shibuya et al. 2014), and  $\text{Ly}\alpha$  escape fraction (Yang et al. 2017), all indicating  $\text{Ly}\alpha$  photons are significantly absorbed by neutral hydrogen gas and dust inside the ISM of massive, highly star-forming galaxies (e.g., Verhamme et al. 2008; Erb et al. 2014; Yang et al. 2017). At high redshift, the  $\text{Ly}\alpha$  EW distribution is usually parameterized as a conditional probability of  $p(W | M_{\text{UV}})$  (Dijkstra & Wyithe 2012; Treu et al. 2012), though dependence on UV spectral slope  $\beta$  has also been considered (Schenker et al. 2014).

We take the  $z \sim 6$  EW distribution from De Barros et al. (2017) and L. Pentericci et al. (2018, in preparation) from a Large Program with VLT/FORS2. This sample contains 127 objects, with UV magnitudes between  $-22.5 \lesssim M_{\text{UV}} \lesssim -17.5$ , of which 63% have  $\text{Ly}\alpha$  detections. We parameterize it as an exponential distribution plus a delta function:

$$p_6(W | M_{\text{UV}}) = \frac{A(M_{\text{UV}})}{W_c(M_{\text{UV}})} e^{-\frac{W}{W_c(M_{\text{UV}})}} H(W) + [1 - A(M_{\text{UV}})] \delta(W), \quad (4)$$

where  $A$  and  $W_c$  account for the fraction of non-emitters, and for the anti-correlation of EW with  $M_{\text{UV}}$ .  $H(W)$  is the Heaviside step function, and  $\delta(W)$  is a Dirac delta function.  $A$  implicitly includes contamination by low redshift interlopers in the photometric selection (the interloper fraction is  $\leq 29\%$  for this sample, assuming all non-detections were low redshift contaminants; De Barros et al. 2017)—that is, we do not distinguish between non-emitters at  $z \sim 6$  and low redshift contaminants when accounting for non-detections in fitting the parameters (see below). Within our framework, this means we assume a similarly small total interloper and non-emitter fraction at  $z \sim 7$ . Recent work by Vulcani et al. (2017) supports this assumption: they found comparably low contamination fractions at  $z \sim 6$  and  $z \sim 7$  in an evaluation of photometric selections.

To find these parameters, we divided the sample into three bins:  $M_{\text{UV}} \leq -21$ ;  $-21 < M_{\text{UV}} < -20$ ; and  $M_{\text{UV}} \geq -20$ . We used Equation (4) as a likelihood ( $p_6(W|A, W_c)$ ) and performed a Bayesian inference to infer  $A$  and  $W_c$  for each



**Figure 3.**  $z \sim 6$  Ly $\alpha$  equivalent width distributions for Lyman Break galaxies given by Equation (4). The dotted lines show the true distribution. For better comparison with the data, we show the PDFs convolved with a  $5 \text{ \AA}$  typical measurement error on  $W$  as solid lines. We plot the PDFs for two values of UV magnitude:  $M_{UV} = -18.5, -21.5$  (blue, orange). UV faint objects tend to have higher EW and a higher duty cycle of Ly $\alpha$  emission, whereas UV bright galaxies are less likely to emit Ly $\alpha$  and have lower EWs. We also plot the observed EW from De Barros et al. (2017) and L. Pentericci et al. (2018, in preparation) in UV bright (orange) and UV faint (blue) bins. In these histograms we plot all upper limits at EW = 0, though note we fully account for upper limits in fitting the EW distribution and the reionization inferences (see Equation (11)).

bin, similar to the methods of Oyarzún et al. (2017), using uniform priors with  $0 < A < 1$  and  $0 < W_c < 100$ . In the inference, we fully account for the non-detections of Ly $\alpha$  (using  $p_6(W < W_{\text{lim}} | A, W_c)$  as the likelihood given an EW limit  $W_{\text{lim}}$ , in the same way as described in more detail in Section 3). The uncertainties and EW limits calculated by De Barros et al. (2017) are obtained using simulations that fully account for incompleteness and wavelength sensitivity. We note that in this framework the EW limits are a conservative minimum that could be measured over the entire wavelength range (see also Section 4.2); future work could incorporate the full wavelength-dependent line flux sensitivity. To allow these parameters to smoothly vary with magnitude in the range of  $-21 < M_{UV} < -20$ , we use a hyperbolic tangent function to connect our inferred parameters, without extrapolating beyond the range of the data. We find  $A = 0.65 + 0.1 \tanh[3(M_{UV} + 20.75)]$  and  $W_c = 31 + 12 \tanh[4(M_{UV} + 20.25)] \text{ \AA}$  from fitting to the data.  $A$  and  $W_c$  vary smoothly with magnitude.

We choose this exponential parameterization of the data because it gives a good description of the data and is easy to treat analytically, and has previously been shown to be an excellent fit to Ly $\alpha$  EW PDFs (e.g., Oyarzún et al. 2017). We do not include uncertainties in these parameters, and we note the parameterization of  $p_6(W | M_{UV})$  is fairly arbitrary but does not qualitatively affect Ly $\alpha$  modeling during the EoR (Treu et al. 2012; Gronke et al. 2015b). Indeed we get the same results, within the uncertainties, if we use the  $p_6(W)$  truncated

Gaussian distribution from Treu et al. (2012) based on the sample presented in Stark et al. (2011).

Example PDFs given by Equation (4) are plotted in Figure 3 for two values of  $M_{UV}$ . We show both the intrinsic PDF and the distribution convolved with a  $5 \text{ \AA}$  typical measurement error which introduces a “bump” around  $W = 0$ , where the underlying distribution is a delta function. We also show histograms of the EW observations of De Barros et al. (2017) and L. Pentericci et al. (2018, in preparation) in two bins corresponding to UV bright and faint LBGs. As shown by, for example, Verhamme et al. (2008), Stark et al. (2010), and Oyarzún et al. (2017), Ly $\alpha$  EW strongly depends on UV magnitude.

## 2.2. IGM and CGM Ly $\alpha$ Radiative Transfer

A Ly $\alpha$  emission line is significantly attenuated by the CGM and IGM as its photons redshift into resonance with abundant neutral hydrogen along the line of sight. Effectively, for a Ly $\alpha$  line at  $z \gtrsim 6$ , all photons emitted blueward of the Ly $\alpha$  resonance ( $1216 \text{ \AA}$ ) are absorbed by the IGM, as even after reionization there is still a fraction of neutral hydrogen within H II regions (Gunn & Peterson 1965). Infalling overdense gas around halos can also increase the opacity of the IGM near the Ly $\alpha$  resonance and onto the red side of the Ly $\alpha$  line (Santos 2004; Dijkstra et al. 2007; Laursen et al. 2011).

For simplicity we assume all Ly $\alpha$  photons emitted below the circular velocity of a halo are absorbed in the CGM, and all redder photons are transmitted (Dijkstra et al. 2011; Laursen et al. 2011). This treatment of the CGM may be crude, but it enables us to investigate the relative difference between observations at  $z \sim 6$  and  $z \sim 7$ , assuming any evidence of a difference is driven by reionization. Future work could incorporate more complex modeling of CGM absorption (e.g., Kakiichi & Dijkstra 2017). Figure 1 shows example model Ly $\alpha$  emission lines, where the dotted black lines correspond to the intrinsic line profile after transmission through the ISM and the black solid lines correspond to the lineshape after resonant absorption in the CGM/IGM, which absorbs the flux blueward of  $v_{\text{circ}}$ .

During reionization, there is an additional opacity to Ly $\alpha$  caused by the presence of cosmic diffuse neutral hydrogen patches, which attenuate the damping wing of the Ly $\alpha$  line cross-section (Miralda-Escude 1998). The transmission of Ly $\alpha$  photons through the reionizing IGM is driven by the global fraction of neutral hydrogen,  $\bar{x}_{\text{H I}}(z)$ .

Thus the total opacity to Ly $\alpha$  due to neutral hydrogen within the IGM is given by

$$\tau_{\text{IGM}}(z, \nu) = \tau_{\text{D}}(z, \nu) + \tau_{\text{H II}}(z, \nu), \quad (5)$$

where  $\tau_{\text{D}}(z, \nu)$  is the damping wing optical depth which is only present during the EoR, and  $\tau_{\text{H II}}(z, \nu)$  is the optical depth due to resonant absorption within the CGM of galaxies (infalling gas) and any neutral hydrogen within the local H II region of a galaxy. For simplicity, we assume  $e^{-\tau_{\text{H II}}} = H(\nu - \nu_{\text{circ}})$  at both  $z \sim 6$  and  $z \sim 7$ .

In this model, we assume the universe is fully ionized at  $z \sim 6$ ; thus the damping wing opacity only becomes important at  $z > 6$ . This may not be exactly the case, but current constraints on  $\bar{x}_{\text{H I}}$  at  $z \sim 6$  suggest the neutral fraction is low



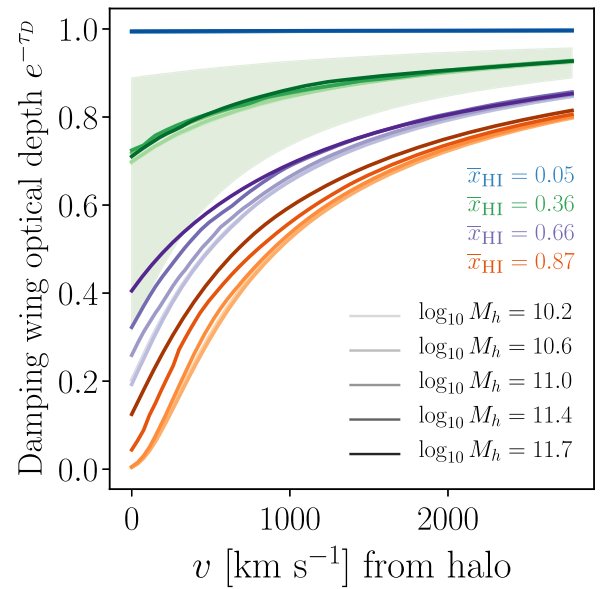
( $\bar{x}_{\text{HI}} < 0.1$ ; McGreer et al. 2014), so the reionization effect on Ly $\alpha$  emission will be small.

To obtain the damping wing optical depths,  $\tau_{\text{D}}(z = 7, v)$  requires a model of the IGM topology during reionization. While observation papers of Ly $\alpha$  emission with reionization inferences have used simple “patchy” or “smooth” IGM topologies (Treu et al. 2012, 2013; Pentericci et al. 2014; Tilvi et al. 2014), for this work, we consider realistic reionization topologies from state-of-the-art theoretical modeling. We obtain Ly $\alpha$  damping optical depths from the public Evolution of 21 cm Structure (EoS) suite of reionization simulations described by Mesinger et al. (2015, 2016).<sup>9</sup>

Due to the strong clustering of the first galaxies, spatial fluctuations in the IGM neutral fraction during reionization existed on scales of tens of Mpc. Accurately modeling these fluctuations and the growth of ionized H II bubbles in the IGM requires cosmological simulations at least 100 Mpc in size (Trac & Gnedin 2011; Mesinger et al. 2015). The EoS reionization simulations use 21CMFASTV2 (Sobacchi & Mesinger 2014), where inhomogeneous recombinations and ionizations in the IGM are treated at a sub-grid level on a density field in a box with sides 1.6 Gpc with a resolution  $1024^3$ . The simulations produce  $\bar{x}_{\text{HI}}$  maps at different redshifts and superimpose them on the  $z \sim 7$  halo field to produce cubes of the  $z \sim 7$  IGM for a range of neutral fractions. For the bulk of reionization, this is analogous to changing the ionization efficiency at a fixed redshift (e.g., McQuinn et al. 2007b; Mesinger & Furlanetto 2008b).

