The great escape: understanding the connection between Ly α emission and LyC escape in simulated JWST analogues

Nicholas Choustikov, Harley Katz, Aayush Saxena, Thibault Garel, Julien Devriendt, Adrianne Slyz, Taysun Kimm, Jeremy Blaizot, Joakim Rosdahl

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **532,** 2463–2484 (2024) https://doi.org/10.1093/mnras/stae1586
Advance Access publication 2024 July 10

The great escape: understanding the connection between Ly α emission and LyC escape in simulated *JWST* analogues

Nicholas Choustikov [1]★ Harley Katz,[1,2] Aayush Saxena [1,3] Thibault Garel,[4] Julien Devriendt,[1] Adrianne Slyz,[1] Taysun Kimm [5] Jeremy Blaizot [6] and Joki Rosdahl [6]

[1]*Sub-department of Astrophysics, University of Oxford, Keble Road, Oxford OX1 3RH, UK*
[2]*Department of Astronomy and Astrophysics, University of Chicago, 5640 S Ellis Avenue, Chicago, IL 60637, USA*
[3]*Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK*
[4]*Observatoire de Genève, Université de Genève, Chemin Pegasi 51, CH-1290 Versoix, Switzerland*
[5]*Department of Astronomy, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea*
[6]*CNRS, Centre de Recherche Astrophysique de Lyon UMR5574, Univ Lyon, Univ Lyon1, Ens de Lyon, F-69230 Saint-Genis-Laval, France*

Accepted 2024 June 19. Received 2024 June 18; in original form 2024 January 19

ABSTRACT

Constraining the escape fraction of Lyman Continuum (LyC) photons from high-redshift galaxies is crucial to understanding reionization. Recent observations have demonstrated that various characteristics of the Ly α emission line correlate with the inferred LyC escape fraction ($f_{\rm esc}^{\rm LyC}$) of low-redshift galaxies. Using a data set of 9600 mock Ly α spectra of star-forming galaxies at $4.64 \leq z \leq 6$ from the SPHINX^{20} cosmological radiation hydrodynamical simulation, we study the physics controlling the escape of Ly α and LyC photons. We find that our mock Ly α observations are representative of high-redshift observations and that typical observational methods tend to overpredict the Ly α escape fraction ($f_{\rm esc}^{\rm Ly\,\alpha}$) by as much as 2 dex. We investigate the correlations between $f_{\rm esc}^{\rm LyC}$ and $f_{\rm esc}^{\rm Ly\,\alpha}$, Ly α equivalent width (W_λ(Ly α)), peak separation ($v_{\rm sep}$), central escape fraction ($f_{\rm cen}$), and red peak asymmetry ($A_f^{\rm red}$). We find that $f_{\rm esc}^{\rm Ly\,\alpha}$ and $f_{\rm cen}$ are good diagnostics for LyC leakage, selecting for galaxies with lower neutral gas densities and less UV attenuation that have recently experienced supernova feedback. In contrast, W_λ(Ly α) and $v_{\rm sep}$ are found to be necessary but insufficient diagnostics, while $A_f^{\rm red}$ carries little information. Finally, we use stacks of Ly α, H α, and F150W mock surface brightness profiles to find that galaxies with high $f_{\rm esc}^{\rm LyC}$ tend to have less extended Ly α and F150W haloes but larger H α haloes than their non-leaking counterparts. This confirms that Ly α spectral profiles and surface brightness morphology can be used to better understand the escape of LyC photons from galaxies during the epoch of reionization.

Key words: galaxies: evolution – galaxies: high-redshift – dark ages, reionization, first stars – early Universe.

1 INTRODUCTION

The Universe transitioned from a predominantly neutral to ionized state by the redshift interval of $5 \lesssim z \lesssim 6$ (Kulkarni et al. 2019a; Keating et al. 2020; Planck Collaboration VI 2020; Becker et al. 2021; Bosman et al. 2022). However, the beginning of this process of reionization, the nature of the sources responsible, and the evolution of the neutral fraction all still remain uncertain. Observational upper and lower limits have been placed on the first of these questions by inferring star formation histories (SFHs) of high-redshift galaxies (e.g. Laporte et al. 2021) as well as by the Cosmic Microwave Background (e.g. Heinrich & Hu 2021). In contrast, model-dependent constraints have been derived for the evolution of the neutral fraction from the damping wings of high-redshift quasars (Davies et al. 2018; Greig, Mesinger & Bañados 2019; Ďurovčíková et al. 2020), the decrease in number densities of Lyman-α (Ly α) emitters (LAEs) (Stark et al. 2010; Pentericci et al. 2011; Mason et al. 2018; Jones et al. 2023) as well as the opacity of the Ly α forest (Fan, Carilli & Keating 2006a; Fan et al. 2006b).

There remains debate about the sources of ionizing photons needed to drive this period of reionization. Generally, due to number density arguments derived from the steep, faint-end slope of the high-redshift UV luminosity function (e.g. Bouwens et al. 2015), the primary candidates tend to be low-mass galaxies. Of these, the star-forming galaxies (SFGs) are typically considered due to the amounts of Lyman Continuum (LyC) photons they produce (e.g. Robertson et al. 2015). Beyond this however, other properties (e.g. mass, metallicity, sizes, morphology) of such galaxies remain largely unknown. This is important as the source model impacts the topology of reionization. Specifically, modifying the shape and amplitude of the 21-cm signal (Zaldarriaga, Furlanetto & Hernquist 2004; McQuinn et al. 2007; Kulkarni et al. 2017), as well as impacting the subsequent evolution of these galaxies themselves (e.g. Efstathiou 1992; Shapiro, Iliev & Raga 2004; Gnedin & Kaurov 2014; Rosdahl et al. 2018; Ocvirk et al. 2020; Katz et al. 2020a; Bird et al. 2022; Borrow et al. 2023). Beyond this, some of the ionizing photon budget is likely to be contributed by active galactic nuclei (AGNs, e.g. Madau & Haardt 2015; Chardin,

★ E-mail: nicholas.choustikov@physics.ox.ac.uk

Puchwein & Haehnelt 2017; Torres-Albà, Bosch-Ramon & Iwasawa 2020) or even more extreme, rare sources, including mini-quasars (Haiman & Knox 1999; Ricotti & Ostriker 2004; Hao, Yuan & Wang 2015) and shocks (Dopita et al. 2011; Wyithe, Mould & Loeb 2011). While all of these sources can produce a large number of ionizing photons, their rarity and possible dust obscuration limits their relative contribution. In the case of AGNs, they are only considered to dominate the ionizing budget at $z \lesssim 4$ (e.g. Kulkarni, Worseck & Hennawi 2019b; Dayal et al. 2020; Trebitsch et al. 2021).

In all cases, the production rate of LyC photons for a given source is given by the UV luminosity function ($\rho_{\rm UV}$), LyC photon production rate per UV luminosity ($\xi_{\rm ion}$), and the LyC photon escape fraction ($f_{\rm esc}^{\rm LyC}$), defined as the fraction of LyC photons that escape the virial radius of their host galaxy. Both $\rho_{\rm UV}$ and $\xi_{\rm ion}$ have been fairly well constrained, due to the fact that $\rho_{\rm UV}$ can be measured from deep imaging surveys (e.g. Bowler et al. 2020; Harikane et al. 2022; Donnan et al. 2023; Varadaraj et al. 2023), while $\xi_{\rm ion}$ can be predicted by stellar population synthesis models (e.g. Leitherer et al. 1999; Stanway & Eldridge 2018) or inferred from Balmer line emission (e.g. Maseda et al. 2020; Saxena et al. 2024). Uncertainties in $\xi_{\rm ion}$ are driven by differences in these models themselves (e.g. binaries, IMF, gas geometry etc.). In contrast, the escape fraction is much more difficult to measure directly. This is due to the fact that it is sensitive to complex non-linear physics in the interstellar medium (ISM), can be highly line-of-sight dependent, and cannot be directly measured at redshifts $\gtrsim 4$ due to the increasingly neutral intergalactic medium (IGM, e.g. Inoue et al. 2014; Worseck et al. 2014). As a result, studies of the LyC escape fraction at redshifts relevant to the epoch of reionization rely on indirect measurements, made using calibrated diagnostics. These diagnostics include investigating Ly α forest transmission through galaxies (e.g. Kakiichi et al. 2018) and studies of low-redshift 'analogue' galaxies (e.g. Leitherer et al. 2016; Schaerer et al. 2016; Izotov et al. 2018b). Recently, the low redshift Lyman continuum survey (LzLCS; Flury et al. 2022a, b) has pushed the frontier of low-redshift studies of LyC escape by observing 35 galaxies with confirmed LyC emission, nearly tripling the number of observed LyC leakers. However, the question of whether these 'analogue' galaxies are representative of high-redshift galaxies is still debated (Katz et al. 2022; Schaerer et al. 2022; Brinchmann 2023; Katz et al. 2023b). Furthermore, observational studies of low-redshift leakers are only able to see line-of-sight LyC emission, which is not necessarily correlated with global (i.e. angle-averaged) LyC emission (Paardekooper, Khochfar & Dalla Vecchia 2015; Fletcher et al. 2019): the quantity which is crucial to reionization. As a result, it is useful to also model LyC escape with cosmological radiation hydrodynamic simulations (e.g. Gnedin, Kravtsov & Chen 2008; Wise & Cen 2009; Kimm & Cen 2014; Xu et al. 2016; Trebitsch et al. 2017; Rosdahl et al. 2018, 2022). These simulations allow us to gain an insight into how such processes occur in high-redshift galaxies, while being informed by observational studies.

While a large number potential diagnostics have been suggested to infer LyC escape fractions (see e.g. Jaskot & Oey 2014; Nakajima & Ouchi 2014; Verhamme et al. 2017; Chisholm et al. 2018; Izotov et al. 2018b; Chisholm et al. 2022; Flury et al. 2022b; Saldana-Lopez et al. 2022; Choustikov et al. 2024 and references therein), some of the most accurate metrics involve Ly α emission. Due to the fact that Ly α is sensitive to the neutral H I opacity of its host galaxy (e.g. Verhamme et al. 2015) and neutral H I is the primary absorber of LyC photons, it follows that Ly α emission can contain information about the leakage of ionizing photons. Furthermore, due to the fact that Ly α is a resonant transition (Dijkstra 2017), the emergent spectrum holds a significant amount of information about the source (e.g. Orlitová et al. 2018), the intervening medium (e.g. Verhamme et al. 2017), as well as the geometry of the system (e.g. Blaizot et al. 2023).

Various properties of Ly α profiles have been suggested which trace ionizing photon production as well as physical conditions in the ISM favourable to LyC leakage. These include the Ly α equivalent width (W_λ(Ly α), e.g. Henry et al. 2015; Yang et al. 2017b; Steidel et al. 2018; Pahl et al. 2021), Ly α peak separation velocity ($v_{\rm sep}$, e.g. Verhamme et al. 2015; Izotov et al. 2018b, 2021), the asymmetry of the red peak ($A_f^{\rm red}$, e.g. Kakiichi & Gronke 2021; Witten, Laporte & Katz 2023), the central escape fraction ($f_{\rm cen}$, Naidu et al. 2022), and the Ly α escape fraction ($f_{\rm esc}^{\rm Ly\,\alpha}$, e.g. Dijkstra, Gronke & Venkatesan 2016; Verhamme et al. 2017; Izotov et al. 2020). This diversity of potential diagnostics highlights the necessity to understand the connection between the physics of the Ly α line and $f_{\rm esc}^{\rm LyC}$.

Numerical simulations have been used to study this interplay, including those of isolated galaxies (e.g. Verhamme et al. 2012; Behrens & Braun 2014), Ly α nebulae (e.g. Trebitsch et al. 2016; Gronke & Bird 2017), molecular clouds (e.g. Kimm et al. 2019; Kakiichi & Gronke 2021; Kimm et al. 2022), zoom-in simulations of individual galaxies (e.g. Laursen et al. 2019), and large box simulations with comparatively low resolution (e.g. Ocvirk et al. 2020; Gronke et al. 2021). Maji et al. (2022) studied the connections between intrinsic and escaped integrated Ly α luminosities for SPHINX galaxies and their LyC escape fractions, particularly finding a correlation between $f_{\rm esc}^{\rm Ly\,\alpha}$ and $f_{\rm esc}^{\rm LyC}$. However, this work was limited by not studying the line-of-sight variability in these quantities, as well as by focusing only on these two properties of Ly α emission. Building on this, Yuan et al. (2024) used the PANDORA suite of high-resolution zoom-in cosmological simulations (Martin-Alvarez et al. 2023) to investigate Ly α as a tracer of feedback-regulated LyC escape from a dwarf galaxy with varying physical models. It was concluded that there is a universal correlation between Ly α spectral shape parameters and $f_{\rm esc}^{\rm LyC}$ for a high time-cadence set of post-processed mock Ly α observations from a large number of sightlines. This reiterates our need to explore the connection between Ly α emission and LyC escape in a statistical sample of simulated epoch of reionization galaxies, while further exploring the physics underpinning each potential diagnostic.

The aim of this work is to investigate the line-of-sight escape and dust-attenuation of LyC and Ly α photons from a statistical sample of SPHINX^{20} galaxies (Rosdahl et al. 2022), inspired by the representative sample of star-forming high-redshift galaxies with well-resolved ISMs. Using the framework presented in Choustikov et al. (2024) to test potential diagnostics for the global LyC escape fraction, we explore the physics controlling the escape of Ly α photons and quantify the efficacy of various Ly α-based diagnostics in predicting the LyC escape fraction.

This work is organized as follows. In Section 2 we outline the numerical methods needed to simulate Ly α emission from and transmission through SPHINX^{20} galaxies. Section 3 presents and contextualizes the Ly α properties of SPHINX^{20} galaxies. In Section 4 we use our framework to contextualize and explain known Ly α diagnostics and explore the use of extended Ly α haloes as a potential diagnostic for the LyC escape fraction in stacked samples. Finally, in Section 5 we discuss caveats in our analysis of Ly α radiative transfer before concluding in Section 6.

2 NUMERICAL SIMULATIONS

There is a high degree of complexity to the physical processes underlying the production and escape of Lyman α (Ly $\alpha \sim$ 1215.67 Å) emission from galaxies. Therefore, we choose to employ state-of-

the-art high-resolution numerical simulations of galaxy formation to study the production of Ly α photons, as well as the resonant line physics involved in their escape. We employ the SPHINX^{20} cosmological radiation hydrodynamics simulation (Rosdahl et al. 2022). This simulation is ideal for such a study as the volume (20^3 cMpc^3) is large enough to capture a diversity of galaxies and it has enough resolution to model atomic cooling haloes (which contribute to LyC escape for reionization) as well as the multiphase ISM of these systems. Full details of the simulations are provided in Rosdahl et al. (2018, 2022). We use mock observations from the SPHINX Public Data Release v1 (SPDRv1; Katz et al. 2023a): a sample of 960 galaxies at redshifts $z \in 4.64, 5, 6$ with 10 Myr−averaged star formation rates (SFRs) ≥ 0.3 M$_\odot$yr^{-1}, representative of those observable by *JWST* (see discussions in sections 3.2 and 3.3.7 of Katz et al. 2023a, as well as Choustikov et al. 2024).

In order to generate the intrinsic emission of Ly α photons for each gas cell, we use the non-equilibrium hydrogen ionization fraction directly from the simulation. For all gas cells not containing unresolved Stromgren spheres, we then calculate the exact recombination and collisional components to Ly α emissivities for each gas cell. For gas cells with unresolved Stromgren spheres, we run a grid of spherical models using CLOUDY v17.03 (Ferland et al. 2017), iterated to convergence across a variety of gas densities, metallicities, ionization parameters, and electron fractions. The total intrinsic luminosity of each galaxy is then the integral over all gas cells within the virial radius. For a deeper discussion of the method used, see discussions in Choustikov et al. (2024) and Katz et al. (2023a).

Due to the fact that Ly α is a resonant line, in order to study its propagation through an attenuating, dusty medium it is crucial to capture the diffusion of Ly α photons both spatially and in frequency space as they escape their host galaxy. This physics is simulated using the RASCAS Monte-Carlo radiative transfer simulation code (Michel-Dansac et al. 2020). As dust evolution is not self-consistently modelled in SPHINX^{20}, we use the phenomenological Small Magellanic Cloud (SMC) dust model from Laursen, Sommer-Larsen & Andersen (2009) to assign dust to each gas cell based on their neutral gas fraction and metallicity. We also use dust asymmetry and albedo properties taken from the SMC dust model of Weingartner & Draine (2001). The propagation of these Ly α photons is followed until they pass the virial radius, at which point they are considered to have escaped into the intergalactic medium. Finally, as Ly α escape is a line-of-sight sensitive process, we use the peeling algorithm (Yusef-Zadeh, Morris & White 1984; Zheng & Miralda-Escudé 2002; Dijkstra 2017) to compute the Ly α spectrum for ten different sightlines. Due to the fact that the cost of the code scales with the number of sightlines, we select ten as a good balance between capturing the line-of-sight-driven uncertainty and computational cost. Recently, the angular dependence of these relations has been studied in more depth (Blaizot et al. 2023; Giovinazzo et al. 2024; Yuan et al. 2024). In the end, this results in a total sample of 9600 mock Ly α spectra and images. We also utilize both global and line-of-sight LyC escape fractions, computed for photons with wavelengths $\lambda < 912$Å and 900Å $< \lambda < 912$Å, respectively. The dust-attenuated SED for each line-of-sight is produced using similar methods (as detailed in Katz et al. 2023a and Choustikov et al. 2024).

Fig. 1 shows the dust-attenuated Ly α spectra (*middle*), composite three-colour images (*top border*), and dust-attenuated Ly α images (*bottom border*) for all ten lines-of-sight of the SPHINX^{20} galaxy with the highest escaping LyC luminosity at $z = 5$. The composite images are produced using *JWST* filters in the same way as the UNCOVER mosaic (Bezanson et al. 2022). Namely, F365W, F444W, and F410M are used for the red channel, F200W, and F277W are used for green, and F115W and F150W are used for blue (see also fig. 1 of Katz et al. 2023a). Border and line colours correspond to the same given line-of-sight. This demonstrates the fact that different sightlines produce drastically different Ly α spectra and images, even for the same galaxy at a given moment in time (see also Blaizot et al. 2023).

3 LY α PROPERTIES OF SPHINX^{20} GALAXIES

We begin by comparing Ly α emission properties of our sample of SPHINX^{20} galaxies with observations. The angle-averaged Ly α luminosity functions have already been discussed (Garel et al. 2021; Katz et al. 2022), while the line-of-sight equivalent was recently presented in fig. 31 of Katz et al. (2023a), finding very good agreement with observational constraints.

Here we discuss the distributions of line-of-sight Ly α luminosities, equivalent widths, and escape fractions.[1] Fig. 2 shows histograms of these quantities for the three redshifts considered compared to observational data spanning the redshift range $z \approx 3-8$ (Roy et al. 2023; Saxena et al. 2024). Where possible, both the intrinsic (dotted) and dust-attenuated (solid) values are shown. The effect of IGM attenuation on these properties is discussed in Appendix A. We find little redshift dependence, apart from the fact that galaxies at $z = 6$ tend to exhibit slightly larger equivalent widths and Ly α escape fractions. We also note that observational studies of LAEs are unlikely to include many galaxies with $f_{\rm esc}^{\rm Ly\,\alpha} \lesssim 1$ per cent. We see that dust attenuation decreases both the total Ly α luminosity and equivalent widths. This behaviour is consistent with the picture outlined in section 4.1 of Verhamme et al. (2012). Here, most UV continuum and Ly α photons are emitted from the cores of dense, star-forming clouds. The UV continuum photons are then strongly extinguished before escaping these birth clouds, as modelled by Charlot & Fall (2000). In contrast, the Ly α photons are scattered many times within the birth clouds before having a chance to escape. This increases the distance travelled by them and thus attenuates them more significantly than the UV photons.

