Constraining reionization in progress at $z = 5.7$ with Lyman-α emitters: voids, peaks, and cosmic variance

Nakul Gangolli,[1]★ Anson D'Aloisio,[1] Fahad Nasir [1] and Zheng Zheng[2]

[1]*Department of Physics and Astronomy, University of California, Riverside, CA 92521, USA*
[2]*Department of Physics and Astronomy, University of Utah, Salt Lake City, UT 84112, USA*

ABSTRACT

A number of independent observations suggest that the intergalactic medium was significantly neutral at $z = 7$ and that reionization was, perhaps, still in progress at $z = 5.7$. The narrowband survey, SILVERRUSH, has mapped over 2000 Lyman-α emitters (LAEs) at these redshifts (G58). Previous analyses have assumed that reionization was over by $z = 5.7$, but this data may actually sample the final stages of reionization when the last neutral islands were relegated to the cosmic voids. Motivated by these developments, we re-examine LAE void and peak statistics and their ability to constrain reionization. We construct models of the LAE distribution in $(1 \, \mathrm{Gpc} \, h^{-1})^3$ volumes, spanning a range of neutral fractions at $z = 5.7$ and 6.6. Models with a higher neutral fraction show an enhanced probability of finding holes in the LAE distribution. When comparing models at fixed mean surface density, however, LAEs obscured by neutral gas in the voids must be compensated by visible LAEs elsewhere. Hence, in these models, the likelihood of finding an overdense peak is also enhanced in the latter half of reionization. Compared to the widely used angular two-point correlation function (2PCF), we find that the void probability function (VPF) provides a more sensitive test of models during the latter half of reionization. By comparison, at neutral fractions ~ 50 per cent, the VPF and a simple peak thresholding statistic are both similar to the 2PCF in constraining power. Lastly, we find that the cosmic variance and large-scale asymmetries observed in the SILVERRUSH fields are consistent with large-scale structure in a ΛCDM universe.

Key words: methods: numerical – intergalactic medium – dark ages, reionization, first stars.

1 INTRODUCTION

Observational measurements have narrowed the range of viable reionization models to those that end around $z = 6$. The *Planck* measurement of the Thomson scattering optical depth to the cosmic microwave background places the bulk of reionization between $z = 6$–12 (Planck Collaboration VI 2020). This timing is consistent with transmission measurements of the Ly α and Ly β forests (Fan et al. 2006; McGreer, Mesinger & D'Odorico 2015; Wang et al. 2019), as well as evidence of quasar damping wings in $z > 7$ quasars (Mortlock et al. 2011; Davies et al. 2018; Wang et al. 2020). Additionally, at $z \gtrsim 5.5$, the Ly α forest exhibits large scatter in its transmission averaged over long path lengths, epitomized by the $110 \, h^{-1}$ comoving Mpc Ly α trough reported by Becker et al. 2015 (see also Fan et al. 2006; Bosman et al. 2018; Eilers, Davies & Hennawi 2018). Some models proposed to explain this scatter place the end of reionization as late as $z = 5$ (Kulkarni et al. 2019; Keating et al. 2020; Nasir & D'Aloisio 2020). Although a broadly consistent picture of late reionization has emerged from all of these observations, the reionization history in detail as well as its sources remain unknown.

Lyman-α emitters (LAEs) have long been an important probe of the Epoch of Reionization (EoR) owing to the resonant nature of the Ly α line to neutral gas. Visible LAEs indicate the presence of ionized gas within the intergalactic medium (IGM) because even small densities of neutral hydrogen nearby can attenuate LAEs significantly. Indeed, the luminosity function of LAEs evolves rapidly at $z > 6$ and the fraction of Lyman Break Galaxies showing strong Ly α emission is also observed to decline (e.g. Kashikawa et al. 2006; Ono et al. 2012; Schenker et al. 2012; Pentericci et al. 2014; Mesinger et al. 2015; Schmidt et al. 2016; Inoue et al. 2018; Mason et al. 2018a; Hu et al. 2019). The drop-off in space density of LAEs can be attributed to an increasingly neutral IGM towards high redshift, an interpretation bolstered by other observations as described above (see, however, Sadoun, Zheng & Miralda-Escudé 2017; Hassan & Gronke 2020). The differential evolution of the bright and faint LAE populations at $z > 7$ (e.g. Caruana et al. 2012, 2014; Ota et al. 2008, 2010; Stark et al. 2017; Hoag et al. 2019; Endsley et al. 2020) appears consistent with expectations from reionization simulations. In fact, models suggest that the brightest LAEs at $z > 7$ may serve as signposts for the earliest ionized structures (Mason et al. 2018b; Weinberger et al. 2018). A noteworthy case is the double-peaked emitter COLA1, for which the presence of a prominent blue wing places a lower limit on the size of the ionized bubble hosting it (Hu et al. 2016; Matthee et al. 2018; Mason & Gronke 2020). LAEs have also been used in combination with Ly α forest data to study the connection between LAEs and their intergalactic environments (Becker et al. 2018; Kakiichi et al. 2018; Meyer et al. 2020).

Given sufficiently large survey area, the angular two-point correlation function (2PCF) is thought to be the most robust probe of reionization among LAE statistics. Simulations show that the obscu-

★ E-mail: ngang002@ucr.edu

ration of LAEs by neutral gas during reionization generally enhances their apparent clustering (Furlanetto, Zaldarriaga & Hernquist 2006; McQuinn et al. 2007; Jensen et al. 2014; Hutter, Dayal & Müller 2015; Sobacchi & Mesinger 2015). An early measurement of the 2PCF at $z = 6.6$ from 207 LAEs in 1 deg^2 constrained their bias, host halo masses, and duty cycle (Ouchi et al. 2010). The recently conducted narrowband survey, SILVERRUSH, greatly expanded the sample of known reionization-era LAEs using the Subaru/Hyper Suprime-Cam. The collaboration has reported over 2000 LAEs at $z = 5.7$ and 6.6 covering areas of 14–21 deg^2 (0.3–0.5 Gpc^2). With this updated sample, Ouchi et al. (2018) found only mild redshift evolution in the 2PCF from $z = 5.7$ to 6.6. Comparing their measurements to previously published models, they estimated a global neutral fraction of $x_{H_I} = 0.15 \pm 0.15$ at $z = 6.6$. At face value, this measurement provides little evidence for the late reionization scenario that has become increasingly supported by independent observations, although the uncertainties in the 2PCF remain large (see e.g. discussion in Inoue et al. 2018).

One of the central questions motivating this paper is whether the 2PCF is the most sensitive statistic for constraining reionization during its final stages. As noted above, there is mounting evidence that existing LAE samples at $z \sim 7$ are probing deep into reionization. Even the most densely sampled data at $z = 5.7$ may be probing the tail end of reionization (Keating et al. 2020; Nasir & D'Aloisio 2020). Radiative transfer simulations indicate that neutral gas becomes increasingly relegated to the voids in the source distribution when the global neutral fraction drops below 50 per cent (e.g. Doussot, Trac & Cen 2019; Giri et al. 2019; Kulkarni et al. 2019; Wu et al. 2019). We therefore expect LAEs located in voids to be preferentially obscured by neutral islands during the last stages of reionization. Motivated by this expectation, we re-examine the use of LAE void statistics for constraining reionization, with an eye towards recent models that place the neutral fraction at ~ 10 per cent at $z = 5.7$. We also explore peaks in the LAE distribution as a complementary method for quantifying reionization's effect on the apparent clustering of LAEs.

LAE void statistics in the context of reionization were considered by McQuinn et al. (2007). This paper adds to this work in three ways: (1) Whereas the simulations in McQuinn et al. (2007) concluded reionization well before $z = 5.7$, we focus explicitly on the possibility of constraining models in which reionization was still ongoing at $z = 5.7$. This updates previous work with the most recent empirical developments and may be of particular interest because LAE samples are currently largest at $z = 5.7$; (2) We employ simulation volumes of (1 Gpc h^{-1})^3, a factor of ≈ 470 bigger than the largest in McQuinn et al. (2007), to model reionization and the LAE population. This allows us to develop statistically representative models of narrowband surveys whose individual fields of view often are comparable to, or larger than, the simulation box sizes of previous studies. The SILVERRUSH fields are striking visually because they exhibit significant field-to-field variation in surface density, as well as clear asymmetries in the LAE distribution over enormous scales ($\sim$200 Mpc/h). Some fields are pocketed with highly clustered regions and nearly empty holes. We exploit our large simulation volumes to address additionally whether the observed features are consistent with expected cosmic variance from large-scale structure, or whether they may be indicative of some other source of variance; (3) We contrast the constraining power of voids and peak statistics with that of the 2PCF. We will find that voids provide a more sensitive test of models than the 2PCF during the late stages of reionization.

The remainder of this paper is structured as follows. In Section 2, we describe our simulation approach and in Section 3 we lay out and

calibrate our late reionization models. Section 4 contains our main results on void and peak statistics, and cosmic variance. In Section 5, we offer concluding remarks. We assume a standard ΛCDM cosmology with $H_0 = 100 \, h$ km s^{-1} Mpc^{-1} and $h = 0.68$, $\Omega_m = 0.3$, and $\Omega_\Lambda = 0.7$, consistent with the latest *Planck* measurements (Planck Collaboration VI 2020). Unless otherwise stated, all distances are reported in comoving units and velocities are in proper units.

2 SIMULATION METHODOLOGY

The visibility of LAEs is set by physical processes spanning an enormous dynamic range, from hundreds of Mpc h^{-1} to subgalactic scales. In addition, narrowband survey fields of view span ($\sim$hundreds of Mpc)^2. Our multiscale approach attempts to bridge the scale gap by using several cosmological simulations in tandem. In this section, we describe each component of our models.

2.1 The underlying LAE population

The basis for our mock LAE catalogues is the publicly available halo catalogues from the Multi-Dark Simulation (MDPL) of Klypin et al. (2016), which affords us a 1 (h^{-1} Gpc)^3 volume. The MDPL simulation is a dark-matter-only *N*-body simulation that was run with a modified version of the GADGET-2 code (Springel 2005) using $N = 3840^3$ particles, which corresponds to a particle mass of $1.51 \times 10^9 \, h^{-1}$ M$_\odot$. We used the friends-of-friends (FoF) halo catalogues provided by the CosmoSim data base.[1] The minimum halo mass in these catalogues is $M_{min} = 3 \times 10^{10} \, h^{-1}$ M$_\odot$.

To populate the MDPL haloes with galaxies, we applied an abundance matching scheme that equates the cumulative number density of haloes with the integrated rest-frame UV luminosity functions of Bouwens et al. 2015 (see also Finkelstein et al. 2015). Following Weinberger, Haehnelt & Kulkarni (2019), we model the duty cycle of these galaxies with the method of Trenti et al. (2010). The duty cycle is parametrized by a star formation time-scale Δt, which we take to be one of the free parameters in our models. Smaller values of Δt result in UV luminous galaxies being assigned to less-massive haloes. In what follows, we adopt $\Delta t = 50$ Myr as our fiducial value. In Section 3, we show that this choice provides a reasonable match to statistics of the observed LAE population.

Given UV luminosities, the Lyα rest-frame equivalent widths (REWs) of galaxies were drawn randomly from the empirically calibrated model distribution of Dijkstra & Wyithe (2012). Scatterings of Lyα photons by the interstellar gas results in a complex frequency structure for the Lyα line as it emerges from source galaxies. We do not attempt to model these physical processes in detail here. Instead, we assume that the structure of the line blue-ward of systemic is completely absorbed by the IGM – a good approximation at the high redshifts of interest, in most cases (see, however, Hu et al. 2016; Matthee et al. 2018; Songaila et al. 2018). We model the red-ward side with a Gaussian profile characterized by a velocity offset, Δv, and width σ_v. We set $\Delta v = \beta v_{circ}$, where v_{circ} is the circular velocity of the halo at its boundary,[2] and β is a free parameter in our model (e.g. Weinberger et al. 2019). Following Mason et al. (2018a), we fix σ_v by setting the full width at half-maximum of the Gaussian equal to v_{circ}. These choices are motivated by radiative transfer calculations

---

[1] https://www.cosmosim.org/
[2] Here, the 'radius' of an FoF halo is defined to be that of a sphere with equal volume.

of the Ly α line structure emerging from simple models of the ISM in star-forming galaxies (Zheng & Miralda-Escudé 2002).

2.2 Opacity of the IGM

The attenuation of Ly α emission lines by intergalactic hydrogen comes from two sources: (1) The neutral hydrogen in yet-to-be reionized gas; (2) Self-shielding regions and residual neutral hydrogen within ionized bubbles. Sections 2.2.1 and 2.2.2, respectively, describe how we model these.

