MNRAS 000, 1–23 (2022) Preprint 5 April 2022 Compiled using MNRAS LATEX style file v3.0

The THESAN project: Lyman-α emission and transmission during the Epoch of Reionization

A. Smith,[1]★ R. Kannan,[2]† E. Garaldi,[3]‡ M. Vogelsberger,[1] R. Pakmor,[3] V. Springel[3] and L. Hernquist[2]
[1]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[2]Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA
[3]Max-Planck Institute for Astrophysics, Karl-Schwarzschild-Str. 1, D-85741 Garching, Germany

arXiv:2110.02966v2 [astro-ph.CO] 2 Apr 2022

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

The visibility of high-redshift Lyman-alpha emitting galaxies (LAEs) provides important constraints on galaxy formation processes and the Epoch of Reionization (EoR). However, predicting realistic and representative statistics for comparison with observations represents a significant challenge in the context of large-volume cosmological simulations. The THESAN project offers a unique framework for addressing such limitations by combining state-of-the-art galaxy formation (IllustrisTNG) and dust models with the AREPO-RT radiation-magnetohydrodynamics solver. In this initial study we present Lyman-alpha centric analysis for the flagship simulation that resolves atomic cooling haloes throughout a $(95.5\,\mathrm{cMpc})^3$ region of the Universe. To avoid numerical artefacts we devise a novel method for accurate frequency-dependent line radiative transfer in the presence of continuous Hubble flow, transferable to broader astrophysical applications as well. Our scalable approach highlights the utility of LAEs and red damping-wing transmission as probes of reionization, which reveal nontrivial trends across different galaxies, sightlines, and frequency bands that can be modelled in the framework of covering fractions. In fact, after accounting for environmental factors influencing large-scale ionized bubble formation such as redshift and UV magnitude, the variation across galaxies and sightlines mainly depends on random processes including peculiar velocities and self-shielded systems that strongly impact unfortunate rays more than others. Throughout the EoR local and cosmological optical depths are often greater than or less than unity such that the $\exp(-\tau)$ behavior leads to anisotropic and bimodal transmissivity. Future surveys will benefit by targeting both rare bright objects and Goldilocks zone LAEs to infer the presence of these (un)predictable (dis)advantages.

Key words: galaxies: high-redshift – cosmology: dark ages, reionization, first stars – radiative transfer – methods: numerical

1 INTRODUCTION

The Epoch of Reionization (EoR) is the time period in the history of the Universe when the radiation from the first stars and galaxies initiated a cosmic phase transition throughout the intergalactic medium (IGM), which went from being cold and neutral to warm and ionized. In recent years we have witnessed significant progress in understanding the reionization process and advancing the current observational and computational frontiers. However, many of the central questions only have tentative answers that may be revised as the constraints from available data improve. For example, there is significant uncertainty about the role and contribution of low- and high-mass galaxies, the sources and escape of ionizing photons, and even the timing, duration, and morphology of reionization itself (Barkana & Loeb 2001; Loeb & Furlanetto 2013; Dayal & Ferrara 2018; Wise 2019).

There is mounting evidence for the so-called late reionization in which most of the IGM is rapidly ionized around redshift $z \sim 7$–8 (Naidu et al. 2020). This understanding comes from the *Planck* measurement of a low optical depth for electron scattering of CMB photons (Planck Collaboration et al. 2020), the declining fraction of Lyman-alpha emitters (LAEs) among the galaxy population at $z \gtrsim 6$ (Stark et al. 2011; Schenker et al. 2014; Mason et al. 2019), the imprint of neutral hydrogen in the IGM as a damping wing absorption feature on the spectrum of high-redshift quasars (Simcoe et al. 2012; Davies et al. 2018), and the spatial fluctuations of the Lyα forest transmission (Oñorbe et al. 2017; Kulkarni et al. 2019). We anticipate a more complete understanding of the EoR from upcoming facilities, including the *James Webb Space Telescope* (*JWST*) for the characterization of high-z galaxies and the Low-Frequency Array (LOFAR), Hydrogen Epoch of Reionization Array (HERA), and Square Kilometer Array (SKA) for 21 cm cosmology measurements to map out the distribution of neutral hydrogen in the Universe. On the theory side, radiation hydrodynamic (RHD) simulations have played an increasingly important role in capturing the non-linear, multiscale physics of reionization although there are still important computational challenges to overcome in the coming decades as well (Ciardi et al. 2000; Iliev et al. 2006; Gnedin 2014; Ocvirk et al. 2016; Pawlik et al. 2017; Rosdahl et al. 2018).

Ultimately, the joint analysis of observational probes sensitive to unique aspects of galaxy formation and IGM properties will provide definitive answers to the main questions about the EoR. It is in this spirit that we pursue an initial study of Lyman-alpha (Lyα) emission

★ E-mail: arsmith@mit.edu; NHFP Einstein Fellow.
† E-mail: rahul.kannan@cfa.harvard.edu
‡ E-mail: egaraldi@mpa-garching.mpg.de

© 2022 The Authors

from atomic hydrogen gas during the EoR from the THESAN suite of large-volume cosmological reionization simulations (Kannan et al. (2022), Garaldi et al. (2022), hereafter Papers I and II). The THESAN project utilizes the adaptive moving mesh magneto-hydrodynamics code AREPO (Springel 2010; Weinberger et al. 2020), in combination with the state-of-the-art IllustrisTNG galaxy formation model (Weinberger et al. 2017; Pillepich et al. 2018a), self-consistent radiation hydrodynamics (Kannan et al. 2019), and dust modelling (McKinnon et al. 2017). The flagship THESAN-1 run resolves atomic cooling haloes throughout a $(95.5\,\mathrm{cMpc})^3$ region of the Universe, providing sufficient particle resolution and halo statistics to bring unique insights about galaxy and IGM properties. Of crucial importance for LAEs specifically, THESAN connects the production of Lyα photons to their subsequent transmission through the IGM. One of the main drawbacks is the subresolution treatment of the interstellar medium (ISM) as a two-phase gas where cold clumps are embedded in a smooth, hot phase produced by supernova explosions (Springel & Hernquist 2003). However, such subgrid modelling comes with the territory of large-volume simulations with demonstrated agreement with observations down to $z = 0$ (Vogelsberger et al. 2020a). Thus, we proceed with our current exploration emphasizing that the THESAN project will be followed up by high-resolution zoom-in resimulations of a wide range of galaxies from the flagship run for self-consistent Lyα radiative transfer modelling from ISM to IGM scales.

The phenomenological impact of the IGM on radiation in the vicinity of the Lyα line is well known (Gunn & Peterson 1965; Miralda-Escudé 1998; Madau & Rees 2000), as are the implications when leveraging LAEs as a probe of reionization (Malhotra & Rhoads 2004; McQuinn et al. 2007; Dijkstra 2014; Kakiichi et al. 2016). In essence, neutral hydrogen far from the source can remove Lyα photons with a single scattering out of the line of sight. For fortunate LAEs residing within ionized bubbles on the order of ~ 0.1–1 physical Mpc (somewhat less stringent for peaks with large red velocity offsets), the light can redshift sufficiently far from resonance to avoid total suppression by the intervening IGM. However, numerical studies exhibit a complex landscape of sightline-to-sightline and galaxy-to-galaxy variations, which hints towards non-trivial dependence on the reionization history, environmental and proximity effects, and even galaxy properties such as the halo mass, specific star formation rate (SFR), and infall and peculiar velocities (Laursen et al. 2011; Dayal & Ferrara 2012; Jensen et al. 2014; Byrohl & Gronke 2020; Garel et al. 2021; Gronke et al. 2021; Park et al. 2021). The growing number of studies working in the context of large-volume cosmological RHD simulations is indicative of the importance of providing higher accuracy predictions for current and upcoming LAE surveys extending into the EoR. Such detailed studies will be invaluable to interpreting the observational signatures of both high-z galaxies (e.g. from the *JWST*) and integrated diffuse emission (e.g. from the *Spectro-Photometer for the history of the Universe, Epoch of Reionization and Ices Explorer: SPHEREx*).

In this study, we provide comprehensive galaxy Lyα emission and IGM transmission catalogues, which will be available with the public release of THESAN. The accessibility of such simulation-based surveys is timely given the current and forthcoming state of Lyα observation data. In fact, narrow-band surveys such as SILVERRUSH have already mapped over 2000 LAEs at $z = 5.7$ and 6.6, revealing target halo masses of $M_{\rm halo} \sim 10^{11}\,\mathrm{M}_\odot$ with per cent level duty cycles (Ouchi et al. 2018, 2020). Likewise, surveys of $3 \lesssim z \lesssim 6$ LAEs observed with the Multi-Unit Spectroscopic Explorer (MUSE) reveal that Lyα haloes are ubiquitous and possibly associated with high-z cosmic structure formation (Leclercq et al. 2017; Gallego et al. 2018). These windows correspond to the tail-end of reionization when most bubbles overlap but cosmic voids still harbour neutral islands capable of boosting the LAE clustering signal. Searches at $z \gtrsim 7$ are still impeded by low number statistics in constraining the reionization history at these epochs (Jung et al. 2020).

Our work is distinct from previous high-z studies in several aspects. In particular, the THESAN project employs a realistic galaxy formation model that for example produces the correct stellar-to-halo mass relation over a wide range of halo masses, includes novel secondary physics such as dust processes and black hole radiation and feedback, and medium resolution physics variations within the suite. The simulations do not require post-processing ionizing radiative transfer as is done for the modern pioneering Lyα IGM transmission analysis by Laursen et al. (2011). Furthermore, our volumes are large enough to avoid global cosmic variance that can bias transmission statistics. Recently, Garel et al. (2021) performed end-to-end Monte Carlo Lyα radiative transfer for thousands of galaxies within a $(10\,\mathrm{cMpc})^3$ SPHINX simulation, providing valuable insights about using LAEs to constrain the global reionization history. Still, the THESAN volumes are almost a thousand times larger, which is essential for capturing a representative number of bright LAEs for comparison with current and upcoming observations. Finally, other recent Lyα transmission studies with comparable volumes as ours are either focused on lower redshifts when the spatially uniform UV background approximation is valid (e.g. Behrens et al. 2018; Byrohl & Gronke 2020) or due to the poor spatial resolution within galaxies the astrophysical connections remain tenuous (e.g. Gronke et al. 2021; Park et al. 2021). Self-consistent simulation-based LAE science represents a monumental challenge with encouraging progress from several fronts, including improving treatments of ISM-to-IGM scale radiative transfer modelling focused on the EoR (e.g. Behrens et al. 2019; Laursen et al. 2019; Smith et al. 2019; Garel et al. 2021). Of course, our intuition and understanding of Lyα radiative transfer has been aided by analytical studies (e.g. Harrington 1973; Neufeld 1990; Hansen & Oh 2006; Lao & Smith 2020), idealized setups (e.g. Dijkstra et al. 2006; Behrens et al. 2014; Gronke et al. 2017; Song et al. 2020; Li et al. 2021), and isolated galaxy simulations (e.g. Verhamme et al. 2012; Behrens & Braun 2014; Smith et al. 2021), which allow better resolution and control of the small-scale ISM physics.

The paper is organized as follows. In Section 2, we briefly describe the flagship THESAN simulation employed throughout this paper. In Section 3, we introduce the Lyα emission catalogues that form the basis of statistical explorations of intrinsic luminosity based properties. In Section 4, we outline the procedure for calculating frequency-dependent transmission curves, including a novel integration scheme to account for continuous Hubble flow within the IGM. We also present our main results exploring the non-trivial dependence on frequency, redshift, and UV magnitude. Finally, in Section 5, we provide a summary and brief perspective utilizing THESAN for Lyα science.

2 THESAN SIMULATIONS

In this section we briefly summarize the main features of the THESAN simulations, which are introduced in detail in Paper I Kannan et al. (2022). THESAN is a suite of radiation-magnetohydrodynamical simulations run with the moving-mesh hydrodynamics code AREPO (Springel 2010; Weinberger et al. 2020)[1]. The code employs a finite-volume method to solve the Euler equations on an unstructured Voronoi tessellation for an accurate treatment of quasi-Lagrangian

[1] Public code access and documentation available at `www.arepo-code.org`.

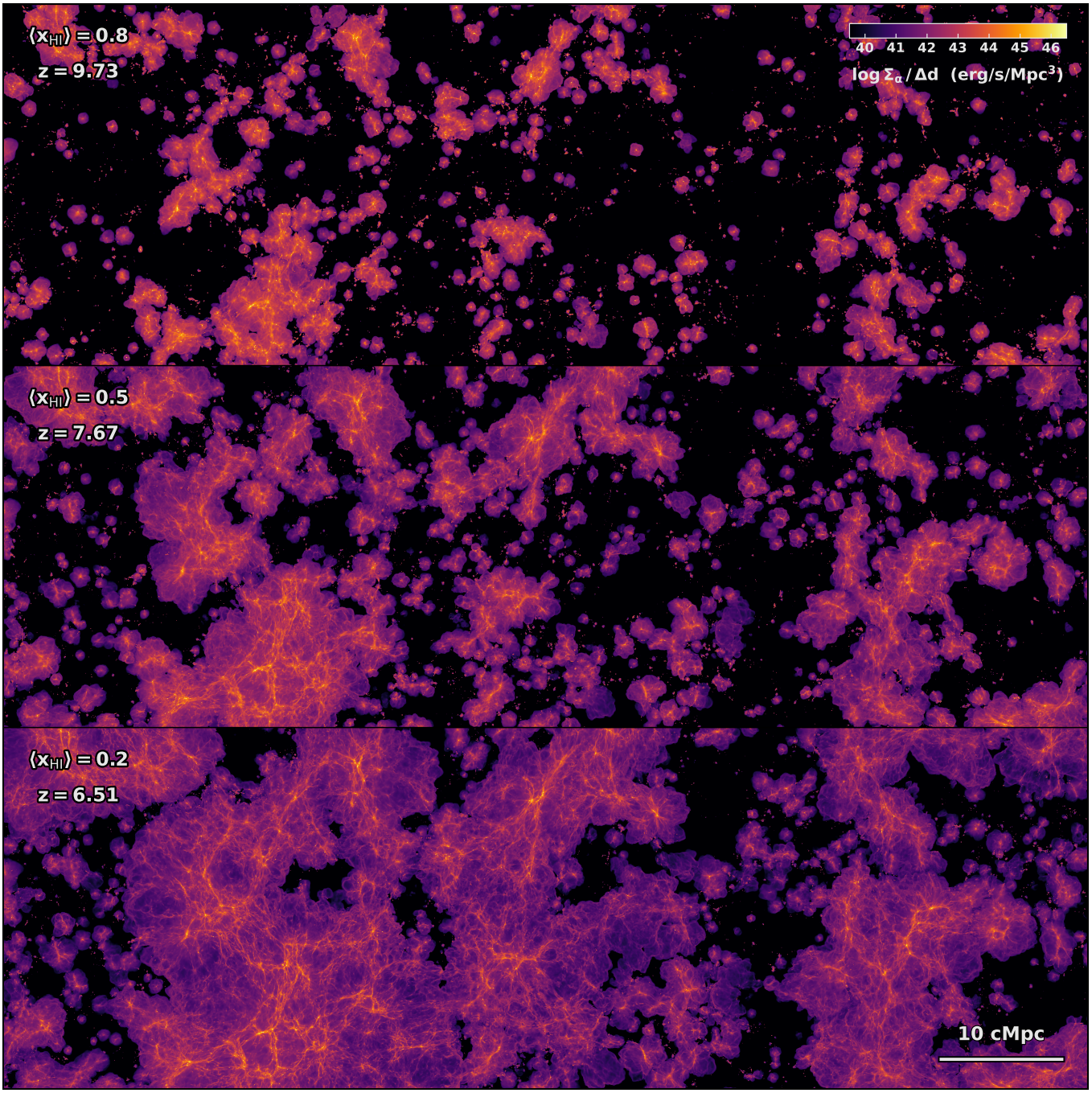

Figure 1. Intrinsic Lyα surface brightness covering the same $90 \times 30 \times 3\,\mathrm{cMpc}^3$ subvolume for snapshots corresponding global neutral hydrogen fractions of $\langle x_{\mathrm{H\,I}} \rangle \approx \{0.8, 0.5, 0.2\}$ or redshifts of $z \approx \{9.73, 7.67, 6.51\}$ from top to bottom, respectively. The images are made with an adaptive quadrature ray-tracing scheme though the Voronoi tessellation to guarantee conservation of the luminosity as given by equations (1)–(3). Although radiative transfer effects on ISM to IGM scales are not included, the Lyα emissivity is clearly connected to the large-scale cosmic structure and topology of reionization.

flows with complex source terms over large dynamic ranges. Gravity calculations utilize a hybrid Tree-PM approach, which splits the force into short- (direct summation) and long-range (particle mesh) contributions computed through an adaptive oct-tree data structure (Barnes & Hut 1986). In addition, the AREPO implementation includes hierarchical time integration of the resolution elements and randomization of the node centres at each domain decomposition as described in the GADGET4 paper (Springel et al. 2021).

For self-consistent radiative transfer, we employ the AREPO-RT extension described by Kannan et al. (2019), which solves the first two moments of the RT equation assuming the M1 closure relation (Levermore 1984). This scheme reaches second-order accuracy by replacing the piecewise constant approximation of the Godunov (1959) scheme with a slope-limited piecewise linear spatial extrapolation utilizing a local least-squares fit for gradient estimates (Pakmor et al. 2016). A first-order time extrapolation based on half time-steps

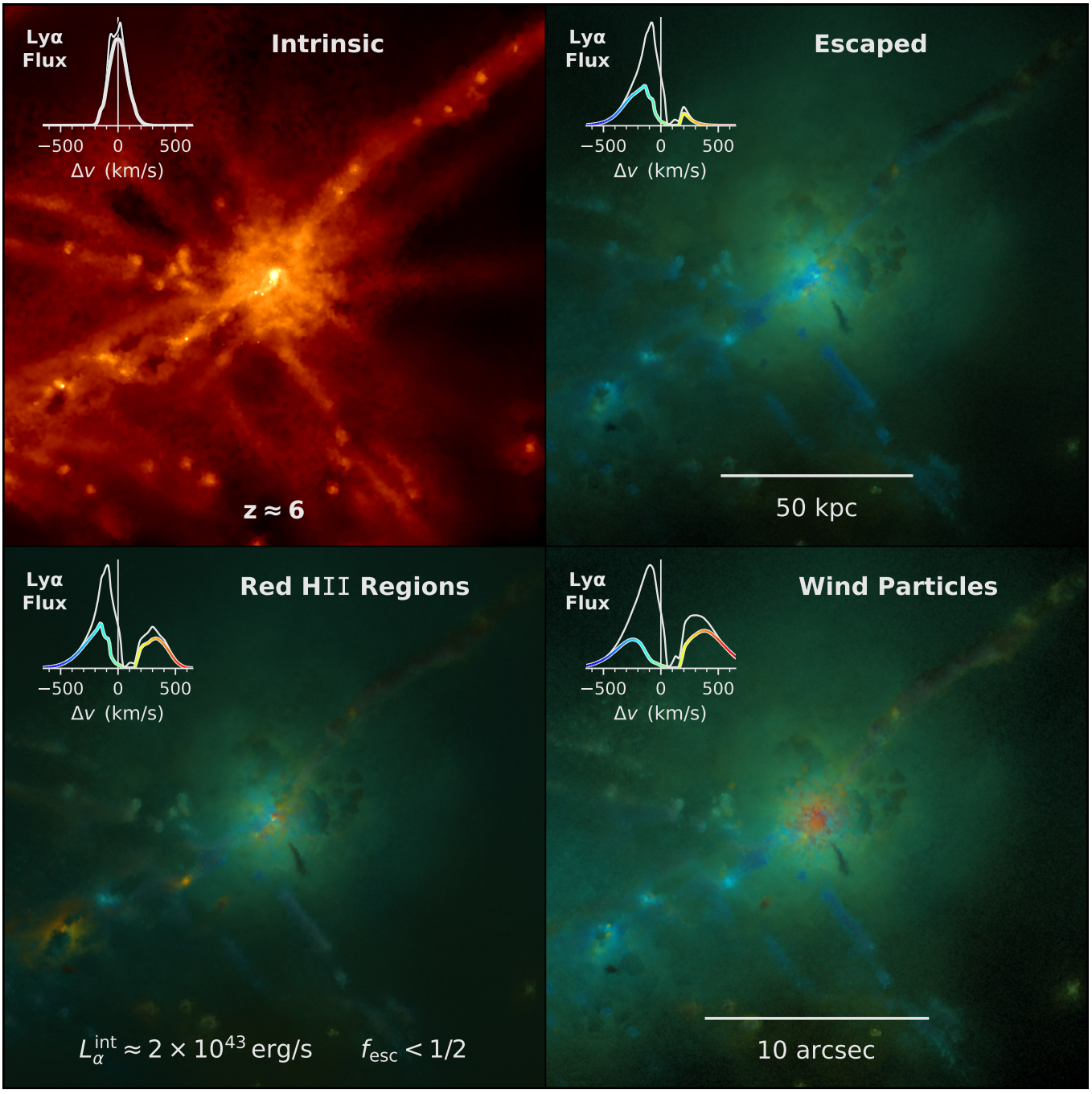

Figure 2. *Upper left-hand panel:* Intrinsic Lyα surface brightness for a galaxy of mass $M_{\rm halo} \approx 10^{11}\,{\rm M}_\odot$ at $z = 6$. *Upper right-hand panel:* False colour rendering of the escaped Lyα emission based on synthetic integral field unit (IFU) data generated with the COLT Monte Carlo radiative transfer code. The spectroscopic information is blended from blue–green to yellow–red with the image opacity encoding the surface brightness. The rest-frame is defined by the frequency centroid of the intrinsic line profile shown in the left-hand panel, while the emergent spectra in the remaining panels serve as velocity offset colour maps. The bold coloured line profiles are taken from a smaller aperture (1″ radius) while the white spectra include (mostly blue peak) photons from a wider area (≈ 20″ radius). *Lower left-hand panel:* An alternative emission model that artificially reddens the initial frequency of unresolved H II regions to explore the impact of local feedback-induced outflows on ISM scales. *Lower right-hand panel:* An alternative transport model that incorporates wind particles to explore the role of galactic winds in shaping line profiles on CGM scales.

is employed to obtain the primitive variables on both sides of the interface (van Leer 1979). For computational efficiency, we choose to only model the ionizing part of the radiation spectrum and discretize photons into energy bins defined by the following thresholds: [13.6, 24.6, 54.4, ∞) eV. Each resolution element tracks the comoving photon number density and flux for each bin. To partially compensate the loss of resolution in the frequency sampling, we assume the radiation within each bin follows the spectrum of a 2 Myr old, quarter-solar metallicity stellar population with amplitude given by the local photon number density. This choice is physically motivated

as young stars contribute most of the ionizing budget. Still, the effective photon properties are relatively insensitive to age and metallicity. The stellar spectra employ the Binary Population and Spectral Synthesis models (BPASS version 2.2.1; Eldridge et al. 2017), assuming a Chabrier IMF (Chabrier 2003). The bin average values of the H I, He I, and He II photoionization cross-sections (σ), energy injected into the gas per interacting photon ($\mathcal{E}$), and mean energy per photon (e) are reported in Table 1 of Paper I. The RT equations are coupled to a non-equilibrium solver that accurately computes the ionization state of hydrogen and helium, as well as the temperature change due to photoheating, atomic, metal and Compton cooling. Finally, we employ a reduced speed of light (RSLA) approximation with an effective value of $\tilde{c} = 0.2\,c$ (e.g. Gnedin 2016). We emphasize that although Ocvirk et al. (2019) argue for higher values, in Appendix A of Paper I we demonstrate that in the context of our model this value is large enough to accurately capture the propagation of ionization fronts and post-reionization gas properties.

