

The ionizing properties of JWST's compact broad-line emitters

Sara Mascia¹*, Jorryt Matthee¹, Alberto Torralba¹, Jenny E. Greene², Anna-Christina Eilers^{3,4}, Edoardo Iani¹, and Rohan P. Naidu^{4,5}

¹ Institute of Science and Technology Austria (ISTA), Am Campus 1, A-3400 Klosterneuburg, Austria

² Department of Astrophysical Sciences, Princeton University, 4 Ivy Lane, Princeton, NJ08544, USA

³ Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

⁴ MIT Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

⁵ Institute for Astronomy, University of Hawai'i, 2680 Woodlawn Drive, Honolulu, HI 96822, USA

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

The recent JWST discovery of a numerous population of broad-line emitters (BLEs) at $z > 4$ has reopened the question of whether sources other than faint galaxies contributed significantly to cosmic reionization. To assess this contribution, we present a systematic census of compact, blue broad-line emitters at $4 \leq z \leq 7$ selected from the Dawn JWST Archive (DJA), designed to isolate sources in which the UV continuum and the broad-line emission originate from the same compact physical region. Starting from a parent sample of 4145 galaxies with NIRSpec/PRISM spectroscopy and NIRCам imaging, we apply criteria on broad H α emission (FWHM $> 2000 \text{ km s}^{-1}$ detected at $S/N \geq 5$, complete at $L_{\text{H}\alpha, \text{broad}} \gtrsim 10^{42.7} \text{ erg s}^{-1}$), blue continuum slopes ($\beta_{\text{UV}} \leq -1.5$, $\beta_{\text{opt}} < 0.5$), and morphological compactness in both the rest-frame UV and optical, returning a final sample of 20 sources. This constitutes $\sim 20\%$ of the PRISM-selected broad-line sample that is dominated by Little Red Dots (LRDs). We find that the compact blue BLEs display hot ionizing continua, high-ionization UV lines (C IV, N IV], He II), and elevated Ly α emitter fractions ($X_{\text{Ly}\alpha} = 50\text{--}67\%$) relative to star-forming galaxies and LRDs. Beyond their similar lack of X-ray emission, their optical spectra closely resemble those of LRDs. This suggests that compact blue BLEs extend this population towards lower column density envelopes, but with similar engines, rather than representing a physically distinct population. Using SIROCCO to model physical configurations that reproduce the observed spectra, we find a high ionizing photon production efficiency ($\log \xi_{\text{ion}}/\text{Hz erg}^{-1} \sim 25.4$) and a weighted average escape fraction $f_{\text{esc}} \sim 0.2$ for the bluest sources, and ~ 0.01 for LRDs. This **implies that broad-line sources do not dominate reionization globally**, but given their luminosity, their contribution can dominate up to few Mpc scales.

Key words. galaxies: active – galaxies: high-redshift – galaxies: evolution – galaxies: nuclei – dark ages, reionization

1. Introduction

The epoch of reionization (EoR) marks the last major phase transition of baryonic matter in the Universe, during which the neutral hydrogen of the intergalactic medium (IGM) was ionized by early galaxies (for a recent review, see Stark et al. 2026). The timeline of reionization is relatively well constrained: hydrogen reionization is largely complete by $z \sim 5.3\text{--}5.5$ as inferred from the statistics of the Ly α forest measured in quasars (e.g., Bosman et al. 2022) and the optical depth to Thomson scattering measured in the cosmic microwave background constrains the average redshift at $z \approx 7.7$ (Planck Collaboration et al. 2020).

The canonical picture is that faint star-forming galaxies (SFGs) are the primary contributors to the reionization photon budget, based on the fact that their comoving number density is much higher than that of active galactic nuclei (AGN) during the EoR, and that even modest Lyman Continuum (LyC) escape fractions ($f_{\text{esc}}^{\text{LyC}} \approx 10\%$) are sufficient to sustain reionization when integrated over the faint end of the UV luminosity function (e.g., Robertson et al. 2015; Finkelstein et al. 2019; Atek et al. 2026; Simmonds et al. 2024; Giovannizzo et al. 2026). Observational support for this picture has grown substantially in recent years, with direct detections of LyC emission from $z \sim 3\text{--}4$ and $z \sim 0.3$ SFGs (e.g., Rivera-Thorsen et al. 2017; Steidel et al. 2018; Izotov et al. 2018; Flury et al. 2022; Gupta et al. 2024)

that have enabled estimates of $f_{\text{esc}}^{\text{LyC}}$ of relatively bright galaxies in the EoR ($M_{\text{UV}} \leq -18$; e.g., Mascia et al. 2024) based on indirect indicators that are in line with the required values. Complementary evidence has emerged from cross-correlation measurements between galaxies and the Ly α forest transmission at the closing of reionization, which provide a statistical, IGM-averaged estimate of the LyC emissivity of the galaxy population as a whole (Kashino et al. 2023, 2026; Kakiichi et al. 2025). Nevertheless, the averaged escape fraction of the full galaxy population remains uncertain because virtually no direct constraints exist on the $f_{\text{esc}}^{\text{LyC}}$ of galaxies fainter than $M_{\text{UV}} > -19$. These account for the vast majority of the galaxy population at these redshifts, contributing an estimated $\gtrsim 75\%$ of the total UV luminosity density at $z \sim 6$ (Atek et al. 2026), and their collective ionizing output is therefore the most poorly anchored component of any reionization budget. Moreover, the role of rare, bright starbursts with potential significant leakage (e.g., Marques-Chaves et al. 2026a) remains unclear (e.g., Sharma et al. 2017; Naidu et al. 2018; Matthee et al. 2022).

