



**HAL**  
open science

## Lyman- $\alpha$ as a tracer of cosmic reionization in the SPHINX radiation-hydrodynamics cosmological simulation

Thibault Garel, Joakim A Rosdahl, Joakim A Rosdahl, Léo Michel-Dansac, Martin G Haehnelt, Harley Katz, Taysun Kimm, Anne Verhamme

► **To cite this version:**

Thibault Garel, Joakim A Rosdahl, Joakim A Rosdahl, Léo Michel-Dansac, Martin G Haehnelt, et al.. Lyman- $\alpha$  as a tracer of cosmic reionization in the SPHINX radiation-hydrodynamics cosmological simulation. Monthly Notices of the Royal Astronomical Society, 2021, 504 (2), pp.1902-1926. 10.1093/mnras/stab990 . hal-03444370

**HAL Id: hal-03444370**

**<https://cnrs.hal.science/hal-03444370>**

Submitted on 23 Nov 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Lyman- $\alpha$ as a tracer of cosmic reionisation in the SPHINX radiation-hydrodynamics cosmological simulation

Thibault Garel<sup>1,2\*</sup>, Jérémy Blaizot<sup>2</sup>, Joakim Rosdahl<sup>2</sup>, Léo Michel-Dansac<sup>2</sup>, Martin G. Haehnelt<sup>3</sup>, Harley Katz<sup>4†</sup>, Taysun Kimm<sup>5</sup> and Anne Verhamme<sup>1</sup>

<sup>1</sup> *Observatoire de Genève, Université de Genève, 51 Ch. des Maillettes, 1290 Versoix, Switzerland*

<sup>2</sup> *Univ Lyon, Univ Lyon1, Ens de Lyon, CNRS, Centre de Recherche Astrophysique de Lyon UMR5574, F-69230, Saint-Genis-Laval, France*

<sup>3</sup> *Kavli Institute for Cosmology and Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK*

<sup>4</sup> *Sub-department of Astrophysics, University of Oxford, Keble Road, Oxford OX1 3RH, UK*

<sup>5</sup> *Department of Astronomy, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea*

9 April 2021

## ABSTRACT

The Ly $\alpha$  emission line is one of the most promising probes of cosmic reionisation but isolating the signature of a change in the ionisation state of the IGM is challenging because of intrinsic evolution and internal radiation transfer effects. We present the first study of the evolution of Ly $\alpha$  emitters (LAE) during the epoch of reionisation based on a full radiation-hydrodynamics cosmological simulation that is able to capture both the large-scale process of reionisation and the small-scale properties of galaxies. We predict the Ly $\alpha$  emission of galaxies in the  $10^3$  cMpc<sup>3</sup> SPHINX simulation at  $6 \leq z \leq 9$  by computing the full Ly $\alpha$  radiation transfer from ISM to IGM scales. SPHINX is able to reproduce many observational constraints such as the UV/Ly $\alpha$  luminosity functions and stellar mass functions at  $z \gtrsim 6$  for the dynamical range probed by our simulation ( $M_{1500} \gtrsim -18$ ,  $L_{\text{Ly}\alpha} \lesssim 10^{42}$  erg s<sup>-1</sup>,  $M_{\star} \lesssim 10^9 M_{\odot}$ ). As intrinsic Ly $\alpha$  emission and internal Ly $\alpha$  escape fractions barely evolve from  $z = 6$  to 9, the observed suppression of Ly $\alpha$  luminosities with increasing redshift is fully attributed to IGM absorption. For most observable galaxies ( $M_{1500} \lesssim -16$ ), the Ly $\alpha$  line profiles are slightly shifted to the red due to internal radiative transfer effects which mitigates the effect of IGM absorption. Overall, the enhanced Ly $\alpha$  suppression during reionisation traces the IGM neutral fraction  $x_{\text{HI}}$  well but the predicted amplitude of this reduction is a strong function of the Ly $\alpha$  peak shift, which is set at ISM/CGM scales. We find that a large number of LAEs could be detectable in very deep surveys during reionisation when  $x_{\text{HI}}$  is still  $\approx 50\%$ .

**Key words:** galaxies: formation – galaxies: evolution – galaxies: high-redshift – methods: numerical.

## 1 INTRODUCTION

Cosmic reionisation is one of the most fundamental stages in the history of the Universe, marking the end of the Dark ages and the formation of the first luminous sources. A patchy scenario in which HII regions expand around ionising sources until filling up the entire Universe is currently favoured but a thorough understanding of this process remains challenging. In spite of intense research over the last decades, there is still no consensus regarding the nature of the objects which reionised the intergalactic medium and the timeline over which it occurred.

While active galactic nuclei certainly contributed to the global ionising photon budget, there is growing evidence that stellar emission within galaxies is the dominant source (Kulkarni et al. 2019; Parsa et al. 2017; Finkelstein et al. 2019). Nevertheless, the relative contribution of low-mass versus massive galaxies still needs to be assessed due to uncertainties in the abundance of faint dwarfs during the epoch of reionisation (EoR; Livermore et al. 2017; Bouwens et al. 2015; Atek et al. 2018; Bhatawdekar et al. 2019) and in the ability of ionising photons to escape (Robertson et al. 2015; Wise et al. 2014; Rosdahl et al. 2018; Ma et al. 2016; Kimm & Cen 2014; Paardekooper et al. 2015). The direct measurement of the ionising Lyman continuum (LyC) escape fraction is impossible at high redshift because of the high opacity of the intergalactic medium (IGM) but observations of low-redshift analogs suggest that the typical

\* Email: thibault.garel@unige.ch

† Visitor

fraction of photons able to escape galaxies is low ( $\lesssim 10\%$ ; Steidel et al. 2018; Izotov et al. 2016; Grazian et al. 2015), even though a handful of strong leakers have been reported ( $\gtrsim 50\%$ ; e.g. Vanzella et al. 2016; Izotov et al. 2018). In parallel, quasar absorption spectra suggest that the Universe was almost fully ionised at  $z \approx 5-6$  (Fan et al. 2006; Mesinger 2010; Kulkarni et al. 2019) and still partially neutral at  $z \gtrsim 7$  (e.g. Bañados et al. 2017; Davies et al. 2018; Āurovčíková et al. 2020).

In addition to future 21 cm observations, one of the most promising routes to probe the EoR resides in Lyman- $\alpha$  (hereafter Ly $\alpha$ ) surveys. It is well known that the strong Ly $\alpha$  line produced in galaxies can be used as an indirect measurement of the neutral IGM component since Ly $\alpha$  photons can be scattered off the line of sight by intervening H I atoms. As the Universe becomes more neutral towards higher redshifts, the visibility of Ly $\alpha$  emitters (hereafter LAEs) will drop and the imprint of reionisation should translate into a shift of the Ly $\alpha$  luminosity function (LF; Haiman & Cen 2005; Dijkstra et al. 2007). Hints for such behaviour have been indeed reported in various narrow-band surveys at  $z \gtrsim 6$  (Ouchi et al. 2010; Konno et al. 2014; Zheng et al. 2017). A similar signature of reionisation is also seen in UV-selected samples where the fraction of objects with strong Ly $\alpha$  emission,  $\chi_{\text{Ly}\alpha}$ , is first found to increase from  $z \approx 3$  to 6 and then to decline at higher redshift (Stark et al. 2010; Pentericci et al. 2018; Hoag et al. 2019). While this trend is often interpreted as a rapid increase of the volumetric IGM neutral fraction ( $x_{\text{HI}}$ ) at  $z \gtrsim 6$ , it is noticeable that the significance and the redshift of the drop often differ from one study to another (see e.g. Kusakabe et al. 2020; Stark et al. 2016; Fuller et al. 2020). This may be a consequence of the patchiness of the reionisation process, or simply due to the different depths, selections and limited statistics of the samples used to compute  $\chi_{\text{Ly}\alpha}$ .

Altogether, these diagnostics can be used to assess the variation of the visibility of LAEs and therefore probe the evolution of the ionisation state of the IGM. However it is not necessarily straightforward to disentangle the impact of IGM attenuation from the intrinsic evolution of the Ly $\alpha$  emission and galactic radiative transfer (RT) effects (Dayal & Ferrara 2012; Laursen et al. 2011; Jensen et al. 2013; Garel et al. 2015; Hassan & Gronke 2021). Intrinsic Ly $\alpha$  luminosities are usually assumed to scale linearly with star formation rate. Still, this relation may evolve at high redshift if recombination of photoionised gas is no longer the dominant production channel of Ly $\alpha$  photons, or if very low metallicities are involved (Laursen et al. 2019; Smith et al. 2018; Raiter et al. 2010). In addition, the Ly $\alpha$  line is very sensitive to resonant scattering in the interstellar medium (ISM) and circumgalactic medium (CGM), i.e. the material inside and close to galaxies. First, the enhanced distance travelled by Ly $\alpha$  photons due to local scatterings increases dust absorption which can significantly suppress the flux emerging from galaxies. Second, RT in the optically thick regime can strongly affect the line profile and shift it away from resonance, especially in non-static media. In the presence of outflows, this effect can tremendously reduce the relative impact of the IGM on the visibility of LAEs (Santos 2004; Dijkstra & Wyithe 2010; Garel et al. 2012; Mason et al. 2018).

The modelling of the LAE population during the EoR is therefore a multi-scale problem which ideally requires to self-consistently describe the production and transfer of Ly $\alpha$  photons at small scales in galaxies as well as their propagation in the intergalactic medium. Such simulations are computationally expensive because of (i) the wide dynamical range involved, (ii) the need for radiation-hydrodynamics (RHD) to account for the interplay between ionising radiation and the gas, and (iii) the full post-processing with Ly $\alpha$  RT.

Several studies have focused on individual objects but neglecting the IGM component (Verhamme et al. 2012; Yajima et al. 2014; Smith et al. 2018). Alternatively, the transmission of the Ly $\alpha$  line through the IGM has been investigated in representative simulation volumes (Dayal et al. 2011; Hutter et al. 2014; Gronke et al. 2020; Jensen et al. 2014; Inoue et al. 2018). This is however at the expense of the physical and mass resolution which is needed to model the Ly $\alpha$  emission and transfer within the ISM and the CGM. To overcome some of these issues, Laursen et al. (2019) have recently built a new hybrid framework to model hundreds of Ly $\alpha$  sources at  $z \approx 9$ . Their approach combines a semi-analytical scheme to predict the halo mass function with high-resolution hydrodynamic zoom simulations in which both ionising and Ly $\alpha$  RT are performed as a post-processing step.

In this paper, we present a new study of the evolution of LAEs during the EoR based on the SPHINX simulation project (Rosdahl et al. 2018). SPHINX is a set of full RHD cosmological simulations of the formation and evolution of galaxies at  $z > 6$ . In the current study, we use exclusively the  $10^3 \text{ cMpc}^3$  version of SPHINX which includes the effect of binary stars with BPASS v2.0 to fully reionise the simulated volume before redshift six<sup>1</sup>. Taking advantage of the adaptive mesh refinement code RAMSES-RT (Teyssier 2002; Rosdahl et al. 2013), we are able to capture a wide range of scales with SPHINX. Here, we intend to assess the relative impact of intrinsic evolution, absorption at galaxy scales and IGM transmission to predict to which extent the visibility of LAEs is tracing the IGM neutral fraction during the EoR. The SPHINX simulation is therefore well-suited since it allows us to investigate the transport of Ly $\alpha$  photons from the ISM to the IGM for a large sample of objects.

The outline of the article is as follows. Section 2 describes the SPHINX simulation project and our modelling of Ly $\alpha$  emission and transfer. In Section 3, we compare our results with statistical observational constraints (stellar mass, UV/Ly $\alpha$  LFs, LAE fraction) and assess the relative evolution of the Ly $\alpha$  IGM transmission compared to Ly $\alpha$  intrinsic emission and escape fraction during the EoR. Then we attempt to characterise the imprint of the IGM on the Ly $\alpha$  LF, equivalent width (EW) distribution, LAE fraction, and spectra as a function of  $x_{\text{HI}}$ . We discuss our results in Section 4 and we give a summary in Section 5.

## 2 SIMULATION AND METHOD

In this section, we describe the SPHINX simulation suite and the RASCAS radiation transfer code that we use to post-process the SPHINX outputs.

### 2.1 The SPHINX simulation

SPHINX is a set of cosmological radiation-hydrodynamics simulations of galaxy formation during the epoch of reionisation. It has been run with the 3D adaptive mesh refinement code RAMSES-RT (Rosdahl et al. 2013) to describe the evolution of dark matter, baryons, and ionising radiation via gravity, hydrodynamics, RT, and non-equilibrium radiative cooling/heating. The SPHINX simulation suite has been presented in Rosdahl et al. (2018) and Katz et al. (2020) and here we recall the main features that are relevant to our study.

<sup>1</sup> Although there are now several SPHINX simulations, we will for simplicity refer to this  $10 \text{ cMpc}$  simulation with BPASS v2.0 throughout this paper as SPHINX.

### 2.1.1 Numerical setup

Rosdahl et al. (2018) have explored several simulations with various sizes, mass resolutions and SED models. Here, we make use of the fiducial simulation of the SPHINX project which describes a  $V_{\text{box}} = 10^3 \text{ cMpc}^3$  volume, and includes the effects of binary stars, a maximum physical resolution of 10.9 pc (at  $z=6$ ), and  $512^3$  dark matter particles of mass  $m_{\text{DM}} = 2.5 \times 10^5 M_{\odot}$ .

In SPHINX, the hydrodynamics are solved using the HLLC Riemann solver (Toro et al. 1994) and a MinMod slope limiter. An adiabatic index of  $5/3$  is assumed to close the relation between gas pressure and internal energy. Gravitational interactions for DM and stellar particles are computed with a particle-mesh solver and cloud-in-cell interpolation following Guillet & Teyssier (2011). The radiation is advected between cells using the M1 closure method (Levermore 1984) and the Global-Lax-Friedrich intercell flux function.

The initial conditions (IC) were generated with MUSIC (Hahn & Abel 2011) assuming cosmological parameters consistent with the Planck results (Ade et al. 2014,  $h_{100} = 0.6711$ ,  $\Omega_{\Lambda} = 0.6825$ ,  $\Omega_{\text{m}} = 0.3175$ ,  $\Omega_{\text{b}} = 0.049$ , and  $\sigma_8 = 0.83$ ) and they were chosen from a set of DM-only simulations so as to obtain a representative sample of sources that minimises the effect of cosmic variance on the ionising radiation budget. This was done using a large number of simulations with different IC realisations and the chosen set of ICs corresponds to the one yielding a sample of objects which generates an average ionizing luminosity budget (Section 2.2.1 in Rosdahl et al. 2018). Regarding the primordial abundance of chemical elements, we have adopted a mixture of hydrogen ( $X = 0.76$ ), helium ( $Y = 0.24$ ) and metals ( $Z = 6.4 \times 10^{-6}$ ) where the initial metallicity value was chosen to account for the lack of molecular cooling at early stages, such that first stars can start forming by  $z \approx 15 - 20$ .

### 2.1.2 Baryonic physics and stellar library

The gas cooling implementation includes the contribution of both primordial species and metal lines, following the prescription presented in Rosdahl et al. (2013). Star formation (SF) is modelled using a recipe adapted from Federrath & Klessen (2012) where turbulent gas motions act as an additional pressure support against gravitational collapse. As described in Rosdahl et al. (2018), stars can form in a grid cell when the local density corresponds to a maximum and is greater than 200 times the cosmological mean, the gas motion is locally convergent, and the turbulent Jeans length is less than one cell width. Gas cells meeting these criteria can produce stars according to a Schmidt law with a varying SF efficiency that depends on the local thermoturbulent properties of the gas (see Kimm et al. 2017, for details). In each simulation cell, the gas is stochastically converted into stellar particles by sampling the Poisson probability distribution for gas to star conversion over the timestep (see Rasera & Teyssier 2005, for details), such that on average, the conversion rate follows the Schmidt law (Eq. 3 in Rosdahl et al. 2018). Initially, stellar particles, each representing a stellar population, are allocated masses equal to integer multiples of  $10^3 M_{\odot}$ . An upper limit is set such that no more than 90% of the cell gas can turn into stars. As shown in Trebitsch et al. (2017), this recipe leads to a much more bursty SF than typical models based on a constant SF efficiency.

Stellar evolution and feedback is modelled following Kimm et al. (2015) by injecting mass, metal and momentum to surrounding gas cells. In practice, Type II SN explosions are stochastically

sampled from the delay-time distribution for the Kroupa (2001) IMF over the first 50 Myr of the lifetime of each star particle. We assume that each star particle hosts 4 SN events per  $100 M_{\odot}$ , which is four times larger than the typical SN frequency computed for the Kroupa IMF (1 per  $100 M_{\odot}$ ), in order to avoid overcooling and reproduce observational constraints at  $z \approx 6$  (Rosdahl et al. 2018).

Spectral energy distributions are computed using the BPASS library (Eldridge et al. 2008) which includes the effect of interacting binary stars (assuming 100% of stars are in binary systems) with metallicities and ages in the range 0.001 – 0.4 and 1 Myr – 10 Gyr respectively. As shown in Rosdahl et al. (2018), this choice of stellar library produces a much earlier reionisation history than an identical simulation with single stars only as it can fully reionise the box by  $z \approx 7$  whereas the IGM is still  $\approx 50\%$  neutral at  $z=6$  with the single star model. The discrepancy is mainly due to two factors. First, the binary model produces more ionising photons for a given stellar population, especially at low metallicities. Second, the ionising emission is prolonged for interacting binaries with respect to single stars (e.g. 25 Myr after a starburst, the ionising luminosity is  $\geq 10$  times larger with binaries) which leaves more time for SN feedback to clear the gas away from dense regions, allowing photons to escape more easily into the IGM.

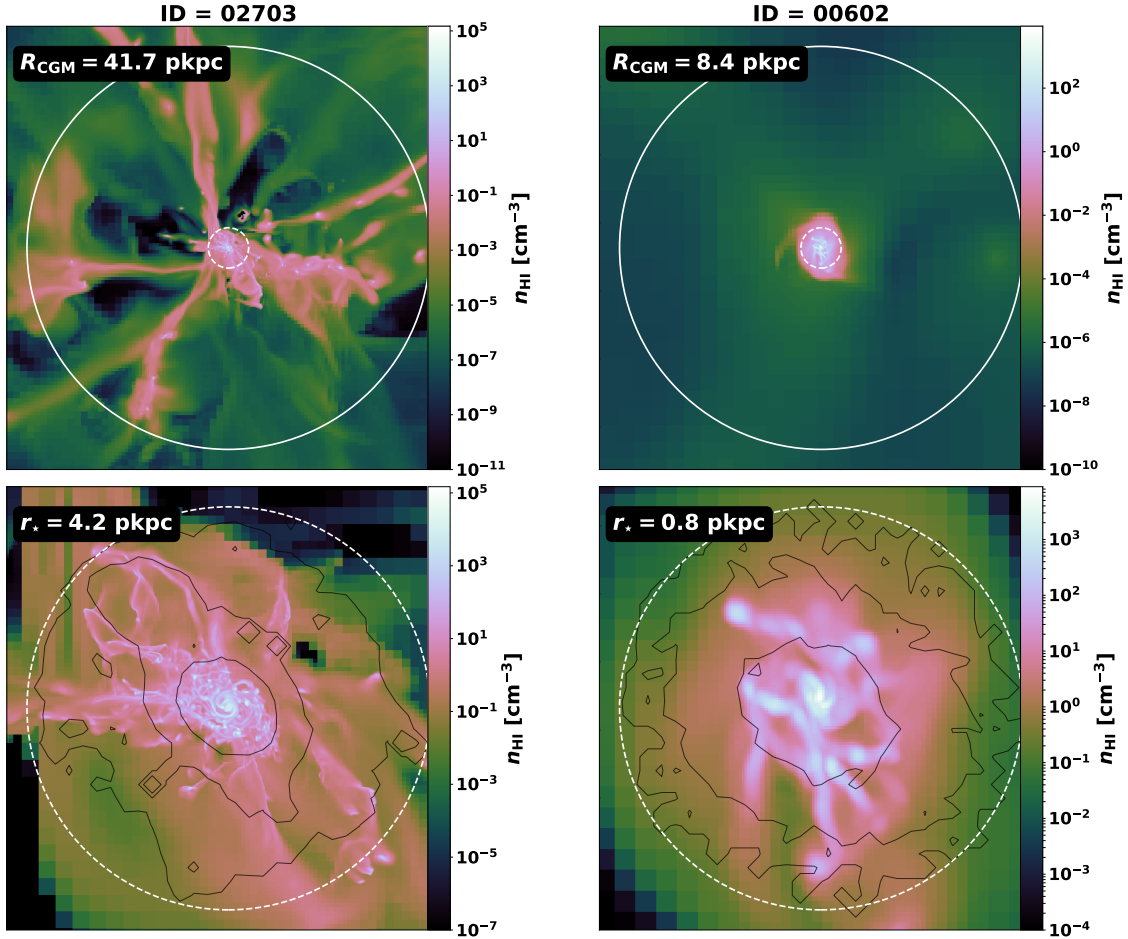
The ionising radiation is injected directly into the cells hosting stellar particles in each simulation step and propagated through the volume using the so-called M1 moment method. The radiation is split into three monochromatic groups bracketed by the H $\gamma$ , H $\epsilon$ , and H $\eta$  ionisation energies. The simulation tracks the local non-equilibrium ionisation fractions of hydrogen and helium and radiation interacts with the gas via photoionisation, heating, and momentum transfer.

As shown in Rosdahl et al. (2018), SPHINX starts ionising the Universe subsequently to the formation of the first stars and reionisation proceeds through the growth of H $\gamma$  bubbles until filling the whole volume with ionised hydrogen by redshift  $\approx 7$ . The reionisation history in SPHINX seems to occur over a similar timescale as estimated from observations but completes slightly too early (i.e. by  $\Delta z \approx 0.5$ ) with respect to these observational constraints (see Figure 9 in Rosdahl et al. 2018).

### 2.1.3 Galaxy catalog

The present study aims at following the evolution of LAEs during the EoR and thus we decide to focus on the four snapshots of the simulation corresponding to  $z = 6, 7, 8$  and  $9$ . At each snapshot, we identify individual galaxies with ADAPTAHOP (Aubert et al. 2004; Tweed et al. 2009) and select groups with at least 100 star particles ( $M_{\star, \text{min}} = 10^5 M_{\odot}$ ) and a local density threshold  $\rho_{\text{th}} = 1000$  following the notation of Aubert et al. (2004). These values have been chosen so as to avoid spurious identifications and to maximise the association of star particles with galaxies. The galaxy size is returned by the galaxy finder and corresponds to the distance from the furthest star particle to the mass center. This value, defined as the stellar radius  $r_{\star}$ , ensures that it encompasses the bulk of the photon budget produced within the interstellar medium in order to compute the Ly $\alpha$  and UV intrinsic emissivities of each galaxy (see Section 2.2.2). This methodology allows to construct a statistical sample of simulated galaxies at each redshift of interest, yielding 2911, 2357, 1867 and 1353 sources at  $z = 6, 7, 8$ , and  $9$  respectively.

