

# Strong C IV emission from star-forming galaxies: a case for high Lyman continuum photon escape

A. Saxena<sup>1</sup>,<sup>\*</sup> E. Cryer,<sup>1</sup> R. S. Ellis,<sup>1</sup> L. Pentericci,<sup>2</sup> A. Calabrò,<sup>2</sup> S. Mascia<sup>2</sup>, A. Saldana-Lopez<sup>3</sup>,  
D. Schaerer,<sup>3,4</sup> H. Katz,<sup>5</sup> M. Llerena<sup>6</sup> and R. Amorín<sup>6,7</sup>

<sup>1</sup>Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK

<sup>2</sup>INAF – Osservatorio Astronomico di Roma, via Frascati 33, I-00078 Monteporzio Catone, Italy

<sup>3</sup>Department of Astronomy, University of Geneva, 51 Chemin Pegasi, CH-1290 Versoix, Switzerland

<sup>4</sup>CNRS, IRAP, 14 avenue E. Belin, F-31400 Toulouse, France

<sup>5</sup>Astrophysics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK

<sup>6</sup>Departamento de Física y Astronomía, Universidad de La Serena, Av. Juan Cisternas 1200 Norte, La Serena, Chile

<sup>7</sup>Instituto de Investigación Multidisciplinar en Ciencia y Tecnología, Universidad de La Serena, Raúl Bitrán, 1305 La Serena, Chile

Accepted 2022 September 22. Received 2022 September 12; in original form 2022 June 10

## ABSTRACT

Finding reliable indicators of Lyman continuum (Ly C) photon leakage from galaxies is essential in order to infer their escape fraction in the epoch of reionization, where direct measurements of Ly C flux are impossible. To this end, here we investigate whether strong C IV  $\lambda\lambda$ 1548, 1550 emission in the rest-frame UV spectra of galaxies traces conditions ripe for ample production and escape of Ly C photons. We compile a sample of 19 star-forming galaxies in the redshift range  $z = 3.1$ – $4.6$  from the VANDELS survey that exhibit strong C IV emission, producing a stacked spectrum where all major rest-UV emission lines are clearly detected. Best-fitting spectral energy distribution models containing both stellar and nebular emission suggest the need for low stellar metallicities ( $Z = 0.1$ – $0.2 Z_{\odot}$ ), young stellar ages ( $\log(\text{age yr}^{-1}) = 6.1$ – $6.5$ ), a high ionization parameter ( $\log U = -2$ ) and little to no dust attenuation ( $E(B - V) = 0.00$ – $0.01$ ). However, these models are unable to fully reproduce the observed C IV and He II line strengths. We find that the Ly  $\alpha$  line in the stacked spectrum is strong and peaks close to the systemic velocity, features that are indicative of significant Ly C photon leakage along the line of sight. The covering fractions of low-ionization interstellar absorption lines are also low, implying Ly C escape fraction in the range  $\approx 0.05$ – $0.30$ , with signatures of outflowing gas. Finally, C IV/C III] ratios of  $> 0.75$  for a subset of individual galaxies with reliable detections of both lines are also consistent with physical conditions that enable significant Ly C leakage. Overall, we report that multiple spectroscopic indicators of Ly C leakage are present in the stacked spectrum of strong C IV emitting galaxies, potentially making C IV an important tracer of Ly C photon escape at  $z > 6$ .

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

## 1 INTRODUCTION

Understand the contribution of star-forming galaxies in governing cosmic reionization, a process whereby the intergalactic medium (IGM) underwent a phase transition from a neutral to completely ionized gas, requires measures of their ionizing photon production efficiencies ( $\xi_{\text{ion}}$ ) the fraction of Lyman continuum (Ly C;  $\lambda_0 < 912 \text{ \AA}$ ) photons that manage to escape from the galaxy ( $f_{\text{esc}}$ ; see Dayal & Ferrara 2018 for a review) and the integrated space density of galaxies, derived from UV luminosity functions (e.g. Robertson et al. 2013, 2015; Bouwens et al. 2015; Finkelstein et al. 2015).

The nature and strength of the ionizing radiation emerging from young stars within star-forming galaxies can be derived from emission lines seen at rest-frame UV and optical wavelengths (e.g. Feltre,

Charlot & Gutkin 2016; Gutkin, Charlot & Bruzual 2016; Xiao, Stanway & Eldridge 2018; Plat et al. 2019), including the Balmer H  $\alpha$  line whose intensity is related to the number of ionizing photons produced through recombination physics (e.g. Shivaee et al. 2018). Detailed spectroscopic studies of galaxies at intermediate redshifts have provided reliable measurements of  $\xi_{\text{ion}}$  and we can expect continued progress within the reionization era ( $z \gtrsim 6$ ) following the successful deployment of the *James Webb Space Telescope* (JWST).

Due to the increasing neutrality of the IGM at higher redshifts, however, direct measurements of Ly C radiation from galaxies become challenging at  $z \gtrsim 4$  (e.g. Inoue et al. 2014). A popular strategy for estimating  $f_{\text{esc}}$  for galaxies in the reionization era is to study ‘analogues’ (i.e. with comparable physical properties) of  $z > 6$  galaxies at intermediate redshifts ( $z \sim 3$ ), for which Ly C leakage can be directly observed (e.g. Shapley et al. 2006). With a detailed understanding of the ionizing output of stars and conditions of the interstellar medium (ISM) that enable significant escape of Ly C photons (e.g. Nakajima et al. 2020; Saxena et al. 2022) in analogous

\* E-mail: [aayush.saxena@ucl.ac.uk](mailto:aayush.saxena@ucl.ac.uk)

intermediate redshift galaxies, it may then be possible to infer which types of sources are likely to be the key drivers of cosmic reionization.

Although several dedicated searches have identified Ly C leaking galaxies from large-area surveys at intermediate redshifts, their number remains quite modest (see Meštrić et al. 2021, for a recent compilation). Inference from ground-based data, in particular, is restricted mostly to luminous galaxies (e.g. Grazian et al. 2016; Guaita et al. 2016; Marchi et al. 2017; Naidu et al. 2018; Steidel et al. 2018; Saxena et al. 2022). A further complication in assessing the statistics of Ly C leakers in any given survey is the expectation from numerical simulations that Ly C leakage may be highly anisotropic, such that detections are only possible when the leaking channels are favourably aligned to the line-of-sight to the observer (e.g. Barrow et al. 2020; Katz et al. 2020; Kimm et al. 2022), with some observational evidence to support this idea (e.g. Vanzella et al. 2021).

Rather than undertaking a systematic survey for Ly C leakage in a sample of photometrically selected star-forming galaxies, it may be more productive to study galaxies whose emission line properties are indicative of emission of copious amounts of ionizing photons with high ionization parameters. Specific examples include sources with high [O III]  $\lambda$ 5007/[O II]  $\lambda$ 3727 ratios (e.g. Izotov et al. 2018; Nakajima et al. 2018a), which may reflect density-bound nebulae from which Ly C leakage may be possible (Zackrisson, Inoue & Jensen 2013; Nakajima & Ouchi 2014), as well as those with high ionization energy lines which may reflect young stellar populations whose associated supernovae can clear channels that permit free passage of Ly C photons (Berg et al. 2019; Nanayakkara et al. 2019; Saxena et al. 2020b; Senchyna et al. 2021; Tang et al. 2021; Vanzella et al. 2021).

Following this strategy, in this paper we focus on consideration of the C IV  $\lambda\lambda$  1548, 1550 doublet, whose ionization energy is 47.9 eV and has already been detected in rest-UV spectra of several  $z > 6$  galaxies (e.g. Stark et al. 2015; Mainali et al. 2017; Schmidt et al. 2017, although it remains unclear whether the C IV emission is purely due to star formation or due to active galactic nuclei or AGNs). In the local universe, strong nebular C IV emission is almost exclusively seen in low-mass galaxies with extremely low metallicities ( $\lesssim 0.1 Z_{\odot}$ , where  $Z_{\odot}$  is the solar metallicity value), young stellar ages ( $\log(\text{age yr}^{-1}) \lesssim 7$ ), and high specific star formation rates ( $\log(\text{sSFR yr}) < -8$ ; Berg et al. 2016, 2018; Senchyna et al. 2017, 2019), properties that are likely common amongst galaxies in the reionization era.

There is also growing evidence that strong C IV emission may be associated with Ly C leakage: Schaerer et al. (2022) found strong C IV emission ubiquitously in a sample of  $z < 0.7$  galaxies, and strong C IV has also been observed in a confirmed Ly C leaker at  $z \sim 3$  (e.g. Vanzella et al. 2016). Schaerer et al. (2022) interpreted their results with an elevated level of  $\xi_{\text{ion}}$  and the presence of density-bound H II regions. At low gas-phase metallicities, these conditions increase both the C IV luminosity as well as the C IV/C III] line ratio. Additionally, as a resonant line, C IV may trace photon escape through high-ionization gas (Berg et al. 2019) and its typical P-Cygni profile is a valuable indicator of outflows driven by massive stars (e.g. Steidel et al. 2016) that are necessary to clear out channels for Ly C escape. All the foregoing suggests C IV emission may be an important pointer to Ly C leakage.

In this paper, we investigate the properties of a sample of star-forming galaxies with spectroscopic redshifts at  $z \sim 3.1$ –4.6 selected purposely to have strong C IV emission. Our aim is to better understand the spectroscopic properties of strong C IV emitting galaxies and via independent spectroscopic measures, investigate the presence

of signatures that may point towards significant Ly C photon leakage. Our study complements ongoing efforts to understand the properties of star-forming galaxies that leak Ly C photons at lower redshifts (e.g. Low-Redshift Lyman Continuum Survey, Flury et al. 2022a,b and the COS Legacy Archive Spectroscopy Survey, Berg et al. 2022). Ultimately, our study aims to provide a reference for an improved understanding and interpretation of *JWST*/NIRSpec spectra of star-forming galaxies at  $z > 6$  in the context of Ly C photon escape.

The layout of this paper is as follows. We describe the spectroscopic data, identification of C IV emitting galaxies, emission line measurements and flagging of potential AGN in Section 2. We present a stacked spectrum of C IV emitting galaxies along with a detailed analysis of other strong rest-frame UV emission lines in Section 3. We explore the nature of the underlying sources of ionization that best describe the observed C IV emission (and other lines) in the stack in Section 4. Finally, we investigate whether significant Ly C photon leakage can be inferred from the stacked spectrum of C IV emitting galaxies using other indirect spectral signatures in Section 5, summarizing our findings in Section 6.

Throughout this paper, we assume a  $\Lambda$ CDM cosmology with  $\Omega_{\text{m}} = 0.3$  and  $H_0 = 67.7 \text{ km s}^{-1} \text{ Mpc}^{-1}$  taken from Planck Collaboration et al. (2016). All logarithms are in base 10, unless otherwise specified. In this paper, we adopt a solar metallicity value of  $Z_{\odot} = 0.02$ .

## 2 DATA

We use spectroscopic data from VANDELS – a deep VIMOS survey of the CANDELS fields – which is a recently completed ESO public spectroscopic survey carried out using the VLT. VANDELS covers two well-studied extragalactic fields, the UKIDSS Ultra Deep Survey (UDS) and the Chandra Deep Field South (CDFSGOODS-S). We refer the readers to McLure et al. (2018) for details about the survey description and target selection, and to Pentericci et al. (2018) for more information about data reduction and spectroscopic redshift determination. The final VANDELS data release, DR4,<sup>1</sup> contains spectra of  $\sim 2100$  galaxies in the redshift range  $1.0 < z < 7.0$ , with on-source integration times ranging from 20 to 80 h, where  $> 70$  per cent of the targets have at least 40 h of integration time (Garilli et al. 2021). The spectral resolution of VANDELS spectra is  $R \sim 600$ .

The reliability of redshifts in the VANDELS data base is recorded using the following flags: 0 – no redshift could be assigned, 1 – 50 per cent probability to be correct, 2 – 70–80 per cent probability to be correct, 3 – 95–100 per cent probability to be correct, 4 – 100 per cent probability to be correct, and 9 – spectrum shows a single emission line. The typical accuracy of spectroscopic redshift measurements is  $\sim 150 \text{ km s}^{-1}$  (Pentericci et al. 2018).

### 2.1 A search for C IV emitters

In this work, we only select spectroscopically confirmed galaxies from VANDELS that have a redshift reliability flag of either 3 or 4, which guarantees that the redshift measured by the VANDELS team has a  $> 95$  per cent probability of being correct. This also ensures reliable detection of other emission or absorption features in the spectrum.

Since our main goal is to explore spectroscopic properties of galaxies that show strong C IV  $\lambda\lambda$  1548, 1550, we limit our focus to the redshift range  $z = 3.1$ –4.6 where additionally the Ly  $\alpha$  line is

<sup>1</sup><http://vandel.inaf.it/dr4.html>

visible. We find 735 galaxies across the CDFS and UDS fields in the redshift range  $z = 3.1\text{--}4.6$  and redshift reliability flags 3 or 4, which constitute the parent sample in this study.