It is important to establish that the Ly α escape fraction discussed in this work is different to that inferred observationally. While we compute $f_{\rm esc}^{\rm Ly\,\alpha}$ by directly calculating the number of Ly α photons to escape using RASCAS, traditionally this is done by inferring the intrinsic Ly α flux from the H α (or some other Balmer) line (e.g. Osterbrock 1989),

$$f_{\rm esc}^{\rm Ly\,\alpha} = \frac{L_{\rm Ly\,\alpha}}{8.7 L_{\rm H\,\alpha} \times 10^{0.4 A_{\rm H\,\alpha}}}, \quad (1)$$

where $L_{\rm Ly\,\alpha}$ and $L_{\rm H\,\alpha}$ are the dust-attenuated luminosity of Ly α and H α, respectively. The denominator is then the estimated intrinsic H α luminosity, inferred using the SMC dust attenuation law, $A_{\rm H\,\alpha}$ (Gordon et al. 2003),[2] converted to the intrinsic Ly α luminosity by means of a canonical factor discussed below. Fig. 3 shows a histogram of the true and inferred [from equation (1)] $f_{\rm esc}^{\rm Ly\,\alpha}$ where each bin is coloured by the average ratio of attenuation-corrected to true intrinsic H α luminosity. We find a mean absolute error between the two methods of 0.43 dex. Here, we see very clearly

[1] Note that when viewed along individual sightlines, the Ly α escape fractions can be > 100 per cent due to resonant scattering of photons emitted away from an observers sightline back into it (see e.g. Verhamme et al. 2012).

[2] This was done using the dust-extinction PYTHON package (v1.2) available at https://github.com/karllark/dust_extinction. The correction for H α is then $A_{\rm H\,\alpha} = 2.17 E(B-V)$, where $E(B-V)$ is measured using the Balmer decrement for each line of sight.

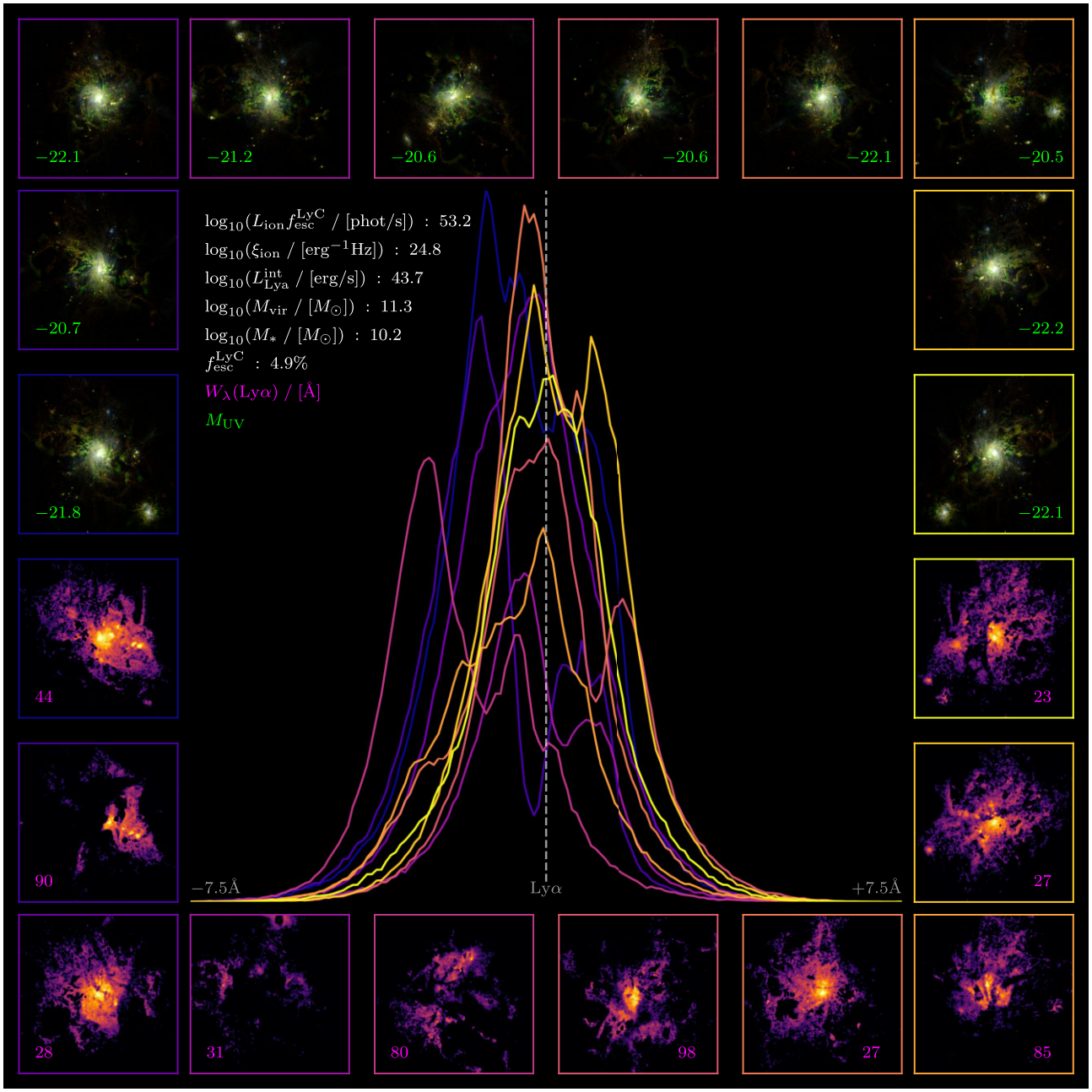

Figure 1. Complete mock data mosaic for the SPHINX[20] galaxy with the highest escaping LyC luminosity at $z = 5$. Dust-attenuated Ly α spectra (*middle*), three colour composite images (top border), and dust-attenuated Ly α images (bottom border) are shown for all ten sightlines used, indicated by the colours used. The RGB images are made using *JWST* filters: F365W, F444W, and F410M for the red channel, F200W and F277W for the green, while F115W and F150W are used for the blue channel [see also Bezanson et al. (2022) and Katz et al. (2023a)].

that this method of inferring $f_{\rm esc}^{{\rm Ly}\,\alpha}$ (which is typically used in observational studies, e.g. Hayes et al. 2005; Sobral & Matthee 2019; Flury et al. 2022b; Roy et al. 2023; Saxena et al. 2024) is only close to accurate when the dust-correction works sufficiently well to recover the intrinsic emission (i.e. for the purple pixels). To understand this, it is important to consider that equation (1) makes use of several key assumptions, which clearly break down for the orange pixels. First, it uses the fact that under case B assumptions, the intrinsic Ly α-to-H α ratio is fixed at some value (here, 8.7). In reality, this value depends on the electron number density as well as particularly the electron temperature (typically $n_e \sim 300\,{\rm cm}^{-3}$ and $T_e \sim 1 \times 10^4$ K are used). These are also assumed to be constant for a given galaxy, whereas fluctuations in such properties across galaxies are known to bias integrated measurements (e.g. Cameron, Katz & Rey 2023). Next, this analysis ignores contributions from the collisional emissivity to Ly α emission. While subdominant, this contribution can be important (see e.g. Kimm et al. 2019; Mitchell et al. 2021; Smith et al. 2022b, for discussions). For example, fig. 8 of Kimm et al. (2019) suggests that collisional radiation contributes about ~ 25 per cent of the intrinsic Ly α emission at the giant molecular cloud scale. This would imply an upper-limit systematic shift of 0.1 dex which would help us to explain the systematic

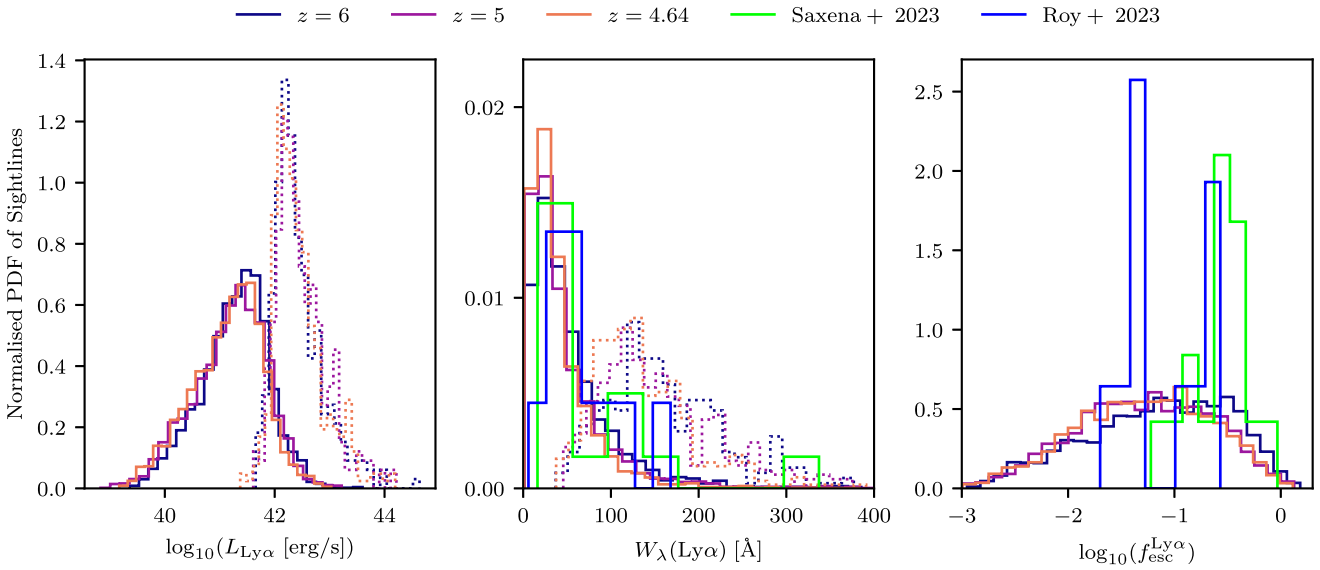

Figure 2. Distributions of Ly α properties of our sample, including the Ly α luminosity (left), Ly α equivalent width (centre), and Ly α escape fraction (right). Solid lines are given by dust-attenuated quantities, while intrinsic values are given where possible by dotted lines. Comparisons are given to observational data from Saxena et al. (2024) and Roy et al. (2023) in green and blue, respectively.

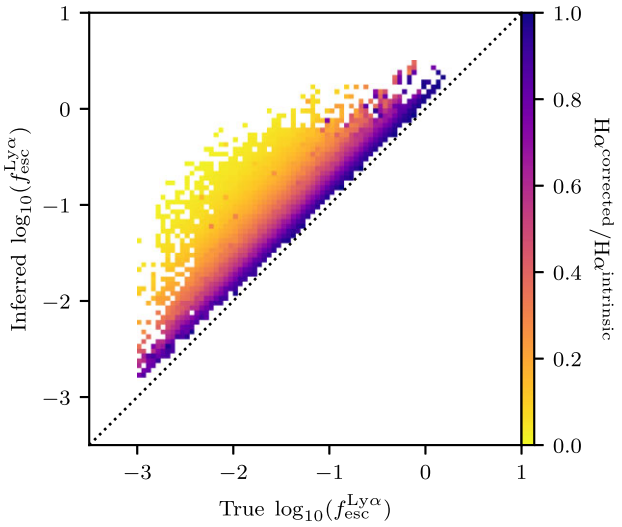

Figure 3. Histogram of Inferred $f_{\mathrm{esc}}^{\mathrm{Ly}\alpha}$ (equation 1) as a function of true $f_{\mathrm{esc}}^{\mathrm{Ly}\alpha}$. The colour of each bin is given by the average ratio of dust-corrected to intrinsic H α contained therein. The one-to-one relation is shown as a dotted line. We find that dusty galaxies with insufficient flux correction tend to overpredict $f_{\mathrm{esc}}^{\mathrm{Ly}\alpha}$. We find a mean absolute error of 0.43 dex.

overestimate in Fig. 3. Moreover, it is important to note that the collisional component is less affected by dust than the recombination component. This is due to the fact that while recombination photons are predominantly produced in dense, cold gas (as discussed above, see also Charlot & Fall 2000), collisional photons are also produced significantly in more diffuse gas where it is easier for them to escape unencumbered. Finally, the geometry of the dust-screen model is important. Methods such as equation (1) typically assume a uniform dust screen. However, it has been shown that clumpy dust screen models better reproduce theoretical line ratios (see Scarlata et al. 2009, and references therein), while also representing the more realistic and complex dust geometry found in simulated galaxies.

Beyond including the modelling of the contributions discussed above, another strategy to better infer $f_{\mathrm{esc}}^{\mathrm{Ly}\alpha}$ at lower redshifts may be to make use of detections of other transitions of hydrogen – most notably the Paschen series. For example, recently Reddy et al. (2023) used JWST/NIRSpec observations of Paschen lines in galaxies at $1 \lesssim z \lesssim 3$ to re-evaluate dust-extinction curves, finding that Balmer-inferred estimates were insufficient. While this evidence is marginal, it certainly points to the need to better understand dust attenuation and dust-obscured regions. Another option would be to exploit the synergy of rest-frame UV/NIR *JWST* observations of galaxies during the epoch of reionization with rest-frame IR/FIR measurements from ALMA, a technique which is beginning to be explored (e.g. Rujopakarn et al. 2023). Here, the total dust emission can be used to better estimate the attenuation of UV/optical emission lines. However, at such redshifts the point-spread function and sensitivity of the instrument can rapidly become an issue.

This overestimate is likely to be unimportant in studies comparing $f_{\mathrm{esc}}^{\mathrm{Ly}\alpha}$ to the escape of ionizing radiation, as sightlines for which the dust-correction works well are likely to have little dust and therefore significant LyC leakage. Otherwise, this bias is clearly important to include in error estimation of reported Ly α escape fractions.

Next, we explore the effect of this overestimate in $f_{\mathrm{esc}}^{\mathrm{Ly}\alpha}$on the $f_{\mathrm{esc}}^{\mathrm{Ly}\alpha} - W_\lambda(\mathrm{Ly}\alpha)$ relation. Fig. 4 shows $f_{\mathrm{esc}}^{\mathrm{Ly}\alpha}$ as a function of Ly α equivalent widths, coloured by the observed Ly α luminosity, compared to high-redshift observations (Roy et al. 2023; Saxena et al. 2024). The empirical estimator from Sobral & Matthee (2019) for $2.2 \lesssim z \lesssim 2.6$ is plotted in dashed black, as well as the running mean for SPHINX[20] in lilac. On the *left*, we show the true $f_{\mathrm{esc}}^{\mathrm{Ly}\alpha}$, while on the *right* we use the H α-inferred $f_{\mathrm{esc}}^{\mathrm{Ly}\alpha}$ as given by equation (1). We find that these three quantities are well correlated. However, we find that comparisons with the relation from Sobral & Matthee (2019) depend strongly on the version of $f_{\mathrm{esc}}^{\mathrm{Ly}\alpha}$ used. Specifically, this estimator tends to systematically overestimate the true Ly α escape fractions of our mock observations by ~ 0.3 dex. As discussed above, in the case of the H α-inferred values however, Ly α escape fractions become overestimated, shifting the distribution closer to the relation from

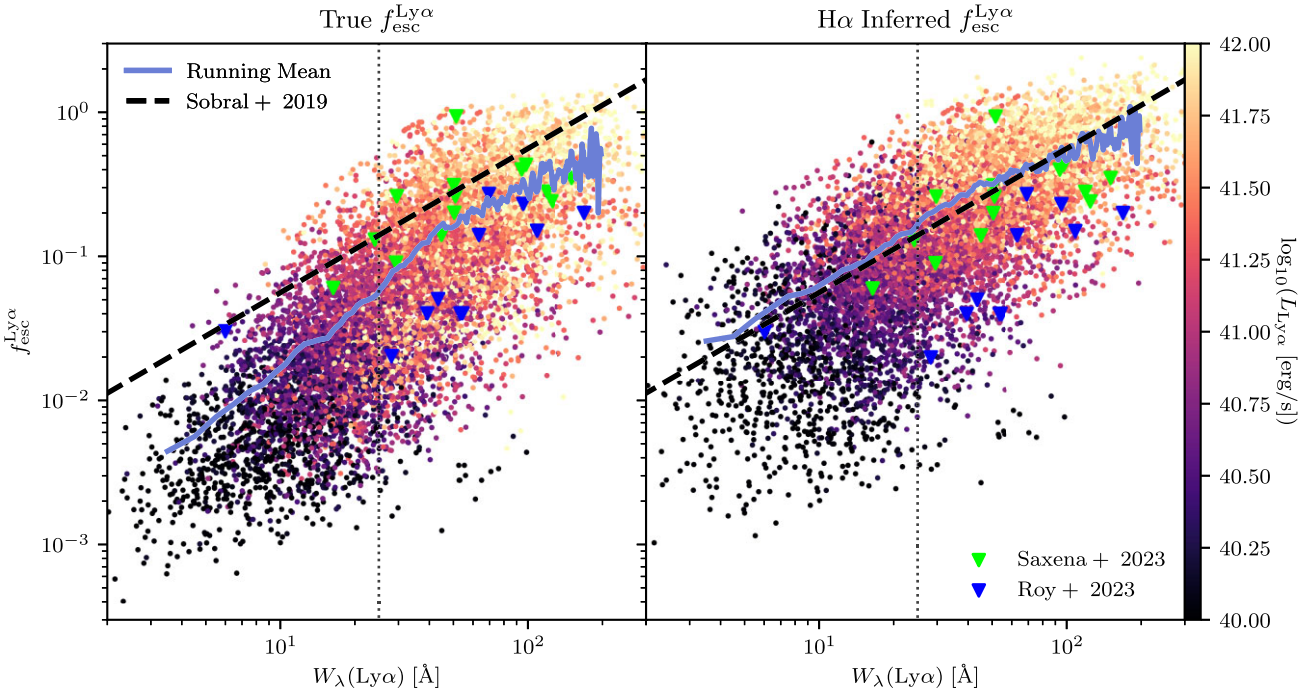

Figure 4. Ly α escape fractions as a function of Ly α equivalent widths for galaxies in SPHINX[20], coloured by their observed Ly α luminosity. On the left, we use the true Ly α escape fraction as computed by RASCAS, while on the *right* we use the H α-inferred value as computed by equation (1). The empirical estimator for $2.2 \lesssim z \lesssim 2.6$ from Sobral & Matthee (2019) is shown in dashed black, along with our running mean in lilac. The typical observational cut for LAEs of $W_\lambda({\rm Ly}\,\alpha) \geq 25$Å is also shown in black. Observational data from Saxena et al. (2024) and Roy et al. (2023) are given as a comparison.

Sobral & Matthee (2019). Furthermore, lines-of-sight that produce lower Ly α equivalent widths and luminosities (and therefore tend to be dustier) tend to more severely underpredict the intrinsic luminosity of H α, therefore overpredicting $f_{\rm esc}^{{\rm Ly}\,\alpha}$ even more. This has the effect of reducing the gradient of the running mean, such that it matches the empirical estimator remarkably well. Overall, our mock observations of SPHINX[20] galaxies reproduce the trends and spread from *JWST* observations of Ly α emission in the high-redshift Universe, as shown by the points in blue and green.

In general, the numerator of the Ly α equivalent width is set by the Ly α escape fraction, implying that we should expect the correlation evident in Fig. 4. However, $W_\lambda({\rm Ly}\,\alpha)$ also depends on the impact of sSFR and local ISM properties on the continuum at 1216 Å. Therefore, one would expect a correlation between the two quantities. However, matching the observational trend perfectly also requires the local state of the ISM to be realistic in order to produce the correct intrinsic emission (Osterbrock 1989). Therefore, we can be cautiously optimistic that the multiphase ISM of SPHINX[20] reproduces conditions in and around real H II regions reasonably well.

4 LY α DIAGNOSTICS FOR THE LYC ESCAPE FRACTION

We are now able to test the efficacy of known Ly α diagnostics for LyC escape from SPHINX[20] galaxies. In order to understand the underlying physics, we utilize the framework presented in Choustikov et al. (2024), which argues that a good diagnostic for high LyC leakage should:

(i) select for galaxies with a high specific star formation rate (sSFR),

(ii) be sensitive to stellar population ages ≥ 3.5 Myr and ≤ 10 Myr to catch galaxies that have undergone supernovae feedback, but are still producing ionizing radiation,

(iii) contain a proxy for the density and neutral state of the ISM.

