The Physics of Indirect Estimators of Lyman Continuum Escape and their Application to High-Redshift JWST Galaxies

Nicholas Choustikov, Harley Katz, Aayush Saxena, Alex Cameron, Julien Devriendt, Adrianne Slyz, Joki Rosdahl, Jeremy Blaizot, Leo Michel-Dansac

The Physics of Indirect Estimators of Lyman Continuum Escape and their Application to High-Redshift JWST Galaxies

Nicholas Choustikov[1]★, Harley Katz[1], Aayush Saxena[1,2], Alex J. Cameron[1], Julien Devriendt[1], Adrianne Slyz[1], Joki Rosdahl[3], Jeremy Blaizot[3], and Leo Michel-Dansac[3]
[1]Sub-department of Astrophysics, University of Oxford, Keble Road, Oxford OX1 3RH, United Kingdom
[2]Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, United Kingdom
[3]CNRS, Centre de Recherche Astrophysique de Lyon UMR5574, Univ Lyon, Univ Lyon1, Ens de Lyon, F-69230 Saint-Genis-Laval, France

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

Reliable indirect diagnostics of LyC photon escape from galaxies are required to understand which sources were the dominant contributors to reionization. While multiple escape fraction ($f_{\rm esc}$) indicators have been proposed to trace favourable conditions for LyC leakage from the interstellar medium of low-redshift "analog" galaxies, it remains unclear whether these are applicable at high redshifts where LyC emission cannot be directly observed. Using a library of 14,120 mock spectra of star-forming galaxies with redshifts $4.64 \leq z \leq 10$ from the SPHINX^{20} cosmological radiation hydrodynamics simulation, we develop a framework for the physics that leads to high $f_{\rm esc}$. We investigate LyC leakage from our galaxies based on the criteria that successful LyC escape diagnostics must *i)* track a high specific star formation rate, *ii)* be sensitive to stellar population age in the range 3.5 – 10 Myr representing the times when supernova first explode to when LyC production significantly drops, and *iii)* include a proxy for neutral gas content and gas density in the interstellar medium. O_{32}, $\Sigma_{\rm SFR}$, $M_{\rm UV}$, and Hβ equivalent width select for one or fewer of our criteria, rendering them either necessary but insufficient or generally poor diagnostics. In contrast, UV slope (β), and E(B − V) match two or more of our criteria, rendering them good $f_{\rm esc}$ diagnostics (albeit with significant scatter). Using our library, we build a quantitative model for predicting $f_{\rm esc}$ based on β, E(B − V), Hβ, $M_{\rm UV}$, R_{23}, and O_{32}. When applied to bright $z > 6$ Lyα emitters observed with JWST, we find that the majority of them have $f_{\rm esc} \lesssim 10\%$.

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

1 INTRODUCTION

While various astrophysical (e.g. Keating et al. 2020; Kulkarni et al. 2019a; Becker et al. 2021) and cosmological (Planck Collaboration et al. 2020) probes indicate that the Universe had transitioned from a neutral to ionized state by the redshift interval $5 \lesssim z \lesssim 6$, the onset of reionization and sources responsible for it, as well as the neutral fraction evolution during the transition remain uncertain. Both upper and lower limits on the onset of cosmic dawn are provided by the Cosmic Microwave Background (e.g. Heinrich & Hu 2021) and star formation histories (SFHs) of high-redshift galaxies (e.g. Laporte et al. 2021). Although often model-dependent, neutral fraction evolution constraints can be derived from the decreasing prevalence of Lyα emitters (e.g. Pentericci et al. 2011; Stark et al. 2010; Mason et al. 2018), the damping wings of high-redshift quasars (e.g. Ďurovčíková et al. 2020; Davies et al. 2018; Greig et al. 2019), and the opacity of the Lyα forest (e.g. Fan et al. 2006; Bosman et al. 2022). However, only limited observational constraints exist on the sources responsible for reionization.

Understanding the sources of reionization is of key importance. The topology of reionization is strongly affected by the source model, which not only impacts the shape and amplitude of the 21 cm signal, (e.g. Zaldarriaga et al. 2004; McQuinn et al. 2007; Kulkarni et al. 2017), but also controls which dwarf galaxies and filaments are regulated by radiation feedback (Katz et al. 2020a). Furthermore, the temperature the IGM reaches during reionization is sensitive to the spectral energy distribution (SED) of the sources responsible. Star-forming galaxies are often considered the primary candidates for providing the bulk of the LyC photons for reionization (e.g. Robertson et al. 2015) due to the strong decline in the cumulative emissivity of AGN (Kulkarni et al. 2019b). Within the galaxy population, it is generally assumed that reionization was driven by dwarf galaxies due to the steep observed faint-end slopes of the high-redshift UV luminosity function (e.g. Bouwens et al. 2022a; Harikane et al. 2023). However this latter assumption is highly dependent on the amount of LyC photons that are produced per unit star formation (or UV luminosity) as well as the fraction of LyC photons that leak ($f_{\rm esc}$) as a function of mass (or UV luminosity).

The former can be estimated from stellar population synthesis models (e.g. Leitherer et al. 1999; Stanway & Eldridge 2018). While uncertainties in the ionizing photon production rate exist due to systematic differences between stellar population models (e.g. binaries, rotation, IMF, etc.), the escape fraction is far less constrained. This is due to the fact that it emerges from complex, highly non-linear physics (describing e.g. the state of the ISM) and, more importantly, it cannot be directly detected during the epoch of reionization due to

★ nicholas.choustikov@physics.ox.ac.uk

the increasingly neutral IGM (Inoue et al. 2014). For these reasons, constraints on $f_{\rm esc}$ are derived indirectly, for example, by observing samples of low-redshift LyC leaking galaxies that are considered "analogs" of those that form during the epoch of reionization (e.g. Izotov et al. 2018a; Flury et al. 2022a), by directly modelling LyC leakage with cosmological radiation hydrodynamics simulations (e.g. Paardekooper et al. 2015; Rosdahl et al. 2018), or by correlating galaxies with Lyα forest transmission (e.g. Kakiichi et al. 2018).

The number of observational measurements of $f_{\rm esc}$ are rapidly growing. LyC photons are directly detectable with space-based facilities, such as at $z \sim 0.3$ with the cosmic origins spectrograph on HST (Green et al. 2012; Leitherer et al. 2016) or at even higher redshifts with AstroSat (Saha et al. 2020). Ground-based and space-based observatories have pushed the redshift frontier of LyC measurements to $z \gtrsim 3$ (e.g. Vanzella et al. 2010; Steidel et al. 2018; Fletcher et al. 2019; Saxena et al. 2022). The Low Redshift Lyman Continuum Survey (LzLCS, Flury et al. 2022a; Flury et al. 2022b) in particular has significantly increased the total number of low-redshift galaxies with detected LyC emission. However, it remains unclear whether these "analogs" are truly representative of the high-redshift galaxy population (Katz et al. 2022b; Brinchmann 2022; Schaerer et al. 2022b; Katz et al. 2023). Moreover, it is not always clear how to generalize results from observations of individual objects or surveys with complex selection functions to the general population of high-redshift galaxies.

Numerical simulations provide a complementary framework to understand the physics of LyC leakage. Self-consistently modelling the production and transfer of LyC photons through a resolved, multiphase ISM remains a technically challenging problem. Nevertheless, simulations of individual or a few high-redshift galaxies are commonplace (e.g. Kimm et al. 2017; Trebitsch et al. 2017; Kimm & Cen 2014; Ma et al. 2020), though these suffer from similar generalizability arguments. Larger volume or full-box cosmological radiative transfer simulations that resolve the ISM for thousands of galaxies are now becoming technically feasible (e.g. Xu et al. 2016; Rosdahl et al. 2018). SPHINX^{20} (Rosdahl et al. 2022) represents such an effort where the connection between $f_{\rm esc}$ and various galaxy properties (such as stellar and halo mass, UV luminosity, star formation rate (SFR), specific star formation rate (sSFR), metallicity, etc.) can be studied across a sample of $> 10,000$ galaxies per snapshot between $4.64 \leq z \lesssim 15$.

Unfortunately, the connection between simulations and observations remains limited. With the exception of Lyα (e.g. Verhamme et al. 2015; Kakiichi & Gronke 2021; Kimm et al. 2019, 2022; Maji et al. 2022), simulations tend to focus on how $f_{\rm esc}$ varies with "unobservable" quantities such as halo mass. Efforts have been made to mock observations (e.g. Mauerhofer et al. 2021; Barrow et al. 2020; Katz et al. 2020b; Zackrisson et al. 2017) and infer LyC escape; however, these remain a small minority. In contrast, observational-based studies often focus on indirect diagnostics of $f_{\rm esc}$. For example, such diagnostics include high [OIII]λ5007/[OII]$\lambda\lambda$3726, 3728 (O_{32}) ratios (e.g. Jaskot & Oey 2013; Nakajima & Ouchi 2014; Izotov et al. 2018b), Lyα peak separation (e.g. Verhamme et al. 2017; Izotov et al. 2020) or Lyα equivalent width (Steidel et al. 2018; Pahl et al. 2021), [Mg II] $\lambda\lambda$2796, 2804 doublet ratios (Chisholm et al. 2020; Katz et al. 2022b), strong [C IV] $\lambda\lambda$1548, 1550 emission (Schaerer et al. 2022a; Saxena et al. 2022), UV slope (Chisholm et al. 2022), and S II deficits (Wang et al. 2021).

In this work, we address the applicability of various indirect, observationally-developed, diagnostics of LyC escape on a statistical sample of simulated high-redshift galaxies that are likely to be observable with JWST. We develop a physically motivated model for the conditions that need to be met to ensure both high $f_{\rm esc}$ and a simultaneous significant production of LyC photons. Using mock observations from Version 1 of the SPHINX Public Data Release (SPDR1), we discuss how various indirect diagnostics fit within our framework to elucidate the physics of why a leakage indicator is successful (or not). We intend our results to be immediately applicable to the large samples of JWST galaxies currently being observed at $z > 6$ (e.g. Curtis-Lake et al. 2022; Finkelstein et al. 2022; Treu et al. 2022; Matthee et al. 2022a).

This work is organized as follows. In Section 2 we outline the numerical methods behind our new SPHINX^{20} data-set. Section 3 presents and tests our new generalised framework for identifying LyC leaking galaxies. In Section 4 we use our framework to contextualise and explain known diagnostics and use our data-set to predict escape fractions from JWST spectra. Finally, caveats are given in Section 5 and we conclude in Section 6.

2 NUMERICAL SIMULATIONS

Due to the observational challenges of both detecting LyC radiation and being limited to individual lines of sight, we employ state-of-the-art numerical simulations to understand both the physics driving LyC leakage and the observational signatures of a high escape fraction. More specifically, we use the SPHINX^{20} cosmological radiation hydrodynamics simulation (Rosdahl et al. 2022). This simulation is ideal for our purposes because the volume (20^3 cMpc^3) is large enough to sample a wide diversity of galaxy masses and properties, while the spatial and mass resolution[1] allow us to simultaneously resolve many of the low mass haloes that may be contributing to reionization as well as much of the multiphase ISM in the simulated galaxies. Full details of the simulation are described in (Rosdahl et al. 2018, 2022).

We select a subset of galaxies from SPHINX^{20} in the snapshots at $z = 10, 9, 8, 7, 6, 5, 4.64$, that have 10 Myr-averaged SFRs > 0.3 M$_\odot$yr^{-1}. The SFR threshold is designed to mimic a flux limited survey and allows us to study a select group of galaxies from SPHINX^{20} that are most likely to be detectable by deep JWST observations. The total sample contains 1,412 galaxies and in Figure 1 we show the SFR versus stellar mass for our entire sample, along with the best-fit main sequence for the subset of galaxies at $z = 4.64$. Due to the SFR threshold, galaxies in the higher redshift snapshots tend to fall above the star formation main sequence. Hereafter, data displayed will contain galaxies from each of these redshift bins.

For each galaxy we compute a total SED by summing emission from the star particles and gas cells. Stellar emission follows the BPASS v2.2.1 SED (Stanway & Eldridge 2018), and is computed based on the mass, age, and metallicity of each star particle. Line emission from each gas cell is computed using an updated version of the method presented in Katz et al. (2022a). For each gas cell, we calculate the total ionizing flux. For cells that host star particles, we sum the contributions from all star particles within the cell[2]. For gas cells without star particles, we use the ionizing fluxes directly from the RT solver in the simulation. We then identify all cells that host star particles where the Stromgren radius ($R_{\rm S}$) is unresolved

[1] The maximum level of refinement corresponds to a physical scale of ~ 10 pc at $z = 6$ while the star and dark matter particles have masses of 400 M$_\odot$ and 2×10^5 M$_\odot$, respectively.

[2] This assumes that the local star particles dominate the ionizing photon budget of the gas cell. This assumption may fail when a cell has neighbours with much higher star formation efficiencies.