C IV emission in this work is identified primarily via visual inspection. We first inspect the 1D spectra in the parent sample to search for C IV emission. If the C IV line coincides with a skyline residual, we discard the source as any line flux measurement would be relatively unreliable. Once C IV emission is identified in the 1D spectrum, we inspect the 2D spectrum to ensure that the emission is real and not due to hot pixels, noise peaks, and/or sky residuals. Two of the co-authors independently performed the visual identification and only those that were identified as C IV emitters by both individuals from 1D and 2D spectra were retained in our sample for further analysis. Out of 735 galaxies in the parent sample, we identified 22 sources as reliable C IV emitters. None of these are detected in the deep *Chandra* X-ray catalogues in either CDFS (Luo et al. 2017) or in UDS (Kocevski et al. 2018) fields, ruling out any clear AGN activity.

We note here that the goal of this study is not to identify a complete sample of C IV emitters in the VANDELs survey, but to explore the spectroscopic properties of sources that show clear evidence of strong C IV emission at intermediate redshifts. Therefore, we do not include galaxies with marginal C IV detections, or possible line emission that may be partly contaminated by a sky residual and would otherwise be considered as a real C IV line. A more statistically complete sample of C IV emitters from the VANDELs survey will be assembled in a future study (Mascia et al., in preparation).

## 2.2 Rest-UV emission line measurements

We measure line fluxes of C IV and other rest-frame UV emission lines visible in the spectra, which include Ly $\alpha$ , He II  $\lambda 1640$ , and O III]  $\lambda\lambda 1660, 1666$ . The line fitting is performed using the PYTHON package MPDAF<sup>2</sup> in a similar fashion to Saxena et al. (2020b). Briefly, we fit the observed emission lines with single or double Gaussian functions and measure the local continuum level in a wavelength range free of other emission or absorption either side of the line, calculating the line fluxes, full width at half maxima (FWHM) and rest-frame equivalent widths ( $EW_0$ ). Below we give a summary of the rest-frame UV lines identified in the individual spectra in this work.

By design, we identify C IV emission in all 22 objects. The integrated line fluxes range from 0.2 to  $6.0 \times 10^{-18} \text{ erg s}^{-1} \text{ cm}^{-2}$ , with rest-frame equivalent widths ranging from  $EW_0 = 0.7\text{--}12.8 \text{ \AA}$ .

18 out of 22 objects show clear Ly $\alpha$  emission with a range of line strengths, widths, and profiles. Three sources do not show any Ly $\alpha$  emission in their spectra and the Ly $\alpha$  line in one source is contaminated by a sky line.

Nine sources also show He II emission with signal-to-noise ratio ( $S/N \geq 2$ ), five of which were also identified by Saxena et al. (2020b) using an earlier data release of VANDELs. The integrated He II line fluxes range from 0.2 to  $7.4 \times 10^{-18} \text{ erg s}^{-1} \text{ cm}^{-2}$  and  $EW_0$  in the range 2.0–18.7  $\text{\AA}$ .

Nine sources show the O III]  $\lambda\lambda 1660, 1666$  emission lines as well, which often appear to be blended and therefore only the total O III] flux is measured and reported. The O III] line fluxes range from 0.4 to  $7.2 \times 10^{-18} \text{ erg s}^{-1} \text{ cm}^{-2}$ , with  $EW_0$  ranging from 1.5–11.3  $\text{\AA}$ .

The wavelength coverage of VANDELs spectra allows the detection of C III]  $\lambda\lambda 1907, 1909$  (which appear to be blended) only at  $z \lesssim$

3.9 in our sample. Additionally, since the C III] line lies in a relatively red part of the observed spectrum, it is more prone to contamination by skyline residuals. Therefore, we only robustly identify C III] in 5 objects, with line fluxes ranging from  $0.7$  to  $4.3 \times 10^{-18} \text{ erg s}^{-1} \text{ cm}^{-2}$ , and  $EW_0$  in the range 3.3–17.2  $\text{\AA}$ , which is comparable to a more statistical measurement of C III] from VANDELs galaxies presented by Llerena et al. (2021).

The full suite of Ly $\alpha$ , C IV, He II, and O III] lines are detected with  $S/N \geq 2$  for only three sources in our sample, making it necessary to perform spectral stacking to boost  $S/N$  of other rest-UV features necessary to characterize the properties of the underlying sources of ionizing photons (see Section 3). However, before producing a stacked spectrum of C IV emitters, we attempt to identify sources that may be dominated by AGN activity based on their C IV and He II line strengths.

## 2.3 Identifying possible AGN

AGNs are capable of producing a much larger number of ionizing photons through the accretion of material on to the central super-massive black hole, which results in a non-thermal spectral energy distribution that can easily produce extremely high energy photons giving rise to higher order transitions of the most common elements, such as C IV, He II, and N V.

Since our sample is selected on the C IV line and the second most common nebular emission line seen in individual galaxy spectra is He II, we use model predictions built around these two lines to identify possible AGNs in our sample. In particular, we employ the diagnostic from Nakajima et al. (2018a), comparing the C IV/He II ratio against  $EW_0(\text{C IV})$ . The Nakajima et al. (2018a) emission line predictions are obtained using the photoionization code CLOUDY (Ferland et al. 2013). To model the spectral energy distribution (SED) of star-forming galaxies, the authors employ BPASS models (Stanway, Eldridge & Becker 2016) that include the effect of interacting binary stars. The AGN models on the other hand assume a power-law ionizing radiation field emanating from the narrow-line region surrounding the active black hole.

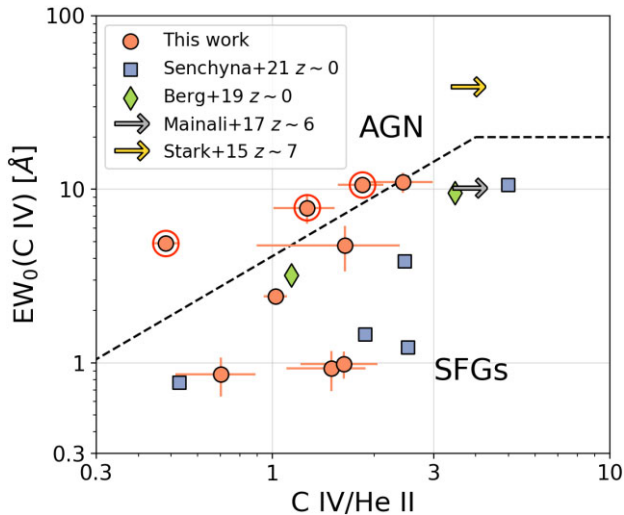
In Fig. 1, we show  $EW_0(\text{C IV})$  and C IV/He II measured in the spectra of galaxies in our sample where both lines were detected with  $S/N > 2$ . The dashed line demarcates the parameter space permitted by ionization due to star formation alone and ionization due to AGNs. For comparison, we also show measurements from low metallicity dwarf galaxies in the local Universe (Berg et al. 2019; Senchyna et al. 2021), as well as C IV measurements from a bright galaxy at  $z \sim 7$  reported by Stark et al. (2015) and at  $z \sim 6.1$  reported by Mainali et al. (2017).

We identify three sources in our sample that are likely to be dominated by AGN, whereas six sources occupy the parameter space where star formation alone is enough to explain the C IV and He II emission. For sources with no He II detection, the resulting lower limit on C IV/He II  $\gtrsim 2.5$  and  $EW_0(\text{C IV}) < 10 \text{ \AA}$  is suggestive of ionization due to star formation. Since our goal is to investigate C IV emission exclusively from star-forming galaxies, we remove these three potential AGN from our sample, leaving us with a final sample of 19 star-forming galaxies that show C IV emission in their spectra. Further, we do not find any individual detections for these sources in the deep X-ray data available in the fields.

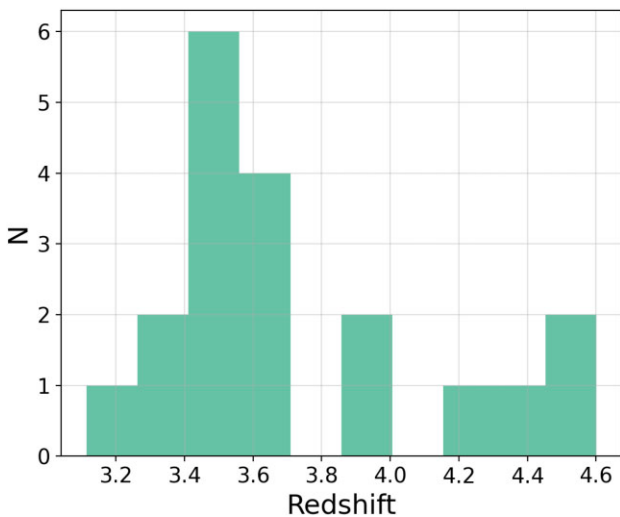
We do note, however, that a few star-forming galaxies in our sample lie very close to the dividing line between photoionization by AGN and star formation. The parameter space occupied by our C IV and He II emitting galaxies is also comparable to the low mass, metal-poor galaxies in the nearby Universe that show high ionization

<sup>2</sup><https://mpdaf.readthedocs.io/en/latest/index.html>





**Figure 1.** Distribution of C IV/He II flux ratio versus C IV rest-frame equivalent width for individual galaxies in this work where both emission lines are securely detected. The division between ionization due to star formation alone (SFG) and ionization due to AGN has been adapted from the modelling of Nakajima et al. (2018a). Also shown for comparison are measurements from local low mass, low metallicity galaxies (Berg et al. 2019; Senchyna et al. 2021), as well as the C IV detection from a galaxy at  $z \sim 7$  reported by Stark et al. (2015) and at  $z \sim 6$  reported by Mainali et al. (2017). We find that three C IV emitters are likely to be dominated by AGN that are marked with red circles, and six galaxies can be explained through star formation activity alone, lying close to the AGN–SFG dividing line implying the presence of sources hard ionizing radiation. The three sources identified as AGN are removed from our final sample.



**Figure 2.** Redshift distribution of 19 C IV emitting star-forming galaxies in our final sample, selected from the VANDELS survey. The galaxies cover a redshift range of  $3.1 < z < 4.6$ , with a median redshift of  $z = 3.6$ .

lines. We also note that the very strong C IV detection in the  $z \sim 7$  galaxy reported by Stark et al. (2015) suggests photoionization due to AGNs; however, C IV measurement from the galaxy at  $z \sim 6.1$  from Mainali et al. (2017) is comparable to what we find in our sample.

A histogram of the redshift distribution of 19 star-forming galaxies with strong C IV emission in our final sample is shown in Fig. 2. The median redshift of our sample is 3.6 and the sources span an observed

$i$ -band magnitude range of 24.4–26.7 AB. Using these 19 galaxies, we now produce a stacked spectrum in the following section.

### 3 A STACKED SPECTRUM OF C IV EMITTERS AT $3 < z < 4.5$

As mentioned earlier, detecting the full suite of rest-UV emission and absorption features requires high S/N of the underlying continuum. Therefore, to boost S/N of these features to study the sample-averaged properties of C IV emitting galaxies and better understand their underlying stellar populations and state of the ISM, in this section we produce a stacked spectrum of C IV emitting galaxies.

#### 3.1 Stacking procedure

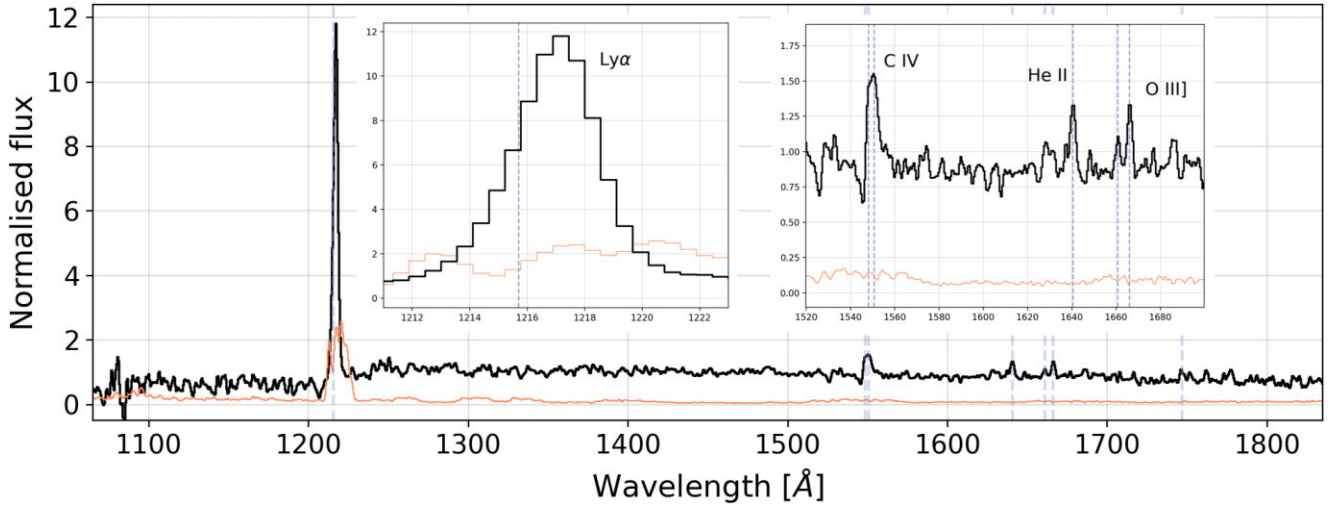
The method to co-add spectra adopted in this study is similar to Saxena et al. (2020b) – the stacking is performed by first deredshifting each spectrum using the ‘systemic’ redshifts derived primarily from the strong C IV emission lines, as well as He II, O III], and C III] lines whenever visible in the individual galaxy spectrum. For galaxies where two or more of the above-mentioned lines were detected, we found that the redshift derived from each line was within  $100 \text{ km s}^{-1}$  of each other, well within the resolution element of the spectrograph. The good agreement between the redshifts measured from these lines suggests the C IV line in emission generally traces systemic redshifts well across our galaxies, however, this may not be universally true as the C IV emission line is prone to resonant scattering.

