Two Modes of LyC Escape From Bursty Star Formation: Implications for [C II] Deficits and the Sources of Reionization

Harley Katz, Aayush Saxena, Joki Rosdahl, Taysun Kimm, Jeremy Blaizot, Thibault Garel, Leo Michel-Dansac, Martin Haehnelt, Richard S. Ellis, Laura Penterrici, et al.

Two Modes of LyC Escape From Bursty Star Formation: Implications for [C II] Deficits and the Sources of Reionization

Harley Katz[1]★, Aayush Saxena[2], Joki Rosdahl[3], Taysun Kimm[4], Jeremy Blaizot[3], Thibault Garel[5], Leo Michel-Dansac[3], Martin Haehnelt[6], Richard S. Ellis[2], Laura Penterrici[7], Julien Devriendt[1], & Adrianne Slyz[1]

[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]*Univ Lyon, Univ Lyon1, Ens de Lyon, CNRS, Centre de Recherche Astrophysique de Lyon UMR5574, F-69230, Saint-Genis-Laval, France*
[4]*Department of Astronomy, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea*
[5]*Observatoire de Genève, Université de Genève, Chemin Pegasi 51, 1290 Versoix, Switzerland*
[6]*Kavli Institute for Cosmology and Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, United Kingdom*
[7]*INAF - Osservatorio Astronomico di Roma, via Frascati 33, 00078, Monteporzio Catone, Italy*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

We use the SPHINX^{20} cosmological radiation hydrodynamics simulation to study how Lyman Continuum (LyC) photons escape from galaxies and the observational signatures of this escape. We define two classes of LyC leaker: Bursty Leakers and Remnant Leakers, based on their star formation rates (SFRs) that are averaged over 10 Myr (SFR_{10}) or 100 Myr (SFR_{100}). Both have $f_{esc} > 20\%$ and experienced an extreme burst of star formation, but Bursty Leakers have $SFR_{10} > SFR_{100}$, while Remnant Leakers have $SFR_{10} < SFR_{100}$. The maximum SFRs in these bursts were typically $\sim$ 100 times greater than the SFR of the galaxy prior to the burst, a rare 2σ outlier among the general high-redshift galaxy population. Bursty Leakers are qualitatively similar to ionization-bounded nebulae with holes, exhibiting high ionization parameters and typical HII region gas densities. Remnant Leakers show properties of density-bounded nebulae, having normal ionization parameters but much lower HII region densities. Both types of leaker exhibit $[CII]_{158\mu m}$ deficits on the [CII]-SFR_{100} relation, while only Bursty Leakers show deficits when SFR_{10} is used. We predict that [CII] luminosity and SFR indicators such as Hα and $M_{1500Å}$ can be combined to identify both types of LyC leaker and the mode by which photons are escaping. These predictions can be tested with [CII] observations of known $z = 3 - 4$ LyC leakers. Finally, we show that leakers with $f_{esc} > 20\%$ dominate the ionizing photon budget at $z \gtrsim 7.5$ but the contribution from galaxies with $f_{esc} < 5\%$ becomes significant at the tail-end of reionization.

Key words: galaxies: evolution, galaxies: formation, galaxies: high-redshift, stars: formation, ISM: evolution, ISM: general

1 INTRODUCTION

While it is well established that the Universe completed reionization sometime in the redshift interval of $z = 5 - 7$ (Fan et al. 2006; Kulkarni et al. 2019), with Lyα forest data suggesting islands of neutral gas extending to $z \sim 5.3$ (Bosman et al. 2022), uncertainties remain on the timing of the onset of reionization, the neutral fraction history, and the sources responsible for the ionizing photons. Empirical constraints on all three are important for understanding the formation of the first stars, metal and dust production in the early Universe, the impact of an emerging UV background on galaxy formation, the visibility of various emission lines (e.g. Lyα) at high-redshift, as well as many other characteristics of galaxy formation at cosmic dawn.

Analytic models of reionization that rely on a star formation rate density, an ionizing photon emissivity per unit star formation (ξ_{ion}), and a Lyman Continuum (LyC) escape fraction (f_{esc}) are often used (e.g. Madau et al. 1999; Robertson et al. 2013) to model the evolution of the cosmic neutral fraction and constrain the reionization history. While the star formation rate density as a function of redshift can be constrained with observations of the UV luminosity function (e.g. Bouwens et al. 2021b), and estimates on ξ_{ion} can be adopted from theoretical stellar evolution models (e.g. Leitherer et al. 1999) or inferred from observations (e.g. Stark et al. 2015; Bouwens et al. 2016), the majority of the uncertainty in these models stems from our inability to constrain f_{esc}. This is due to the fact that f_{esc} cannot be calculated analytically as it is subject to the detailed properties of the interstellar medium (ISM) and the distribution of the sources within.

Because of the intervening intergalactic medium (IGM), it is nearly impossible to directly observe escaping LyC radiation during the epoch of reionization. Rather observational studies of f_{esc} often target lower redshift analogues of high-redshift galaxies at $z \sim 0$ (e.g. Flury et al. 2022), $z \sim 3$ (e.g. Fletcher et al. 2019), and more recently at $z \sim 1.5$ (e.g. Saha et al. 2020). The limited numbers of observed LyC leakers has historically inhibited a detailed study of their galaxy properties in relation to the general galaxy population.

★ E-mail: harley.katz@physics.ox.ac.uk

Recent large scale surveys are now making this possible (e.g. Flury et al. 2022); however, the total number of confirmed LyC leakers is still only ∼ 100. Thus any biases in selection function and low number statistics may still contribute significantly to any observed trends between galaxy properties and LyC leakage. For this reason, numerical simulations that resolve the ISM of galaxies remain an invaluable tool for understanding the physics that controls LyC leakage and the relation to galaxy properties (e.g. Katz et al. 2020).

Zackrisson et al. (2013) envisioned two mechanisms by which LyC photons escape from galaxies: a radiation bounded nebula with holes, where the escape fraction is set by the covering fraction of the holes, and a density bounded nebula, where the escape fraction is set by the optical depth of the nebula. These scenarios are not necessarily mutually exclusive; nevertheless, they demonstrate that there exist certain observational signatures that differentiate these mechanisms, especially when dust is included. High-resolution cosmological simulations seem to indicate that $f_{\rm esc}$ is a feedback-regulated quantity (e.g. Trebitsch et al. 2017; Kimm et al. 2017; Rosdahl et al. 2018; Barrow et al. 2020). Bursts of star formation that generate a large quantity of LyC photons can heat and reduce the density of the ISM, occasionally creating holes in the gas distribution where LyC photons escape. This process is then followed by supernova (SN) feedback that can clear even larger channels or destroy the structure of the ISM entirely. Hence, numerical simulations also predict two modes of $f_{\rm esc}$. In the first scenario, early stellar feedback in the form of ionizing radiation and perhaps SNe from the most massive stars creates the first channels through which LyC photons can escape. Because the stellar populations are still young, the LyC production efficiency remains high (e.g. Eldridge et al. 2008; Stanway et al. 2016). In the second scenario, later stellar feedback in the form of SNe create super bubbles, clearing out further channels for LyC photons to escape. However, because the LyC production efficiency drops significantly as a function of the age of the stellar population, it is not clear which phase results in more LyC photons leaking into the IGM. Once again, these two modes are not mutually exclusive and it is often the case that one is followed by the other (e.g. Wise & Cen 2009; Kimm & Cen 2014; Trebitsch et al. 2017; Kimm et al. 2017; Rosdahl et al. 2018). Rosdahl et al. (2018) demonstrate this in their Figure 12, where, for a particular galaxy, they show a 50 Myr time series of the evolution of $f_{\rm esc}$, the LyC luminosity, and the structure of the ISM after a strong burst of star formation. The two modes of $f_{\rm esc}$ can also be seen sequentially in the multi-peaked distribution of $f_{\rm esc}$ over time in Figure 8 of Kimm et al. (2017).

While numerical simulations have predicted the mechanisms by which LyC photons escape galaxies during reionization, they have also demonstrated that the process is inefficient — at any given time, only a very small fraction of galaxies exhibit high $f_{\rm esc}$ (e.g. Paardekooper et al. 2015) and only a few percent of the ionizing photons produced by galaxies escape into the IGM. Hence it is not surprising that observers struggle with finding large populations of LyC leakers (e.g. Leitet et al. 2013). Similarly, simulations also show that the escape fraction is highly viewing angle-dependent (e.g. Cen & Kimm 2015). Thus, even if a galaxy is a LyC leaker, the probability of directly observing LyC photons can significantly decrease due to geometrical effects.

Of the numerous methods that have been suggested as indirect tracers of LyC leakage, the vast majority rely on emission lines at UV or optical wavelengths. These include high O_{32} ratios (e.g. Izotov et al. 2018), Lyα peak separations (e.g. Verhamme et al. 2015, 2017), SiII deficits (Wang et al. 2019), MgII doublet flux ratios (Chisholm et al. 2020), strong C IV emission (Schaerer et al. 2022; Saxena et al. 2022a), the combination of Hβ and UV slope (Zackrisson et al. 2013), etc. Because such lines are subject to dust attenuation, and in the case of resonant lines, the gas distribution of the relevant species, line strengths and shapes might also be highly orientation angle-dependent. In contrast, IR emission lines are not as sensitive to geometrical effects. Katz et al. (2020) explored the use of the $[\rm CII]_{158\mu m}$ and $[\rm OIII]_{88\mu m}$ lines as probes of $f_{\rm esc}$ because numerical simulations show that at high-redshift [CII] emission correlates with the presence of neutral gas (Pallottini et al. 2017; Katz et al. 2019; Lupi et al. 2020) while [OIII] emission traces star formation and feedback (e.g. Katz et al. 2019; Lupi et al. 2020). Thus the ratio of the two emission lines is expected to provide insight into $f_{\rm esc}$ (Inoue et al. 2016; Katz et al. 2020, 2022b). While there are now more than 100 galaxies at high-redshift with [CII] observations (e.g. Le Fèvre et al. 2020; Bouwens et al. 2021b), observations of [OIII] at $z \gtrsim 4$ remain limited (e.g. Carniani et al. 2020). The lack of known [OIII] emitters motivates the study of the prospect of using only [CII] as a means of both identifying potential leakers and the mechanisms by which LyC photons escape. Since the vast majority of known [CII] emitters are at $z < 6$, they can, in principle, be followed up with direct observations in the LyC bands (although IGM transmission can be a problem even at intermediate redshifts (Inoue & Iwata 2008)). Furthermore, with the recent launch of the James Webb Space Telescope (JWST), there is potential to follow up many of the galaxies observed in [CII] with observations in the rest-frame UV and optical. For this reason, in this work, we use cosmological simulations to study how [CII] combined with various star formation rate indicators at intermediate and high redshift can be used to identify LyC leakers and differentiate the mechanisms by which LyC photons escape.

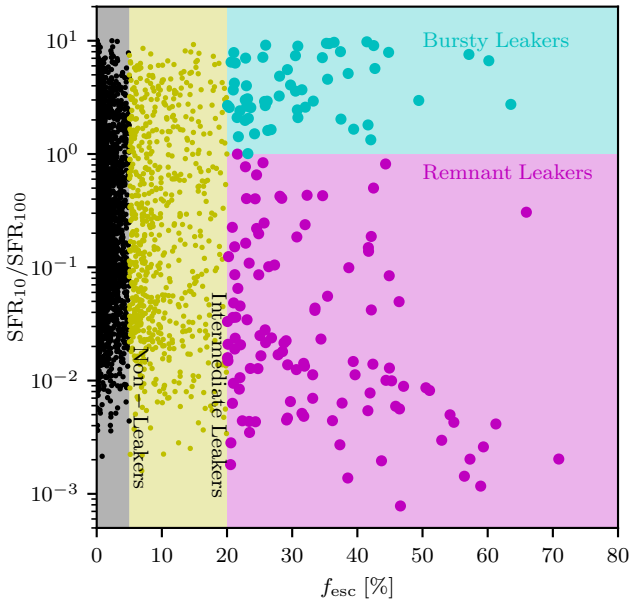

Figure 1. Ratio of 10 Myr-averaged SFR to 100 Myr-averaged SFR versus LyC escape fraction for SPHINX^{20} galaxies at $z = 4.64$. The different coloured regions represent the different classifications, as labelled on the diagram.

This work is organized as follows. In Section 2 we briefly review the SPHINX^{20} simulation that resolves 3×10^4 galaxies at $z \leq 6$. In Section 3 we discuss the utility of [CII] as an $f_{\rm esc}$ indicator and highlight synergies between ALMA and other observational facilities that can be used to constrain the properties of the high-redshift ISM. Finally, in Section 4, we present our discussion and conclusions.

