



## LJMU Research Online

Heintz, KE, Brammer, GB, Watson, D, Oesch, PA, Keating, LC, Hayes, MJ, Abdurro'uf, KZ, Arellano-Cordova, KZ, Carnall, AC, Christiansen, CR, Cullen, F, Dave, R, Dayal, P, Ferrara, A, Finlator, K, Fynbo, JPU, Flury, SR, Gelli, V, Gillman, S, Gottumukkala, R, Gould, K, Greve, TR, Hardin, SE, Hsiao, TY-Y, Hutter, A, Jakobsson, P, Killi, M, Khosravaninezhad, N, Laursen, P, Lee, MM, Magdis, GE, Matthee, J, Naidu, RP, Narayanan, D, Pollock, C, Prescott, MKM, Rusakov, V, Shuntov, M, Sneppen, A, Smit, R, Tanvir, NR, Terp, C, Toft, S, Valentino, F, Vijayan, AP, Weaver, JR, Wise, JH and Witstok, J

The JWST-PRIMAL archival survey A JWST/NIRSpec reference sample for the physical properties and Lyman- $\alpha$  absorption and emission of  $\sim 600$  galaxies at  $z = 5.0 - 13.4$

<http://researchonline.ljmu.ac.uk/id/eprint/25318/>

### Article

**Citation** (please note it is advisable to refer to the publisher's version if you intend to cite from this work)

**Heintz, KE, Brammer, GB, Watson, D, Oesch, PA, Keating, LC, Hayes, MJ, Abdurro'uf, KZ, Arellano-Cordova, KZ, Carnall, AC, Christiansen, CR, Cullen, F, Dave, R, Dayal, P, Ferrara, A, Finlator, K, Fynbo, JPU, Flury, SR, Gelli, V, Gillman, S, Gottumukkala, R, Gould, K, Greve, TR, Hardin, SE, Hsiao, TY-Y.**

LJMU has developed **LJMU Research Online** for users to access the research output of the University more effectively. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LJMU Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain.

<http://researchonline.ljmu.ac.uk/>

The version presented here may differ from the published version or from the version of the record. Please see the repository URL above for details on accessing the published version and note that access may require a subscription.

For more information please contact [researchonline@ljmu.ac.uk](mailto:researchonline@ljmu.ac.uk)

# The JWST-PRIMAL archival survey

## A JWST/NIRSpec reference sample for the physical properties and Lyman- $\alpha$ absorption and emission of $\sim 600$ galaxies at $z = 5.0\text{--}13.4$

K. E. Heintz<sup>1,2,3,\*</sup>, G. B. Brammer<sup>1,2</sup>, D. Watson<sup>1,2</sup>, P. A. Oesch<sup>3,1,2</sup>, L. C. Keating<sup>4</sup>, M. J. Hayes<sup>5</sup>, Abdurro'uf<sup>6,7</sup>, K. Z. Arellano-Córdova<sup>4</sup>, A. C. Carnall<sup>4</sup>, C. R. Christiansen<sup>1,2</sup>, F. Cullen<sup>4</sup>, R. Davé<sup>4</sup>, P. Dayal<sup>8</sup>, A. Ferrara<sup>9</sup>, K. Finlator<sup>10,1,2</sup>, J. P. U. Fynbo<sup>1,2</sup>, S. R. Flury<sup>11</sup>, V. Gelli<sup>1,2</sup>, S. Gillman<sup>1,12</sup>, R. Gottumukkala<sup>1,2</sup>, K. Gould<sup>1,2</sup>, T. R. Greve<sup>1,12</sup>, S. E. Hardin<sup>13</sup>, T. Y.-Y. Hsiao<sup>6,7</sup>, A. Hutter<sup>1,2</sup>, P. Jakobsson<sup>14</sup>, M. Killi<sup>15</sup>, N. Khosravanezhad<sup>16,1,2</sup>, P. Laursen<sup>1,2</sup>, M. M. Lee<sup>1,12</sup>, G. E. Magdis<sup>1,12,2</sup>, J. Matthee<sup>17</sup>, R. P. Naidu<sup>18,\*\*</sup>, D. Narayanan<sup>19,1,2</sup>, C. Pollock<sup>4</sup>, M. K. M. Prescott<sup>10</sup>, V. Rusakov<sup>1,2</sup>, M. Shuntov<sup>1,2</sup>, A. Sneppen<sup>1,2</sup>, R. Smit<sup>20</sup>, N. R. Tanvir<sup>21</sup>, C. Terp<sup>1,2</sup>, S. Toft<sup>1,2</sup>, F. Valentino<sup>22,1</sup>, A. P. Vijayan<sup>1,12</sup>, J. R. Weaver<sup>11</sup>, J. H. Wise<sup>13</sup>, and J. Witstok<sup>23,24</sup>

(Affiliations can be found after the references)

Received 4 April 2024 / Accepted 11 September 2024

### ABSTRACT

**Context.** One of the surprising early findings with JWST has been the discovery of a strong “roll-over” or a softening of the absorption edge of Ly $\alpha$  in a large number of galaxies at  $z \geq 6$ , in addition to systematic offsets from photometric redshift estimates and fundamental galaxy scaling relations. This has been interpreted as strong cumulative damped Ly $\alpha$  absorption (DLA) wings from high column densities of neutral atomic hydrogen (HI), signifying major gas accretion events in the formation of these galaxies.

**Aims.** To explore this new phenomenon systematically, we assembled the JWST/NIRSpec PRImordial gas Mass AssembLy (PRIMAL) legacy survey of 584 galaxies at  $z = 5.0\text{--}13.4$ , designed to study the physical properties and gas in and around galaxies during the reionization epoch.

**Methods.** We characterized this benchmark sample in full and spectroscopically derived the galaxy redshifts, metallicities, star formation rates, and ultraviolet (UV) slopes. We defined a new diagnostic, the Ly $\alpha$  damping parameter  $D_{\text{Ly}\alpha}$ , to measure and quantify the net effect of Ly $\alpha$  emission strength, the HI fraction in the intergalactic medium, or the local HI column density for each source. The JWST-PRIMAL survey is based on the spectroscopic DAWN JWST Archive (DJA-Spec). We describe DJA-Spec in this paper, detailing the reduction methods, the post-processing steps, and basic analysis tools. All the software, reduced spectra, and spectroscopically derived quantities and catalogs are made publicly available in dedicated repositories.

**Results.** We find that the fraction of galaxies showing strong integrated DLAs with  $N_{\text{HI}} > 10^{21} \text{ cm}^{-2}$  only increases slightly from  $\approx 60\%$  at  $z \approx 6$  up to  $\approx 65\text{--}90\%$  at  $z > 8$ . Similarly, the prevalence and prominence of Ly $\alpha$  emission is found to increase with decreasing redshift, in qualitative agreement with previous observational results. Strong Ly $\alpha$  emitters (LAEs) are predominantly found to be associated with low-metallicity and UV faint galaxies. By contrast, strong DLAs are observed in galaxies with a variety of intrinsic physical properties, but predominantly at high redshifts and low metallicities.

**Conclusions.** Our results indicate that strong DLAs likely reflect a particular early assembly phase of reionization-era galaxies, at which point they are largely dominated by pristine HI gas accretion. At  $z = 8\text{--}10$ , this gas gradually cools and forms into stars that ionize their local surroundings, forming large ionized bubbles and producing strong observed Ly $\alpha$  emission at  $z < 8$ .

**Key words.** galaxies: evolution – galaxies: formation – galaxies: general – galaxies: high-redshift – galaxies: ISM – dark ages, reionization, first stars

## 1. Introduction

The first epoch of galaxy formation is believed to have occurred at  $z \sim 15\text{--}20$  (Hashimoto et al. 2018; Robertson 2022), some 100–200 million years after the Big Bang. This process was primarily driven by the accretion and gravitational collapse of primordial, neutral gas onto dark matter halos (White & Rees 1978; Blumenthal et al. 1984; White & Frenk 1991; Kereš et al. 2005; Schaye et al. 2010; Dayal & Ferrara 2018), which eventually led to the formation of stars in galaxies. The strong ultraviolet (UV) radiation originating from the first stars and supermassive black holes gradually ionized their immediate and then large-scale surroundings, initi-

ating the reionization epoch, which is currently estimated to have ended by  $z \approx 5\text{--}6$  (e.g., Stark 2016; Dayal & Ferrara 2018; Keating et al. 2020; Robertson 2022; Bosman et al. 2022; Fan et al. 2023). When these first stars ended their lives as supernovae, the pristine gas was chemically enriched with elements that formed through explosive nucleosynthesis (Hoyle 1954; Cameron 1957; Woosley & Weaver 1995). Charting these independent components, their interplay through the “baryon cycle” (Tumlinson et al. 2017; Péroux & Howk 2020), and their evolution with cosmic time is at the heart of contemporary galaxy formation and evolution studies.

The launch of the *James Webb* Space Telescope (JWST) (Gardner et al. 2006, 2023) has now enabled us to peer deep into this early cosmic epoch, probing the rest-frame optical and UV emission of galaxies, potentially all the way back to

\* Corresponding author; keheintz@nbi.ku.dk

\*\* NASA Hubble Fellow.

redshifts  $z \gtrsim 15$  (Castellano et al. 2022; Naidu et al. 2022; Harikane et al. 2023; Atek et al. 2023; Donnan et al. 2023; Bouwens et al. 2023; Wang et al. 2023b; Austin et al. 2023). This leap in near-infrared (NIR) capability now enables detailed characterization of the physical properties and baryonic components of galaxies, reaching the first epoch of galaxy formation, the cosmic dawn. This epoch seems to contain a larger population of more luminous ( $M_{UV} < -20$ ) and more massive galaxies than expected from theoretical models (Labbé et al. 2023; Finkelstein et al. 2023; Franco et al. 2023; Harikane et al. 2023; Adams et al. 2024; Casey et al. 2024; Bouwens et al. 2023; Chemerynska et al. 2024; McLeod et al. 2024). Various possibilities have been considered to resolve this discrepancy, including a varying initial mass function (IMF; Haslbauer et al. 2022; Trinca et al. 2023, though see Rasmussen Cueto et al. 2024), radiation pressure pushing dust away from star-forming regions (Ferrara et al. 2023), bursty star formation histories (Sun et al. 2023), or an unexpected overabundance of active galactic nuclei (AGN; Inayoshi et al. 2022; Pacucci et al. 2022; Dayal et al. 2024). The apparent discrepancy may be exaggerated by selection biases toward the youngest, most highly star-forming galaxies with more bursty star formation histories at  $z > 10$  (Mason et al. 2023; Mirocha & Furlanetto 2023; Shen et al. 2023), making the galaxies appear temporarily brighter in the UV. As neutral gas accretion is the primary driver and fuel for star formation, a robust measure of this component is key to resolving the debate.

Recent JWST spectroscopic studies have pushed the limit of the most distant galaxies confirmed to  $z \approx 11\text{--}13$  (Curtis-Lake et al. 2023; Wang et al. 2023a; Fujimoto et al. 2023b; Arrabal Haro et al. 2023; Harikane et al. 2024; Zavala et al. 2024; Castellano et al. 2024). The metal content of galaxies during the reionization epoch can now also be readily measured based on direct  $T_e$ -based methods for sufficiently bright galaxies (e.g., Schaerer et al. 2022; Arellano-Córdova et al. 2022; Taylor et al. 2022; Brinchmann 2023; Curti et al. 2023; Katz et al. 2023; Rhoads et al. 2023; Trump et al. 2023; Heintz et al. 2023a; Nakajima et al. 2023; Sanders et al. 2024) or inferred through rest-frame optical strong-line calibrations (Heintz et al. 2023b; Langeroodi et al. 2023; Curti et al. 2024; Matthee et al. 2023). These observations show a substantial offset from the fundamental-metallicity relation (FMR, Mannucci et al. 2010; Lara-López et al. 2010b; Curti et al. 2020; Sanders et al. 2021) established at  $z \approx 0\text{--}3$  by up to 0.5 dex at  $z > 7$  (Heintz et al. 2023b), suggesting copious pristine HI gas inflows onto galaxies at this epoch before chemical enrichment from star formation can catch up to equilibrium (e.g., Torrey et al. 2018). The exact redshift for this transition is still debated (Curti et al. 2024; Nakajima et al. 2023).

Constraining the inflow of neutral, pristine HI gas is key to understanding the assembly of the primordial matter and the buildup of stars and metals in the first galaxies. However, the circumstantial evidence from the FMR for excessive HI gas accretion needs to be corroborated by directly measuring the neutral, atomic hydrogen (HI) – the key missing element in the baryonic matter budget of high-redshift galaxies. Due to the weakness of the hyperfine HI 21 cm line, HI has historically been measured at redshifts  $z \gtrsim 2$  through Lyman- $\alpha$  ( $Ly\alpha$ ) absorption in bright background sources such as quasars (Wolfe et al. 1986, 2005; Prochaska & Wolfe 2009; Noterdaeme et al. 2012; Péroux & Howk 2020) or gamma-ray bursts (Jakobsson et al. 2006; Totani et al. 2006; Prochaska et al. 2007; Fynbo et al. 2009; Tanvir et al. 2019; Heintz et al. 2023c). However, these only probe HI in narrow, pencil-beam sight-

lines and often at large impact parameters.  $Ly\alpha$  absorption has also been observed directly in integrated spectra of Lyman-break galaxies at  $z \approx 3$  (Pettini et al. 2000; Shapley et al. 2003; Steidel et al. 2010; Cooke & O’Meara 2015; Lin et al. 2023), but only in rare, extreme cases.

Recently, Heintz et al. (2024) reported the discovery of extremely strong ( $N_{HI} \gtrsim 10^{22} \text{ cm}^{-2}$ ) damped  $Ly\alpha$  absorption systems (DLAs) in star-forming galaxies at  $z > 8$  (see also Umeda et al. 2023; D’Eugenio et al. 2024; Chen et al. 2023; Hainline et al. 2024a). These seem to represent a galaxy population showing more prominent and substantially more prevalent DLAs than observed in Lyman-break galaxies at  $z \approx 3$  (e.g. Shapley et al. 2003). The shape of the  $Ly\alpha$  damping wings of galaxies had until these discoveries mainly been thought to trace the bulk HI of the intergalactic medium (IGM) at  $z \gtrsim 6$  (Miralda-Escudé 1998; McQuinn et al. 2008), and had been used to infer the reionization history of the Universe with JWST spectroscopy (Chen 2023; Keating et al. 2023a,b; Umeda et al. 2023). The discovery of these massive interstellar or circumgalactic HI gas reservoirs carry important implications for our study of the early Universe: they constitute the first direct evidence for the mass accretion of primordial intergalactic gas into ordinary galaxies, that is, these DLAs represent the creation of the baryonic component of galaxies. They also strongly affect observations of early galaxies because they change the shape of the  $Ly\alpha$  damping wings, altering and complicating inferences on the state of the bulk IGM. Further, they hinder the escape of ionizing photons (Laursen et al. 2011; Verhamme et al. 2015; Steidel et al. 2018; Hu et al. 2023; Hayes & Scarlata 2023), and may also be responsible for the bias in photometric redshifts compared to spectroscopic measurements (Heintz et al. 2024; Finkelstein et al. 2024; Fujimoto et al. 2023a; Hainline et al. 2024b) of galaxies during the reionization era. Their discovery, however, also presents a fortuitous new avenue to directly probe the buildup of pristine HI in galaxies at this critical epoch.

In this work, we systematically characterize these effects by measuring the prevalence and prominence of strong  $Ly\alpha$  emission and damped  $Ly\alpha$  absorption in a large spectroscopic sample of star-forming galaxies at  $z = 5.0\text{--}13.4$  observed with JWST/NIRSpec. The observational data for these galaxies are all presented here as part of the DAWN JWST Archive (DJA), which includes fully reduced and post-processed 1D and 2D spectra of all public JWST/NIRSpec data. The large JWST/NIRSpec archive forms the basis of the JWST/NIRSpec PRIMordial gas Mass AssemblY (PRIMAL) legacy survey presented here. In this first part of the survey we aim to disentangle the  $Ly\alpha$  damping wings produced from HI gas in the immediate surroundings of galaxies (ISM or CGM) from the effects of an increasingly neutral IGM, and establish statistical correlations for these features and the presence of  $Ly\alpha$  emission with the physical properties of the galaxies. We further make the reduced spectra, source catalogs, and the spectroscopically derived results for each galaxy in the JWST-PRIMAL sample publicly available on dedicated webpages.

We have structured the paper as follows. In Sect. 2 we present the observations, describe the spectroscopic reduction and post-processing, and outline the JWST-PRIMAL sample compilation. In Sect. 3, we detail the spectroscopic analysis of  $Ly\alpha$ , the nebular emission line fluxes and equivalent widths, and the inferred physical properties of the sample galaxies. In Sect. 4, we consider the cosmic evolution of galaxy DLAs and  $Ly\alpha$  emitters and quantify the physical correlations driving or hampering the observed  $Ly\alpha$  emission and excess damped

Ly $\alpha$  absorption. In Sect. 5, we discuss and conclude our work and provide a future outlook. All the spectroscopic data products and the analysis and results presented in this work are made publicly available on dedicated webpages. Throughout the paper, we assume concordance flat  $\Lambda$ CDM cosmology, with  $H_0 = 67.4 \text{ km s}^{-1} \text{ Mpc}^{-1}$ ,  $\Omega_m = 0.315$ , and  $\Omega_\Lambda = 0.685$  (Planck Collaboration VI 2020).

## 2. Observations, data processing, and sample selection

### 2.1. The DAWN JWST Archive (DJA) – Spectroscopy

The data considered in this work are processed as part of the DAWN JWST Archive (DJA), an online repository containing reduced images, photometric catalogs, and spectroscopic data for public JWST data products<sup>1</sup>. Here, we detail the spectroscopic data reduction and post-processing only, but see Valentino et al. (2023) for details on the imaging and photometric data. The DJA spectroscopic archive (DJA-Spec) is, at the time of writing, comprised of observations taken from some of the large Early Release Science (ERS), General Observer (GO) and Guaranteed Time (GTO) Cycle 1 & 2 programs, such as CEERS (ERS-1345; Finkelstein et al. 2023), GLASS-DDT (DDT-2756; Treu et al. 2022; Roberts-Borsani et al. 2023) (ERS-1324; Treu et al. 2022), JADES (GTO-1180, 1210, GO-3215; Bunker et al. 2024; Eisenstein et al. 2023), and UNCOVER (GO-2561; Bezanson et al. 2022). Many of the spectra analyzed in this work have been presented in dedicated survey papers cited above and in others (e.g., Arrabal Haro et al. 2023; Atek et al. 2024; Cameron et al. 2023), which also describe the sample selection and observational design of the different programs. With DJA-Spec we reduce and extract all of the spectra from this diverse array of observational programs with a single standardized pipeline.

### 2.2. DJA spectroscopic reduction and data processing

All data processing is performed with the publicly available GRIZLI (Brammer 2023a)<sup>2</sup> and MSAEXP (Brammer 2023b)<sup>3</sup> software modules, with the procedure as follows. The spectroscopic analysis begins with reprocessing the individual uncalibrated (unCAL) exposures retrieved from the Mikulski Archive for Space Telescopes (MAST) with the JWST Detector1Pipeline<sup>4</sup>. The pipeline is executed with the default parameters, but with snowblind<sup>5</sup> run between the jump and ramp\_fit steps for improved masking of the bright cosmic ray “snowballs” (Rigby et al. 2023). We remove a column average from the count-rate (rate) files produced by the first step to remove the  $1/f$  noise. We also find that scaling the read noise extension of the calibrated count-rate images by a factor of order  $\approx 1.4$  derived separately for each exposure is necessary to explain the variance of pixels in un-illuminated portions of the NIRSpec detectors. The corrected rate files are then run through the calwebb\_spec2 pipeline with its default parameters up through the photometric calibration (photom) step. The products at this stage are flux- and wavelength-calibrated 2D spectra for every source indicated with an open shutter in the attached MSA metadata, in a frame

cut out from the original detector pixel grid with curved spectral traces. These 2D SlitModel cutouts are saved to individual multi-extension FITS files, and it is here that the subsequent processing with MSAEXP departs significantly from the standard JWST pipeline.