Recent JWST observations have added a new layer of complexity to this picture. The detection of Ly α emission at $z > 10$ in several sources (Witstok et al. 2025; Bunker et al. 2023) and the apparent weak damping wing in MoM-z14 at $z = 14.4$ (Naidu et al. 2025) are difficult to reconcile with the high neutral hydrogen fraction expected at these redshifts, pointing to local ionization by exceptionally luminous sources. Do such rare,

* Corresponding author: sara.mascia@ista.ac.at

compact sources that may be capable of carving out large ionized bubbles in an otherwise predominantly neutral IGM play a disproportionate role in reionization?

AGN have been considered subdominant contributors to reionization primarily because the space density of quasars declines sharply above $z \sim 3$ (Hopkins et al. 2007; Kulkarni et al. 2019). However, AGN may play a more significant role than previously assumed. A faint AGN population had been proposed as a significant driver of reionization on the basis of pre-JWST X-ray and UV constraints (Madau & Haardt 2015; Giallongo et al. 2019), but observational support for such a population remained limited. Deep spectroscopic surveys with JWST have revealed a numerous population of moderate luminosity, broad-line emitters at $4 \lesssim z \lesssim 9$ ($M_{\text{UV}} \approx -18$ to -21 ; e.g., Kocevski et al. 2023; Harikane et al. 2023; Greene et al. 2024; Matthee et al. 2024; Maiolino et al. 2024a; Juodžbalis et al. 2025), accounting for $\sim 1\%$ among UV-selected galaxies at $z \sim 4 - 7$. These number densities have prompted a renewed reassessment of AGN-dominated reionization scenarios. In particular, Madau et al. (2024) showed that, under certain assumptions on the AGN fraction, escape fraction (Cristiani et al. 2016; Grazian et al. 2018; Romano et al. 2019), and UV spectral slope, a population of faint Type-1 AGN can reproduce both the hydrogen and helium reionization history and the Thomson scattering optical depth.

However, a key challenge is that the broad-line emitter population identified by JWST is far from homogeneous. A substantial fraction consists of the so-called *Little Red Dots* (LRDs; Matthee et al. 2024), characterized by blue UV and red optical colors, compact optical morphology and broad Balmer lines (Kocevski et al. 2023; Greene et al. 2024; Kokorev et al. 2024; Matthee et al. 2024; Labbe et al. 2025). These are interpreted as AGN highly covered by dense gas (e.g., Torralba et al. 2026b; Naidu et al. 2025; Inayoshi et al. 2025; de Graaff et al. 2025b) from which few UV photons escape (but see Tang et al. 2026; Ji et al. 2026; Torralba et al. 2026a, for indications of holes in the envelope through which some UV emission may escape). Potentially more relevant for reionization is that a fraction of broad-line emitters display blue UV continua and no Balmer absorption characteristic of a dense, high-column gas envelope (Übler et al. 2023; Brazzini et al. 2026; Geris et al. 2026; Matthee et al. 2026). However, the fraction of the galaxy population that comprises such sources has not been systematically explored yet. Whether the newly identified broad-line source contributed to reionization depends on their $f_{\text{esc}}^{\text{LyC}}$: direct measurements are precluded at $z \gtrsim 4$ by IGM absorption (Inoue et al. 2014), and indirect constraints from analogous sources at lower redshift, where the LyC is observable, do not yet exist either.

An alternative approach to measuring $f_{\text{esc}}^{\text{LyC}}$ consists of modeling the observed spectra of these sources under assumptions on their geometry and gas distribution (Sneppen et al. 2026b; Gentile et al. 2026). Such models have reproduced key observables of LRDs including the strength of the Balmer break and the $H\alpha$ line profile. Yet, no study has used this framework to estimate the ionizing output from these sources.

In this paper, we take a step toward filling this gap by presenting a systematic study of blue, compact broad-line emitters at $4 \leq z \leq 7$ drawn from the Dawn JWST Archive (DJA; Heintz et al. 2024; de Graaff et al. 2025a), selected to have simultaneous coverage of $\text{Ly}\alpha$ and $H\alpha$ with JWST/NIRSpec. Our analysis addresses four key questions: *i.* What fraction of sources exhibit bright, blue UV emission that is predominantly powered by the same region as the broad lines? *ii.* What are the UV continuum and emission line properties of the blue, compact broad-

line emitters as a population? *iii.* What spectral diagnostics provide the most effective constraints on the nature and hardness of the ionizing continuum, and can AGN accretion and stellar processes be distinguished using the available data? *iv.* What is the contribution of this population to the reionization photon budget and how sensitive is this estimate to the assumed powering mechanism, surrounding gas geometry, and structure?

The paper is structured as follows. We describe the data and observational setup in Sect. 2. The sample selection criteria are detailed in Sect. 3. The spectral characterization of the sample is presented in Sect. 4. We present photoionization modeling with SIRIO in Sect. 5. The contribution to reionization is quantified in Sect. 6, where we also discuss the implications of our results. We summarize our conclusions in Sect. 7. Throughout this paper, we assume a flat Λ CDM cosmology with $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_M = 0.3$, and $\Omega_\Lambda = 0.7$.

2. Data

2.1. NIRSpec

Our spectroscopic analysis is based on spectra obtained with the NIRSpec multi-shutter assembly (MSA) on JWST. We use observations compiled in version 4.5 of the DJA¹. This dataset includes major public surveys such as CANUCS (Sarrouh et al. 2025), CAPERS (GO-6368; PI: Dickinson), CEERS (Finkelstein et al. 2025), JADES (Eisenstein et al. 2023; Curtis-Lake et al. 2025; Scholtz et al. 2025), NEXUS (Shen et al. 2024), NIRSpec GTO-Wide (Maseda et al. 2024), RUBIES (de Graaff et al. 2025a), UNCOVER (Bezanson et al. 2024; Price et al. 2025), and GO-8204 (PIs Greene and Labbé) and several smaller programs. We also included observations from NIRSpec IFU programs, GA-NIFS (GO-1216, PI Lützgendorf), and GO-5664 (PI: Matthee). All spectra were reduced uniformly using the msaexp software package (Brammer 2023b), implementing techniques consistent with recent analysis of deep NIRSpec datasets (Heintz et al. 2024; de Graaff et al. 2025a). Where feasible, we employed local background subtraction utilizing nodded exposure sequences to minimize contamination from sky emission and neighboring sources.