The THESAN simulations were designed to simultaneously capture the assembly of primeval galaxies and their impact on reionization as realistically as possible. For this reason, we employ the state-of-the-art IllustrisTNG galaxy formation model, which updates the previous Illustris model (Genel et al. 2014; Vogelsberger et al. 2014a,b) to include subresolution physics tuned to reproduce a wide range of galaxy properties that are consistent with available observations down to low redshift (Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018; Pillepich et al. 2018b; Springel et al. 2018). This choice ensures that, although the THESAN simulations are only evolved to $z \simeq 5.5$, the physical model can be trusted throughout the history of the Universe (e.g. for especially relevant high-z galaxy predictions see Shen et al. 2020, 2022; Vogelsberger et al. 2020b). Additionally, the only new free parameter is the stellar escape fraction $f_{\rm esc}^{\rm ion}$[2], which we calibrate to match constraints for the global reionization history[3] (0.37 for the flagship simulation). Furthermore, we also include a state-of-the-art dust model developed by McKinnon et al. (2017).

The full THESAN simulation set and parameters are catalogued in Table 2 of Paper I. All runs follow the evolution of a cubic patch of the universe with linear comoving size $L_{\rm box} = 95.5\,{\rm cMpc}$. The initial conditions employ a method in which the initial Fourier mode amplitudes are fixed to the ensemble average power spectrum to suppress variance (Angulo & Pontzen 2016). Throughout this paper, we employ a Planck Collaboration et al. (2016) cosmology with simulation constants of $h = 0.6774$, $\Omega_0 = 0.3089$, and $\Omega_{\rm b} = 0.0486$, where all the symbols have the usual meaning.

The flagship THESAN-1 simulation is designed to resolve atomic cooling haloes with virial temperatures of $T_{\rm vir} \gtrsim 10^4\,{\rm K}$ and masses of $M_{\rm halo} \gtrsim 10^8\,h^{-1}{\rm M}_\odot$ (see e.g. Bromm & Yoshida 2011). Thus, the total number of dark matter and (initial) gas particles is $N_{\rm particles} = 2100^3$ each with mass resolutions of $m_{\rm DM} = 3.1 \times 10^6\,{\rm M}_\odot$ and $m_{\rm gas} = 5.8{\times}10^5\,{\rm M}_\odot$, respectively. The gravitational softening length for the star and dark matter particles is set to 2.2 ckpc, while the gas cells employ adaptive softening according to the cell radius. Gas cells are (de-)refined to ensure masses remain within a factor of two from the target mass, thus the minimum cell radius at $z = 5.5$ is ~ 10 pc for a dynamic resolution range of six orders of magnitude.

To quantify the resolution in the IGM we define the effective radius of each cell to be $r_{\rm cell} \equiv (3V_{\rm cell}/4\pi)^{1/3}$, where $V_{\rm cell}$ denotes the cell volume. We calculate the average cell radius as $\langle r_{\rm cell}\rangle_V \equiv \sum r_{\rm cell} V_{\rm cell} / \sum V_{\rm cell} \approx 5.28\,(2.82)\,{\rm kpc}$ at $z = 5.5\,(10)$, noting that due to the Lagrangian nature of the code the spatial resolution is significantly better for higher density gas near galaxies and IGM structures where most absorption occurs. In particular, Rahmati & Schaye (2018) showed that the main H I sinks of ionizing radiation, namely Lyman-limit systems with expected sizes of 1–10 kpc, are well resolved in their reference runs that have 3.5 times coarser baryonic mass resolution than the THESAN-1 simulation used in this study (see also Paper II). Thus, THESAN provides a state-of-the-art framework for connecting resolved galaxy and IGM properties throughout the EoR, ideal for this study of Lyα transmission and other topics relevant to the high-redshift Universe. However, there is also the question of achieving adequate spatial resolution throughout the extended CGM of galaxies. Although it is beyond the current state-of-the-art to uniformly require $\lesssim 1\,{\rm kpc}$ resolution for such large-volume simulations, it may be worth making steps in this direction through various optimization trade offs to study the impact of CGM resolution on reionization, especially as this has been found to increase the covering fraction of Lyman-limit systems around isolated Milky Way-mass galaxies (van de Voort et al. 2019). In this paper, we focus exclusively on the main THESAN-1 simulation, deferring comparisons with the other simulations to a future study.

Table 1. Brief description of fields in the Lyα group and subhalo catalogues.

Field	Units	Description
L_α	erg s^{-1}	Total Lyα luminosity ($L_\alpha = L_\alpha^{\rm rec} + L_\alpha^{\rm col} + L_\alpha^{\rm stars}$)
$L_\alpha^{\rm rec}$	erg s^{-1}	Lyα luminosity from resolved recombination
$L_\alpha^{\rm col}$	erg s^{-1}	Lyα luminosity from collisional excitation
$L_\alpha^{\rm stars}$	erg s^{-1}	Lyα luminosity from unresolved H II regions
$L_{\lambda,1216}$	erg s^{-1} Å^{-1}	Stellar continuum spectral luminosity at 1216 Å
$L_{\lambda,1500}$	erg s^{-1} Å^{-1}	Stellar continuum spectral luminosity at 1500 Å
$L_{\lambda,2500}$	erg s^{-1} Å^{-1}	Stellar continuum spectral luminosity at 2500 Å
$L_{\rm ion}^{\rm AGN}$	erg s^{-1}	Ionizing luminosity from active galactic nuclei
$\boldsymbol{r}_\alpha$	kpc	Centre of Lyα luminosity position in the box
$\boldsymbol{v}_\alpha$	km s^{-1}	Centre of Lyα luminosity peculiar velocity
σ_α	km s^{-1}	Centre of Lyα luminosity 1D velocity dispersion

3 GALAXY EMISSION CATALOGUES

For all snapshots we produce Lyα catalogues directly mirroring the friends-of-friends (FoF) halo catalogues and SUBFIND subhalo catalogues. These post-processing files provide supplemental data for each group and subhalo, corresponding to identifications of dark matter haloes and galaxies, respectively. There is a single `Lya_*` HDF5 file for each snapshot containing the following groups: Header, Group, Subhalo, Inner, Outer, and Total. For convenience, the header contains information about the simulation and the others contain global sums of various Lyα luminosities over all groups, subhaloes, inner/outer "fuzz" of unbound particles, and the entire simulation box. More importantly, we provide local sums for each individual group or subhalo, summarized in Table 1 and explained below.

[2] Employing a constant value for the ionizing escape fraction at the level of the birth cloud is intended to approximately capture subresolution internal neutral hydrogen and dust self-absorption. In reality, the local escape fraction transitions from zero early on to order unity after feedback disperses the high-density star-forming gas (e.g. Kimm et al. 2019). The variation in time- and rate-averaged values is still an open question, and likely depends on complex environmental properties requiring improved resolution and ISM modelling.

[3] For reference, the global reionization history is included in Fig. 13.

3.1 Lyα production

The total Lyα luminosity L_α gives the intrinsic emission from recombinations, collisional excitation, and local stars. We calculate the resolved luminosity due to radiative recombination as

$$L_\alpha^{\rm rec} = h\nu_\alpha \int P_{\rm B}(T)\alpha_{\rm B}(T)\, n_e n_p \, {\rm d}V \,, \quad (1)$$

where $h\nu_\alpha = 10.2\,$eV, the Lyα conversion probability per recombination event is $P_{\rm B} \approx 0.68$ (Cantalupo et al. 2008), the case B recombination coefficient is $\alpha_{\rm B}$ (Hui & Gnedin 1997), and the number densities n_e and n_p are for free electrons and protons, respectively (Dijkstra 2019). In addition, we calculate the resolved contribution of radiative de-excitation of collisional excitation of neutral hydrogen by free electrons as

$$L_\alpha^{\rm col} = h\nu_\alpha \int q_{1s2p}(T)\, n_e n_{\rm H\,I} \, {\rm d}V \,, \quad (2)$$

where the temperature-dependent rate coefficient q_{1s2p} is taken from Scholz & Walters (1991). Due to the uncertainties surrounding the effective equation of state (EoS) for cold gas above the density threshold $n_{\rm H} \approx 0.13\,{\rm cm}^{-3}$, we isolate recombinations and collisional excitation emission from non-EoS cells as identified by the SFR being identically zero, noting that these components each contribute at the $\sim 10\%$ level. The non-equilibrium radiation hydrodynamics solver provides accurate thermal and ionization states for such gas. However, we note that the ordering of the black hole energy injection routine leads to hot gas above the EoS that is artificially neutral for half a time-step. Therefore, for each cell we compare the neutral hydrogen fraction state from the simulation output to the maximum expected value assuming collisional ionization equilibrium (CIE), updating the ionization states if $x_{\rm H\,I} > x_{\rm H\,I,CIE}$. We note that we employ iteration so the rate coefficients reflect the correct temperatures. We find this pre-conditioning step robustly eliminates unphysical collisional excitation emission that will easily be avoided in future THESAN simulations by adjusting the ordering of the cooling physics. Of course, there will be additional Lyα cooling emission in the unresolved EoS cells, therefore our collisional excitation luminosities are conservative values. Furthermore, due to the RHD coupling of the thermochemistry a fraction of the ionizing budget is lost to the EoS cells. To account for this we also track the EoS recombination emission assuming these cells become ionized by local sources during radiation subcycles. We show later that this approach provides good agreement with the expectation for the global production of Lyα photons assuming an escape fraction of zero.

We account for unresolved H II regions in the vicinity of stellar populations by converting self-absorbed ionizing photons to sources of Lyα emission at the subgrid level as

$$L_\alpha^{\rm stars} = 0.68 h\nu_\alpha (1 - f_{\rm esc}^{\rm ion}) \dot{N}_{\rm ion} \,. \quad (3)$$

Here, the factor 0.68 is the fiducial conversion probability assuming a temperature for emitting gas of $10^4\,$K, $f_{\rm esc}^{\rm ion}$ denotes the escape fraction of ionizing photons, and $\dot{N}_{\rm ion}$ is the emission rate of ionizing photons from stars. The latter is a complex function of age and metallicity taken from the BPASS models (v2.2.1), which include binary stellar poplulations accounting for mass transfer, common envelope phases, binary mergers, and quasi-homogeneous evolution at low metallicities (Eldridge et al. 2017). Thus, our Lyα emission catalogues correspond to the same model as the intrinsic sources of reionization within the simulations. Overall, the relative contribution of resolved and unresolved sources provides some intuition about the uncertainties in our Lyα emission modelling.

To better understand the intrinsic production of Lyα photons in the THESAN simulations in Fig. 1 we show the Lyα surface brightness for a $90 \times 30 \times 3\,{\rm cMpc}^3$ sub-volume for snapshots corresponding global neutral hydrogen fractions of $\langle x_{\rm H\,I} \rangle \approx \{0.8, 0.5, 0.2\}$ or redshifts of $z \approx \{9.73, 7.67, 6.51\}$, respectively. We employ an adaptive quadrature ray-tracing scheme though the Voronoi tessellation to ensure conservation of the luminosity as given by equations (1–3). Specifically, convergence is achieved via the iterative refinement algorithm described in Appendix A of Yang et al. (2020). Although radiative transfer effects on ISM to IGM scales are not included, the Lyα emissivity is clearly connected to the large-scale structure (LSS) and topology of reionization. This visualization also emphasizes the strong impact of cosmic variance on Lyα studies based on small ($\lesssim 30\,$cMpc) reionization simulations as they may lead to biased statistics for certain quantities such as clustering and IGM transmission (for a discussion of implications for observing the post-reionization cosmic web in Lyα emission see Witstok et al. 2021). The THESAN resolution and box size are sufficient to provide representative galaxy populations for Lyα emission and correctly capture EoR fluctuations (Gnedin & Kaurov 2014; Iliev et al. 2014).

In Fig. 2 we demonstrate that THESAN is fully capable of capturing IGM scale effects, but detailed LAE modelling is strongly affected by ISM scale sourcing and radiation transport through the circumgalactic medium (CGM). In the left-hand panel, we illustrate the intrinsic Lyα emissivity for a galaxy of mass $M_{\rm halo} \approx 10^{11}\,{\rm M}_\odot$ at $z = 6$. In the remaining panels we exhibit false colour renderings of the escaped Lyα emission based on synthetic integral field unit (IFU) data generated with the Cosmic Lyα Transfer code (COLT; Smith et al. 2015, 2019, 2021, which describe the physics and code implementation in detail)[4]. The Monte Carlo radiative transfer calculations employ 10^8 photon packets sourced according to equations (1)–(3) with photons drawn within cell volumes for resolved recombinations (32%) and collisional excitation (8%) or from star particle locations for unresolved H II regions (60%). Ray integrations are performed using the native Voronoi cells with the neutral hydrogen and dust densities, gas temperatures, and bulk velocities taken directly from the simulation. We assume a dust opacity of $\kappa_{\rm d} = 5.8 \times 10^4\,{\rm cm}^2\,{\rm g}^{-1}$ of dust, scattering albedo of $A = 0.325$, and asymmetry parameter of $\langle \cos\theta \rangle = 0.676$, based on the Milky Way dust model from Weingartner & Draine (2001). All images and spectra are for the same line-of-sight, although we have checked that all positive and negative coordinate direction observations result in similar qualitative conclusions. The spectroscopic data are blended in colour space with the image opacity encoding the surface brightness. The rest-frame is defined with respect to the frequency centroid of the intrinsic Hα line profile inset in the first panel. The emergent flux is dominated by a blue peak shaped by the cosmological environment, which reflects the insufficient resolution and inconsistencies with the density, temperature, velocity, and ionization state regulated by the effective EoS model, but also the wide integration aperture ($\approx 20''$ radius). To isolate this last effect, we show that in comparison line profiles taken from a smaller aperture ($1''$ radius) have reduced blue peaks.

In the lower panels, we explore two alternative scenarios designed to produce spectra with enhanced red peaks in better agreement with observations. First, we artificially redden the initial frequency of unresolved H II regions (i.e. photons from star particles are injected as a red peak) to model local feedback-induced outflows on subgrid ISM scales. We note that the resolved recombination and collisional excitation contributions ($\sim$ half) remain unchanged (including their frequency initialization). Lastly, we incorporate hydrodynamically

[4] For public code access and documentation see `colt.readthedocs.io`.

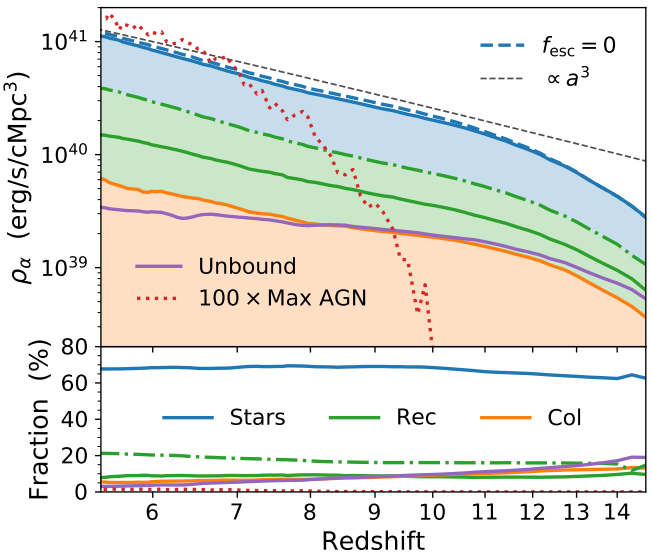

Figure 3. *Top:* Redshift evolution of the global Lyα intrinsic luminosity density including cumulative contributions from resolved collisional excitation (orange) and recombination (green; the dash-dotted curve includes EoS cells) emission and unresolved H II regions (blue). The total emission is well described by a power law of $\rho_\alpha = 8.3 \times 10^{40}\,[(z+1)/7]^{-3}\,\mathrm{erg\,s^{-1}\,cMpc^{-3}}$ over the range $z \in (5.5, 11)$, offset for visualization purposes. For comparison we show that AGN (red dotted) contribute at the per cent level if ionizing photons are efficiently converted to Lyα photons according to $\rho_\alpha^{\rm AGN} \approx 0.68 h\nu_\alpha \dot{N}_{\rm ion}^{\rm AGN}$. *Bottom:* The fraction of luminosity from each channel illustrating that local stars dominate the emission budget. For clarity, we include separate contributions from recombinations, collisional excitation, and sources not bound to any subhalo (purple).

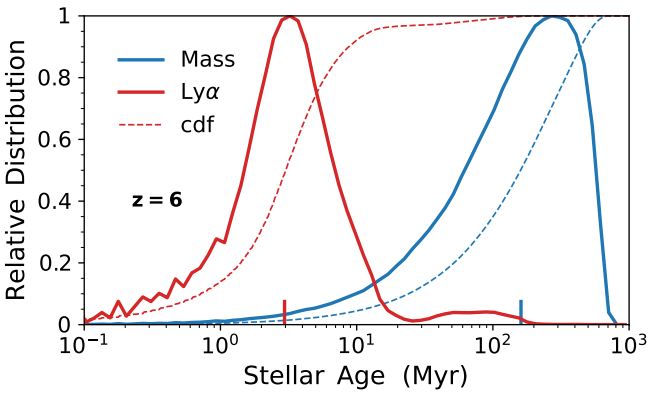

Figure 4. Relative distribution of ages contributing to the intrinsic ionizing radiation that is locally reprocessed into Lyα photons, with cumulative distribution functions shown as dashed curves and vertical markers denoting medians. The emission traces the youngest stellar populations with a median age of roughly 3 Myr, while the mass-weighted median age is 160 Myr at $z = 6$. Thus, the star formation history acts as the primary driver of ionizing sources rather than the stellar mass.

decoupled wind particles (i.e. as additional gas cells) to emulate a clumpy outflow on resolved scales tracing galactic winds out to the CGM. Multiphase winds help regulate star formation and transport metals out of galaxies, but we reduce the speeds by a factor of four to better track the cold neutral components. Encouragingly, in each case the red peak is enhanced reflecting the physically motivated improvements, but further adjustments and calibrations may be required for comparison to LAE surveys. These may include dust rescaling, subgrid clumping, bulk velocities informed by winds, or modifications to EoS cell properties, each affecting the robustness of predicted spectra (see also Li et al. 2020; Byrohl et al. 2021; Gronke et al. 2021). We are pursuing more accurate Lyα radiative transfer studies from zoom-in simulations of THESAN galaxies, which will set anchor points on these uncertainties. In fact, our preliminary results reveal that red peak dominated spectra naturally arise in the context of realistic multiphase ISM and CGM environments shaped by feedback in these simulations. For now, the Lyα line profile results in Fig. 2 provide important insights and warnings to the community regarding galaxy scale Lyα radiative transfer calculations from large-volume reionization simulations. Specifically, it remains difficult to produce red peak dominated spectra and ultimately the solution must rely on improved galaxy formation models rather than empirical corrections. As IGM transmission can be highly sensitive to the emergent spectra, it is non-trivial to robustly connect back to the intrinsic Lyα emission. Therefore, the remainder of this paper focuses on emission and transmission, allowing dedicated follow-up explorations to more fully address radiative transfer uncertainties. We return to this example after presenting our main IGM transmission findings in Section 4.7.