2.2.1 Reionization simulations

We modelled the distribution of neutral hydrogen during reionization with a numerical implementation of the excursion-set model of reionization (ESMR; Furlanetto, Zaldarriaga & Hernquist 2004; Mesinger, Furlanetto & Cen 2011). Following previous implementations (see e.g. Zahn et al. 2011), we smoothed the MDPL halo field on to a grid with dimension $N = 512^3$ to obtain the collapsed fraction field, $f_{coll}(\mathbf{r})$. The latter was then smoothed over a hierarchy of scales with a top-hat filter in Fourier space. Cells were marked as ionized when the condition $\zeta f_{coll}(\mathbf{r}) \geq 1$ was satisfied, where ζ is a free parameter quantifying the efficiency at which the sources deliver ionizing photons to the IGM. As we describe in Section 3, we adjusted ζ to obtain different global ionized fractions at $z = 5.7$ and $z = 6.6$.

It is important to account for the enhancement of the ionizing ultraviolet background (UVB) intensity around bright and/or clustered LAEs. We modelled spatial fluctuations in the UVB with a simple attenuation model, separately from the ESMR reionization simulations. We computed the hydrogen photoionization rate, Γ_{HI}, on a uniform grid with $N = 512^3$ cells assuming that each source at location $\vec{x}_j$ makes a contribution $\propto L_{912} \exp(-|\vec{x}_i - \vec{x}_j|/\lambda_{912}^{mfp})/|\vec{x}_i - \vec{x}_j|^2$ to the cell at $\vec{x}_i$, where L_{912} is the source's specific luminosity at 912 Å, and λ_{912}^{mfp} is the mean free path. For simplicity, we adopt a uniform $\lambda_{912}^{mfp} = 30(15)\, h^{-1}$ Mpc at $z = 5.7(6.6)$, which is motivated by extrapolating the observational measurements of Worseck et al. (2014) at $z \lesssim 5.2$. Assuming a constant λ_{912}^{mfp} affords us the ability to compute Γ_{HI} fields efficiently using Fast Fourier Transforms, at the expense of neglecting potentially large spatial variations in λ_{912}^{mfp} at these redshifts (Davies & Furlanetto 2016; D'Aloisio et al. 2018). Small local values of λ_{912}^{mfp} – for example, in cosmic voids with a dearth of sources – leads to more attenuation of the local LAE population by self-shielding systems (Bolton & Haehnelt 2013; Sadoun et al. 2017), introducing another source of fluctuation in the LAE distribution. Another effect that our method misses is the shadowing of the UVB that leads to a suppression of Γ_{HI} in the vicinity of neutral hydrogen (Nasir & D'Aloisio 2020).

2.2.2 Absorptions within ionized bubbles

Even within an ionized bubble, residual neutral hydrogen in the cosmic web can contribute significantly to the attenuation of Ly α emission (e.g. Bolton & Haehnelt 2013; Mesinger et al. 2015). We model this component of the opacity using 1D skewers extracted from an Eulerian cosmological hydrodynamics simulation run with a modified version of the code of Trac & Pen (2004). The details of our code implementation have been described elsewhere (D'Aloisio et al. 2018; Nasir & D'Aloisio 2020). Here, we summarize the salient points (including updates) for this paper.

Our simulation used a periodic box with side length $L_{box} = 20\, h^{-1}$ Mpc, and 2×2048^3 gas and dark matter resolution elements. The simulation was initialized at $z = 300$ using first-order perturbation

theory and transfer functions from the CAMB software (Lewis, Challinor & Lasenby 2000). The gas was flash ionized at redshift $z_{reion} = 8.5$ by a uniform ionizing radiation background with specific intensity $J_\nu \propto \nu^{-1.5}$ (where ν is frequency) between 1 and 4 Ry. At this time, the gas was also impulsively heated to $T = 20\,000$ K. The normalization of the UVB was fixed to give a global mean hydrogen photoionization rate of $\Gamma_{HI} = 10^{-13}$ s^{-1}. However, to incorporate spatial fluctuations in the UVB, we re-scaled Γ_{HI} in post-processing along the skewers using the UVB models from the last section. To model self-shielding in over-dense regions, we implemented the model of Rahmati et al. (2013) with parameters updated by Chardin, Kulkarni & Haehnelt (2018). We traced 20 000 skewers at random angles from haloes[3] with masses $M \geq 10^{10}$ M$_\odot$. As we describe in the next section, we used the gas data from these skewers to model attenuation of the Ly α line by self-shielding regions along the sightlines to LAEs.

2.3 Construction of mock surveys

We piece together all the elements described above to construct mock LAE catalogues. The first step is to create 1000 mock fields of view (FoVs) by taking randomly located and oriented sub-volumes of the MDPL simulation – utilizing the periodic boundary conditions. Our sub-volumes have widths of 180(200) h^{-1} Mpc at $z = 5.7(6.6)$, and depths of 29 h^{-1} Mpc. The widths correspond approximately to the average FoV scale of SILVERRUSH, and the depths represent the selection windows of the NB816 (NB921) filters (Ouchi et al. 2018). We then populate the haloes in these sub-volumes with LAEs using the method outlined in Section 2.1. This step produces the 'intrinsic' LAE population.

The next step is to attenuate Ly α luminosities and REWs by applying the opacity of the IGM redward of systematic Ly α. We trace skewers of length 150 h^{-1} Mpc from each LAE along the line of sight. The skewers consist of uniform grid points at the resolution of the hydrodynamics simulation, $\Delta x = 9.77\, h^{-1}$ kpc, or $\Delta v = 1.41(1.50)$ km s^{-1} at $z = 5.7(6.6)$. Gas properties are assigned to these skewers from randomly chosen skewers through the hydrodynamics simulation. Our ESMR simulations provide the morphology of the neutral gas around the LAEs. For a given reionization model, wherever a skewer intersected neutral regions in the corresponding ESMR field, the neutral fraction of the gas was set to unity. We emphasize that the ESMR fields and the LAEs share the same underlying density field (from the MDPL box), but the $L = 20\, h^{-1}$ Mpc hydrodynamics simulation does not. Similar multiscale approaches have been employed by Mesinger et al. (2015) and Davies et al. (2018).

By integrating along the line of sight, we compute the IGM transmission fraction redward of systemic,

$$T_{Ly\alpha}^{IGM} = \frac{\int d\nu\, J(\nu) e^{-\tau(\nu)}}{\int d\nu\, J(\nu)}, \tag{1}$$

where $J(\nu)$ is the line profile normalized to integrate to unity, and $\tau(\nu)$ is the opacity including contributions from the ionized and fully neutral regions (Mesinger et al. 2015; Weinberger et al. 2019). After attenuating the LAEs, we apply cuts on the Ly α luminosity ($L_{Ly\alpha}$) and REW. Unless specified otherwise, we select $L_{Ly\alpha} > 6.3 \times 10^{42}$ (7.9×10^{42}) erg s^{-1} and REW $>10(14)$ Å at $z = 5.7(6.6)$. These

---

[3]Haloes were identified with a spherical-overdensity criterion in which the mass enclosed is $M_{200} = 200\bar{\rho}_m(4\pi/3)R_{200}^3$, where $\bar{\rho}_m$ is the cosmic mean matter density and R_{200} is the radius within which the mean matter density is $200\bar{\rho}_m$.

Table 1. Summary of our model parameters. Columns 2–6 list the five free parameters in our models. The last column shows the resulting mean LAE surface densities, $\langle \Sigma \rangle$, which have been tuned to match (approximately) the observed values in SILVERRUSH. Here, $\bar{x}_{\mathrm{HI}}$ is the mean neutral fraction of the IGM, β is the proportionality constant between halo circular velocity and Ly α line offset, $\lambda_{912\text{Å}}^{\mathrm{mfp}}$ is the mean free path in ionized regions, $\langle \Gamma_{\mathrm{HI}} \rangle$ is the mean hydrogen photoionization rate in ionized regions, and Δt is the duty cycle star formation time-scale.

z	$\bar{x}_{\mathrm{HI}}$	β	$\lambda_{912\text{Å}}^{\mathrm{mfp}}$ (cMpc h^{-1})	$\langle \Gamma_{\mathrm{HI}} \rangle$ ($\times 10^{-12}$ s^{-1})	Δt (Myr)	$\langle \Sigma \rangle$ (deg^{-2})
5.7	0.0	1.2	30	0.37	50	72
	0.1	1.25	30	0.50	50	72
	0.3	1.4	30	0.74	50	71
6.6	0.0	1.2	15	0.37	50	32
	0.25	1.4	15	0.50	50	32
	0.5	1.8	15	0.74	50	32

cuts from Weinberger et al. (2019) approximate those employed by SILVERRUSH, though we note that different sample cuts are applied in the SILVERRUSH analyses depending on the application. We also note that different SILVERRUSH fields have different limiting depths.

The end result of this process is a sample of 1000 mock survey FoVs with properties similar to the SILVERRUSH fields. We will use these mocks to explore the effect of reionization on various LAE statistics in Section 4.

3 MODELS

A summary of parameters for each of our models is provided in Table 1. We generated mock LAE populations for three global neutral fractions ($\bar{x}_{\mathrm{HI}}$) at $z = 5.7$, and three at $z = 6.6$. Since the neutral fraction is the key distinction between our models, we will label them with $\bar{x}_{\mathrm{HI}}$ hereafter. The values of $\bar{x}_{\mathrm{HI}}$ are loosely motivated by the 'late' and 'early' reionization scenarios considered in Nasir & D'Aloisio (2020). We note that a neutral fraction of $\bar{x}_{\mathrm{HI}} = 0.3$ at $z = 5.7$ is in mild tension with the Ly α forest dark pixel constraints of McGreer et al. (2015). One of the main motivations of this work is to explore whether LAE measurements could provide a complementary test of such a model.

The other parameters were fixed by matching to observational constraints provided by SILVERRUSH and quasar absorption spectra measurements. To fix the LAE parameters, we considered the mean LAE surface density, REW distribution, Ly α luminosity function, and 2PCF. Most importantly for the ensuing discussions, at a given redshift, all of our models were calibrated to have approximately the same mean LAE surface density, as shown in the last column of Table 1. To achieve this, we adjusted the proportionality between halo circular velocity and line offset (β) to match our models approximately to SILVERRUSH surface densities. Our model surface densities are closest to those of the 'Homogeneous sample' described in Ouchi et al. (2018).[4] We found that a reasonable match could be obtained for all the other LAE constraints by fixing the duty cycle to $\Delta t = 50$ Myr.

The mean ionizing intensity in our fluctuating UVB models (Section 2.2.1) were calibrated by Ly α forest measurements where available. At $z = 5.7$, we constructed synthetic Ly α forest sight-lines through our simulations and rescaled the intensity, assuming photoionization equilibrium, to match the mean flux measurement of Bosman et al. (2018). Currently, forest segments at $z = 6.6$ are too saturated and sparse to provide a reliable estimate of the mean flux. For simplicity, we held the mean intensities fixed with redshift between the $x_{\mathrm{HI}} = 0$ models, the $x_{\mathrm{HI}} = 0.1$ and 0.25 models, and the $x_{\mathrm{HI}} = 0.3$ and 0.5 models. This can be seen in the mean hydrogen photoionization rates in ionized regions provided in Table 1. We tested models in which $\langle \Gamma_{\mathrm{HI}} \rangle$ is a factor of 5 lower at $z = 6.6$ than at $z = 5.7$. We found that $\langle \Gamma_{\mathrm{HI}} \rangle$ is almost completely degenerate with β, motivating our choice to hold the former fixed with redshift. The mean free path through ionized gas, $\lambda_{912}^{\mathrm{mfp}}$, is another key parameter in our UVB model. We fixed this parameter by extrapolating to higher redshift the empirical fit of Worseck et al. (2014). We note that λ^{mfp} during reionization could be considerably shorter than the values adopted here owing to the potentially lower Γ_{HI} as well as relaxation effects (D'Aloisio et al. 2020).