The rest-frame spectra are normalized using the mean flux density value in the wavelength range 1460–1540 Å. Each spectrum is then assigned a weight based on the standard deviation of flux in this wavelength range, where the weight is inversely proportional to the measured variance on the flux density at  $\sim 1500$  Å. Such an approach to calculating weights for stacking helps alleviate uncertainties that may be introduced by weighting the stacking using pixel-by-pixel standard deviations, which could be highly variable across individual sources. Assigning a weight based on the SNR at 1500 Å, which is almost identical to luminosity weighted stacking, gives comparable results to that obtained by weighting using the standard deviation of pixels (e.g. Calabrò et al. 2022).

The spectra are then re-sampled to a uniform wavelength grid ranging from 1050 to 1820 Å, which is the rest-frame wavelength range most commonly probed by observed spectra in our sample, with a step size of 0.56 Å, which is the wavelength resolution obtained at a redshift of 3.6, the median redshift of our final sample.

The wavelengths in the observed individual spectra that do not fall within the rest-frame wavelength grid for the stacked spectrum are masked on a source by source basis to avoid incomplete sampling of the fluxes across galaxies. This means that the C III] line is unfortunately not covered in the stacked spectrum. We also mask residual sky lines and hot pixels in the spectra by employing a  $>20\sigma$  clipping. A stacked spectrum is then produced using a weighted averaging procedure.

The standard deviation of the stacked spectrum will have contributions both from S/N limitations of individual spectra as well as source-by-source spectral variation in the individual sources (e.g. Jones, Stark & Ellis 2012). Therefore, to capture both these effects, we use a bootstrapping method to calculate uncertainties on the stacked spectrum. To do this, we take our sample of 19 C IV emitters and replace one randomly selected C IV emitter by a C IV non-emitter selected from the parent VANDELS sample. We then produce a stacked spectrum of these 19 objects, with 18 C IV emitters and one



**Figure 3.** Stacked spectrum (black) of 19 C IV emitting star-forming galaxies in our final sample, with  $1\sigma$  uncertainties (orange) obtained from bootstrapping. Multiple rest-frame UV emission lines are clearly detected in the stacked spectrum with high S/N. We show insets with the Ly  $\alpha$  line (left), and C IV  $\lambda\lambda$  1548, 1550, He II  $\lambda$  1640, and O III]  $\lambda\lambda$  1660, 1666 lines (right). We additionally mark the position of the blended N III] feature at  $\sim\lambda$  1747 line in the spectrum, which is not as strong as the other lines highlighted. The equivalent widths and FWHMs of emission lines in the stacked spectrum are given in Table 1.

non-emitter. We repeat this process of random replacement 500 times, consequently producing 500 stacked spectra with 18 C IV emitters and one non-emitter.

In this way, any strong spectral feature that is being contributed by an outlying C IV emitter can be accounted for, and the resulting standard deviation will be an accurate measure of the ‘true’ deviation of the stacked spectrum. The standard deviation of these bootstrapped spectra is used to calculate the  $1\sigma$  uncertainty on the final stacked spectrum of C IV emitters. The final stacked spectrum along with  $1\sigma$  errors is shown in Fig. 3.

A range of emission and absorption features are clearly detected in the stacked spectrum. In the sections that follow, we investigate the spectroscopic properties inferred from this stacked spectrum, and below we briefly summarize the observed properties of some of the brightest emission lines seen in the stack.

### 3.2 Emission lines

The emission lines in the stack are measured in a manner similar to that discussed in Section 2.2. The errors on both the line flux and width measurements are a consequence of the  $1\sigma$  uncertainty on the stacked spectrum calculated via bootstrapping.

The Ly  $\alpha$  line in the stack is strong and has a symmetric profile, with non-zero flux bluewards of the Ly  $\alpha$  peak. The peak of the Ly  $\alpha$  line is only slightly redshifted compared to the systemic redshift of the stacked spectrum. Such a line profile is typically associated with Ly C emitting galaxies across redshifts (e.g. Verhamme et al. 2017; Gazagnes et al. 2020), which we investigate further in Section 5.1. A single Gaussian function provides a good fit to the line, with a measured FWHM of  $820.0 \pm 17.0 \text{ km s}^{-1}$ , and a high  $\text{EW}_0 = 37.5 \pm 0.9 \text{ \AA}$ , placing it in the regime of strong Ly  $\alpha$  emitting galaxies (LAEs).

The C IV emission in the stacked spectrum is clear. Due to the relatively low spectral resolution the two components at  $\lambda$  1548 and  $\lambda$  1550 appear to be marginally blended, requiring a double Gaussian function fit to recover the full line flux. From this fit we recover a total  $\text{EW}_0(\text{C IV}) = 5.2 \pm 0.5$ . A P-Cygni absorption feature immediately bluewards of the  $\lambda$  1548 peak is also visible.

**Table 1.** Properties of emission lines clearly detected in the stacked spectrum.

Line/Vacuum $\lambda$ ( $\text{\AA}$ )	$\text{EW}_0$ ( $\text{\AA}$ )	FWHM ( $\text{km s}^{-1}$ )
Ly $\alpha$ $\lambda$ 1216	$37.5 \pm 0.7$	$830 \pm 18$
C IV $\lambda\lambda$ 1548, 1550	$5.2 \pm 0.5$	–
He II $\lambda$ 1640	$2.2 \pm 0.1$	$592 \pm 23$
O III] $\lambda$ 1660	$1.1 \pm 0.3$	$547 \pm 34$
O III] $\lambda$ 1666	$2.0 \pm 0.1$	$502 \pm 31$
N III] $\lambda$ 1747	$1.2 \pm 0.3$	$440 \pm 120$

*Note.* Since the C IV doublet appears to be blended and is fitted using a combination of two Gaussian functions, we do not report the FWHM.

We identify strong He II emission in the stacked spectrum and the emission line is fitted using a single Gaussian function, giving an  $\text{FWHM}(\text{He II}) = 591.6 \pm 22.6 \text{ km s}^{-1}$  and  $\text{EW}_0(\text{He II}) = 2.2 \pm 0.1 \text{ \AA}$ . With He<sup>+</sup> ionizing potential being  $\approx 54.4 \text{ eV}$ , He II emission is indicative of the presence of young stellar populations with low metallicities, capable of producing highly energetic photons.

We identify both O III]  $\lambda$  1660, 1666 emission lines, with  $\lambda$  1660 being fainter than  $\lambda$  1666. The doublet is sufficiently separated in wavelength such that two independent Gaussian functions can be fit to the lines. We measure a FWHM of  $518.3 \pm 25.9 \text{ km s}^{-1}$  for  $\lambda$  1660 and  $611.0 \pm 30.8 \text{ km s}^{-1}$  for  $\lambda$  1666. The total  $\text{EW}_0(\text{O III])}$  is  $3.2 \pm 0.3 \text{ \AA}$ , with a O III]  $\lambda$  1660/O III]  $\lambda$  1666 ratio of  $\approx 0.5$ .

We also find N III] emission around 1748  $\text{\AA}$ , which is a quintuplet with lines at 1746, 1748, 1749, 1750, and 1752  $\text{\AA}$ , but appears to be blended in the spectrum. The N III] emission has a lower ionizing potential of  $\approx 26.9 \text{ eV}$  compared to O III] and other stronger lines visible in the stacked spectrum. The presence of N III] emission has not been widely reported – including from stacked spectra – in star-forming galaxies in the literature (but see Le Fèvre et al. 2019). However, the presence of N III] is not surprising considering that the sources were selected on emission lines requiring much higher ionizing energies.

The  $\text{EW}_0$  and FWHM of the above-mentioned lines are summarized in Table 1. In the following section, we compare line measurements from our stacks to other stacked spectra of star-forming galaxies produced at comparable redshifts, as well as C IV

emitting galaxies identified both in the local Universe as well as at  $z > 6$ .

### 3.3 Comparison with literature

The  $EW_0(\text{Ly } \alpha)$  from our stack is comparable to that measured in the stacked spectrum of LBGs at  $z \sim 3$  from Steidel et al. (2018) that were classified as LAEs ( $EW_0(\text{Ly } \alpha) = 44.1 \text{ \AA}$ ), and the stack of the quartile of their sources with the strongest  $\text{Ly } \alpha$  emission, dubbed WQ4 ( $EW_0(\text{Ly } \alpha) = 43.2 \text{ \AA}$ ). Both LAEs and WQ4 galaxies in Steidel et al. (2018) were identified to be the strongest LyC leakers using spectroscopic measurements. Interestingly, the strength of the C IV emission also appears to increase with increasing  $\text{Ly } \alpha$  in the stacked spectra of Steidel et al. (2018), but no measure of C IV emission across stacks is given.

Comparing with the stacks of narrow-band selected LAEs reported by Nakajima et al. (2018b), we find that our  $EW_0(\text{Ly } \alpha)$  is closest to that measured from the stack of UV-luminous LAEs having  $EW_0(\text{Ly } \alpha) = 38\text{--}40 \text{ \AA}$ . Prominent C IV features were also reported in their stacks of LAEs with large  $EW_0(\text{Ly } \alpha)$  and fainter UV magnitudes, with  $EW_0(\text{C IV})$  in the range  $2.9\text{--}3.9 \text{ \AA}$ . The  $EW_0(\text{C IV})$  measured by Nakajima et al. (2018b) even for their strongest LAEs is lower than what we observe.

We then compare our measurements with the stacked spectra of LAEs at  $3 < z < 4.6$  in the MUSE HUDF reported by Feltre et al. (2020). The C IV doublet in emission is seen in their stack of LAEs that have low FWHM( $\text{Ly } \alpha$ ), high  $EW_0(\text{Ly } \alpha)$  and faint UV magnitudes. The  $EW_0(\text{C IV})$  ranged from  $1.95$  to  $4.74 \text{ \AA}$ . A slightly weaker He II emission with  $EW_0(\text{He II}) \approx 1\text{--}1.3 \text{ \AA}$  was also reported from their stacks.

Only a handful of C IV observations currently exist at  $z > 6$ . The C IV emitter identified by Stark et al. (2015) at  $z \sim 7$  shows a higher  $EW_0(\text{Ly } \alpha) = 65 \pm 12 \text{ \AA}$  compared to our stack, a significantly higher  $EW_0(\text{C IV}) \approx 40 \text{ \AA}$  with  $EW_0(\text{He II}) < 11.4 \text{ \AA}$ . However, there are suggestions that this source may be powered by an AGN, as we demonstrated in Section 2.3. The lensed C IV emitter at  $z = 6.11$  from Schmidt et al. (2017) also has a higher  $EW_0(\text{Ly } \alpha) = 68 \pm 6 \text{ \AA}$  and higher  $EW_0(\text{C IV}) = 24 \pm 4 \text{ \AA}$  compared to our stack. Another lensed C IV emitter at  $z \sim 6.1$  reported by Mainali et al. (2017) has a comparable  $EW_0(\text{Ly } \alpha) = 40 \pm 5 \text{ \AA}$  with  $EW_0(\text{C IV}) \approx 10$ , which is higher than what is seen in our stack.

C IV emission was identified in an ultra-deep ( $\approx 140$  h) MUSE spectrum of a galaxy at  $z = 4.77$  (Matthee et al. 2022) with  $EW_0(\text{C IV}) \approx 5 \text{ \AA}$ . C IV emission has also been found in a lensed galaxy at  $z = 4.88$  (Witstok et al. 2021) with  $EW_0(\text{C IV}) \approx 18 \text{ \AA}$ . These galaxies also show strong  $\text{Ly } \alpha$  emission with  $EW_0(\text{Ly } \alpha) \approx 62 \text{ \AA}$  and  $\approx 143 \text{ \AA}$ , respectively.

Turning our attention to comparable observations in the local Universe, Senchyna et al. (2019) and Berg et al. (2019) reported the detection of strong C IV emission with  $EW_0(\text{C IV}) \approx 3\text{--}10 \text{ \AA}$  from two low-mass, metal-poor galaxies at  $z \sim 0$  that also show strong He II emission. Strong and narrow He II emission has also been detected in these galaxies with  $EW_0(\text{He II}) \approx 2.8$ . Both studies concluded that stellar metallicities less than 10 per cent solar (e.g. Senchyna et al. 2021) are required to explain this emission, and that these galaxies are likely analogues of metal-poor star-forming systems in the reionization era.

Interestingly, rest-frame UV spectroscopy of LyC leakers at  $z < 0.7$  has also resulted in the detection of strong C IV emission in all galaxies with  $f_{\text{esc}} > 0.1$  (Schaerer et al. 2022), with some of the highest  $EW_0(\text{C IV})$  seen in low- $z$  star-forming galaxies.

The  $EW_0(\text{C IV})$  for a majority of their LyC leaking sources are comparable to that of our stacked spectrum.

