The Impact of Scatter in the Galaxy UV Luminosity to Halo Mass Relation on Lyα Visibility During the Epoch of Reionization

Lily R. Whitler,[1]★ Charlotte A. Mason,[2]† Keven Ren,[3,4] Mark Dijkstra,
Andrei Mesinger,[5] Laura Pentericci,[6] Michele Trenti,[3,4] and Tommaso Treu[7]

[1]*School of Earth and Space Exploration, Arizona State University, 781 Terrace Mall, Tempe, AZ 85287, USA*
[2]*Center for Astrophysics | Harvard & Smithsonian, 60 Garden St., Cambridge, MA 02138, USA*
[3]*School of Physics, University of Melbourne, Parkville, Victoria, Australia*
[4]*ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D)*
[5]*Scuola Normale Superiore, Piazza dei Cavalieri 7, I-56126 Pisa, Italy*
[6]*INAF, Osservatorio Astronomico di Roma, via Frascati 33, I-00040 Monteporzio, Italy*
[7]*Department of Physics and Astronomy, University of California, Los Angeles, CA 90095-1547, USA*

ABSTRACT

The reionization of hydrogen is closely linked to the first structures in the universe, so understanding the timeline of reionization promises to shed light on the nature of these early objects. In particular, transmission of Lyman alpha (Lyα) from galaxies through the intergalactic medium (IGM) is sensitive to neutral hydrogen in the IGM, so can be used to probe the reionization timeline. In this work, we implement an improved model of the galaxy UV luminosity to dark matter halo mass relation to infer the volume-averaged fraction of neutral hydrogen in the IGM from Lyα observations. Many models assume that UV-bright galaxies are hosted by massive dark matter haloes in overdense regions of the IGM, so reside in relatively large ionized regions. However, observations and N-body simulations indicate that scatter in the UV luminosity–halo mass relation is expected. Here, we model the scatter (though we assume the IGM topology is unaffected) and assess the impact on Lyα visibility during reionization. We show that UV luminosity–halo mass scatter reduces Lyα visibility compared to models without scatter, and that this is most significant for UV-bright galaxies. We then use our model with scatter to infer the neutral fraction, $\overline{x}_{\mathrm{H\,I}}$, at $z \sim 7$ using a sample of Lyman-break galaxies in legacy fields. We infer $\overline{x}_{\mathrm{H\,I}} = 0.55^{+0.11}_{-0.13}$ with scatter, compared to $\overline{x}_{\mathrm{H\,I}} = 0.59^{+0.12}_{-0.14}$ without scatter, a very slight decrease and consistent within the uncertainties. Finally, we place our results in the context of other constraints on the reionization timeline and discuss implications for future high-redshift galaxy studies.

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

1 INTRODUCTION

The Epoch of Reionization (EoR) is currently one of the most poorly understood eras in cosmic history. During the EoR, neutral hydrogen in the intergalactic medium (IGM) was (re)ionized by the first sources of light in the universe – stars and accreting black holes in galaxies. The properties of these sources are presently only loosely constrained, but a detailed understanding of the timeline of cosmic reionization will shed light on the characteristics of these early luminous objects (e.g. Barkana & Loeb 2001; Robertson et al. 2010; Mesinger 2016; Dayal & Ferrara 2018; Mason et al. 2019b).

In the last decade, observations of the cosmic microwave background (CMB), gamma ray bursts, quasars, and galaxies have begun to probe this era. For example, *Planck* measurements of the Thomson optical depth imply that hydrogen was 50 per cent neutral at $z \sim 7.8 - 8.8$ (Planck Collaboration et al. 2016a,b). Absorption features in quasar spectra, such as Gunn-Peterson troughs (Gunn & Peterson 1965) in $z \gtrsim 6$ spectra (Fan et al. 2006) and the dark pixel fraction in the Lyman alpha (Lyα; 1216 Å) and Lyman beta (Lyβ)

★ E-mail: lwhitler@asu.edu
† Hubble Fellow

forest at $z \sim 5-6$ (McGreer et al. 2015), suggest that reionization was largely complete by $z \sim 6$, although there is growing evidence that reionization was incomplete as late as $z \sim 5.5$ (e.g. Lidz et al. 2007; Mesinger 2010; Kulkarni et al. 2019a). Meanwhile, Lyα damping wing absorption in two quasar spectra at $z \gtrsim 7$ indicate that the universe was still partially neutral at $z \sim 7$ (e.g. Greig et al. 2017, 2019; Bañados et al. 2018; Davies et al. 2018).

However, probing the EoR with galaxies has advantages that CMB and quasar measurements lack. The CMB provides only integral constraints on the reionization history (Komatsu et al. 2011), and bright quasars at $z \gtrsim 7$ are both extremely rare (e.g. Parsa et al. 2018; Kulkarni et al. 2019b) and potentially biased towards high-density regions (Lidz et al. 2007). In contrast, galaxies can target specific stages of the EoR, and are more common and less biased than quasars at $z \gtrsim 7$.

In particular, Lyα emission from high-redshift galaxies is a key EoR probe. Lyα has an extremely large cross section for scattering by neutral hydrogen, making it sensitive to the ionization state of the IGM. Spectroscopic follow-up of galaxies selected as photometric dropouts (Lyman-break galaxies, or LBGs) show larger fractions of LBGs emitting Lyα with increasing redshift until $z \sim 6$ (Stark et al. 2010; Hayes et al. 2011; Curtis-Lake et al. 2012; Cassata et al. 2015), likely due to decreasing dust fractions and thus less attenuation in the interstellar medium (ISM) (Hayes et al. 2011). However, at $z \gtrsim 6$, the fraction of Lyα emitters begins to drop rapidly, suggesting increasing absorption of Lyα in the IGM as the universe becomes increasingly neutral (e.g. Stark et al. 2010; Pentericci et al. 2011; Ono et al. 2012; Schenker et al. 2012; Treu et al. 2012). Moreover, the clustering of Lyα-emitting galaxies also provides insights into the EoR. Lyα emission is preferentially transmitted through large ionized regions, boosting the clustering signal of Lyα emitters during reionization (Sobacchi & Mesinger 2015; Ouchi et al. 2018; Weinberger et al. 2019).

Quantitatively constraining the IGM ionization state with Lyα is nontrivial, however. The resonance of Lyα with neutral hydrogen makes it sensitive to galactic properties (e.g. neutral hydrogen column density and covering fraction) as well as the IGM (e.g. Haiman & Spaans 1999; Malhotra & Rhoads 2004; Santos 2004; McQuinn et al. 2007; Dijkstra 2014; Mason et al. 2018a). Therefore, models must account for physics ranging from pc to Gpc scales.

In this work, we build on the methods of Mason et al. (2018a) (hereafter M18a) to use Lyα observations to improve our understanding of the timeline of the EoR. M18a model Lyα visibility during reionization by combining semi-numerical simulations of the IGM with empirical models of radiative transfer effects in the ISM. They obtain simulated cubes of the reionizing IGM and the distribution of dark matter haloes from the Evolution of 21 cm Structure simulations (Mesinger et al. 2016), then use physically-motivated models to populate haloes with galaxy properties – UV luminosities and Lyα line profiles – to study transmission of Lyα through the reionizing IGM.

A key assumption of the M18a models is a one-to-one mapping between UV luminosity and halo mass. M18a use the model of Mason et al. (2015), which calibrates a star-formation efficiency at a single redshift with abundance matching, and evolves the UV luminosity–halo mass relation

over cosmic time assuming star formation rates are proportional to halo mass accretion rates. This evolution preserves the one-to-one relation between UV luminosity and halo mass from the calibration redshift, and captures the simple physical intuition that UV luminous galaxies reside in high-mass haloes in more overdense regions, which reionize early.

To test this assumption, we now model Lyα visibility during the EoR with scatter in the galaxy UV luminosity–halo mass relation, which allows haloes of a single mass to host galaxies with a distribution of UV luminosities. Scatter is expected due to (at least) variation in halo mass assembly times and the stochastic nature of star formation, both of which increase with increasing redshift. Ren et al. (2018) investigated the impact of scatter in the UV luminosity–halo mass relation in the context of the Mason et al. (2015) modelling framework and calculated a minimum log-normal scatter between UV luminosity and halo mass at $z \sim 7$ of ~ 0.5 mag. This is consistent with scatter obtained in semi-analytical models built on N-body simulations (e.g. Behroozi et al. 2010, 2013; Moster et al. 2013; Ren et al. 2018; Tacchella et al. 2018), estimated from comparing observed luminosity and stellar mass functions (Finkelstein et al. 2015), and satellite kinematics (More et al. 2009). This scatter is also consistent with the observed $z \gtrsim 6$ luminosity functions (Ren et al. 2019).

Scatter in the UV luminosity–halo mass relation alters the distribution of galaxies within the reionizing IGM, and therefore impacts Lyα transmission through the IGM. As the most massive haloes are extremely rare, including scatter greatly increases the probability of a very luminous galaxy being hosted in a more common, lower-mass halo. In contrast, abundance matching places UV-bright galaxies in high-mass haloes, which tend to reside in overdense regions with abundant neighboring haloes, creating larger ionized bubbles (e.g. Furlanetto et al. 2004, 2006; McQuinn et al. 2007) and facilitating higher transmission of Lyα through the IGM. But if the most UV-luminous galaxies do not necessarily live in the most massive haloes surrounded by large ionized regions, and rather can reside in less massive haloes surrounded by smaller ionized bubbles, this could reduce the visibility of Lyα from these galaxies. On average, if the degree of absorption in the IGM is fixed, this would tend to reduce the quantity of neutral hydrogen in the IGM necessary to attenuate Lyα to the same extent. That is, the UV luminosity–halo mass relation without scatter requires a higher volume-averaged fraction of neutral hydrogen in the IGM to suppress Lyα than when scatter is included.

This paper is structured as follows. We describe our model for Lyα visibility during the EoR, including scatter in the UV luminosity–halo mass relation, in Section 2. In Section 3, we present our results on Lyα visibility and our inferred value of the fraction of neutral hydrogen in the IGM at $z \sim 7$, accounting for scatter. We discuss our results in Section 4, and summarise and conclude in Section 5.

We adopt a ΛCDM cosmology with parameters $\Omega_\Lambda = 0.69$, $\Omega_m = 0.31$, $\Omega_b = 0.048$, $H_0 = 68\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $n_s = 0.97$, and $\sigma_8 = 0.81$ (Planck Collaboration et al. 2016a). We use the Sheth et al. (2001) halo mass function and all magnitudes are given in the AB system.