In Figure 1, we present examples of a bright/massive galaxy (left) and a relatively faint and less massive galaxy (right) from SPHINX at  $z = 6$ . The images show their H $\gamma$  density maps at the CGM scale (top) and at the ISM scale (bottom). The black contours



**Figure 1.** Projections of two galaxies from SPHINX at  $z = 6$  showing the HI density maps at the CGM scale (top) and at the ISM scale (bottom). The left panels correspond to one of the most massive galaxies in our simulation (ID2703) while the right panels present a relatively low-mass object (ID602). The black contours in the lower panels indicate the distribution of stars within the ISM, with stellar surface density levels of  $10^5$ ,  $10^6$  and  $10^7$   $M_{\odot} \text{ kpc}^{-2}$ . The white solid and dashed circles depict the CGM radius ( $R_{\text{CGM}}$ ) and the stellar radius ( $r_{\star}$ ) respectively (the values are given in the legend of each panel; see Section 2.2.3 for the definition of  $R_{\text{CGM}}$ ). The stellar mass and star formation rate (computed over the last 10 Myr) for ID2703 (resp. ID602) are  $1.5 \times 10^9 M_{\odot}$  (resp.  $10^7 M_{\odot}$ ) and  $3 M_{\odot} \text{ yr}^{-1}$  (resp.  $0.005 M_{\odot} \text{ yr}^{-1}$ ).

in the lower panels represent the distribution of stars within the ISM, with increasing levels of stellar surface density from  $10^5$  to  $10^7 M_{\odot} \text{ kpc}^{-2}$ . Figure 1 highlights the high level of details that can be resolved in the internal structure of our galaxies and their surrounding medium, as well as the large dynamical range that can be probed with gas densities spanning many orders of magnitudes. Note that galaxies in our simulation display a wide diversity of morphologies so these two objects, which have been chosen arbitrarily, are not necessarily representative of the global population.

## 2.2 Ly $\alpha$ and UV post-processing

### 2.2.1 The RASCAS code

The emission and transport of Ly $\alpha$  and (non-ionising) UV photons is performed in post-processing using the 3D Monte-Carlo RT code RASCAS (Michel-Dansac et al. 2020). RASCAS has been specifically designed to ingest large simulations like SPHINX using full MPI parallelization, domain decomposition and adaptive load balancing in order to predict intrinsic emissivities from the gas/stars

and the transfer of resonant lines (as well as non-resonant lines or continuum) in the presence of dust.

RASCAS generates the intrinsic emission for each source (i.e. a gas cell or a star particle) of interest with a given number of photon packets according to its luminosity, each photon packet being assigned a constant weight. Photon packets are cast isotropically from the source with a probability  $P$  which is given by  $P = \dot{N}_{\lambda}^{\text{intr}} / \dot{N}_{\lambda, \text{tot}}^{\text{intr}}$  where  $\dot{N}_{\lambda}^{\text{intr}}$  is the true number of emitted photons per unit time by the source and  $\dot{N}_{\lambda, \text{tot}}^{\text{intr}}$  is the sum over all sources, such that  $\dot{N}_{\lambda, \text{tot}}^{\text{intr}} = \sum_i \dot{N}_{\lambda, i}^{\text{intr}}$ .

The subsequent propagation of photon packets through the mesh is performed based on a Monte-Carlo procedure which includes the core-skipping algorithm of Smith et al. (2015). The interaction with matter is set by the optical depth of a mixture of hydrogen and dust (see Section 3 in Michel-Dansac et al. 2020). While Ly $\alpha$  photons can interact with HI atoms and dust, UV continuum photons only interact with the latter. When a photon interacts with a dust grain, it can either be absorbed or scattered with a probability set by the albedo,  $A$ . Following Li & Draine (2001), we assume  $A = 0.32$  at Ly $\alpha$  and  $A = 0.38$  at  $1500 \text{ \AA}$ .

The formation of dust grains is not modelled in SPHINX so we use the default dust model implemented in RASCAS, and based on the formulation of Laursen et al. (2009), to compute the effective dust content of each cell. With this prescription, the dust absorption coefficient is given by  $(n_{\text{HI}} + f_{\text{ion}}n_{\text{HII}})\sigma_{\text{dust}}(\lambda)Z/Z_0$  in each cell, where  $Z$  is the gas metallicity and  $f_{\text{ion}} = 0.01$  is a free parameter representing the relative dust abundance in ionised gas. The effective dust cross-section per H atom  $\sigma_{\text{dust}}(\lambda)$  and the  $Z_0 = 0.005$  parameter are normalised to the Small Magellanic Cloud (SMC) extinction curve, as in e.g. Laursen et al. (2009) and Smith et al. (2018). As noted in Laursen et al. (2009), the SMC is hosting younger stellar populations than the Milky-Way (MW) or Large Magellanic Cloud (LMC) so the SMC normalisation is presumably more appropriate when applied to low-mass galaxies at high redshift like in SPHINX (see also Reddy et al. 2012). These authors also show that the Ly $\alpha$  escape fraction from galaxies varies only by a few percent when switching from the SMC to the LMC normalisation<sup>2</sup>.

In the present study, we run RASCAS on all galaxies identified in the catalog at  $z = 6, 7, 8,$  and  $9$ , both for Ly $\alpha$  and the UV continuum 1500 Å band in order to compare SPHINX results with existing observational data. The main goal of the current study being the analysis of the co-evolution of the Ly $\alpha$  intrinsic properties, internal attenuation by dust and IGM transmission, we explicitly describe these three steps separately in the following subsections.

### 2.2.2 Ly $\alpha$ and UV emission

The intrinsic emission of Ly $\alpha$  and UV photons from each galaxy is computed from the gas and stars within  $r_*$  respectively. The Ly $\alpha$  production occurs through two different channels, namely recombinations and collisions, arising from the gas cells. The total number of isotropically-emitted Ly $\alpha$  photons per unit time in a gas cell is given by  $\dot{N}_{\text{Ly}\alpha}^{\text{intr}} = \dot{N}_{\text{Ly}\alpha,\text{rec}}^{\text{intr}} + \dot{N}_{\text{Ly}\alpha,\text{coll}}^{\text{intr}}$  where:

$$\dot{N}_{\text{Ly}\alpha,\text{rec}}^{\text{intr}} = n_e n_{\text{HII}} \epsilon_{\text{Ly}\alpha}^{\text{B}}(T) \alpha_{\text{B}}(T) dV \quad (1)$$

$$\dot{N}_{\text{Ly}\alpha,\text{coll}}^{\text{intr}} = n_e n_{\text{HI}} C_{\text{Ly}\alpha}(T) dV$$

For the recombination term  $\dot{N}_{\text{Ly}\alpha,\text{rec}}^{\text{intr}}$ ,  $n_e$  and  $n_{\text{HII}}$  are respectively the electron and proton number densities, directly predicted by the simulation.  $\alpha_{\text{B}}(T)$  is the case B recombination coefficient (Hui & Gnedin 1997),  $\epsilon_{\text{Ly}\alpha}^{\text{B}}(T)$  is the fraction of recombinations leading to a Ly $\alpha$  emission (Cantalupo et al. 2008), and  $dV$  is the volume of the cell. For the collision term  $\dot{N}_{\text{Ly}\alpha,\text{coll}}^{\text{intr}}$ ,  $n_{\text{HI}}$  is the number density of neutral H atoms, and  $C_{\text{Ly}\alpha}(T)$  is the rate of collisional excitations from  $1s$  to  $2p$ . In practice, we cast 20,000 photon packets per galaxy and sample the frequencies in the rest-frame of the cells according to a Gaussian distribution with a width set by the local thermal velocity of the gas and centred on the Ly $\alpha$  resonance wavelength  $\lambda_{\alpha} = 1215.67$  Å.

For the UV continuum, the intrinsic stellar emission (directly given by the BPASS library) is distributed over  $10^6$  photon packets per galaxy emitted in the rest-frames of the star particles. A detailed description of the spatial and spectral sampling procedures is given in Section 2 of Michel-Dansac et al. (2020).

<sup>2</sup> Shallower extinction curves (based on e.g. the MW or SN-like dust formation scenarios; Gallerani et al. 2010) would further increase escape fractions compared to the SMC or LMC cases by a few percent, so the impact on our results would be almost negligible.

### 2.2.3 Internal Ly $\alpha$ and UV radiative transfer

To follow the radiation transport through the ISM and CGM, we define a characteristic radius,  $R_{\text{CGM}} = 10r_*$ , at which we evaluate the escape fraction for Ly $\alpha$  and the UV continuum before photons enter the IGM. The choice of this particular value for the CGM radius is twofold. First, this radius needs to be large enough such that the escape fraction is converged and that dust attenuation is no longer effective beyond  $R_{\text{CGM}}$ . The second reason is inherent to Ly $\alpha$  RT numerical experiments in the IGM in which we assume that a Ly $\alpha$  photon is removed from the line of sight (i.e. *not transmitted* to the observer) if it scatters once along its path in the IGM (see the next section). It is well-known that Ly $\alpha$  photons keep being resonantly scattered in and out of the line of sight in the densest parts of the CGM, leading to the observed extended emission around star forming galaxies (Steidel et al. 2011; Wisotzki et al. 2016). Thus, a common assumption is to choose  $R_{\text{CGM}}$  such that the number of photons that still scatter at this scale becomes small (see e.g. Laursen et al. 2011; Gronke et al. 2020). We perform a series of convergence tests that are presented in Appendix A from which we choose to set  $R_{\text{CGM}}$  to  $10r_*$  to separate the internal RT (i.e. ISM and CGM) and the IGM RT regions. Note that the shape of the Ly $\alpha$  profiles and the Ly $\alpha$  escape fractions are only weakly dependent on our choice of the  $R_{\text{CGM}}$  values (see Section 3.6).

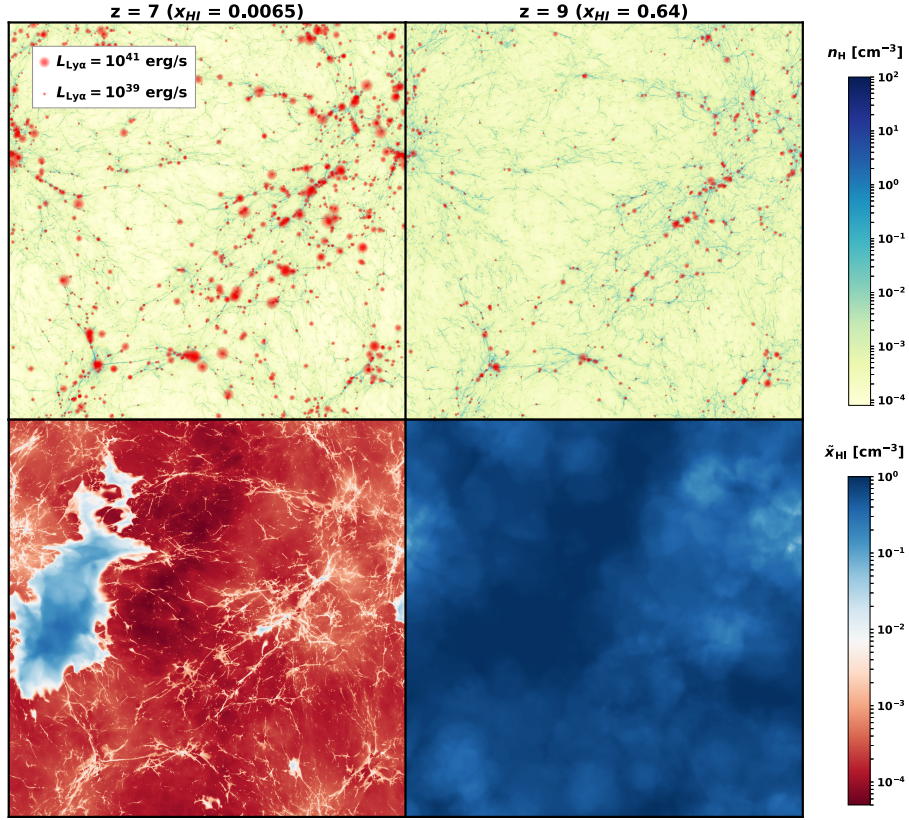
In practice, the angle-averaged escape fractions after internal transfer in the ISM/CGM are computed as follows:  $f_{\text{esc}} = \int \dot{N}_{\lambda}^{\text{CGM}}(hc/\lambda)d\lambda / \int \dot{N}_{\lambda}^{\text{intr}}(hc/\lambda)d\lambda = L_{\lambda}^{\text{CGM}}/L_{\lambda}^{\text{intr}}$  where  $\dot{N}_{\lambda}^{\text{intr}}$  and  $\dot{N}_{\lambda}^{\text{CGM}}$  are the total number of emitted photons per unit time and the total number of emitted photons per unit time that escape at  $R_{\text{CGM}}$ .  $L_{\lambda}^{\text{intr}}$  and  $L_{\lambda}^{\text{CGM}}$  refer to the intrinsic luminosity and the dust-attenuated luminosity respectively. We do not choose any particular direction to estimate the escape fractions, so  $f_{\text{esc}}$  is an angle-average quantity computed by summing photons over all directions. Therefore,  $L_{\lambda}^{\text{CGM}}$  corresponds to the mean escaping luminosity.

### 2.2.4 IGM Ly $\alpha$ radiative transfer

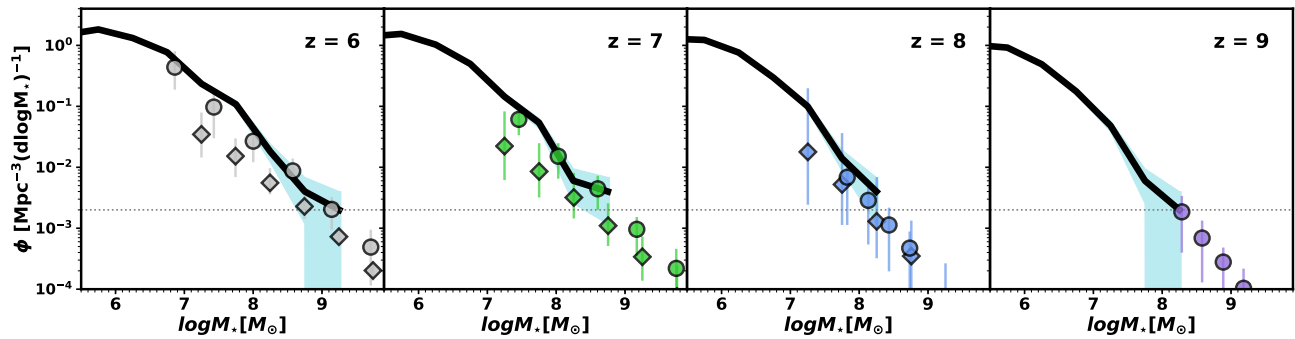
Once they reach  $R_{\text{CGM}}$ , Ly $\alpha$  photons continue their propagation through the IGM in the whole simulation volume using the periodic boundary conditions. We modify the RASCAS code to account for the Hubble flow by adding a velocity component,  $V_{\text{hub}}$ , to the cell gas velocity. The Ly $\alpha$  scattering probability is a sharp function centred at the Ly $\alpha$  resonance (i.e. a Voigt profile  $\Phi(x)$ ) that varies sensitively with  $x$ , the frequency shift expressed in Doppler units. It is therefore important to compute it at the correct  $x$ , especially when the Hubble flow within a given cell becomes non negligible compared to the thermal gas velocity  $V_{\text{th}}$  (see e.g. Jensen et al. 2013; Behrens et al. 2019). Hence, we introduce an adaptive scheme to propagate photons within cells in order to accurately evaluate the probability of interaction between a Ly $\alpha$  photon and an H atom. In practice, photons are walked over substeps in velocity space that remain small compared to the gas thermal motion in the cell and the local variation of the Voigt profile.

If a photon ever happens to scatter, we consider that it is removed from the line of sight and will not transmit to the observer. Alternatively, a photon is transmitted if it can travel a sufficiently large distance without being absorbed. To assess this stopping criterion, we follow Loeb & Rybicki (1999) and compute the proper distance,  $d_{\text{tr}}(z)$ , at which an expanding, homogeneous, and neutral IGM becomes transparent to Ly $\alpha$  photons (i.e. where the opacity equals one). Interestingly, this distance is nearly independent of  $z$  and corresponds to  $\approx 1$  pMpc at  $z = 6 - 9$ , i.e. roughly the physical





**Figure 2.** Projected maps of the SPHINX volume at  $z = 7$  (left) and  $z = 9$  (right). The top panels represent the hydrogen mass-weighted density distribution ( $n_{\text{H}}$ ). LAEs are painted on top of the density map as red dots. The size of the dots scales with the  $\text{Ly}\alpha$  luminosity after internal and IGM transfer, ranging from  $10^{38}$  to  $10^{42}$   $\text{erg s}^{-1}$ . The bottom panels show the local volume-weighted hydrogen neutral fraction,  $\bar{x}_{\text{HI}}$ .



**Figure 3.** Stellar mass functions at  $z = 6, 7, 8$  and  $9$ . We compare our results (black curves) with observational data from Song et al. (2016) (diamonds) and Bhatawdekar et al. (2019) (circles). The cyan shaded area in each panel represents the statistical error ( $\propto \sqrt{N}$ ) on the number counts in each bin of  $\log M_*$ . The grey horizontal dotted line indicates our volume limit of one object per bin.

size of the SPHINX box. We have performed a series of tests to make sure that our  $\text{Ly}\alpha$  transmissions are not affected by the exact  $d_{\text{tr}}(z)$  value. We find that the results are well converged if we use  $10d_{\text{tr}}$  (i.e. a Hubble-velocity shift of  $V_{\text{hub}} \gtrsim 6000 \text{ km s}^{-1}$ ) and we thus opt for this value. For photons emerging blueward of  $\text{Ly}\alpha$  from the CGM, we require an additional travelled distance,  $d_{\text{blue}\rightarrow\text{red}}$ , corresponding to the time needed to red-shift past the resonance such that blue photons need to travel  $(10d_{\text{tr}} + d_{\text{blue}\rightarrow\text{red}})$  to be transmitted.

As mentioned in the previous section, the escape fractions

from the CGM are computed by averaging over all directions. Similarly, we define the IGM transmission as the ratio of the total transmitted luminosity to the total of escaped luminosity:  $T_{\text{IGM}} = \int \dot{N}_{\text{Ly}\alpha}^{\text{IGM}} (hc/\lambda) d\lambda / \int \dot{N}_{\text{Ly}\alpha}^{\text{CGM}} (hc/\lambda) d\lambda$  where  $\dot{N}_{\text{Ly}\alpha}^{\text{IGM}}$  is the total number of IGM transmitted  $\text{Ly}\alpha$  photons per unit time. The transmitted  $\text{Ly}\alpha$  luminosity is therefore given by  $L_{\text{Ly}\alpha}^{\text{IGM}} = T_{\text{IGM}} f_{\text{esc}} L_{\text{Ly}\alpha}^{\text{intr}}$ .

### 3 RESULTS

To begin with, we present visualisations of the SPHINX simulation at  $z = 7$  and  $z = 9$  (Figure 2). The top panels illustrate the filamentary structure of the hydrogen gas density distribution over the 10 cMpc scale spanned by our simulation. The red dots represent individual LAEs with the dot sizes reflecting the observed Ly $\alpha$  intensity of each object (i.e. after internal and IGM transfer) which vary from  $10^{38}$  to  $10^{42}$  erg s $^{-1}$  here. Along with Figure 1, these images emphasise the broad range of physical scales probed by SPHINX. The two bottom panels of Figure 2 illustrate the patchy reionization process captured by SPHINX by showing maps of the volume-weighted hydrogen neutral fraction. While the IGM is still highly neutral at  $z = 9$  ( $x_{\text{HI}} \approx 0.64$ ), it ionises rapidly over about 200 Myr to reach  $x_{\text{HI}} \approx 0.007$  by  $z = 7$ .

In the following, we present the main results of the Ly $\alpha$  post-processing of SPHINX at  $z = 6 - 9$ , starting with an analysis of the galaxy properties and a comparison with statistical observational constraints. Then we focus on the redshift evolution of relevant Ly $\alpha$  quantities (luminosity function, EWs, LAE fraction and spectra) and assess the relative effects of IGM transmission and dust attenuation on the visibility of LAEs in the context of cosmic reionisation in SPHINX.

#### 3.1 Stellar mass and UV luminosity functions

As explained in Rosdahl et al. (2018), SPHINX is calibrated on the stellar mass-to-halo mass relation at  $z = 6$  by boosting the number of SN explosions compared to the fiducial value for a Kroupa IMF. Here, we extend the comparison to observational constraints by presenting the stellar mass function (SMF) and the dust-attenuated UV luminosity functions (LF) at  $z = 6, 7, 8$  and  $9$ .

From Figure 3, we see that SPHINX can well reproduce the abundance of galaxies in the stellar mass range probed by SPHINX. Indeed, because of the limited box size of our simulation, rare bright/massive galaxies are missed which means that we do not predict the massive end of the SMF for  $M_{\star} \gtrsim 2 \times 10^9 M_{\odot}$  ( $M_{\star} \gtrsim 2 \times 10^8 M_{\odot}$ ) at  $z = 6$  ( $z = 9$ ). The lack of massive objects is highlighted by the shaded regions in Figure 3 which represent the statistical error in each bin of  $\log M_{\star}$ .

In Figure 4, we show the UV luminosity functions before dust attenuation (dashed lines) and after dust attenuation (solid lines) and compare with existing constraints (top panels). At all redshifts, the dust-attenuated LF is in good agreement with the observational data at magnitudes  $M_{1500} \gtrsim -18$ . Due to the same finite-volume effect already mentioned above, the brightest intrinsic magnitudes found in SPHINX are  $\approx -20$ . Nevertheless, recent deep surveys have pushed the observational limit down to extremely faint magnitudes ( $M_{1500} \gtrsim -13$ ) which allows us to compare our results over a wide dynamical range ( $\approx 6$  mag; Livermore et al. 2017; Bouwens et al. 2015). We find that the abundance of galaxies increases steeply towards faint magnitudes which is in good agreement with observations, although error bars remain large at  $M_{1500} \gtrsim -15$ . Here it is worth pointing out that the apparent flattening of the simulated UV LF (and SMF) at the faint (low-mass) end does not necessarily represent a physical turnover, and may in part be due to mass resolution effects. At the faint-end, the LFs are incomplete because of our selection on stellar mass (we only analyse galaxies with more than  $10^5 M_{\odot}$  in stars). Concerning the low-mass end of the SMF, as discussed in Rosdahl et al. (2018) and Katz et al. (2020) the simulation only barely describes the formation of galaxies in halos

at the atomic cooling limit, which are resolved with only  $\approx 100$  DM particles, and we may thus miss some of the smallest objects.