The timeline and topology of reionization is determined by the mass of galaxies that dominate reionization and the ionization efficiency,  $\zeta \propto f_{\text{esc}} \times f_{\star}$ , where  $f_{\text{esc}}$  is the fraction of ionizing photons which escape galaxies into the IGM, and  $f_{\star}$  is the stellar mass fraction in galaxies. As both of these parameters are expected to scale with halo mass in complementary ways (i.e., low mass halos host galaxies with a low stellar mass fraction and high escape fraction; e.g., Kimm et al. 2017; Trebitsch et al. 2017), over the relevant range of halo masses that host galaxies that dominate reionization ( $M_h \lesssim 10^{11} M_{\odot}$ ; e.g., Kimm et al. 2017),  $\zeta$  is assumed to be constant in the EoS simulations. The simulations use a free parameter that sets the minimum mass of halos capable of hosting star formation, and then adjust  $\zeta$  to produce a Thompson scattering optical depth to the CMB consistent with the Planck Collaboration (2015).

We use the fiducial “Faint Galaxies” run, which corresponds the primary drivers of reionization (being low mass star-forming galaxies) with an atomic cooling threshold of  $T_{\text{vir}} \gtrsim 10^4$  K, with  $\zeta = 20$ , producing IGM morphologies characterized by small HII regions. While the EoS simulations have another run, “Bright Galaxies,” where reionization is driven by more massive galaxies, producing larger HII regions, it has been shown that information about the reionization morphology is mostly smeared out when using galaxies spread in redshift ( $\Delta z \gtrsim 0.1$  bin; e.g., Sobacchi & Mesinger 2015, though with large spectroscopic samples,  $\Delta z \sim 0.01$ , the sensitivity does increase), as is the case for our sample (see Section 4.2), so we do not expect a significant change in our results if we were to use an alternative run. Indeed, Mason et al. (2018) uses both simulation runs but shows that the transmission of Ly $\alpha$  from galaxies



**Figure 4.** Median Ly $\alpha$  IGM damping wing optical depths due to cosmic H I patches during reionization as a function of velocity offset from the center of the source halos. We plot optical depths for five different mass halos (indicated by tone of the line, where darkest lines are the highest mass halos) and for four volume-averaged neutral fractions  $\bar{x}_{\text{HI}}$  (indicated by color). We plot the median optical depth for each halo from the thousands of simulated sightlines. For  $\bar{x}_{\text{HI}} = 0.36$ , we plot the  $1\sigma$  confidence region for the optical depths from all the sightlines to the  $\log_{10} M_h = 10.2$  halos as a shaded area, showing the large variation across sightlines.

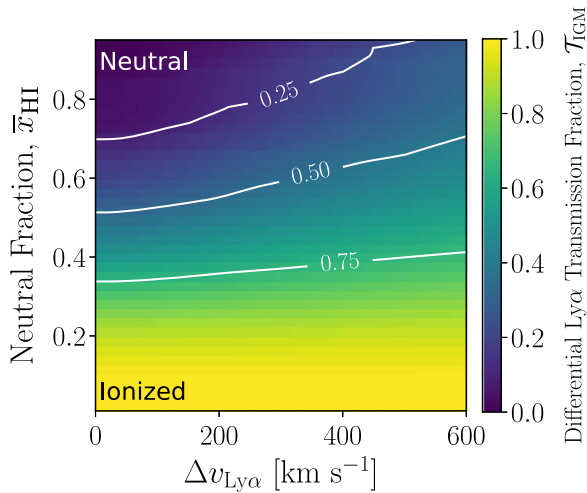
$M_h \sim 10^{10}$ – $10^{12}$  is relatively independent of the reionization morphology. Similarly, Greig et al. (2017) show QSO damping wing effects are not particularly sensitive to the reionization morphology.

Halos are located in the same density field as the IGM simulation. We ignore absorption from Damped Ly $\alpha$  Absorbers (DLAs) inside the cosmic H II regions (Bolton & Haehnelt 2013), which has been shown to have a minor impact on the Ly $\alpha$  fraction when self-shielding is calculated more accurately (Mesinger et al. 2015). We refer the reader to Mesinger et al. (2016) for more details of the simulation. For this work we focus on  $z \sim 7$ , where large samples of LBGs have spectroscopic follow-up (Pentericci et al. 2014; Schmidt et al. 2016), but it is easy to extend the work to any other redshift.

We take thousands of sightlines emanating from halos with masses  $\sim 10^{10}$ – $10^{12} M_{\odot}$  (comparable to typical  $z \gtrsim 5$  halo masses for  $-22 \lesssim M_{\text{UV}} \lesssim -18$  galaxies; Barone-Nugent et al. 2014; Harikane et al. 2016, 2017) and compute the damping wing optical depth,  $\tau_{\text{D}}$ , for Ly $\alpha$  emission as a function of velocity offset from the systemic redshift of the source halos in boxes with a range of global neutral fractions. Median values of  $\exp[-\tau_{\text{D}}]$  along  $\sim 50$  (to the rarest high mass halos) to  $\gtrsim 4000$  (to typical  $10^{10.5} M_{\odot}$  halos) sightlines are plotted in Figure 4 for a range of halo masses and  $\bar{x}_{\text{HI}}$ . The optical depths are smooth functions of velocity and clearly damp Ly $\alpha$  more effectively for higher  $\bar{x}_{\text{HI}}$ . In general, higher mass halos have lower optical depths to Ly $\alpha$ , as their large bias means they are more likely to live in the centers of large H II regions, relatively more distant from the cosmic H I patches that produce the damping wing absorption during the EoR.

For a given sightline, the final fraction of Ly $\alpha$  photons emitted by a galaxy in a halo with mass  $M_h$ , which are

<sup>9</sup> <http://homepage.sns.it/mesinger/EOS.html>



**Figure 5.** Median fraction of Ly $\alpha$  photons transmitted through the IGM,  $\mathcal{T}_{\text{IGM}}$ , as a function of  $\bar{x}_{\text{HI}}$  and  $\Delta v$  computed with Equation (6) from  $\sim 5000$  sightlines to halos with mass  $10^{10} M_{\odot}$ , assuming  $\mathcal{T}_{\text{IGM},6} = 1$ . Contours show transmission fractions of 25%, 50%, and 75%. In a predominantly neutral universe, Ly $\alpha$  photons have a higher probability of escape through predominately ionized IGM and if emitted at high velocity offsets from their originating galaxies.

transmitted through the IGM,  $\mathcal{T}_{\text{IGM}}$ , is given by

$$\mathcal{T}_{\text{IGM}}(\bar{x}_{\text{HI}}, M_h, \Delta v) = \int_0^{\infty} dv J_{\alpha}(\Delta v, M_h, v) e^{-\tau_{\text{IGM}}(\bar{x}_{\text{HI}}, M_h, v)}, \quad (6)$$

where  $\Delta v$  is the velocity offset of the Ly $\alpha$  line center from the systemic redshift of the source galaxy (which depends on the galaxy’s ISM, as described in Section 2.1) and  $J_{\alpha}(\Delta v, M_h, v)$  is the line profile of Ly $\alpha$  photons escaping from the galaxy as function of velocity  $v$ .

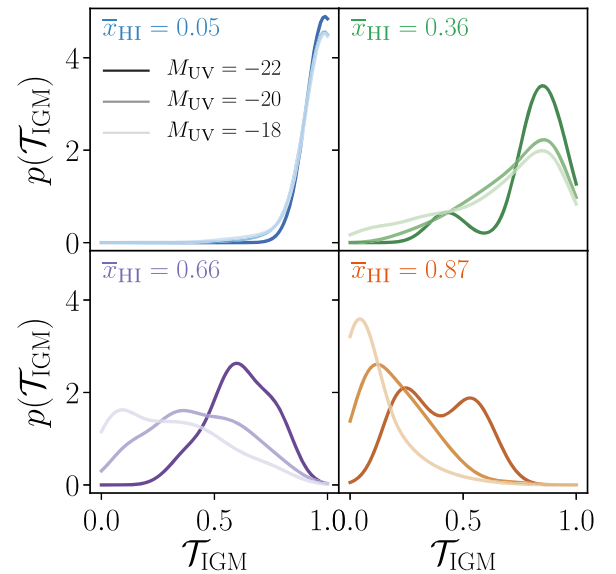
As any photons emitted bluer than the halo circular velocity will be resonantly absorbed by intervening neutral hydrogen (Gunn & Peterson 1965; Dijkstra et al. 2007; Zheng et al. 2010; Laursen et al. 2011; Schroeder et al. 2013),  $J_{\alpha}$  is described as

$$J_{\alpha}(\Delta v, M_h, v) \propto \begin{cases} \frac{1}{\sqrt{2\pi}\sigma_{\alpha}} e^{-\frac{(v-\Delta v)^2}{2\sigma_{\alpha}^2}} & \text{if } v \geq v_{\text{circ}}(M_h) \\ 0 & \text{otherwise} \end{cases}. \quad (7)$$

If  $J_{\alpha}$  is normalized, then  $\mathcal{T}_{\text{IGM},6} = 1$ , as we assume  $\tau_{\text{D}} = 0$  at  $z \sim 6$ . Compared to the intrinsic emitted line,  $\mathcal{T}_{\text{IGM},6}$  can be very low (Dijkstra et al. 2007; Zheng et al. 2010; Laursen et al. 2011). For ease of notation we refer to the differential transmission at  $z \sim 7$ ,  $\mathcal{T}_{\text{IGM},7}/\mathcal{T}_{\text{IGM},6}$ , as  $\mathcal{T}_{\text{IGM}}$ .

Example intrinsic and transmitted emission lines are plotted in Figure 1. Sightline median values of  $\mathcal{T}_{\text{IGM}}(\bar{x}_{\text{HI}}, \Delta v)$  at fixed halo mass are plotted in Figure 5. As expected, as the neutral fraction increases, the transmission fraction of Ly $\alpha$  decreases smoothly. While at low neutral fractions the velocity offset of Ly $\alpha$  has little impact, in a predominantly neutral universe, ( $\bar{x}_{\text{HI}} \gtrsim 0.6$ ) lines are more easily transmitted if they were emitted at high velocity offset.

In Figure 6 we plot probability distribution functions for  $\mathcal{T}_{\text{IGM}}$  for three different values of  $M_{\text{UV}}$ , where we have transformed from halo mass to  $M_{\text{UV}}$  using the Mason et al. (2015) LF model as above and drawn  $\Delta v$  values for halos using the distribution in Equation (2). The transmission distributions evolve smoothly with neutral fraction and UV magnitude.



**Figure 6.** Distributions of differential Ly $\alpha$  transmission fractions  $\mathcal{T}_{\text{IGM}}$  at  $z \sim 7$  for simulated galaxies of different UV luminosities (UV bright = darkest lines), for a range of IGM neutral fractions  $\bar{x}_{\text{HI}}$ . As described in Section 2.2, this is the ratio of Ly $\alpha$  transmission at  $z \sim 7$  to that at  $z \sim 6$ , where there is already significant absorption within the ionized IGM (Dijkstra et al. 2007; Zheng et al. 2010; Laursen et al. 2011). The transmission fractions evolve smoothly with the neutral fraction, though the evolution is more gradual for UV bright galaxies.

Transmission of Ly $\alpha$  evolves more slowly for the brightest galaxies, due to a combination of their increased velocity offsets and their locations in the most overdense regions, far from the cosmic HI regions, which cause the damping wing absorption.

Galaxies in high mass halos ( $M_h > 10^{12} M_{\odot}$ , corresponding to approximately  $M_{\text{UV}} < -22$ ) require special attention. First, they are rare, and lines of sight to such objects in the simulations are not well-sampled, leading to large statistical errors. Second, the correlation between  $M_{\text{UV}}$  and  $M_h$  is particularly uncertain in this regime. Third, such bright galaxies have been observed to buck the trend in the declining Ly $\alpha$  emission fraction at  $z \gtrsim 7$  at  $z > 7.5$  (Curtis-Lake et al. 2012; Stark et al. 2017). For these reasons, they require special attention, especially because they are prime targets for detailed spectroscopic follow-up. Since they are intrinsically rare, they would contribute negligibly to the analysis presented in this paper. Thus we leave their analysis for future work (Mason et al. 2018) and exclude them from the sample considered here.