Here, we define sSFR as the 10 Myr-averaged star formation rate normalized by the stellar mass, we use a mean stellar population age weighted by the ionizing luminosity contribution of each star particle (in order to focus on the population crucial to producing LyC photons) and, following Choustikov et al. (2024), define the composite parameter $\zeta_{\rm ISM} = {\rm E(B-V)} \times \langle n_{\rm HI} \rangle_{\rm [OII]} / [{\rm cm}^{-3}]$ as the product of the UV attenuation as well as the [OII] $\lambda\lambda$3726, 3728-weighted density of neutral hydrogen. This is used for demonstrative purposes to isolate information about the density and neutral state of the ISM. This proxy is used as it is difficult to find a suitable property which is uncorrelated with the age or SFR of the galaxy. It is important to note that galaxies falling outside of these criteria are **not** omitted from any analysis or discussion in this work. We use this framework to understand correlations (or lack thereof) between Ly α properties and LyC leakage.

When comparing with direct observational data, we use the line-of-sight values for the LyC escape fraction ($f_{\rm esc,\,LOS}^{\rm LyC}$). Our goal however is to understand correlations with the global (i.e. angle-averaged) LyC escape fraction ($f_{\rm esc}^{\rm LyC}$) as this is the relevant quantity for reionization. Scatter between these two quantities is discussed in Appendix B. All Ly α quantities discussed are dust-attenuated using RASCAS and observed along a given sightline.

4.1 Ly α escape fraction, $f_{\rm esc}^{{\rm Ly}\,\alpha}$

It is well-established that the Ly α and LyC escape fractions are well-correlated (e.g. Dijkstra et al. 2016; Verhamme et al. 2017; Steidel et al. 2018; Gazagnes et al. 2020; Izotov et al. 2021; Pahl et al. 2021;

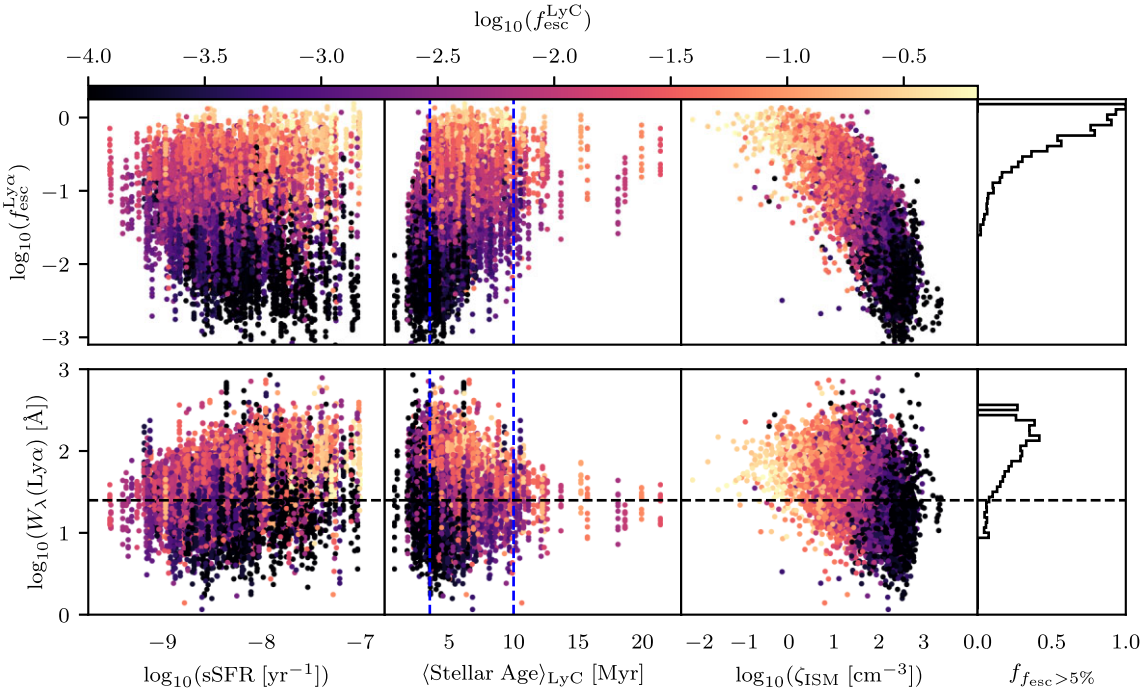

Figure 5. (Top) Ly α escape fraction, $f_{\rm esc}^{{\rm Ly}\,\alpha}$, as a function of sSFR (left), LyC luminosity-weighted mean stellar age (*centre*), and neutral gas attenuation parameter $\zeta_{\rm ISM}$ (right) coloured by the global LyC escape fraction, $f_{\rm esc}^{\rm LyC}$. The attached histogram (right) shows the fraction of galaxies for a given $f_{\rm esc}^{{\rm Ly}\,\alpha}$ with $f_{\rm esc}^{\rm LyC} > 5$ per cent for each bin that contains at least five such galaxies. While not correlating well with sSFR, high $f_{\rm esc}^{{\rm Ly}\,\alpha}$ selects for galaxies with mean stellar ages > 3.5 Myr and low $\zeta_{\rm ISM}$. Therefore, $f_{\rm esc}^{{\rm Ly}\,\alpha}$ is a good indicator for LyC escape. (Bottom) Same as above, but for Ly α equivalent widths. $W_\lambda({\rm Ly}\,\alpha)$ correlates weakly with sSFR but does not trace the mean stellar age. However, larger Ly α equivalent widths select for galaxies with lower $\zeta_{\rm ISM}$. Thus, $W_\lambda({\rm Ly}\,\alpha)$ is a necessary but insufficient diagnostic for $f_{\rm esc}^{\rm LyC}$.

Katz et al. 2022; Maji et al. 2022). This is expected, given the fact that both of these photons can be absorbed by similar components of the ISM. However, LyC photons can be absorbed by both dust and H I while Ly α is mainly absorbed by dust[3] and can be scattered back into a given sightline. As a result, is clear that one would naively expect $f_{\rm esc}^{{\rm Ly}\,\alpha}$ to be greater than $f_{\rm esc}^{\rm LyC}$ for a given sightline.

In the top row of Fig. 5, we plot the Ly α escape fraction as a function of sSFR (*left*), mean stellar population age (*centre*), and the composite ISM parameter $\zeta_{\rm ISM}$ (*right*), coloured by $f_{\rm esc}^{\rm LyC}$. While $f_{\rm esc}^{{\rm Ly}\,\alpha}$ does not correlate with sSFR, we know from Fig. 4 that it correlates with $L_{{\rm Ly}\,\alpha}$, which itself does correlate with SFR (e.g. Sobral et al. 2018). As a result, selecting for galaxies with very high $f_{\rm esc}^{{\rm Ly}\,\alpha}$ does marginally systematically select for systems with higher sSFR. Next, we find that observations with high $f_{\rm esc}^{{\rm Ly}\,\alpha}$ match very well with galaxies within the correct stellar population age, particularly selecting against galaxies with ages $\lesssim 3.5$ Myr where the birth clouds are yet to be disrupted by SNe. Finally, as expected we find that $f_{\rm esc}^{{\rm Ly}\,\alpha}$ correlates with the state of the ISM. Lines-of-sight with less neutral hydrogen and dust attenuation tend to produce higher Ly α escape fractions (e.g. Verhamme et al. 2015; Dijkstra et al. 2016; Dijkstra 2017; Verhamme et al. 2017). As a result, $f_{\rm esc}^{{\rm Ly}\,\alpha}$ selects for galaxies with the optimal mean stellar ages and traces $\zeta_{\rm ISM}$, thus satisfying two criteria very well, making it a good indicator for high LyC escape fractions. This is seen particularly easily in the histogram (*top-right* of Fig. 5), which shows the fraction of galaxies in each bin of $f_{\rm esc}^{{\rm Ly}\,\alpha}$ which have $f_{\rm esc}^{\rm LyC} \geq 5$ per cent.

[3]Deuterium and molecular hydrogen can also absorb Ly α photons, but their effect is minimal.

Fig. 6 shows both variations in the line-of-sight LyC escape fraction, $f_{\rm esc,\,LOS}^{\rm LyC}$ (*left*) as well as the global LyC escape fraction, $f_{\rm esc}^{\rm LyC}$ (*right*) as a function of the Ly α escape fraction, coloured by $f_{\rm esc}^{\rm LyC}\xi_{\rm ion}$. We also include observational data (Flury et al. 2022b; Izotov et al. 2022, 2024) in green, black, and cyan respectively, as well as the line of best fit from Izotov et al. (2024) in black. We find that our mock observations reproduce the observational trend very well. Indeed, as discussed above it is clear that $f_{\rm esc}^{{\rm Ly}\,\alpha}$ is a suitable diagnostic for the LyC escape fraction, with $f_{\rm esc}^{{\rm Ly}\,\alpha}$ being a somewhat stronger predictor of the global than line-of-sight LyC escape fraction. Furthermore, we find that in general, $f_{\rm esc}^{\rm LyC} < f_{\rm esc}^{{\rm Ly}\,\alpha}$, with the systems for which this is not the case having the largest $f_{\rm esc}^{\rm LyC}\xi_{\rm ion}$. In both cases we find that mock observations with larger Ly α escape fraction tend to have larger $f_{\rm esc}^{\rm LyC}\xi_{\rm ion}$, suggesting that galaxies with high Ly α escape fractions might contribute decisively to reionization. However, it is clear from the *right* panel that there is significant scatter in this relation.

This scatter is driven predominantly by the line-of-sight-dependent nature of Ly α, where Ly α photons can scatter out of or into a given aperture. However, it is important to note that the sources of Ly α (gas) and LyC (mostly stars) photons are distributed differently in galaxies, and therefore have different escape channels. For example, Maji et al. (2022) found that six of their analysed SPHINX galaxies had $f_{\rm esc}^{\rm LyC} > f_{\rm esc}^{{\rm Ly}\,\alpha}$, corresponding to galaxies with dusty, low H I density escape channels close to their centres. Indeed, in Fig. 6 we find a number of such systems too, particularly when global $f_{\rm esc}^{\rm LyC}$ is considered, agreeing too with low-redshift observations.

There is a built-in dependence for the global LyC escape fraction on metallicity due to our implementation of dust. Namely, in systems with little metallicity and dust, the Ly α escape fraction will be close

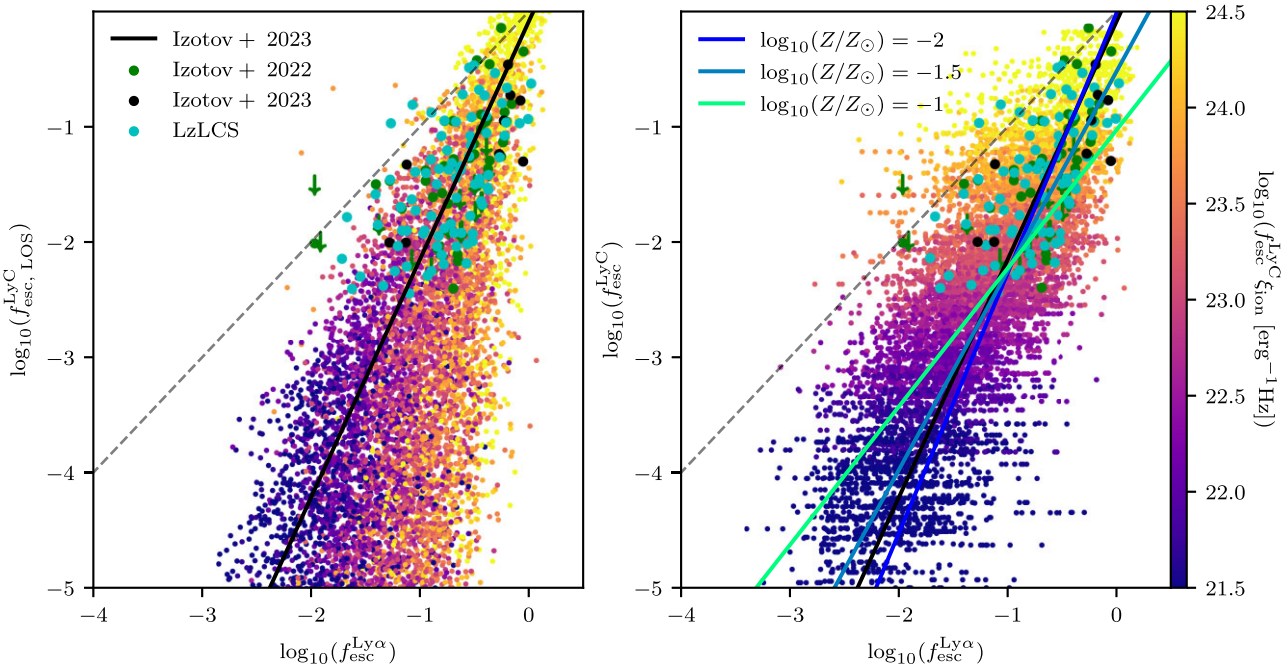

Figure 6. (Left) Line-of-sight LyC escape fraction versus line-of-sight Ly α escape fraction, coloured by $f_{\rm esc}^{\rm LyC}\xi_{\rm ion}$. The one-to-one relation is shown as a dashed line. We overplot observational data from Flury et al. (2022b), Izotov et al. (2022), and Izotov et al. (2024) in cyan, green, and black, respectively. (Right) Global LyC escape fraction as a function of line-of-sight $f_{\rm esc}^{{\rm Ly}\,\alpha}$, coloured by $f_{\rm esc}^{\rm LyC}\xi_{\rm ion}$. We show the one-to-one relation, the relation given by Izotov et al. (2024), as well as lines of best fit given by equation (2) for $\log_{10}(Z/Z_\odot) \in \{-2, -1.5, -1\}$. In both cases, we find a good correlation between the two quantities. A clear gradient is visible. Galaxies with larger $f_{\rm esc}^{{\rm Ly}\,\alpha}$ effectively contribute more to reionization of the IGM, as they have the largest global values of $f_{\rm esc}^{\rm LyC}\xi_{\rm ion}$.

to 100 per cent, irrespective of the LyC escape fraction (which can be orders of magnitude lower). To this end, we include a best-fitting relation depending on both the Ly α escape fraction and metallicity given by:

$$\log_{10}(f_{\rm esc}^{\rm LyC}) = \left[0.1429 - 1.057\log_{10}(Z/Z_\odot)\right]\log_{10}(f_{\rm esc}^{{\rm Ly}\,\alpha}) - 1.015\log_{10}(Z/Z_\odot) - 2.014, \quad (2)$$

Lines are overplotted for $\log_{10}(Z/Z_\odot) \in \{-2, -1.5, -1\}$. Here, we can see that the gradient of this relation is most strongly affected, steepening for metal-poor populations with less dust, in line with the above discussion. As expected, the line for low-intermediate metallicities of $\sim 0.03 Z_\odot$ visually fits our data best.

4.2 Ly α equivalent width, W_λ(Ly α)

Previously, the Ly α equivalent width, W_λ(Ly α) has been found to depend strongly on the H I covering fraction, as well as on the optical depth (e.g. Reddy et al. 2016; Gazagnes et al. 2018). As a result, it is reasonable to expect that greater W_λ(Ly α) should correlate with increased $f_{\rm esc}^{\rm LyC}$. It has also been found that Ly α doublets with large red to blue ratios tend to have higher equivalent widths (Blaizot et al. 2023). Given the fact that such signals tend to correspond to outflows (signatures of effective feedback), it further suggests that this trend should exist provided that LyC leakage is a feedback-regulated quantity (e.g. Trebitsch et al. 2017). Steidel et al. (2018) and Pahl et al. (2021) found a very strong correlation between $f_{\rm esc}^{\rm LyC}$ and W_λ(Ly α), while a slightly weaker correlation was found by the LzLCS (Flury et al. 2022b), for the Green Peas (GPs; Henry et al. 2015; Yang et al. 2017b), and galaxies from the VANDELS survey (Saldana-Lopez et al. 2023).

Fig. 5 (*bottom row*) shows W_λ(Ly α) as a function of sSFR, mean stellar age and $\zeta_{\rm ISM}$ coloured by $f_{\rm esc}^{\rm LyC}$. We find that W_λ(Ly α) correlates very weakly with sSFR. Furthermore, greater equivalent widths tend to select for galaxies with stellar ages $\lesssim 10$ Myr, while selecting for systems with less neutral gas and UV attenuation. As a result, W_λ(Ly α) satisfies two criteria weakly, making it a helpful but insufficient diagnostic for LyC leakage. We do note however that the typical selection cut of equivalent widths greater than 25 Å produces a sample with a much larger number of strong LyC leaking galaxies, as seen in the Ly α histogram (*bottom-right*) of Fig. 5. We also see a slight decrease in the fraction of galaxies with $f_{\rm esc}^{\rm LyC} > 5$ per cent in bins with equivalent widths of W_λ(Ly α) $\gtrsim 250$ Å. This is due to the fact that such a sample is contaminated by galaxies with young stellar populations ($\lesssim 3.5$ Myr) and large $\zeta_{\rm ISM}$ that have yet to clear the dusty ISM surrounding the stellar nursery in order to let LyC photons to escape.

Fig. 7 shows the line-of-sight LyC escape fraction as a function of W_λ(Ly α) coloured by $f_{\rm esc}^{\rm LyC}\xi_{\rm ion}$, compared to observational data (Pahl et al. 2021; Flury et al. 2022b) in cyan and blue respectively. We find some correlation (albeit with a lot of scatter) between the two quantities, with virtually all strong leakers being bright LAEs ($W_\lambda > 25$ Å). While we find some good agreement with Flury et al. (2022b), we find poor agreement with the rest of the observational data as well as the relation from Pahl et al. (2021), particularly for bright LAEs. This may be due to selection effects as the galaxies used in stacks by Steidel et al. (2018) and Pahl et al. (2021) are significantly more UV-bright than those generally found in SPHINX[20], filling out the region of galaxies with moderate $f_{\rm esc,\,LOS}^{\rm LyC}$ and low W_λ(Ly α). We find that the SPHINX[20] galaxies that best compare to these stacks tend to have the strongest UV flux and youngest mean stellar population ages, corresponding to starburst galaxies that are yet to experience strong feedback.

There is another effect to consider. If the LyC escape fraction is close to 100 per cent, then the neutral hydrogen column density

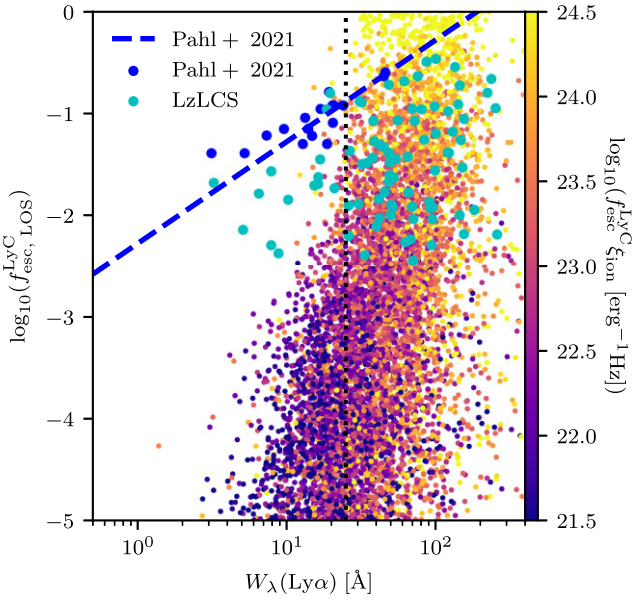

Figure 7. Line-of-sight LyC escape fraction as a function of Ly α equivalent widths, coloured by $f_{\rm esc}^{\rm LyC}\xi_{\rm ion}$. This is compared to observational data from Pahl et al. (2021), and Flury et al. (2022b) in blue and cyan, respectively. The relation from Pahl et al. (2021) is also shown as a dashed blue line. We find that observations with large $f_{\rm esc,\,LOS}^{\rm LyC}$ all have W_λ(Ly α) > 25 Å, and are therefore bright LAEs. This however is not definitively indicative of LyC leakage.

must be very low. As a result, this case should produce lower Ly α equivalent widths (e.g. Nakajima & Ouchi 2014; Steidel et al. 2018) as fewer Ly α photons are produced. However, Ly α equivalent widths also depend strongly on the stellar continuum flux and stellar population source model. In reality, it is unclear whether we should expect such a bimodal distribution.

