

# The physics of Lyman $\alpha$ escape from high-redshift galaxies

Aaron Smith <sup>1,2</sup>★†, Xiangcheng Ma <sup>3,4</sup>, Volker Bromm,<sup>2</sup> Steven L. Finkelstein,<sup>2</sup>  
Philip F. Hopkins <sup>4</sup>, Claude-André Faucher-Giguère<sup>5</sup> and Dušan Kereš<sup>6</sup>

<sup>1</sup>Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

<sup>2</sup>Department of Astronomy, The University of Texas at Austin, Austin, TX 78712, USA

<sup>3</sup>Department of Astronomy and Theoretical Astrophysics Center, University of California Berkeley, Berkeley, CA 94720, USA

<sup>4</sup>TAPIR, MC 350-17, California Institute of Technology, Pasadena, CA 91125, USA

<sup>5</sup>Department of Physics and Astronomy and CIERA, Northwestern University, Evanston, IL 60647, USA

<sup>6</sup>Department of Physics, University of California at San Diego, La Jolla, CA 92093, USA

Accepted 2018 December 19. Received 2018 December 11; in original form 2018 October 18

## ABSTRACT

Lyman  $\alpha$  ( $\text{Ly}\alpha$ ) photons from ionizing sources and cooling radiation undergo a complex resonant scattering process that generates unique spectral signatures in high-redshift galaxies. We present a detailed  $\text{Ly}\alpha$  radiative transfer study of a cosmological zoom-in simulation from the Feedback In Realistic Environments (FIRE) project. We focus on the time, spatial, and angular properties of the  $\text{Ly}\alpha$  emission over a redshift range of  $z = 5\text{--}7$ , after escaping the galaxy and being transmitted through the intergalactic medium (IGM). Over this epoch, our target galaxy has an average stellar mass of  $M_\star \approx 5 \times 10^8 M_\odot$ . We find that many of the interesting features of the  $\text{Ly}\alpha$  line can be understood in terms of the galaxy's star formation history. The time variability, spatial morphology, and anisotropy of  $\text{Ly}\alpha$  properties are consistent with current observations. For example, the rest-frame equivalent width has an  $\text{EW}_{\text{Ly}\alpha,0} > 20 \text{ \AA}$  duty cycle of 62 per cent with a non-negligible number of sightlines with  $> 100 \text{ \AA}$ , associated with outflowing regions of a starburst with greater coincident UV continuum absorption, as these conditions generate redder, narrower (or single-peaked) line profiles. The lowest equivalent widths correspond to cosmological filaments, which have little impact on UV continuum photons but efficiently trap  $\text{Ly}\alpha$  and produce bluer, broader lines with less transmission through the IGM. We also show that in dense self-shielding, low-metallicity filaments and satellites,  $\text{Ly}\alpha$  radiation pressure can be dynamically important. Finally, despite a significant reduction in surface brightness with increasing redshift,  $\text{Ly}\alpha$  detections and spectroscopy of high- $z$  galaxies with the upcoming *James Webb Space Telescope* is feasible.

**Key words:** radiative transfer – galaxies: formation – galaxies: high-redshift.

## 1 INTRODUCTION

Lyman  $\alpha$  ( $\text{Ly}\alpha$ ) emission from stellar populations and active galactic nuclei is a powerful diagnostic of high- $z$  objects, enabling the effective identification and spectroscopic confirmation of these sources (Bromm & Yoshida 2011; Finkelstein 2016). However, due to the complex physics of the  $\text{Ly}\alpha$  radiative transfer process, it has proven difficult to interpret the results from theoretical models and observational data (Dijkstra 2014). For example, the fraction of  $\text{Ly}\alpha$  emitters (LAEs) among the galaxy population has been found to first increase with redshift, and then drop dramatically at  $z > 6$  (Stark, Ellis & Ouchi 2011; Schenker et al. 2014).

Intriguingly, there is evidence that the redshift evolution is less dramatic than previously thought (Pentericci et al. 2018), and the main difference between LAEs and non-LAEs is that the latter are significantly dustier (De Barros et al. 2017). The  $\text{Ly}\alpha$  escape fractions from low mass  $z < 6$  galaxies can therefore be quite high, especially given the tendency of supernova explosions and radiation pressure to drive galactic outflows, generating a redshifted  $\text{Ly}\alpha$  line (Stark et al. 2017). At lower redshifts, where more data is available, LAEs with high equivalent widths exhibit bluer UV slopes and are younger, less massive, and have lower star formation rates (e.g. Hayes 2015; Trainor et al. 2016). Given the increasing quantity and quality of high- $z$  data, we anticipate additional progress in using LAE visibility and quasar absorption spectra as probes of reionization (Furlanetto & Pritchard 2006; Dayal, Maselli & Ferrara 2011; Jensen et al. 2013; Mason et al. 2018),  $\text{Ly}\alpha$  haloes to study the circumgalactic medium (CGM; Steidel et al. 2011;

\* E-mail: arsmith@mit.edu

† Einstein Fellow.

Hayes et al. 2013; Momose et al. 2016; Leclercq et al. 2017; Erb, Steidel & Chen 2018), and spatially/spectrally resolved properties to better understand the formation and evolution of galaxy populations (Dayal & Ferrara 2018).

On the theoretical side, accurate Monte Carlo radiative transfer (MCRT) simulations allow us to study the resonant scattering of Ly $\alpha$  photons within the interstellar medium (ISM) of galaxies and their subsequent transmission through the intergalactic medium (IGM). Substantial progress has been made in understanding Ly $\alpha$  escape in homogeneous media (Harrington 1973; Neufeld 1990), a clumpy multiphase ISM (Neufeld 1991; Hansen & Oh 2006; Dijkstra & Kramer 2012; Laursen, Duval & Östlin 2013; Duval et al. 2014; Gronke & Dijkstra 2014; Gronke et al. 2016), expanding shell environments (Dijkstra, Haiman & Spaans 2006; Verhamme, Schaerer & Maselli 2006; Gronke, Bull & Dijkstra 2015; Hashimoto et al. 2015), and idealized anisotropic environments (Behrens, Dijkstra & Niemeyer 2014; Zheng & Wallace 2014; Smith et al. 2015). These models have been used to interpret observed spectra with varying success (e.g. Verhamme et al. 2008; Orlitova et al. 2018). On the other hand, MCRT has been applied to galaxy simulations in post-processing from cosmological initial conditions (Tasitsiomi 2006; Laursen, Razoumov & Sommer-Larsen 2009a; Faucher-Giguère et al. 2010; Zheng et al. 2010; Barnes et al. 2011; Yajima et al. 2012; Smith et al. 2015; Trebitsch et al. 2016) and isolated disc galaxy configurations (Verhamme et al. 2012; Behrens & Braun 2014). However, the analysis and interpretation of MCRT results from hydrodynamical simulations is further complicated by the inability to isolate physical effects by adjusting parameters one at a time. Furthermore, successful models will eventually need to simultaneously and self-consistently resolve the sub-structure of the ISM and the large-scale structure of the increasingly neutral IGM in the high- $z$  Universe. Still, the field is maturing as multiple groups are incorporating Ly $\alpha$  MCRT into state-of-the-art adaptive resolution simulation pipelines, fostering high-level theoretical and observational synergies.

Ideally, by matching Ly $\alpha$  simulations and observations, we can extract additional missing information about high- $z$  galaxies. This requires more robust models for individual LAEs and the emergent statistical properties of galaxy populations. The main goal of this paper is to explore the physics of Ly $\alpha$  escape from individual high- $z$  galaxies as comprehensively as possible. Such realistic radiative transfer modelling can provide more accurate predictions for future observations and more meaningful constraints from currently available data. For example, Laursen et al. (2018) have recently demonstrated that the UltraVISTA survey, with a narrowband filter tuned to detect Ly $\alpha$  emission at  $z = 8.8$ , only has a  $\sim 10$  per cent probability of detecting any LAEs at all once the survey has finished, which is a significantly smaller success rate than earlier predictions. This is attributed to more realistic (i.e. simulation-based) models and a lower than expected survey depth. Additionally, understanding Ly $\alpha$  escape is highly relevant for 21-cm cosmology (Pritchard & Loeb 2012). In fact, the recent tentative detection by the Experiment to Detect the Global Epoch of Reionization Signature (EDGES) of a global absorption trough centred at  $z \approx 17$ , presumably an imprint of the first luminous sources at cosmic dawn, provides a case in point (Bowman et al. 2018; Madau 2018). There, the physics of Ly $\alpha$  photon escape from high- $z$  sources is crucial to understand the local radiation background needed to achieve strong coupling of the spin and kinetic temperatures.

In this work, we present a detailed Ly $\alpha$  radiative transfer study of a star-forming galaxy over the redshift range of  $z = 5-7$ , selected to be sufficiently bright to have detectable Ly $\alpha$  with the upcoming

*James Webb Space Telescope (JWST)* but not too massive to be very dusty and less important for reionization. In Section 2, we describe the high-resolution cosmological simulation and various radiative transfer methodologies, including for ionizing, Ly $\alpha$ , and UV continuum photons. In Section 3, we discuss the time, spatial, and angular-resolved properties of the Ly $\alpha$  photons immediately after emission, after scattering through the ISM and CGM, and after transmission through the IGM. We isolate properties for the recombination and collisional excitation mechanisms. Our simulations provide insights on the Ly $\alpha$  escape fraction, angular anisotropy, red-to-blue flux ratio, half-light radius, halo scale height, frequency moments, peak offset, full-width at half-maximum, and equivalent width of the emergent spectra. Beyond this, we also characterize and illustrate quantities related to the surface brightness, line flux, moment maps, azimuthal variations, and angular power spectra. In Section 4, we consider the observability of our target galaxy to better inform LAE survey strategies. In Section 5, we perform a statistical analysis to construct a model predicting the escape fraction of Ly $\alpha$  photons from high- $z$  galaxies based on the local emission environment. In Section 6, we calculate the role of Ly $\alpha$  radiation pressure from the post-processed cosmological simulations. Finally, in Section 7, we briefly discuss the implications of our work and the desirable requirements of future Ly $\alpha$  radiative transfer studies.

## 2 METHODS

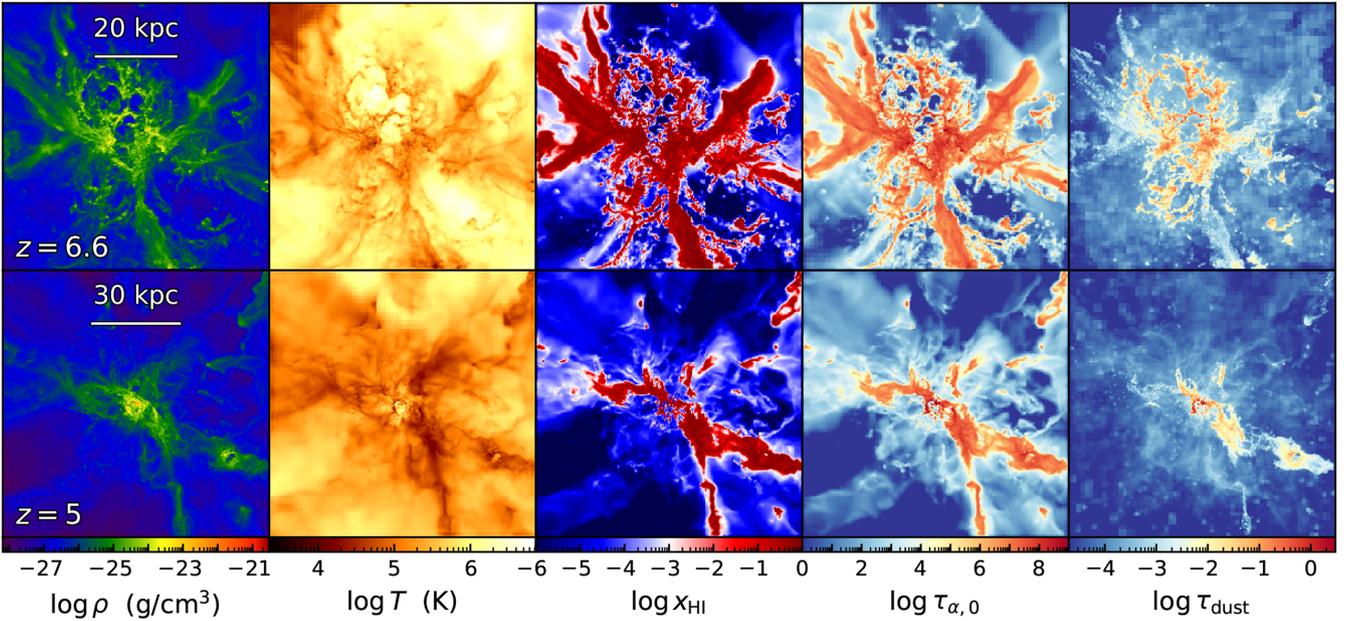
We now describe the cosmological simulation and post-processing methods to accurately calculate Ly $\alpha$  observables. In particular, we provide the context for the results of our study in Section 3, regarding the spectral properties and evolution of high-redshift LAEs.

### 2.1 Simulation setup

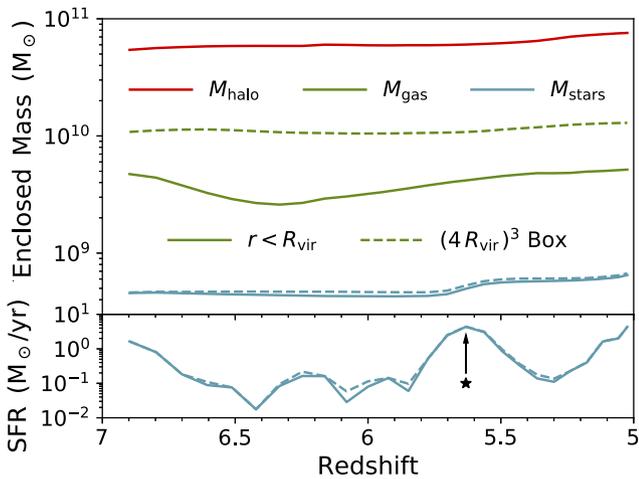
In this work, we examine a cosmological zoom-in simulation from the suite of galaxies presented in Ma et al. (2018), specifically the galaxy labeled *z5m11c*, which was selected based on being a typical low-mass LAE at  $z = 5$ . The simulation is a part of the Feedback In Realistic Environments project<sup>1</sup> (FIRE-2; Hopkins et al. 2017), which is an updated version of the first implementation (FIRE-1; Hopkins et al. 2014). The simulation employs the meshless finite-mass (MFM) method in the GIZMO hydrodynamics code (Hopkins 2015), with initial conditions and halo selection criteria as described in Ma et al. (2018). The multiscale zoom-in process ensures zero contamination from low-resolution particles within  $2R_{\text{vir}}$  at  $z = 5$ . The initial particle masses are  $m_{\text{gas}} = 7126.5 M_{\odot}$  for gas and  $m_{\text{DM}} = 3.9 \times 10^4 M_{\odot}$  for dark matter, while the minimum Plummer-equivalent force softening lengths of gas, star, and high-resolution dark matter particles are  $\epsilon_{\text{gas}} = 0.42$  pc,  $\epsilon_{\star} = 2.1$  pc, and  $\epsilon_{\text{DM}} = 42$  pc, respectively. The softening lengths are fixed for stars and dark matter but adaptive for gas, with the minimum value in comoving units at  $z > 9$  and physical units thereafter.

In Fig. 1, we show the spatial distribution of the gas based on mass-weighted projections of hydrodynamical quantities related to Ly $\alpha$  radiative transfer, including the density, temperature, ionization state, and dust content. Furthermore, in Fig. 2, we show the redshift evolution of the halo mass  $M_{\text{halo}}$  of the target galaxy, along with the total stellar mass  $M_{\star}$  and gas mass  $M_{\text{gas}}$  within the virial radius  $R_{\text{vir}}$  and  $(4 R_{\text{vir}})^3$  MCRT simulation domain. As a summary reference,

<sup>1</sup> See the FIRE project web site at: <http://fire.northwestern.edu> <http://fire.northwestern.edu>.



**Figure 1.** Projected images of the target galaxy at  $z = 6.6$  and  $z = 5$ , illustrating the redshift evolution of the line-of-sight mass-weighted density  $\rho$ , temperature  $T$ , neutral hydrogen fraction  $x_{\text{H I}} \equiv n_{\text{H I}}/n_{\text{H}}$ , Ly $\alpha$  optical depth at line centre  $\tau_{\alpha,0}$ , and dust optical depth  $\tau_{\text{dust}}$  (see Section 2.3 regarding the last two quantities). The density, temperature, and ionization states are extremely anisotropic and undergo significant morphological changes with redshift.



**Figure 2.** The redshift evolution of the target galaxy’s halo mass  $M_{\text{halo}}$ , along with the total stellar mass  $M_{\star}$  and gas mass  $M_{\text{gas}}$ , within the virial radius  $R_{\text{vir}}$ . The total gas and stellar content within the  $(4 R_{\text{vir}})^3$  MCRT simulation domain remains fairly constant, although at  $z \approx 5.63$  the galaxy undergoes significant starburst activity, i.e. star formation rates  $\gtrsim 1 M_{\odot} \text{ yr}^{-1}$ .

we provide several time-averaged quantities in Table 1, both over the entire redshift range and further divided into intervals to indicate any evolution. The time-averaged masses are  $\langle M_{\text{halo}} \rangle \approx 6.2 \times 10^{10} M_{\odot}$ ,  $\langle M_{\text{gas}} \rangle \approx 1.1 \times 10^{10} M_{\odot}$ , and  $\langle M_{\star} \rangle \approx 5.2 \times 10^8 M_{\odot}$ . The corresponding virial radius for the central halo is  $\langle R_{\text{vir}} \rangle \approx 19 \text{ kpc}$ , with a maximum circular velocity of  $\langle V_{\text{max}} \rangle \approx 127 \text{ km s}^{-1}$  reached at a radius of  $\langle R_{\text{max}} \rangle \approx 11 \text{ kpc}$ . We use the Binary Population and Spectral Synthesis (BPASS) models (version 2.0; Eldridge, Izzard & Tout 2008; Stanway, Eldridge & Becker 2016) to compute the spectral energy distribution (SED) of each star particle based on its age and metallicity. The BPASS models include binary stellar populations, taking into account effects from mass transfer, common envelope phases, binary mergers, and quasi-homogeneous evolution

at low metallicities. The inclusion of binaries is often invoked to reproduce the nebular line properties in  $z \approx 2\text{--}3$  galaxies (Steidel et al. 2016). In this work, we adopt a Kroupa (2002) initial mass function (IMF) from  $0.1\text{--}100 M_{\odot}$ , with IMF slopes of  $-1.30$  from  $0.1\text{--}0.5 M_{\odot}$  and  $-2.35$  from  $0.5\text{--}100 M_{\odot}$ . Under these conditions, we calculate an intrinsic ionizing ( $> 13.6 \text{ eV}$ ) photon emission rate of  $\langle \dot{N}_{\text{ion}} \rangle \approx 1.4 \times 10^{53} \text{ s}^{-1}$  and a luminosity of  $\langle L_{\text{ion}} \rangle \approx 4.9 \times 10^{42} \text{ erg s}^{-1}$ .

## 2.2 Intrinsic emission

The dominant source of Ly $\alpha$  photons is recombination radiation due to nebular emission. In a study by Ma et al. (2015), the approximate prescription for on-the-fly photoionization in the FIRE simulations was found to be fairly consistent with more accurate post-processing calculations. However, to ensure the reliability of the line emissivities and scattering opacities necessary for Ly $\alpha$  radiative transfer, we must first perform Lyman continuum (LyC) radiative transfer to update the local ionization states.<sup>2</sup> To retain the high resolution of the simulation, we deposited the unstructured mesh data on to an adaptive octree grid structure with a refinement criterion of no more than two particles per leaf cell. We then employed an octree version of the Monte Carlo code described in Ma et al. (2015, 2016) based on the SEDONA code (Kasen, Thomas & Nugent 2006), which iteratively solves the ionization states assuming ionization equilibrium. The code accounts for photoionization from each star particle and a uniform, redshift-dependent meta-galactic ionizing background and collisional ionization (Faucher-Giguère

<sup>2</sup>Even though the ionization states change, we keep the gas temperature the same as the original simulations. This leads to some overheated H II regions, but these cells would only make a small change to the total Ly $\alpha$  emissivity as the recombination coefficient depends weakly on temperature. In the halo, the hydrodynamical simulations already fairly capture both photoheating by the ionizing background and shock-heating of accreted gas.

**Table 1.** Time-averaged quantities over the redshift interval  $z = 5-7$ , with additional subdivisions illustrating the evolution. The type column denotes intrinsic properties of the galaxy (INT) and quantities that rely on the ionizing radiative transfer calculations (ION). Specifically,  $M_{\text{gas}}$ ,  $M_*$ ,  $N_{\text{ion}}$ ,  $L_{\text{ion}}$ ,  $\langle x_{\text{HII}} \rangle_V$ , and  $\langle x_{\text{HII}} \rangle_\rho$  are calculated for the gas within the virial radius, while all other quantities are for the  $(4 R_{\text{vir}})^3$  MCRT simulation domain to capture the extended Ly $\alpha$  halo. We also report the  $1\sigma$  standard deviation, defined as  $\sigma_f^2 = \int (f(t) - \mu_f)^2 dt / \int dt$ , where the time-average is  $\mu_f = \int f(t) dt / \int dt$ . In cases with LOS fluctuations, we provide the median and asymmetric  $1\sigma$  (68.27%) time-weighted confidence levels, based on binning the escaped photon packets into 3072 healpix directions of equal solid angle.