### 3. Bayesian Inference

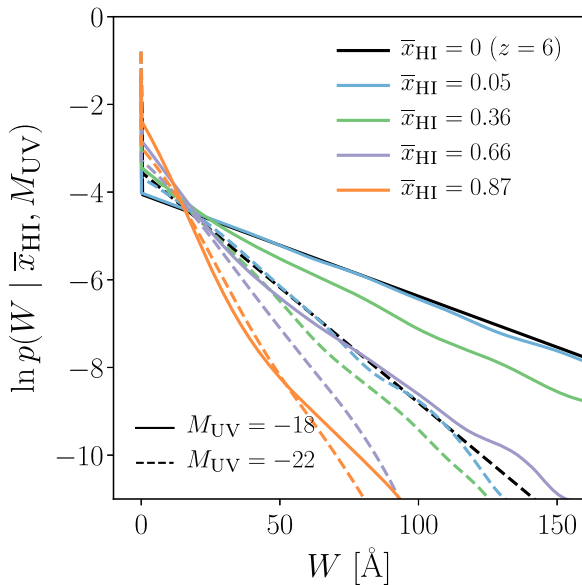
Bayes’ theorem enables us to infer the posterior distribution of model parameters of interest,  $\theta$ , given our observed data  $Y$  from the likelihood of obtaining the data, given our model and our prior information of the model parameters. The posterior probability of  $\theta$  is written as

$$p(\theta | Y) = \frac{p(Y | \theta)p(\theta)}{p(Y)}, \quad (8)$$

where  $p(Y | \theta)$  is the likelihood function,  $p(\theta)$  is the prior, and  $p(Y)$  is the Bayesian evidence that normalizes the posterior.

We want to obtain the posterior distribution of the volume-averaged fraction of neutral hydrogen,  $\bar{x}_{\text{HI}}$ , a global IGM property, given the observed data: measurements of Ly $\alpha$





**Figure 7.** Simulated observed distribution of Ly $\alpha$  equivalent widths (the likelihoods for our model) for a range of neutral fractions (colors), for faint (solid line) and bright (dashed line) UV magnitudes. The intrinsic distributions (Equation (4)) are shown as black lines. The EW distribution evolves significantly for the UV faint galaxies with increasing  $\bar{x}_{\text{HI}}$ , while the distribution for UV bright galaxies evolves more slowly.

equivalent widths  $W$  and galaxy rest-frame UV absolute magnitudes  $M_{\text{UV}}$ . As described in Section 2, we model both IGM and ISM effects on Ly $\alpha$  transmission and produce forward models of the observed Ly $\alpha$  equivalent widths for galaxies of a given UV magnitude.

Using Bayes’ theorem we can write the posterior probability for  $\bar{x}_{\text{HI}}$  inferred from one observation in the absence of noise as

$$p(\bar{x}_{\text{HI}} | W, M_{\text{UV}}) \propto p(W | \bar{x}_{\text{HI}}, M_{\text{UV}})p(\bar{x}_{\text{HI}}), \quad (9)$$

where  $p(W | \bar{x}_{\text{HI}}, M_{\text{UV}})$  is the likelihood of observing a Ly $\alpha$  equivalent width, given our forward model of the ISM and IGM, and  $p(\bar{x}_{\text{HI}})$  is the prior on the neutral fraction, which we assume is uniform between 0 and 1.

Usually the likelihood function is obtained from a model with an analytic form (e.g., a normal distribution); however, due to including simulated IGM cubes, our model is complex and does not have an analytic parameterization. We therefore generate the likelihood by sampling  $10^6$  realizations of galaxies in our model at a given  $(\bar{x}_{\text{HI}}, M_{\text{UV}})$  and then perform a Kernel Density Estimation (Rosenblatt 1956; Parzen 1962) to fit a smooth probability density function to the sampled distribution. Examples of the likelihood function are shown in Figure 7. Generation of the likelihoods is described in more detail in Section 3.1.

In reality, our observations will always have measurement uncertainties, and some observations can only place an upper limit on a measurement, given a noise level. When we include noise, our likelihood for measuring an equivalent width  $W_i$  with Gaussian noise level  $\sigma_i$  becomes

$$p(W_i | \bar{x}_{\text{HI}}, M_{\text{UV}}) = \int_0^\infty dW \frac{e^{-\frac{(W-W_i)^2}{2\sigma_i^2}}}{\sqrt{2\pi}\sigma_i} p(W | \bar{x}_{\text{HI}}, M_{\text{UV}}), \quad (10)$$

and the likelihood for upper limits,  $W_i < \mathcal{W}$ , is given by

$$\begin{aligned} p(W_i < \mathcal{W} | \bar{x}_{\text{HI}}, M_{\text{UV}}) &= \int_{-\infty}^{\mathcal{W}} dW p(W_i | \bar{x}_{\text{HI}}, M_{\text{UV}}) \\ &= \int_0^\infty dW \frac{1}{2} \operatorname{erfc}\left(\frac{W - \mathcal{W}}{\sqrt{2}\sigma_i}\right), \quad (11) \\ &\quad \times p(W | \bar{x}_{\text{HI}}, M_{\text{UV}}) \end{aligned}$$

where  $\operatorname{erfc}(x)$  is the complementary error function for  $x$ .

In this work we consider samples with good redshift completeness (i.e., the probability of a Ly $\alpha$  line falling within the observable range is close to one; see Section 4.2). Thus this inference framework does not include information about redshift in the likelihood; this is left for future work.

We can combine the inference from a set of independent observations (i.e., individual galaxies) by simply multiplying the posteriors:

$$p(\bar{x}_{\text{HI}} | \{W, M_{\text{UV}}\}) \propto \prod_i^{N_{\text{gals}}} p(W_i | \bar{x}_{\text{HI}}, M_{\text{UV},i}) p(\bar{x}_{\text{HI}}). \quad (12)$$

### 3.1. Generating the Likelihood

Our observed data are a set of Ly $\alpha$  equivalent widths (and limits) and absolute magnitudes from galaxies at a given redshift:  $\{W, M_{\text{UV}}\}$ . Due to the complexity of the IGM topology, there is no simple analytic model to express the likelihood of obtaining these data, given a neutral fraction  $\bar{x}_{\text{HI}}$ . Thus we use our model to generate large samples of mock observations that provide a non-analytic likelihood.

We take IGM simulations with global neutral fractions  $0.01 \leq \bar{x}_{\text{HI}} \leq 0.95$  ( $\Delta \bar{x}_{\text{HI}} \sim 0.02$ ) and a population of halos with masses  $10^{10} \lesssim M_h [M_\odot] \lesssim 10^{12}$  with  $\Delta \log M_h \sim 0.1$ . This mass range corresponds to UV magnitudes of  $-16 \gtrsim M_{\text{UV}} \gtrsim -22$  at  $z \sim 7$  (Mason et al. 2015). The likelihood is computed in the following way:

1. Obtain the Ly $\alpha$  damping wing optical depths (see Section 2.2) along thousands of different sightlines to individual halos in each simulation, to account for the inhomogeneous nature of reionization.
2. For a grid of UV magnitudes  $-22 \leq M_{\text{UV}} \leq -17$ , we nearest-neighbor match the simulation halo masses with UV magnitudes at  $z \sim 7$ , given by the relation in Mason et al. (2015), which is consistent with  $M_{\text{UV}}-M_h$  measurements from clustering at  $z \sim 7$  (Barone-Nugent et al. 2014; Harikane et al. 2016, 2017). We do not add scatter to this matching, but note the halo mass step in the simulations ( $\sim 0.13$  dex) is not dissimilar to the scatter in the inferred  $M_{\text{UV}}-M_h$  relation for galaxies around  $M_{\text{UV}}^*$  (e.g., 0.3 dex; Finkelstein et al. 2015b), so some  $M_{\text{UV}}-M_h$  scatter is included. Furthermore we note the optical depth scatter between sightlines for a given halo mass is far greater than the scatter among sightlines between halos of different masses (compare lines and shaded region in Figure 4); thus the  $M_{\text{UV}}-M_h$  scatter is sub-dominant.
3. Populate these model galaxies with Ly $\alpha$  line velocity offsets from the distribution  $p(\Delta v | M_h)$ , as described by Equation (2), including the scatter  $\sigma_v$ , and the Ly $\alpha$  equivalent widths for an ionized universe (we which assumed to be the same as at  $z \sim 6$ ),  $p_6(W_{\text{em}} | M_{\text{UV}})$

described in Section 2.1, creating  $10^6$  realizations of a galaxy in each halo.

4. We compute the differential Ly $\alpha$  transmission fraction,  $\mathcal{T}_{\text{IGM}}$ , with Equation (6) along sightlines through the IGM to every model galaxy and the observed equivalent width, where  $W_{\text{obs}} = \mathcal{T}_{\text{IGM}} \times W_{\text{em}}$ .
5. The distributions of model observed  $W_{\text{obs}}$  at fixed  $(\bar{x}_{\text{HI}}, M_{\text{UV}})$  are described by the form

$$p(W | \bar{x}_{\text{HI}}, M_{\text{UV}}) = A(M_{\text{UV}})f(W, \bar{x}_{\text{HI}})H(W) + [1 - A(M_{\text{UV}})]\delta(W), \quad (13)$$

where  $f(W, \bar{x}_{\text{HI}})$  describes the evolution of the equivalent width distribution as the neutral fraction evolves and is fitted with a Gaussian Kernel Density Estimator (Rosenblatt 1956; Parzen 1962), and  $A(M_{\text{UV}})$  denotes the fraction of non-emitters and contaminants, as described in Equation (4), that does not change as the neutral fraction increases ( $\mathcal{T}_{\text{IGM}} \neq 0$  exactly).

These distributions  $p(W | \bar{x}_{\text{HI}}, M_{\text{UV}})$  are the likelihoods for the observed data. Some examples are plotted in Figure 7. For increasing neutral fraction, the EW distribution becomes steeper, as more Ly $\alpha$  is damped by cosmic neutral patches. The evolution of  $p(W | \bar{x}_{\text{HI}}, M_{\text{UV}})$  is slower for more UV bright (more massive) galaxies because the transmission functions evolve more slowly with increasing neutral fraction (see Section 2.2 and Figure 6).

We chose to marginalize out  $\Delta\nu$  at this stage to ease computation by reducing a degree of freedom, but it is possible to produce the likelihood conditional on  $\Delta\nu$ :  $p(W | \bar{x}_{\text{HI}}, M_{\text{UV}}, \Delta\nu)$ . It is then possible to infer  $\Delta\nu$  for an individual observed galaxy, or if  $\Delta\nu$  is already known, recover a narrower posterior on the neutral fraction.

## 4. Results

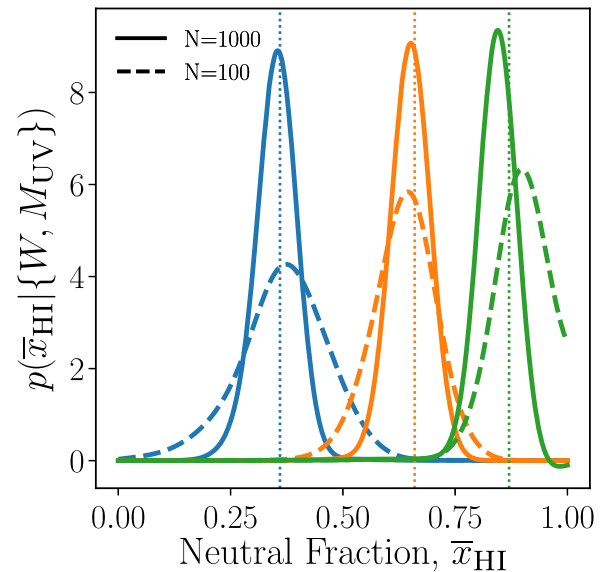
In this section we describe the key results and predictions from our model. In Section 4.1 we show our method can accurately recover the neutral fraction for simulated data sets. We perform inference on current data from Pentericci et al. (2014) in Section 4.2. In Section 4.3 we make predictions for future surveys with *JWST*.