Version 4 of the DJA implements empirical wavelength calibration corrections based on source centroid positions within MSA shutters, addressing systematic offsets between PRISM and higher-resolution modes (de Graaff et al. 2025b), and incorporates updated flux calibration reference files that extend extraction beyond nominal wavelength ranges (Valentino et al. 2025; Pollock et al. 2025).

1D spectra were extracted using an optimal extraction algorithm (Horne 1986), which weights pixels according to the light profile of the object in the spectrum to maximize the signal to noise. The resulting extraction kernel, combined with precise source positioning within each shutter, enabled wavelength-dependent slit loss corrections which, for point-like sources, have been demonstrated to perform reliably (e.g., de Graaff et al. 2025a; Hviding et al. 2025) and were therefore applied without additional flux adjustments. All reduced spectra underwent comprehensive visual inspection to verify spectral quality and validate automated redshift determinations (see de Graaff et al. 2025b, for further details). We selected only sources with grade = 3 redshifts (robust redshifts, de Graaff et al. 2025b).

For sources observed in gravitationally lensed fields, we applied magnification corrections to all derived luminosity measurements. Specifically, we adopted the Abell 2744 lens model

¹ <https://dawn-cph.github.io/dja/>

from Furtak et al. (2023) and Price et al. (2025) for UNCOVER sources, while magnification estimates in CANUCS cluster fields were released in Sarrouh et al. (2025).

2.2. NIRCam

Deep, high resolution imaging data in the near-infrared were obtained from JWST/NIRCam mosaics compiled in version 7 of the DJA, constructed at a pixel scale of $0.04''$. Reductions were performed using the *grizli* pipeline (Brammer 2023a), with detailed processing methodology documented in Valentino et al. (2023). For this work, we focus on both rest-frame UV and rest-frame optical morphological properties of our sources. We extracted imaging data in all available NIRCam filters from the aforementioned JWST programs. This includes the following filters where available: F070W, F090W, F115W, F150W, F200W, F277W, F356W, and F444W.

We note that temporal offsets between spectroscopic observations and subsequent NIRCam imaging campaigns can introduce minor astrometric discrepancies. This is particularly relevant for early NIRSpec programs that relied on HST-based target selection prior to the availability of JWST imaging. To mitigate potential systematic offsets, we cross-matched our spectroscopic catalog against comprehensive photometric catalogs maintained on the DJA platform, obtaining refined source centroid positions based on NIRCam detections. Photometric measurements were then extracted from image cutouts centered on these cross-matched positions using circular apertures with radii of $0.1''$ and $0.2''$, consistent with standard approaches for compact source photometry (e.g. Kokorev et al. 2024; Labbe et al. 2025).

3. Sample selection

Our parent sample consists of 4145 sources drawn from the DJA spectroscopic compilation described in Sect. 2, selected to have simultaneous rest-frame spectral coverage of $\text{Ly}\alpha$ to $\text{H}\alpha$ at $4 \leq z \leq 7$. To isolate compact, blue sources with a significant broad $\text{H}\alpha$ emission component, we apply three sequential selection criteria which we describe in detail below. The number of sources retained at each step is summarized in Table 1.

3.1. Broad $\text{H}\alpha$ emission

The first step is to select sources with strong, broad $\text{H}\alpha$ emission, regardless of their broad-band colors. To maximize our statistics we use PRISM spectra to identify these broad lines, which implies that our sample is limited to sources with significant contribution from broad lines.

We fit the $\text{H}\alpha$ emission line of each source in the parent sample with two competing models: a single Gaussian plus continuum, and a two-component narrow-plus-broad Gaussians plus continuum. The continuum is in both cases modeled as a power-law fitted locally over two sidebands on either side of $\text{H}\alpha$, over rest-frame windows of $6150\text{--}6450\text{ \AA}$ and $6650\text{--}6750\text{ \AA}$. Both models are fitted to the PRISM spectra using a Levenberg–Marquardt least-squares minimization. The instrumental line spread function of NIRSpec/PRISM is approximated as a wavelength-dependent Gaussian with σ_{LSF} computed from the point-source resolution curve, and convolved in quadrature with the intrinsic line profile. Parameter uncertainties are propagated via Monte Carlo sampling of the best-fit parameter covariance matrix, drawing $N_{\text{MC}} = 1000$ realizations per source.

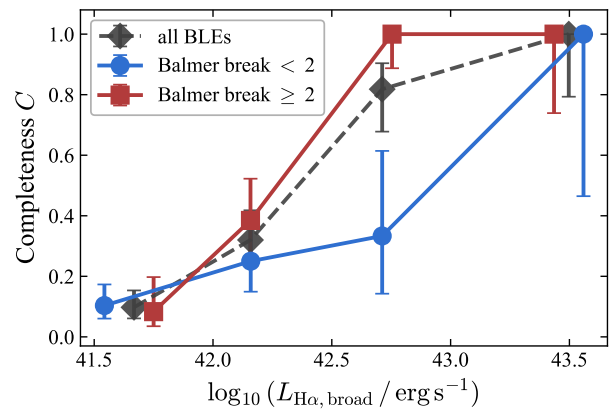


Fig. 1: The completeness of detecting broad lines in PRISM data as a function of broad- $\text{H}\alpha$ luminosity $L_{\text{H}\alpha,\text{broad}}$. The recovered fraction of grating-confirmed broad-line emitters is shown for all sources (black, dashed), for the population with weak (< 2) Balmer breaks (blue circles), and for the sources with Balmer breaks ≥ 2 (red squares). Points mark the median luminosity of each bin; error bars are 68% binomial confidence intervals.