Quantity	Type	$z = 5-7$	$z = 6-7$	$z = 5-6$
$\log M_{\text{halo}} [M_\odot]$	INT	$10.79 \pm 0.04$	$10.77 \pm 0.01$	$10.81 \pm 0.04$
$\log M_{\text{gas}} [M_\odot]$	INT	$9.58 \pm 0.10$	$9.50 \pm 0.08$	$9.64 \pm 0.06$
$\log M_* [M_\odot]$	INT	$8.69 \pm 0.06$	$8.65 \pm 0.01$	$8.72 \pm 0.06$
$R_{\text{vir}} [\text{kpc}]$	INT	$18.8 \pm 2.0$	$16.9 \pm 0.7$	$20.1 \pm 1.4$
$R_{\text{max}} [\text{kpc}]$	INT	$11.2 \pm 1.9$	$12.5 \pm 1.4$	$10.4 \pm 1.7$
$V_{\text{max}} [\text{km s}^{-1}]$	INT	$127.0 \pm 4.4$	$126.9 \pm 5.6$	$127.0 \pm 3.5$
$\log N_{\text{ion}} [\text{s}^{-1}]$	INT	$53.1 \pm 0.5$	$52.8 \pm 0.4$	$53.3 \pm 0.5$
$\log L_{\text{ion}} [\text{erg s}^{-1}]$	INT	$42.64 \pm 0.5$	$42.4 \pm 0.4$	$42.8 \pm 0.5$
$\langle x_{\text{HII}} \rangle_V [\%]$	ION	$97.9 \pm 1.1$	$97.4 \pm 1.0$	$98.2 \pm 1.1$
$\langle x_{\text{HII}} \rangle_\rho [\%]$	ION	$4.7 \pm 1.0$	$5.2 \pm 1.0$	$4.4 \pm 0.9$
$f_{\text{esc}}^{\text{LyC}} [\%]$	ION	$13.3 \pm 9.6$	$16.6 \pm 9.3$	$11.2 \pm 9.2$
$f_{\text{esc}}^{\text{UV}} [\%]$	ION	$92.8^{+3.9}_{-17.1}$	$96.0^{+1.9}_{-3.2}$	$87.1^{+7.6}_{-18.7}$
$M_{\text{UV}}$	ION	$-18.3^{+0.5}_{-1.0}$	$-18.2^{+0.3}_{-0.8}$	$-18.4^{+0.7}_{-0.8}$
$R_{1/2, \text{UV}} [\text{kpc}]$	ION	$2.9^{+0.8}_{-1.3}$	$3.2^{+0.6}_{-0.7}$	$2.4^{+1.0}_{-1.1}$
$\log L_{\text{ion}} [\text{erg s}^{-1}]$	ION	$41.7 \pm 0.7$	$41.6 \pm 0.6$	$41.7 \pm 0.7$

et al. 2009). Each of these mechanisms contributes to the resolved Ly $\alpha$  emissivity in the simulation. Therefore, the Ly $\alpha$  luminosity due to recombination is

$$L_\alpha^{\text{rec}} = h\nu_\alpha \int P_B(T) \alpha_B(T) n_e n_p dV, \quad (1)$$

where  $h\nu_\alpha = 10.2$  eV, the Ly $\alpha$  conversion probability per recombination event is  $P_B(T) \approx 0.68$ , the case B recombination coefficient is  $\alpha_B$ , and the number densities  $n_e$  and  $n_p$  are for free electrons and protons, respectively (Cantalupo, Porciani & Lilly 2008; Dijkstra 2014). At these redshifts, we also model the radiation arising from collisions between free electrons and neutral hydrogen atoms, which populate excited states and produce additional Ly $\alpha$  ‘cooling’ emission. The Ly $\alpha$  luminosity due to collisional excitation is

$$L_\alpha^{\text{col}} = h\nu_\alpha \int q_{1s2p}(T) n_e n_{\text{HI}} dV, \quad (2)$$

where the temperature-dependent rate coefficient  $q_{1s2p}(T)$  is taken from Scholz & Walters (1991). Due to the exponential dependence on temperature around  $T \sim 10^4$  K, this term can be highly uncertain, and thus requires a proper treatment of the thermal effects of ionizing radiation, e.g. as demonstrated in Faucher-Giguère et al. (2010).

As expected, during these redshifts, the gas within the virial radius  $R_{\text{vir}}$  is mostly ionized by volume but neutral by density. Specifically, after the post-processing calculations, the time-averaged ionized fraction is  $\langle x_{\text{HII}} \rangle_V \approx 97.9$  per cent, when weighted by volume, and  $\langle x_{\text{HII}} \rangle_\rho \approx 4.7$  per cent, when weighted by density (Table 1), where  $x_{\text{HII}} \equiv n_{\text{HII}}/n_{\text{H}}$  is the ratio of ionized and total hydrogen number densities. The escape fraction of ionizing photons from high- $z$  galaxies can fluctuate significantly in response to the specific history of mergers, cosmological cold gas flows, and starburst

activity. We calculate the time-averaged LyC escape fraction as  $\langle f_{\text{esc}}^{\text{LyC}} \rangle \approx 13.3$  per cent, such that the ionizing luminosity escaping beyond the CGM ( $\sim 2 R_{\text{vir}}$ ) is  $\langle L_{\text{ion,esc}} \rangle \approx 4.7 \times 10^{41}$  erg s $^{-1}$ .

The corresponding intrinsic Ly $\alpha$  luminosity within the  $(4 R_{\text{vir}})^3$  MCRT simulation domain based on equations (1) and (2) is  $\langle L_\alpha \rangle \approx 1.3 \times 10^{42}$  erg s $^{-1}$ , but also fluctuates in proportion to the intrinsic ionizing luminosity with a peak of  $L_\alpha \approx 10^{43}$  erg s $^{-1}$  at  $z \approx 5.63$ . In Fig. 3, we illustrate the relative Ly $\alpha$  luminosity across the density–temperature phase space. Most of the emission comes from moderate density H II regions with a temperature of  $T \sim 10^4$  K, characteristic of Ly $\alpha$  cooling. We have verified that our approach of modelling the recombination radiation directly is consistent with the luminosity calculated from star particles, as the LyC code is photon conserving. The spatial distribution of recombination emission is also more realistic within such resolved H II region morphologies than emitting Ly $\alpha$  radiation from star particles directly.

### 2.3 Ly $\alpha$ radiative transfer

We perform Ly $\alpha$  MCRT with the Cosmic Ly $\alpha$  Transfer code (COLT; Smith et al. 2015). Photon packets are inserted according to the recombination and collisional luminosities from equations (1) and (2). This is achieved by constructing the cumulative distribution function from individual cells, drawing a random number to find the source, and assigning the emission position uniformly within the cell volume. The subsequent transport of photon packets follows the Monte Carlo procedure with a local Ly $\alpha$  absorption coefficient of

$$k_\alpha = n_{\text{HI}} \sigma_\alpha(\nu), \quad (3)$$

where the Ly $\alpha$  cross-section  $\sigma_\alpha(\nu)$  is proportional to the Voigt line profile. At line centre, the scattering cross-section obtains a maximum value of  $\sigma_0 = 5.898 \times 10^{-14} [T/(10^4 \text{ K})]^{-1/2}$  cm $^2$ . For a visual comparison, we also show the projected line centre optical depth  $\tau_{\alpha,0} = \int n_{\text{HI}} \sigma_0 d\ell$  in Fig. 1.

We follow the prescription of Laursen, Sommer-Larsen & Andersen (2009b) for calculating the dust content within each cell. Specifically, we compute the effective local dust absorption coefficient as

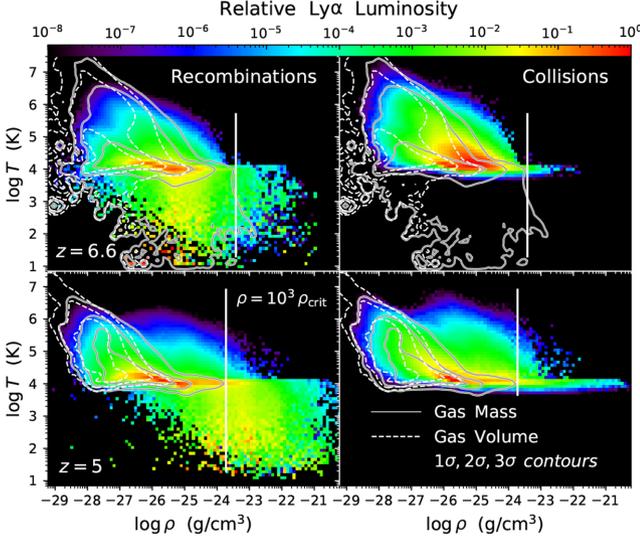
$$k_d = (x_{\text{HI}} + f_{\text{ion}} x_{\text{HII}}) n_{\text{H}} \sigma_{\text{d,SMC}} \frac{Z}{Z_{\text{SMC}}}, \quad (4)$$

where  $n_{\text{H}}$  denotes the hydrogen number density such that  $x_{\text{HI}} \equiv n_{\text{HI}}/n_{\text{H}}$  and  $x_{\text{HII}} \equiv n_{\text{HII}}/n_{\text{H}}$  are the neutral and ionized fractions, respectively. We assume the dust survival fraction within ionized regions  $f_{\text{ion}}$  is 1 per cent, although in reality the dust abundance also depends on the temperature of the ionized medium. Furthermore, this model assumes Small Magellanic Cloud (SMC) type dust with an effective cross-section per hydrogen atom of  $\sigma_{\text{d,SMC}} \approx 3.95 \times 10^{-22}$  cm $^2$ , where  $\log(Z_{\text{SMC}}/Z_\odot) \approx -0.6$  with  $Z_\odot \approx 0.0134$ . Because this particular FIRE simulation does not include on-the-fly subgrid metal diffusion, we use a smoothed version of the metallicity  $Z$  based on a cubic spline kernel over the 32 nearest neighbor particles before mapping to the octree. Finally, the dust radiative transfer also follows the methodology outlined by Laursen et al. (2009b), i.e. with a fiducial dust scattering albedo of  $A = 0.32$  and Henyey–Greenstein phase function with an asymmetry parameter of  $g = (\cos \theta) = 0.73$ . We show the projected dust optical depth  $\tau_d = \int k_d d\ell$  in Fig. 1.

We now describe the binning method to obtain line-of-sight (LOS) statistics for a large number of directions. We perform separate COLT simulations for recombination and collisional excitation emission, each with  $10^7$  photon packets. The final state of

**Table 1.** *continued.* Time-averaged properties of the intrinsic Ly $\alpha$  emission over the redshift interval  $z = 5-7$ , with subdivisions summarizing the evolution. We also show separate columns for photons originating from recombinations and collisional excitation.

Quantity	Type	All Ly $\alpha$ photons			Ly $\alpha$ from recombinations			Ly $\alpha$ from collisional excitation		
		$z = 5-7$	$z = 6-7$	$z = 5-6$	$z = 5-7$	$z = 6-7$	$z = 5-6$	$z = 5-7$	$z = 6-7$	$z = 5-6$
$\log L_\alpha$ [erg s $^{-1}$ ]	ION	$42.2 \pm 0.4$	$41.9 \pm 0.3$	$42.3 \pm 0.4$	$42.0 \pm 0.5$	$41.7 \pm 0.3$	$42.2 \pm 0.5$	$41.5 \pm 0.2$	$41.5 \pm 0.2$	$41.6 \pm 0.2$
$L_\alpha^{\text{rec,col}}/L_\alpha$ [%]	ION	–	–	–	$70.8 \pm 13.4$	$63.1 \pm 9.8$	$75.8 \pm 13.1$	$29.2 \pm 13.4$	$36.9 \pm 9.8$	$24.2 \pm 13.1$
$F_{\text{red}}/F_{\text{blue}}$	ION	$1^{+0.81}_{-0.45}$	$1^{+0.71}_{-0.42}$	$1^{+0.94}_{-0.48}$	$1^{+1.30}_{-0.57}$	$1^{+1.26}_{-0.56}$	$1^{+1.34}_{-0.57}$	$1^{+0.42}_{-0.30}$	$1^{+0.41}_{-0.29}$	$1^{+0.43}_{-0.30}$
$\langle \Delta v \rangle$ [km s $^{-1}$ ]	ION	$0 \pm 21.1$	$0 \pm 20.7$	$0 \pm 21.4$	$0 \pm 27.1$	$0 \pm 30.1$	$0 \pm 25.5$	$0 \pm 13.6$	$0 \pm 14.3$	$0 \pm 13.1$
$\sigma_{\Delta v}$ [km s $^{-1}$ ]	ION	$68.4^{+24.8}_{-17.3}$	$80.2^{+20.8}_{-22.5}$	$62.4^{+19.4}_{-14.0}$	$66.0^{+27.6}_{-19.7}$	$81.7^{+21.8}_{-26.7}$	$58.7^{+22.0}_{-14.9}$	$71.7^{+17.2}_{-12.2}$	$74.3^{+21.2}_{-15.4}$	$70.8^{+11.7}_{-10.8}$
$\text{EW}_{\text{Ly}\alpha}$ [Å]	ION	$107 \pm 56$	$63 \pm 16$	$135 \pm 54$	$83 \pm 59$	$42 \pm 17$	$110 \pm 62$	$24 \pm 7$	$21 \pm 3$	$25 \pm 8$



**Figure 3.** Relative Ly $\alpha$  luminosity in the density–temperature phase plane at  $z = 6.6$  (top) and  $z = 5$  (bottom). Ly $\alpha$  photons are produced by recombination of ionized hydrogen (left) and collisional excitation of hydrogen atoms by free electrons (right). The two mechanisms show significant overlap at warm temperatures ( $\sim 10^4$  K) and moderate densities ( $\sim 10^{-27}$ – $10^{-23}$  g cm $^{-3}$ ), tracing photoionized H II regions and shock-heated gas in the CGM. The recombination emission at  $T < 10^4$  K is likely associated with relic H II region cooling and inefficient photoheating, whereas collisional excitation requires higher temperatures. However, see footnote 2 regarding the Ly $\alpha$  emissivity from underheated H II regions. The solid (dashed) contours show the phases containing most of the mass (volume). We also provide reference overdensities of  $\rho/\rho_{\text{crit}} = 10^3$  as vertical lines in each panel.

all escaped and absorbed photons is recorded, i.e. the frequency  $\nu_i$ , position  $\mathbf{r}_i$ , and direction  $\hat{\mathbf{k}}_i$ , where  $i$  denotes the photon index. Although Poisson noise becomes significant at high angular and spatial resolution, we smooth the packet discretization by a weighting kernel with a specific estimated error criterion. For spatially integrated quantities in a given direction  $\hat{\mathbf{n}}$ , we calculate the directional cosine of each photon as  $\mu_i \equiv \cos \theta_i = \hat{\mathbf{n}} \cdot \hat{\mathbf{k}}_i$  and  $\Delta\mu_i \equiv 1 - \mu_i$ . For reference, a tophat filter has constant weight if  $\mu \geq \mu_{\text{min}}$  and zero elsewhere, for an effective photon number of  $N_{\text{ph,eff}} \approx N_{\text{ph}} \Delta\mu_{\text{min}}/2$ . Under an isotropic distribution, the approximate error per pixel is  $\approx 1/\sqrt{N_{\text{ph,eff}}}$ , such that 1 per cent noise is obtained for  $10^7$  photons with  $\Delta\mu_{\text{min}} \approx 0.002$  or  $\Delta\theta_{\text{min}} \approx 3:6$  defining the effective resolution. However, in this work, we employ a Gaussian filter to increase the sensitivity and avoid edge effects, represented as  $W_G(\mu) \propto \exp(-\Delta\mu^2/2\sigma_\mu^2)$  if  $\mu \geq 0$  and

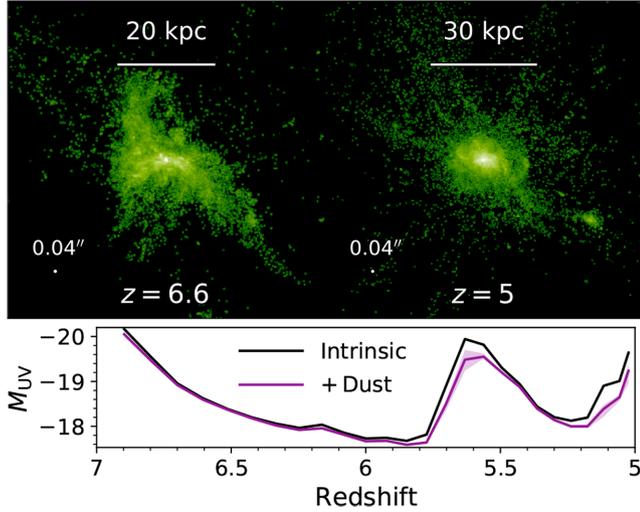
zero elsewhere, which when integrated gives an effective number of photons of  $N_{\text{ph,eff}} \approx N_{\text{ph}} \sqrt{\frac{\pi}{2}} \frac{\sigma_\mu}{2} \text{erf}(1/\sqrt{2}\sigma_\mu)$ . Numerically solving the analogous relation yields an approximate error per pixel of 1 per cent for  $10^7$  photons with a standard deviation of  $\sigma_\mu \approx 0.0016$  or  $\sigma_\theta \approx 3:2$  as the effective resolution. We find that this prescription provides reasonable control on the Poisson noise due to the Monte Carlo discretization. We also note that uncertainty from noise has the greatest impact on sightlines that are less likely to be observable.

## 2.4 IGM transmission

The COLT output represents the emergent Ly $\alpha$  observables without accounting for subsequent scattering in the IGM. To include this important effect, we use the frequency- and redshift-dependent transmission curves of Laursen, Sommer-Larsen & Razoumov (2011), which have been kindly provided to us by the author. Specifically, we use their ‘benchmark’ Model 1 results as described in relation to their Figs. 2 and 3. The curves provide the median and  $1\sigma$  statistics based on a large number of sightlines ( $\gtrsim 10^3$ ) cast from several hundreds of galaxies through a simulated cosmological volume. Ly $\alpha$  photons are generally free streaming beyond a few virial radii, so in this work ‘ISM’ denotes escape from the  $(4 R_{\text{vir}})^3$  MCRT simulation domain, while ‘IGM’ means that each photon is reweighted by  $\exp[-\tau_{\text{IGM}}(\nu, z)]$  from the transmission model. Any IGM model implemented in this manner should be considered in the context of nontrivial variations depending on the particular galaxy and sightline. We are not currently aware of any Ly $\alpha$  radiative transfer simulations that self-consistently follow the resonant scattering through both the highly resolved ISM of individual galaxies and the more diffuse IGM throughout cosmological volumes. Such a treatment is beyond the scope of the analysis of this work but will be pursued in a future work.

## 2.5 UV continuum

We also calculate the escape of non-ionizing UV continuum photons, which is necessary for observational comparison to the Ly $\alpha$  rest-frame equivalent width  $\text{EW}_{\text{Ly}\alpha,0}$ . Given the uncertainties in this study, it is sufficient to employ single-scattering monochromatic dust attenuation with an absorption coefficient given by equation (4), and we expect a more accurate treatment with multiple scattering and frequency dependence to only slightly modify our results. Specifically, the intrinsic continuum luminosity,  $L_\lambda$  in units of erg s $^{-1}$  Å $^{-1}$ , is found by integrating the SED of each stellar population over a filter with an effective wavelength of  $\lambda_{\text{UV}} = 1500$  Å. Images and other LOS observables in the  $\hat{\mathbf{n}}$  direction are calculated by direct ray-tracing from each star particle, such that  $f_{\lambda,\text{obs}}(\hat{\mathbf{n}}) \equiv \sum_i L_{\lambda,i} \exp(-\tau_{d,i})/(4\pi(1+z)d_L^2)$ , where the dust optical depth



**Figure 4.** *Top:* Spatial images of the UV continuum for the target galaxy at  $z = 6.6$  and  $z = 5$ , showing the evolving distribution of stellar populations. The images are rendered after absorption by dust and Gaussian smoothing with  $\text{FWHM} \approx 0.04''$  to simulate the *JWST* NIRCcam aperture. *Bottom:* The evolution of the absolute magnitude  $M_{\text{UV}}$ , with  $1\sigma$  LOS variations.