2 MODELLING Lyα VISIBILITY

To model the visibility of Lyα during the EoR, we combine simulations of the IGM from the Evolution of 21 cm Structure[1] (EOS) project (Mesinger et al. 2016) with empirical models of ISM properties. We have revised the treatment of the galaxy UV luminosity to halo mass relation from the methods described by M18a, though we refer the reader there for most details. Below, we provide a brief overview of the full model and a full description of our implementation of the galaxy UV luminosity–halo mass relation with scatter, which follows the conditional luminosity function modelling of Ren et al. (2019) (hereafter R19).

2.1 Lyα Transmission Through the ISM and IGM

We use the same simulations of the reionizing IGM and dark matter haloes as those used by M18a, obtained from the EOS simulations (Mesinger et al. 2016). The EOS simulations use 21CMFASTv2 (Sobacchi & Mesinger 2014), which incorporates sub-grid prescriptions for inhomogeneous recombinations and photoheating feedback, to simulate cubes (1.6 comoving Gpc on a side with resolution of 1024^3) of the IGM ionization state and 21 cm signal during Cosmic Dawn and the EoR (Mesinger et al. 2016). A specific reionization history is not assumed; rather, the ionization field at a given redshift is generated directly from a nonlinear density field assuming an ionizing efficiency parameter. Additionally, we note that we do not expect recombinations to considerably impact the morphology of reionization, since they only contribute significantly in the final $\lesssim 10$ per cent of reionization (Sobacchi & Mesinger 2015).

The ionizing efficiency is proportional to the product of the escape fraction of ionizing photons and the stellar mass fraction in a halo. Since the escape fraction likely decreases with increasing halo mass while the stellar mass fraction decreases, the ionizing efficiency is approximately constant when averaged over all haloes. Then, the ionizing efficiency is chosen to ensure that the simulated Thompson scattering optical depth is consistent with that measured by Planck Collaboration et al. (2016a). See Mesinger & Furlanetto (2007); Mesinger et al. (2011) and Mesinger et al. (2015) for further details.

To obtain various values of $\overline{x}_{\rm H\,I}$ at a $z \sim 7$, we apply ionization fields from different redshifts to the density field at $z \sim 7$, analogous to varying the escape fraction of ionizing photons and the timing of reionization (e.g. McQuinn et al. 2007; Jensen et al. 2014; Sobacchi et al. 2016). While this is not fully self-consistent, the topology of reionization is highly redshift-independent and instead depends primarily on the global neutral fraction (McQuinn et al. 2007).

The sources of reionization are taken to be star-forming galaxies hosted by dark matter haloes of some spatially-dependent minimum mass, $M_{\rm min}(\boldsymbol{x}, z)$. The minimum mass is determined by taking the maximum of the halo mass corresponding to the virial temperature required for cooling by atomic hydrogen ($M_{\rm cool}$, assuming a virial temperature of $T_{\rm vir} \gtrsim 2 \times 10^4\,{\rm K}$), the halo mass required to form stars inside H II regions ($M_{\rm photo}$), and the halo mass determined

by supernova feedback ($M_{\rm SNe}$), since more efficient feedback suppresses star formation for higher-mass haloes (Mesinger et al. 2016).

The role of supernova feedback in star formation during reionization is largely unconstrained, so Mesinger et al. (2016) take $M_{\rm SNe}$ to be a free parameter and choose two extreme values to approximately encompass the parameter space of the sources that drive reionization. In the 'Faint Galaxies' run, supernova feedback is inefficient, $M_{\rm min}(\boldsymbol{x}, z)$ is fixed by $M_{\rm cool}$ and $M_{\rm photo}$, and the primary drivers of reionization are low-mass galaxies of halo mass $\sim 10^8 - 10^9\,{\rm M}_\odot$ (see Figure 1 of Mesinger et al. 2016) that produce small ionized regions. In contrast, the 'Bright Galaxies' run corresponds to extremely efficient supernova feedback and is characterised by more massive galaxies (halo mass $\sim 10^{10}\,{\rm M}_\odot$) that produce larger ionized patches. In this work, we choose the 'Faint Galaxies' run as our fiducial model, as the difference in the inferred volume-averaged neutral fraction between the two is negligible for current observational samples (Mason et al. 2018a,b; Greig et al. 2019).

We populate the simulated dark matter haloes (with mass $10^{10} \leq M_{\rm h}/{\rm M}_\odot \leq 10^{12}$) with UV magnitudes ($M_{\rm UV}$), and Lyα line profiles and emitted Lyα equivalent widths (EWs) drawn from empirical models, and calculate the observed Lyα EW after transmission along sightlines through the reionizing IGM. We describe our method for assigning UV magnitudes, updated from the models of M18a, in Section 2.2. Otherwise, we follow the procedure presented by M18a in section 2.1 (and which we briefly outline here) to supply the haloes with emitted Lyα EWs in the rest frame and Lyα line profiles.

To model the observed Lyα EW distribution at $z \sim 7$, we first assume that the observed $z \sim 6$ EW distribution is equal to the emitted $z \sim 7$ EW distribution. That is, any change in the observed Lyα EW distribution between $z \sim 7$ and $z \sim 6$ is entirely due to reionization and not evolution of galactic properties such as dust. This is likely a simplification, as lower-redshift galaxies show a trend towards increasing Lyα EWs with increasing redshift as dust fraction decreases (Hayes et al. 2011). However, the time-scale between $z \sim 7$ and $z \sim 6$ is short ($\lesssim 200$ Myr) and we do not expect significant change (Treu et al. 2012). Moreover, if the intrinsic emitted Lyα EW distribution does evolve significantly between $z \sim 7$ and $z \sim 6$, it will likely be towards higher EWs at $z \sim 7$ due to decreasing dust (Hayes et al. 2011), requiring a higher neutral fraction than we find to suppress the modelled Lyα EW distribution to that which is observed. Given this caveat (discussed further in Section 4.1), we proceed with the model for the emitted Lyα EW distribution at $z \sim 7$, $p({\rm EW}_{\rm emit} \mid M_{\rm UV})$, derived by M18a from a fit to the observed $z \sim 6$ sample by De Barros et al. (2017) (the largest homogeneous sample of $z \sim 6$ galaxies with deep spectroscopic observations probing the presence of Lyα for ${\rm EW}_{\rm Ly\alpha} \geq 25$ Å).

We use the differential Lyα transmission fraction, $\mathcal{T}_{\rm IGM}$, to compare the Lyα EW distributions at $z \sim 7$ and $z \sim 6$. $\mathcal{T}_{\rm IGM}$ is defined as the fraction of emitted Lyα photons that are transmitted through the IGM, $\mathcal{T}_{\rm IGM} = {\rm EW}_{\rm obs}/{\rm EW}_{\rm emit}$, and can be calculated as

$$\mathcal{T}_{\rm IGM}(\overline{x}_{\rm H\,I}, M_{\rm h}, \Delta v) = \int_{-\infty}^{\infty} {\rm d}v\, J_{\rm Ly\alpha}(\Delta v, M_{\rm h}, v)\, {\rm e}^{-\tau_{\rm IGM}(\overline{x}_{\rm H\,I}, M_{\rm h}, v)},$$

---

[1] http://homepage.sns.it/mesinger/EOS.html

$$\tag{1}$$

where $\mathcal{T}_{\mathrm{IGM}}$ is normalised to $\mathcal{T}_{\mathrm{IGM}}(z=6)=1$. $J_{\mathrm{Ly}\alpha}$ is the line profile of Lyα photons escaping from galaxies as a function of velocity, v, including the velocity offset of the Lyα line from the systemic velocity of the galaxy (Δv, see below), and τ_{IGM} is the optical depth to Lyα photons through the IGM. τ_{IGM} is comprised of two components: the Lyα damping wing absorption in the partially neutral IGM, $\tau_{\mathrm{D}}(\bar{x}_{\mathrm{H\,I}}, M_{\mathrm{h}}, v)$, and resonant absorption due to residual neutral gas in the local ionized region around the source, $\tau_{\mathrm{H\,II}}(M_{\mathrm{h}}, v)$ (see e.g. Dijkstra 2014, for a review). While τ_{D} only arises during reionization, $\tau_{\mathrm{H\,II}}$ can be significant at all redshifts, and increases with redshift due the increasing density of gas (Laursen et al. 2011). In addition, infall of over-dense gas around haloes causes significant resonant absorption of Lyα photons on the red side of line center (Santos 2004; Dijkstra et al. 2007; Laursen et al. 2011; Weinberger et al. 2018). In this work, following Santos (2004) and M18a, we model $\tau_{\mathrm{H\,II}}$ as a step function, truncated below the halo's circular velocity to model resonant absorption in the infalling circumgalactic medium (CGM) and the ionized IGM.

Dijkstra et al. (2011) first showed that radiative transfer effects within galaxies, particularly scattering in galactic outflows, strongly impacts the emerging Lyα line profile, which in turn affects the transmission of Lyα through the reionizing IGM. In this work, we model radiative transfer in the ISM using the methods of M18a, which models the line shape of Lyα emerging from the ISM as a Gaussian centred at a velocity offset, Δv, from the systemic redshift of the source. In our empirical model, Δv and halo mass are correlated, as lower-mass galaxies are expected to have a lower column density of neutral hydrogen than high-mass galaxies (e.g. Yang et al. 2017), so scattering in the ISM is decreased (e.g., Neufeld 1990). Additionally, Δv is often correlated with the outflow velocity (Verhamme et al. 2006, 2008), and thus the halo's circular velocity, which is also regulated by the halo mass. We note that in this formulation, we have assumed that the evolution of $\mathcal{T}_{\mathrm{IGM}}$ is only due to the changing Lyα damping wing optical depth. That is, we do not model any redshift evolution (at fixed M_{h}) of absorption by residual neutral hydrogen in the ionized component of the IGM and CGM, or due to evolving Δv. See Section 4.1 for further discussion.

We compute Lyα optical depths, τ_{IGM}, for $\sim 10^4$ sightlines through the simulated IGM for a range of neutral fractions from $0.01 \leq \bar{x}_{\mathrm{H\,I}} \leq 0.95$ with $\Delta\bar{x}_{\mathrm{H\,I}} \approx 0.02$ (see M18a for details) and evaluate $\mathcal{T}_{\mathrm{IGM}}$ with Equation 1. We then calculate the observed Lyα EW, $\mathrm{EW}_{\mathrm{obs}}$, as $\mathrm{EW}_{\mathrm{obs}} = \mathcal{T}_{\mathrm{IGM}} \times \mathrm{EW}_{\mathrm{emit}}$, where $\mathrm{EW}_{\mathrm{emit}}$ is drawn from the M_{UV}-dependent emitted Lyα EW distribution for 10^5 simulated galaxies.