Broad-line emitters (BLEs) are identified following the three sequential criteria of Greene et al. (2024). First, the improvement in χ^2 between the two models must satisfy $\Delta\chi^2 > 11.5$, (C1), corresponding to the 3σ significance threshold for four additional degrees of freedom. Second, the FWHM of the broad component must exceed 2000 km s^{-1} (C2). Third, the signal-to-noise (S/N) ratio of the broad component, must satisfy $S/N_b \geq 5$ (C3). Sources satisfying C1 and C2 but failing C3 are flagged as candidate BLEs and excluded from subsequent analysis. Only sources satisfying all three criteria simultaneously are retained as confirmed broad-line emitters. All the PRISM-selected sources with usable grating observations are confirmed BLEs (100% purity), i.e. broad $\text{H}\alpha$ is detected with $S/N \geq 3$ in both dispersers. However, our PRISM selection recovers only the 30% of the grating-selected BLEs. The incompleteness depends strongly on both the $\text{H}\alpha$ luminosity and line width (Fig. 1): the recovery rate of grating-confirmed BLEs in the PRISM rises from 11% at $L_{\text{H}\alpha,\text{broad}} < 10^{42}\text{ erg s}^{-1}$ to 100% above $10^{43}\text{ erg s}^{-1}$, and because the PRISM resolution is redshift-dependent, from 18% at $z < 5$ to 67% at $z > 6$. These trends are not independent: more luminous broad- $\text{H}\alpha$ lines are also intrinsically broader, and both properties make the broad component easier to detect at PRISM resolution. The sources we preferentially recover are therefore the luminous, broad-lined emitters, which are also those with the strongest Balmer breaks (Fig. 1). Overall, our selection reaches 50% completeness at $L_{\text{H}\alpha,\text{broad}} \approx 10^{42.3}\text{ erg s}^{-1}$ and 90% completeness at $\approx 10^{42.7}\text{ erg s}^{-1}$, and is substantially incomplete below these luminosities, where we miss the majority of the fainter broad- $\text{H}\alpha$ population.

3.2. UV and optical continuum slopes

The UV and optical continuum slopes are measured by fitting a power law $F_\lambda \propto \lambda^\beta$ to the rest-frame PRISM spectra over two wavelength windows: $1300\text{--}3000\text{ \AA}$ for the UV slope β_{UV} , and $5000\text{--}6500\text{ \AA}$ for the optical slope β_{opt} . The regression is performed by weighted least-squares in log–log space, with known emission lines masked within $\pm 100\text{ \AA}$ of their rest-frame wavelengths ($\text{Ly}\alpha$, C iv, He ii, O iii], C iii], Mg ii in the UV window;

H β , [O III], H α , [N II] in the optical window; at PRISM resolution several of these are blended), leaving sufficient line-free continuum for a robust power-law fit in both windows. We verified that the resulting slopes are consistent with those derived independently from broadband NIRCcam photometry (F090W–F200W for β_{UV} ; F277W–F444W for β_{opt}) for sources where both measurements are available. We additionally measure the Balmer break strength as the ratio of the median flux density in the 5400–5700 Å window to that in the 3500–3650 Å window (Wang et al. 2025; de Graaff et al. 2025c).

Kocevski et al. (2025) select LRD candidates with $-2.8 < \beta_{UV} < -0.37$, where the lower limit excludes brown dwarf contaminants and the upper limit avoids dust-reddened systems (Brazzini et al. 2026; Geris et al. 2026). Our broad-H α selection (Sect. 3.1) already removes the possible contaminants from the sample, so the role of the β_{UV} cut is not to identify AGN but to isolate sources in which the UV continuum reaches the observer with minimal attenuation. We therefore adopt a more stringent threshold of $\beta_{UV} \leq -1.5$, motivated by the empirical observed slope of unreddened quasar composites (Vanden Berk et al. 2001): any source redder than $\beta_{UV} \sim -1.5$ is unlikely to be observed without significant dust attenuation along the line of sight, and its UV slope would no longer reliably trace the intrinsic continuum of the central source. Of the 99 confirmed broad-line emitters, 35 have $\beta_{UV} > -1.5$; these are predominantly LRDs (31), whose redder UV slopes reflect attenuation by the dense gas envelope, and a small number of classical broad-line AGN (4) reddened by dust along the line of sight. We further require $\beta_{opt} < 0.5$ to exclude the reddest sources in the optical. Based on a visual inspection of their spectra, sources above this threshold in our sample are either obscured quasars (2) or LRDs (25) not compact in the UV, therefore not representative of the compact blue BLE population we aim to characterize. The measured slopes and Balmer break strengths for the final sample are reported in Table 2. Fig. 2 shows the distribution of β_{UV} versus β_{opt} for the compact blue BLEs, the other BLE sample, and the SFG comparison sample.

3.3. Morphological compactness

Morphological compactness has been widely used to identify high-redshift AGN candidates and LRDs, with most studies relying on a single compactness criterion measured in the rest-frame optical, typically the F444W band (e.g., Kocevski et al. 2025; Hviding et al. 2025; Kokorev et al. 2024). However, a compact optical morphology alone does not guarantee equally compact rest-frame UV emission, which can be contaminated by an extended host galaxy and bias our study of the UV emission from the broad-line region (Chen et al. 2025; Rinaldi et al. 2025; Baggen et al. 2026; Ishikawa et al. 2026). To mitigate this contamination, require compactness in both the rest-frame UV and optical, ensuring that the blue continuum genuinely originates from a point-like source.