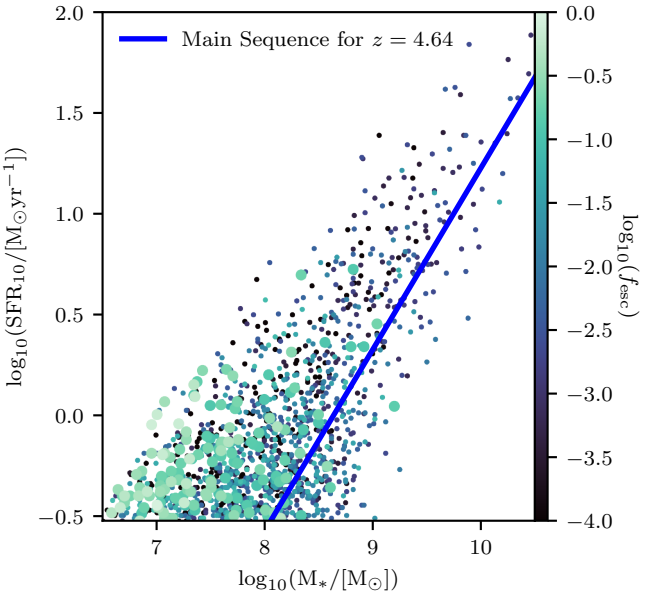

Figure 1. Star formation rate (averaged over 10 Myr) as a function of stellar mass. Each data-point is coloured by the LyC escape fraction, with those with fractions above 10% enlarged. This shows that LyC leakers tend to fall above the main sequence fitted for galaxies at $z = 4.64$, representing starburst periods.

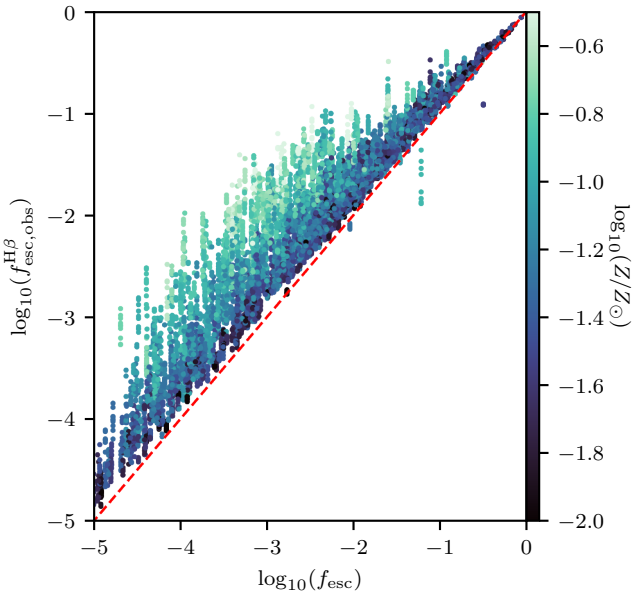

Figure 2. Line-of-sight Hβ-derived LyC escape fractions as a function of true LyC escape fraction for SPHINX^{20} galaxies, coloured by the gas metallicity. Metal-rich systems tend to over-predict f_{esc} as compared to the true value.

(i.e. $R_{\mathrm{S}} < \Delta x/2$). For cells without an unresolved Stromgren sphere, line emissivities are calculated with CLOUDY v17.03 (Ferland et al. 2017). We tabulate a grid of one-zone slab models varying the gas density, metallicity, ionization parameter, and electron fraction. All models are iterated to convergence and the shape of the SED varies with metallicity but is assumed to have an age of 10 Myr. To calculate the emission from unresolved Stromgren spheres, we run a second grid of CLOUDY models assuming a spherical geometry, varying stellar age, metallicity, gas density, and total ionizing luminosity. Stellar age is computed as the ionizing luminosity-weighted average stellar age in each gas cell. The total intrinsic emission line luminosity of a galaxy is then the sum of all gas cells within the virial radius.

To better compare with observations we also account for dust scattering and absorption. Following Katz et al. (2022b), we use RASCAS (Michel-Dansac et al. 2020) and the effective SMC dust model from Laursen et al. (2009) to attenuate the SED and line emission. Since the attenuation depends on viewing angle, we employ the peeling algorithm (Yusef-Zadeh et al. 1984; Zheng & Miralda-Escudé 2002; Dijkstra 2017) to compute the full dust-attenuated SED for ten different uniformly distributed directions. Hence our full data set consists of 14,120 simulated spectra, though we will in some cases discuss angle-averaged versions of these quantities. We then use these spectra to extract dust-attenuated observables, including line luminosities, equivalent widths, UV spectral index (β, by fitting to $f_\lambda \propto \lambda^\beta$ around 1500 Å), UV attenuation given by the Balmer decrement (E(B − V)), UV magnitude (M_{UV}), and the star formation rate surface density (Σ_{SFR}). In order to compare like-for-like, Σ_{SFR} is measured by a SFR converted from Hβ luminosity (Kennicutt 1998) as well as the dust attenuated half-light radius at 1500 Å.

We have also post-processed the simulation to measure f_{esc} for all galaxies in our sample. Angle-averaged LyC escape fractions are calculated with RASCAS (Michel-Dansac et al. 2020) by ray-tracing LyC emission from star particles (see Rosdahl et al. 2022). While this value can be used to measure the instantaneous contribution of each halo to reionization, it cannot be easily measured with observations due to both the wavelength coverage and the deep exposure times needed. For this reason, we measure a second value of f_{esc} for each line of sight using the Hβ luminosity such that (Izotov et al. 2016):

$$f^{\mathrm{H}\beta}_{\mathrm{esc,obs}} = \frac{f_{\mathrm{esc}} L_{\mathrm{LyC}}}{L_{\mathrm{H}\beta}/4.86 \times 10^{-13} + f_{\mathrm{esc}} L_{\mathrm{LyC}}}. \quad (1)$$

This allows us to make a more fair comparison with observational surveys such as LzLCS (Flury et al. 2022b) where such a method is used. A comparison between the two f_{esc} measurements is shown in Figure 2. While there is a strong correlation between the two quantities for the highest values of f_{esc}, in general, the Hβ method tends to over-predict the true value. This is due to the fact that the normalization constant of 4.86×10^{-13} which is commonly used in the literature (e.g. Matthee et al. 2022a) is not fully representative of the value in our simulation[3]. This bias will not qualitatively impact the general trends we find between f_{esc} and various observational quantities. Henceforth, any reference to f_{esc} invokes the angle-averaged value derived by RASCAS, while $f^{\mathrm{H}\beta}_{\mathrm{esc}}$ refers to the value derived by Equation 1.

3 A GENERALISED FRAMEWORK FOR LYC LEAKAGE

Numerous diagnostics for identifying galaxies with high f_{esc} have been suggested in the literature (see Section 1); however the vast majority have been shown to be "necessary but insufficient" conditions for LyC leakage (Flury et al. 2022b). We therefore aim to provide a more general framework describing the conditions necessary for

[3] By fitting intrinsic Hβ to total ionizing flux for our galaxies we find a value of 4.05×10^{-13}. However, in order to best compare to observational methods we use the theoretical value. (e.g. Schaerer 2003)

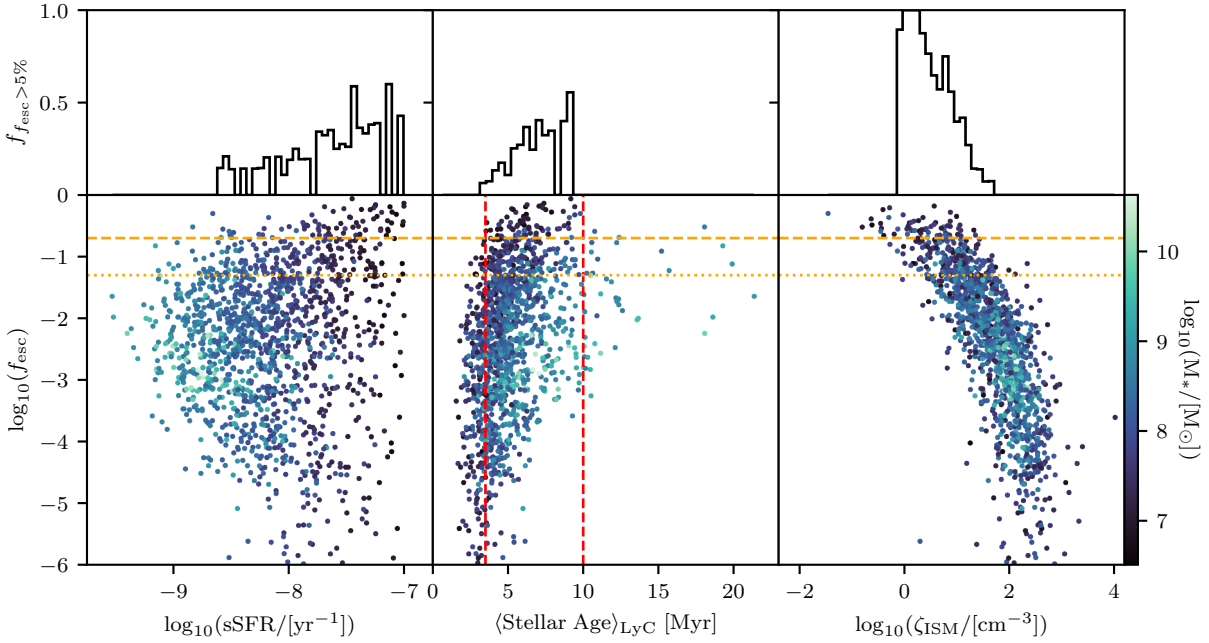

Figure 3. LyC escape fraction as a function of specific star formation rate (left), ionizing flux-weighted mean stellar age (center) and angle-averaged composite neutral gas attenuation parameter $\zeta_{\mathrm{ISM}} = \mathrm{E(B-V)} \times \langle \mathrm{n_H} \rangle_{\mathrm{[OII]}}$ (right), all coloured by stellar mass. Escape fractions of 20% and 5% are marked with dashed and dotted horizontal orange lines, respectively. The top panels show a histogram of the fraction of galaxies with $f_{\mathrm{esc}} > 5\%$ in each bin with more than 5 such galaxies. Systems with the highest escape fractions tend to have high sSFR, fall within 3.5 Myr and 10 Myr (indicated by red lines), low neutral gas attenuation parameter, and have low stellar masses. However, each of these requirements is a necessary but insufficient diagnostic for LyC leakers.

galaxies to both produce and leak a significant amount of ionizing LyC radiation.

In order for a galaxy to meaningfully contribute to reionization it must simultaneously be producing copious amounts of ionizing photons and a fraction of those photons must be able to leak into the low-density IGM where the recombination timescale is longer than the Hubble time. While simulations have yet to quantitatively agree on the average escape fractions of galaxies as a function of various galaxy properties (e.g. mass, Ma et al. 2020; Rosdahl et al. 2022), qualitatively they nearly all find that f_{esc} is a feedback-related quantity (e.g. Trebitsch et al. 2017). Energetic feedback from stars (e.g. supernovae (SN)) disrupt the ISM, creating ionized channels through which photons leak into the IGM. Thus, based on such a scenario we propose that in order to produce a significant amount of LyC leakage there needs to be:

- A strong burst of star formation such that a significant quantity of ionizing photons are produced.
- Either a significant reduction in the neutral gas content of the ISM such that photons can leak relatively isotropically (density-bounded case) or a creation of ionized channels (ionization-bounded case). This is achieved through photoionization and mechanical feedback.
- A timescale synchronization such that stars continue producing significant amounts of LyC photons after feedback has disrupted the ISM.

Therefore, we argue that a good diagnostic to identify LyC leakers should simultaneously encapsulate:

(i) High sSFR so that significant numbers of ionizing photons are produced and feedback has a chance of disrupting the ISM.

(ii) A stellar population age $\gtrsim$ 3.5 Myr such that SN have had time to explode but $\lesssim$ 10 Myr so that the LyC production rate remains high.

(iii) A proxy for neutral gas content and the ionization state of the ISM to identify when feedback has efficiently coupled to gas in order to create ionized channels or disrupt/ionize the entire medium.

In Figure 3 we demonstrate that each of these conditions alone are insufficient to select a sample of only LyC leaking galaxies. In the left panel we show f_{esc} versus sSFR (coloured by total stellar mass) and while many of the leakers have high sSFR, in general there is no trend between the two quantities (see also Rosdahl et al. 2022). Similar behaviour is also seen in observations (e.g. Flury et al. 2022b; Saxena et al. 2022). The centre panel shows f_{esc} versus mean stellar age weighted by LyC luminosity (to highlight the age of stars which contribute to LyC flux). In order to reach an escape fraction of 20% (above the dashed orange line), the age must be $\gtrsim$ 3.5 Myr as indicated by the left-most vertical red line. However, selecting by age alone clearly results in significant contamination. Our final criteria is a representative proxy for the state of the ISM. It is difficult to find a proxy that only contains information about the ISM and not the SFR or age as these are highly coupled. Nevertheless, for demonstrative purposes we have chosen a composite parameter $\zeta_{\mathrm{ISM}} = \mathrm{E(B-V)} \times \langle \mathrm{n_H} \rangle_{\mathrm{[OII]}} / [\mathrm{cm}^{-3}]$ where we multiply the angle-averaged UV attenuation E(B − V), (which has already been shown by Saldana-Lopez et al. (2022) to empirically correlate with f_{esc}) by the gas density weighted by the intrinsic [O II] $\lambda\lambda$3727 luminosity[4]. While there is clearly a strong trend and all galaxies with $f_{\mathrm{esc}} >$ 20% have a value $\log_{10}(\zeta_{\mathrm{ISM}}) \lesssim 1.3$, there are many non-leakers with such low values as well. These findings are reinforced by the

[4] Note that we find qualitatively similar results when replacing density with the intrinsic ratio of [O II] λ3729/[O II] λ3726.

histogram above each subplot showing the fraction of galaxies with escape fractions higher than 5% (i.e. above the dotted orange line) in each bin containing more than five such leakers.