### 4.1. Large Samples of Galaxies Can Accurately Constrain the Neutral Fraction

To test our inference framework we perform simulated surveys of LBG follow-up. We draw a realistic sample of LBGs at  $z \sim 7$  from the Mason et al. (2015) UV luminosity function model (which is consistent with all observations, including new deep data from the Hubble Frontier Fields at  $z \gtrsim 7$ ; e.g., Atek et al. 2015b; McLeod et al. 2016; Bouwens et al. 2017a; Livermore et al. 2017). We populate these galaxies with an EW given by our simulated  $p(W | \bar{x}_{\text{HI}}, M_{\text{UV}})$  (see Section 3.1) for several test values of the neutral fraction.

We assume an apparent magnitude limit of  $m_{\text{AB}} = 28.5$ , corresponding to  $M_{\text{UV}} \sim -18.5$  and a  $5\sigma$  flux limit of  $10^{-18}$  erg s $^{-1}$  cm $^{-2}$ . We draw samples of 100 and 1000 total galaxies, and perform the inference on the full samples, including upper limits.

In Figure 8 we plot the resulting posterior distributions for  $\bar{x}_{\text{HI}}$ . With large samples we can clearly recover the input neutral fraction well. With small samples the posterior



**Figure 8.** Posterior distributions for  $\bar{x}_{\text{HI}}$  from simulated samples of Ly $\alpha$  detections from 1000 (solid lines) and 100 (dashed lines) galaxies, for a simulation input value of  $\bar{x}_{\text{HI}} = [0.36, 0.66, 0.87]$  (blue, orange, green—the input value is shown by the vertical dotted line). With large samples, the input neutral fraction is recovered well. With smaller samples, the posterior is wider, but includes the true value within  $1\sigma$  uncertainty.

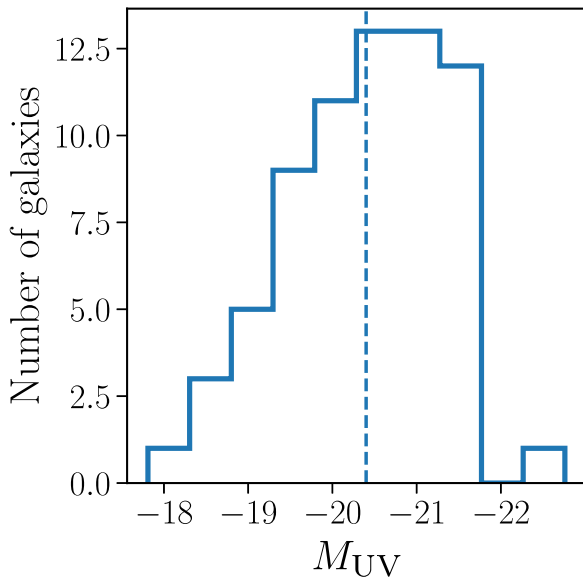
distribution is broader as we sample less of the likelihood, but the posteriors still include the input value within  $1\sigma$ .

### 4.2. Inference from Current Data

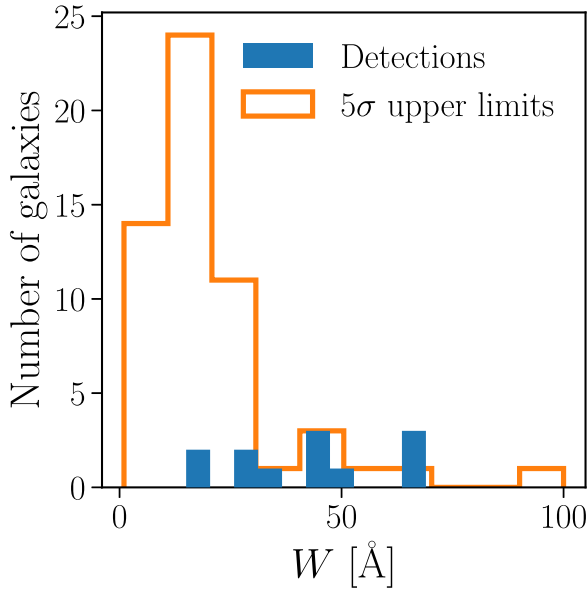
We use the inference framework described above to infer the neutral fraction from current observations. We take the largest published sample of LBGs at  $z \sim 7$  with spectroscopic follow-up to date, presented in Pentericci et al. (2014). These data comprise 68 galaxies spanning UV magnitudes  $-22.75 \lesssim M_{\text{UV}} \lesssim -17.8$  and include 10 intrinsically faint objects gravitationally lensed behind the Bullet Cluster (Bradač et al. 2012) as well as observations in deep *HST* legacy fields (Fontana et al. 2010; Vanzella et al. 2011; Ono et al. 2012; Schenker et al. 2012).

In total, the sample comprises eight independent lines of sight with field areas  $\sim 50$ – $100$  arcmin $^2$  each. The detections are spread over these fields. Pentericci et al. (2014) quantified the cosmic variance in this sample is very low ( $\sim 6\%$  uncertainty in the optical depth to Ly $\alpha$ ; see also Trenti & Stiavelli 2008). Of the 68 LBGs, 12 Ly $\alpha$  lines were spectroscopically confirmed. In Figure 9 we plot the UV magnitude, and in Figure 10 the EW distributions for this sample. As for the De Barros et al. (2017)  $z \sim 6$  sample, the EW limits are obtained by inserting simulated lines of varying flux, FWHM, and redshift into raw data, and then trying to recover them. A conservative minimum flux that could be measured over the entire wavelength range is used for the EW limit. Our framework utilizes the fact that the non-detections must be fainter than this conservative limit: fainter lines could be observed, for example, in regions free of sky lines. Future work could include the wavelength-dependent line flux sensitivity in the likelihood for non-detections (Equation (11)).

The majority of targets were  $z_{850}$ -dropouts selected primarily using a color criteria technique, described in detail in Grazian et al. (2012), to find targets with a high probability of having redshifts  $6.5 \lesssim z \lesssim 7.5$ . The median redshift for this selection

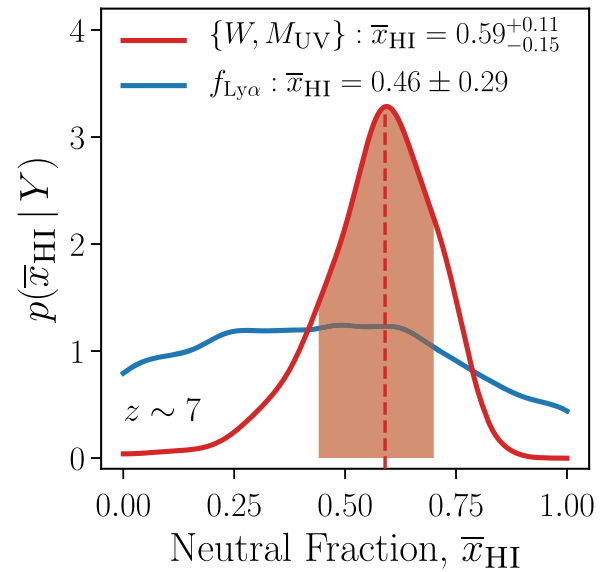


**Figure 9.** UV magnitude distributions for the  $z \sim 7$  sample used for the inference. We plot the median UV magnitude of the sample as a dashed line ( $M_{UV} = -20.4$ ).



**Figure 10.** EW distribution for the  $z \sim 7$  sample used for the inference. We show both the Ly $\alpha$  EW measurements (filled blue) and  $5\sigma$  upper limits for the non-detections (orange line).

function was  $z = 6.9$  (see Figure 1 in Grazian et al. 2012). For literature targets not directly observed by the Pentericci et al. (2014) group, but included in the sample, they only included  $z$ -dropouts with colors consistent with the color selection criteria. As noted by Pentericci et al. (2014), the probability of galaxies being outside of the observable range for their setup ( $z \sim 7.3$ ) is negligible, except for the 10 objects in the Bullet Cluster (Bradač et al. 2012), where  $\sim 48\%$  of objects could be above this redshift due to the broad J filter used for selection of that sample (Hall et al. 2012). To test the impact of these few objects potentially being at higher redshifts, we ran the inference excluding the Bullet Cluster and found it does not significantly impact the results.



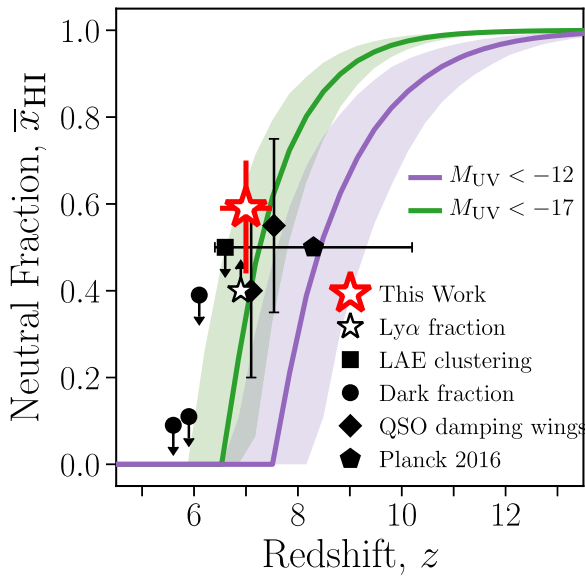
**Figure 11.** Posterior distribution for  $\bar{x}_{\text{HI}}$  from the data set of 68 galaxies at  $z \sim 7$  (including 12 with detected Ly $\alpha$  emission) from Pentericci et al. (2014). In red we plot the posterior distribution obtained from the full sample of  $\{W, M_{UV}\}$  measurements as described in Section 3, and infer a neutral fraction of  $\bar{x}_{\text{HI}} = 0.59^{+0.11}_{-0.15}$  (16%–84%). The dashed line shows the median value, and the shaded region shows the (16<sup>th</sup> and 84<sup>th</sup>) percentile bounds. For comparison, in blue we plot the posterior for  $\bar{x}_{\text{HI}}$  obtained if we used only the fraction of galaxies emitting Ly $\alpha$  with  $W > 25 \text{ \AA}$ ,  $f_{\text{Ly}\alpha}$ . In this case, we infer  $\bar{x}_{\text{HI}} = 0.46 \pm 0.29$ . Using the full distribution of EW provides much more information about the evolving IGM compared to  $f_{\text{Ly}\alpha}$  and allows for tighter constraints on the neutral fraction.

For each galaxy in this sample, we compute the likelihoods for obtaining the observed equivalent width or upper limit using Equations (10) and (11) for every value of the neutral fraction in our simulations. We exclude the brightest objects ( $M_{UV} < -22$ , 1 object) due to the insufficient sampling of very massive halos in the simulations (see Section 2.2) and the uncertainty in their intrinsic EW evolution (Stark et al. 2017), but note this does not affect the inferred neutral fraction for our sample because the UV bright objects are so rare. We use an MCMC sampler (Foreman-Mackey et al. 2013) to infer the posterior distribution of  $\bar{x}_{\text{HI}}$  from these data, which is shown in Figure 11. We infer a neutral fraction of  $\bar{x}_{\text{HI}} = 0.59^{+0.11}_{-0.15}$  (16–84%).