Overall, the properties of emission lines in the stacked spectrum of C IV emitters presented in this study appear to be consistent with those seen in strong LAEs at  $z \sim 3\text{--}5$ , and are somewhat representative of the limited detections of rest-UV emission from bright galaxies at  $z > 6$ . The line strengths in our sample also resemble those from extremely low-metallicity galaxies in the local Universe, as well as strong LyC leakers at  $z \lesssim 1$ .

## 4 UNDERLYING SOURCES OF IONIZING PHOTONS

Having produced a stacked spectrum of strong C IV emitters and identified prominent rest-frame UV lines, in this section we compare the properties of the stacked spectrum with photoionization models in a bid to constrain the dominant mechanism of ionizing photon production within C IV emitting galaxies.

### 4.1 Spectral energy distribution (SED) fitting

To understand the nature of ionizing sources that could give rise to strong C IV emission (as well as other lines), we now find the best-fitting SED model for our stacked spectrum. Since the main aim of fitting SEDs to our stacked spectrum of C IV emitters is to broadly investigate what physical conditions within galaxies may give rise to strong C IV (and other line) emission, we choose to compare with fairly simple stellar population synthesis models, opting to use SEDs containing stellar continuum, nebular continuum and nebular line emission produced by the BPASS team (e.g. Xiao et al. 2018). These models are built on stellar SEDs generated using BPASS v2.2.1 for a single age starburst, with stellar ages in the range  $\log(\text{age yr}^{-1}) = 6\text{--}7.5$  (Stanway & Eldridge 2018; Xiao et al. 2018). These SEDs are generated using the default BPASS initial mass function (IMF), which is a broken power law with a slope of  $-1.30$  for  $M_* < 0.5 M_\odot$  and a slope of  $-2.35$  for  $M_* > 0.5 M_\odot$ , with an upper mass cutoff at  $300 M_\odot$ .

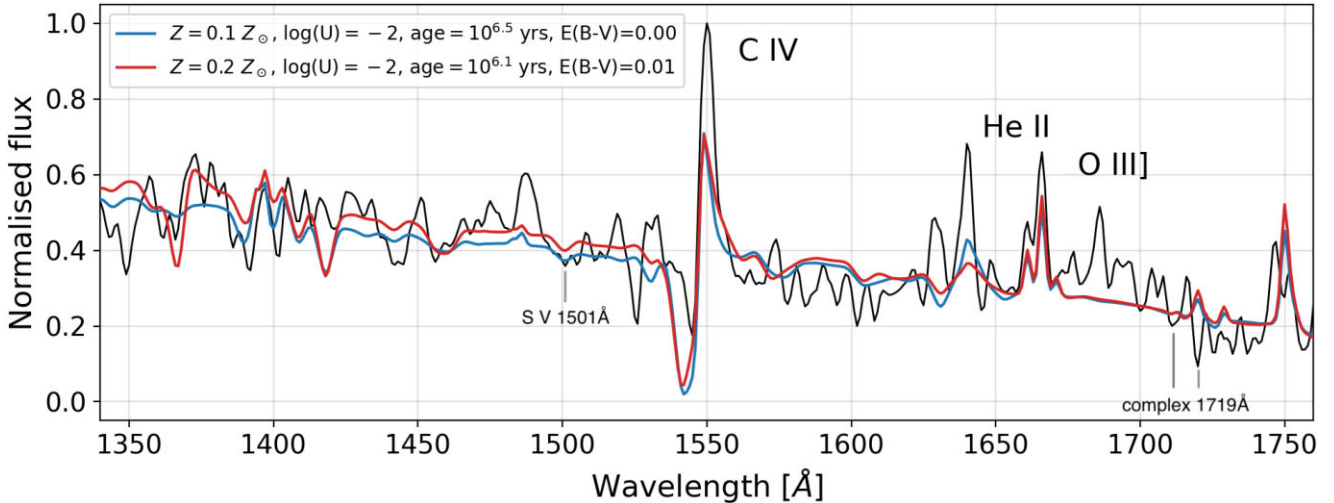
These simple stellar populations models are then processed using the photoionization code CLOUDY, assuming a nebular gas cloud with density  $\log(n_{\text{H}}/\text{cm}^{-3}) = 2.3$ , a spherical geometry and the same metallicity of stars and nebular gas to calculate nebular line fluxes.<sup>3</sup> For this analysis, we consider CLOUDY processed models containing stellar and nebular emission with stellar metallicities in the range  $Z = 0.005 Z_\odot$  to  $Z = 2 Z_\odot$  and the dimensionless ionization parameter,  $\log(U)$ ,<sup>4</sup> in the range  $[-1.0, -4.0]$ . We additionally include a dust attenuation prescription derived by Reddy et al. (2016a) for  $z \sim 3$  galaxies at rest-frame UV wavelengths, with the same attenuation applied to both nebular lines and the stellar continuum, considering  $E(B - V)$  values in the range  $[0, 1]$  in steps of 0.01.

Since the  $\text{Ly } \alpha$  line is prone to high levels of scatter and due to the increased noise levels at the edges of the stacked spectrum, we focus our model comparison on the wavelength range  $1300\text{--}1820 \text{ \AA}$  that contains C IV, He II, and O III] lines. It is important to match the

<sup>3</sup><https://flexiblelearning.auckland.ac.nz/bpass/4.html>

<sup>4</sup>SED models often tend to employ the dimensionless ionization parameter,  $U$ , which gives the ratio of the density of ionizing photons to the density of hydrogen. The ionizing photon production efficiency or  $\xi_{\text{ion}}$ , on the other hand is the number of ionizing photons produced per unit UV luminosity. Under assumptions of hydrogen density for a given UV luminosity,  $U$  and  $\xi_{\text{ion}}$  are closely related to one another and broadly trace the ionizing photon production capabilities of a source.





**Figure 4.** Zoom-in of the wavelength range covering C IV, He II, and O III] emission lines in the stacked spectrum of C IV emitters. The spectrum has been normalized using minimum–maximum scaling to aid comparisons with SED models (see text in Section 4.1), which have also been smoothed to the VANDELS spectral resolution. We show the two best-fitting spectral energy distribution (SED) models. Model 1 (blue) has a stellar metallicity of  $Z = 0.1 Z_{\odot}$ ,  $\log(\text{age yr}^{-1}) = 6.5$ , an ionization parameter of  $\log(U) = -2.0$  and no dust attenuation. Model 2 (red) has a higher stellar metallicity of  $Z = 0.2 Z_{\odot}$ , a lower stellar age of  $\log(\text{age yr}^{-1}) = 6.1$ , an ionization parameter of  $\log(U) = -2.0$  and a small dust attenuation of  $E(B - V) = 0.01$ . We find that both models are able to reproduce the P-Cygni profile of C IV, but underpredict the C IV and He II lines. The models match the O III] doublet strength as well as their ratio reasonably well. Overall, the best-fitting SED models suggest that young and low-metallicity starbursts with a high ionization parameter are needed to explain the observed C IV (and He II) in the stacked spectrum. We additionally mark the locations of the absorption features used to independently measure stellar metallicities in Section 4.2: S V at 1501 Å and a blend of N IV, Si IV, Al II, and Fe IV at 1719 Å. Finally, we note that the emission feature at 1486.5 Å is from N IV] and the feature spotted near  $\approx 1680$  Å is likely a noise fluctuation.

normalization of observed spectra with synthetic SEDs generated by models to enable accurate comparison. To this end, in this study we employ the *MinMaxScaler* routine,<sup>5</sup> which scales and translates the features in a given array on to a finite range, in this case between zero and one. When this scaling is applied consistently across the observed spectrum and the model SEDs, an accurate comparison between data and models becomes possible.

To find the best-fitting SED to our stacked spectrum, we employ the root-mean-square deviation (RMSD) estimator. The RMSD estimator measures the differences between values predicted by a model and data, where the deviations represent the residuals. Low RMSD values indicate the closest agreement between models and data. We populate a grid of models with varying combinations of metallicities, ages, ionization parameters and dust attenuation in the ranges described above, calculating the RMSD for each model compared to the stacked spectrum. We find that two SEDs in particular give comparably good fits to the stacked spectrum, having the lowest RMSD values, which are described below and shown in Fig. 4.

We note that when fitting SEDs we do not mask faint ISM absorption lines, which are not present in the models that we use. Owing to the presence of strong emission lines and relatively weak ISM absorption, we find that the principal components driving the model fit to the data are the multiple strong emission lines in the stacked spectrum, followed by the UV slope that covers a broad range of wavelengths.

The first model (Model 1) has a stellar metallicity of  $Z = 0.1 Z_{\odot}$ , a stellar age of  $\log(\text{age yr}^{-1}) = 6.5$ , an ionization parameter  $\log(U) = -2.0$  with no dust attenuation,  $E(B - V) = 0.0$ , shown in blue

**Table 2.** Properties of the best-fitting BPASS+LOUDY SED models.

	$Z/Z_{\odot}$	$\log(\text{age yr}^{-1})$	$\log(U)$	$E(B - V)$	RMSD
Model 1	0.1	6.5	-2.0	0.00	0.098
Model 2	0.2	6.1	-2.0	0.01	0.099

in Fig. 4. Model 1 broadly reproduces the C IV emission line’s P-Cygni profile but underpredicts the line flux and overpredicts the absorption bluewards of the peak. The He II emission in the model is also not enough to match the observed line emission; however, the O III] doublet line strengths and ratios, as well as N III] emission are well reproduced, with a good match to the observed UV slope of the stacked spectrum.

The second model (Model 2) has a higher stellar metallicity of  $Z = 0.2 Z_{\odot}$  compared to Model 1, with a lower stellar age of  $\log(\text{age yr}^{-1}) = 6.1$ ,  $\log(U) = -2.0$  and a small dust attenuation of  $E(B - V) = 0.01$ , shown in red in Fig. 4. The C IV line profile in Model 2 is nearly identical to Model 1, underpredicting the emission but overpredicting the absorption. The He II emission in Model 2 is lower when compared to Model 1, but the O III] and N III] emission are well reproduced. Model 2 also matches the observed UV slope in the spectrum. Both Models 1 and 2 have highly comparable RMSD values, and their properties are summarized in Table 2.

Both well-fitting SEDs imply that stellar populations with metallicities of  $0.1\text{--}0.2 Z_{\odot}$ , low stellar ages and relatively high ionization parameters are needed to reproduce the vast majority of the observed rest-UV emission lines and match the UV slope of the stacked spectrum of C IV emitting sources. Interestingly, the largest discrepancy between observations and model predictions are for emission lines requiring higher ionization energies. For example, the O III] and N III] lines that require energies of 35.1 and 29.6 eV, respectively, are relatively well produced by both models, but the C IV and He II

<sup>5</sup><https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html>

lines requiring energies of 47.9 and 54.4 eV are underpredicted. We note here that further increasing the ionization parameter of these models results in a catastrophic mismatch between the observed and predicted UV slopes, which is not significantly improved by increased dust attenuation, leading to sub-optimal fits.

The inability of these models to accurately match the observed C IV and He II emission suggests that the best-fitting SEDs are not producing ionizing radiation fields that are ‘hard’ enough, such that a large amount of photons with extremely high energies are produced, which we discuss further in Section 4.3. Further, the observed C IV absorption blueward of the peak is much weaker in the stacked spectrum compared to what both models predict, indicating that the best-fitting models may be more metal rich than the observations might suggest. Therefore, to obtain constraints on stellar metallicities we use additional spectroscopic indicators in the following section.

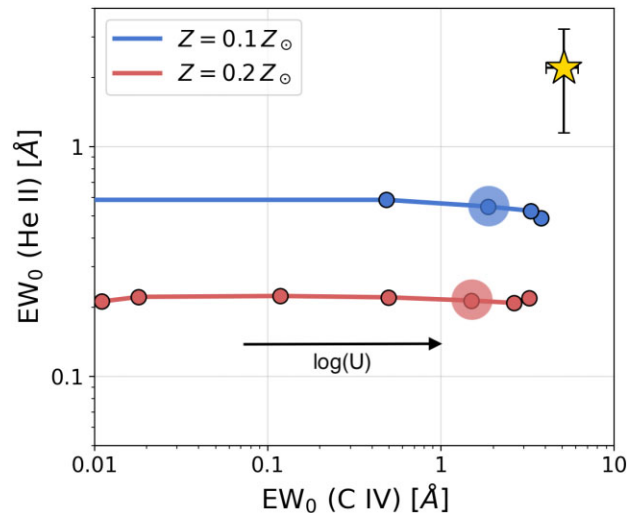
#### 4.2 Stellar metallicity from absorption indices

We now also measure the stellar metallicity of the stacked spectrum in a more direct manner using rest-frame UV absorption lines. We note the presence of absorption features due to ionized S V at  $\approx 1501 \text{ \AA}$  and due to a blend of N IV, Si IV, Al II, and Fe IV at  $\approx 1719 \text{ \AA}$  in the stack. These features arise as a result of absorption in the stellar photospheres of young, hot stars, and the strength of absorption can be used as a reliable tracer of stellar metallicity, independent of age of the stars or their initial mass function (e.g. Calabrò et al. 2021, and references therein) assuming a constant star formation history.