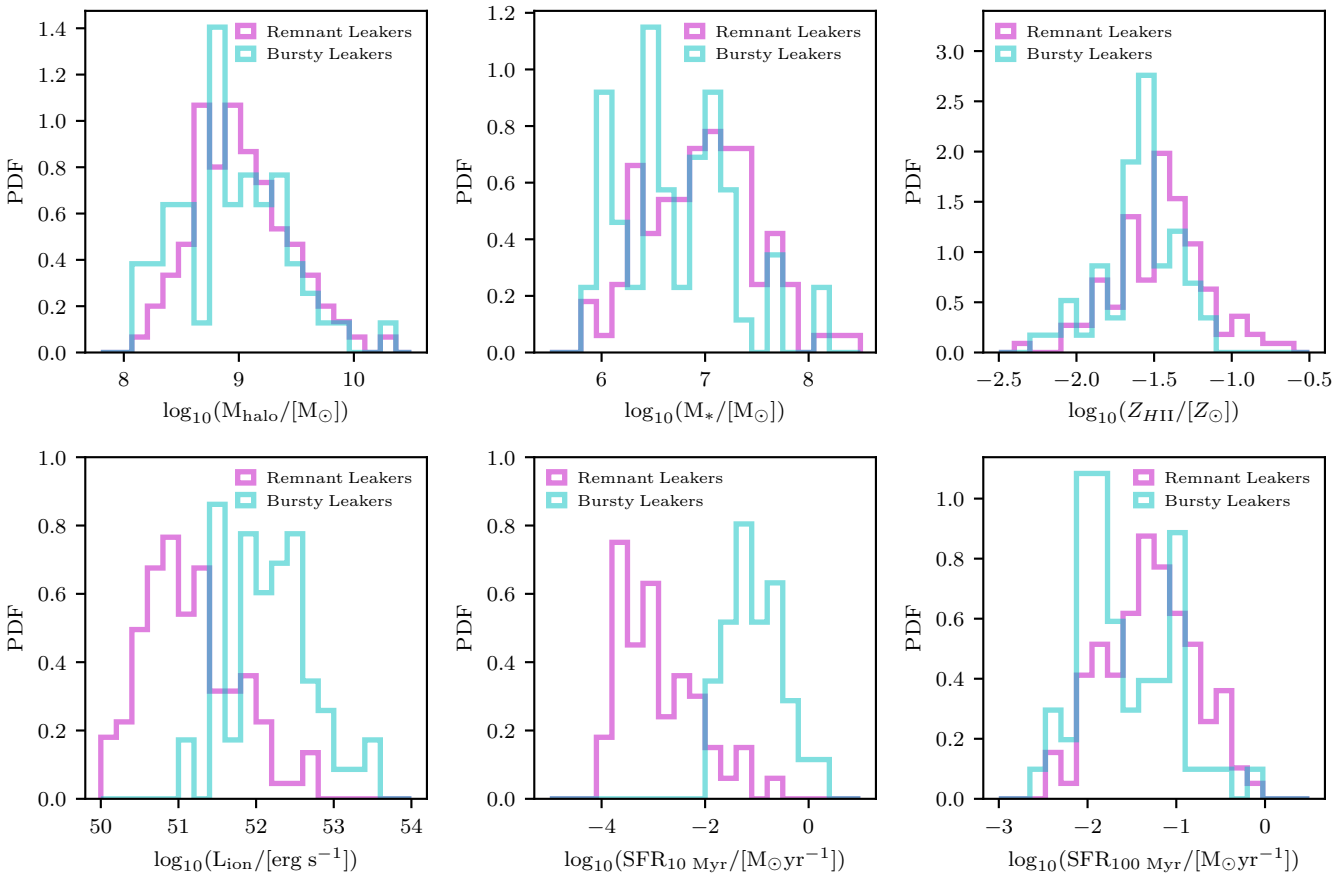

Figure 2. Probability distribution functions (PDFs) comparing various galaxy and halo properties for the Bursty Leakers (cyan) with the Remnant Leakers (magenta) at $z = 4.64$. We compare halo mass, stellar mass, HII region metallicity, total intrinsic ionizing luminosity, SFR_{10}, and SFR_{100}. Besides SFR_{10} and ionizing luminosity, the two classifications of leakers have very similar properties because they represent similar galaxies in different evolutionary phases.

2 METHOD

We employ the SPHINX^{20} simulation (Rosdahl et al. 2022), the largest volume run of all simulations in the SPHINX suite of cosmological radiation hydrodynamics simulations (Rosdahl et al. 2018; Katz et al. 2020, 2021, 2022b). SPHINX^{20} was run with the radiation hydrodynamics extension (Rosdahl et al. 2013; Rosdahl & Teyssier 2015) of the adaptive mesh refinement code RAMSES (Teyssier 2002). The simulation incorporates state-of-the-art models for star formation (Kimm et al. 2017) and stellar feedback (Kimm et al. 2015) and employs the variable speed of light approximation (Katz et al. 2017) to capture the motion of ionization fronts through the ISM and IGM. With a maximum spatial resolution of ~ 10 pc, the simulation is able to model a multi-phase ISM structure and the low density channels through which LyC photons escape. The large volume of 20^3 cMpc^3 allows us to sample a wide range of galaxy properties, resolving haloes of 10^8 $M_\odot$ by 400 dark matter particles. Full details of the physics included in the simulations is described in detail in Rosdahl et al. (2022) and our methods for calculating line emission, in particular [CII], as well as any minor changes between SPHINX^{10} and SPHINX^{20} are described in Katz et al. (2022b). Escape fractions are calculated in post-processing by using Monte Carlo radiative transfer (RASCAS, Michel-Dansac et al. 2020) to follow 912Å photons from star particles to the virial radius of each halo (see Katz et al. 2022b). Not all simulations use the same radius to measure the escape fraction; however, our choice is consistent with all other work on the SPHINX simulations (Rosdahl et al. 2022). For each galaxy 10^7 photon packets are distributed among the star particles with initial positions randomly sampled from a multinomial distribution based on the location and ionizing emissivity. We measure escape fractions along individual lines of sight as well as the angle-averaged values. By measuring f_{esc} in two ways, we can better constrain the impact of anisotropic leakage in prospective observations. However, we note that the angle-averaged value is the important quantity for measuring the impact of individual galaxies on reionization.

In this work, we primarily study the $z = 4.64$ snapshot, the final snapshot of the simulation, for which we calculated [CII] emission for the nearly 30,000 galaxies with halo masses $\geq 10^8$ $M_\odot$. There are significantly more galaxies observed in [CII] at $z < 6$ and, due to the neutral IGM at high redshift, direct LyC detections are only possible at $z < 6$. This motivates our study of the lowest redshift snapshot of the simulation. However, for our results to matter for reionization, we must also show that the trends between f_{esc} and galaxy properties at $z = 4.64$ also hold for $z \geq 6$. For this reason, we will then link the results to $z = 6$ where the simulation resolves similar numbers of galaxies.

Emission lines are calculated on a cell-by-cell basis in post-processing by running CLOUDY models (Ferland et al. 2017) on all cells in the simulation based on their gas density, metallicity, temperature, dust content, and local radiation field. We have adopted the solar abundance pattern model from Katz et al. (2022b). Due to the large volume and high resolution of the simulation, at $z = 4.64$,

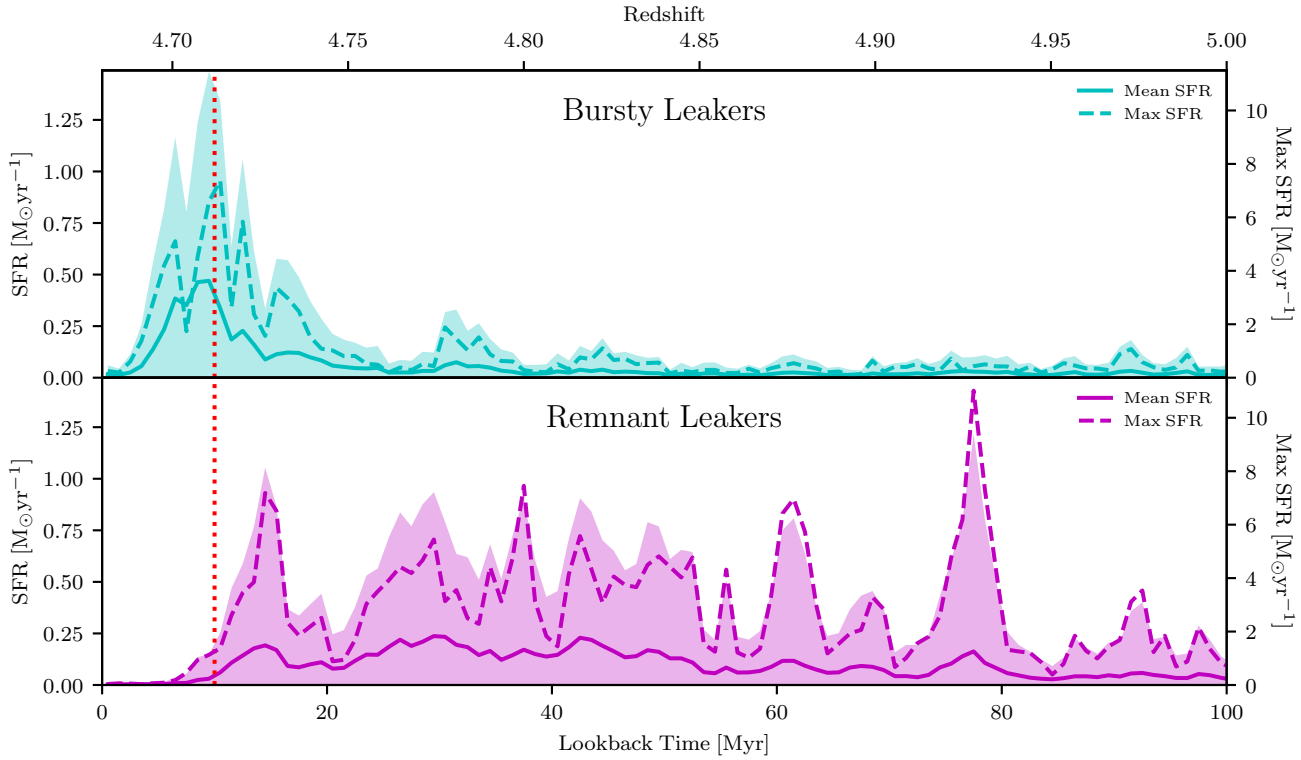

Figure 3. Star formation histories of Bursty Leakers (top) and Remnant Leakers (bottom) for the 100 Myr prior to $z = 4.64$. The solid line and shaded region show the mean and 1σ scatter of the SFR as a function of time with values as given on the y-axis on the left side of the plot. The dashed lines represent the maximum SFR among the population as a function of time and correspond to the y-axis on the right side of the plot. The vertical red dotted line denotes 10 Myr. We include all star particles within the virial radius of each halo and assume a Δt of 1 Myr to compute the SFR. As populations, the Bursty Leakers exhibit very little star formation beyond 20 Myr since $z = 4.64$, while the remnant leakers exhibit almost no star formation in the previous 10 Myr since $z = 4.64$.

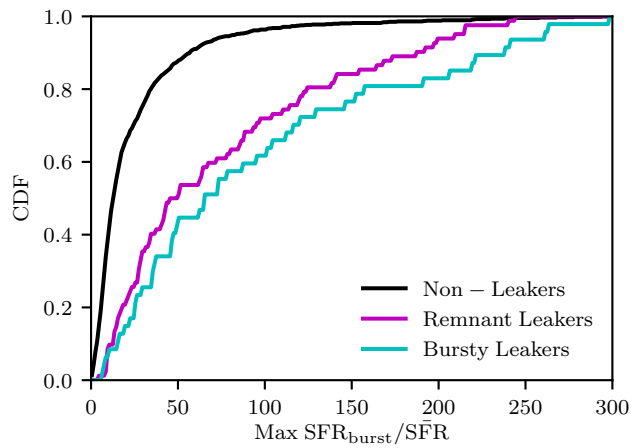

Figure 4. Cumulative distribution function of the maximum SFR in the burst to the typical SFR of the galaxy prior to the burst for Bursty leakers (cyan), Remnant leakers (magenta), and non-leakers (black.)