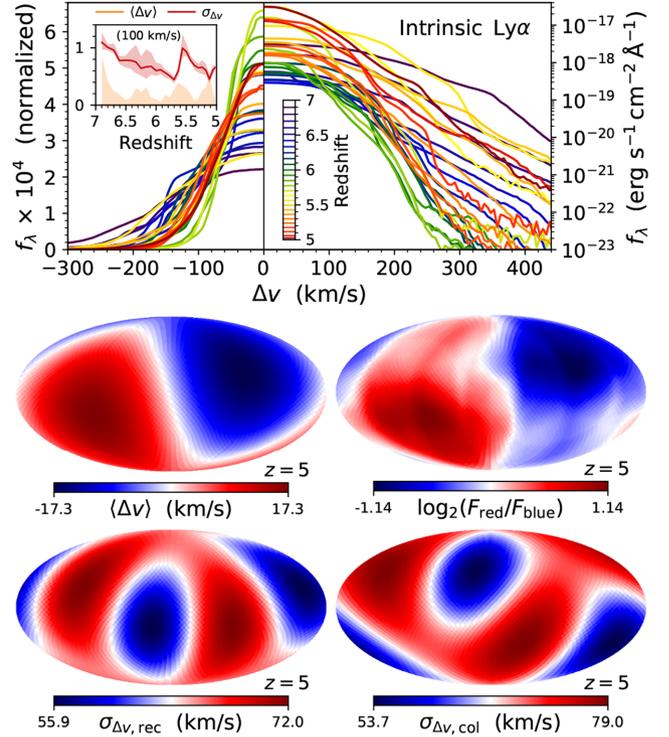
from star  $i$  is  $\tau_{d,i} \equiv \int_{i,\tilde{n}} k_d d\ell$  and  $d_L$  is the luminosity distance to the galaxy. In Fig. 4, we show the escaped UV continuum for the target galaxy at  $z = 6.6$  and  $z = 5$ . We also apply Gaussian smoothing with a full width at half maximum of  $\text{FWHM} \approx 0.04''$  to simulate the *JWST* NIRCcam aperture. We find the time-weighted median LOS escape fraction for non-ionizing UV continuum radiation to be  $\langle f_{\text{esc}}^{\text{UV}} \rangle \approx 93$  per cent, which decreases at lower redshift as the mass and total dust content of the galaxy increases. The dust-attenuated rest-frame UV absolute magnitude is  $\langle M_{\text{UV}} \rangle \approx -18.25$  with a half-light radius of  $\langle R_{1/2,\text{UV}} \rangle \approx 3 \text{ kpc} \approx 0.5''$ , which becomes brighter and more concentrated with time.

### 3 LY $\alpha$ RADIATIVE TRANSFER RESULTS

#### 3.1 Intrinsic spectra

We first discuss the spectral properties of the intrinsic Ly $\alpha$  emission. Under the Case B approximation, the H $\alpha$  (H $\beta$ ) hydrogen Balmer transition at  $\lambda_{\text{H}\alpha} = 6562.8 (4861.3) \text{ \AA}$  closely traces the intrinsic Ly $\alpha$  recombination emission, albeit with a factor of  $F_{\text{Ly}\alpha}/F_{\text{H}\alpha} \approx 0.68 v_{\text{Ly}\alpha}/0.45 v_{\text{H}\alpha} \approx 8 (23)$  reduction in flux (Dijkstra 2014). For methodological consistency, we run COLT without any absorption or scattering with  $10^8$  photon packets. Interestingly, the proper motion of the galaxy induces a substantial Doppler shift and complex spectral morphology for the LOS line, e.g. asymmetric peaks and plateaus. In fact, the peak frequency and full width at half maximum (FWHM) are not always robust characterizations of the line profile. Therefore, we instead calculate the flux-weighted frequency centroid defined as  $\langle \Delta v \rangle \equiv \int \Delta v f_{\lambda} d\lambda / \int f_{\lambda} d\lambda$  and the standard deviation of the line as  $\sigma_{\Delta v} \equiv ((\Delta v^2) - \langle \Delta v \rangle^2)^{1/2}$ .

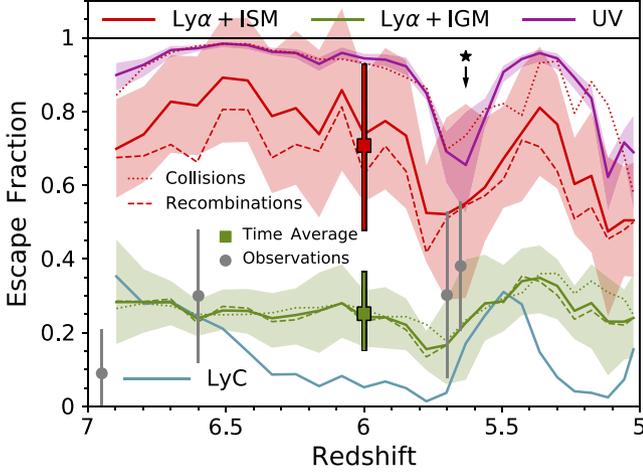
In Fig. 5, we show the redshift evolution of the intrinsic Ly $\alpha$  line profile. Although the angular-averaging generates symmetric profiles about the line centre, the observed spectra are much more complex due to the aggregate LOS motion of the emitting gas. We also illustrate the angular distribution of the frequency moments



**Figure 5.** *Top:* Angular-averaged intrinsic Ly $\alpha$  flux as a function of Doppler velocity  $\Delta v = c\Delta\lambda/\lambda$  for each simulation snapshot. The mean profiles are symmetric about the line centre; however, observed spectra are more complex due to the aggregate LOS motion of the emitting gas. In the left hand-panel, we show the evolving shape of the normalized lines while the right-hand panel shows the fluctuating luminosity. In the inset, we include the redshift evolution of the flux-weighted frequency centroid  $\langle \Delta v \rangle \equiv \int \Delta v f_{\lambda} d\lambda / \int f_{\lambda} d\lambda$  and standard deviation  $\sigma_{\Delta v} \equiv ((\Delta v^2) - \langle \Delta v \rangle^2)^{1/2}$  as a measure of the velocity offset and width of the line, respectively. *Bottom:* We illustrate the angular distribution of these quantities, which exhibit a dipole moment in  $\langle \Delta v \rangle$  and  $F_{\text{red}}/F_{\text{blue}}$  from the galaxy's proper motion and a quadrupole modulation of  $\sigma_{\Delta v}$  from triaxial velocity dispersion. There is also a slight kinematic misalignment between recombination and collisional emission.

with healpix maps. There is a dipole moment<sup>3</sup> ( $\gtrsim 10 \text{ km s}^{-1}$ ) due to the proper motion of the galaxy, with higher-order residual angular variations ( $\lesssim 1 \text{ m s}^{-1}$ ) from satellite galaxies and other substructures representing a small fraction of the total luminosity ( $\lesssim 10^{-4}$ ). This is mirrored in the red-to-blue-flux ratio  $F_{\text{red}}/F_{\text{blue}}$ , which varies by a factor of  $\approx 4$ . The line width also has a significant quadrupole modulation due to axially-asymmetric velocity dispersion, which induces broader lines along sightlines with higher directed turbulent motions. Specifically, we calculate a time- and angular-averaged value of  $\langle \sigma_{\Delta v} \rangle \approx 68 \pm 21 \text{ km s}^{-1}$  ( $\pm 30$  per cent), corresponding to  $\langle \text{FWHM} \rangle \approx 161 \pm 50 \text{ km s}^{-1}$ , assuming the line is approximately Gaussian. We also note that Ly $\alpha$  photons produced by recombinations and collisionals have a kinematic misalignment of their  $\langle \Delta v \rangle$

<sup>3</sup>The results in this work are given in the centre of mass velocity frame based on the dark matter and gas within a  $3R_{\text{vir}}$  sphere. We note that the dipole signal in the intrinsic flux depends on the reference frame so may not be a robust prediction. One could instead measure Ly $\alpha$  properties based on the systemic redshifts of optically-thin lines. However, the Ly $\alpha$  line couples to the CGM so without a more self-consistent treatment of IGM transmission the cosmological reference frame is the least biased over all sightlines.



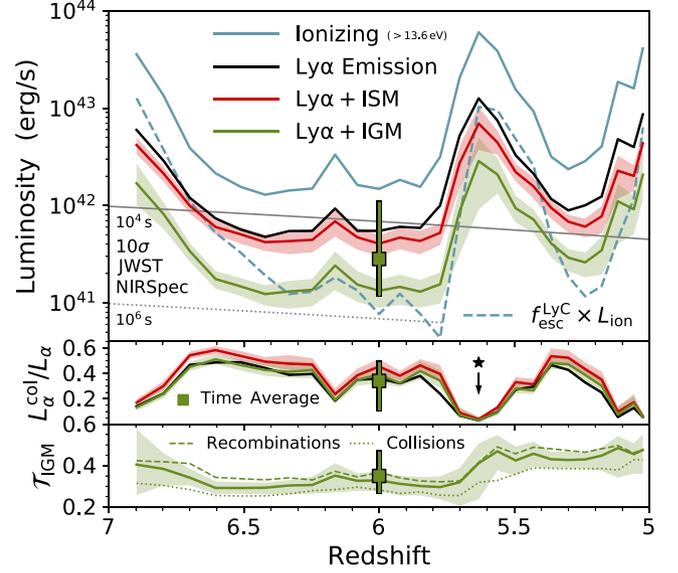
**Figure 6.** The redshift evolution of the Ly $\alpha$ , non-ionizing UV (1500 Å), and LyC escape fractions from the target galaxy. The red curve is after dust absorption in the ISM while the green curves are after IGM transmission based on the model of Laursen et al. (2011). The shaded regions show the  $1\sigma$  confidence levels considering different viewing angles. The grey data points are the observed Ly $\alpha$  escape fraction estimates reported by Hayes et al. (2011). The square points at  $z = 6$  are the time-weighted averages from Tables 1–3. Although the escape fraction of Ly $\alpha$  photons originating from collisions is higher than from recombinations, the transmitted escape fractions are roughly the same. The star symbol marks the main starburst.

dipole and  $\sigma_{\Delta v}$  quadrupole moments due to occupying slightly different regions of  $\rho$ – $T$  phase space (see Fig. 3).

### 3.2 Escape fraction

One of the primary motivations for this study is to determine how the Ly $\alpha$  escape fraction  $f_{\text{esc}}^{\text{Ly}\alpha}$  evolves for a typical LAE during the epoch of reionization. In Fig. 6, we show the fraction of photons emerging from the  $(4 R_{\text{vir}})^3$  region centered on the galaxy. The red curve represents the result after absorption by dust during scattering through the ISM, while the green curve takes into account the subsequent transmission through the IGM, based on the model by Laursen et al. (2011, see Section 2.4). To understand the uncertainty in the IGM model, we compared the results obtained from three subsamples based on the circular velocity of the host galaxy environment as being “small” ( $V_c < 55 \text{ km s}^{-1}$ ), “intermediate”, and “large” ( $V_c > 80 \text{ km s}^{-1}$ ), finding they yield similar results. Thus, for simplicity, we employ the model based on their entire sample of galaxies. For reference, in Fig. 6, we also show the escape fraction of 1500 Å UV photons  $f_{\text{esc}}^{\text{UV}}$  (purple curve) and the escape fraction of ionizing photons  $f_{\text{esc}}^{\text{LyC}}$  (blue curve).

The shaded regions in Fig. 6 illustrate the  $1\sigma$  variation (16<sup>th</sup> and 84<sup>th</sup> percentiles) based on binning the escaped photon packets into 3072 sightlines, corresponding to equal area healpix directions of the unit sphere. Including multiple viewing angles allows us to mimic observations of different galaxies with similar properties. Although the scatter due to orientation is not equivalent to cosmic variance, e.g. environmental and large-scale structure effects, which induce further variation across different parts of the sky, it is certainly useful to consider both effects and their corresponding biases. Finally, we include several observed escape fraction estimates over our redshift range from Hayes et al. (2011) as grey data points along with the reported errors, which is based on samples with comparable galaxy halo masses. Given the large



**Figure 7.** *Top:* The redshift evolution of ionizing and Ly $\alpha$  emission. The escaping luminosity corresponds to the respective escape fractions shown in Fig. 6 with similar styling. At  $z \approx 5.63$ , the galaxy experiences significant starburst activity, which is mirrored by a high escaping LyC luminosity,  $f_{\text{esc}}^{\text{LyC}} \times L_{\text{ion}}$ . We also show approximate visibility curves for the JWST NIRSpec instrument assuming a  $10\sigma$  detection after  $10^4$  and  $10^6$  s of exposure time. *Centre:* The fraction of luminosity originating from collisional excitation. *Bottom:* The overall transmission of Ly $\alpha$  photons through the IGM, defined as  $\mathcal{T}_{\text{IGM}} = f_{\text{esc,IGM}}^{\text{Ly}\alpha} / f_{\text{esc,ISM}}^{\text{Ly}\alpha}$ . Ly $\alpha$  photons from collisions experience lower IGM transmission than from recombinations due to the overall bluer spectral shape, i.e. infall signatures with less resonant broadening.

uncertainties in these measurements and the variation between sightlines, both sets of values are consistent over the redshift range considered.

Furthermore, the escape fraction exhibits significant fluctuation across redshift, lagging slightly behind the star formation activity. In Fig. 7, illustrate the evolution of the observed Ly $\alpha$  luminosity along with the fraction of luminosity originating from collisional excitation and the overall transmission of Ly $\alpha$  photons through the IGM, defined as  $\mathcal{T}_{\text{IGM}} = f_{\text{esc,IGM}}^{\text{Ly}\alpha} / f_{\text{esc,ISM}}^{\text{Ly}\alpha}$ . To summarize our results, we provide the corresponding time- and angular-averaged quantities in Tables 2 and 3, again over full and half intervals to indicate any evolution. Although the perceived trends are primarily due to the dynamical changes of our target galaxy, the evolution is representative of the expected behaviour of similar high- $z$  LAEs. It is significant that the intrinsic and transmitted Ly $\alpha$  escape fractions are relatively high compared to ionizing photons, with median values of  $\langle f_{\text{esc}}^{\text{Ly}\alpha} \rangle_{\text{ISM}} \approx 71$  per cent and  $\langle f_{\text{esc}}^{\text{Ly}\alpha} \rangle_{\text{IGM}} \approx 25$  per cent, such that the IGM transmission is  $\langle \mathcal{T}_{\text{IGM}} \rangle \approx 35$  per cent. The average observed luminosity is  $\langle L_{\alpha,\text{obs}} \rangle \approx 3 \times 10^{41} \text{ erg s}^{-1}$  with a peak of  $L_{\alpha,\text{obs}} \approx 3 \times 10^{42} \text{ erg s}^{-1}$  corresponding to the  $z \approx 5.6$  starburst. This implies that galaxies can go in and out of visibility depending on the observational sensitivity. In Fig. 7, we include curves for the JWST NIRSpec instrument capabilities, assuming a  $10\sigma$  point source detection after  $10^4$  and  $10^6$  s of exposure time at a medium spectral resolution of  $R \equiv \lambda/\Delta\lambda \approx 1000$  (Gardner et al. 2006).

**Table 2.** Time-averaged Ly $\alpha$  properties after accounting for ISM scattering (ISM) with a similar layout as Table 1.

Quantity	Type	All Ly $\alpha$ photons			Ly $\alpha$ from recombinations			Ly $\alpha$ from collisional excitation		
		$z = 5-7$	$z = 6-7$	$z = 5-6$	$z = 5-7$	$z = 6-7$	$z = 5-6$	$z = 5-7$	$z = 6-7$	$z = 5-6$
$\log L_\alpha$ [erg s $^{-1}$ ]	ISM	41.9 $^{+0.6}_{-0.2}$	41.7 $^{+0.3}_{-0.1}$	42.0 $^{+0.5}_{-0.3}$	41.6 $^{+0.7}_{-0.3}$	41.4 $^{+0.4}_{-0.2}$	41.8 $^{+0.7}_{-0.4}$	41.4 $^{+0.3}_{-0.2}$	41.4 $^{+0.4}_{-0.2}$	41.5 $^{+0.3}_{-0.2}$
$L_\alpha^{\text{rec,col}}/L_\alpha$ [%]	ISM	–	–	–	60.7 $^{+25.4}_{-13.6}$	52.1 $^{+19.8}_{-7.8}$	65.7 $^{+24.3}_{-14.6}$	39.3 $^{+13.6}_{-25.4}$	47.9 $^{+7.8}_{-19.8}$	34.3 $^{+14.6}_{-24.3}$
$f_{\text{esc}}^{\text{Ly}\alpha}$ [%]	ISM	70.8 $^{+21.3}_{-22.5}$	80.6 $^{+17.8}_{-17.8}$	63.1 $^{+23.0}_{-20.4}$	62.6 $^{+22.0}_{-20.7}$	70.9 $^{+20.5}_{-18.5}$	56.2 $^{+22.6}_{-19.6}$	87.8 $^{+20.2}_{-21.3}$	94.8 $^{+18.6}_{-17.1}$	82.2 $^{+21.2}_{-20.6}$
$F_{\text{LOS}}/F_\Omega$	ISM	1 $^{+0.24}_{-0.26}$	1 $^{+0.21}_{-0.22}$	1 $^{+0.26}_{-0.29}$	1 $^{+0.28}_{-0.30}$	1 $^{+0.25}_{-0.26}$	1 $^{+0.30}_{-0.33}$	1 $^{+0.21}_{-0.21}$	1 $^{+0.19}_{-0.19}$	1 $^{+0.22}_{-0.23}$
$F_{\text{red}}/F_{\text{blue}}$	ISM	1.0 $^{+0.4}_{-0.3}$	1.1 $^{+0.3}_{-0.2}$	0.9 $^{+0.4}_{-0.3}$	1.2 $^{+0.6}_{-0.4}$	1.4 $^{+0.7}_{-0.4}$	1.1 $^{+0.5}_{-0.4}$	0.8 $^{+0.2}_{-0.2}$	0.9 $^{+0.2}_{-0.2}$	0.7 $^{+0.2}_{-0.2}$
$R_{1/2}$ [kpc]	ISM	7.2 $^{+5.7}_{-3.4}$	10.6 $^{+4.4}_{-4.7}$	5.9 $^{+4.1}_{-3.0}$	6.6 $^{+4.1}_{-3.2}$	8.9 $^{+3.3}_{-3.8}$	5.6 $^{+3.6}_{-2.8}$	8.4 $^{+6.4}_{-3.5}$	12.1 $^{+5.8}_{-4.8}$	6.9 $^{+4.5}_{-2.9}$
$R_{\alpha, \text{h}}$ [kpc]	ISM	21.4 $^{+8.5}_{-4.5}$	22.2 $^{+9.1}_{-7.9}$	21.2 $^{+7.6}_{-3.4}$	20.9 $^{+8.4}_{-4.2}$	18.8 $^{+8.3}_{-3.8}$	21.1 $^{+8.7}_{-3.5}$	21.5 $^{+11.0}_{-5.9}$	24.7 $^{+13.5}_{-12.1}$	20.5 $^{+8.2}_{-4.1}$
$\langle \Delta v \rangle$ [km s $^{-1}$ ]	ISM	–5 $^{+38}_{-47}$	6 $^{+32}_{-32}$	–14 $^{+42}_{-52}$	15 $^{+44}_{-52}$	28 $^{+47}_{-38}$	4 $^{+42}_{-56}$	–39 $^{+37}_{-48}$	–19 $^{+25}_{-34}$	–56 $^{+43}_{-44}$
$\sigma_{\Delta v}$ [km s $^{-1}$ ]	ISM	215 $^{+24}_{-25}$	206 $^{+17}_{-28}$	222 $^{+26}_{-25}$	208 $^{+25}_{-23}$	200 $^{+19}_{-20}$	215 $^{+27}_{-25}$	225 $^{+27}_{-30}$	208 $^{+27}_{-38}$	237 $^{+22}_{-26}$
$\text{EW}_{\text{Ly}\alpha}$ [Å]	ISM	68.8 $^{+47.4}_{-25.3}$	49.1 $^{+20.3}_{-12.8}$	83.2 $^{+69.6}_{-23.3}$	40.7 $^{+56.2}_{-19.0}$	25.1 $^{+22.8}_{-8.4}$	52.6 $^{+84.4}_{-20.9}$	22.6 $^{+9.1}_{-6.9}$	21.0 $^{+5.9}_{-4.4}$	24.8 $^{+10.1}_{-10.2}$

**Table 3.** Time-averaged Ly $\alpha$  properties after accounting for IGM transmission (IGM) with a similar layout as Tables 1 and 2.