Using our sample of $\mathcal{T}_{\mathrm{IGM}}$ values, we calculate probability density functions for $\mathcal{T}_{\mathrm{IGM}}$ given M_{UV}, $\bar{x}_{\mathrm{H\,I}}$, and scatter in the UV luminosity–halo mass relation (σ) (discussed further in Section 2.2): $p(\mathcal{T}_{\mathrm{IGM}} \mid M_{\mathrm{UV}}, \bar{x}_{\mathrm{H\,I}}, \sigma)$. We calculate $p(\mathcal{T}_{\mathrm{IGM}} \mid M_{\mathrm{UV}}, \bar{x}_{\mathrm{H\,I}}, \sigma)$ on a grid of neutral fractions and UV magnitudes ($-21.9 \leq M_{\mathrm{UV}} \leq -16$, $\Delta M_{\mathrm{UV}} = 0.1$). To do this, we first evaluate $p(\mathcal{T}_{\mathrm{IGM}} \mid M_{\mathrm{h}}, \bar{x}_{\mathrm{H\,I}})$ on a grid of neutral fractions and halo masses ($10.2 \lesssim \log M_{\mathrm{h}} \lesssim 11.9$, $\Delta \log M_{\mathrm{h}} \approx 0.1$), and map from halo mass to UV magnitude using the method described below.

2.2 Incorporating Scatter in the UV Luminosity–Halo Mass Relation

To model Lyα emission from galaxies, we map from observed UV luminosity to emitted Lyα EW and differential Lyα transmission fraction, based on dark matter halo mass. This requires modelling the relationship between halo mass and UV luminosity. M18a use the one-to-one model described by Mason et al. (2015) (which successfully reproduces $0 \lesssim z \lesssim 10$ UV luminosity functions assuming star formation rates are proportional to halo mass accretion rates modulo a mass-dependent but redshift-independent efficiency) to map from halo mass to UV luminosity. We now include scatter in the galaxy UV luminosity–halo mass relation using the conditional luminosity function (CLF) approach of R19, which introduces scatter in M_{UV} given a halo mass, M_{h}.

The CLF presented by R19 is a log-normal distribution of galaxy UV luminosities, L, given a host halo mass, M_{h}, with a median UV luminosity–halo mass relation, L_c, and a log-normal dispersion, $\sigma_{\log L}$. We have converted the CLF of R19 from log-normal in luminosity to normal in magnitude, so $\log L$ becomes M_{UV}, $\log L_c$ becomes $M_{\mathrm{UV},c}$, and the standard deviation is $\sigma = \sigma_{M_{\mathrm{UV}}} = 2.5\,\sigma_{\log L}$. Thus, our CLF is as follows:

$$p(M_{\mathrm{UV}} \mid M_{\mathrm{h}}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-[M_{\mathrm{UV}} - M_{\mathrm{UV},c}(M_{\mathrm{h}}, \sigma, z)]^2}{2\sigma^2} \right). \tag{2}$$

The dispersion, σ, was originally introduced to explain scatter observed in the Tully–Fisher relation (Yang et al. 2005), and is a free parameter in the CLF.

The standard UV luminosity function, $\Phi(M_{\mathrm{UV}})$, is derived from the CLF by convolving with the halo mass function:

$$\Phi(M_{\mathrm{UV}}) = \int_0^\infty \mathrm{d}M_{\mathrm{h}}\, p(M_{\mathrm{UV}} \mid M_{\mathrm{h}}) \frac{\mathrm{d}n}{\mathrm{d}M_{\mathrm{h}}}, \tag{3}$$

where $p(M_{\mathrm{UV}} \mid M_{\mathrm{h}})$ is given by Equation 2 and $\mathrm{d}n/\mathrm{d}M_{\mathrm{h}}$ is the halo mass function (we use the Sheth et al. (2001) halo mass function). The luminosity function is calibrated to the observed UV luminosity function at $z \sim 5$ (Bouwens et al. 2015) and extrapolated based on the galaxy evolution model of Mason et al. (2015). For $\sigma > 0$, the median UV luminosity to halo mass relation is then a function of halo mass, standard deviation (σ), and redshift, and a critical flattening threshold is imposed based on either mass or luminosity. The flattening threshold can be linked to feedback processes by active galactic nuclei; an extended discussion is given in R19.

The degree of scatter, σ, is a free parameter in the CLF model. However, R19 explored a range of values for σ and found that $\sigma = 0.5$ mag is consistent with the observed luminosity functions for $z = 6-10$ in the framework of the Mason et al. (2015) UV luminosity function evolution model. Figure 4 of R19 shows a comparison of the modelled luminosity functions with scatter of $\sigma = 0, 0.5, 1$ mag for selected redshifts from $z \sim 6 - 12$, demonstrating that both no scatter and scatter of $\sigma = 0.5$ mag are consistent with observations, while $\sigma = 1$ mag overpredicts the bright end. Thus, we use $\sigma = 0.5$ mag as our fiducial model to ensure that we reproduce the observed UV luminosity function at $z \sim 7$.

The introduction of scatter into the Mason et al. (2015) evolution model with the CLF requires a critical flattening
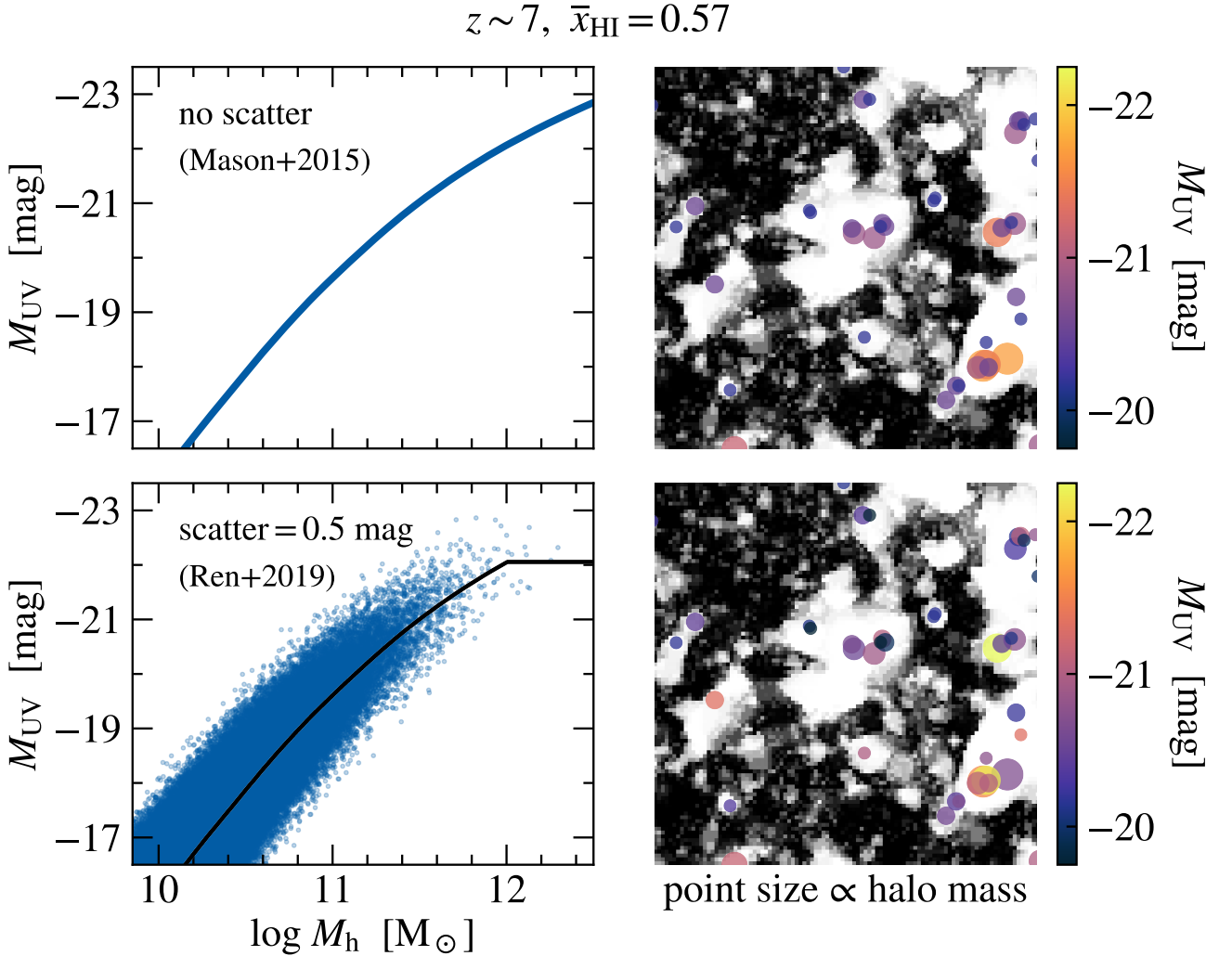
Figure 1. The spatial distribution of dark matter haloes in a $150 \times 150 \times 5$ Mpc slice of the reionizing IGM at $z \sim 7$ (with volume-averaged fraction of neutral hydrogen of $\bar{x}_{\mathrm{H\,I}} = 0.57$), and the UV magnitudes of their corresponding galaxies with and without scatter in the UV magnitude to halo mass relation. Left: The UV magnitude–halo mass relations at $z \sim 7$. The relation without scatter (top) is derived from the technique based on abundance matching described by Mason et al. (2015), and the relation with scatter of $\sigma = 0.5$ mag (bottom) follows the conditional luminosity function presented by Ren et al. (2019), where the solid black line is the median UV magnitude–halo mass relation. Right: The spatial distribution of haloes (point size is proportional to halo mass) in the reionizing IGM with $\bar{x} = 0.57$ (white is ionized and black is neutral), simulated with 21cmFAST (Mesinger & Furlanetto 2007; Mesinger et al. 2011). For clarity, only haloes of mass $M_{\mathrm{h}} \geq 10^{11} \, \mathrm{M_\odot}$ are shown. The galaxy UV magnitudes, assigned with the UV magnitude–halo mass relation in the corresponding left panel, are shown as the colour scale. Without scatter, the most massive haloes in the largest ionized bubbles host the most UV-bright galaxies. In contrast, lower-mass haloes in smaller ionized regions can host more UV-luminous galaxies when scatter is included.

threshold to maintain consistency with observations. This flattening can be regulated by either a critical mass threshold or a critical luminosity threshold, and the predicted luminosity functions are consistent with the observed luminosity functions for $z = 6 - 10$ for both (see figure 4 of R19). We tested the impact on the inference in Section 3.2 of imposing a critical luminosity flattening threshold instead of a critical mass threshold, and found the difference in the inferred neutral fraction was negligible. This is expected, as at $z \sim 7$, there is little difference in the median UV luminosity–halo mass relation between the two flattening criteria (R19).

In this work, we choose to use the critical mass flattening threshold, though either criteria would suffice since the modelled UV luminosity function and our inferred neutral fraction is insensitive to the choice of flattening threshold.