We note that these calibrations may need to be revised for other star formation histories, such as those predicting a larger fraction of older stars. However, given the smaller time-scales involved when modelling galaxies at  $z > 3$  older stars may not play a big role (see e.g. Cullen et al. 2019). Another possible source of uncertainty is the spectral resolution of the observations, which, as Calabrò et al. (2021) showed can introduce errors of up to  $0.2 Z_{\odot}$  in the inferred metallicities. Encouragingly, the spectral resolution used to calibrate the metallicity indicators in Calabrò et al. (2021) are at the VANDELS resolution, thereby minimizing its effects in this study.

We measure  $EW_0(1501) = -1.0 \pm 0.5 \text{ \AA}$  and  $EW_0(1719) = -1.3 \pm 1.0 \text{ \AA}$  and using the metallicity calibrations from Calabrò et al. (2021) that use BPASS models without any nebular emission, we obtain a stellar metallicity of  $Z = 0.0036 \pm 0.0025$  from the 1501 index and a comparable  $Z = 0.0041 \pm 0.0031$  from the 1719 index. Using the solar metallicity value of  $Z_{\odot} = 0.02$  as before, both these independent metallicity measurements from best fits imply  $Z \approx 0.2 Z_{\odot}$  within the uncertainties.

The best-fitting metallicity measurement is consistent with Model 2 presented in Section 4.1, although given the large uncertainties on the measurement a lower metallicity of  $\approx 0.1 Z_{\odot}$  suggested by Model 1 is also possible. From the best fits it appears that galaxies with strong C IV emission requiring high-energy ionizing photons do not necessarily have abnormally low metallicities when compared with measurements for galaxies at  $z \sim 3.5$  from VANDELS, which have stellar metallicities in the range  $Z \approx 0.1\text{--}0.2 Z_{\odot}$  (Cullen et al. 2020; Calabrò et al. 2021). The picture is different in the local Universe, however, where known C IV and He II emitters almost always have extremely low stellar (and gas phase) metallicities of the order  $Z \lesssim 0.1 Z_{\odot}$  (e.g. Berg et al. 2019; Senchyna et al. 2021). However, given the large uncertainties as well as the considerably stronger C IV absorption features in the best-fitting SEDs compared to observations, a very low metallicity solution for C IV emitters cannot be fully ruled out.



**Figure 5.** Comparison of the  $EW_0(\text{C IV})$  and  $EW_0(\text{He II})$  measured in the stacked spectrum (gold star) with predictions from best-fitting models: Model 1 with  $Z = 0.1 Z_{\odot}$ ,  $\log(\text{age yr}^{-1}) = 6.5$  (blue) and Model 2:  $Z = 0.2 Z_{\odot}$ ,  $\log(\text{age yr}^{-1}) = 6.1$  (red). The best-fitting ionization parameter values of  $\log(U) = -2$  for both models are highlighted. Particularly for Model 2, which has a consistent metallicity with that inferred from absorption features, the He II line strength is underpredicted by an order of magnitude and the C IV line strength by a factor of  $\sim 3$ . Additional sources of high-energy ionizing photons may be needed to explain the observed emission line strengths.

#### 4.3 Underprediction of C IV and He II

We note again that neither of the two best-fitting SEDs presented in Section 4.1 is able to fully reproduce the observed C IV and/or He II emission in the stacked spectrum. This is also shown clearly in Fig. 5, where the ionization parameter increases from right to left, with the highest value  $\log(U) = -1.0$ . The best-fitting  $\log(U)$  value of  $-2$  obtained from SED fitting has been highlighted for both models.

Looking at Model 2 with  $Z = 0.2 Z_{\odot}$ , the He II  $EW_0$  is underpredicted by an order of magnitude at  $\gtrsim 3\sigma$  significance, whereas the C IV line is underpredicted by a factor of  $\sim 3$ , also with  $\gtrsim 3\sigma$ . The discrepancy between observed and predicted equivalent widths also exists when comparing with Model 1, but Model 1 produces stronger He II in comparison to Model 2, bringing the tension down to  $\sim 2\sigma$ . The discrepancy with the C IV line, however, remains at  $\sim 3\sigma$ .

An underprediction of He II equivalent widths even by the lowest metallicity models from BPASS was previously also reported by Saxena et al. (2020b). To account for the ‘missing’ He II ionizing photons within such galaxies, Saxena et al. (2020b) suggested the inclusion of additional sources of ionization such as faint AGNs, stripped binary stars (e.g. Göteborg et al. 2019) or ultraluminous X-ray sources (e.g. Schaerer, Fragos & Izotov 2019; Saxena et al. 2020a; Simmonds, Schaerer & Verhamme 2021; Umeda et al. 2022). One additional scenario that may explain strong C IV and He II emission is a recent starburst event (preferentially from metal-poor gas) within galaxies that harbour relatively enriched ( $Z \gtrsim 0.2 Z_{\odot}$ ) stellar populations.

Evidence for the presence of both young and evolved stellar populations in high-redshift galaxies that show strong emission lines has been recently discussed: broad-band SED fitting and Atacama Large Millimetre Array (ALMA) observations of hyperfine transition metal lines of galaxies at  $z \gtrsim 9$  have suggested that these may already harbour evolved stellar populations (e.g. Roberts-Borsani, Ellis & Laporte 2020; Laporte et al. 2021). Recently, Tang, Stark & Ellis



(2022) also reported the presence of evolved stellar populations in extreme [O III]  $\lambda 5007$  emitting galaxies at redshifts 1.3–3.7, where this evolved population can be up to  $\sim 40$  times more massive than the young starburst associated with the extreme line emission.

Therefore, galaxies undergoing periods of starburst activity may temporarily be able to produce copious amounts of high-energy photons (i.e. periods of high  $\xi_{\text{ion}}$ ), giving rise to strong emission lines such as C IV, and He II, as well as [O III]  $\lambda 5007$ . Radiative transfer calculations in zoom-in simulations of galaxy formation from Barrow et al. (2020) showed that periods of high  $\xi_{\text{ion}}$  are often coincident with periods of extreme emission line strengths driven by starburst events. Periods of high Ly C  $f_{\text{esc}}$  tend to then follow, once the gas has been blown out and channels of low absorption have been established over a time-scale of  $\sim \text{Myrs}$  due to rampant supernova activity.

Having seen possible evidence of elevated ionizing photon production in C IV emitting galaxies, in the next section we investigate whether C IV emitters may also trace conditions that enable a higher fraction of Ly C photons to escape into the IGM to assess whether C IV emission in the reionization epoch may effectively trace Ly C leaking galaxies.

## 5 C IV EMITTERS AS TRACERS OF HIGH IONIZING PHOTON ESCAPE

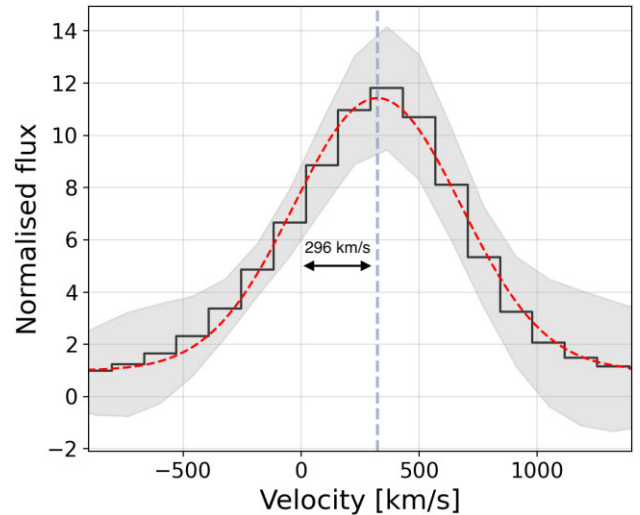
The primary goal of this paper is to investigate whether C IV emitting galaxies at intermediate redshifts trace conditions that might enable a high Ly C  $f_{\text{esc}}$ , which we test in this section using spectroscopic indicators in the stack.

### 5.1 Inference from Ly $\alpha$ strength and velocity offset

Useful information regarding the presence of channels in the ISM, allowing for a high escape fraction of Ly C photons, can be obtained from both the strength (e.g. Dijkstra 2014), the profile and the velocity offset compared to systemic of the Ly  $\alpha$  emission line (e.g. Verhamme et al. 2015). The redshift range of our targets ensures that the Ly  $\alpha$  line is visible for all galaxies, and in this section we use the observed Ly  $\alpha$  emission in the stacked spectrum to explore the possibility of high Ly C  $f_{\text{esc}}$  from C IV emitting galaxies.

It has been shown that galaxies with high Ly C  $f_{\text{esc}}$  are expected to have the peak of their Ly  $\alpha$  emission line close to the systemic redshift, with non-zero Ly  $\alpha$  flux bluewards of the systemic redshift (Verhamme et al. 2015; Dijkstra, Gronke & Venkatesan 2016). In this scenario, both Ly  $\alpha$  and Ly C photons would be able to escape the H II regions within which they are produced with relative ease through a porous ISM. This has also been verified observationally through Ly  $\alpha$  profiles of low redshift Ly C leakers, with an anticorrelation observed between Ly C  $f_{\text{esc}}$  and the separation of the Ly  $\alpha$  blue and red peaks as well as velocity offset from systemic (e.g. Izotov et al. 2018, 2021).

We find that the velocity offset of the Ly  $\alpha$  line in the stacked spectrum compared to the systemic redshift is relatively small at  $v \approx 296 \pm 20 \text{ km s}^{-1}$ , as shown in Fig. 6. We also note that the Ly  $\alpha$  emission is strong and appears to be symmetric, reproduced well by a single Gaussian function with non-zero flux bluewards of the Ly  $\alpha$  peak (see also Section 3.2). However, the emergent symmetrical profile in the stacked spectrum may be a consequence of stacking objects with a variety of Ly  $\alpha$  profiles and velocity offsets, with low spectral resolution additionally resolving out any intrinsic multiplex structure of Ly  $\alpha$  emission seen in individual objects. The strength and the low velocity offset compared to systemic, however,



**Figure 6.** Zoom-in of the Ly  $\alpha$  emission line in the stacked spectrum in units of velocity, with  $v = 0$  tracing the systemic velocity constrained using the C IV  $\lambda 1550$  feature in the stack. The velocity offset of the Ly  $\alpha$  peak is  $\approx 300 \text{ km s}^{-1}$  compared to the systemic velocity and there is significant flux bluewards of the Ly  $\alpha$  peak. The line profile is highly symmetric. All of these Ly  $\alpha$  features are consistent with expectations from galaxies that may or do indeed show a high escape fraction of Ly C photons (e.g. Verhamme et al. 2015, 2017).

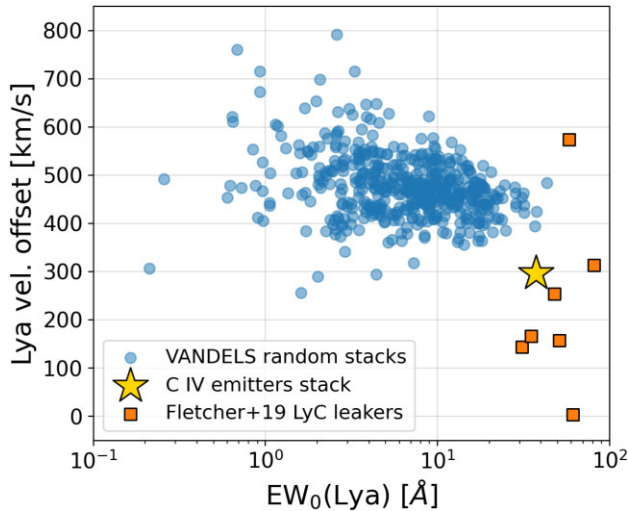
are consistent with predictions from Verhamme et al. (2015) of high Ly  $\alpha$  and Ly C photon escape.

To put the observed Ly  $\alpha$  strength and velocity offset observed in the stacked spectrum of C IV emitters in the global context of star-forming galaxies at  $z \sim 3.6$ , we compare our measurements to stacked spectra of 19 randomly selected star-forming galaxies from the VANDELS parent sample with the same range of UV luminosities, redshifts, and redshift quality flags as the C IV emitters. We repeat the stacking process 500 times following the methodology outlined in Section 3, obtaining a distribution of stacked spectra of non-C IV emitters randomly drawn from VANDELS. We then measure the strength of Ly  $\alpha$  emission in each stack as well as the offset of the peak of Ly  $\alpha$  emission from the expected ‘systemic’ redshift, using the spectroscopic redshifts compiled by the VANDELS team as reference.

We show the results of this exercise in Fig. 7, where it is clear that the stack of C IV emitters not only shows a  $\text{EW}_0(\text{Ly } \alpha)$  value that is  $5\sigma$  higher than the median Ly  $\alpha$  EW measured from randomly stacking VANDELS spectra, its velocity offset compared to systemic is also  $3\sigma$  lower than the median offset. We also find that stronger Ly  $\alpha$  lines tend to peak closer to the systemic redshift, which has also been observed in the literature (see Erb et al. 2014, for example).