Quantity	Type	All Ly $\alpha$ photons			Ly $\alpha$ from recombinations			Ly $\alpha$ from collisional excitation		
		$z = 5-7$	$z = 6-7$	$z = 5-6$	$z = 5-7$	$z = 6-7$	$z = 5-6$	$z = 5-7$	$z = 6-7$	$z = 5-6$
$\log L_\alpha$ [erg s $^{-1}$ ]	IGM	41.5 $^{+0.6}_{-0.4}$	41.2 $^{+0.4}_{-0.2}$	41.6 $^{+0.5}_{-0.4}$	41.2 $^{+0.7}_{-0.4}$	41.0 $^{+0.5}_{-0.2}$	41.4 $^{+0.7}_{-0.5}$	40.9 $^{+0.4}_{-0.2}$	40.8 $^{+0.4}_{-0.2}$	41.0 $^{+0.3}_{-0.3}$
$L_\alpha^{\text{rec,col}}/L_\alpha$ [%]	IGM	–	–	–	65.9 $^{+22.0}_{-13.5}$	58.4 $^{+19.1}_{-7.6}$	70.4 $^{+21.9}_{-15.0}$	34.1 $^{+13.5}_{-22.0}$	41.6 $^{+7.6}_{-19.1}$	29.6 $^{+15.0}_{-21.9}$
$f_{\text{esc}}^{\text{Ly}\alpha}$ [%]	IGM	25.2 $^{+10.8}_{-9.3}$	25.6 $^{+8.3}_{-7.3}$	24.7 $^{+12.9}_{-11.0}$	24.2 $^{+12.1}_{-9.9}$	25.0 $^{+9.9}_{-8.1}$	23.5 $^{+13.8}_{-11.3}$	26.8 $^{+9.0}_{-7.5}$	25.7 $^{+7.4}_{-5.6}$	27.8 $^{+10.0}_{-9.2}$
$\tau_{\text{IGM}}$ [%]	IGM	35.0 $^{+10.9}_{-7.4}$	30.9 $^{+6.4}_{-4.0}$	39.9 $^{+8.5}_{-11.0}$	38.4 $^{+11.0}_{-8.8}$	34.5 $^{+8.1}_{-5.4}$	42.5 $^{+9.0}_{-12.1}$	29.8 $^{+9.8}_{-5.3}$	27.2 $^{+3.8}_{-3.3}$	34.7 $^{+6.8}_{-9.3}$
$F_{\text{LOS}}/F_\Omega$	IGM	1 $^{+0.31}_{-0.35}$	1 $^{+0.28}_{-0.29}$	1 $^{+0.34}_{-0.40}$	1 $^{+0.35}_{-0.39}$	1 $^{+0.32}_{-0.35}$	1 $^{+0.38}_{-0.43}$	1 $^{+0.26}_{-0.27}$	1 $^{+0.25}_{-0.23}$	1 $^{+0.27}_{-0.30}$
$\log (F_{\text{red}}/F_{\text{blue}})$	IGM	3.5 $^{+1.9}_{-2.1}$	4.8 $^{+8.9}_{-0.7}$	2.1 $^{+1.4}_{-0.9}$	3.6 $^{+2.0}_{-2.1}$	4.8 $^{+9.2}_{-0.7}$	2.3 $^{+1.4}_{-1.0}$	3.4 $^{+1.8}_{-2.1}$	4.7 $^{+8.7}_{-0.8}$	1.9 $^{+1.6}_{-0.9}$
$R_{1/2}$ [kpc]	IGM	6.0 $^{+5.1}_{-3.1}$	8.8 $^{+4.3}_{-4.2}$	5.2 $^{+3.6}_{-2.8}$	5.6 $^{+3.8}_{-2.9}$	7.5 $^{+3.1}_{-3.4}$	4.8 $^{+3.2}_{-2.5}$	7.5 $^{+6.1}_{-3.1}$	10.7 $^{+5.6}_{-4.7}$	6.4 $^{+4.3}_{-2.9}$
$R_{\alpha, \text{h}}$ [kpc]	IGM	17.5 $^{+10.8}_{-4.6}$	17.5 $^{+11.6}_{-8.5}$	17.5 $^{+10.1}_{-3.3}$	15.7 $^{+10.8}_{-5.2}$	14.1 $^{+10.9}_{-6.7}$	16.3 $^{+11.2}_{-3.8}$	17.6 $^{+12.5}_{-5.6}$	19.6 $^{+16.1}_{-11.8}$	16.9 $^{+9.8}_{-4.0}$
$\langle \Delta v \rangle$ [km s $^{-1}$ ]	IGM	202 $^{+30}_{-35}$	210 $^{+40}_{-15}$	194 $^{+28}_{-41}$	205 $^{+33}_{-33}$	216 $^{+41}_{-17}$	195 $^{+29}_{-36}$	197 $^{+30}_{-42}$	201 $^{+37}_{-16}$	194 $^{+27}_{-57}$
$\sigma_{\Delta v}$ [km s $^{-1}$ ]	IGM	127 $^{+27}_{-19}$	115 $^{+26}_{-14}$	133 $^{+34}_{-18}$	128 $^{+27}_{-18}$	118 $^{+28}_{-13}$	133 $^{+31}_{-16}$	123 $^{+38}_{-22}$	107 $^{+22}_{-15}$	135 $^{+49}_{-25}$
$\Delta v_{\text{peak}}$ [km s $^{-1}$ ]	IGM	132 $^{+33}_{-32}$	135 $^{+21}_{-16}$	126 $^{+47}_{-36}$	132 $^{+36}_{-36}$	136 $^{+29}_{-18}$	124 $^{+49}_{-38}$	135 $^{+33}_{-28}$	135 $^{+20}_{-15}$	136 $^{+44}_{-39}$
FWHM [km s $^{-1}$ ]	IGM	158 $^{+34}_{-29}$	146 $^{+30}_{-20}$	166 $^{+33}_{-32}$	150 $^{+35}_{-29}$	145 $^{+30}_{-23}$	155 $^{+36}_{-34}$	160 $^{+37}_{-32}$	137 $^{+33}_{-21}$	173 $^{+35}_{-27}$
$\text{EW}_{\text{Ly}\alpha}$ [Å]	IGM	24.7 $^{+22.1}_{-11.9}$	15.1 $^{+10.2}_{-4.8}$	32.8 $^{+30.6}_{-12.8}$	16.0 $^{+22.6}_{-8.8}$	8.6 $^{+9.9}_{-3.1}$	21.9 $^{+34.6}_{-10.1}$	6.8 $^{+4.1}_{-2.5}$	5.6 $^{+2.1}_{-1.4}$	8.2 $^{+4.4}_{-3.9}$

### 3.3 Equivalent width

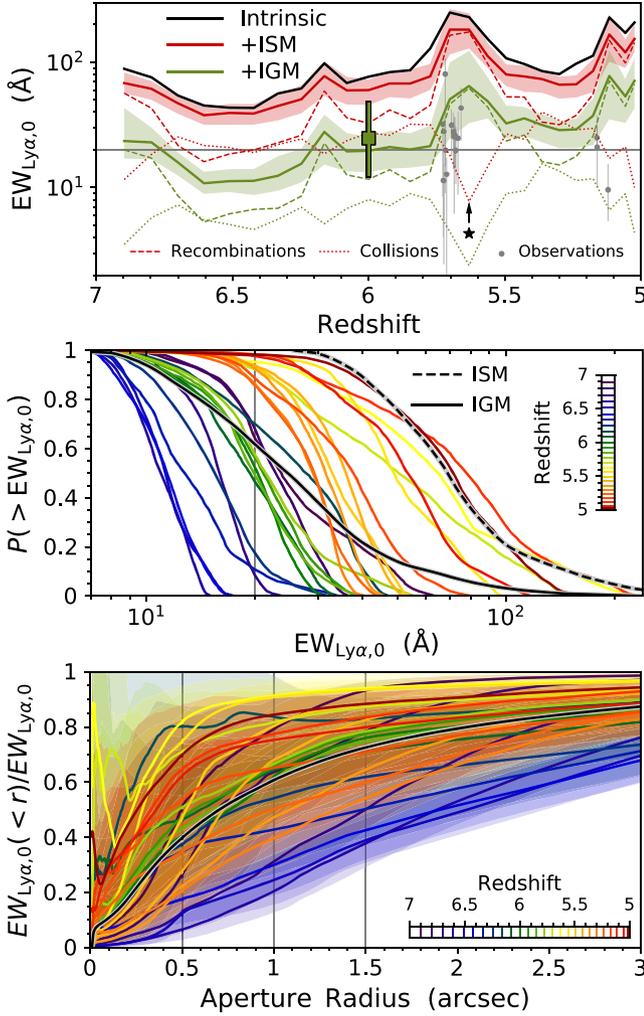
The Ly $\alpha$  equivalent width characterizes the strength of the observed Ly $\alpha$  line relative to the continuum flux, and is defined as:

$$\text{EW}_{\text{Ly}\alpha} \equiv \int \frac{f_{\lambda, \text{Ly}\alpha} - f_{\lambda, \text{UV}}}{f_{\lambda, \text{UV}}} d\lambda \approx \frac{F_{\text{Ly}\alpha}}{f_{\lambda, \text{UV}}}, \quad (5)$$

where the bolometric Ly $\alpha$  flux is  $F_{\text{Ly}\alpha} \equiv \int f_{\lambda, \text{Ly}\alpha} d\lambda$ , and in the rest-frame  $\text{EW}_{\text{Ly}\alpha,0} = \text{EW}_{\text{Ly}\alpha}/(1+z)$ . We calculate an intrinsic time-averaged value of  $\langle \text{EW}_{\text{Ly}\alpha,0} \rangle_{\text{INT}} \approx 110 \text{ \AA}$  (with no dust extinction). The median LOS value decreases to  $\langle \text{EW}_{\text{Ly}\alpha,0} \rangle_{\text{ISM}} \approx 70 \text{ \AA}$  after ISM scattering (with dust) and  $\langle \text{EW}_{\text{Ly}\alpha,0} \rangle_{\text{IGM}} \approx 25 \text{ \AA}$  after IGM transmission (assuming no effect for continuum photons). In Fig. 8, we show that the redshift evolution of  $\text{EW}_{\text{Ly}\alpha,0}$  is primarily driven by the star formation activity in our simulations. We also show several representative observations from the study of Mallery et al. (2012) as grey data points, using selection criteria of  $z > 5$  and  $\log(M_*/M_\odot) < 9.7$  for a similar mass range as our galaxy. In the centre panel, we provide the cumulative probability distributions for observing an equivalent width greater than a given value. During

and after the main starburst, we identify a non-negligible number of high equivalent-width sightlines ( $\text{EW}_{\text{Ly}\alpha,0} \gtrsim 100 \text{ \AA}$ ), even after accounting for IGM transmission (cf. Hashimoto et al. 2017). Finally, in the bottom panel, we show the impact of aperture size on the observed equivalent width, which can be greatly reduced within the central  $\sim 1''$  regions (Laursen et al. 2018). However, the fraction  $\text{EW}_{\text{Ly}\alpha,0}(< r)/\text{EW}_{\text{Ly}\alpha,0}$  converges fairly quickly when the source is brightest due to the corresponding reduced Ly $\alpha$  half-light radius.

Our galaxy has a moderate virial mass  $M_{\text{halo}} < 10^{11} M_\odot$ , such that the star formation efficiency is regulated by strong feedback, resulting in  $M_*/M_{\text{gas}} < 5$  per cent. The dominant starburst corresponds to  $z \approx 5.6$ , although the luminosity is also quite strong at  $z \approx 7$  and  $z \approx 5$  (see Fig. 7). Specifically, the galaxy has duty cycles for the IGM transmitted (intrinsic) equivalent width with  $\text{EW}_{\text{Ly}\alpha,0} > 20 \text{ \AA}$  of 62 per cent (100 per cent) and the luminosity with  $L_\alpha > 10^{42} \text{ erg s}^{-1}$  of 17 per cent (52 per cent). The equivalent width distribution for more massive galaxies may be driven by different mechanisms. For example, star forming regions enshrouded by dusty gas could disproportionately attenuate UV



**Figure 8.** *Top:* The redshift evolution of the Ly $\alpha$  rest-frame equivalent width for the target galaxy. The black curve is the intrinsic value, while the red and green curves are for median sightlines after ISM scattering and IGM transmission, respectively. The shaded regions show the  $1\sigma$  confidence levels considering different viewing angles. The grey data points represent observed values from Mallery et al. (2012) selected with  $z > 5$  and  $\log(M_*/M_\odot) < 9.7$ . *Centre:* The cumulative probability for observing an equivalent width greater than a given value based on viewing angle. The solid (dashed) black curve is the time-weighted distribution after (before) IGM transmission. *Bottom:* The impact of aperture size on the observed equivalent width, which is greatly reduced in the central  $\sim 1''$  regions.

continuum photons leading to an equivalent width enhancement (Finkelstein et al. 2008, 2009), especially if the Ly $\alpha$  resonant scattering process avoids such pathways (Neufeld 1991; Hansen & Oh 2006). Furthermore, Ly $\alpha$  photons originate from more diffuse emission and subsequent scattering in the CGM leads to relatively isotropic low surface brightness Ly $\alpha$  haloes.

Finally, the equivalent widths due to recombination and collisional excitation are anti-correlated. During a starburst the cooling radiation only contributes a small fraction of the total budget, while there is a delayed boost to around  $\sim 50$  per cent afterwards (compare  $L_{\alpha}^{\text{col}}/L_{\alpha}$  over  $z \approx 5.6$ – $5.3$  in Fig. 7). The time variability is mainly regulated by young, massive stars that dominate the production of ionizing photons. However, the deposition of energy by post-starburst galactic winds also enhances the collisional contribution, as predicted for UV metal lines in the CGM (Sravan et al. 2016).

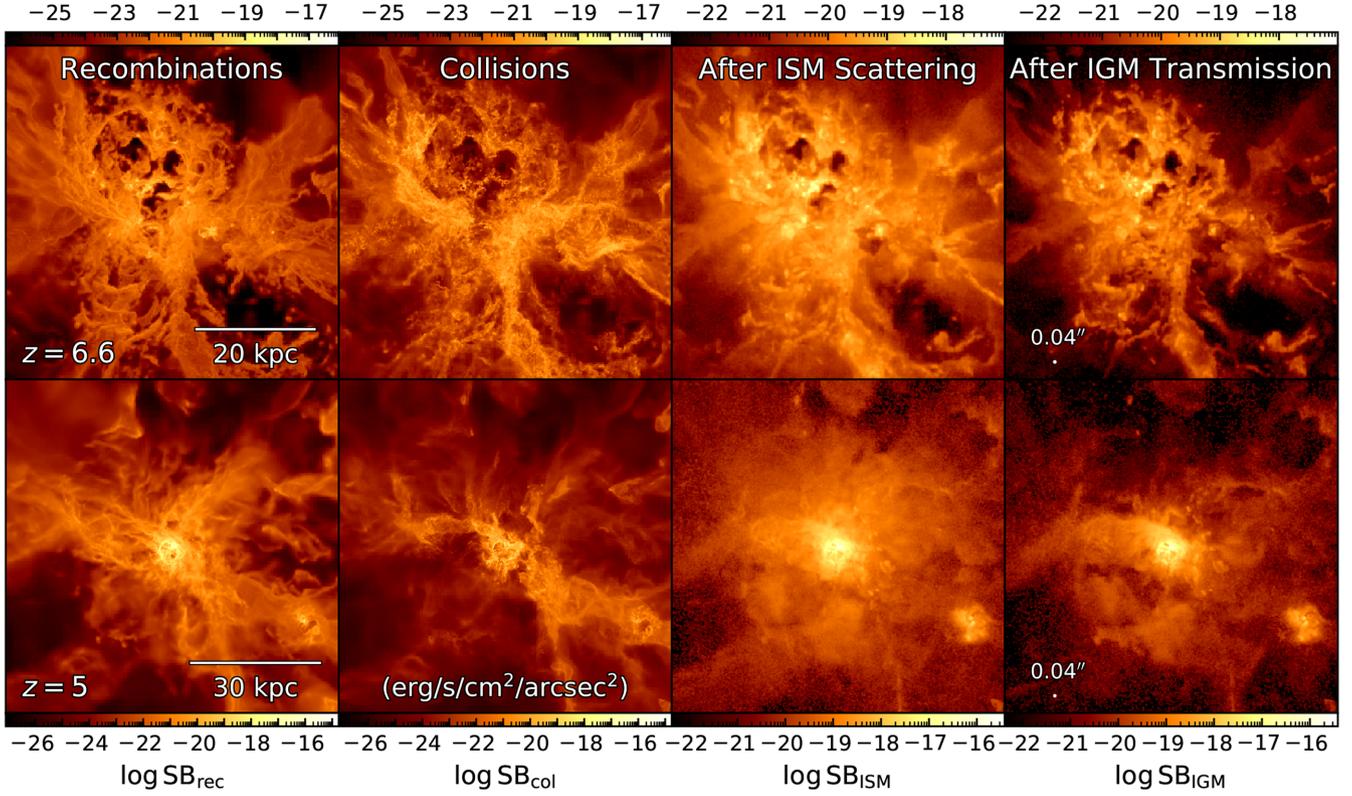
Therefore, it is possible that a subset of LAEs is powered by cooling radiation rather than nebular emission, although this may require a different environment, halo mass, assembly history, or redshift than the galaxy in these simulations.

### 3.4 Surface brightness

We employ the next-event estimator method in COLT to construct LOS surface brightness images for the target galaxy. The simulated images are composed of square pixels at the resolution of the *JWST* NIRCam instrument corresponding to  $\Omega_{\text{pix,NIRCam}} \approx 10^{-3}$  arcsec $^2$ . When also considering the full surface brightness spectral density we employ a Doppler resolution of  $\Delta v \approx 10$  km s $^{-1}$ , corresponding to a spectral resolution of  $R \equiv \lambda/\Delta\lambda \approx 30\,000$ . This is achievable with large-aperture ground-based telescopes equipped with adaptive optics, but significantly exceeds the capability of the *JWST* NIRSpec instrument. Thus, our resolution is sufficient for comparison with both photometric and spectroscopic observations.

In Fig. 9, we illustrate the different stages of Ly $\alpha$  escape from high- $z$  galaxies with spatially resolved images of the (i) intrinsic Ly $\alpha$  recombination and collisional luminosity, (ii) escaped Ly $\alpha$  radiation after scattering through the ISM, and (iii) observed surface brightness after transmission through the IGM. The morphology changes in a nontrivial manner at each stage, with the final surface brightness being much more diffuse and fainter than the intrinsic emission due to the resonant scattering process. For some viewing angles there are dark fluffy streaks illustrating obscuration from high neutral hydrogen column density clouds blocking particular sightlines. Such inhomogeneities also provide preferred channels of escape. To quantify the net effect of Ly $\alpha$  directional boosting, we calculate the time-averaged deviation from isotropy as  $\langle F_{\text{LOS}}/F_{\Omega} \rangle_{\text{ISM}} \approx 1 \pm 0.25$  after ISM scattering and  $\langle F_{\text{LOS}}/F_{\Omega} \rangle_{\text{IGM}} \approx 1 \pm 0.33$  after IGM transmission, where  $F_{\text{LOS}}$  and  $F_{\Omega}$  denote the spatially integrated LOS and angular-averaged flux, respectively. We also note that there is a slight increase in anisotropy with time as the gas structure becomes more defined and the CGM is more ionized (Tables 2 and 3).

In Fig. 10, we show the median radial surface brightness after IGM transmission for each snapshot coloured according to redshift. The shaded regions represent the  $1\sigma$  confidence levels based on the variation between different sightlines for the 3072 healpix directions. The radial profiles illustrate the cuspy nature of the central Ly $\alpha$  source and the transition to an exponentially damped halo. We fit the logarithmic slopes of the haloes as  $\text{SB}(r) \propto \exp(-r/R_{\alpha,h})$  beyond 10 kpc, plotting the characteristic scale lengths of the Ly $\alpha$  haloes  $R_{\alpha,h}$  both after ISM scattering and IGM transmission as an inset within the figure. In Fig. 10 we also present the median normalized integrated light within a given radius, which may be thought of as  $P(<r) \propto \int_0^r \text{SB}(r')r'dr'$ . We employ linear radial bins with  $\sim 1$  kpc resolution for  $\text{SB}(r)$  and higher resolution logarithmic radial bins to calculate the cumulative distribution function  $P(<r)$ . As with the radial surface brightness profiles we also provide the  $1\sigma$  confidence levels based on viewing angle uncertainties. We also show the redshift evolution of the half-light radius, defined as  $P(<R_{1/2}) = 1/2$  normalized such that  $P(<2R_{\text{vir}}) = 1$ , both after ISM scattering and IGM transmission as an inset within the figure. For reference, the Ly $\alpha$  halo scale length is comparable to the virial radius, i.e.  $R_{\alpha,h} \gtrsim R_{\text{vir}}$ , and the half-light radius is within the radius of maximal circular velocity, i.e.  $R_{1/2} \lesssim R_{\text{max}}$ . Specifically, the halo slopes are steeper after IGM transmission with time-averaged values of  $\langle R_{\alpha,h} \rangle_{\text{ISM}} \approx 21.4$  kpc and  $\langle R_{\alpha,h} \rangle_{\text{IGM}} \approx 17.5$  kpc, also displaying slight flattening with redshift. Likewise, the half-



**Figure 9.** The Ly $\alpha$  surface brightness at three stages of the escape process: intrinsic recombination and collisional emission, total escaped radiation after ISM scattering, and observed surface brightness after IGM transmission. This illustrates the impact of resonant scattering as the final images are noticeably fainter and more diffuse than the intrinsic emission. The images are shown at  $z = 6.6$  (top) and  $z = 5$  (bottom), to compare the evolving morphologies. The panels for ISM scattering and IGM transmission use the same colour scale.

light radius decreases after IGM transmission with time-averaged values of  $\langle R_{1/2} \rangle_{\text{ISM}} \approx 7.2$  kpc and  $\langle R_{1/2} \rangle_{\text{IGM}} \approx 6.0$  kpc, becoming increasingly concentrated with redshift (see Tables 2 and 3). The Ly $\alpha$  haloes are significantly more extended than the UV continuum emission with  $\langle R_{1/2}/R_{1/2,\text{UV}} \rangle_{\text{ISM}} \approx 2.7$  and  $\langle R_{1/2}/R_{1/2,\text{UV}} \rangle_{\text{IGM}} \approx 2.2$ . After accounting for the  $(1+z)^{-4}$  dependence of the Ly $\alpha$  surface brightness along with potential variations due to environmental factors and IGM reprocessing, our predicted profiles are consistent with the Ly $\alpha$  haloes generically observed around star-forming galaxies at lower redshifts, e.g.  $R_{\alpha,h} \approx 25$  kpc in Steidel et al. (2011).