To implement the scatter, we use the HMF[2] package (Murray et al. 2013) to sample millions of haloes ranging from $\sim 10^{10} - 10^{12.5} \, \mathrm{M_\odot}$ from the Sheth et al. (2001) halo mass function at $z \sim 7$, and assign each halo a UV magnitude

---

[2] https://github.com/steven-murray/hmf

according to the CLF (Equation 2; R19). Fig. 1 shows the spatial distribution of haloes in a $150 \times 150 \times 5\,\mathrm{Mpc}$ slice of the partially reionized IGM at $z \sim 7$, simulated with 21cm-fast[3] (Mesinger & Furlanetto 2007; Mesinger et al. 2011), where the ionizing efficiency is set to obtain an average neutral fraction of $\overline{x}_{\mathrm{H\,I}} \approx 0.60$ at $z \sim 7$. We note this is only an illustration of our model, as the EOS simulations we use for this work are larger (1.6 comoving Gpc on a side) and include prescriptions for inhomogeneous recombinations and photoheating feedback; see section 2.1 of Mesinger et al. (2016). We also show the UV magnitudes of the galaxies hosted by the haloes, assigned with and without scatter, as the colour scale. Without scatter, UV-bright galaxies reside in high-mass haloes surrounded by large bubbles of ionized hydrogen. Conversely, with scatter, UV-bright galaxies can be hosted in lower-mass haloes surrounded by smaller ionized regions.

2.2.1 *Lyα Transmission Probability*

As M_{UV} is a real observable, we wish to obtain $p(\mathcal{T}_{\mathrm{IGM}} \mid M_{\mathrm{UV}}, \overline{x}_{\mathrm{H\,I}}, \sigma)$ (where we have chosen $\sigma = 0.5\,\mathrm{mag}$), but the outputs of our model are $p(\mathcal{T}_{\mathrm{IGM}} \mid M_{\mathrm{h}}, \overline{x}_{\mathrm{H\,I}})$ (as described in Section 2.1). Accordingly, we marginalise over halo mass to calculate probability distributions of the differential Lyα transmission fraction, $\mathcal{T}_{\mathrm{IGM}}$, for fixed UV magnitude, M_{UV}, neutral fraction, $\overline{x}_{\mathrm{H\,I}}$, and scatter, σ:

$$p(\mathcal{T}_{\mathrm{IGM}} \mid M_{\mathrm{UV}}, \overline{x}_{\mathrm{H\,I}}, \sigma) = \tag{4}$$
$$\int_0^\infty \mathrm{d}M_{\mathrm{h}}\, p(\mathcal{T}_{\mathrm{IGM}} \mid M_{\mathrm{h}}, \overline{x}_{\mathrm{H\,I}})\, p(M_{\mathrm{h}} \mid M_{\mathrm{UV}}, \sigma),$$

where $p(\mathcal{T}_{\mathrm{IGM}} \mid M_{\mathrm{h}}, \overline{x}_{\mathrm{H\,I}})$ is the probability distribution of $\mathcal{T}_{\mathrm{IGM}}$ given a halo mass, M_{h}, and neutral fraction, $\overline{x}_{\mathrm{H\,I}}$, as generated by M18a, and we use Equation 2 to find $p(M_{\mathrm{h}} \mid M_{\mathrm{UV}}, \sigma)$.

We use a numerical approach to calculate $p(M_{\mathrm{h}} \mid M_{\mathrm{UV}}, \sigma)$ from Equation 2 for $\sigma = 0.5\,\mathrm{mag}$. We create a two-dimensional histogram of $p(M_{\mathrm{UV}} \mid M_{\mathrm{h}})$ with bin sizes of $\Delta \log M_{\mathrm{h}} \approx 0.1$ to match the EOS simulation halo masses, and $\Delta M_{\mathrm{UV}} = 0.1$ to match our chosen M_{UV} grid. To obtain $p(M_{\mathrm{h}} \mid M_{\mathrm{UV}}, \sigma)$, we fix M_{UV} and extract the corresponding one-dimensional histogram.

Sample distributions of $p(M_{\mathrm{h}} \mid M_{\mathrm{UV}}, \sigma)$ for a range of UV magnitudes are shown in Fig. 2 (note that these distributions do not show relative number densities of different UV magnitudes). Bright galaxies are more likely than faint galaxies to reside in high-mass haloes, but with significant dispersion around the halo mass with peak probability. Additionally, there is slight broadening in the probability distributions from faint to bright galaxies, where bright galaxies may be hosted in a wider range of halo masses than fainter galaxies.

It should be noted that we have assumed that when scatter in the UV luminosity–halo mass relation is included, the distribution of galaxies within the IGM changes, but the IGM topology does not. This is reasonable if low-mass, faint galaxies dominate reionization, but if bright galaxies contribute significantly, the topology may change. Namely,
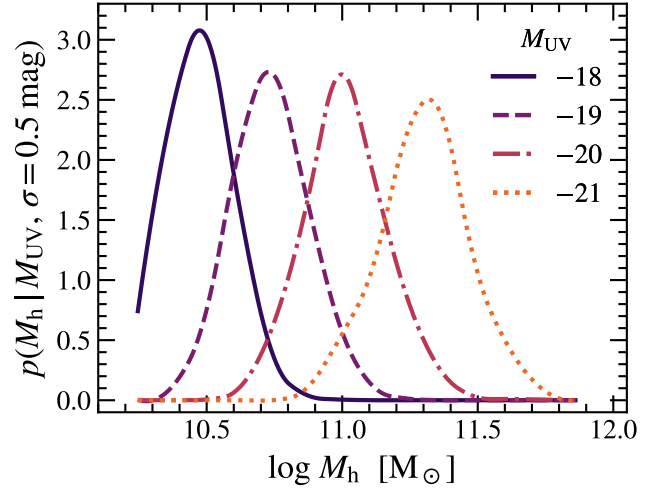
Figure 2. Normalised probability density functions of halo mass, M_{h}, given UV magnitude, M_{UV}, and scatter of $\sigma = 0.5\,\mathrm{mag}$ (note that these distributions do not show relative number densities of galaxies of various UV magnitudes, which follow the UV luminosity function). Colours and linestyles correspond to galaxies of UV magnitudes $M_{\mathrm{UV}} = -18, -19, -20, -21$, where darker colours indicate fainter galaxies. Though brighter galaxies are more likely than fainter galaxies to have high-mass hosts, there is significant scatter around their peak probability. There is also some broadening in the distributions, where bright galaxies may reside in a wider range of halo masses than faint ones.

bright galaxies in low-mass haloes may reside in larger ionized regions than we model. Thus, our results demonstrate the maximum expected deviation from the case without scatter. We discuss this assumption in more detail in Section 4.1.

3 RESULTS

In this section, we describe the key effects on Lyα visibility of modelling scatter in the UV luminosity to halo mass relation. In Section 3.1, we compare our results with scatter with the findings of M18a without scatter in the UV luminosity to halo mass relation. In Section 3.2, we infer the neutral fraction with observations presented by Pentericci et al. (2014) and compare with the neutral fraction inferred by M18a.

3.1 Lyα Visibility

In Fig. 3, we plot probability density functions for $\mathcal{T}_{\mathrm{IGM}}$ for two values of M_{UV}, mapping from M_{h} to M_{UV} with and without incorporating scatter in the UV luminosity–halo mass relation. To transform from M_{h} to M_{UV} without scatter, we use the one-to-one models of M18a, which in turn use the UV luminosity function model of Mason et al. (2015). To transform from M_{h} to M_{UV} with scatter, we use the CLF approach of R19 described in Section 2.2, where we have chosen a dispersion of $\sigma = 0.5\,\mathrm{mag}$. With scatter, Lyα transmission tends to decrease for UV-bright galaxies compared to the case without scatter.

Fig. 4 shows the probability distributions of observed
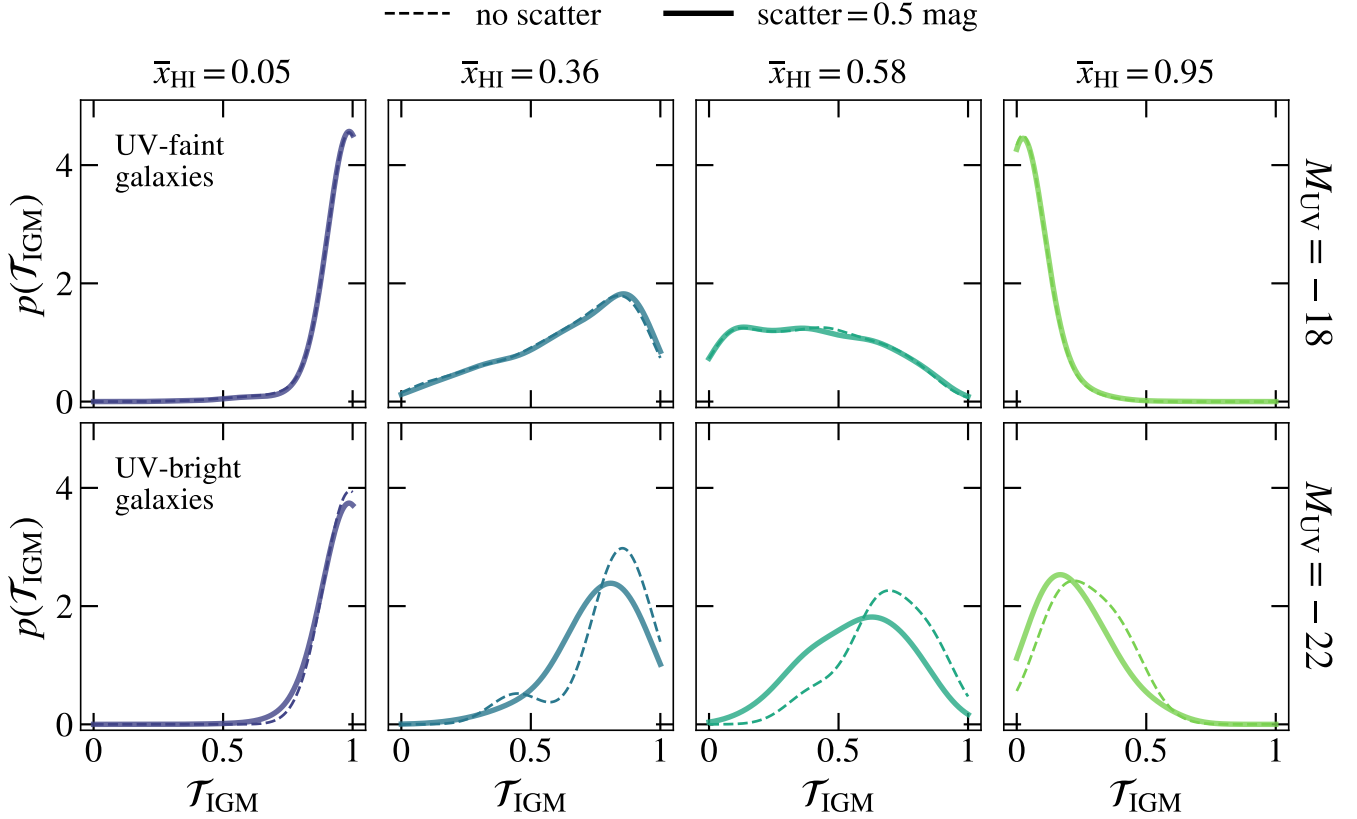
Figure 3. Comparison of the probability distributions of differential Lyα transmission fractions through the IGM, $\mathcal{T}_{\mathrm{IGM}}$, with scatter (solid) and without scatter (dashed) in the UV luminosity to halo mass relation at $z \sim 7$. Each column and line colour shows a different neutral fraction, where darker colours correspond to lower neutral fractions (i.e. a more ionized IGM). UV-faint ($M_{\mathrm{UV}} = -18$) galaxies are in the top row and UV-bright ($M_{\mathrm{UV}} = -22$) galaxies are in the bottom row. Introducing scatter tends to increase the probability of lower transmission, particularly for UV-bright galaxies.