This constraint is much tighter than previous measurements of the neutral fraction from Ly $\alpha$  observations (e.g., Pentericci et al. 2014; Mesinger et al. 2015) because we use the full distribution of equivalent widths,  $p(W|M_{UV})$ , in our inference. Previous analyses used only the fraction of galaxies emitting Ly $\alpha$  with  $W > 25 \text{ \AA}$ ,  $f_{\text{Ly}\alpha}$ , to constrain the neutral fraction. In Figure 11 we also plot the posterior distribution obtained if we had used only  $f_{\text{Ly}\alpha}$ , that is, the posterior is  $p(\bar{x}_{\text{HI}}|f_{\text{Ly}\alpha})$ , where we compare the simulation  $f_{\text{Ly}\alpha}(\bar{x}_{\text{HI}})$  derived from Equation (13) with the fraction obtained in Pentericci et al. (2014):  $f_{\text{Ly}\alpha} = 0.29^{+0.20}_{-0.15}$  (for their faint sample,  $20.25 < M_{UV} < 18.75$ ). With just the Ly $\alpha$  fraction, we infer a neutral fraction of  $\bar{x}_{\text{HI}} = 0.46 \pm 0.29$ . Clearly, using the full distribution of EW enables us to constrain the neutral fraction much more accurately, and now that large samples of LBGs with spectroscopic follow-up are available, it should become the statistic of choice for Ly $\alpha$  reionization inferences.

Where does this constraint sit in our consensus picture of reionization? In Figure 12 we plot constraints derived from observations of Ly $\alpha$  emission from galaxies (Mesinger et al. 2015); the clustering of Ly $\alpha$  emitting galaxies (Ouchi





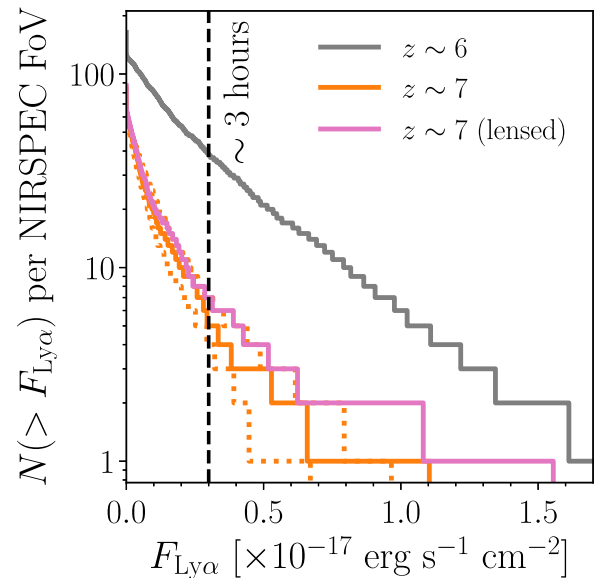
**Figure 12.** Fraction of neutral hydrogen as a function of redshift. Our new constraint is plotted as a red open star. We plot constraints derived from observations of previous estimates from the fraction of LBGs emitting Ly $\alpha$  (open black star; Mesinger et al. 2015); the clustering of Ly $\alpha$  emitting galaxies (square; Ouchi et al. 2010; Sobacchi & Mesinger 2015); Ly $\alpha$  and Ly $\beta$  forest dark fraction (circle; McGreer et al. 2014); and QSO damping wings (diamond; Bañados et al. 2017; Greig & Mesinger 2017b). We offset the constraints at  $z \sim 7$  (QSO ULASJ1120+0641 damping wing, Greig et al. 2017, Ly $\alpha$  fraction and our new constraint) by  $\delta z = 0.1$  for clarity. We also plot the Planck Collaboration (2016) redshift range of instantaneous reionization (black hatched region). We show as shaded regions the reionization history from integrating the Mason et al. (2015) UV luminosity function to two magnitude limits of  $M_{UV} = -17$  (green) and  $M_{UV} = -12$  (purple), drawing from uniform distributions for the ionizing photon escape fraction 10%–30% and clumping factor  $C = 1 - 6$ , and log-normal distribution for the ionizing efficiency  $\xi_{ion}$  with mean 25.2 and standard deviation 0.15 dex. There are many uncertainties in obtaining the reionization history from luminosity functions, so these should not be taken as real constraints on the neutral fraction, but given that galaxies fainter than  $M_{UV} = -17$  likely exist (e.g., Kistler et al. 2009; Bouwens et al. 2017b; Livermore et al. 2017; Weisz & Boylan-Kolchin 2017), our result suggests high escape fractions may not be necessary for reionization.

et al. 2010; Sobacchi & Mesinger 2015); Ly $\alpha$  and Ly $\beta$  forest dark fraction (McGreer et al. 2014); and QSO ULASJ1120+0641 damping wings (Greig & Mesinger 2017b). We also plot the neutral hydrogen fraction as a function of redshift, using the Mason et al. (2015) UV luminosity function model, assuming galaxies are the source of ionizing photons and using two limiting magnitudes for the galaxy population:  $M_{UV} < -17$  (currently detectable galaxies) and  $M_{UV} < -12$  (ultra-faint undetected galaxies). The uncertainties in the Mason et al. (2015) reionization histories come from the range of possible reionization parameters (e.g., ionizing photon escape fraction, IGM clumping factor, number of ionizing photons per UV photon).

Our constraint is consistent within  $1\sigma$  with the other constraints at  $z \sim 7$ , providing more strong evidence that reionization is ongoing at  $z \sim 7$ . Our constraint lies  $\Delta \bar{x}_{HI} \sim 0.2$  higher than the constraint from the  $z = 7.1$  QSO ULASJ1120+0641 damping wings (Greig et al. 2017), but is still consistent within the uncertainties.

#### 4.3. Predictions for JWST

JWST will be uniquely equipped to observe Ly $\alpha$  and rest-frame optical emission lines into Cosmic Dawn, with extremely



**Figure 13.** Predicted cumulative number counts of LAEs with JWST NIRSpect at  $z \sim 6$  (gray) and  $z \sim 7$  (orange) using our recovered neutral fraction  $\bar{x}_{HI} = 0.59_{-0.15}^{+0.11}$ . Galaxies are drawn from the Mason et al. (2015) UV luminosity function model and populated with equivalent widths via  $p(W | M_{UV}, \bar{x}_{HI})$ , the likelihood described in Section 3.1. The number counts obtained within the (16%–84%) regions on  $\bar{x}_{HI}$  are shown as dotted orange lines. We also show the cumulative number counts for a gravitationally lensed field where we assume a uniform magnification factor of  $\mu = 2$  (pink line), which would reveal more emission lines. We obtain the Ly $\alpha$  fluxes using Equation (14). The dashed black line shows the flux limit for a  $\sim 3$  hr exposure at  $R = 1000$ , with JWST NIRSpect F070LP/G140M at 1–1.5  $\mu\text{m}$ , calculated with the JWST ETC (<https://jwst.eti.stsci.edu>).

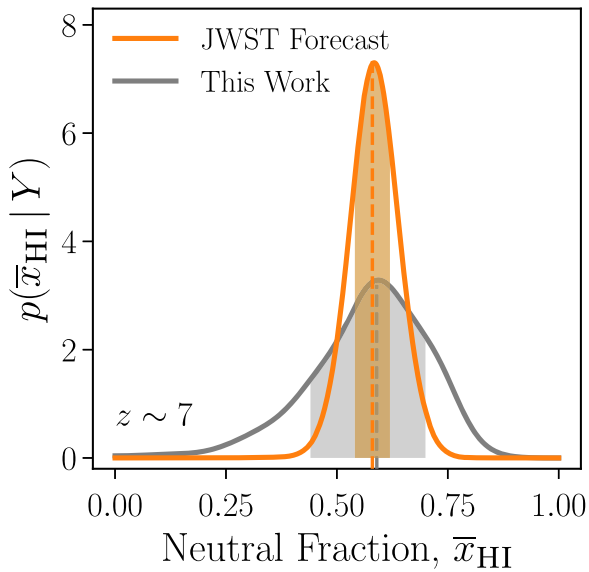
sensitive spectrometers NIRSpect and NIRISS covering 1–5  $\mu\text{m}$  in a large field of view (Gardner et al. 2006). This will enable direct measurement of the Ly $\alpha$   $\Delta v$  and detailed studies of the ISM properties of galaxies during reionization.

Using our inferred value of  $\bar{x}_{HI} = 0.59_{-0.15}^{+0.11}$  for the neutral fraction at  $z \sim 7$ , we predict the number of Ly $\alpha$  emitters detectable in one NIRSpect pointing ( $\sim 9$  sq arcmins) by drawing galaxies from the Mason et al. (2015) UV luminosity function and populating them with EW given by our simulated  $p(W | \bar{x}_{HI}, M_{UV})$ . We transform Ly $\alpha$  equivalent width  $W$  to flux using the relation

$$f(W, m, z) = W f_0 10^{-0.4m_{UV}} \frac{c}{\lambda_\alpha^2 (1+z)} \left( \frac{\lambda_{UV}}{\lambda_\alpha} \right)^{-\beta-2}, \quad (14)$$

where  $f_0 = 3.631 \times 10^{-20} \text{ erg s}^{-1} \text{ Hz}^{-1} \text{ cm}^{-2}$ ,  $m_{UV}$  is the apparent magnitude of the UV continuum,  $c$  is the speed of light,  $\lambda_\alpha$  is the rest-frame wavelength of Ly $\alpha$ ,  $\lambda_{UV}$  is the rest-frame wavelength of the UV continuum (usually 1500 Å), and  $\beta$  is the UV slope. For simplicity we assume  $\beta = -2$ , consistent with observations of  $z \sim 7$  objects (e.g., Bouwens et al. 2012), though very UV faint galaxies likely have steeper slopes due to extremely low metallicities (Vanzella et al. 2016).

We plot the predicted number counts in Figure 13, where we assume a  $5\sigma$  UV continuum flux limit of  $m_{AB} > 29$  ( $M_{UV} \sim -18$ , corresponding to  $\sim 1$  hr integration in JWST NIRCAM). We predict a 3 hr exposure in one pointing ( $\sim 9$  sq arcmins) with JWST NIRSpect will detect  $\sim 6 \pm 3$   $z \sim 7$  Ly $\alpha$  lines with a  $5\sigma$  flux limit of  $\sim 3 \times 10^{-18} \text{ erg s}^{-1} \text{ cm}^{-2}$  (calculated using the JWST ETC), from a total of  $\sim 80$  LBG



**Figure 14.** Posterior distribution of  $\bar{x}_{\text{HI}}$  for a simulated 10 pointing *JWST* NIRSpec survey (orange), which is able to tightly constrain the IGM neutral fraction compared to our inference on current observations (red, the same as in Figure 11). Dashed lines show the median of the distributions; the shaded regions show the 16%–84% regions. We take a 10 pointing *JWST*/NIRSpec mock survey at  $z \sim 7$ , which assumes  $\bar{x}_{\text{HI}} = 0.58$ , as described in Section 4.3, and perform Bayesian inference, assuming a  $5\sigma$  flux limit of  $> 3 \times 10^{-18} \text{ erg s}^{-1} \text{ cm}^{-2}$ . We show the posterior distribution for  $\bar{x}_{\text{HI}}$  inferred from current data (as described in Section 4.2) for comparison. In this example, *JWST* could reduce the uncertainty on the neutral fraction by  $\sim 70\%$ .

dropout detections. We also show the forecast for a cluster lensing survey (e.g., GLASS, Treu et al. 2015; Schmidt et al. 2016), assuming a simple uniform magnification factor of  $\mu = 2$  due to gravitational lensing (i.e.,  $p(\mu) = \delta(\mu - 2)$ ). In this case, all fluxes are magnified by  $\mu$  while the area decreases by  $1/\mu$ , and assuming the same flux limit as above, we predict  $\sim 10 \pm 2$  Ly $\alpha$  lines from a total of  $\sim 90$  LBG detections. The NIRSpec field of view is still small compared to large-scale structure at  $z \sim 7$ , so wide area random pointing surveys will be essential to estimate the global  $\bar{x}_{\text{HI}}$ .