The sky background of each source spectrum is removed by taking differences of the 2D cutout spectra taken at different “nod” positions of the telescope. Typically there are three nod offsets that shift the source by  $0''.5 \approx 5$  pixels within the MSA slitlet (the so-called 3-Shutter-Slitlet pattern), though the pipeline calculates the groupings automatically to handle other configurations of the heterogeneous archival observations. Average 2D source and sky spectra (and associated variance images) are calculated from exposures at a given nod position and from those at all other offset positions, respectively. We fit a 2D Gaussian profile to the source – sky spectra where the width is the quadrature sum of a free parameter and the wavelength-dependent width of the point spread function estimated from WebbPSF (Perrin et al. 2014)<sup>6</sup> and where a centroid shift is fit starting from an initial guess provided by the exposure MSA metadata and the catalog used to design the MSA mask. We note that the distribution of the centroid shifts follows approximately a Gaussian with  $\sigma_{\text{NMAD}} = 0.55$  pix (50 mas), which, while consistent with the expected precision of the input catalog astrometry and telescope pointing, is nontrivial and measurable and required for robust extraction of the 1D spectra. The sky-subtracted 2D spectra cutout at different offset positions are combined in a single rectified frame with orthogonal wavelength and cross-dispersion pixel axes by binning with a 2D histogram algorithm. Finally, the one-dimensional spectra and uncertainties are extracted with “optimal” weighting (Horne 1986) using the fitted 2D Gaussian profile. We note that Roberts-Borsani et al. (2024) have recently used the same pipeline presented here and developed for this work, presenting a similar effort to spectroscopically characterize galaxies at  $z > 5$  observed with NIRSpec Prism. We do not perform a photometric correction of the spectra, which is either nonexistent or comprised of a varying depth and filter configuration, but note that both the absolute and color-gradient of the final 1D spectra typically match to within 10–15% of the photometric data points, when available (see also Schaerer et al. 2024).

Currently (Apr. 2, 2024) DJA-Spec includes 7,319 individual combined spectra taken with the NIRSpec Prism/CLEAR and 1,665 combined spectra with NIRSpec medium and high-resolution gratings. A full overview can be found on the dedicated webpage<sup>7</sup>, where the following data are available: best-fit and inspected redshifts, the median signal-to-noise ratios (S/N) of the spectra, grades from visual inspection on the quality of the spectra, Hubble Space Telescope (HST) and JWST/NIRCam imaging thumbnails, and optimally reduced and extracted 2D and 1D spectra. We find that, depending on the survey,  $\sim 60$ – $70\%$  secure spectroscopic redshifts can be extracted for the sources in a given mask.

In this work, we focus on the JWST/NIRSpec Prism spectroscopic data from DJA due to the higher continuum sensitivity which is essential for quantifying the evidence for Ly $\alpha$  emission or absorption damping wings. JWST’s NIRSpec Prism configuration covers the entire near-infrared passband from  $0.6 \mu\text{m}$  to  $5.3 \mu\text{m}$ , with a varying spectral resolution from the minimum  $\mathcal{R} \approx 30$  at  $1.1 \mu\text{m}$  to  $\mathcal{R} \approx 330$  in the reddest end (Jakobsen et al.

<sup>1</sup> <https://dawn-cph.github.io/dja>

<sup>2</sup> <https://github.com/gbrammer/grizli>

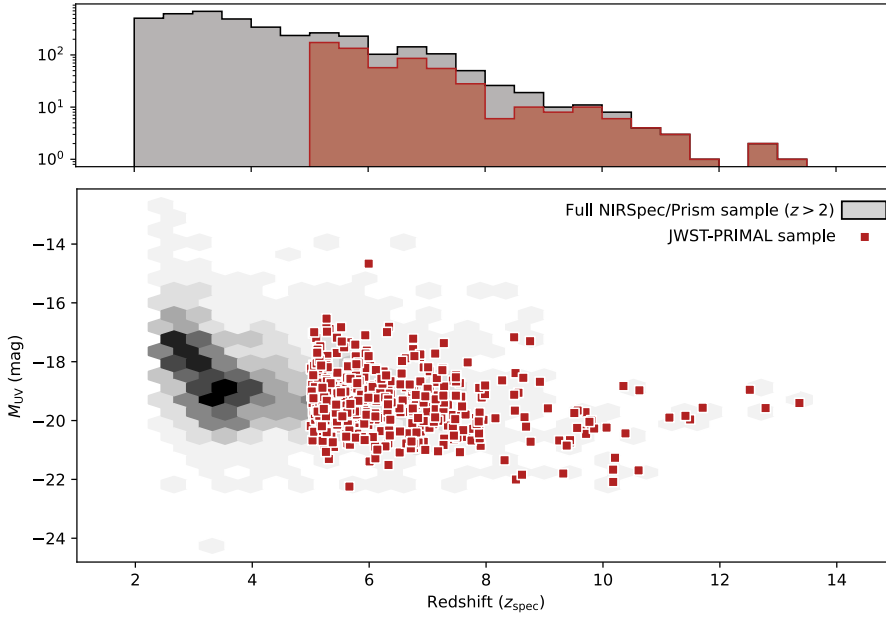
<sup>3</sup> <https://github.com/gbrammer/msaexp>

<sup>4</sup> JWST pipeline version 1.12.5; CRDS\_CONTEXT = jwst\_1180.pmap

<sup>5</sup> <https://github.com/mpi-astronomy/snowblind>

<sup>6</sup> The wavelength-dependent PSF is included in the MSAEXP repository.

<sup>7</sup> [https://s3.amazonaws.com/msaexp-nirspec/extractions/nirspec\\_graded\\_v2.html](https://s3.amazonaws.com/msaexp-nirspec/extractions/nirspec_graded_v2.html)



**Fig. 1.** Absolute rest-frame UV magnitude ( $M_{UV}$ ) as a function of the spectroscopic redshift for the JWST/NIRSpec Prism sources at  $z > 2$  from DJA with robust spectroscopic redshifts (grade 3) shown in gray. The JWST-PRIMAL sources are highlighted by the red squares, selected by requiring  $z_{\text{spec}} > 5.0$ , an integrated S/N  $> 3$  around the redshifted Ly $\alpha$  wavelength region, and full NIRSpec Prism spectroscopic wavelength coverage from 0.6–5.3  $\mu\text{m}$  as outlined in Sect. 2.3.

2022). We have carefully reduced and processed the spectra in DJA in a homogeneous way, limiting any biases between reduction and calibration codes across surveys. Further, this optimizes the consistency and uniformity of the JWST-PRIMAL sample measurements and the DJA-Spec sample as a whole.

### 2.3. JWST-PRIMAL sample selection

The main objective of this paper is to accurately model the Ly $\alpha$  damping wings and/or measure strong Ly $\alpha$  emission, and spectroscopically derive the physical properties of galaxies during the reionization epoch. For this, we require the following criteria for the sources to enter our archival JWST-PRIMAL sample:

1. A robust, spectroscopic redshift measurement at  $z > 5.0$  from a minimum of one emission line detected at  $>3\sigma$  (in addition to the Ly $\alpha$  break).
2. An integrated signal-to-noise over the redshifted Ly $\alpha$  and rest-frame UV (1550  $\text{\AA}$ ) regions of S/N  $> 3$ .
3. Flux density measurements over the entire NIRSpec Prism wavelength coverage from 0.6–5.3  $\mu\text{m}$ .

These criteria ensure that the redshift of the foreground HI gas absorber, from interstellar to intergalactic scales, and the shape of the Ly $\alpha$  line profile can be accurately measured. Further, the requirement of the full spectral coverage enables a full characterization of each source in the JWST-PRIMAL sample, including robust spectroscopic modeling of the stellar continuum and rest-frame UV spectral slope, emission line fluxes and equivalent widths EWs, and direct constraints on the star formation rates (SFRs), metallicities, and ionization parameters for all sources.

In total, 584 sources from DJA-Spec meet these criteria. Fig. 1 shows the absolute UV magnitude,  $M_{UV}$ , as a function of redshift for the JWST-PRIMAL sources and compared to the underlying DJA-Spec sample (at  $z > 2$ ). Table 1 summarizes the target list, including their coordinates, original program and source ID, and emission-line redshifts.

## 3. Analysis and results

Here we detail the spectroscopic measurements derived for the full JWST-PRIMAL sample. We focus on the Ly $\alpha$  damping

wings, but also detail the measurements and basic physical properties of the sample galaxies. All the spectroscopically derived quantities are made available on a dedicated webpage<sup>8</sup>.

### 3.1. The Lyman- $\alpha$ damping parameter

During the reionization epoch at  $z \gtrsim 5$ , several factors add to the line shape of Ly $\alpha$  on integrated galaxy spectra: They can show strong Ly $\alpha$  emission (LAE; Matthee et al. 2018; Mason et al. 2018; Witstok et al. 2024a), damping wings from an increasingly neutral IGM (Miralda-Escudé 1998; Keating et al. 2023a; Chen 2023), excess continuum flux in the wings related to the size of the immediate ionized bubbles (McQuinn et al. 2008; Castellano et al. 2016, 2018; Fujimoto et al. 2023b; Umeda et al. 2023; Hayes & Scarlata 2023), strong damped Ly $\alpha$  absorption (DLA) from local HI gas (Heintz et al. 2024; D’Eugenio et al. 2024), intrinsic variations due to a changing velocity and density distribution of gas and dust in the ISM and CGM (Dayal et al. 2011; Verhamme et al. 2015, 2017; Gronke et al. 2017; Hutter et al. 2023) or even possibly be dominated by two-photon emission processes (Steidel et al. 2016; Chisholm et al. 2019; Cameron et al. 2024), though see Terp et al. (2024). Many of these effects are degenerate, so we have to disentangle them statistically. For this, we define a new simple diagnostic, which we denote the Ly $\alpha$  damping parameter:

$$D_{\text{Ly}\alpha} \equiv \int_{\lambda_{\text{Ly}\alpha,\text{low}}}^{\lambda_{\text{Ly}\alpha,\text{up}}} (1 - F_{\lambda}/F_{\text{cont}}) d\lambda / (1 + z_{\text{spec}}), \quad (1)$$

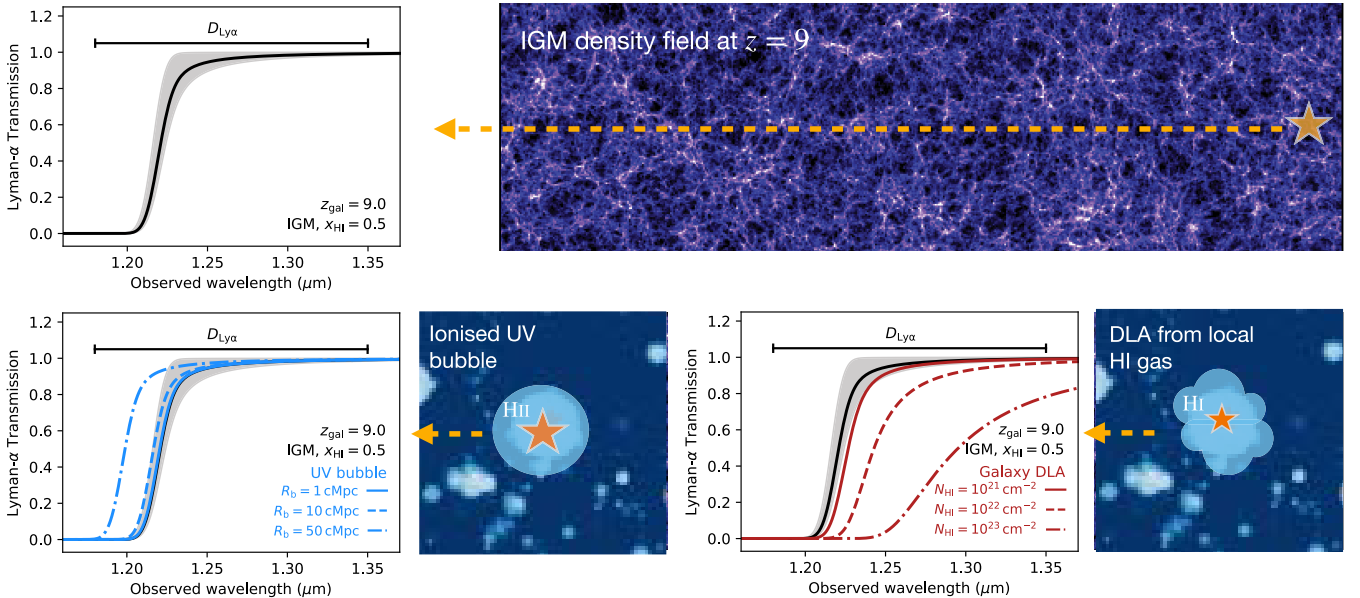
which is the equivalent width (EW) of the transmitted flux density,  $F_{\lambda}$  over the wavelength region covered by the instrumentally broadened Ly $\alpha$  transition. The continuum flux,  $F_{\text{cont}}$ , over the same region is estimated by extrapolating the rest-frame UV slope,  $\beta_{UV}$ , which is derived directly in each spectra from rest-frame 1400–2600  $\text{\AA}$  (see Sect. 3.4 for further details). This approximation of the stellar continuum is generally consistent with models of the continuum flux using galaxy templates (Heintz et al. 2024). We define the integration limits from

<sup>8</sup> <https://github.com/keheintz/jwst-primal>

**Table 1.** Archival references for the JWST-PRIMAL sample.

R.A. (deg) (1)	Decl. (deg) (2)	Prog. ID (3)	Src. ID (4)	S/N <sub>UV</sub> (5)	$z_{\text{spec}}$ (6)	Ref. (7)
214.94391	52.85004	1345	1374	62.87	5.0019	–
53.11629	–27.78944	6541	118675	37.10	5.0049	(Bu23)
3.58842	–30.34992	2561	42644	17.36	5.0151	–
3.55293	–30.40388	2756	160345	10.36	5.0175	–
189.02753	62.25374	1181	38493	29.5478	5.0212	(Bu23)
⋮	⋮	⋮	⋮	⋮	⋮	⋮
214.90663	52.945504	2750	10	25.564	11.49062	(AH23)
53.164762	–27.774626	3215	20130158	61.61	11.7060	(B23, ECL23)
53.166346	–27.821557	3215	20096216	42.17	12.5119	(ECL23)
3.513563	–30.35680	2561	38766	24.95	12.7822	(W23)
53.149881	–27.776502	3215	20128771	46.07	13.3605	(ECL23)

**Notes.** Column (1): Right ascension in degrees (J 2000). Column (2): Declination in degrees (J 2000). Column (3): Program ID under which the object was observed. Column (4): Designated MSA source ID. Column (5): Rest-frame UV signal-to-noise ratio. Column (6): Spectroscopic redshift. Column (7): The original survey paper references. A full version of this table can be found online at [github.com/keheintz/jwst-primal](https://github.com/keheintz/jwst-primal). References. (Bu23) Bunker et al. (2024); (AH23) Arrabal Haro et al. (2023); (ECL23) Curtis-Lake et al. (2023); (W23) Wang et al. (2023a).



**Fig. 2.** Schematic highlighting the intrinsic physical models affecting the shape of the Ly $\alpha$  transmission from an example galaxy at  $z = 9.0$ , all convolved by the nominal JWST/NIRSpec Prism spectral resolution. The default model with  $x_{\text{HI}} = 0.5$  is shown at the top left, where the grey-shaded region represents the effect of a partly ( $x_{\text{HI}} = 0.01$ ) to fully neutral ( $x_{\text{HI}} = 1.0$ ) IGM. For illustrative purposes the expected IGM density field at  $z = 9$  is shown in the top right, extracted from the ASTRAEUS simulations. In the bottom panels are shown the combined effects of the default IGM model and various sizes of the ionized UV bubble (left, blue:  $R_b = 1, 10, 50$  cMpc) and DLAs from local HI gas reservoirs (right, red:  $N_{\text{HI}} = 10^{21}, 10^{22}, 10^{23}$  cm $^{-2}$ ). The integration region for  $D_{\text{Ly}\alpha}$  is marked by the top line in all the Ly $\alpha$  transmission curve figures and is defined to encapsulate all the predicted physical scenarios.

$\lambda_{\text{Ly}\alpha, \text{low}} = 1180 \text{ \AA}$  to  $\lambda_{\text{Ly}\alpha, \text{up}} = 1350 \text{ \AA}$  (rest-frame) to capture most of the damping feature for the most extreme cases and to simultaneously limit the contamination from absorption or emission lines at longer wavelengths. We tested various lower limits and found this to be the bluest wavelength at which information is not lost at the resolution of NIRSpec Prism. We note that for potential strong AGN, there may be a contribution from the N  $\nu\lambda 1238, 1242$  doublet emission in this wavelength range. These features are generally not resolved and appear to contribute minimally to the continuum spectra in the majority of sources.

A set of physical models, shown as imprints on the Ly $\alpha$  transmission curves, are visualized in Fig. 2 including: varying

ionized bubble sizes from  $R_b = 1\text{--}50$  cMpc, neutral hydrogen fractions of the IGM from  $x_{\text{HI}} = 0.1\text{--}1.0$ , and DLA HI column densities  $N_{\text{HI}} = 10^{21}\text{--}10^{22}$  cm $^{-2}$ , all affecting the shape of the Ly $\alpha$  damping wings. The IGM models are based on the prescription by Miralda-Escudé (1998), using the approximation described in Totani et al. (2006), and the UV bubble sizes are included by varying the upper bound of the IGM integration region, such that  $z_{\text{IGM, upper}} < z_{\text{gal}}$  (e.g., McQuinn et al. 2008). These models are only meant as simply illustrations to guide the interpretation of  $D_{\text{Ly}\alpha}$ . A more careful comparison of the effect of the IGM on the Ly $\alpha$  damping wing and the  $D_{\text{Ly}\alpha}$  measure from simulations is provided in Sect. 5.2. The DLAs with varying HI

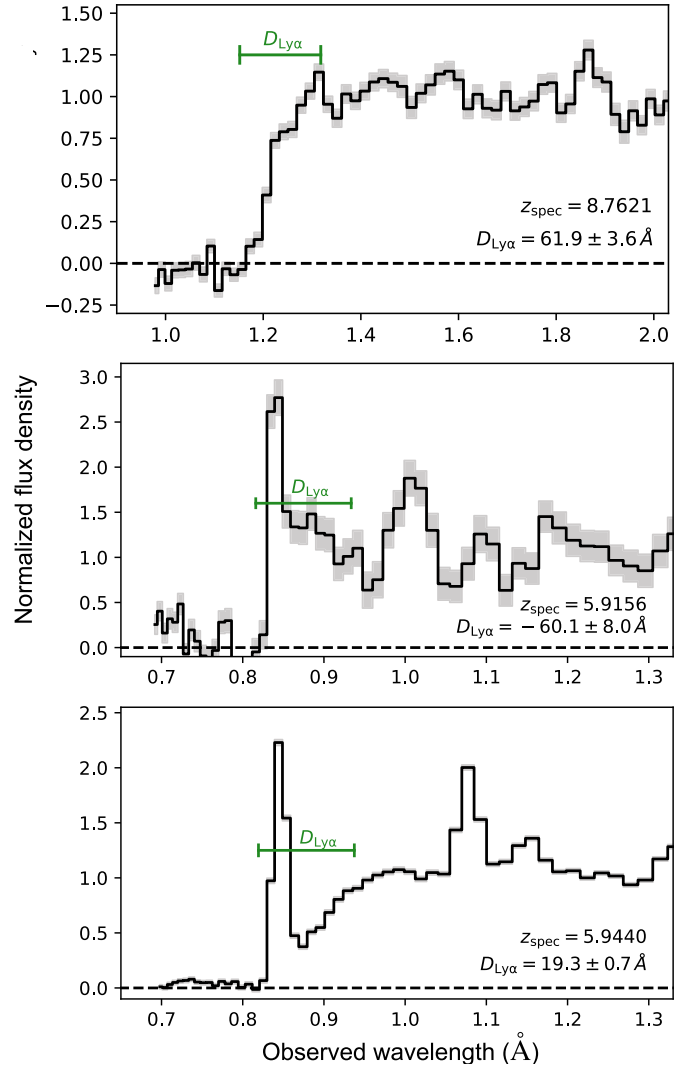
column densities are derived from the Voigt-profile approximation by [Tepper-García \(2006\)](#). The corresponding Ly $\alpha$  damping parameter  $D_{\text{Ly}\alpha}$  is listed for each model. The UV bubble and DLA models all assume a benchmark IGM fraction of  $x_{\text{HI}} = 0.5$ , and may therefore vary with the same scatter in  $D_{\text{Ly}\alpha}$  as observed for the IGM-only models.