As highlighted in the bottom panels of Figure 4, the effect of dust is stronger for bright sources. This is a consequence of UV bright galaxies being on average more massive, more star-forming and therefore more metal- and gas-rich. The median dust attenuation  $A_{1500}$ , represented by the curves, is approximately 0.5 dex at the bright end while it becomes negligible at the faint end. This trend is similar to the observed one reported for bright Lyman-Break galaxies at high redshift where the magnitude attenuation evolves from 0.5 dex at  $M_{1500} \approx -19$  to 1.5 dex at  $M_{1500} \approx -22$  (Bouwens et al. 2016). Interestingly, despite this correlation, the  $A_{1500}$  values are widely spread around the median value at all magnitudes and redshifts, and galaxies as faint as  $M_{1500} \approx -14$  can suffer an attenuation up to  $\approx 2.5$ . These outliers typically correspond to objects which experienced a very recent starburst ( $t \lesssim 5$  Myr), indicating the presence of high gas densities and an ongoing production of metals in the SF sites, and thus increasing the attenuation according to our dust model (see Sec. 2.2.3).

#### 3.2 Ly $\alpha$ luminosity functions

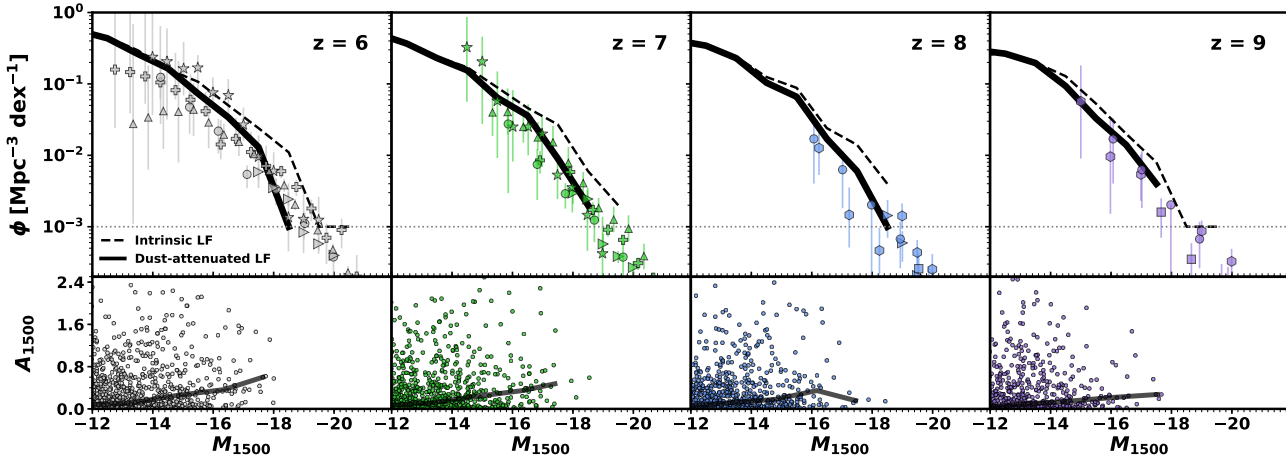
The Ly $\alpha$  luminosity function (Ly $\alpha$  LF) is a fundamental quantity used to probe cosmic reionisation since the Ly $\alpha$  line is expected to be increasingly suppressed by the neutral IGM towards higher redshifts. Statistical samples of LAEs at  $z \gtrsim 6$  have allowed us to put constraints on the bright-end of the Ly $\alpha$  LF, i.e.  $L_{\text{Ly}\alpha} \gtrsim 10^{42}$  erg s $^{-1}$ . While mild evolution is seen below  $z \approx 6$  (Ouchi et al. 2010; Cassata et al. 2011), the characteristic luminosity parameter  $L^*$  appears to drop by a factor 1.4 at  $z = 6.6$  and by a factor 2–3 at  $z \approx 7-7.5$  (Zheng et al. 2017; Itoh et al. 2018) compared to  $z = 5.7$ . Here, we present our predicted Ly $\alpha$  LFs at  $z = 6 - 9$  before dust attenuation, after dust attenuation and after IGM transmission. We estimate the relative impact of the IGM on the redshift evolution of the Ly $\alpha$  LF in our simulation to assess to which extent the observed suppression of the LF can be interpreted as an imprint of reionisation.

##### 3.2.1 Intrinsic Ly $\alpha$ emission

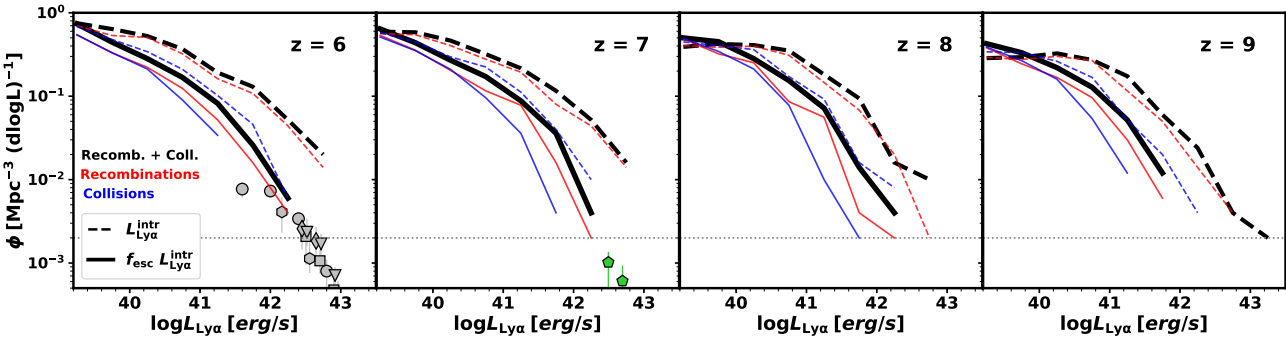
We begin with Figure 5 that shows our predicted Ly $\alpha$  LFs at  $z = 6, 7, 8$  and  $9$  ignoring the effect of IGM. In each panel, the black dashed curves represent the intrinsic LF. Although there are no very massive objects in our sample, we see that intrinsically bright LAEs can be produced, with Ly $\alpha$  luminosities as high as  $\approx 10^{43}$  erg s $^{-1}$ . This is mostly caused by (i) the burstiness of star formation in our simulation (see e.g. Trebitsch et al. 2017) which gives rise to brief but intense Ly $\alpha$  emission episodes and (ii) the use of the BPASS stellar library which boosts the ionising photon budget for a given SF event compared to stellar evolution models without binary stellar systems, and hence the Ly $\alpha$  production under case B recombination. For a constant SFR and a Kroupa IMF (with single stars only), the intrinsic Ly $\alpha$  luminosity from recombination is often estimated to be  $1.7 \times 10^{42} \times (\text{SFR}/[M_{\odot}\text{yr}^{-1}])$  erg s $^{-1}$  (e.g. Dijkstra 2017). For the two reasons mentioned above, we instead find an average relation of  $L_{\text{Ly}\alpha}^{\text{intr}} \approx 3 - 4 \times 10^{42} \times (\text{SFR}/[M_{\odot}\text{yr}^{-1}])$  erg s $^{-1}$  in SPHINX.

As explained in Section 2.2.2, Ly $\alpha$  photons are emitted through two different channels in our study. In Figure 5, we also show the relative contribution of recombinations (in blue) and collisions (in





**Figure 4.** UV luminosity functions (LF; 1500 Å rest-frame) at  $z = 6, 7, 8$  and  $9$  (top). The dashed and solid black curves correspond to the intrinsic and dust-obscured LFs respectively. We compare with observations represented by the symbols: Bouwens et al. (plus signs 2015, 2017), Bhatwdekar et al. (circles 2019), Finkelstein et al. (right triangles 2015), Livermore et al. (2017, stars), Atek et al. (2015, 2018, upward triangles), Ishigaki et al. (hexagons 2018), and Oesch et al. (squares 2013). The data points of Atek et al. are shifted by  $-0.1$  mag for clarity. The grey horizontal dotted line indicates our volume limit of one object per bin. We also show the dust attenuation,  $A_{1500}$ , as a function of UV magnitude (bottom). Dots correspond to individual galaxies and the line shows the median  $A_{1500}$  per  $M_{1500}$  bin.



**Figure 5.**  $\text{Ly}\alpha$  luminosity functions at  $z = 6, 7, 8$  and  $9$  without IGM transmission. In each panel, the dashed and solid black lines account for the total  $\text{Ly}\alpha$  luminosity before and after dust attenuation respectively. The relative contribution of recombinations (in red) and collisions (in blue) are also highlighted. The grey data points represent  $z \approx 5.5 - 6.5$  observational data from Konno et al. (2017, squares), Herenz et al. (2019, diamonds), Cassata et al. (2011, hexagons), and Santos et al. (2016, downward triangles). At this redshift, the only constraints on the faint-end of the  $\text{Ly}\alpha$  LF which are deep enough to be directly comparable to our predictions come from the MUSE-Deep survey (Drake et al. 2017, grey circles) ( $L_{\text{Ly}\alpha}^{\text{intr}} \gtrsim 3 \times 10^{41} \text{ erg s}^{-1}$ ). The green pentagons show the LF measured by Konno et al. (2017) at  $z \approx 6.6$ . The grey horizontal dotted line indicates our volume limit of one object per bin.

red) to the  $\text{Ly}\alpha$  LFs. At all redshifts, recombinations strongly dominate the  $\text{Ly}\alpha$  intrinsic budget over collisions in brighter LAEs ( $L_{\text{Ly}\alpha} \gtrsim 10^{41} \text{ erg s}^{-1}$ ) whereas both channels contribute equivalently in fainter objects. Note that the collisional excitation rate,  $C_{\text{Ly}\alpha}(T)$ , is highly sensitive to the temperature so its exact contribution will depend on the subgrid physics that can affect the thermal properties of the gas, in particular the feedback model. In addition, the collisional excitation rate is poorly estimated in gas cells where the net cooling time is small compared to the simulation timestep. We therefore make the conservative approximation of setting  $\text{Ly}\alpha$  collisional emission to zero in cells which net cooling time is less than five times the timestep value. We tested that our results are not sensitive to this choice and that it has a minor impact on the total budget of  $\text{Ly}\alpha$  collisional emission (Blaziot et al., in prep).

### 3.2.2 $\text{Ly}\alpha$ transfer in the ISM and CGM

Due to the complex nature of the  $\text{Ly}\alpha$  resonant line, it is paramount to account for the radiative transfer of  $\text{Ly}\alpha$  photons in the ISM and CGM (which we refer to as internal RT for simplicity) to realistically model the evolution of LAEs during reionisation. Based on the procedure detailed in Section 2.2.3, we construct the dust-attenuated  $\text{Ly}\alpha$  LF and show our results in Figure 5 (black solid curves).

At all redshifts, the internal RT suppresses  $\text{Ly}\alpha$  emission by a factor 1.5 – 3 on average. Most of SPHINX LAEs are too faint to be compared with observations except at  $z = 6$ , where our LF is in reasonable agreement with the deep MUSE constraints ( $L_{\text{Ly}\alpha} \lesssim 10^{42} \text{ erg s}^{-1}$ ), though slightly above (but we remind that we have ignored IGM transmission for now). Similarly to the UV LF, we predict that the LF keeps rising steeply at  $L_{\text{Ly}\alpha} \lesssim 10^{42} \text{ erg s}^{-1}$  despite the mass resolution effect discussed earlier which implies that the number density is even under-estimated at the very faint end ( $L_{\text{Ly}\alpha} \lesssim 10^{40} \text{ erg s}^{-1}$ ).

$\text{Ly}\alpha$  photons produced by recombinations dominate the bright-

end of our LF after RT in the ISM and the CGM, as was already the case for the intrinsic emission. Nevertheless, Ly $\alpha$  radiation emitted through collisional excitation has a somewhat higher escape fraction. This is because recombinations mainly occur in dense, metal-rich, star-forming regions where dust extinction is generally strong while collisional emission can also be generated in the more diffuse and metal-poor parts of the ISM.

### 3.2.3 Impact of IGM transmission on the Ly $\alpha$ LF

As discussed in the introduction, only a fraction of the Ly $\alpha$  flux escaping galaxies can reach the observer due to H $\text{I}$  absorption by the IGM. Yet, understanding how much of the observed Ly $\alpha$  suppression is connected to the IGM neutral fraction at a given redshift remains elusive. In the current and following sections, we intend to quantify the impact of IGM transmission on the Ly $\alpha$  LF.

From Figure 6, we see that the IGM has quite a significant impact on the LF (solid orange curves), in particular towards higher redshifts. Of course, this is expected because  $x_{\text{HI}}$  increases from  $10^{-4}$  at  $z = 6$  to 0.6 at  $z = 9$  in the simulation. At  $z = 6$ , the IGM transmission  $T_{\text{IGM}}$  is about 50% whereas it drops to  $\approx 5\text{-}10\%$  at  $z = 9$ , clearly reflecting the evolution of the ionisation state of the diffuse IGM (see Section 3.5).

Although the comparison with observational data is obviously dubious at these high redshifts and low luminosities, we note that our  $z = 6$  IGM-attenuated LF falls near the MUSE-deep constraints at  $L_{\text{Ly}\alpha} \lesssim 10^{42} \text{ erg s}^{-1}$ . Due to our limited box size, it is impossible to draw any conclusion regarding the bright-end but it is worth noting that a crude extrapolation "by eye" of our LFs at  $z = 6$  and 7 does not seem inconsistent with the data at  $L_{\text{Ly}\alpha} \approx 10^{42-43} \text{ erg s}^{-1}$ . While there is no compelling observational constraints at  $z = 8-9$ , we nevertheless plot as a guide the Schechter and power-law fits derived by Matthee et al. (2014) and extrapolated to  $L_{\text{Ly}\alpha} \approx 10^{41} \text{ erg s}^{-1}$ . Based on this (uncertain) comparison, our LF falls in the expected range of densities at such low luminosities.

Our predicted LFs indicate that numerous LAEs should be detectable during the heart of reionisation era, assuming that detection limits are pushed further down by a couple orders of magnitude. The identification of LAEs in typical narrow-band surveys is however often based on colour selections, or equivalently, EW thresholds. We therefore also show in Figure 6 the effect of EW cuts ( $EW_{\text{thres}} > 25 \text{ \AA}$  and  $> 50 \text{ \AA}$ ). As discussed in Appendix C, the Ly $\alpha$  EWs are larger for brighter LAEs in our simulation (see Figure C1), so that EW selections will predominantly remove galaxies at the faint end of the LF. Even so, taking a conservative cut of  $EW_{\text{thres}} > 50 \text{ \AA}$ , we predict a high number density of LAEs at  $L_{\text{Ly}\alpha} \approx 10^{40} \text{ erg s}^{-1}$  at  $z = 8$  where the IGM neutral fraction is still  $\approx 35\%$ .

### 3.2.4 Ly $\alpha$ LF evolution with redshift

In Figure 7, we highlight the redshift evolution of the Ly $\alpha$  LF by plotting together the LFs based on intrinsic luminosities (top panel), dust-attenuated luminosities (middle panel), and IGM-transmitted luminosities (bottom panel). On the one hand, we clearly see that the intrinsic and dust-attenuated LFs remain nearly constant from  $z = 6$  to  $z = 9$ , highlighting the very weak evolution of the internal properties driving Ly $\alpha$  emission and escape from galaxies during the EoR. On the other hand, the substantial effect of IGM is completely dominating the variation of the visibility of LAEs during this period. While the Ly $\alpha$  LF is unchanged at  $z \lesssim 7$  as long as the IGM is highly ionised ( $x_{\text{HI}} < 0.01$ ), the Ly $\alpha$  transmission drops

significantly from  $z = 7$  to  $z = 9$ . This strong suppression is directly due to the IGM neutral fraction increasing rapidly at  $z \gtrsim 7$  in SPHINX, i.e.  $x_{\text{HI}} \approx 0.35$  at  $z = 8$  and  $x_{\text{HI}} \approx 0.65$  at  $z = 9$  (see Figure 9 of Rosdahl et al. 2018).

In order to get a more quantitative assessment of the impact of the IGM, we define the decrement of the observed Ly $\alpha$  luminosity density (in  $\text{erg s}^{-1} \text{ Mpc}^{-3}$ , Inoue et al. 2018):

$$\Delta\rho_{\text{IGM}\rightarrow\text{CGM}}^{\text{Ly}\alpha} = \log\rho_{\text{IGM}}^{\text{Ly}\alpha} - \log\rho_{\text{CGM}}^{\text{Ly}\alpha} \quad (2)$$

where  $\log\rho_{\text{IGM}}^{\text{Ly}\alpha}$  and  $\log\rho_{\text{CGM}}^{\text{Ly}\alpha}$  are the Ly $\alpha$  luminosity densities after IGM transmission and after internal transfer, integrated down to our completeness limit ( $L_{\text{Ly}\alpha} \approx 10^{40} \text{ erg s}^{-1}$ ). The orange curve in Figure 8 shows that the decrement is nearly constant from  $z = 6$  to  $z = 7$  and decreases significantly by  $\approx 0.5$  dex from  $z = 7$  to  $z = 8$  and by  $\approx 1$  dex from  $z = 8$  to  $z = 9$  due to reduced IGM transmission.

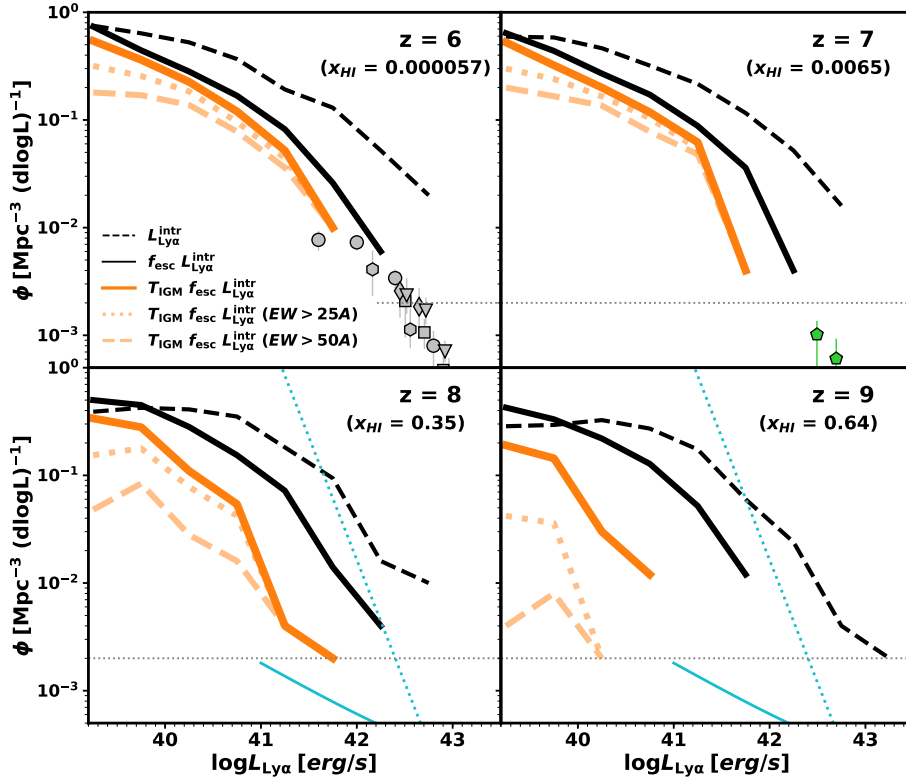
For comparison, we also plot the decrements of the dust-attenuated Ly $\alpha$  and UV luminosity densities relative to the intrinsic ones ( $\Delta\rho_{\text{CGM}\rightarrow\text{intr}}^{\text{Ly}\alpha}$  and  $\Delta\rho_{\text{CGM}\rightarrow\text{intr}}^{\text{UV}}$ ; black and blue curves respectively). Both remain nearly unchanged from  $z = 6$  to  $z = 9$  which suggests that any significant detectable evolution in the Ly $\alpha$  LF during the EoR should be fully attributed to a rapid increase of  $x_{\text{HI}}$ . Finally, we note that the offset between  $\Delta\rho_{\text{CGM}\rightarrow\text{intr}}^{\text{Ly}\alpha}$  and  $\Delta\rho_{\text{CGM}\rightarrow\text{intr}}^{\text{UV}}$  reflects the differential escape fractions from galaxies between Ly $\alpha$  and UV photons. Stellar (non-ionising) UV continuum usually escapes galaxies more easily than Ly $\alpha$ , especially for more massive, dustier sources, which is a direct consequence of the enhanced probability of resonant Ly $\alpha$  photons to be destroyed by dust grains on their way out of the galaxy (Verhamme et al. 2008; Hayes et al. 2011; Garel et al. 2015).

### 3.2.5 Abundance of very faint LAEs

Once internal RT and IGM transmission are accounted for, the dynamical range spanned by LAEs in our  $10^3 \text{ cMpc}^3$  simulation is restricted to Ly $\alpha$  luminosities below  $L_{\text{Ly}\alpha} \approx 10^{42} \text{ erg s}^{-1}$ . Thanks to its fine mass-resolution, SPHINX is however able to resolve low-mass systems, allowing us to investigate the very faint-end of the Ly $\alpha$  LF. As shown in Figure 6, Ly $\alpha$  emitters at such low levels are unfortunately still out of reach in current surveys and it is not clear to which extent the LF keeps rising at the faint-end. Still, the recent detection of extended Ly $\alpha$  emission at  $> 1 \text{ cMpc}$  scale at  $z \approx 3-5$  in the MUSE Extremely Deep Field provides clues for the existence a numerous population of ultra-faint LAEs, possibly down to  $L_{\text{Ly}\alpha} \approx 10^{37} \text{ erg s}^{-1}$  and assuming a steep LF slope (Bacon et al. 2021).

Such sources should sit predominantly in low-mass DM haloes but, as extensively discussed in the literature, the feedback from stellar radiation can prevent the formation of galaxies in these systems due to photoheating and gas inflow suppression (Okamoto et al. 2008). Using a smaller SPHINX simulation run than ours (but with the same baryonic physics and BPASS library), Katz et al. (2020) have shown that reionization has a significant impact on the gas content of dwarf galaxies at  $z \gtrsim 6$  but that, meanwhile, most haloes below the atomic cooling limit can remain self-shielded against ionising radiation and can thus keep forming stars even after the end of reionization.