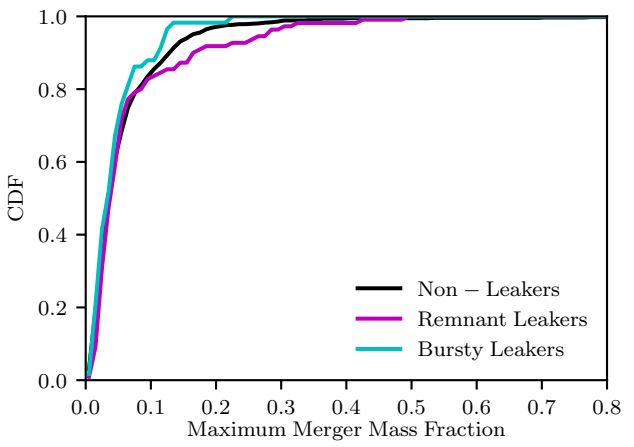

Figure 5. Cumulative distribution function of the maximum fractional change in dark matter mass in the 100 Myr time period before $z = 4.64$ for Bursty leakers (cyan), Remnant leakers (magenta), and non-leakers (black.)

halo masses range from 10^8 M$_\odot$ – $10^{11.7}$ M$_\odot$, stellar masses span $10^{2.5}$ M$_\odot$ – $10^{10.5}$ M$_\odot$, SFRs (averaged over 10 Myr) vary between 0 M$_\odot$yr^{-1} – $10^{1.9}$ M$_\odot$yr^{-1}, and finally HII region metallicities range between $10^{-3.3}$ $Z_\odot$ – $10^{0.1}$ $Z_\odot$. We expect these values to be rather typical of any 20^3 cMpc^3 volume as the initial conditions in the simulation were chosen so that the $z = 6$ halo mass function was most representative of the average of multiple random realizations.

We classify SPHINX^{20} galaxies into four different groups based on their angle-averaged LyC escape fractions and star formation histories: Non-Leakers, Intermediate Leakers, Bursty Leakers, and Remnant Leakers. Non-leakers are defined as galaxies with angle-averaged $f_{\rm esc} \leq 5\%$, Intermediate Leakers have $5\% < f_{\rm esc} < 20\%$, Bursty Leakers have $f_{\rm esc} \geq 20\%$ and their star formation rate (SFR) averaged over the previous 10 Myr (SFR$_{10}$) is greater than their SFR averaged over the previous 100 Myr (SFR$_{100}$), and finally, Remnant

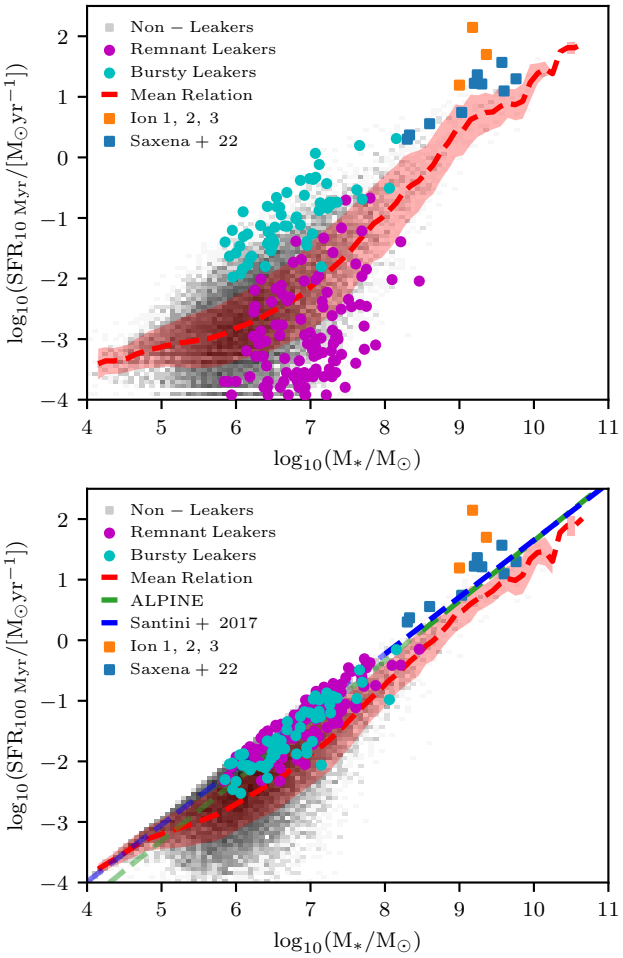

Figure 6. Star formation main sequence using SFR_{10} (top) or SFR_{100} (bottom) for SPHINX^{20} galaxies at $z = 4.64$. Non-Leakers are shown as the grey histogram while Bursty and Remnant Leakers are shown in cyan and magenta, respectively. The dashed red line and shaded region represent the mean and 1σ scatter for SPHINX^{20} galaxies. In the bottom panel, we compare our simulated main sequence with observational constraints from Santini et al. (2017) and Khusanova et al. (2021) (ALPINE) and find good agreement, especially at high stellar masses. The lines become translucent at low stellar masses where observational inferences have been extrapolated. We also show known $z = 3 - 3.5$ LyC leakers from Saxena et al. (2022b) and *Ion 1, 2* and *3* from Vanzella et al. (2012, 2016, 2018) to demonstrate that all fall above the main sequence. Note that there are no direct constraints on SFR_{10} and SFR_{100} for the observed samples as they are derived from the SED and UV, so we show them only for qualitative comparison.

Leakers have $f_{esc} \geq 20\%$ and $SFR_{100} \geq SFR_{10}$[1]. A schematic of this distribution can be seen in Figure 1. It should be noted that neither SFR_{10} nor SFR_{100} are observable quantities. They must be inferred from observations, for example by fitting the SED or comparing with IR or line emission, with each indicator potentially probing star formation on different time scales (e.g. Calzetti et al. 2007; Kennicutt & Evans 2012). We will show that the trends we find for Bursty and Remnant leakers should persist if we replace SFR_{10} and SFR_{100} with SFR indicators that are sensitive to star formation on very short and slightly longer time scales. Hα luminosity and infrared luminosity are such indicators (e.g. Kennicutt 1992).

Our classification scheme is related to the different modes of f_{esc}. Bursty Leakers are subject to strong photoionization feedback as well as SNe feedback from the most massive stars. The Remnant Leakers have had a burst of star formation sometime in the recent past compared to their current SFR. It is likely that the SNe feedback from the previous burst disrupted the ISM enough to shut down star formation. Interestingly, the galaxies with $SFR_{10}/SFR_{100} < 2\times10^{-3}$ are all Remnant Leakers with $f_{esc} > 20\%$. Thus the first mode is high f_{esc} due to radiation and early SN feedback while the latter mode is after significant SN feedback has occurred. Note that our two modes are distinct from whether the radiation leaks through holes or more uniformly due to $\tau < 1$ (e.g. Zackrisson et al. 2013); however, there are parallels that we discuss below.

The cut at an escape fraction of 20% is arbitrary and about twenty times[2] as much as the global (i.e. ionizing luminosity-weighted) escape fraction at $z = 4.64$ in SPHINX^{20}; however, our results are not fundamentally different if we employ other thresholds[3]. In total, at $z = 4.64$, selecting only galaxies with [CII] luminosities $> 10^2 L_\odot$ and $SFR_{10} > 10^{-4}\ M_\odot yr^{-1}$ we find 58 Bursty Leakers, 111 Remnant Leakers, 695 Intermediate Leakers, and 16,085 Non-Leakers.

3 RESULTS

3.1 Galaxy Property Comparison of Bursty and Remnant Leakers

In Figure 2 we compare various galaxy properties of the two classes of leakers. Bursty and Remnant leakers have very similar distributions of halo mass, stellar mass, and HII region[4] metallicity. Most of the leakers in the simulation have halo masses of $\sim 10^9\ M_\odot$. More massive galaxies in SPHINX^{20} are less likely to be leakers. The 10 Myr-averaged SFRs of the Bursty Leakers are significantly greater than the Remnant Leakers, which is unsurprising given how the two populations are defined. More specifically, the median SFR_{10} of the Bursty Leakers is $0.056\ M_\odot yr^{-1}$, more than an order of magnitude greater than the median of $0.001\ M_\odot yr^{-1}$ for the Remnant Leakers. In contrast, both Bursty and Remnant Leakers exhibit very similar[5] 100 Myr-averaged SFRs, indicating that the total amount of star formation over the past 100 Myr is also similar for the two populations. Because of the significantly enhanced recent star formation, Bursty Leakers also exhibit much higher ionizing luminosities compared to remnant leakers. We find more than an order of magnitude difference between the median ionizing luminosities of the two groups of galaxies.

[1] In practice, there is a smooth transition between the Bursty and Remnant Leaker populations in terms of SFR_{10}/SFR_{100}, but for simplicity, we have introduced a hard cutoff.

[2] The luminosity-weighted escape fraction in SPHINX^{20} at $z = 4.64$ is 1%. This is consistent with the upper limits on the "average" escape fraction at $z = 3.3$ measured by Grazian et al. (2017) using ultra-deep U-band imaging.

[3] We have tested that all of the trends we present hold down to an f_{esc} threshold of 5% (i.e. removing the intermediate leaker bin completely). Similar thresholds were recently employed in observational studies of Lyα emitters at low-redshift (e.g. Naidu et al. 2022; Matthee et al. 2022). Using a threshold $\gg 20\%$ results in too few galaxies being classified as leakers. The number of leakers in each class decreases approximately linearly when varying the threshold f_{esc} from 5% to 20%.

[4] Throughout this paper, HII regions are defined as gas cells with HII fractions $> 50\%$, temperatures $< 10^{5.5}$ K, and gas densities $> 1\ cm^{-3}$.

[5] The median SFR_{100} for the Bursty leakers is slightly smaller (i.e. 70%) of the median value of the Remnant Leakers. This is well within the sampling uncertainty of the two distributions as the tension is $< 1\sigma$.

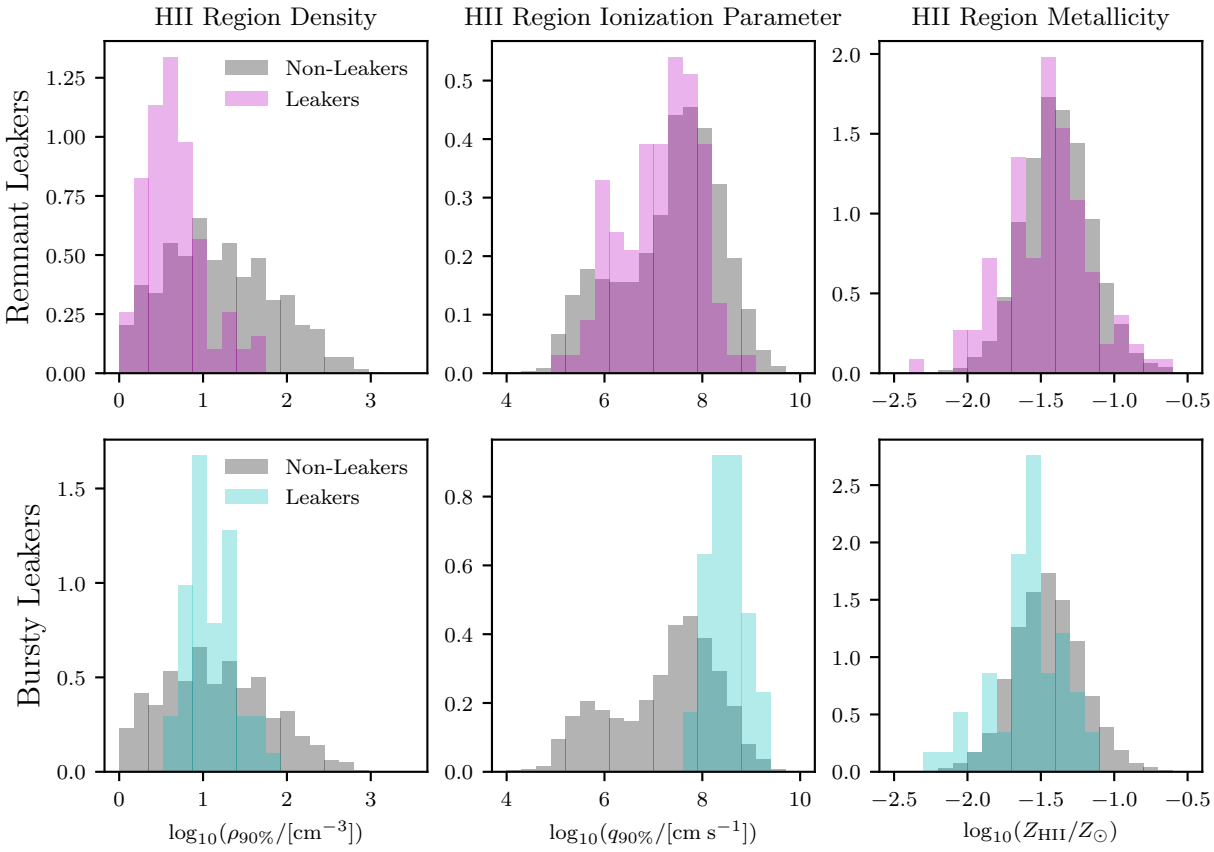

Figure 7. Comparison of ISM properties of Remnant Leakers (top, magenta) and Bursty Leakers (bottom, cyan) with control samples of Non-Leakers that were stellar mass-selected to exhibit the same distribution of stellar masses as each leaker population. From left to right, we compare probability distribution functions of 90th percentile HII region gas density ($\rho_{90\%}$), 90th percentile HII region ionization parameter ($q_{90\%}$, defined as ionizing photon flux divided by gas density), and HII region metallicity.