Interestingly, the performance of W_λ(Ly α) as an indicator for LyC leakage is similar to that of the equivalent width of Balmer lines such as Hβ (W_λ(Hβ)). Comparing this discussion with that of W_λ(Hβ) in a previous study (see section 4.1.5 of Choustikov et al. 2024), we find that larger Ly α equivalent widths correlate worse with sSFR but are less polluted with galaxies with mean stellar ages younger than 3.5 Myr. We conclude that the equivalent widths of Hydrogen lines should only be used in combined diagnostics for LyC escape fractions, such as the W_λ(Hβ) − β diagram (Zackrisson, Inoue & Jensen 2013; Zackrisson et al. 2017). In general, we find that large Ly α equivalent widths are a necessary but insufficient criterion for strong LyC leakage. Selecting for bright LAEs (as indicated by the vertical dotted line in Fig. 7) manages to capture all strong line-of-sight LyC leakers. Therefore, W_λ(Ly α) can be used as a useful first stage in any attempt to find LyC leakers from a sample of galaxies with confirmed Ly α emission.

4.3 Peak separation, $v_{\rm sep}$

Ly α photons must scatter away from line centre in order to escape. As a result, this process often drives the production of a double peak in Ly α spectra. In such cases, the velocity shift corresponding to the separation between the two peaks of the doublet, $v_{\rm sep}$ is a natural quantity to investigate. Due to the fact that scattering in frequency space depends on the H I column density, from a theoretical basis $N_{\rm HI}$ should correlate with $v_{\rm sep}$ (e.g. Dijkstra, Haiman & Spaans 2006; Verhamme et al. 2015; Kakiichi & Gronke 2021) and therefore we expect $v_{\rm sep}$ to correlate with $f_{\rm esc}^{\rm LyC}$ (see also Kimm et al. 2019; Yuan et al. 2024). Observationally, the relationship between $v_{\rm sep}$ and $f_{\rm esc}^{\rm LyC}$ has been well explored (e.g. Verhamme et al. 2017; Izotov et al. 2018a; Flury et al. 2022b).

The *top row* of Fig. 8 shows $v_{\rm sep}$ as a function of sSFR, mean stellar age, and $\zeta_{\rm ISM}$ coloured by the global $f_{\rm esc}^{\rm LyC}$. We find no trend between $v_{\rm sep}$ and sSFR. Next, we find that spectra with large peak separations ($> 250\,{\rm km\,s^{-1}}$) select for younger mean stellar populations with more UV attenuation and a higher neutral gas density. Therefore, these systems are unlikely to be LyC leakers, agreeing with previous work (e.g. Verhamme et al. 2015; Flury et al. 2022b; Naidu et al. 2022). As a result, we find that Ly α peak separations $\lesssim 250\,{\rm km\,s^{-1}}$ weakly fulfill two of the criteria, making it a potentially useful but insufficient diagnostic for LyC leakage.

Fig. 9 shows $f_{\rm esc,\,LOS}^{\rm LyC}$ versus $v_{\rm sep}$[4] for our mock observations as a function of $f_{\rm esc}^{\rm LyC}\xi_{\rm ion}$. In order to compare to observations, we overplot data from the LzLCS (Flury et al. 2022b) and Naidu et al. (2022) in cyan and green, respectively, as well as the relation from Izotov et al. (2018a) in blue. We find some agreement (with a large amount of scatter), along with the fact that the majority of strong leakers have $v_{\rm sep} < 250\,{\rm km\,s^{-1}}$. However, there is no strong trend present. Fig. 9 also indicates the existence of an envelope, suggesting that it is very unlikely to find galaxies with large $v_{\rm sep}$ and high $f_{\rm esc}^{\rm LyC}$. Furthermore, there is a perhaps unexpectedly large population of galaxies with low $v_{\rm sep}$ and low $f_{\rm esc}^{\rm LyC}$. This, along with the fact that our mock observations tend to have systematically lower peak separations is likely due to the fact that SPHINX^{20} galaxies are altogether less massive and less UV-bright than our comparative sample. This is discussed further in Appendix C, but suggests that selection effects may have a role to play.

4.4 Central escape fraction, $f_{\rm cen}$

Recently, Naidu et al. (2022) proposed a new Ly α indicator for LyC escape, $f_{\rm cen}$. This represents the ratio of Ly α flux within $\pm 100\,{\rm km\,s^{-1}}$ of line centre to the total Ly α flux.[5] This is crucial, as line-centre emission[6] (corresponding to large values of $f_{\rm cen}$) may be indicative of Ly α photons that escape at line centre with little scattering, thus tracing low-column density channels (Behrens, Dijkstra & Niemeyer 2014). These may also be optically thin to LyC photons (e.g. Harrington 1973; Neufeld 1991), thus suggesting that the two quantities should be well correlated.

As before, the *middle row* of Fig. 8 shows $f_{\rm cen}$ as a function of sSFR, mean stellar age and $\zeta_{\rm ISM}$. For clarity, we also include the cut suggested by Naidu et al. (2022) in cyan. We find no strong trend between $f_{\rm cen}$ and sSFR, beyond that empirically, lines-of-sight with $f_{\rm cen} \gtrsim 50$ per cent tend to have systematically higher sSFRs. Next, we find no strong relationship between $f_{\rm cen}$ and mean stellar age, apart from the fact that lines-of-sight with $f_{\rm cen} < 10$ per cent tend to have young stellar populations and lines-of-sight with $f_{\rm cen} \gtrsim 40$ per cent tend to have stellar populations in the correct age range (indicated by vertical blue lines). As a result, we see that observations with $f_{\rm cen} > 10$ per cent are significantly more likely to have LyC leakage, while observations with $f_{\rm cen} \gtrsim 40$ per cent are

[4]Ly α spectra with single peaks are assigned a $v_{\rm sep}$ of zero so that they are not omitted.
[5]Specifically, we follow Naidu et al. (2022) and use the excess flux within a window of $\pm 1000\,{\rm km\,s^{-1}}$ to capture the total flux.
[6]Due to the fact that each galaxy has a peculiar velocity, we assume that H α provides the true redshift and shift the spectra accordingly.

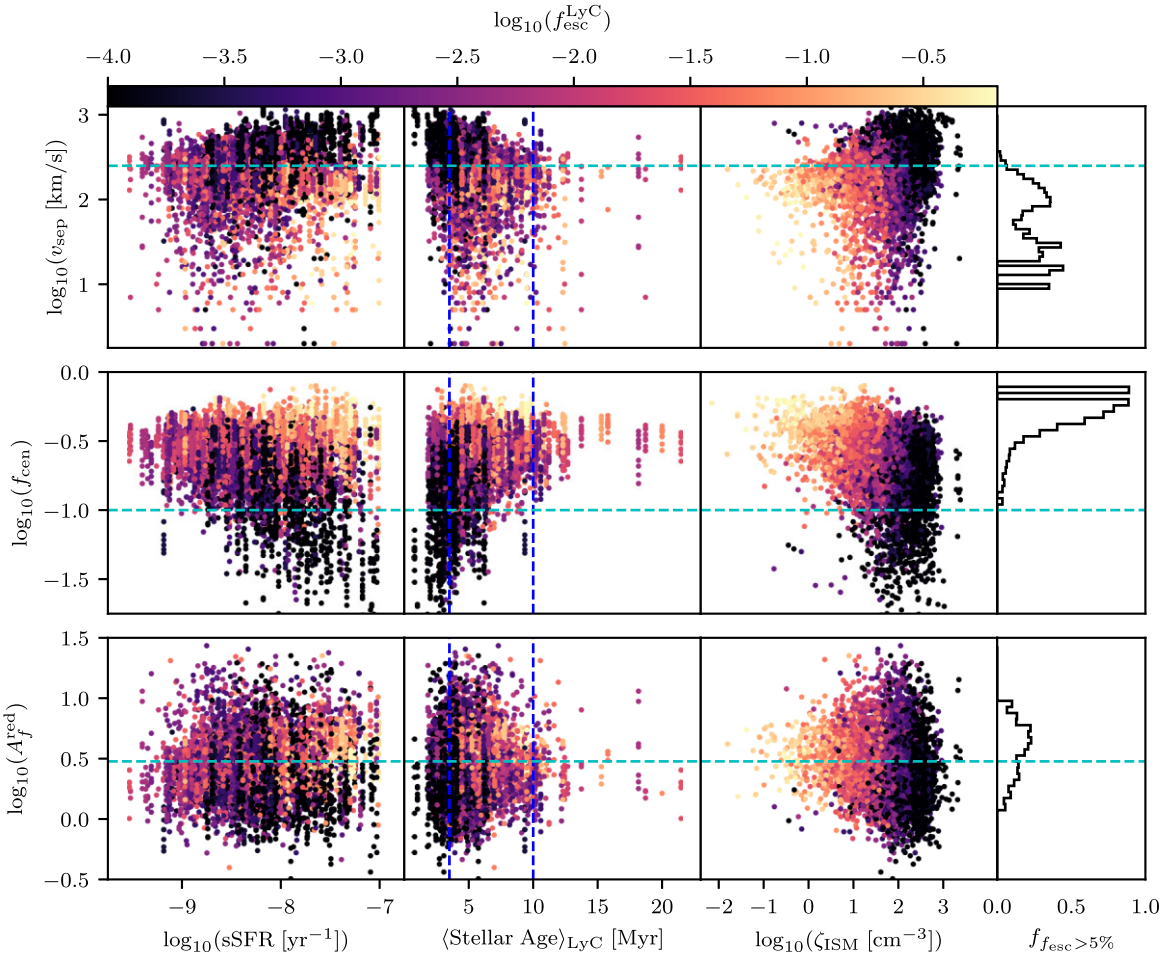

Figure 8. (Top) As in Fig. 5, but for the Lyα peak separation. Mock observations with $v_{\rm sep} > 250\,{\rm km s^{-1}}$ (cyan line) select for galaxies with young stellar populations and large $\zeta_{\rm ISM}$. Therefore, low peak separations are found in galaxies with higher LyC escape fractions. (Middle) As above, but for the central escape fraction. We find that large values of $f_{\rm cen}$[[compared to the 10 per cent cut suggested by Naidu et al. (2022) in cyan] select for galaxies with high sSFR, stellar ages in the range required and low $\zeta_{\rm ISM}$. Thus, for large values of $f_{\rm cen}$ it is a suitable diagnostic for the LyC escape fraction. (Bottom) As above, but for the red peak asymmetry. We find that $A_f^{\rm red}$ does not trace sSFR or the mean stellar age. However, galaxies with the lowest $\zeta_{\rm ISM}$ tend to be clustered around $A_f^{\rm red} \sim 3$ (shown in cyan). Thus, $A_f^{\rm red}$ is not by itself a good indicator for the presence of LyC escape, though it can be used to infer the method of escape (Kakiichi & Gronke 2021).

overwhelmingly likely to be strong LyC leakers. Finally, galaxies with larger $f_{\rm cen}$ tend to correspond to sightlines with less dust attenuation and neutral gas. Combining these facts, we find that the central escape fraction marginally satisfies all three criteria when $f_{\rm cen} \gtrsim 40$ per cent, making it a potentially good diagnostic for LyC leakage for this range of values, when also combined with other information. This is also seen in the histogram for $f_{\rm cen}$ (*middle-right*), where bins of larger $f_{\rm cen}$ tend to have a higher fraction of galaxies with $f_{\rm esc}^{\rm LyC} > 5$ per cent.

Fig. 10 shows $f_{\rm esc,\,LOS}^{\rm LyC}$ as a function of $f_{\rm cen}$ for our mock observations coloured by $f_{\rm esc}^{\rm LyC}\xi_{\rm ion}$. We include observational data from Naidu et al. (2022) in green for comparison. We find a large amount of scatter between the two quantities, however galaxies with larger $f_{\rm cen}$ tend to leak more LyC photons. Interestingly, we find in practice that there is a marginally stronger correlation between $f_{\rm cen}$ and $f_{\rm esc}^{\rm LyC}$ than that with $f_{\rm esc,\,LOS}^{\rm LyC}$. Furthermore, while the cut of $f_{\rm cen} > 10$ per cent discussed above captures the vast majority of strong leakers, we find that this is an insufficient diagnostic as the sample is highly impure (see discussion in Section 4.7). We note that the fact that our average $f_{\rm cen}$ values are skewed systematically larger than those reported by Naidu et al. (2022). This offset is also likely due to the fact that SPHINX^{20} galaxies are less massive and have smaller H I masses than the comparative observational sample. In fact, this offset is also consistent with the offset found in $v_{\rm sep}$, as discussed further in Appendix C. As a result, while $f_{\rm cen}$ has the potential to be a useful diagnostic for LyC leakage, we caution that there is likely a hidden H I mass dependence that needs to be accounted for.

4.5 Red peak asymmetry, $A_f^{\rm red}$

Recently, Kakiichi & Gronke (2021) suggested the red peak asymmetry as a diagnostic of LyC leakage. This asymmetry can be quantified as $A_f^{\rm red}$,

$$A_f^{\rm red} = \frac{\int_{v_{\rm red}}^{\infty} f_\lambda \, {\rm d}\lambda}{\int_{v_{\rm valley}}^{v_{\rm red}} f_\lambda \, {\rm d}\lambda}, \tag{3}$$

where $v_{\rm red}$ and $v_{\rm valley}$ are the velocity shifts of the red peak and central valley, respectively.

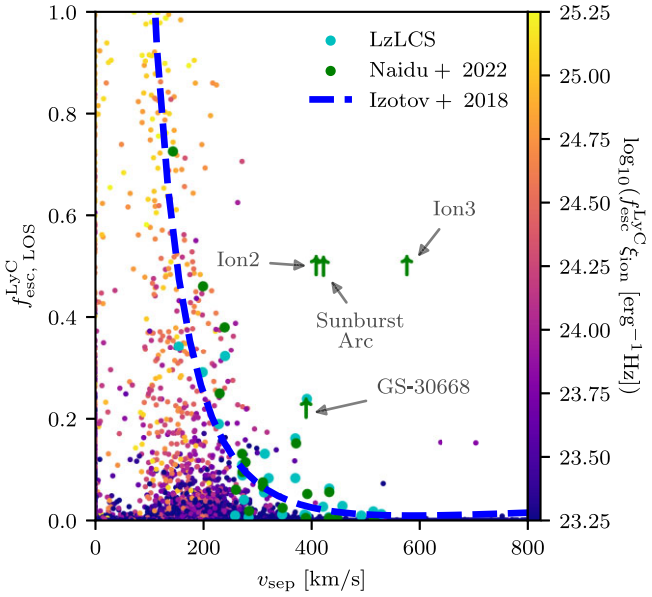

Figure 9. Line-of-sight LyC escape fraction as a function of Ly α peak separation (for all cases where there are two peaks in the Ly α spectrum) coloured by $f_{\rm esc}^{\rm LyC}\xi_{\rm ion}$. We include observational data from Flury et al. (2022b) and Naidu et al. (2022) in cyan and green, respectively, as well as the relation from Izotov et al. (2018a) as a dashed blue line. We find that mock observations with $v_{\rm sep} < 250\,{\rm km s^{-1}}$ are much more likely to be significant LyC leakers, but this is not a sufficient condition.

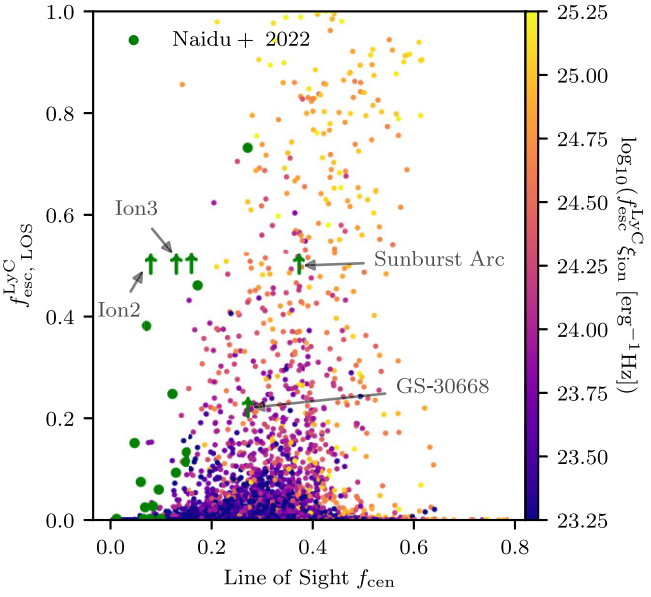

Figure 10. Line-of-sight LyC escape fraction as a function of central Ly α escape fraction, coloured by $f_{\rm esc}^{\rm LyC}\xi_{\rm ion}$. We include observational data from Naidu et al. (2022) in cyan. We find that in general, galaxies with larger $f_{\rm cen}$ have higher LyC escape fractions, as well as contributing more to reionization (by having a larger value of $f_{\rm esc}^{\rm LyC}\xi_{\rm ion}$). Particularly, we find that contamination from non-leaking galaxies drops off after $f_{\rm cen} \gtrsim 0.4$.

Using RHD simulations of giant molecular clouds with turbulence and radiative feedback, Kakiichi & Gronke (2021) showed that in a turbulent H II region, the flux of ionizing photons from the central source will produce a mix of density- and ionization-bounded channels, based on the low or high column density of neutral hydrogen, respectively. LyC photons traversing the ionization-bounded medium will then have a much larger probability of escape (i.e. contributing a larger $f_{\rm esc}^{\rm LyC}$) as compared to those passing through a density-bounded region. From the perspective of Ly α photons, these two types of channels represent different types of escape (e.g. Gronke et al. 2016, 2017). In the ionization-bounded (or high $N_{\rm HI}$) case, the gas remains optically thick far into the wings of the line. As a result, Ly α photons must scatter multiple times in order to diffuse sufficiently far in frequency space as to escape. In contrast, in the density-bounded (or low $N_{\rm HI}$) case, the gas is optically thick around line centre, becoming optically thin in the Lorentzian wings. Therefore, Ly α photons are able to escape after a single interaction only, having been frequency-shifted sufficiently far from the Doppler core (e.g. Dijkstra 2017). Here, Kakiichi & Gronke (2021) found that systems with either extreme case of leakage produced Ly α signals with low red peak asymmetry ($A_f^{\rm red} \lesssim 3$) with a mix of small and large $f_{\rm esc}^{\rm LyC}$. In contrast, configurations with a mixture of the two cases (i.e. with low column density 'holes') produced larger red peak asymmetries ($A_f^{\rm red} \gtrsim 3$) with a mixture of low to intermediate LyC escape fractions.

The *bottom* row of Fig. 8 gives $A_f^{\rm red}$ as a function of sSFR, mean stellar age, and $\zeta_{\rm ISM}$, coloured by $f_{\rm esc}^{\rm LyC}$. The turnover value of $A_f^{\rm red} = 3$ from Kakiichi & Gronke (2021) is highlighted in cyan. We find no significant dependence on sSFR or mean stellar age, but note that sightlines with lower $\zeta_{\rm ISM}$ tend to be within the region of $A_f \sim 2-5$. Furthermore, the ISM resolution in SPHINX^{20} is insufficient to capture the escape of photons through small channels created by turbulence. This explains the drop-off in $f_{\rm esc}^{\rm LyC}$ for $A_f^{\rm red} \gtrsim 6$ observed in Fig. 11. We note that while our results show some strong discrepancies with those of Kakiichi & Gronke (2021), one significant difference is the fact that SPHINX^{20} galaxies have a CGM. This extra stage of reprocessing of Ly α photons has been shown to modify Ly α signals, notably by increasing the asymmetry of red peaks (e.g. Blaizot et al. 2023), thus explaining the relative shift in our data-points in Fig. 11.