In Fig. 3, we show the redshift evolution of the global Lyα intrinsic luminosity density including contributions from unresolved H II regions and resolved collisional excitation and recombination emission. The total emission is well described by a power law of $\rho_\alpha = 8.3 \times 10^{40}\,[(z+1)/7]^{-3}\,\mathrm{erg\,s^{-1}\,cMpc^{-3}}$ over the range $z \in (5.5, 11)$. Local stars (blue) dominate the emission budget followed by recombination (green) and cooling radiation (orange), thus radiative transfer effects are expected to result in drastic reprocessing of the observed emission beyond what we are able to capture here. Interestingly, the fraction from sources not bound to any subhalo (purple), i.e. the inner and outer "fuzz", declines with time to a few per cent by $z \lesssim 6$. Likewise, AGN only contribute at the per cent level even if ionizing photons are efficiently converted to Lyα photons according to $\rho_\alpha^{\rm AGN} \approx 0.68 h\nu_\alpha \dot{N}_{\rm ion}^{\rm AGN}$ (red dotted; see Section 3.2), which may be boosted by a factor of a few due to harder ionizing spectra inducing multiple ionizations (Raiter et al. 2010). In fact, the total emission budget agrees with the expectation of a resolution agnostic model ($f_{\rm esc} = 0$) assuming every ionizing photon eventually produces Lyα emission (blue dashed).

To explore the sources of Lyα emission further, in Fig. 4 we provide the age distribution of stars emitting ionizing radiation that is locally reprocessed into Lyα photons at $z = 6$. The emission traces the youngest stellar populations with a median age of roughly 3 Myr, while the mass-weighted median age is 160 Myr at $z = 6$, emphasizing the important role of the star formation history beyond the stellar mass alone. Similarly, in Fig. 5 we illustrate the relative Lyα intrinsic luminosity originating from resolved recombination and collisional excitation emission as functions of hydrogen number density and temperature ($n_{\rm H}$–T) at $z = 6$. Lyα photons are mainly produced by gas around the effective equation of state threshold of $\sim 0.13\,\mathrm{cm^{-3}}$. The majority of the emission ($> 90\%$) is from photoheated gas slightly above 10^4 K. Overall, the properties of Lyα production are in line with our expectations considering the limitations of the galaxy formation model on ISM scales.

3.2 Additional Lyα-centric fields

We also provide the stellar continuum spectral luminosities $L_{\lambda,\rm cont}$ at $\lambda_{\rm cont} = \{1216, 1500, 2500\}$ Å, derived by taking the logarith-

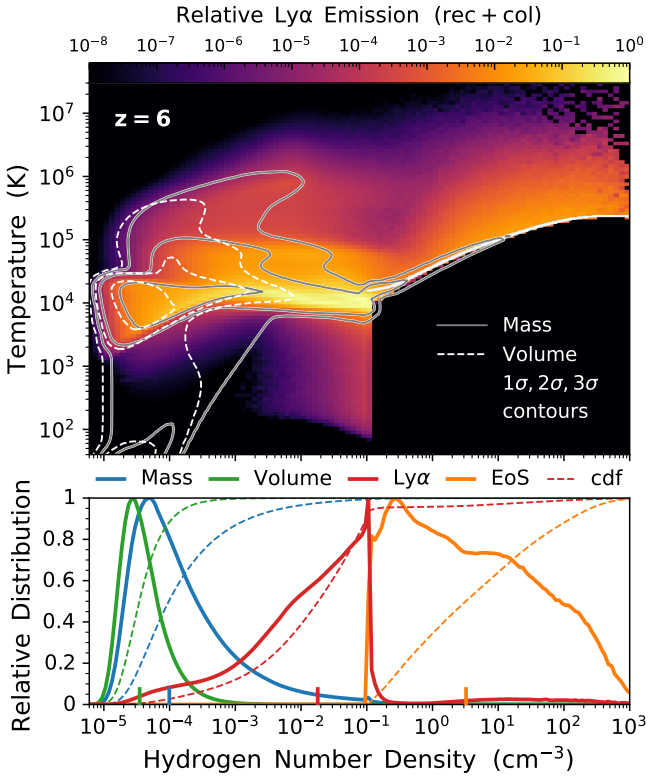

Figure 5. *Top:* Relative Lyα intrinsic luminosity originating from resolved recombination and collisional excitation emission in the gas hydrogen number density and temperature ($n_{\rm H}$–T) phase plane at $z = 6$. *Bottom:* Distributions showing that Lyα photons are mainly produced by gas around the effective equation of state threshold of $\sim 0.13\,{\rm cm}^{-3}$, dominated by ionized gas heated to $\sim 10^4$ K. The various curves are for mass (blue), volume (green), and Lyα emission separated into resolved (red) and EoS (orange) components. The dashed curves are cumulative distribution functions, with vertical markers denoting the median volume, mass, and Lyα luminosity-weighted number densities at 3.5×10^{-5}, 9.9×10^{-5}, 0.018, and $3.2\,{\rm cm}^{-3}$, respectively.

mic average of the BPASS SEDs over windows of $\{50, 20, 20\}$ Å around the reference wavelengths. The larger window around $\lambda_\alpha = 1215.67$ Å reduces the sensitivity to Lyα absorption features in the BPASS spectra. For convenience, the absolute magnitude as if viewing the stars from a distance of 10 pc at these wavelengths is $M_{\rm cont} = -2.5\log[(\lambda_{\rm cont}/\text{Å})^2 L_{\lambda,\rm cont}/(\text{erg/s/Å})] + 97.78683$, which is used throughout this study. Likewise, the rest-frame equivalent width characterizing the strength of the Lyα line relative to the continuum flux is given by ${\rm EW}_{\alpha,0} \approx L_\alpha / L_{\lambda,1216}$. When the local continuum is not detected in observations, the equivalent width may be derived by extrapolating continuum values based on the UV slope assuming a power-law $L_\lambda \propto \lambda^\beta$ (Hashimoto et al. 2017), such that $\beta = \log(L_{\lambda,2500}/L_{\lambda,1500})/\log(2500/1500)$ and the estimated continuum around the Lyα line is $L^{\rm est}_{\lambda,1216} = L_{\lambda,1500}\,(1215.67/1500)^\beta$. The relative difference of these two approaches is explored in Appendix A where we find that extrapolations systematically underpredict intrinsic Lyα equivalent widths by approximately 30 per cent.

Similarly, the ionizing luminosity from active galactic nuclei (AGN) $L^{\rm AGN}_{\rm ion}$ gives the escaping emission from supermassive black holes. We note that the AGN spectral energy distribution (SED) uses the Lusso et al. (2015) parametrization with 35.5 per cent of the bolometric AGN luminosity at energies above 13.6 eV. After converting this quantity to c.g.s. units the equivalent number of ionizing photons is determined by dividing by the average energy per photon, i.e. 5.29×10^{-11} erg ≈ 33 eV. In this paper we do not explore the properties of AGN bright galaxies beyond confirming that the global contribution is subdominant (see Fig. 3). However, defining the fraction of ionizing photons originating from AGN as $f_{\rm AGN} \equiv \dot{N}^{\rm AGN}_{\rm ion}/(\dot{N}^{\rm AGN}_{\rm ion} + \dot{N}^{\rm stars}_{\rm ion})$, we find at $z = \{5.5, 6, 7, 8, 9\}$ there are $\{23, 6, 4, 0, 1\}$ galaxies with $f_{\rm AGN} \geq 0.5$ and $\{170, 104, 41, 15, 7\}$ with $f_{\rm AGN} \geq 0.1$. Thus, by the end of the simulation rare but very luminous quasars can be dominant for a handful of galaxies, which becomes even more relevant for larger boxes and is worth pursuing in detail in future investigations.

Finally, we also calculate the centre of Lyα luminosity position $\boldsymbol{r}_\alpha$ within the periodic box, peculiar velocity $\boldsymbol{v}_\alpha$, and 1D velocity dispersion σ_α for each group and subhalo. For notational convenience we define the Lyα luminosity-weighted average of a quantity as $\langle f \rangle_\alpha = \sum_i f_i L_{\alpha,i} / \sum_i L_{\alpha,i}$, where the sum is over all Lyα sources including gas and stars. Thus, the centre of Lyα luminosity quantities are respectively $\boldsymbol{r}_\alpha \equiv \langle \boldsymbol{r} \rangle_\alpha$, $\boldsymbol{v}_\alpha \equiv \langle \boldsymbol{v} \rangle_\alpha$, and $\sigma^2_\alpha \equiv (\langle \boldsymbol{v} \cdot \boldsymbol{v} \rangle_\alpha - \langle \boldsymbol{v} \rangle_\alpha \cdot \langle \boldsymbol{v} \rangle_\alpha)/3$, where the individual $\boldsymbol{r}$ and $\boldsymbol{v}$ are centre of mass positions and peculiar velocities. We note that these quantities are weighted by the intrinsic Lyα emission so are not directly observable due to radiative transfer effects. However, they are still useful for connecting Lyα emission to other galaxy properties, or to account for spatial and velocity offsets and line broadening as the $n^2_{\rm H}$ dependence for emission can lead to biases here (discussed in Section 3.4).

3.3 Luminosity functions

We now consider the evolution of the occurrence and luminosities of galaxies throughout the EoR. In Fig. 6, we show galaxy halo and stellar mass functions over the redshift range $z \in [6, 10]$ covering four orders of magnitude in resolved structure formation ($M_{\rm halo} \gtrsim 10^8\,{\rm M}_\odot$) down to the clustering resolution of the simulation ($M_{\rm stars} \gtrsim 10^6\,{\rm M}_\odot$). For reference, we include a horizontal dashed line representing a volume limit of 10 objects within the simulation box. Similarly, in Fig. 7 we provide galaxy far UV (rest-frame 1500 Å) and Lyα intrinsic luminosity functions for the same redshift range. The evolution is smooth and relatively steep due to the continual formation of young stars as galaxies assemble, merge, and are fed by streams of cold gas accretion. The distributions for L_α closely follow that of UV magnitude M_{1500} with the caveat that galaxies can be brighter in Lyα due to the recombination and collisional excitation emission. To provide a sense of population convergence, e.g. above $M_{1500} \approx -15.5$ or $L_\alpha \approx 3 \times 10^{41}\,{\rm erg\,s}^{-1}$, we also show the normalized cumulative luminosity from haloes above a given brightness threshold. The details of these luminosity functions may depend on modelling choices and resolution but the shape is robust as it follows the star formation density evolution, which IllustrisTNG is designed to match for currently available observational data at lower redshifts.

We note that the Springel & Hernquist (2003) model employs stochastic sampling to determine when new star particles are created in the simulation, with comparable mass resolutions for star and gas particles. This prescription is correct in the global sense and favourable for avoiding numerical artefacts, e.g. with gravitational force calculations being more robust against artificial mass segregation due to different resolutions for stars and gas (Ludlow et al. 2019). However, the age discretization leads to characteristically bursty star formation histories (SFHs) in marginally resolved haloes (Iyer et al. 2020). To extend the reliability of halo-by-halo predictions for galaxies at the faint-end of the luminosity function we employ the following SFH smoothing procedure. We first calculate

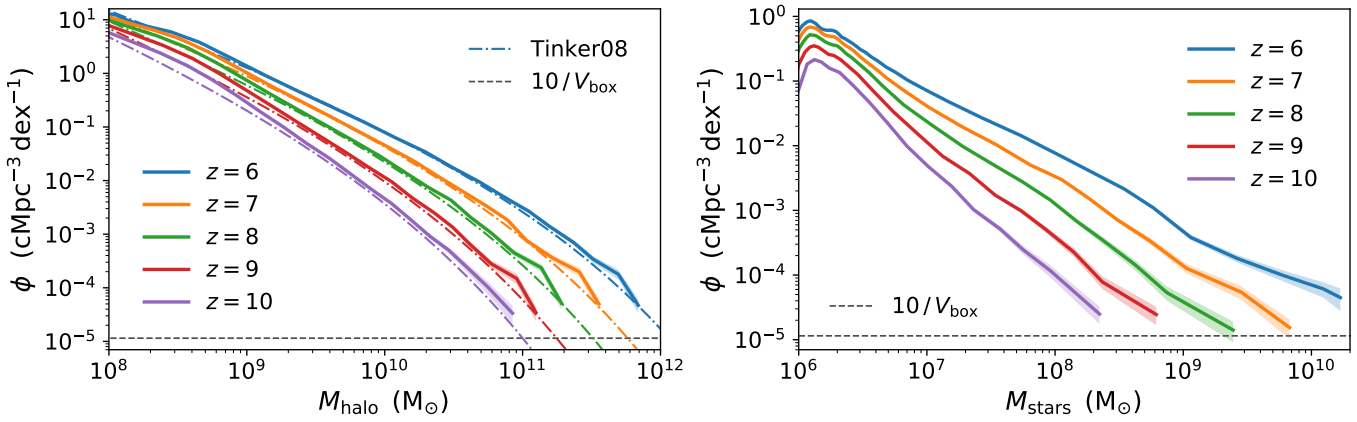

Figure 6. *Left-hand panel:* Galaxy halo mass functions for integer redshifts over the range $z \in [6, 10]$ showing the evolution of resolved structures ($M_{\rm halo} \gtrsim 10^8\,{\rm M}_\odot$). *Right-hand panel:* Galaxy stellar mass functions down to the clustering resolution of the simulation ($M_{\rm stars} \gtrsim 10^6\,{\rm M}_\odot$). The shaded regions represent the Poisson error ($\propto \sqrt{N}$) on the number counts in each bin. For reference we include a grey dashed line representing 10 objects within the simulation box. For comparison, in the left-hand panel we also include the analytic halo mass function model from Tinker et al. (2008), which was calibrated at higher masses and lower redshifts, specifically they optimized the agreement for halo masses of $M_{\rm halo} \sim 10^{11-15}\,{\rm M}_\odot\,h^{-1}$ at $z \sim 0$.

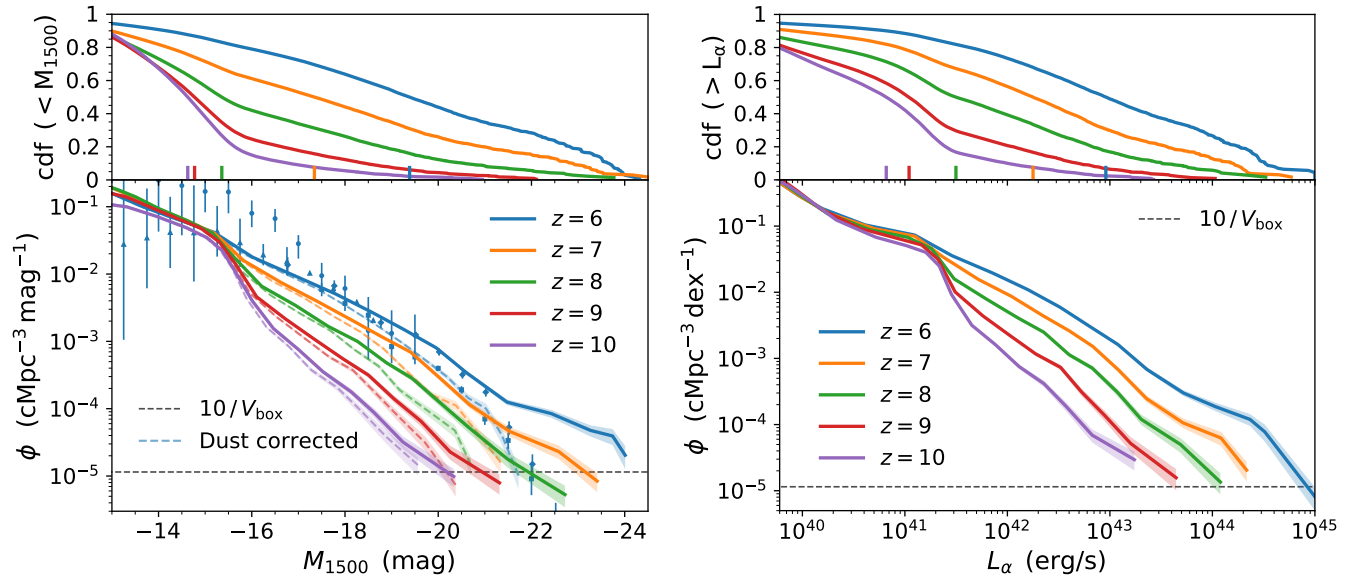

Figure 7. *Left-hand panel:* Galaxy UV (rest-frame 1500 Å) luminosity functions for integer redshifts over the range $z \in [6, 10]$ showing the emergence of bright galaxies and moderately steep faint-end slopes near the resolution limit ($M_{1500} \gtrsim -15$). The simulated results match observational estimates from Bouwens et al. (2015, diamonds), Finkelstein et al. (2015, squares), Livermore et al. (2017, circles) and Atek et al. (2018, triangles) over a wide range of magnitudes after applying the empirical relation presented in Gnedin (2014) to account for dust attenuation, which strongly affects luminous sources. The dust opacity is scaled according to the redshift-dependent dust-to-metal ratio given in Vogelsberger et al. (2020a). *Right-hand panel:* Galaxy intrinsic Lyα luminosity functions showing smooth evolution mirroring the UV luminosity functions but with additional recombination and collisional excitation emission. To give a sense of population convergence, in the upper panels we also show the normalized cumulative luminosity from haloes above a given brightness threshold, including vertical markers for the medians. As in Fig. 6 the shaded regions represent the Poisson error ($\propto \sqrt{N}$) and the horizontal dashed lines represent 10 objects within the simulation box.

the total mass of young stars (< 5 Myr) in each subhalo and combine this with the instantaneous SFR to define a duration over which we expect these stars to have formed: $\Delta t_{\rm SFH} = M_{\rm stars}(< 5\,{\rm Myr})/{\rm SFR}$. In the event that this duration is longer than 5 Myr we reassign the young stars as having all been formed with a constant SFR with the expected ages. This is done with 100 equal age bins in the interval $[0, \Delta t_{\rm SFH}]$, retaining the mass and metallicity distribution of each stellar population. This prescription helps us to smooth out an artificial bump around $M_{1500} \sim -15$ or $L_\alpha \sim 10^{41}\,{\rm erg\,s^{-1}}$, corresponding to newly spawned $\sim 5 \times 10^5\,{\rm M}_\odot$ star particles in low mass haloes, while the rest of the luminosities are essentially unaltered. Unless stated otherwise, we employ this SFH smoothing method for all quantities derived from star luminosities when required for individual subhaloes, including Lyα-centric fields.