We simulate a 10 pointing NIRSpec survey with F070LP/G140M ( $R = 1000$ ), with 3 hr exposures in each field, by again sampling the Mason et al. (2015) luminosity function in a larger area. We perform the inference on these mock *JWST* observations at  $z \sim 7$ , assuming  $\bar{x}_{\text{HI}} = 0.59$ . This yields  $\sim 60$  detections from  $\sim 800$  LBGs. Again, we assume a  $5\sigma$  flux limit of  $> 3 \times 10^{-18} \text{ erg s}^{-1} \text{ cm}^{-2}$ . The posterior distribution obtained from the *JWST* mock observations is shown in Figure 14, with the posterior from the current observations (Section 4.2) shown for comparison. We obtain  $\bar{x}_{\text{HI}} = 0.60^{+0.02}_{-0.06}$ , a  $\sim 70\%$  reduction in uncertainty compared to the current sample. We note this is an average forecast, and a more realistic survey forecast would require sampling the simulation directly (e.g., Mesinger & Furlanetto 2008a). We also caution our mock survey assumes 100% completeness, and maximized filling of NIRSpec slits, but, nevertheless, observations with NIRSpec will constrain the neutral fraction much more tightly than current observations.

## 5. Discussion

In this section we discuss our result in the context of other probes of reionization (Section 5.1), and we discuss the

implications of the mass-dependent Ly $\alpha$  velocity offset on the evolving Ly $\alpha$  fraction for average (Section 5.2) and UV bright (Section 5.3) galaxies.

### 5.1. The Global Reionization History

Robust constraints on the reionization history are challenging. While quasars provide high S/N information about individual (but rare) lines of sight, they are likely to be biased to overdense and more ionized regions (Barkana & Loeb 2004; Mesinger 2010; Decarli et al. 2017), and the number densities of bright quasars drop dramatically at  $z > 6$  (Fan et al. 2001; Manti et al. 2016; Parsa et al. 2018). Constraining reionization with large samples of galaxies clearly avoids these problems; with the help of gravitational lensing in clusters (e.g., the Frontier Fields; Lotz et al. 2017), we know there are large populations of faint galaxies at  $z > 6$  (Yue et al. 2014; Atek et al. 2015a; Livermore et al. 2017; Vanzella et al. 2017), and GRB host galaxy searches indicate far fainter galaxies must also exist (Kistler et al. 2009; Trenti et al. 2012).

Ly $\alpha$  emission from galaxies has long been recognized as a probe of reionization (Haiman & Spaans 1999; Malhotra & Rhoads 2004; Santos 2004; Verhamme et al. 2006; McQuinn et al. 2007a; Dijkstra 2014), and the framework presented in this paper provides a direct constraint on the IGM neutral fraction from observations of Ly $\alpha$  emission from galaxies, incorporating both realistic galaxy properties and realistic IGM topologies for the first time.

Our constraint on the neutral fraction,  $\bar{x}_{\text{HI}} = 0.59^{+0.11}_{-0.15}$ , is consistent with other robust probes of IGM neutrality at  $z \sim 7$  (Mesinger et al. 2015; Greig et al. 2017), demonstrating the power of Ly $\alpha$  follow-up of LBGs to constrain the neutral fraction, and providing more strong evidence that the IGM is undergoing significant reionization at  $z \sim 7$ . Using the full distribution of observed  $\{W, M_{\text{UV}}\}$  as inputs to our inference provides much tighter constraints than using the standard “Ly $\alpha$  fraction,” as we demonstrated in Figure 11.

Our median value lies  $\Delta \bar{x}_{\text{HI}} \sim 0.2$  higher than that inferred by Greig et al. (2017) from the QSO ULASJ1120+0641 damping wings at  $z = 7.1$ , which was obtained using the same IGM simulations, though our posterior distribution is marginally skewed to lower values (see Figure 11). This offset is not significant given the uncertainties, and does not require us to invoke any additional evolution in galaxy properties. Within the next few years larger samples, as demonstrated in our mock survey with *JWST* described in Section 4.3, will greatly reduce the uncertainties in our constraints from Ly $\alpha$  detections and non-detections.

With large samples, it will be possible to measure the variations in  $\bar{x}_{\text{HI}}$  over the sky, and cross-correlate with other constraints from quasars and eventually 21 cm observations (Lidz et al. 2009; Mesinger et al. 2016; Mirocha et al. 2016; Sobacchi et al. 2016; Vrbanec et al. 2016; Greig & Mesinger 2017a) to directly observe the inhomogeneous process of reionization. Furthermore, with tighter constraints on the timeline of reionization, it will be possible to better constrain the sources of ionizing photons, as the ionizing photon budget from galaxies depends on, for example, the minimum mass/luminosity of galaxies and the rate of ionizing photons per unit UV luminosity.

### 5.2. A Sudden Drop in Ly $\alpha$ Emission—Redshift Evolution of $\Delta v$ ?

In our model, we include empirically calibrated relations for both the intrinsic dependence of Ly $\alpha$  EW on UV magnitude and ISM radiative transfer in galaxies of a given halo mass (UV magnitude at fixed redshift), which builds in a simple redshift evolution assuming galaxies of the same UV magnitude live in less massive halos at higher redshifts. In this framework, UV faint galaxies have intrinsically higher EW than UV bright galaxies and lower Ly $\alpha$  velocity offsets.

These correlations are motivated by numerous observations of Ly $\alpha$  emission from galaxies at a range of redshifts, including very low redshift samples where detailed spatial and spectral observations are possible (Hayes et al. 2013; Yang et al. 2016). It is likely the density and distribution of neutral gas in the ISM plays a key role in the mediation of Ly $\alpha$  propagation through galaxies: an ISM with high column densities of neutral hydrogen,  $N_{\text{HI}}$ , scatters Ly $\alpha$  photons more significantly, spectrally, and spatially (Verhamme et al. 2006; Zheng et al. 2010). Observations of  $z < 4$  galaxies confirm high  $N_{\text{HI}}$  correlates with high Ly $\alpha$  velocity offset (Hashimoto et al. 2015; Henry et al. 2015; Yang et al. 2016), and more Ly $\alpha$  extended halos (Hayes et al. 2013; Guaita et al. 2017).

With increasing redshift, where galaxies are less massive (Lacey & Cole 1993), Ly $\alpha$  should escape more easily with high EW: these galaxies will contain less dust (as ALMA and Plateau de Bure Interferometer [PdBI] results are suggesting; Walter et al. 2012; Ouchi et al. 2013; Ota et al. 2014; Capak et al. 2015; Maiolino et al. 2015; Schaerer et al. 2015; Bouwens et al. 2016a; Pentericci et al. 2016) and neutral gas than at low redshifts. Additionally, the covering fraction of neutral hydrogen may evolve with galaxy mass, star formation rate, stellar populations, and/or redshift. Hard ionizing spectra from low metallicity stars (which may be significant at high redshifts; Stark et al. 2015, 2017; Mainali et al. 2017; Schmidt et al. 2017) can create more ionized holes through the ISM, reducing the covering fraction, an effect that is enhanced for low mass galaxies (Trebitsch et al. 2017). A low covering fraction would facilitate Ly $\alpha$  escape closer to the galaxy systemic velocity, and some observations have indicated a decreasing covering fraction with redshift (Jones et al. 2013; Leethochawalit et al. 2016).

All these factors, and the correlation of  $\Delta v$  with halo mass, as shown in Figure 2, suggest velocity offsets should decrease with increasing redshift for galaxies at fixed UV magnitude. These low velocity offsets are correlated with reduced scattering within the ISM and thus a higher EW of Ly $\alpha$ . This should increase the visibility of Ly $\alpha$  until the IGM starts to become neutral and these low  $\Delta v$  lines are easily attenuated by nearby neutral hydrogen. As was noted by Mesinger et al. (2015) and Choudhury et al. (2015), this offers a simple explanation for evolving galaxy properties which may accelerate the decline in Ly $\alpha$  in UV faint galaxies.

### 5.3. Ly $\alpha$ from UV Bright Galaxies—Redshifted away from Resonance?

A high fraction of Ly $\alpha$  observed in some UV bright ( $M_{\text{UV}} < -21.5$ ) galaxies at  $z > 6$  (Curtis-Lake et al. 2012; Stark et al. 2017, though cf. Treu et al. 2013 for non-detections of Ly $\alpha$  in slightly fainter galaxies) is surprising for several reasons: the electron scattering optical depth from the Planck

Collaboration (2016) favors a significant IGM neutral fraction at these redshifts, with instantaneous reionization occurring at  $z = 7.8$ – $8.8$ ; and the observed fraction of UV faint galaxies appears to steadily decrease at the same redshifts (Pentericci et al. 2014; Schenker et al. 2014). Why can we more easily see Ly $\alpha$  in some UV bright galaxies into the Epoch of Reionization?

The most UV bright galaxies at high redshift probably reside in halos with mass  $\gtrsim 10^{11} M_{\odot}$ , which may already have stable gaseous disks, as suggested by recent ALMA observations of two UV bright galaxies at  $z \sim 6$  (Smit et al. 2018) and observations of stable rotation in low mass galaxies at  $z \sim 1$ – $2$  (Mason et al. 2016; Stott et al. 2016). Thus it is likely Ly $\alpha$  photons traveling from these galaxies will experience significant radiative transfer effects with the ISM.

The enhanced Doppler shift of the emerging Ly $\alpha$  photons in UV bright galaxies provides some explanation for the high fraction of Ly $\alpha$  observations for these populations compared to UV faint galaxies at  $z \sim 7$ . As shown in Figures 6 and 7, we predict the transmission of UV bright galaxies evolves more slowly with the evolving IGM compared to fainter objects, making them visible far into the epoch of reionization and thus prime targets for spectroscopic confirmation. With this in mind, note their underlying EW distribution is likely much steeper and has a higher peak of non-emitters than for UV faint galaxies. When Ly $\alpha$  is emitted from UV bright objects, it is likely to have low EW, as the photons are dispersed spatially and spectrally.

However, this effect is also highly correlated with the large-scale environment in which these galaxies reside; assessing the relative contributions of evolving galaxy properties and environment to this apparent increase in the Ly $\alpha$  fraction is explored by Mason et al. (2018). The high Ly $\alpha$  transmission of UV bright galaxies make them ideal targets for spectroscopic follow-up to understand the star formation processes occurring in the early universe.

## 6. Summary and Conclusions

We have developed a flexible Bayesian inference framework to infer the IGM neutral fraction during reionization by forward-modeling the observed equivalent width distribution of Ly $\alpha$  emission from LBGs. Our model incorporates sightlines through realistic IGM simulations to model galaxies with realistic ISM properties.

Our main conclusions are as follows:

- (i) The Ly $\alpha$  line profile emerging from the ISM has a huge impact on the probability of transmission through the IGM (Dijkstra et al. 2011), and is related to the properties of the source galaxy. This must be systematically accounted for in reionization inference.
- (ii) We introduce a simple empirical relation between the halo mass of a galaxy (or UV luminosity at fixed redshift) and its Ly $\alpha$  line peak velocity offset, where the most massive galaxies have the largest velocity offsets likely due to increased  $N_{\text{HI}}$  in the ISM, higher halo circular velocities, and/or the presence of star formation induced outflows.
- (iii) This relation predicts that with increasing redshift, Ly $\alpha$  velocity offsets will decrease for galaxies at fixed UV luminosity, making Ly $\alpha$  lines more susceptible to absorption in the IGM. This effect would accelerate the



decline in Ly $\alpha$  emission compared to other reionization probes and be a factor in explaining the sudden drop of Ly $\alpha$  emission observed at  $z > 6$ .