It is clear that none of these three conditions alone are sufficient for identifying a contamination-free sample of LyC leakers. However, together, they provide a robust framework for identifying LyC leakers. In Figure 4 we plot angle-averaged $\zeta_{\rm ISM}$ versus sSFR for SPHINX^{20} galaxies with $3.5 \leq \langle \text{Stellar Age} \rangle_{\rm LyC} / [\text{Myr}] < 10$. The galaxies with high $f_{\rm esc}$ are biased towards having high sSFR and low $\zeta_{\rm ISM}$. By selecting galaxies with

$$\log_{10}(\zeta_{\rm ISM}/[\text{cm}^{-3}]) < 0.4 \times \log_{10}(\text{sSFR}/[\text{yr}^{-1}]) + 4.3, \quad (2)$$

we generate a sample based on angle-averaged quantities that is biased towards having high $f_{\rm esc}$. In the bottom panel of Figure 4 we show a cumulative distribution function for the escape fractions of galaxies that satisfy our selection criteria. This consists of 227 galaxies (16% of our sample), $\sim$ 74% of which have $f_{\rm esc} > 5\%$. These leakers account for $\sim$ 65% of all such leakers in our sample. This can be contrasted with LzLCS where only 12/66 galaxies (18%) have $f_{\rm esc} > 5\%$. While the efficiency of our selection is not yet perfect, our framework provides a theoretically motivated model with minimal contamination from galaxies with low $f_{\rm esc}$.

Despite the effectiveness of such a selection method, it is important to note that thus far this discussion has focused on galaxy properties that are not directly observable. Nevertheless, for completeness, in Figure 5 we plot $f_{\rm esc}$ versus halo mass, stellar mass, metallicity, 10 Myr-averaged SFR, 100 Myr-averaged SFR and the sSFR surface density. As a subsample of SPHINX^{20} galaxies, the data set presented here displays many of the trends described in Rosdahl et al. (2022) relating to which galaxy properties correlate with $f_{\rm esc}$. However, there are subtle differences because we have selected only star-forming galaxies. There is a minor tendency for $f_{\rm esc}$ to decrease with increasing halo mass and stellar mass, consistent with Razoumov & Sommer-Larsen (2010); Kimm & Cen (2014); Paardekooper et al. (2015); Xu et al. (2016); Ma et al. (2020); Rosdahl et al. (2022). Because of our cut in SFR, we do not sample a downturn in $f_{\rm esc}$ that is seen at lower stellar and halo masses in some simulations (Ma et al. 2020; Rosdahl et al. 2022).

It is well established observationally that gas-phase metallicity scales with stellar mass (e.g. Lequeux et al. 1979; Tremonti et al. 2004). Such a trend holds in SPHINX^{20} and for this reason, the trend of $f_{\rm esc}$ with metallicity is also similar to that of $f_{\rm esc}$ with stellar mass. Likewise, there exists a star formation main sequence that correlates stellar mass and SFR (e.g. Brinchmann et al. 2004; Salim et al. 2007), as shown for our sample in Figure 1. Hence we see similar behaviour between SFR_{10} and $f_{\rm esc}$ as for stellar mass and $f_{\rm esc}$. Though we see the same broad behaviour for SFR_{100} (albeit with more scatter), low longer-term SFR seem to better select for LyC leakers.

While few trends exist between these fundamental galaxy properties and $f_{\rm esc}$, the histogram of each sub-Figure shows that there are certain regions of parameter space where one is more likely to find LyC leakers. For example, there are a higher fractions of leakers at low virial masses, stellar masses, and gas metallicities, indicating that the conditions needed for leakage are more often accessible in these environments. Similar behaviour is also observed in LzLCS where even though a galaxy property may not be predictive of the value of $f_{\rm esc}$, the detection fractions of LyC emitters may be higher when certain conditions are met (e.g. high O_{32}, high EW(Hβ), etc.). JWST data will be key for determining whether such conditions are more often met in the high-redshift Universe compared to locally (e.g. Katz et al. 2023; Cameron et al. 2023).

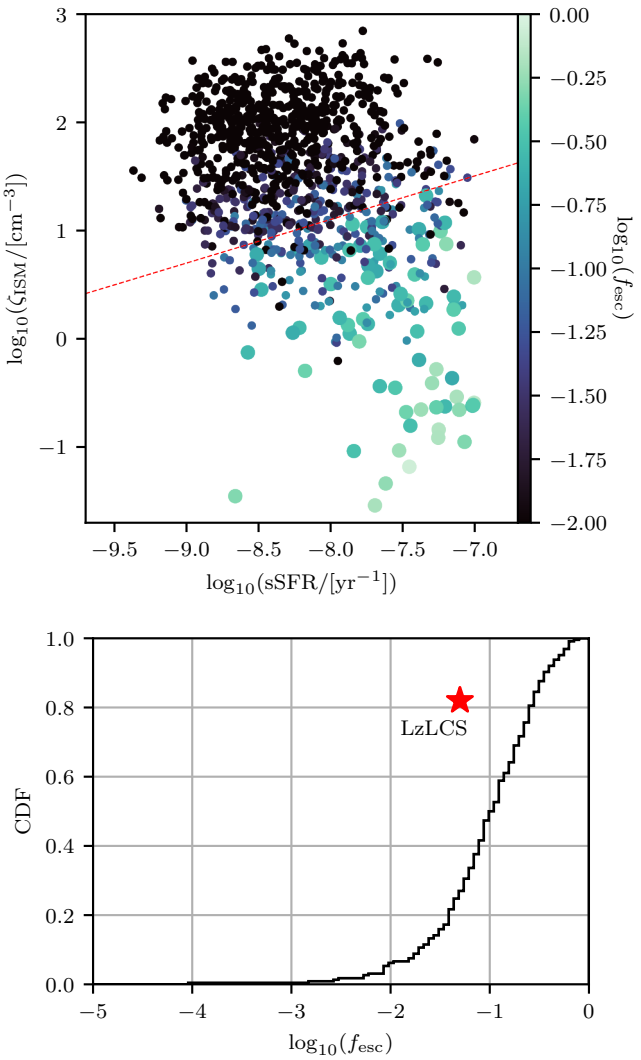

Figure 4. (Top) angle-averaged neutral gas parameter $\zeta_{\rm ISM}$ as a function of specific star formation rate for galaxies with mean stellar ages between 3.5 Myr and 10 Myr coloured by LyC escape fraction. Systems with $f_{\rm esc} > 20\%$ are enlarged. A selection criteria (given by Equation 2) to produce a sample highly enriched with leakers, 74% of which have $f_{\rm esc} > 5\%$ is shown as the red line. (Bottom) Cumulative distribution for escape fractions for galaxies that satisfy our selection criteria. For comparison, the red star shows that 82% of galaxies selected to be part of LzLCS have $f_{\rm esc} < 5\%$, while 26% of SPHINX^{20} galaxies within these criteria have $f_{\rm esc} < 5\%$.

4 DISCUSSION

In this section, we contextualize numerous indirect observational diagnostics for $f_{\rm esc}$ and develop a new model that can be used to quantitatively infer $f_{\rm esc}$ from high-redshift observations.

4.1 Contextualising Existing Diagnostics

Using the theoretically motivated model for LyC leakage presented in Section 3 we proceed to predict whether individual indirect $f_{\rm esc}$ diagnostics suggested in the literature work well based on whether they match our three criteria.

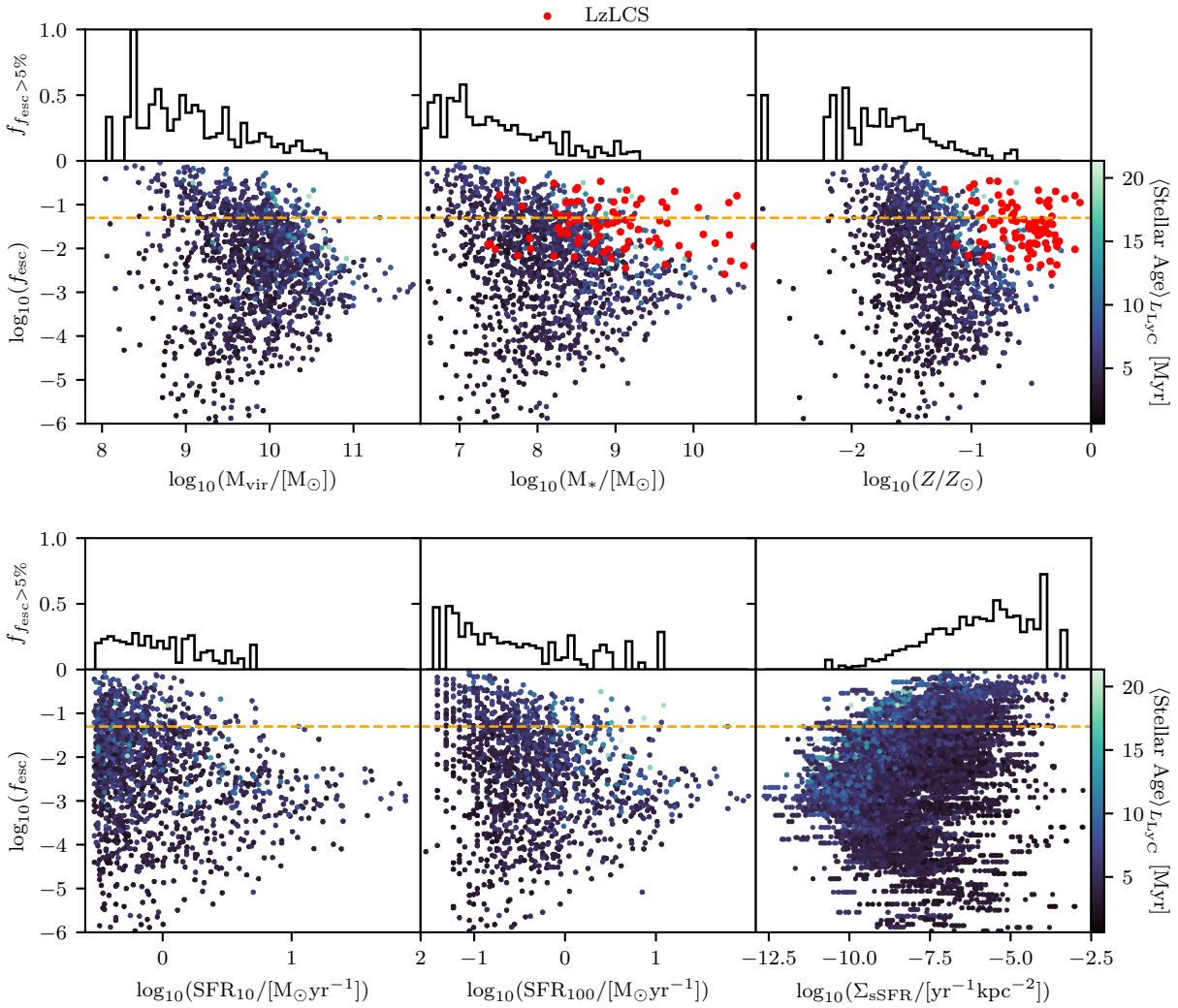

Figure 5. LyC escape fraction as a function of halo virial mass (top-left), stellar mass (top-center), gas metallicity (top-right), 10 Myr averaged SFR (bottom-left), 100 Myr averaged SFR (bottom-center), and sSFR surface density (bottom-right) coloured by the LyC luminosity-weighted mean stellar age. Where available, LzLCS data is shown in red for comparison. For each quantity, a histogram is given for the number density of galaxies with $f_{esc} > 5\%$ in each bin (shown by the dashed orange line).

4.1.1 O_{32}

There remains significant debate in the literature on the applicability of O_{32} as a diagnostic of f_{esc}. While 1D ISM models (e.g. Jaskot & Oey 2013; Nakajima & Ouchi 2014) and certain observations (e.g. Faisst 2016; Izotov et al. 2018b) indicate a strong correlation between O_{32} and f_{esc}, numerous theoretical models (e.g. Katz et al. 2020b; Barrow et al. 2020) and other observations (Bassett et al. 2019; Nakajima et al. 2020; Flury et al. 2022b) demonstrate that there are complications with viewing angle, ionization parameter, and metallicity such that high O_{32} is perhaps a necessary but insufficient condition for LyC leakage. Within our framework, O_{32} is particularly complex.

As an ionization parameter diagnostic (e.g. Kewley & Dopita 2002), high O_{32} is likely indicative of high sSFRs as seen in the top-left panel of Figure 6. There is significant scatter in the relation based on geometric effects, as well as variations with metallicity and other ISM properties, nevertheless, the highest observed values of O_{32} in our simulation traces the highest sSFR, and hence O_{32} satisfies our first criteria.