In summary, Bursty Leakers and Remnant Leakers exhibit very similar fundamental galaxy properties except for the fact that the burst of star formation responsible for the LyC leakage is currently underway or happened very recently for the Bursty Leakers, whereas it happened further in the past for Remnant Leakers.

We continue the comparison in Figure 3 where we show the mean and maximum SFRs as a function of time in the 100 Myr period prior to $z = 4.64$ for the Bursty Leakers (top) and the Remnant Leakers (bottom). In both populations of leakers, star formation has significantly decreased at $z = 4.64$, signifying the impact of strong stellar feedback (both radiative and SNe). The peak in the mean SFR for the Bursty Leakers occurred $\sim$ 8 Myr prior to $z = 4.64$; however for individual leakers in this population, the maximum SFR could have occurred up to 18 Myr prior to the snapshot. In contrast, the mean SFR of the Remnant Leakers remains relatively constant between 15 Myr and 80 Myr prior to $z = 4.64$ due to the fact that we are averaging the star formation histories of multiple galaxies that have had bursts at different times.

In both populations, the galaxies undergo extreme bursts of star formation which is highly correlated with having a high escape fraction. The primary difference between the two is the timing of the burst. The severity of the burst can be calculated by separating the SFR in the range $\pm$10 Myr before and after the maximum recorded SFR and comparing the SFR in the burst to the typical SFR of the galaxy in the 50 Myr time period prior[6] to the burst. In this experiment we measure the SFRs over 1 Myr intervals. For both Bursty and Remnant Leakers, the maximum SFRs in the burst are $\sim$ 90×-100× the typical SFR of the galaxy (as given by the median of the two populations). Rather than taking the maximum SFR in the burst, we can compare the mean SFR in the burst window and still find that the enhancement in SFR is $\sim$ 14× the typical SFR of the galaxy.

In order to estimate the probability of having such a strong burst of star formation, we construct a control sample of non-leakers by randomly selecting $\sim$ 1000 galaxies such that the stellar mass distribution of this control sample is the same as the combination of Bursty and Remnant Leakers. The maximum SFRs in the burst of control sample galaxies are typically only 13× the typical SFR and the median SFR in the burst window is only 3× the typical SFR. This can easily be seen in Figure 4 where we show the cumulative distribution function of the ratio of the maximum SFR in the burst to the typical SFR of the galaxy prior to the burst for Bursty leakers, Remnant leakers, and non-leakers. Here we see that the strength of the bursts for

[6] The choice of time window to measure the "typical" SFR of a galaxy is arbitrary. We have experimented with choosing windows that account for the periods both before an after the burst and our results are qualitatively similar. We have chosen to only measure the SFR prior to the burst in order to remove the time period where star formation is significantly suppressed due to the burst.

the Bursty and Remnant leakers are, in general, significantly stronger than what is observed for non-leakers. This confirms previous claims (e.g. Trebitsch et al. 2017; Kimm et al. 2017) that f_{esc} is feedback regulated and extreme bursts of star formation are required to disrupt the ISM enough so that LyC photons can efficiently escape. Among the control sample, only 6% of galaxies exhibit maximum SFRs as bursty as what is found for the leaker sample in the 100 Myr prior to $z = 4.64$. Similarly only 8% of galaxies in the control sample exhibit mean SFRs in the 20 Myr burst window that are $\sim 13\times$ the typical SFR of the galaxy. Both statistics indicate star bursts of this strength are rare.

Because mergers are more common at high redshift (Hopkins et al. 2010), it is interesting to consider whether mergers are responsible for the starbursts that lead to high f_{esc}. To assess this, we look in the 100 Myr window prior to $z = 4.64$ and record the maximum fractional change in dark matter mass (i.e. $(M_{DM,t+1} - M_{DM,t})/M_{DM,t}$) between simulation outputs. Galaxies with large fractional changes close to 1 will have undergone a near equal mass merger. In Figure 5 we show the cumulative distribution function of these maximum fractional changes for the Bursty leakers, Remnant leakers, and the randomly selected 1000 non-leakers. We do not find a significant excess of mergers in the Bursty or Remnant leakers compared to the non-leakers. For all three galaxy populations, the maximum merger mass ratio was 20:1 or less for $\sim 50\%$ indicating that a major merger is not required for high f_{esc}.

The fact that the Non-Leaker population contains some star bursts as strong as what we see in the leaker population introduces the question of why some star bursts lead to high f_{esc} while others do not. 30% of the sample of Non-Leakers that exhibit strong starbursts have either had a very recent burst (within the past 5 Myr) or the star burst occurred in the range between 70 – 100 Myr prior to $z = 4.64$. In the former case, feedback has not had enough time to clear channels in the HI distribution. In the latter, there may have been enough time for the galaxy to re-collapse and form a dense ISM structure with a low escape fraction. For the remaining 70%, there are two possibilities. Either the haloes never exhibited a high f_{esc}, despite the star burst, or the star burst did efficiently clear channels in the ISM but the gas re-collapsed and settled much faster than what we see in the Remnant Leaker population, thus their f_{esc} just happens to be low at $z = 4.64$. As we are primarily interested in developing methods to find current LyC leakers and constrain the mechanisms by which photons escape, we leave this question open for future work.

Another way to quantify these bursts is by comparing the Bursty and Remnant Leakers with the Non-Leaker galaxy population on the galaxy formation main sequence (i.e. stellar mass versus SFR). In the top panel of Figure 6 we show stellar masses of SPHINX^{20} galaxies at $z = 4.64$ compared with SFR_{10}. In general, the Bursty Leakers populate a region significantly higher than the main sequence while the Remnant Leakers mostly fall significantly below. As we discussed earlier, the Bursty Leakers have had large bursts of star formation in the past 10 Myr while the Remnant Leakers have had their star formation nearly completely shut down so it is unsurprising that we find strong differentiation on the galaxy main sequence. In contrast, if we show the main sequence but replace SFR_{10} with SFR_{100}, as shown in the bottom panel of Figure 6, both populations of leakers now fall above the mean relation. This is because both Bursty and Remnant Leakers on average have exhibited extreme bursts of star formation in the past 100 Myr. Consistent with our simulations, candidate LyC leakers at $z = 3 - 3.5$ with $f_{esc} > 0.2$ from Saxena et al. (2022b) as well as other strong LyC leaking galaxies, *Ion 1, 2* and *3* (Vanzella et al. 2012, 2016, 2018), also fall above the main sequence.

We have compared our simulated galaxy main sequence using SFR_{100} with observational estimates from the ALPINE survey (Khusanova et al. 2021). Data in the IR was used to make estimates of SFR, which is comparable to star formation over the last 100 Myrs. We find very good agreement between our simulations and the observational constraints. The key result from this exercise is that the location of a LyC leaker on the galaxy formation main sequence is highly dependent on the time scale over which an SFR indicator is sensitive.

3.2 ISM Property Comparison of Bursty and Remnant Leakers

The mechanism (i.e. radiation-bounded with holes or density bounded) by which photons are leaking in each population of LyC leaker is still unclear. One method for elucidating this physics is to compare the ISM properties of each population conjointly as well as with control samples of Non-Leakers that are selected to exhibit the same stellar mass distribution as each leaker population[7].

In Figure 7 we compare distributions of 90th percentile[8] HII region gas density ($\rho_{90\%}$), 90th percentile HII region ionization parameter ($q_{90\%}$), and HII region metallicity with the control samples of stellar mass-selected Non-Leaker galaxies. Beginning with similarities, we find no difference in HII region metallicity between any of the samples. The peaks of the distribution occur at $\sim 3\%\ Z_\odot$. Where our stellar masses overlap, the mass-metallicity relation of SPHINX^{20} galaxies is in good agreement with observational estimates from Faisst et al. (2016) at $z = 5$ so we expect our predicted high-redshift leaker metallicities to be reasonably robust. We find no differentiation between the Bursty Leakers, Remnant Leakers, and Non-Leakers on the stellar mass-metallicity relation (not shown).

Continuing with the differences, the middle panels of Figure 7 compare the ionization parameters of the Bursty and Remnant Leakers with the stellar mass-selected control samples from the Non-Leaker population. While the distribution of ionization parameters is very consistent between Remnant Leakers and Non-Leakers, we find a significant enhancement in the dimensional ionization parameter (q) for the Bursty Leakers compared to the other two populations. This reflects the strong enhancement in ionizing luminosity observed in Figure 2 and is certainly due to the much younger stellar populations in the Bursty Leakers as can be seen in Figure 3 as this is the primary source if ionizing radiation in the galaxies.

If we compare HII region gas density, we find that the Bursty Leakers exhibit similar gas densities to the Non-Leaker control sample, perhaps lacking some of the densest gas, although this may be due to limited sample size. This indicates that the ISM in the Bursty Leakers is not fully disrupted, but the combination of early SNe and radiation feedback must be clearing channels in the ISM. The fact that the Bursty leakers have typical gas densities also further demonstrates that the enhancement in ionization parameter is due to an enhancement in radiation and not a reduction in density. In contrast, the Remnant Leakers exhibit significantly lower ISM gas

[7] We note that there is no clear method for how to properly choose a control sample. We have controlled for stellar mass but it might also be appropriate to control for star formation history, star formation rate, halo mass, or any other galaxy property.

[8] Describing the properties of the ISM by a single number is not a well defined problem. We have chosen the 90% percentile value of the distribution for the gas density and ionization parameter to be consistent with Katz et al. (2022b). The results are not fundamentally different if other percentiles are used. The values are computed by locating all gas cells that are part of the HII regions of the galaxies and measuring the 90th percentile values of the distribution.

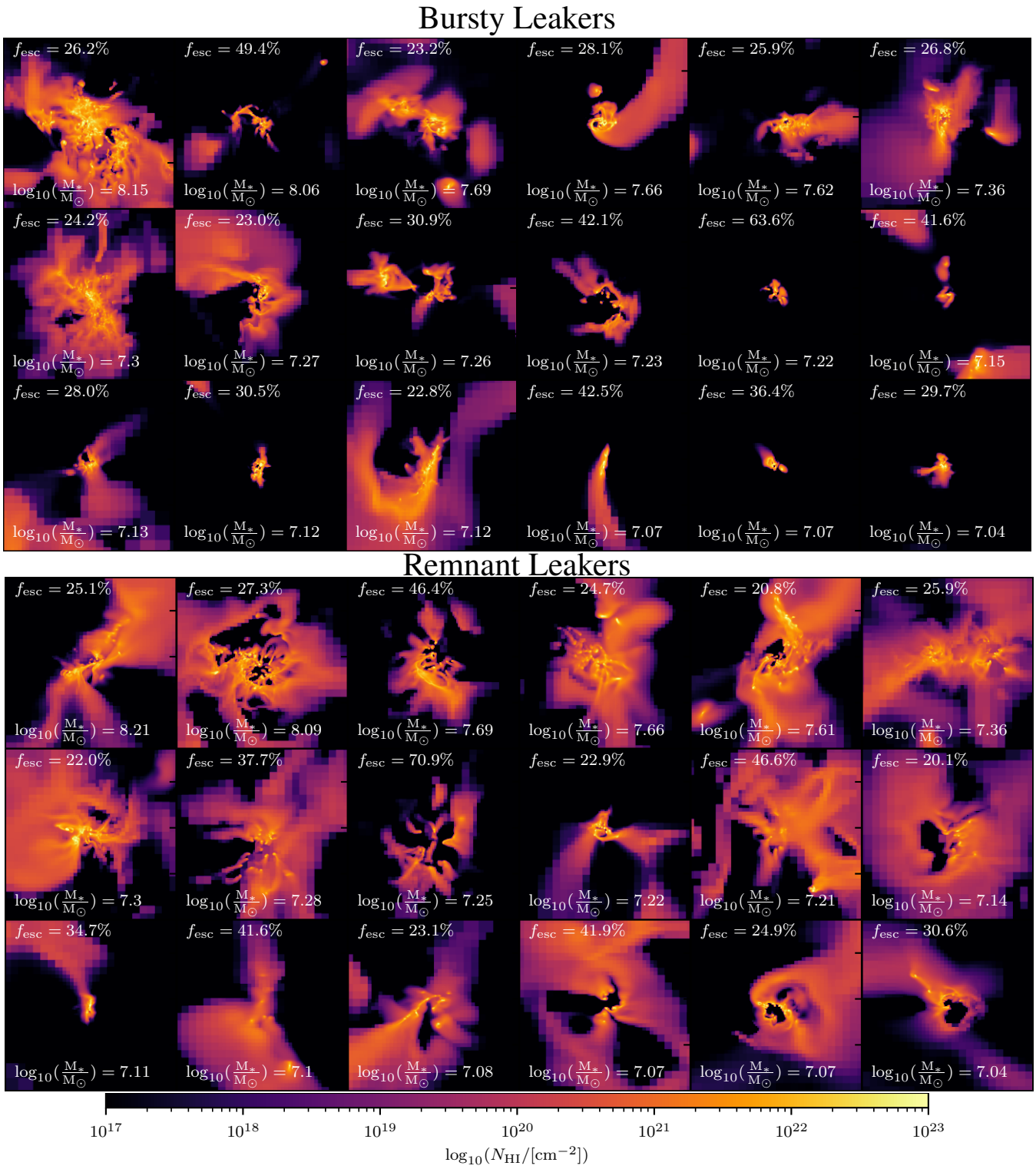

Figure 8. HI column density maps for Bursty Leakers (top) and Remnant Leakers (bottom) at $z = 4.64$. All maps are 10 physical kpc in width and we list the escape fraction and stellar mass of each object. We show the densest Bursty Leakers (representing 31% of the total sample) with a corresponding stellar mass selected sample of Remnant Leakers. In the top panel, we see that the ISM is still intact for many Bursty Leakers but the strong radiation field has often ionized a significant amount of the gas in the CGM. In contrast, Remnant Leakers tend to show large holes is a disrupted ISM.