Fig. 1 shows that our models are in reasonable agreement with measurements of the REW distribution (left), luminosity function (middle), and 2PCF (right) from SILVERRUSH observations. The top and bottom rows correspond to $z = 5.7$ and 6.6, respectively. In all top (bottom) panels, the red, blue, and green curves, respectively, correspond to $\bar{x}_{\mathrm{HI}} = 0.0$ (0.0), 0.1 (0.25), and 0.3 (0.5). Additionally, the grey curves show the intrinsic statistics before attenuation by the IGM. For comparison, the data points show some of the most recent observational measurements. IGM opacity acts to skew the REW distribution towards lower values. It also changes the normalization and shape of the luminosity function, and increases the observed clustering of LAEs. These trends are consistent with previous findings (e.g. McQuinn et al. 2007; Jensen et al. 2014; Weinberger et al. 2019). Similar to the results of Weinberger et al. (2019), our 2PCFs exhibit more clustering compared to the SILVERRUSH measurements at $z = 6.6$. This may owe to the relatively massive haloes that our LAEs inhabit. Our models have $M_{\mathrm{min}} = 3 \times 10^{10}\,h^{-1}$ $\mathrm{M}_\odot$ and an average mass of $\langle M \rangle = 1.5 \times 10^{11}\,h^{-1}\,\mathrm{M}_\odot$. In contrast, SILVERRUSH reports $M_{\mathrm{min}} \sim 2.1 \times 10^9 (8.8 \times 10^8)\,h^{-1}\,\mathrm{M}_\odot$, and average masses of $\sim 8.7 \times 10^{10} (4.4 \times 10^{10})\,h^{-1}\,\mathrm{M}_\odot$ at $z = 5.7 (6.6)$ (Ouchi et al. 2018). These masses were derived by applying the halo occupation distribution (HOD) modelling of Harikane et al. (2016) to the measured 2PCF. Another effect at play could be the lower surface densities in our models compared to the Homogenous sample used to measure the 2PCF.

Figs 2 and 3 show mock survey fields of size $(500\,h^{-1}\,\mathrm{Mpc})^2$ at $z = 5.7$ and 6.6, respectively. For all panels, we impose the survey cuts described in Section 2.3. The top-left panels show the intrinsic distribution of LAEs in the slices, i.e. without attenuation by the IGM. The subsequent plots (moving clockwise) correspond to increasing global neutral fractions, where the shading shows the projected neutral gas density. The different dot-colours correspond to Ly α magnitude bins, as denoted in the plot legends. In the latter half of reionization, the neutral gas is increasingly confined to voids in the LAE distribution. This gas attenuates the Ly α lines of background LAEs, obscuring some of them from view. Importantly, because we fix the surface density of LAEs to be approximately the same between our models, these missing LAEs are compensated by an increased density of LAEs in the ionized regions. This effect contributes to the enhanced clustering seen in the 2PCFs (Fig. 1), and we will see its consequences for peak statistics later in this paper.
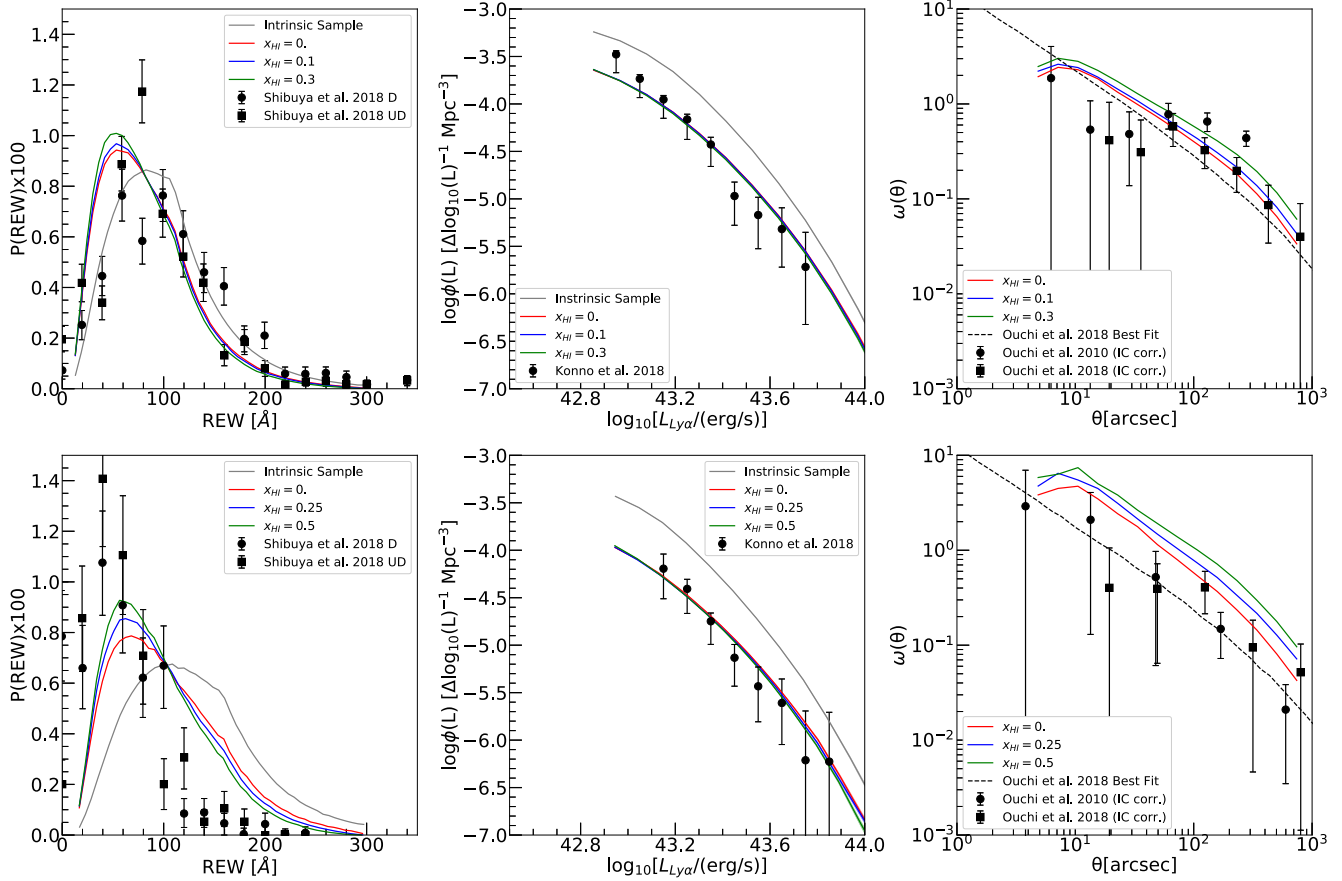
---

[4]For reference, the Homogeneous sample from Ouchi et al. (2018) has surface densities of 70 and 41 deg^{-2} at $z = 5.7$ and 6.6, respectively.

Figure 1. Comparison of our models to recent observational measurements. The left, middle, and right columns correspond to the REW distribution, Ly α luminosity function, and angular 2PCF, respectively. The top and bottom rows show results for $z = 5.7$ and 6.6, respectively. Our models are distinguished by the mean neutral fraction of the IGM, as indicated in the plot legends. In the left-hand and middle panels, the light grey curves correspond to the 'intrinsic' LAE population, before any attenuation from the IGM/neutral regions is applied. In the right-hand panels, the dashed curves show the best-fitting power-law models of Ouchi et al. (2018). In the left-hand panel legends, 'D' and 'UD' denote the Deep and the Ultra-Deep fields, respectively. Our model parameters have been tuned to provide reasonable agreement with these observations. In Section 4, we use these models to explore the efficacy of void and peak statistics for constraining reionization.

4 RESULTS

4.1 Cosmic variance

We begin with a broad comparison of our simulated ensembles to the SILVERRUSH fields. We have applied homogeneous luminosity and colour cuts to the publicly available SILVERRUSH catalogues (see text in Section 4.1.1). The LAE surface density of each field is reported in Table 2. The data exhibit significant field-to-field variation. At $z = 5.7$, D-ELAISN1 has a surface density of just ≈ 31 deg^{-2}, whereas UD-SXDS is a factor of 3 denser with ≈ 98 deg^{-2}. Additionally, some fields, such as D-ELAISN1 at $z = 5.7$ and D-DEEP2-3 at $z = 6.6$, show conspicuous asymmetries over ~ 200 h^{-1} Mpc scales (see e.g. figs 4 and 5 of Ouchi et al. 2018). Some regions remain largely empty of LAEs, while others contain significant overdensities. In this section, we explore whether the observed variations are consistent with expectations from large-scale structure formation alone, or if some other source of variance is required. We also quantify the contribution of reionization to the expected cosmic variance.

4.1.1 LAE surface density

In Fig. 4, we show the LAE surface density distributions in our reionization models. The top and bottom panels correspond to $z =$

5.7 and $z = 6.6$, respectively. The red, blue, and green histograms represent different neutral fractions, as denoted in the plot legends. For each model, the distributions consist of 1000 mock FoVs. To facilitate comparisons with the observations, we have matched the sizes of our fields to the mean sizes of the SILVERRUSH fields (see Section 2.3). This is not ideal because the SILVERRUSH fields vary in size, from 2–6 deg^2 (Konno et al. 2018; Shibuya et al. 2018). However, we will find that the observed cosmic variance can be reproduced reasonably with our simplistic approach. To remove the effect of small differences in the mean surface densities in our models, we plot the surface density contrast, $\delta = (\Sigma - \langle\Sigma\rangle)/\langle\Sigma\rangle$, where $\langle\Sigma\rangle$ is the average surface density. Fig. 4 shows that reionization increases the cosmic variance in δ quite subtly. For reference, the standard deviations of δ between our models are 0.14(0.20), 0.15(0.23), and 0.17(0.25) for neutral fractions of $\bar{x}_{\rm HI} = 0.0(0.0)$, 0.1(0.25), and 0.3(0.5) at $z = 5.7(6.6)$, respectively.

The black vertical lines in Fig. 4 illustrate the cosmic variance in the SILVERRUSH data. Each line is annotated with its respective field name (see Table 2). The luminosity and colour cuts applied in the published SILVERRUSH analyses vary, depending on the application. Hence, we downloaded the public SILVERRUSH catalogues[5]
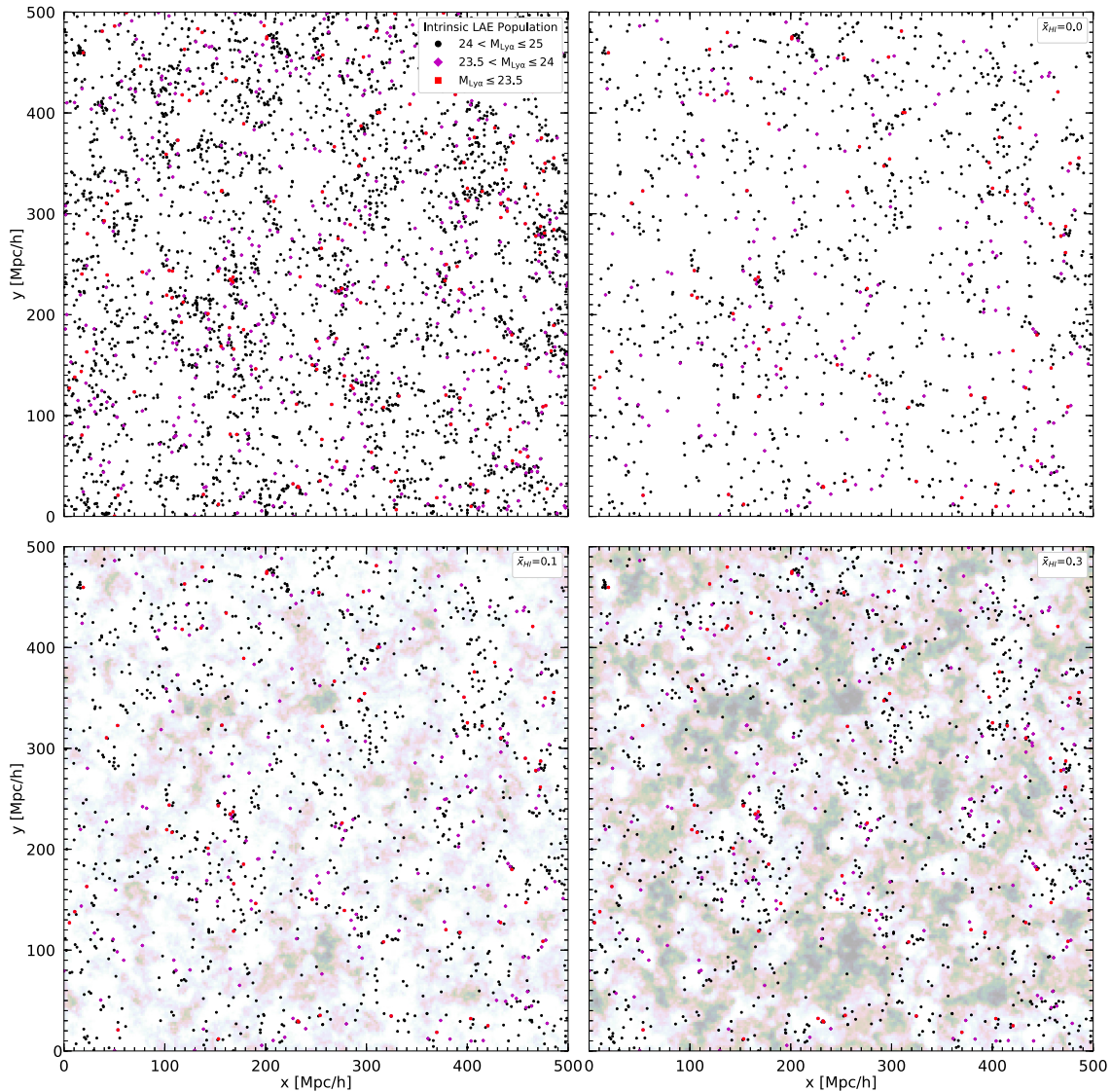
[5] http://cos.icrr.u-tokyo.ac.jp/rush.html