O_{32} is expected to strongly vary with stellar cluster age due to the evolution of the ionizing sources (both in terms of brightness and spectral shape). This behaviour is shown in Figure 2 of Barrow et al. (2020) and is sensitive to the chosen SED model as well as the presence of Wolf-Rayet stars. The galaxies with the highest O_{32} also have the youngest LyC luminosity-weighted stellar ages (see the top-second panel of Figure 6). O_{32} peaks at ages < 3.5 Myr and thus fails our second criterion as it preferentially traces objects that have yet to surpass the SN time scale. Similarly, O_{32} is not an indicator of neutral ISM density and only marginally traces dust, thus also failing our third criteria (see the top-third panel of Figure 6). For these reasons, by itself, O_{32} is not a good predictor of f_{esc}.

In Figure 7 we show the observed and de-reddened O_{32} versus $f^{H\beta}_{esc,obs}$. As expected, no obvious trend emerges. We find reasonably good agreement with LzLCS (red points) in that there is a preference for galaxies with high f_{esc} to be biased towards high O_{32}, but high O_{32} does not imply high $f^{H\beta}_{esc,obs}$. Hence this confirms our previous

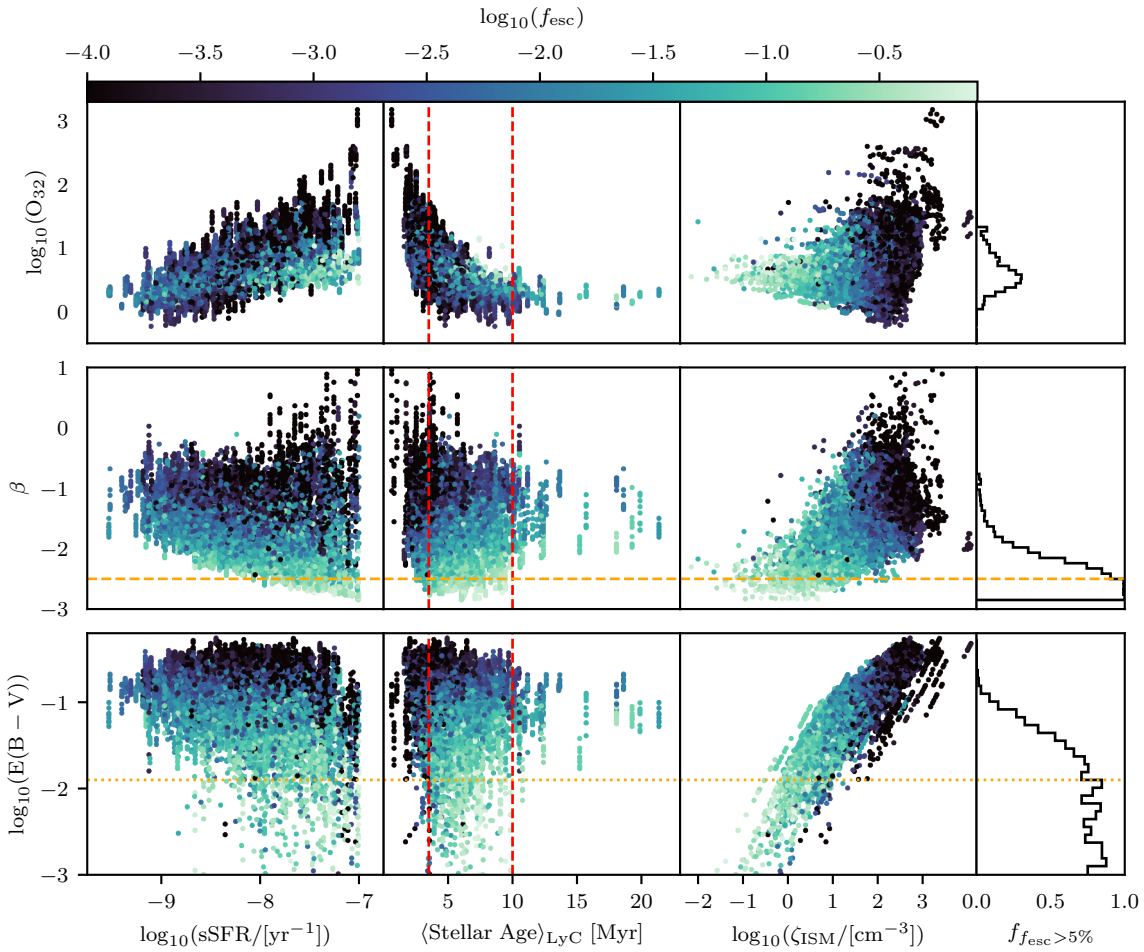

Figure 6. (Top) Observed dust-corrected (by the Balmer decrement) O_{32} as a function of sSFR (left), LyC luminosity-weighted mean stellar age (second) and neutral gas attenuation parameter $\zeta_{\rm ISM}$ (third) coloured by the LyC escape fraction. The histogram (right) shows the fraction of galaxies for a given O_{32} with $f_{\rm esc} > 5\%$ for each bin which contains at least 5 such galaxies. There is a preference for leakers to have $3 < \log_{10}(O_{32}) < 10$. Higher values of O_{32} correlate with high sSFR, yet select for younger stellar populations and do not trace the neutral gas content of the ISM. Hence, O_{32} by itself does not reliably predict $f_{\rm esc}$. (Middle) Same as above, but for spectral index β. Values of $\beta < -2.5$ are highlighted (dashed orange line) as they fulfill all three criteria, empirically selecting for high sSFR, ideal stellar age, and lower neutral gas densities in the ISM. These galaxies are the strongest leakers, as can be seen in the histogram. (Bottom) Same as above, but for the UV attenuation E(B − V). Values of $E(B-V) < 0.01$ are highlighted (dotted orange line) as they appear to fulfill 2/3 criteria, empirically selecting for galaxies with the correct mean stellar population age and low $\zeta_{\rm ISM}$.

assertion that high O_{32} is a necessary but insufficient condition for high $f_{\rm esc}$. We argue that the reason why there is a bias for leakers to have high observed O_{32} is two-fold. First there is a clear, albeit with significant scatter, correlation between O_{32} and sSFR which represents one of our three conditions. Second, for galaxies older than 3.5 Myr, high ionization parameter can help make feedback more efficient. Nevertheless, there are a few galaxies with high $f_{\rm esc}$ and lower O_{32} compared to the typical SPHINX^{20} galaxy. The majority of these galaxies appear as bright blue points on Figure 7 because they have the oldest stellar ages. Such galaxies tend to be intrinsically much fainter than the galaxies with high $f_{\rm esc}$ and high O_{32} and represent the population of post-starburst Remnant Leakers described in Katz et al. (2022c).

We highlight that there seems to be a preferred O_{32} between 3 and 10 where galaxies are most likely to have significant LyC leakage. This range corresponds to the typical values of O_{32} that a galaxy reaches when it has evolved past the SN timescale, and can be easily seen in the top histogram of Figure 6. The exact details of this peak are sensitive to Wolf-Rayet modelling as well as dust treatment, ISM gas densities, and star formation model in the simulation; all of which will need to improve before we are confident in the robustness of this particular scale.

4.1.2 *Spectral Index β*

Recently, Chisholm et al. (2022) have suggested that the UV slope, β, is a strong indicator of $f_{\rm esc}$, such that galaxies with bluer β have higher $f_{\rm esc}$. β is a readily observable quantity in the high-redshift Universe as it can be estimated for large samples of galaxies from both photometry or more accurately with JWST spectroscopy (e.g. Topping et al. 2022; Endsley et al. 2022; Cullen et al. 2023). Hence, should it be a good diagnostic of $f_{\rm esc}$, β may be an exciting probe of LyC leakeage directly in the EoR.

While β is not necessarily a strong indicator of sSFR, in order to have a steep slope, the young stellar population must outshine the older stellar populations in the galaxy. We empirically find (middle-

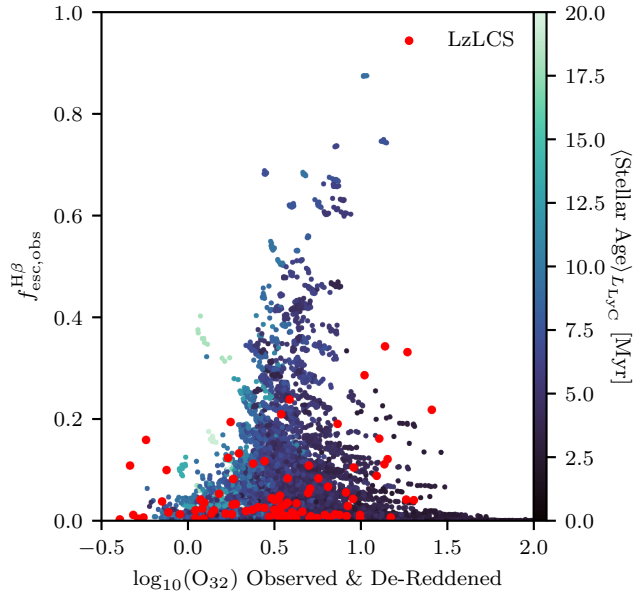

Figure 7. LyC escape fraction as a function of observed and de-reddened O_{32}, coloured by the ionizing luminosity weighted mean stellar age in each galaxy. Over-plotted are observational results from the LzLCS. We find that LyC escape fractions peak qualitatively at values of $\log_{10}(O_{32})$ between 3 and 10. Larger values of O_{32} are dominated by young stellar populations which have yet to disrupt the ISM, producing low escape fractions.

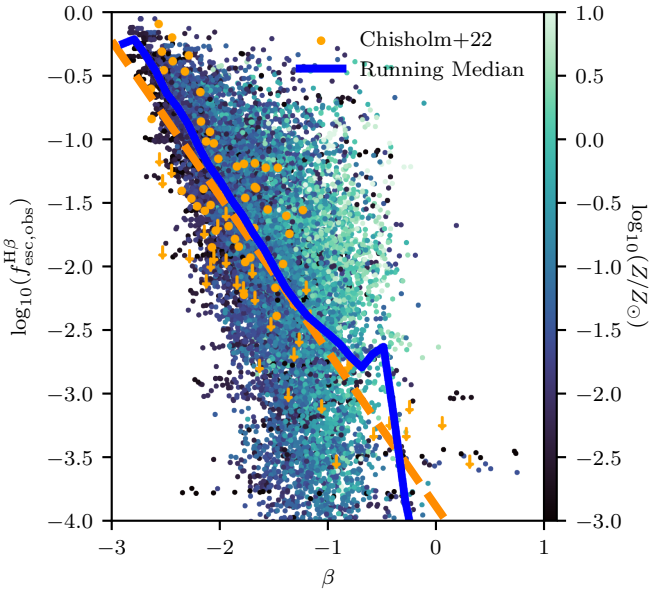

Figure 8. LyC escape fraction as a function of observed spectral index β and coloured by gas metallicity. Over-plotted are observational data-points and predicted relation (orange) from Chisholm et al. (2022), as well as our median relation (blue). We find that our data agrees well with observational data, confirming that β strongly anti-correlates with LyC escape fraction.

left panel of Figure 6) that galaxies with the bluest β (i.e. < -2.5) are also very strongly biased towards having the highest sSFRs. A high sSFR does not guarantee a blue β and there exists strong scatter due to dust; however as an f_{esc} indicator, β satisfies our first criteria.

The UV slope is predicted to stay approximately constant for the first 10 Myr of evolution (Stanway et al. 2016). Hence a blue β does not reveal much about the stellar population age apart from the fact that it may still be producing significant quantities of ionizing photons. Thus β marginally satisfies our second criteria in that it picks out galaxies that can contribute to reionization but we expect some contamination from low f_{esc} galaxies with stellar ages younger than 3.5 Myr (middle-second panel of Figure 6).

Finally, the observed β is strongly sensitive to dust. While not a density indicator, β can easily select for galaxies with very low E(B − V) and hence β also marginally satisfies our third criteria. For these reasons, in a metal-enriched environment we expect β to be a relatively good indicator of f_{esc}, albeit with significant scatter.

In Figure 8 we plot $f^{H\beta}_{esc,obs}$ versus the observed β for SPHINX[20] galaxies and we find a strong trend that high $f^{H\beta}_{esc,obs}$ galaxies tend to have bluer β (as can be clearly seen in the middle histogram of Figure 6). We do find systems that also have redder β and high $f^{H\beta}_{esc,obs}$ but some of this is a sight-line effect and likewise, Remnant Leakers, which do not contribute meaningfully to reionization, will populate this region of the plot. The orange points and dashed orange line show both the data and relation predicted by Chisholm et al. (2022) which when compared to our median relation (blue) shows very good agreement.

We emphasize that this result holds only when metals/dust are present as without dust obscuration, our third criterion would not be satisfied. Indeed the galaxies with the lowest f_{esc} at the bluest β are the most metal-poor galaxies in our simulated sample. Our model naturally predicts that this could become problematic at $Z/Z_\odot < 0.01$. Here we have assumed that the dust-to-gas mass ratio scales linearly with metallicity; however, observations show that dust content may fall off as a power-law with decreasing metallicity (e.g. Rémy-Ruyer et al. 2014). In that case, we expect that the critical metallicity where β no longer becomes a good f_{esc} indicator occurs at $Z/Z_\odot > 0.01$. Hence to be conservative, we advocate that observed β is likely to be a good f_{esc} indicator at $Z/Z_\odot > 0.1$ and the details at lower metallicity (and the ability to apply this relation in the epoch of reionization) are sensitive to how the dust-to-gas mass ratio scales with metallicity and the timescales of dust formation at high-redshift.