Fig. 11 shows $f_{\rm esc,\,LOS}^{\rm LyC}$ as a function of $A_f^{\rm red}$ for all of our mock observations, coloured by $f_{\rm esc}^{\rm LyC}\xi_{\rm ion}$. We also include the simulation data from Kakiichi & Gronke (2021) as well as observational data from Izotov et al. (2016, 2018a, b) in red and green, respectively. We find that our data agrees with the observational data, but that our strong LyC leakers tend to have larger $A_f^{\rm red}$ than those produced by Kakiichi & Gronke (2021). Furthermore, though we do find evidence for a characteristic transition between the two behaviours, this occurs at a larger red peak asymmetry (at ~ 6) than predicted by Kakiichi & Gronke (2021) (as indicated by the vertical line on Fig. 11). Here, it is important to note that not all H II regions in SPHINX^{20} galaxies are fully resolved. In these cases, $A_f^{\rm red}$ is likely to be reduced as Ly α emission is surrounded by the uniform medium of an underresolved cell.

4.6 Extended Ly α haloes

Extended Ly α haloes (LAHs) have now been measured across a large range of redshifts. For example, studies using the Lyman Alpha Reference Sample (LARS; Hayes et al. 2014; Östlin et al. 2014) of $z \lesssim 0.2$ galaxies, found a large number of LAHs which extended up to four times further than the effective radii of H α, far UV (Hayes et al. 2013; Rasekh et al. 2022) and UV (Yang et al. 2017a) emission. At intermediate redshifts ($z \sim 2.65$), narrow band imaging of Lyman break galaxies (LBGs) has been successful in finding Ly α profiles exceeding their UV counterparts by a factor of 5–10 (Steidel et al. 2011). Sensitive integral field spectrographs, e.g. MUSE, have also been leveraged to detect individual LAHs at intermediate redshifts

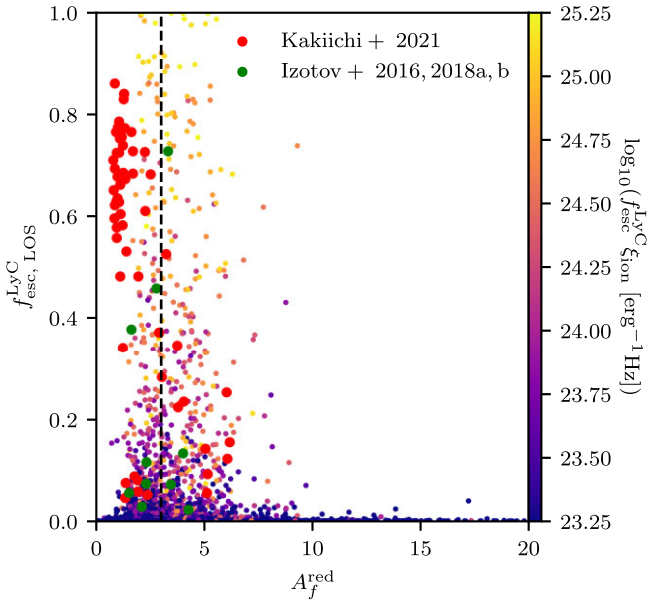

Figure 11. Line-of-sight LyC escape fraction as a function of Ly α red peak asymmetry, coloured by $f_{\rm esc}^{\rm LyC}\xi_{\rm ion}$. We include observational data from Izotov et al. (2016, 2018a, b) in green as well as data from simulations in red (Kakiichi & Gronke 2021). We find that galaxies with significant LyC leakage tend to cluster around $A_f^{\rm red} \sim 3$, but that such a sample is contaminated by non-leakers. The strongest LyC leakers have larger values of $A_f^{\rm red}$ as compared to those from simulated giant molecular clouds. This is due to effects introduced by later reprocessing of Ly α photons in the CGM (Blaizot et al. 2023).

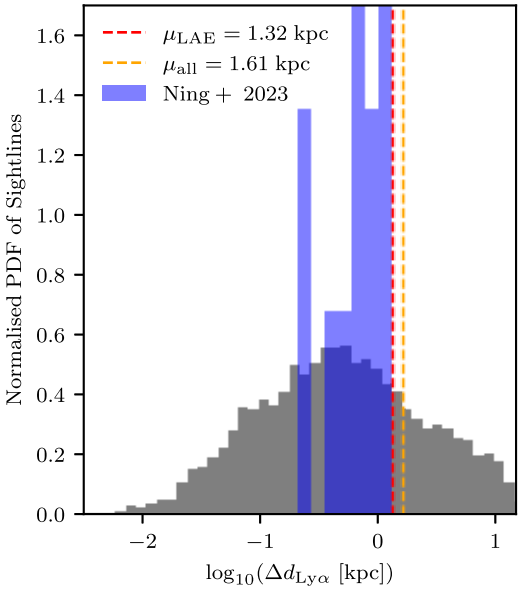

Figure 12. Normalized histogram of the physical offset between the centroids of brightest Ly α and F150W emission for our mock images. We include observational measurements from Ning et al. (2024) as a comparison. We find a mean offset of 1.32 kpc for LAEs and 1.61 kpc for all galaxies.

($2 \lesssim z \lesssim 6$, e.g. Wisotzki et al. 2016; Leclercq et al. 2017, 2020; Erb, Steidel & Chen 2018). *JWST* has detected extended Ly α emission at redshifts of $z \sim 7.47 - 7.75$ (Jung et al. 2023) and $z = 10.6$ (Bunker et al. 2023). Despite these observations, there is little evidence for evolution in the relative sizes of extended Ly α or H α (or UV) haloes as a function of redshift (e.g. Runnholm et al. 2023).

Due to the resonant nature of Ly α, the source of extended emission remains unclear. For example, Kim et al. (2020) used polarimetry to study extended Ly α emission from a Ly α nebula at $z = 2.656$ containing an obscured, embedded AGN. They found that escaping Ly α emission (sourced by AGN-photo-ionized gas) is scattered by the cloud at large radii back into the sightline. Furthermore, post-processed simulations of star-forming galaxies from the IllustrisTNG50 simulation were used to infer that the majority of photons observed in LAHs are re-scattered photons from star-forming regions (Byrohl et al. 2021), as opposed to photons emitted *in-situ* or by satellite galaxies. Mitchell et al. (2021) also found similar results, studying a cosmological radiation hydrodynamical zoom simulation of a single galaxy. Here, they found that these three emission components contributed equally at a radius of ~ 10 kpc. At very large radii however, these profiles flatten, with the majority of the contributions coming from nearby galaxies, haloes, and cooling radiation rather than diffuse emission. Median-stacked observations at $1.9 \lesssim z \lesssim 3.5$ from HETDEX (Lujan Niemeyer et al. 2022) as well as at $3 \lesssim z \lesssim 4$ from the MUSE Extremely Deep Field (Guo et al. 2023) seem to both confirm this prediction, finding similar surface brightness profiles.

The presence of an extended halo of neutral gas means that Ly α emission can be scattered, producing larger, more extended LAHs. Moreover, ionizing photons that escape the central source are also able to photoionize neutral hydrogen in the CGM, leading to *in-situ* emission of Ly α called fluorescence (Furlanetto et al. 2005; Nagamine, Choi & Yajima 2010). This process leads to more concentrated, centrally peaked LAHs (Mas-Ribas & Dijkstra 2016). Finally, Ly α emission can also be powered by gravitational cooling, caused by the accretion of neutral hydrogen onto the galaxy. In contrast, while H α is also emitted predominantly by recombination transitions in starburst-powered H II regions and can be produced by fluorescence, it is not a resonant transition. This leads H α haloes to generally be much smaller than their LAH counterparts. The UV continuum is even simpler, as it is only emitted in the ISM surrounding ionizing sources, making it a direct tracer of star formation. All of this, as well as the relative contributions of satellite haloes at larger impact parameters to Ly α emission (Mas-Ribas et al. 2017a) led Mas-Ribas et al. (2017b) to claim that extended Ly α, H α, and continuum emission can be used to infer the escape fraction of ionizing radiation from a central source into the CGM. We build on this, using the fact that other Ly α-based diagnostics (as discussed above) tend to work by indicating the existence of escape channels for ionizing radiation to suggest that the relative size of extended LAHs is inversely proportional to the global escape of LyC photons.

In order to test this with our simulation, we measure the surface brightness profiles using mock images taken along all 10 lines-of-sight of each galaxy in Ly α, H α, and UV (using the F150W JWST/NIRCam filter) emission. Because Ly α emission is not necessarily co-spatial with H α or UV flux, we choose to first re-centre all images onto the centroid of the brightest region in the smoothed F150W image, as segmented by PHOTUTILS (Bradley et al. 2023). Doing so, we find that a large number of our mock images have a significant offset between the brightest Ly α and UV emission ($\Delta d_{{\rm Ly}\alpha}$), in agreement with recent results from Ning et al. (2024), who find an average offset of $\Delta d_{{\rm Ly}\alpha} \sim 1$ kpc for a sample of 14 LAEs at $z \sim 6$ imaged using JWST/NIRCam. Fig. 12 shows a normalized histogram of these offsets for all galaxies in our mock sample, as well as those from Ning et al. (2024). We find an average offset between Ly α and F150W emission of $\langle \Delta d_{{\rm Ly}\alpha} \rangle = 1.32$ kpc for galaxies with $W_\lambda > 25$Å and 1.61 kpc for

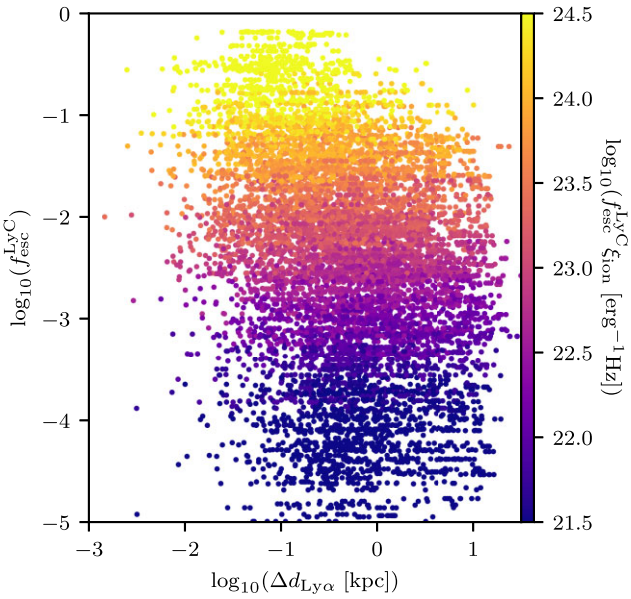

Figure 13. Global LyC escape fraction as a function of the line-of-sight physical offset between the centroids of brightest Ly α and F150W emission, coloured by $f_{\rm esc}^{\rm LyC}\xi_{\rm ion}$. No clear trend is visible, but galaxies which contribute the most to reionization tend to have offsets of $\Delta d_{{\rm Ly}\alpha} \lesssim 1$ kpc.

all galaxies. Fig. 13 shows the angle-averaged LyC escape fraction as a function of the physical offset between centroids of Ly α and F150W emission, coloured by $f_{\rm esc}^{\rm LyC}\xi_{\rm ion}$. We find that galaxies that contribute most to reionization tend to have $\Delta d_{{\rm Ly}\alpha} \lesssim 1$ kpc, but that there is no clear trend between $f_{\rm esc}^{\rm LyC}$ and $\Delta d_{{\rm Ly}\alpha}$. Following this re-centring process, we then calculate the surface brightness in circular bins, before normalizing by the average surface brightness within the central 10 kpc.

We have stacked surface brightness profiles based on the value of $f_{\rm esc}^{\rm LyC}$ for each galaxy. Fig. 14 shows the median normalized surface brightness profiles of Ly α (black), H α (red), and F150W (blue), with shaded regions indicating the 16th and 84th percentiles of the respective distributions. Ly α profiles have no central peak due to the offset discussed above. Galaxies with larger angle-averaged $f_{\rm esc}^{\rm LyC}$ tend to have less extended Ly α and F150W profiles, while also having more extended H α surface brightness profiles. However, changes in F150W and H α are much smaller than those for Ly α. While the physics behind Ly α profiles has been discussed, it is believed that in LyC leakers, H α profiles become more extended due to fluorescence exciting H α emission in the CGM (e.g. Mas-Ribas et al. 2017a), while the F150W profiles become increasingly steep due to the presence of nuclear starbursts. This indicates the possibility that for galaxies with similar bulk properties, morphological differences might be indicative of LyC leakage.

In order to quantify the relationship between these surface brightness profiles and $f_{\rm esc}^{\rm LyC}$, we introduce the integral ratio, $\mathcal{I}({\rm Ly}\alpha, X)$, given by:

$$\mathcal{I}({\rm Ly}\alpha, X) = \frac{\int_0^{5\,{\rm kpc}} \langle \Sigma_{{\rm Ly}\alpha} \rangle {\rm d}r}{\int_0^{5\,{\rm kpc}} \langle \Sigma_X \rangle {\rm d}r}, \quad (4)$$

where $\langle \Sigma_X \rangle$ is the normalized median surface brightness profile for an image in filter X. We note that the Ly α surface brightness profiles (particularly the *top row* of Fig. 14) near $r_{\rm vir}$ should be taken as lower limits as there are clear edge effects appearing due to the fact that we truncate the radiative transfer at the virial radius. Realistically, to better estimate these profiles out to such distances, it is necessary to simulate Ly α radiative transfer through a significantly larger volume, which is a computational limitation of our work. As a result, we use an upper limit of 5 kpc in equation (4) to avoid uncertainties due to these effects. We note that while changing this upper limit does affect the value $\mathcal{I}$, it does not strongly affect the trends discussed below.

Fig. 15 shows the integral ratio calculated using both H α and F150W with respect to Ly α for each bin in $f_{\rm esc}^{\rm LyC}$. The number of lines-of-sight in each bin is printed above the respective point, along with error bars representing values calculated using the 16th and 84th percentiles. Here, we see clearly that systems with higher $f_{\rm esc}^{\rm LyC}$ have smaller integral ratios, tending to a theoretical value of one (corresponding to the LAH having the same size as the halo imaged in the corresponding wavelength). We find that this trend appears for both H α and F150W, albeit slightly weaker for the latter and with more scatter.

In reality, we are unfortunately unable to use $f_{\rm esc}^{\rm LyC}$ to stack observed galaxies at high redshift. However, given the fact that the stacked integral ratio correlates with $f_{\rm esc}^{\rm LyC}$, it can be used as an independent method to investigate the quality of other diagnostics for LyC leakage. Fig. 16 shows the average integral ratios for bins in the UV stellar continuum slope, β (*top-left*), nebular UV attenuation, E(B − V) (*top-right*), $f_{\rm esc}^{{\rm Ly}\alpha}$ (*bottom-left*), and [OIII]λ5007/[OII]λλ3726, 3728, O_{32} (*bottom-right*) as above in Fig. 15. In order to guide the eye, we also include values for the integral ratio corresponding to a bin with $f_{\rm esc}^{\rm LyC} > 10$ per cent as horizontal lines. These act as effective thresholds. Here, we can use the fact that lower values of $\mathcal{I}({\rm Ly}\alpha, X)$ correspond to bins with greater $f_{\rm esc}^{\rm LyC}$ to test each diagnostic. In the case of β, we find that bluer UV slopes smoothly correlate with LyC leakers in agreement with previous works (e.g. Chisholm et al. 2022; Flury et al. 2022b; Choustikov et al. 2024). Next, galaxies with less dust attenuation also have lower integral ratios, corresponding to larger $f_{\rm esc}^{\rm LyC}$ (e.g. Saldana-Lopez et al. 2022; Choustikov et al. 2024). As discussed in Section 4.1, galaxies with significant Ly α leakage tend to also have significant LyC leakage. Next, we can compare integral ratios for each quantity with the threshold lines shown on each figure. In this way, we find that β is the best of these diagnostics, as the two stacks with the most negative UV slopes both fell below the threshold for strong LyC leakage. This is in agreement with the results of Choustikov et al. (2024). Finally, we find that this method suggests that galaxies with $0.2 \lesssim \log_{10}(O_{32}) \lesssim 0.6$ tend to have the largest $f_{\rm esc}^{\rm LyC}$, with larger O_{32} ratios corresponding to larger integral ratios and thus lower LyC escape fractions. We find that these trends also appear when using the F150W images, albeit much weaker.

These findings indicate that there is a wealth of information contained in extended photometric data that (while realistically difficult to obtain at high redshift) can help us to understand the efficacy of LyC diagnostics at low and intermediate redshifts.

4.7 Finding LyC leakers with Ly α

Having explored how well Ly α emission properties can be used to select for LyC leakage individually, we now continue to study how these properties can be used in conjunction with other observable quantities. In order to more fairly compare with observations, in this section we follow results from Section 4.2 and exclude all simulated galaxies from the analysis that have Ly α EWs < 25 Å, leaving a reduced sample of 63 per cent of our SPHINX[20] galaxies. We also note that in practice, the ability to measure many of these Ly α derivative quantities is contingent on the quality of continuum subtraction (particularly in the case of $f_{\rm cen}$ e.g. Naidu et al. 2022).

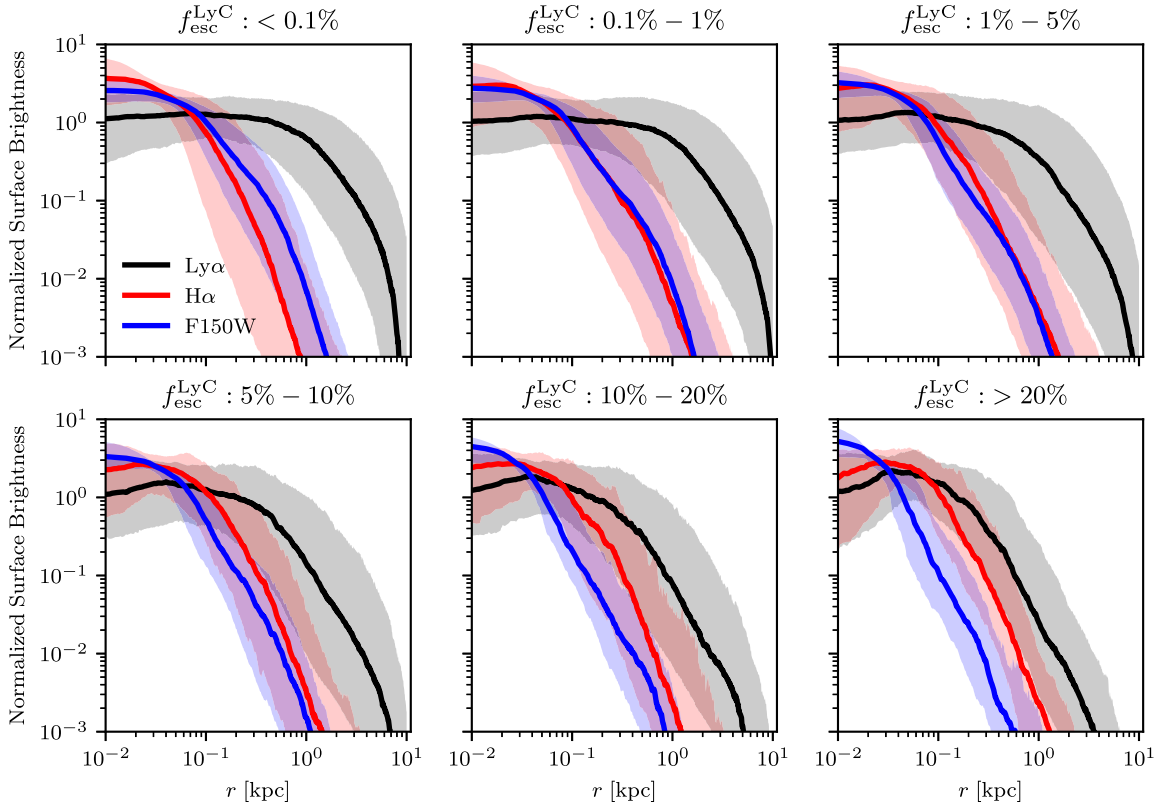

Figure 14. Median normalized surface brightness profiles for Ly α (black), H α (red), and F150W (blue) images for our mock observations binned by their LyC escape fractions. 16th and 84th percentiles are also shown to demonstrate the spread. We find that in galaxies with significant LyC leakage ($f_{\rm esc}^{\rm LyC} > 20$ per cent), the Ly α surface brightness profile is not significantly extended as compared to its H α and F150W profiles, in contrast to stacks of galaxies with low $f_{\rm esc}^{\rm LyC}$.