Figure 2. Example $500 \times 500\,(h^{-1}\,\mathrm{Mpc})^2$ slices through our simulated LAE distributions at $z = 5.7$. The different dot-colours correspond to the Ly α apparent narrowband magnitude $(M_{\mathrm{Ly}\alpha})$ ranges shown in the top-left panel. The top-left panel shows the intrinsic LAE sample before any attenuation by the IGM is applied. The remaining panels correspond to our models with $\bar{x}_{\mathrm{HI}} = 0.0$ (top-right), $\bar{x}_{\mathrm{HI}} = 0.1$ (bottom-left), and $\bar{x}_{\mathrm{HI}} = 0.3$ (bottom-right). The shading shows the integrated neutral fraction along the line of sight, which is $30\,h^{-1}$ Mpc deep. Darker shades indicate regions with higher neutral fractions. These three panels have approximately the same mean LAE surface density, $\langle\Sigma\rangle \approx 72\,(\mathrm{deg})^{-2}$. At fixed $\langle\Sigma\rangle$, a higher neutral fraction obscures LAEs in the voids, but these must be compensated by more visible LAEs elsewhere. Hence, the presence of neutral gas leads to more apparent clustering of the LAEs. Notably, the neutral gas is relegated to the voids in the tail end of reionization, suggesting that LAE void statistics may provide an informative test of models in this regime.

and applied our own cuts with the goal of matching our simulated catalogues to the extent possible. Specifically, at $z = 5.7$, we selected Ly α luminosities $>6.3 \times 10^{42}\,\mathrm{erg\,s^{-1}}$ and applied a colour cut of i $-\mathrm{NB816} > 1.2$. This yields $\langle\Sigma\rangle = 72\,\mathrm{deg}^{-2}$, in agreement with our models (see Table 1). At $z = 6.6$, we used $>1.41 \times 10^{43}\,\mathrm{erg\,s^{-1}}$ and $z - \mathrm{NB921} > 1.8$. The higher luminosity limit adopted for $z = 6.6$ ensures that a homogeneous cut can be applied across all fields in the public SILVERRUSH data. Note that this limit is much higher than in our fiducial $z = 6.6$ catalogue used throughout the rest of this paper, and results in our models having $\langle\Sigma\rangle = 9.5\,\mathrm{deg}^{-2}$ at $z = 6.6$. The colour cuts, which are from Konno et al. (2018), roughly match the REW cuts applied to our simulated catalogues. For comparison, the red/dashed vertical lines show the 2σ limits of the model distributions with $\bar{x}_{\mathrm{HI}} = 0$. (The 2σ limits for the other models are similar.)

We find that the observed field-to-field variation is almost entirely consistent with expectations from large-scale structure alone, and that reionization is unlikely to play a significant role in increasing this variation. At $z = 5.7$, our models, however, cannot account for the dearth of LAEs observed in D-ELAISN1. However, as noted by Konno et al. (2018), the deficit may owe to poor seeing for that field. Interestingly, D-ELAISN1 also shows a strong asymmetry in the LAE distribution, although there is no evidence of spatial variations in the data quality (Ouchi et al. 2018). In the next section, we explore the nature of such asymmetries.

4.1.2 Split-screen contrast

We quantify half-plane variations in the LAE fields using a statistic that we term the 'split-screen contrast.' The FoV is split, either
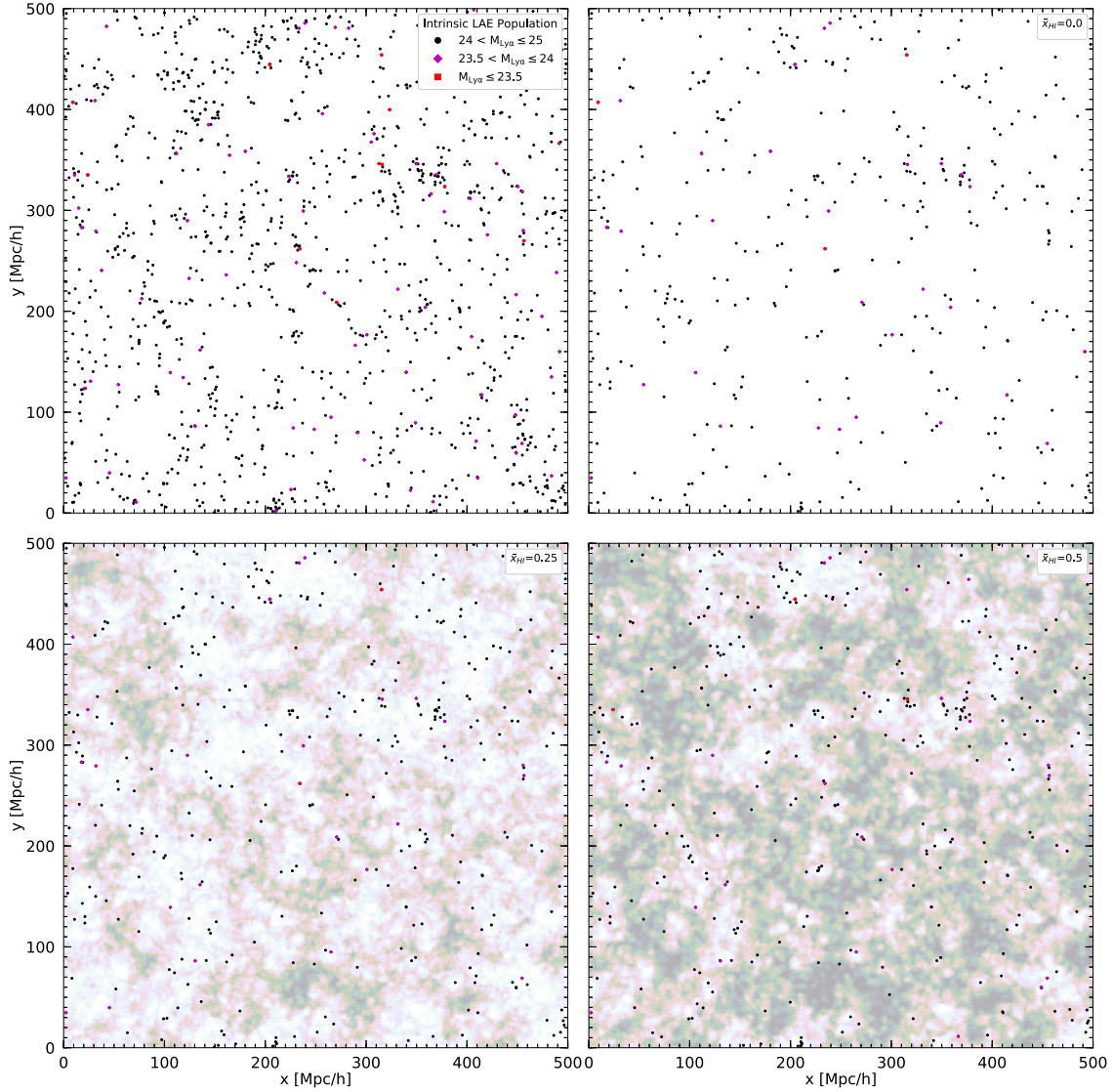
Figure 3. Same as Fig. 2, except at $z = 6.6$. The top-right ($\bar{x}_{HI} = 0.0$), bottom-left ($\bar{x}_{HI} = 0.25$), and bottom-right ($\bar{x}_{HI} = 0.5$) panels have a fixed $\langle \Sigma \rangle \approx 32$ (deg)$^{-2}$.

Table 2. Surface densities (third column) and split-screen contrasts (fourth column) measured from our sample of the SILVERRUSH data. Our luminosity and colour cuts are described in Section 4.1.1. These data are compared against our models in Figs 4 and 5.

z	Field	Σ [deg^{-2}]	$\Delta\Sigma / \Sigma$
5.7	UD-COSMOS	75.6	0.27, 0.14
	UD-SXDS	98.4	0.33, 0.03
	D-DEEP2-3	84.4	0.28, 0.46, 0.18
	D-ELAISN1	31.1	0.20, 0.18, 0.76, 0.02, 0.55, 0.58
6.6	UD-COSMOS	12.2	0.75, 0.32
	UD-SXDS	9.41	0.11, 0.55
	D-COSMOS	8.27	0.75, 0.44, 0.31
	D-DEEP2-3	8.16	0.31, 0.15, 0.08, 0.15, 0.23, 0.08
	D-ELAISN1	9.38	0.15, 0.07, 0.44, 0.22, 0.29, 0.51

vertically, horizontally, or diagonally to create regions of roughly equal area. We then calculate the contrast,

$$\frac{\Delta\Sigma}{\Sigma} = \frac{|\Sigma_1 - \Sigma_2|}{\Sigma}, \tag{2}$$

where Σ_1 and Σ_2 are the surface densities of the two halves (or 'split screens') and Σ is the surface density of the field.

In Fig. 5, we show the cumulative probability distribution of split-screen contrasts in our models, $P(< \Delta\Sigma/\Sigma)$, or the probability of finding $\Delta\Sigma/\Sigma$ less than a given value. Reionization widens the distribution somewhat. At $z = 5.7$, the standard deviation increases by 30 per cent between a completely ionized IGM and $\bar{x}_{HI} = 0.3$. At $z = 6.6$, the differences are more subtle, a 12 per cent increase in width between $\bar{x}_{HI} = 0.0$ and 0.5, owing to the relative dominance of shot noise from a lower mean surface density. Generally speaking, a larger neutral fraction extends the high-contrast tail of the distribution, enhancing the likelihood of finding more extreme half-plane asymmetries.
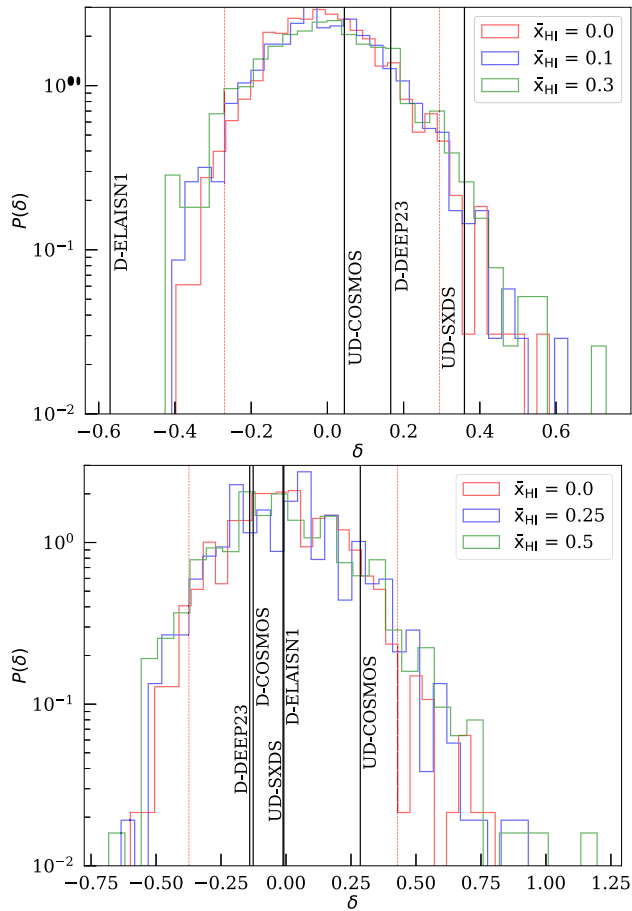
Figure 4. Cosmic variance in the surface density of SILVERRUSH-sized fields of view. Here, we show the probability distribution of the surface density contrast, $\delta = \Sigma/\langle\Sigma\rangle - 1$. The top and bottom panels correspond to $z = 5.7$ and $z = 6.6$, respectively. The red, blue, and green histograms represent our models with different global neutral fractions. The vertical black lines represent δ measured from the publicly available SILVERRUSH catalogues (Shibuya et al. 2018), with uniform luminosity and colour cuts applied. The red, thin-dashed lines show the 2σ range of our surface density distribution in a completely ionized scenario. Black lines are annotated with individual field names. A more neutral IGM yields only a slight increase in the cosmic variance. Our model distributions are consistent with the data, aside from D-ELAISN1, which is likely affected by poor seeing (Konno et al. 2018).