4.1.3 E(B − V)

The problem of whether dust attenuation strongly affects f_{esc} is not completely understood. Chisholm et al. (2018) suggest that even small dust attenuation removes significant numbers of ionizing photons. However, simulations have shown that dust tracks neutral hydrogen which has a much more important impact on f_{esc} (Katz et al. 2022b). Therefore, it is natural to explore the use of UV attenuation (e.g. E(B − V), derived from the Balmer decrement) as a potential f_{esc} diagnostic. Saldana-Lopez et al. (2022) found a strong anti-correlation between the UV dust-attenuation and LyC escape fraction for galaxies in the LzLCS sample, suggesting that LyC leakers tend to have a dust-poor ISM.

Unsurprisingly, we find no significant dependence of E(B − V) on sSFR (bottom-left panel of Figure 6), with any residual relationship introduced by the fact that stellar mass is the denominator of sSFR and high-mass galaxies tend to be more metal-enriched and thus have more dust. E(B − V) does not pass our first criterion. Interestingly, we find that galaxies outside our stellar age criterion tend to have significantly more UV attenuation (bottom-second panel of Figure 6). This points to the fact that SNe are able to destroy dust (in our case by destroying neutral hydrogen) through mechanical feedback (e.g.

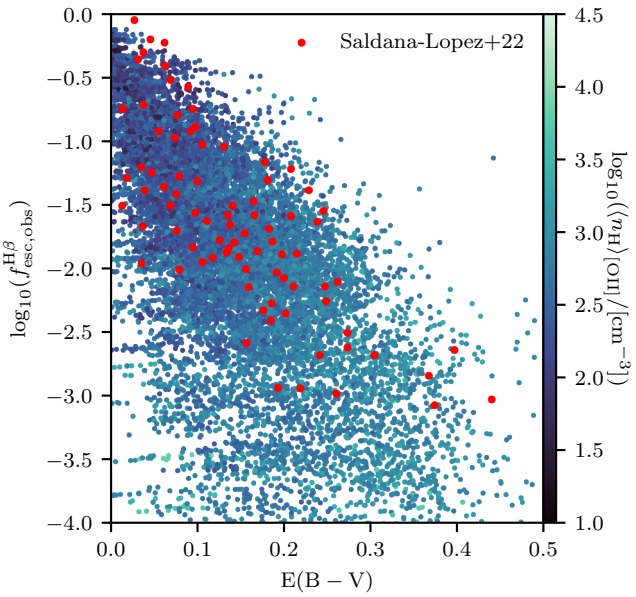

Figure 9. Observed LyC escape fraction as a function of dust-attenuation at 912 Å coloured by the mean neutral gas density weighted by the O II doublet ratio, compared to data from the LzLCS (Saldana-Lopez et al. 2022).

Priestley et al. 2021). Therefore, E(B − V) marginally satisfies our second criterion. Finally, it is clear that (by construction), there is a strong correlation between E(B − V) and $\zeta_{\rm ISM}$ (bottom-third panel of Figure 6). Therefore, we conclude that E(B − V) should be a good indicator of LyC leakage.

In Figure 9 we show the observed $f_{\rm esc}$ as a function of E(B − V) compared to LzLCS galaxies from Saldana-Lopez et al. (2022). We find a strong trend between the two quantities. Furthermore, galaxies with lower $\langle n_{\rm H}\rangle_{\rm [OII]}$ exhibit less scatter. However, it is likely that such comparisons are sensitive to the dust model used in the simulation. Similarly, Saldana-Lopez et al. (2022) assume a uniform dust screen which is not representative of the dust distribution in our simulation.

4.1.4 $\Sigma_{\rm SFR}$

Star formation rate surface density is perhaps intuitively a good indicator of $f_{\rm esc}$. Since $f_{\rm esc}$ is predicted to be feedback-regulated (e.g. Trebitsch et al. 2017; Kimm et al. 2017), concentrated star formation may help increase the local efficiency of mechanical feedback, creating optically thin, low-density channels in the ISM. With limited empirical data, Sharma et al. (2017); Naidu et al. (2020) assumed that galaxies with the highest $\Sigma_{\rm SFR}$ also have the highest $f_{\rm esc}$, which can result in a reionization scenario dominated by the most massive galaxies.

As SFR is the numerator of $\Sigma_{\rm SFR}$, there is unsurprisingly a strong correlation between sSFR and $\Sigma_{\rm SFR}$ (top-left panel of Figure 10). This is due to the weaker dependence of galaxy size on stellar mass (Kawamata et al. 2018; Bouwens et al. 2022b). As with previous diagnostics, we similarly find significant scatter due to the variable impact of dust on the Hβ emission and projected galaxy size as a function of sight line. $\Sigma_{\rm SFR}$ undoubtedly passes our first criterion.

In contrast we find that the highest values of $\Sigma_{\rm SFR}$ tend to correspond to galaxies with ages younger than the SN timescale (top-second panel of Figure 10), although there are a significant number of galaxies that have ages > 3.5 Myr and $\Sigma_{\rm SFR} > 10$ M$_\odot$/yr/kpc^2 (the canonical value reported in Flury et al. (2022b) as the threshold for strong leakage). Galaxy size in our simulations is also relatively stable beyond 10 Myr. Thus, $\Sigma_{\rm SFR}$ alone does not satisfy our second criterion and may be biased towards galaxies with too young stellar ages. Finally, we find no relation between $\Sigma_{\rm SFR}$ and the state of the ISM (top-third panel of Figure 10). We then might expect that galaxies with high $f_{\rm esc}$ might also have high $\Sigma_{\rm SFR}$ due to the correlation with sSFR, but we expect there to be no strong correlation.

Indeed in Figure 11 we find no trend between the quantities. The relation suggested by Naidu et al. (2020) does not envelope our data, nor does it represent the LzLCS galaxies, which are consistent with those presented here. Similarly, the assumption by Sharma et al. (2017) that all galaxies with $\Sigma_{\rm SFR} > 0.1$ M$_\odot$/yr/kpc^2 have $f_{\rm esc} = 20\%$ is clearly an inaccurate representation of both SPHINX^{20} galaxies and LzLCS. Interestingly, in our SFR-limited sample, the lowest halo mass galaxies have the highest $\Sigma_{\rm SFR} > 0.1$ M$_\odot$/yr/kpc^2 as they tend to fall above the main sequence of star formation.

Flury et al. (2022b) do note a trend between $\Sigma_{\rm SFR}$ and $f_{\rm esc}$. It is possible that one emerges due to their selection criteria which are not fully representative of the general galaxy population. Hence trying to reproduce their correlation coefficient with SPHINX^{20} may result in the correct value for the wrong reasons. Finally, we find that by selecting with $\Sigma_{\rm sSFR}$ one is more likely to find a galaxy with significant leakage, which can be seen by comparing the histogram in Figures 5 and the top histogram of 10. This is in contrast with findings in the LzLCS, which report that the addition of stellar mass did not improve selection power (Flury et al. 2022b).

4.1.5 Hβ Equivalent Width

The connection between Hβ equivalent width (EW(Hβ)) and $f_{\rm esc}$ remains debated. Green Pea galaxies are among the most studied low-redshift galaxy populations that contain a significant number of LyC leakers (e.g. Izotov et al. 2016, 2018b,a) and these systems often exhibit extreme emission line ratios and equivalent widths (e.g. Yang et al. 2017). In contrast, as the escape fraction approaches 100%, EW(Hβ) should tend towards zero as none of the ionizing photons are absorbed (Zackrisson et al. 2017). While LzLCS find no strong trend between EW(Hβ) and $f_{\rm esc}$, galaxies with high $f_{\rm esc}$ tend to also have high EW(Hβ).

EW(Hβ) is a very strong tracer of sSFR (middle-left panel of Figure 10) as Balmer lines have long been known to track SFR and the strength of the continuum is sensitive to total stellar mass (e.g. Kennicutt 1998). However, like O_{32}, the highest values of EW(Hβ) trace ages < 3.5 Myr and there is no correlation between EW(Hβ) and the state of the ISM (middle-second and middle-third panels of Figure 10 Hence we expect no strong correlation between EW(Hβ) and $f_{\rm esc}$; although, the connection with sSFR would explain the trend seen in LzLCS that the LyC leaker fractions increases with EW(Hβ). We note that similar to our findings for O_{32}, there appears to be a characteristic EW(Hβ) $\sim$ 100Å for which galaxies tend to show elevated $f_{\rm esc}$ (see the middle histogram of Figure 10). However, for the same reasons as before it is difficult to claim a robust value.

4.1.6 $M_{\rm UV}$

Flury et al. (2022b) recently reported a weak correlation between $f_{\rm esc}$ and $M_{\rm UV}$, in agreement with other observations that indicate LyC leakers tend to be lower mass, fainter galaxies (Steidel et al. 2018; Pahl et al. 2021). In contrast, (Rosdahl et al. 2022; Saxena et al. 2022) find no relation between between $f_{\rm esc}$ and $M_{\rm UV}$, although they agree that lower luminosity galaxies are likely the sources of reionization.

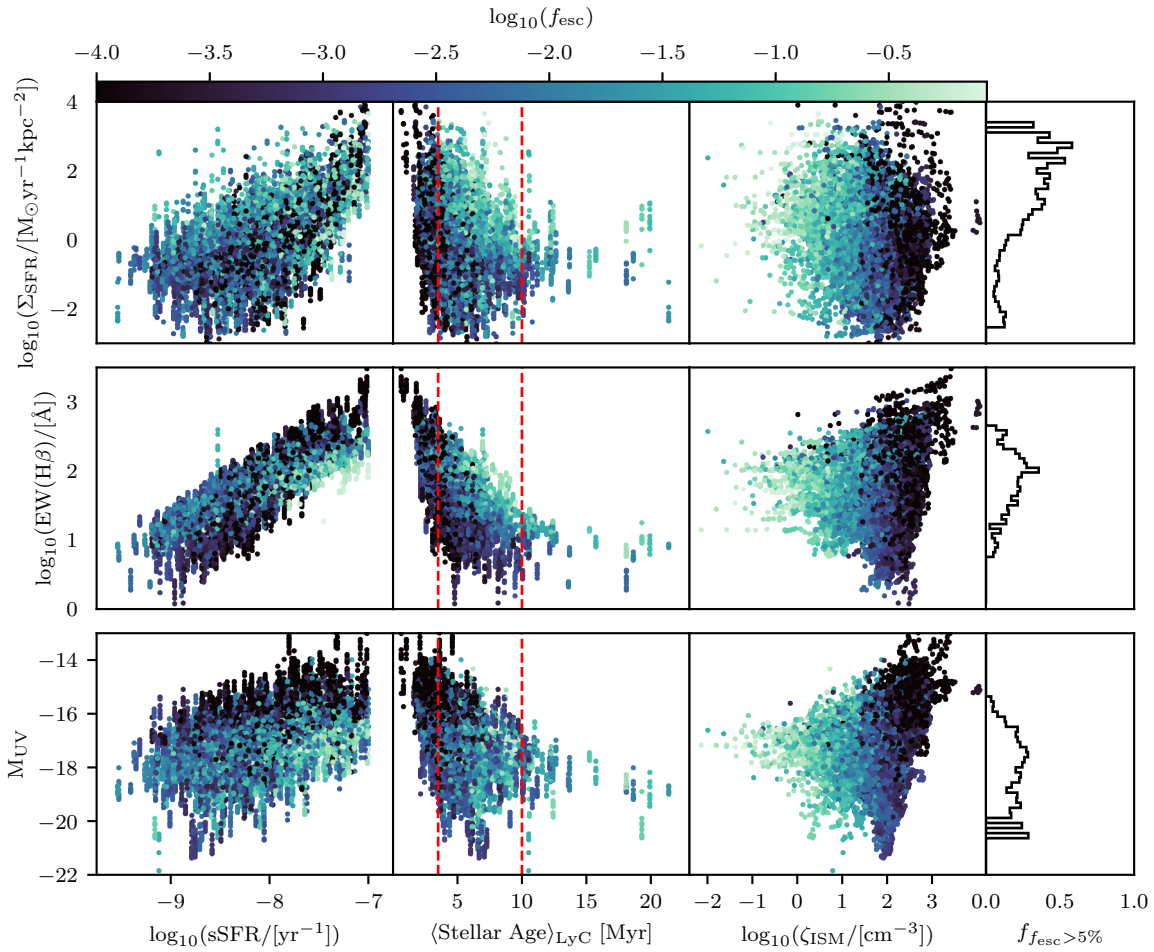

Figure 10. (Top) Observed $\Sigma_{\rm SFR}$ (as defined in Section 2) as a function of sSFR (left), LyC luminosity-weighted mean stellar age (second) and neutral gas attenuation parameter $\zeta_{\rm ISM}$ (third) coloured by the LyC escape fraction. The histogram (right) shows the fraction of galaxies for a given $\Sigma_{\rm SFR}$ with $f_{\rm esc} > 5\%$ for each bin that contains at least 5 such galaxies. Though $\Sigma_{\rm SFR}$ correlates weakly with sSFR (as both depend explicitly on SFR), it selects for young stellar ages and does not trace $\zeta_{\rm ISM}$. Therefore, $\Sigma_{\rm SFR}$ by itself is not a reliable diagnostic for the LyC escape fraction. (Middle) Same as above, but for Hβ equivalent widths. EW(Hβ) traces sSFR very well, but greater values select for stellar ages < 3.5 Myr and do not trace the state of the ISM. We therefore expect no strong relation with the LyC escape fraction, though values of 100 Å are weakly preferred. (Bottom) Same as above, but for $M_{\rm UV}$. We find that $M_{\rm UV}$ does not trace the sSFR, but bright magnitudes select for the correct stellar ages. However, $M_{\rm UV}$ shows no dependence on ISM density. As a result, leakers show a weak bias towards brighter magnitudes, but $M_{\rm UV}$ is not a useful indicator of the LyC escape fraction.