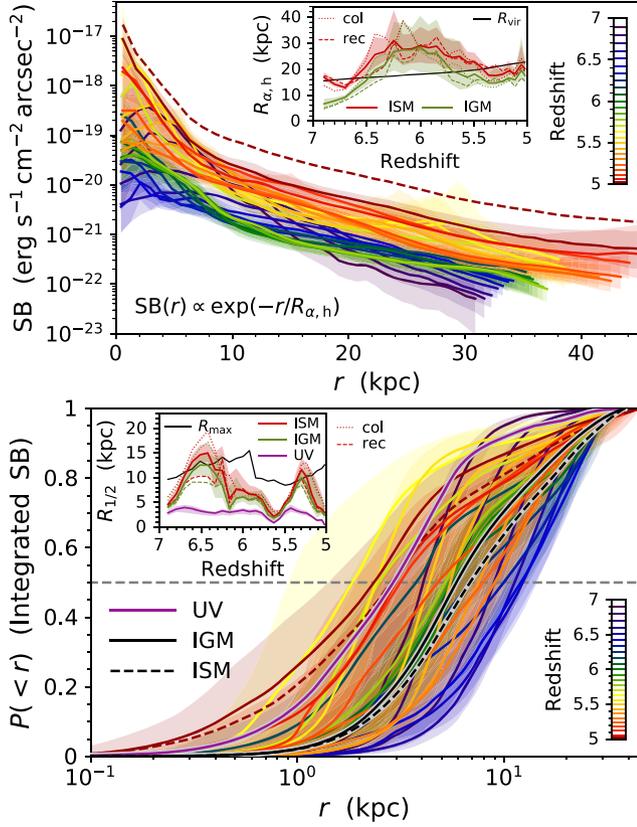
### 3.5 Line flux

The spectral line profile is perhaps the most distinguishing observable feature from LAEs. To summarize the redshift evolution of its characteristic properties, in Fig. 11, we display the median LOS emergent Ly $\alpha$  flux density with  $1\sigma$  confidence regions, as a function of Doppler velocity  $\Delta v = c\Delta\lambda/\lambda$  for each simulation snapshot considering both ISM scattering (dashed) and IGM transmission (solid). The top panel employs a logarithmic scale to show the time-dependent luminosity while the middle panel utilizes a linear scale to illustrate the evolving shape of the normalized lines. We note that individual LOS spectra are generally quite asymmetric with little emission at the systemic velocity. However, even though the line profiles exhibit significant diversity, including single, double, and triple peaked morphologies, the median profile washes out some of these features, especially if there is a moderate proper motion from the cosmological environment. The bottom panel

shows the cumulative distribution function of each profile. To further condense the information, we also include the redshift dependence of the peak velocity offset  $\Delta v_{\text{peak}}$  and full width of the observed line at half of the maximum (FWHM), both given as median LOS quantities with viewing angle uncertainties. Specifically, after IGM reprocessing, these quantities are  $\langle \Delta v_{\text{peak}} \rangle_{\text{IGM}} \approx 132$  km s $^{-1}$  and  $\langle \text{FWHM} \rangle_{\text{IGM}} \approx 158$  km s $^{-1}$ . We also consider the redshift dependence of the flux-weighted frequency centroid  $\langle \Delta v \rangle \equiv \int \Delta v f_{\lambda} d\lambda / \int f_{\lambda} d\lambda$  and standard deviation of the line  $\sigma_{\Delta v} \equiv (\langle \Delta v^2 \rangle - \langle \Delta v \rangle^2)^{1/2}$ . The time-averages after escaping the galaxy are  $\langle \langle \Delta v \rangle \rangle_{\text{ISM}} \approx -5$  km s $^{-1}$  and  $\langle \sigma_{\Delta v} \rangle_{\text{ISM}} \approx 215$  km s $^{-1}$ , or  $\langle \langle \Delta v \rangle \rangle_{\text{IGM}} \approx 202$  km s $^{-1}$  and  $\langle \sigma_{\Delta v} \rangle_{\text{IGM}} \approx 126$  km s $^{-1}$  after IGM transmission. We note that there is some evolution towards broader and less offset lines with decreasing redshift, mostly due to the changing IGM transmission (see Tables 2 and 3).

#### 3.5.1 Red and blue flux ratio

The ratio of red and blue fluxes  $F_{\text{red}}/F_{\text{blue}}$  can potentially act as a diagnostic for the inflow/outflow kinematics of galaxies, particularly at lower redshifts where the blue peak is less affected by IGM transmission. Therefore, we focus our discussion on the red-to-blue flux ratio of Ly $\alpha$  photons immediately escaping from the target galaxy. In Fig. 12, we show the median LOS values as a function of radius and redshift. Interestingly, there is an apparent evolution towards lower red-to-blue flux ratios with time, due to the decaying influence of feedback between star forming episodes. In fact, prior to the starburst at  $z \approx 5.6$ , the infall feature shows

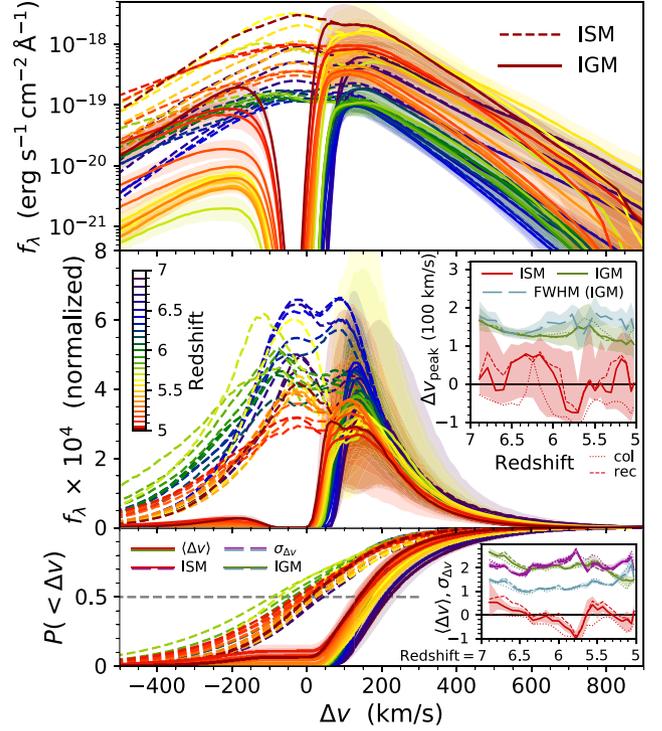


**Figure 10.** *Top:* Radial Ly $\alpha$  surface brightness profiles after transmission through the IGM. The curves represent median LOS values with  $1\sigma$  confidence regions, all coloured according to redshift. The inset shows the evolution of the Ly $\alpha$  halo scale length  $R_{\alpha,h}$ , which is calculated by fitting to an exponential function and is comparable to the virial radius  $R_{\text{vir}}$ . *Bottom:* The integrated light within a given radius,  $P(< r) \propto \int_0^r SB(r')r'dr'$ . The inset shows the half-light radius, defined as  $P(< R_{1/2}) = 1/2$ , and is within the radius of maximal circular velocity  $R_{\text{max}}$ .

significant blue dominance, changing to red dominance as radiative feedback and supernovae launch new galactic winds. To highlight this transition in Fig. 12, we illustrate the spatial distribution of the red-to-blue-flux ratio before and after the starburst, providing  $1\sigma$  and  $2\sigma$  (68 percent and 95 percent) Ly $\alpha$  surface brightness contours for reference. For presentation purposes, the red and blue channel maps have each been smoothed to simulate a FWHM  $\approx 0.3$  aperture prior to taking the ratio. The more extended collisional cooling emission almost always exhibits a blue signature (inflow), while the more centralized recombination emission exhibits a red signature (outflow). Specifically, we calculate  $\langle F_{\text{red}}/F_{\text{blue}} \rangle_{\text{ISM}} \approx 1$  with  $\langle F_{\text{red}}/F_{\text{blue}} \rangle_{\text{ISM,rec}} \approx 1.2$  and  $\langle F_{\text{red}}/F_{\text{blue}} \rangle_{\text{ISM,col}} \approx 0.8$ , explaining the lower transmission of collisional emission through the IGM (see Table 2). These results are consistent with an observation by Erb et al. (2018) of a low mass ( $M_* \approx 5 \times 10^8 M_\odot$ ), low metallicity ( $Z \approx 0.25 Z_\odot$ ) star-forming galaxy at  $z = 2.3$ , which exhibits a dominant red peak in the centre with  $F_{\text{red}} \approx F_{\text{blue}}$  in the outskirts of the Ly $\alpha$  halo.

### 3.5.2 Moment maps

Images of the frequency moments before IGM transmission provide additional insight beyond the spatially resolved ratio of red and blue fluxes. In Fig. 13 we show the flux-weighted frequency centroid

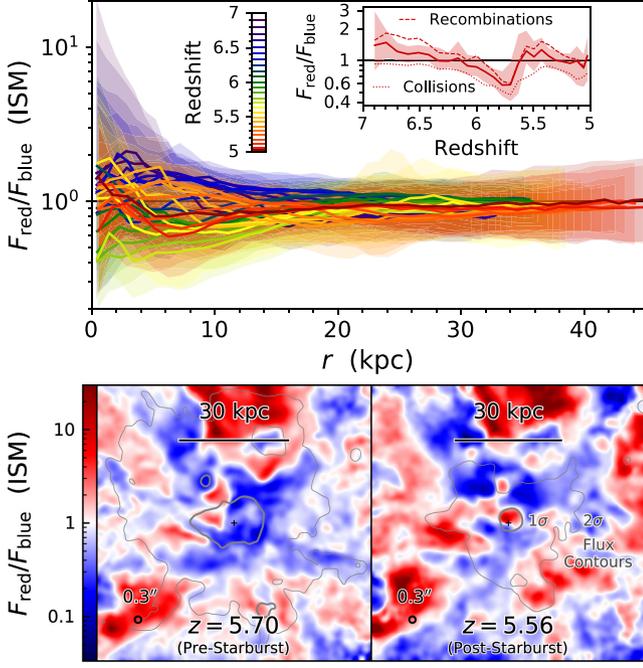


**Figure 11.** Median LOS Ly $\alpha$  flux density as a function of Doppler velocity  $\Delta v = c\Delta\lambda/\lambda$  for each simulation snapshot after ISM scattering and IGM transmission. The top panel shows the fluctuating flux while the middle panel illustrates the evolving shape of the (IGM) normalized lines. The bottom panel shows the corresponding cumulative distribution functions. We also include insets showing the redshift dependence of the peak velocity offset  $\Delta v_{\text{peak}}$ , full width at half maximum (FWHM), frequency centroid ( $\Delta v$ ), and standard deviation  $\sigma_{\Delta v}$ , all given as median LOS quantities.

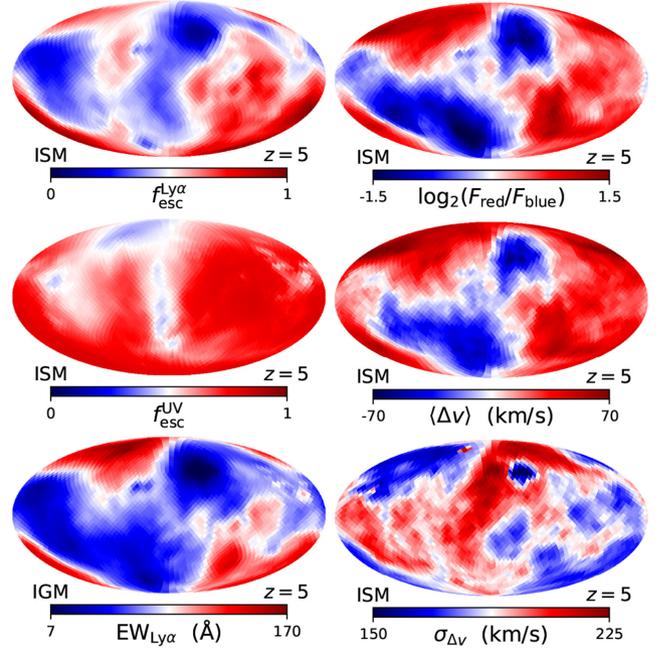
$\langle \Delta v \rangle$  and standard deviation  $\sigma_{\Delta v}$  of the escaped Ly $\alpha$  radiation at  $z = 6.6$  (top) and  $z = 5$  (bottom), for direct comparison with many of the other figures in this paper. There are substantial morphological differences between the two redshifts, e.g. the galaxy is more disturbed at  $z = 6.6$  and more compact and structured at  $z = 5$ . However, at all redshifts the moment maps exhibit evidence for effects due to CGM scale bulk rotation with large segmented red and blue regions, radial inflows and outflows as concentrated regions with particularly strong red or blue flux, and significant line broadening from resonant scattering in the optically thick regions. There is a very close correspondence between the LOS gas velocity and the frequency centroid, and likewise between the gas density and the line width, so we include mass-weighted projections of each of these quantities alongside the moment maps. For reference, in Fig. 13 we include  $1\sigma$  and  $2\sigma$  Ly $\alpha$  surface brightness contours in each panel to provide a sense for the size and value of the region dominating the spatially-integrated flux. We also illustrate the perpendicular gas flow with streamlines in the leftmost panel. The arrows indicate the directionality and the thickness denotes the relative magnitude of the velocity, i.e. as being proportional to  $v_{\text{gas},\perp}$ .

### 3.5.3 Angular variations

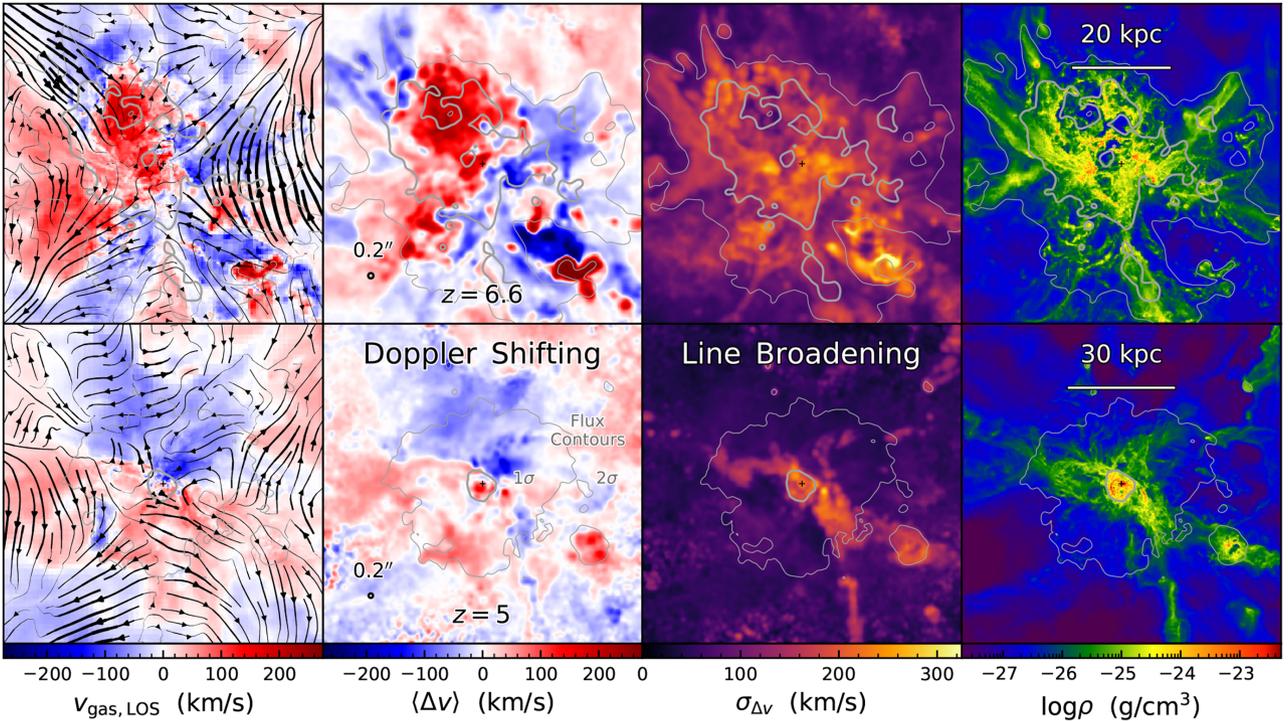
We now consider the angular variations of the spatially integrated flux for all lines of sight. In Fig. 14, we show the escape fraction of Ly $\alpha$  photons and other quantities in each of the 3072 healpix



**Figure 12.** *Top:* Radial red-to-blue-flux ratio  $F_{\text{red}}/F_{\text{blue}}$  before transmission through the ISM. The curves are median LOS values with  $1\sigma$  confidence regions, all coloured according to redshift. The inset shows the redshift evolution of the spatially-integrated flux ratio. *Bottom:* An illustration of the extreme change from a blue (infall) to red (outflow) dominated signature within the  $1\sigma$  Ly $\alpha$  surface brightness contours before and after the main starburst. For presentation purposes, the red and blue channel maps have been smoothed to simulate a FWHM  $\approx 0''.3$  aperture.



**Figure 14.** The angular distributions of various Ly $\alpha$  properties after scattering through the ISM, including  $f_{\text{esc}}^{\text{Ly}\alpha}$ ,  $f_{\text{esc}}^{\text{UV}}$ ,  $\text{EW}_{\text{Ly}\alpha,0}$  (after IGM transmission),  $F_{\text{red}}/F_{\text{blue}}$ ,  $\langle \Delta v \rangle$ , and  $\sigma_{\Delta v}$ . There are hints of connections between galaxy kinematics, dust absorption, and line broadening due to resonant scattering. In particular, the highest rest-frame equivalent widths also correspond to outflowing regions with narrower (or single peaked) lines and greater UV continuum attenuation.



**Figure 13.** *From left to right:* Spatial images of the LOS gas velocity, flux-weighted frequency centroid  $\langle \Delta v \rangle$ , standard deviation  $\sigma_{\Delta v}$  of the escaped Ly $\alpha$  radiation, and density at  $z = 6.6$  (top) and  $z = 5$  (bottom), before transmission through the IGM. The moment maps reveal hints of effects due to CGM scale bulk rotation, radial inflows and outflows, and significant line broadening from resonant scattering in the optically thick regions. For reference, we include  $1\sigma$  and  $2\sigma$  Ly $\alpha$  surface brightness contours in each panel. Likewise, the streamlines in the leftmost panel illustrate the directionality of the perpendicular flow, with the thickness denoting the relative magnitude, i.e.  $\propto v_{\text{gas}, \perp}$ .

directional bins. There are hints of correlations between the galaxy kinematics, dust absorption, and line broadening due to resonant scattering. Although the detailed radiative transfer is far from trivial, it is not surprising that the highest rest-frame equivalent widths also correspond to outflowing regions with a narrow dominant red peak and greater UV continuum attenuation. These regions are natural consequences of strong radiative feedback, which more easily affects the gas in low density channels.

For the most part, significant differences are mainly apparent on large angular scales. Heuristically, we can associate each multipole mode  $\ell$  with a typical angular size of  $\theta \sim 180^\circ/\ell$ . The global escape is smoothed out on small angular scales by the multiple scatterings during the transition from optically thick gas to the free streaming regime beyond the CGM. The low  $\ell$  mode fluctuations may also be viewed as the projection of galactic and cosmic web structures. For example, regions with the lowest escape fractions likely correspond to cosmological filaments and to a lesser degree high opacity clouds that produce a shadowing effect. In this sense, we find that the covering factor is quite small, such that the global Ly $\alpha$  escape is fairly isotropic. However, the rest-frame equivalent width can vary on much smaller angular scales as UV continuum photons do not undergo the same multiple scattering.

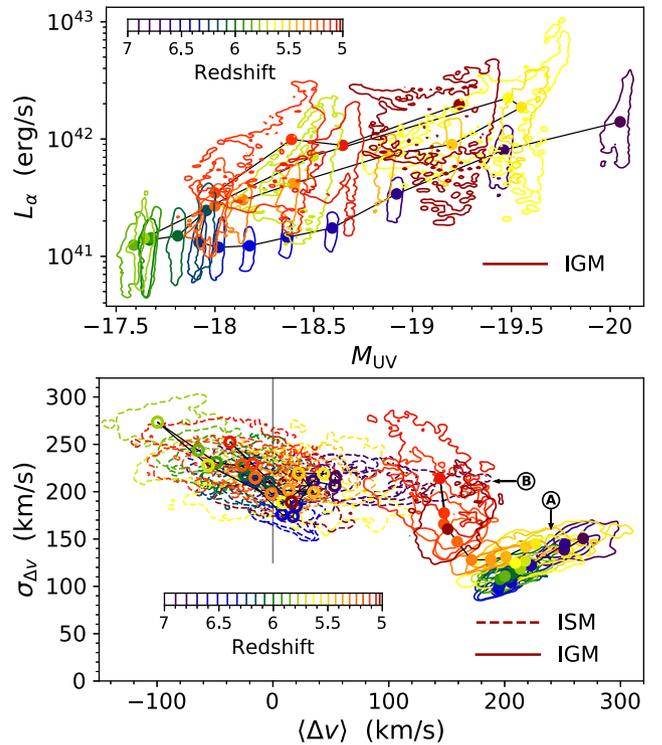
See Appendix A1 for an analysis of the angular power spectra of these quantities. Likewise, see Appendix A2 for similar insights gained by considering the continuous dependence of Ly $\alpha$  properties as a function of rotation angle for equatorial sightlines.