Lyα EWs with and without scatter for faint and bright galaxies at one neutral fraction, $\overline{x}_{\mathrm{HI}} = 0.58$, calculated as described in Section 2.1. The Lyα EW distribution for UV-bright galaxies tends towards slightly lower EWs, reflecting the decreased Lyα transmission for bright galaxies seen in Fig. 3.

For a fixed neutral fraction and UV luminosity, high Lyα transmissions and EWs are less probable when scatter in the UV luminosity–halo mass relation is included, especially for more UV-luminous objects. In general, modelling scatter allows UV-bright objects to reside in less massive haloes, and so less overdense regions. Hence, on average, they are surrounded by smaller ionized regions than would be predicted by an abundance matching technique. This tends to increase the attenuation of Lyα by neutral hydrogen along the line of sight.

3.2 Inferred Neutral Fraction at $z \sim 7$

We infer the neutral fraction, $\overline{x}_{\mathrm{HI}}$, at $z \sim 7$ with the flexible Bayesian framework described by M18a. We provide a brief description below, but refer the reader to M18a for further details.

By Bayes' Theorem, we can write the posterior probability of $\overline{x}_{\mathrm{HI}}$ inferred from one observation of a galaxy with measured Lyα EW (or upper limit), EW_i, and UV magnitude, $M_{\mathrm{UV},i}$, given the UV magnitude–halo mass scatter, σ, as the following:

$$p(\overline{x}_{\mathrm{HI}} \mid EW_i, M_{\mathrm{UV},i}, \sigma) \propto p(EW_i \mid M_{\mathrm{UV},i}, \overline{x}_{\mathrm{HI}}, \sigma)p(\overline{x}_{\mathrm{HI}}), \quad (5)$$

where $p(EW_i \mid M_{\mathrm{UV},i}, \overline{x}_{\mathrm{HI}}, \sigma)$ is the *likelihood* of observing a Lyα EW (or upper limit) given our model of Lyα visibility (including our chosen scatter of $\sigma = 0.5$ mag in the UV luminosity to halo mass relation) and $p(\overline{x}_{\mathrm{HI}})$ is our *prior* on $\overline{x}_{\mathrm{HI}}$, which we take to be uniform between 0 and 1.

We use a non-analytic prescription to obtain the likelihood of observing our data (sets of $\{EW, M_{\mathrm{UV}}\}$ for galaxies at $z \sim 7$) given our model of Lyα transmission through the ISM and IGM, as the complex topology of the IGM prevents an analytic formulation. We use our models, described in Section 2.2, to generate large samples of mock observed Lyα EWs on a grid of neutral fractions, $\overline{x}_{\mathrm{HI}}$, and UV magnitudes, M_{UV}, and use a Gaussian Kernel Density Estimator (Rosenblatt 1956; Parzen 1962) to generate smooth estimates for $p(EW \mid M_{\mathrm{UV}}, \overline{x}_{\mathrm{HI}}, \sigma)$. We also account for uncertainties and upper limits of Lyα EW measurements as described by M18a. Finally, we combine the inference from
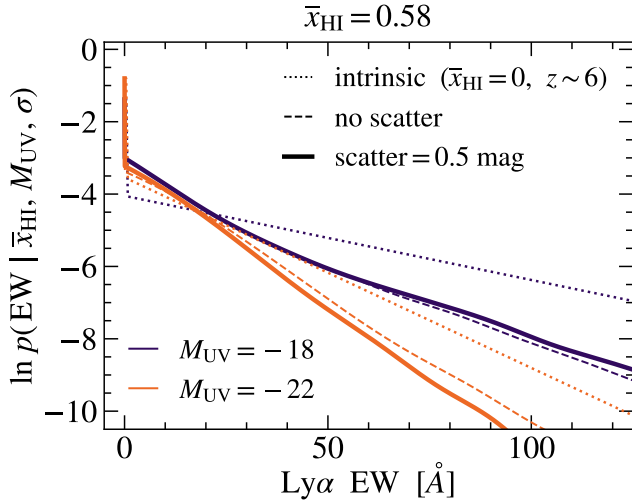
$$\overline{x}_{\rm HI} = 0.58$$

Figure 4. Probability distributions of Lyα EWs with (solid) and without (dashed) scatter of $\sigma = 0.5$ mag in the UV luminosity–halo mass relation for UV-faint and UV-bright galaxies at one neutral fraction, $\overline{x}_{\rm H\,I} = 0.58$. UV-faint ($M_{\rm UV} = -18$) galaxies are shown in purple and UV-bright ($M_{\rm UV} = -22$) galaxies are shown in orange. The intrinsic distributions in a fully ionized universe ($\overline{x}_{\rm H\,I} = 0$) at $z \sim 6$ are shown as dotted lines. With scatter, bright galaxies tend to slightly lower Lyα EWs than without scatter, reflecting the decrease in Lyα transmission with scatter (Fig. 3).

a set of $N_{\rm gals}$ uncorrelated observations by multiplying the individual posterior distributions:

$$p(\overline{x}_{\rm H\,I} \mid \{{\rm EW}, M_{\rm UV}\}, \sigma) \propto \prod_{i=1}^{N_{\rm gals}} p({\rm EW}_i \mid M_{{\rm UV},i}, \overline{x}_{\rm H\,I}, \sigma) p(\overline{x}_{\rm H\,I}).$$

(6)

We use this framework to infer the neutral fraction at $z \sim 7$ from the sample of Lyman-break galaxies (LBGs) presented by Pentericci et al. (2014). These data are comprised of 68 LBGs with spectroscopic follow-up in legacy fields, spanning a range of UV magnitudes $-22.75 \lesssim M_{\rm UV} \lesssim -17.8$, of which 12 had Lyα emission detected (Pentericci et al. 2014). M18a used the same sample for their inference, so we directly compare our new inferred neutral fraction with their result.

In Fig. 5, we show the posterior distributions of $\overline{x}_{\rm H\,I}$ inferred from the Pentericci et al. (2014) sample with and without accounting for scatter in the UV luminosity–halo mass relation. With scatter, we infer a neutral fraction of $\overline{x}_{\rm H\,I} = 0.55^{+0.11}_{-0.13}$ (the errors denote the 68 per cent credible interval). Compared to the neutral fraction inferred without scatter of $\overline{x}_{\rm H\,I} = 0.59^{+0.12}_{-0.14}$ (M18a), our result is slightly lower (by < 8 per cent) and largely consistent within the $\sim 20-25$ per cent uncertainties. However, it should be noted that though the two values are consistent within the uncertainties, we have used the same data (Pentericci et al. 2014) and simulated dark matter haloes (see Section 2.1) to infer the neutral fraction both with and without scatter. Thus, the change in the inferred neutral fraction is due exclusively to our updated method of computing the likelihood, as described in Section 2.2.
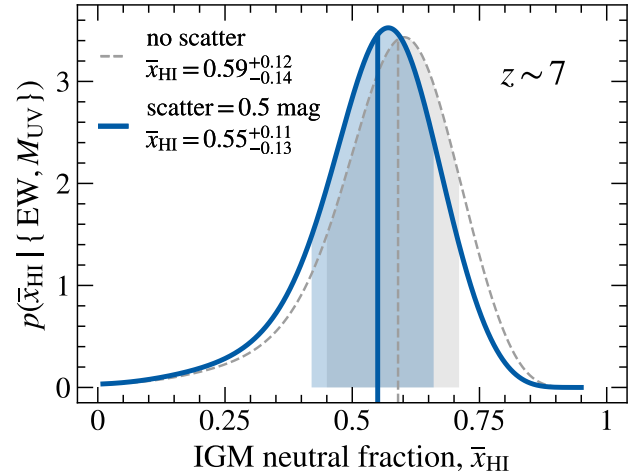
Figure 5. Posterior probability density functions for the volume-averaged fraction of neutral hydrogen in the IGM, $\overline{x}_{\rm H\,I}$, at $z \sim 7$ with scatter of $\sigma = 0.5$ mag (solid blue) and without scatter (dashed grey) in the UV luminosity–halo mass relation, inferred from the Pentericci et al. (2014) sample of LBGs. Vertical lines are the medians and the shaded regions are the 68 per cent credible intervals. With scatter, we infer a neutral fraction of $\overline{x}_{\rm H\,I} = 0.55^{+0.11}_{-0.13}$. Without scatter, the inferred neutral fraction is $\overline{x}_{\rm H\,I} = 0.59^{+0.12}_{-0.14}$ (M18a). Including scatter in the UV luminosity–halo mass relation slightly decreases the median neutral fraction inferred without scatter (i.e. we infer a more ionized IGM), but does not significantly affect the width of the posteriors.

Including scatter in the UV luminosity–halo mass relation tends to decrease Lyα visibility for galaxies of given UV luminosity at a fixed neutral fraction. This is especially true for UV-bright galaxies while reionization is ongoing, and the universe is neither extremely neutral nor extremely ionized; see Fig. 3. Thus, as scatter reduces the observed Lyα EW distribution, a somewhat lower neutral fraction must be invoked to explain the decline in the observed Lyα EWs at $z \gtrsim 6$ when scatter is included, compared to the inference without scatter.

Fig. 6 shows our result in the context of the redshift evolution of the volume-averaged fraction of neutral hydrogen in the IGM, $\overline{x}_{\rm H\,I}$. We show measurements derived from observations of the evolution of the Lyα EW distribution without scatter (Mason et al. 2018a, 2019a; Hoag et al. 2019), the clustering of Lyα emitters (Ouchi et al. 2010; Sobacchi & Mesinger 2015), Lyα and Lyβ forest dark pixel fraction (Mc-Greer et al. 2015), and quasar damping wings (Davies et al. 2018; Greig et al. 2019). We also show the 68 and 95 per cent credible intervals of the reionization history obtained from fitting the Planck Collaboration et al. (2018) CMB optical depth and dark pixel fraction constraints (Mason et al. 2019b).