For comparison, we attempt to quantify the split-screen contrasts in the SILVERRUSH data using the same luminosity and colour cuts described in the last section. We note that the fields have different sizes and non-trivial shapes. To account for this, we inspected the fields and found axes of symmetry along which to perform the bisections, maintaining roughly equal areas between the split screens. The black histograms in Fig. 5 show the results of our crude measurements, which are reported in Table 2. (Individual fields have more than one split-screen configuration.) The largest contrasts are found in D-ELAISN1 at $z = 5.7$, and in D-COSMOS and UD-COSMOS at $z = 6.6$. Note that we have applied our own homogeneous luminosity and colour cuts to the data, so the split-screen contrasts that we report here may not be visually apparent in published visualizations of the fields (e.g. Ouchi et al. 2018).

Our measurements from SILVERRUSH lie comfortably within the dispersion of our models. A Kolmogorov–Smirnov test indicates that the observed distributions are consistent with all of the models.
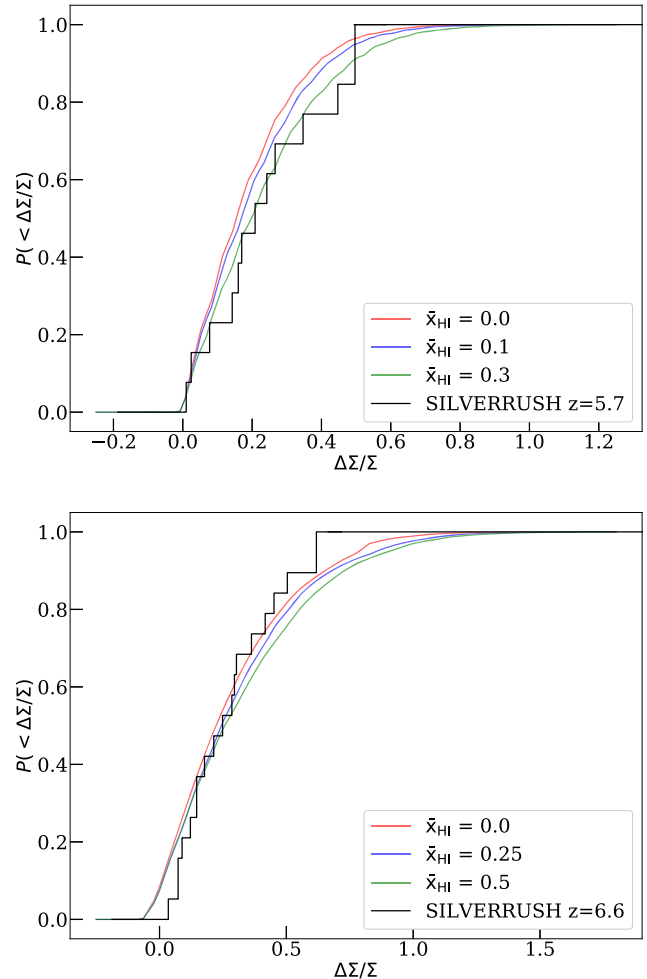

Figure 5. Cumulative probability distribution of the split-screen contrast, $\Delta\Sigma/\Sigma$. The split-screen contrast quantifies half-plane asymmetry in the LAE distribution (see main text for more details). The top and bottom panels correspond to $z = 5.7$ and $z = 6.6$, respectively. The coloured curves correspond to our reionization models. The black histograms show measurements from the publicly available SILVERRUSH catalogues (Shibuya et al. 2018), with uniform luminosity and colour cuts applied. The likelihood of finding a field with high $\Delta\Sigma/\Sigma$ increases slightly with neutral gas fraction. However, the visually apparent half-plane asymmetries seen in some of the SILVERRUSH fields are consistent with expectations from large-scale structure formation alone.

For example, at $z = 5.7$, we find p-values of 0.8 and 0.3 for $\bar{x}_{HI} = 0.3$ and 0.0, respectively. We conclude that the half-plane asymmetries present in the SILVERRUSH data are consistent with the expected large-scale structure alone. Patchy reionization does, however, enhance the probability of observing larger split-screen contrasts. For example, the probabilities of finding $\Delta\Sigma/\langle\Sigma\rangle \geq 0.5$ are 4(21) per cent, 6(23) per cent, 10(27) per cent in our models with $\bar{x}_{HI} = 0.0(0.0)$, 0.1(0.3), and 0.3(0.5) at $z = 5.7(6.6)$, respectively.

In conclusion, we have found that our ensembles of simulated fields are consistent with the cosmic variance and half-plane asymmetries observed in the SILVERRUSH data.

4.2 Voids

Clustering statistics of LAEs provide an important observational window into the reionization process. Previous studies have focused on using the 2PCF to constrain the neutral fraction of the IGM

(McQuinn et al. 2007; Ouchi et al. 2010, 2018; Jensen et al. 2014; Hutter et al. 2015; Sobacchi & Mesinger 2015). However, Fig. 2 illustrates that, towards the end of reionization, neutral gas is mostly relegated to voids in the LAE distribution. This suggests that void statistics could be more informative in the final stages of reionization, when the neutral gas disproportionately obscures LAEs in underdense regions. Also working in favour of this idea is the fact that the space density of $z = 5.7$ LAEs is considerably larger than at e.g. $z = 6.6$. In what follows, we examine the prospect of constraining reionization with the statistics of voids.

We quantify reionization's effect on LAE void statistics using the void probability function (VPF), or the probability of finding a circle of radius R containing no LAEs (Tinker, Weinberg & Warren 2006; McQuinn et al. 2007; Tinker et al. 2008; Perez et al. 2020). The VPF is complementary to the 2PCF, encoding correlations at orders beyond the two-point function. A key advantage of the VPF is that it is simple and can be applied even to sparse LAE fields, although see Perez et al. (2020) for important caveats on its application. We compute the VPF by randomly placing down circles with radius R in our mock fields and then finding the fraction of circles that do not contain LAEs.

In Fig. 6, we show our VPF results at $z = 5.7$ (top) and $z = 6.6$ (bottom). The red, blue, and green curves show our reionization scenarios as denoted in the plot legends. For comparison, the black curves show the VPF in the case of a uniform-random distribution of LAEs, $P(R) = \exp(-\pi R^2 \langle \Sigma \rangle)$, with the same mean surface density (see Table 1 for $\langle \Sigma \rangle$). A comparison between the black and red curves shows that large-scale structure clearly enhances the VPF above the Poisson expectation. Consistent with previous works, we find that an ongoing reionization process further enhances the VPF. Neutral gas in the cosmic voids attenuates Lyα emission, making holes in the LAE distribution appear larger (cf. Figs 2 and 3). A higher global neutral fraction enhances the probability of finding empty circles on larger scales. At $z = 6.6$, we find that the probability of finding larger empty circles is enhanced significantly. (Note the different scales in the panels.) This owes to two effects. First, there are simply fewer LAEs at $z = 6.6$, which results in larger random gaps in the distribution (see D'Aloisio & Furlanetto 2007 for a discussion). Secondly, the larger neutral fraction results in more LAEs being obscured.

Even on the $L \sim 200\ h^{-1}$ Mpc scales of our fields, large-scale power can lead to substantial cosmic variance in our void statistics. Such variations can pose a challenge for constraining reionization with voids. We quantify the cosmic variance by bootstrap resampling our simulated FoV ensemble. If our mock survey contains N fields, we draw N samples with replacement and calculate the VPF from these samples. The resulting VPF constitutes one realization of a mock survey, and we generate 10 000 of these. We have checked that our results are well converged in the number of realizations. In what follows, we consider two survey configurations; (1) The current SILVERRUSH size: $N = 4$ and 5 for $z = 5.7$ and 6.6, respectively; (2) A future SILVERRUSH-like survey covering twice the area, $N = 10$. The error bars in Fig. 6 show the 1σ dispersion from cosmic variance in the VPF assuming the futuristic $N = 10$ field survey. From here on, all cosmic variance error bars assume $N = 10$.

Can the VPF provide stronger constraints than the 2PCF on the tail end of reionization, as we have argued above? We address this question with a simple χ^2 analysis. We gauge the significance at which a statistic can rule out our model with $\bar{x}_{\mathrm{HI}} = 0.3$ (0.5) at $z = 5.7$ (6.6), assuming that the true model is $\bar{x}_{\mathrm{HI}} = 0$. To this end, we used our bootstrap samples of the $\bar{x}_{\mathrm{HI}} = 0$ model to compute the
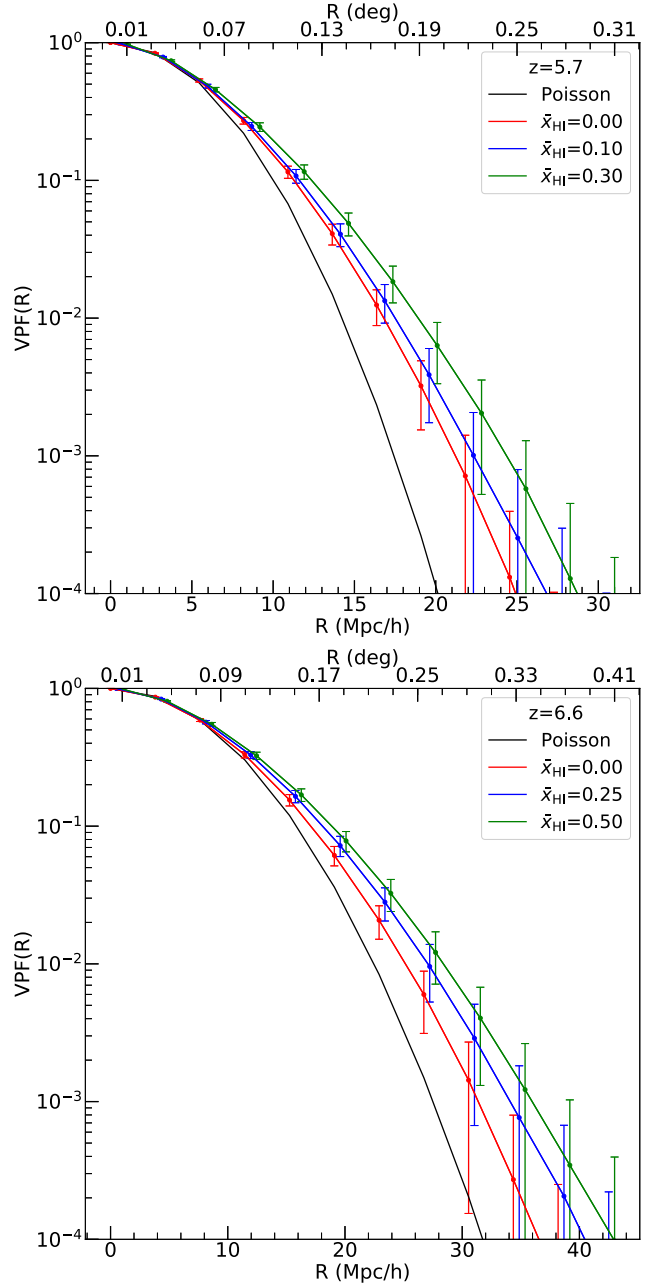

Figure 6. The VPF at $z = 5.7$ (top) and $z = 6.6$ (bottom) for our models with different global neutral fractions. The black curves show the analytic result for a uniform-random distribution of LAEs with matched surface density. The error bars show the 1σ dispersion from cosmic variance in a 10-field survey, or a SILVERRUSH-like survey with roughly double the size. A larger neutral fraction increases the probability of finding larger empty circles. This is a consequence of LAEs being obscured in the low-density regions of the IGM, which are the last places to be reionized.

covariance matrix,[6] taking into account correlations between radial bins. We then calculated $\chi^2 = \boldsymbol{P}^T C^{-1} \boldsymbol{P}$, where C is the covariance matrix and $\boldsymbol{P}$ is the difference vector between the 0.3(0.5) model and a mock survey drawn from the $\bar{x}_{\mathrm{HI}} = 0$ model. Finally, we calculated

---

[6]We have checked that our calculation is well converged in number of realizations.