Source compactness is assessed independently in the rest-frame UV and rest-frame optical using the NIRCcam imaging described in Sect. 2.2. We quantify compactness via the ratio of fluxes enclosed within circular apertures of 0.2'' and 0.1'' radius centered on the source position, $C \equiv \frac{F_{0.2''}}{F_{0.1''}}$, measured on the NIRCcam mosaics using PHOTUTILS (Bradley et al. 2024). No PSF aperture correction is applied to the flux ratio: our aperture sums are extracted directly from the NIRCcam mosaic pixel data, so no de-correction step is needed before computing C . Sources are classified as compact if $C \leq 1.7$ in both the rest-frame UV and

Table 1: Sample selection.

Selection step	N
Parent sample ($4 \leq z \leq 7$)	4145
Confirmed broad-line emitters	99
Compact in rest-frame UV ($C_{UV} \leq 1.7$)	71
Compact in rest-frame optical ($C_{opt} \leq 1.7$)	65
Compact in UV and optical, $\beta_{UV} \leq -1.5, \beta_{opt} \leq 0.5$	20

optical. In the rest-frame UV, we adopt a filter fallback chain in order of preference: F115W, F150W, F200W, selecting the first available filter for each source. In the rest-frame optical we use F444W. We note that our compactness criterion is more tolerant than a pure point-source criterion to allow for closely separated clumps. In the case of clumpy systems, we verify through visual inspection that there is blue, UV bright, compact emission co-spatial with the broad Balmer line-emission.

Flux uncertainties are estimated from empty-aperture statistics: 200 random apertures of the same radius are placed in source-free regions of each image cutout, excluding a zone of radius 0.3'' around the source centroid, and the standard deviation of their flux is adopted as σ_{flux} . When the cutout geometry does not permit a sufficient number of valid sky placements, we fall back to a background RMS estimated from sigma-clipped border pixels, scaled by the square root of the aperture area.

3.4. Final sample

Fig. 2 shows the UV and optical colors of the 99 confirmed broad-line emitters, together with the star-forming galaxy (SFG) sample. Broad line sources are color-coded by their Balmer break strength, as this has been widely studied in the context of LRDs (e.g., Matthee et al. 2026; Sneppen et al. 2026b).

The diagram reveals a broad diversity of spectral types within the BLE population. Sources in the upper-right region ($\beta_{UV} \gtrsim -0.5, \beta_{opt} \gtrsim 2$) have red slopes across the UV to optical range, which is indicative of attenuation by dust. The majority of the 99 PRISM-selected BLEs ($\approx 75\%$) occupy an intermediate region with red optical slopes and moderate-to-strong Balmer breaks whereas their UV slopes are generally blue but probe a relatively wide range ($-2 \lesssim \beta_{UV} \lesssim -1$). By visual inspection and the change in continuum slope, we classify 74 of these sources as LRDs. A small subset of sources, classified as classical broad-line AGN, are highlighted in yellow. Among these, a single source stands out with a very blue optical slope: a luminous quasar observed through the ASPIRE program (GO 2078, PI Wang). The distribution of SFGs peaks around $\beta_{UV} \approx -2, \beta_{opt} \approx -2$, which highlights that the main difference between SFGs and broad-line sources is the optical continuum slope.

We designed the selection criteria $\beta_{UV} \leq -1.5$ and $\beta_{opt} < 0.5$ (dashed lines in Fig. 2) to isolate the 20 compact blue BLEs in the lower-left corner of the diagram, where sources are blue in both the UV and the optical and have uniformly weak Balmer breaks ($F_{\nu,red}/F_{\nu,blue} \lesssim 2$ for the majority). These cuts are motivated by the requirement of a hot ionizing radiation field that is uncontaminated by evolved stellar populations or dust reddening, and are complemented by the compactness criteria $C_{UV}, C_{opt} \leq 1.7$, which ensure that both the UV and optical continua originate from a single compact physical region rather than from an extended host. Our selection recovers most LRDs from Geris et al. (2026) but few blue broad-line sources, whose faint broad H α falls below the PRISM S/N threshold.

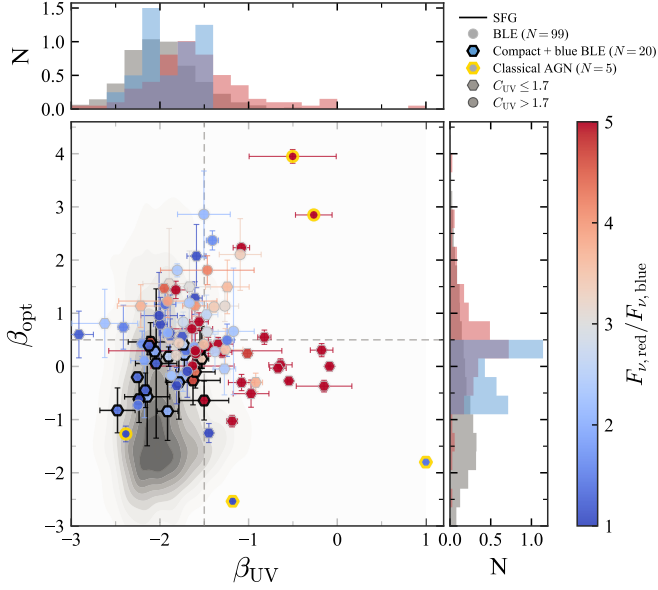


Fig. 2: Spectroscopic UV versus optical β slopes for the population of star-forming galaxies and the broad-line emitters identified in this work. *Main panel*: each BLE is color-coded by Balmer break strength $F_{\nu,\text{red}}/F_{\nu,\text{blue}}$, shown by the color bar on the right, and the marker shape encodes UV compactness: hexagons indicate sources with $C_{\text{UV}} \leq 1.7$, circles $C_{\text{UV}} > 1.7$. The 20 compact blue BLEs selected in this work are highlighted with black outlines; sources classified as classical broad-line AGN with yellow outlines. Grey contours show the distribution of DJA star-forming galaxies at $4 \leq z \leq 7$. Dashed lines mark the selection thresholds $\beta_{\text{UV}} = -1.5$ and $\beta_{\text{opt}} = 0.5$. *Top and right panels*: normalized distributions of β_{UV} and β_{opt} for star-forming galaxies (grey), BLEs (red), and compact blue BLEs (blue).