As can be seen from the cumulative Ly $\alpha$  LFs (after IGM) at  $z = 7$  and 9 (Figure 9), very faint LAEs do exist in our simulation and we find that their cumulative number density keeps rising until  $L_{\text{Ly}\alpha} < 10^{37} \text{ erg s}^{-1}$  which confirms that low-mass haloes keep



**Figure 6.** Effect of IGM transmission on the Ly $\alpha$  luminosity functions at  $z = 6, 7, 8$  and  $9$  shown by the solid orange curves. For comparison, we plot again the Ly $\alpha$  LF before and after dust attenuation (dashed and solid black curves respectively). To mimic roughly the narrow-band selection in wide-field surveys, we also construct the IGM-transmitted LFs with two different EW cuts:  $EW_{\text{thresh}} > 25 \text{ \AA}$  (dotted orange line) and  $> 50 \text{ \AA}$  (dashed orange line). The observed data points at  $z \approx 6$  and  $7$  are identical to those shown in Figure 5. There are no constraints on the faint-end of the Ly $\alpha$  LF at  $z \geq 8$  but, as a guide, we add the power-law and Schechter function best-fits reported by Matthee et al. (2014) extrapolated at low luminosities (dotted and solid cyan curves respectively). The grey horizontal dotted line indicates our volume limit of one object per bin.

forming stars efficiently. We compare our predicted Ly $\alpha$  LF with the best-fit Schechter functions measured by Santos et al. (2016) at  $z \approx 7$ , assuming three different faint-end slopes,  $\alpha$ . Our  $z = 7$  LF seems to be more consistent with moderately steep values ( $\alpha \approx -1.5$ ) but it is difficult to assess because we have restricted our sample to galaxies more massive than  $10^5 M_{\odot}$ . As shown in Figure C1, the brightest Ly $\alpha$  luminosities in galaxies at our stellar mass threshold correspond to roughly  $L_{\text{Ly}\alpha} \approx 10^{40} \text{ erg s}^{-1}$  at all redshifts considered here. This means that our LAE sample is incomplete below this value such that our LFs appear shallower than they should. Based on our simulation, the expected numbers of LAEs at  $L_{\text{Ly}\alpha} \gtrsim 10^{37} \text{ erg s}^{-1}$  at  $z = 7$  ( $x_{\text{HI}} = 0.007$ ;  $\approx 2$  per  $\text{cMpc}^3$ ) and  $z = 9$  ( $x_{\text{HI}} = 0.64$ ;  $\approx 0.8$  per  $\text{cMpc}^3$ ) may therefore be seen as lower limits, suggesting that the abundance of extremely faint LAEs is high towards the end of the EoR.

We note that our predicted LFs are in slight disagreement with the results of Laursen et al. (2019) who simulated the visibility of LAEs at  $z \approx 9$  using zoom-in hydrodynamics simulations applied to a large cosmological DM run. As our study, they follow the Ly $\alpha$  radiation from their emission sites through the ISM, CGM and IGM allowing for accurate estimation of the internal RT and IGM transmission. Two relevant differences though relate to the ionising transfer, which they perform as a post-processing step, and their dynamical range which covers more massive haloes than ours on average. Our  $z = 9$  LF (after IGM) is only overlapping with the one of Laursen et al. (2019) at  $L_{\text{Ly}\alpha} \approx 10^{41} \text{ erg s}^{-1}$  where it roughly matches their intrinsic LF. Once they account for Ly $\alpha$  RT, they pre-

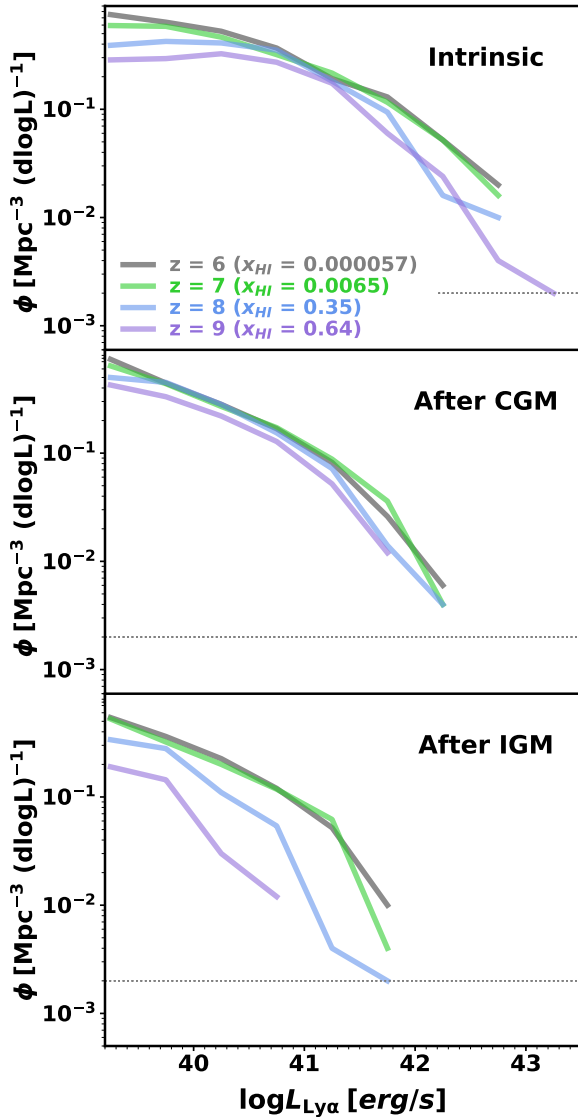
dict an abundance of LAEs significantly smaller than in SPHINX at this particular luminosity. The reasons for the discrepancy are unclear and could arise from incompleteness at the faint-end in the sample of Laursen et al. (2019) or from cosmic variance effects that can be significant especially in moderate volume sizes like in SPHINX.

Based on their simulation, Laursen et al. (2019) predict that very few LAEs can be detected in the UltraVISTA survey with a 168h exposure (i.e. their probability of detecting more than one LAE is 1%), corresponding to Ly $\alpha$  detection limit of  $\approx 10^{43} \text{ erg s}^{-1}$  at  $z = 8.8$ . We cannot make number count predictions at such bright Ly $\alpha$  luminosities with SPHINX but the significantly higher LAE number density that we predict at  $L_{\text{Ly}\alpha} \approx 10^{41} \text{ erg s}^{-1}$  compared to Laursen et al. (2019) suggests that more optimistic numbers of detections can be achieved with such deep surveys during the EoR.

### 3.3 Ly $\alpha$ equivalent widths

Defined as the ratio of Ly $\alpha$  emission over UV continuum, the equivalent width (EW) encodes valuable information about galaxies such as the metallicity and age of the underlying stellar population (e.g. Hashimoto et al. 2017). During the EoR, the differential evolution of EWs can also be used as a proxy for IGM neutrality (Mason et al. 2018; Jung et al. 2020).

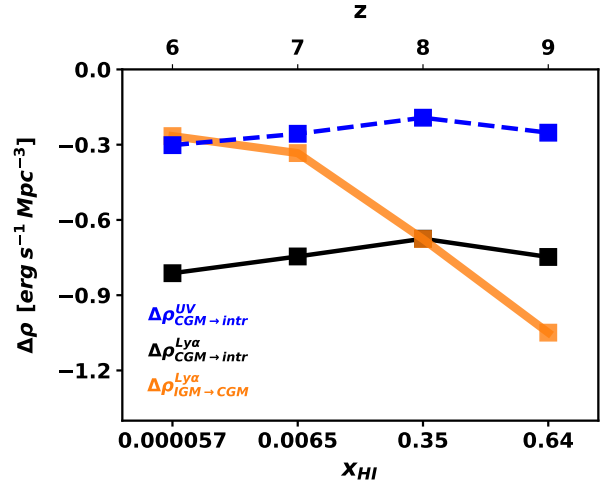
As is often done with observational datasets, we compute the EWs by estimating the continuum around the Ly $\alpha$  wavelength from far-UV bands to measure the UV slope,  $\beta_{\text{UV}}$ , and extrapolating



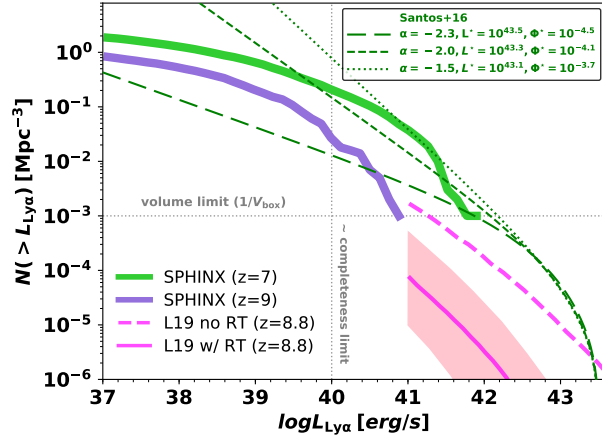
**Figure 7.** Redshift evolution of the Ly $\alpha$  luminosity functions highlighting the contribution of intrinsic emission, internal absorption and IGM transmission. The grey horizontal dotted line indicates our volume limit of one object per bin.

the flux level at 1216 Å (Hashimoto et al. 2017). In practice, we predict the intrinsic and dust-attenuated emissivities at 1500Å and 2500Å and we measure  $\beta_{UV}$  before and after internal RT in order to compute the intrinsic and dust-attenuated continuum luminosity densities at 1216 Å,  $L_{1216}$ . The Ly $\alpha$  EWs are simply obtained as the ratio of the Ly $\alpha$  intrinsic and dust-attenuated luminosities by these values:  $EW^{\text{intr}} = L_{Ly\alpha}^{\text{intr}}/L_{1216}^{\text{intr}}$  and  $EW^{\text{CGM}} = L_{Ly\alpha}^{\text{CGM}}/L_{1216}^{\text{CGM}}$ . To estimate the IGM-transmitted EW, we multiply the dust-attenuated EW by the IGM transmission such that  $EW^{\text{IGM}} = T_{\text{IGM}}EW^{\text{CGM}}$ .

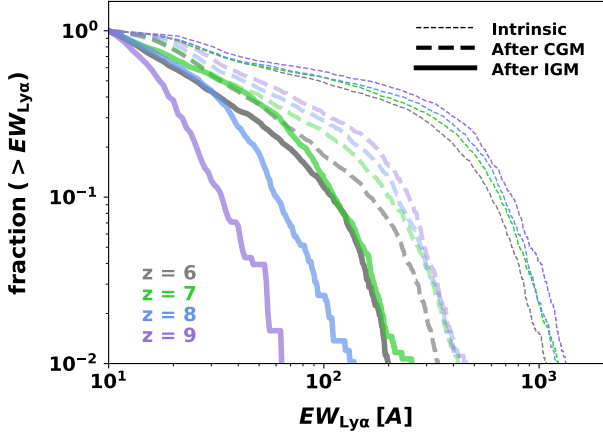
Figure 10 shows the resulting cumulative distributions at  $z = 6, 7, 8,$  and  $9$  normalised to the total number of LAEs in each snapshot. A first striking aspect is the high intrinsic EW values that are reached in some galaxies. About ten percent of LAEs have  $EW$  greater than 500Å and  $\approx$  one percent of them produce  $EW$  above 1000 Å. For a standard IMF and solar metallicity, the maximum  $EW$  produced through recombination in star-forming regions is about 250 Å. However, with metallicities of  $0.02Z_{\odot}$  that are plausibly



**Figure 8.** Evolution of the Ly $\alpha$  and UV luminosity density decrement as a function of the IGM neutral fraction,  $x_{\text{HI}}$ . The top x-axis shows the redshifts of the corresponding snapshots of the SPHINX simulation. The Ly $\alpha$  and UV density have been integrated down to  $L_{Ly\alpha} = 10^{40}$  erg s $^{-1}$  and  $M_{1500} = -12$  respectively which approximately corresponds to our completeness limit.



**Figure 9.** Cumulative Ly $\alpha$  luminosity function at  $z = 7$  (green) and  $z = 9$  (purple) after IGM transmission in SPHINX. We show that the Ly $\alpha$  LF keeps rising down to very faint Ly $\alpha$  luminosities during the EoR. The horizontal and vertical grey lines indicate our volume and Ly $\alpha$ -luminosity completeness limits respectively. These imply that we cannot predict the number density of LAEs below  $10^{-3}$  Mpc $^{-3}$  due to our limited simulation volume, and that we under-estimate the number density of LAEs below  $L_{Ly\alpha} \approx 10^{40}$  erg s $^{-1}$  because we choose to only identify galaxies more massive than  $M_{\star} = 10^5 M_{\odot}$  (i.e. 100 star particles) in our study. At  $z = 9$ , we compare our prediction with the LF of Laursen et al. (2019) who model bright-end LAEs using cosmological hydrodynamic simulations post-processed with ionising and Ly $\alpha$  transfer at  $z = 8.8$ . Their intrinsic LF is represented by the dashed magenta curve. The solid magenta curve shows their predicted LF after Ly $\alpha$  RT in the haloes and in the IGM and is therefore comparable to our LF at  $z = 9$ . The pink shaded area corresponds to the  $1\sigma$  directional variation. The dotted, dashed, and solid dark green lines are the extrapolated Schechter best-fits of the observed Ly $\alpha$  LFs from Santos et al. (2016) at  $z \approx 7$  (see legend for the Schechter parameter values).



**Figure 10.** Redshift evolution of the cumulative Ly $\alpha$  EW distribution showing the effects of dust attenuation and IGM transmission from  $z = 6$  to  $z = 9$ . The thin dotted curves represent the fraction of galaxies with an intrinsic Ly $\alpha$  EW larger than a given value while the thick dashed and solid curves include the effects of internal RT (after CGM) and IGM transmission (after IGM) respectively. Here, we use all galaxies from each snapshot without any UV magnitude selection.

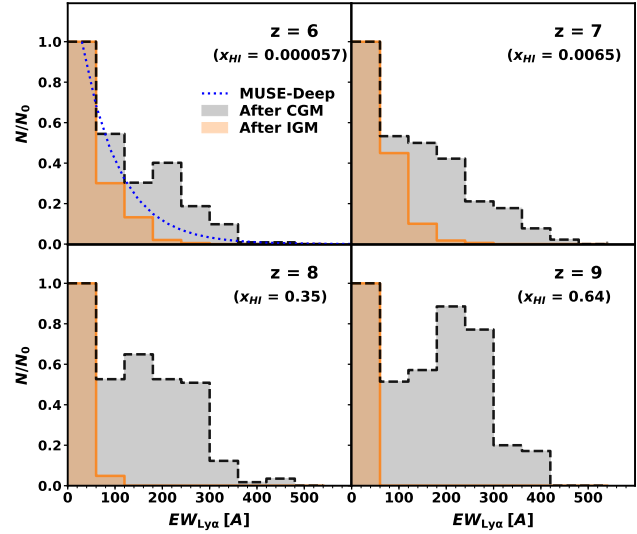
more representative of low-mass galaxies at high redshift, stellar synthesis models can easily produce EWs as large as 400 Å (e.g. Hashimoto et al. 2017). In our simulation, the gas-phase metallicities are comprised between  $0.1Z_{\odot}$  and  $0.001Z_{\odot}$  (see Figure C2) so values of 400 Å are indeed expected. The other two factors able to boost even further the EW above 1000 Å in our simulated galaxies are (i) the use of BPASS which increases the Ly $\alpha$  emissivity for a given SF episode (see Section 3.2.1), and (ii) the contribution of collisional emission that can increase the global Ly $\alpha$  photon budget (see Figure 5). After internal RT, the median EW is about 50 Å but a small fraction of galaxies harbour very large values ( $> 400\text{Å}$ ) at all redshifts. This seems consistent with the recent measurements of Kerutt et al. (in prep.) who report EWs up to 900 Å in deep MUSE observations at  $z = 3 - 6.5$ .

The intrinsic and dust-attenuated EW distributions do not show a strong evolution with redshift but we note that the fraction of high values becomes slightly larger towards higher redshifts (see Figures 10 and C1). A more drastic evolution is seen when looking at the redshift evolution of the distribution of IGM-attenuated equivalent widths. While the  $z = 6$  and  $z = 7$  distributions evolve similarly after IGM transmission, the high-EW tail is cut off when the IGM neutral fraction becomes significant (i.e. at  $z \gtrsim 7-8$  in our simulation). This behaviour can be interpreted as the Ly $\alpha$  LF evolution discussed in Section 7 where the increasing IGM neutrality starts suppressing the Ly $\alpha$  line when  $x_{\text{HI}}$  becomes greater than  $\approx 1\%$ . Overall, our results support the idea that the evolution of the Ly $\alpha$  EW distribution at  $z > 6$  can be used to probe the IGM neutrality during the EoR (Mason et al. 2018).

### 3.4 LAE fraction

The measurement of the fraction of UV-selected galaxies that emit Ly $\alpha$  is a commonly used diagnostic of reionisation. The LAE fraction, or  $X_{\text{Ly}\alpha}$ , is defined as follows:

$$X_{\text{Ly}\alpha}(z) = \frac{N_{\text{LAE}}(z, M_{1500}, EW)}{N_{1500}(z, M_{1500})} \quad (3)$$



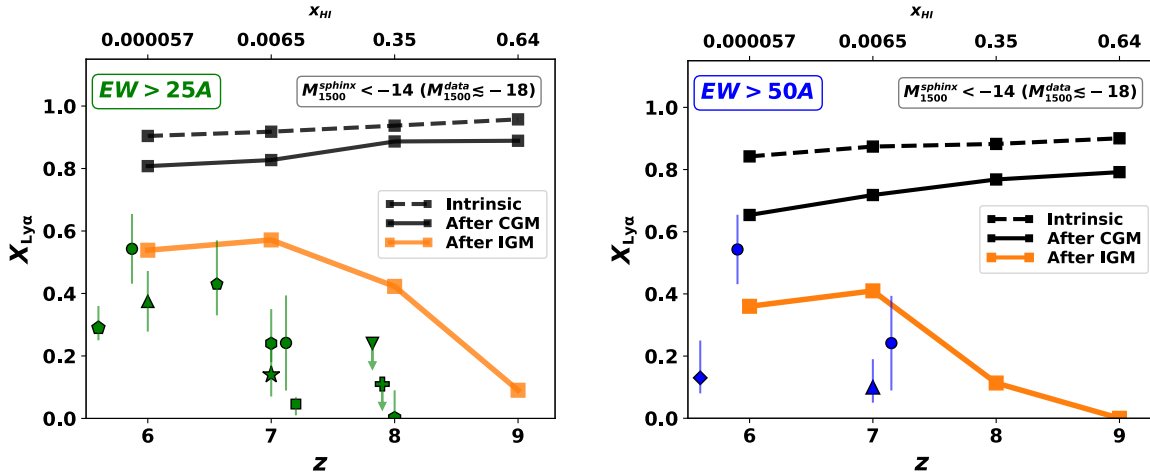
**Figure 11.** Equivalent width distributions at  $z = 6, 7, 8,$  and  $9$ . The grey histograms correspond to the Ly $\alpha$  EW after internal RT while IGM-transmitted values are in orange. For comparison, we overplot the best-fit exponential distributions from the MUSE-Deep survey at  $4.5 < z < 6.6$  (Hashimoto et al. 2017) ( $N \propto \exp(-EW/w_0)$ , where  $w_0 = 79$ ). In this figure, we use a UV magnitude cut of  $-14$  in order to have similar statistics as in the MUSE-Deep data (i.e. at least 100 galaxies per snapshot). The comparison is however mainly illustrative because, in spite of the  $> 10\text{h}$  MUSE exposure-time combined with exquisite HST counterpart data, MUSE-Deep LAEs usually have  $L_{\text{Ly}\alpha} \gtrsim 10^{41} \text{ erg s}^{-1}$  and  $M_{1500} \lesssim -16$  whereas most of our simulated sources are fainter than these values.

where  $N_{1500}(z, M_{1500})$  is the number of galaxies brighter than a fixed UV magnitude limit in a given redshift bin.  $N_{\text{LAE}}(z, M_{1500}, EW)$  is a subsample of  $N_{1500}(z, M_{1500})$  that corresponds to LAEs, i.e. sources with a Ly $\alpha$  equivalent width greater than a typical threshold value ( $EW > 25$  or  $50$  Å are two commonly used values in LAE surveys). Under the assumption that the IGM is the main cause of the apparent fading of the Ly $\alpha$  line at  $z \gtrsim 6$ ,  $X_{\text{Ly}\alpha}$  should decline when  $x_{\text{HI}}$  increases. Such a trend has been reported by many surveys (e.g. Schenker et al. 2012; Pentericci et al. 2018; Fuller et al. 2020) based on samples of galaxies brighter than  $M_{1500} \lesssim -18.5$ . Given that there are only a handful of such bright sources in our simulation, we can only compute  $X_{\text{Ly}\alpha}$  with a lower UV magnitude limit. Using a somewhat arbitrary cut of  $M_{1500} = -14$  allows us to have sufficient statistics (i.e. at least 100 galaxies at each snapshot) to produce a sample size comparable to observational studies such as the MUSE-Deep survey (Hashimoto et al. 2017). We show in Figure 11 the resulting EW distributions where the black and orange histograms represent the EWs after CGM and after IGM respectively. Despite the different UV selection, our predicted EW distributions reproduce reasonably well the one from the MUSE-Deep survey so we keep  $-14$  as our UV detection limit for our study the LAE fraction<sup>3</sup>.

In Figure 12, we compare our  $X_{\text{Ly}\alpha}$  for weak emitters (i.e.  $EW > 25$  Å) and strong emitters (i.e.  $EW > 50$  Å) in the left and

<sup>3</sup> We note that a correlation exists (with a large scatter) between EW and  $M_{1500}$  (see Figure C2) so varying the UV magnitude cut will affect the selected EW distribution, and hence the resulting LAE fraction. Nevertheless, we have checked that setting lower or higher  $M_{1500}$  limits only impacts the overall amplitude of  $X_{\text{Ly}\alpha}$ , and not the shape of its redshift evolution, so this choice does not change our main conclusions (see Figure B1).





**Figure 12.** Fraction of LAEs in the total sample of galaxies,  $X_{\text{Ly}\alpha}$ , as a function of redshift. In the left and right panels, we construct the fraction of galaxies with  $M_{1500} < -14$  (see text) exhibiting an EW greater than 25 and 50 respectively. The LAE fractions based on intrinsic (dashed black lines) and dust-attenuated (solid black lines) UV magnitudes and  $\text{Ly}\alpha$  EWs remain almost constant with  $z$ . The orange curves show  $X_{\text{Ly}\alpha}(z)$  with the inclusion of the effect of IGM transmission in the computation of the  $\text{Ly}\alpha$  EW. The symbols correspond to the following observational measurements: Stark et al. (2010, circles), De Barros et al. (2017, upward triangle), Pentericci et al. (2018, star), Schenker et al. (2012, hexagons), Tilvi et al. (2014, downward triangle), Fuller et al. (2020, pentagons), Hoag et al. (2019, square), Kusakabe et al. (2020, diamond).

right panel respectively. Ignoring the effect of dust and IGM transmission (dashed black curve), we find for both cases that  $X_{\text{Ly}\alpha}$  does not evolve with redshift, indicating that, on average, the intrinsic  $\text{Ly}\alpha$  strength of galaxies remains unchanged relatively to the stellar continuum. The solid black curve represents the LAE fraction by only accounting for dust (i.e. galaxies selected based on their dust-attenuated magnitudes and  $\text{Ly}\alpha$  EW). Again,  $X_{\text{Ly}\alpha}$  does not decline but remains constant (or even slightly increases) at  $z \geq 6$ . It is only when IGM transmission is included to compute the  $\text{Ly}\alpha$  EW that  $X_{\text{Ly}\alpha}$  starts to drop sharply around  $z \approx 7$  (orange curve), which corresponds to the transition between a fully ionised to a partially ionised Universe in our simulation. While the UV continuum is not affected by the IGM, the  $\text{Ly}\alpha$  line can be strongly altered leading to a significant reduction of the EW, and hence a clear drop in  $X_{\text{Ly}\alpha}$ . Nevertheless, we note that  $X_{\text{Ly}\alpha}$  is not extremely sensitive to the evolution of  $x_{\text{HI}}$ . For weak emitters for instance,  $X_{\text{Ly}\alpha}$  is reduced by a factor 5 between  $z = 7$  and  $z = 9$  while the IGM neutrality has increased by a factor of  $\approx 100$  over this period.