We derive damping parameters for the full sample in the range  $D_{\text{Ly}\alpha} = -400$  to  $150 \text{ \AA}$ , with mean and median values for the full sample of  $52.6 \text{ \AA}$  and  $59.0 \text{ \AA}$ , respectively. We define four parameter spaces probing different physical regimes: (i) cases with excess continuum flux suggesting contributions from ionized bubbles or weak LAEs have  $D_{\text{Ly}\alpha} = 0\text{--}35 \text{ \AA}$ ; (ii) the effect on the Ly $\alpha$  damping wings of an increasing neutral IGM results in  $D_{\text{Ly}\alpha} = 35\text{--}55 \text{ \AA}$ ; (iii) excess damped Ly $\alpha$  absorption from local HI (with  $N_{\text{HI}} > 10^{21} \text{ cm}^{-2}$ ) have  $D_{\text{Ly}\alpha} > 55 \text{ \AA}$ ; and (iv) strong LAEs are typically observed with  $D_{\text{Ly}\alpha} < 0 \text{ \AA}$ . Fig. 3 shows three example spectra from the JWST-PRIMAL sample. One shows an example of a broad Ly $\alpha$  damping wing consistent with local HI absorption with  $D_{\text{Ly}\alpha} = 61.9 \pm 3.6 \text{ \AA}$ , consistent with  $N_{\text{HI}} \approx 10^{22} \text{ cm}^{-2}$ . Another shows strong Ly $\alpha$  emission with a measured  $D_{\text{Ly}\alpha} = -60.1 \pm 8.0 \text{ \AA}$ .

This simple diagnostic now allows us to quantify the number of sources in each distinct population. We caution, however, that the low spectral resolution of the JWST/NIRSpec Prism configuration may conceal weak Ly $\alpha$  emission in the continuum, which requires higher resolution spectroscopy to resolve (see e.g. the case for GN-z11; [Bunker et al. 2023](#)). This will likely cause a slight decrease in  $D_{\text{Ly}\alpha}$ . Further, as demonstrated in the bottom panel of Fig. 3, rare cases ( $\lesssim 1\%$  of the sample) that show both a substantial damped Ly $\alpha$  broadening and simultaneously a strong Ly $\alpha$  emission line will not be accurately identified, here with  $D_{\text{Ly}\alpha} = 19.3 \pm 0.7 \text{ \AA}$ , misleadingly suggesting mild excess Ly $\alpha$  continuum flux (the continuum flux of this particular source might be dominated by two-photon nebular emission, e.g., [Cameron et al. 2024](#), but see [Terp et al. 2024](#)). The JWST/NIRSpec shutter may also not cover the regions of the strongest Ly $\alpha$  emission, thereby underestimating the total Ly $\alpha$  photon output of the target source ([Jiang et al. 2023](#); [Ning et al. 2024](#)). The  $D_{\text{Ly}\alpha}$  parameter is thus most powerful in identifying the most extreme cases, such as strong Ly $\alpha$  emitters or galaxy DLAs with  $N_{\text{HI}} > 10^{21} \text{ cm}^{-2}$ , which uniquely have  $D_{\text{Ly}\alpha} < 0 \text{ \AA}$  and  $D_{\text{Ly}\alpha} > 50 \text{ \AA}$ , respectively.

### 3.2. Line flux measurements

For each source in the JWST-PRIMAL sample, we measure the spectroscopic redshifts from the most prominent nebular emission lines, when detected, including: the [O II]  $\lambda\lambda 3726, 3729$  and [O III]  $\lambda\lambda 4960, 5008$  doublets, [Ne III]  $\lambda 3870$ , He I  $\lambda 3889$ , the [S II]  $\lambda\lambda 6718, 6732$  doublet, and the Balmer lines H $\alpha$ , H $\beta$ , H $\gamma$ , and H $\delta$ . The [O II]  $\lambda\lambda 3726, 3729$  and [S II]  $\lambda\lambda 6718, 6732$  doublets are unresolved at all wavelengths in the NIRSpec Prism spectra, whereas the [O III]  $\lambda\lambda 4960, 5008$  doublet is marginally resolved at  $z \approx 6$  and fully resolved at  $z \gtrsim 7.5$  due to the increasing spectral resolution with wavelength in this particular configuration. H $\alpha$  and [N II]  $\lambda 6585$  are unresolved at all redshifts. [Ne III]  $\lambda 3870$  and He I  $\lambda 3889$  are also generally unresolved, except for the highest redshifts at  $z \gtrsim 10$ . Due to the broad wavelength coverage (0.6–5.3  $\mu\text{m}$ ), H $\alpha$  and [S II]  $\lambda\lambda 6718, 6732$  can be detected up to  $z \approx 7$ , H $\beta$  and the [O III]  $\lambda\lambda 4960, 5008$  doublet up to  $z \approx 9.5$ , and Ly $\alpha$  can be detected at all redshifts from  $z \gtrsim 4$ . The auroral [O III]  $\lambda 4363$  line is also in the majority of cases

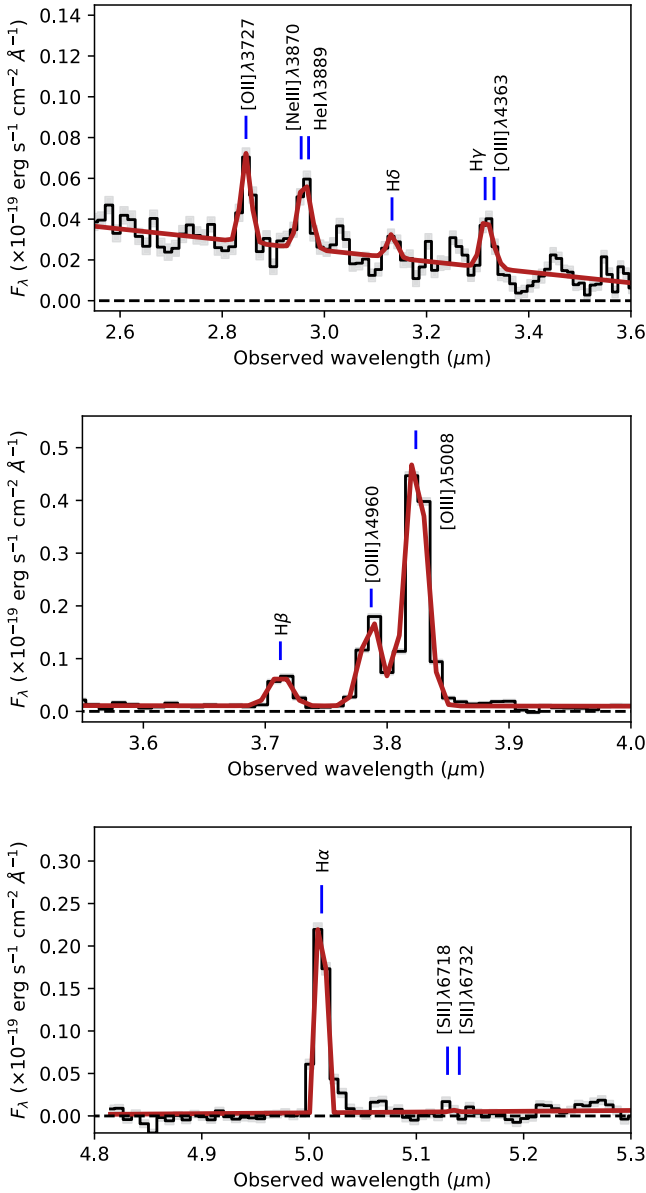


**Fig. 3.** Three examples of normalized NIRSpec Prism spectra from the JWST-PRIMAL sample. The top panel shows an example of a strong DLA from local HI absorption, the middle panel a strong LAE, and the bottom panel a combination of both. The redshifts,  $D_{\text{Ly}\alpha}$  measurements, and  $D_{\text{Ly}\alpha}$  integration regions are highlighted for each case.

blended with H $\gamma$  but becomes resolved at  $z \gtrsim 9.5$ , enabling robust measurements of this important transition ([Heintz et al. 2024](#); [Hsiao et al. 2023](#)). Unfortunately, the [O III]  $\lambda\lambda 4960, 5008$  doublet is redshifted out of the NIRSpec Prism coverage at approximately the same redshifts, hindering direct  $T_e$ -based metallicity measurements with the JWST/NIRSpec Prism configuration, except for a narrow redshift range ( $z \approx 9.3\text{--}9.5$ ).

To determine the physical properties of the star-forming regions in each of the target galaxies, we measure or derive upper bounds on the line fluxes for each of these nebular emission lines when covered by the spectra. We fit the continuum around the lines with a simple polynomial and superimpose a set of redshifted Gaussian line profiles at the rest-frame wavelength of each transition and fit for the redshift  $z_{\text{spec}}$ , line equivalent widths (EWs) and fluxes. The line widths (full-width-at-half-maximum; FWHM) are in most cases directly proportional to the line-spread function of the NIRSpec Prism configuration ([Jakobsen et al. 2022](#)) and is modeled as such. For each case, we tie  $z_{\text{spec}}$  and the FWHM corrected by the spectral resolution,





**Fig. 4.** Example of line flux modeling for one of the PRIMAL sources at  $z = 6.6345$ . The JWST/NIRSpec Prism spectrum is shown in black and the associated error spectrum in gray. The best fit continuum and Gaussian line model is shown in red.

effectively assuming that the nebular emission lines all originate from the same HII regions. An example of the line model fits is shown in Fig. 4. All the main identified lines, their derived line fluxes, or  $2\sigma$  upper limits, are summarized in Table 2 for the JWST-PRIMAL sample and provided in full online<sup>9</sup>, which also includes an expanded line list.

### 3.3. [O III]+H $\beta$ equivalent widths

A simple diagnostic of the specific star formation rate (sSFR) of galaxies is the emission line [O III]+H $\beta$  EWs. These have been found to increase with increasing redshifts (Khostovan et al. 2016; Endsley et al. 2021; Matthee et al. 2023), and toward lower stellar masses and metallicities (Malkan et al. 2017; Reddy et al. 2018b), due to the more intense ionization fields and

active star formation in early galaxies. We derive the [O III]+H $\beta$  EW for each of the sources in our sample that have these lines covered by the spectra (311 sources at  $z < 9.5$ ). As demonstrated in Fig. 5, we integrate the flux density over the spectral region covering the H $\beta$  and [O III]  $\lambda\lambda 4960, 5008$  lines, normalized by the continuum flux, and correct the redshifted EWs to the rest frame. For each source, we extrapolate the continuum flux at the position of the lines by fitting a set of polynomials to the spectral regions covering the entirety of the line complex.

In Fig. 6, we show the [O III]+H $\beta$  EW distribution of the full JWST-PRIMAL sample and divided into redshift bins:  $z = 5.0$ – $6.0$ ,  $z = 6.0$ – $7.0$ ,  $z = 7.0$ – $8.0$ , and  $z = 8.0$ – $9.5$ . The median of the full distribution is  $1045 \text{ \AA}$ , with 16th and 84th distribution percentiles from  $356 \text{ \AA}$  to  $2543 \text{ \AA}$ . We observe no apparent evolution with redshift ( $p$ -value is 0.07 from a two-sided KS test between the highest and lowest redshift bins), but note that the number statistics is largely dominated by the lowest redshift bins. We also compare the full distribution to the much larger, albeit photometrically derived distribution from the CEERS survey of galaxies at  $z = 6.5$ – $8.0$  (Endsley et al. 2023). The two distributions are in good agreement, though the peak of the spectroscopic distribution is slightly shifted towards higher EWs. We note that the spectroscopically derived [O III]+H $\beta$  EWs are much more accurate than photometric estimates, however, the latter potentially being underestimated by up to  $\approx 30\%$  (Duan et al. 2023). The fact that EW distribution of the JWST-PRIMAL sample is only slightly shifted, but otherwise follow the same distribution, indicates that there are no strong biases in our spectroscopic sample toward strong emission-line sources from the underlying galaxy population.

### 3.4. Ultraviolet spectral slopes, magnitudes, and luminosities

The UV spectral slope,  $\beta_{UV}$ , where  $F_\lambda \propto \lambda^{\beta_{UV}}$ , of the stellar continuum at wavelengths  $\lambda_{rest} = 1250$ – $2600 \text{ \AA}$  encodes key information regarding dust attenuation (Calzetti et al. 1994; Meurer et al. 1999), stellar metallicity (Cullen et al. 2021), the average stellar population age (Zackrisson et al. 2011), and the escape fraction of ionizing photons (Chisholm et al. 2022) from galaxies. The minimum value expected for standard stellar populations with a Galactic IMF and with negligible dust attenuation is  $\beta_{UV} \approx -2.5$  (Cullen et al. 2017; Reddy et al. 2018a). Steeper spectral power-law indices of  $\beta_{UV} \approx -3.0$  would require a very recent onset of star formation ( $\lesssim 2 \text{ Myr}$ ) or near-pristine, neutral gas and negligible nebular processing. JWST has enabled measurements of  $\beta_{UV}$  of galaxies at  $z > 10$ , hinting at even steeper spectral slopes, particularly for the faintest, less massive systems (Topping et al. 2023; Cullen et al. 2023; Austin et al. 2024). These results have, however, been inferred from photometric data alone, which may be contaminated by varying Ly $\alpha$  or metal emission (or absorption) line strengths or even low-luminosity AGN.

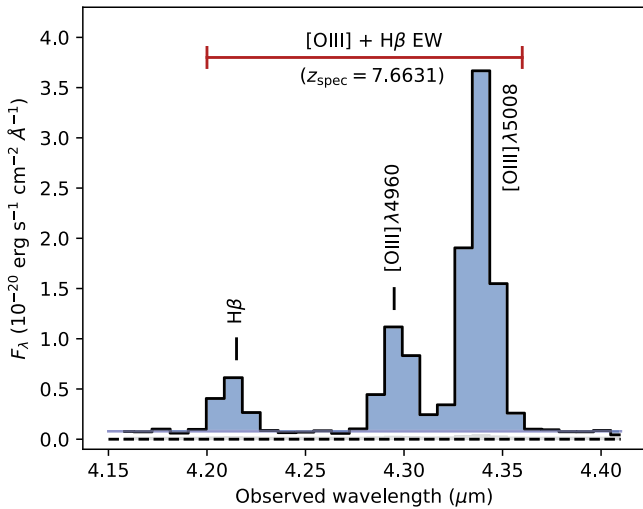
Here, we measure  $\beta_{UV}$  directly from the spectra for each of the JWST-PRIMAL sources at  $z = 5.0$ – $13.4$ , carefully masking any identified emission lines or broad Ly $\alpha$  absorption troughs. We use the emcee minimizer within the LMFIT framework and recover the parameter covariances to estimate the median and 16th to 84th percentiles of the posterior distributions on  $\beta_{UV}$  and the normalization constant derived at UV  $1550 \text{ \AA}$  rest-frame,  $m_{UV}$  of the best-fit model,  $F_\lambda = F_{UV,1550} \lambda^\beta$ . An example of one of the model fits is shown in Fig. 7. For our full sample, we recover UV continuum slopes ranging from  $\beta_{UV} = -1$  down to a few cases with  $\beta_{UV} \lesssim -3.0$ . This suggests that this set of rare galaxies are particularly dust- and metal-poor

<sup>9</sup> <https://github.com/keheintz/jwst-primal>

**Table 2.** Line flux measurements for the most prominent nebular emission lines JWST-PRIMAL sample.

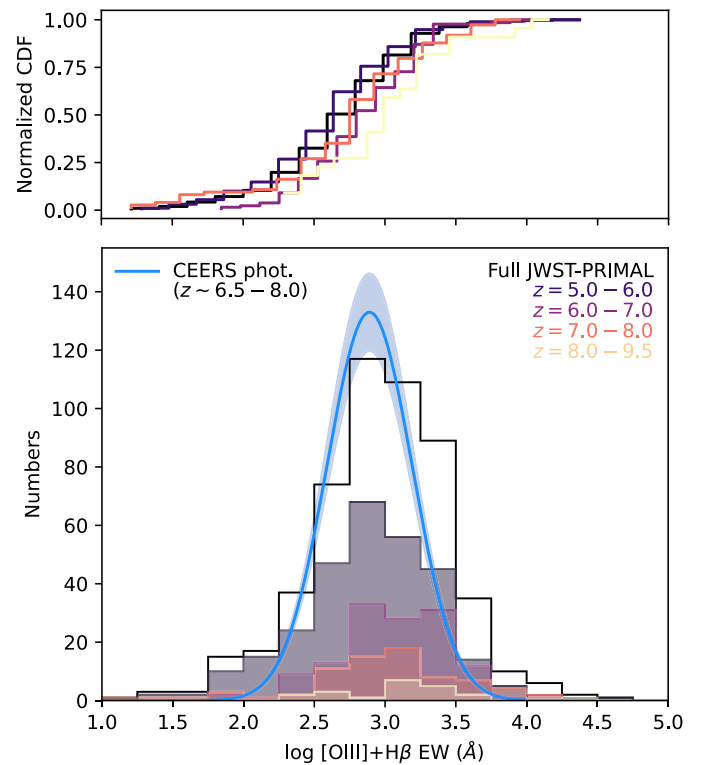
$z_{\text{spec}}$	[O II] $\lambda 3727$	[Ne III] $\lambda 3870$ + He I $\lambda 3889$	H $\delta$	H $\gamma$ + [O III] $\lambda 4363$	H $\beta$	[O III] $\lambda 5008$	H $\alpha$ + [N II] $\lambda 6585$	[S II] 6725
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
5.0019	424.9 $\pm$ 30.6	246.9 $\pm$ 41.6	88.6 $\pm$ 22.7	105.0 $\pm$ 24.7	329.9 $\pm$ 19.1	2072.9 $\pm$ 33.9	1004.5 $\pm$ 21.8	76.4 $\pm$ 12.7
5.0049	146.5 $\pm$ 15.7	<40.1	<22.9	<26.0	59.8 $\pm$ 8.4	249.1 $\pm$ 9.5	367.4 $\pm$ 9.4	49.1 $\pm$ 6.3
5.0151	<39.6	<52.8	<29.6	52.0 $\pm$ 16.6	22.2 $\pm$ 10.1	119.3 $\pm$ 9.7	121.6 $\pm$ 8.6	<14.2
5.0175	<108.0	<142.6	<83.4	<97.4	65.0 $\pm$ 28.9	290.6 $\pm$ 30.1	212.2 $\pm$ 26.1	<48.4
5.0212	137.9 $\pm$ 23.6	109.4 $\pm$ 31.1	54.4 $\pm$ 17.5	76.9 $\pm$ 19.4	138.5 $\pm$ 13.2	797.5 $\pm$ 21.5	409.4 $\pm$ 15.3	<21.4
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
11.4906	17.0 $\pm$ 5.0	<36.0, <36.0	<36.0	<45.0	–	–	–	–
11.7060	<8.7	7.3 $\pm$ 2.3, <9.6	<11.0	–	–	–	–	–
12.5119	9.1 $\pm$ 2.7	9.3 $\pm$ 3.1, <11	–	–	–	–	–	–
12.7822	<10.0	18 $\pm$ 6, <10.0	–	–	–	–	–	–
13.3606	6.0 $\pm$ 2.0	–	–	–	–	–	–	–

**Notes.** All measurements are reported in units of  $10^{-20}$  erg  $s^{-1}$   $\text{\AA}^{-1}$ . Uncertainties on measurements are stated at  $1\sigma$  standard deviations and upper limits at  $2\sigma$  confidence limits. Column (1): Spectroscopic redshift. Column (2): Unresolved [O II]  $\lambda 3726, 3729$  doublet line flux. Column (3): Unresolved [Ne III]  $\lambda 3870$  + He I  $\lambda 3889$  line flux measurements, except for the highest redshift sources ( $z > 10$ ). Column (4): Balmer H $\delta$ . Column (5): Unresolved Balmer H $\gamma$  + auroral [O III]  $\lambda 4363$  line flux measurements, except for the highest redshift sources ( $z > 9.5$ ). Column (6): Balmer H $\beta$ . Column (7): [O III]  $\lambda 5008$  line flux, where [O III]  $\lambda 4960$  is assumed to be 1/3 of this. Column (8): Unresolved Balmer H $\alpha$ + [N II]  $\lambda 6585$ . Column (9): Unresolved [S II]  $\lambda \lambda 6718, 6732$  doublet line flux. A full version of this table, including an expanded set of emission line identifications and measurements, can be found online.



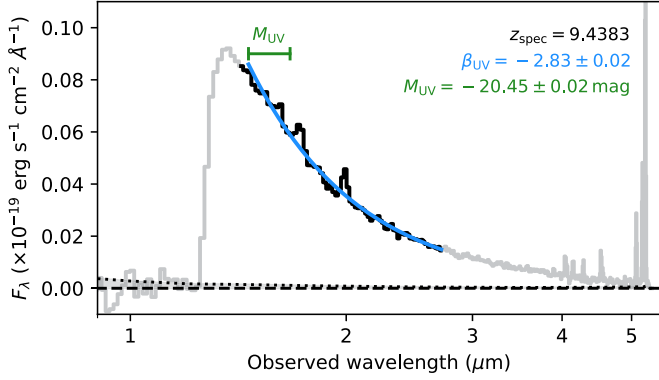
**Fig. 5.** Example of a [O III]+H $\beta$  line equivalent width measurement. The blue-shaded region marks the lines integrated over, also represented by the top red bar. The continuum flux is measured by fitting a linear polynomial to the spectrum on each side of the lines.