In general, $M_{\rm UV}$ is a weak indicator of sSFR (bottom-left panel of Figure 10) while high $M_{\rm UV}$ could also indicate ages < 3.5 Myr (bottom-second panel of Figure 10). Similarly we find no correlation between $M_{\rm UV}$ and ISM state (bottom-third panel of Figure 10) so our model would predict no correlation between $M_{\rm UV}$ and $f_{\rm esc}$ as shown in (Rosdahl et al. 2022). Figure 12 demonstrates this, comparing our galaxies to those of the LzLCS (Flury et al. 2022b). Here, we find that the majority of this scatter is introduced by a strong dependence of $M_{\rm UV}$ on stellar mass.

When stellar mass is fixed, high intrinsic $M_{\rm UV}$ correlates with a high sSFR. However, like the observed β at fixed stellar mass, observed $M_{\rm UV}$ becomes a strong indicator of dust attenuation which represents one of the tracers of our combination parameter on ISM state. Thus at fixed stellar mass the observed $M_{\rm UV}$ satisfies two out of three criteria with scatter introduced due to stellar age. A strong colour gradient is present in Figure 13 where we compare $M_{\rm UV}$ with M_* and colour points by $f_{\rm esc}$ such that brighter galaxies at fixed stellar mass are biased towards having higher $f_{\rm esc}$. This demonstrates that sample selection is key for the emergence of trends between $f_{\rm esc}$ and various galaxy properties.

4.2 A New Combined Diagnostic

Inspired by the success of our theoretical selection criteria in Figure 4 (top), we now aim to reproduce the ability of a three-point criterion at isolating systems with high LyC leakage, albeit with quantities which are directly observable. Furthermore, we can make use of the fact that several quantities satisfy multiple criteria to find the best set of observables from which to construct our diagnostic.

Given the fact that galaxies with blue dust-attenuated β marginally satisfy criteria 1 and 2, while strongly satisfying 3 (see middle row of Figure 6) we choose to start here. We continue with our second-strongest individual diagnostic, E(B – V), which weakly indicates criterion 2 and strongly satisfies criterion 3 (see bottom row of Figure 6). Finally, within this set of diagnostics we note that the most loosely constrained is mean stellar age, for which we now select the Hα-to-

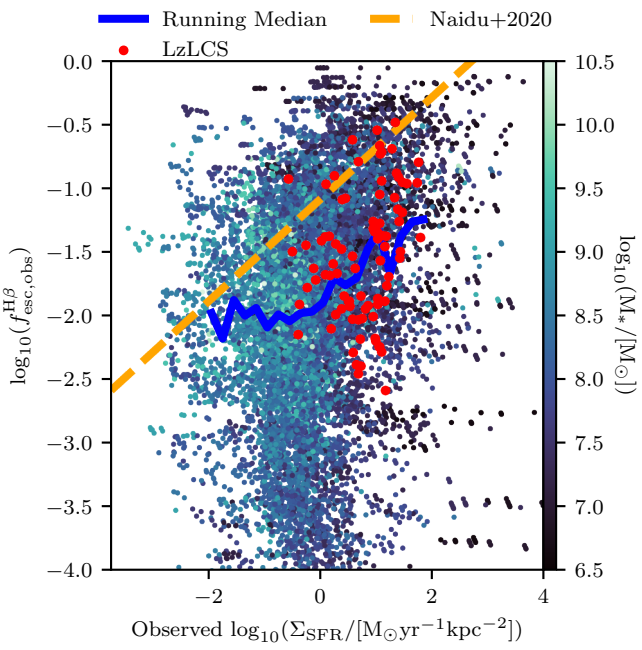

Figure 11. LyC escape fraction as a function of $\Sigma_{\rm SFR}$ coloured by stellar mass. LzLCS data is shown in red, while the relation suggested by Naidu et al. (2020) is shown in orange. We find that SPHINX^{20} galaxies agree well with LzLCS observations, while disagreeing with the proposed relation of Naidu et al. (2020).

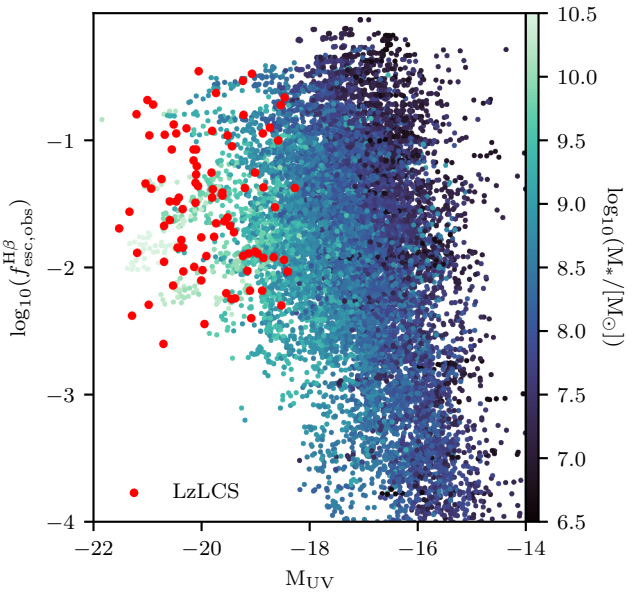

Figure 12. LyC escape fraction as a function of line-of-sight $\rm M_{UV}$ coloured by the stellar mass. Data from LzLCS is shown in red, agreeing with bright SPHINX^{20} galaxies. We find that when all masses are considered, $\rm M_{UV}$ is not a reliable diagnostic for the LyC escape fraction.

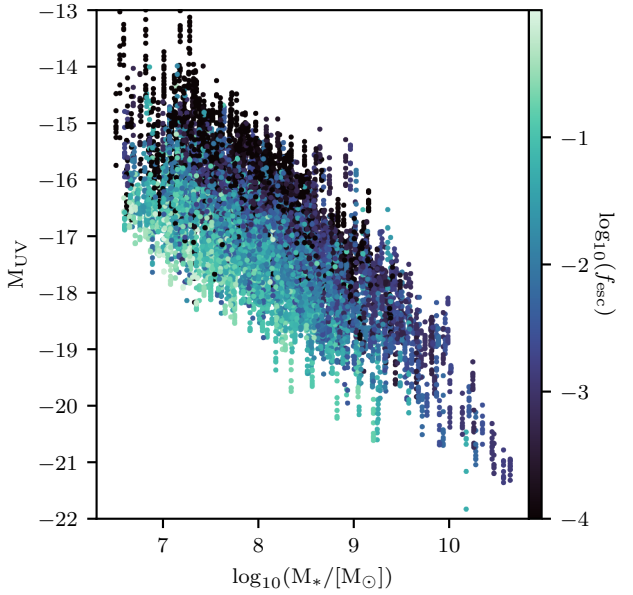

Figure 13. Line-of-sight observed UV magnitude for each mock observation as a function of the galaxy stellar mass, coloured by the LyC escape fraction. We find that more massive galaxies tend to be more luminous in the UV, and that for a fixed galaxy mass bin, brighter galaxies tend to have higher LyC escape fractions. Therefore, at constant stellar mass, $\rm M_{UV}$ can be used as a LyC diagnostic.

FUV flux ratio, given the fact that it has been previously shown to indicate stellar age both observationally (Weisz et al. 2012) and in simulations (Sparre et al. 2017), though this is by no means settled (c.f. Rezaee et al. 2022). Therefore, we predict that this set of three diagnostics can be used to reliably select a sample of galaxies which are greatly enriched with LyC leakers.

Figure 14 shows this, plotting line-of-sight β as a function of E(B − V) for all galaxies with $\log_{10}({\rm H}\alpha/{\rm L}_{1500}) < 1.6$[5] coloured by $f_{\rm esc}$, with galaxies with $f_{\rm esc} > 20\%$ shown as larger points. Here, we find that by selecting galaxies with

$$\log_{10}({\rm E(B-V)}) < -1.1\beta - 3.3, \quad (3)$$

we generate a sample that is highly enriched with LyC leakers. Namely, this reduced set consists of 1931 observations (16.9% of our sample), ∼ 67% of which have $f_{\rm esc} > 5\%$. This accounts for 62% of all such galaxies in our sample. In contrast, the theoretical selection criteria given in Figure 4 is 74% enriched by such galaxies, accounting for 65% of the overall sample. We stress that this is a theoretically motivated set of directly observable diagnostics which successfully selects for the majority of LyC leakers in our sample.

4.3 Predicting $f_{\rm esc}$ for bright Lyα emitters observed with JWST

There remains debate in the literature over the contribution of bright Lyα emitters to reionization. For example, based on lower-redshift stacks, Naidu et al. (2022) assumed that bright Lyα emitters with low peak velocity separation and line-centre flux all have escape fractions

[5] We note that while using a stronger constraint, e.g. $\log_{10}({\rm H}\alpha/{\rm L}_{1500}) < 1.4$ will produce a more enriched sample of galaxies, it will inevitably represent a smaller set of all LyC leaking systems.

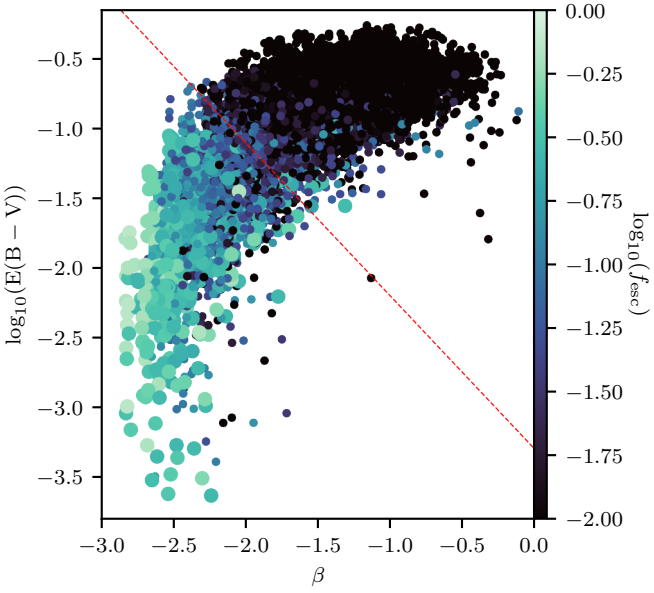

Figure 14. Line-of-sight measurements of β as a function of E(B − V) for all observations of galaxies in our sample with observed $\log_{10}(H\alpha/L_{1500}) < 1.6$ coloured by the true LyC escape fraction. Systems with escape fractions greater than 20% are enlarged. A selection has been made (red, given by Equation 3) to produce a sample of highly enriched leakers, 67% of which have $f_{esc} > 5\%$. Those selected by this set of criteria account for 62% of the total population of such systems in our sample.

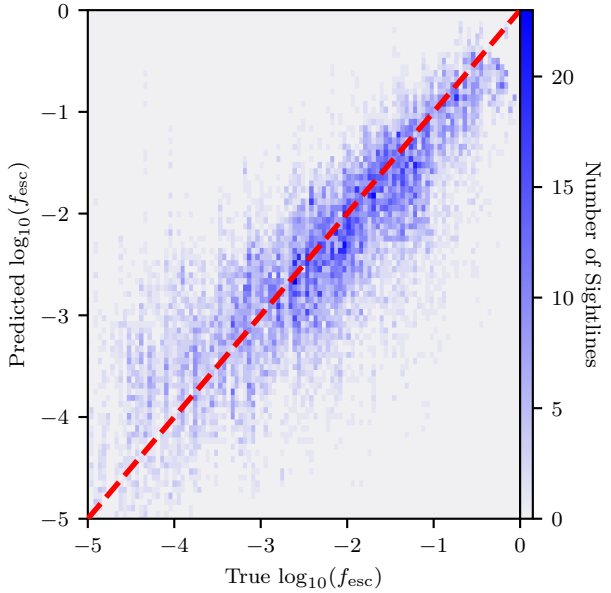

Figure 15. Histogram of predicted f_{esc} as a function of true f_{esc} for our entire sample, as estimated by the generalized linear model given by Equation 4 with parameters from Table 1. Points are coloured by the number of sightlines in each bin. The one-to-one relation is shown in red.

of 20%. Furthermore, Matthee et al. (2022b) showed that if half of bright Lyα emitters have an escape fraction of 50%, then one can match observational constraints on the neutral fraction evolution. In contrast, based on a small stack of $z > 6$ galaxies, Witten et al. (2023) inferred that bright Lyα emitters have $f_{esc} \lesssim 10\%$.