Finally, before proceeding further we emphasize that the results presented in this paper are all without accounting for galaxy-scale Lyα radiative transfer effects. At the bright end, the observed luminosity function is expected to differ by up to two orders of magnitude

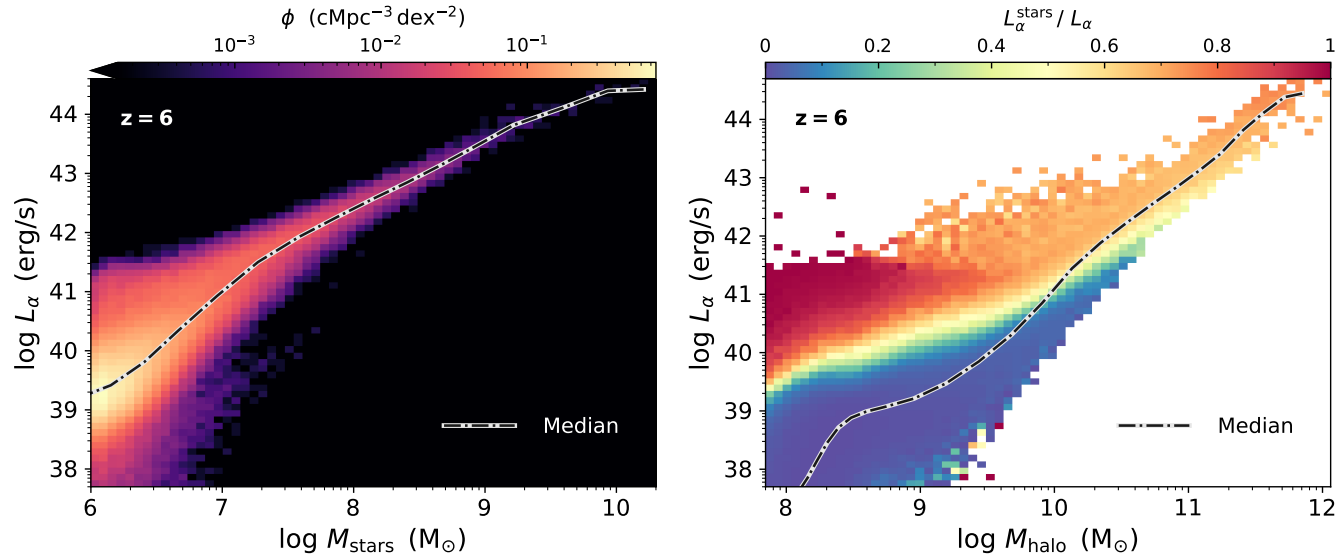

Figure 8. *Left-hand panel:* Total intrinsic Lyα luminosity L_α as a function of galaxy stellar mass, $M_{\rm stars}$, at $z = 6$. The colour axis shows the number density of haloes, which provides a sense for the scatter around the median (dash–dotted curve) caused by the wide range of star formation histories. *Right-hand panel:* Total intrinsic Lyα luminosity L_α and galaxy halo mass, $M_{\rm halo}$, at $z = 6$. The colour gradient in the fraction of emission from unresolved H II regions ($\sum L_\alpha^{\rm stars}/\sum L_\alpha$) within each bin) mirrors the physical diversity of star formation across different haloes. The median reveals where most of the haloes reside within the relation.

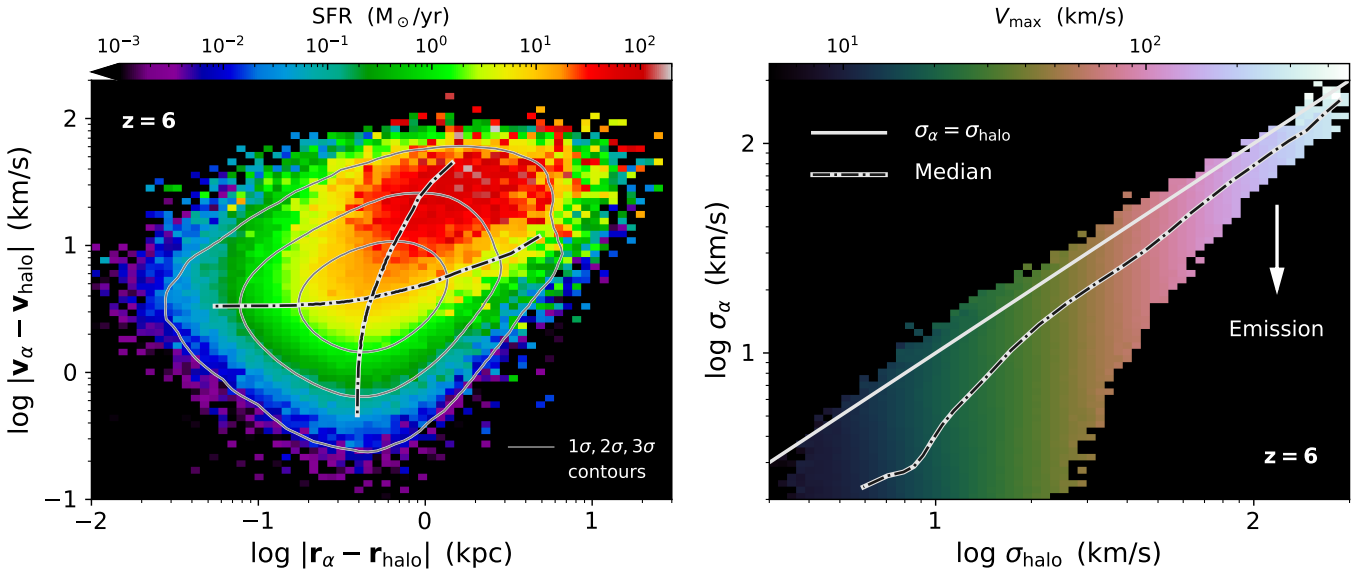

Figure 9. *Left-hand panel:* Position and velocity offsets between the intrinsic Lyα luminosity and mass centroids for each halo at $z = 6$. Galaxies with higher SFRs (denoted by colour) can differ by up to ~ 10 kpc and $\sim 100\,{\rm km\,s^{-1}}$, affecting the systemic position $\boldsymbol{r}_\alpha$ and velocity $\boldsymbol{v}_\alpha$ for IGM transmission. We show halo number density contours and medians of position and velocity offset bins. *Right-hand panel:* 1D velocity dispersions with intrinsic Lyα luminosity (σ_α) and mass ($\sigma_{\rm halo}$) weights for each halo. For reference, we include a line of equality, median halo counts, and bin-averaged colours showing the maximum of the rotation curve $V_{\rm max}$. The intrinsic Lyα line widths before resonant scattering are up to ≈ 2 times narrower than would be inferred by $\sigma_{\rm halo}$ or $V_{\rm max}$.

(comparing intrinsic values to Taylor et al. 2020, 2021) highlighting the crucial role of a proper RT treatment (Laursen et al. 2019; Garel et al. 2021). A comparison of luminosity functions in the literature reveals that such predictions remain highly uncertain, even from studies that carry out careful radiative transfer calculations. Given this, it would be surprising to see agreement between simulation groups given subtle but important differences in resolution, convergence, algorithm implementations, environmental effects, cosmic variance, and star formation, ISM, or dust modelling choices.

3.4 Galaxy correlations

We finish this section by exploring various Lyα correlations across galaxy populations. In Fig. 8 we show the tight power-law relationship between the total intrinsic Lyα luminosity L_α and galaxy stellar

mass $M_{\rm stars}$ at $z = 6$. The colour axis illustrates the number density of haloes, which provides a sense for the variation around the median within stellar mass bins (shown by the dash–dotted curve). In particular, we find a relatively large scatter in low mass haloes mostly as a result of the wide range of star formation histories. To investigate this further, we also show the relationship between the Lyα luminosity and galaxy halo mass $M_{\rm halo}$. There is a slightly larger scatter mirroring the physical diversity of star formation across different haloes. However, the colours provide information about the fraction of Lyα luminosity originating from unresolved H II regions, i.e. $\sum L_\alpha^{\rm stars} / \sum L_\alpha$ within each bin. In this case, the median curve is important to show where most haloes reside. At the low mass end there is a clear gradient from haloes almost entirely dominated by stars due to a recent starburst (red) to haloes with very little recent star formation such that recombination and cooling emission powers the luminosity (blue). At the high mass end the ratio matches the global value with minor deviations (see Fig. 3). This picture is consistent with star formation duty cycles and suppression prior to transitioning into a steady and sustained growth mode for higher mass haloes.

In examining the additional Lyα-centric fields we focus on the most relevant quantities for the study of IGM transmissivity in the next section. In Fig. 9 we illustrate the position and velocity offsets between the intrinsic Lyα luminosity centroids ($\boldsymbol{r}_\alpha$, $\boldsymbol{v}_\alpha$) and equivalent centre-of-mass quantities ($\boldsymbol{r}_{\rm halo}$, $\boldsymbol{v}_{\rm halo}$) for each halo at $z = 6$. While there is only a weak correlation and the majority of haloes are in reasonable agreement, we find that galaxies with higher SFRs can differ by up to ~ 10 kpc and $\sim 100\,{\rm km\,s^{-1}}$ as can be seen by the gradient towards the red average colour ($\gtrsim 10\,{\rm M_\odot/yr}$). As this can have an impact on IGM transmission calculations, in this study we prefer to set the systemic position $\boldsymbol{r}_\alpha$ and velocity $\boldsymbol{v}_\alpha$ based on these values from the Lyα catalogue. To provide further information about the statistics of galaxy populations, the contours show the halo number density and the dash–dotted curves are medians of position and velocity offset bins, respectively. In the right-hand panel of Fig. 9 we compare the 1D velocity dispersions calculated with intrinsic Lyα luminosity (σ_α) and mass ($\sigma_{\rm halo}$) weights for each halo. For reference, we include a line denoting equality ($\sigma_\alpha = \sigma_{\rm halo}$), median halo counts, and bin-averaged colours showing the maximum value of the spherically averaged rotation curve $V_{\rm max}$. In general, we find the intrinsic Lyα line widths before resonant scattering and other RT effects can be up to ≈ 2 times narrower than would be inferred by either $\sigma_{\rm halo}$ or $V_{\rm max}$. Of course, this is the expected behavior for emission processes due to both the clustered nature of starbursts and the selection bias towards high density gas for two-body ($\propto \rho^2$) recombination and collisional excitation emission.

4 TRANSMISSION CURVES FROM GALAXIES

The strong absorption of photons near the Lyα line provides a powerful probe of the structure and evolution of reionization. We thus explore various connections between galaxies and the IGM via Lyα transmission statistics from the flagship THESAN simulation.

4.1 Transmission catalogue procedures

In this work we provide a catalogue of Lyα IGM transmission curves that are designed to be as robust as possible for reuse with future studies. While we follow many of the procedures suggested by previous and concurrent authors, including Laursen et al. (2011), Jensen et al. (2014), Byrohl & Gronke (2020), Gronke et al. (2021), Garel et al. (2021), and Park et al. (2021), we briefly describe our parameter choices in case they differ. In particular, for each selected galaxy we extract 768 radially outward rays corresponding to equal area healpix directions of the unit sphere. To cut back on the amount of data storage and computation we focus on the central galaxies of each group. We expect satellites to generally be fainter than the central and have similar cosmological scale IGM transmission with the exception of minor localized sightline and velocity offset effects. Our catalogue includes all centrals with at least 32 star particles, but also includes less resolved centrals if they are among the 90% most massive centrals. Thus, our initial sample covers a large range of environments and galaxy histories for 44 700 galaxies and 3.43×10^7 ray extractions at $z = \{6, 7, 8, 9, 10, 11, 13\}$ – see Fig. 10 for cumulative number counts with halo mass. In fact, if satellites inherit transmission properties from the central halo then the catalogue can be viewed as nearly complete for all galaxies with resolved star formation histories.

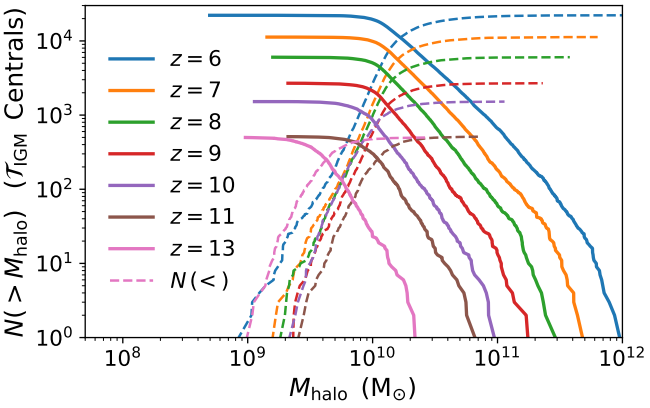

Figure 10. Cumulative number of central galaxies above (solid curves) and below (dashed curves) a given halo mass $M_{\rm halo}$ that are included in the initial IGM transmission catalogues for each redshift $z = \{6, 7, 8, 9, 10, 11, 13\}$.

As we focus on central galaxies, we start the rays at initial distances of $R_{\rm vir}$ for the entire group, defined as the radius within which the mean density becomes 200 times the cosmic value (R_{200}). Our choice is a compromise between sufficient proximity to capture the local imprint of the CGM and distance to ensure most photons will not scatter back into the line-of-sight, and in Appendix B we show that the resulting median statistics are unaffected by doubling this value. We take the systemic location and rest-frame from the subhalo catalogue described in Section 3, which accounts for the intrinsic Lyα luminosity averaged position and velocity including all emission sources, i.e. recombinations, collisional excitation, and stars. Although the Lyα spectra will be altered in the process of escaping galaxies, we believe these spatial and spectral anchors are slightly more accurate than using the halo position and velocity directly. We select a broad wavelength range of $\Delta v \in [-2000, 2000]\,{\rm km\,s^{-1}}$ sampled at a high spectral resolution of $5\,{\rm km\,s^{-1}}$ or a resolving power of $R \approx 60\,000$ for convergence and to cover a variety of future use cases. To ensure the bluemost frequencies are redshifted well into the red wing we perform the integrations out to a distance of $4000\,{\rm km\,s^{-1}}/H(z) \approx 40\,{\rm cMpc}\,[(1+z)/7]^{-1/2}$.

4.2 Ray extractions

Radiative transfer on Voronoi meshes has recently been adopted in the Lyα community (Smith et al. 2017b; Byrohl & Gronke 2020; Byrohl et al. 2021; Camps et al. 2021; Smith et al. 2021). However, while there are many approaches to integration for optically thin

radiative transfer, we have developed a low memory method for exact ray-tracing through the native Voronoi unstructured mesh data (implemented as an additional module within CoLT). This was done to avoid constructing the tessellation for the entire simulation box each time a new set of rays is desired, especially as this may not be feasible on local in-house computing facilities. The idea is to select particles from a cylindrical region around the ray, construct a localized tessellation from these points, and output 1D data objects based on ray-tracing through the smaller tessellation. This process produces identical results compared to ray-tracing through the full grid as long as each intersecting cell is included in the selection. Thus, we first calculate the maximum cell volume $V_{\rm max}$ over the entire simulation and set the cylinder impact radius as $r_{\rm max} = \eta V_{\rm max}^{1/3}$, where η is a factor accounting for the geometry. Based on convergence tests we use $\eta = 0.75$, slightly larger than the value for spherical cells of $(3/4\pi)^{1/3} \approx 0.62$. In most cases cells are significantly smaller than this value so an adaptive impact parameter would likely be more efficient but we did not experiment with such an approach.

The extraction process is simplified by recentring the box so that the origin is the same for all rays under consideration, and we assume this is true in the equations that follow. We note that periodic boundary conditions are implemented by checking all possible tilings and duplicating particle data as many times as is necessary, although the rays for our Lyα transmission study are smaller than half the box size so this is not necessary here. The distance from each point $\boldsymbol{r} = (x, y, z)$ to the line defined by the unit vector direction $\boldsymbol{n} = (n_x, n_y, n_z)$ passing through the origin is $b = \|\boldsymbol{r} - (\boldsymbol{r}\cdot\boldsymbol{n})\boldsymbol{n}\|$. The squared distance along the ray is then $d^2 = r^2 - b^2$, where the distance from the point to the origin is $r = \|\boldsymbol{r}\|$ and the sign of d is the same as that of $\boldsymbol{r}\cdot\boldsymbol{n}$. Thus a point is within the cylindrical region if $b < r_{\rm max}$ and $d \in (r_0 - r_{\rm max}, r_0 + l + r_{\rm max})$ where r_0 and l denote the starting radial offset and length of the ray, respectively. For convenience with ray-tracing after extraction we rotate all points to align with the z-axis and shift the start of the ray to the origin. The new points are located at $\boldsymbol{r}' = (x', y', z') = (x - n_x C_z, y - n_y C_z, \boldsymbol{r}\cdot\boldsymbol{n} - r_0)$, where the rotation constant is $C_z = (\boldsymbol{r}\cdot\boldsymbol{n} + z)/(1 + n_z)$. Finally, after selection and ray-tracing the raw particle data is written as a compact file containing the ray properties, origins, directions, length segments, and raw data for each cell in the traversed order.

4.3 Integration with continuous Hubble flow

For frequency-dependent optically thin Lyα radiative transfer it is convenient to convert to the dimension-less frequency

$$x \equiv \frac{\nu - \nu_0}{\Delta\nu_{\rm D}}\,, \tag{4}$$

where $\nu_0 = 2.466 \times 10^{15}$ Hz denotes the frequency at line centre and $\Delta\nu_{\rm D} \equiv (v_{\rm th}/c)\nu_0$ the Doppler width of the profile. The frequency dependence of the absorption coefficient is given by the Voigt profile $\phi_{\rm Voigt}$. For convenience we define the Hjerting–Voigt function $H(x) = \sqrt{\pi}\Delta\nu_{\rm D}\phi_{\rm Voigt}(\nu)$ as the dimension-less convolution of Lorentzian and Maxwellian distributions,[5]

$$H(x) = \frac{a}{\pi}\int_{-\infty}^{\infty}\frac{e^{-y^2}\,{\rm d}y}{a^2 + (y-x)^2} \approx \begin{cases} e^{-x^2} & \text{'core'} \\ \dfrac{a}{\sqrt{\pi}x^2} & \text{'wing'}, \end{cases} \tag{5}$$
$$= {\rm Re}\left(e^{(a-ix)^2}\,{\rm erfc}(a-ix)\right) \approx e^{-x^2} + \frac{2a}{\sqrt{\pi}}\left(2xF(x) - 1\right).$$

Here the 'damping parameter', $a \equiv \Delta\nu_L/2\Delta\nu_D \approx 4.7 \times 10^{-4}\,T_4^{-1/2}$, describes the relative broadening compared to the natural line width $\Delta\nu_{\rm L} = 9.936 \times 10^7$ Hz. The final approximation is the first order expansion in a, and the (complex) complementary error function is related to the area under a Gaussian by ${\rm erfc}(z) \equiv 1 - 2\int_0^z e^{-y^2}\,{\rm d}y/\sqrt{\pi}$ and the Dawson integral is $F(x) \equiv \int_0^x e^{y^2 - x^2}\,{\rm d}y$.

We now introduce a new method for calculating the traversed optical depth based on continuous Doppler shifting due to velocity gradients encountered during propagation. For the specific case of cosmological Hubble flow the expansion induces a constant redshift per unit distance such that the change in velocity is $\Delta v = H(z)\Delta\ell$. Thus, photons experience continuous Doppler shifting that can be modelled by a position-dependent frequency as

$$\frac{\Delta\lambda}{\lambda} = \frac{\Delta v}{c} \quad \Rightarrow \quad x' = x - \frac{H(z)}{v_{\rm th}}\Delta\ell\,, \tag{6}$$

where we have used the relation between Doppler frequency and velocity: $x = -\Delta v/v_{\rm th}$. The resulting optical depth in this case is

$$\tau = k_0\int_0^{\ell} H\left(x - \mathcal{K}\ell'\right)\,{\rm d}\ell' \tag{7}$$
$$\approx \frac{\sqrt{\pi}k_0}{2\mathcal{K}}\left[{\rm erf}(x) - {\rm erf}(x - \mathcal{K}\ell)\right] + \frac{2ak_0}{\sqrt{\pi}\mathcal{K}}\left[F(x - \mathcal{K}\ell) - F(x)\right],$$

where $\mathcal{K} \equiv H(z)/v_{\rm th}$ and the absorption coefficient at line centre is $k_0 \equiv n_{\rm H\,I}\sigma_0$ with cross-section of $\sigma_0 = f_{12}\sqrt{\pi}e^2/(m_e c\Delta\nu_{\rm D})$ and oscillator strength of $f_{12} = 0.4162$. The final expression employs the first-order expansion from equation (5), which is sufficiently accurate for the Lyα line although it is possible to include higher order terms in a if desired. We note that for numerical stability if $\mathcal{K}\ell \ll 1$, corresponding to $\ell \ll 18.4\,{\rm kpc}\,T_4^{1/2}[(1+z)/7]^{-3/2}$, then it is suitable to use the static approximation for the optical depth: $\tau = k_0 H(x)\ell$. Finally, this scheme can also be used to incorporate exact Doppler shifting for other scenarios such as underresolved galactic winds by incorporating the local LOS velocity gradient relative to the comoving frame of the gas. A similar gridless Monte Carlo radiative transfer scheme has also been employed to accurately integrate through arbitrary density gradients (Lao & Smith 2020).

As we have already extracted the ray segments into small data arrays the IGM absorption approximation results in a series of independent radiative transfer calculations for each input frequency and segment. Specifically, we adopt notation for the $i^{\rm th}$ frequency index and $j^{\rm th}$ path segment such that for a given initial velocity offset Δv_i (i.e. reference frequency), unit direction $\boldsymbol{n}$, LOS peculiar velocity relative to the systemic velocity $v_j \equiv \boldsymbol{n}\cdot(\boldsymbol{v}_{{\rm pec},j} - \boldsymbol{v}_{\rm sys})$, and segment starting distance r_j (all in physical units), the Doppler frequency becomes $x_{i,j} = -(\Delta v_i + v_j + H(z)r_j)/v_{{\rm th},j}$. The optical depth $\Delta\tau_{i,j}$ contributed by each segment is the result of evaluating equation (7) with this frequency over the path length $\Delta\ell_j$. In practice, the total optical depth $\tau(\Delta v_i) = \sum_j \Delta\tau_{i,j}$ defines the frequency-dependent

[5] We employ the standard notation of H for both the Hubble parameter and the line profile, but the meaning can be understood from the context.