3.3 Biases in the Neutral Fraction Inference

Future observational campaigns with instruments such as the *Wide Field Infrared Survey Telescope* may access statistical samples of Lyα from UV-bright LBGs at $z \gtrsim 7$ (Spergel et al. 2015). Thus, we simulate luminosity-selected samples
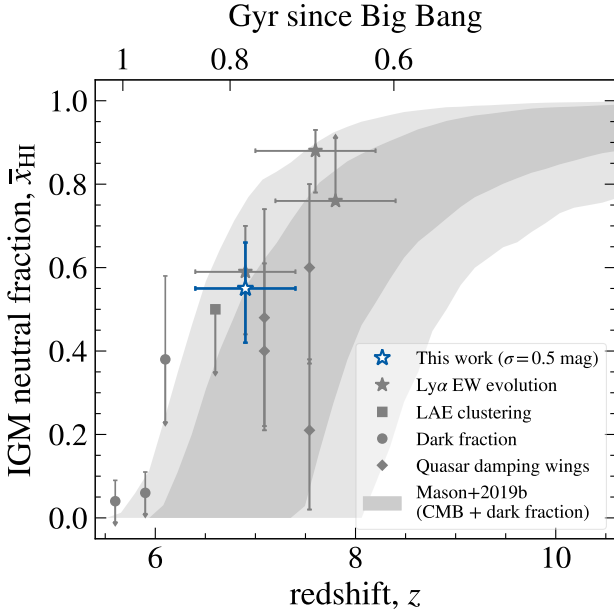
Figure 6. The redshift evolution of the volume-averaged fraction of neutral hydrogen in the IGM, $\overline{x}_{\rm HI}$. Our new measurement is shown as the blue star. We also plot measurements derived from observations of the evolving Lyα EW distribution without accounting for UV luminosity–halo mass relation scatter (grey filled stars; Mason et al. 2018a, 2019a; Hoag et al. 2019), the clustering of Lyα emitting galaxies (square; Ouchi et al. 2010; Sobacchi & Mesinger 2015), Lyα and Lyβ forest dark pixel fraction (circles; McGreer et al. 2015), and quasar damping wings (diamonds; Davies et al. 2018; Greig et al. 2019). We also plot the 68 and 95 per cent credible intervals of the reionization history obtained from fitting the Planck Collaboration et al. (2018) CMB optical depth and dark pixel fraction constraints (grey shaded regions; Mason et al. 2019b).

to investigate how not accounting for scatter in the UV luminosity–halo mass relation may bias the inferred neutral fraction. We simulate Lyα observations from galaxies, including UV luminosity–halo mass scatter, and compare the neutral fraction inferred with and without accounting for scatter.

We draw samples of galaxies from the UV luminosity function (Mason et al. 2015) and populate their Lyα EWs from the probability distribution $p({\rm EW} \mid M_{\rm UV}, \overline{x}_{\rm HI}, \sigma)$ with scatter of $\sigma = 0.5$ mag at a fixed neutral fraction, $\overline{x}_{\rm HI} = 0.49$. We also add Gaussian noise with a standard deviation of 5 Å to the Lyα EWs. We draw from the UV luminosity function with two limiting magnitudes (one bright, $M_{\rm UV} \leq -21.8$, and one faint, $M_{\rm UV} \leq -19.5$), each with two different sample sizes. We then perform the inference on the simulated samples with and without accounting for scatter in the UV luminosity–halo mass relation and compare the resulting posterior distributions (see Appendix A for further details).

For the bright sample, without accounting for scatter, the inferred neutral fraction is biased to slightly higher values than the simulated input neutral fraction, $\overline{x}_{\rm HI} = 0.49$, while we recover the correct neutral fraction when scatter is accounted for. While the input neutral fraction is recovered within the uncertainty both with and without scatter for the

smaller sample, the inference without scatter for the large, bright sample becomes inconsistent with the input neutral fraction as the precision on the measurement increases. See Appendix A for further discussion.

4 DISCUSSION

In this section, we consider caveats of our model and discuss implications of our results for future high-redshift galaxy studies. In Section 4.1, we discuss modelling caveats. We discuss implications for the clustering of Lyα emitters during reionization in Section 4.2, and implications for Lyα in UV-bright galaxies during the EoR in Section 4.3. Finally, we examine the potential for constraining the scatter in the CLF in Section 4.4.

4.1 Modelling Caveats

In this work, we make several significant assumptions. First, following M18a, we assume that the intrinsic emitted (i.e. after processing the ISM) Lyα EW distribution at $z \sim 7$ is the same as the observed Lyα EW distribution at $z \sim 6$ (see Section 2.1). This is likely a simplification, as trends at lower redshifts show an evolution towards higher Lyα EWs with redshift as dust fraction decreases (Hayes et al. 2011). However, this assumption also likely results in a lower inferred neutral fraction than if the Lyα EW distribution at $z \sim 7$ is not the same as the observed Lyα EW distribution at $z \sim 6$. If the emitted Lyα EW distribution evolves significantly in the short time between $z \sim 7$ and $z \sim 6$ ($\lesssim 200$ Myr), it will probably continue to follow the lower-redshift trend and have higher Lyα EWs at $z \sim 7$. This would require a larger neutral fraction than we infer to suppress our modelled Lyα EW distribution to the observed Lyα EW distribution. However, we note that the evolution of the intrinsic Lyα EW distribution is not fully understood: De Barros et al. (2017) found tentative evidence for decreasing Lyα EWs between $z \sim 5 - 6$ which could be due to intrinsic evolution, or an indication there is still neutral gas in some IGM sightlines at $z \sim 5 - 6$ (as suggested by an extended Gunn-Peterson trough at $z \sim 5.5$; Becker et al. 2015; Keating et al. 2019).

We also consider only the distribution of galaxies within the IGM and do not model any change in the IGM topology when scatter in the UV luminosity–halo mass relation is included. This is reasonable if low-mass, faint galaxies dominate reionization, as including scatter has a smaller impact on these galaxies. However, if brighter galaxies do contribute significantly to reionization, scatter could change the IGM topology. For example, UV-bright galaxies in low-mass haloes may be bright due to a burst of star formation, and ionizing photons produced in that burst could more easily escape from a low-mass halo than from a high-mass halo due to the smaller potential well (e.g. Wise et al. 2014; Paardekooper et al. 2015). Hence, UV-bright galaxies in low-mass haloes could reside in larger ionized regions than we consider, particularly in the earliest stages of reionization; in later stages, the EoR topology depends primarily on the cumulative ionizing contribution from all sources. Thus, here we model the maximum expected difference in Lyα visibility from the UV luminosity–halo mass relation without scatter.

Accounting for the changing IGM topology would likely increase Lyα visibility, and therefore the IGM neutral fraction, compared to our results in this work.

Future improvements to the model to account for these assumptions would, on average, most likely result in higher observed Lyα EWs at fixed neutral fraction (particularly for UV-bright galaxies), and therefore we would infer a higher neutral fraction than we find in this work.

Finally, this work assumes that the evolution of Lyα visibility between $z \sim 7$ and $z \sim 6$ is due to the evolution of the Lyα damping wing optical depth of the diffuse neutral component of the IGM. We do not model redshift evolution of Lyα transmission in the ionized component of the IGM and CGM at fixed halo mass, which can depend strongly on the velocity offset from the systemic redshift of the source (e.g. Dijkstra et al. 2011; Laursen et al. 2011; Choudhury et al. 2015). This assumption is reasonable if the Lyα line profiles emerging from LBGs are dominated by scattering off outflows (e.g., Verhamme et al. 2008), which produces high Δv (as at lower redshifts, e.g. Steidel et al. 2010; Hashimoto et al. 2013; Shibuya et al. 2014). However, if Lyα velocity offsets decrease significantly at fixed halo mass at $z > 6$, decrease in the transmission of Lyα through the ionized IGM and CGM may be significant (Dijkstra et al. 2011; Choudhury et al. 2015), and lessen the requirement for a rapid evolution in $\overline{x}_{\mathrm{H\,I}}$. We will explore the impact on $\overline{x}_{\mathrm{H\,I}}$ inferences of redshift-dependent velocity offsets and the evolution of resonant absorption in the local ionized region around the source, $\tau_{\mathrm{H\,II}}$, in future work.

4.2 Clustering of Lyα Emitters

In this work, we examine the evolution of the observed Lyα EW distribution between $z \sim 6$ and $z \sim 7$ due to absorption by neutral hydrogen in the IGM during the EoR. However, the clustering signal of Lyα emitters (LAEs, selected on their Lyα flux rather than UV continuum) is also sensitive to attenuation by neutral hydrogen; Lyα emission is preferentially transmitted through the large ionized bubbles surrounding clusters of LAEs, therefore increasing the clustering signal (e.g. Sobacchi & Mesinger 2015; Weinberger et al. 2019).

Recent measurements of the LAE angular correlation function by Ouchi et al. (2018) show no significant evolution from $z \sim 5.7$ to $z \sim 6.6$, a somewhat unexpected result given the expected amplification of the clustering signal during reionization. Now, with scatter in the UV luminosity–halo mass relation, galaxies are hosted by lower mass halos on average, thus the probability of a low-mass halo hosting an LAE increases, implying some weakening of the LAE clustering signal. However, this depends on the relationship between M_{UV} and Lyα flux: if LAEs are biased towards UV-luminous galaxies, the decrease in clustering due to UV luminosity–halo mass scatter may be negligible. We leave quantitative study of the effects of scatter in the UV luminosity–halo mass relation on LAE clustering for future work.

4.3 Lyα in UV-bright Galaxies During the EoR

Recent detections of Lyα at $z \gtrsim 7.5$ from UV-bright galaxies (e.g. Oesch et al. 2015; Zitrin et al. 2015; Roberts-Borsani et al. 2016; Stark et al. 2017) are somewhat surprising, since at lower redshifts, UV-bright galaxies are unlikely to have strong Lyα emission (e.g. Stark et al. 2010). The inhomogeneous nature of reionization likely accounts for some of this apparent increase in Lyα from UV-bright galaxies, due to higher Lyα transmission through the reionizing IGM from bright galaxies residing in large ionized bubbles (Fig. 3). However, Mason et al. (2018b) found that merely increased Lyα transmission through the IGM for extremely UV-bright galaxies ($M_{\mathrm{UV}} \lesssim -22$) is insufficient to explain the evolution of the Lyα fraction for these galaxies between $6 \lesssim z \lesssim 8$ (Stark et al. 2017). This suggests that the emitted Lyα EWs from these bright galaxies are higher than expected. In this work, we find a decrease in Lyα transmission for UV-bright galaxies from the models of M18a (Fig. 3), implying that the emitted Lyα EWs of the Roberts-Borsani et al. (2016); Stark et al. (2017) sample are even higher than previously required by Mason et al. (2018b) to explain their observed emission. The implied high intrinsic EW of this sample, relative to other galaxies of a similar UV luminosity, is likely due to atypical intrinsic galactic properties, selection effects, and/or that these sources have hard ionizing spectra which may have highly ionized their surroundings, reducing the optical depth to Lyα in the ionized IGM (see Stark et al. 2017; Mason et al. 2018b, for further discussion).