As explained earlier, our results are not directly comparable to observations due to the different UV magnitude selection. Nevertheless, we note that the overall shape of the LAE fraction evolution is well recovered by the simulation, especially for  $EW > 25 \text{ \AA}$  where the constraints are the tightest. In that particular case, the observed  $X_{\text{Ly}\alpha}$  declines from  $\approx 50\%$  at  $z \approx 6$  to  $\approx 10\%$  at  $z \approx 8$ . The simulated  $X_{\text{Ly}\alpha}$  spans a similar range as the observations but with a horizontal shift of about  $\Delta z = 1$ , most probably due to the reionisation history in SPHINX. Indeed, as shown in Figure 9 of Rosdahl et al. (2018), the SPHINX neutral fraction is only  $\approx 0.005$  at  $z = 7$  and rapidly increases to  $\approx 0.35$  at  $z = 8$  while the observationally estimated  $x_{\text{HI}}$  value is already  $\approx 0.3 - 0.4$  at  $z = 7$ . This just reflects that cosmic reionisation is achieved too early in SPHINX. The main and remarkable point is that the predicted evolution of  $X_{\text{Ly}\alpha}$  and the amplitude of its decline is clearly tracing the change in the global neutral fraction of the IGM.

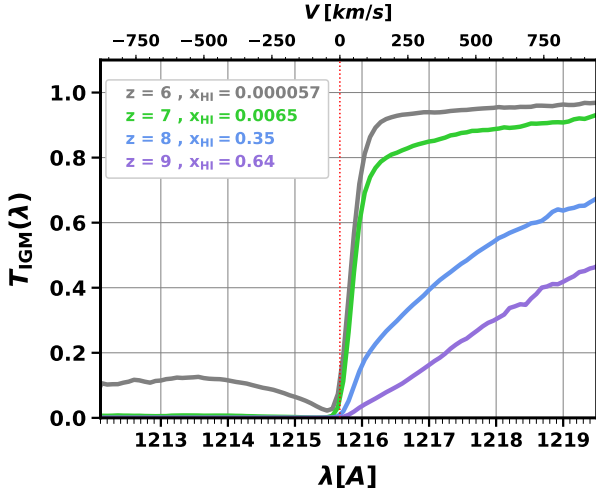
### 3.5 $\text{Ly}\alpha$ IGM transmission

In this section, we first investigate how the IGM transmission varies with respect to the velocity shift from the line center. Then we compare the evolution of the global, blue, and red transmissions as a function of redshift and galaxy properties.

#### 3.5.1 Transmission curve as a function of wavelength

Here we focus on the redshift evolution of our simulated  $\text{Ly}\alpha$  IGM transmission. Figure 13 presents the wavelength dependence of  $T_{\text{IGM}}$  for our four snapshots computed as the mean IGM transmission of  $\text{Ly}\alpha$  photons that escaped the galaxies,  $T_{\text{IGM}}(\lambda)$ . It is worth pointing out that this definition of  $T_{\text{IGM}}(\lambda)$  is somewhat different from what has been used previously in the literature. For instance, Laursen et al. (2011) compute the  $\text{Ly}\alpha$  IGM transmission by casting sightlines in random directions from the border of the haloes (see also Gronke et al. 2020). While this method allows to estimate accurately the average isotropic  $\text{Ly}\alpha$  transmission from a given location through the IGM, it does not account for the possibility that  $\text{Ly}\alpha$  photons escape galaxies along particular lines-of-sight, and that the direction of escape may be correlated with the local IGM distribution. By propagating only photons that can emerge from the CGM, we therefore estimate the effective IGM transmission of  $\text{Ly}\alpha$  photons for each galaxy, as opposed to the formulation of Laursen et al. (2011).

At all redshifts, we measure a strong variation of  $T_{\text{IGM}}$  as a function of  $\lambda$  with the blue side being much more suppressed than the red side. This is a well-known consequence of the  $\text{Ly}\alpha$  RT in the Hubble flow : blue photons unavoidably redshift past the resonance along their propagation through the IGM such that they will be scattered off the line-of-sight as soon as HI is present at the corresponding distance,  $d = V/H(z)$  (where  $V$  is the velocity offset from line centre of a blue photon in the Hubble flow). In a fully neutral IGM, the medium is extremely optically thick and the blue part will be fully absorbed. In a partially ionised IGM, a diffuse neutral component may remain but the  $\text{Ly}\alpha$  transmission is also strongly affected by surviving dense self-shielded clouds and



**Figure 13.** Ly $\alpha$  IGM transmission as a function of rest-frame wavelength at  $z = 6, 7, 8,$  and  $9$ . The vertical red dotted line indicates the line centre ( $V = 0$ ). The evolution of  $T_{\text{IGM}}$  with redshift reflects the increase of the volumetric neutral fraction  $x_{\text{HI}}$  towards higher  $z$  (see legend).

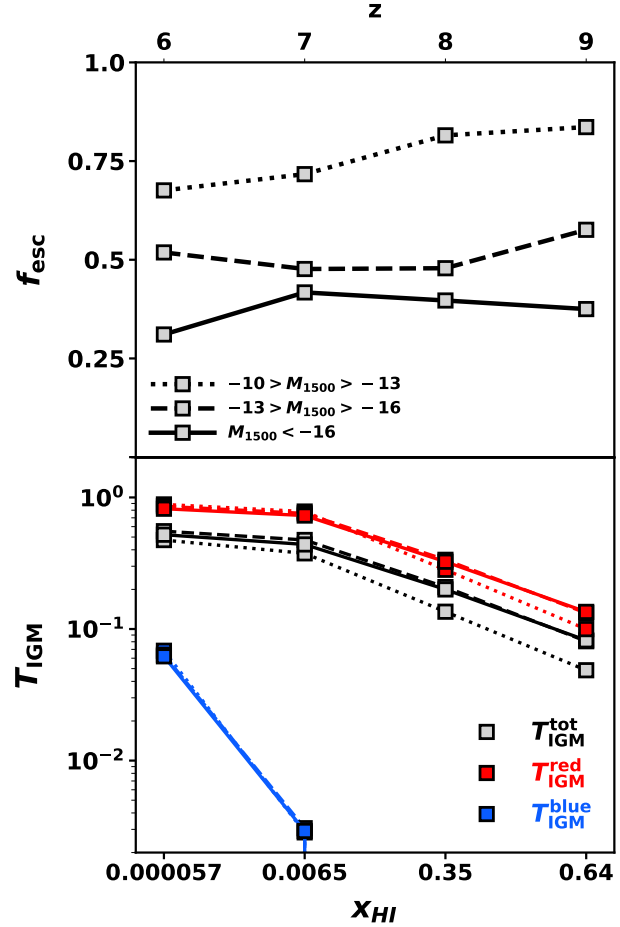
residual H $\text{I}$  within ionised bubbles (Dijkstra 2017; Gronke et al. 2020). As can be seen from Figure 13, a small (but non-negligible) fraction of flux is transmitted at  $z = 6$  blueward of Ly $\alpha$  even though the Universe is almost fully ionised ( $x_{\text{HI}} \approx 6 \times 10^{-5}$ ). We note that  $T_{\text{IGM}}$  reaches a minimum at  $V \approx -100 \text{ km s}^{-1}$  which is attributed to the increase of gas density in the vicinity of galaxies (see section 5.1 in Laursen et al. 2011, for a detailed discussion of this effect). When the volumetric neutral fraction becomes less than about 1% (i.e.  $z \approx 7$  in SPHINX), the blue part of the spectrum is nearly fully erased and only red photons can be transmitted.

At  $\lambda > \lambda_{\alpha}$ , the IGM transmission increases towards higher velocity offsets as a result of the wing absorption profile of the diffuse neutral component. Closer to the line centre, the red transmission,  $T_{\text{IGM}}^{\text{red}}$ , can be further decreased by infalling H $\text{I}$  clouds which are able to resonantly scatter photons leading to an IGM absorption. We recall that here we are showing the mean transmissions but there is a strong dispersion of  $T_{\text{IGM}}$  from one galaxy to another, especially near the line centre due to the directional variation of the occurrence rate of optically thick H $\text{I}$  in the neighbourhood of galaxies.

Overall, the IGM transmission is unsurprisingly dominated by the red part at all redshifts. We also clearly recover a strong evolution with the velocity offset at  $\lambda > \lambda_{\alpha}$ , with  $T_{\text{IGM}}^{\text{red}}$  at  $V = 150 \text{ km s}^{-1}$  reaching  $\approx 80\%$  ( $\approx 10 - 20\%$ ) at  $z \lesssim 7$  ( $z = 8 - 9$ ). As we will discuss in Section 3.6, this emphasises the need for realistic modelling of the internal RT since gas outflows in typical high redshift star-forming galaxies (Cassata et al. 2020) can alter and shift the Ly $\alpha$  line at similar velocity offsets (Verhamme et al. 2008).

### 3.5.2 Evolution of Ly $\alpha$ $T_{\text{IGM}}$ and $f_{\text{esc}}$ with galaxy properties and redshift

Figure 14 summarises the redshift evolution of the median Ly $\alpha$  internal escape fraction  $f_{\text{esc}}$  and IGM transmission  $T_{\text{IGM}}$  of galaxies split according to their UV magnitude. As mentioned earlier, the ability of Ly $\alpha$  photons to escape through the ISM and CGM does not vary much from  $z = 6$  to  $z = 9$  on average but it strongly depends on UV magnitude, or equivalently stellar mass (see Figure C2 for the correlation between  $M_{1500}$  and  $M_{\star}$ ). We find that



**Figure 14.** Evolution of the median Ly $\alpha$  escape fraction (top) and the median IGM transmission (bottom) as a function of IGM neutral fraction and redshift. The different line styles correspond to different bins of dust-attenuated UV magnitude. In the bottom panel, we also highlight the relative evolution of IGM transmission blueward ( $T_{\text{IGM}}^{\text{blue}}$ ) and redward of Ly $\alpha$  ( $T_{\text{IGM}}^{\text{red}}$ ).

$f_{\text{esc}}$  increases from  $\approx 25\%$  for bright UV sources ( $M_{1500} \leq -16$ ) to  $\approx 75\%$  at the very faint end ( $M_{1500} \geq -13$ ).

The IGM transmission on the other hand weakly varies as a function of UV magnitude.  $T_{\text{IGM}}$  is slightly larger for the bright UV sample than for the UV faint one. There does not seem to be a very strong correlation between  $T_{\text{IGM}}$  and galaxy properties, or as one could have expected, with environment. In patchy reionisation scenarios, as is the case in SPHINX, brighter galaxies form at density peaks so that they can blow larger H $\text{II}$  bubbles around them which can ease the transmission of Ly $\alpha$  photons. The fact that we detect only a small environment dependency in SPHINX is likely due to the fact our box is representative of an *average* field which, by construction, contains neither big voids nor significant overdensities.

In Figure 14, we also show separately the blue and red median transmissions ( $T_{\text{IGM}}^{\text{blue}}$  and  $T_{\text{IGM}}^{\text{red}}$ ) computed respectively from  $[-1000; 0]$  and  $[0; 1000] \text{ km s}^{-1}$ . As already mentioned in the previous section,  $T_{\text{IGM}}$  is mainly determined by the transmission redward of Ly $\alpha$ , especially at  $z \gtrsim 7$  when  $T_{\text{IGM}}^{\text{blue}}$  drops to zero. At  $z = 6$ , most of the red part of the spectrum is transmitted (80 – 90%) and  $T_{\text{IGM}}^{\text{red}}$  starts to decrease rapidly above  $z \approx 7$  to reach only 10 – 15% at  $z \approx 9$ .

Altogether, this suggests that (i) a small but non-negligible



fraction of Ly $\alpha$  radiation can be transmitted to the observer even when  $x_{\text{HI}}$  is large, and that (ii) the global Ly $\alpha$  IGM transmission is fully dominated by the contribution of the red part of the spectrum during the EoR. This reinforces the statement made in previous studies that special care must be given to the modelling of the Ly $\alpha$  transfer at smaller scales in order to assess realistically the visibility of LAEs.

### 3.6 Ly $\alpha$ spectra

In light of the former section, we now turn our interest to the spectral shapes of the Ly $\alpha$  line profiles. As discussed in detail in the literature (e.g. Santos 2004; Dijkstra et al. 2007; Laursen et al. 2011), the impact of the IGM is highly dependent on the spectral morphology of the line emerging from the galaxy, especially on the velocity offset of the Ly $\alpha$  peak with respect to the line center. From low and intermediate redshift observations, we know that typical LAEs harbour a single red asymmetric profile<sup>4</sup>, sometimes associated with a smaller blue peak (a.k.a. a blue-bump). As recently shown by Hayes et al. (2020), this general trend seems to hold up to  $z \approx 5$  and the amplitude of the blue peak appears to diminish with increasing redshift because of the IGM opacity. Indeed, IGM absorption is expected to significantly impede the Ly $\alpha$  transmission on the blue side of the resonance due to the Hubble flow. In addition, the transmission of the red part strongly varies with velocity shift over a few hundreds of  $\text{km s}^{-1}$  (see Figure 13). This velocity range corresponds to the typical speeds of galactic outflows that are thought to alter the shape, amplitude and peak shift of the Ly $\alpha$  line. In this context, assessing the spectral shapes of LAEs after internal RT is therefore essential in order to correctly predict the Ly $\alpha$  IGM transmission during the EoR and investigate its connection with the neutral fraction  $x_{\text{HI}}$ .

#### 3.6.1 Relative impact of internal RT and IGM

In Figure 15, we present the median angle-averaged Ly $\alpha$  spectra of our simulated galaxies at  $z = 6, 7, 8$  and  $9$  in the galaxy frame (from top to bottom) which are split into three bins of dust-attenuated UV magnitudes ( $-10 \geq M_{1500} \geq -13$ ,  $-13 \geq M_{1500} \geq -16$ , and  $M_{1500} \leq -16$  from left to right). In each panel, the thin dashed curves are indicative of the intrinsic Gaussian profiles centred on the Ly $\alpha$  line centre ( $V = 0$ ).

We first notice from Figure 15 that, in all cases, spectra after internal RT (thick solid lines) are broader than the intrinsic lines ( $FWHM \approx 300 - 600 \text{ km s}^{-1}$ ) and exhibit significant flux on both blue and red sides. This signature is typical of resonant scattering in optically thick and low-velocity media (Neufeld 1990; Verhamme et al. 2006), suggesting that galaxies host a dense, slow, H $\text{I}$  gas component in the ISM or in their local environment. That said, we are showing here spectra summed over all directions which erases any directional variation and may further broaden the spectral shapes, such that nearly symmetric angle-averaged spectra do not necessarily indicate RT in static media in our case. Preliminary analysis of the spectra along individual sightlines suggest that there is a very strong variability in terms of spectral morphologies for our

<sup>4</sup> These peculiar line shapes are plausibly a consequence of internal RT effects (e.g. back-scatterings in outflowing gas) which redistribute Ly $\alpha$  photons redward of the line centre and, consequently, ease their escape from galaxies (Verhamme et al. 2006; Ahn et al. 2003; Dijkstra et al. 2006).

galaxies but that flux blueward of Ly $\alpha$  is nearly always present in the simulation (Blaizot et al. in prep).

Interestingly, the median line shapes after internal RT presented in Figure 15 show very little variation with redshift. This indicates that the physical conditions (e.g. gas density, ionisation state, etc) at the ISM/CGM scale does not evolve much from  $z = 9$  to  $z = 6$ . Regarding the variation with UV magnitude however, we find a much more significant trend. While Ly $\alpha$  profiles in UV faint sources display nearly symmetric double-peaks centred on  $V = 0$  (left panels), the blue peak becomes strongly reduced towards brighter UV magnitudes. For the brightest UV bin, it is nearly completely suppressed such that the Ly $\alpha$  line resembles a redshifted asymmetric line. It is worth pointing out that velocity offset of the red peak is about  $150 \text{ km s}^{-1}$ , independently of the UV magnitude, which is the typical value measured in LAEs (e.g. Hashimoto et al. 2015). From the legend of Figure 15, we see that the median Ly $\alpha$  escape fractions become smaller towards brighter sources. Nonetheless, UV-brighter objects still correspond to higher Ly $\alpha$  luminosities after internal RT. Altogether, our results predict that galaxies that are observable in current surveys ( $M_{1500} \leq -16$  and  $L_{\text{Ly}\alpha}^{\text{intr}} \geq 10^{41} \text{ erg s}^{-1}$ ) have most of their Ly $\alpha$  flux emerging from the CGM redward of Ly $\alpha$ , peaking at  $\approx 150 \text{ km s}^{-1}$ .

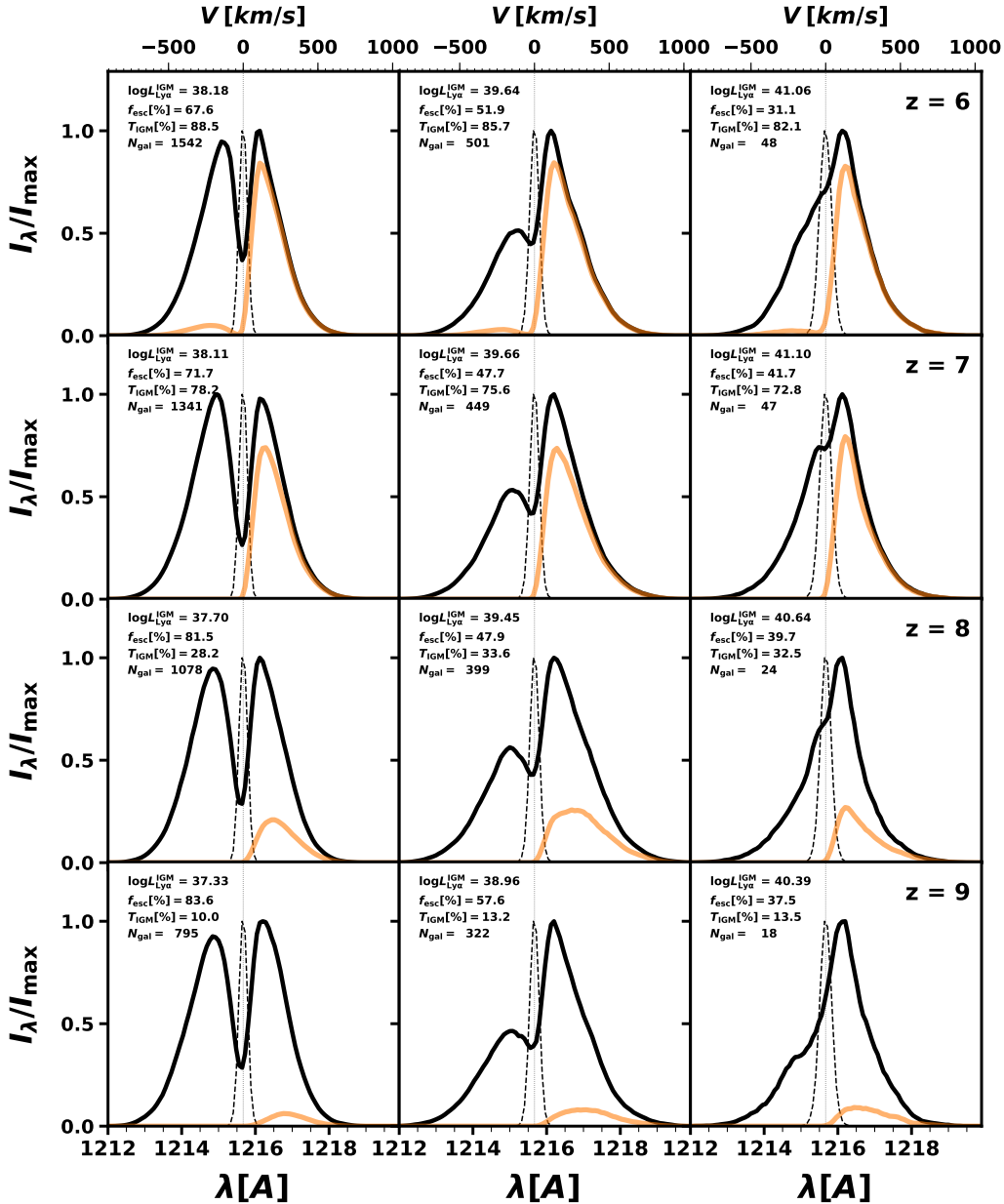
The effect of the IGM on the spectra is visible in Figure 15 (orange lines): at  $z = 6$ , the red part is almost fully transmitted whereas only a small fraction of blue photons typically remains. At this redshift, the median spectra resemble the typical observed lines, i.e. redward asymmetric or *blue-bump* profiles. Unsurprisingly, the IGM-attenuated spectra strongly evolve with redshift as the IGM opacity becomes increasingly high. From  $z = 7$ , blue photons can no longer be transmitted because the occurrence of clear sightlines drops as soon as the IGM neutral fraction starts rising. A  $z \geq 8$ , even the red peak becomes strongly suppressed and only a weak red peak is transmitted (see Section 3.5).

Figure 16 summarises the evolution of the red-to-total Ly $\alpha$  flux ratio at  $z = 7$  and  $z = 9$  with UV magnitude for intrinsic emission (left), escaping emission (middle) and IGM-transmitted emission. We clearly see that internal RT is a major cause of the frequency redistribution of Ly $\alpha$  photons, preferentially towards the red in UV bright galaxies. At  $z = 7$ , the IGM significantly favors the transmission of Ly $\alpha$  photons on the red side, although a non negligible fraction of blue photons manage to be transmitted along clear sightlines. At  $z = 9$  however, only photons with a sufficient red-shift avoid IGM absorption and can reach the observer.

#### 3.6.2 Variation of the spectral shape as a function of radius

As detailed in the previous section, our angle-average spectra can only reproduce the observed typical shapes once we account for IGM transmission. This feature is commonly seen in Ly $\alpha$  RT experiments in cosmological hydrodynamical simulations (Laursen et al. 2009; Smith et al. 2018; Mitchell et al. 2020). However, single-peak red-shifted profiles are also commonly observed at low redshift (where the impact of IGM is negligible). Therefore, the failure of hydrodynamic simulations at predicting these line shapes after internal RT is most likely related to gas outflows, and especially to the lack of fast-moving neutral hydrogen predicted by state-of-art simulations of galaxy formation. As shown by Barnes et al. (2011), the Ly $\alpha$  line shapes (as well as the Ly $\alpha$  spatial distribution) are very sensitive to the underlying galactic wind properties.