## 4 OBSERVABILITY

### 4.1 Correlations between observables

The relationships between the various observables considered in this paper are highly nontrivial, and would likely be most meaningful for large samples of simulated and observed LAEs covering a wide range of possible masses, environments, and star formation histories. Therefore, we simply attempt to validate our intuition regarding the redshift evolution and uncertainties from our target galaxy. In particular, in Fig. 15, we show the median values with  $1\sigma$  LOS contours for the observed Ly $\alpha$  luminosity  $L_\alpha$  and UV continuum absolute magnitude  $M_{UV}$ . This illustrates the roles of both viewing angle and starburst activity in the Ly $\alpha$  and UV continuum visibility. There are also hints of a positive correlation between different sightlines at each redshift, however, at low metallicity the spread in  $f_{esc}^{Ly\alpha}$  dominates over the spread in  $f_{esc}^{UV}$ . Still, the median observed  $L_\alpha$  can vary by almost an order of magnitude between different redshifts with the same  $M_{UV}$ . Thus, due to the efficient production of Ly $\alpha$  photons from young massive stars, there is an overall increase in the equivalent width while ramping up to the peak brightness of a starburst and a corresponding decrease in the aftermath with the aging of the star populations (see also Fig. 8).

In Fig. 15, we also show a potential relationship between the line width standard deviation  $\sigma_{\Delta\nu}$  and frequency centroid  $\langle\Delta\nu\rangle$ , which could help constrain models of galaxy kinematics and IGM transmission. However, because our IGM model is statistical we warn against over interpretation as the most robust aspect is that IGM transmission reprocesses the Ly $\alpha$  line to be fainter, redder, and narrower. Still, there is an intriguing correlation between the IGM transmitted observables with higher velocity offsets also yielding broader lines (see the region marked ‘A’), which follows our intuition from analytical considerations (Neufeld 1990), empirical results from simulations (Schaerer et al. 2011; Zheng & Wallace 2014), and observations of LAEs across a wide redshift range

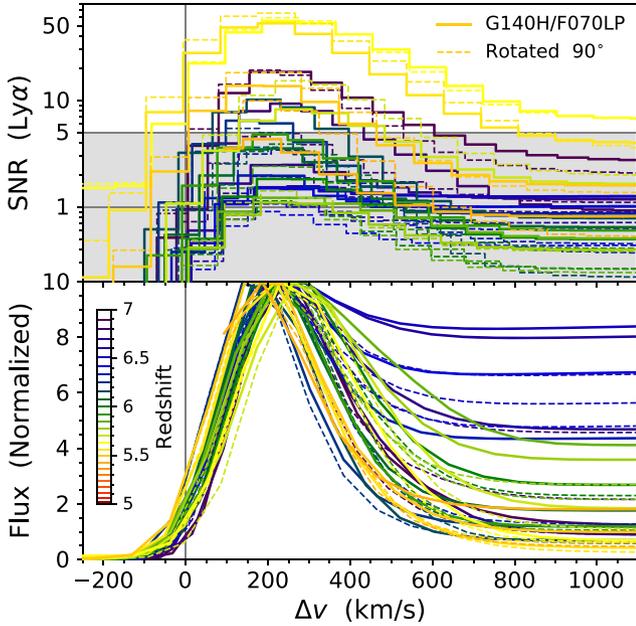


**Figure 15.** The redshift evolution for potentially correlated observables shown as the median LOS values along with  $1\sigma$  contours illustrating the variation across 3072 sightlines. *Top:* The Ly $\alpha$  luminosity  $L_\alpha$  and UV continuum absolute magnitude  $M_{UV}$ , highlighting the roles of both viewing angle and starburst activity in the  $L_\alpha$  and  $M_{UV}$  visibility. *Bottom:* The line width standard deviation  $\sigma_{\Delta\nu}$  and frequency centroid  $\langle\Delta\nu\rangle$ , which could help constrain models of galaxy kinematics and IGM transmission. The dashed (open) and solid (filled) contours (medians) denote values after ISM scattering and IGM transmission, respectively. The regions labeled ‘A’ and ‘B’ are to guide the discussion in the text.

(Verhamme et al. 2018). It seems that for the most realistic radiative transfer models the velocity offset and line width correlation is primarily driven by the H I opacity of the ISM, CGM, and subsequent transmission through the IGM, with secondary effects from galaxy kinematics, dust, and 3D geometry. The upturn in Fig. 15 towards broader lines at lower redshifts is due to a significant increase in the fraction of double peaked spectra (see the region marked ‘B’), in which case one might use the separation of blue and red peaks to infer the systemic redshifts of the observed galaxies.

### 4.2 JWST NIRSpec observations

At high redshifts there is a significant reduction in surface brightness as Ly $\alpha$  photons scatter to larger distances from the galaxy. This has important implications for deciding on optimal Ly $\alpha$  observation strategies. However, one of the preferred observing modes with the upcoming JWST will be NIRSpec multi-object spectroscopy (MOS) using the micro-shutter assembly (MSA), as LAE spectroscopic surveys will undoubtedly benefit from the capability of obtaining simultaneous spectra of multiple targets within each  $3.6' \times 3.4'$  field of view exposure. Therefore, in this section we focus on a three shutter MSA configuration with a size of  $(3 \times 0.46)'' \times 0.2'' = 0.276 \square''$  centered on the UV continuum flux centroid. We note that other modes with larger collecting areas could provide higher signal to noise ratios for individual Ly $\alpha$  follow-up observations, e.g. the

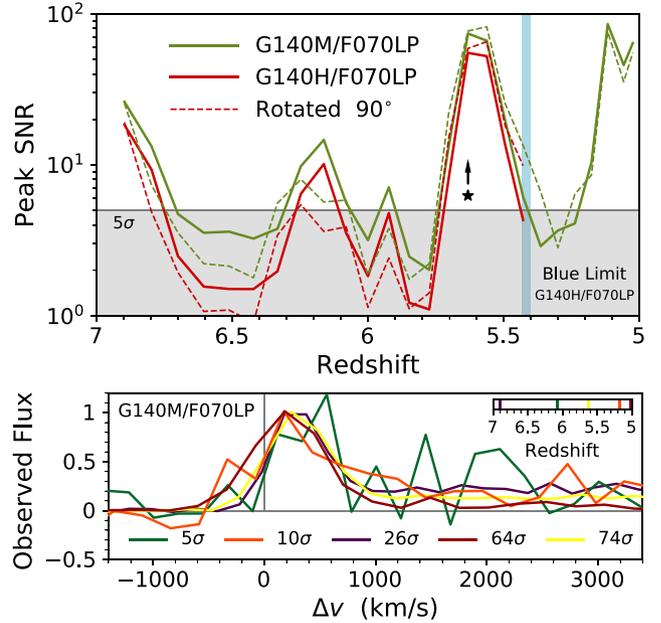


**Figure 16.** *Top:* Spectral signal to noise ratio (SNR) after  $10^4$  seconds of exposure time for the *JWST* NIRSpec three shutter MSA high resolution G140H/F070LP spectral configuration. *Bottom:* For comparison, we also show the flux normalized to the peak value as a function of Doppler velocity  $\Delta v = c\Delta\lambda/\lambda$  for each simulation snapshot (see also Fig. 11). The dashed curves correspond to the same direction but with the slit rotated by  $90^\circ$  to demonstrate the approximate variation due to orientation.

S400A1 and S1600A1 fixed slits have areas that are respectively 5.5 and 9.3 times larger than the three shutter MSA configuration.

In Fig. 16 we illustrate the spectral signal to noise ratio (SNR) and line flux normalized to the peak value for each simulation snapshot, while in Fig. 17 we show the peak SNR as a function of redshift. The curves were calculated with the *JWST* exposure time calculator using the total flux extracted from the MSA region after applying a Gaussian point spread function with  $\text{FWHM} \approx 0''.04$  (Pontoppidan et al. 2016). Each curve corresponds to the same direction and orientation, although we also show the results obtained when the slit is rotated by  $90^\circ$  to demonstrate the approximate variation due to the slit orientation. We find that the SNR and peak locations are fairly robust as this is dominated by the central emission. However, the orientation can induce subtle differences in the UV continuum flux due to the obliquity of stellar orbits, and as seen in Section 3.5 the details of the line profile are affected by the density and kinematics of the observed regions.

The SNR and flux profiles are for the high-resolution G140H/F070LP spectral configuration, which provides a nominal resolving power of  $R \approx 2,700$  over a wavelength range of  $\lambda = 0.78\text{--}1.27\ \mu\text{m}$ . In Fig. 17, we also show the results under the medium resolution G140M/F070LP configuration with  $R \approx 1,000$  down to  $\lambda \approx 0.7\ \mu\text{m}$ . We note that the restricted wavelength coverage implies a blue lower limit for the Ly $\alpha$  line corresponding to  $z \gtrsim 5.4$  (4.8) for the high (medium) resolution mode, although at wavelengths shorter than  $0.9\ \mu\text{m}$  the filter throughput is already below 20 per cent. For our calculations, we have chosen an exposure time of  $10^4$  seconds, which is a realistic depth that nicely demonstrates the source coming in and out of visibility. Unfortunately, it seems that spectroscopic continuum detections for relatively low mass LAEs, such as the one in this study, would require significantly deeper exposures. However, these measurements can

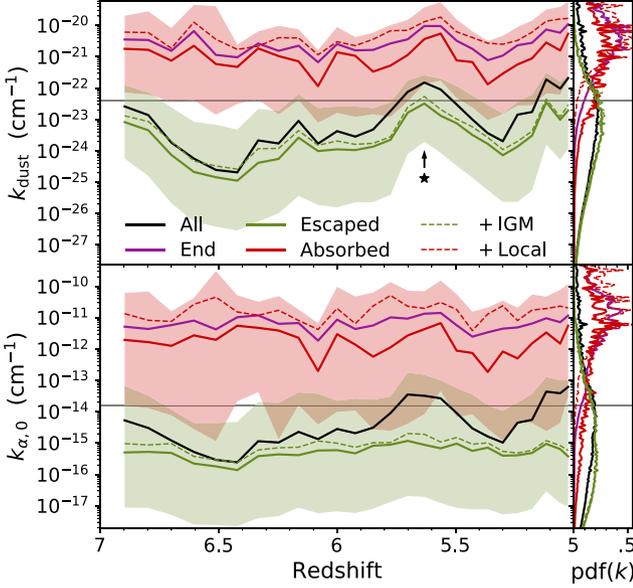


**Figure 17.** *Top:* Redshift evolution of the peak SNR corresponding to Fig. 16, which is closely correlated with the luminosities from Fig. 7. The restricted wavelength coverage of the high-resolution G140H/F070LP configuration implies a blue lower limit of  $z \gtrsim 5.4$  for LAE detections, while the medium resolution G140M/F070LP mode has higher SNRs and allows coverage down to  $z \gtrsim 4.8$ . *Bottom:* We show example observed spectra inclusive of noise, selected to illustrate both high and low peak SNRs.

still be obtained with *JWST* NIRCcam photometry or the low resolution PRISM/CLEAR configuration with  $R \approx 100$ . In any case, with a time-averaged peak SNR of  $18\sigma$  ( $11\sigma$ ) and a  $>5\sigma$  detection duty cycle of 37 per cent (25 per cent) for  $10^4$  seconds of exposure time using the medium (high) resolution configuration, it seems that Ly $\alpha$  detections and spectroscopy of high- $z$  galaxies is quite feasible. The medium (high) resolution configuration only achieves  $\Delta v \approx 300$  ( $100$ )  $\text{km s}^{-1}$ , roughly the width of the IGM transmitted lines (see the spectra with noise in Fig. 17). Ly $\alpha$  surveys at higher spectral resolution are possible with large aperture ground-based telescopes, however, even with next-generation facilities follow-up observations are subject to the night sky background. The high likelihood of encountering a sky emission line typically restricts LAE searches to narrow redshift windows of reduced contamination. Still, diagnostics from other lines and cross-correlation studies will also help unravel the detailed properties of high- $z$  LAEs.

## 5 PREDICTING THE ESCAPE OF INDIVIDUAL PHOTONS

The Monte Carlo procedure also naturally generates a mapping from the local emission of each photon packet to the corresponding outcome, motivating predictive models for the Ly $\alpha$  escape fractions of galaxies in cosmological simulations. The full radiative transfer calculations only need to be carried out for select training datasets and then the computationally inexpensive local model can be applied to a larger number of simulated galaxies to probe the statistical properties of LAEs, although we defer this particular application to a future study.



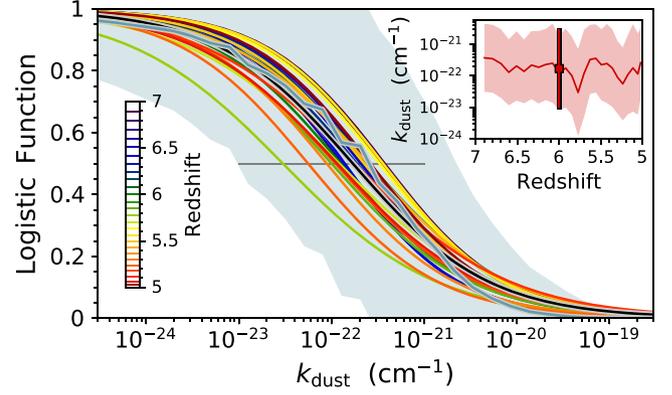
**Figure 18.** The local absorption coefficients for dust  $k_{\text{dust}}$  and Ly $\alpha$   $k_{\alpha,0}$  at line centre are fairly accurate predictors for the Ly $\alpha$  escape fraction. The median curves show the evolution and time-averaged probability distribution functions for the categories described in Section 5 and summarized in Table A1, which also provides statistics on the probability bisect (grey line), overlap, and separation distance for the absorbed and escaped distributions.

### 5.1 Separation between absorbed and escaped distributions

For the analysis in this section we track the initial position and cell index of each photon packet, along with the final states of escaping and absorbed photons. At this point we differentiate between six categories of photon trajectories: (i) ‘All’, to describe the local environments of all emitted photon packets, (ii) ‘Escaped’, to select only escaped photons, (iii) ‘+IGM’, to further weight escaped photons according to the IGM transmission based on the escaping frequency, (iv) ‘Absorbed’, to select only absorbed photons, (v) ‘+Local’, to further distinguish between photons that are absorbed in the same cell from which they were emitted, and (vi) ‘End’, to describe the environments where photons are eventually absorbed. In Fig. 18, we illustrate the properties of each of these categories by considering the local absorption coefficients for dust  $k_{\text{dust}}$  and Ly $\alpha$   $k_{\alpha,0}$  at line centre. We show the redshift evolution of the median values for each category, along with the  $1\sigma$  confidence levels for absorbed (red) and escaped (green) photons. We note that the final absorption environment typically exhibits higher absorption coefficients (purple), especially for photons absorbed “locally” (dashed red). Also, the total distribution represents an escape fraction weighted average between the absorbed and escaped distributions. Finally, we include the time-averaged normalized probability distribution functions for each category within inset right panels. See Appendix A3 for additional details about comparing these probability distributions.

### 5.2 Logistic regression analysis

Based on the apparent separation between the absorbed and escaped distributions, it seems reasonable to employ the local dust content as a predictor of the global Ly $\alpha$  escape fraction. Therefore, we employ a linear logistic regression analysis to find a function representing the conditional probability of escape given the local



**Figure 19.** Logistic functions for the Ly $\alpha$  (ISM) escape fraction at different redshifts for a model based on the local dust absorption coefficient  $k_{\text{dust}}$  as a single predictor. The black curve represents the time-weighted model. The blue curve illustrates the average and standard deviation of the simulated escape fractions after binning according to  $k_{\text{dust}}$ . The inset shows the redshift dependence of the 50 per cent transition region with  $1\sigma$  confidences.

**Table 4.** Time-averaged logistic functions for models with predictor variables in the set  $\{k_{\text{dust}}, k_{\alpha,0}, \rho, T\}$  along with the relative error of the estimates. See equation (6) for the relation to the conditional escape probability and Fig. 19 for an illustration.

Quantity	Logit: $f(\mathbf{x}) = \beta \log x + \beta_0$	$\langle  1 - f_{\text{esc,est}}^{\text{Ly}\alpha} / f_{\text{esc}}^{\text{Ly}\alpha}  \rangle$
$k_{\text{dust}}$ [ $\text{cm}^{-1}$ ]	$-1.370 \log k_{\text{dust}} - 29.82$	4.91%
$k_{\alpha,0}$ [ $\text{cm}^{-1}$ ]	$-0.899 \log k_{\alpha,0} - 11.49$	6.66%
$\rho$ [ $\text{g cm}^{-3}$ ]	$-1.656 \log \rho - 37.00$	5.69%
$T$ [K]	$2.202 \log T - 6.97$	8.99%

emission environment. Specifically, we consider a model with one or more predictor variables in the set  $\mathbf{X} = \{k_{\text{dust}}, k_{\alpha,0}, \rho, T\}$ , with one binary target variable  $Y \in \{0, 1\}$  which is described by a (sigmoid) logistic function,

$$P(Y = 1 | \mathbf{X} = \mathbf{x}) = \sigma(f(\mathbf{x})) = \frac{1}{1 + \exp(-f(\mathbf{x}))} \quad (6)$$

where the logit (log odds) is given by  $f(\mathbf{x}) = \sum \beta_i \log x_i + \beta_0$ . An illustration of this is given in Fig. 19 for a model based on the local dust absorption coefficient  $k_{\text{dust}}$  as a single predictor. We report the time-weighted results obtained from the `scikit-learn` logistic regression python package in Table 4. The logistic function may be used as a classifier by choosing a cutoff value, e.g. the 50 per cent transition region is defined via  $f(\mathbf{x}) > 0$ . Additionally, the global escape fraction may be estimated without expensive radiative transfer calculations by the luminosity-weighted average over the simulation, i.e.  $f_{\text{esc,est}}^{\text{Ly}\alpha} \approx \sum L_{\alpha,j} \sigma(f(\mathbf{x}_j)) / L_{\alpha}$  where  $j$  denotes the cell index. With this simple model, we achieve a time-averaged relative error in the estimated escape fractions of only  $\langle |1 - f_{\text{esc,est}}^{\text{Ly}\alpha} / f_{\text{esc}}^{\text{Ly}\alpha}| \rangle \approx 5$  per cent, although it is unclear how sensitive this result is to galaxy mass, metallicity, or even the resolution and sub-grid models of particular simulations. Still, Ly $\alpha$  escape based on local conditions could be useful as a correlation tool applied to a large number of similar simulations, or for comparison across simulation frameworks. Either way, we caution against simply applying our model to large-scale simulations without careful recalibration.

## 6 LY $\alpha$ RADIATION PRESSURE

### 6.1 Energy density

We employ Monte Carlo estimators based on the traversed path lengths to calculate the Ly $\alpha$  energy density  $U_\alpha$  within each cell (see Smith, Bromm & Loeb 2017a). This allows us to illustrate the internal 3D structure of the resonant scattering process, which is not directly observable. For comparison, in Fig. 20, we show the LOS mass-weighted projections of the energy density for ionizing and Ly $\alpha$  radiation at  $z = 6.6$  (top) and  $z = 5$  (bottom), isolating the contributions due to the recombination and collisional excitation components. The ionizing photons produce discernible cone-like features corresponding to directions of lower H I column density near the central star-forming regions. On the other hand, the resonant scattering process leads to a much smoother Ly $\alpha$  radiation field. The mass-weighting clearly reveals the presence of shadowing behind dense structures, trapping environments within the galaxy, and filamentary self-shielding regions. On the other hand, a volume-weighted projection adds extra emphasis to the radiation field outside the galaxy, which reveals the highly isotropic halo background. Comparing the energy densities across snapshots shows an increasingly centralized configuration with time. Finally, the Ly $\alpha$  emission originating from recombinations is more concentrated and experiences more trapping than Ly $\alpha$  from collisional excitation.

### 6.2 Eddington factor

We also employ Monte Carlo estimators based on the traversed opacity to calculate the momentum imparted on the gas due to multiple scattering of Ly $\alpha$  photons. The radiation trapping time within a given volume is given by

$$t_{\text{trap}} = \frac{\int U_\alpha dV}{\int j_\alpha dV}, \quad (7)$$

where  $j_\alpha$  denotes the Ly $\alpha$  volume emissivity. Over the entire simulation domain we calculate the time-averaged ratio of trapping to light crossing times to be  $\langle t_{\text{trap}}/t_{\text{light}} \rangle \approx 1.1$ , where  $t_{\text{light}} \approx 2 R_{\text{vir}}/c$ . Therefore, the Ly $\alpha$  photons are well within the optically thin free streaming regime by the time they leave the domain, and all radiative transfer effects due to trapping occur on much smaller scales. However, the trapping ratio increases significantly within the high opacity regions of the galaxy and filaments. Furthermore, such regions also trap radiation from outside sources, thereby producing an artificially strong radiation field when compared to the local intrinsic emission. Thus, a more meaningful measure of the net effect of resonant scattering on the gas is the force multiplier, which is given by

$$M_F = \frac{\int \|\mathbf{a}_{\text{Ly}\alpha}\| \rho dV}{\int j_\alpha/c dV}, \quad (8)$$

where  $\mathbf{a}_{\text{Ly}\alpha}$  denotes the local acceleration due to Ly $\alpha$  radiation pressure. We calculate a time-averaged force multiplier of  $\langle M_F \rangle \approx 114$  over the entire simulation domain. This is certainly significant, especially as the calculation accounts for vector cancellation of momentum within individual cells.