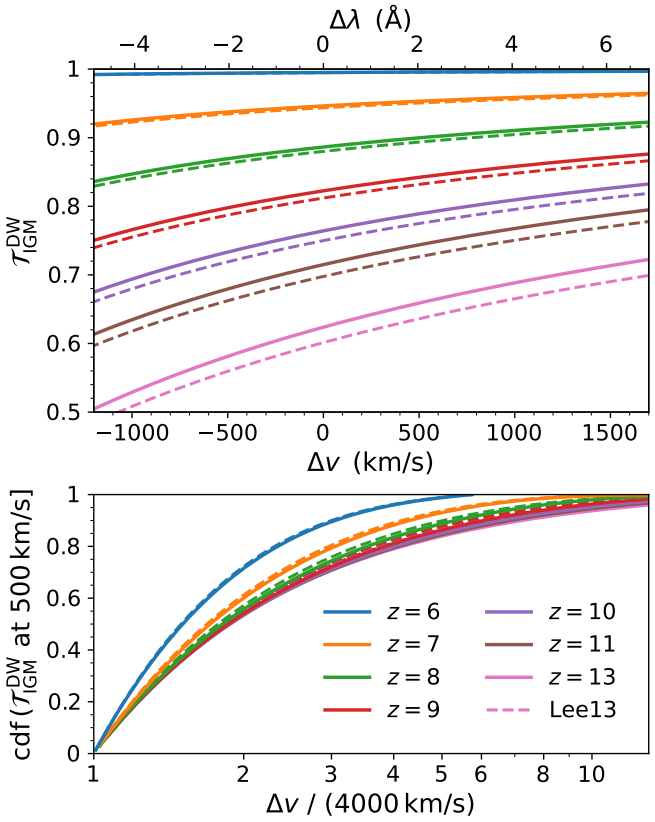

Figure 11. *Top:* Damping-wing IGM transmission $\mathcal{T}_{\rm IGM}^{\rm DW} = \exp(-\tau_{\rm DW})$ beyond the local rays of $\Delta v_s = 4000\,{\rm km\,s^{-1}}$ as a function of velocity offset Δv and rest-frame wavelength offset $\Delta\lambda$. This large-scale contribution cannot be ignored at pre-reionization redshifts. The solid curves assume a standard Lorentzian profile while the dashed curves include quantum-mechanical corrections to the Voigt profile (Lee 2013). *Bottom:* Cumulative distribution functions of the damping-wing absorption as a function of traversed velocity offset. Full convergence requires light-cone distances of up to ∼ 10 times longer than the detailed local calculations considered in this study.

transmission function for each ray:

$$\mathcal{T}(\Delta v) \equiv \exp\left[-\tau(\Delta v)\right], \tag{8}$$

which describes the fraction of flux not attenuated by the IGM after escaping the halo. Thus, we have a statistical framework for Lyα transmission directly connected to the galaxy catalogues.

4.4 Damping-wing absorption

Similar to Park et al. (2021), we choose to incorporate the distant IGM beyond the local rays in a statistical sense based on the global reionization history. A more accurate treatment requires considering the time-evolution of ionized bubbles through $\gtrsim 200\,{\rm cMpc}$ light cones, which we will explore in a future study. We emphasize that this additional damping-wing absorption has a very minor impact after the EoR but is increasingly important at higher redshifts (Miralda-Escudé 1998). We choose to frame the integration in terms of the source redshift z_s, initial velocity offset Δv, and global volume-averaged neutral hydrogen fraction $\langle x_{\rm H\,I}\rangle$. The cosmological neutral hydrogen number density is then

$$\langle n_{\rm H\,I}\rangle = \frac{\Omega_b X}{m_{\rm H}}\frac{3H_0^2}{8\pi G}\langle x_{\rm H\,I}\rangle(1+z)^3\,, \tag{9}$$

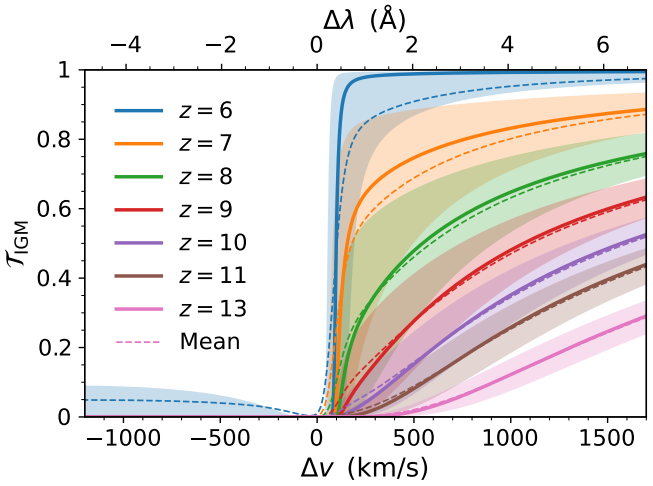

Figure 12. IGM transmission $\mathcal{T}_{\rm IGM}$ as a function of velocity offset Δv and rest-frame wavelength offset $\Delta\lambda$ around the Lyα line at each redshift. The solid (dashed) curves show the catalogue median (mean) statistics and shaded regions give the 1σ confidence levels. With equal weight to all galaxies and sightlines, this view is biased towards lower mass that dominate the catalogues by halo count. However, this overview nicely illustrates blue peak suppression and red damping-wing absorption throughout the EoR.

where the hydrogen mass fraction is $X = 0.76$ and Hubble constant is $H_0 = 100\,h\,{\rm km/s/Mpc}$. The wing cross-section depends only on the photon frequency, which in terms of velocity offsets becomes

$$\sigma(\Delta v) \approx \frac{a\sigma_0}{\sqrt{\pi}x^2} = \frac{f_{12}ce^2\Delta\nu_{\rm L}}{2m_e\nu_0^2\Delta v^2}\,. \tag{10}$$

However, cosmological redshifting continuously changes the frequency such that if we ignore peculiar motions then the relation $\nu(z) = \nu(z_s)(1+z)/(1+z_s)$ translates to offsets of

$$\frac{\Delta v(z)}{c} = \frac{\nu_0 - \nu(z)}{\nu_0} = \frac{z_s - z + (1+z)\Delta v/c}{1+z_s}\,. \tag{11}$$

The effective redshift corresponding to the end of the local rays is $z_{s,\rm eff} = z_s - (1+z_s)\Delta v_s/c$, where the ray length is characterized by the local velocity offset, which in our case is $\Delta v_s = 4000\,{\rm km\,s^{-1}}$. Thus, the traversed optical depth in physical units is

$$\tau_{\rm DW} = \int_0^{z_{s,\rm eff}} \langle n_{\rm H\,I}\rangle\sigma\left[\Delta v(z)\right]\frac{c\,{\rm d}z}{(1+z)H(z)}\,. \tag{12}$$

If the reionization history is a step function then exact analytical expressions can be found for both pure Lorentzian wings and with the common modification of an additional $(\nu/\nu_0)^4$ dependence appropriate for Rayleigh scattering (Miralda-Escudé 1998). However, we employ a numerical integration over the high redshift resolution ($\Delta z \lesssim 0.005$) reionization history directly from the simulations. Furthermore, we adopt the complete first-order quantum-mechanical correction to the Voigt profile presented by Lee (2013),

$$\sigma_{\rm Lee}(\Delta v) \approx \sigma(\Delta v)\ (1 + 1.792\,\Delta v/c)\,, \tag{13}$$

which strengthens the red wing due to positive interference of scattering from all other levels.

In Fig. 11 we show this transmission $\mathcal{T}_{\rm IGM}^{\rm DW} = \exp(-\tau_{\rm DW})$ beyond the local rays as a function of velocity offset Δv and rest-frame wavelength offset $\Delta\lambda$. In the idealized model given by equation (12) the statistical absorption depends mainly on the redshift, reionization history, and local ray lengths. In detail, we expect a range of values

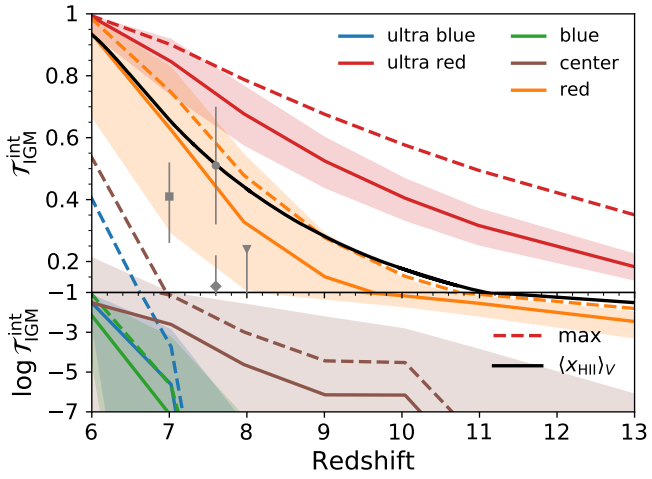

Figure 13. Evolution of the integrated IGM transmission $\mathcal{T}^{\rm int}_{\rm IGM}$ over five broad wavelength ranges from (ultra) blue to red (see the text for details). The median blue peak suppression remains strong until the universe is fully ionized, while red peaks are sensitive probes of the global reionization history (black curve). We also show the median maximum transmission spike in each spectral window yielding more optimistic prospects (dashed curves). For reference, we also include observational constraints from the detection of Lyα emission in Lyman break selected galaxies, which may exhibit properties of both the red and central bands (Mason et al. 2018a – square; Mason et al. 2019 – triangle; Hoag et al. 2019 – diamond; Jung et al. 2020 – circle).

due to variations in the neutral hydrogen density of individual light cones. We also warn that our assumption of a homogeneous Universe in the damping-wing calculation will result in too much absorption on average. While we plan to investigate this further in a future study, we expect our conservative estimates will be most uncertain around the midpoint of reionization. This is because the clumping factor in the IGM remains close to unity beforehand ($C_{100} \approx 1$; see fig. 17 in Paper I) while the damping-wing transmission saturates to unity afterwards ($\mathcal{T}^{\rm DW}_{\rm IGM} \approx 1$). Still, this large-scale contribution cannot be ignored at high redshifts and is therefore included in all of the analysis in this paper, including the quantum-mechanical correction in equation (13). To explore where this extra opacity comes from we also plot the corresponding cumulative distribution functions against the traversed velocity offset. About half of the large-scale damping-wing scattering optical depth originates within an additional $\approx 3000\,{\rm km\,s^{-1}}$. However, we emphasize that achieving per cent level accuracy for the EoR damping-wing opacity requires a light-cone procedure in simulations with distances of up to ~ 10 times longer than the detailed local calculations considered carried out in this study.

4.5 Redshift dependence

We first consider the redshift evolution of global catalogue statistics, in which we give equal weight to all galaxies and sightlines. In Fig. 12 we show the IGM transmission $\mathcal{T}_{\rm IGM}$ including the local and damping-wing contributions as a function of velocity offset Δv and rest-frame wavelength offset $\Delta\lambda$ around the Lyα line at redshifts of $z = \{6, 7, 8, 9, 10, 11, 13\}$. The solid (dashed) curves show the full median (mean) statistics while the shaded regions give the 1σ confidence levels. We emphasize that this view is biased towards low mass haloes that dominate the number counts. Therefore, our focus is primarily to provide a reference comparison for similar works. Still, the full frequency dependence clearly and intuitively illustrates the blue peak suppression and red damping-wing absorption throughout the EoR. In the following subsections we explore the rich diversity of transmission properties in greater detail.

For a complementary perspective of the rapid change in transmissivity throughout the EoR we consider band averages. In particular, we focus on five velocity offset Δv spectral windows that are relevant for Lyα related science. The 'ultra blue' band of $(-2000, -500)\,{\rm km\,s^{-1}}$ represents a baseline for absorption where most of the flux is free from the immediate proximity of the host halo before redshifting into resonance. The 'blue' band of $(-500, -100)\,{\rm km\,s^{-1}}$ is sensitive to both the local halo overdensity and the ionized bubble statistics. The 'centre' band of $(-100, 100)\,{\rm km\,s^{-1}}$ is meant to give wiggle room for systemic and peculiar velocities that blend the blue and red bands. The 'red' band of $(100, 500)\,{\rm km\,s^{-1}}$ is most relevant for LAE surveys as this corresponds to the location of observed red peaks at low and high redshifts (e.g. Ouchi et al. 2020). The 'ultra red' band of $(500, 2000)\,{\rm km\,s^{-1}}$ is also mostly detached from proximity effects and thus reflects damping-wing absorption in the context of LSS and bubble size distributions.

To summarize the band properties we consider the following metrics taken over each wavelength range: first, the integrated or mean transmission defined as $\mathcal{T}^{\rm int}_{\rm IGM} \equiv \int \mathcal{T}_{\rm IGM}\,{\rm d}\Delta v / \int {\rm d}\Delta v$, and secondly, the maximum transmission defined as $\mathcal{T}^{\rm max}_{\rm IGM} \equiv \max(\mathcal{T}_{\rm IGM})$. In Fig. 13 we show the redshift evolution of the integrated and maximum IGM transmission over each band. Although neither of these statistics are directly observable, as transmission also depends on the input Lyα spectra emerging from the galaxy, we expect these roughly trace what would be inferred by isolating IGM effects, especially for the non-saturated red and ultra red bands. We find that the median blue peak suppression remains strong until the universe is fully ionized, with slightly higher transmission for the ultra blue band compared to the blue band due to the reduced proximity effects. Overall, the maximum transmission statistics yield systematically more optimistic prospects for LAE detectability, especially if these transmission spikes coincide with emergent Lyα peaks (shown by the dashed curves). We find it encouraging that the red band seems to provide a sensitive probe of the global reionization history (shown by the black curve). The ultra red band does the same for sources in the damping wing such as faint quasars or gamma-ray burst afterglows found in futuristic high-redshift surveys (e.g. Lidz et al. 2021).

A more equitable view of the catalogue is given by considering the distributions of band values across all haloes. In Fig. 14 we quantify the relative probability for a given integrated ($\mathcal{T}^{\rm int}_{\rm IGM}$, bottom panels) and maximum ($\mathcal{T}^{\rm max}_{\rm IGM}$, top panels) band transmission at each redshift. For reference we also show the cumulative distribution functions as dashed curves and the median and 1σ summary statistics in the middle panels. It is extremely unlikely to have non-negligible transmission spikes in the blue bands at $z \gtrsim 8$, although this is certainly allowed (int) or even common (max) at $z \lesssim 6$. When also considering the central band in rare cases it is plausible to witness multiple-peaked or otherwise complex spectral line profiles (e.g. see the discussions by Byrohl & Gronke 2020; Mason & Gronke 2020; Gronke et al. 2021; Park et al. 2021). This also stresses the need for systemic tracers beyond Lyα to distinguish IGM signatures and pinpoint the origins of various spectral features. Perhaps more significant is the broad distribution in the red band. The transmission is relatively high after the midpoint of reionization $z \lesssim 7.67$, and roughly follows the global ionized fraction (see Fig. 13). However, as will be shown later the exponential sensitivity on optical depth ($\mathcal{T}_{\rm IGM} = e^{-\tau_{\rm IGM}}$) allows the same galaxy to have both large and small values, i.e. the bimodal-

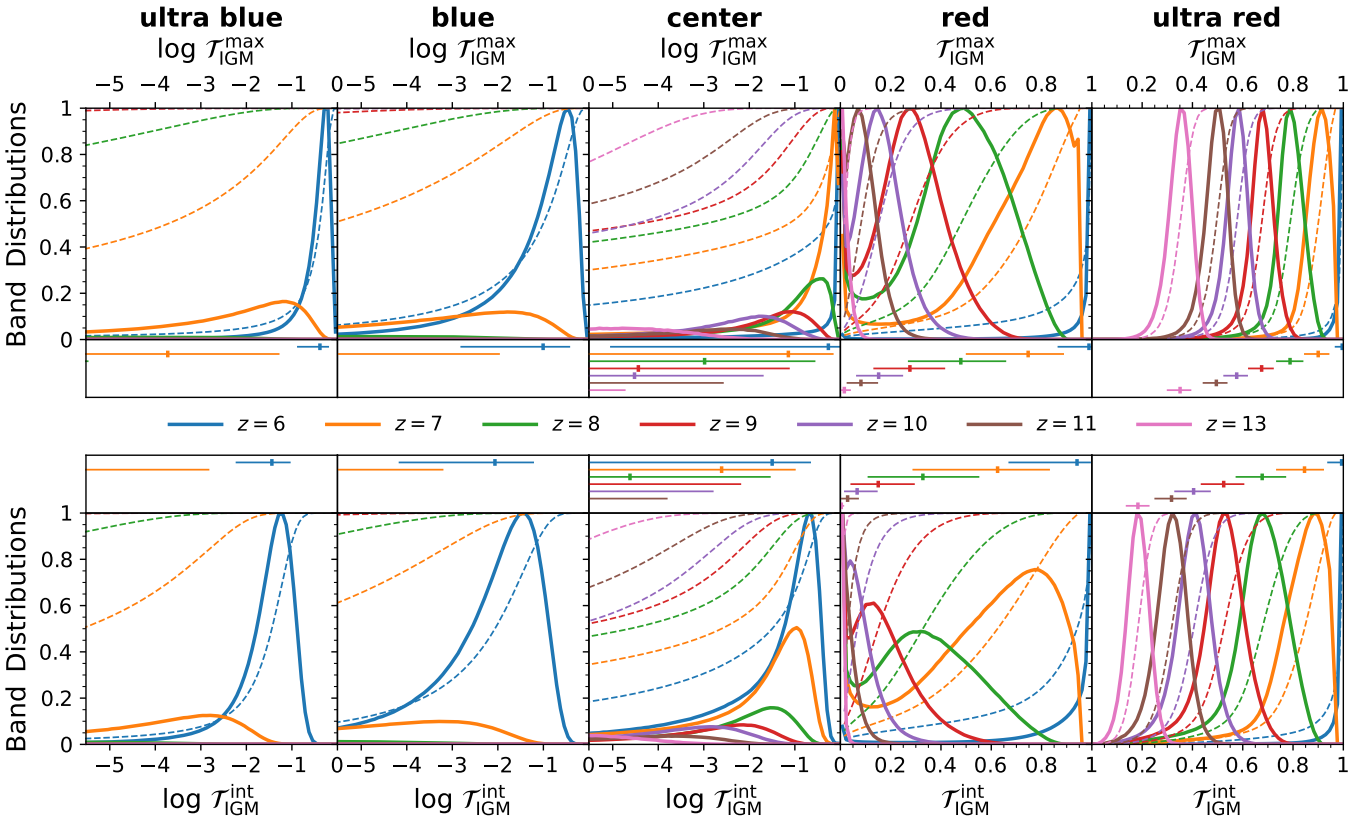

Figure 14. Relative probability distributions for a given integrated ($\mathcal{T}_{\rm IGM}^{\rm int}$, bottom panels) and maximum ($\mathcal{T}_{\rm IGM}^{\rm max}$, top panels) band transmission at each redshift (different colour curves) considering all haloes. For reference we also show the cumulative distribution functions as dashed curves and the median and 1σ summary statistics in the middle panels. See the text for further discussion, but this perspective reveals a complex landscape of broad, skewed, or bimodal distributions.