Also shown in Fig. 7 are Ly  $\alpha$  measurements from Ly C leakers identified by Fletcher et al. (2019) from a sample of narrow-band selected LAEs, which have comparable Ly  $\alpha$  line strengths and offsets from the systemic velocity to those seen in our stacked spectrum.<sup>6</sup> This suggests that galaxies with strong C IV emission preferentially show stronger Ly  $\alpha$  emission peaking close to the

<sup>6</sup>Here, we use all Ly C leakers from Fletcher et al. (2019) that have  $\text{EW}_0(\text{Ly } \alpha)$  measurements from spectroscopy, with accurate systemic redshifts from other rest-UV lines. These galaxies belong to both *Gold* (4 galaxies) and *Silver* (3 galaxies) sub-samples.



**Figure 7.** Distribution of  $\text{Ly}\alpha$   $\text{EW}_0$  and velocity offset from systemic redshift measured from the stacked spectrum of C IV emitters, compared with measurements from 500 iterations of stacking 19 randomly drawn galaxies from the VANDELS survey occupying the same redshift range as the C IV emitters. The  $\text{Ly}\alpha$  strength and velocity offset of C IV emitters is an outlier compared to the distribution inferred from randomly selected VANDELS galaxies, implying that the  $\text{Ly}\alpha$  emission line from C IV emitting galaxies is indicative of high  $\text{Ly}\alpha$   $f_{\text{esc}}$  (e.g. Verhamme et al. 2015). We additionally show  $\text{Ly}\alpha$  measurements from Ly C leakers at  $z \sim 3$  presented in Fletcher et al. (2019), which also exhibit strong  $\text{Ly}\alpha$  as well as low velocity offset from the systemic redshift.

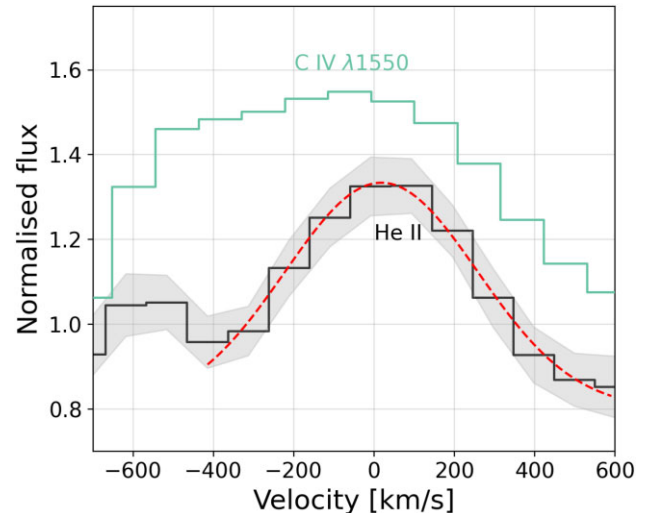
systemic redshift, as is often seen in the spectra of Ly C leaking galaxies across redshifts.

Other known Ly C leaking galaxies at high redshifts in the literature show similar  $\text{Ly}\alpha$  strengths and profiles, with  $\text{Ly}\alpha$  from the *Sunburst Arc* at  $z = 2.4$  (Rivera-Thorsen et al. 2019), *Ion2* at  $z = 3.2$  (Vanzella et al. 2016), and *Ion3* at  $z = 4.0$  (Vanzella et al. 2018) peaking close to systemic  $v = 0$  with non-zero flux bluewards of the peak. High-resolution spectra for such galaxies have revealed multiple peak  $\text{Ly}\alpha$  morphologies, indicative of ionizing photon escape channels in the neutral H I gas (e.g. Vanzella et al. 2018).

The  $\text{Ly}\alpha$  line seen in our stacked spectrum is also comparable to that of the stacked spectrum of the subsets of LAEs at  $z \sim 2$  that are likely leaking significant Ly C radiation presented by Naidu et al. (2022). Those authors also find C IV emission in the stacked spectrum of candidate Ly C leakers, with no C IV emission detected in the stack of LAEs that are unlikely to be leaking Ly C photons.

We note here that 3 out of 19 C IV emitters in our final sample do not show any  $\text{Ly}\alpha$  emission in their spectra, which is reminiscent of no  $\text{Ly}\alpha$  emission being observed in the spectrum of a strong Ly C leaker *Ion1* at  $z = 3.79$  (Ji et al. 2020, and references therein). Finding no  $\text{Ly}\alpha$  emission in a Ly C leaking galaxy is puzzling, but our current understanding of the connection between  $\text{Ly}\alpha$  strength and Ly C leakage, especially at intermediate redshifts, is also based on very limited samples. A likely explanation of the absence of  $\text{Ly}\alpha$  in the presence of other strong UV lines and Ly C leakage is that strong  $\text{Ly}\alpha$  and Ly C leakage may not be coincident (see e.g. Keenan et al. 2017), which is especially true for galaxies with clumpy morphologies (e.g. Rivera-Thorsen et al. 2019), or the possibility of resonant scattering removing  $\text{Ly}\alpha$  from our direct line-of-sight (Ji et al. 2020).

Due to the relatively low spectral resolution of our stacked spectrum we are unable to resolve the  $\text{Ly}\alpha$  line to explore in more



**Figure 8.** Velocity profile of He II emission (black) in the stacked spectrum shown alongside the C IV emission (green), with  $v = 0$  tracing the systemic velocity. The He II line is well fit with a single Gaussian (red dashed line) with the peak of the Gaussian within  $\approx 50 \text{ km s}^{-1}$  of the systemic velocity. The presence of both C IV and He II emission lines requires the production of extremely high-energy photons ( $E > 47.9$  and  $E > 54.4$  eV for C IV and He II, respectively). We find a relatively high He II/C IV ratio of  $\approx 0.5$ , which is indicative of extremely high ionizing photon production efficiencies of  $\log(\xi_{\text{ion}}) \gtrsim 25.6 \text{ erg}^{-1} \text{ Hz}$  (e.g. Berg et al. 2019; Schaerer et al. 2022).

detail indication of significant Ly C  $f_{\text{esc}}$ . However, future higher resolution observations of the  $\text{Ly}\alpha$  line profile of C IV emitting galaxies would be valuable to further investigate the presence of Ly  $\alpha$ /Ly C escape channels.

## 5.2 Presence of strong He II

We also find strong He II emission in addition to C IV emission in the stacked spectrum, with He II/C IV  $\approx 0.5$ , indicative of the presence of sources capable of producing hard ionizing radiation fields (see also Berg et al. 2018, 2019). Schaerer et al. (2022) also showed the presence of He II emission in strong Ly C leaking galaxies at  $z < 0.7$ , with high equivalent widths of 3–8 Å, which can be explained with relatively high ionizing photon production efficiencies,  $\log(\xi_{\text{ion}}) \approx 25.6\text{--}25.8 \text{ erg}^{-1} \text{ Hz}$ . These line strengths are higher than  $\text{EW}_0(\text{He II}) = 2.6$  that we measure from our stacked spectrum, but the relatively high He II/C IV ratio we find is comparable to that seen in the spectra of Ly C leakers presented by Schaerer et al. (2022), indicative of similarly high  $\xi_{\text{ion}}$ .

In Fig. 8, we show the velocity profile of the observed He II emission with respect to the C IV line, finding that the peak of He II emission is within  $\approx 50 \text{ km s}^{-1}$  to that of C IV, tracing the alignment of channels through which photons with much higher energies may also escape in addition to the relatively lower energy channels traced by  $\text{Ly}\alpha$ . With C IV being a resonant line, the peaks of C IV and He II being coincident suggests that the resonant scattering does not dramatically alter the energies of escaping photons, pointing towards the presence of well-defined columns in the multiphase ISM facilitating high-energy ionizing (and possibly Ly C) photon escape.

Further, Naidu et al. (2022) also found He II emission in the stacked spectra of candidate Ly C leaking LAEs at  $z \sim 2$  with  $\text{EW}_0(\text{He II}) \approx 2 \text{ \AA}$ , which is highly comparable to our measurement. However, Marques-Chaves et al. (2022) reported the detection of comparable He II  $\lambda 4686$  emission from Ly C leakers and non-leakers  $z \sim 0.2\text{--}0.4$ ,

noting that metallicity and not Ly C escape is the dominant factor in setting He II line strengths. Their study implies that Ly C leaking galaxies may not show systematically ‘harder’ ionizing spectra compared to non-leakers at similar metallicities, casting doubt on the presence of He II emission as a stand-alone indicator of strong Ly C leakage. None the less, strong He II emission in our stacked spectrum of C IV emitters, when combined with other indicators, may still favour high  $f_{\text{esc}}$ .

### 5.3 Insights from low-ionization interstellar absorption

In this section, we attempt to infer Ly C  $f_{\text{esc}}$  from both the depths and the profiles of absorption features in the stacked spectrum of C IV emitters arising from singly ionized species in the ISM. A low covering fraction ( $f_c < 1$ ) of metal-enriched gas, and consequently of that of H I gas, is expected to be a necessary but not sufficient condition for significant Ly C photon escape, and several studies have investigated this relationship using rest-UV absorption lines (Jones et al. 2013; Henry et al. 2015; Leethochawalit et al. 2016; Reddy et al. 2016b; Chisholm et al. 2018; Steidel et al. 2018; Reddy et al. 2022; Saldana-Lopez et al. 2022).

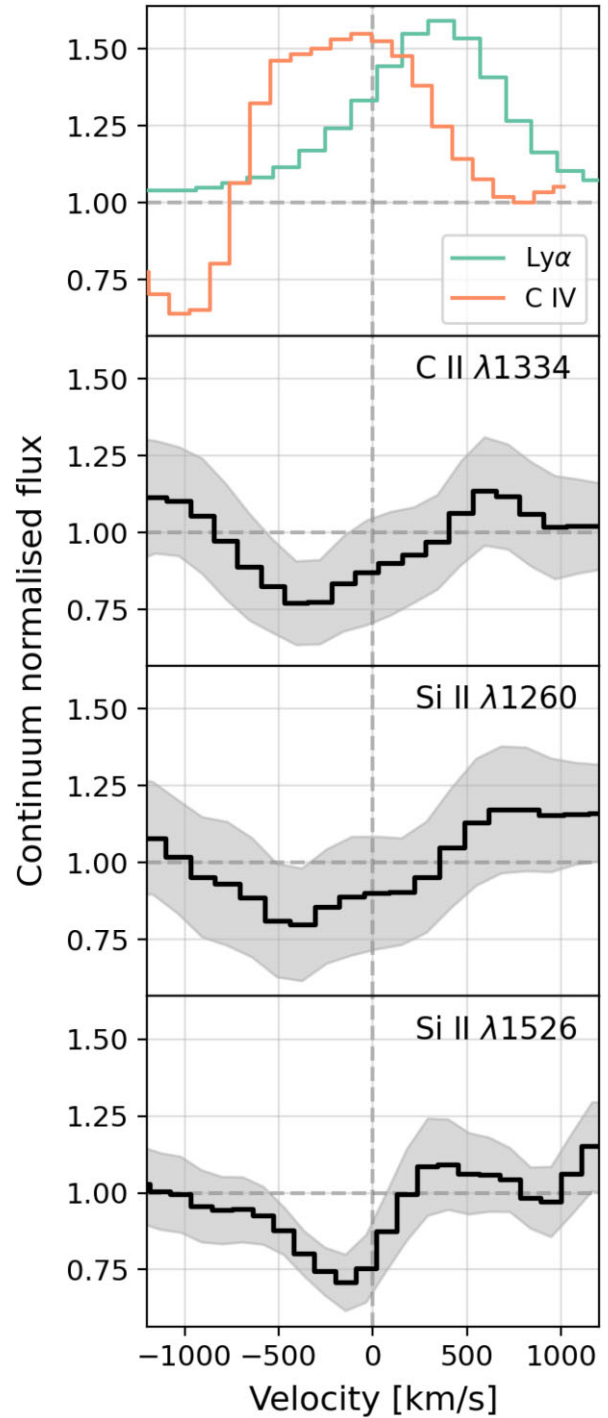
Since the wavelength range of our stacked spectrum does not cover the H I Lyman-series transitions of Ly  $\beta$  and beyond, we focus instead on other well-studied low-ionization interstellar (LIS) absorption features at rest-UV wavelengths, C II  $\lambda 1334$  and Si II  $\lambda 1260$  and  $\lambda 1526$ . We do not use the Si II  $\lambda 1304$  feature as it appears to be contaminated by O I  $\lambda 1302$ .

Following Jones et al. (2013), we calculate the covering fraction of the above-mentioned transitions assuming a ‘picket fence’ like distribution of optically thick gas clouds and optically thin ‘holes’ in the ISM (e.g. Steidel et al. 2018). Within the gas clouds, the column density of gas is assumed to be high enough such that it appears optically thick ( $\tau \gg 1$ ) at the absorbing wavelength, saturating the absorption lines. In this picture, the covering fraction may be calculated as  $f_c = 1 - I(\lambda)/I_0$ , where  $I_0$  is the local continuum level and  $I(\lambda)$  is the residual intensity in the spectrum (see also Saldana-Lopez et al. 2022). We find an average covering fraction of LIS lines to be  $\approx 0.2$ . Following the best-fitting relation obtained by Saldana-Lopez et al. (2022) for the covering fraction of LIS lines and H I, we estimate H I covering fractions of  $\approx 0.7$  resulting in limits of Ly C  $f_{\text{esc}} \lesssim 0.3$ .