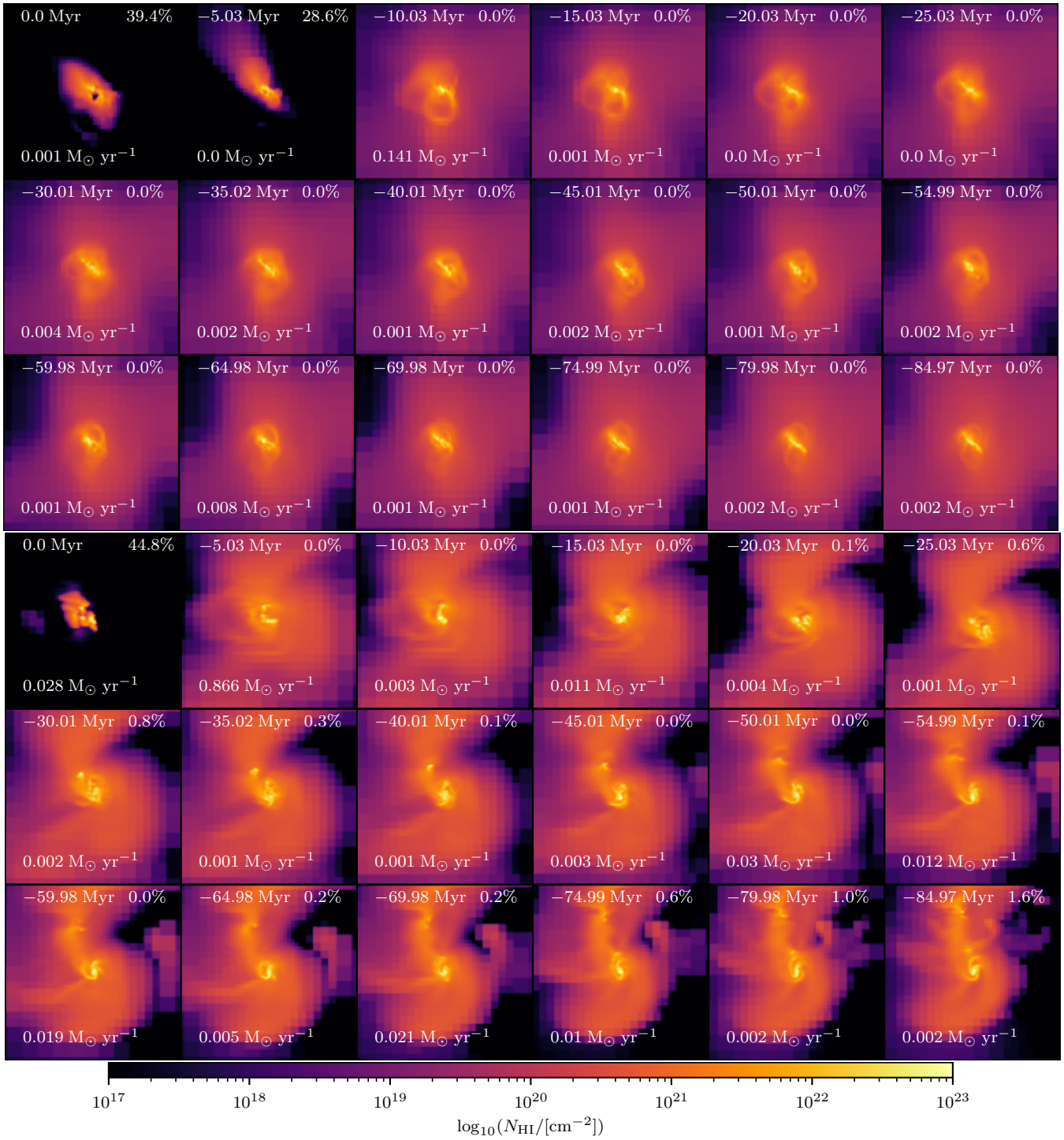

Figure 9. Maps for the HI column density in a time series of the 85 Myr prior to $z = 4.64$ for two Bursty Leakers with halo masses of $10^{8.3}$ $M_\odot$ and $10^{9.1}$ $M_\odot$ and stellar masses of $10^{6.1}$ $M_\odot$ and $10^{6.7}$ $M_\odot$, respectively. The three labels indicate the time of the snapshot, the LyC escape fraction, and the star formation rate measured over the previous 1 Myr.

densities compared to the Non-Leaker control sample or the Bursty Leakers, indicating that their ISM is nearly fully disrupted. Returning to the Zackrisson et al. (2013) models, the ISM of Bursty Leakers is more akin to an ionization-bounded nebula with holes while the ISM of Remnant Leakers seem to be more representative of a density-bounded nebula.

We further demonstrate the differences in gas distribution between Bursty and Remnant Leakers in Figure 8 where we show HI column density maps for 18 Bursty Leakers (top rows) and 18 Remnant Leakers (bottom rows). The Remnant Leaker maps have been selected by stellar mass to closely match those of the Bursty Leakers that are shown. Each image is 10 physical kpc in width. It is clear that galaxies in both leaker categories exhibit a wide diversity in HI morphology and in all systems, there is evidence for disruption in

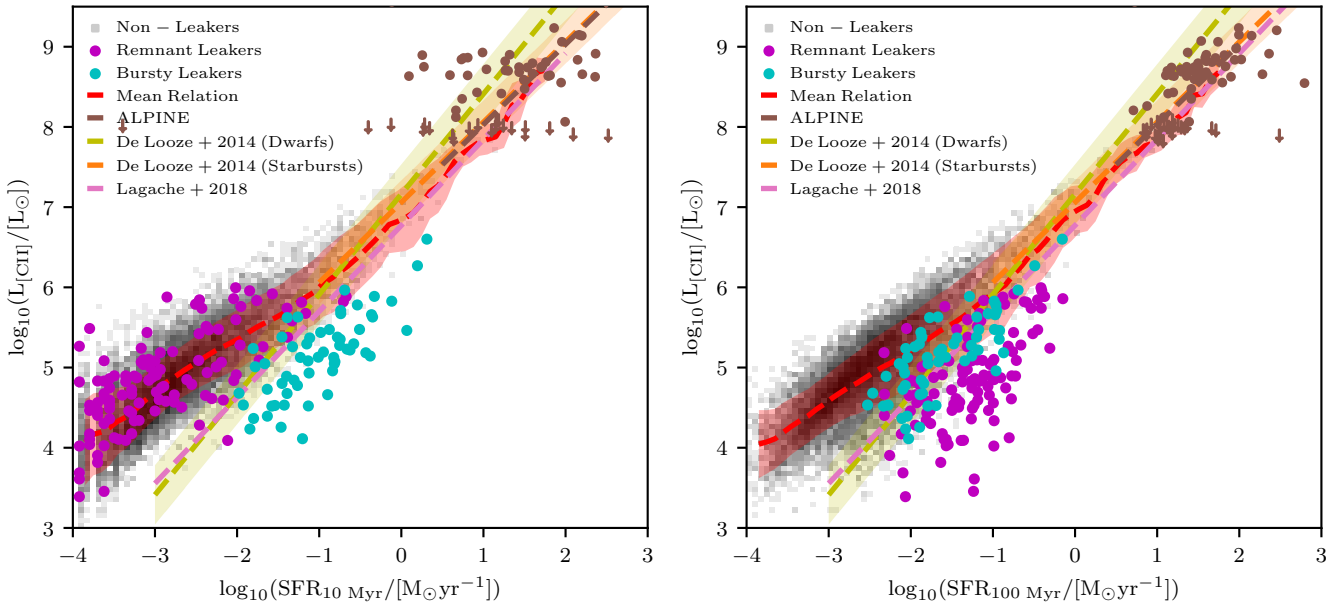

Figure 10. [CII]-SFR$_{10}$ (left) and [CII]-SFR$_{100}$ (right) relation for SPHINX^{20} galaxies at $z = 4.64$ compared with fits to various types of low-redshift galaxies (olive-green and orange, De Looze et al. 2014) as well as high-redshift observations (brown, Schaerer et al. 2020) and models (pink, Lagache et al. 2018). Bursty Leakers are shown in cyan while Remnant Leakers are shown in magenta. All SPHINX^{20} Non-Leakers are shown as the grey background histogram. The mean relation from the simulation is shown as the red dashed line and the red shaded region represents the 1σ scatter. Individual $z < 5.5$ galaxies from ALPINE are shown in brown with points representing detections and arrows representing 3σ upper limits on [CII] luminosity. ALPINE SFRs that are calculated from Hα inferred from *Spitzer* photometry and using SED fitting (Faisst et al. 2020) against their corresponding [CII] luminosities (Béthermin et al. 2020) are shown in the left and right panels, respectively.

the ISM. Nevertheless, among the Bursty Leaker population, there seem to be substantially more systems with dense central clouds of neutral gas compared to the Remnant Leaker population. In contrast, the Remnant Leaker galaxy population often exhibits large cavities of ionized gas that can only be created by substantial SNe feedback. Qualitatively, there are morphological differences between the neutral gas distributions in each class of leaker.

In Figure 9 we show a time series of the HI distribution for two example Bursty Leakers in the 85 Myr prior to $z = 4.64$. In both cases, the galaxy is in an idle state until a strong burst of star formation occurs. For the first galaxy in Figure 9, we see no evidence of any strong dynamical interactions that is driving the burst while for the second galaxy, a small gaseous clump can be seen migrating to the center, which may help drive the starburst. It is clear that the feedback from star formation is driving the increase in $f_{\rm esc}$ in both galaxies. Such behaviour was also reported in (Trebitsch et al. 2017; Kimm et al. 2017; Rosdahl et al. 2018).

3.3 Implications for [CII] 158μm Emission

From a theoretical viewpoint, differentiating the mechanisms by which LyC photons escape galaxies can improve our understanding of the epoch of reionization and the sources responsible. Testing the predictions from our simulations against observations is a necessary step to determine the reliability of our model. Identifying the observational consequences of each leakage mode is thus key for testing such predictions. As the [CII] 158μm emission line is one of the brightest emission lines at high-redshift (e.g Carilli & Walter 2013), we focus this Section on the implications of different LyC leakage modes on [CII] emission. We reiterate that it is already well established that the ratio of [OIII] 88μm to [CII] emission positively correlates with $f_{\rm esc}$ (e.g. Inoue et al. 2016; Katz et al. 2020, 2022b); however, this ratio does not describe how $f_{\rm esc}$ relates only to [CII], which is particularly interesting at $z < 6$ where [OIII] is more difficult to observe with ALMA due to it falling in Band 9 or Band 10 that has a limited observational window.

In Figure 10 we show the [CII]-SFR relation for SPHINX^{20} galaxies at $z = 4.64$ compared with low-redshift observational estimates from De Looze et al. (2014), high-redshift observational estimates from the ALPINE [CII] survey at $z \approx 4.4 - 5.9$ (Béthermin et al. 2020; Schaerer et al. 2020; Faisst et al. 2016) where only $z < 5.5$ galaxies from ALPINE are shown, and the high-redshift models from Lagache et al. (2018). The left panel shows the relation using SFR$_{10}$ while the right panel uses SFR$_{100}$. We find very little difference in the mean [CII]-SFR relation (shown as dashed red lines) regardless of which time period the SFR is measured over. At SFR > 1 M$_\odot$yr^{-1} we find very good agreement between SPHINX^{20} and both low- and high-redshift observations. However, the situation is fundamentally different at lower SFR.