However, determining the potential impact of Ly $\alpha$  radiation pressure requires a direct comparison with other dynamical forces.<sup>4</sup>

<sup>4</sup>We also note that post-processing estimates can be inaccurate if the trapping becomes comparable to dynamical time-scales.

Therefore, we compute the local Eddington factor as the ratio of magnitudes for Ly $\alpha$  radiative and total accelerations (including gravity and pressure gradients) experienced by the gas,<sup>5</sup> i.e.

$$f_{\text{Edd}} \equiv \frac{\|\mathbf{a}_{\text{Ly}\alpha}\|}{\|\mathbf{a}_{\text{tot}}\|}, \quad (9)$$

with (anti)parallel and perpendicular components given by

$$f_{\text{Edd},\parallel} \equiv -\frac{\mathbf{a}_{\text{Ly}\alpha} \cdot \mathbf{a}_{\text{tot}}}{\|\mathbf{a}_{\text{tot}}\|^2} \quad (10)$$

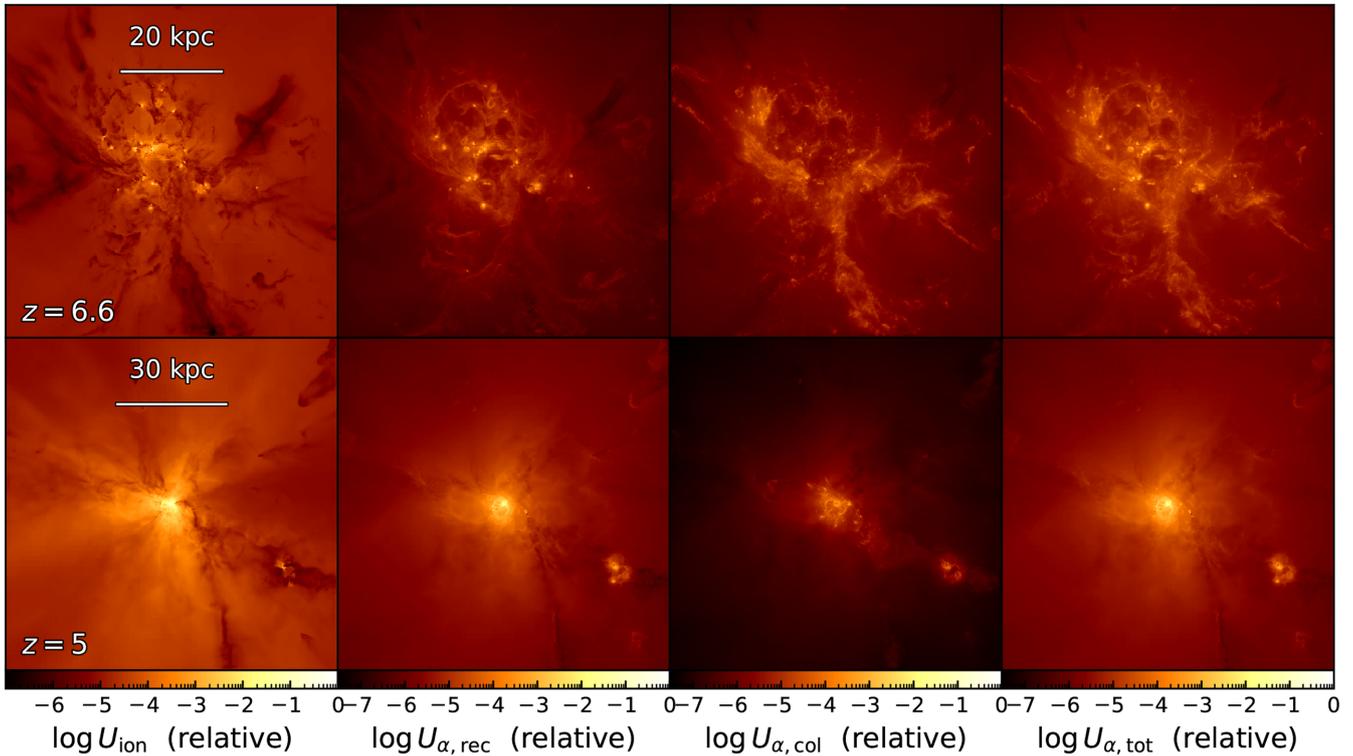
and

$$f_{\text{Edd},\perp} \equiv \sqrt{f_{\text{Edd}}^2 - f_{\text{Edd},\parallel}^2}. \quad (11)$$

We note that if the directions of the radiative and gravitational forces are uncorrelated then  $\langle |f_{\text{Edd},\parallel}| \rangle / \langle f_{\text{Edd}} \rangle \approx \int_0^1 \mu d\mu = 1/2$  and  $\langle f_{\text{Edd},\perp} \rangle / \langle f_{\text{Edd}} \rangle \approx \int_0^1 \sqrt{1 - \mu^2} d\mu = \pi/4$ , which is more or less the case in our simulations for the mass-weighted and volume-weighted time-averages. In Fig. 21, we illustrate the 3D structure of the regions most affected by Ly $\alpha$  trapping with a LOS mass-weighted projection of the Eddington factor, choosing a snapshot corresponding to the main starburst when the Ly $\alpha$  radiation field is strong. It is clear that in this system Ly $\alpha$  radiation pressure is likely to play only a minor role in the overall galactic dynamics. This is a result of the highly inhomogeneous gas distribution, which allows radiation to escape through low-opacity channels. However, the Eddington factor can be quite large along the cosmological filaments feeding the growth of the galaxy. To understand the extent of the high Eddington factor regime, in Fig. 21 we also show the cumulative mass function of gas above a given value of  $f_{\text{Edd}}$ . We find that  $M_{\text{gas}}(f_{\text{Edd}} > 1)$  fluctuates significantly with redshift in the range of  $10^6$ – $10^9 M_\odot$ , or 0.01–10 percent of the total gas mass, primarily driven by the total luminosity of the galaxy. While only a small fraction of the gas has a very high Eddington factor, from Fig. 21 it is clear that local trapping can affect the pressure of the infalling filamentary gas, which can have interesting consequences for the nature of filamentary gas infall as a major channel of growth in high-redshift galaxies (e.g. Kereš et al. 2005).

Previous studies have considered the effects of Ly $\alpha$  radiation pressure in highly idealized geometries or with approximate methods for the radiative transfer. In particular, analytic order of magnitude calculations regarding the relative role compared to other feedback mechanisms have been explored by several authors (e.g. Cox 1985; Haehnelt 1995; Oh & Haiman 2002; McKee & Tan 2008; Milosavljević et al. 2009; Wise et al. 2012). With the aid of MCRT, Dijkstra & Loeb (2008, 2009) found that multiple scattering within high H I column density shells is capable of significantly enhancing the effective Ly $\alpha$  force. Later Smith, Bromm & Loeb (2016); Smith et al. 2017a performed self-consistent MCRT Ly $\alpha$  radiation hydrodynamics (RHD) calculations of galactic winds in the first galaxies, finding that direct collapse black holes (DCBHs), which form in primordial gas, foster an environment that is especially susceptible to Ly $\alpha$  feedback. It has subsequently been found that trapped Ly $\alpha$  cooling radiation potentially affects the initial collapse of these massive black hole seeds through chemical (Johnson & Dijkstra 2017) and thermal (Ge & Wise 2017) feedback. To address the limitations of 1D geometry, Smith et al. (2017b) performed a post-processing radiative feedback analysis of a DCBH assembly

<sup>5</sup>Most regions in the simulation are not self-gravitating, so we consider the total acceleration due to gravity and hydrodynamical (thermal, turbulent, and radiation) pressure gradients from the simulation.



**Figure 20.** Projected mass-weighted energy density for ionizing and Ly $\alpha$  photons at  $z = 6.6$  (top) and  $z = 5$  (bottom), with the recombination and collisional emission colour scaled to the maximum projected  $U_{\alpha, \text{tot}}$ . Although these quantities are not directly observable, they provide intuition as ionizing and Ly $\alpha$  radiation have significantly different photon trajectories, resulting in visually distinctive escape morphologies. The radiation fields reveal regions affected by shadowing, trapping, and self-shielding. Furthermore, Ly $\alpha$  from recombinations is more concentrated and experiences more trapping than Ly $\alpha$  from collisional excitation. The ionization rate includes local stellar sources, collisional ionization, and a uniform redshift-dependent meta-galactic ionizing background.

environment, concluding that fully coupled 3D Ly $\alpha$  RHD will be crucial to consider in future DCBH simulations.

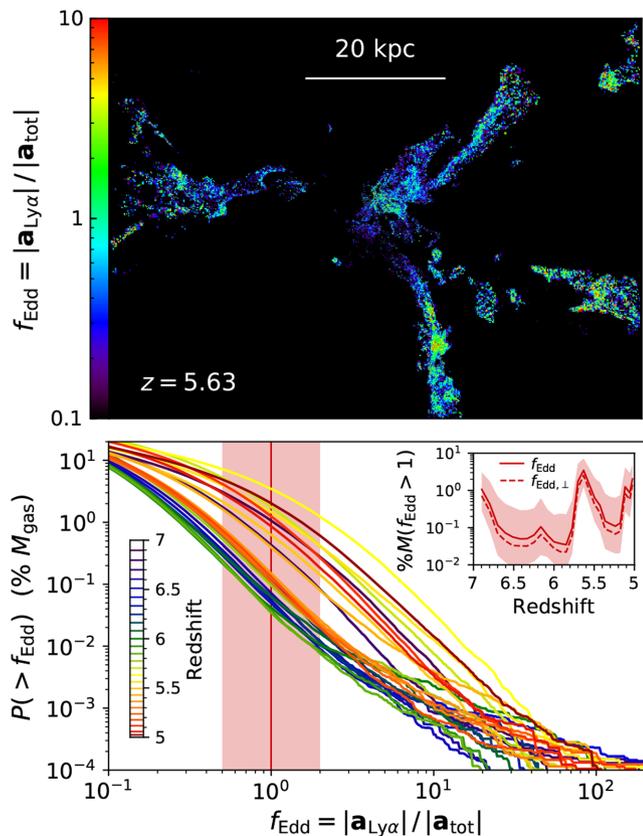
Recently, Kimm et al. (2018) incorporated a local subgrid model for Ly $\alpha$  momentum transfer into 3D RHD simulations of an isolated metal-poor dwarf galaxy. The authors found that Ly $\alpha$  feedback can regulate the dynamics of star-forming clouds before the onset of supernova explosions, thereby suppressing star formation and indirectly weakening galactic outflows. Our results further demonstrate the need for full 3D Ly $\alpha$  RHD in cosmological simulations to determine the extent to which Ly $\alpha$  trapping in low-metallicity environments is capable of impeding the growth of the first stars and supermassive black holes, disrupting cold gas accretion flows, driving winds in dwarf galaxies, or supplying additional turbulence to regulate star formation. Such simulations based on full solutions of the radiation transport equation will be possible in the near future with acceleration schemes such as the resonant Discrete Diffusion Monte Carlo (rDDMC; Smith et al. 2018).

## 7 SUMMARY AND DISCUSSION

In this work, we presented a comprehensive Ly $\alpha$  radiative transfer study of a cosmological zoom-in simulation from the FIRE project (Hopkins et al. 2017; Ma et al. 2018). We focused on the physics of Ly $\alpha$  escape from an individual galaxy over a redshift range of  $z = 5-7$ . Although our results are subject to cosmic variance and only represent one possible assembly and star formation history, our approach still allows us to treat each LOS as separate observations to quantify the variation due to viewing angle. We also achieve high spatial and angular resolution in the context of the galaxy’s

redshift evolution, which provides numerous insights about how Ly $\alpha$  observables change in response to starburst activity. Our main conclusions are as follows:

- (i) The dominant driver of the Ly $\alpha$  radiation field is the star formation history, and many properties are susceptible to fluctuations before and after star forming episodes.
- (ii) High equivalent width sightlines are rare, and are typically associated with outflows with additional coincident UV continuum absorption. The lowest equivalent widths correspond to cosmological filaments.
- (iii) Individual sources come in and out of visibility during their lifetimes. Thus, multi-object spectroscopic surveys will need to be quite deep to achieve high completeness of all phases of LAE populations for a given mass range. Still, even with  $10^4$  s exposure times Ly $\alpha$  detections and spectroscopy of high- $z$  galaxies with the *JWST* is feasible. However, the high (medium) resolution configuration only achieves  $\Delta v \approx 100$  (300)  $\text{km s}^{-1}$ , which is the same order of magnitude as the observed line widths after severe IGM reprocessing. As a consequence, we expect that diagnostics from other lines and cross-correlation studies will be necessary to unravel the detailed properties of high- $z$  LAEs.
- (iv) The local dust opacity is anti-correlated with Ly $\alpha$  escape, and logistic regression based on the local emission environment can predict the Ly $\alpha$  escape fraction to within 5 percent error. Similar models may thus provide efficient alternatives to obtain LAE statistics from hydrodynamical simulations.
- (v) Ly $\alpha$  radiation pressure can be dynamically important in dense, neutral, low-metallicity filaments and satellites.



**Figure 21.** *Top:* Projected mass-weighted Ly $\alpha$  Eddington factor  $f_{\text{Edd}} = \|\mathbf{a}_{\text{Ly}\alpha}\|/\|\mathbf{a}_{\text{tot}}\|$  shown during the main starburst at  $z = 5.63$ . The highest factors are along the cosmological filaments feeding the galaxy. *Bottom:* The cumulative mass function of gas above a given Eddington factor for each snapshot, with an inset showing the redshift evolution for  $f_{\text{Edd}} > 1^{+1.0}_{-0.5}$ , which is primarily driven by the total luminosity of the galaxy.

Future Ly $\alpha$  radiative transfer studies should require many of the elements included in this and previous works. For example, post-processing simulations must accurately account for full radiation hydrodynamics of ionizing radiation and other forms of feedback. Results with this approach are increasingly in agreement with observations, although the next major advances may require significantly higher spatial resolution to self-consistently model the multiphase ISM and CGM for statistical samples of simulated high- $z$  galaxies. Capturing the 3D structure of galaxies is vital, and assessing the physics from idealized analytical models is subject to limitations. However, we will continue to benefit from dedicated models and simulations probing radiative transfer effects on small and large scales. Furthermore, IGM treatments specific to individual sightlines within the same cosmological environment may be necessary to self-consistently account for double peaked spectra at  $z \gtrsim 6$  (e.g. Matthee et al. 2018). Some interesting physics to consider is cosmic ray feedback, which may leave an imprint on the Ly $\alpha$  spectra as the resultant outflows are smoother, colder, and denser than supernova driven winds (Gronke et al. 2018). As the observational frontier moves to higher redshifts and lower mass galaxies, Ly $\alpha$  radiation hydrodynamics will be increasingly relevant (Smith et al. 2017a; Kimm et al. 2018), possibly requiring additional acceleration schemes such as the rDDMC method (Smith et al. 2018). Certainly, an understanding of the intermittency and response to elevated star formation rates is key in interpreting Ly $\alpha$  spectral signatures. Furthermore, the physics of Ly $\alpha$  escape is highly

relevant for 21-cm cosmology and reionization studies, and we anticipate the increasing capability of simulations and abundance of data to facilitate the effective confluence of theory and observation at the high-redshift frontier.

## ACKNOWLEDGEMENTS

We thank the referee Luke Barnes for constructive comments that improved the content of this paper. The authors thank Peter Laursen who kindly provided IGM transmission data and helpful correspondence. AS benefited from numerous discussions with Benny Tsang, Intae Jung, Miloš Milosavljević, and Yao-Lun Yang. AS also thanks Jérémy Blaizot, Max Gronke, Dawn Erb, Anne Verhamme, Andrea Ferrara, Edward Robinson, Paul Shapiro, Alaina Henry, and Jorjyt Matthee for insightful conversations. Support for Program number *HST*-HF2-51421.001-A was 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. VB acknowledges support from NSF grant AST-1413501. CAFG was supported by NSF through grants AST-1412836, AST-1517491, AST-1715216, and CAREER award AST-1652522, by NASA through grant NNX15AB22G, by STScI through grant *HST*-AR-14562.001, and by a Cottrell Scholar Award from the Research Corporation for Science Advancement. DK was supported by NSF grant AST-1715101 and the Cottrell Scholar Award from the Research Corporation for Science Advancement. Support for PFH was provided by an Alfred P. Sloan Research Fellowship, NSF Collaborative Research Grant #1715847 and CAREER grant #1455342, and NASA grants NNX15AT06G, JPL 1589742, 17-ATP17-0214. Numerical calculations were run on the Caltech compute cluster “Wheeler,” allocations from XSEDE TG-AST130039 and PRAC NSF.1713353 supported by the NSF, and NASA HEC SMD-16-7592. The authors acknowledge the Texas Advanced Computing Center (TACC) at the University of Texas at Austin for providing HPC resources.

## REFERENCES

- Barnes L. A., Haehnelt M. G., Tescari E., Viel M., 2011, *MNRAS*, 416, 1723
- Behrens C., Braun H., 2014, *A&A*, 572, A74
- Behrens C., Dijkstra M., Niemeyer J. C., 2014, *A&A*, 563, A77
- Bowman J. D., Rogers A. E. E., Monsalve R. A., Mozdzen T. J., Mahesh N., 2018, *Nature*, 555, 67
- Bromm V., Yoshida N., 2011, *ARA&A*, 49, 373
- Cantalupo S., Porciani C., Lilly S. J., 2008, *ApJ*, 672, 48
- Cox D. P., 1985, *ApJ*, 288, 465
- Dayal P., Maselli A., Ferrara A., 2011, *MNRAS*, 410, 830
- De Barros S. et al., 2017, *A&A*, 608, A123
- Dijkstra M., 2014, *PASA*, 31, e040
- Dijkstra M., Kramer R., 2012, *MNRAS*, 424, 1672
- Dijkstra M., Loeb A., 2008, *MNRAS*, 391, 457
- Dijkstra M., Loeb A., 2009, *MNRAS*, 396, 377
- Dijkstra M., Haiman Z., Spaans M., 2006, *ApJ*, 649, 14
- Duval F., Schaerer D., Östlin G., Laursen P., 2014, *A&A*, 562, A52
- Eldridge J. J., Izzard R. G., Tout C. A., 2008, *MNRAS*, 384, 1109
- Erb D. K., Steidel C. C., Chen Y., 2018, *ApJ*, 862, L10
- Faucher-Giguère C.-A., Lidz A., Zaldarriaga M., Hernquist L., 2009, *ApJ*, 703, 1416
- Faucher-Giguère C.-A., Kereš D., Dijkstra M., Hernquist L., Zaldarriaga M., 2010, *ApJ*, 725, 633
- Finkelstein S. L., 2016, *PASA*, 33, e037