- (iv) We conduct a Bayesian inference from current observations at  $z \sim 7$  from Pentericci et al. (2014) and infer the first direct constraint on the neutral fraction from Ly $\alpha$  transmission of  $\bar{x}_{\text{H I}} = 0.59^{+0.11}_{-0.15}$ , which is consistent with other robust probes of the neutral fraction and confirms reionization is ongoing at  $z \sim 7$ .
- (v) Using the full distribution of Ly $\alpha$  equivalent width measurements enables us to provide much tighter constraints on the neutral fraction compared to the standard “Ly $\alpha$  fraction,”  $P(W > 25\text{\AA})$ , used in previous analyses.
- (vi) We make predictions for spectroscopic surveys with *JWST* and find a  $\sim 30$  hr LBG follow-up survey with *JWST*/NIRSpec could reduce the uncertainty in  $\bar{x}_{\text{H I}}$  by  $\sim 70\%$ .

Future near-IR spectrographs in space, such as *JWST* NIRSpec and NIRISS, will be able to observe both Ly $\alpha$  and rest-frame optical lines for galaxies to  $z \lesssim 12$  and to measure SFRs and Ly $\alpha$  velocity offsets for these objects, enabling us to further understand the interactions between star-forming regions, the ISM, and the reionizing IGM. It will soon be possible to apply our framework to large samples, free of cosmic variance, to get accurate universal constraints on the evolution of the neutral fraction.

The authors thank Dawn Erb and Dan Stark for providing their observational data. We thank Simon Birrer, Fred Davies, Max Gronke, Joe Hennawi, and Crystal Martin for useful discussions.

C.M. acknowledges support by NASA Headquarters through the NASA Earth and Space Science Fellowship Program Grant NNX16AO85H. A.M. acknowledges support from the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation program (grant agreement No 638809 AIDA). This research was partially supported by the Australian Research Council through awards FT130101593 and CE170100013. This work was supported by the *HST* BoRG grants GO-12572, 12905, and 13767, and the *HST* GLASS grant GO-13459.

This work made use of the following open source software: IPython (Pérez & Granger 2007), matplotlib (Hunter 2007), NumPy (Van Der Walt et al. 2011), SciPy (Oliphant 2007), Astropy (Robitaille et al. 2013), and EMCEE (Foreman-Mackey et al. 2013).

### ORCID iDs

Charlotte A. Mason  <https://orcid.org/0000-0002-3407-1785>  
 Tommaso Treu  <https://orcid.org/0000-0002-8460-0390>  
 Michele Trenti  <https://orcid.org/0000-0001-9391-305X>  
 Laura Pentericci  <https://orcid.org/0000-0001-8940-6768>  
 Stephane de Barros  <https://orcid.org/0000-0002-5877-8793>  
 Eros Vanzella  <https://orcid.org/0000-0002-5057-135X>

### References

Atek, H., Richard, J., Jauzac, M., et al. 2015a, *ApJ*, 814, 69  
 Atek, H., Richard, J., Kneib, J.-P., et al. 2015b, *ApJ*, 800, 18  
 Bañados, E., Venemans, B. P., Mazzucchelli, C., et al. 2017, *Natur*, 553, 473  
 Barkana, R., & Loeb, A. 2004, *ApJ*, 601, 64

Barone-Nugent, R. L., Trenti, M., Wyithe, J. S. B., et al. 2014, *ApJ*, 793, 17  
 Becker, G. D., Bolton, J. S., Madau, P., et al. 2015, *MNRAS*, 447, 3402  
 Behroozi, P. S., Marchesini, D., Wechsler, R. H., et al. 2013a, *ApJL*, 777, L10  
 Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013b, *ApJ*, 770, 57  
 Bolton, J. S., & Haehnelt, M. G. 2013, *MNRAS*, 429, 1695  
 Bouwens, R. J., Aravena, M., Decarli, R., et al. 2016a, *ApJ*, 833, 72  
 Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2012, *ApJ*, 754, 83  
 Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2014, *ApJ*, 793, 115  
 Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2015a, *ApJ*, 811, 140  
 Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2015b, *ApJ*, 803, 34  
 Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2017a, *ApJ*, 843, 41  
 Bouwens, R. J., Illingworth, G. D., Rosati, P., et al. 2003, *ApJ*, 595, 589  
 Bouwens, R. J., Oesch, P. A., Illingworth, G. D., Ellis, R. S., & Stefanon, M. 2017b, *ApJ*, 843, 129  
 Bouwens, R. J., Smit, R., Labbé, I., et al. 2016b, *ApJ*, 831, 176  
 Bradač, M., Garcia-Appadoo, D., Huang, K.-H., et al. 2017, *ApJL*, 836, L2  
 Bradač, M., Vanzella, E., Hall, N., et al. 2012, *ApJL*, 755, L7  
 Bradley, L., Trenti, M., Oesch, P. A., et al. 2012, *ApJ*, 760, 108  
 Bunker, A. J., Stanway, E. R., Ellis, R. S., & McMahon, R. G. 2004, *MNRAS*, 355, 374  
 Calvi, V., Trenti, M., Stiavelli, M., et al. 2016, *ApJ*, 817, 120  
 Capak, P. L., Carilli, C., Jones, G., et al. 2015, *Natur*, 522, 455  
 Caruana, J., Bunker, A. J., Wilkins, S. M., et al. 2012, *MNRAS*, 427, 3055  
 Caruana, J., Bunker, A. J., Wilkins, S. M., et al. 2014, *MNRAS*, 443, 2831  
 Cassata, P., Tasca, L. A. M., Le Fèvre, O., et al. 2015, *A&A*, 573, A24  
 Chonis, T. S., Blanc, G. A., Hill, G. J., et al. 2013, *ApJ*, 775, 99  
 Choudhury, T. R., Puchwein, E., Haehnelt, M. G., & Bolton, J. S. 2015, *MNRAS*, 452, 261  
 Curtis-Lake, E., McLure, R. J., Pearce, H. J., et al. 2012, *MNRAS*, 422, 1425  
 De Barros, S., Pentericci, L., Vanzella, E., et al. 2017, *A&A*, 608, A123  
 Decarli, R., Walter, F., Venemans, B. P., et al. 2017, *Natur*, 545, 457  
 Dijkstra, M. 2014, *PASA*, 31, e040  
 Dijkstra, M., Lidz, A., & Wyithe, J. S. B. 2007, *MNRAS*, 377, 1175  
 Dijkstra, M., Mesinger, A., & Wyithe, J. S. B. 2011, *MNRAS*, 414, 2139  
 Dijkstra, M., & Wyithe, J. S. B. 2012, *MNRAS*, 419, 3181  
 Dijkstra, M., Wyithe, S., Haiman, Z., Mesinger, A., & Pentericci, L. 2013, *MNRAS*, 440, 3309  
 Erb, D. K. 2015, *Natur*, 523, 169  
 Erb, D. K., Steidel, C. C., Trainor, R. F., et al. 2014, *ApJ*, 795, 33  
 Faisst, A. L., Capak, P., Carollo, C. M., Scarlata, C., & Scoville, N. 2014, *ApJ*, 788, 87  
 Fan, X., Narayanan, V. K., Lupton, R. H., et al. 2001, *AJ*, 122, 2833  
 Fan, X., Strauss, M. A., Becker, R. H., et al. 2006, *AJ*, 132, 117  
 Finkelstein, S. L., Papovich, C., Dickinson, M., et al. 2013, *Natur*, 502, 524  
 Finkelstein, S. L., Papovich, C., Ryan, R. E., et al. 2012a, *ApJ*, 758, 93  
 Finkelstein, S. L., Papovich, C., Salmon, B., et al. 2012b, *ApJ*, 756, 164  
 Finkelstein, S. L., Ryan, R. E., Papovich, C., et al. 2015a, *ApJ*, 810, 71  
 Finkelstein, S. L., Song, M., Behroozi, P., et al. 2015b, *ApJ*, 814, 95  
 Fontana, A., Vanzella, E., Pentericci, L., et al. 2010, *ApJL*, 725, L205  
 Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, 125, 306  
 Furlanetto, S. R., Zaldarriaga, M., & Hernquist, L. 2006, *MNRAS*, 365, 1012  
 Gardner, J. P., Mather, J. C., Clampin, M., et al. 2006, *SSRv*, 123, 485  
 Giallongo, E., Grazian, A., Fiore, F., et al. 2015, *A&A*, 578, A83  
 Grazian, A., Castellano, M., Fontana, A., et al. 2012, *A&A*, 547, A51  
 Grazian, A., Giallongo, E., Paris, D., et al. 2017, *A&A*, 602, A18  
 Greig, B., & Mesinger, A. 2017a, *MNRAS*, 472, 2651  
 Greig, B., & Mesinger, A. 2017b, *MNRAS*, 465, 4838  
 Greig, B., Mesinger, A., Haiman, Z., & Simcoe, R. A. 2017, *MNRAS*, 466, 4239  
 Gronke, M., Bull, P., & Dijkstra, M. 2015a, *ApJ*, 812, 123  
 Gronke, M., & Dijkstra, M. 2016, *ApJ*, 826, 14  
 Gronke, M., Dijkstra, M., McCourt, M., & Oh Peng, S. 2016, *ApJL*, 833, L26  
 Gronke, M., Dijkstra, M., Trenti, M., & Wyithe, S. 2015b, *MNRAS*, 449, 1284  
 Guaita, L., Talia, M., Pentericci, L., et al. 2017, *A&A*, 606, A19  
 Gunn, J. E., & Peterson, B. A. 1965, *ApJ*, 142, 1633  
 Haiman, Z., & Spaans, M. 1999, *ApJ*, 518, 138  
 Hall, N., Bradač, M., Gonzalez, A. H., et al. 2012, *ApJ*, 745, 155  
 Harikane, Y., Ouchi, M., Ono, Y., et al. 2016, *ApJ*, 821, 123  
 Harikane, Y., Ouchi, M., Ono, Y., et al. 2017, *PASJ*, 70, S11  
 Hashimoto, T., Ouchi, M., Shimasaku, K., et al. 2013, *ApJ*, 765, 70  
 Hashimoto, T., Verhamme, A., Ouchi, M., et al. 2015, *ApJ*, 812, 157  
 Hayes, M., Östlin, G., Schaerer, D., et al. 2013, *ApJL*, 765, L27  
 Hayes, M., Schaerer, D., Östlin, G., et al. 2011, *ApJ*, 730, 8  
 Hayward, C. C., & Hopkins, P. F. 2017, *MNRAS*, 465, 1682  
 Henry, A., Scarlata, C., Martin, C. L., & Erb, D. 2015, *ApJ*, 809, 19