The final sample is likely a lower limit on the true number of compact blue broad-line emitters in the full BLE population, for two reasons. First, in Sect. 3.1 we discuss the incompleteness of our sample: PRISM broad-line selection recovers only the $\sim 30\%$ of the grating-selected broad-line sample, with the incompleteness rising steeply toward low $H\alpha$ luminosity, narrow line width, and low redshift. Blue broad-line emitters are, on average, fainter in broad $H\alpha$ and narrower-lined than LRDs, they are recovered less efficiently (Fig. 1), so the PRISM selection preferentially misses the blue population rather than the LRDs. Second, sources that pass the optical compactness criterion but fail the UV compactness cut could in principle host a genuine hot ionizing source whose UV emission is outshone or blended with extended host-galaxy star formation at the spatial resolution and depth of the available imaging. Disentangling the central ionizing source from the host UV emission in these cases would require higher spatial resolution observations capable of resolving the UV morphology on sub-kpc scales. Our selection therefore provides a clean, conservative sample in which the UV properties can be unambiguously attributed to the central source, at the cost of potentially missing sources where the host contribution to the UV is non-negligible.

We caution that any statement about LRDs dominating the broad-line population is luminosity-dependent. The PRISM broad-line selection is complete only for the most luminous emitters and, at fainter luminosities, more easily recovers the reddest sources (Sect. 3.1); the LRD fraction it returns is there-

fore representative only at the bright end and increasingly biased toward fainter luminosities. A robust characterization of the composition of the broad-line population below our completeness limit, will require a systematic analysis of the higher-resolution grating spectra, which can recover the fainter broad lines that the PRISM misses. The NIRCам images and the NIR-Spec PRISM spectra for the 20 sources in the compact, blue BLE sample are shown in Fig. 3 and Fig. A.1.

3.5. Example sources

To illustrate the diversity of the sample, we briefly describe a subset of sources that are individually visible in Fig. 3 and Fig. A.1, and that connect to results in the literature.

GN-16813 was first identified as a broad $H\alpha$ emitter by Matthee et al. (2024), who noted its compact morphology and broad Balmer emission. Subsequent IFU follow-up revealed a blue UV continuum with strong UV lines, unveiling its nature as a compact blue broad-line emitter rather than a classical LRD with a strong Balmer break. In the present sample, GN-16813 is the only source with medium-resolution grating spectroscopy covering the full rest-frame UV (from the DIVER program, GO-8018, PI Lin; see Sect. 4.4), making it, in terms of the UV, the best-characterized object in the sample.

GS_3073 was first studied in detail by Vanzella et al. (2010), who detected strong $\text{Ly}\alpha$ and $\text{N IV}] \lambda 1486$ emission in the rest-frame UV. It was subsequently identified as an AGN at $z \approx 5.55$ by Grazian et al. (2020), and by Übler et al. (2023) via JWST/NIRSpec IFU spectroscopy, based on broad hydrogen and helium emission and a large $\text{He II } \lambda 4686$ equivalent width ($\text{EW} \sim 20 \text{ \AA}$). Brazzini et al. (2026) identified it as the prototypical “Rosetta Stone” of the Little Blue Dot (LBD) class – the blue, compact counterpart to LRDs – finding no Balmer absorption, no time variability, and strong broad He II emission supporting their interpretation as BLR stratification and a relatively unobscured sight line to the central engine. In the present sample, its blue UV slope, negligible Balmer break, and low Balmer decrement ($F(H\alpha)/F(H\beta) = 3.21 \pm 0.05$) make it one of the least obscured sources and a natural anchor point for the population-level analysis in Sect. 4. We note that while GS_3073 shows a very clumpy morphology in the rest-frame UV NIRCам image, the dominant UV emission is co-spatial with the $H\alpha$ emission.

rubies-egs53_4233_50052 corresponds to CEERS-2782, analyzed by Kocevski et al. (2023) as one of the first spectroscopically confirmed broad-line AGN candidates at $z > 5$ in the CEERS field. With $\beta_{\text{UV}} = -2.16 \pm 0.06$ and detections of both $\text{N IV}]$ and C IV in the PRISM spectrum (Sect. 4.4), it is among the most UV-bright sources in the sample. Similar to many LRDs (Baggen et al. 2026), it has a compact, nearby blue clump.

abell2744-greene_8204_69688 is located behind the Abell 2744 cluster and is gravitationally lensed with a magnification of $\mu \sim 2$ (Furtak et al. 2023). It was selected as a broad line emitters from ALT (Naidu et al. 2024) and presented in Matthee et al. (2025) and observed with JWST/NIRSpec MSA as part of GO program 8204 (PIs Greene & Labbé), and is analyzed in Matthee et al. (2026). Its UV morphology appears clumpy, suggestive of ongoing interactions. Its PRISM spectrum reveals exceptionally strong $\text{N IV}]$ and $\text{N III}]$ emission.