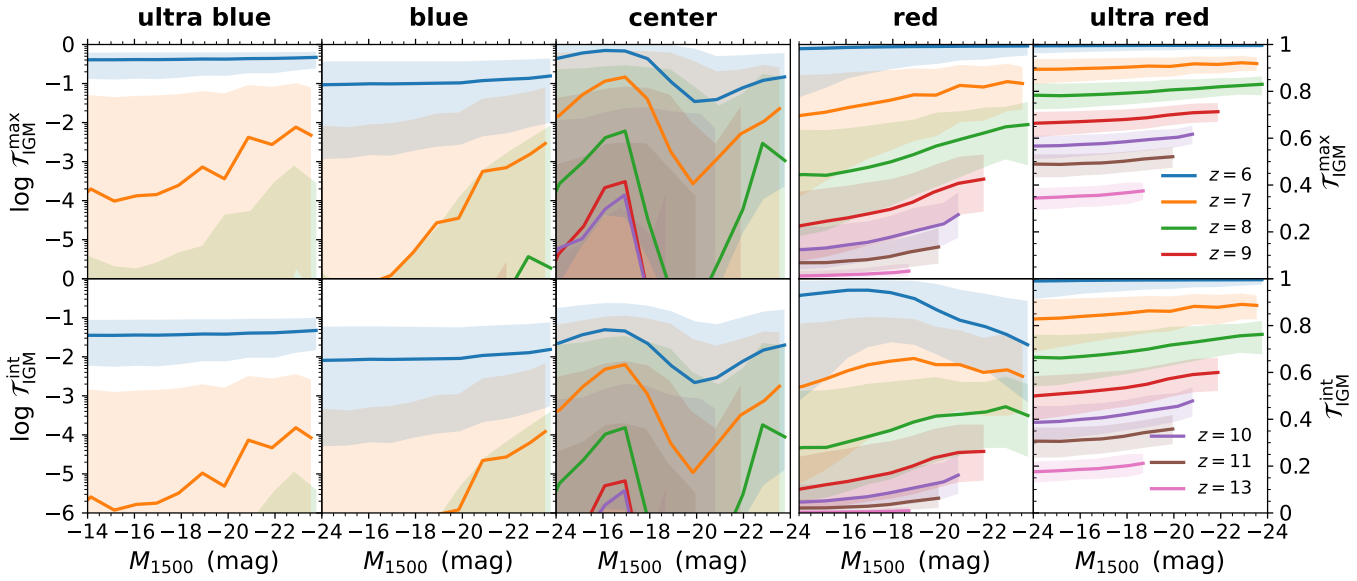

Figure 15. Band integrated ($\mathcal{T}_{\rm IGM}^{\rm int}$, bottom panels) and maximum ($\mathcal{T}_{\rm IGM}^{\rm max}$, top panels) IGM transmission median and 1σ statistics as a function of UV magnitude M_{1500} for each redshift. The highly suppressed blue bands are shown with a log scale axis, while the red bands employ a linear axis. We find a clear trend of high transmission for UV bright galaxies, which is especially evident for $\mathcal{T}_{\rm IGM}^{\rm max}$. However, there is significant absorption that encroaches into the red band for $\mathcal{T}_{\rm IGM}^{\rm int}$ due to the complex distribution of high-velocity neutral gas structure around these galaxies.

ity is not due to halo mass or environment alone. In fact, there will always be sightlines with very low transmission due to filaments and other self-shielding structures common at high-z. The location and broad nature of the higher transmission peak is redshift dependent, which is an important consideration when using LAEs as probes of reionization. We emphasize that the sharp cutoffs below $\mathcal{T}_{\rm IGM} \approx 1$ are not physical but are due to the global treatment of the long-range damping-wing absorption. For example, at $z = 7$ the red band values cannot exceed $\mathcal{T}_{\rm IGM} \approx 0.95$ as every sightline on average experiences at least 5% absorption via cosmological integration throughout the remainder of the EoR. Finally, we note that the ultra red band is less susceptible to resonant scattering and halo proximity effects and is therefore a cleaner probe of the global state of the IGM if this can be robustly measured with deep spectroscopy for large numbers of high-z galaxies.

4.6 Dependence on UV magnitude

We now explore the dependence on UV magnitude (without any dust correction), which serves as an observational indicator for young stellar populations. In fact, UV bright galaxies are expected to have high ionizing photon budgets and therefore promote inside out reionization of their local bubbles. On the other hand, these same galaxies may give rise to relatively low and sightline-dependent Lyman continuum escape fractions. High dust contents, crowded environments, and cosmic streams of infalling gas introduce additional complexity that can offset the large bubble advantage for IGM transmission. In Fig. 15 we explore the band integrated ($\mathcal{T}_{\rm IGM}^{\rm int}$, bottom panels) and maximum ($\mathcal{T}_{\rm IGM}^{\rm max}$, top panels) IGM transmission as a function of UV magnitude M_{1500} for each redshift based on the median and 1σ statistics. We find that at $z \gtrsim 7$ the highly suppressed blue bands exhibit less transmission for fainter galaxies, although this seems to wash out by $z = 6$ at the tail end of reionization. However, the most interesting aspect of the UV dependence is in the red band, where the maximum statistic $\mathcal{T}_{\rm IGM}^{\rm max}$ clearly shows increasing transmissivity for brighter galaxies at all redshifts. This means that fainter galaxies are more likely to be universally suppressed across the entire red spectral window, or conversely that brighter galaxies have an advantage for transmission somewhere within $\leq 500\,{\rm km\,s^{-1}}$. We note that the effect is present but less dramatic for the ultra red band.

Somewhat counter-intuitively, the integrated red band transmission $\mathcal{T}_{\rm IGM}^{\rm int}$ seems to level off or even dip for the brightest haloes. This can be understood in terms of a step function model in the frequency dependence with an increasing fraction of galaxies and sightlines being cut off above the lower edge of $100\,{\rm km\,s^{-1}}$. The basic argument is that if a photon originates on the blue side of line centre then given the high Gunn–Peterson absorption it will be eliminated as soon as it is Hubble redshifted into resonance. However, if the peculiar gas velocity is infalling near the source then even photons redwards of the systemic line centre may be viewed as blue in the comoving frame of the gas. Thus, the frequency range likely to encounter a resonance point is extended to approximately the circular velocity, which is often used to estimate gravitational infall velocities (for further discussion see e.g. Santos 2004; Dijkstra et al. 2007):

$$V_c = \sqrt{\frac{GM_{\rm halo}}{R_{\rm vir}}} \approx 144\,{\rm km\,s^{-1}} \left(\frac{M_{\rm halo}}{10^{11}\,{\rm M}_\odot}\right)^{1/3} \left(\frac{1+z}{7}\right)^{1/2}, \quad (14)$$

assuming the *Planck* cosmological parameters and a standardized overdensity constant of $\Delta_c = 200$. Still, despite this encroaching absorption feature, observational studies are consistent with no evolution in massive galaxies while faint ones exhibit a drop in the LAE

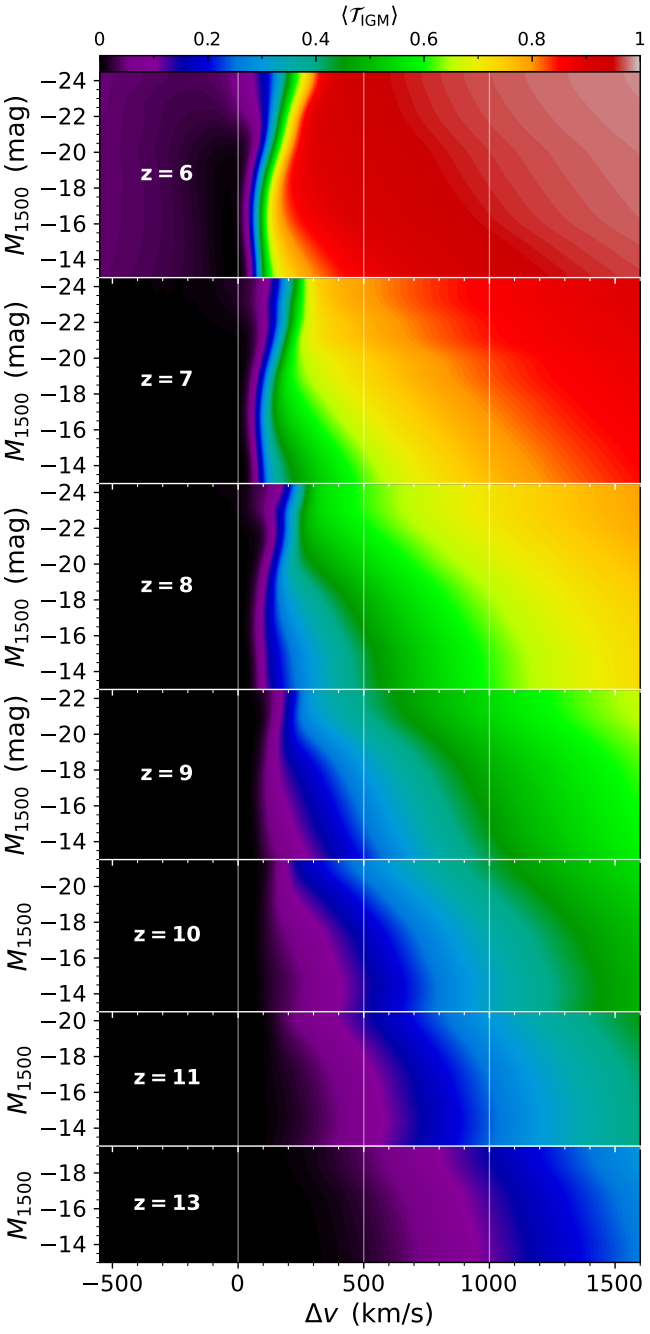

Figure 16. Average IGM transmission $\langle \mathcal{T}_{\rm IGM} \rangle$ as a function of velocity offset Δv and UV magnitude M_{1500}. This perspective illustrates the non-trivial dependence on both of these quantities. Brighter galaxies provide a clear advantage for red wing photons. At the same time, the sharp transition frequency can reach higher velocity offsets, undercutting IGM transparency for Lyα photons escaping near line centre in these environments.

population fraction (e.g. Stark et al. 2011; Endsley et al. 2021). This is consistent with our results if bright galaxies have larger velocity offsets and broader line widths, as is expected based on lower redshift observations (Yang et al. 2016; Verhamme et al. 2018) and a number of theoretical arguments in the literature (also recall Fig. 9).

Unfortunately, the THESAN volumes are not large enough to assess the statistical behavior of the brightest galaxies at $z \gtrsim 10$. In the pre-reionized Universe the ensemble average detectability of first galaxy

LAEs may be discouragingly low as more and more photons are lost to the diffuse Lyα background (Loeb & Rybicki 1999; Visbal & McQuinn 2018). Furthermore, even with the increased sensitivity of next-generation surveys it may be difficult to probe large enough volumes with sufficient depth to detect enough rare bright galaxies to discern between various IGM transmission models, which can differ by an order of magnitude in the predicted emission and transmissivity (Smith et al. 2015, 2017a). However, the combination of a sharp frequency cutoff and gradual wing opacity is also what makes Lyα (non-)detections a rich probe of EoR physics. In summary, based on extrapolating the current trends there is room for cautious optimism in which fortunate sightlines, bubble sizes, and emergent line profiles facilitate minimal local and damping-wing absorption.

To further investigate this behavior, in Fig. 16 we show the average IGM transmission $\langle \mathcal{T}_{\rm IGM} \rangle$ as a function of velocity offset Δv and UV magnitude M_{1500}. This image vividly explains the turnover in the red band relation at $M_{1500} \lesssim -20$ seen in Fig. 15. One of the key features is that brighter galaxies provide a clear advantage for red wing photons at all redshifts, due to residing within larger ionized bubbles. At the same time, the sharp transition frequency tends to curve towards higher velocity offsets with a large spread by $z = 6$. This encroachment into the red band acts to undercut IGM transparency for Lyα photons escaping near line centre in these environments. Another important feature is that (on average) blue photon transmission is permitted by $z = 6$. In fact, lower redshift studies also exhibit the pattern of higher absorption at line centre before transitioning to bluer wavelengths that are less susceptible to halo proximity effects (see e.g. Laursen et al. 2011).

4.7 Covering fractions

The concept of covering fractions provides a fundamentally different perspective on the mechanisms responsible for hampering Lyα visibility at the tail end of reionization. Intriguingly, we find that IGM transmission in wavelength ranges where we expect Lyα line emission can be highly anisotropic. In general, emission will be observed as close to line centre as allowed by the halo escape and IGM transmission physics. To quantify this effect, we define the covering fraction of each individual galaxy as the fraction of sightlines with transmission below 20 per cent, i.e. $P(\mathcal{T}_{\rm IGM} < 0.2)$. In Fig. 17 we show the median and 1σ range of halo covering fractions as a function of UV magnitude M_{1500} (upper panels). For comparison, we provide these statistics at the physically relevant velocity offsets of $\Delta v = 200$ and $400\,{\rm km\,s^{-1}}$, averaged over wavelength windows of $50\,{\rm km\,s^{-1}}$ to match typical spectroscopic instrument capabilities. The velocity offset windows are chosen to capture the impact of IGM transmission on observed red peaks nearer and farther from line centre. The general trend is that fainter galaxies have larger covering fractions, corresponding to more isotropic suppression. Interestingly, there is also an upturn for bright galaxies ($M_{1500} \lesssim -20$) at $200\,{\rm km\,s^{-1}}$ caused by anisotropic cold gas accretion around these clustered environments. As expected, this effect washes out by $400\,{\rm km\,s^{-1}}$ where the red peak transmission is less affected by the local high velocity infall. Thus, there is a relatively large optimal M_{1500} range for observing transmission. In Table 2 we summarize the redshift and frequency dependence of covering fractions by providing $P(\mathcal{T}_{\rm IGM} < 0.2)$ including all galaxies with UV brightness $M_{1500} < -19$ for a grid of several observationally relevant velocity offsets, which also demonstrates the relative (in)sensitivity to the chosen velocity offset windows.

We find that these qualitative features are quite robust, although the details depend on the threshold criterion. The value of 20 per cent represents a substantial (but unsaturated) degree of absorption at $z \lesssim 9$. To better understand the impact of redshift evolution we also show a measure of bimodality at $200\,{\rm km\,s^{-1}}$ (bottom panel). This calculation is based on a common test employing moments around the mean. Specifically, if the mean is $m_1 \equiv \langle \mathcal{T}_{\rm IGM} \rangle$ and unstandardized moments are $m_k \equiv \langle (\mathcal{T}_{\rm IGM} - m_1)^k \rangle$, then the standard deviation is $\sigma \equiv \sqrt{m_2}$, skewness is $\gamma \equiv m_3/\sigma^3$, and kurtosis is $\kappa \equiv m_4/\sigma^4$, such that the inverted bimodality coefficient is $b^{-1} = 1/(\kappa - \gamma^2) \in [0, 1]$. We note that we choose to plot b^{-1} as this more intuitively depicts higher bimodality with larger values. Our results demonstrate that brighter galaxies exhibit a more on/off absorption morphology in comparison to less bright galaxies ($M_{1500} \sim -18$). Furthermore, we also find the bimodality is much less significant at higher redshifts ($z \gtrsim 10$) as low opacity sightlines all but disappear. Of course, this result should not be overinterpreted as this does not imply bimodality in optical depths, but rather the presence of viewing angles with $\tau_{\rm IGM}^{200}$ greater than and less than unity around the same halo.

Finally, to more clearly visualize the covering fraction results, in Fig. 18 we show the angular distributions of the IGM transmission $\mathcal{T}_{\rm IGM}^{200}$ around a velocity offset of $200\,{\rm km\,s^{-1}}$ for a prototypical subset of galaxies. We select the most massive haloes at $z = \{6, 7, 8\}$ as well as haloes with masses 10 and 100 times smaller as listed in the figure. We provide the mean (left-hand side) and median with 1σ confidence intervals (right-hand side) statistics below each image based on 3072 healpix directions of equal solid angle. The maps reveal a

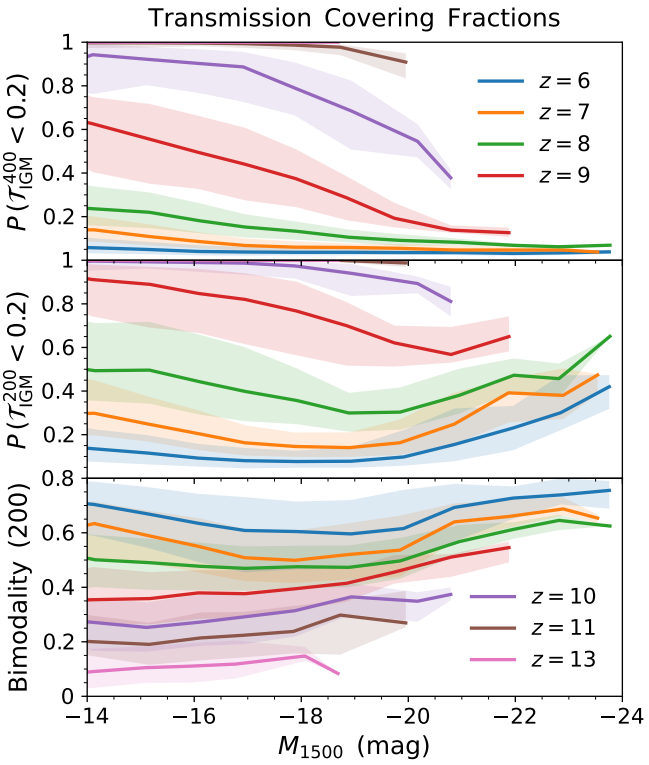

Figure 17. IGM transmission covering fractions defined as the fraction of sightlines around each galaxy with transmission below 20 per cent, i.e. $P(\mathcal{T}_{\rm IGM} < 0.2)$. The curves and shaded regions give the median and 1σ variation as a function of UV magnitude M_{1500} at velocity offsets of $\Delta v = 200\,{\rm km\,s^{-1}}$ (top panel) and $\Delta v = 200\,{\rm km\,s^{-1}}$ (middle panel). We also show a measure of bimodality to explore the increasing coverings around bright galaxies (bottom panel). See the text for explanations of various trends with redshift and brightness in terms of, e.g., anisotropy and environment.

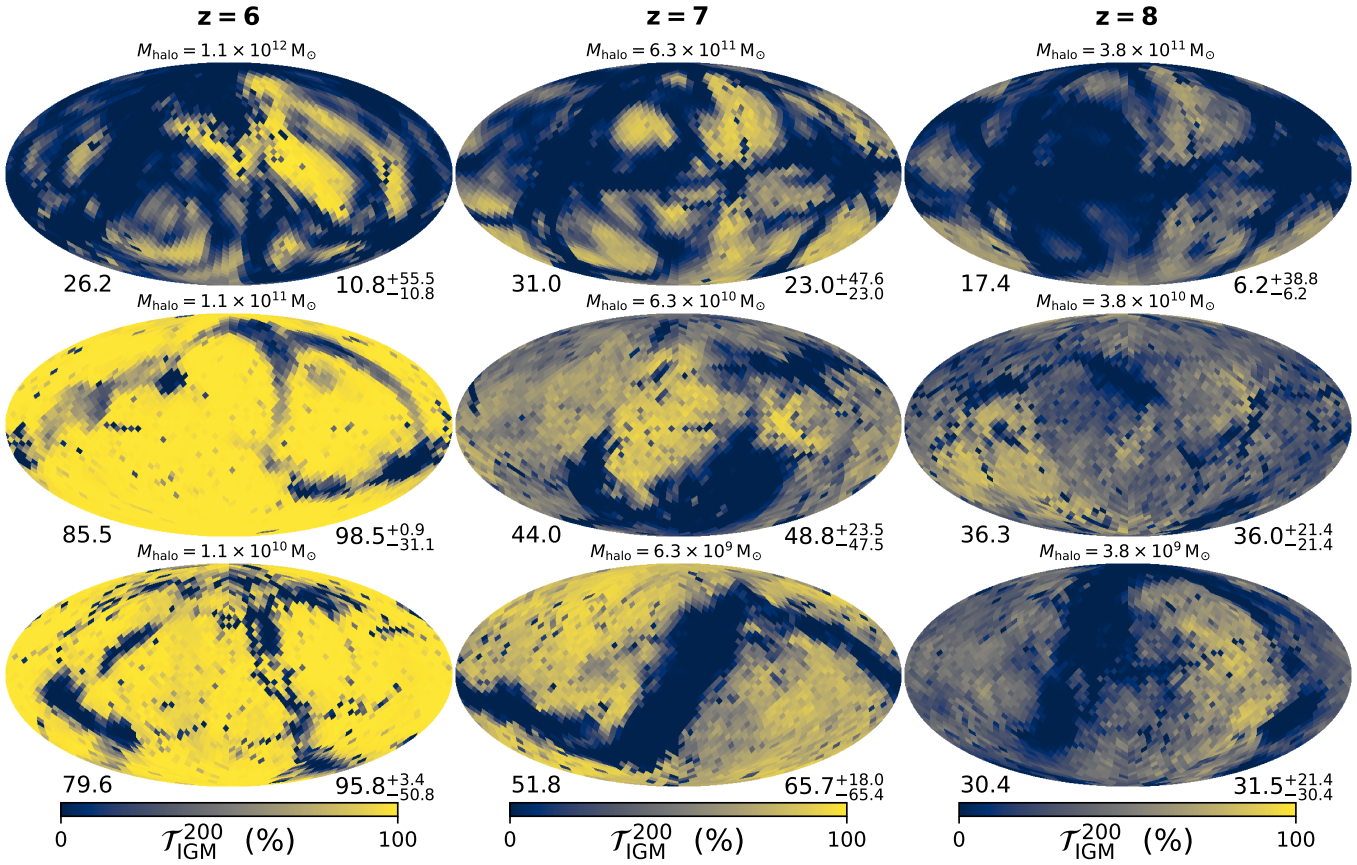

Figure 18. Example angular distributions of the IGM transmission at a velocity offset of $200\,\mathrm{km\,s^{-1}}$, $\mathcal{T}^{200}_{\mathrm{IGM}}$. We provide the mean and median statistics based on 3072 healpix directions of equal solid angle, included as values on the lower left and right of each map, respectively. We select the most massive haloes at $z = \{6, 7, 8\}$ as well as haloes with masses 10 and 100 times smaller. The wide variety of structures include smooth background fluctuations, continent-sized dimming and brightening, and interconnected node and filamentary features from a cosmic web of cold gas streams in projection. The most massive haloes have large covering fractions and dense knots of octopus-like absorption tracks.