We further measure  $\text{EW}_0(\text{LIS})$  following the methodology of Saldana-Lopez et al. (2022) by first locating the minimum depth of the absorption line in question and then integrating over a velocity range of  $\pm 1250 \text{ km s}^{-1}$ , dividing the measured flux by the local stellar continuum. Since we use the convention whereby emission lines have positive equivalent widths, the equivalent width measured for absorption features in this work are negative. We find  $\text{EW}_0(\text{Si II } \lambda 1260) = -0.61 \text{ \AA}$ ,  $\text{EW}_0(\text{Si II } \lambda 1526) = -0.80 \text{ \AA}$  and  $\text{EW}_0(\text{C II } \lambda 1334) = -0.69 \text{ \AA}$ .

Using the average  $\text{EW}_0(\text{LIS}) \approx -0.70 \pm 0.8 \text{ \AA}$  and the relation between  $f_{\text{esc}}$  and  $\text{EW}_0(\text{LIS})$  obtained by Saldana-Lopez et al. (2022), the  $f_{\text{esc}}$  is estimated to be in the range  $\sim 0.05\text{--}0.30$ . Finally, from only  $\text{EW}_0(\text{C II } \lambda 1334)$  using the Mauerhofer et al. (2021) relation we infer  $f_{\text{esc}} > 0.1$ . All of these estimates point towards significant  $f_{\text{esc}}$  from galaxies that show strong C IV emission in their spectra.

These LIS absorption features along with strong Ly  $\alpha$  and C IV lines are shown in Fig. 9. Once again,  $v = 0$  is set to be at the peak of the C IV  $\lambda 1550$  emission. The shaded regions show  $1\sigma$  uncertainties calculated using bootstrapping in Section 3. Here, the flux density of the stacked spectrum is normalized to have a value of 1 at  $1500 \text{ \AA}$ , and the Ly  $\alpha$  emission has been re-scaled to aid visualization.



**Figure 9.** LIS absorption of C II  $\lambda 1334$ , Si II  $\lambda 1334$ , and  $\lambda 1526$  shown along with strong Ly  $\alpha$  and C IV emission in the stacked spectrum. Here, the continuum is set to 1 at  $1500 \text{ \AA}$ . The shaded regions around the LIS lines are  $1\sigma$  uncertainties calculated using bootstrapping. The systemic velocity  $v = 0$  is set using C IV  $\lambda 1550$ . All features have been continuum normalized, with the Ly  $\alpha$  line re-scaled for visualization. The low covering fractions of LIS absorption features are indicative of relatively high  $f_{\text{esc}}$  (Mauerhofer et al. 2021; Saldana-Lopez et al. 2022), which also appear to be blueshifted compared to the systemic velocity indicative of outflowing gas with velocities in the range  $-500 \lesssim v \lesssim -200 \text{ km s}^{-1}$ , consistent with measurements from known Ly C leakers (e.g. Chisholm et al. 2017; Steidel et al. 2018).



We further note that the peaks of the absorption features are blueshifted with respect to the systemic velocity, indicative of the presence of outflowing gas (e.g. Jones et al. 2013). We infer outflow velocities in the range  $-500 \lesssim v \lesssim -200 \text{ km s}^{-1}$ , which are within one resolution element of the VANDELS spectra. The maximum covering fractions for the LIS gas are measured at  $v \approx 150\text{--}400 \text{ km s}^{-1}$ .

It is important to note that the scatter on the relation between  $f_{\text{esc}}$  and  $\text{EW}_0(\text{LIS})$  can be large, and the relation is further sensitive to the spectral resolution as demonstrated by Saldana-Lopez et al. (2022). Those authors showed that these effects may increase the error on the inferred covering fractions by 5–20 per cent, with a compounded effect on the uncertainty on  $f_{\text{esc}}$ . We further note that stacking the spectra of individual galaxies with varying outflow velocities may also muddle the absorption troughs by artificially broadening the absorption features, which has an important implication that  $f_{\text{esc}}$  measured from the covering fraction of LIS lines from stacked spectra will be an upper limit.

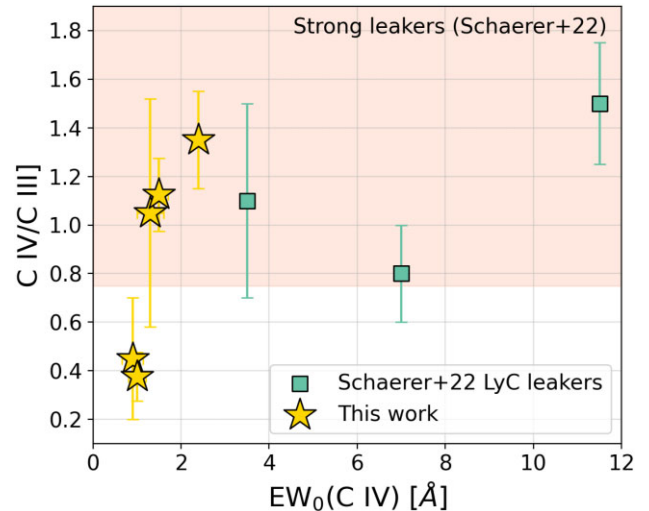
However, the widths of the LIS absorption features in the stacked spectrum appear to be fairly consistent with the spectral resolution of VANDELS, indicating that the individual outflow velocities do not vary dramatically across our sample. The outflow velocities for these lines that we infer are consistent with those measured for confirmed Ly C leakers by Chisholm et al. (2017). We further note the remarkable resemblance of both the absorption profiles as well as the outflow velocities with the high  $f_{\text{esc}}$  sub-samples from Steidel et al. (2018), who also reported an increase in the depth of these absorption features with decreasing Ly C  $f_{\text{esc}}$ .

#### 5.4 C IV/C III] ratios for individual galaxies

In Section 2.2, we noted the presence of C III] emission in 5 galaxies in our sample, and fortunately all 5 of these sources are likely to be star-forming galaxies (and not AGNs). Schaerer et al. (2022) recently reported the detection of both C IV and C III] from a sample of confirmed Ly C leaking galaxies at  $z < 0.7$ , and found strong Ly C leakers ( $f_{\text{esc}} > 0.1$ ) to have C IV/C III] ratios in excess of 0.75. Therefore, in this section we explore whether strong Ly C leakage can be inferred, at least qualitatively, from galaxies in our sample with both C IV and C III] line detections.

Before comparing with measurements from low- $z$  leakers, we note that in the analysis of Schaerer et al. (2022) the C IV emission was likely purely nebular in origin, owing to a lack of P-Cygni absorption feature that is indicative of a stellar origin because of stellar photospheric absorption. We do note the presence of some absorption blueward of the C IV line both in individual sources and the stacked spectrum, and we now attempt to capture the fraction of C IV flux that could be attributed to stellar emission. To closely replicate the estimation of the stellar and nebular components in the C IV line in the Schaerer et al. (2022) analysis, we fit the individual galaxy spectra using SEDs that only contain stellar emission using the methodology of Saldana-Lopez et al. (2022).

Briefly, these stellar-only SEDs use STARBURST99 single star models (Leitherer et al. 2011) that include stellar rotation across a range of ages and metallicities. The individual spectra are converted to rest frame and the model SEDs are convolved with a Gaussian kernel to match the VANDELS spectral resolution. Additionally, employing a uniform foreground dust attenuation model, the observed spectra are fitted with a linear combination of STARBURST99 models (that only include light from the stellar continuum). We refer the readers to Saldana-Lopez et al. (2022) for more details about the fitting procedure. From this exercise, we find that on average  $\approx 25$  per cent of the C IV flux may be attributed to stellar origin across our sample.



**Figure 10.** Ratio of nebular C IV and C III] versus  $\text{EW}_0(\text{C IV})$  for individual galaxies in this study where both lines are reliably detected. Also shown for comparison are measurements for confirmed Ly C leaking galaxies at  $z < 0.7$  from Schaerer et al. (2022). The shaded area represents the parameter space where Schaerer et al. (2022) found Ly C leaking galaxies in their sample to lie. We find that 3 out of 5 galaxies in our sample that have both C IV and C III] detections show comparable C IV/C III] ratios to that of Ly C leakers from Schaerer et al. (2022). We note that the  $\text{EW}_0(\text{C IV})$  for our galaxies are lower than that of low redshift leakers.

When plotting the C IV/C III] ratio in Fig. 10, we have removed the stellar contribution to C IV and only show the ratio of the nebular component of these lines as a function of  $\text{EW}_0(\text{C IV})$ . We also show measurements from strong Ly C leakers from Schaerer et al. (2022). Interestingly, based on this simple diagnostic we infer that 3 out of 5 galaxies in our sample that show both these lines lie in the regime of strong Ly C leakage. We do note that the strength of C IV emission from galaxies in our sample is systematically lower than what Schaerer et al. (2022) find for their Ly C leakers.

Unfortunately, the C III] line does not fall within the observed wavelength range of all galaxies in our final sample, and lying in a relatively redder part of the wavelength range of the spectrograph, the C III] line is often contaminated by skyline residuals. Therefore, we choose not to consider C III] emission measures drawn from our stacked spectrum. However, from a very simple C IV/C III] ratio based diagnostic it is clear that strong Ly C leakage may be expected from a fraction of C IV emitting galaxies in our sample, especially when considered in combination with other spectroscopic indicators in the stacked spectrum.

#### 5.5 Concluding remarks on Ly C escape

We conclude this section by reiterating the numerous lines of evidence supporting the contention that galaxies selected by their strong C IV emission have ISM conditions favourable for the escape of ionizing Ly C photons. These include strong Ly  $\alpha$  emission line peaking close to the systemic redshift with non-zero flux bluewards of the peak, the presence of high-ionization lines such as He II and low covering fractions of LIS absorption lines and their blueshifted absorption troughs. We also find elevated C IV/C III] ratios observed in individual C IV emitters where both lines are robustly detected, which is also expected under conditions of significant Ly C leakage.

The remarkable presence of all of these features together strongly indicates, albeit indirectly, that C IV emitting star-forming galaxies

may be good candidates of strongly Ly C leaking galaxies. With NIR-Spec on-board *JWST* it will be possible to detect C IV, He II and C III] emission from galaxies at  $z > 6$ , while probing absorption features from metals as well as neutral hydrogen gas in galaxies at even higher redshifts, possibly providing reliable observables to infer Ly C  $f_{\text{esc}}$  in the epoch of reionization. We have shown that galaxies exhibiting strong C IV have both high  $\xi_{\text{ion}}$  as well as potentially high  $f_{\text{esc}}$ , the product of which is a key ingredient to understand the role of star-forming galaxies towards the cosmic reionization budget at  $z > 6$ .

## 6 SUMMARY

In this work, we have identified 19 C IV emitting star-forming galaxies from the VANDELS survey spanning a redshift range  $z = 3.1\text{--}4.6$  and presented their stacked spectrum (Sections 2 and 3). Some individual C IV emitters show other rest-frame UV lines such as He II and O III] along with Ly  $\alpha$ , and all of these lines are securely detected in the stacked spectrum allowing for a detailed analysis of the average properties of the underlying stellar populations as well as the interstellar medium in C IV emitting galaxies.

We show that the inferred rest-frame UV line fluxes and ratios from the stacked spectrum of C IV emitting galaxies at  $z \sim 3.6$  suggest that they are comparable in ionization properties to local C IV (and He II) emitting metal-poor galaxies, that have long been touted as analogues of reionization era galaxies. We also find that the line strengths in the stack are similar to a handful of known C IV emitting sources at  $z > 6$  (Section 3.3).

For the stacked spectrum of C IV emitters, we find that the best-fitting spectral energy distribution (SED) models incorporating both stellar continuum and nebular line emission have low stellar metallicities of  $Z = 0.1\text{--}0.2 Z_{\odot}$ , a young stellar ages of  $\log(\text{age yr}^{-1}) = 6.1\text{--}6.5$ , a high ionization parameter  $\log(U) = -2.0$  and little to no dust ( $E(B - V) = 0.00\text{--}0.01$ ). This suggests that the presence of young, metal-poor stellar populations is necessary to explain the strong C IV (and other rest-UV) line emission seen across our sample (Section 4.1).

We also measure the average stellar metallicity of C IV emitters from the stacked spectrum using absorption indices at 1501 and 1719 Å that are sensitive to the metal content of stars. Both indices give a stellar metallicity of  $Z \approx 0.2 Z_{\odot}$  within errors, which is consistent with the metallicity of SED model with  $\log(\text{age yr}^{-1}) = 6.1$  and a small dust attenuation of  $E(B - V) = 0.01$  (Section 4.2).

However, we find that the SED with  $Z \approx 0.2 Z_{\odot}$  underpredicts the C IV equivalent width by a factor of 3, and the He II equivalent width by a factor of 10. Since extremely young/metal-poor stars are needed to increase the predicted C IV and He II line fluxes from models, the observations of these strong lines and relatively higher metallicities suggest that a relatively young starburst event within a galaxy containing older populations may be able to explain the observed spectral features of our stacked spectrum (Section 4.3).