Table 3. Simple χ^2 analysis to gauge the efficacy of void and peak statistics for testing reionization models. The reference models have $\bar{x}_{\mathrm{H\,I}} = 0$ and we compute the probabilities of obtaining χ^2 at least as extreme as those of the $\bar{x}_{\mathrm{H\,I}} = 0.3$ and 0.5 models at $z = 5.7$ and 6.6, respectively, assuming cosmic variance uncertainties. These p-values are shown for two survey configurations: (1) A SILVERRUSH-like survey with 4 and 5 fields at $z = 5.7$ and $z = 6.6$, respectively; (2) An expanded SILVERRUSH-like survey with 10 fields (roughly doubled in area). We consider the 2PCF, VPF (Section 4.2), Pixel Threshold Fraction (PTF; Section 4.3.1), and Peak Probability Function (PPF; Section 4.3.3). The numbers shown in parentheses for the 2PCF and VPF assume that 10 per cent or 30 per cent of the LAEs are misidentified foreground sources. All other numbers assume no contamination. The VPF is particularly well suited for probing the tail end of reionization, where LAE surface densities are highest.

z	Statistic	p-value (SILVERRUSH-like)	p-value (10 fields)
5.7	2PCF	0.26	0.18
	10 per cent contamination	(0.28)	(0.22)
	30 per cent contamination	(0.34)	(0.32)
	VPF	0.18	0.07
	10 per cent contamination	(0.23)	(0.12)
	30 per cent contamination	(0.31)	(0.23)
	PTF	0.42	0.30
	PPF	0.63	0.51
6.6	2PCF	0.19	0.08
	10 per cent contamination	(0.26)	(0.12)
	30 per cent contamination	(0.32)	(0.21)
	VPF	0.21	0.09
	10 per cent contamination	(0.25)	(0.14)
	30 per cent contamination	(0.34)	(0.23)
	PTF	0.19	0.10
	PPF	0.58	0.25

p-values to assess the rarity of the obtained χ^2. Given the non-Gaussianity exhibited in some of the radial bins, we integrated the numerical χ^2-distribution from the bootstrapping to compute the p-values.

A summary of our simple χ^2 analysis is reported in Table 3. There we show p-values for the 2PCF, VPF, as well as two of the peak-based statistics that we explore in the next section. To ensure that our conclusions are not biased by a random fluctuation, the p-values reported there are the average values over 1000 mock observations. Results are given for current and 'futuristic' (10-field) survey configurations. We note that our analysis adopts cosmic variance error bars, ignoring all other observational uncertainties. The numbers in parentheses correspond to cases in which we assume that a fraction of the LAEs are actually misidentified foreground interlopers. (See discussion below.) Let us consider first the ideal case with no contamination. We find that the VPF outperforms the 2PCF at testing late reionization models in which the tail end extends below $z = 5.7$. On the other hand, for both survey configurations, the 2PCF and VPF provide similar constraining power for a model with $\bar{x}_{\mathrm{H\,I}} = 0.5$ at $z = 6.6$. These results have an intuitive interpretation. The neutral gas is mostly in the voids at $z = 5.7$, leaving a stronger signature in the VPF compared to the 2PCF.

In the absence of full spectroscopic confirmation, however, narrowband surveys are inevitably contaminated by misidentified lower redshift sources. The contamination could be particularly problematic for the VPF, which is exponentially sensitive to interlopers (McQuinn et al. 2008). To gauge this effect, we selected a fraction of our LAEs and shuffled them to random locations in the field. This procedure ignores the clustering of the interlopers, which is

likely to be a small effect for the low number densities that we assume here. Based on spectroscopic followups, Ouchi et al. (2018) estimated contamination rates of $\approx$ 8 per cent and $\approx$ 14 per cent in their $z = 5.7$ and $z = 6.6$ samples, respectively. Ouchi et al. (2018) also found that the contamination rate depends on source brightness. They reported a higher rate of $\approx$ 33 per cent for sources brighter than 24 magnitudes in their $z = 6.6$ sample. In Table 3, the p-values in parentheses assume contamination fractions of 10 per cent and 30 per cent. We find that the contamination significantly degrades the sensitivities of both the VPF and 2PCF. However, for the 10-field configuration, we find that the VPF maintains an advantage over the 2PCF at $z = 5.7$, even in the presence of significant contamination.

Our results suggest that the VPF can provide information on reionization that is complementary to the 2PCF. We find that the former is a more sensitive test of the final stages of reionization in which neutral gas is contained primarily in the voids, and LAE surface densities are highest.

4.3 Peak statistics

LAE peak statistics offer a complementary picture to the voids. When comparing models with $\bar{x}_{\mathrm{H\,I}} = 0$ and 0.3 at fixed $\langle \Sigma \rangle$, any LAEs that are obscured in the voids of the latter model must be compensated by other LAEs unveiled elsewhere. Thus, we also expect the statistics of LAE overdensities to be different between the two models (see Section 3). We quantify this effect here. We explore three different approaches, starting with the simplest statistic based on thresholding the LAE field.

4.3.1 Pixel threshold fraction

We begin by counting the fraction of pixels exceeding an overdensity threshold in a smoothed map of the LAE distribution. The LAE distribution is first smoothed on to a uniform 2D grid using a Gaussian filter with a smoothing scale of $\sigma = 10\ h^{-1}$ Mpc. We employ a grid with $N = 512^2$, but our conclusions are broadly robust to the exact choice. The size of the smoothing scale was adopted from previous studies of galaxy proto-clusters at $z = 5.7$ and 6.6 (e.g. Chiang et al. 2017). It is roughly the spatial extent of overdense regions that later collapse into clusters at $z \sim 0$. We then count the fraction of pixels above a given surface overdensity $\Sigma_{10}(\mathbf{x})/\langle \Sigma \rangle$, where the subscript denotes that we have smoothed over a scale of $10\ h^{-1}$ Mpc. We set the minimum threshold to be the 95th percentile in the $\bar{x}_{\mathrm{H\,I}} = 0$ models, since we find that differences between the models are strongest at the largest overdensities. For reference, this minimum threshold corresponds to $2\langle \Sigma \rangle$ and $2.5\langle \Sigma \rangle$ at $z = 5.7$ and 6.6, respectively.[7] We refer to this statistic, $F(> \Sigma_{10}/\langle \Sigma \rangle)$, as the PTF.

We show our results for the PTF in Fig. 7. Consistent with expectations, the fraction of overdense pixels increases with the neutral fraction for fixed $\langle \Sigma \rangle$. This is particularly prominent among our models at $z = 6.6$. The differences in the PTF reflect an enhancement in the apparent clustering of LAEs due to patchy reionization, offering an alternative to the 2PCF for quantifying this

---

[7]Our thresholds are lower than those of previous studies that define clusters and proto-clusters with surface-densities greater than 4–5σ over the mean density (e.g. Inoue et al. 2018; Harikane et al. 2019). In this section, we use the term 'LAE peak' to distinguish between our liberal definition of a high-density region and the more stringent classification of proto-clusters used in those works (Inoue et al. 2018; Harikane et al. 2019).
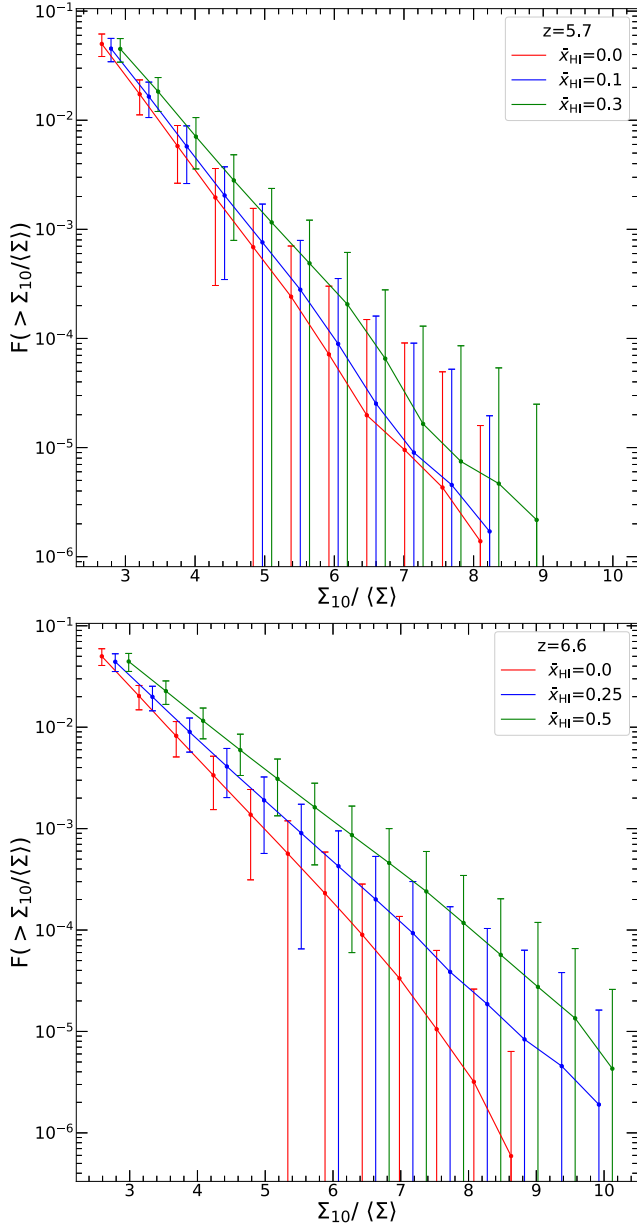
Figure 7. The PTF, or fraction of pixels with LAE surface densities above $\Sigma_{10}/\langle\Sigma\rangle$. Here, Σ_{10} is the surface density smoothed over a scale of $\sigma = 10 \; h^{-1}$ Mpc with a Gaussian filter. The error bars show the 1σ dispersion from cosmic variance in an expanded (10-field) SILVERRUSH-like survey. At a given redshift, all models shown have a fixed $\langle\Sigma\rangle$ (see Table 1). In this case, a more neutral IGM leads to a larger fraction of overdense pixels, a consequence of conservation of LAE number. LAEs that are obscured in voids are compensated by more LAEs being visible elsewhere, which leads to an apparent enhancement in clustering.

effect. The error bars in Fig. 7 show the 1σ cosmic variance obtained from the bootstrap method described in Section 4.2. The results of our χ^2 analysis on the PTF are also shown in Table 3. At $z = 6.6$, the PTF can discriminate the $\bar{x}_{HI} = 0.5$ and $\bar{x}_{HI} = 0.0$ models at a similar level to the 2PCF and the VPF. However, at $z = 5.7$, the PTF is significantly weaker. This behaviour may owe to the fact that the LAEs in our models are more highly biased at $z = 6.6$.

4.3.2 Peak overdensity function

Next, we consider a more sophisticated approach that groups together high surface density pixels ('peak pixels'). This approach has the added benefit that we can analyse the effect of reionization on the morphology of LAE peaks. Peak pixels are identified in our smoothed surface density maps using the same threshold as in the last section. They are then grouped into connected sets that we call peaks. To avoid edge effects, we eliminate peaks whose boundaries are one pixel from any edge of the field. We count the number of LAEs contained within each peak and eliminate peaks with less than 8(5) LAEs at $z = 5.7(6.6)$, which, we find, reduces the effects of shot noise and grid resolution on our results. This procedure yields catalogues of bounded regions of the highest intensity pixels.

In Fig. 8, we show example applications of our peak finding algorithm to mock fields at $z = 5.7$ (top row) and $z = 6.6$ (bottom row). The columns correspond to different neutral fractions as denoted in the plot labels. In each panel, the dots show the LAE distributions, while the colour maps correspond to the smoothed surface density contrasts. The solid curves bound the LAE peaks selected by our algorithm while the dashed curves show regions that did not make our final cut owing to low LAE occupancy (i.e. not satisfying the minimum occupancy of 8(5) at $z = 5.7$ and 6.6).

We have quantified the shapes of LAE peaks by computing moments of their surface density distributions. We found that the shapes (e.g. ellipticities) and sizes of the peaks are statistically quite similar between our models, with variations ~ 5 per cent. We do not consider these statistics further here. Instead we focus on the abundance and occupancy of peaks, which do show significant differences. At fixed $\langle\Sigma\rangle$, a more neutral IGM leads to more LAEs per peak. In particular, models with a higher neutral fraction contain more peaks with high LAE occupancy. To demonstrate this, we consider the number of peaks per unit area, per unit surface density, $dN/d\Delta_\Sigma$, where Δ_Σ is the surface density of a peak in units of the global mean. We term $dN/d\Delta_\Sigma$ the Peak Overdensity Function (POF).