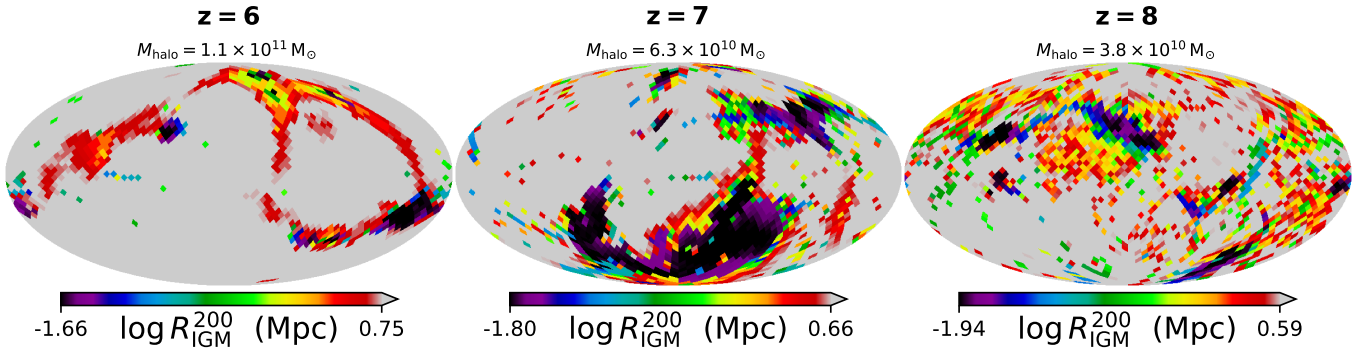

Figure 19. Angular distributions of the distance to the $\tau = 1$ scattering surface R^{200}_{IGM} at a velocity offset of $200\,\mathrm{km\,s^{-1}}$ for haloes corresponding to the middle row of Fig. 18. These images highlight that the obstruction morphology can be shaped by nearby gas but pushed over the threshold at cosmological scales ($\gtrsim 1$ Mpc). The leftmost galaxy at $z = 6$ is also shown in Fig. 2, providing a visual connection between the angular and spatial representations.

variety of structures including both easily understood and non-trivial dependence on redshift, halo mass, environment, and viewing angle. Broadly speaking the smooth backdrop in each image captures the inhomogeneous reionization process with large patches of dimming and brightening, likely due to moderate to large-scale bubble variations. On the other hand, nearby absorption features highlight shortcomings when separating ISM- and IGM-scale radiative transfer.

Beyond this there are relatively large and interconnected node and filamentary absorption features representing the cosmic web of cold gas streams in projection. It is clear that the environments around the most massive galaxies are significantly different than those of lower

masses. In fact, the strong gravitational potential and crowded local volumes give rise to dense knots of octopus-like absorption tracks. Such structures are at the heart of the elevated covering fractions discussed above, and become essentially transparent for red wing photons ($\Delta v \gtrsim 400\,\mathrm{km\,s^{-1}}$). We also observe punctuated extinction from small to moderate solid angle self-shielding clumps, corresponding to small self-shielded clumps and distant damped Lyman-alpha systems in the intervening IGM that persist for red wing photons too (see the discussion by Park et al. 2021). To quantify this further, in Fig. 19 we show the distances to the $\tau = 1$ scattering surface $R_{\rm IGM}^{200}$ at a velocity offset of $200\,\mathrm{km\,s^{-1}}$ for haloes corresponding to the middle row of Fig. 18. These images highlight that the obstruction morphology can be shaped by nearby neutral gas but pushed over the opacity threshold at cosmological scales ($\gtrsim 1$ Mpc).

A proper treatment incorporating radiative transfer effects down to ISM and CGM scales is expected to smooth out some of these features while enhancing others. For example, clumpy media has been shown to produce directional Lyα equivalent width boosting, which vanishes for spatially extended sources once the projected shadow sizes for emitting regions exceed the mean cloud separation distances (Gronke & Dijkstra 2014). Overall, Lyα resonant scattering tends to reduce the flux anisotropy compared to continuum radiation, but even directional dependence favouring blue or red photons can provide non-trivial variations when accounting for IGM transmission. Similar signatures are found in the viewing angle dependence of high-resolution cosmological Lyα radiative transfer simulations (Behrens et al. 2019; Smith et al. 2019; Kimock et al. 2021; Mitchell et al. 2021). The galaxy and IGM directional effects are certainly correlated to some degree, but it is standard to conceptually include them independently (Laursen et al. 2011). Ultimately, incorporating IGM transmission for individual LAEs is most accurately captured by extending resonant scattering calculations to well beyond the virial radius, especially at increasingly high redshifts. In fact, nearly saturated IGM transmission covering fractions, i.e. $P(\mathcal{T}_{\rm IGM} < 0.2) \approx 1$, are indicative that scattering back into the line of sight is increasingly important, implying that our results are more affected by not taking radiative transfer into account at $z = 10$ compared to $z = 6$. In Fig. 20 we come full circle with the same example $z = 6$ galaxy from Fig. 2 by showing the angular distributions of the red-to-blue flux ratio $F_{\rm red}/F_{\rm blue}$ and the fraction of observed flux taken as the product of the dust escape fraction $f_{\rm esc}$ and IGM transmission $\mathcal{T}_{\rm IGM}$ of the emergent spectra. As expected, there is a clear correlation in the structure of the spectral behaviour and observed flux as a result of connecting the ISM- and IGM-scale radiative transfer. For simplicity, we only show results for the wind model which produces more realistic enhanced red peaks than the fiducial model. For quantitative comparison the fraction of flux emerging on the red side of line centre is $F_{\rm red}/F_{\rm tot} \approx 57.4^{+3.9}_{-6.1}(11.3^{+7.5}_{-5.0})$, the dust escape fraction is $f_{\rm esc} \approx 46.8^{+5.3}_{-8.6}(58.1^{24.3}_{13.6})$, and the final observed fraction is $f_{\rm esc} \times \mathcal{T}_{\rm IGM} \approx 26.7^{+4.9}_{-7.2}(7.3^{+3.8}_{-3.4})$ for the wind (fiducial) models, respectively. Overall, this confirms that Lyα radiative transfer effects are important for the observability of high-z LAEs.

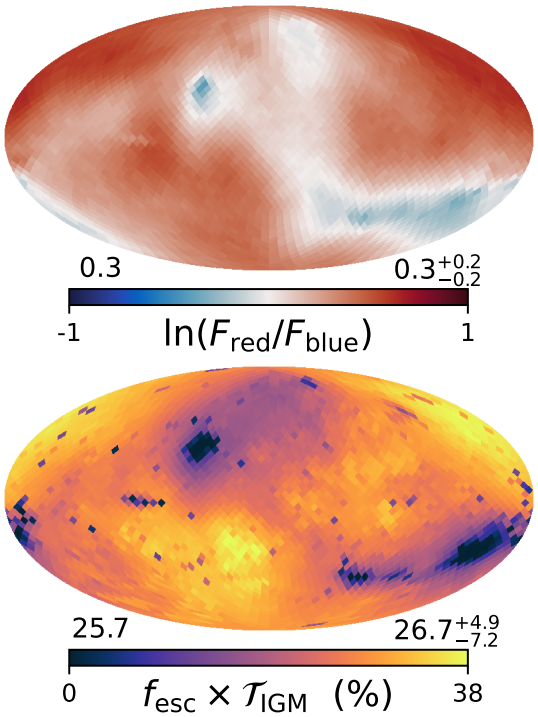

Figure 20. Angular distributions of the red-to-blue flux ratio $F_{\rm red}/F_{\rm blue}$ (top) and fraction of observed flux (bottom) taken as the product of the dust escape fraction $f_{\rm esc}$ and IGM transmission $\mathcal{T}_{\rm IGM}$ of the emergent spectra based on Monte Carlo Lyα radiative transfer calculations. This is the same $z = 6$ galaxy ($M_{\rm halo} \approx 10^{11}\,\mathrm{M_\odot}$) as in Figs. 2, 18, and 19 for the wind model, and demonstrates a clear connection between the spectral behaviour and observed flux connecting ISM- and IGM-scale radiative transfer.

4.8 Ionization statistics

In Fig. 21, we show several ionization properties for each central galaxy in our catalogue. To better connect the transmission statistics to the large-scale reionization morphology (beyond the example angular distributions), from top to bottom we plot the line-of-sight distance to the $\tau = 1$ scattering surface $R_{\rm IGM}^{400}$ at a velocity offset of $400\,\mathrm{km\,s^{-1}}$, the effective size of the local ionized bubble $R_{\rm H\,II}$, and the neutral fraction within this radius $\langle x_{\rm HI}\rangle(< R_{\rm H\,II})$, all as a function of the intrinsic UV magnitude M_{1500}. Specifically, $R_{\rm H\,II}$ is calculated as the nearest distance along each ray where the neutral fraction exceeds $x_{\rm H\,I} > 0.9$, i.e. the closest Lyman-limit system apart from gas within the virial radius. As before, the curves and shaded regions give the median and 1σ variation, which we emphasize includes 768 measurements per halo as incorporating the sightline variance is important for connecting back to observations. For the neutral fraction we show both volume- and mass-weighted averages represented by solid and dashed curves, respectively. Together $R_{\rm H\,II}$ and $\langle x_{\rm HI}\rangle$ reveal clear trends that the line-of-sight environments around brighter galaxies exhibit larger ionized zones and lower residual neutral fractions than fainter ones, in agreement with observed discoveries of very large ionized bubbles (e.g. Jung et al. 2020; Endsley & Stark 2022). These results also demonstrate that red damping-wing absorption distances $R_{\rm IGM}^{400}$ generally exceed the local bubble sizes by up to a factor of a few, although this is no longer the case near line centre where photons experience resonant absorption within ionized gas. This starts to be significant by $200\,\mathrm{km\,s^{-1}}$ around bright, massive haloes with strong infall motions. We also note that at lower redshifts $R_{\rm IGM}$ saturates such that the IGM is optically thin as bubbles become large enough for the Hubble flow to efficiently transmit red peaks to the observer. Thus, the visibility of a Lyα emitter is given mainly by the redshift according to the local and global progress of reionization in terms of bubble sizes and residual neutral fractions. Beyond this, the location of absorption can be strongly affected according to the covering fraction of nearby infalling or self-shielding systems.

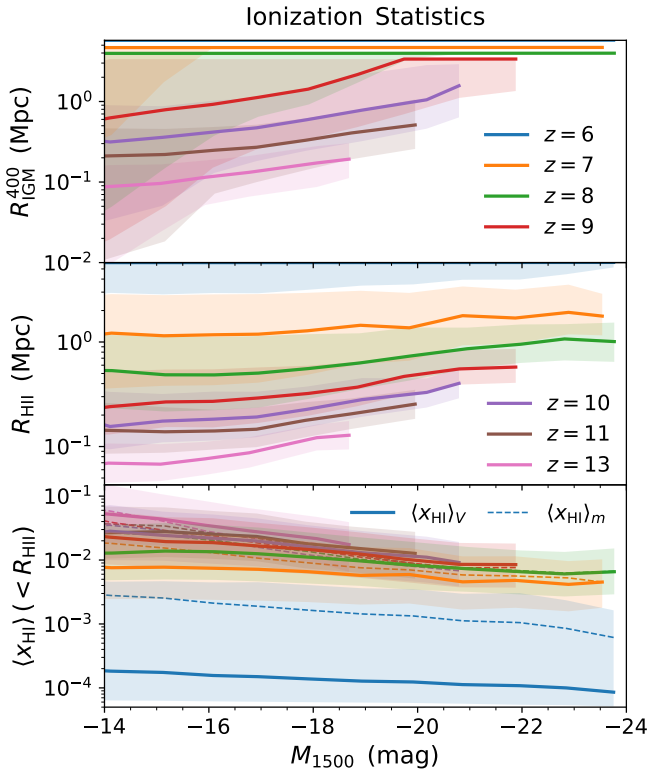

Figure 21. *Top:* Distance to the $\tau = 1$ scattering surface $R_{\rm IGM}^{400}$ (physical Mpc) at a velocity offset of 400 km s^{-1} as a function of UV magnitude M_{1500}. The curves and shaded regions give the median and 1σ variation. *Middle:* Size of the local ionized bubble $R_{\rm H\,II}$ defined as the line-of-sight distance where the neutral fraction exceeds $x_{\rm H\,I} > 0.9$, ignoring gas within the virial radius. *Bottom:* Average neutral fraction along each ray within the local ionized bubble $\langle x_{\rm HI}\rangle(< R_{\rm H\,II})$, with volume- and mass-weighted averages shown as solid and dashed curves, respectively. Overall, red damping-wing absorption distances generally exceed the local bubble sizes, and that brighter galaxies produce larger ionized bubbles and lower residual neutral fractions.

Table 2. *Top:* Lyα IGM transmission $\mathcal{T}_{\rm IGM}$ tabulated as functions of redshift and frequency, averaged over wavelength windows of 50 km s^{-1} to match observations. The summary statistics are calculated including all central galaxies with UV brightness $M_{1500} < -19$ and are given in per cent units with median and asymmetric 1σ confidence regions. *Bottom:* Lyα transmission covering fractions on the same grid defined as the fraction of sightlines for each galaxy with transmission below 20 per cent, i.e. $P(\mathcal{T}_{\rm IGM} < 0.2)$.

$\mathcal{T}_{\rm IGM}$	200 km s^{-1}	300 km s^{-1}	400 km s^{-1}	500 km s^{-1}
$z = 6$	$95.2^{+4.1}_{-66.6}$	$98.7^{+0.8}_{-13.0}$	$99.1^{+0.4}_{-8.6}$	$99.2^{+0.3}_{-7.0}$
$z = 7$	$61.5^{+22.4}_{-51.6}$	$74.1^{+14.1}_{-27.4}$	$77.2^{+12.1}_{-21.8}$	$79.2^{+10.8}_{-18.4}$
$z = 8$	$32.8^{+23.7}_{-30.2}$	$46.2^{+18.4}_{-23.8}$	$51.5^{+16.1}_{-20.1}$	$55.5^{+14.3}_{-17.4}$
$z = 9$	$15.0^{+16.4}_{-14.2}$	$25.5^{+14.6}_{-16.0}$	$31.2^{+13.3}_{-15.0}$	$36.0^{+12.0}_{-13.9}$
$z = 10$	$6.5^{+10.3}_{-6.0}$	$13.1^{+10.7}_{-8.8}$	$17.9^{+10.3}_{-9.6}$	$22.4^{+9.8}_{-9.8}$
$z = 11$	$2.4^{+5.6}_{-2.3}$	$5.8^{+6.9}_{-4.6}$	$9.4^{+7.4}_{-5.9}$	$13.1^{+7.5}_{-6.8}$
$P(\mathcal{T}_{\rm IGM} < 0.2)$	200 km s^{-1}	300 km s^{-1}	400 km s^{-1}	500 km s^{-1}
$z = 6$	$10.1^{+14.8}_{-4.5}$	$5.1^{+2.5}_{-1.7}$	$3.5^{+1.6}_{-1.2}$	$2.6^{+1.2}_{-0.9}$
$z = 7$	$17.0^{+12.7}_{-6.1}$	$8.1^{+2.6}_{-2.2}$	$5.6^{+2.0}_{-1.6}$	$4.0^{+1.4}_{-1.2}$
$z = 8$	$30.9^{+11.7}_{-8.2}$	$14.1^{+4.0}_{-3.5}$	$9.2^{+2.7}_{-2.3}$	$6.5^{+2.1}_{-1.6}$
$z = 9$	$63.4^{+9.7}_{-11.9}$	$34.4^{+14.2}_{-11.2}$	$21.6^{+8.3}_{-7.3}$	$12.8^{+4.3}_{-4.0}$
$z = 10$	$91.5^{+5.5}_{-10.4}$	$77.9^{+10.0}_{-17.0}$	$60.7^{+11.4}_{-16.6}$	$39.3^{+13.7}_{-11.2}$
$z = 11$	$99.4^{+0.5}_{-1.7}$	$97.6^{+1.5}_{-3.7}$	$94.2^{+1.8}_{-8.2}$	$83.4^{+5.1}_{-6.3}$

5 SUMMARY AND DISCUSSION

In this paper, we have constructed catalogues for Lyα emission and IGM transmission for the flagship THESAN simulation. In a forthcoming study, we will also present a detailed comparison of the Lyα properties from the remaining suite of lower resolution simulations (presented in Paper I), including reionization histories affected by halo mass-dependent escape fractions, alternative dark matter, and numerical convergence. In a further study, we will also make predictions for Lyα intensity mapping and other diffuse cosmological radiative transfer statistics. This will utilize the high time cadence on-the-fly redshift-space renderings of Lyα properties to conveniently map out the ensemble radiation field throughout the EoR. We also leave empirical modelling of LAEs to a future study, e.g. constraining idealized galaxy source parameters such as spectral profiles and Lyα escape fractions (as in Jensen et al. 2013; Weinberger et al. 2019; Gangolli et al. 2021). None the less, our initial explorations presented herein showcase the strengths and weaknesses of pursuing LAE science from state-of-the-art large-volume cosmological reionization simulations. In fact, by augmenting the IllustrisTNG galaxy formation model with fully coupled radiation hydrodynamics and dust physics, THESAN already incorporates most of the essential ingredients for realistic and representative simulation-based Lyα surveys. Thus, we welcome collaborations on related science topics or general utilization of Lyα-centric catalogues following the upcoming public data release.

One of the main drawbacks is the sub-resolution treatment of the ISM as a two-phase gas (Springel & Hernquist 2003). The predictive power of Lyα radiative transfer calculations based on simulations using the SH03 model without significant modification, and especially in combination with a low escape fraction of ionizing photons from the birth cloud, is reduced. However, we are pursuing a variety of approaches to more closely relate Lyα emission, absorption, and scattering from ISM to IGM scales. Ultimately, the THESAN project will include high-resolution zoom-in resimulations of a wide range of galaxies including a multiphase ISM framework (Marinacci et al. 2019; Kannan et al. 2020) and self-consistent meso-scale reionization environment inherited directly from the flagship simulation. Beyond this, we emphasize that statistical IGM transmission studies also suffer from stitching and resolution effects when attempting to connect to Lyα emission sources (e.g. see Gronke et al. 2021). However, while the emergent spectra are certainly tied to the local ($\lesssim 1$ Mpc) radiation transport via environmental and directional biases, the more distant absorption features may be understood as a combination of redshift evolution, UV brightness, and stochastic covering fractions. Different sightlines from the same galaxy may correspond to underdense regions or self-shielded cosmological filaments, with the exponential sensitivity ($\mathcal{T}_{\rm IGM} = e^{-\tau_{\rm IGM}}$) acting to increase the variation and bimodality in IGM transmissivity. As a result, LAE surveys will require large sample sizes to get a handle on the statistics (Park et al. 2021).

The THESAN simulations confirm that large ionized bubbles form around the brightest galaxies at any epoch. Such accelerated reionization regions provide an advantage for Lyα transmission of red peaks, thereby boosting the visibility of LAEs around UV bright galaxies (Mason et al. 2018b). However, it is also clear that infall motion plays a significant role in shaping the IGM transmissivity

(Santos 2004; Dijkstra et al. 2007; Sadoun et al. 2017; Park et al. 2021). For example, the truncation frequency of the transmission curve is typically set by the maximum infall velocity along the line of sight, which roughly corresponds to the circular velocity of the galaxy $V_c = \sqrt{GM_{\rm halo}/R_{\rm vir}}$. Of course, ISM and galaxy scale bulk motions set the intrinsic line profile prior to resonant scattering, and the emergent Lyα spectra from the galaxy is modulated by the 3D geometry of the neutral hydrogen and dust distributions (Verhamme et al. 2018; Smith et al. 2019). Thus, while infall motions and bubble sizes are crucial for IGM transmission, it is necessary to combine Lyα radiative transfer modelling with IGM reprocessing for realistic luminosity functions (e.g. as in Garel et al. 2021).

Of course, we do not expect a clean separation of ISM and IGM scale Lyα radiative transfer effects into the EoR as these are increasingly coupled for higher redshift galaxies. For example, it is well understood that resonant scattering leads to ubiquitous extended Lyα haloes (e.g. Wisotzki et al. 2018), which only become more exaggerated as the global neutral hydrogen density increases (Loeb & Rybicki 1999). Still, Lyα signatures are shaped from small to large scales with significant correlated physics along the way. For example, feedback can drive outflows through low column density channels that induce redshifting and boost the transmission through ionized windows between clusters of bright galaxies, or vice versa. On the other hand, serendipitous and deleterious circumstances add significant scatter, but sufficient statistics and complementary probes such as non-resonant lines should help distinguish between physics and random processes. Such richness and complexity calls for rigorous theoretical efforts to accurately interpret results from current and upcoming LAE surveys. With this outlook, the THESAN project offers a unique framework for studying high-redshift galaxies and the impact of cosmic reionization.