We find a kink in the relation at SFR$_{10,100} \sim 0.1$ M$_\odot$yr^{-1} such that the typical SPHINX^{20} galaxy exhibits much higher [CII] luminosities at lower SFR compared to the extrapolations of the observed relations. The kink in the [CII]-SFR relation is not unique to our simulation as it is also seen in Pallottini et al. (2022).

Depending on whether we use SFR$_{10}$ or SFR$_{100}$ for the [CII]-SFR relation, the Bursty Leakers and Remnant Leakers populate different portions of the diagram. Beginning with SFR$_{10}$, the Bursty Leakers exhibit large [CII] deficits for their given SFR while Remnant Leakers tend to fall on the relation. Switching SFR$_{10}$ for SFR$_{100}$ (as shown in the right panel of Figure 10), both Bursty and Remnant Leakers exhibit deficits in the [CII]-SFR relation, with the Remnant Leakers exhibiting slightly larger deficits.

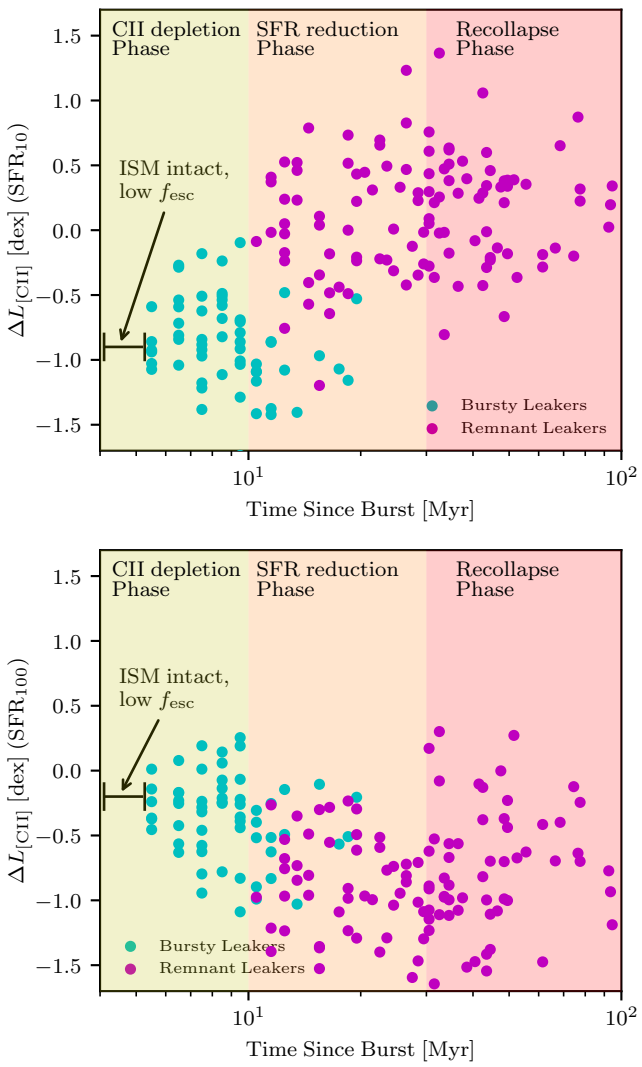

Figure 11. Deviation of the [CII] luminosity from the mean value on the [CII]-SFR$_{10}$ relation (top) or [CII]-SFR$_{100}$ relation (bottom) as a function of time since the peak SFR for SPHINX^{20} leakers. Bursty and Remnant Leakers are shown in cyan and magenta, respectively. No leakers populate the first 5 Myr of the plot as it takes time for feedback to disrupt the ISM of the star forming clouds. In the top panel, galaxies initially show deficits due to radiation feedback and early SNe feedback destroying PDRs. We propose that once the feedback becomes effective enough, the SFR drops, which moves galaxies closer to the [CII]-SFR$_{10}$ relation. Once star formation is completely shut down, galaxies appear to have a [CII] excess due to residual neutral gas. Finally galaxies approach the relation as the ISM reforms and star formation begins again. In the bottom panel, the [CII] deficit continues to increase with time because the SFR is fixed over the 100 Myr interval, but SNe continue to explode over 50 Myr reducing the neutral gas content.

We emphasize that both the emission line signatures and SFRs of individual galaxies are dynamic and highly evolving with time (e.g. Barrow et al. 2020). In fact, we view the evolution of an individual galaxy on the [CII]-SFR relation as a potential probe of state of the ISM in concert with star formation. This is highlighted in the top panel of Figure 11 where we plot the time elapsed in Myr since the burst of star formation[9] that caused LyC photons to leak against the [CII]-deficit (defined as the difference between the [CII] emission of a galaxy and the mean [CII] emission of all SPHINX^{20} galaxies at a fixed SFR) for the [CII]-SFR$_{10}$ relation for both Bursty and Remnant Leakers. We propose the following illustrative model. For the first 4 – 5 Myr after the burst, f_{esc} is low as the low column density channels have yet to form. Hence the ISM remains intact and we have no leakers in this regime. However, once enough stars form, the ionizing radiation can create holes in the ISM and reduce the neutral gas content. We describe this phase as the CII depletion phase[10] and it is qualitatively shown in yellow in Figure 11. Once the feedback becomes efficient, the SFR begins to decrease, which moves the galaxy to the left on the [CII]-SFR$_{10}$ relation, back towards the mean relation, and has the effect of reducing the [CII] deficit. At the same time SNe are exploding, which can further reduce the neutral gas content and can balance the reduction in [CII] deficit caused by reducing the SFR. Once star formation has been substantially reduced (i.e. by a few orders of magnitude), the galaxies may overshoot the [CII]-SFR relation as some neutral gas likely persists. The galaxies will loiter in this SFR reduction phase (shown in orange on Figure 11) until the gas in the CGM can recollapse and inflows from the IGM bring fresh gas into the system. As the galaxies cool down, the ISM will begin to reform and approach the mean relation as both the SFR and [CII] emission begin to increase again (shown in red on Figure 11 as the Recollapse phase).

The behaviour is fundamentally different when following the same evolution using the deficit on the [CII]-SFR$_{100}$ relation. In this case, the SFR for both Bursty and Remnant Leakers is enhanced compared to the typical galaxy of the same stellar mass. This moves the galaxies to the right on the [CII]-SFR$_{100}$ relation, resulting in a [CII] deficit. As the ionizing photons destroy the neutral gas content, the [CII] deficit increases. As SNe explode, the deficit continues to increase, which is why the Remnant Leakers continue the trend of increasing [CII] deficit with time. The SNe explode for $\sim$ 50 Myr and as this process subsides, the galaxies can begin to recollapse. Thus, it is important to note that the total CII content of a galaxy will decrease throughout both the CII depletion and SFR reduction phases, even if the galaxy moves closer to the [CII]-SFR$_{10}$ relation during this phase.

The qualitative evolution that we have described is subject to significant scatter due to various galaxy properties; hence, the [CII] deficit is not a perfect one-to-one relation with the time since the burst. For example, the length and strength of the bursts will play a role in how quickly the photodissociation regions (PDRs) are destroyed and star formation is reduced. The initial state and structure of the ISM as well as the local efficiency of star formation will also help determine how efficient the feedback is and how quickly it impacts [CII] and subsequent star formation (Kimm et al. 2019; Kim et al. 2019; Kimm et al. 2022). The strength of galactic inflows and the cooling rate in the CGM will impact the length of the loitering/SFR reduction phase as well as how quickly the gas can recollapse. There

[9] The time of the burst is measured at the time of peak star formation rate within the 100 Myr prior to $z = 4.64$.

[10] Note that there is a prior phase that we call the Burst phase where the SFR is increasing. This also causes [CII] deficits as discussed earlier. It is however not shown on the plot because we plot time since the maximum SFR (i.e. the period after the initial Burst phase) where the instantaneous SFR begins decreasing again. One must keep in mind that SFR$_{10}$ can be long enough to average over the most of the burst (i.e. both part of the increase and decrease, see Figure 3) so it takes slightly longer for SFR$_{10}$ to decrease compared to the instantaneous SFR (i.e. that averaged over 1 Myr).

is already a substantial amount of scatter in the [CII]-SFR relation that encapsulates many of these galaxy formation processes.

We conclude this Section by noting that the idea that strong bursts of star formation lead to [CII] deficit is not new. Since the early discovery of potential [CII] deficits (e.g. Maiolino et al. 2015) and the latter confirmation of some (e.g. Laporte et al. 2019; Carniani et al. 2020), various explanations have been proposed for this behaviour. These include the impact of the CMB and low metallicity (e.g. Vallini et al. 2015; Pallottini et al. 2017), radiation field intensity (e.g. Lagache et al. 2018), and bursty star formation (e.g. Ferrara et al. 2019; Pallottini et al. 2022). We confirm that on the [CII]-SFR_{10} relation, young bursty leakage leads to [CII] deficits while older starbursts that lead to LyC leakage typically result in normal or even enhanced [CII] for a given SFR. In all cases, strong bursts lead to deficits on the [CII]-SFR_{100} relation, consistent with other work in the literature (e.g. Ferrara et al. 2019; Pallottini et al. 2022).

3.4 Finding LyC Leakers in and out of the Epoch of Reionization with [CII]

We have shown that the location of a LyC leaker on both the [CII]-SFR_{10} and [CII]-SFR_{100} relations can differentiate the type of LyC leakage that is occurring in the galaxy. The primary motivation for studying this physics at $z < 6$ is because IGM attenuation prevents direct observations of LyC emission during the epoch of reionization (EoR). Furthermore, our predictions can be tested at lower redshift with known LyC leakers. However, as our ultimate goal is to understand LyC leakage in the EoR, we continue our analysis by showing that our results at $z = 4.64$ also hold at $z = 6$ and that current large ALMA programs (e.g. Le Fèvre et al. 2020; Bouwens et al. 2021a) focusing on [CII] emitters at high-redshift may be able to detect LyC leakers. In order to best compare with observations, we focus our analysis on the $\sim$ 700 most massive haloes (i.e. haloes with virial masses $> 3\times10^9 M_\odot$) in SPHINX^{20} at $z = 6$. In this sample, we find 8 Bursty Leakers, 20 Remnant Leakers, and 515 Non-Leakers. The distribution of Bursty to Remnant leakers is similar to the $z = 4.64$ snapshot.

In the top left and top centre panels of Figure 12 we show the [CII]-SFR_{10} and [CII]-SFR_{100} relations for the massive SPHINX^{20} galaxies at $z = 6$. The behaviour is identical to that seen in Figure 10 for $z = 4.64$. When using SFR_{10}, Bursty Leakers show [CII] deficits, while Remnant Leakers populate similar regions as Non-leakers. In contrast, when using SFR_{100} both types of leaker exhibit [CII] deficits. For comparison with observations, the top centre panel also contains detections as well as 3σ limits on [CII] for spectroscopically confirmed galaxies at $z > 5.5$ from the ALPINE survey and at $z > 6$ from the compilation by Matthee et al. (2019).

In the bottom left and bottom right panels of Figure 12, we show the [CII]-Hα[11] and [CII]-Hβ relations, respectively. We find a strong linear trend between the log values of [CII] and both Balmer emission lines, which reflects the fact that both are considered SFR indicators (e.g. Kennicutt 1992; De Looze et al. 2014), despite the fact that they probe different gas (i.e. neutral versus ionized). Bursty leakers once again exhibit [CII] deficits while remnant leakers are more consistent with the bulk of the SPHINX^{20} galaxy population at $z = 6$. To probe longer time scale star formation, in the top right panel of Figure 12, we compare [CII] emission with the 1500Å UV magnitude[12] of the galaxy. We now see that both Bursty and Remnant Leakers exhibit [CII] deficits compared to other SPHINX^{20} galaxies at the same UV magnitude. Thus the combination of [CII] luminosity, UV magnitude, and Hα and Hβ emission are very powerful for identifying Bursty and Remnant leakers, and potentially differentiating the two.