Fig. 9 shows the POF for our models at $z = 5.7$ (top) and $z = 6.6$ (bottom). We find that Δ_Σ ranges from 3 to 5 at $z = 5.7$ and 4 to 7 at $z = 6.6$, generally reflecting the stronger bias of LAEs at higher redshifts. The number of overdense peaks clearly increases with the neutral fraction, consistent with our arguments above. However, the error bars show considerable cosmic variance in even a future 10-field SILVERRUSH-like survey. Compared to the PTF, we find that the POF offers a much weaker discriminator between our models owing to the rarity of peaks. Indeed, Fig. 8 illustrates that only 2–3 peaks are identified per FoV. Thus, we did not consider the POF in our χ^2 comparison.

4.3.3 Peak probability function

The last statistic that we consider is an analogue to the VPF of Section 4.2. We define the PPF to be the probability that a randomly placed circle of radius R has an overdensity of LAEs greater than or equal to some threshold. We randomly place down circles of radius R, find the number of LAEs (N) contained within the circle, and then calculate the surface density, $\Sigma_R = N/\pi R^2$. Using the same thresholds from Section 4.3.1, we count only circles with $\Sigma_R > 2(2.5)\langle\Sigma\rangle$ at $z = 5.7(6.6)$. We only consider R, such that there is a minimum of 10 LAE contained within a circle. We then divide the number of circles making the cut by the total number drawn to find the probability as a function of R.

We show our results in Fig. 10. Consistent with the previous statistics, we find that the probability of obtaining a peak increases
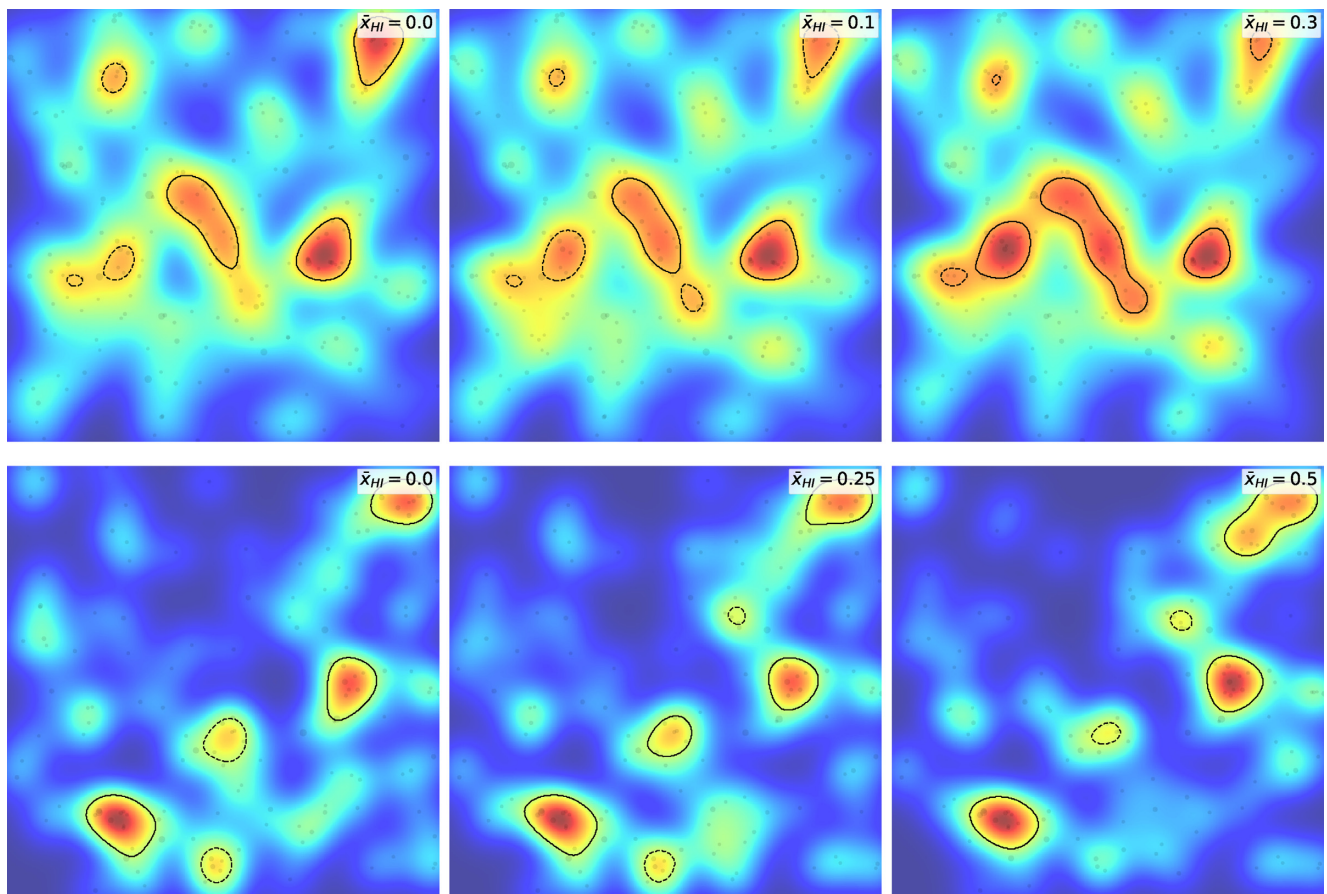
Figure 8. Examples of our peak finding algorithm described in Section 4.3.2. The colour maps show the surface density contrast of the LAE field smoothed on a $\sigma = 10~h^{-1}$ Mpc scale with a Gaussian filter. The top and bottom rows correspond to $z = 5.7$ and $z = 6.6$, respectively, while the columns correspond to models with different global neutral fractions. The top panels are $180 \times 180~(h^{-1}$ Mpc$)^2$ while the bottom panels are $200 \times 200~(h^{-1}$ Mpc$)^2$. The solid black contours enclose regions identified as peaks in the LAE distribution. For reference, the dashed contours show overdense regions that did not fulfill our LAE occupancy cut (see main text for details).

with the neutral fraction. This further cements the conclusion that, at fixed $\langle \Sigma \rangle$, a higher neutral fraction yields a higher abundance of overdense peaks in the LAE distribution. The results of our χ^2 analysis for the PPF are given in Table 3. Perhaps surprisingly, the PPF is not as strong of a test of our models as the other statistics we considered, and is significantly weaker than the 2PCF. We attribute this finding to strong correlations between the radial bins, which are reflected in the highly off-diagonal structure of the covariance matrix for the PPF. The physical picture for this is as follows. Owing to the bias of LAEs, if a peak of a given radius R is found, then the probability of finding additional peaks with other values of R is enhanced significantly. On other hand, the correlations between radial bins of the VPF are weaker. Finding a random hole in the LAE distribution does not necessarily enhance the probability of finding a larger hole, for example.

5 CONCLUSION

In this paper, we have explored the use of narrowband LAE surveys to constrain late reionization scenarios, with an eye towards recently proposed models in which reionization ends as late as $z \approx 5$ (Kulkarni et al. 2019; Keating et al. 2020; Nasir & D'Aloisio 2020). During the tail end of reionization, with neutral fractions ~ 10 per cent, the last remaining neutral islands in the Universe were likely relegated

to voids in the galaxy distribution. The neutral islands would have preferentially obscured LAEs in underdense regions, motivating us to explore LAE void statistics further in this paper. When comparing reionization models at a fixed mean surface density, LAEs obscured in the voids must be compensated by visible LAEs elsewhere in the spatial distribution. Thus, we have also explored peak statistics as a complementary discriminator of reionization models.

For several reionization scenarios spanning global neutral fractions of $\bar{x}_{\mathrm{HI}} = 0$–0.3 at $z = 5.7$, and $\bar{x}_{\mathrm{HI}} = 0$–0.5 at $z = 6.6$, we have constructed mock LAE surveys using one of the $L = (1~h^{-1}\mathrm{Gpc})^3$ Multi-Dark simulations as a basis. Our models were tuned to match the LAE surface densities observed in the recently published SILVERRUSH survey, as well as a number of other observational constraints. The enormous simulation volume of Multi-Dark allowed us to capture the $\sim$10–100 Mpc characteristic scales of reionization, and provided representative models for comparison against SILVERRUSH, which covers an area of 14–21 deg^2 (0.3–0.5 Gpc^2). One of our primary focuses was to quantify the cosmic variance expected in a SILVERRUSH-like survey. To this end, we constructed a statistical ensemble of 1000 mock fields of view.

The SILVERRUSH data exhibit significant field-to-field variation in surface density, and some fields have visual asymmetries in the LAE distribution over $\sim$200 Mpc scales. We compared the SILVERRUSH data against our models and found that the surface density
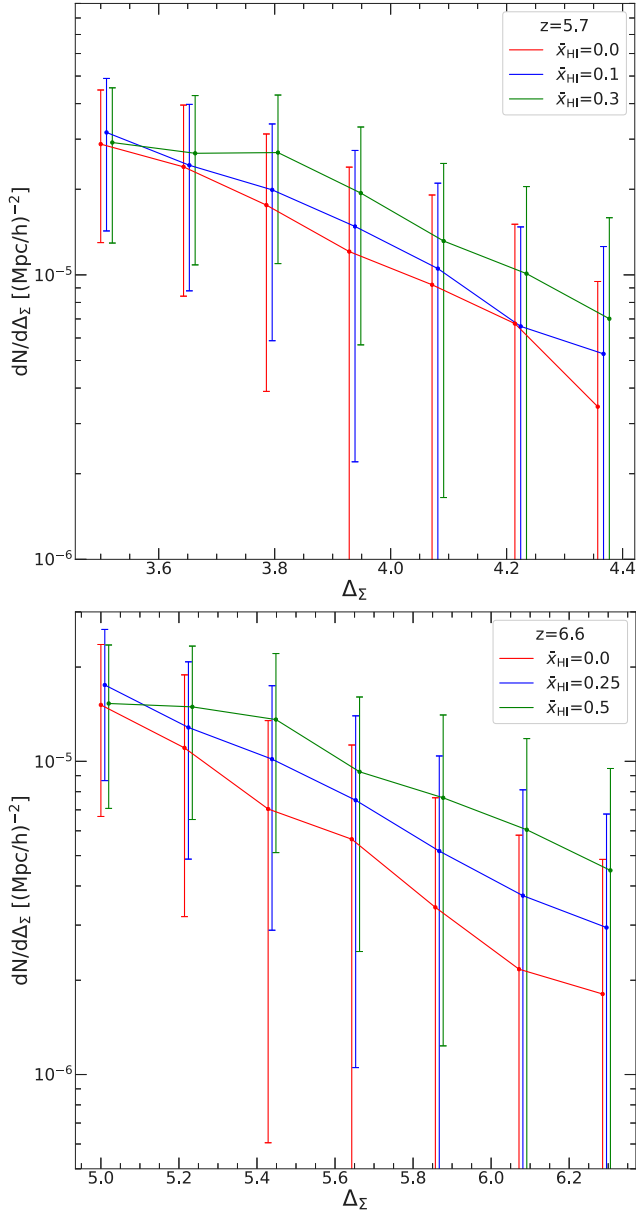
Figure 9. The POF, $dN/d\Delta_\Sigma$, or number of peaks per unit area, per unit surface overdensity, $\Delta_\Sigma = \Sigma/\langle\Sigma\rangle$. For this statistic, peaks are identified as connected sets of pixels satisfying a minimum overdensity threshold, as well as a minimum LAE occupancy (see main text for details). Error bars show the 1σ dispersion from cosmic variance in an expanded (10-field) SILVERRUSH-like survey. At fixed $\langle\Sigma\rangle$, the overdense LAE peaks are more numerous in models with higher neutral fraction.
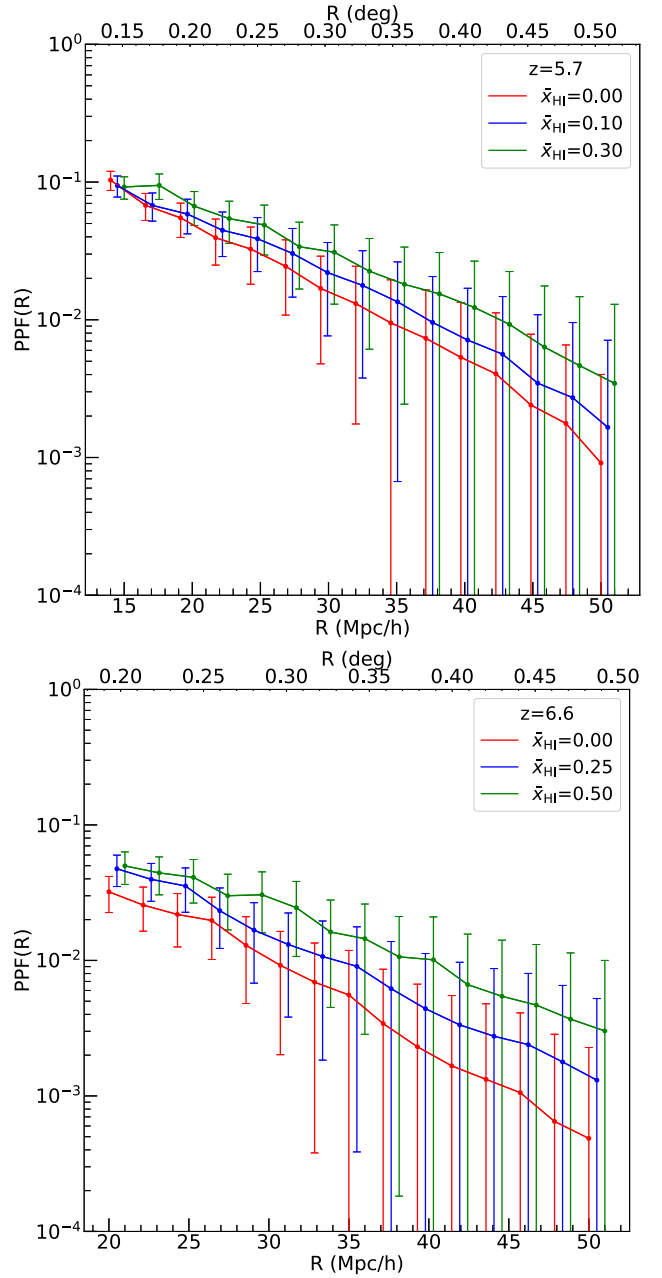