Single-peak red profiles with various peak shift, skewness and width seem to only be reproduced in more idealised Ly $\alpha$  numerical

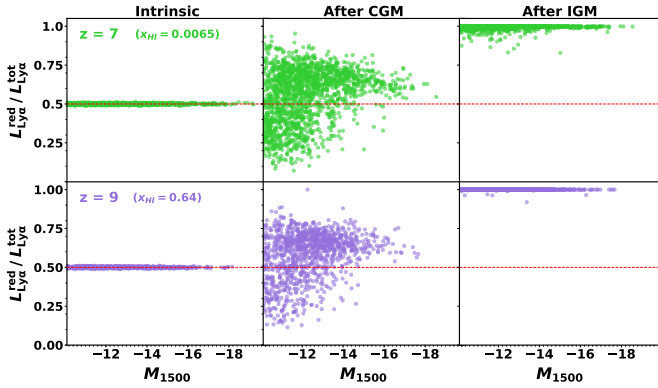


**Figure 15.** Median angle-averaged Ly $\alpha$  spectra at  $z = 6, 7, 8$  and  $9$  (from top to bottom). The three columns correspond to various dust-attenuated UV magnitude ranges: from left to right,  $-10 \geq M_{1500} \geq -13$ ,  $-13 \geq M_{1500} \geq -16$ , and  $M_{1500} \leq -16$ . In each panel, the thin black dashed line represents the intrinsic emission while the thick solid black and orange curves show the profiles after CGM and IGM attenuation respectively. The spectra for intrinsic and dust-obscured emission are normalised to their respective maxima whereas the IGM-attenuated spectra are normalised to the maximum of the dust-obscured line profile. The vertical dotted grey line corresponds to the Ly $\alpha$  line center. The legend in the top left of each panel gives the median observed Ly $\alpha$  luminosity, escape fraction, IGM transmission and the total number of galaxies used to compute the corresponding median spectrum.

experiments in which the input Ly $\alpha$  line propagates through high-velocity ( $\geq 100 \text{ km s}^{-1}$ ) and dense ( $\geq 10^{19} \text{ cm}^{-2}$ ) H $\alpha$  outflows (Verhamme et al. 2008; Hashimoto et al. 2015; Gronke 2017). Assessing if single red peaks after internal RT also prevail at very high redshift is still an open question. It is therefore unclear if (i) the small-scale H $\alpha$  distribution/kinematics predicted by cosmological simulations is somehow unrealistic (due to e.g. poorly constrained subgrid feedback models), or if (ii) there is a redshift evolution of the ISM/CGM conditions yielding double-peak profiles which are then suppressed on the blue side by the increasingly neutral IGM. It is worth pointing out that, in the latter case, the Ly $\alpha$  line shape

could represent a very powerful probe of the IGM topology during reionisation (Gronke et al. 2020).

Adding complexity to the problem, we do not fully understand either how the red single-peak Ly $\alpha$  profile is formed in low redshift galaxies, i.e. whether it is mainly via radiation transfer effects in the ISM or in the CGM. Recently, Kimm et al. (2019) investigated the escape of Ly $\alpha$  photons from turbulent ISM clouds simulated with RAMSES-RT at sub-pc resolution. Quite interestingly, their findings suggest a strong variability of the Ly $\alpha$  line shape emerging from the clouds, sometimes producing a red-dominated profile with a less prominent blue bump, which suggests that Ly $\alpha$  spectra



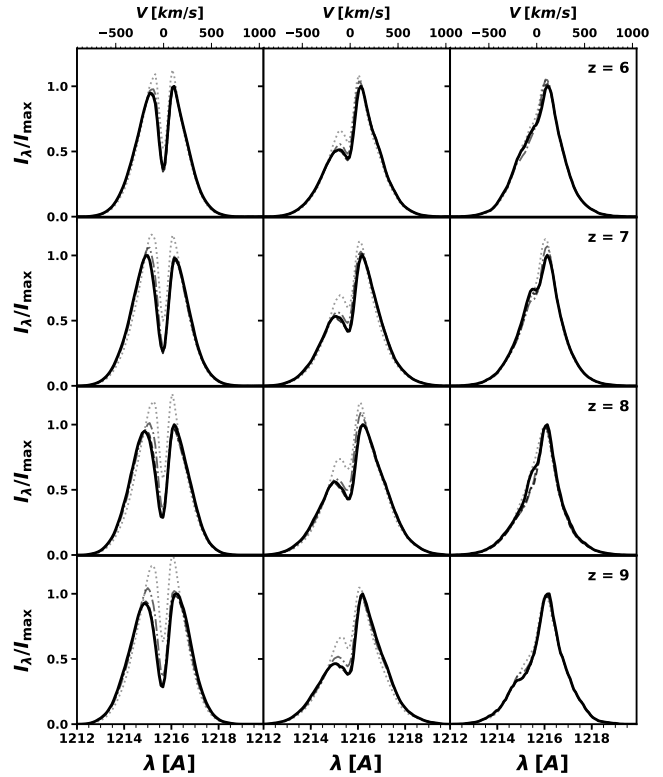
**Figure 16.** Ratio of the red-to-total Ly $\alpha$  flux as a function of dust-attenuated UV magnitude at  $z = 7$  (top) and  $z = 9$  (bottom). Each dot corresponds to an individual galaxy. The left, middle and right panels show this ratio for the intrinsic, dust-attenuated, and IGM transmitted emission respectively.

may be already (at least partially) in place at very small scale (see also Kakiichi & Gronke 2019). While the physical resolution in SPHINX prevents us from resolving such fine structure, we can still quantify the evolution of the Ly $\alpha$  profiles from ISM to CGM scale. To do so, we plot in Figure 17, the median spectra after internal RT computed at  $r_*$ ,  $2r_*$ ,  $5r_*$ , and  $10r_*$  ( $= R_{\text{CGM}}$ ). Overall, we find very little variation in terms of shape and amplitude whatever the redshift or UV magnitude range. This demonstrates that, in our simulation, the frequency distribution of Ly $\alpha$  photons is mostly set in the ISM. A closer look at Figure 17 reveals that the peaks are slightly broader with a larger separation at larger radii. This suggests that Ly $\alpha$  photons do keep scattering in the CGM (as shown in Appendix A) but that the density, kinematics, and/or covering fraction of the neutral gas in the CGM are not sufficient to alter significantly the emergent spectral shapes. Besides, Figure 17 demonstrates that the shape of Ly $\alpha$  spectra is fairly independent of the exact value of CGM scale which further validates our choice of choosing  $R_{\text{CGM}} = 10r_*$ . In light of Kimm et al. (2019)’s study, it is worth pointing out that achieving higher resolution in the ISM (or also in the CGM; see Tumlinson et al. 2017; Gronke et al. 2017, for instance) would affect more strongly the typical line shapes. Altogether, our results and the above discussion highlight the importance of internal Ly $\alpha$  radiative transport for interpreting LAE observations during the EoR as well as the uncertainties related to that matter.

### 3.6.3 Velocity shift of the Ly $\alpha$ line and IGM transmission

To illustrate the possible impact of the Ly $\alpha$  internal RT on the visibility of LAEs during the EoR discussed in the previous section, we introduce a toy model for the Ly $\alpha$  LF in which the spectrum emerging from the CGM is arbitrarily modified. As discussed in the previous section, our Ly $\alpha$  spectra after CGM RT are double-peaked with a peak separation of  $\approx 300\text{km s}^{-1}$  for fainter galaxies and single-peaked with an offset of  $\approx +150\text{km s}^{-1}$  for bright ones. Here, we assess by how much the IGM transmission, and therefore the observed Ly $\alpha$  LF, would change if different spectral shapes were assumed.

To do so, we keep the dust-attenuated Ly $\alpha$  luminosities the same in our toy model but we replace the Ly $\alpha$  lines after internal RT by single-peaked Gaussian profiles with various rms widths,  $\sigma_v$ , and positive velocity peak offsets,  $V_{\text{peak}}$ . This assumption on the line shape is quite simplistic because most observed Ly $\alpha$  lines usually



**Figure 17.** Variation of the median angle-averaged Ly $\alpha$  spectra as a function of radius at  $z = 6, 7, 8$  and  $9$  (from top to bottom). The three columns correspond to various dust-attenuated UV magnitude ranges: from left to right,  $-10 \geq M_{1500} \geq -13$ ,  $-13 \geq M_{1500} \geq -16$ , and  $M_{1500} \leq -16$ . In each panel, the thick solid black curves show the profiles evaluated at  $R_{\text{CGM}} = 10r_*$ , as in Figure 15. To assess the radial evolution of the spectral shape, we also show the median angle-averaged profiles computed at  $r_*$  (dotted grey line),  $2r_*$  (dot-dashed grey line), and  $5r_*$  (dashed grey line). All spectra are normalised to the maximum of the profile measured at  $R_{\text{CGM}}$ .

appear either asymmetric or double-peaked but, to first order, the two parameters  $\sigma_v$  and  $V_{\text{peak}}$  are sufficient to investigate the overall effect of line broadening and red-shift on the Ly $\alpha$  transmission by IGM. The individual IGM transmission  $T_{\text{IGM}}(\lambda)$  of each galaxy is applied to each profile to compute the IGM-transmitted luminosities.

As can be seen from Figure 18, the observed Ly $\alpha$  LF can dramatically change depending on the shape of the Ly $\alpha$  line emerging from galaxies and the effect becomes stronger towards higher redshifts. For the parameter values assumed here, the dispersion induced on the Ly $\alpha$  luminosities amounts to  $\approx 0.5$  dex at  $z = 6$  to  $\approx 2$  dex at  $z = 9$ . Small  $V_{\text{peak}}$  values tend to significantly reduce the IGM transmission because, in this case, most of the Ly $\alpha$  flux escapes galaxies near the line center where the IGM absorption is maximal (Figure 13). Conversely, large velocity offsets (up to  $500 \text{ km s}^{-1}$  in our toy model) greatly favour the transmission of Ly $\alpha$  photons to the observer. This is particularly true at  $z \lesssim 7$  where  $T_{\text{IGM}}$  resembles a step-function around  $V = 0$  where the red part is nearly fully transmitted. At these redshifts the IGM neutral fraction is less than 0.01 and we find that the IGM becomes fully transparent to Ly $\alpha$  photons for  $V_{\text{peak}} \gtrsim 300\text{km s}^{-1}$ . Note that this trend holds for the two  $\sigma_v$  assumed here, 50 and  $200 \text{ km s}^{-1}$ . These rms widths correspond to full-width-at-half-maximum of  $\approx 120$  and  $470 \text{ km s}^{-1}$ , typical of faint high-redshift LAEs (Kerutt et al., in prep). For

a narrow line ( $\sigma_v = 50 \text{ km s}^{-1}$ ), the IGM transmission almost only depends on  $V_{\text{peak}}$  which leads to the large dispersion in the resulting LFs (top panel of 18). For a broader line (bottom panel), more Ly $\alpha$  photons can be transmitted to the red whatever the velocity peak offset, so varying  $V_{\text{peak}}$  has a milder effect on the LFs (as long as  $\sigma_v$  remains the same order of magnitude as  $V_{\text{peak}}$ ).

## 4 DISCUSSION

### 4.1 IGM Ly $\alpha$ transmission and neutral fraction

The Ly $\alpha$  emission from galaxies has long been put forward as a possible probe of reionisation (Miralda-Escude 1998; Haiman 2002; Furlanetto et al. 2006). Observations show evidence that Ly $\alpha$  emission becomes increasingly suppressed at  $z \gtrsim 6$  as can be inferred from the evolution of the Ly $\alpha$  LF, LAE clustering and  $\chi_{\text{Ly}\alpha}$  (Zheng et al. 2017; Itoh et al. 2018; Ouchi et al. 2017; Schenker et al. 2012).

In agreement with our present study, this is often interpreted as the imprint of the reionisation of the inter-galactic medium, but alternative explanations have been suggested. A possible scenario is of course the co-evolution of galaxy and Ly $\alpha$  properties towards high redshift, plausibly due to variations of the gas and dust content, distribution, and kinematics (Dayal & Ferrara 2012; Jensen et al. 2013; Garel et al. 2015; Hassan & Gronke 2021). The incidence of optically thick systems in the vicinity of galaxies can also have a dramatic effect on the Ly $\alpha$  visibility (Bolton & Haehnelt 2012) and reduce the number of strong emitters towards higher redshift. Sadoun et al. (2017) pointed out the possible impact of a rapidly evolving UV background on the ionisation state of the CGM of the galaxies themselves. As the infall region becomes more self-shielded towards higher redshift, Ly $\alpha$  scattering would produce a more extended emission component that can be partially missed by observing apertures and thus artificially induce a drop of  $\chi_{\text{Ly}\alpha}$ . Finally, cosmic variance is undoubtedly a source of uncertainties in deep surveys, especially for measurements of  $\chi_{\text{Ly}\alpha}$  that are based on spectroscopic samples. Nevertheless, this effect is unlikely to fully dominate the observed evolution of Ly $\alpha$  properties at  $z \gtrsim 6$  (Taylor & Lidz 2013).

Our results suggest that the contributions of intrinsic galaxy properties and dust attenuation are not driving the apparent evolution of LAEs during the EoR and that only IGM transmission is at play. This does not necessarily mean however that  $T_{\text{IGM}}$  is an obvious probe of the reionisation process. In SPHINX, galaxies start ionising the Universe by  $z \approx 15$  and  $x_{\text{HI}}$  drops to 50% at  $z \approx 8.5$  and below 1% at  $z \approx 7$ . This phase change is sustained by the inclusion of binary stars which boost the escape of LyC photons. As shown in Figure 19, the mean ionising escape fraction changes only insignificantly between  $z = 6$  and 9 (5 – 10%) which, interestingly, echoes the redshift evolution found for the Ly $\alpha$  escape fraction (Figure 14). Note however that the escape of Ly $\alpha$  photons from the ISM/CGM is set by dust attenuation (which is enhanced by resonant scattering in optically thick gas) whereas the escape of ionising photons is fully driven by the H I opacity within galaxies in SPHINX<sup>5</sup>.

Figure 19 also compares the evolution of  $T_{\text{IGM}}$ , now in linear

<sup>5</sup> The impact of dust on ionising RT was neglected in the SPHINX simulations (Rosdahl et al. 2018) but it would only have a very minor effect on the escape of ionising photons (i.e. probably delaying reionisation very slightly) because H I Lyman-continuum opacities largely dominate over dust optical depths in our galaxies.

scale, with other fundamental quantities related to cosmic reionisation. The global Ly $\alpha$  IGM transmission computed from all galaxies at each snapshot decreases from  $\approx 45\%$  at  $z = 6$  to  $\approx 5\%$  only at  $z = 9$  (black circles). Over the same redshift range, the volumetric neutral fraction of the IGM  $x_{\text{HI}}$  varies from  $\lesssim 1\%$  to  $\approx 65\%$  (green line; see Figure 13 in Rosdahl et al. 2018).  $T_{\text{IGM}}(z)$  appears to be nicely anti-correlated to  $x_{\text{HI}}(z)$  and it can be well fit by a functional form (grey curve).

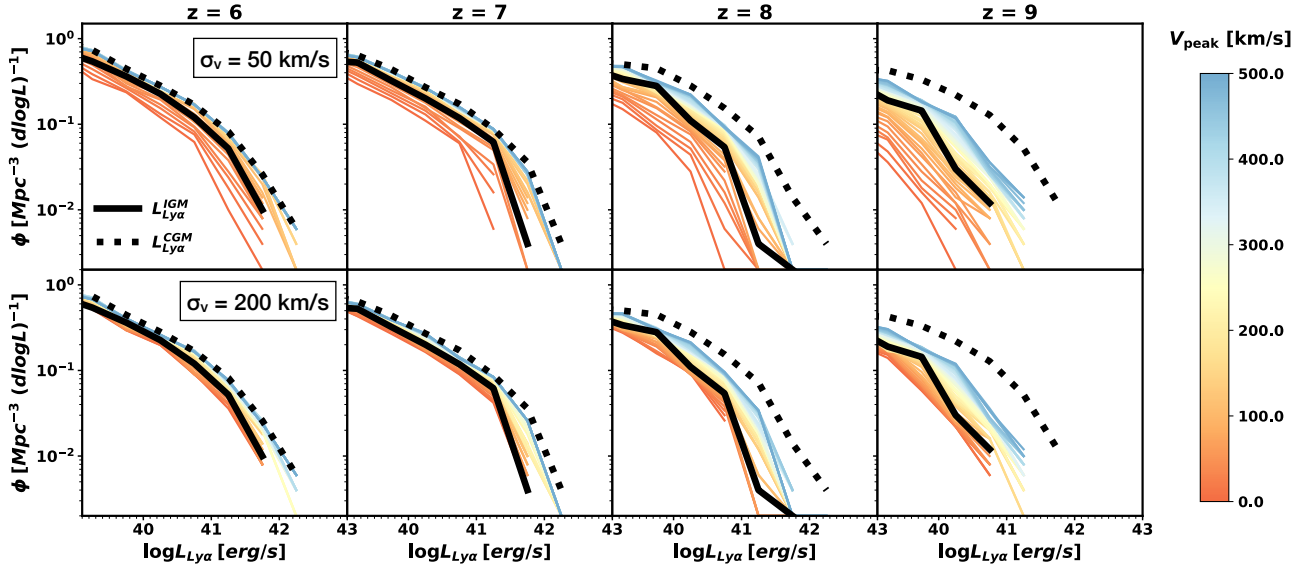
As discussed in Section 3.5.2, the IGM transmission of blue photons quickly falls to zero at  $z \gtrsim 7$ . However, the red transmission  $T_{\text{IGM}}^{\text{red}}$  evolves strongly with redshift (red circles) and traces fairly well the global IGM transmission but boosted by a factor  $\approx 2$ . We discussed in Section 3.5.1 how Ly $\alpha$  photons can be transmitted redward of the line centre. While photons may be scattered by infalling H I clouds close to resonance, the damping wing of the neutral component of the IGM is mainly responsible for the overall shape of  $T_{\text{IGM}}^{\text{red}}(\lambda)$ . The connection between Ly $\alpha$  transmission and  $x_{\text{HI}}$  therefore strongly depends on the velocity shift after internal RT and the distance to nearby neutral patches, i.e. the size of the H I bubble in which a galaxy is embedded.

We further illustrate this aspect by showing the expected Gunn-Peterson (GP) transmission for photons redward of Ly $\alpha$ ,  $e^{-\tau_{\text{GP}}}$ , where the GP opacity  $\tau_{\text{GP}}$  is inversely proportional to the velocity shift from resonance  $\Delta V$  in a partially neutral IGM ( $\tau_{\text{GP}} \propto (\Delta V)^{-1} (1+z)^{3/2}$ ; Miralda-Escude 1998; Dijkstra & Wyithe 2010)<sup>6</sup>. Here,  $\Delta V$  corresponds to the offset seen by the neutral IGM so it therefore includes both the contribution of outflows and Hubble flow. In Figure 19, we plot the GP transmission redward of Ly $\alpha$  for three values of  $\Delta V$  (100, 300, and 600  $\text{km s}^{-1}$ ; red dotted lines from bottom to top) which span a similar range of velocities as our simulated spectra (see Figure 15). We see that the  $T_{\text{IGM}}^{\text{red}}$  predicted by the analytic model varies strongly depending on  $\Delta V$  but it reproduces well the general redshift evolution of the simulated transmission. While our mean simulated spectra peak at  $V \approx 150 \text{ km s}^{-1}$ , it is the model with  $V = 300 \text{ km s}^{-1}$  that provides the best match to our results. In addition to the impact of outflows, this also plausibly highlights the connection between IGM transmission and the topology of H I bubbles in the environment of galaxies during the EoR. This key aspect was studied recently by Gronke et al. (2020) and we intend to investigate it in a forthcoming paper.

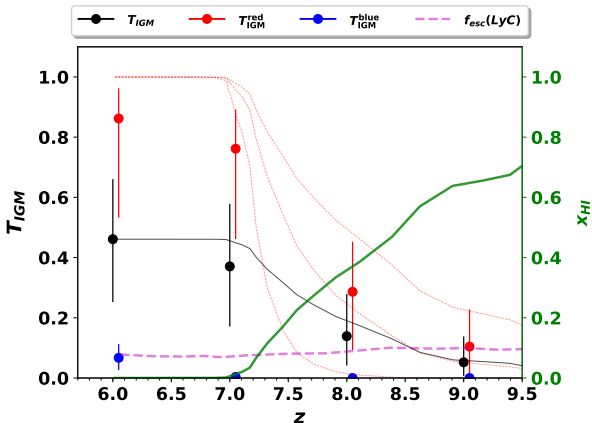
### 4.2 Model assumptions and potential caveats

As detailed in Section 2.1, SPHINX provides an unprecedented trade-off between resolution (in terms of mass and physical sampling) and box size for cosmological RHD simulations which allows us to capture large-scale reionization as well as the physics and radiation transfer in resolved galaxies. The  $10^3 \text{ cMpc}^3$  simulated volume remains nevertheless relatively small which has a number of implications. First, our simulation does not contain large scale overdensities so we miss the contribution of massive galaxies (as well as active galactic nuclei) to reionisation. Regarding LAEs, bright sources are rare in SPHINX and the highest observed Ly $\alpha$  luminosities at  $z = 6$  reach  $L_{\text{Ly}\alpha} \approx 10^{42} \text{ erg s}^{-1}$  which prevents us from comparing our results with most observational data.

<sup>6</sup> As detailed in Dijkstra (2017),  $\tau_{\text{GP}} = 2.3 x_{\text{D}} \left( \frac{\Delta V}{600 \text{ km s}^{-1}} \right)^{-1} \left( \frac{1+z}{10} \right)^{3/2}$  where  $x_{\text{D}}$  corresponds to the "patch-averaged" neutral fraction which depends on the volumetric neutral fraction  $x_{\text{HI}}$  in a non-trivial way. Since we only intend to illustrate the inverse scaling between GP opacity and velocity offset in Figure 19, we assume here for simplicity that  $x_{\text{D}} = x_{\text{HI}}$ .



**Figure 18.** Impact of the Ly $\alpha$  line shape on the IGM transmission. The four panels show our fiducial Ly $\alpha$  luminosity functions at  $z = 6, 7, 8$  and  $9$  for dust-attenuated (black dotted lines) and IGM-transmitted emission (solid black lines) as in Figure 6. The coloured lines correspond to our toy model in which the line shape emerging from each galaxy (i.e. after CGM RT) is artificially replaced by a Gaussian profile with varying rms width ( $\sigma_v$ ) and velocity peak offset ( $V_{\text{peak}}$ ). The IGM-transmitted Ly $\alpha$  luminosities ( $L_{\text{Ly}\alpha}^{\text{IGM}}$ ) are then computed from the individual IGM transmission  $T_{\text{IGM}}(\lambda)$  of each galaxy. The top and bottom panels correspond to  $\sigma_v = 50$  and  $200 \text{ km s}^{-1}$  respectively and the  $V_{\text{peak}}$  values are given by the colorbar.



**Figure 19.** Comparison of the redshift evolution of  $T_{\text{IGM}}$  with the ionising escape fraction  $f_{\text{esc}}(\text{LyC})$  and the volumetric IGM neutral fraction  $x_{\text{HI}}$ . The black, blue, and red circles correspond to our median  $T_{\text{IGM}}$ ,  $T_{\text{IGM}}^{\text{blue}}$ , and  $T_{\text{IGM}}^{\text{red}}$  respectively, and the error bars are the 10-90th percentiles. Note that the red and blue circles are shifted by  $+0.05$  for clarity. The solid grey line shows the fit to  $T_{\text{IGM}}$  as a function  $x_{\text{HI}}$  assuming the following functional form :  $T_{\text{IGM}} = T_{\text{IGM}}^{\bar{z}=6} \times (1 - x_{\text{HI}})^2$ . The red dotted curves represent the Gunn-Peterson transmission for red photons assuming a velocity shift of 100, 300, and 600  $\text{km s}^{-1}$  (from bottom to top; see text). The green curve shows the redshift evolution of  $x_{\text{HI}}$  in SPHINX. The dashed magenta curve is the average LyC escape fractions of galaxies as measured in Rosdahl et al. (2018).