Finally, we note that the selection boundary between compact blue BLEs and classical LRDs is not a sharp one, and a handful of sources in the sample occupy the transition region between the two populations. Sources such as **nexus-obs3_5105_10835**, **rubies-uds23_4233_172350**, and

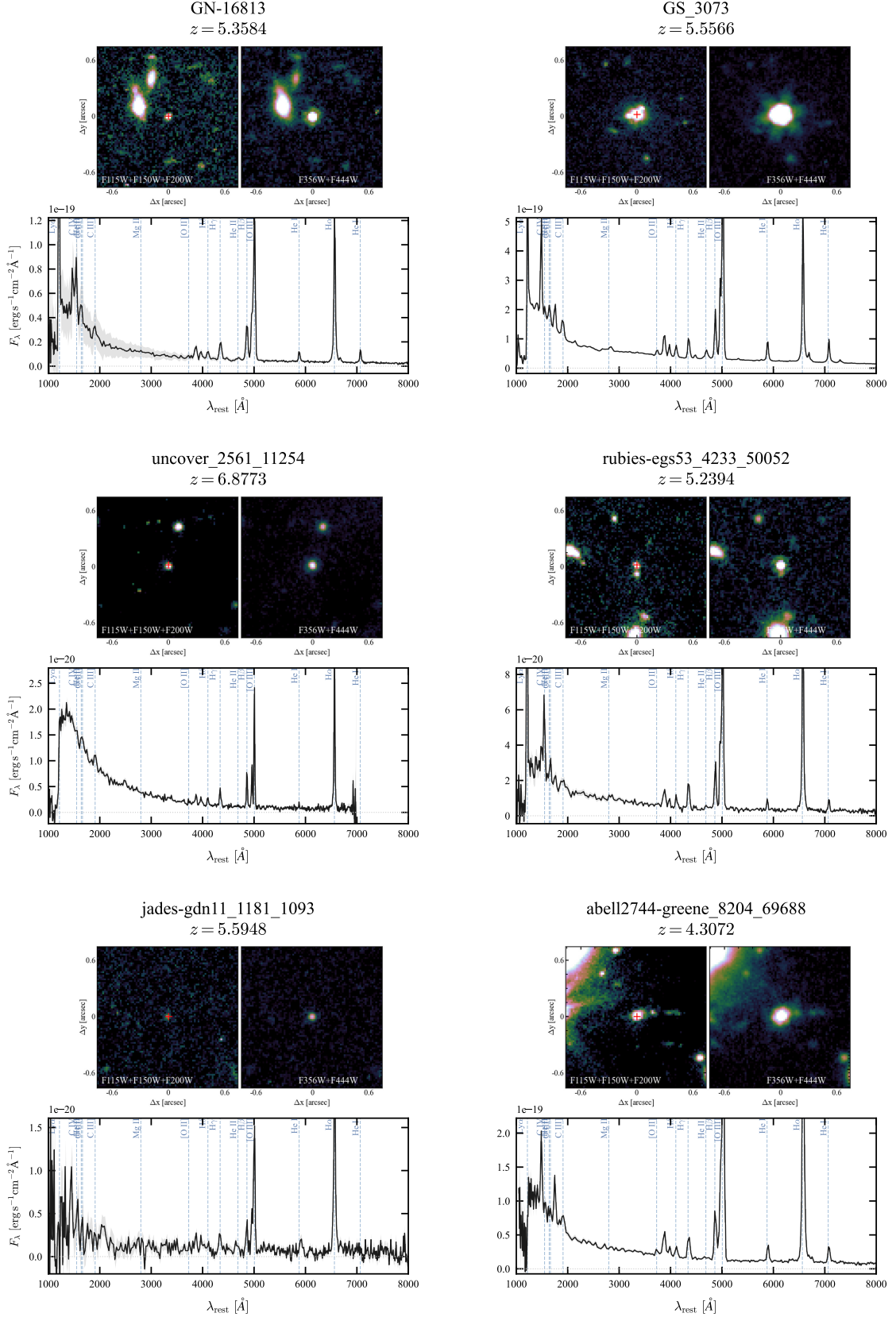


Fig. 3: NIRCam stamps and NIRSpect/PRISM spectra for 6 compact blue BLEs. For each source, the upper panel shows NIRCam composites in the rest-frame UV (F115W+F150W+F200W, left) and optical (F356W+F444W, right), with the optical centroid marked by a red cross on the UV stamp. The lower panel shows the PRISM spectrum (black) with the 1σ uncertainty band (grey). Vertical dashed lines mark the positions of prominent emission lines. Sources are ordered by increasing β_{UV} (blue to red from top to bottom), with the redshift indicated above each stamp. The remaining 14 sources are shown in Fig. A.1.

Table 2: Properties of the final sample of 20 compact, blue broad-line emitters.