( $Z_{\text{gas}}/Z_{\odot} < 1\%$ ), have mean stellar populations with ages  $\lesssim 2$  Myr (e.g., Cullen et al. 2024), and likely have ionizing photon escape fractions of  $\geq 10\%$  (Chisholm et al. 2022). In Fig. 8, we show the derived  $\beta_{\text{UV}}$  spectral slopes as a function of redshift. We observe a marginal tendency for galaxies to have steeper UV slopes at higher redshifts, though still consistent with a non-evolution within the scatter in the sample distributions. We further note that distribution of dust and H II regions may account for the large scatter in the population around the expected slope for a standard stellar population with negligible dust and maximum nebular continuum contribution (e.g., Calzetti et al. 1994; Vijayan et al. 2024).



**Fig. 6.** [O III]+H $\beta$  emission line equivalent width distribution (bottom) and normalized cumulative distribution function (top). The full sample up to  $z \approx 9.5$ , where these lines can be detected, is shown by the black step function, and divided into redshift bins as indicated by the colors. The observed distribution of the spectroscopic JWST-PRIMAL sample is compared to the larger photometric CEERS survey results from Endsley et al. (2023).

To determine the UV luminosity,  $L_{\text{UV}}$ , and absolute magnitude,  $M_{\text{UV}}$ , for each source we derive the flux density and



**Fig. 7.** Example of spectral fitting of the UV power-law index,  $\beta_{UV}$ , where  $F_\lambda \propto \lambda^{\beta_{UV}}$ , of the stellar continuum. The grey curve shows the full NIRSpect Prism spectrum of a galaxy at  $z = 9.4383$ , with the error spectrum shown by the dotted line. The black part of the spectrum highlights the rest-frame fitting region from  $\lambda_{rest} = 1250\text{--}2600 \text{ \AA}$ . The top green bar marks the integration region for the  $M_{UV}$  measurements,  $\lambda = 1400\text{--}1700 \text{ \AA}$  in the rest frame.

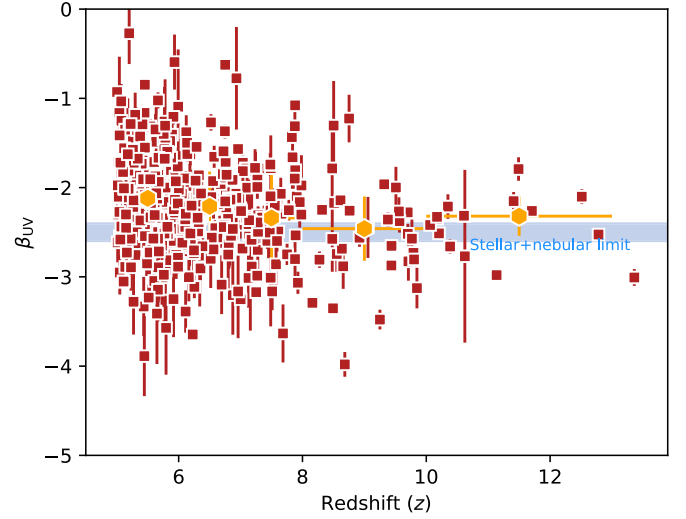
corresponding UV magnitude,  $m_{UV}$  at rest-frame  $1550 \text{ \AA}$  from the spectral power-law model. We convert the apparent  $m_{UV}$  to the intrinsic, absolute brightness,  $M_{UV}$ , via  $M_{UV} = m_{UV} - \mu(z) + 2.5 \times \log(1 + z)$ , where  $\mu(z)$  is the distance modulus at a given redshift derived from the `astropy.cosmology` module. The derived  $M_{UV}$  for each source are summarized in Table 3, not corrected for potential magnification factors for galaxies drawn from lensing cluster surveys. Our full sample spans absolute UV magnitudes of  $M_{UV} = -22$  to  $-16$  mag, respectively (see Fig. 1).

### 3.5. Star formation rate

The Balmer recombination lines originating from star-forming H II regions, H $\alpha$  in particular, are robust tracers of star formation on short ( $\lesssim 20$  Myr) timescales. The SFR has been found to increase at a given stellar mass (Whitaker et al. 2012; Speagle et al. 2014; Salmon et al. 2015; Thorne et al. 2021; Sandles et al. 2022) and rest-frame UV size (van der Wel et al. 2014; Ward et al. 2023; Langeroodi & Hjorth 2023) for galaxies at increasing redshifts, following the evolution in the peak of the stellar mass halo mass relation with redshift (e.g., Behroozi et al. 2019). The SFR can now be measured directly from the Balmer recombination lines for galaxies at  $z > 6$  with JWST, revealing intense star-forming galaxies during the reionization epoch (e.g., Heintz et al. 2023b; Shapley et al. 2023; Fujimoto et al. 2023a; Sanders et al. 2024; Nakajima et al. 2023; Curti et al. 2024). Given that H $\alpha$  is only detected for the subset of the sample at  $z < 7$ , and may be contaminated by [N II], we instead derive the SFR based on H $\beta$  for the majority of sources at  $z < 10$  as:

$$\text{SFR}_{H\beta} (M_\odot \text{ yr}^{-1}) = 5.5 \times 10^{-42} L_{H\beta} (\text{erg s}^{-1}) \times f_{H\alpha/H\beta} \quad (z < 10) \quad (2)$$

following Kennicutt (1998), but here assuming the more top-heavy initial mass function (IMF) from Kroupa (2001), which is likely more representative of the high-redshift galaxy population (e.g., Steinhardt et al. 2023). We derive the H $\beta$  line luminosity via  $L_{H\beta} = F_{H\beta} \times 4\pi D_L^2(z)$  with  $D_L(z)$  the luminosity distance at the given redshift and assume  $f_{H\alpha/H\beta} = 2.86$  from the Case B recombination scenario at  $T_e = 10^4 \text{ K}$  (Osterbrock & Ferland 2006). We propagate the uncertainty of  $\approx 20\text{--}30\%$  from the exact choice of the IMF in the relation, in addition to the dependence



**Fig. 8.** Galaxy UV spectral slope,  $\beta_{UV}$ , as a function of redshift. The red points mark the JWST-PRIMAL sample from this work, and the shown quantities are all derived directly from the NIRSpect Prism spectroscopy. The yellow hexagons show the mean  $\beta_{UV}$  in bins denoted by the horizontal errorbars, and the vertical errorbars represent the 16th to 84th percentile of the distribution in the respective bins. The blue shaded region represents the minimum value  $\beta_{UV} = -2.6$  to  $-2.4$  for a standard stellar population with negligible dust and maximum nebular continuum contribution (see Sect. 3.4 and Cullen et al. 2024).

on the electron density and temperature of the H II region which is only of the order  $\lesssim 5\%$ .

For the sources at  $z > 10$ , where H $\beta$  is redshifted out of the NIRSpect Prism spectral coverage, we instead rely on the [O II]  $\lambda 3727$  line strength

$$\text{SFR}_{[\text{O II}]} (M_\odot \text{ yr}^{-1}) = 1.0 \times 10^{-41} L_{[\text{O II}]} (\text{erg s}^{-1}) \quad (z > 10) \quad (3)$$

again following Kennicutt (1998), and assuming the Kroupa (2001) IMF. For the full sample, we derive SFRs in the range  $0.1\text{--}100 M_\odot \text{ yr}^{-1}$ , not corrected for the relevant magnification factors or dust extinction, as summarized in Table 3. The SFRs can also be derived from overall UV luminosity of the sources following Madau & Dickinson (2014), where

$$\text{SFR}_{UV} (M_\odot \text{ yr}^{-1}) = 1.0 \times 10^{-28} L_{UV} (\text{erg s}^{-1} \text{ Hz}^{-1}) \quad (4)$$

assuming 10% solar metallicity which traces star formation on a longer timescale ( $\sim 100$  Myr) compared to the Balmer lines ( $\sim 10$  Myr, see Calzetti 2013). We find that the  $\text{SFR}_{UV}$  estimate is consistent with that derived from the Balmer recombination lines or [O II] for the full sample, when allowing for up to  $A_V \approx 0.5$  mag of dust attenuation.

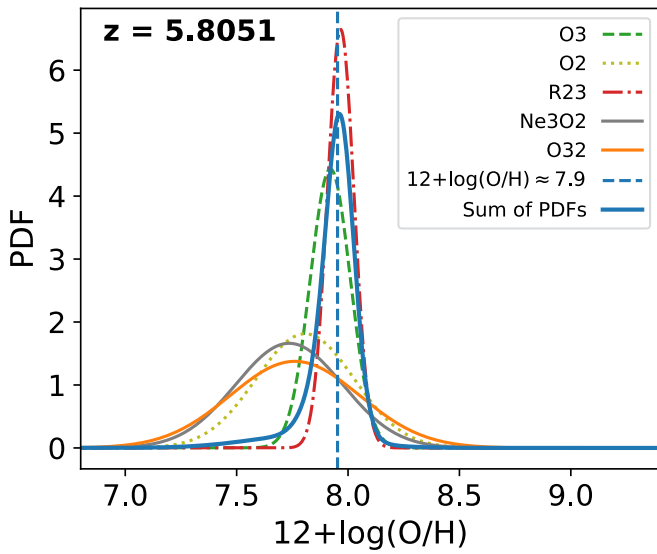
### 3.6. Gas-phase metallicity

One of the most fundamental characteristics of galaxies is their metallicity. This is notoriously difficult to measure directly, in particular for the highest redshift galaxies, motivating the use of strong-line diagnostics of the most prominent nebular emission lines (see e.g. Kewley et al. 2019; Maiolino & Mannucci 2019, for recent reviews). In rare cases, the auroral [O III]  $\lambda 4363$  emission line can be detected directly, which enables  $T_e$  sensitive measurements of the gas-phase metallicity (known as the direct  $T_e$ -method), typically quantified as the oxygen abundance,  $12 + \log(\text{O}/\text{H})$  (e.g., Izotov et al. 2006). This particular feature

**Table 3.** Spectroscopically derived physical properties of the JWST-PRIMAL sample.

$z_{\text{spec}}$	$M_{\text{UV}}$ (mag)	[O III]+H $\beta$ EW ( $\text{\AA}$ )	$\beta_{\text{UV}}$	$D_{\text{Ly}\alpha}$ ( $\text{\AA}$ )	SFR $_{\text{H}\beta, [\text{O II}]}$ ( $M_{\odot} \text{ yr}^{-1}$ )	$O_{32}$	$12 + \log(\text{O}/\text{H})$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
5.0019	-20.19	1714.7	$-2.11 \pm 0.06$	$12.1 \pm 2.5$	$14.1 \pm 0.8$	$4.9 \pm 0.4$	$7.71^{+0.09}_{-0.09}$
5.0049	-18.23	180.7	$-0.93 \pm 0.01$	$41.6 \pm 9.0$	$2.6 \pm 0.4$	$1.7 \pm 0.2$	$7.43^{+0.11}_{-0.19}$
5.0151	-18.26	856.6	$-2.61 \pm 0.22$	$36.8 \pm 8.3$	$1.0 \pm 0.4$	$7.5 \pm 9.2$	$7.45^{+0.10}_{-0.07}$
5.0175	-19.01	1144.0	$-1.97 \pm 0.01$	$85.7 \pm 13.7$	$2.8 \pm 1.2$	$6.1 \pm 6.9$	$7.29^{+0.10}_{-0.07}$
5.0212	-19.2	2284.2	$-2.30 \pm 0.13$	$-65.4 \pm 5.2$	$6.0 \pm 0.6$	$5.8 \pm 1.0$	$7.55^{+0.09}_{-0.08}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
11.4906	-19.96	–	$-1.79 \pm 0.13$	$35.6 \pm 7.1$	–	–	–
11.7060	-19.57	–	$-2.26 \pm 0.06$	$48.4 \pm 2.6$	$2.9 \pm 0.7$	–	$7.28 \pm 0.24$
12.5119	-18.96	–	$-2.10 \pm 0.08$	$90.4 \pm 3.9$	$0.4 \pm 0.2$	–	$7.60 \pm 0.24$
12.7822	-19.57	–	$-2.53 \pm 0.01$	$32.1 \pm 7.3$	$< 2.2$	–	–
13.3605	-19.40	–	$-3.01 \pm 0.09$	$56.7 \pm 3.1$	$1.5 \pm 0.5$	–	–

**Notes.** Column (1): Spectroscopic redshift. Column (2): Absolute UV (rest-frame  $\approx 1500 \text{\AA}$ ) magnitude. Column (3): Rest-frame [O III]+H $\beta$  equivalent width. Column (4): Rest-frame UV (1250–2600  $\text{\AA}$ ) spectral slope. Column (5): The Ly $\alpha$  damping parameter. Column (6): The star formation rate derived from H $\beta$  ( $z < 10$ ) or [O II] ( $z > 10$ ), **not corrected for dust**. Column (7): The [O III]  $\lambda 5008$ / [O II]  $\lambda 3727$  line flux ratio. Column (8): Gas-phase metallicity inferred from the joint PDF of the different strong-line calibrations from Sanders et al. (2024). A full version of this table can be found online.



**Fig. 9.** Example of joint gas-phase metallicity estimate for one of the sources. The blue-colored function shows the normalized joint probability density function (PDF) based on the normalized PDFs of the individual line-ratio diagnostics from Sanders et al. (2024). The blue dashed line marks the median of the joint PDF.

has now been detected in star-forming galaxies at  $z > 6$  with JWST (e.g., Schaerer et al. 2022; Curti et al. 2023; Heintz et al. 2023a; Nakajima et al. 2023; Sanders et al. 2024), even out to  $z \approx 10$  (Hsiao et al. 2023; Heintz et al. 2023b, 2024). The auroral [O III]  $\lambda 4363$  line feature, however, will only be resolved from H $\gamma$  in the minority of sources due to the increasing spectral resolution towards the red end of the spectra.

We therefore base our main metallicity measurements on strong-line calibrations. To preserve the homogeneity of the measurements, we use a single set of line diagnostics from Sanders et al. (2024) for the different line ratios, derived for galaxies at  $z = 2$ –9 through the direct  $T_e$ -method. Specifi-

cally, we consider the strong-line ratios:  $O3 = [\text{O III}] \lambda 5008 / \text{H}\beta$ ,  $O2 = [\text{O II}] \lambda 3727 / \text{H}\beta$ ,  $R23 = ([\text{O III}] \lambda \lambda 4960, 5008 + [\text{O II}] \lambda 3727) / \text{H}\beta$ ,  $\text{Ne3O2} = [\text{Ne III}] \lambda 3870 / [\text{O II}] \lambda 3727$ , and  $O32 = [\text{O III}] \lambda 5008 / [\text{O II}] \lambda 3727$ , which are available at most of the targeted redshifts. Based on the line-flux measurements and errors (representing the likelihood distributions of fluxes), we construct the probability density function (PDF) for each diagnostic ratio where the width denotes the scatter in the relation. We then combine the PDFs from the available calibrations, inversely weighted by their scatter, for each individual source to determine the gas-phase metallicity through the median and 16th to 84th percentiles of the joint PDF, as illustrated in Fig. 9. This more conservative approach takes into account the added uncertainty from sources with potential high ionization parameters, typically quantified via  $O_{32}$ , thus minimizing the uncertainty from assuming a single line-ratio.

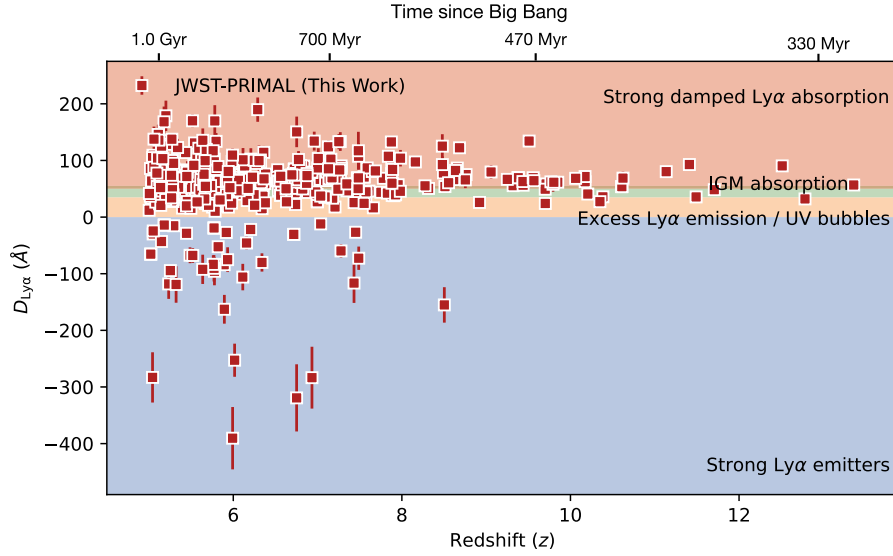
We derive gas-phase metallicities ranging from  $12 + \log(\text{O}/\text{H}) = 6.5$  to  $8.2$ , corresponding to  $0.6\%$  to  $30\%$  of solar abundance, respectively (given  $12 + \log(\text{O}/\text{H})_{\odot} = 8.69$ ; Asplund et al. 2009). The relevant line fluxes used for these measurements are provided in Table 2 and the derived metallicities are listed for each source in Table 3.

#### 4. Lyman- $\alpha$ emission and absorption in galaxies during the reionization epoch

With the spectroscopically derived physical properties for the full set of JWST-PRIMAL sample sources we can now investigate and chart the redshift evolution and the physical driver of strong galaxy DLAs produced by massive H I gas reservoirs to prominent ionized bubbles and Ly $\alpha$  emission during the reionization epoch.

##### 4.1. Charting $D_{\text{Ly}\alpha}$ as a function of redshift

To obtain a first overview of the prevalence and prominence of DLAs and LAEs across time in the reionization epoch, we show  $D_{\text{Ly}\alpha}$  as a function of redshift for the full sample in Fig. 10. This



**Fig. 10.**  $\text{Ly}\alpha$  damping parameter,  $D_{\text{Ly}\alpha}$ , for the full JWST-PRIMAL sample as a function of redshift. The corresponding age of the Universe assuming the cosmological parameters from [Planck Collaboration VI \(2020\)](#) is given at the top. The colored backgrounds indicate the typical physical regimes probed for a given  $D_{\text{Ly}\alpha}$ : Strong  $\text{Ly}\alpha$  emission ( $D_{\text{Ly}\alpha} < 0 \text{ \AA}$ ; blue), extended ionized bubbles or weak  $\text{Ly}\alpha$  emission ( $D_{\text{Ly}\alpha} = 0\text{--}35 \text{ \AA}$ ; orange), IGM absorption from  $x_{\text{HI}} = 0\text{--}1$  ( $D_{\text{Ly}\alpha} = 35\text{--}55 \text{ \AA}$ ; green), and strong galaxy integrated DLAs with  $N_{\text{HI}} > 10^{21} \text{ cm}^{-2}$  from local HI gas ( $D_{\text{Ly}\alpha} > 55 \text{ \AA}$ ; red).