ACKNOWLEDGEMENTS

We thank the referee for constructive comments and suggestions which have improved the quality of this work. We thank Hyunbae Park, Ewald Puchwein, Sandro Tacchella, Hui Li, and Paul Torrey for insightful discussions related to this work. AS acknowledges support for Program number *HST*-HF2-51421.001-A provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, incorporated, under NASA contract NAS5-26555. MV acknowledges support through NASA ATP grants 16-ATP16-0167, 19-ATP19-0019, 19-ATP19-0020, 19-ATP19-0167, and NSF grants AST-1814053, AST-1814259, AST-1909831 and AST-2007355. The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (`www.gauss-centre.eu`) for funding this project by providing computing time on the GCS Supercomputer SuperMUC-NG at Leibniz Supercomputing Centre (`www.lrz.de`). Additional computing resources were provided by the Extreme Science and Engineering Discovery Environment (XSEDE), at Stampede2 through allocation TG-AST200007 and by the NASA High-End Computing (HEC) Program through the NASA Advanced Supercomputing (NAS) Division at Ames Research Center. We are thankful to the community developing and maintaining software packages extensively used in our work, namely: `matplotlib` (Hunter 2007), `numpy` (Walt et al. 2011) and `scipy` (Jones et al. 2001).

DATA AVAILABILITY

All simulation data, including snapshots, group and subhalo catalogues (with Lyα-centric data), merger trees, and high time cadence Cartesian outputs will be made publicly available in the near future. Data will be distributed via `www.thesan-project.com`. Before the public data release, data underlying this article will be shared on reasonable request to the corresponding author(s).

REFERENCES

Angulo R. E., Pontzen A., 2016, MNRAS, 462, L1
Atek H., Richard J., Kneib J.-P., Schaerer D., 2018, MNRAS, 479, 5184
Barkana R., Loeb A., 2001, Phys. Rep., 349, 125
Barnes J., Hut P., 1986, Nature, 324, 446
Behrens C., Braun H., 2014, A&A, 572, A74
Behrens C., Dijkstra M., Niemeyer J. C., 2014, A&A, 563, A77
Behrens C., Byrohl C., Saito S., Niemeyer J. C., 2018, A&A, 614, A31
Behrens C., Pallottini A., Ferrara A., Gallerani S., Vallini L., 2019, MNRAS, 486, 2197
Bouwens R. J., et al., 2015, ApJ, 803, 34
Bromm V., Yoshida N., 2011, ARA&A, 49, 373
Byrohl C., Gronke M., 2020, A&A, 642, L16
Byrohl C., et al., 2021, MNRAS,
Camps P., Behrens C., Baes M., Kapoor A. U., Grand R., 2021, ApJ, 916, 39
Cantalupo S., Porciani C., Lilly S. J., 2008, ApJ, 672, 48
Chabrier G., 2003, ApJ, 586, L133
Ciardi B., Ferrara A., Governato F., Jenkins A., 2000, MNRAS, 314, 611
Davies F. B., et al., 2018, ApJ, 864, 142
Dayal P., Ferrara A., 2012, MNRAS, 421, 2568
Dayal P., Ferrara A., 2018, Phys. Rep., 780, 1
Dijkstra M., 2014, Publ. Astron. Soc. Australia, 31, e040
Dijkstra M., 2019, Saas-Fee Advanced Course, 46, 1
Dijkstra M., Haiman Z., Spaans M., 2006, ApJ, 649, 14
Dijkstra M., Lidz A., Wyithe J. S. B., 2007, MNRAS, 377, 1175
Eldridge J. J., Stanway E. R., Xiao L., McClelland L. A. S., Taylor G., Ng M., Greis S. M. L., Bray J. C., 2017, Publ. Astron. Soc. Australia, 34, e058
Endsley R., Stark D. P., 2022, MNRAS, 511, 6042
Endsley R., Stark D. P., Charlot S., Chevallard J., Robertson B., Bouwens R. J., Stefanon M., 2021, MNRAS, 502, 6044
Finkelstein S. L., et al., 2015, ApJ, 810, 71
Gallego S. G., et al., 2018, MNRAS, 475, 3854
Gangolli N., D'Aloisio A., Nasir F., Zheng Z., 2021, MNRAS, 501, 5294
Garaldi E., Kannan R., Smith A., Springel V., Pakmor R., Vogelsberger M., Hernquist L., 2022, MNRAS,
Garel T., Blaizot J., Rosdahl J., Michel-Dansac L., Haehnelt M. G., Katz H., Kimm T., Verhamme A., 2021, MNRAS, 504, 1902
Genel S., et al., 2014, MNRAS, 445, 175
Gnedin N. Y., 2014, ApJ, 793, 29
Gnedin N. Y., 2016, ApJ, 833, 66
Gnedin N. Y., Kaurov A. A., 2014, ApJ, 793, 30
Godunov S. K., 1959, Mat. Sb., Nov. Ser., 47, 271
Gronke M., Dijkstra M., 2014, MNRAS, 444, 1095
Gronke M., Dijkstra M., McCourt M., Oh S. P., 2017, A&A, 607, A71
Gronke M., et al., 2021, MNRAS,
Gunn J. E., Peterson B. A., 1965, ApJ, 142, 1633
Hansen M., Oh S. P., 2006, MNRAS, 367, 979
Harrington J. P., 1973, MNRAS, 162, 43
Hashimoto T., et al., 2017, A&A, 608, A10
Hoag A., et al., 2019, ApJ, 878, 12
Hui L., Gnedin N. Y., 1997, MNRAS, 292, 27
Hunter J. D., 2007, Computing In Science & Engineering, 9, 90
Iliev I. T., Mellema G., Pen U. L., Merz H., Shapiro P. R., Alvarez M. A., 2006, MNRAS, 369, 1625
Iliev I. T., Mellema G., Ahn K., Shapiro P. R., Mao Y., Pen U.-L., 2014, MNRAS, 439, 725
Iyer K. G., et al., 2020, MNRAS, 498, 430

Jensen H., Laursen P., Mellema G., Iliev I. T., Sommer-Larsen J., Shapiro P. R., 2013, MNRAS, 428, 1366
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/
Jung I., et al., 2020, ApJ, 904, 144
Kakiichi K., Dijkstra M., Ciardi B., Graziani L., 2016, MNRAS, 463, 4019
Kannan R., Vogelsberger M., Marinacci F., McKinnon R., Pakmor R., Springel V., 2019, MNRAS, 485, 117
Kannan R., Marinacci F., Vogelsberger M., Sales L. V., Torrey P., Springel V., Hernquist L., 2020, MNRAS, 499, 5732
Kannan R., Garaldi E., Smith A., Pakmor R., Springel V., Vogelsberger M., Hernquist L., 2022, MNRAS, 511, 4005
Kimm T., Blaizot J., Garel T., Michel-Dansac L., Katz H., Rosdahl J., Verhamme A., Haehnelt M., 2019, MNRAS, 486, 2215
Kimock B., et al., 2021, ApJ, 909, 119
Kulkarni G., Keating L. C., Haehnelt M. G., Bosman S. E. I., Puchwein E., Chardin J., Aubert D., 2019, MNRAS, 485, L24
Lao B.-X., Smith A., 2020, MNRAS, 497, 3925
Laursen P., Sommer-Larsen J., Razoumov A. O., 2011, ApJ, 728, 52
Laursen P., Sommer-Larsen J., Milvang-Jensen B., Fynbo J. P. U., Razoumov A. O., 2019, A&A, 627, A84
Leclercq F., et al., 2017, A&A, 608, A8
Lee H.-W., 2013, ApJ, 772, 123
Levermore C. D., 1984, J. Quant. Spectrosc. Radiative Transfer, 31, 149
Li Y., Gu M. F., Yajima H., Zhu Q., Maji M., 2020, MNRAS,
Li Z., Steidel C. C., Gronke M., Chen Y., 2021, MNRAS, 502, 2389
Lidz A., Chang T.-C., Mas-Ribas L., Sun G., 2021, ApJ, 917, 58
Livermore R. C., Finkelstein S. L., Lotz J. M., 2017, ApJ, 835, 113
Loeb A., Furlanetto S. R., 2013, The First Galaxies in the Universe. Princeton Univ. Press, Princeton, NJ
Loeb A., Rybicki G. B., 1999, ApJ, 524, 527
Ludlow A. D., Schaye J., Schaller M., Richings J., 2019, MNRAS, 488, L123
Lusso E., Worseck G., Hennawi J. F., Prochaska J. X., Vignali C., Stern J., O'Meara J. M., 2015, MNRAS, 449, 4204
Madau P., Rees M. J., 2000, ApJ, 542, L69
Malhotra S., Rhoads J. E., 2004, ApJ, 617, L5
Marinacci F., et al., 2018, MNRAS, 480, 5113
Marinacci F., Sales L. V., Vogelsberger M., Torrey P., Springel V., 2019, MNRAS, 489, 4233
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
Mason C. A., et al., 2019, MNRAS, 485, 3947
McKinnon R., Torrey P., Vogelsberger M., Hayward C. C., Marinacci F., 2017, MNRAS, 468, 1505
McQuinn M., Hernquist L., Zaldarriaga M., Dutta S., 2007, MNRAS, 381, 75
Miralda-Escudé J., 1998, ApJ, 501, 15
Mitchell P. D., Blaizot J., Cadiou C., Dubois Y., Garel T., Rosdahl J., 2021, MNRAS, 501, 5757
Naidu R. P., Tacchella S., Mason C. A., Bose S., Oesch P. A., Conroy C., 2020, ApJ, 892, 109
Naiman J. P., et al., 2018, MNRAS, 477, 1206
Nelson D., et al., 2018, MNRAS, 475, 624
Neufeld D. A., 1990, ApJ, 350, 216
Oñorbe J., Hennawi J. F., Lukić Z., 2017, ApJ, 837, 106
Ocvirk P., et al., 2016, MNRAS, 463, 1462
Ocvirk P., Aubert D., Chardin J., Deparis N., Lewis J., 2019, A&A, 626, A77
Ouchi M., et al., 2018, PASJ, 70, S13
Ouchi M., Ono Y., Shibuya T., 2020, ARA&A, 58, 617
Pakmor R., Springel V., Bauer A., Mocz P., Munoz D. J., Ohlmann S. T., Schaal K., Zhu C., 2016, MNRAS, 455, 1134
Park H., et al., 2021, arXiv e-prints, p. arXiv:2105.10770
Pawlik A. H., Rahmati A., Schaye J., Jeon M., Dalla Vecchia C., 2017, MNRAS, 466, 960
Pillepich A., et al., 2018a, MNRAS, 473, 4077
Pillepich A., et al., 2018b, MNRAS, 475, 648
Planck Collaboration et al., 2016, A&A, 594, A13
Planck Collaboration et al., 2020, A&A, 641, A6
Rahmati A., Schaye J., 2018, MNRAS, 478, 5123
Raiter A., Schaerer D., Fosbury R. A. E., 2010, A&A, 523, A64
Rosdahl J., et al., 2018, MNRAS, 479, 994
Sadoun R., Zheng Z., Miralda-Escudé J., 2017, ApJ, 839, 44
Santos M. R., 2004, MNRAS, 349, 1137
Schenker M. A., Ellis R. S., Konidaris N. P., Stark D. P., 2014, ApJ, 795, 20
Scholz T. T., Walters H. R. J., 1991, ApJ, 380, 302
Shen X., et al., 2020, MNRAS, 495, 4747
Shen X., Vogelsberger M., Nelson D., Tacchella S., Hernquist L., Springel V., Marinacci F., Torrey P., 2022, MNRAS, 510, 5560
Simcoe R. A., Sullivan P. W., Cooksey K. L., Kao M. M., Matejek M. S., Burgasser A. J., 2012, Nature, 492, 79
Smith A., Safranek-Shrader C., Bromm V., Milosavljević M., 2015, MNRAS, 449, 4336
Smith A., Bromm V., Loeb A., 2017a, MNRAS, 464, 2963
Smith A., Becerra F., Bromm V., Hernquist L., 2017b, MNRAS, 472, 205
Smith A., Ma X., Bromm V., Finkelstein S. L., Hopkins P. F., Faucher-Giguère C.-A., Kereš D., 2019, MNRAS, 484, 39
Smith A., et al., 2021, arXiv e-prints, p. arXiv:2111.13721
Song H., Seon K.-I., Hwang H. S., 2020, ApJ, 901, 41
Springel V., 2010, MNRAS, 401, 791
Springel V., Hernquist L., 2003, MNRAS, 339, 289
Springel V., et al., 2018, MNRAS, 475, 676
Springel V., Pakmor R., Zier O., Reinecke M., 2021, MNRAS, 506, 2871
Stark D. P., Ellis R. S., Ouchi M., 2011, ApJ, 728, L2
Taylor A. J., Barger A. J., Cowie L. L., Hu E. M., Songaila A., 2020, ApJ, 895, 132
Taylor A. J., Cowie L. L., Barger A. J., Hu E. M., Songaila A., 2021, ApJ, 914, 79
Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, ApJ, 688, 709
Verhamme A., Dubois Y., Blaizot J., Garel T., Bacon R., Devriendt J., Guiderdoni B., Slyz A., 2012, A&A, 546, A111
Verhamme A., et al., 2018, MNRAS, 478, L60
Visbal E., McQuinn M., 2018, ApJ, 863, L6
Vogelsberger M., et al., 2014a, MNRAS, 444, 1518
Vogelsberger M., et al., 2014b, Nature, 509, 177
Vogelsberger M., Marinacci F., Torrey P., Puchwein E., 2020a, Nature Reviews Physics, 2, 42
Vogelsberger M., et al., 2020b, MNRAS, 492, 5167
Walt S. v. d., Colbert S. C., Varoquaux G., 2011, Computing in Science & Engineering, 13, 22
Weinberger R., et al., 2017, MNRAS, 465, 3291
Weinberger L. H., Haehnelt M. G., Kulkarni G., 2019, MNRAS, 485, 1350
Weinberger R., Springel V., Pakmor R., 2020, ApJS, 248, 32
Weingartner J. C., Draine B. T., 2001, ApJ, 548, 296
Wise J. H., 2019, Contemporary Physics, 60, 145
Wisotzki L., et al., 2018, Nature, 562, 229
Witstok J., Puchwein E., Kulkarni G., Smit R., Haehnelt M. G., 2021, A&A, 650, A98
Yang H., Malhotra S., Gronke M., Rhoads J. E., Dijkstra M., Jaskot A., Zheng Z., Wang J., 2016, ApJ, 820, 130
Yang Y.-L., et al., 2020, ApJ, 891, 61
van Leer B., 1979, Journal of Computational Physics, 32, 101
van de Voort F., Springel V., Mandelker N., van den Bosch F. C., Pakmor R., 2019, MNRAS, 482, L85

APPENDIX A: CONTINUUM FLUX EXTRAPOLATION

We briefly explore the difference between the true and extrapolated Lyα continuum flux levels. This has important consequences when estimating Lyα equivalent widths from both simulated and observed

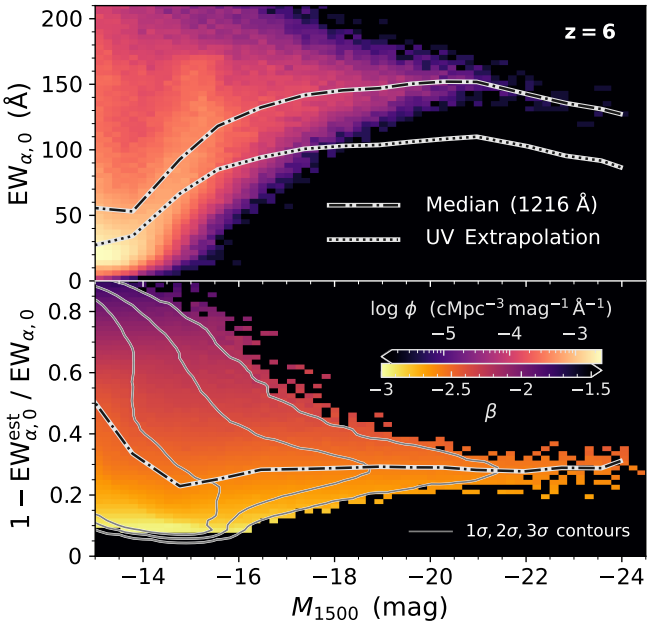

Figure A1. *Upper panel:* Rest-frame Lyα equivalent widths $\mathrm{EW}_{\alpha,0}$ as a function of UV magnitude M_{1500} for each halo at $z = 6$. The colour axis gives the halo number density while the curves show the median for the true and extrapolated values. *Lower panel:* Relative difference in the estimated equivalent widths with the colour denoting the UV slopes β. Estimates based on extrapolations of the intrinsic UV slope systematically underpredict intrinsic Lyα equivalent widths by approximately 30 per cent.

data. Specifically, in Fig. A1 we show rest-frame Lyα equivalent widths $\mathrm{EW}_{\alpha,0}$ as a function of UV magnitude M_{1500} for each halo at $z = 6$. We find that estimates based on extrapolations of the intrinsic UV slope systematically underpredict intrinsic Lyα equivalent widths by approximately 30 per cent. We expect a similar degree of uncertainty to remain or be enhanced after dust scattering and absorption. Thus, while a proper radiative transfer treatment is necessary to assess this fully, we caution that extrapolations in general likely incur biases for Lyα equivalent width measurements and predictions.

APPENDIX B: INITIAL INTEGRATION RADIUS

As a start of integration we use a fixed value of 1 R_{vir} (i.e. R_{200} of the entire group). Beyond this we assume that Lyα scatterings back into the aperture line-of-sight are negligible. While this choice is both physically motivated and informed by previous studies (e.g. Laursen et al. 2011; Byrohl & Gronke 2020), it is important to evaluate the impact for the present simulation and redshift range. Therefore, in Fig. B1 we show the difference in IGM transmission $\Delta\mathcal{T}_{\mathrm{IGM}}$ between $2\,R_{\mathrm{vir}}$ and $1\,R_{\mathrm{vir}}$, which demonstrates that changes in the median (solid) and mean (dashed) curves are quite small. The peaks of the differences are around $\Delta v \approx 100\,\mathrm{km\,s^{-1}}$ with plateaus on the red side of line centre, while the blue side is essentially unchanged due to already being highly suppressed. Interestingly, the mean variations ($\Delta\mathcal{T}_{\mathrm{IGM}} \sim 1\%$) are higher than the median ones ($\Delta\mathcal{T}_{\mathrm{IGM}} \sim 0.01\%$) due to outlier sightlines with significant localized absorption. In the extreme case of maximally bimodal distributions the mean variation can be viewed as the fraction of sightlines that are impacted by the starting radius value, especially in the context of describing IGM transmission with a covering fraction model that increases sightline

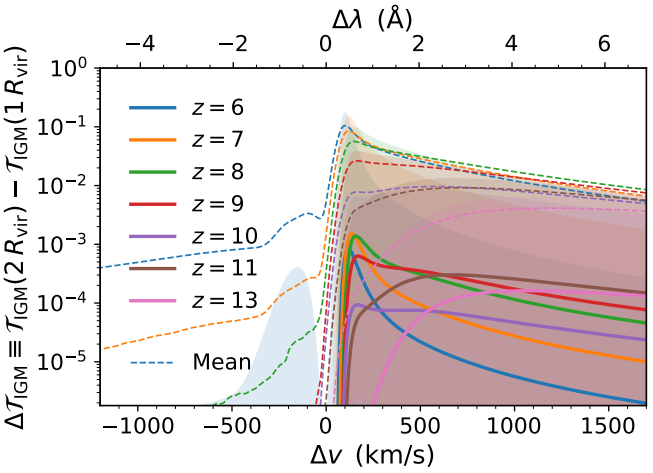

Figure B1. Difference in IGM transmission $\Delta\mathcal{T}_{\mathrm{IGM}}$ between 1 R_{vir} and 2 R_{vir} as a function of velocity offset Δv and rest-frame wavelength offset $\Delta\lambda$ around the Lyα line at each redshift. The solid (dashed) curves show the catalogue median (mean) statistics and shaded regions give the 1σ confidence levels. Overall, the choice of starting integration radius mainly affects results on the red side of line centre as blue photons are already highly suppressed. The mean values ($\Delta\mathcal{T}_{\mathrm{IGM}} \sim 1\%$) are higher than the median ones ($\Delta\mathcal{T}_{\mathrm{IGM}} \sim 0.01\%$) due to outliers that experience significant local absorption.

variability as reionization progresses. We expect such variations to be mirrored in end-to-end radiative transfer calculations, so our main conclusions are largely unaffected by our choice of initial integration radius.

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