However, we highlight that there are potential caveats with this approach. The observed LyC escape fraction is highly viewing angle-dependent (e.g. Cen & Kimm 2015). Not all galaxies with [CII] deficits will be observed as leakers, both because there are some galaxies with deficits that are truly Non-Leakers (i.e. the [CII] deficit galaxies are not purely comprised of leakers), and due to the viewing angle dependence, an observer may be unlucky and be positioned along an optically thick line of sight. Only large samples of galaxies will be able to disentangle this degeneracy and this is further discussed below. Thus to confirm our predictions, we recommend initially studying known LyC leakers and investigating whether they exhibit the relevant [CII] deficits. However, this also introduces biases as the population of known LyC leakers is subject to various selection effects; nevertheless, we expect [CII] deficits should persist, even for biased samples of leakers, when using a SFR indicator sensitive to long time scales. We have also assumed that the intrinsic Hα and Hβ emission can be derived with the appropriate dust corrections while the angle-averaged 1500Å attenuation is representative of what one would observe along a typical line of sight. Viewing angle effects can also introduce additional scatter into the relations. Since [CII] emission originates primarily in neutral gas at these redshifts (Pallottini et al. 2017; Katz et al. 2019; Lupi et al. 2020), Hα and Hβ emission comes from mostly ionized (or partially ionized) gas, and LyC photons and 1500Å photons come from stars, differences in attenuation for each of these sources along various lines of sight can be important. The simplistic dust modelling employed in SPHINX^{20} in post-processing can account for this (Katz et al. 2022a); however, future simulations with more self-consistent modelling will be required to further assess this additional scatter.

4 DISCUSSION & CONCLUSIONS

4.1 Recommended Observing Strategy

Our results excite great optimism that [CII] can be combined with other commonly observed quantities (e.g. UV magnitude or Hα luminosity) to not only identify potential LyC leakers, but also gauge the mechanisms by which the LyC radiation is leaking. However, we emphasize that our results represent population averages and results for individual haloes may vary due to numerous observational (e.g. orientation angle) and physical effects. For this reason, we argue that the best way to test our model is to follow up known low-redshift LyC leakers with [CII] observations. This will remove issues related to line of sight effects, as all leakers (i.e. regardless of being Bursty or Remnant) must show [CII] deficits at some level when using a SFR indicator sensitive to $\sim$ 100 Myr time scales.

ALMA is sensitive to [CII] in Band 8 at $3 \lesssim z \lesssim 4$. The relatively restrictive transmission function of ALMA Band 8 means that [CII] observations from galaxies in only certain redshift bands are possible. Nevertheless, with the increasingly large samples of known LyC leakers in this redshift interval (see Meštrić et al. 2021, for a

[11] Here we use the intrinsic Hα emission rather than the dust attenuated value. This assumes that the dust content of the galaxy can be inferred and attenuation corrected for. We do the same for Hβ.

[12] Here we are using the angle-averaged dust attenuated value for UV magnitude. The results are the same if we use the intrinsic magnitude rather than the dust attenuated value.

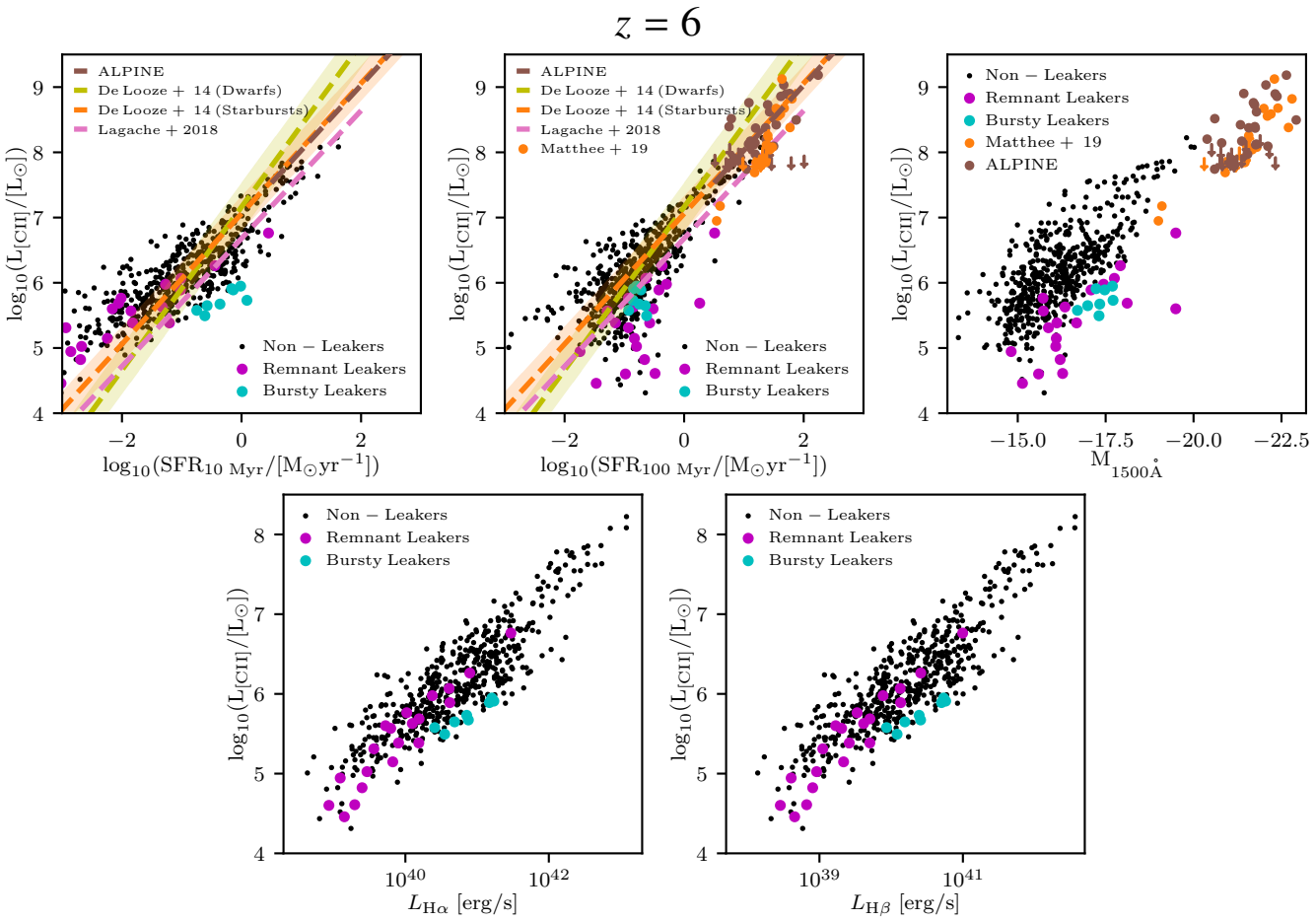

Figure 12. [CII]-SFR$_{10}$ (top left), [CII]-SFR$_{100}$ (top centre), [CII]-1500Å UV magnitude (top right), [CII]-Hα (bottom left), and [CII]-Hβ (bottom right) relations for SPHINX^{20} galaxies at $z = 6$ compared with fits to various types of low-redshift galaxies (olive-green and orange, De Looze et al. 2014) as well as high-redshift observations (brown, Schaerer et al. 2020) and models (pink, Lagache et al. 2018). Bursty Leakers are shown in cyan while Remnant Leakers are shown in magenta. Non-Leakers are shown in black. Galaxies are only included if their halo mass is $> 3 \times 10^9 M_\odot$. Individual $z > 5.5$ ALPINE galaxies are shown as brown points for detections or arrows representing 3σ upper limits on [CII] luminosity. Orange points represent the high-redshift data compilation from Matthee et al. (2019).

recent compilation), there are enough galaxies to begin testing the predictions made in this work.

We highlight a few important considerations to keep in mind when testing our predictions. We note that not all galaxies with [CII] deficits are LyC leakers. In Figure 13 we show the probability of a galaxy being a LyC leaker as a function of [CII] deficit for both the [CII]-SFR$_{10}$ and [CII]-SFR$_{100}$ relations binned by SFR. It is clear that this fraction varies significantly between SFR bins and as a function of [CII] deficit. On the [CII]-SFR$_{100}$ relation, a [CII] deficit seems to be a necessary, but insufficient condition for being a LyC leaker. Furthermore, the direct detection of LyC leakage suffers from both orientation angle effects and IGM transmissivity. This means that even if a galaxy exhibits high [CII] deficits, significant LyC leakage may not be detectable for a particular galaxy. Finally, when following up known LyC leakers with [CII] observations, it is also important to consider orientation angle effects as well because an individual sight line in a galaxy may be optically thin even if the angle-averaged LyC f_{esc} is low, in which case no [CII] deficit may be present. For example, we find that the number of lines of sight with $f_{esc} > 20\%$ that are hosted by galaxies that have angle-averaged $f_{esc} > 20\%$ is a ratio of 2 to 1 (see Katz et al. 2022a). For these reasons, large samples of galaxies will be needed to overcome the scatter introduced by these biases.

4.2 Contribution to the Reionization Budget

We have focused primarily on two types of LyC leakers that were arbitrarily defined to have LyC escape fractions $> 20\%$. An open question is what role do these systems play in the reionization of the Universe and whether it changes between $z \approx 4.64$ and $z \approx 6$. To understand this, we quantify the contribution of each type of leaker to the reionization process as a function of redshift. In Figure 14 we show the total luminosity of ionizing photons escaping into the IGM as a function of redshift compared to the contribution from Bursty and Remnant Leakers.

At all redshifts, Remnant Leakers represent a subdominant contribution to the total escaping ionizing luminosity budget. Despite their higher number densities, the lack of recent star formation means they have much lower intrinsic ionizing luminosities compared to Bursty Leakers. Although subdominant, the Remnant Leaker photon budget is non-negligible and represents up to $\sim 10\%$ of the total at any redshift. Bursty Leakers contribute significantly more ionizing photons, contributing up to $\sim 47\%$ at high-redshift. The relative contribution

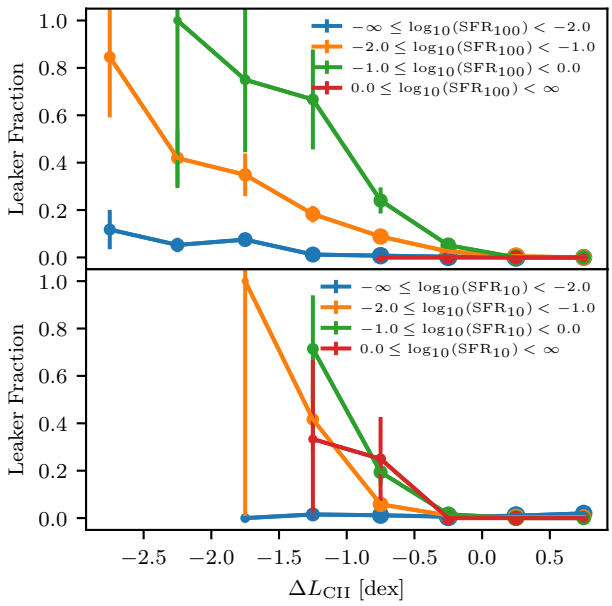

Figure 13. Fraction of galaxies that have $f_{\mathrm{esc}} > 20\%$ as a function of deviation from the [CII]-SFR relation for different bins of star formation rate. The top and bottom panels show the results using SFR_{100} and SFR_{10}, respectively.

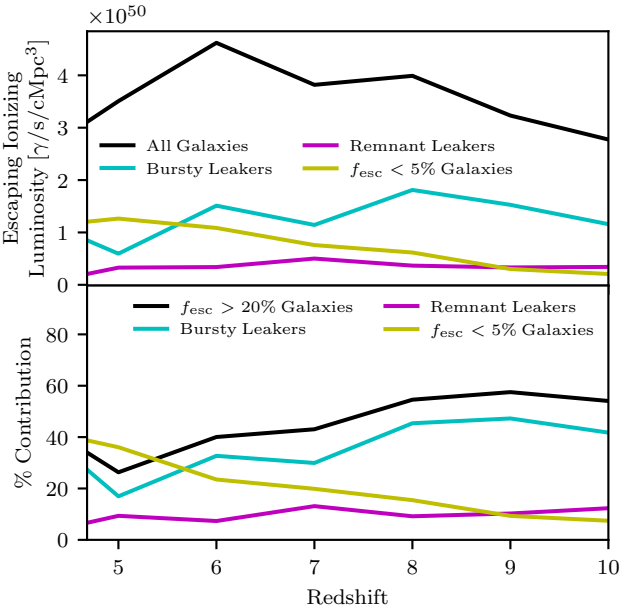

Figure 14. (Top) Ionizing luminosity escaping into the IGM as a function of redshift in SPHINX^{20}. The black line shows the contribution from all galaxies while the cyan and magenta lines show the total escaping ionizing luminosities of Bursty and Remnant Leakers, respectively. The olive-green line shows the contribution from galaxies with $f_{\mathrm{esc}} < 5\%$. (Bottom) Percentage contribution to the total escaping ionizing budget for Bursty and Remnant Leakers as a function of redshift.

from both Bursty and Remnant Leakers decreases with decreasing redshift.