One of the primary advancements of JWST compared to earlier observations is the ability to obtain high resolution spectra of the rest-frame UV and optical for large numbers of galaxies at $z > 6$. While our previous discussion has focused on which indirect indicators are likely to identify enriched samples of LyC leakers, here we use the observable properties of galaxies to quantitatively predict the value of the escape fraction, particularly for a sample of high-redshift Lyα emitters. A similar exercise was performed in Maji et al. (2022); however, the main difference here is that we focus only on observable quantities such that our models are immediately applicable to available galaxy spectra.

We consider eight observable quantities: β, E(B − V), Hβ, EW(Hβ), M_{UV}, R_{23}, O_{32}, and the half-light radius measured at 1500 Å. We then run a generalized linear model on scaled data[6] using L1 regularization to limit the number of needed measurements and maintain simplicity so that the model is interpretable. To avoid over-fitting, we have split[7] the data such that the model is trained on 80% and the remaining 20% is used for validation. Six of the eight initial parameters have non-zero coefficients, with both EW(Hβ) and the half-light radius proving irrelevant for our model. Using these six

[6] i.e. mean of zero and unit variance.

[7] When we split the data, we ensure that all viewing angles for each galaxy are part of the same class to avoid information leakage.

Table 1. Coefficients and constants required to solve Equation 4 to quantitatively predict the value of f_{esc} from an observed spectrum.

i	p_i	$\bar{p}_i$	σ_{p_i}	C_i
1	β	−1.528	0.526	−0.641
2	E(B − V)	0.139	0.094	−0.391
3	$\log_{10}(H\beta/[\mathrm{erg\ s^{-1}}])$	40.630	0.342	0.030
4	M_{UV}	−17.055	1.110	−0.021
5	$\log_{10}(R_{23})$	0.661	0.235	0.031
6	$\log_{10}(O_{32})$	0.680	0.357	−0.424

parameters, f_{esc} can be estimated as:

$$\log_{10}(f_{esc}) = -2.5 + \sum_{i=1}^{6} C_i \frac{p_i - \bar{p}_i}{\sigma_{p_i}}. \quad (4)$$

Values for all coefficients are listed in Table 1. The median absolute error on the training and validation sets are identical at 0.39 dex indicating that the model generalizes well. This can also be seen in Figure 15, where we show predicted f_{esc} as a function of true f_{esc} for our sample.

The coefficients provide insight into how each parameter correlates with f_{esc}. For example, the coefficients for β and E(B − V) are strongly negative indicating that blue UV slopes and low dust content are strong indicators of leakage. Counter-intuitively, O_{32} negatively scales with f_{esc} in our model. We believe this is due to the fact that once galaxies with blue UV slopes and low dust are selected, O_{32} becomes an age indicator and lower O_{32} indicates older ages. We emphasize that this anti-correlation exists only in the framework that contains these other parameters.

As a first application, we apply this relation to JADES-GS-z7-LA, a $z = 7.3$ Lyα emitter (Saxena et al. 2023) recently discovered as part of the JWST JADES GTO program. Based on a variety of spectroscopic features, the authors concluded that the current value of f_{esc} is not substantially high, despite the high Lyα f_{esc}. Using our model, we predict a value of 3% which is considerably lower than

the value needed to inflate an ionized bubble such that Lyα is not completely attenuated by the IGM (Saxena et al. 2023). Hence it is more likely that faint nearby dwarf galaxies are likely responsible for the local ionized bubble.

Beyond $z \sim 9.5$ [O III] λ5007 drops out of NIRSpec; however, GN-z11, a spectroscopically confirmed $z = 10.6$ galaxy was recently discovered to be a bright Lyα emitter (Bunker et al. 2023). Because Hβ, O_{32}, and R_{23} have not been measured for this galaxy, we create a custom model using Hγ and [O III] λ4363 rather than [O III] λ5007. We find an escape fraction of 11%, significantly greater than the 4% reported as the Lyα escape fraction. While it is theoretically difficult to obtain a LyC escape fraction greater than that of Lyα and similarly this is rarely observed (e.g. Izotov et al. 2022; Verhamme et al. 2017), there is undoubtedly scatter in our model, which more than accounts for this discrepancy. IGM attenuation can also play a role in extinguishing the observed Lyα. Hayes & Scarlata (2023) argue that the Lyα escape from GN-z11 into the IGM is as high as 50%, which makes our measured LyC escape fraction fully consistent. Similarly the SED fit for GN-z11 shows a marginal $A_V = 0.17$. If we input this into our model (rather than the fiducial parameters that assumed no dust as suggested in Bunker et al. 2023), the estimated LyC escape fraction decreases to 6%.

While these two galaxies represent single-object detections, Tang et al. (2023) recently published a sample of six $z > 7$ Lyα emitters from CEERS (Finkelstein et al. 2022). Among these six, four galaxies (CEERS-1019, CEERS-1027, CEERS-698, & CEERS-44) have measurements for all of the quantities we need to apply our model. We find LyC escape fractions of 4%, 0.6%, 0.7%, and 10%, respectively. Compared to the Lyα escape fractions for these galaxies (4%, 9%, 5%, and 34%), our measured LyC f_{esc} values are once again consistent with Lyα escape being higher than LyC escape.

In general, our model seems to point to bright Lyα emitters having LyC escape fractions $\lesssim$ 10%. This result agrees with Witten et al. (2023) but contradicts Naidu et al. (2022) where it was assumed that bright Lyα emitters with low peak velocity separation and line-centre flux all have $f_{esc} = 20\%$. The origin of this difference seems to be related to UV slope. Within the context of our model, β is the strongest indicator of f_{esc}. We re-emphasize that this seemingly also holds true for low-redshift "analogs" (Chisholm et al. 2022). The β values between the assumed high and low f_{esc} galaxies in Naidu et al. (2022) are formally consistent within the scatter, with a slight preference for the higher alleged f_{esc} sample to be more blue. Using the stacks they provide, we have estimated f_{esc} for their two samples and found values of 4% and 0.4% for the stacks with low peak velocity separation and high line-centre flux and high peak velocity separation and low line-centre flux, respectively. Applying the model from (Chisholm et al. 2022), which only depends on β, we derive escape fraction values of 5% and 3%, respectively. The difference in f_{esc} in our model is primarily driven by the E(B – V) difference between the two samples.

Our result does not indicate that bright Lyα emitters are insignificant for reionization. Despite their estimated lower escape fractions, their intrinsic production of ionizing photons is very high, and if 5% of the ionizing photons leak, this may represent an important contribution to the emissivity budget. For the galaxies considered in Matthee et al. (2022b) to dominate reionization one would need to increase the assumed ξ_{ion} to reconcile the lower escape fractions. Future JWST observations will undoubtedly provide new constraints on both the f_{esc} of Lyα emitters and ξ_{ion}.

4.4 Comparison with Other Simulations

SPHINX^{20} has not been the only attempt to study escape fractions in a cosmological context. Numerous works have been carried out studying how ionizing photons leak out of galaxies (e.g. Xu et al. 2016; Barrow et al. 2020; Trebitsch et al. 2021; Rosdahl et al. 2022; Hassan et al. 2022). Our approach however has focused primarily on observational signatures. This is important, as we can directly study mock-observed galaxy properties, accounting for the anisotropic nature of LyC leakage.

Simulations regularly show that f_{esc} is sensitive to stellar age. For example, haloes with stellar populations younger than 5 Myr in the FIBY simulations have higher escape fractions across a larger solid angle (Paardekooper et al. 2015). FIRE-2 find a lag between the timing of a star burst and an increase in f_{esc}, due to the time needed for feedback to clear channels (Ma et al. 2020), similar to what we find in SPHINX^{20}. They also found a telltale geometry when f_{esc} is high. Star-forming regions are surrounded by an accelerated, dense gas shell. Within this shell, young stars with ages $3 - 10$ Myr are able to ionize low column-density channels through which radiation can leak. Kimm & Cen (2014) also find evidence for lags between star formation and high f_{esc}. However, in their model, the lag was 10 Myr. This highlights the sensitivity of this exact time delay to the SN model being used. In Kimm & Cen (2014), star particles undergo SNe exactly 10 Myr after their birth, following Schaller et al. (1992). In contrast, the feedback recipe followed in SPHINX^{20} includes a number of staggered SNe more accurately representing a single stellar population (Kimm et al. 2015). Specifically, these begin as early as $\sim$ 3 Myr.

The need for ISM disruption has also been explored at great length. Ma et al. (2020) find that star particles in galaxies with high escape fractions tend to be situated in regions with low column densities out to the virial radius. This is also corroborated by the findings of Trebitsch et al. (2017) (see Figure 14). Moreover, Paardekooper et al. (2015) show that the neutral gas column density within 10 pc of a source is the defining quantity of escape fractions. Kimm & Cen (2014) agree, finding that galaxies with the lowest escape fractions tend to have the highest optical depths out to 100 pc and that the location of feedback is of vital importance. Particularly, the inclusion of runaway OB stars increases average escape fractions, due to the fact that these stars tend to move to lower density regions where the efficiency of feedback is greater. In our model, local gas density is included in our ζ_{ISM} parameter. Because we have weighted the gas density by the [O II] $\lambda\lambda$3727 luminosity, we specifically pick out the densities in star-forming regions. The anti-correlation we find between ζ_{ISM} and f_{esc} in SPHINX^{20} is in agreement with these previous models.

5 CAVEATS

Like all numerical simulations, SPHINX^{20} employs a series of subgrid models for star formation, feedback, and ISM processes that could impact our results. For example, SPHINX^{20} samples a distribution of SN time scales rather than assuming a fixed value which is why we find a critical time scale of 3.5 Myr for LyC leakage to begin. While assuming a delay time distribution is likely more realistic than a fixed value, a different delay time distribution would undoubtedly change which galaxies in SPHINX^{20} are leakers, shifting the exact stellar age dependence stated. Likewise, our model for star formation assumes a variable efficiency of conversion from gas to stars. Changing this value will impact the clustering of stars and SN as well as

the gas densities near young star particles. This will simultaneously affect the intrinsic emission line luminosities, the observed values (through a changing amount of dust as it is tied to the gas), and the efficiency of SN and radiative feedback.

Since SPHINX^{20} does not follow the formation and distribution of dust, we have employed an effective model where the dust-to-metal ratio is fixed and dust primarily tracks neutral gas (Laursen et al. 2009). In contrast, observations show that the dust-to-gas mass ratio decreases following a power-law as a function of metallicity (Rémy-Ruyer et al. 2014). While the dust is often a sub-dominant contribution to the optical depth to ionizing photons at these redshifts (e.g. Katz et al. 2022b) and therefore does not impact our escape fractions, the dust model does affect emission line luminosities and UV slopes. This motivates the study of IR lines as a probe of $f_{\rm esc}$ (e.g. Katz et al. 2020b; Ramambason et al. 2022) as they are significantly less sensitive to dust content.

In this work, we have employed the BPASS SED which crucially extends the period over which ionizing photons are released due to binary interactions, compared to other SEDs. Similarly, one could change the model for Wolf-Rayet stars or include X-ray binaries which would similarly impact emission line fluxes. Our models also assume that metal abundance ratios match that of solar, whereas observations demonstrate that they likely vary as a function of metallicity. This will impact gas cooling and the state of the ISM as well as the emission line luminosities. For this reason, we have only worked with oxygen emission lines and their ratios which are likely well captured by our assumptions.

While currently the state-of-the-art for full-box reionization simulations, SPHINX^{20} remains subject to limited spatial and mass resolution. This most importantly manifests in our inability to always fully resolve the Stromgren spheres of star particles. We have attempted to remedy this by using complex post-processing methods. However, the gas cells are still limited to have a fixed density on $\sim$ 10 pc scales. Subgrid density structure will impact emission line fluxes as well as dust attenuation and potentially change the impact of pre-SN feedback.

SPHINX^{20} does not resolve the ISM to the same degree as simulations such as those in Kimm et al. (2019, 2022). In particular, Kimm et al. (2019) suggest that radiation feedback begins to disrupt the ISM of giant molecular clouds to the point of allowing LyC leakage at around 2 Myr, before SNe begin. In such a scenario, the window for effective LyC leakage would begin even earlier, inviting the need for further work with higher resolution simulations and better ISM and dust physics. However, it remains unclear whether gas outside the immediate molecular cloud prevents LyC photons from escaping into the IGM.

SPHINX^{20} also neglects some potentially important physical processes such as stellar winds and cosmic rays. This could help lower the local densities around star particles and provide local metal enrichment. Both can impact emission line luminosities and the escape fraction. Likewise, we have neglected the nebular continuum. While properties such as O_{32} and $\Sigma_{\rm SFR}$ are unaffected, the nebular continuum can reduce observed equivalent widths and make β appear redder. This is unimportant when $f_{\rm esc} \sim 100\%$. Furthermore, the highest EW galaxies in our sample are non-leakers because they have not yet reached the SN time scale and it is these galaxies where the nebular continuum will be the most important.

Despite these caveats, SPHINX^{20} has been successful in reproducing numerous observations of the high-redshift Universe such as the UV luminosity function (Rosdahl et al. 2022) and Lyα luminosity function (Garel et al. 2021). The agreement we find with LzLCS is a promising sign that in many ways SPHINX^{20} provides an adequate representation of the physics that leads to LyC leakage.