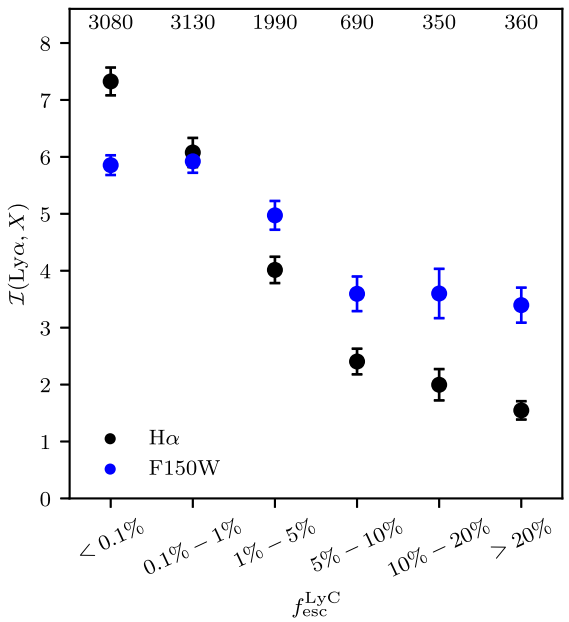

Figure 15. Integral ratios (as defined by equation 4) for median surface brightness profiles in Ly α, H α, and F150W images of galaxies, binned with respect to their LyC escape fraction. Error bars indicate the standard error of each measurement using the 16th and 84th percentiles as shown in Fig. 14. The number of galaxies in each bin is labelled above each point. We find that when stacked, bins of galaxies with higher LyC escape fractions tend to have smaller integral ratios.

As a result, beginning with such a selection in W_λ(Ly α) is a good way to ensure the quality of subsequent inferences.

So far, we have neglected the contribution of the nebular continuum to the UV slope, β. In doing so, it is expected that this inclusion tends to redden the UV slope. However, the relative contributions of stellar and nebular continua to β can, in principle, be disentangled with SED fitting. Nevertheless, it can be useful to derive a good combined diagnostic for LyC leakage that does not require the stellar and nebular contributions to the continuum to be disentangled. To this end, based on Fig. 16, it is clear that the best combination which suitably covers the three criteria for LyC leakage would be $f_{\rm esc}^{\rm Ly\,\alpha}$ and E(B − V). Fig. 17 shows line-of-sight Ly α escape fractions as a function of line-of-sight E(B − V), coloured by the global LyC escape fraction. We also include observational data from Hayes et al. (2010), Kornei et al. (2010), and Blanc et al. (2011) in cyan, lime, and green, respectively, as well as the best fit from Hayes et al. (2011) in cyan. We find very good agreement, as demonstrated by our running mean which is also shown in red. We find that mock observations with exceedingly low levels of UV attenuation (i.e. with E(B − V) $\lesssim$ 0.05) as well as with the highest Ly α escape fractions ($f_{\rm esc}^{\rm Ly\,\alpha} > 30$ per cent) also have the largest global LyC escape fractions. We have developed a selection criterion given by:

$$\log_{10}(f_{\rm esc}^{\rm Ly\,\alpha}) \geq 8.4\ E(B-V) + \gamma, \tag{5}$$

where γ is a free parameter that can be used to inform the stringency of the cut. More specifically, we find that lower values of γ tend to give reduced samples that are more complete with respect to LyC leakers, while also giving lower purities.

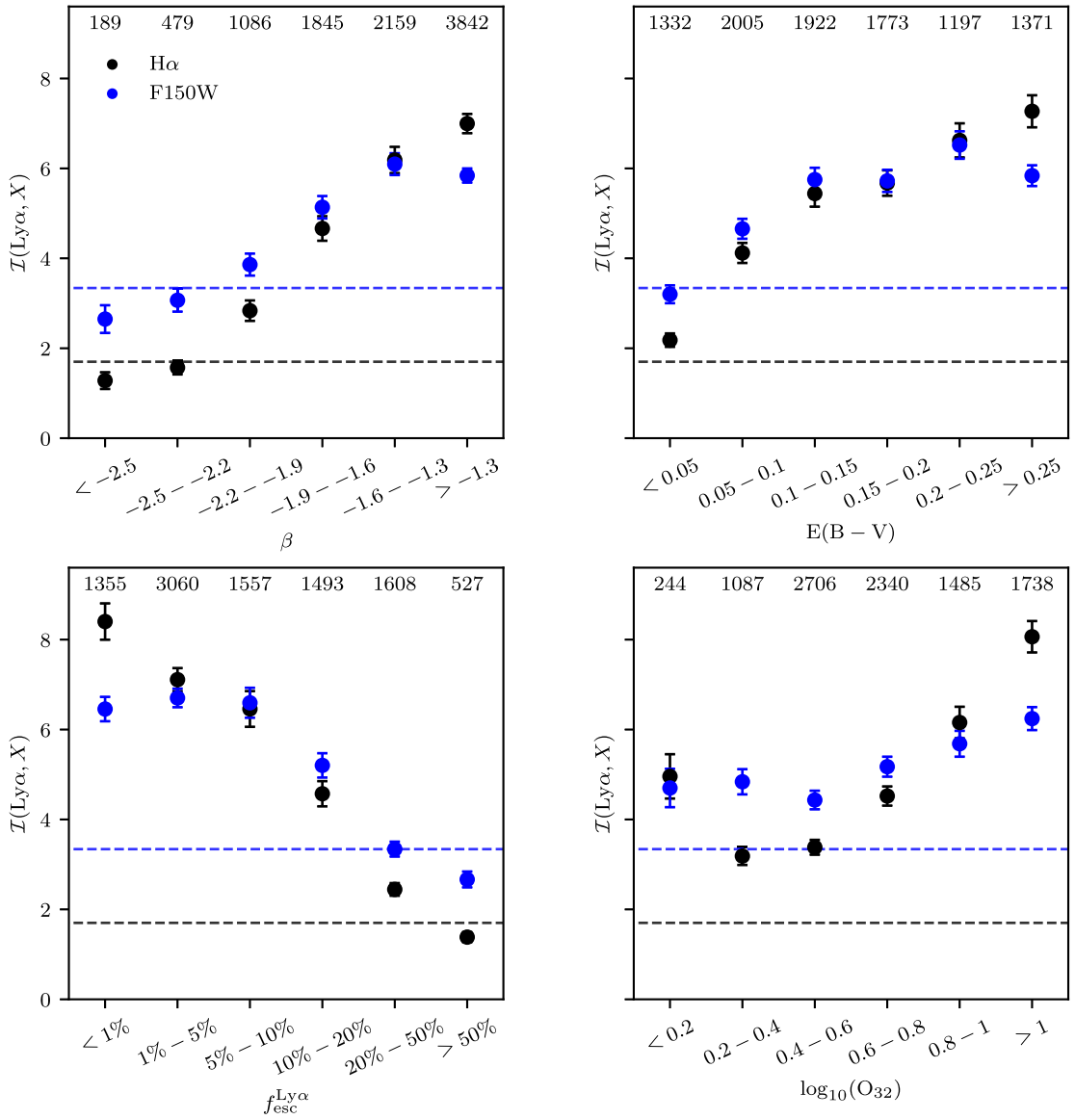

Figure 16. (Top-left) As in Fig. 15, but for the UV stellar continuum slope, β. We also indicate the integral ratio values corresponding to bins with $f_{\mathrm{esc}}^{\mathrm{LyC}} > 10$ per cent with horizontal lines. We find a strong correlation between the integral ratio and β, suggesting that stacks of galaxies with small integral ratios (and thus high LyC escape fractions) tend to have negative UV slopes, in agreement with observations (e.g. Chisholm et al. 2022; Flury et al. 2022b). (Top-right) As above, but for the UV nebular attenuation, E(B − V). We find that bins with smaller integral ratios tend to have less UV attenuation. This agrees with previous work suggesting that galaxies with less dust have higher LyC escape fractions (e.g. Saldana-Lopez et al. 2022). (*Bottom-left*) As above, but for the Ly α escape fraction. We find no strong correlation for $f_{\mathrm{esc}}^{\mathrm{Ly}\alpha} \lesssim 5$ per cent − 10 per cent, but with a very strong dependence for galaxies leaking more Ly α photons. As before, this confirms the fact that high $f_{\mathrm{esc}}^{\mathrm{Ly}\alpha}$ is an indicator for LyC leakage. (*Bottom-right*) As before, but for O_{32}. We find no strong correlation, albeit a weak indication that bins with $0.2 < O_{32} < 0.4$ should have the highest LyC escape fractions. This disagrees with the notion that larger O_{32} ratios directly correspond to higher $f_{\mathrm{esc}}^{\mathrm{LyC}}$ (e.g. Nakajima & Ouchi 2014; Izotov et al. 2018a), though the connection between O_{32} and LyC escape is by no means completely understood (c.f. Barrow et al. 2020; Katz et al. 2020b; Flury et al. 2022b).

This is exemplified in Table 1 where the percentage completeness and purity are given for $\gamma \in \{-2.4, -1.2, -0.6\}$ with respect to mild LyC leakers ($f_{\mathrm{esc}}^{\mathrm{LyC}} \geq 5$ per cent) as well as strong LyC leakers ($f_{\mathrm{esc}}^{\mathrm{LyC}} \geq 20$ per cent). These cuts are also shown as dashed lines in Fig. 17. Here, we find that using $\gamma = -2.4$ gives a sample that is fully complete with respect to both populations, sacrificing purity, being only 29 per cent and 8 per cent complete with respect to the mild and strong LyC leakers, respectively. In contrast, using a value of $\gamma = -0.6$ gives a reduced sample that is less complete (33 per cent and 68 per cent for mild and strong LyC leakers, respectively). However, these samples are significantly more pure with respect to LyC leakers, being 85 per cent and 45 per cent pure with mild and strong LyC leakers. We also note that galaxies selected by these three cuts are responsible for 76 per cent, 50 per cent, and 26 per cent, respectively of all ionizing radiation released into the IGM. As a result, this suggests that $f_{\mathrm{esc}}^{\mathrm{Ly}\alpha}$, β, and $E(B-V)$ can be used as a powerful combined diagnostic for galaxies with significant LyC leakage, as predicted by the framework proposed in Choustikov et al. (2024) and discussed above.

Table 1. Percentage completeness and purity of LyC leakers in samples as cut using equation (5). We define completeness as $N(\geq f_{\rm esc}^{\rm LyC} \mid {\rm cut})/N(\geq f_{\rm esc}^{\rm LyC})$, while purity is defined as $N(\geq f_{\rm esc}^{\rm LyC} \mid {\rm cut})/N({\rm all} \mid {\rm cut})$. We note that galaxies selected by these three cuts are responsible for 76 per cent, 50 per cent, and 26 per cent, respectively of all ionizing radiation released into the IGM.

	$f_{\rm esc}^{\rm LyC} \geq 5$ per cent		$f_{\rm esc}^{\rm LyC} \geq 20$ per cent	
γ	Per cent completeness	Per cent purity	Per cent completeness	Per cent purity
−2.4	99	29	100	8
−1.2	70	57	92	19
−0.6	33	85	68	45

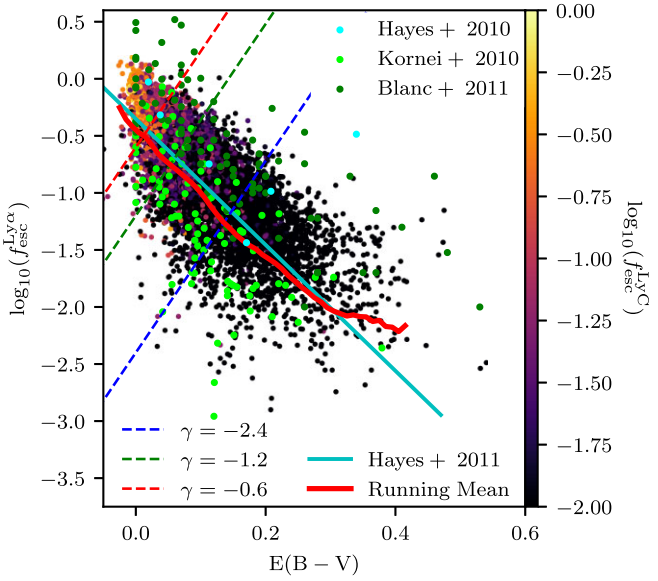

Figure 17. Line-of-sight Ly α escape fraction as a function of UV attenuation for all mock observations in our sample, coloured by the global LyC escape fraction of the given galaxy. Observational data from Hayes et al. (2010); Kornei et al. (2010); Blanc et al. (2011) are included in cyan, lime, and green respectively, as well as the best fit from Hayes et al. (2011) in cyan. Our running mean is shown in red. We include cuts given by equation (5) with $\alpha \in -2.4, -1.2, -0.6$ as dashed lines in red, green, and blue, respectively. The completeness and purity of each cut are given in Table 1. We find that strong leakers can be identified by combining these two quantities, with lines-of-sight with large $f_{\rm esc}^{\rm Ly\,\alpha}$ and low $E(B-V)$ being significantly more likely to have high LyC escape fractions.

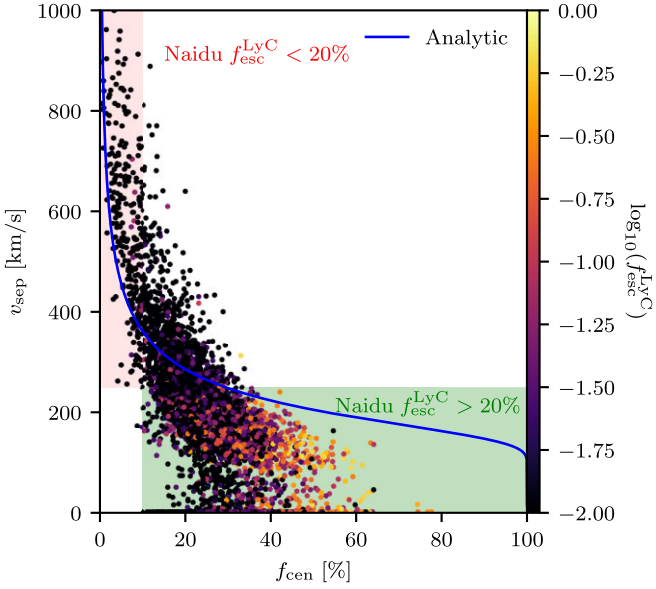

Figure 18. Ly α peak separation as a function of central escape fractions for all bright LAEs ($W_\lambda({\rm Ly}\,\alpha) > 25$Å) coloured by the global LyC escape fraction. The simple analytic model from Appendix C is also included. We include suggested selection criteria from Naidu et al. (2022), in particular suggesting that $f_{\rm cen} > 10$ per cent and $v_{\rm sep} < 250\,{\rm km\,s^{-1}}$ will select the majority of strong LyC leakers. While this is true (in the case of strong leakers producing a Ly α doublet), it is still a highly contaminated sample (94 per cent have $f_{\rm esc}^{\rm LyC} < 20$ per cent).

Naidu et al. (2022) suggested that galaxies with $f_{\rm cen} > 10$ per cent and $v_{\rm sep} < 250\,{\rm km\,s^{-1}}$ have $f_{\rm esc}^{\rm LyC} > 20$ per cent. In order to test this, Fig. 18 shows $v_{\rm sep}$ as a function of $f_{\rm cen}$ for all of our lines-of-sights, coloured by the global LyC escape fraction. Here, the green region indicates the space of galaxies predicted to have large LyC escape fractions, while the red region is reserved for non-leakers. We find that the green region captures 69 per cent of all galaxies with $f_{\rm esc}^{\rm LyC} > 20$ per cent in our sample, mainly due to the fact that the remaining strong leakers do not have multiple peaks.[7] However, this region of the observable space is also heavily contaminated by non-leakers, as 94 per cent of these lines-of-sight have $f_{\rm esc}^{\rm LyC} < 20$ per cent. Thus, we note that while this set of diagnostics certainly works to identify a sample containing the majority of strong LyC leakers (with an average LyC escape fraction of 4.0 per cent), we do not reproduce the purity of 80 per cent quoted by Naidu et al. (2022). Nevertheless, we include the caveat that there may be a hidden H I mass dependence underlying both of these quantities. To this end, we have also overplotted the analytic results for the simple analytical model derived in Appendix C, where we discuss this potential selection effect further.

[7]This number increases to 99 per cent if we include lines-of-sight producing single peaks with $f_{\rm cen} > 10$ per cent.

5 CAVEATS

The emergent emission spectrum for resonant transmission lines like Ly α are very sensitive to small scale structures and fluctuations in the density and velocity field of the emitting and attenuating medium. As all cosmological simulations, SPHINX[20] has a finite spatial resolution (~ 10 pc at $z = 6$). Therefore, while it is able to resolve the multiphase nature of the ISM, it cannot completely capture the small-scale dynamics and feedback processes that are inherent to the ISM or giant molecular clouds (e.g. Kakiichi & Gronke 2021; Kimm et al. 2022). While predominantly felt at small scales, these effects (particularly including turbulence, stellar winds, radiation pressure etc.) will change the scattering process of Ly α

and therefore are known to modify the emergent spectral profile in both low and high resolution simulations (e.g. Camps et al. 2021). Even so, a similar model to that used in this paper has recently been shown to reproduce the plethora of observed galaxy Ly α spectral profiles (Blaizot et al. 2023).

Due to the fact that the SPHINX^{20} simulation does not self-consistently follow the formation and evolution of dust, we use an effective phenomenological dust law, where the dust to metallicity ratio is held constant and dust predominantly traces the neutral gas in our simulation (Laursen et al. 2009). While it is an effective model that has been shown to reproduce observational trends (e.g. Katz et al. 2022, 2023a; Choustikov et al. 2024), it is important to note the effects that such a model can have, particularly due to the role that dust plays in absorbing Ly α photons. This differential distribution of dust will have two key effects. The first, is that it may absorb LyC photons before they have had a chance to ionize pockets of neutral hydrogen, thus modifying the spatial distribution of Ly α emitting gas cells. Secondly, changes to the spatial distribution of Ly α-absorbing dust will slightly modify the subsequent Ly α spectral profiles, particularly in terms of their overall asymmetry (e.g. Verhamme, Schaerer & Maselli 2006, see also Smith et al. 2022b). However, we do not consider this to be a major effect. Finally, the exact dust-to-gas mass relationship is debated, with some observations suggesting that it should follow a power law with metallicity (Rémy-Ruyer et al. 2014). However, studying these effects is beyond the scope of this paper.

Next, while we have included transmission effects due to Ly α photons travelling through the CGM (i.e. out to the virial radius), we have not included the full effect of the IGM. While the IGM can reduce the overall visibility of LAEs during the epoch of reionization (e.g. Jeeson-Daniel et al. 2012; Behrens & Niemeyer 2013; Schenker et al. 2014; Kusakabe et al. 2020; Garel et al. 2021, see also Appendix A), it can also modify the emergent Ly α spectral profile, particularly attenuating the blue peak (see fig. 12 of Smith et al. 2022a). However, given the fact that we have focused on studying galaxies at redshifts $z \leq 6$, we believe this to only be a minor effect (Garel et al. 2021). Moreover, our work is also applicable to galaxies contained within large ionized bubbles at higher redshifts (e.g. Saxena et al. 2023), as their Ly α profiles are modified only by transmission through the local ISM and CGM.

The context and caveats to the physics and emission line modelling behind this SPHINX^{20} data set have been explored extensively before. Therefore, we direct interested readers to consult Katz et al. (2023a) and Choustikov et al. (2024) for discussions about sub-grid modelling in SPHINX^{20}, as well as comparisons to other works utilizing simulations such as these.