**Table 4.**  $\text{Ly}\alpha$  absorption vs. emission statistics.

Redshift range	5.0 – 6.0	6.0–7.0	7.0–8.0	8.0 – 10.0	10.0 – 13.4
LAEs	36/354 ( $10^{+2}_{-1}\%$ )	12/169 ( $7^{+3}_{-2}\%$ )	7/103 ( $7^{+3}_{-2}\%$ )	1/46 (2%)	0/17 (0%)
$\text{Ly}\alpha$ exc.	46/354 ( $13^{+2}_{-2}\%$ )	17/169 ( $9^{+3}_{-3}\%$ )	6/103 ( $4^{+3}_{-1}\%$ )	2/46 (4%)	2/17 (6%)
IGM abs.	93/354 ( $26^{+3}_{-2}\%$ )	39/169 ( $17^{+3}_{-3}\%$ )	26/103 ( $22^{+5}_{-3}\%$ )	6/46 (7%)	7/17 (29%)
DLAs	179/318 ( $56^{+3}_{-3}\%$ )	101/157 ( $64^{+4}_{-4}\%$ )	64/96 ( $67^{+4}_{-5}\%$ )	37/45 ( $82^{+4}_{-7}\%$ )	8/17 ( $47^{+12}_{-11}\%$ )

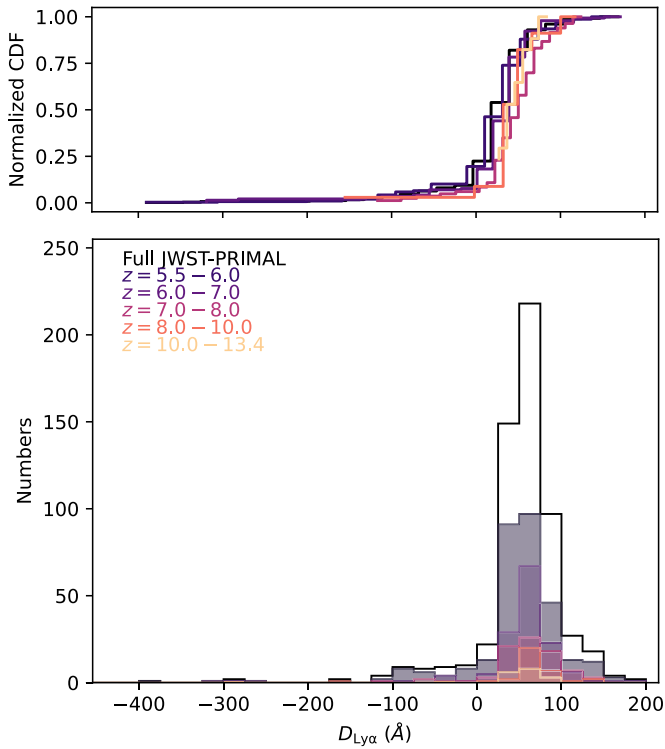
**Notes.** Errorbars denote the  $1\sigma$  confidence interval ( $c = 0.683$ ) on the computed fraction derived from the formalism by [Cameron \(2011\)](#). DLA fractions are computed by excluding the strong LAEs from the sample at each redshift bin.

represents the evolution from  $z = 13.4$  to  $z = 5.0$ , corresponding to approx. 300 Myr to 1 Gyr after the Big Bang. The  $D_{\text{Ly}\alpha}$  parameter space has been color-coded to highlight the regimes driven by the physical models for the expected dominant contribution to the observed  $\text{Ly}\alpha$  transmission curve. The number statistics of sources located in each region are summarized in Table 4, divided into redshift bins of  $z = 5.0\text{--}6.0$  (354 sources),  $z = 6.0\text{--}7.0$  (169),  $z = 7.0\text{--}8.0$  (103),  $z = 8.0\text{--}10.0$  (46), and  $z = 10.0\text{--}13.4$  (17).

We find that there are 56 strong LAEs in the full JWST-PRIMAL sample (i.e., 8%), based on  $D_{\text{Ly}\alpha} < 0 \text{ \AA}$ . None of these are found at  $z > 9$ . At  $z < 8$ , we find that 46 out of 533 of sources show strong LAEs (9%) in approximately equal fractions from  $z = 5.0\text{--}6.0$ ,  $z = 6.0\text{--}7.0$ , and  $z = 7.0\text{--}8.0$ , albeit with the most prominent LAEs quantified by the lowest  $D_{\text{Ly}\alpha}$  at  $z < 7$ . This fraction of LAEs could be as high as  $\approx 20\text{--}25\%$  if we include all sources with excess  $\text{Ly}\alpha$  flux as potential weakly emitting LAEs. These results are slightly lower than found previous literature studies, showing a typical fraction of LAEs of  $\approx 30\%$  at  $z \approx 6\text{--}8$  (e.g., [Saxena et al. 2023](#); [Nakane et al. 2023](#); [Simmonds et al. 2023](#); [Witstok et al. 2024a](#); [Jung et al. 2023](#); [Jones et al. 2024](#); [Witten et al. 2024](#); [Napolitano et al. 2024](#)). However, they were also found to be much more sparsely populated beyond  $z \gtrsim 8$  (though see [Tang et al. 2023](#); [Bunker et al. 2023](#); [Fujimoto et al. 2023b](#); [Witstok et al. 2024b](#)), consistent with our results, and as

also recovered by simulations ([Hutter et al. 2023](#)). We caution that the presence or fraction of LAEs are typically reported based on the intrinsic  $\text{Ly}\alpha$  line EW,  $\text{Ly}\alpha_{\text{EW},0} \gtrsim 20\text{--}25 \text{ \AA}$ , which likely causes the observed discrepancy. These findings would also naturally explain the null detections of LAEs at  $z = 8.8$  in large photometric surveys using narrow-band imaging such as the UltraVISTA (e.g., [Laursen et al. 2019](#)).

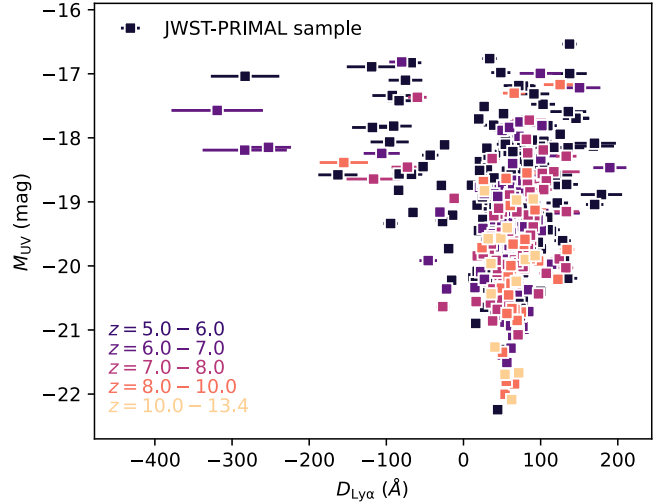
Similarly, we consider the redshift evolution of strong DLAs, here classified as sources with  $D_{\text{Ly}\alpha} > 55 \text{ \AA}$ , corresponding to HI column densities  $N_{\text{HI}} \gtrsim 10^{21} \text{ cm}^{-2}$ . Here, we exclude the strong LAEs to avoid biasing the result due to the known redshift evolution of the fraction of LAEs as also recovered here. We find that the fraction of galaxy DLAs only slightly increases from  $\approx 55\%$  at  $z \approx 5\text{--}6$  up to  $\approx 65\text{--}90\%$  at  $z > 8$ . A similar conclusion is reached by [Umeda et al. \(2023\)](#), who find an ubiquitous presence of extremely strong DLAs ( $N_{\text{HI}} \gtrsim 10^{22} \text{ cm}^{-2}$ ) at  $z > 10$ . This evolution is more clearly highlighted in Fig. 11 where we show the histograms of  $D_{\text{Ly}\alpha}$  for the five different redshift ranges, demonstrating higher prevalence of galaxy-integrated DLAs at increasing redshifts. These results are in sharp contrast to previous studies of Lyman-break galaxies near the peak of cosmic star formation at  $z \sim 2\text{--}3$  ([Pettini et al. 2000](#); [Steidel et al. 2010](#)), where only a small subset of the population ( $<25\%$ ) show integrated galaxy line-of-sight HI column densities of  $N_{\text{HI}} \approx 3 \times 10^{20} \text{ cm}^{-2}$  ([Shapley et al. 2003](#)). This corroborates



**Fig. 11.** Histogram of the Ly $\alpha$  damping parameter,  $D_{\text{Ly}\alpha}$ , distribution (bottom) and normalized cumulative distribution function (top). The full JWST-PRIMAL sample is represented by the black step function, and divided into sub-bins according to their redshift as indicated by the colors. Strong DLAs are more prevalent in galaxy spectra at increasing redshift.

our findings here of an increasing probability of having high HI column densities or larger fraction of DLAs at higher redshifts, down to  $z \approx 2-3$ . This may indicate that we are probing the epoch of first-galaxy assembly with excessive amounts of accreting or overdense pristine HI gas that are yet to be processed into stars or ionized by the first stellar sources. Indeed, the average gas mass density of galaxies have been inferred to increase by an order of magnitude through accretion from  $z = 10$  to 6 (see [Walter et al. 2020](#); [Heintz et al. 2022](#), and Sect. 5 for further discussion). Fig. 11 also demonstrates an excess of galaxies with strong LAEs ( $D_{\text{Ly}\alpha} < 00 \text{ \AA}$ ) at  $z = 5.0-6.0$  from the full distribution, distinct for the epoch when the universe is fully ionized.

We note that the fraction of  $\gtrsim 70\%$  galaxy DLAs at  $z > 9$  may be a lower bound, given that our selection criteria on the  $S/N_{\text{Ly}\alpha}$  will preferentially be biased towards strong Ly $\alpha$  emission. On the other hand, fainter, less massive sources are likely to have stronger Ly $\alpha$  EWs ([Saxena et al. 2024](#)) and contribute to most of the ionizing photons ([Lin et al. 2024](#)). To investigate this, we compare  $D_{\text{Ly}\alpha}$  to the spectroscopically derived  $M_{\text{UV}}$  in Fig. 12. We confirm the trend of the intrinsically faintest sources showing the highest Ly $\alpha$  EWs, with the strongest LAEs with  $D_{\text{Ly}\alpha} < -100 \text{ \AA}$  showing luminosities predominantly at  $M_{\text{UV}} > -19$  mag. This further suggests that the large HI gas overdensities are preferentially located close to bright, massive sources. In the following sections, we further investigate the underlying physical properties driving the prevalence and prominence of strong Ly $\alpha$  emission and absorption in these reionization era sources.



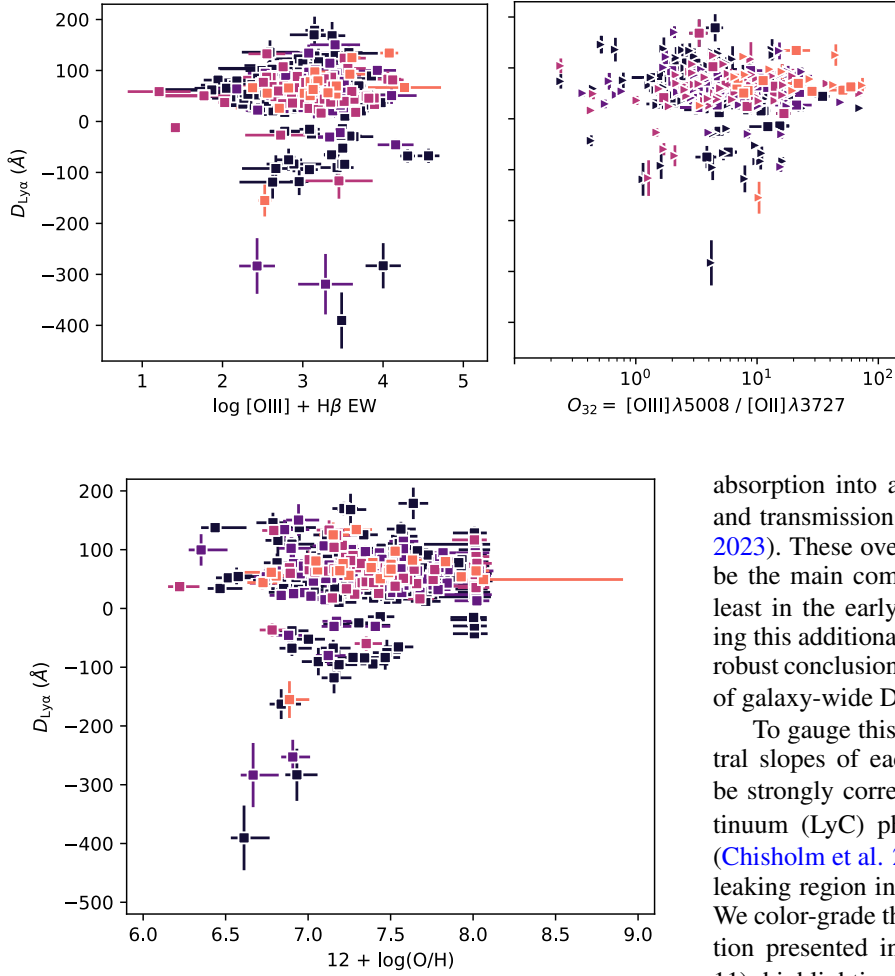
**Fig. 12.** Absolute UV brightness,  $M_{\text{UV}}$ , of the JWST-PRIMAL sources as a function of  $D_{\text{Ly}\alpha}$ . The sources with the highest Ly $\alpha$  EWs (negative  $D_{\text{Ly}\alpha}$ ) are observed to represent the UV faintest galaxy population at  $z > 5$ .

#### 4.2. The role of stellar ionization fields on Ly $\alpha$ emission and absorption

The intense UV radiation fields produced by younger, more massive stars are likely to play a key role in the ionization of HI and the production and escape of Ly $\alpha$  photons. To test this and quantify it in terms of our observations, we compare  $D_{\text{Ly}\alpha}$  to two common tracers of active star formation and high ionization: The [O III]+H $\beta$  EW, and the ionization parameter  $O_{32} = [\text{O III}] \lambda 5008 / [\text{O II}] \lambda 3727$  in Fig. 13. We find no apparent correlations between  $D_{\text{Ly}\alpha}$  and any of the two quantities. For example, strong galaxy DLAs are observed in galaxies spanning the entire dynamical range of [O III]+H $\beta$  EWs and  $O_{32}$  ionization. Similarly, strong LAEs are observed in galaxies spanning [O III]+H $\beta$  EWs of  $100-10^4 \text{ \AA}$ , and are not directly associated with sources with high or low  $O_{32}$  ratios. This is at odds with studies of local “green pea” galaxies at  $z \approx 0$ , believed to be robust analogs to high- $z$  star-forming galaxies, that find that larger Ly $\alpha$  emission EWs and escape fractions are commensurate with higher, more intense stellar ionization fields ([Yang et al. 2017](#); [Jaskot et al. 2019](#); [Hayes et al. 2023](#)). We note, however, that the galaxies at  $z > 8$  predominantly show high  $D_{\text{Ly}\alpha}$  absorption and  $O_{32}$  ionization and instead typically represent the large-scale environment or HI gas covering fraction at these redshifts. Consequently, our results disfavor a scenario where the HI abundance (or the lack thereof) is driven by the ionization output of the stellar sources in the galaxy.

#### 4.3. Gas-phase metallicity and Ly $\alpha$ absorption and emission

The gas-phase metallicity of galaxies is one of the key regulators of the gas cooling efficiency, the ionization state of the gas, and the escape fraction of Ly $\alpha$  and LyC photons. Lower metallicities are often associated with higher ionization and temperatures, effectively reducing the observed line-of-sight HI column density. By contrast, high metallicity and likely thereby also high dust content, will enable less efficient ionization of HI and as a consequence more efficiently scatter the output ionizing photons, effectively suppressing the intrinsic and observed escape fraction of Ly $\alpha$  photons ([Dayal et al. 2011](#); [Henry et al.](#)



**Fig. 14.** Gas-phase metallicity,  $12 + \log(\text{O}/\text{H})$ , as a function of the Ly $\alpha$  damping parameter,  $D_{\text{Ly}\alpha}$ . The strength of the Ly $\alpha$  absorption is observed to increase with increasing metallicity, whereas Ly $\alpha$  emission is most efficiently produced or allowed to escape in low-metallicity galaxies.

2015; Dijkstra et al. 2016; Verhamme et al. 2017; Yang et al. 2017; Izotov et al. 2018; Jaskot et al. 2019). We therefore expect a strong correlation between the gas-phase metallicity and  $D_{\text{Ly}\alpha}$  for our high-redshift sample sources.

To test this scenario, we show  $D_{\text{Ly}\alpha}$  as a function of gas-phase metallicity,  $12 + \log(\text{O}/\text{H})$ , for each of the sources in the JWST-PRIMAL sample in Fig. 14. The strongest LAEs ( $D_{\text{Ly}\alpha} < 0 \text{ \AA}$ ) are predominantly observed in galaxies with low metallicities, and ubiquitously found at  $12 + \log(\text{O}/\text{H}) < 7.7$ , corresponding to  $< 10\%$  solar metallicity. These results are consistent with the expectations that prominent Ly $\alpha$  emission can only escape from low-metallicity systems. By contrast, we find that galaxies with strong DLAs ( $D_{\text{Ly}\alpha} > 50 \text{ \AA}$ ) span the entire dynamical range of inferred metallicities,  $12 + \log(\text{O}/\text{H}) = 6.5\text{--}8.2$  (i.e.  $0.6\text{--}30\%$  solar).

#### 4.4. The impact of galaxy DLAs on the reionization history

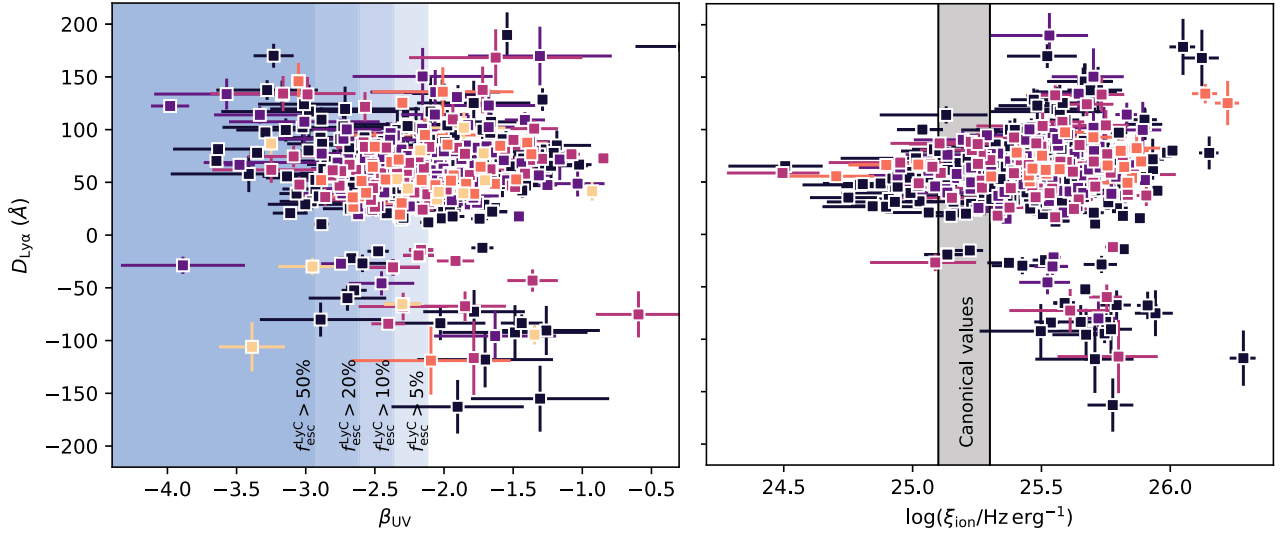
One of the major implications of high HI-column density DLAs observed in galaxy-integrated spectra at  $z \gtrsim 6$  is their effect on the escape fraction of Ly $\alpha$  and LyC photons, key to constrain the reionization history and topology. Numerical modelling and reionization simulations have previously only taken IGM

**Fig. 13.** Correlations of  $D_{\text{Ly}\alpha}$  with the galaxy properties that represent the average the specific SFR ( $[\text{O III}] + \text{H}\beta$  EW, left), and the intensity of the stellar ionization field ( $O_{32} = [\text{O III}] \lambda 5008 / [\text{O II}] \lambda 3727$ , right). The square symbols show the JWST-PRIMAL sources at  $z < 9.5$ , where  $[\text{O III}]$  can be detected, and a color-coded according to their redshift (identical to Fig. 12).

absorption into account when predicting the radiative transfer and transmission of Ly $\alpha$  (e.g., Laursen et al. 2019; Hutter et al. 2023). These overabundant local HI gas reservoirs are likely to be the main component in the transfer of ionizing photons, at least in the early evolutionary stages of galaxies, so quantifying this additional local ISM or CGM absorption is vital to draw robust conclusions and physically interpret the output and impact of galaxy-wide DLAs in simulations.

To gauge this effect, we first compare  $D_{\text{Ly}\alpha}$  to the  $\beta_{\text{UV}}$  spectral slopes of each galaxy in Fig. 15.  $\beta_{\text{UV}}$  has been found to be strongly correlated with the escape fraction of Lyman Continuum (LyC) photons,  $f_{\text{esc}}^{\text{LyC}}$ , in low-redshift galaxy samples (Chisholm et al. 2022; Flury et al. 2022) and for a small, LyC-leaking region in the Sunburst arc at  $z = 2.4$  (Kim et al. 2023). We color-grade the  $\beta_{\text{UV}}$  parameter space with the empirical relation presented in their paper Chisholm et al. (2022, their Eq. 11), highlighting the inferred escape fractions from  $f_{\text{esc}}^{\text{LyC}} = 5\%$  to  $50\%$ . We find no apparent correlations between  $\beta_{\text{UV}}$  or the inferred  $f_{\text{esc}}^{\text{LyC}}$  with the strength of the Ly $\alpha$  emission or DLA absorption. This indicates that the hardness of the UV spectral shape is not the main driver of the ionization of HI, and that larger HI abundances do not translate directly into lower escape fractions of ionizing photons. This is somewhat at odds with theoretical predictions and local observations, since both Ly $\alpha$  and LyC photons are affected by resonant scattering from abundant interstellar or circumgalactic HI gas columns (e.g., Flury et al. 2022; Xu et al. 2022; Begley et al. 2024), though other recent high-redshift results confirm the same scenario (Napolitano et al. 2023; Saxena et al. 2024). We caution, however, that the relations for  $f_{\text{esc}}^{\text{LyC}}$  derived from the local galaxy sample at  $z \approx 0.3$  studied by Chisholm et al. (2022) may not be universally representative for galaxies at  $z \gtrsim 6$  (Choustikov et al. 2024; Pahl et al. 2024). Since the high- $z$  galaxy population show even more abundant HI gas reservoirs than galaxies at equivalent metallicities or  $O_{32}$  ionization parameters (Heintz et al. 2024), the escape fraction of Ly $\alpha$  and LyC photons is likely correspondingly lower. Indeed, comparing the sources from Chisholm et al. (2022) that are mostly absorbed by the neutral HI gas, the inferred LyC escape fraction could be  $\sim 10\times$  lower.