Due to its limited size, our simulation also underestimates the large scale fluctuations and peculiar motions of the IGM (see Iliev et al. 2014, for a discussion on the scale needed to capture the global topology of reionisation), which may have an impact on our derived Ly $\alpha$  transmissions. However, our mean  $T_{\text{IGM}}$  at a given  $z$  seems to weakly depend on the UV magnitude (Figure 14; we have also tested that the trend is similar by taking stellar mass bins instead of  $M_{1500}$ ).

Since  $M_{1500}$  is, on average, tracing the environment (i.e. brighter UV galaxies are preferentially located in denser environments), our mean IGM transmissions do not appear to be very sensitive to the IGM topology fluctuations at the 10 cMpc scale. In addition, the effect of IGM on the Ly $\alpha$  visibility is dominated by Ly $\alpha$  absorption in the close environment of LAEs in SPHINX. Indeed, we checked that varying the stopping criterion for the IGM Ly $\alpha$  RT from 1 to  $10L_{\text{box}}$  (see Section 2.2.4) gives very similar results (e.g. similar Ly $\alpha$  LFs). Altogether this suggests that the average Ly $\alpha$  IGM transmissions in SPHINX are unlikely to be significantly affected by the box size for the population of galaxies we are looking at.

Another potential caveat in our study is that intrinsic Ly $\alpha$  emission is only arising from the ISM ( $r < r_{\star}$ ) as the contribution of CGM emission (at  $r_{\star} < r < R_{\text{CGM}}$ ) is ignored on purpose. Including photons produced in-situ in the CGM for each source up to  $R_{\text{CGM}}$  would be problematic in our setup because a non-negligible number of gas cells would be associated with more than one source. This would be particularly significant in low-mass galaxies that are strongly clustered around more massive objects. Overcounting such cells would artificially, and incorrectly, boost the intrinsic emission budget of many LAEs. We bypass this potential issue by restricting the emission region to the ISM ( $r < r_{\star}$ ). Our intrinsic Ly $\alpha$  luminosities can be seen as conservative values but we have checked that the bulk of intrinsic emission is coming from the ISM and therefore is accounted for, especially for bright galaxies for which the amount of missed Ly $\alpha$  emission is negligible (see also Mitchell et al. 2020).

Finally, a strong hypothesis in our SPHINX run is the use of BPASS (v2.0) with 100% binary stars. As shown in Rosdahl et al. (2018), this SED model has the net advantage of leading to a efficient and rapid ionisation of the Universe whereas a more standard library based on single-stars only fails to do so. We note that the Universe is almost fully ionised by  $z = 7$  in SPHINX which is about

$\Delta z = 0.5$  earlier than suggested by observations (Figure 9 in Rosdahl et al. 2018). Interestingly, Figure 12 shows that our predicted LAE fraction disagrees with the observed  $\chi_{\text{Ly}\alpha}$ . Instead, our LAE fraction starts decreasing at  $z \gtrsim 7$  which corresponds to the epoch where the IGM neutral fraction becomes non-negligible in our simulation, suggesting that  $\chi_{\text{Ly}\alpha}$  traces well the evolution of the IGM neutral phase. BPASS also significantly boosts the intrinsic budget of Ly $\alpha$  photons produced through case B recombination (see Section 3.2.1). Considering 100% binary stars is perhaps somewhat extreme and we note that models with a slightly lower fraction of binaries, and thus generating less ionising and Ly $\alpha$  photons per starburst, may be more realistic.

While it is important to keep these aspects in mind to guide future work and to make quantitative predictions for forthcoming Ly $\alpha$  surveys during the EoR, none of these caveats is likely to significantly affect our main results, i.e. that the LAE evolution with redshift is predominantly due to IGM absorption and not to an evolution in the ISM or CGM.

## 5 SUMMARY

Using the non-zoom cosmological radiation-hydrodynamics SPHINX simulation, we have investigated the redshift evolution of the Ly $\alpha$  signatures of galaxies during the epoch of reionisation. We have mainly focused on the relative impacts of intrinsic evolution, dust-attenuation at ISM/CGM scales and IGM transmission on the visibility of LAEs at  $z \geq 6$ . The unique ability of SPHINX to capture both reionisation at Mpc scale and the production and escape of ionising radiation within galaxies allows us to attempt for the first time to fill the gap between simulations of Ly $\alpha$  RT within individual galaxy environments and Ly $\alpha$  propagation through the IGM.

In order to study the imprint of reionisation on Ly $\alpha$  observables, we post-process four different snapshots with RASCAS (at  $z = 6, 7, 8,$  and  $9$ ) and compute Ly $\alpha$  angle-averaged properties of galaxies. The SPHINX volume being fully reionised by  $z = 6 - 7$  (predominantly thanks to the inclusion of binary stars in the SED modelling), our study covers a period of  $\approx 400$  Myr over which the IGM neutral fraction evolves from 60% to nearly 0%.

Even though our study is mostly restricted to faint and low-mass objects, we show that our simulation can reproduce a number of observational constraints at high redshift, in particular the stellar mass and UV/Ly $\alpha$  luminosity functions. The detailed analysis of the Ly $\alpha$  LFs and EW distributions in the different snapshots tells us that the redshift evolution of intrinsic and dust-attenuated properties is very mild from  $z = 6$  to  $9$ , if not null. Hence, the contribution of these processes to the observed Ly $\alpha$  suppression at  $z \gtrsim 6$  is predicted to be completely sub-dominant. We find however a significant reduction in terms of Ly $\alpha$  fluxes and EW due to the increase of the IGM opacity with redshift. The inclusion of the IGM transmission provides good agreement with observational data at  $z = 6$ . We also measure the redshift evolution of the LAE fraction  $\chi_{\text{Ly}\alpha}$  and find that it is indeed a promising diagnostic to probe reionisation. While the value of  $\chi_{\text{Ly}\alpha}$  varies with the UV magnitude of the selected galaxy population, its evolution with redshift is almost fully driven by the change of the ionisation state of the Universe. Looking back in time,  $\chi_{\text{Ly}\alpha}$  is found to be nearly constant as long as the IGM volumetric neutral fraction  $x_{\text{HI}}$  is less than one percent, and to decline once the Universe becomes more and more neutral.

In our simulation, the typical Ly $\alpha$  escape fraction from galaxies is on average 25% (75%) for bright/massive sources (faint/low-mass) but remains constant from  $z = 6$  to  $9$ . The global Ly $\alpha$  IGM

transmission  $T_{\text{IGM}}$  drops from 45% to 5% between  $z = 6$  and  $z = 9$  and we find a positive, but barely significant, trend with UV magnitude which may suggest that the Ly $\alpha$  observability is slightly enhanced in more massive/overdense environments. While a small fraction of the flux blueward of Ly $\alpha$  (5 – 10%) is transmitted at  $z = 6$ , the blue side of the spectrum is completely erased by the IGM at  $z \gtrsim 7$ . Individual clear sightlines might nevertheless exist at these redshifts but we decided to focus on mean properties and to leave the study of the directional variation to future work.

The global IGM transmission is found to be mainly driven by the red part of the spectrum. The red transmission is however a strong function of the velocity shift of Ly $\alpha$  photons that emerge from the galaxy. While the angle-averaged line profiles after internal RT are nearly symmetric for faint galaxies, brighter observable sources exhibit single-peak spectra red-shifted by  $\approx 150$  km s $^{-1}$  on average, most likely because of outflows at the ISM/CGM scale. At  $z \approx 6$ , a significant fraction of the flux redward of Ly $\alpha$  can be transmitted ( $\approx 80\%$ ). At  $z \approx 9$ , where the neutral fraction is already  $\approx 60\%$ , we find a mean red transmission of approximately 10% which suggests that a non-negligible number of intrinsically bright LAEs may still be visible during the EoR. Interestingly, we do not predict any break or turnover of the Ly $\alpha$  LF at the faint-end, and we instead find that the observable number of LAEs keeps rising down to  $L_{\text{Ly}\alpha} = 10^{37}$  erg s $^{-1}$  at least.

This first study of the full Ly $\alpha$  modelling in a cosmological RHD simulation highlights that careful modelling of the internal Ly $\alpha$  RT is essential to assess the impact of the IGM on the observability of LAEs during the EoR. Nonetheless, further improvements are required to draw more general conclusions. First, higher physical resolution may be needed to describe the Ly $\alpha$  radiation transfer process in the ISM and in the CGM. Although the resolution reached in SPHINX is already substantial for a cosmological simulation, we still have to rely on idealised or zoom simulations to assess the impact of small-scale structure on the Ly $\alpha$  RT (e.g. Kimm et al. 2019). Second, the relatively small volume of SPHINX is insufficient to capture the larger modes of structure formation and the large-scale topology of reionisation. Building upon the present study, we intend to make a step forward by extending our analysis to a new (eight times bigger) simulation (Rosdahl et al., in prep.).

## ACKNOWLEDGEMENTS

We thank the anonymous referee for helpful comments. TG and AV are supported by the ERC Starting Grant 757258 "TRIPLE". The results of this research have been achieved using the PRACE Research Infrastructure resource SuperMUC based in Garching, Germany (PRACE project ID 2016153539). We are grateful for the excellent technical support provided by the SuperMUC staff. The radiation transfer simulations and analysis were also performed at the Common Computing Facility (CCF) of the LABEX Lyon Institute of Origins (ANR-10-LABX-0066). TK was supported in part by the National Research Foundation of Korea (NRF-2019K2A9A1A0609137711 and NRF-2020R1C1C100707911) and in part by the Yonsei University Future-leading Research Initiative (RMS2-2019-22-0216). This work was supported by the Programme National Cosmology et Galaxies (PNCG) of CNRS/INSU with INP and IN2P3, co-funded by CEA and CNES. MGH acknowledges support from the UKRI Science and Technology Facilities Council (grant numbers ST/N000927/1 and ST/S000623/1).



## DATA AVAILABILITY

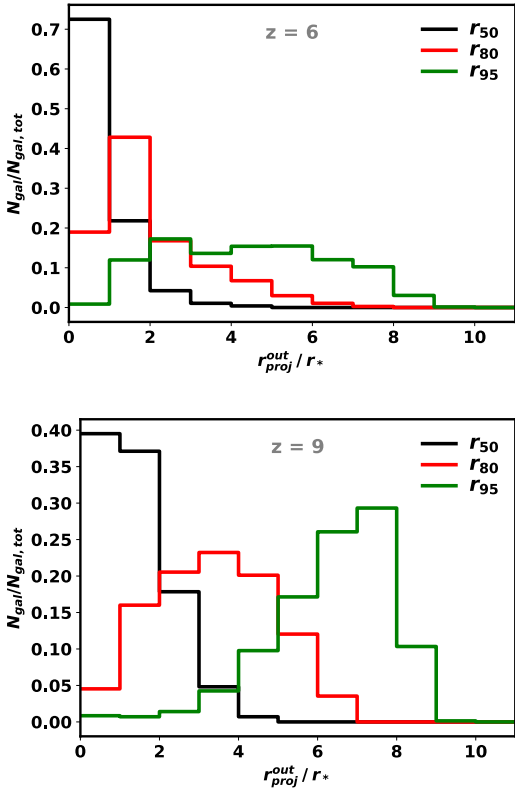
The data underlying this article will be shared on reasonable request to the corresponding author.

## REFERENCES

- Ade P. A. R. et al., 2014, *Astronomy & Astrophysics*, 571, A1
- Ahn S.-H., Lee H.-W., Lee H. M., 2003, *MNRAS*, 340, 863
- Atek H. et al., 2015, *The Astrophysical Journal*, 814, 69
- Atek H., Richard J., Kneib J.-P., Schaerer D., 2018, *Monthly Notices of the Royal Astronomical Society*, 479, 5184
- Aubert D., Pichon C., Colombi S., 2004, *Monthly Notices of the Royal Astronomical Society*, 352, 376
- Bacon R. et al., 2021, *Astronomy & Astrophysics*, 647, A107
- Bañados E. et al., 2017, *Nature*, 553, 473
- Barnes L. A., Haehnelt M. G., Tescari E., Viel M., 2011, *MNRAS*, 416, 1723
- Behrens C., Pallottini A., Ferrara A., Gallerani S., Vallini L., 2019, *Monthly Notices of the Royal Astronomical Society*, 486, 2197
- Bhatawdekar R., Conselice C. J., Margalef-Bentabol B., Duncan K., 2019, *Monthly Notices of the Royal Astronomical Society*, 486, 3805
- Bolton J. S., Haehnelt M. G., 2012, *MNRAS*, p. 412
- Bouwens R. et al., 2016, *The Astrophysical Journal*, 833, 72
- Bouwens R. J. et al., 2015, *The Astrophysical Journal*, 803, 34
- Bouwens R. J., Oesch P. A., Illingworth G. D., Ellis R. S., Stefanon M., 2017, *The Astrophysical Journal*, 843, 129
- Cantalupo S., Porciani C., Lilly S. J., 2008, *The Astrophysical Journal*, 672, 48
- Cassata P. et al., 2011, *A&A*, 525, A143
- Cassata P. et al., 2020, *Astronomy & Astrophysics*, 643, A6
- Davies F. B. et al., 2018, *The Astrophysical Journal*, 864, 142
- Dayal P., Ferrara A., 2012, *MNRAS*, 421, 2568
- Dayal P., Maselli A., Ferrara A., 2011, *MNRAS*, 410, 830
- De Barros S. et al., 2017, *Astronomy & Astrophysics*, 608, A123
- Dijkstra M., 2017, *Saas-Fee Lecture Notes: Physics of Lyman Alpha Radiative Transfer*
- Dijkstra M., Haiman Z., Spaans M., 2006, *ApJ*, 649, 14
- Dijkstra M., Lidz A., Wyithe J. S. B., 2007, *MNRAS*, 377, 1175
- Dijkstra M., Wyithe J. S. B., 2010, *MNRAS*, 408, 352
- Dijkstra M., Wyithe J. S. B., Haiman Z., 2007, *MNRAS*, 379, 253
- Drake A. B. et al., 2017, *Astronomy & Astrophysics*, 608, A6
- Đurovičková D., Katz H., Bosman S. E. I., Davies F. B., Devriendt J., Slyz A., 2020, *Monthly Notices of the Royal Astronomical Society*, 493, 4256
- Eldridge J. J., Izzard R. G., Tout C. A., 2008, *Monthly Notices of the Royal Astronomical Society*, 384, 1109
- Fan X. et al., 2006, *AJ*, 132, 117
- Federrath C., Klessen R. S., 2012, *The Astrophysical Journal*, 761, 156
- Finkelstein S. L. et al., 2019, *The Astrophysical Journal*, 879, 36
- Finkelstein S. L. et al., 2015, *The Astrophysical Journal*, 810, 71
- Fuller S. et al., 2020, *The Astrophysical Journal*, 896, 156
- Furlanetto S. R., Zaldarriaga M., Hernquist L., 2006, *MNRAS*, 365, 1012
- Gallerani S. et al., 2010, *Astronomy & Astrophysics*, 523, A85
- Garel T., Blaizot J., Guiderdoni B., Michel-Dansac L., Hayes M., Verhamme A., 2015, *MNRAS*, 450, 1279
- Garel T., Blaizot J., Guiderdoni B., Schaerer D., Verhamme A., Hayes M., 2012, *MNRAS*, 422, 310
- Grazian A. et al., 2015, *Astronomy & Astrophysics*, 585, A48
- Gronke M., 2017, *Astronomy & Astrophysics*, 608, A139
- Gronke M., Dijkstra M., McCourt M., Peng Oh S., 2017, *Astronomy & Astrophysics*, 607, A71
- Gronke M. et al., 2020, *Lyman-alpha transmission properties of the intergalactic medium in the CoDaII simulation*
- Guillet T., Teyssier R., 2011, *Journal of Computational Physics*, 230, 4756
- Hahn O., Abel T., 2011, *Monthly Notices of the Royal Astronomical Society*, 415, 2101
- Haiman Z., 2002, *The Astrophysical Journal*, 576, L1
- Haiman Z., Cen R., 2005, *ApJ*, 623, 627
- Hashimoto T. et al., 2017, *Astronomy & Astrophysics*, 608, A10
- Hashimoto T. et al., 2015, *The Astrophysical Journal*, 812, 157
- Hassan S., Gronke M., 2021, *The Astrophysical Journal*, 908, 219
- Hayes M., Schaerer D., Östlin G., Mas-Hesse J. M., Atek H., Kunth D., 2011, *ApJ*, 730, 8
- Hayes M. J., Runnholm A., Gronke M., Scarlata C., 2020, *Spectral shapes of the Lyman-alpha emission from galaxies: I. blueshifted emission and intrinsic invariance with redshift*
- Herenz E. C. et al., 2019, *Astronomy & Astrophysics*, 621, A107
- Hoag A. et al., 2019, *The Astrophysical Journal*, 878, 12
- Hui L., Gnedin N. Y., 1997, *Monthly Notices of the Royal Astronomical Society*, 292, 27
- Hutter A., Dayal P., Partl A. M., Müller V., 2014, *Monthly Notices of the Royal Astronomical Society*, 441, 2861
- Iliev I. T., Mellema G., Ahn K., Shapiro P. R., Mao Y., Pen U.-L., 2014, *Monthly Notices of the Royal Astronomical Society*, 439, 725
- Inoue A. K. et al., 2018, *Publications of the Astronomical Society of Japan*, 70
- Ishigaki M., Kawamata R., Ouchi M., Oguri M., Shimasaku K., Ono Y., 2018, *The Astrophysical Journal*, 854, 73
- Itoh R. et al., 2018, *The Astrophysical Journal*, 867, 46
- Izotov Y. I., Schaerer D., Thuan T. X., Worseck G., Guseva N. G., Orlitová I., Verhamme A., 2016, *Monthly Notices of the Royal Astronomical Society*, 461, 3683
- Izotov Y. I., Worseck G., Schaerer D., Guseva N. G., Thuan T. X., Fricke Verhamme A., Orlitová I., 2018, *Monthly Notices of the Royal Astronomical Society*, 478, 4851
- Jensen H., Hayes M., Iliev I. T., Laursen P., Mellema G., Zackrisson E., 2014, *Monthly Notices of the Royal Astronomical Society*, 444, 2114
- Jensen H., Laursen P., Mellema G., Iliev I. T., Sommer-Larsen J., Shapiro P. R., 2013, *MNRAS*, 428, 1366
- Jung I. et al., 2020, *The Astrophysical Journal*, 904, 144
- Kakiichi K., Gronke M., 2019, *Lyman Radiation Hydrodynamics of Turbulent H II Regions in Molecular Clouds: A Physical Origin of LyC Leakage and the Associated Ly $\pm$  Spectra*
- Katz H. et al., 2020, *Monthly Notices of the Royal Astronomical Society*, 494, 2200
- Kimm T., Blaizot J., Garel T., Michel-Dansac L., Katz H., Rosdahl J., Verhamme A., Haehnelt M., 2019, *Monthly Notices of the Royal Astronomical Society*, 486, 2215
- Kimm T., Cen R., 2014, *The Astrophysical Journal*, 788, 121
- Kimm T., Cen R., Devriendt J., Dubois Y., Slyz A., 2015, *Monthly Notices of the Royal Astronomical Society*, 451, 2900
- Kimm T., Katz H., Haehnelt M., Rosdahl J., Devriendt J., Slyz A., 2017, *Monthly Notices of the Royal Astronomical Society*, p. stx052
- Konno A. et al., 2014, *The Astrophysical Journal*, 797, 16
- Konno A. et al., 2017, *Publications of the Astronomical Society of Japan*, 70