6 SUMMARY AND CONCLUSIONS

We have post-processed a sample of 960 observable star-forming galaxies at $4.64 \lesssim z \lesssim 6$ from the SPHINX^{20} cosmological radiation hydrodynamical simulation with CLOUDY and RASCAS to produce a library of 9600 resonant-scatter and dust-attenuated Ly α spectra. We also use RASCAS to simulate the LyC escape fractions for all 9600 lines-of-sight. We combine these with global LyC escape fractions from the SPHINX Public Data Release v1 (Katz et al. 2023a) to carry out the first complete test of the viability of using properties of observed Ly α emission to infer LyC leakage from epoch of reionization galaxies in a cosmological simulation.

It is confirmed that Ly α properties of SPHINX^{20} galaxies are representative of observations of the high-redshift Universe made by *JWST*. We also found that the typical method of estimating the Ly α escape fraction produces overestimates (by as much as two orders of magnitude in extreme cases), particularly for dusty-sightlines where attenuation corrections are sometimes insufficient.

We have investigated the viability of using spectroscopic properties of Ly α emission as diagnostics to infer global LyC leakage. The framework for observational diagnostics of LyC leakage proposed by Choustikov et al. (2024) has also been used to explore the physical reasons behind why each diagnostic is successful (or not). This framework states that a good diagnostic for high LyC leakage should select for galaxies with high sSFRs, mean stellar population ages in the range $3.5 \leq \langle \tau_{\rm Stellar} \rangle / [{\rm Myr}] \leq 10$ and should contain a proxy for the density and neutral state of the galaxy's ISM.

Using this we find that the line-of-sight Ly α escape fraction, $f_{\rm esc}^{{\rm Ly}\,\alpha}$ is a good diagnostic for LyC leakage, due to the fact that while a weak indicator of sSFR, $f_{\rm esc}^{{\rm Ly}\,\alpha}$ can be a very good indicator for whether the mean age of a stellar population of a given galaxy is $\gtrapprox 3.5$ Myr and unsurprisingly traces the neutral, dusty phase of the ISM well. Next, increased Ly α equivalent widths, $W_\lambda({\rm Ly}\,\alpha)$ are a weak indicator of sSFR as well as the dust attenuation and neutral gas density of the ISM. However, large $W_\lambda({\rm Ly}\,\alpha)$ do not trace the stellar population age. As a result, by satisfying 2/3 criteria weakly, we find that $W_\lambda({\rm Ly}\,\alpha)$ is a necessary but insufficient diagnostic for LyC leakage. Next, large Ly α peak separations, $v_{\rm sep}$ were found to select for stellar populations too young to clear channels in their ISM, correlating with UV attenuation and neutral gas density. As a result, we find that strong LyC leakers tend to have $v_{\rm sep} < 250\,{\rm km s^{-1}}$. However, given the fact that $v_{\rm sep}$ does not correlate with sSFR, we find that this is a necessary but insufficient diagnostic for LyC leakage. The fraction of Ly α photons escaping near line centre, $f_{\rm cen}$ was found to correlate strongly with the density of the dusty ISM. Furthermore, we find that for $f_{\rm cen} \gtrsim 30$ per cent, $f_{\rm cen}$ correlates with sSFR and for $f_{\rm cen} \gtrsim 40$ per cent selects galaxies with mean stellar population ages in the correct range for effective LyC leakage. As a result, we find that $f_{\rm cen}$ has the possibility of being a very useful diagnostic for LyC leakage, albeit with the caveat that the $f_{\rm cen} - f_{\rm esc}^{\rm LyC}$ relationship is far from trivial as there is likely a hidden galaxy mass dependence. Finally, while the asymmetry of the red peak, $A_f^{\rm red}$ has been explored as a useful tool for investigations of the exact method of LyC leakage on small scales, we find that it does not correlate with sSFR or the mean stellar population age. Interestingly, we find that the strongest leakers tend to be clustered around $A_f^{\rm red} \sim 3$ (with a slight bias towards larger values) due to the fact that such lines-of-sight tended to have less dense or dusty ISMs. However, given the fact that $A_f^{\rm red}$ only marginally informs us of 1/3 criteria, we conclude that it is an unsuitable indicator for LyC leakage by itself.

Building on the work of Mas-Ribas et al. (2017b), we have used cosmological simulations to investigate the connection between extended Ly α, H α, UV continuum (F150W) emission, and LyC escape. Studying re-centred and stacked mock images of our galaxies at these wavelengths, we find that strong LyC leakers tend to have contracted Ly α and UV haloes with similar surface brightness profiles to their H α haloes (which in contrast are slightly more extended). In contrast, stacked samples with significantly extended Ly α haloes tend to have low LyC escape fractions. This follows from the fact that the majority of extended Ly α emission is believed to be re-scattered light from the central star-forming regions as well as fluorescence, implying the significant presence of neutral hydrogen in the CGM. Using the integral ratio as defined in equation (4), we have also explored how stacked Ly α profiles compare to their H α and F150W counterparts when stacked in bins according to other potential $f_{\rm esc}^{\rm LyC}$ diagnostics, including the UV slope, β, UV attenuation, $E(B-V)$ and O_{32}. We find that this method

independently verifies previous results that β and $E(B-V)$ are good diagnostics for $f_{\rm esc}^{\rm LyC}$, while O_{32} is a necessary but insufficient diagnostic (Choustikov et al. 2024). This exercise confirms the fact that Ly α surface brightness morphology can be used to understand the leakage of ionizing radiation from the centres of galaxies.

Finally, the possibility of using properties of Ly α emission to infer large LyC escape fractions was also explored. Given the discussions above, it was found that $f_{\rm esc}^{{\rm Ly}\,\alpha}$ is the most promising feature, despite the fact that it is often overpredicted in observational studies (see Section 3). A combined criterion which should be unaffected by pollution from nebular continuum emission is proposed by combining $f_{\rm esc}^{{\rm Ly}\,\alpha}$ with $E(B-V)$. Here, the cut given in equation (5) is found to provide a flexible method to select LyC leaker enriched samples with desired completeness and purity. In general, it is found that a combination of $f_{\rm esc}^{{\rm Ly}\,\alpha}$, β, and $E(B-V)$ are best at selecting for LyC leaking galaxies.

We have explored the feasibility of various Ly α-based indirect diagnostics for galaxies with high LyC escape fractions. By using a rich data set of mock Ly α observations of simulated high-redshift galaxies with multiphase ISMs, as well as a physically motivated theoretical framework for the physics driving LyC leakage, we have found that properties of Ly α spectral and surface brightness profiles can indeed be used as reliable tracers for LyC leakage. This is in agreement with recent observational studies of the local and high-redshift Universe. We recognize that these results depend on the resolution limits of SPHINX[20] as well as the sub-grid physics used, despite the fact that it is a state-of-the-art simulation of galaxy formation during the epoch of reionization. Nevertheless, this work highlights the potential of *JWST* data to find and understand the sources of cosmological reionization, further deepening our understanding of the cosmic dawn.

ACKNOWLEDGEMENTS

NC thanks Alex J. Cameron and Sophia Flury for insightful discussions. NC and HK also thank Jonathan Patterson for helpful support on Glamdring throughout the project.

NC acknowledges support from the Science and Technology Facilities Council (STFC) for a Ph.D. studentship. HK acknowledges support from the Beecroft Fellowship. TK is supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2020R1C1C1007079 and 2022R1A6A1A03053472).

This work used the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). The equipment was funded by BEIS capital funding via STFC capital grants ST/P002293/1, ST/R002371/1, and ST/S002502/1, Durham University and STFC operations grant ST/R000832/1. DiRAC is part of the National e-Infrastructure.This work was performed using the DiRAC Data Intensive service at Leicester, operated by the University of Leicester IT Services, which forms part of the STFC DiRAC HPC Facility (www.dirac.ac.uk). The equipment was funded by BEIS capital funding via STFC capital grants ST/K000373/1 and ST/R002363/1 and STFC DiRAC Operations grant ST/R001014/1. DiRAC is part of the National e-Infrastructure.

Computing time for the SPHINX project was provided by the Partnership for Advanced Computing in Europe (PRACE) as part of the 'First luminous objects and reionization with SPHINX (cont.)' (2016153539, 2018184362, 2019215124) project. We thank Philipp Otte and Filipe Guimaraes for helpful support throughout the project and for the extra storage they provided us. We also thank GENCI for providing additional computing resources under GENCI grant A0070410560. Resources for preparations, tests, and storage were also provided by the Common Computing Facility (CCF) of the LABEX Lyon Institute of Origins (ANR-10-LABX-0066) and PSMN (Pôle Scientifique de Modélisation Numérique) at ENS de Lyon.

Author contributions: The main roles of the authors were, using the CRediT (Contribution Roles Taxonomy) system[8]:

NC: Conceptualization; formal analysis; writing – original draft; methodology; visualization. **HK**: Conceptualization; formal analysis; writing – original draft; methodology. **AS**: Conceptualization; writing – review and editing. **TG**: Software; writing – review and editing. **JD**: Resources; supervision; writing – review and editing. **AS**: Resources; supervision; writing – review and editing. **TK**: Writing – review and editing. **JB**: Writing – review and editing. **JR**: Writing – review and editing.

DATA AVAILABILITY

The SPHINX[20] data used in this article is available as part of the SPHINX Public Data Release v1 (SPDRv1; Katz et al. 2023a).

REFERENCES

Barrow K. S. S., Robertson B. E., Ellis R. S., Nakajima K., Saxena A., Stark D. P., Tang M., 2020, ApJ, 902, L39
Becker G. D., D'Aloisio A., Christenson H. M., Zhu Y., Worseck G., Bolton J. S., 2021, MNRAS, 508, 1853
Behrens C., Braun H., 2014, A&A, 572, A74
Behrens C., Niemeyer J., 2013, A&A, 556, A5
Behrens C., Dijkstra M., Niemeyer J. C., 2014, A&A, 563, A77
Bezanson R. et al., 2022, preprint (arXiv:2212.04026)
Bird S., Ni Y., Di Matteo T., Croft R., Feng Y., Chen N., 2022, MNRAS, 512, 3703
Blaizot J. et al., 2023, MNRAS, 523, 3749
Blanc G. A. et al., 2011, ApJ, 736, 31
Borrow J., Kannan R., Garaldi E., Smith A., Vogelsberger M., Pakmor R., Springel V., Hernquist L., 2023, MNRAS, 525, 5932
Bosman S. E. I. et al., 2022, MNRAS, 514, 55
Bouwens R. J. et al., 2015, ApJ, 803, 34
Bowler R. A. A., Jarvis M. J., Dunlop J. S., McLure R. J., McLeod D. J., Adams N. J., Milvang-Jensen B., McCracken H. J., 2020, MNRAS, 493, 2059
Bradley L. et al., 2023, astropy/photutils: 1.9.0, Zenodo
Brinchmann J., 2023, MNRAS, 525, 2087
Bunker A. J. et al., 2023, A&A, 677, A88
Byrohl C. et al., 2021, MNRAS, 506, 5129
Cameron A. J., Katz H., Rey M. P., 2023, MNRAS, 522, L89
Camps P., Behrens C., Baes M., Kapoor A. U., Grand R., 2021, ApJ, 916, 39
Chardin J., Puchwein E., Haehnelt M. G., 2017, MNRAS, 465, 3429
Charlot S., Fall S. M., 2000, ApJ, 539, 718
Chisholm J. et al., 2018, A&A, 616, A30
Chisholm J. et al., 2022, MNRAS, 517, 5104
Choustikov N. et al., 2024, MNRAS, 529, 3751
Davies F. B. et al., 2018, ApJ, 864, 142
Dayal P. et al., 2020, MNRAS, 495, 3065
Dijkstra M., 2017, preprint (arXiv:1704.03416)
Dijkstra M., Haiman Z., Spaans M., 2006, ApJ, 649, 14
Dijkstra M., Gronke M., Venkatesan A., 2016, ApJ, 828, 71
Donnan C. T. et al., 2023, MNRAS, 518, 6011

[8]https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/credit.html

Dopita M. A., Krauss L. M., Sutherland R. S., Kobayashi C., Lineweaver C. H., 2011, Ap&SS, 335, 345
Ďurovčíková D., Katz H., Bosman S. E. I., Davies F. B., Devriendt J., Slyz A., 2020, MNRAS, 493, 4256
Efstathiou G., 1992, MNRAS, 256, 43P
Erb D. K., Steidel C. C., Chen Y., 2018, ApJ, 862, L10
Fan X., Carilli C. L., Keating B., 2006a, ARA&A, 44, 415
Fan X. et al., 2006b, AJ, 132, 117
Ferland G. J. et al., 2017, Rev. Mex. Astron. Astrofis., 53, 385
Fletcher T. J., Tang M., Robertson B. E., Nakajima K., Ellis R. S., Stark D. P., Inoue A., 2019, ApJ, 878, 87
Flury S. R. et al., 2022a, ApJS, 260, 1
Flury S. R. et al., 2022b, ApJ, 930, 126
Furlanetto S. R., Schaye J., Springel V., Hernquist L., 2005, ApJ, 622, 7
Garel T., Blaizot J., Rosdahl J., Michel-Dansac L., Haehnelt M. G., Katz H., Kimm T., Verhamme A., 2021, MNRAS, 504, 1902
Gazagnes S., Chisholm J., Schaerer D., Verhamme A., Rigby J. R., Bayliss M., 2018, A&A, 616, A29
Gazagnes S., Chisholm J., Schaerer D., Verhamme A., Izotov Y., 2020, A&A, 639, A85
Giovinazzo E., Trebitsch M., Mauerhofer V., Dayal P., Oesch P. A., 2024, preprint (arXiv:2402.17635)
Gnedin N. Y., Kaurov A. A., 2014, ApJ, 793, 30
Gnedin N. Y., Kravtsov A. V., Chen H.-W., 2008, ApJ, 672, 765
Gordon K. D., Clayton G. C., Misselt K. A., Landolt A. U., Wolff M. J., 2003, ApJ, 594, 279
Greig B., Mesinger A., Bañados E., 2019, MNRAS, 484, 5094
Gronke M., Bird S., 2017, ApJ, 835, 207
Gronke M., Dijkstra M., McCourt M., Oh S. P., 2016, ApJ, 833, L26
Gronke M., Dijkstra M., McCourt M., Oh S. P., 2017, A&A, 607, A71
Gronke M. et al., 2021, MNRAS, 508, 3697
Guo Y. et al., 2023, preprint (arXiv:2309.05513)
Haiman Z., Knox L., 1999, in de Oliveira-Costa A., Tegmark M., eds, ASP Conf. Ser. Vol. 181, Microwave Foregrounds. Astron. Soc. Pac., San Francisco, p. 227
Hao J.-M., Yuan Y.-F., Wang L., 2015, MNRAS, 451, 1875
Harikane Y. et al., 2022, ApJS, 259, 20
Harrington J. P., 1973, MNRAS, 162, 43
Hayes M., Östlin G., Mas-Hesse J. M., Kunth D., Leitherer C., Petrosian A., 2005, A&A, 438, 71
Hayes M. et al., 2010, Nature, 464, 562
Hayes M., Schaerer D., Östlin G., Mas-Hesse J. M., Atek H., Kunth D., 2011, ApJ, 730, 8
Hayes M. et al., 2013, ApJ, 765, L27
Hayes M. et al., 2014, ApJ, 782, 6
Heinrich C., Hu W., 2021, Phys. Rev. D, 104, 063505
Henry A., Scarlata C., Martin C. L., Erb D., 2015, ApJ, 809, 19
Inoue A. K., Shimizu I., Iwata I., Tanaka M., 2014, MNRAS, 442, 1805
Izotov Y. I., Schaerer D., Thuan T. X., Worseck G., Guseva N. G., Orlitová I., Verhamme A., 2016, MNRAS, 461, 3683
Izotov Y. I., Schaerer D., Worseck G., Guseva N. G., Thuan T. X., Verhamme A., Orlitová I., Fricke K. J., 2018a, MNRAS, 474, 4514
Izotov Y. I., Worseck G., Schaerer D., Guseva N. G., Thuan T. X., Fricke Verhamme A., Orlitová I., 2018b, MNRAS, 478, 4851
Izotov Y. I., Schaerer D., Worseck G., Verhamme A., Guseva N. G., Thuan T. X., Orlitová I., Fricke K. J., 2020, MNRAS, 491, 468
Izotov Y. I., Worseck G., Schaerer D., Guseva N. G., Chisholm J., Thuan T. X., Fricke K. J., Verhamme A., 2021, MNRAS, 503, 1734
Izotov Y. I., Chisholm J., Worseck G., Guseva N. G., Schaerer D., Prochaska J. X., 2022, MNRAS, 515, 2864
Izotov Y. I., Thuan T. X., Guseva N. G., Schaerer D., Worseck G., Verhamme A., 2024, MNRAS, 527, 281
Jaskot A. E., Oey M. S., 2014, ApJ, 791, L19
Jeeson-Daniel A., Ciardi B., Maio U., Pierleoni M., Dijkstra M., Maselli A., 2012, MNRAS, 424, 2193
Jones G. C. et al., 2023, A&A, 683, A238
Jung I. et al., 2023, ApJ, 967, 73
Kakiichi K., Gronke M., 2021, ApJ, 908, 30
Kakiichi K. et al., 2018, MNRAS, 479, 43
Katz H. et al., 2020a, MNRAS, 494, 2200
Katz H. et al., 2020b, MNRAS, 498, 164
Katz H. et al., 2022, MNRAS, 515, 4265
Katz H. et al., 2023a, Open J. Astrophys., 6, 44
Katz H. et al., 2023b, MNRAS, 518, 592
Keating L. C., Weinberger L. H., Kulkarni G., Haehnelt M. G., Chardin J., Aubert D., 2020, MNRAS, 491, 1736
Kim E., Yang Y., Zabludoff A., Smith P., Jannuzi B., Lee M. G., Hwang N., Park B.-G., 2020, ApJ, 894, 33
Kimm T., Cen R., 2014, ApJ, 788, 121
Kimm T., Blaizot J., Garel T., Michel-Dansac L., Katz H., Rosdahl J., Verhamme A., Haehnelt M., 2019, MNRAS, 486, 2215
Kimm T., Bieri R., Geen S., Rosdahl J., Blaizot J., Michel-Dansac L., Garel T., 2022, ApJS, 259, 21
Kornei K. A., Shapley A. E., Erb D. K., Steidel C. C., Reddy N. A., Pettini M., Bogosavljević M., 2010, ApJ, 711, 693
Kulkarni G., Choudhury T. R., Puchwein E., Haehnelt M. G., 2017, MNRAS, 469, 4283
Kulkarni G., Keating L. C., Haehnelt M. G., Bosman S. E. I., Puchwein E., Chardin J., Aubert D., 2019a, MNRAS, 485, L24
Kulkarni G., Worseck G., Hennawi J. F., 2019b, MNRAS, 488, 1035
Kusakabe H. et al., 2020, A&A, 638, A12
Laporte N., Meyer R. A., Ellis R. S., Robertson B. E., Chisholm J., Roberts-Borsani G. W., 2021, MNRAS, 505, 3336
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, A&A, 627, A84
Leclercq F. et al., 2017, A&A, 608, A8
Leclercq F. et al., 2020, A&A, 635, A82
Leitherer C. et al., 1999, ApJS, 123, 3
Leitherer C., Hernandez S., Lee J. C., Oey M. S., 2016, ApJ, 823, 64
Lujan Niemeyer M. et al., 2022, ApJ, 929, 90
Madau P., Haardt F., 2015, ApJ, 813, L8
Maji M. et al., 2022, A&A, 663, A66
Martin-Alvarez S., Sijacki D., Haehnelt M. G., Farcy M., Dubois Y., Belokurov V., Rosdahl J., Lopez-Rodriguez E., 2023, MNRAS, 525, 3806
Mas-Ribas L., Dijkstra M., 2016, ApJ, 822, 84
Mas-Ribas L., Dijkstra M., Hennawi J. F., Trenti M., Momose R., Ouchi M., 2017a, ApJ, 841, 19
Mas-Ribas L., Hennawi J. F., Dijkstra M., Davies F. B., Stern J., Rix H.-W., 2017b, ApJ, 846, 11
Maseda M. V. et al., 2020, MNRAS, 493, 5120
Mason C. A., Treu T., Dijkstra M., Mesinger A., Trenti M., Pentericci L., de Barros S., Vanzella E., 2018, ApJ, 856, 2
McQuinn M., Lidz A., Zahn O., Dutta S., Hernquist L., Zaldarriaga M., 2007, MNRAS, 377, 1043
Michel-Dansac L., Blaizot J., Garel T., Verhamme A., Kimm T., Trebitsch M., 2020, A&A, 635, A154
Mitchell P. D., Blaizot J., Cadiou C., Dubois Y., Garel T., Rosdahl J., 2021, MNRAS, 501, 5757
Nagamine K., Choi J.-H., Yajima H., 2010, ApJ, 725, L219
Naidu R. P., Tacchella S., Mason C. A., Bose S., Oesch P. A., Conroy C., 2020, ApJ, 892, 109
Naidu R. P. et al., 2022, MNRAS, 510, 4582
Nakajima K., Ouchi M., 2014, MNRAS, 442, 900
Neufeld D. A., 1991, ApJ, 370, L85
Ning Y. et al., 2024, ApJ, 963, L38
Ocvirk P. et al., 2020, MNRAS, 496, 4087
Orlitová I., Verhamme A., Henry A., Scarlata C., Jaskot A., Oey M. S., Schaerer D., 2018, A&A, 616, A60
Osterbrock D. E., 1989, Astrophysics of Gaseous Nebulae and Active Galactic Nuclei. University Science Books, Herndon, VA
Östlin G. et al., 2014, ApJ, 797, 11
Paardekooper J.-P., Khochfar S., Dalla Vecchia C., 2015, MNRAS, 451, 2544
Pahl A. J., Shapley A., Steidel C. C., Chen Y., Reddy N. A., 2021, MNRAS, 505, 2447
Parkash V., Brown M. J. I., Jarrett T. H., Bonne N. J., 2018, ApJ, 864, 40