To more directly relate the amount of photons capable of ionizing HI to the presence and strength of Ly $\alpha$  emission or DLAs, we also compare the ionizing photon production efficiency,  $\xi_{\text{ion}}$ , to  $D_{\text{Ly}\alpha}$  in Fig. 15.  $\xi_{\text{ion}}$  can be derived through the Balmer recombination line strengths, such as  $\text{H}\beta$ , since these accurately trace the ionizing photon production rate from massive stars on short ( $\lesssim 10$  Myr) timescales (Bouwens et al. 2016).



**Fig. 15.** Comparison of  $D_{\text{Ly}\alpha}$  to the spectral  $\beta_{\text{UV}}$  slope (left) and the ionizing photon production efficiency,  $\xi_{\text{ion}}$  (right). The JWST-PRIMAL sources are shown by the red squares and color-coded according to redshift (identical to Fig. 12). The light- to dark-blue colored regions in the left panel indicates increasing escape fractions of LyC photons,  $f_{\text{esc}}^{\text{LyC}}$ , based on the low-redshift empirical relation between  $\beta_{\text{UV}}$  and  $f_{\text{esc}}^{\text{LyC}}$  from Chisholm et al. (2022). The grey-colored band in the right panel shows the canonical reionization values  $\log(\xi_{\text{ion}}/\text{Hz erg}^{-1}) = 25.1\text{--}25.3$  (Robertson et al. 2013). Strong LAEs are observed with a large variety of  $\beta_{\text{UV}}$  slopes, but are predominantly associated with galaxies with high ionizing photon production efficiencies. Galaxies with strong DLAs span the entire probed range in  $\xi_{\text{ion}}$ .

Following Matthee et al. (2023), we define

$$\xi_{\text{ion}} (\text{Hz erg}^{-1}) = \frac{L_{\text{H}\beta} (\text{erg s}^{-1})}{c_{\text{H}\beta} (\text{erg}) L_{\text{UV}} (\text{erg s}^{-1} \text{Hz}^{-1})} \quad (5)$$

where  $c_{\text{H}\beta} = 4.86 \times 10^{-13}$  erg is the  $\text{H}\beta$  line-emission coefficient, assuming a Case B recombination scenario with  $T_e = 10^4$  K (e.g., Schaerer 2003). We find from Fig. 15 that galaxies with strong DLAs ( $D_{\text{Ly}\alpha} > 50 \text{ \AA}$ ) occupy most of the probed  $\xi_{\text{ion}}$  parameter space, though with a median of  $\log(\xi_{\text{ion}}/\text{Hz erg}^{-1}) = 25.55$ , higher than the canonical values,  $\log(\xi_{\text{ion}}/\text{Hz erg}^{-1}) = 25.1\text{--}25.3$  (Robertson et al. 2013; Atek et al. 2024). The large scatter in  $\xi_{\text{ion}}$  at  $D_{\text{Ly}\alpha} > 50 \text{ \AA}$  could indicate bursty star formation, particularly since  $\xi_{\text{ion}}$  traces massive O-stars, and may reflect typical younger stellar populations that are yet to ionize the neutral H I gas in their local surroundings. By contrast, galaxies that are strong LAEs tend to exhibit high ionizing photon production efficiencies, occupying the region with  $\log(\xi_{\text{ion}}/\text{Hz erg}^{-1}) > 25.5$ , higher than the canonical value. Since these are predominantly found at  $z < 8$ , we argue that it is mainly this particular subset of galaxies with high  $\xi_{\text{ion}}$  and late in the reionization epoch that provides the strongest contribution to this last phase transition.

## 5. Discussion

To place our results into context of the large-scale efforts to quantify the galaxy mass assembly and reionization timeline, we describe in this section a physical scenario connecting the observed prevalence and prominence of Ly $\alpha$  emission and DLA absorption from abundant, pristine H I gas in galaxies through redshift and as a function of their intrinsic physical properties. We further quantify the effect on  $D_{\text{Ly}\alpha}$  from the neutral hydrogen fraction based on more sophisticated IGM simulations and estimate the effect of strong galaxy DLAs on photometric redshifts measurements at  $z > 8$ , which appear to be systematically overestimated.

### 5.1. H I gas assembly and reionization across cosmic time

In this work, we quantified the strength of DLA absorption (Ly $\alpha$  emission) via the positive (negative) strength of  $D_{\text{Ly}\alpha}$ . We found that galaxies with high  $D_{\text{Ly}\alpha}$ , indicating substantial H I gas column densities in the ISM or CGM of galaxies, are ubiquitously observed at  $z > 9$  and in galaxies spanning the entire dynamical range of physical properties probed here. The fact that there appears to be no correlation between the amount of local H I gas (or the lack thereof) and the ionizing output of the galaxies themselves may suggest that DLAs trace particularly young galaxies, at all redshifts, that are yet to ionize the surrounding H I gas or convert it into molecules and eventually stars. For galaxies with  $D_{\text{Ly}\alpha} > 50 \text{ \AA}$ , equivalent to  $N_{\text{HI}} > 10^{21} \text{ cm}^{-2}$ , the absorbing H I gas is completely self-shielding of ionizing photons, even with substantial clumping with density variations down  $\Delta N_{\text{HI}} \approx 10^{-3}$  supporting this scenario.

On the contrary, strong LAEs are only observed at  $z \lesssim 8$ . This indicates that only after  $z \approx 8\text{--}9$ , on average, is the surrounding large-scale IGM sufficiently ionized for a prominent fraction of Ly $\alpha$  photons to escape. Indeed, the observed redshift evolution of  $D_{\text{Ly}\alpha}$  is in good qualitative agreement with the evolution of the Ly $\alpha$  escape fraction, as compiled by Saxena et al. (2024), Hayes & Scarlata (2023). Consequently, this favors more late and rapid reionization models (e.g., Naidu et al. 2020), over a smooth, early transition (e.g., Finkelstein et al. 2019). A similar rapid evolution seems to be suggested by damped Ly $\alpha$  wings in quasar spectra (Greig et al. 2017, 2019; Davies et al. 2018; Bañados et al. 2018; Yang et al. 2020; Wang et al. 2021; Ocvirk et al. 2021; Āurovčková et al. 2020, 2024) and Ly $\alpha$  emission statistics of galaxies (Ouchi et al. 2010; Sobacchi & Mesinger 2015; Mason et al. 2018, 2019), indicating a near-fully neutral IGM at  $z \approx 8$ .

The strong LAEs are predominantly associated with UV faint ( $M_{\text{UV}} > -19$  mag) galaxies with high ionizing photon production efficiencies,  $\log(\xi_{\text{ion}}/\text{Hz erg}^{-1}) > 25.5$ . This could either indicate that these particular systems were the most effi-



cient at ionizing their immediate surroundings, or that viewing angles play a larger effect at low ( $z < 8$ ) redshifts compared to their higher redshift counterparts. In this scenario, galaxies at  $z > 8$  are deeply embedded into dense, neutral gas overdensities with large covering fraction, showing strong absorption independent of the viewing angle. At lower redshifts, low-density  $N_{\text{HI}}$  channels are carved by ionizing photons and we thus observe a broader span in  $D_{\text{Ly}\alpha}$  at  $z < 8$ . This scenario would be commensurate with the non-dependence of the strength of LAEs with the  $O_{32}$  ionization parameter,  $[\text{O III}] + \text{H}\beta$  EW, the  $\beta_{\text{UV}}$  spectral slope and the inferred escape fraction of LyC photons. In fact, strongly directionally dependent escape of ionizing radiation has been proposed by simulations (Cen & Kimm 2015; Rosdahl et al. 2022; Yeh et al. 2023). Since only  $\approx 10$ –20% of the galaxies at  $z < 8$  have strong LAEs, we argue that it is either this particular subset of faint, low mass galaxies that drive reionization or that ionizing photons from galaxies at this epoch will only be able to escape through the  $\lesssim 20\%$  low-density sightlines.

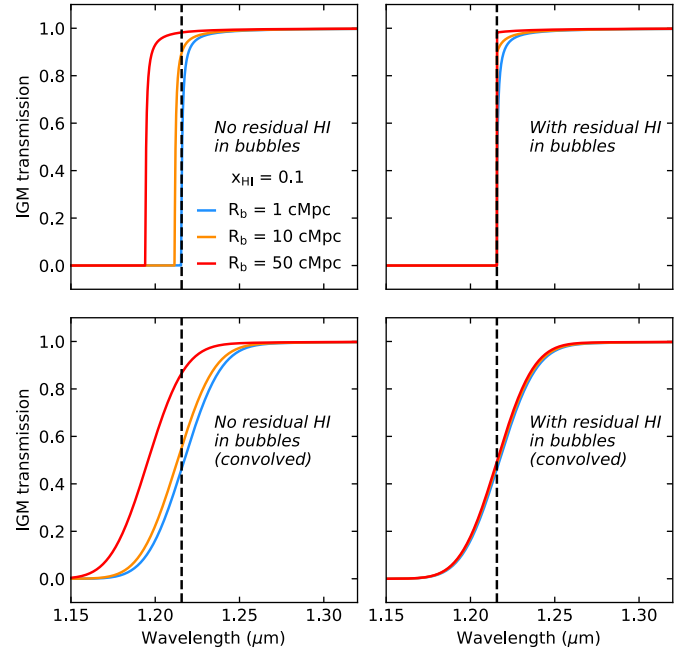
## 5.2. Insights into the $D_{\text{Ly}\alpha}$ parameter from simulations

So far, the IGM regime of  $D_{\text{Ly}\alpha}$  has been investigated assuming a homogeneously ionized IGM with an average value of the ionized fraction  $x_{\text{HI}}$  everywhere. It is therefore interesting to investigate how the  $D_{\text{Ly}\alpha}$  parameter evolves in simulations of patchy reionization, to better understand the allowed range of  $D_{\text{Ly}\alpha}$  and what some limitations may be.

Attenuation by the intergalactic medium is expected to play a significant role in regulating visibility of Ly $\alpha$  emission until the Universe was fully reionized (McQuinn et al. 2007), and will impact the  $D_{\text{Ly}\alpha}$  parameter to some extent over all of the regimes identified in Fig. 10. A full comparison of observed and simulated  $D_{\text{Ly}\alpha}$  across the whole parameter space would therefore require detailed modelling of the redshift evolution of ionized bubbles, as well as the detailed shape and of the Ly $\alpha$  emission profile (e.g., Hayes & Scarlata 2023). We instead focus here on identifying the allowed range of  $D_{\text{Ly}\alpha}$  for which attenuation by the IGM dominates the measurement. We measure  $D_{\text{Ly}\alpha}$  from IGM transmission spectra normalized by the continuum, i.e., with no contribution from Ly $\alpha$  emission.

First, we show how the Ly $\alpha$  transmission curves change based on the residual fraction of HI inside an ionized bubble. As discussed in several works, even a small amount of neutral gas can completely saturate absorption of the Ly $\alpha$  forest (Mesinger & Haiman 2007; Mason et al. 2018; Keating et al. 2023a). The difference in accounting for this residual neutral gas is shown in Fig. 16, which shows IGM transmission curves calculated using the Miralda-Escudé (1998) analytic model. In the top left panel, the bubbles are assumed to be completely ionized, allowing for transmission on the blue side of Ly $\alpha$ . When this neutral gas is accounted for, a sharp cutoff in IGM transmission blueward of the Ly $\alpha$  emission is observed. The bottom panel shows the same transmission profiles, but now convolved with the NIRSPEC PRISM instrumental profile. It is difficult to statistically disentangle the difference between the models with different bubble sizes once the residual HI is included.

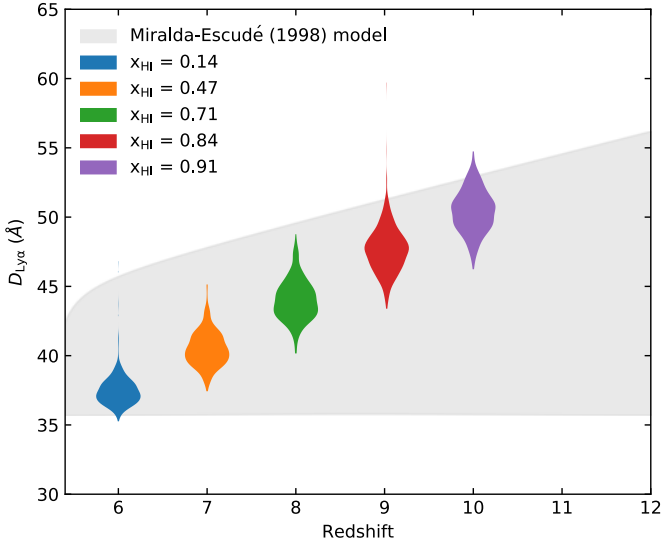
Based on these results, we estimate an allowed range of IGM-dominated  $D_{\text{Ly}\alpha}$  parameters, shown in Fig. 17. The lower value of  $D_{\text{Ly}\alpha}$  is given by a step function, where there is no transmission blueward of Ly $\alpha$  and complete transmission redward of Ly $\alpha$ . This mimics the limit of large bubbles containing a small amount of residual HI. The upper value of  $D_{\text{Ly}\alpha}$  is calculated from the Miralda-Escudé (1998) analytic model assuming



**Fig. 16.** Impact of considering residual neutral hydrogen on IGM transmission curves for models with ionized bubble sizes of 1 cMpc (blue), 10 cMpc (orange) and 50 cMpc (red) and a volume-averaged IGM neutral fraction  $x_{\text{HI}} = 0.1$  at redshift 9. The vertical black dashed line marks the observed Ly $\alpha$  wavelength at this redshift. The left column shows IGM damping wings calculated using the Miralda-Escudé (1998) model, assuming that the bubbles are completely ionized. The right column shows the effect of accounting for a small amount of residual neutral gas in the bubbles, which can completely saturate the absorption in the Ly $\alpha$  forest. The top panel shows the models before convolving with the NIRSPEC PRISM instrumental profile and the bottom panel shows the models after convolution.

the Universe is completely neutral and assuming a bubble size  $R_b = 0$  cMpc. We see that the maximum value of  $D_{\text{Ly}\alpha}$  grows with redshift, due to the evolution of the mean density of the Universe (Keating et al. 2023b).

For comparison, we also measure IGM-dominated  $D_{\text{Ly}\alpha}$  parameters directly from damping wings constructed from a simulation of inhomogeneous reionization. We use the spectra presented and described in Keating et al. (2023a), which were generated from sightlines through a simulation from the Sherwood-Relics simulation suite (Puchwein et al. 2023). We analyze results from five snapshots from this simulation at  $z = 6$ –10. The distributions of  $D_{\text{Ly}\alpha}$  parameters we measure are shown as the violin plots in Fig. 17. We find that the range of allowed IGM  $D_{\text{Ly}\alpha}$  parameters is well captured by the grey shaded region we defined. At  $z = 6$ , some sightlines have smaller  $D_{\text{Ly}\alpha}$  parameters as the Ly $\alpha$  forest starts to show transmission at this redshift. At  $z = 10$ , some sightlines have larger  $D_{\text{Ly}\alpha}$  parameters due to effects not captured in the analytic model for IGM damping wings, such as dense clumps of gas close to the host halo or the effect of infalling material. In general, however, the range of  $D_{\text{Ly}\alpha}$  parameters for which the reionization of the IGM plays a dominant role are in good agreement with the range of  $D_{\text{Ly}\alpha} = 35$ –50 Å as defined in Section 3.1. However, we note that the IGM will also play an important role in regulating the fraction of flux transmitted by galaxies showing evidence of Ly $\alpha$  emission, and hence falling into the regime of smaller values of  $D_{\text{Ly}\alpha}$ .



**Fig. 17.** Redshift evolution of the damping parameter for continuum-normalized spectra dominated by IGM absorption. The grey shaded region shows the range of damping parameters calculated from a step function at the observed Ly $\alpha$  wavelength (which sets the lower limit) and the IGM damping wing calculated using the Miralda-Escudé (1998) model in a completely neutral Universe (which sets the upper limit). The colored violin plots represent the distribution of damping parameters measured from IGM damping wings generated from the Sherwood-Relics simulation of an inhomogeneously reionized IGM as presented in Keating et al. (2023a). The volume-averaged HI fraction  $x_{\text{HI}}$  for each simulation snapshot is indicated in the legend.

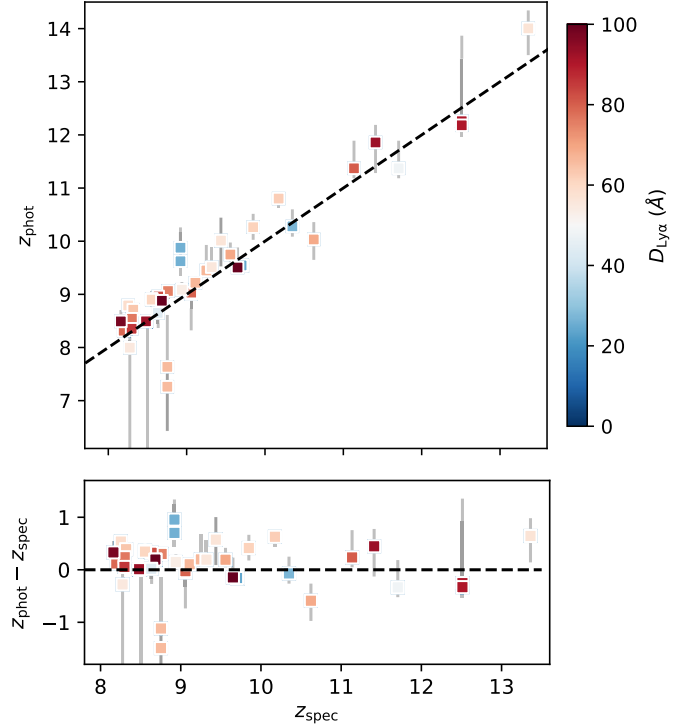
### 5.3. The effect of Ly $\alpha$ damping wings on high- $z$ photometric redshift estimates

Some of the first efforts to photometrically identify high-redshift galaxy candidates seemed to generally overpredict their redshifts when compared to the subsequent spectroscopic redshift determination of the sources at  $z > 8$  based on strong nebular emission lines (e.g., Finkelstein et al. 2024; Fujimoto et al. 2023a; Hainline et al. 2024b; Steinhardt et al. 2023; Serjeant & Bakx 2023). In Heintz et al. (2024), we proposed that this may be due to the strong damping wings observed in the large fraction of galaxies at these redshifts, suppressing the derived flux density in the blue-most photometric bands and thereby mimicking a slightly larger Ly $\alpha$  “break redshift” (e.g., Arrabal Haro et al. 2023).

To test this, we show the derived photometric redshifts for the subset of sources where these are available in the DJA (Valentino et al. 2023)<sup>10</sup> relative to the spectroscopically derived redshifts in Fig. 18. We find that the photometric redshifts are systematically overestimated for galaxies at  $z > 8$ , with median  $z_{\text{phot}} - z_{\text{spec}} = 0.11$ . Considering only the galaxies with strong DLAs ( $D_{\text{Ly}\alpha} > 50 \text{ \AA}$ ) the discrepancy becomes slightly larger, with median  $z_{\text{phot}} - z_{\text{spec}} = 0.15$ . To optimize future photometric redshift estimates, the effects of strong Ly $\alpha$  damping wings need to be incorporated into spectral energy distribution fitting algorithms such as EAZY (Brammer et al. 2008).