Our analysis in Section 3.3 and Appendix A indicates that using UV-bright galaxies ($M_{\mathrm{UV}} \lesssim -22$) to constrain $\overline{x}_{\mathrm{H\,I}}$ *without* accounting for UV luminosity–halo mass scatter can lead to biased measurements of $\overline{x}_{\mathrm{H\,I}}$. As the magnitude of the scatter itself is uncertain (see below), extremely UV-bright galaxies should only be used as probes of the IGM if the impact of UV luminosity–halo mass scatter is marginalised over. However, it should be noted that the scatter is only a significant systematic uncertainty for large samples ($N \gtrsim 1000$) of UV-bright galaxies. For samples that include fainter galaxies, the neutral fraction is recovered well with or without scatter, even for samples as large as $N = 1000$.

4.4 Constraining UV Luminosity–Halo Mass Scatter

The degree of scatter in the CLF is a free parameter, and we chose $\sigma = 0.5\,\mathrm{mag}$ as our fiducial value, since R19 found that this scatter is consistent with observations at $z \sim 6-10$.

To maintain consistency with observations, R19 also applied a critical flattening threshold to the median UV magnitude–halo mass relation based on either mass or luminosity, which can be related to feedback processes by active galactic nuclei. R19 explored alternative values of σ with both mass and luminosity flattening thresholds, and found that for high-scatter cases ($\sigma > 0.5\,\mathrm{mag}$), scatter and flattening criteria were degenerate when using only the UV luminosity function to discriminate between cases. This degeneracy arises because the CLF does not fully recover the

brightest end of the observed luminosity function[4] (see R19 for further discussion).

R19 suggest that an independent constraint on the scatter, such as measurements of the local clustering strength around UV-bright galaxies, would break this degeneracy. Alternatively, a precise measurement of the EoR timeline could constrain the degree of the scatter. The transmission of Lyα through the IGM is sensitive to variations in the column density of neutral hydrogen along the line of sight, which in turn depends on the distribution of the ionizing sources (i.e. galaxies). Thus, our model discriminates between larger and smaller scatter, though it is largely insensitive to the flattening threshold at $z \sim 7$. However, determining the correct scatter would require a precision measurement of the EoR timeline by another method (for example, via measurements of the redshifted 21 cm line from neutral hydrogen).

5 SUMMARY AND CONCLUSIONS

We have updated a model of galaxy properties during the EoR to include physically-motivated scatter in the relation between galaxy UV luminosity and dark matter halo mass. We have modelled Lyα visibility during the EoR, inferred the fraction of neutral hydrogen in the IGM at $z \sim 7$ with the Pentericci et al. (2014) sample of LBGs, and compared with results from models without scatter in the UV luminosity–halo mass relation (M18a). Our primary conclusions are as follows:

(i) Scatter in the galaxy UV luminosity to halo mass relation alters the distribution of galaxies within the IGM, and thus impacts Lyα visibility from galaxies.

(ii) Scatter reduces Lyα visibility for very UV-luminous objects. This is due to the increased probability of a UV-bright galaxy residing in a lower-mass halo surrounded by a smaller ionized region, thus increasing the probability of absorption by neutral hydrogen along the line of sight. This effect is less significant for fainter galaxies, as they are much more numerous (so introducing outliers has a smaller effect) and already reside in low-mass haloes surrounded by small ionized regions.

(iii) Scatter slightly decreases the median neutral fraction inferred at $z \sim 7$ from the Pentericci et al. (2014) sample of 68 LBGs compared to the inference without accounting for scatter. With scatter, we infer a neutral fraction of $\overline{x}_{\mathrm{H\,I}} = 0.55^{+0.11}_{-0.13}$, compared to the neutral fraction inferred without scatter of $\overline{x}_{\mathrm{H\,I}} = 0.59^{+0.12}_{-0.14}$, (M18a). However, scatter does not impact the width of the probability distribution for the neutral fraction.

(iv) Using samples of UV-bright galaxies to constrain $\overline{x}_{\mathrm{H\,I}}$ without accounting for scatter in the UV luminosity–halo

mass relation leads to overestimates of $\overline{x}_{\mathrm{H\,I}}$. The discrepancy becomes more significant for larger sample sizes.

More refinements to our model, such as including the evolution of the intrinsic Lyα EW distribution, accounting for the changing IGM during the EoR, modelling the evolution of Δv with redshift, and fully constraining the scatter in the galaxy UV luminosity–halo mass relation, will enable us to further understand the nature of scatter and its impact on Lyα visibility during reionization. Along with observational campaigns with, e.g. the *James Webb Space Telescope*, we will be able to place precision constraints on the timeline and topology of the EoR and understand the properties of the sources that drive reionization.

REFERENCES

Astropy Collaboration et al., 2013, A&A, 558, A33
Astropy Collaboration et al., 2018, AJ, 156, 123
Bañados E., et al., 2018, Nature, 553, 473
Barkana R., Loeb A., 2001, Phys. Rep., 349, 125
Becker G. D., Bolton J. S., Madau P., Pettini M., Ryan-Weber E. V., Venemans B. P., 2015, MNRAS, 447, 3402
Behroozi P. S., Conroy C., Wechsler R. H., 2010, ApJ, 717, 379
Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57
Bouwens R. J., et al., 2015, ApJ, 803, 34
Cassata P., et al., 2015, A&A, 573, A24
Choudhury T. R., Puchwein E., Haehnelt M. G., Bolton J. S., 2015, MNRAS, 452, 261
Curtis-Lake E., et al., 2012, MNRAS, 422, 1425
Davies F. B., et al., 2018, ApJ, 864, 142
Dayal P., Ferrara A., 2018, Phys. Rep., 780, 1

---

[4] A smaller dispersion with a critical mass flattening threshold is difficult to distinguish from a larger dispersion combined with a critical luminosity flattening threshold. A critical mass threshold tends to flatten the median UV luminosity–halo mass relation at brighter luminosities than a critical luminosity threshold does (R19), so a critical mass threshold with smaller dispersion predicts approximately the same abundance of bright objects as a critical luminosity threshold with larger dispersion.

De Barros S., et al., 2017, A&A, 608, A123

Dijkstra M., 2014, Publ. Astron. Soc. Australia, 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

Fan X., Carilli C. L., Keating B., 2006, ARA&A, 44, 415

Finkelstein S. L., et al., 2015, ApJ, 814, 95

Furlanetto S. R., Zaldarriaga M., Hernquist L., 2004, ApJ, 613, 1

Furlanetto S. R., Zaldarriaga M., Hernquist L., 2006, MNRAS, 365, 1012

Greig B., Mesinger A., Haiman Z., Simcoe R. A., 2017, MNRAS, 466, 4239

Greig B., Mesinger A., Bañados E., 2019, MNRAS, 484, 5094

Gunn J. E., Peterson B. A., 1965, ApJ, 142, 1633

Haiman Z., Spaans M., 1999, ApJ, 518, 138

Hashimoto T., Ouchi M., Shimasaku K., Ono Y., Nakajima K., Rauch M., Lee J., Okamura S., 2013, ApJ, 765, 70

Hayes M., Schaerer D., Östlin G., Mas-Hesse J. M., Atek H., Kunth D., 2011, ApJ, 730, 8

Hoag A., et al., 2019, ApJ, 878, 12

Hunter J. D., 2007, Computing in Science and Engineering, 9, 90

Jensen H., Hayes M., Iliev I. T., Laursen P., Mellema G., Zackrisson E., 2014, MNRAS, 444, 2114

Jones E., Oliphant T., Peterson P., et al., 2001, SciPy: Open source scientific tools for Python, http://www.scipy.org/

Keating L. C., Weinberger L. H., Kulkarni G., Haehnelt M. G., Chardin J., Aubert D., 2019, arXiv e-prints, p. arXiv:1905.12640

Komatsu E., et al., 2011, ApJS, 192, 18

Kulkarni G., Keating L. C., Haehnelt M. G., Bosman S. E. I., Puchwein E., Chardin J., Aubert D., 2019a, MNRAS, 485, L24

Kulkarni G., Worseck G., Hennawi J. F., 2019b, MNRAS, 488, 1035

Kullback S., 1968, Information theory and statistics. Dover Publications

Laursen P., Sommer-Larsen J., Razoumov A. O., 2011, ApJ, 728, 52

Lidz A., McQuinn M., Zaldarriaga M., Hernquist L., Dutta S., 2007, ApJ, 670, 39

Malhotra S., Rhoads J. E., 2004, ApJ, 617, L5

Mason C. A., Trenti M., Treu T., 2015, ApJ, 813, 21

Mason C. A., Treu T., Dijkstra M., Mesinger A., Trenti M., Pentericci L., de Barros S., Vanzella E., 2018a, ApJ, 856, 2

Mason C. A., et al., 2018b, ApJ, 857, L11

Mason C. A., et al., 2019a, MNRAS, 485, 3947

Mason C. A., Naidu R. P., Tacchella S., Leja J., 2019b, MNRAS, 489, 2669

McGreer I. D., Mesinger A., D'Odorico V., 2015, MNRAS, 447, 499

McQuinn M., Lidz A., Zahn O., Dutta S., Hernquist L., Zaldarriaga M., 2007, MNRAS, 377, 1043

Mesinger A., 2010, MNRAS, 407, 1328

Mesinger A., 2016, Understanding the Epoch of Cosmic Reionization: Challenges and Progress. Vol. 423, Springer, doi:10.1007/978-3-319-21957-8

Mesinger A., Furlanetto S., 2007, ApJ, 669, 663

Mesinger A., Furlanetto S., Cen R., 2011, MNRAS, 411, 955

Mesinger A., Aykutalp A., Vanzella E., Pentericci L., Ferrara A., Dijkstra M., 2015, MNRAS, 446, 566

Mesinger A., Greig B., Sobacchi E., 2016, MNRAS, 459, 2342

More S., van den Bosch F. C., Cacciato M., Mo H. J., Yang X., Li R., 2009, MNRAS, 392, 801