Pentericci L. et al., 2011, ApJ, 743, 132
Planck Collaboration VI, 2020, A&A, 641, A6
Rasekh A. et al., 2022, A&A, 662, A64
Reddy N. A., Steidel C. C., Pettini M., Bogosavljević M., Shapley A. E., 2016, ApJ, 828, 108
Reddy N. A., Topping M. W., Sanders R. L., Shapley A. E., Brammer G., 2023, ApJ, 948, 83
Rémy-Ruyer A. et al., 2014, A&A, 563, A31
Ricotti M., Ostriker J. P., 2004, MNRAS, 352, 547
Robertson B. E., Ellis R. S., Furlanetto S. R., Dunlop J. S., 2015, ApJ, 802, L19
Rosdahl J. et al., 2018, MNRAS, 479, 994
Rosdahl J. et al., 2022, MNRAS, 515, 2386
Roy N. et al., 2023, ApJ, 952, L14
Rujopakarn W. et al., 2023, ApJ, 948, L8
Runnholm A. et al., 2023, MNRAS, 522, 4275
Saldana-Lopez A. et al., 2022, A&A, 663, A59
Saldana-Lopez A. et al., 2023, MNRAS, 522, 6295
Saxena A. et al., 2023, A&A, 678, A68
Saxena A. et al., 2024, A&A, 684, A84
Scarlata C. et al., 2009, ApJ, 704, L98
Schaerer D., Izotov Y. I., Verhamme A., Orlitová I., Thuan T. X., Worseck G., Guseva N. G., 2016, A&A, 591, L8
Schaerer D., Marques-Chaves R., Barrufet L., Oesch P., Izotov Y. I., Naidu R., Guseva N. G., Brammer G., 2022, A&A, 665, L4
Schenker M. A., Ellis R. S., Konidaris N. P., Stark D. P., 2014, ApJ, 795, 20
Shapiro P. R., Iliev I. T., Raga A. C., 2004, MNRAS, 348, 753
Smith A., Kannan R., Garaldi E., Vogelsberger M., Pakmor R., Springel V., Hernquist L., 2022a, MNRAS, 512, 3243
Smith A. et al., 2022b, MNRAS, 517, 1
Sobral D., Matthee J., 2019, A&A, 623, A157
Sobral D., Santos S., Matthee J., Paulino-Afonso A., Ribeiro B., Calhau J., Khostovan A. A., 2018, MNRAS, 476, 4725
Stanway E. R., Eldridge J. J., 2018, MNRAS, 479, 75
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, ApJ, 869, 123
Torres-Albà N., Bosch-Ramon V., Iwasawa K., 2020, A&A, 635, A57
Trebitsch M., Verhamme A., Blaizot J., Rosdahl J., 2016, A&A, 593, A122
Trebitsch M., Blaizot J., Rosdahl J., Devriendt J., Slyz A., 2017, MNRAS, 470, 224
Trebitsch M. et al., 2021, A&A, 653, A154
Varadaraj R. G., Bowler R. A. A., Jarvis M. J., Adams N. J., Häußler B., 2023, MNRAS, 524, 4586
Verhamme A., Schaerer D., Maselli A., 2006, A&A, 460, 397
Verhamme A., Dubois Y., Blaizot J., Garel T., Bacon R., Devriendt J., Guiderdoni B., Slyz A., 2012, A&A, 546, A111
Verhamme A., Orlitová I., Schaerer D., Hayes M., 2015, A&A, 578, A7
Verhamme A., Orlitová I., Schaerer D., Izotov Y., Worseck G., Thuan T. X., Guseva N., 2017, A&A, 597, A13
Weingartner J. C., Draine B. T., 2001, ApJS, 134, 263
Wise J. H., Cen R., 2009, ApJ, 693, 984
Wisotzki L. et al., 2016, A&A, 587, A98
Witten C. E. C., Laporte N., Katz H., 2023, ApJ, 944, 61
Worseck G. et al., 2014, MNRAS, 445, 1745
Wyithe J. S. B., Mould J., Loeb A., 2011, ApJ, 743, 173
Xu H., Wise J. H., Norman M. L., Ahn K., O'Shea B. W., 2016, ApJ, 833, 84
Yang H., Malhotra S., Rhoads J. E., Leitherer C., Wofford A., Jiang T., Wang J., 2017a, ApJ, 838, 4
Yang H. et al., 2017b, ApJ, 844, 171
Yuan Y., Martin-Alvarez S., Haehnelt M. G., Garel T., Sijacki D., 2024, MNRAS, in press
Yusef-Zadeh F., Morris M., White R. L., 1984, ApJ, 278, 186
Zackrisson E., Inoue A. K., Jensen H., 2013, ApJ, 777, 39
Zackrisson E. et al., 2017, ApJ, 836, 78
Zaldarriaga M., Furlanetto S. R., Hernquist L., 2004, ApJ, 608, 622
Zheng Z., Miralda-Escudé J., 2002, ApJ, 578, 33

APPENDIX A: IGM ATTENUATION

In Section 3, we explored the Ly α properties of our sample of mock observations, including comparisons to *JWST* observations. Due to the fact that these galaxies are observed during the epoch of reionization, it is important to account for absorption due to neutral hydrogen in the IGM.

To explore the effect of including IGM attenuation, we apply the analytic model for the transmission function of the IGM given in Inoue et al. (2014) to our Ly α spectra and re-compute the, Ly α luminosities, equivalent widths, and escape fractions. In Fig. A1 we show histograms of each of these quantities, comparing values

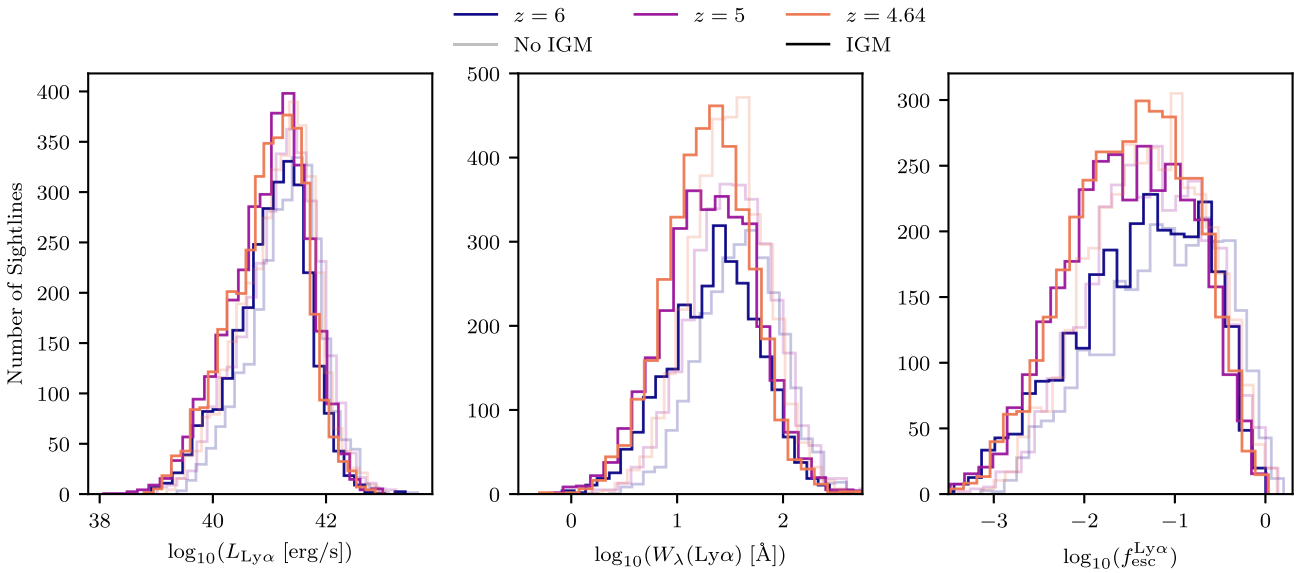

Figure A1. Histograms of Ly α luminosity (left), equivalent width (centre) and escape fraction (right) with (*solid*) and without (*faint*) IGM attenuation, using the analytic expression for IGM optical depths from Inoue et al. (2014).

with (*solid*) and without (*faint*) the IGM correction. We find that as expected, the effect of IGM attenuation is to reduce each of these quantities. However in practice, these IGM-attenuated observables are still entirely consistent with observations shown in Fig. 2 (Roy et al. 2023; Saxena et al. 2024).

APPENDIX B: LINE-OF-SIGHT VERSUS GLOBAL LYC ESCAPE FRACTIONS

It is clear that the processes leading to leakage of LyC radiation are profoundly chaotic and depend on local processes of galaxy evolution. As a result, the amount of ionizing radiation which escapes has a strong line-of-sight dependence (e.g. Gnedin et al. 2008; Zackrisson et al. 2013; Fletcher et al. 2019; Katz et al. 2022; Kimm et al. 2022). As a result, it is important to understand whether line-of-sight measurements of LyC leakage are representative of the global escape fraction for a given galaxy. This is particularly crucial given the fact that observational studies which are able to directly measure LyC emission from galaxies (e.g. Flury et al. 2022a) are limited by only being able to observe galaxies from a single perspective. As a result, their samples may be polluted by fortuitous lines-of-sight with uncharacteristically high, or low, LyC escape fractions.

To this end, we compare the global LyC escape fractions, $f_{\rm esc}^{\rm LyC}$ to the 10 line-of-sight LyC escape fractions, $f_{\rm esc,\,LOS}^{\rm LyC}$ measured for each galaxy in our mock sample. Fig. B1 shows a histogram of $f_{\rm esc}^{\rm LyC}$ versus all 10 $f_{\rm esc,\,LOS}^{\rm LyC}$ values for each galaxy. We also include the angle averaged line-of-sight values for each galaxy in black, finding that while we are still affected by small number statistics for a single galaxy, we do recover the expected one-to-one relation (shown in red) for the full mock sample. Interestingly, we find that galaxies with the largest global escape fractions (i.e. $\gtrsim 5$ per cent) tend to show more isotropic leakage. On the other hand, galaxies with intermediate global leakage (i.e. ~ 1 per cent) tend to be dominated by a few channels with effective leakage.

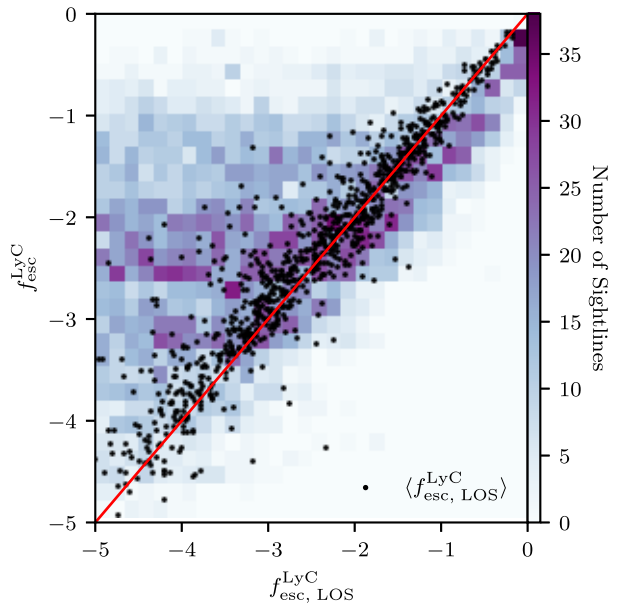

Figure B1. Histogram of global LyC escape fractions, $f_{\rm esc}^{\rm LyC}$ against the 10 line-of-sight LyC escape fractions, $f_{\rm esc,\,LOS}^{\rm LyC}$ measured for each galaxy. We also include the angle averaged line-of-sight values for each true global value as black points. The one-to-one relation is shown in red.

APPENDIX C: OFFSETS DUE TO H I MASS OF SPHINX^{20} GALAXIES

In Section 4, we found that compared to observational data from the LzLCS (Flury et al. 2022b) and Naidu et al. (2022), the Ly α spectra of SPHINX^{20} galaxies tended to exhibit systematically lower $v_{\rm sep}$ and higher $f_{\rm cen}$. Here we explore the origin of this discrepancy. Similar discrepancies were found in cosmological zoom-in simulations of a dwarf galaxies by Yuan et al. (2024).

It has already been established that SPHINX^{20} galaxies are less massive and less UV-bright than those discussed by both the LzLCS and Naidu et al. (2022). As a result, it is reasonable to assume that they will have smaller H I masses (Parkash et al. 2018) and therefore lower H I column densities.

In order to understand the impact of H I column density on $f_{\rm cen}$ and $v_{\rm sep}$, it is instructive to use a simple analytical model. We consider a source of Ly α at the centre of a homogeneous, neutral spherical cloud, with optical depth τ_0. The emergent normalized spectrum has the form (Dijkstra et al. 2006)

$$J(x) = \frac{\pi^{3/2}}{\sqrt{6}a\tau_0}\left[\frac{x^2}{1+\cosh\left(\sqrt{\frac{2\pi^3}{27}}\frac{|x^3|}{a\tau_0}\right)}\right], \tag{C1}$$

where $x = (\nu - \nu_0)/\Delta\nu_D$ is the dimensionless frequency relative to the Ly α line centre at 2.47×10^{15} Hz, $\Delta\nu_D = v_{\rm th}\nu_0/c$ for the thermal velocity $v_{\rm th} = \sqrt{2k_BT/m_p}$. Here, k_B is the Boltzmann constant, T

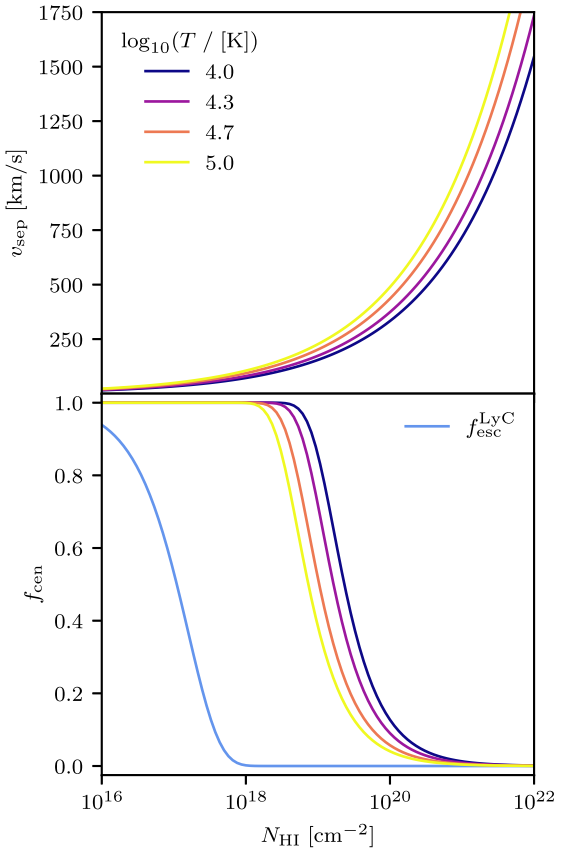

Figure C1. Ly α peak separation, $v_{\rm sep}$, and central escape fraction, $f_{\rm cen}$ for a central source in a static, homogeneous sphere of gas, as given by equations (C3) and (C4).

is the gas temperature, and m_p is the proton mass. Finally, $a = A_{21}/4\pi\Delta v_D$ is the Voigt parameter, where A_{21} is the Einstein A-coefficient for the transition. Furthermore, under these assumptions

$$\tau_0 = N_{\rm HI}\sigma_0 \approx 8.3 \times 10^6 \left(\frac{N_{\rm HI}}{2 \times 10^{20}\ {\rm cm}^{-2}}\right)\left(\frac{T}{2 \times 10^4\ {\rm K}}\right), \quad \text{(C2)}$$

allowing us to connect this theoretical spectrum to the H I column density discussed before. Based on this analytical expression (which we note is symmetric about $x = 0$), we are able to derive expressions for $v_{\rm sep}$ and $f_{\rm cen}$ directly as a function of $N_{\rm HI}$. Doing so, we find

$$v_{\rm sep} = \frac{2cu}{1-u^2}, \quad u = \frac{x_p \Delta v_D}{v_0}, \quad x_p = 0.92(a\tau_0)^{1/3}, \quad \text{(C3)}$$

and

$$f_{\rm cen} = \tanh\left[\frac{\pi^{3/2}}{3\sqrt{6a\tau_0}}\left(\frac{v_0}{\Delta v_D}\frac{V_{\rm cut}}{c - V_{\rm cut}}\right)^3\right], \quad \text{(C4)}$$

where we use a cut-off at $V_{\rm cut} = 100\,{\rm km s}^{-1}$ to compare to the method used by Naidu et al. (2020). We note that these expressions only hold for the static, simple geometry that we have considered here.

Fig. C1 shows $v_{\rm sep}$ and $f_{\rm cen}$ as functions of $N_{\rm HI}$ calculated using equations (C3) and (C4) for various gas temperatures. We also include the LyC escape fraction for this simplified setup to help guide the eye. It is clear that both $v_{\rm sep}$ and $f_{\rm cen}$ can vary strongly with H I column density. For the SPHINX^{20} data set, we have average values of $\langle v_{\rm sep}\rangle = 252\ {\rm km\,s}^{-1}$ and $\langle f_{\rm cen}\rangle = 26$ per cent, respectively. Inverting equations (C3) and (C4), these correspond to an H I column density of $\sim 10^{19.5} - 10^{20}\ {\rm cm}^{-2}$. Furthermore, an average H I column density only 2.5 times greater than that of the SPHINX^{20} average would correspond to a $v_{\rm sep}$ of $\sim 350\ {\rm km\,s}^{-1}$ and $f_{\rm cen}$ of ~ 10 per cent, consistent with the data from the LzLCS (Flury et al. 2022b) and Naidu et al. (2022). Indeed, in Fig. 18 we can see that even such a simple model is able to capture the crux of the relationship between two such complex quantities reasonably well. Therefore, the discrepancies discussed above are likely caused by differences in the H I column densities of these observations compared to the SPHINX^{20} simulations.

As a side-note, it is also clear from Fig. C1 that while this simple model can be instructive, it is clearly insufficient to capture the intricacies of the systems being studied. Namely, this model predicts that $v_{\rm sep}$ will only be a viable diagnostic below H I column densities of $\sim 10^{18}\ {\rm cm}^{-2}$ and that $f_{\rm cen}$ could never be a diagnostic for LyC escape, in contrast to the findings of Section 4. Therefore, we can conclude that these two diagnostics function best when the ISM is in-homogeneous and contains dust.

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