We interpret this to mean that galaxies with high f_{esc} are a necessary and important part of reionization, but low f_{esc} galaxies are equally important for reionization. To demonstrate this, we show the contribution of galaxies with $f_{\mathrm{esc}} < 5\%$ to the reionization budget as olive-green lines. At the end of reionization, the low f_{esc} galaxies contribute equal numbers of ionizing photons into the IGM as the high f_{esc} galaxies. It should be noted that this does not necessarily imply that there are numerous galaxies with low f_{esc} all contributing equally. At any given redshift, the reionization budget is dominated by a small fraction of halos. Since intrinsic ionizing luminosities increase with decreasing redshift (as galaxies become more massive), low f_{esc} galaxies can still provide substantial numbers of ionizing photons to the IGM. This is further discussed in Rosdahl et al. (2022). However, a potential caveat to these conclusions is that SPHINX^{20} reionizes slightly later than the reionization history inferred from observations (e.g. Fan et al. 2006; Kulkarni et al. 2019). A detailed discussion into the numerics of this is beyond the scope of this paper; however, different solutions will have varying impacts on Figure 14 and this warrants further exploration.

4.3 Caveats

As our work is based on cosmological radiation hydrodynamics simulations, various caveats should be considered when interpreting our results. A detailed discussion of these can be found in Katz et al. (2022b) and we summarize the primary uncertainties of our model below.

Like all simulations, SPHINX^{20} has finite spatial and mass resolution and thus quantities such as the LyC escape fraction can only be measured on the scale (i.e. 10 pc) at which they are injected into the simulation. Important ISM physics may be occurring at scales that are not resolved by the simulation. SPHINX^{20} includes various sub-grid models for star formation and feedback. Such models are designed to reasonably reproduce the behaviour of this physics on the scales that are resolved, but similar luminosity functions or stellar mass-halo mass relations may be obtained with another choice of model. These choices can all impact [CII] emission, SFRs, and LyC escape. Furthermore, due to the volume of SPHINX^{20}, the stellar mass range probed by the simulation does not yet overlap with observed high-redshift galaxy samples (e.g. ALPINE or REBELS). SPHINX^{20} predicts that high mass galaxies have very low f_{esc} in general, so observations may need to push to lower stellar masses for our predictions to be robustly tested.

Extracting [CII] emission from simulations is highly non-trivial (see e.g. Olsen et al. 2018). Because SPHINX^{20} does not compute the non-equilibrium CII abundance or level populations self-consistently, such measurements must be made in post-processing. While we have attempted to use as much of the information as possible from the simulation in post-processing, there are multiple methods for extracting [CII] luminosity that may result in different values. Our results are in good agreement with observational constraints, especially at high stellar masses, which should provide confidence in our method; however future simulations (e.g. Katz 2022) that follow the non-equilibrium physics in a more self-consistent manner will be needed to confirm our results.

4.4 Summary

We have analyzed the SPHINX^{20} simulation at $z = 4.64$ and $z = 6$ to understand the observational signatures of escaping LyC pho-

tons. Focusing primarily on $[CII]_{158\mu m}$ emission, we identified two classes of galaxies with high f_{esc} (i.e. $f_{esc} > 20\%$): Bursty Leakers and Remnant Leakers. The former are categorized as having a very recent burst of star formation within the past 5-15 Myr, exceptionally high ionization parameters and typical gas densities. The latter are remnants of a much older burst of star formation (within the past 100 Myr) and exhibit very little current star formation, low ISM gas densities, and normal ionization parameters. The two classes of leakers are qualitatively similar to ionization bounded nebulae with holes and density bounded nebula, respectively. We compared and contrasted each type of leaker with each other and the Non-Leaker population and our conclusions can be summarised as follows:

- To reach a LyC escape fraction of $> 20\%$, galaxies typically undergo bursts of star formation that are $90\times - 100\times$ higher (when averaged in 1 Myr intervals) than the SFR of the galaxy prior to the burst. Not all bursts of star formation of such magnitude are guaranteed to result in efficient LyC leakage. The primary difference between whether a galaxy is a Bursty Leaker or a Remnant Leaker is the timing of the burst of star formation.
- When compared with Non-Leakers on the galaxy formation main sequence ($SFR - M_*$), both Bursty and Remnant Leakers populate the regions above the mean when using an SFR indicator that is sensitive to timescales of $\sim$ 100 Myr, which is consistent with observations. For shorter time scale indicators (e.g. Hα), only the Bursty Leakers appear as outliers above the mean.
- When using short time scale SFR indicators (e.g. Hα), only Bursty Leakers show [CII] deficits. In contrast, when using a longer time scale indicator (e.g. UV magnitude), both types of leaker show [CII] deficits. The combination of observations of [CII] as well as short (e.g. Hα and Hβ) and long time scale SFR indicators (e.g. M_{UV}) is a powerful combination for not only identifying potential LyC leakers, but also to constrain the mode by which the photons are leaking.
- Although [CII] deficits are promising tracers of significant LyC escape, we note that several considerations must be made when inferring [C II] deficits from LyC leakers and vice versa. Importantly, the fraction of galaxies with high [CII] deficits that also show strong LyC leakage is dependent on the SFR. Orientation and line-of-sight effects can impact the measurement of LyC leakage but are unlikely to affect [CII] luminosities. Finally, although an observed LyC leaking galaxy may exhibit a high escape fraction along the line-of-sight aligned with the observer, its angle-averaged LyC escape fraction may still be low, potentially weakening the [CII] deficit observed.
- Despite having LyC escape fractions of $> 20\%$, the total contribution of Bursty and Remnant Leakers only peaks at $\sim 60\%$ at $z = 9$ and decreases thereafter. At $z \lesssim 5.5$, galaxies with $f_{esc} < 5\%$ contribute more ionizing photons to the UV background than galaxies with $f_{esc} > 20\%$. Thus, it is the high f_{esc} galaxies that dominate early reionization, but low f_{esc} galaxies that complete and maintain it.

The ideas presented in this work can be tested with a reasonable set of [CII] observations in the intermediate-redshift Universe. Constraining models for LyC leakage is paramount for understanding the reionization era and galaxy formation at $z \geq 6$. Thus connecting observations and simulations in this regard should be considered a top priority, especially in the context of upcoming JWST observations.

ACKNOWLEDGEMENTS

We thank the referee for their comments which improved the manuscript. AS and RSE acknowledge financial support from European Research Council Advanced Grant FP7/669253. TG is supported by the ERC Starting grant 757258 'TRIPLE'. TK was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (No. 2020R1C1C1007079 and No. 2022R1A6A1A03053472). Computing time for this work 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. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 693024). Some of 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. Some of 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.

DATA AVAILABILITY

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

REFERENCES

Barrow K. S. S., Robertson B. E., Ellis R. S., Nakajima K., Saxena A., Stark D. P., Tang M., 2020, ApJ, 902, L39
Béthermin M., et al., 2020, A&A, 643, A2
Bosman S. E. I., et al., 2022, MNRAS, 514, 55
Bouwens R. J., Smit R., Labbé I., Franx M., Caruana J., Oesch P., Stefanon M., Rasappu N., 2016, ApJ, 831, 176
Bouwens R. J., et al., 2021a, arXiv e-prints, p. arXiv:2106.13719
Bouwens R. J., et al., 2021b, AJ, 162, 47
Calzetti D., et al., 2007, ApJ, 666, 870
Carilli C. L., Walter F., 2013, ARA&A, 51, 105
Carniani S., et al., 2020, MNRAS, 499, 5136
Cen R., Kimm T., 2015, ApJ, 801, L25
Chisholm J., Prochaska J. X., Schaerer D., Gazagnes S., Henry A., 2020, MNRAS, 498, 2554
De Looze I., et al., 2014, A&A, 568, A62
Eldridge J. J., Izzard R. G., Tout C. A., 2008, MNRAS, 384, 1109
Faisst A. L., et al., 2016, ApJ, 822, 29
Faisst A. L., et al., 2020, ApJS, 247, 61
Fan X., et al., 2006, AJ, 132, 117
Ferland G. J., et al., 2017, Rev. Mex. Astron. Astrofis., 53, 385
Ferrara A., Vallini L., Pallottini A., Gallerani S., Carniani S., Kohandel M., Decataldo D., Behrens C., 2019, MNRAS, 489, 1
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., 2022, arXiv e-prints, p. arXiv:2201.11716
Grazian A., et al., 2017, A&A, 602, A18
Hopkins P. F., et al., 2010, ApJ, 724, 915
Inoue A. K., Iwata I., 2008, MNRAS, 387, 1681
Inoue A. K., et al., 2016, Science, 352, 1559
Izotov Y. I., Worseck G., Schaerer D., Guseva N. G., Thuan T. X., Fricke Verhamme A., Orlitová I., 2018, MNRAS, 478, 4851
Katz H., 2022, MNRAS, 512, 348
Katz H., Kimm T., Sijacki D., Haehnelt M. G., 2017, MNRAS, 468, 4831
Katz H., et al., 2019, MNRAS, 487, 5902
Katz H., et al., 2020, MNRAS, 498, 164
Katz H., et al., 2021, MNRAS, 507, 1254
Katz H., et al., 2022a, MNRAS,
Katz H., et al., 2022b, MNRAS, 510, 5603
Kennicutt Robert C. J., 1992, ApJ, 388, 310
Kennicutt R. C., Evans N. J., 2012, ARA&A, 50, 531
Khusanova Y., et al., 2021, A&A, 649, A152
Kim J.-G., Kim W.-T., Ostriker E. C., 2019, ApJ, 883, 102
Kimm T., Cen R., 2014, ApJ, 788, 121
Kimm T., Cen R., Devriendt J., Dubois Y., Slyz A., 2015, MNRAS, 451, 2900
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., Keating L. C., Haehnelt M. G., Bosman S. E. I., Puchwein E., Chardin J., Aubert D., 2019, MNRAS, 485, L24
Lagache G., Cousin M., Chatzikos M., 2018, A&A, 609, A130
Laporte N., et al., 2019, MNRAS, 487, L81
Le Fèvre O., et al., 2020, A&A, 643, A1
Leitet E., Bergvall N., Hayes M., Linné S., Zackrisson E., 2013, A&A, 553, A106
Leitherer C., et al., 1999, ApJS, 123, 3
Lupi A., Pallottini A., Ferrara A., Bovino S., Carniani S., Vallini L., 2020, MNRAS, 496, 5160
Madau P., Haardt F., Rees M. J., 1999, ApJ, 514, 648
Maiolino R., et al., 2015, MNRAS, 452, 54
Matthee J., et al., 2019, ApJ, 881, 124
Matthee J., et al., 2022, MNRAS, 512, 5960
Meštrić U., Ryan-Weber E. V., Cooke J., Bassett R., Prichard L. J., Rafelski M., 2021, MNRAS, 508, 4443
Michel-Dansac L., Blaizot J., Garel T., Verhamme A., Kimm T., Trebitsch M., 2020, A&A, 635, A154
Naidu R. P., et al., 2022, MNRAS, 510, 4582
Olsen K., et al., 2018, Galaxies, 6, 100
Paardekooper J.-P., Khochfar S., Dalla Vecchia C., 2015, MNRAS, 451, 2544
Pallottini A., Ferrara A., Gallerani S., Vallini L., Maiolino R., Salvadori S., 2017, MNRAS, 465, 2540
Pallottini A., et al., 2022, arXiv e-prints, p. arXiv:2201.02636
Robertson B. E., et al., 2013, ApJ, 768, 71
Rosdahl J., Teyssier R., 2015, MNRAS, 449, 4380
Rosdahl J., Blaizot J., Aubert D., Stranex T., Teyssier R., 2013, MNRAS, 436, 2188
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
Santini P., et al., 2017, ApJ, 847, 76
Saxena A., et al., 2022a, arXiv e-prints, p. arXiv:2206.06161
Saxena A., et al., 2022b, MNRAS, 511, 120
Schaerer D., et al., 2020, A&A, 643, A3
Schaerer D., et al., 2022, A&A, 658, L11
Stanway E. R., Eldridge J. J., Becker G. D., 2016, MNRAS, 456, 485
Stark D. P., et al., 2015, MNRAS, 454, 1393
Teyssier R., 2002, A&A, 385, 337
Trebitsch M., Blaizot J., Rosdahl J., Devriendt J., Slyz A., 2017, MNRAS, 470, 224
Vallini L., Gallerani S., Ferrara A., Pallottini A., Yue B., 2015, ApJ, 813, 36
Vanzella E., et al., 2012, ApJ, 751, 70
Vanzella E., et al., 2016, ApJ, 825, 41
Vanzella E., et al., 2018, MNRAS, 476, L15
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., Heckman T. M., Leitherer C., Alexandroff R., Borthakur S., Overzier R. A., 2019, ApJ, 885, 57
Wise J. H., Cen R., 2009, ApJ, 693, 984
Zackrisson E., Inoue A. K., Jensen H., 2013, ApJ, 777, 39

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