Moster B. P., Naab T., White S. D. M., 2013, MNRAS, 428, 3121

Murray S. G., Power C., Robotham A. S. G., 2013, Astronomy and Computing, 3, 23

Neufeld D. A., 1990, ApJ, 350, 216

Oesch P. A., et al., 2015, ApJ, 804, L30

Oliphant T. E., 2006, A guide to NumPy. Vol. 1, Trelgol Publishing USA

Ono Y., et al., 2012, ApJ, 744, 83

Ouchi M., et al., 2010, ApJ, 723, 869

Ouchi M., et al., 2018, PASJ, 70, S13

Paardekooper J.-P., Khochfar S., Dalla Vecchia C., 2015, MNRAS, 451, 2544

Parsa S., Dunlop J. S., McLure R. J., 2018, MNRAS, 474, 2904

Parzen E., 1962, Ann. Math. Statist., 33, 1065

Pentericci L., et al., 2011, ApJ, 743, 132

Pentericci L., et al., 2014, ApJ, 793, 113

Perez F., Granger B. E., 2007, Computing in Science and Engineering, 9, 21

Planck Collaboration et al., 2016a, A&A, 594, A13

Planck Collaboration et al., 2016b, A&A, 596, A108

Planck Collaboration et al., 2018, arXiv e-prints, p. arXiv:1807.06209

Ren K., Trenti M., Mutch S. J., 2018, ApJ, 856, 81

Ren K., Trenti M., Mason C. A., 2019, ApJ, 878, 114

Roberts-Borsani G. W., et al., 2016, ApJ, 823, 143

Robertson B. E., Ellis R. S., Dunlop J. S., McLure R. J., Stark D. P., 2010, Nature, 468, 49

Rosenblatt M., 1956, Proceedings of the National Academy of Science, 42, 43

Santos M. R., 2004, MNRAS, 349, 1137

Schenker M. A., Stark D. P., Ellis R. S., Robertson B. E., Dunlop J. S., McLure R. J., Kneib J.-P., Richard J., 2012, ApJ, 744, 179

Sheth R. K., Mo H. J., Tormen G., 2001, MNRAS, 323, 1

Shibuya T., et al., 2014, ApJ, 788, 74

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

Spergel D., et al., 2015, arXiv e-prints, p. arXiv:1503.03757

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

Stark D. P., et al., 2017, MNRAS, 464, 469

Steidel C. C., Erb D. K., Shapley A. E., Pettini M., Reddy N., Bogosavljević M., Rudie G. C., Rakic O., 2010, ApJ, 717, 289

Tacchella S., Bose S., Conroy C., Eisenstein D. J., Johnson B. D., 2018, ApJ, 868, 92

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

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

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

Weinberger L. H., Kulkarni G., Haehnelt M. G., Choudhury T. R., Puchwein E., 2018, MNRAS, 479, 2560

Weinberger L. H., Haehnelt M. G., Kulkarni G., 2019, MNRAS, 485, 1350

Wise J. H., Demchenko V. G., Halicek M. T., Norman M. L., Turk M. J., Abel T., Smith B. D., 2014, MNRAS, 442, 2560

Yang X., Mo H. J., Jing Y. P., van den Bosch F. C., 2005, MNRAS, 358, 217

Yang H., et al., 2017, ApJ, 844, 171

Zitrin A., et al., 2015, ApJ, 810, L12

van der Walt S., Colbert S. C., Varoquaux G., 2011, Computing in Science and Engineering, 13, 22

APPENDIX A: COMPARISON OF BIASES IN THE NEUTRAL FRACTION INFERENCE

We compare the inferred neutral fraction with and without accounting for scatter for four samples of simulated galaxies: two samples ($N = 68$ and $N = 1000$ galaxies) biased towards bright ($M_{UV} \leq -21.8$) galaxies, and two fainter samples with
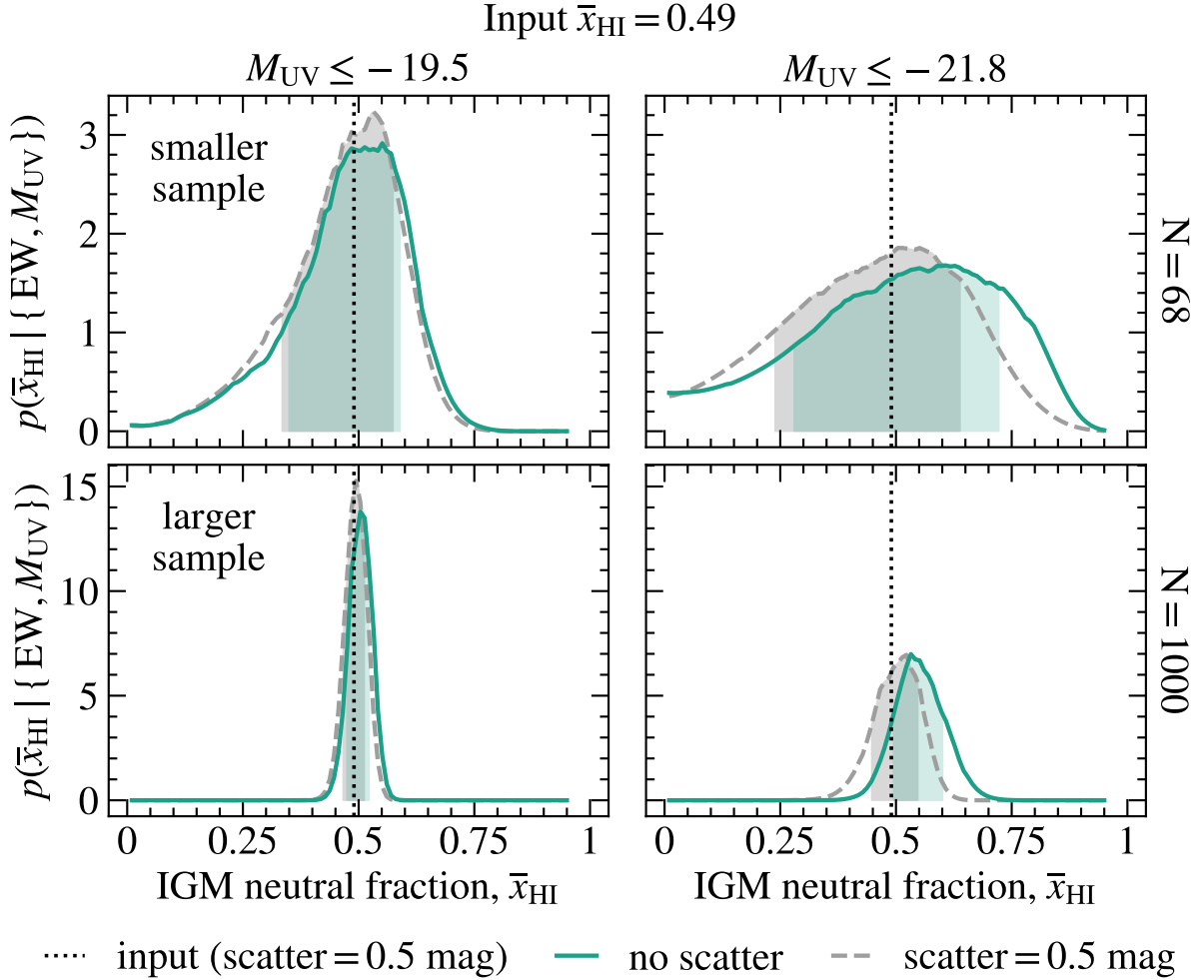
Figure A1. Posterior distributions for $\overline{x}_{\mathrm{HI}}$ inferred from four simulated samples of galaxies with various sample sizes and limiting absolute magnitudes with (dashed grey) and without (solid green) accounting for scatter in the UV luminosity–halo mass relation. Columns correspond to samples with limiting magnitudes of $M_{\mathrm{UV}} \leq -19.5$ (left) and $M_{\mathrm{UV}} \leq -21.8$ (right), and rows correspond to sample sizes of $N = 68$ (top) and $N = 1000$ (bottom). Black dotted lines denote the input neutral fraction ($\overline{x}_{\mathrm{HI}} = 0.49$) and the shaded regions are the 68 per cent credible intervals. With the smaller sample, the inferences both with and without scatter recover the input neutral fraction within errors. However, for the larger samples, the posterior becomes more precise, and the inference with limiting magnitude $M_{\mathrm{UV}} = -21.8$ without scatter overestimates the neutral fraction. Thus, accounting for UV luminosity–halo mass relation scatter becomes important for large samples of bright galaxies.

$M_{\mathrm{UV}} \leq -19.5$. We draw galaxies from the Mason et al. (2015) UV luminosity function at $z \sim 7$ and populate them with Lyα EWs from the probability distribution $p(EW \mid M_{\mathrm{UV}}, \overline{x}_{\mathrm{HI}}, \sigma)$ with scatter of $\sigma = 0.5\,\mathrm{mag}$ at a neutral fraction of $\overline{x}_{\mathrm{HI}} = 0.49$. We furthermore add Gaussian noise with a standard deviation of $5\,\text{Å}$ to the Lyα EWs. We then infer the neutral fraction with and without accounting for scatter in the UV luminosity–halo mass relation.

Fig. A1 shows the posterior distributions for $\overline{x}_{\mathrm{HI}}$ inferred with and without accounting for scatter in the UV luminosity–halo mass relation for all four samples. For the smaller sample size of $N = 68$, the inferences both with and without scatter recover the input neutral fraction of $\overline{x}_{\mathrm{HI}} = 0.49$ within the uncertainties. However, for the larger sample of $N = 1000$, the inference without scatter performs more poorly as the posteriors become more precise: the

inference without scatter for the samples of galaxies with $M_{\mathrm{UV}} \leq -21.8$ systematically overestimates the input neutral fraction, which is more significant for the larger sample size. In contrast, the inference without scatter using the sample with $M_{\mathrm{UV}} \leq -19.5$ recovers the neutral fraction well for both sample sizes. This is because scatter in the UV luminosity–halo mass relation primarily impacts bright galaxies, but fainter galaxies are much more numerous and thus contribute significantly to the inference when included. Thus, the effects of scatter on UV-bright galaxies do not bias neutral fraction inferences when the observed sample include galaxies $M_{\mathrm{UV}} > M_{\mathrm{UV}}^{\star}$.

To further assess the impact of including scatter in the UV luminosity–halo mass relation, we evaluate the Kullback-Leibler divergence (KL divergence; Kullback 1968) between the posterior distributions with and without ac-

counting for scatter for each sample of simulated galaxies. The KL divergence between the posteriors with and without scatter is larger for the sample with the brighter magnitude limit, reflecting the increased impact of scatter on Lyα transmission through the reionizing IGM for brighter galaxies, which biases the inferred neutral fraction. The KL divergence between the two posteriors increases with increasing sample size, reflecting the more precise constraints on the neutral fraction obtained from larger samples.

This paper has been typeset from a TEX/LATEX file prepared by the author.