- Herenz, E. C., Urrutia, T., Wisotzki, L., et al. 2017, *A&A*, 606, A12
- Hunter, J. D. 2007, *CSE*, 9, 99
- Iliev, I. T., Mellema, G., Ahn, K., et al. 2014, *MNRAS*, 439, 725
- Illingworth, G. D., Magee, D., Oesch, P. A., et al. 2013, *ApJS*, 209, 6
- Inoue, A. K., Tamura, Y., Matsuo, H., et al. 2016, *Sci*, 352, 1559
- Jensen, H., Laursen, P., Mellema, G., et al. 2013, *MNRAS*, 428, 1366
- Jones, T. A., Ellis, R. S., Schenker, M. A., & Stark, D. P. 2013, *ApJ*, 779, 52
- Kakiichi, K., & Dijkstra, M. 2017, arXiv:1710.10053
- Kimm, T., Katz, H., Haehnelt, M., et al. 2017, *MNRAS*, 466, 4826
- Kistler, M. D., Yüksel, H., Beacom, J. F., Hopkins, A. M., & Wyithe, J. S. B. 2009, *ApJL*, 705, L104
- Konno, A., Ouchi, M., Shibuya, T., et al. 2017, *PASJ*, 70, S16
- Lacey, C., & Cole, S. 1993, *MNRAS*, 262, 627
- Laursen, P., Sommer-Larsen, J., & Razoumov, A. O. 2011, *ApJ*, 728, 52
- Leethochawalit, N., Jones, T. A., Ellis, R. S., Stark, D. P., & Zitrin, A. 2016, *ApJ*, 831, 152
- Lehnert, M. D., & Bremer, M. 2003, *ApJ*, 593, 630
- Lidz, A., Zahn, O., Furlanetto, S. R., et al. 2009, *ApJ*, 690, 252
- Livermore, R. C., Finkelstein, S. L., & Lotz, J. M. 2017, *ApJ*, 835, 113
- Lotz, J. M., Koekemoer, A., Coe, D., et al. 2017, *ApJ*, 837, 97
- Ma, X., Kasen, D., Hopkins, P. F., et al. 2015, *MNRAS*, 453, 960
- Madau, P., & Haardt, F. 2015, *ApJL*, 813, L8
- Mainali, R., Kollmeier, J. A., Stark, D. P., et al. 2017, *ApJL*, 836, L14
- Maiolino, R., Carniani, S., Fontana, A., et al. 2015, *MNRAS*, 452, 54
- Malhotra, S., & Rhoads, J. E. 2004, *ApJL*, 617, L5
- Manti, S., Gallerani, S., Ferrara, A., Greig, B., & Feruglio, C. 2016, *MNRAS*, 466, 1160
- Mason, C. A., Trenti, M., & Treu, T. 2015, *ApJ*, 813, 21
- Mason, C. A., Treu, T., de Barros, S., et al. 2018, *ApJL*, submitted (arXiv:1801.01891)
- Mason, C. A., Treu, T., Fontana, A., et al. 2016, *ApJ*, 838, 14
- McGreer, I. D., Mesinger, A., & D'Odorico, V. 2014, *MNRAS*, 447, 499
- McLean, I. S., Steidel, C. C., Epps, H. W., et al. 2012, *Proc. SPIE*, 8446, 84460J
- McLeod, D. J., McLure, R. J., & Dunlop, J. S. 2016, *MNRAS*, 459, 3812
- McLinden, E. M., Finkelstein, S. L., Rhoads, J. E., et al. 2011, *ApJ*, 730, 136
- McLure, R. J., Dunlop, J. S., Cirasuolo, M., et al. 2010, *MNRAS*, 403, 960
- McQuinn, M., Hernquist, L., Zaldarriaga, M., & Dutta, S. 2007a, *MNRAS*, 381, 75
- McQuinn, M., Lidz, A., Zahn, O., et al. 2007b, *MNRAS*, 377, 1043
- Mesinger, A. 2010, *MNRAS*, 407, 1328
- Mesinger, A. (ed.) 2016, *Understanding the Epoch of Cosmic Reionization* (Vol. 423; Cham: Springer)
- Mesinger, A., Aykutalp, A., Vanzella, E., et al. 2015, *MNRAS*, 446, 566
- Mesinger, A., & Furlanetto, S. R. 2008a, *MNRAS*, 385, 1348
- Mesinger, A., & Furlanetto, S. R. 2008b, *MNRAS*, 386, 1990
- Mesinger, A., Greig, B., & Sobacchi, E. 2016, *MNRAS*, 459, 2342
- Miralda-Escude, J. 1998, *ApJ*, 501, 15
- Mirocha, J., Furlanetto, S. R., & Sun, G. 2016, *MNRAS*, 464, 1365
- Mitra, S., Roy Choudhury, T., & Ferrara, A. 2015, *MNRAS*, 454, L76
- Mostardi, R. E., Shapley, A. E., Nestor, D. B., et al. 2013, *ApJ*, 779, 65
- Moster, B. P., Naab, T., & White, S. D. M. 2017, arXiv:1705.05373
- Oesch, P. A., Dokkum, P. G. V., Illingworth, G. D., et al. 2015, *ApJL*, 804, L30
- Oliphant, T. E. 2007, *CSE*, 9, 10
- Ono, Y., Ouchi, M., Mobasher, B., et al. 2012, *ApJ*, 744, 83
- Onoue, M., Kashikawa, N., Willott, C. J., et al. 2017, *ApJL*, 847, L15
- Ota, K., Iye, M., Kashikawa, N., et al. 2017, *ApJ*, 844, 85
- Ota, K., Walter, F., Ohta, K., et al. 2014, *ApJ*, 792, 34
- Ouchi, M., Ellis, R., Ono, Y., et al. 2013, *ApJ*, 778, 102
- Ouchi, M., Shimasaku, K., Furusawa, H., et al. 2010, *ApJ*, 723, 869
- Oyarzún, G. A., Blanc, G. A., González, V., et al. 2016, *ApJL*, 821, L14
- Oyarzún, G. A., Blanc, G. A., González, V., Mateo, M., & Bailey, J. I. 2017, *ApJ*, 843, 133
- Papovich, C., Finkelstein, S. L., Ferguson, H. C., Lotz, J. M., & Gialalisco, M. 2011, *MNRAS*, 412, 1123
- Parsa, S., Dunlop, J. S., & McLure, R. J. 2018, *MNRAS*, 474, 2904
- Parzen, E. 1962, *Ann. Math. Stat.*, 33, 1065
- Pentericci, L., Carniani, S., Castellano, M., et al. 2016, *ApJL*, 829, L11
- Pentericci, L., Vanzella, E., Fontana, A., et al. 2014, *ApJ*, 793, 113
- Pérez, F., & Granger, B. E. 2007, *CSE*, 9, 21
- Planck Collaboration, Adam, R., Ade, P. A. R., et al. 2015, *A&A*, 594, A8
- Planck Collaboration, Adam, R., Aghanim, N., et al. 2016, *A&A*, 596, A108
- 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
- Robertson, B. E., Furlanetto, S. R., Schneider, E., et al. 2013, *ApJ*, 768, 71
- Robitaille, T. P., Tollerud, E. J., Greenfield, P., et al. 2013, *A&A*, 558, A33
- Rosenblatt, M. 1956, *Ann. Math. Stat.*, 27, 832
- Santos, M. R. 2004, *MNRAS*, 349, 1137
- Schaerer, D., Boone, F., Zamojski, M., et al. 2015, *A&A*, 574, A19
- Schenker, M. A., Ellis, R. S., Konidaris, N. P., & Stark, D. P. 2013a, *ApJ*, 777, 67
- Schenker, M. A., Ellis, R. S., Konidaris, N. P., & Stark, D. P. 2014, *ApJ*, 795, 20
- Schenker, M. A., Robertson, B. E., Ellis, R. S., et al. 2013b, *ApJ*, 768, 196
- Schenker, M. A., Stark, D. P., Ellis, R. S., et al. 2012, *ApJ*, 744, 179
- Schmidt, K. B., Huang, K.-H., Treu, T., et al. 2017, *ApJ*, 839, 17
- Schmidt, K. B., Treu, T., Bradač, M., et al. 2016, *ApJ*, 818, 38
- Schmidt, K. B., Treu, T., Trenti, M., et al. 2014, *ApJ*, 786, 57
- Schroeder, J., Mesinger, A., & Haiman, Z. 2013, *MNRAS*, 428, 3058
- Shapley, A. E., Steidel, C. C., Pettini, M., & Adelberger, K. L. 2003, *ApJ*, 588, 65
- Shibuya, T., Ouchi, M., Nakajima, K., et al. 2014, *ApJ*, 788, 74
- Shull, J. M., Harness, A., Trenti, M., & Smith, B. D. 2012, *ApJ*, 747, 100
- Smit, R., Bouwens, R. J., Carniani, S., et al. 2018, *Natur*, 553, 178
- Sobacchi, E., & Mesinger, A. 2014, *MNRAS*, 440, 1662
- Sobacchi, E., & Mesinger, A. 2015, *MNRAS*, 453, 1843
- Sobacchi, E., Mesinger, A., & Greig, B. 2016, *MNRAS*, 459, 2741
- 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., Ellis, R. S., & Ouchi, M. 2011, *ApJL*, 728, L2
- Stark, D. P., Walth, G., Charlot, S., et al. 2015, *MNRAS*, 454, 1393
- Steidel, C. C., Rudie, G. C., Strom, A. L., et al. 2014, *ApJ*, 795, 165
- Stott, J. P., Swinbank, A. M., Johnson, H. L., et al. 2016, *MNRAS*, 457, 1888
- Tilvi, V., Papovich, C., Finkelstein, S. L., et al. 2014, *ApJ*, 794, 5
- Trac, H., Cen, R., & Loeb, A. 2008, *ApJL*, 689, L81
- Trac, H. Y., & Gnedin, N. Y. 2011, *ASL*, 4, 228
- Trainor, R. F., Steidel, C. C., Strom, A. L., & Rudie, G. C. 2015, *ApJ*, 809, 89
- Trebtsch, M., Blaizot, J., Rosdahl, J., Devriendt, J., & Slyz, A. 2017, *MNRAS*, 470, 224
- Trenti, M., Bradley, L. D., Stiavelli, M., et al. 2011, *ApJL*, 727, L39
- Trenti, M., Perna, R., Levesque, E. M., Shull, J. M., & Stocke, J. T. 2012, *ApJL*, 749, L38
- Trenti, M., & Stiavelli, M. 2008, *ApJ*, 676, 767
- Treu, T., Schmidt, K. B., Brammer, G. B., et al. 2015, *ApJ*, 812, 114
- Treu, T., Schmidt, K. B., Trenti, M., Bradley, L. D., & Stiavelli, M. 2013, *ApJL*, 775, L29
- Treu, T., Trenti, M., Stiavelli, M., Auger, M. W., & Bradley, L. D. 2012, *ApJ*, 747, 27
- Van Der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, *CSE*, 13, 22
- Vanzella, E., Barros, S. D., Cupani, G., et al. 2016, *ApJL*, 821, L27
- Vanzella, E., Castellano, M., Meneghetti, M., et al. 2017, *ApJ*, 842, 47
- Vanzella, E., Pentericci, L., Fontana, A., et al. 2011, *ApJL*, 730, L35
- Venemans, B. P., Bañados, E., Decarli, R., et al. 2015, *ApJL*, 801, L11
- Verhamme, A., Schaerer, D., Atek, H., & Tapken, C. 2008, *A&A*, 491, 89
- Verhamme, A., Schaerer, D., & Maselli, A. 2006, *A&A*, 460, 397
- Vrbanc, D., Ciardi, B., Jelić, V., et al. 2016, *MNRAS*, 457, 666
- Vulcani, B., Trenti, M., Calvi, V., et al. 2017, *ApJ*, 836, 239
- Walter, F., Decarli, R., Carilli, C., et al. 2012, *ApJ*, 752, 93
- Weisz, D. R., & Boylan-Kolchin, M. 2017, *MNRAS*, 469, L83
- Willott, C. J., Carilli, C. L., Wagg, J., & Wang, R. 2015, *ApJ*, 807, 180
- Yan, H., & Windhorst, R. A. 2004, *ApJL*, 600, L1
- Yang, H., Malhotra, S., Gronke, M., et al. 2016, *ApJ*, 820, 130
- Yang, H., Malhotra, S., Gronke, M., et al. 2017, *ApJ*, 844, 171
- Yue, B., Ferrara, A., Vanzella, E., & Salvaterra, R. 2014, *MNRAS*, 443, 1
- Zheng, Z., Cen, R., Trac, H., & Miralda-Escudé, J. 2010, *ApJ*, 716, 574
- Zheng, Z., & Miralda-Escudé, J. 2002, *ApJ*, 578, 33
- Zheng, Z.-Y., Wang, J., Rhoads, J., et al. 2017, *ApJL*, 842, L22
- Zitrin, A., Labbé, I., Belli, S., et al. 2015, *ApJL*, 810, L12