## 6. Summary and future outlook

In this work, we have compiled and systematically characterized 584 sources observed with JWST/NIRSpec during the reioniza-



**Fig. 18.** Photometric redshift,  $z_{\text{phot}}$ , estimates relative to the spectroscopic redshifts,  $z_{\text{spec}}$ , for the JWST-PRIMAL sample at  $z > 8$ . The sources are color-coded as a function of  $D_{\text{Ly}\alpha}$ , as indicated by the colorbar. The photometric redshifts systematically overpredicts the actual spectroscopically derived redshifts with median  $z_{\text{phot}} - z_{\text{spec}} = 0.11$  ( $z_{\text{phot}} - z_{\text{spec}} = 0.15$  for  $D_{\text{Ly}\alpha} > 50 \text{ \AA}$ ).

tion epoch, at  $z = 5.0\text{--}13.4$ , introducing the JWST-PRIMAL archival legacy survey. The main goal was to statistically quantify the presence and redshift evolution of strong damped Ly $\alpha$  absorption in galaxies, signifying abundant local HI gas reservoirs. To disentangle these systems from sources with damping parameters probing mainly a largely neutral IGM, or residing in extended ionized bubbles, we defined a new simple diagnostic, the Ly $\alpha$  damping parameter,  $D_{\text{Ly}\alpha}$ . This allowed us to statistically chart the HI gas assembly history, the onset and prevalence of strong Ly $\alpha$ -emitting galaxies and to explore the main underlying physical properties driving the emission and escape of ionizing and Ly $\alpha$  photons.

We found that the majority ( $\approx 65\text{--}90\%$ ) of galaxies at  $z > 8$  are dominated by large HI gas reservoirs in their local surroundings (ISM or CGM) with galaxy-integrated HI column densities,  $N_{\text{HI}} > 10^{21} \text{ cm}^{-2}$ . These strong galaxy DLAs appear to exist throughout the reionization epoch, even though the fraction was observed to decrease to  $\approx 60\%$  by  $z = 6$  of the overall galaxy population. Similar HI column densities have been theoretically predicted from cosmological simulations (e.g., Finlator et al. 2018; D’Odorico et al. 2018; Pallottini et al. 2022). Further, we found that these strong DLA galaxies represent a large variety of intrinsic physical properties in terms of their UV luminosity, gas-phase metallicity  $12 + \log(\text{O}/\text{H})$ , UV spectral slope  $\beta_{\text{UV}}$ , nebular [O III]+H $\beta$  line EWs,  $O_{32}$  ionization parameters, and ionizing photon production efficiency,  $\xi_{\text{ion}}$ . We thus surmised that strong galaxy DLAs are a sign of early galaxies in the process of formation, actively accreting and assembling large amount of primordial HI gas, that are yet to be ionized or processed into molecular gas and stars.

<sup>10</sup> Available at <https://dawn-cph.github.io/dja/imaging/v7/>

From the measured redshift evolution of  $D_{\text{Ly}\alpha}$  we were able to quantify the onset and increasing strength of  $\text{Ly}\alpha$ , enabling us to decode the reionization history directly. We found that at  $z \approx 8\text{--}10$  galaxy spectra mainly probe a substantially neutral CGM, whereas evidence for low  $\text{Ly}\alpha$  escape or significant ionized bubbles around these galaxies were only observed at later cosmic times ( $z < 8$ ). The fraction of LAEs was estimated to be  $\approx 20\%$  at  $z = 6$ . The strongest LAEs at  $z > 7$  have been found to reside in particular overdense or ionized regions (Saxena et al. 2023; Witstok et al. 2024a; Witten et al. 2024). From theoretical considerations and cosmological simulations (Pallottini et al. 2022), we expect most of the galaxies studied here to be formed within the first 200–300 Myr of cosmic time at  $z \approx 18\text{--}14$  and rapidly build up their stellar mass and on-going star formation via substantial HI gas accretion and assembly. Over the next 100–300 Myr, down to  $z = 8\text{--}10$ ,

The emergence of these sources showing strong “roll-overs” near the  $\text{Ly}\alpha$  edge in the first era of JWST/NIRSpec Prism spectra have important practical implications for future studies and observations: First, they complicate statistical inferences of the reionization history by tracing the increasing neutral fraction of the IGM from galaxy damping wing measurements (Keating et al. 2023a). Here, we provided the first step forward to statistically model the galaxy DLA population which will eventually be possible to remove to uncover the underlying IGM damping effects. Second, the first efforts in photometrically identifying high-redshift galaxy candidates seemed to show a systematic positive offset in  $z_{\text{phot}} - z_{\text{spec}}$  towards higher redshifts (Finkelstein et al. 2024; Fujimoto et al. 2023a; Hainline et al. 2024b; Steinhardt et al. 2023; Serjeant & Bakx 2023). This may naturally be explained by the increasing fraction of galaxies with stronger  $\text{Ly}\alpha$  damping wings at higher redshifts. This we quantified here and estimated a mean of 0.2 dex overestimations of the photometric redshifts of galaxies at  $z > 8$ . This effect would need to be incorporated into photo- $z$  codes to optimize the accuracy on the  $z_{\text{phot}}$  priors for the earliest galaxy population.

The next natural avenues to explore the early galaxy DLAs in more depth are to examine the effect of clustering, their correlations with physical sizes and SFR surface densities, and the HI gas content; a potential cause of the observed offset in the otherwise fundamental-metallicity relation (FMR). Similar to how strong LAEs are typically found to reside in high-density groups of galaxies, galaxy DLAs may trace overdense regions via their early assembly of HI. Moreover, the subset of galaxies showing strong DLAs integrated over their relative small UV sizes indicate exceptional dense gas surface densities (Kennicutt & Evans 2012). This, in combination with the inferred SFRs, may be key to understand the intense star formation and bright UV luminosities of early  $z > 9$  galaxies (e.g., Finkelstein et al. 2023; Franco et al. 2023; Harikane et al. 2023; Adams et al. 2024; Casey et al. 2024; Bouwens et al. 2023; Chemerynska et al. 2024; Mason et al. 2023; McLeod et al. 2024). Finally, the discovered offset from the FMR towards lower metallicities of galaxies at  $z = 7\text{--}10$  have been interpreted as evidence for excessive pristine HI gas inflow (Heintz et al. 2023b), which have later been confirmed with larger galaxy samples, though the exact redshift for this transition is still debated (e.g., Curti et al. 2024; Nakajima et al. 2023). This scenario will be possible to test directly by correlating the HI column densities of galaxies to their offset from the FMR, moving the 3D plane of galaxy properties to 4 dimensions. Obtaining higher-resolution spectroscopic data of these sources will further help quantifying the HI covering fraction and disentangle the effects on the  $\text{Ly}\alpha$  transmission curve from weak  $\text{Ly}\alpha$  emission to extended ionized bub-

bles and expand the probed physical properties to include the electron densities and temperatures of the gas.

While previous efforts with ALMA to quantify the HI gas mass content using far-infrared tracers such as  $[\text{C II}]\text{--}158\mu\text{m}$  and  $[\text{O I}]\text{--}63, 145\mu\text{m}$  have been successfully applied to galaxies in the epoch of reionization (Heintz et al. 2021, 2022; Vizgan et al. 2022; Liang et al. 2024; Wilson et al. 2024), the results presented here provide the most direct and statistical census of the abundance and assembly of the primordial HI gas in early Universe. This naturally paves the way and provide an important benchmark for next-generation radio facilities, targeting HI of the first galaxies through the 21-cm transition, such as the square kilometre array (SKA; Dewdney et al. 2009).

*Acknowledgements.* We would like to thank the referee for a detailed and constructive report, greatly improving the presentation of the results in this work. We would like to thank Peter Jakobsen for his vision and heroic endeavor in optimally designing the JWST/NIRSpec instrument and some of its first on-sky observations and for enlightening discussions about the intricacies of the NIRSpec data. Further, we would like to thank John Chisholm for helpful clarifications and discussions related to the escape fraction of ionizing photons and Aayush Saxena for enlightening conversations on the escape and absorption of Lyman- $\alpha$  photons. This work has received funding from the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number MB22.00072. The Cosmic Dawn Center (DAWN) is funded by the Danish National Research Foundation under grant DNRF140. The data products presented herein were retrieved from the DAWN JWST Archive (DJA). DJA is an initiative of the Cosmic Dawn Center, which is funded by the Danish National Research Foundation under grant DNRF140. P.D. acknowledges support from the NWO grant 016.VIDI.189.162 (“ODIN”) and warmly thanks the European Commission’s and University of Groningen’s CO-FUND Rosalind Franklin program. Support from the ERC Advanced Grant INTERSTELLAR H2020/740120 is kindly acknowledged (A.F.). S.G. acknowledges financial support from the Villum Young Investigator grants 37440 and 13160 and the Cosmic Dawn Center. M.K. was supported by the ANID BASAL project FB210003. G.E.M. acknowledges financial support from the Villum Young Investigator grants 37440 and 13160 and the Cosmic Dawn Center. J.W. acknowledges support from the Science and Technology Facilities Council (STFC), by the ERC through Advanced Grant 695671 “QUENCH”, by the UKRI Frontier Research grant RISEandFALL. Support for this work was provided by NASA through the NASA Hubble Fellowship grant HST-HF2-51515-001-A awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Incorporated, under NASA contract NAS5-26555. F.C. acknowledges support from a UKRI Frontier Research Guarantee Grant (PI Cullen; grant reference EP/X021025/1). J.H.W. acknowledges support by NSF grant AST-2108020 and NASA grants 80NSSC20K0520 and 80NSSC21K1053. NRT acknowledges support through STFC consolidated grant ST/W000857/1. M.J.H. is supported by the Swedish Research Council, Vetenskapsrådet, and is fellow of the Knut & Alice Wallenberg foundation. This work is based in part on observations made with the NASA/ESA/CSA *James Webb* Space Telescope. The data were obtained from the Mikulski Archive for Space Telescopes (MAST) at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-03127 for JWST. Software: This work made use of and acknowledge the following software: NumPy (Harris et al. 2020), Matplotlib (Hunter 2007), LMFIT (Newville et al. 2014), SciPy (Virtanen et al. 2020), grizli (Brammer 2023a), Astrodrizzle (Gonzaga et al. 2012), and MsaExp (v0.3; Brammer 2023b).

## Data availability

All the data presented in this work are made publicly available through DJA: <https://dawn-cph.github.io/dja/>, and are made easy to inspect and query in a dedicated webpage: [https://s3.amazonaws.com/msaexp-nirspec/extractions/nirspec\\_graded\\_v2.html](https://s3.amazonaws.com/msaexp-nirspec/extractions/nirspec_graded_v2.html). The spectroscopic data have been processed using the custom-made reduction pipeline MSAEXP, publicly available here: <https://github.com/gbrammer/msaexp>. All the data and catalogs related to JWST-PRIMAL are also available on a dedicated webpage: <https://github.com/keheintz/jwst-primal>.

## References

- Adams, N. J., Conselice, C. J., Austin, D., et al. 2024, *ApJ*, **965**, 169
- Arellano-Córdova, K. Z., Berg, D. A., Chisholm, J., et al. 2022, *ApJ*, **940**, L23
- Arrabal Haro, P., Dickinson, M., Finkelstein, S. L., et al. 2023, *Nature*, **622**, 707
- Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, *ARA&A*, **47**, 481
- Atek, H., Chemerynska, I., Wang, B., et al. 2023, *MNRAS*, **524**, 5486
- Atek, H., Labbé, I., Furtak, L. J., et al. 2024, *Nature*, **626**, 975
- Austin, D., Adams, N., Conselice, C. J., et al. 2023, *ApJ*, **952**, L7
- Austin, D., Conselice, C. J., & Adams, N. J. 2024, arXiv e-prints [arXiv:2404.10751]
- Bañados, E., Venemans, B. P., Mazzucchelli, C., et al. 2018, *Nature*, **553**, 473
- Begley, R., Cullen, F., McLure, R. J., et al. 2024, *MNRAS*, **527**, 4040
- Behroozi, P., Wechsler, R. H., Hearin, A. P., & Conroy, C. 2019, *MNRAS*, **488**, 3143
- Bezanson, R., Labbe, I., Whitaker, K. E., et al. 2022, arXiv e-prints [arXiv:2212.04026]
- Blumenthal, G. R., Faber, S. M., Primack, J. R., & Rees, M. J. 1984, *Nature*, **311**, 517
- Bosman, S. E. I., Davies, F. B., Becker, G. D., et al. 2022, *MNRAS*, **514**, 55
- Bouwens, R. J., Smit, R., Labbé, I., et al. 2016, *ApJ*, **831**, 176
- Bouwens, R. J., Stefanon, M., Brammer, G., et al. 2023, *MNRAS*, **523**, 1036
- Brammer, G. 2023a, <https://doi.org/10.5281/zenodo.1146904>
- Brammer, G. 2023b, <https://doi.org/10.5281/zenodo.7299500>
- Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, *ApJ*, **686**, 1503
- Brinchmann, J. 2023, *MNRAS*, **525**, 2087
- Bunker, A. J., Saxena, A., Cameron, A. J., et al. 2023, *A&A*, **677**, A88
- Bunker, A. J., Cameron, A. J., Curtis-Lake, E., et al. 2024, *A&A*, **690**, A288
- Calzetti, D. 2013, *Secular Evolution of Galaxies* (Cambridge, UK: Cambridge University Press), 419
- Calzetti, D., Kinney, A. L., & Storchi-Bergmann, T. 1994, *ApJ*, **429**, 582
- Cameron, A. G. W. 1957, *PASP*, **69**, 201
- Cameron, E. 2011, *PASA*, **28**, 128
- Cameron, A. J., Saxena, A., Bunker, A. J., et al. 2023, *A&A*, **677**, A115
- Cameron, A. J., Katz, H., Witten, C., et al. 2024, *MNRAS*, **534**, 523
- Casey, C. M., Akins, H. B., Shuntov, M., et al. 2024, *ApJ*, **965**, 98
- Castellano, M., Dayal, P., Pentericci, L., et al. 2016, *ApJ*, **818**, L3
- Castellano, M., Pentericci, L., Vanzella, E., et al. 2018, *ApJ*, **863**, L3
- Castellano, M., Fontana, A., Treu, T., et al. 2022, *ApJ*, **938**, L15
- Castellano, M., Napolitano, L., Fontana, A., et al. 2024, *ApJ*, **972**, 143
- Cen, R., & Kimm, T. 2015, *ApJ*, **801**, L25
- Chemerynska, I., Atek, H., Furtak, L. J., et al. 2024, *MNRAS*, **531**, 2615
- Chen, H. 2023, arXiv e-prints [arXiv:2307.04797]
- Chen, Z., Stark, D. P., Mason, C., et al. 2023, arXiv e-prints [arXiv:2311.13683]
- Chisholm, J., Rigby, J. R., Bayliss, M., et al. 2019, *ApJ*, **882**, 182
- Chisholm, J., Saldana-Lopez, A., Flury, S., et al. 2022, *MNRAS*, **517**, 5104
- Choustikov, N., Katz, H., Saxena, A., et al. 2024, *MNRAS*, **532**, 2463
- Cooke, J., & O'Meara, J. M. 2015, *ApJ*, **812**, L27
- Cullen, F., McLure, R. J., Khochfar, S., Dunlop, J. S., & Dalla Vecchia, C. 2017, *MNRAS*, **470**, 3006
- Cullen, F., Shapley, A. E., McLure, R. J., et al. 2021, *MNRAS*, **505**, 903
- Cullen, F., McLure, R. J., McLeod, D. J., et al. 2023, *MNRAS*, **520**, 14
- Cullen, F., McLeod, D. J., McLure, R. J., et al. 2024, *MNRAS*, **531**, 997
- Curti, M., Mannucci, F., Cresci, G., & Maiolino, R. 2020, *MNRAS*, **491**, 944
- Curti, M., D'Eugenio, F., Carniani, S., et al. 2023, *MNRAS*, **518**, 425
- Curti, M., Maiolino, R., Curtis-Lake, E., et al. 2024, *A&A*, **684**, A75
- Curtis-Lake, E., Carniani, S., Cameron, A., et al. 2023, *Nat. Astron.*, **7**, 622
- Davies, F. B., Hennawi, J. F., Bañados, E., et al. 2018, *ApJ*, **864**, 142
- Dayal, P., & Ferrara, A. 2018, *Phys. Rep.*, **780**, 1
- Dayal, P., Maselli, A., & Ferrara, A. 2011, *MNRAS*, **410**, 830
- Dayal, P., Volonteri, M., Greene, J. E., et al. 2024, A&A submitted [arXiv:2401.11242]
- D'Eugenio, F., Maiolino, R., Carniani, S., et al. 2024, *A&A*, **689**, A152
- Dewdney, P. E., Hall, P. J., Schilizzi, R. T., & Lazio, T. J. L. W. 2009, *IEEE Proc.*, **97**, 1482
- Dijkstra, M., Gronke, M., & Venkatesan, A. 2016, *ApJ*, **828**, 71
- D'Odorico, V., Feruglio, C., Ferrara, A., et al. 2018, *ApJ*, **863**, L29
- Donnan, C. T., McLeod, D. J., Dunlop, J. S., et al. 2023, *MNRAS*, **518**, 6011
- Duan, Q., Conselice, C. J., Li, Q., et al. 2023, arXiv e-prints, [arXiv:2309.14961]
- Đurovičková, D., Katz, H., Bosman, S. E. I., et al. 2020, *MNRAS*, **493**, 4256
- Đurovičková, D., Eilers, A. C., Chen, H., et al. 2024, arXiv e-prints [arXiv:2401.10328]
- Eisenstein, D. J., Willott, C., Alberts, S., et al. 2023, arXiv e-prints [arXiv:2306.02465]
- Endsley, R., Stark, D. P., Chevallard, J., & Charlot, S. 2021, *MNRAS*, **500**, 5229
- Endsley, R., Stark, D. P., Whittore, L., et al. 2023, *MNRAS*, **524**, 2312
- Fan, X., Bañados, E., & Simcoe, R. A. 2023, *ARA&A*, **61**, 373
- Ferrara, A., Pallottini, A., & Dayal, P. 2023, *MNRAS*, **522**, 3986
- Finkelstein, S. L., D'Aloisio, A., Paardekooper, J.-P., et al. 2019, *ApJ*, **879**, 36
- Finkelstein, S. L., Bagley, M. B., Ferguson, H. C., et al. 2023, *ApJ*, **946**, L13
- Finkelstein, S. L., Leung, G. C. K., Bagley, M. B., et al. 2024, *ApJ*, **969**, L2
- Finlator, K., Keating, L., Oppenheimer, B. D., Davé, R., & Zackrisson, E. 2018, *MNRAS*, **480**, 2628
- Flury, S. R., Jaskot, A. E., Ferguson, H. C., et al. 2022, *ApJ*, **930**, 126
- Franco, M., Akins, H. B., Casey, C. M., et al. 2023, arXiv e-prints [arXiv:2308.00751]
- Fujimoto, S., Arrabal Haro, P., Dickinson, M., et al. 2023a, *ApJ*, **949**, L25
- Fujimoto, S., Wang, B., Weaver, J., et al. 2023b, arXiv e-prints, [arXiv:2308.11609]
- Fynbo, J. P. U., Jakobsson, P., Prochaska, J. X., et al. 2009, *ApJS*, **185**, 526
- Gardner, J. P., Mather, J. C., Clampin, M., et al. 2006, *Space Sci. Rev.*, **123**, 485
- Gardner, J. P., Mather, J. C., Abbott, R., et al. 2023, *PASP*, **135**, 068001
- Gonzaga, S., Hack, W., Fruchter, A., & Mack, J. 2012, *The DrizzlePac Handbook*, *HST Data Handbook*
- Greig, B., Mesinger, A., Haiman, Z., & Simcoe, R. A. 2017, *MNRAS*, **466**, 4239
- Greig, B., Mesinger, A., & Bañados, E. 2019, *MNRAS*, **484**, 5094
- Gronke, M., Dijkstra, M., McCourt, M., & Oh, S. P. 2017, *A&A*, **607**, A71
- Hainline, K. N., D'Eugenio, F., & Jakobsen, P. 2024a, arXiv e-prints, [arXiv:2404.04325]
- Hainline, K. N., Johnson, B. D., Robertson, B., et al. 2024b, *ApJ*, **964**, 71
- Harikane, Y., Ouchi, M., Oguri, M., et al. 2023, *ApJS*, **265**, 5
- Harikane, Y., Nakajima, K., Ouchi, M., et al. 2024, *ApJ*, **960**, 56
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Nature*, **585**, 357
- Hashimoto, T., Laporte, N., Mawatari, K., et al. 2018, *Nature*, **557**, 392
- Haslbauer, M., Kroupa, P., Zonoozi, A. H., & Haghi, H. 2022, *ApJ*, **939**, L31
- Hayes, M. J., & Scarlata, C. 2023, *ApJ*, **954**, L14
- Hayes, M. J., Runholm, A., Scarlata, C., Gronke, M., & Rivera-Thorsen, T. E. 2023, *MNRAS*, **520**, 5903
- Heintz, K. E., Watson, D., Oesch, P. A., Narayanan, D., & Madden, S. C. 2021, *ApJ*, **922**, 147
- Heintz, K. E., Oesch, P. A., Aravena, M., et al. 2022, *ApJ*, **934**, L27
- Heintz, K. E., Giménez-Arteaga, C., Fujimoto, S., et al. 2023a, *ApJ*, **944**, L30
- Heintz, K. E., Brammer, G. B., Giménez-Arteaga, C., et al. 2023b, *Nat. Astron.*, **7**, 1517
- Heintz, K. E., De Cia, A., Thöne, C. C., et al. 2023c, *A&A*, **679**, A91
- Heintz, K. E., Watson, D., Brammer, G., et al. 2024, *Science*, **384**, 890
- Henry, A., Scarlata, C., Martin, C. L., & Erb, D. 2015, *ApJ*, **809**, 19
- Home, K. 1986, *PASP*, **98**, 609
- Hoyle, F. 1954, *ApJS*, **1**, 121
- Hsiao, T. Y. Y., Abdurro'uf, Coe, D., et al. 2023, arXiv e-prints [arXiv:2305.03042]
- Hu, W., Martin, C. L., Gronke, M., et al. 2023, *ApJ*, **956**, 39
- Hunter, J. D. 2007, *Comput. Sci. Eng.*, **9**, 90
- Hutter, A., Trebitsch, M., Dayal, P., et al. 2023, *MNRAS*, **524**, 6124
- Inayoshi, K., Harikane, Y., Inoue, A. K., Li, W., & Ho, L. C. 2022, *ApJ*, **938**, L10
- Izotov, Y. I., Stasińska, G., Meynet, G., Guseva, N. G., & Thuan, T. X. 2006, *A&A*, **448**, 955
- Izotov, Y. I., Worseck, G., Schaerer, D., et al. 2018, *MNRAS*, **478**, 4851
- Jakobsen, P., Ferruit, P., Alves de Oliveira, C., et al. 2022, *A&A*, **661**, A80
- Jakobsen, P., Fynbo, J. P. U., Ledoux, C., et al. 2006, *A&A*, **460**, L13
- Jaskot, A. E., Dowd, T., Oey, M. S., Scarlata, C., & McKinney, J. 2019, *ApJ*, **885**, 96
- Jiang, H., Wang, X., Cheng, C., et al. 2023, arXiv e-prints [arXiv:2312.04151]
- Jones, G. C., Bunker, A. J., Saxena, A., et al. 2024, *A&A*, **683**, A238
- Jung, I., Finkelstein, S. L., Arrabal Haro, P., et al. 2023, arXiv e-prints [arXiv:2304.05385]
- Katz, H., Saxena, A., Cameron, A. J., et al. 2023, *MNRAS*, **518**, 592
- Keating, L. C., Weinberger, L. H., Kulkarni, G., et al. 2020, *MNRAS*, **491**, 1736
- Keating, L. C., Bolton, J. S., Cullen, F., et al. 2023a, arXiv e-prints [arXiv:2308.05800]
- Keating, L. C., Puchwein, E., Bolton, J. S., Haehnelt, M. G., & Kulkarni, G. 2023b, arXiv e-prints [arXiv:2308.11709]
- Kennicutt, R. C. Jr 1998, *ARA&A*, **36**, 189
- Kennicutt, R. C., & Evans, N. J. 2012, *ARA&A*, **50**, 531
- Kereš, D., Katz, N., Weinberg, D. H., & Davé, R. 2005, *MNRAS*, **363**, 2
- Kewley, L. J., Nicholls, D. C., & Sutherland, R. S. 2019, *ARA&A*, **57**, 511
- Khostovan, A. A., Sobral, D., Mobasher, B., et al. 2016, *MNRAS*, **463**, 2363
- Kim, K. J., Bayliss, M. B., Rigby, J. R., et al. 2023, *ApJ*, **955**, L17
- Kroupa, P. 2001, *MNRAS*, **322**, 231
- Labbé, I., van Dokkum, P., Nelson, E., et al. 2023, *Nature*, **616**, 266
- Langeroodi, D., & Hjorth, J. 2023, arXiv e-prints [arXiv:2307.06336]
- Langeroodi, D., Hjorth, J., Chen, W., et al. 2023, *ApJ*, **957**, 39
- Lara-López, M. A., Cepa, J., Bongiovanni, A., et al. 2010a, *A&A*, **521**, L53