6 CONCLUSIONS

We have post-processed a sample of 1,412 star-forming galaxies in the SPHINX^{20} cosmological radiation hydrodynamics simulation with CLOUDY and RASCAS to produce a diverse library of 14,120 simulated and dust-attenuated high-redshift galaxy spectra. These galaxies have been specifically selected to potentially be bright enough to be observable with JWST. Using this data-set, which represents part of SPHINX Public Data Release v1, we presented a new generalised framework for observational signatures of LyC leakage. Specifically, we argue that a good diagnostic to identify galaxies with significant LyC leakage should:

- track high sSFR;
- select for stellar populations with ages $3.5\,{\rm Myr} \lesssim \langle \text{Stellar Age} \rangle_{\rm LyC} \lesssim 10\,{\rm Myr}$;
- include a proxy diagnostic for neutral gas content as well as the state of the ISM.

This framework can successfully identify samples of galaxies that are highly enriched with LyC leakers. By applying our method to existing indirect $f_{\rm esc}$ diagnostics, we can predict the reasons why each diagnostic will be successful (or fail) and why. For example, we find that high O_{32} is a necessary but insufficient criterion for high escape fractions, due to the fact that it traces high sSFR, but also selects for galaxies with dense, dusty ISMs with stellar populations that are too young to disrupt it. Observed UV slope, β, is empirically found to marginally or strongly satisfy all three criteria for $Z/Z_\odot > 0.01$. It is thus a relatively good diagnostic for the escape fraction. Similarly, E(B − V) is found to satisfy 2/3 criteria and thus traces $f_{\rm esc}$ reasonably well, albeit with significant scatter. In contrast, galaxy properties such as EW(Hβ), $\Sigma_{\rm SFR}$, $\Sigma_{\rm sSFR}$, $M_{\rm UV}$, sSFR, and M_* are all found to be poor indicators of $f_{\rm esc}$ if used in isolation as they satisfy one or none of our criteria.

We can also satisfy all three criteria with multi-dimensional diagnostics. Selecting galaxies with $\log_{10}({\rm H}\beta/{\rm L}_{1500}) < 1.6$ while combining β and E(B − V) (with Equation 3) produces a sample of galaxies of which 67% have $f_{\rm esc} > 5\%$, accounting for 62% of all such galaxies in our data-set. Similarly, we have constructed a generalized linear model that utilizes spectral properties of galaxies to quantitatively predict $f_{\rm esc}$ (see Equation 4). Applying our model to high-redshift Lyα emitters observed with JWST, we find LyC escape fractions less than or equal to the observationally estimated Lyα escape fractions. Our results suggest that bright Lyα emitters tend to have LyC escape fractions $\lesssim 10\%$.

Though our framework for the physics of indirect estimators of LyC escape has been tested by a robust data-set of mock observations, we recognise that such mock data are dependent on the nebular emissivity and dust absorption models used, inviting future work on better-resolved cosmological simulations with more realistic subgrid physics. The hardest of our three criteria to select for is the correct mean stellar population age, suggesting that this needs to be explored in the context of, for example, SED fitting. Nevertheless, our framework highlights the potential of existing and future JWST data to understand the physics of LyC escape and cosmological reionization.

ACKNOWLEDGEMENTS

N.C. and H.K. thank Mengtao Tang for kindly sharing data on CEERS Lyα emitters. N.C. and H.K. also thank Jonathan Patterson for helpful support on Glamdring throughout the project.

N.C. acknowledges support from the Science and Technology Facilities Council (STFC) for a Ph.D. studentship.

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]:

Nicholas Choustikov: Conceptualization; Formal analysis; Writing - original draft; Methodology. **Harley Katz**: Conceptualization; Formal analysis; Writing - original draft; Methodology. **Aayush Saxena**: Conceptualization; Writing - review and editing. **Alex J. Cameron**: Conceptualization; Writing - review and editing. **Julien Devriendt**: Resources; Supervision; Writing - review and editing. **Adrianne Slyz**: Resources; Supervision; Writing - review and editing. **Joki Rosdahl**: Writing - review and editing. **Jeremy Blaizot**: Writing - review and editing. **Leo Michel-Dansac**: Writing - review and editing.

DATA AVAILABILITY

The SPHINX[20] data used in this article will be made public in Summer 2023.

REFERENCES

Barrow K. S. S., Robertson B. E., Ellis R. S., Nakajima K., Saxena A., Stark D. P., Tang M., 2020, ApJ, 902, L39
Bassett R., et al., 2019, MNRAS, 483, 5223
Becker G. D., D'Aloisio A., Christenson H. M., Zhu Y., Worseck G., Bolton J. S., 2021, MNRAS, 508, 1853
Bosman S. E. I., et al., 2022, MNRAS, 514, 55
Bouwens R. J., et al., 2022a, arXiv e-prints, p. arXiv:2211.02607
Bouwens R. J., Illingworth G. D., van Dokkum P. G., Oesch P. A., Stefanon M., Ribeiro B., 2022b, ApJ, 927, 81
Brinchmann J., 2022, arXiv e-prints, p. arXiv:2208.07467
Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, MNRAS, 351, 1151
Bunker A. J., et al., 2023, arXiv e-prints, p. arXiv:2302.07256
Cameron A. J., et al., 2023, arXiv e-prints, p. arXiv:2302.04298
Chisholm J., et al., 2018, A&A, 616, A30
Chisholm J., Prochaska J. X., Schaerer D., Gazagnes S., Henry A., 2020, MNRAS, 498, 2554
Chisholm J., et al., 2022, MNRAS, 517, 5104
Cullen F., et al., 2023, MNRAS, 520, 14
Curtis-Lake E., et al., 2022, arXiv e-prints, p. arXiv:2212.04568
Davies F. B., et al., 2018, ApJ, 864, 142
Dijkstra M., 2017, Saas-Fee Lecture Notes: Physics of Lyman Alpha Radiative Transfer, doi:10.48550/ARXIV.1704.03416, https://arxiv.org/abs/1704.03416
Endsley R., Stark D. P., Whitler L., Topping M. W., Chen Z., Plat A., Chisholm J., Charlot S., 2022, arXiv e-prints, p. arXiv:2208.14999
Faisst A. L., 2016, ApJ, 829, 99
Fan X., et al., 2006, AJ, 132, 117
Ferland G. J., et al., 2017, Rev. Mex. Astron. Astrofis., 53, 385
Finkelstein S. L., et al., 2022, arXiv e-prints, p. arXiv:2211.05792
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, The Astrophysical Journal Supplement Series, 260, 1
Flury S. R., et al., 2022b, ApJ, 930, 126
Garel T., Blaizot J., Rosdahl J., Michel-Dansac L., Haehnelt M. G., Katz H., Kimm T., Verhamme A., 2021, MNRAS, 504, 1902
Green J. C., et al., 2012, ApJ, 744, 60
Greig B., Mesinger A., Bañados E., 2019, MNRAS, 484, 5094
Harikane Y., et al., 2023, ApJS, 265, 5
Hassan S., Davé R., McQuinn M., Somerville R. S., Keating L. C., Anglés-Alcázar D., Villaescusa-Navarro F., Spergel D. N., 2022, ApJ, 931, 62
Hayes M. J., Scarlata C., 2023, arXiv e-prints, p. arXiv:2303.03160
Heinrich C., Hu W., 2021, Phys. Rev. D, 104, 063505
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., Chisholm J., Worseck G., Guseva N. G., Schaerer D., Prochaska J. X., 2022, MNRAS, 515, 2864
Jaskot A. E., Oey M. S., 2013, ApJ, 766, 91
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., 2022a, MNRAS, 510, 5603
Katz H., et al., 2022b, Monthly Notices of the Royal Astronomical Society, 515, 4265
Katz H., et al., 2022c, Monthly Notices of the Royal Astronomical Society, 518, 270
Katz H., et al., 2023, MNRAS, 518, 592
Kawamata R., Ishigaki M., Shimasaku K., Oguri M., Ouchi M., Tanigawa S., 2018, ApJ, 855, 4
Keating L. C., Weinberger L. H., Kulkarni G., Haehnelt M. G., Chardin J., Aubert D., 2020, MNRAS, 491, 1736
Kennicutt Robert C. J., 1998, ARA&A, 36, 189

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

Kewley L. J., Dopita M. A., 2002, ApJS, 142, 35
Kimm T., Cen R., 2014, ApJ, 788, 121
Kimm T., Cen R., Devriendt J., Dubois Y., Slyz A., 2015, Towards simulating star formation in turbulent high-z galaxies with mechanical supernova feedback, doi:10.48550/ARXIV.1501.05655, https://arxiv.org/abs/1501.05655
Kimm T., Katz H., Haehnelt M., Rosdahl J., Devriendt J., Slyz A., 2017, MNRAS, 466, 4826
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
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
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, The Astrophysical Journal, 704, 1640
Leitherer C., et al., 1999, ApJS, 123, 3
Leitherer C., Hernandez S., Lee J. C., Oey M. S., 2016, ApJ, 823, 64
Lequeux J., Peimbert M., Rayo J. F., Serrano A., Torres-Peimbert S., 1979, A&A, 80, 155
Ma X., Quataert E., Wetzel A., Hopkins P. F., Faucher-Giguère C.-A., Kereš D., 2020, MNRAS, 498, 2001
Maji M., et al., 2022, A&A, 663, A66
Mason C. A., Treu T., Dijkstra M., Mesinger A., Trenti M., Pentericci L., de Barros S., Vanzella E., 2018, ApJ, 856, 2
Matthee J., Mackenzie R., Simcoe R. A., Kashino D., Lilly S. J., Bordoloi R., Eilers A.-C., 2022a, arXiv e-prints, p. arXiv:2211.08255
Matthee J., et al., 2022b, MNRAS, 512, 5960
Mauerhofer V., Verhamme A., Blaizot J., Garel T., Kimm T., Michel-Dansac L., Rosdahl J., 2021, A&A, 646, A80
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, Astronomy & Astrophysics, 635, A154
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
Nakajima K., Ellis R. S., Robertson B. E., Tang M., Stark D. P., 2020, ApJ, 889, 161
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
Pentericci L., et al., 2011, ApJ, 743, 132
Planck Collaboration et al., 2020, A&A, 641, A6
Priestley F. D., Chawner H., Matsuura M., De Looze I., Barlow M. J., Gomez H. L., 2021, MNRAS, 500, 2543
Ramambason L., et al., 2022, A&A, 667, A35
Razoumov A. O., Sommer-Larsen J., 2010, ApJ, 710, 1239
Rémy-Ruyer A., et al., 2014, A&A, 563, A31
Rezaee S., et al., 2022, arXiv e-prints, p. arXiv:2208.12371
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
Saha K., et al., 2020, Nature Astronomy, 4, 1185
Saldana-Lopez A., et al., 2022, A&A, 663, A59
Salim S., et al., 2007, ApJS, 173, 267
Saxena A., et al., 2022, MNRAS, 517, 1098
Saxena A., et al., 2023, arXiv e-prints, p. arXiv:2302.12805
Schaerer D., 2003, A&A, 397, 527
Schaerer D., et al., 2022a, A&A, 658, L11
Schaerer D., Marques-Chaves R., Barrufet L., Oesch P., Izotov Y. I., Naidu R., Guseva N. G., Brammer G., 2022b, A&A, 665, L4
Schaller G., Schaerer D., Meynet G., Maeder A., 1992, A&AS, 96, 269
Sharma M., Theuns T., Frenk C., Bower R. G., Crain R. A., Schaller M., Schaye J., 2017, MNRAS, 468, 2176
Sparre M., Hayward C. C., Feldmann R., Faucher-Giguère C.-A., Muratov A. L., Kereš D., Hopkins P. F., 2017, MNRAS, 466, 88
Stanway E. R., Eldridge J. J., 2018, Monthly Notices of the Royal Astronomical Society, 479, 75
Stanway E. R., Eldridge J. J., Becker G. D., 2016, MNRAS, 456, 485
Stark D. P., Ellis R. S., Chiu K., Ouchi M., Bunker A., 2010, MNRAS, 408, 1628
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
Tang M., et al., 2023, arXiv e-prints, p. arXiv:2301.07072
Topping M. W., Stark D. P., Endsley R., Plat A., Whitler L., Chen Z., Charlot S., 2022, ApJ, 941, 153
Trebitsch M., Blaizot J., Rosdahl J., Devriendt J., Slyz A., 2017, MNRAS, 470, 224
Trebitsch M., et al., 2021, A&A, 653, A154
Tremonti C. A., et al., 2004, ApJ, 613, 898
Treu T., et al., 2022, ApJ, 935, 110
Vanzella E., et al., 2010, ApJ, 725, 1011
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
Wang B., et al., 2021, ApJ, 916, 3
Weisz D. R., et al., 2012, ApJ, 744, 44
Witten C. E. C., Laporte N., Katz H., 2023, ApJ, 944, 61
Xu H., Wise J. H., Norman M. L., Ahn K., O'Shea B. W., 2016, ApJ, 833, 84
Yang H., et al., 2017, ApJ, 844, 171
Yusef-Zadeh F., Morris M., White R. L., 1984, ApJ, 278, 186
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, The Astrophysical Journal, 578, 33
Ďurovčíková D., Katz H., Bosman S. E. I., Davies F. B., Devriendt J., Slyz A., 2020, MNRAS, 493, 4256

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