- Finkelstein S. L., Rhoads J. E., Malhotra S., Grogan N., Wang J., 2008, *ApJ*, 678, 655
- Finkelstein S. L., Rhoads J. E., Malhotra S., Grogan N., 2009, *ApJ*, 691, 465
- Furlanetto S. R., Pritchard J. R., 2006, *MNRAS*, 372, 1093
- Gardner J. P. et al., 2006, *Space Sci. Rev.*, 123, 485
- Ge Q., Wise J. H., 2017, *MNRAS*, 472, 2773
- Gronke M., Dijkstra M., 2014, *MNRAS*, 444, 1095
- Gronke M., Bull P., Dijkstra M., 2015, *ApJ*, 812, 123
- Gronke M., Dijkstra M., McCourt M., Oh S. P., 2016, *ApJ*, 833, L26
- Gronke M., Girichidis P., Naab T., Walch S., 2018, *ApJ*, 862, L7
- Haehnelt M. G., 1995, *MNRAS*, 273, 249
- Hansen M., Oh S. P., 2006, *MNRAS*, 367, 979
- Harrington J. P., 1973, *MNRAS*, 162, 43
- Hashimoto T. et al., 2015, *ApJ*, 812, 157
- Hashimoto T. et al., 2017, *A&A*, 608, A10
- Hayes M. et al., 2013, *ApJ*, 765, L27
- Hayes M., 2015, *PASA*, 32, e027
- Hayes M., Schaerer D., Östlin G., Mas-Hesse J. M., Atek H., Kunth D., 2011, *ApJ*, 730, 8
- Hopkins P. F. et al., 2018, *MNRAS*, 480, 800
- Hopkins P. F., 2015, *MNRAS*, 450, 53
- Hopkins P. F., Kereš D., Oñorbe J., Faucher-Giguère C.-A., Quataert E., Murray N., Bullock J. S., 2014, *MNRAS*, 445, 581
- Jensen H., Laursen P., Mellema G., Iliiev I. T., Sommer-Larsen J., Shaprio P. R., 2013, *MNRAS*, 428, 1366
- Johnson J. L., Dijkstra M., 2017, *A&A*, 601, A138
- Kasen D., Thomas R. C., Nugent P., 2006, *ApJ*, 651, 366
- Kereš D., Katz N., Weinberg D. H., Davé R., 2005, *MNRAS*, 363, 2
- Kimm T., Haehnelt M., Blaizot J., Katz H., Michel-Dansac L., Garel T., Rosdahl J., Teyssier R., 2018, *MNRAS*, 475, 4617
- Kroupa P., 2002, *Science*, 295, 82
- Laursen P., Razoumov A. O., Sommer-Larsen J., 2009a, *ApJ*, 696, 853
- Laursen P., Sommer-Larsen J., Andersen A. C., 2009b, *ApJ*, 704, 1640
- Laursen P., Sommer-Larsen J., Razoumov A. O., 2011, *ApJ*, 728, 52
- Laursen P., Duval F., Östlin G., 2013, *ApJ*, 766, 124
- Laursen P., Sommer-Larsen J., Milvang-Jensen B., Fynbo J. P. U., Razoumov A. O., 2018, preprint (arXiv:1806.07392)
- Leclercq F. et al., 2017, *A&A*, 608, A8
- Ma X. et al., 2018, *MNRAS*, 478, 1694
- Ma X., Kasen D., Hopkins P. F., Faucher-Giguère C.-A., Quataert E., Kereš D., Murray N., 2015, *MNRAS*, 453, 960
- Ma X., Hopkins P. F., Kasen D., Quataert E., Faucher-Giguère C.-A., Kereš D., Murray N., Strom A., 2016, *MNRAS*, 459, 3614
- Madau P., 2018, *MNRAS*, 480, L43
- Mallery R. P. et al., 2012, *ApJ*, 760, 128
- Mason C. A., Treu T., Dijkstra M., Mesinger A., Trenti M., Pentericci L., de Barros S., Vanzella E., 2018, *ApJ*, 856, 2
- Matthee J., Sobral D., Gronke M., Paulino-Afonso A., Stefanon M., Röttgering H., 2018, *A&A*, 619, A136
- McKee C. F., Tan J. C., 2008, *ApJ*, 681, 771
- Milosavljević M., Bromm V., Couch S. M., Oh S. P., 2009, *ApJ*, 698, 766
- Momose R. et al., 2016, *MNRAS*, 457, 2318
- Neufeld D. A., 1990, *ApJ*, 350, 216
- Neufeld D. A., 1991, *ApJ*, 370, L85
- Oh S. P., Haiman Z., 2002, *ApJ*, 569, 558
- Orlitova I., Verhamme A., Henry A., Scarlata C., Jaskot A., Oey S., Schaerer D., 2018, *A&A*, 616, A60
- Pentericci L. et al., 2018, *A&A*, 619, A147
- Pontoppidan K. M. et al., 2016, in *Observatory Operations: Strategies, Processes, and Systems VI*. p. 991016
- Pritchard J. R., Loeb A., 2012, *Rep. Prog. Phys.*, 75, 086901
- Ramdas A., Trillos N., Cuturi M., 2017, *Entropy*, 19, 47
- Schaerer D., Hayes M., Verhamme A., Teyssier R., 2011, *A&A*, 531, A12
- 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
- Smith A., Safranek-Shrader C., Bromm V., Milosavljević M., 2015, *MNRAS*, 449, 4336
- Smith A., Bromm V., Loeb A., 2016, *MNRAS*, 460, 3143
- Smith A., Bromm V., Loeb A., 2017a, *MNRAS*, 464, 2963
- Smith A., Becerra F., Bromm V., Hernquist L., 2017b, *MNRAS*, 472, 205
- Smith A., Tsang B. T. H., Bromm V., Milosavljević M., 2018, *MNRAS*, 479, 2065
- Sravan N. et al., 2016, *MNRAS*, 463, 120
- Stanway E. R., Eldridge J. J., Becker G. D., 2016, *MNRAS*, 456, 485
- Stark D. P. et al., 2017, *MNRAS*, 464, 469
- Stark D. P., Ellis R. S., Ouchi M., 2011, *ApJ*, 728, L2
- Steidel C. C., Bogosavljević M., Shapley A. E., Kollmeier J. A., Reddy N. A., Erb D. K., Pettini M., 2011, *ApJ*, 736, 160
- Steidel C. C., Strom A. L., Pettini M., Rudie G. C., Reddy N. A., Trainor R. F., 2016, *ApJ*, 826, 159
- Tasitsiomi A., 2006, *ApJ*, 645, 792
- Trainor R. F., Strom A. L., Steidel C. C., Rudie G. C., 2016, *ApJ*, 832, 171
- Trebitsch M., Verhamme A., Blaizot J., Rosdahl J., 2016, *A&A*, 593, A122
- Verhamme A. et al., 2018, *MNRAS*, 478, L60
- Verhamme A., Schaerer D., Maselli A., 2006, *A&A*, 460, 397
- Verhamme A., Schaerer D., Atek H., Tapken C., 2008, *A&A*, 491, 89
- Verhamme A., Dubois Y., Blaizot J., Garel T., Bacon R., Devriendt J., Guiderdoni B., Slyz A., 2012, *A&A*, 546, A111
- Wise J. H., Abel T., Turk M. J., Norman M. L., Smith B. D., 2012, *MNRAS*, 427, 311
- Yajima H., Li Y., Zhu Q., Abel T., 2012, *MNRAS*, 424, 884
- Zheng Z., Wallace J., 2014, *ApJ*, 794, 116
- Zheng Z., Cen R., Trac H., Miralda-Escudé J., 2010, *ApJ*, 716, 574

## APPENDIX A: SUPPLEMENTARY ANALYSIS

### A1 Angular power spectra

We quantify the structure of the angular variations in Section 3.5.3 by considering the spherical harmonic decomposition,

$$f(\theta, \varphi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(\theta, \varphi), \quad (\text{A1})$$

where the spectral coefficients  $a_{\ell m}$  represent the contribution from each of the real harmonics  $Y_{\ell m}$  and are found by integration as  $a_{\ell m} = \int f(\theta, \varphi) Y_{\ell m}(\theta, \varphi) d\Omega$ . The angular power spectrum is

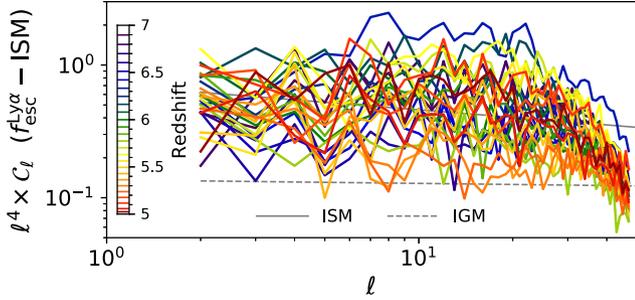
$$C_{\ell} = \frac{1}{2\ell + 1} \sum_{m=-\ell}^{\ell} |a_{\ell m}|^2, \quad (\text{A2})$$

which is well approximated by a power law with time-averaged slopes of  $d \log C_{\ell} / d \log f_{\text{esc}}^{\text{Ly}\alpha} \approx -4.2$  after ISM scattering and  $\approx -4.0$  after IGM transmission (shown in Fig. A1). Similarly,  $d \log C_{\ell} / d \log (F_{\text{red}} / F_{\text{blue}}) \approx -3.7$  (ISM),  $d \log C_{\ell} / d \log f_{\text{esc}}^{\text{UV}} \approx -3.6$ ,  $d \log C_{\ell} / d \log R_{1/2} \approx -3.9$  (ISM) and  $\approx -3.6$  (IGM), and  $d \log C_{\ell} / d \log \text{EW}_{\text{Ly}\alpha,0} \approx -4.1$  (ISM) and  $\approx -4.0$  (IGM). We restrict the fit range to  $2 \leq \ell \leq 35$ , which minimizes the bias due to numerical damping from the Gaussian filter used in the directional binning of photon packets. The damping of the angular power spectrum means that at increasingly small scales the deviations in  $f_{\text{esc}}^{\text{Ly}\alpha}$  are decreasing even faster, which is a natural consequence of diffusion, i.e. the multiple scattering of Ly $\alpha$  photons.

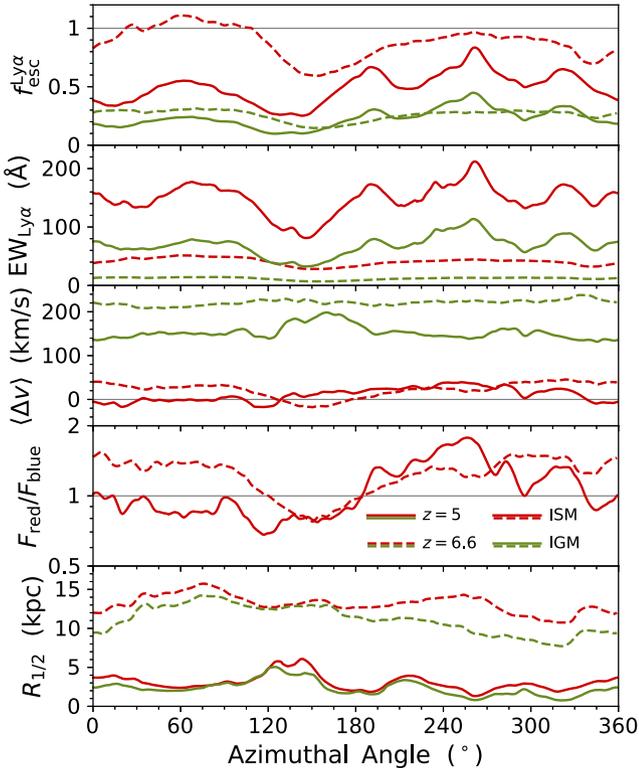
### A2 Azimuthal variations

To explore the line profile in further detail, in Fig. A2 we plot the Ly $\alpha$  escape fraction  $f_{\text{esc}}^{\text{Ly}\alpha}$ , rest-frame equivalent width  $\text{EW}_{\text{Ly}\alpha,0}$ , flux-weighted velocity centroid  $\langle \Delta v \rangle$ , red-to-blue flux ratio  $F_{\text{red}} / F_{\text{blue}}$ , and half-light radius  $R_{1/2}$ , as a function of azimuthal angle for an arbitrary rotation axis. This demonstrates both the range of variability in LOS properties and the correlations between these observables. For example, at  $z = 5$  the highest (lowest) inflections

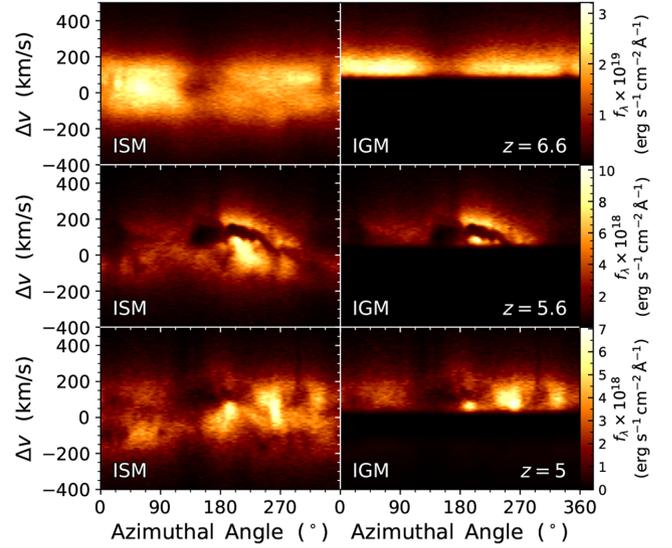
in equivalent width around  $260^\circ$  ( $145^\circ$ ) also exhibit red (blue) dominance in the red-to-blue flux ratio and a smaller (larger) half-light radius. In Fig. A3, we also show how the full spectral line profile smoothly changes across these viewing angles. In particular, the linear colour scale illustrates the impact of directional dependence on the peak velocity offset due to neutral hydrogen cloud obscuration, dust absorption, and Doppler shifting from bulk velocity flows, which affects the escape fraction both after ISM scattering and IGM transmission.



**Figure A1.** The angular power spectrum of the Ly $\alpha$  escape fraction before transmission through the IGM, which approximately follows a power law with a slope of  $d \log C_\ell / d \log f_{\text{esc}}^{\text{Ly}\alpha} \approx -4$ . The downturn at the highest  $\ell$  modes is caused by numerical damping from the Gaussian filter with an effective resolution of  $\sigma_\theta \approx 3.2$ , corresponding to  $\sigma_\ell \approx 56$ , used in the directional binning of photon packets.



**Figure A2.** Continuous dependence of Ly $\alpha$  properties for equatorial sightlines as a function of azimuthal angle, including the escape fraction  $f_{\text{esc}}^{\text{Ly}\alpha}$ , rest-frame equivalent width  $\text{EW}_{\text{Ly}\alpha,0}$ , flux-weighted velocity centroid  $\langle \Delta v \rangle$ , red-to-blue flux ratio  $F_{\text{red}}/F_{\text{blue}}$ , and half-light radius  $R_{1/2}$ .



**Figure A3.** Continuous dependence of the LOS spectra as a function of azimuthal angle, illustrating the complex diversity of line profiles along the rotation axis after ISM scattering and IGM transmission at  $z = \{5, 5.6, 6.6\}$ .

### A3 Comparing probability distributions

It is important to understand the validity and robustness of comparing the absorbed and escaped probability distributions in the predictive models of Section 5. To this end, we consider several statistical quantities to compare the time-averaged absorbed and escaped probability distribution functions, which we denote by  $p_{\text{abs}}(x)$  and  $p_{\text{esc}}(x)$  with  $x$  the random variable, e.g.  $x = k_{\text{dust}}$  for the dust absorption coefficient. First, we define the symmetric probability bisect  $x_B$  as the point dividing equal chances of absorption at lower values and escape at higher values,

$$\text{Probability Bisect: } \int_{-\infty}^{x_B} p_{\text{abs}}(x) dx = \int_{x_B}^{\infty} p_{\text{esc}}(x) dx. \quad (\text{A3})$$

We next consider the overlap of the distributions, based on their smoothed histograms to ensure common support in  $x$ . This statistic provides the probability that samples could be represented by either outcome, and is therefore a value in the unit interval. The minimum overlap  $\omega$  is the intersection between probability densities,

$$\omega \equiv \int_{-\infty}^{\infty} \min(p_{\text{abs}}(x), p_{\text{esc}}(x)) dx. \quad (\text{A4})$$

On the other hand, the  $L^2$ -norm overlap  $\omega_2$  considers regions of mutually significant probability with higher weight, which provides information about the similarity of their shapes,

$$\omega_2 \equiv \left( \int_{-\infty}^{\infty} p_{\text{abs}}(x) p_{\text{esc}}(x) dx \right)^{1/2}. \quad (\text{A5})$$

Likewise, the (squared) Hellinger distance also quantifies the similarity between the probability distributions,

$$H^2 \equiv \frac{1}{2} \int_{-\infty}^{\infty} \left( \sqrt{p_{\text{abs}}(x)} - \sqrt{p_{\text{esc}}(x)} \right)^2 dx. \quad (\text{A6})$$

However, neither quantity encapsulates a measure of how far apart the distributions are. Thus, we also compute the first Wasserstein distance  $W_1$ , also known as the earth mover's distance as it can be viewed as the minimum amount of work in terms of distribution weight multiplied by distance to transform  $p_{\text{abs}}(x)$  into  $p_{\text{esc}}(x)$ . A

**Table A1.** Time-averaged median and  $1\sigma$  statistics over the redshift interval  $z = 5-7$  for (i) ‘All’ = local environment of all emitted Ly $\alpha$  photon packets, (ii) ‘Escaped’ = subset for escaped photons, (iii) ‘+IGM’ = escaped photons weighted by their IGM transmission, (iv) ‘Absorbed’ = subset for absorbed photons, (v) ‘+ Local’ = photons absorbed in the same cell from which they were emitted, and (vi) ‘End’ = environment at the absorption locations. We illustrate the evolution and separation of the absorbed and escaped distributions in Fig. 18 for the absorption coefficients for dust  $k_{\text{dust}}$  and Ly $\alpha$   $k_{\alpha,0}$  at line centre. We also provide the symmetric probability bisect  $x_B$ , the minimum overlap  $\omega$ ,  $L^2$ -norm overlap  $\omega_2$ , Hellinger distance  $H$ , first Wasserstein distance  $W_1$ , and number of standard deviations  $N_1^\sigma$  separating the distributions (see equations (A3)–(A8)). The rows are in decreasing order of discriminating significance.

Quantity	All	Escaped	+ IGM	Absorbed	+ Local	End	Bisect	$\omega$	$\omega_2$	$H$	$W_1$	$N_1^\sigma$
$\log k_{\text{dust}} [\text{cm}^{-1}]$	$-23.2^{+1.8}_{-2.0}$	$-23.6^{+1.4}_{-1.8}$	$-23.4^{+1.3}_{-1.8}$	$-21.1^{+1.2}_{-1.5}$	$-20.3^{+0.8}_{-0.8}$	$-20.6^{+0.9}_{-1.1}$	-22.40	0.39	0.30	0.58	2.56	1.74
$\log k_{\alpha,0} [\text{cm}^{-1}]$	$-14.6^{+2.1}_{-2.2}$	$-15.3^{+1.8}_{-1.8}$	$-15.1^{+1.8}_{-1.8}$	$-11.8^{+1.4}_{-2.5}$	$-10.8^{+1.0}_{-1.0}$	$-11.2^{+1.0}_{-1.6}$	-13.82	0.40	0.26	0.56	3.15	1.67
$\log \rho [\text{g cm}^{-3}]$	$-23.8^{+1.7}_{-1.2}$	$-23.9^{+1.3}_{-1.1}$	$-23.7^{+1.3}_{-1.2}$	$-21.9^{+0.7}_{-1.1}$	$-21.6^{+0.5}_{-0.7}$	$-22.0^{+0.7}_{-1.2}$	-22.78	0.41	0.36	0.54	1.73	1.68
$\log T [\text{K}]$	$4.0^{+0.2}_{-0.8}$	$4.0^{+0.2}_{-0.6}$	$4.0^{+0.2}_{-0.6}$	$3.1^{+0.7}_{-1.2}$	$2.8^{+0.7}_{-1.2}$	$3.1^{+0.8}_{-1.1}$	3.66	0.44	0.66	0.48	0.92	1.39
$\log x_{\text{H I}}$	$-0.91^{+0.77}_{-2.37}$	$-1.90^{+1.63}_{-1.78}$	$-1.86^{+1.60}_{-1.82}$	$-0.35^{+0.26}_{-2.29}$	$-0.18^{+0.11}_{-0.30}$	$-0.11^{+0.10}_{-0.25}$	-0.87	0.62	0.58	0.37	0.95	0.69
$\log r [\text{kpc}]$	$0.72^{+0.56}_{-0.49}$	$0.73^{+0.59}_{-0.51}$	$0.70^{+0.56}_{-0.50}$	$0.45^{+0.38}_{-0.42}$	$0.44^{+0.36}_{-0.43}$	$0.46^{+0.38}_{-0.42}$	0.59	0.72	0.77	0.28	0.28	0.56
$\log Z [Z_\odot]$	$-1.38^{+0.41}_{-0.67}$	$-1.37^{+0.41}_{-0.72}$	$-1.32^{+0.38}_{-0.65}$	$-1.21^{+0.27}_{-0.40}$	$-1.17^{+0.25}_{-0.36}$	$-1.19^{+0.27}_{-0.39}$	-1.27	0.80	0.83	0.24	0.27	0.53

practical definition for computing this metric is

$$W_1 = \int_0^1 |P_{\text{abs}}^{-1}(t) - P_{\text{esc}}^{-1}(t)| dt, \quad (\text{A7})$$

where  $P_{\text{abs}}(t)^{-1}$  and  $P_{\text{esc}}(t)^{-1}$  denote inverse cumulative distribution functions or quartile functions of  $p_{\text{abs}}(x)$  and  $p_{\text{esc}}(x)$ , respectively (Ramdas, Trillos & Cuturi 2017). Finally, we transform this distance into units of significance by normalizing according to the average standard deviations, i.e. we define

$$N_1^\sigma \equiv \frac{2 W_1}{\sigma_{\text{abs}} + \sigma_{\text{esc}}}. \quad (\text{A8})$$

We interpret this as the approximate number of standard deviations separating the distributions, and employ  $N_1^\sigma$  to rank the power of a given local quantity to predict the fate of Ly $\alpha$  photons.

In Table A1, we summarize each of these statistics for the following quantities in order of decreasing discriminating significance based on  $N_1^\sigma$ : dust absorption coefficient  $k_{\text{dust}}$ , Ly $\alpha$  absorption coefficient at line centre  $k_{\alpha,0}$ , density  $\rho$ , temperature  $T$ , neutral fraction  $x_{\text{H I}} \equiv n_{\text{H I}}/n_{\text{H}}$ , radial distance from the galactic centre  $r$ , and metallicity  $Z$ .

This paper has been typeset from a T<sub>E</sub>X/L<sup>A</sup>T<sub>E</sub>X file prepared by the author.