- Lara-López, M. A., Cepa, J., Bongiovanni, A., et al. 2010b, *A&A*, **519**, A31
- Laursen, P., Sommer-Larsen, J., & Razoumov, A. O. 2011, *ApJ*, **728**, 52
- Laursen, P., Sommer-Larsen, J., Milvang-Jensen, B., Fynbo, J. P. U., & Razoumov, A. O. 2019, *A&A*, **627**, A84
- Liang, L., Feldmann, R., Murray, N., et al. 2024, *MNRAS*, **528**, 499
- Lin, X., Cai, Z., Zou, S., et al. 2023, *ApJ*, **944**, L59
- Lin, Y.-H., Scarlata, C., Williams, H., et al. 2024, *MNRAS*, **527**, 4173
- Madau, P., & Dickinson, M. 2014, *ARA&A*, **52**, 415
- Maiolino, R., & Mannucci, F. 2019, *A&ARv*, **27**, 3
- Malkan, M. A., Cohen, D. P., Maruyama, M., et al. 2017, *ApJ*, **850**, 5
- Mannucci, F., Cresci, G., Maiolino, R., Marconi, A., & Gnerucci, A. 2010, *MNRAS*, **408**, 2115
- Mason, C. A., Treu, T., Dijkstra, M., et al. 2018, *ApJ*, **856**, 2
- Mason, C. A., Fontana, A., Treu, T., et al. 2019, *MNRAS*, **485**, 3947
- Mason, C. A., Trenti, M., & Treu, T. 2023, *MNRAS*, **521**, 497
- Matthee, J., Sobral, D., Gronke, M., et al. 2018, *A&A*, **619**, A136
- Matthee, J., Mackenzie, R., Simcoe, R. A., et al. 2023, *ApJ*, **950**, 67
- McLeod, D. J., Donnan, C. T., McLure, R. J., et al. 2024, *MNRAS*, **527**, 5004
- McQuinn, M., Hernquist, L., Zalzarriaga, M., & Dutta, S. 2007, *MNRAS*, **381**, 75
- McQuinn, M., Lidz, A., Zalzarriaga, M., Hernquist, L., & Dutta, S. 2008, *MNRAS*, **388**, 1101
- Mesinger, A., & Haiman, Z. 2007, *ApJ*, **660**, 923
- Meurer, G. R., Heckman, T. M., & Calzetti, D. 1999, *ApJ*, **521**, 64
- Miralda-Escudé, J. 1998, *ApJ*, **501**, 15
- Mirocha, J., & Furlanetto, S. R. 2023, *MNRAS*, **519**, 843
- Naidu, R. P., Tacchella, S., Mason, C. A., et al. 2020, *ApJ*, **892**, 109
- Naidu, R. P., Oesch, P. A., Setton, D. J., et al. 2022, arXiv e-prints [arXiv:2208.02794]
- Nakajima, K., Ouchi, M., Isobe, Y., et al. 2023, arXiv e-prints [arXiv:2301.12825]
- Nakane, M., Ouchi, M., Nakajima, K., et al. 2023, arXiv e-prints [arXiv:2312.06804]
- Napolitano, L., Pentericci, L., Calabrò, A., et al. 2023, *A&A*, **677**, A138
- Napolitano, L., Pentericci, L., Santini, P., et al. 2024, *A&A*, **688**, A106
- Newville, M., Stensitzki, T., Allen, D. B., & Ingarola, A. 2014, <https://doi.org/10.5281/zenodo.11813>
- Ning, Y., Cai, Z., Lin, X., et al. 2024, *ApJ*, **963**, L38
- Noterdaeme, P., Petitjean, P., Carithers, W. C., et al. 2012, *A&A*, **547**, L1
- Ocvirk, P., Lewis, J. S. W., Gillet, N., et al. 2021, *MNRAS*, **507**, 6108
- Osterbrock, D. E., & Ferland, G. J. 2006, *Astrophysics of Gaseous Nebulae and Active Galactic Nuclei* (CA: University Science Books)
- Ouchi, M., Shimasaku, K., Furusawa, H., et al. 2010, *ApJ*, **723**, 869
- Pacucci, F., Dayal, P., Harikane, Y., Inoue, A. K., & Loeb, A. 2022, *MNRAS*, **514**, L6
- Pahl, A. J., Shapley, A. E., Steidel, C. C., et al. 2024, arXiv e-prints [arXiv:2401.09526]
- Pallottini, A., Ferrara, A., Gallerani, S., et al. 2022, *MNRAS*, **513**, 5621
- Péroux, C., & Howk, J. C. 2020, *ARA&A*, **58**, 363
- Perrin, M. D., Sivaramakrishnan, A., Lajoie, C.-P., et al. 2014, *Proc. SPIE*, **9143**, 91433X
- Pettini, M., Steidel, C. C., Adelberger, K. L., Dickinson, M., & Gialvalisco, M. 2000, *ApJ*, **528**, 96
- Planck Collaboration VI. 2020, *A&A*, **641**, A6
- Prochaska, J. X., & Wolfe, A. M. 2009, *ApJ*, **696**, 1543
- Prochaska, J. X., Chen, H.-W., Dessauges-Zavadsky, M., & Bloom, J. S. 2007, *ApJ*, **666**, 267
- Puchwein, E., Bolton, J. S., Keating, L. C., et al. 2023, *MNRAS*, **519**, 6162
- Rasmussen Cueto, E. R., Hutter, A., Dayal, P., et al. 2024, *A&A*, **686**, A138
- Reddy, N. A., Oesch, P. A., Bouwens, R. J., et al. 2018a, *ApJ*, **853**, 56
- Reddy, N. A., Shapley, A. E., Sanders, R. L., et al. 2018b, *ApJ*, **869**, 92
- Rhoads, J. E., Wold, I. G. B., Harish, S., et al. 2023, *ApJ*, **942**, L14
- Rigby, J., Perrin, M., McElwain, M., et al. 2023, *PASP*, **135**, 048001
- Roberts-Borsani, G., Treu, T., Chen, W., et al. 2023, *Nature*, **618**, 480
- Roberts-Borsani, G., Treu, T., Shapley, A., et al. 2024, arXiv e-prints, [arXiv:2403.07103]
- Robertson, B. E. 2022, *ARA&A*, **60**, 121
- Robertson, B. E., Furlanetto, S. R., Schneider, E., et al. 2013, *ApJ*, **768**, 71
- Rosdahl, J., Blaizot, J., Katz, H., et al. 2022, *MNRAS*, **515**, 2386
- Salmon, B., Papovich, C., Finkelstein, S. L., et al. 2015, *ApJ*, **799**, 183
- Sanders, R. L., Shapley, A. E., Jones, T., et al. 2021, *ApJ*, **914**, 19
- Sanders, R. L., Shapley, A. E., Topping, M. W., Reddy, N. A., & Brammer, G. B. 2024, *ApJ*, **962**, 24
- Sandles, L., Curtis-Lake, E., Charlot, S., Chevallard, J., & Maiolino, R. 2022, *MNRAS*, **515**, 2951
- Saxena, A., Robertson, B. E., Bunker, A. J., et al. 2023, *A&A*, **678**, A68
- Saxena, A., Bunker, A. J., Jones, G. C., et al. 2024, *A&A*, **684**, A84
- Schaerer, D. 2003, *A&A*, **397**, 527
- Schaerer, D., Marques-Chaves, R., Barrufet, L., et al. 2022, *A&A*, **665**, L4
- Schaerer, D., Marques-Chaves, R., Xiao, M., & Korber, D. 2024, *A&A*, **687**, L11
- Schaye, J., Dalla Vecchia, C., Booth, C. M., et al. 2010, *MNRAS*, **402**, 1536
- Serjeant, S., & Baxx, T. J. L. C. 2023, *Nat. Astron.*, **7**, 1143
- Shapley, A. E., Steidel, C. C., Pettini, M., & Adelberger, K. L. 2003, *ApJ*, **588**, 65
- Shapley, A. E., Sanders, R. L., Reddy, N. A., Topping, M. W., & Brammer, G. B. 2023, *ApJ*, **954**, 157
- Shen, X., Vogelsberger, M., Boylan-Kolchin, M., Tacchella, S., & Kannan, R. 2023, *MNRAS*, **525**, 3254
- Simmonds, C., Tacchella, S., Maseda, M., et al. 2023, *MNRAS*, **523**, 5468
- Sobacchi, E., & Mesinger, A. 2015, *MNRAS*, **453**, 1843
- Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, *ApJS*, **214**, 15
- Stark, D. P. 2016, *ARA&A*, **54**, 761
- Steidel, C. C., Erb, D. K., Shapley, A. E., et al. 2010, *ApJ*, **717**, 289
- Steidel, C. C., Strom, A. L., Pettini, M., et al. 2016, *ApJ*, **826**, 159
- Steidel, C. C., Bogosavljević, M., Shapley, A. E., et al. 2018, *ApJ*, **869**, 123
- Steinhardt, C. L., Kokorev, V., Rusakov, V., Garcia, E., & Sneppen, A. 2023, *ApJ*, **951**, L40
- Sun, G., Faucher-Giguère, C.-A., Hayward, C. C., et al. 2023, *ApJ*, **955**, L35
- Tang, M., Stark, D. P., Chen, Z., et al. 2023, *MNRAS*, **526**, 1657
- Tanvir, N. R., Fynbo, J. P. U., de Ugarte Postigo, A., et al. 2019, *MNRAS*, **483**, 5380
- Taylor, A. J., Barger, A. J., & Cowie, L. L. 2022, *ApJ*, **939**, L3
- Tepper-García, T. 2006, *MNRAS*, **369**, 2025
- Terp, C., Heintz, K. E., Watson, D., et al. 2024, *A&A*, **690**, A70
- Thorne, J. E., Robotham, A. S. G., Davies, L. J. M., et al. 2021, *MNRAS*, **505**, 540
- Topping, M. W., Stark, D. P., Endsley, R., et al. 2023, arXiv e-prints [arXiv:2307.08835]
- Torrey, P., Vogelsberger, M., Hernquist, L., et al. 2018, *MNRAS*, **477**, L16
- Totani, T., Kawai, N., Kosugi, G., et al. 2006, *PASJ*, **58**, 485
- Treu, T., Roberts-Borsani, G., Bradac, M., et al. 2022, *ApJ*, **935**, 110
- Trinca, A., Schneider, R., Valiante, R., et al. 2023, arXiv e-prints [arXiv:2305.04944]
- Trump, J. R., Arrabal Haro, P., Simons, R. C., et al. 2023, *ApJ*, **945**, 35
- Tumlinson, J., Peebles, M. S., & Werk, J. K. 2017, *ARA&A*, **55**, 389
- Umeda, H., Ouchi, M., Nakajima, K., et al. 2023, arXiv e-prints [arXiv:2306.00487]
- Valentino, F., Brammer, G., Gould, K. M. L., et al. 2023, *ApJ*, **947**, 20
- van der Wel, A., Franx, M., van Dokkum, P. G., et al. 2014, *ApJ*, **788**, 28
- Verhamme, A., Orlitová, I., Schaerer, D., & Hayes, M. 2015, *A&A*, **578**, A7
- Verhamme, A., Orlitová, I., Schaerer, D., et al. 2017, *A&A*, **597**, A13
- Vijayan, A. P., Thomas, P. A., Lovell, C. C., et al. 2024, *MNRAS*, **527**, 7337
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *Nature Methods*, **17**, 261
- Vizgan, D., Heintz, K. E., Greve, T. R., et al. 2022, *ApJ*, **939**, L1
- Walter, F., Carilli, C., Neeleman, M., et al. 2020, *ApJ*, **902**, 111
- Wang, F., Yang, J., Fan, X., et al. 2021, *ApJ*, **907**, L1
- Wang, B., Fujimoto, S., Labbe, I., et al. 2023a, arXiv e-prints [arXiv:2308.03745]
- Wang, B., Leja, J., Labbé, I., et al. 2023b, arXiv e-prints [arXiv:2310.01276]
- Ward, E. M., de la Vega, A., Mobasher, B., et al. 2023, arXiv e-prints [arXiv:2311.02162]
- Whitaker, K. E., van Dokkum, P. G., Brammer, G., & Franx, M. 2012, *ApJ*, **754**, L29
- White, S. D. M., & Frenk, C. S. 1991, *ApJ*, **379**, 52
- White, S. D. M., & Rees, M. J. 1978, *MNRAS*, **183**, 341
- Wilson, S. N., Heintz, K. E., Jakobsson, P., et al. 2024, *A&A*, **685**, A30
- Witstok, J., Smit, R., Saxena, A., et al. 2024a, *A&A*, **682**, A40
- Witstok, J., Jakobsen, P., Maiolino, R., et al. 2024b, arXiv e-prints [arXiv:2408.16608]
- Witten, C., Laporte, N., Martin-Alvarez, S., et al. 2024, *Nat. Astron.*, **8**, 384
- Wolfe, A. M., Turnshek, D. A., Smith, H. E., & Cohen, R. D. 1986, *ApJS*, **61**, 249
- Wolfe, A. M., Gawiser, E., & Prochaska, J. X. 2005, *ARA&A*, **43**, 861
- Woosley, S. E., & Weaver, T. A. 1995, *ApJS*, **101**, 181
- Xu, X., Henry, A., Heckman, T., et al. 2022, *ApJ*, **933**, 202
- Yang, H., Malhotra, S., Gronke, M., et al. 2017, *ApJ*, **844**, 171
- Yang, J., Wang, F., Fan, X., et al. 2020, *ApJ*, **904**, 26
- Yeh, J. Y. C., Smith, A., Kannan, R., et al. 2023, *MNRAS*, **520**, 2757
- Zackrisson, E., Rydberg, C.-E., Schaerer, D., Östlin, G., & Tuli, M. 2011, *ApJ*, **740**, 13
- Zavala, J. A., Castellano, M., Akins, H. B., et al. 2024, arXiv e-prints [arXiv:2403.10491]

- 
- <sup>1</sup> Cosmic Dawn Center (DAWN), Denmark
- <sup>2</sup> Niels Bohr Institute, University of Copenhagen, Jagtvej 128, 2200 Copenhagen N, Denmark
- <sup>3</sup> Department of Astronomy, University of Geneva, Chemin Pegasi 51, 1290 Versoix, Switzerland
- <sup>4</sup> Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh EH9 3HJ, UK
- <sup>5</sup> Stockholm University, Department of Astronomy and Oskar Klein Centre for Cosmoparticle Physics, AlbaNova University Centre, SE-10691 Stockholm, Sweden
- <sup>6</sup> Department of Physics and Astronomy, The Johns Hopkins University, 3400 N Charles St., Baltimore, MD 21218, USA
- <sup>7</sup> Space Telescope Science Institute (STScI), 3700 San Martin Drive, Baltimore, MD 21218, USA
- <sup>8</sup> Kapteyn Astronomical Institute, University of Groningen, 9700, AV Groningen, The Netherlands
- <sup>9</sup> Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126 Pisa, Italy
- <sup>10</sup> New Mexico State University, Las Cruces 88003, NM, USA
- <sup>11</sup> Department of Astronomy, University of Massachusetts Amherst, Amherst, MA 01002, United States
- <sup>12</sup> DTU Space, Technical University of Denmark, Elektrovej, Building 328, 2800 Kgs. Lyngby, Denmark
- <sup>13</sup> Center for Relativistic Astrophysics, School of Physics, Georgia Institute of Technology, 837 State Street, Atlanta, GA 30332, USA
- <sup>14</sup> Centre for Astrophysics and Cosmology, Science Institute, University of Iceland, Dunhagi 5, 107 Reykjavik, Iceland
- <sup>15</sup> Instituto de Estudios Astrofísicos, Facultad de Ingeniería y Ciencias, Universidad Diego Portales, Av. Ejército 441, Santiago 8370191, Chile
- <sup>16</sup> Department of Physics and Astronomy, University of California, Riverside, 900 University Avenue, Riverside, CA 92521, USA
- <sup>17</sup> Institute of Science and Technology Austria (ISTA), Am Campus 1, 3400 Klosterneuburg, Austria
- <sup>18</sup> MIT Kavli Institute for Astrophysics and Space Research, 77 Massachusetts Ave., Cambridge, MA 02139, USA
- <sup>19</sup> Department of Astronomy, University of Florida, 211 Bryant Space Sciences Center, Gainesville, FL 32611, USA
- <sup>20</sup> Astrophysics Research Institute, Liverpool John Moores University, Liverpool L35 UG, UK
- <sup>21</sup> School of Physics and Astronomy, University of Leicester, University Road, Leicester LE1 7RH, UK
- <sup>22</sup> European Southern Observatory, Karl-Schwarzschild-Str. 2, 85748 Garching, Germany
- <sup>23</sup> Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
- <sup>24</sup> Cavendish Laboratory, University of Cambridge, 19 JJ Thomson Avenue, Cambridge CB3 0HE, UK