Figure 10. The Peak Probabiliy Function (PPF), or probability of finding an LAE overdensity greater than some threshold within a circle of radius R. We adopt a threshold of $2(2.5)\langle\Sigma\rangle$ at $z = 5.7(6.6)$ (see main text for details). Error bars show the 1σ dispersion from cosmic variance in an expanded (10-field) SILVERRUSH-like survey. At fixed $\langle\Sigma\rangle$, higher neutral fractions yield increased probabilities of finding LAE peaks.

variations as well as the large-scale asymmetries of the former are well within expectations from large-scale structure formation alone. An exception to this conclusion is the field D-ELAISN1, which has ≈ 20 per cent fewer LAEs than the lowest surface density observed in our ensemble. Interestingly, D-ELAISN1 also exhibits among the largest half-plane asymmetries in its LAE distribution. However, Konno et al. (2018) have attributed the exceptional features of D-ELAISN1 to poor seeing, perhaps reconciling it with our models. We generally find that patchy reionization contributes only slightly to the cosmic variance and large-scale (~200 Mpc) asymmetry of

fields, and is an unlikely explanation for the characteristics of D-ELAISN1.

Next, we considered the VPF as a probe of reionization models. A more neutral IGM yields a higher probability of finding large-scale empty regions in the LAE distribution. We found with a simple χ^2 calculation that the VPF is a more sensitive test of models during the last stages of reionization than the widely used angular 2PCF. In spite of the VPF being exponentially sensitive to interlopers, this conclusion holds even if we assume that 10–30 per cent of the sources are misidentified foreground objects.

We also explored three approaches to quantifying the statistics of peaks in the LAE distribution. Using a method that groups overdense pixels in surface density maps smoothed over $10\ h^{-1}$ Mpc scales, we found that patchy reionization has little effect statistically on the shapes of LAE peaks. When comparing models at fixed mean surface density, the predominant effect of neutral gas is to increase the probability of finding LAE overdensities. We found that this effect can be quantified straightforwardly by thresholding smoothed LAE maps and counting the fraction of overdense pixels. Our simplistic χ^2 analysis showed that this peak statistic offers roughly the same sensitivity as the 2PCF and VPF for discriminating between models with $\bar{x}_{H_I} = 0.0$ and $\bar{x}_{H_I} = 0.5$. However, it is generally less sensitive during the last stages of reionization.

In this work, we have focused on constraining the latter half of reionization with a SILVERRUSH-like survey. However, SILVER-RUSH and The Lyman Alpha Galaxies in the Epoch of Reionization (Zheng et al. 2017) project will continue to expand the sample of known LAEs at ≥ 7. Future work should address in further detail the efficacy for $z \sim 7$ LAEs to constrain reionization, as well as the trade-off between survey area and depth (see G60). We note that all of our calculations have assumed that reionization proceeded in an inside-out manner, consistent with seminumeric models and more detailed radiative transfer simulations. Future studies might further explore whether a combination of statistics could provide more fundamental insight into the connection between reionization and the underlying density field.

In recent years, the sample of known LAEs at $z \geq 5.7$ has grown to over 2000, raising the prospects for detecting the last stages of reionization in the clustering of LAEs. Our calculations suggest that the statistics of voids and peaks offer complimentary approaches to the 2PCF, and may even be more optimal under some conditions.

ACKNOWLEDGEMENTS

DATA AVAILABILITY

The data and code utilized in this work will be shared on a request to the corresponding author.

REFERENCES

Becker G. D., Bolton J. S., Madau P., Pettini M., Ryan-Weber E. V., Venemans B. P., 2015, MNRAS, 447, 3402
Becker G. D., Davies F. B., Furlanetto S. R., Malkan M. A., Boera E., Douglass C., 2018, ApJ, 863, 92
Bolton J. S., Haehnelt M. G., 2013, MNRAS, 429, 1695
Bosman S. E. I., Fan X., Jiang L., Reed S., Matsuoka Y., Becker G., Haehnelt M., 2018, MNRAS, 479, 1055
Bouwens R. J. et al., 2015, ApJ, 803, 34
Caruana J., Bunker A. J., Wilkins S. M., Stanway E. R., Lacy M., Jarvis M. J., Lorenzoni S., Hickey S., 2012, MNRAS, 427, 3055
Caruana J., Bunker A. J., Wilkins S. M., Stanway E. R., Lorenzoni S., Jarvis M. J., Ebert H., 2014, MNRAS, 443, 2831
Chardin J., Kulkarni G., Haehnelt M. G., 2018, MNRAS, 478, 1065
Chiang Y.-K., Overzier R. A., Gebhardt K., Henriques B., 2017, ApJ, 844, L23
D'Aloisio A., Furlanetto S. R., 2007, MNRAS, 382, 860
D'Aloisio A., McQuinn M., Davies F. B., Furlanetto S. R., 2018, MNRAS, 473, 560
D'Aloisio A., McQuinn M., Trac H., Cain C., Mesinger A., 2020, ApJ, 898, 149
Davies F. B., Furlanetto S. R., 2016, MNRAS, 460, 1328
Davies F. B. et al., 2018, ApJ, 864, 142
Dijkstra M., Wyithe J. S. B., 2012, MNRAS, 419, 3181
Doussot A., Trac H., Cen R., 2019, ApJ, 870, 18
Eilers A.-C., Davies F. B., Hennawi J. F., 2018, ApJ, 864, 53
Endsley R., Stark D. P., Chevallard J., Charlot S., Robertson B., Bouwens R. J., Stefanon M., 2020, preprint (arXiv:2010.03566)
Fan X. et al., 2006, AJ, 132, 117
Finkelstein S. L. et al., 2015, ApJ, 810, 71
Furlanetto S. R., Zaldarriaga M., Hernquist L., 2004, ApJ, 613, 1
Furlanetto S. R., Zaldarriaga M., Hernquist L., 2006, MNRAS, 365, 1012
Giri S. K., Mellema G., Aldheimer T., Dixon K. L., Iliev I. T., 2019, MNRAS, 489, 1590
Harikane Y. et al., 2016, ApJ, 821, 123
Harikane Y. et al., 2019, ApJ, 883, 142
Harris C. R. et al., 2020, Nature, 585, 357
Hassan S., Gronke M., 2020, preprint (arXiv:2010.00023)
Hoag A. et al., 2019, ApJ, 878, 12
Hu E. M., Cowie L. L., Songaila A., Barger A. J., Rosenwasser B., Wold I. G. B., 2016, ApJ, 825, L7
Hu W. et al., 2019, ApJ, 886, 90
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Hutter A., Dayal P., Müller V., 2015, MNRAS, 450, 4025
Inoue A. K. et al., 2018, PASJ, 70, 3
Jensen H., Hayes M., Iliev I. T., Laursen P., Mellema G., Zackrisson E., 2014, MNRAS, 444, 2114
Kakiichi K. et al., 2018, MNRAS, 479, 43
Kashikawa N. et al., 2006, ApJ, 648, 7
Keating L. C., Weinberger L. H., Kulkarni G., Haehnelt M. G., Chardin J., Aubert D., 2020, MNRAS, 491, 1736
Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, MNRAS, 457, 4340
Konno A. et al., 2018, PASJ, 70, S16
Kulkarni G., Keating L. C., Haehnelt M. G., Bosman S. E. I., Puchwein E., Chardin J., Aubert D., 2019, MNRAS, 485, L24
Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473
McGreer I. D., Mesinger A., D'Odorico V., 2015, MNRAS, 447, 499
McQuinn M., Hernquist L., Zaldarriaga M., Dutta S., 2007, MNRAS, 381, 75
McQuinn M., Lidz A., Zaldarriaga M., Hernquist L., Dutta S., 2008, MNRAS, 388, 1101
Mason C. A., Gronke M., 2020, MNRAS, 499, 1395
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
Matthee J., Sobral D., Gronke M., Paulino-Afonso A., Stefanon M., Röttgering H., 2018, A&A, 619, A136
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
Meyer R. A. et al., 2020, MNRAS, 494, 1560
Mortlock D. J. et al., 2011, Nature, 474, 616
Nasir F., D'Aloisio A., 2020, MNRAS, 494, 3080

Ono Y. et al., 2012, ApJ, 744, 83
Ota K. et al., 2008, ApJ, 677, 12
Ota K. et al., 2010, ApJ, 722, 803
Ouchi M. et al., 2010, ApJ, 723, 869
Ouchi M. et al., 2018, PASJ, 70, S13
Pentericci L. et al., 2014, ApJ, 793, 113
Perez L. A., Malhotra S., Rhoads J. E., Tilvi V., 2020, ApJ, 906, 17
Planck Collaboration VI, 2020, A&A, 641, A6
Rahmati A., Pawlik A. H., Raičević M., Schaye J., 2013, MNRAS, 430, 2427
Sadoun R., Zheng Z., Miralda-Escudé J., 2017, ApJ, 839, 44
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
Schmidt K. B. et al., 2016, ApJ, 818, 38
Shibuya T. et al., 2018, PASJ, 70, S14
Sobacchi E., Mesinger A., 2015, MNRAS, 453, 1843
Songaila A., Hu E. M., Barger A. J., Cowie L. L., Hasinger G., Rosenwasser B., Waters C., 2018, ApJ, 859, 91
Springel V., 2005, MNRAS, 364, 1105
Stark D. P. et al., 2017, MNRAS, 464, 469
Tinker J. L., Weinberg D. H., Warren M. S., 2006, ApJ, 647, 737
Tinker J. L., Conroy C., Norberg P., Patiri S. G., Weinberg D. H., Warren M. S., 2008, ApJ, 686, 53
Trac H., Pen U.-L., 2004, New Astron., 9, 443
Trenti M., Stiavelli M., Bouwens R. J., Oesch P., Shull J. M., Illingworth G. D., Bradley L. D., Carollo C. M., 2010, ApJ, 714, L202
van der Walt S. et al., 2014, PeerJ, 2, e453
Virtanen P. et al., 2020, Nat. Methods, 17, 261
Wang F. et al., 2019, ApJ, 884, 30
Wang F. et al., 2020, ApJ, 896, 23
Weinberger L. H., Kulkarni G., Haehnelt M. G., Choudhury T. R., Puchwein E., 2018, MNRAS, 479, 2564
Weinberger L. H., Haehnelt M. G., Kulkarni G., 2019, MNRAS, 485, 1350
Wes McKinney, 2010, in van der Walt S., Millman J., eds, Proc. 9th Python in Science Conference. p. 56
Worseck G. et al., 2014, MNRAS, 445, 1745
Wu X., McQuinn M., Kannan R., D'Aloisio A., Bird S., Marinacci F., Davé R., Hernquist L., 2019, MNRAS, 490, 3177
Zahn O., Mesinger A., McQuinn M., Trac H., Cen R., Hernquist L. E., 2011, MNRAS, 414, 727
Zheng Z., Miralda-Escudé J., 2002, ApJ, 578, 33
Zheng Z.-Y. et al., 2017, ApJ, 842, L22

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