We then investigate whether galaxies showing strong C IV emission may exhibit significant hydrogen ionizing Ly C photon leakage into the intergalactic medium using a variety of indicators (Section 5). First, we find that the strength and shape of the Ly  $\alpha$  emission in the stack is indicative of significant Ly C leakage, as the Ly  $\alpha$  line peaks close to the systemic velocity and contains non-zero flux blueward of the peak (Section 5.1). The presence of strong He II emission is indicative of substantial production of high-energy photons from young stars, and is consistent with what has been observed in the spectra of known Ly C leaking galaxies (Section 5.2). The low equivalent widths and outflow velocities of low-ionization interstellar absorption features are indicative of low column density channels through which Ly C photons may escape, with their low covering

fractions suggesting Ly C  $f_{\text{esc}} \approx 0.05\text{--}0.30$  (Section 5.3). Finally, the C IV/C III] ratio of a fraction of C IV emitting galaxies is comparable to measurements from other known Ly C leakers in the low-redshift Universe (Section 5.4).

We therefore conclude that C IV emitting galaxies harbour young stellar populations, tracing recent starburst events that leads to the production of copious amounts of ionizing photons. This starburst phase is needed to displace neutral as well as low-ionization gas in the ISM of galaxies, potentially creating holes in the ISM through which Ly C photons may be able to escape into the IGM. We find indirect evidence of significant Ly C leakage from C IV emitting galaxies, suggesting that such galaxies could be important contributors towards cosmic reionization at  $z > 6$ . Conditions leading to strong C IV emission may be ubiquitous across galaxies at very high redshifts when the stellar populations were young and star formation from metal-deficient gas was widespread.

At  $z \gtrsim 6$  when the increased neutrality of the IGM attenuates Ly  $\alpha$  photons along the line-of-sight, the C IV line may offer a reliable alternative to identify galaxies with significant  $f_{\text{esc}}$ , which will be possible for statistical samples using *JWST*/NIRSpec. However, careful radiative transfer modelling of C IV and Ly C is needed to theoretically back up any dependence of C IV and  $f_{\text{esc}}$ . Therefore, the presence of strong C IV emission combined with other rest-UV indicators in the spectra of galaxies at  $z > 6$  can help establish whether they are likely contributors towards cosmic reionization at early epochs.

## ACKNOWLEDGEMENTS

We thank the referee for constructive feedback that helped improve the quality of this work. AS thanks Silvia Genovese for assistance with plots and figures. AS and RSE acknowledge financial support from European Research Council Advanced Grant FP7/669253. ASL and DS acknowledge support from Swiss National Science Foundation.

This work makes use of VANDELS data. The VANDELS collaboration acknowledges the invaluable role played by ESO staff for successfully carrying out the survey. This work has made extensive use of Jupyter and IPython (Pérez & Granger 2007) notebooks, ASTROPY (Astropy Collaboration 2018), HOKI (Stevance, Eldridge & Stanway 2020), MATPLOTLIB (Hunter 2007), and SEABORN (Waskom & the seaborn development team 2020). This work would not have been possible without the countless hours put in by members of the open-source developing community all around the world.

## DATA AVAILABILITY

The VANDELS Data Release 4 (DR4) is now publicly available and can be accessed using the VANDELS data base at <http://vandels.inaf.it/dr4.html>, or through the ESO archives. The data analysis code was written in PYTHON and may be shared upon reasonable request to the corresponding author.

## REFERENCES

- Astropy Collaboration, 2018, *AJ*, 156, 123
- Barrow K. S. S., Robertson B. E., Ellis R. S., Nakajima K., Saxena A., Stark D. P., Tang M., 2020, *ApJ*, 902, L39
- Berg D. A., Skillman E. D., Henry R. B. C., Erb D. K., Carigi L., 2016, *ApJ*, 827, 126
- Berg D. A., Erb D. K., Auger M. W., Pettini M., Brammer G. B., 2018, *ApJ*, 859, 164

- Berg D. A., Chisholm J., Erb D. K., Pogge R., Henry A., Olivier G. M., 2019, *ApJ*, 878, L3
- Berg D. A. et al., 2022, *ApJS*, 261, 31
- Bouwens R. J., Illingworth G. D., Oesch P. A., Caruana J., Holwerda B., Smit R., Wilkins S., 2015, *ApJ*, 811, 140
- Calabrò A. et al., 2021, *A&A*, 646, A39
- Calabrò A. et al., 2022, preprint([arXiv:2206.14918](https://arxiv.org/abs/2206.14918))
- Chisholm J., Orlitová I., Schaerer D., Verhamme A., Worseck G., Izotov Y. I., Thuan T. X., Guseva N. G., 2017, *A&A*, 605, A67
- Chisholm J. et al., 2018, *A&A*, 616, A30
- Cullen F. et al., 2019, *MNRAS*, 487, 2038
- Cullen F. et al., 2020, *MNRAS*, 495, 1501
- Dayal P., Ferrara A., 2018, *Phys. Rep.*, 780, 1
- Dijkstra M., 2014, *PASA*, 31, e040
- Dijkstra M., Gronke M., Venkatesan A., 2016, *ApJ*, 828, 71
- Erb D. K. et al., 2014, *ApJ*, 795, 33
- Feltre A., Charlot S., Gutkin J., 2016, *MNRAS*, 456, 3354
- Feltre A. et al., 2020, *A&A*, 641, A118
- Ferland G. J. et al., 2013, *Rev. Mex. Astron. Astrofis.*, 49, 137
- Finkelstein S. L. et al., 2015, *ApJ*, 810, 71
- Fletcher T. J., Tang M., Robertson B. E., Nakajima K., Ellis R. S., Stark D. P., Inoue A., 2019, *ApJ*, 878, 87
- Flury S. R. et al., 2022a, *ApJS*, 260, 1
- Flury S. R. et al., 2022b, *ApJ*, 930, 126
- Garilli B. et al., 2021, *A&A*, 647, A150
- Gazagnes S., Chisholm J., Schaerer D., Verhamme A., Izotov Y., 2020, *A&A*, 639, A85
- Götberg Y., de Mink S. E., Groh J. H., Leitherer C., Norman C., 2019, *A&A*, 629, A134
- Grazian A. et al., 2016, *A&A*, 585, A48
- Guaita L. et al., 2016, *A&A*, 587, A133
- Gutkin J., Charlot S., Bruzual G., 2016, *MNRAS*, 462, 1757
- Henry A., Scarlata C., Martin C. L., Erb D., 2015, *ApJ*, 809, 19
- Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90
- Inoue A. K., Shimizu I., Iwata I., Tanaka M., 2014, *MNRAS*, 442, 1805
- Izotov Y. I., Schaerer D., Worseck G., Guseva N. G., Thuan T. X., Verhamme A., Orlitová I., Fricke K. J., 2018, *MNRAS*, 474, 4514
- Izotov Y. I., Worseck G., Schaerer D., Guseva N. G., Chisholm J., Thuan T. X., Fricke K. J., Verhamme A., 2021, *MNRAS*, 503, 1734
- Ji Z. et al., 2020, *ApJ*, 888, 109
- Jones T., Stark D. P., Ellis R. S., 2012, *ApJ*, 751, 51
- Jones T. A., Ellis R. S., Schenker M. A., Stark D. P., 2013, *ApJ*, 779, 52
- Katz H. et al., 2020, *MNRAS*, 498, 164
- Keenan R. P., Oey M. S., Jaskot A. E., James B. L., 2017, *ApJ*, 848, 12
- Kimm T., Bieri R., Geen S., Rosdahl J., Blaizot J., Michel-Dansac L., Garel T., 2022, *ApJS*, 259, 21
- Kocevski D. D. et al., 2018, *ApJS*, 236, 48
- Laporte N., Meyer R. A., Ellis R. S., Robertson B. E., Chisholm J., Roberts-Borsani G. W., 2021, *MNRAS*, 505, 3336
- Le Fèvre O. et al., 2019, *A&A*, 625, A51
- Leethochawalit N., Jones T. A., Ellis R. S., Stark D. P., Zitrin A., 2016, *ApJ*, 831, 152
- Leitherer C. et al., 2011, *Astrophysics Source Code Library*, record ascl:1104.003
- Llerena M. et al., 2021, *A&A*, 659, A16
- Luo B. et al., 2017, *ApJS*, 228, 2
- Mainali R., Kollmeier J. A., Stark D. P., Simcoe R. A., Walth G., Newman A. B., Miller D. R., 2017, *ApJ*, 836, L14
- Marchi F. et al., 2017, *A&A*, 601, A73
- Marques-Chaves R. et al., 2022, *A&A*, 663, L1
- Matthee J. et al., 2022, *A&A*, 660, A10
- Mauerhofer V., Verhamme A., Blaizot J., Garel T., Kimm T., Michel-Dansac L., Rosdahl J., 2021, *A&A*, 646, A80
- McLure R. J. et al., 2018, *MNRAS*, 479, 25
- Meštrić U., Ryan-Weber E. V., Cooke J., Bassett R., Prichard L. J., Rafelski M., 2021, *MNRAS*, 508, 4443
- Naidu R. P., Forrest B., Oesch P. A., Tran K.-V. H., Holden B. P., 2018, *MNRAS*, 478, 791
- Naidu R. P. et al., 2022, *MNRAS*, 510, 4582
- Nakajima K., Ouchi M., 2014, *MNRAS*, 442, 900
- Nakajima K., Fletcher T., Ellis R. S., Robertson B. E., Iwata I., 2018b, *MNRAS*, 477, 2098
- Nakajima K. et al. 2018a, *A&A*, 612, A94
- Nakajima K., Ellis R. S., Robertson B. E., Tang M., Stark D. P., 2020, *ApJ*, 889, 161
- Nanayakkara T. et al., 2019, *A&A*, 624, A89
- Pentericci L. et al., 2018, *A&A*, 616, A174
- Pérez F., Granger B. E., 2007, *Comput. Sci. Eng.*, 9, 21
- Planck Collaboration XIII, 2016, *A&A*, 594, A13
- Plat A., Charlot S., Bruzual G., Feltre A., Vidal-García A., Morisset C., Chevallard J., Todt H., 2019, *MNRAS*, 490, 978
- Reddy N. A., Steidel C. C., Pettini M., Bogosavljević M., 2016a, *ApJ*, 828, 107
- Reddy N. A., Steidel C. C., Pettini M., Bogosavljević M., Shapley A. E., 2016b, *ApJ*, 828, 108
- Reddy N. A. et al., 2022, *ApJ*, 926, 31
- Rivera-Thorsen T. E. et al., 2019, *Science*, 366, 738
- Roberts-Borsani G. W., Ellis R. S., Laporte N., 2020, *MNRAS*, 497, 3440
- Robertson B. E. et al., 2013, *ApJ*, 768, 71
- Robertson B. E., Ellis R. S., Furlanetto S. R., Dunlop J. S., 2015, *ApJ*, 802, L19
- Saldana-Lopez A. et al., 2022, *A&A*, 663, A59
- Saxena A. et al., 2020a, *MNRAS*, 496, 3796
- Saxena A. et al., 2020b, *A&A*, 636, A47
- Saxena A. et al., 2022, *MNRAS*, 511, 120
- Schaerer D., Fragos T., Izotov Y. I., 2019, *A&A*, 622, L10
- Schaerer D. et al., 2022, *A&A*, 658, L11
- Schmidt K. B. et al., 2017, *ApJ*, 839, 17
- Senchyna P. et al., 2017, *MNRAS*, 472, 2608
- Senchyna P., Stark D. P., Chevallard J., Charlot S., Jones T., Vidal-García A., 2019, *MNRAS*, 488, 3492
- Senchyna P., Stark D. P., Charlot S., Chevallard J., Bruzual G., Vidal-García A., 2021, *MNRAS*, 503, 6112
- Shapley A. E., Steidel C. C., Pettini M., Adelberger K. L., Erb D. K., 2006, *ApJ*, 651, 688
- Shivaei I. et al., 2018, *ApJ*, 855, 42
- Simmonds C., Schaerer D., Verhamme A., 2021, *A&A*, 656, A127
- Stanway E. R., Eldridge J. J., 2018, *MNRAS*, 479, 75
- Stanway E. R., Eldridge J. J., Becker G. D., 2016, *MNRAS*, 456, 485
- Stark D. P. et al., 2015, *MNRAS*, 454, 1393
- Steidel C. C., Strom A. L., Pettini M., Rudie G. C., Reddy N. A., Trainor R. F., 2016, *ApJ*, 826, 159
- 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
- Stevance H., Eldridge J., Stanway E., 2020, *J. Open Source Softw.*, 5, 1987
- Tang M., Stark D. P., Chevallard J., Charlot S., Endsley R., Congiu E., 2021, *MNRAS*, 501, 3238
- Tang M., Stark D. P., Ellis R. S., 2022, *MNRAS*, 513, 5211
- Umeda H., Ouchi M., Nakajima K., Isobe Y., Aoyama S., Harikane Y., Ono Y., Matsumoto A., 2022, *ApJ*, 930, 37
- Vanzella E. et al., 2016, *ApJ*, 821, L27
- Vanzella E. et al., 2018, *MNRAS*, 476, L15
- Vanzella E. et al., 2021, *A&A*, 646, A57
- 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
- Waskom M., the seaborn development team, 2020, mwaskom/seaborn
- Witstok J., Smit R., Maiolino R., Curti M., Laporte N., Massey R., Richard J., Swinbank M., 2021, *MNRAS*, 508, 1686
- Xiao L., Stanway E. R., Eldridge J. J., 2018, *MNRAS*, 477, 904
- Zackrisson E., Inoue A. K., Jensen H., 2013, *ApJ*, 777, 39