- Kroupa P., 2001, *Monthly Notices of the Royal Astronomical Society*, 322, 231
- Kulkarni G., Keating L. C., Haehnelt M. G., Bosman S. E. I., Puchwein E., Chardin J., Aubert D., 2019, *Monthly Notices of the Royal Astronomical Society: Letters*, 485, L24
- Kulkarni G., Worseck G., Hennawi J. F., 2019, *Monthly Notices of the Royal Astronomical Society*, 488, 1035
- Kusakabe H. et al., 2020, *Astronomy & Astrophysics*, 638, A12
- Laursen P., Sommer-Larsen J., Andersen A. C., 2009, *ApJ*, 704, 1640
- Laursen P., Sommer-Larsen J., Milvang-Jensen B., Fynbo J. P. U., Razoumov A. O., 2019, *Astronomy & Astrophysics*, 627, A84
- Laursen P., Sommer-Larsen J., Razoumov A. O., 2011, *ApJ*, 728, 52
- Levermore C., 1984, *Journal of Quantitative Spectroscopy and Radiative Transfer*, 31, 149
- Li A., Draine B. T., 2001, *The Astrophysical Journal*, 554, 778
- Livermore R. C., Finkelstein S. L., Lotz J. M., 2017, *The Astrophysical Journal*, 835, 113
- Loeb A., Rybicki G. B., 1999, *The Astrophysical Journal*, 524, 527
- Ma X., Hopkins P. F., Kasen D., Quataert E., Faucher-Giguère C.-A., Kereš D., Murray N., Strom A., 2016, *Monthly Notices of the Royal Astronomical Society*, 459, 3614
- Mason C. A., Treu T., Dijkstra M., Mesinger A., Trenti M., Pentericci L., de Barros S., Vanzella E., 2018, *The Astrophysical Journal*, 856, 2
- Matthee J. J. A. et al., 2014, *Monthly Notices of the Royal Astronomical Society*, 440, 2375
- Mesinger A., 2010, *Monthly Notices of the Royal Astronomical Society*, 407, 1328
- Michel-Dansac L., Blaizot J., Garel T., Verhamme A., Kimm T., Trebitsch M., 2020, *Astronomy & Astrophysics*, 635, A154
- Miralda-Escude J., 1998, *ApJ*, 501, 15
- Mitchell P., Blaizot J., Cadiou C., Dubois Y., , 2020, *Tracing the simulated high-redshift circum-galactic medium with Lyman alpha emission*
- Neufeld D. A., 1990, *ApJ*, 350, 216
- Oesch P. A. et al., 2013, *The Astrophysical Journal*, 773, 75
- Okamoto T., Gao L., Theuns T., 2008, *MNRAS*, 390, 920
- Ouchi M. et al., 2017, *Publications of the Astronomical Society of Japan*, 70
- Ouchi M. et al., 2010, *ApJ*, 723, 869
- Paardekooper J.-P., Khochfar S., Dalla Vecchia C., 2015, *Monthly Notices of the Royal Astronomical Society*, 451, 2544
- Parsa S., Dunlop J. S., McLure R. J., 2017, *Monthly Notices of the Royal Astronomical Society*, 474, 2904
- Pentericci L. et al., 2018, *Astronomy & Astrophysics*, 619, A147
- Raiter A., Schaerer D., Fosbury R. A. E., 2010, *Astronomy & Astrophysics*, 523, A64
- Rasera Y., Teyssier R., 2005, *Astronomy & Astrophysics*, 445, 1
- Reddy N. A., Pettini M., Steidel C. C., Shapley A. E., Erb D. K., Law D. R., 2012, *The Astrophysical Journal*, 754, 25
- Robertson B. E., Ellis R. S., Furlanetto S. R., Dunlop J. S., 2015, *The Astrophysical Journal*, 802, L19
- Rosdahl J., Blaizot J., Aubert D., Stranex T., Teyssier R., 2013, *Monthly Notices of the Royal Astronomical Society*, 436, 2188
- Rosdahl J. et al., 2018, *Monthly Notices of the Royal Astronomical Society*
- Sadoun R., Zheng Z., Miralda-Escudé J., 2017, *The Astrophysical Journal*, 839, 44
- Santos M. R., 2004, *MNRAS*, 349, 1137
- Santos S., Sobral D., Matthee J., 2016, *Monthly Notices of the Royal Astronomical Society*, 463, 1678
- Schenker M. A., Stark D. P., Ellis R. S., Robertson B. E., Dunlop J. S., McLure R. J., Kneib J.-P., Richard J., 2012, *ApJ*, 744, 179
- Smith A., Ma X., Bromm V., Finkelstein S. L., Hopkins P. F., Faucher-Giguère C.-A., Kereš D., 2018, *Monthly Notices of the Royal Astronomical Society*, 484, 39
- Smith A., Safranek-Shrader C., Bromm V., Milosavljević M., 2015, *Monthly Notices of the Royal Astronomical Society*, 449, 4336
- Song M. et al., 2016, *The Astrophysical Journal*, 825, 5
- Stark D. P. et al., 2016, *Monthly Notices of the Royal Astronomical Society*, 464, 469
- Stark D. P., Ellis R. S., Chiu K., Ouchi M., Bunker A., 2010, *MNRAS*, 408, 1628
- 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., Bogosavljević M., Shapley A. E., Reddy N. A., Rudie G. C., Pettini M., Trainor R. F., Strom A. L., 2018, *The Astrophysical Journal*, 869, 123
- Taylor J., Lidz A., 2013, *Monthly Notices of the Royal Astronomical Society*, 437, 2542
- Teyssier R., 2002, *A&A*, 385, 337
- Tilvi V. et al., 2014, *The Astrophysical Journal*, 794, 5
- Toro E. F., Spruce M., Speares W., 1994, *Shock Waves*, 4, 25
- Trebitsch M., Blaizot J., Rosdahl J., Devriendt J., Slyz A., 2017, *Monthly Notices of the Royal Astronomical Society*, 470, 224
- Tumlinson J., Peebles M. S., Werk J. K., 2017, *Annual Review of Astronomy and Astrophysics*, 55, 389
- Tweed D., Devriendt J., Blaizot J., Colombi S., Slyz A., 2009, *A&A*, 506, 647
- Vanzella E. et al., 2016, *The Astrophysical Journal*, 825, 41
- Verhamme A., Dubois Y., Blaizot J., Garel T., Bacon R., Devriendt J., Guiderdoni B., Slyz A., 2012, *A&A*, 546, A111
- Verhamme A., Schaerer D., Atek H., Tapken C., 2008, *A&A*, 491, 89
- Verhamme A., Schaerer D., Maselli A., 2006, *A&A*, 460, 397
- Wise J. H., Demchenko V. G., Halicek M. T., Norman M. L., Turk M. J., Abel T., Smith B. D., 2014, *Monthly Notices of the Royal Astronomical Society*, 442, 2560
- Wisotzki L. et al., 2016, *Astronomy & Astrophysics*, 587, A98
- Yajima H., Li Y., Zhu Q., Abel T., Gronwall C., Ciardullo R., 2014, *MNRAS*, 440, 776
- Zheng Z.-Y. et al., 2017, *The Astrophysical Journal*, 842, L22



**Figure A1.** Distribution of the last-scattering radii of Ly $\alpha$  photons in the CGM.  $r_{50}$ ,  $r_{80}$ , and  $r_{95}$  are the scales at which 50, 80, and 95% of the Ly $\alpha$  photons undergo their last scattering for  $z = 6$  (top) and  $z = 9$  (bottom). For each galaxy, we compute the normed projected radius,  $r_{\text{proj}}^{\text{out}}/r_*$ , at which a given fraction of Ly $\alpha$  photons last-scatter before escaping the CGM.

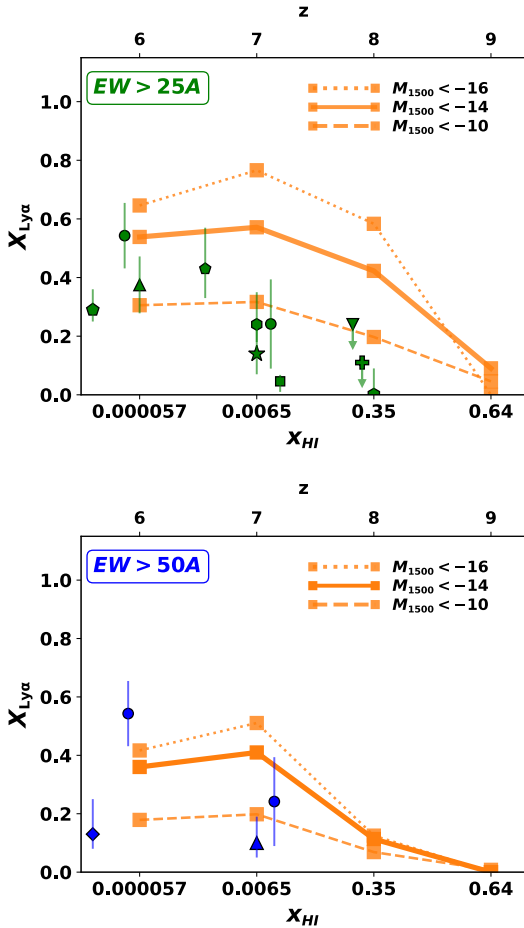
### APPENDIX A: LAST SCATTERINGS IN THE CGM

As discussed in Section 2.2.3, we need to define an arbitrary size for the CGM of galaxies in order to separate internal RT from IGM RT. Although we are interested in angle-averaged quantities in this study, we compute the IGM RT by assuming that any photon being scattered during its propagation in the IGM is removed from the line-of-sight, and therefore not transmitted to the observer. However, at CGM scale, photons that scatter have a probability to be re-directed back and forth on a given sightline, as demonstrated by the large projected extent of Ly $\alpha$  emission around high- $z$  star-forming galaxies (Steidel et al. 2011; Wisotzki et al. 2016). Choosing a CGM scale that is too small would lead us to remove photons during the IGM RT that still have a chance to scatter back towards the observer.

By choosing  $R_{\text{CGM}} = 10r_*$ , the number of scatterings in the relatively dense CGM occurring beyond this scale should be small, as suggested by Figure A1. This figure shows the distribution of distances at which  $n\%$  of the Ly $\alpha$  photons undergo their last scattering. In practice, for each galaxy, we compute the projected map of the last scatterings and then compute their distance to the center of the galaxy,  $r_{\text{proj}}^{\text{out}}$ . This is equivalent to imaging the CGM in Ly $\alpha$  and stacking over all directions. We then calculate the 3D radii at which 50, 80, 90, and 95% of the Ly $\alpha$  photons have their last scattering and plot the distribution for all galaxies at  $z = 6$  (top) and  $z = 9$  (bottom). The Ly $\alpha$  half-light radii (black curves) of galaxies correspond to  $\approx 1 - 2r_*$  for most galaxies. Interestingly, photons seem to scatter further out in the CGM at  $z = 9$  compared to  $z = 6$ , plausibly due the CGM being more neutral towards higher redshifts. Nevertheless, we see that the radius at which at least 95% of Ly $\alpha$  photons last-scatter is always below  $10r_*$  both at  $z = 6$  and  $z = 9$ . It is therefore reasonable to treat Ly $\alpha$  photon interactions with hydrogen atoms at  $r > R_{\text{CGM}}$  as IGM absorptions.

## APPENDIX B: VARIATION OF THE LAE FRACTION WITH UV MAGNITUDE

In Section 3.4, we showed the LAE fraction  $\chi_{\text{Ly}\alpha}$  predicted by SPHINX using a UV magnitude cut of  $M_{1500} = -14$ . In Figure B1, we test other values :  $M_{1500} = -16$  and  $-10$ . We see that using a brighter (fainter) cut increases (decreases) the amplitude of  $\chi_{\text{Ly}\alpha}$  at all redshifts. This is because the Ly $\alpha$  EW are, on average, correlated with UV magnitude in our simulation (Figure C2). Nevertheless, it is remarkable that the general trend of  $\chi_{\text{Ly}\alpha}$  as a function of redshift remains unchanged whatever the UV and EW cuts, i.e. the fraction of LAEs drops towards  $z = 8 - 9$  due to the IGM becoming more neutral.



**Figure B1.** Variation of the LAE fraction  $\chi_{\text{Ly}\alpha}$  with UV magnitude cuts.  $\chi_{\text{Ly}\alpha}$  is computed from IGM-transmitted Ly $\alpha$  luminosities and dust-attenuated UV magnitudes (i.e. the "after IGM" case). The top (bottom) panel shows  $\chi_{\text{Ly}\alpha}$  with LAEs selected with a Ly $\alpha$  EW threshold of 25 Å (50 Å). We show  $\chi_{\text{Ly}\alpha}$  for three different UV magnitude cuts :  $M_{1500} = -16$  (dotted line),  $M_{1500} = -14$  (solid line), and  $M_{1500} = -10$  (dashed line).

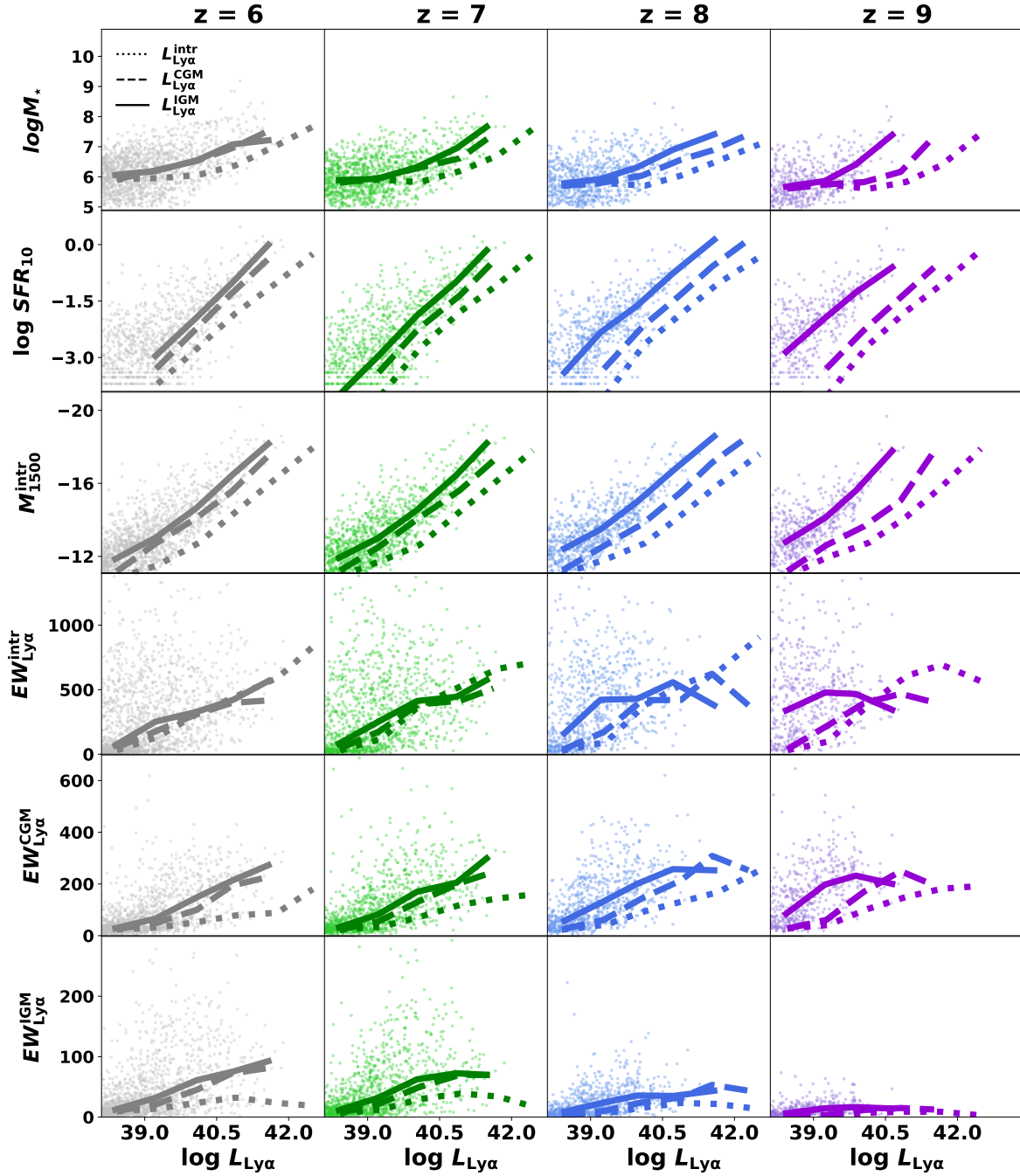
## APPENDIX C: SCALING RELATIONS

Figure C1 shows that intrinsic Ly $\alpha$  luminosities are positively correlated with the main galaxy properties in SPHINX, i.e. stellar mass, SFR, and UV magnitude. Interestingly, these scaling relations hold when we account for Ly $\alpha$  internal RT and IGM attenuation, except that Ly $\alpha$  luminosities are shifted to lower values. As discussed in Section 3.2.4, the  $M_{\star}$ - $L_{\text{Ly}\alpha}$  relation gives us indications on the Ly $\alpha$  luminosity completeness of our simulation due to our limited mass resolution. Star particles correspond to  $10^3 M_{\odot}$  and we chose, for the sake of this study, to only select galaxies with more than a hundred particles. From the top panel of Figure C1, we see that the highest IGM-attenuated Ly $\alpha$  luminosities allowed in objects at our stellar mass limit ( $10^5 M_{\odot}$ ) can reach  $\approx 10^{40}$  erg s $^{-1}$ , so we consider this value as our Ly $\alpha$  completeness limit.

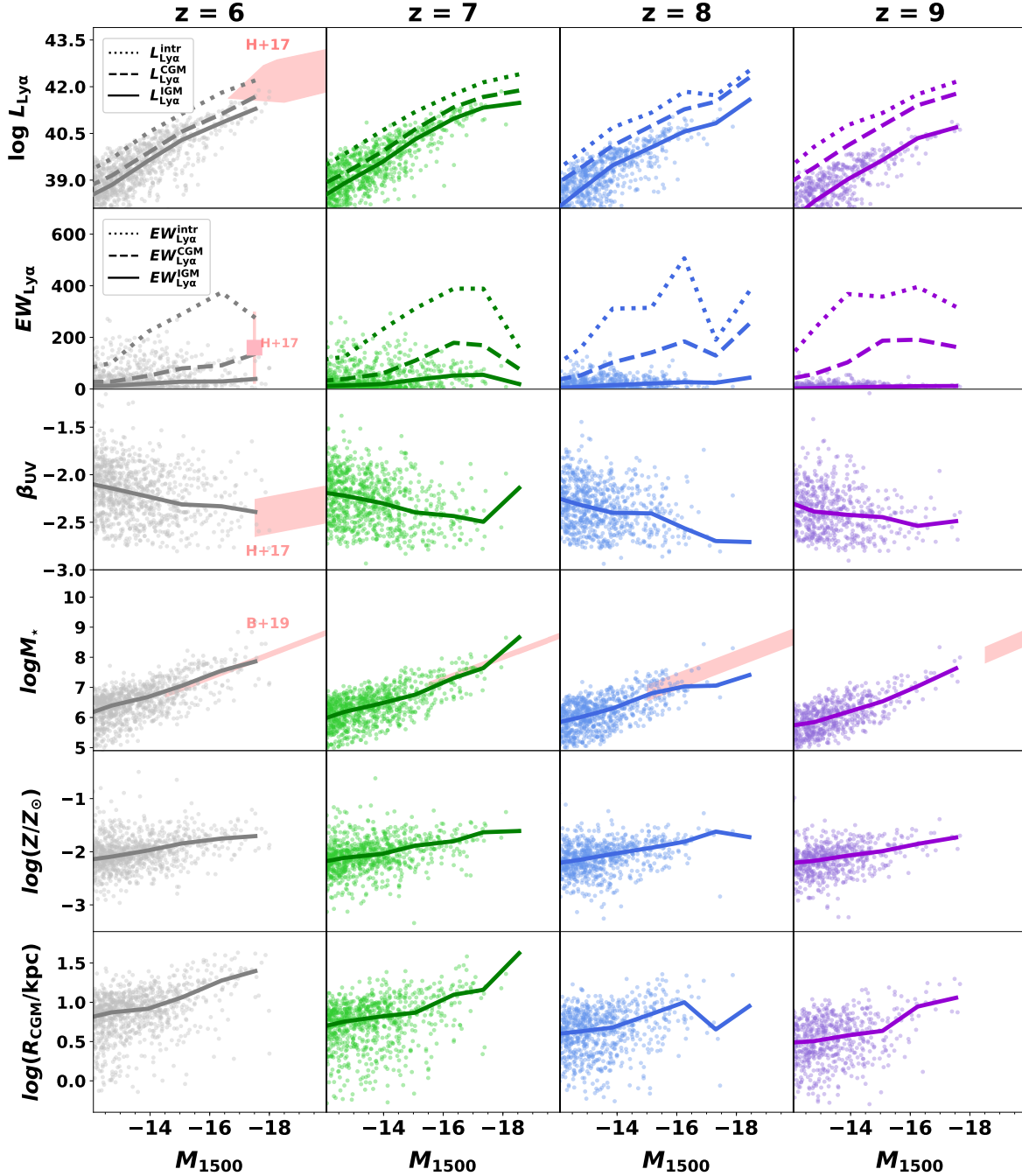
Ly $\alpha$  luminosities appear to be correlated with Ly $\alpha$  EW when looking at the median values (curves in Figure C1). This has the effect of predominantly reducing the number of objects at the faint-end of the Ly $\alpha$  LF rather than the bright-end when selecting LAEs above fixed EW cuts (see Section 3.2.3). Note that there is a very strong dispersion from one object to another in the  $L_{\text{Ly}\alpha}$ -EW relation (dots). This is mainly because intrinsic EW values are sensitive to metallicity (see Figure C2) and to variations of the recent SF histories of galaxies, where the continuum traces young stars over the last  $\lesssim 100$  Myr whereas Ly $\alpha$  is tracing hot massive stars at shorter timescales.

Figure C2 presents additional scaling relations and comparisons of SPHINX galaxy properties with observational data. It shows the relations between the dust-attenuated UV magnitudes and Ly $\alpha$  luminosities, Ly $\alpha$  EWs, UV slopes, stellar masses, and gas metallicities. Only our brightest sources can be compared with observations ( $M_{1500} \approx -18$ ) but we find very good agreement with existing constraints on the link between  $M_{1500}$  and  $L_{\text{Ly}\alpha}$ , EW,  $\beta_{\text{UV}}$  and  $M_{\star}$  (Hashimoto et al. 2017; Bhatwadekar et al. 2019).

The last row of Figure C2 shows the CGM radius of each individual galaxy. The  $R_{\text{CGM}}$  values range from a bit less than one pkpc to  $\approx 30$  pkpc and, on average, are larger for brighter sources. This is of the order of the DM halo virial radii in SPHINX (i.e. with masses between  $\approx 10^8$  and  $10^{11} M_{\odot}$ ; Rosdahl et al. 2018) which span a range between  $\approx 1$  and 20 pkpc at  $z = 6-9$ . In addition, typical instrument apertures usually have  $2''$  diameter, which corresponds to  $\approx 10$  pkpc at  $z = 6-9$  so some objects may be more extended than these typical apertures, hence inducing potential flux losses. However, we do not expect our observed Ly $\alpha$  luminosities to vary much as a function of the aperture size. Based on the discussion in Section 4.2 and 3.6.2, the Ly $\alpha$  emission is only arising from the ISM (i.e. at  $r < r_{\star}$ ) in our simulation, and radiative transfer in the CGM has a relatively small effect on the Ly $\alpha$  line shapes and intensities. Moreover, the radius encompassing 80% of the escaping Ly $\alpha$  flux ( $r_{80}$ ; see Figure A1) is 2-3 times smaller than  $R_{\text{CGM}}$  for the vast majority of galaxies. Therefore, the Ly $\alpha$  luminosities emerging from our galaxies should only weakly depend on the exact CGM scale at which they are measured and the expected aperture flux losses are thus moderate.



**Figure C1.** Scaling relations between Ly $\alpha$  luminosities and various galaxy properties at  $z = 6, 7, 8,$  and  $9$  (columns from left to right). The dotted, dashed, and solid lines represent the median galaxy properties per bin of intrinsic, dust-attenuated, and IGM-transmitted Ly $\alpha$  luminosities respectively. In each panel, the data points correspond to the IGM-transmitted luminosity of individual sources.  $M_{\star}$  : stellar mass (in  $M_{\odot}$ ),  $SFR_{10}$  : star formation rate over 10 Myr ( $M_{\odot} \text{ yr}^{-1}$ ),  $M_{1500}^{\text{intr}}$  : intrinsic UV magnitude,  $EW_{\text{Ly}\alpha}^{\text{intr/CGM/IGM}}$  : intrinsic/dust-attenuated/IGM-transmitted Ly $\alpha$  equivalent width (in  $\text{\AA}$ ; rest-frame).



**Figure C2.** Scaling relations between the dust-attenuated UV magnitude,  $M_{1500}$ , and various galaxy properties at  $z = 6, 7, 8,$  and  $9$  (columns from left to right). In each panel, solid lines show the median relations while data points represent individual sources. In the top two rows, the dotted, dashed, and solid lines show the median relation between  $M_{1500}$  and the intrinsic, dust-attenuated, and IGM-transmitted  $\text{Ly}\alpha$  luminosities (1st row) and EW (2nd row) respectively. Data points correspond to the IGM-transmitted  $\text{Ly}\alpha$  properties in these first two rows. The pink square and shaded areas represent observational constraints from Hashimoto et al. (2017) (H+17;  $5.5 < z < 6.7$ ) and Bhatawdekar et al. (2019) (B+19;  $z = 6, z = 7, z = 8, z = 9$ ), as labelled on the figure.  $L_{\text{Ly}\alpha}$ :  $\text{Ly}\alpha$  luminosity (in  $\text{erg s}^{-1}$ ),  $EW_{\text{Ly}\alpha}$ :  $\text{Ly}\alpha$  equivalent width (in  $\text{\AA}$ ; rest-frame),  $\beta_{\text{UV}}$ : UV slope,  $M_{\star}$ : stellar mass (in  $M_{\odot}$ ),  $Z$ : gas-phase metallicity in units of solar metallicity  $Z_{\odot}$ ,  $R_{\text{CGM}} (= 10r_{\star})$  in physical kpc.