Source	z	R.A. [deg]	Dec. [deg]	β_{UV}	β_{opt}	$F_{\nu,red}/F_{\nu,blue}$	$F(H\alpha)/F(H\beta)$	C_{opt}	C_{UV}
jades-gdn09_1181_73488	4.133	189.1974	62.1772	-1.73 ± 0.06	0.41 ± 0.15	2.60	9.71	1.63	1.38
abell2744-green_8204_69688	4.307	3.5649	-30.3482	-2.12 ± 0.04	0.39 ± 0.17	1.26	6.29	1.70	1.37
capers-egs47_6368_19300	4.543	215.0221	52.9208	-1.53 ± 0.08	0.14 ± 0.17	3.37	14.31	1.61	0.92
nexus-obs3_5105_10835	4.650	268.4377	65.1675	-1.63 ± 0.31	-0.26 ± 0.46	4.67	13.66	1.61	1.24
rubies-egs62_4233_42232	4.952	214.8868	52.8554	-1.50 ± 0.28	-0.64 ± 0.37	6.57	11.45	1.61	1.60
uncover_2561_38108	4.969	3.5300	-30.3580	-1.90 ± 0.08	0.18 ± 0.18	2.54	11.37	1.54	1.17
ceers-ddt_2750_1768	5.088	214.9258	52.9457	-2.11 ± 0.25	0.46 ± 0.36	4.53	11.84	1.60	0.90
rubies-egs53_4233_50052	5.239	214.8235	52.8303	-2.16 ± 0.06	-0.45 ± 0.33	1.29	4.64	1.63	1.38
capers-egs61_6368_20952	5.283	214.8802	52.8126	-1.72 ± 0.09	0.22 ± 0.29	2.20	9.83	1.63	1.53
GN-16813	5.358	189.1793	62.2925	-2.48 ± 0.20	-0.83 ± 0.42	1.41	3.46	1.58	1.12
GS_3073	5.557	53.0789	-27.8842	-2.25 ± 0.02	-0.20 ± 0.01	1.35	3.21	1.70	1.37
rubies-uds23_4233_172350	5.581	34.3690	-5.1039	-1.60 ± 0.22	-0.10 ± 0.30	4.35	11.20	1.57	1.06
jades-gdn11_1181_1093	5.595	189.1797	62.2246	-2.14 ± 0.25	-0.57 ± 0.92	2.14	7.49	1.58	1.19
macs1423_1208_4103248	5.775	215.9612	24.0580	-1.55 ± 0.20	0.32 ± 0.36	6.32	11.70	1.58	1.55
jades-gdn198_1181_38147	5.870	189.2707	62.1484	-1.53 ± 0.10	0.35 ± 0.39	1.64	7.53	1.68	1.33
gto-wide-uds12_1215_1259	5.962	34.4223	-5.2507	-1.91 ± 0.12	-0.84 ± 0.55	2.03	6.15	1.60	1.13
rubies-uds32_4233_966096	6.166	34.4667	-5.1009	-2.04 ± 0.23	0.05 ± 1.40	1.22	18.77	1.59	1.19
capers-cos07_6368_35805	6.528	150.0554	2.2916	-2.06 ± 0.03	0.28 ± 0.17	1.99	7.65	1.70	1.56
jades-gdn2_1181_38509	6.672	189.0915	62.2281	-1.79 ± 0.14	-0.30 ± 0.70	1.64	5.81	1.70	1.21
uncover_2561_11254	6.877	3.5804	-30.4050	-2.24 ± 0.03	-0.62 ± 0.44	0.94	4.25	1.70	1.14

Notes. $F_{\nu,red}/F_{\nu,blue}$, the Balmer break strength, is defined as the ratio of the median flux density at 5400–5700 Å to that at 3500–3650 Å. $C \equiv F_{0.2''}/F_{0.1''}$ is the compactness ratio measured in the rest-frame optical (F444W) and UV (F115W, F150W, or F200W; see Sect. 3.3). The Balmer decrement $F(H\alpha)/F(H\beta)$ is measured from the total line fluxes as described in Sect. 4.1.

macs1423_1208_4103248 display spectra that, while satisfying our selection criteria, show properties more reminiscent of LRDs: a redder UV continuum slope, a more pronounced Balmer break, and a higher Balmer decrement. As an illustrative example, **rubies-uds23_4233_172350** has been selected as LRD from [Hviding et al. \(2025\)](#) and shows $\beta_{UV} \sim -1.60$ and $F(H\alpha)/F(H\beta) \sim 11.2$, placing it close to the LRD locus in the β_{UV} versus β_{opt} diagram (Fig. 2). Rather than excluding such borderline cases, we retain them in the sample as they naturally probe the continuum of properties bridging the two samples. This edge population is qualitatively consistent with the viewing-angle interpretation of Sect. 5.2, in which sources observed at larger θ , intercepting more of the equatorial envelope, would naturally display redder slopes and stronger Balmer absorption, transitioning smoothly into the LRD regime, or is challenged by luminous host galaxy emission. In this case, the separation between LRDs and compact blue BLEs is placed at a relatively arbitrary threshold that reflects a specific gas column density (e.g., [Matthee et al. 2026](#)). Table A.1 lists all 20 sources, their observing programs, and the available data products.

4. Spectral analysis

In this section, we characterize the spectral properties of the 20 compact blue broad-line emitters selected in Sect. 3, with the aim of constraining two quantities central to their role in cosmic reionization: the production rate of ionizing photons and the fraction that escape into the IGM.

Constraining these quantities in this population is non-trivial. Standard calibrations of the ionizing photon production efficiency ξ_{ion} rely on $H\alpha$ under the case B recombination assumption, but the compact blue BLEs may not satisfy this condition because high gas densities give rise to non case B conditions, for example due to a high Balmer optical depth (e.g., [Scarlata et al. 2024](#)). Similarly, standard $Ly\alpha$ -based estimates of the escape fraction and $H I$ column density rely on the line velocity offset and profile shape, which are inaccessible at the resolution of

NIRSpec/PRISM (Sect. 4.3). We therefore use these diagnostics not to derive ξ_{ion} or f_{esc}^{LyC} directly, but to build a consistent observational picture that motivates and constrains the self-consistent photoionization modeling of Sect. 5.

4.1. $H\alpha$ luminosity and Balmer decrement

The $H\alpha$ luminosity and Balmer decrement $F(H\alpha)/F(H\beta)$ provide complementary constraints on the nebular conditions in the compact blue BLEs. While $H\alpha$ traces the total recombination rate and sets a lower bound on the ionizing photon budget, the Balmer decrement encodes information on the physical conditions of the gas and the dust attenuation. In LRDs, Balmer decrements are typically very high (with sample averages of $F(H\alpha)/F(H\beta) \approx 8.7$; [de Graaff et al. 2025b](#)) and correlate with the Balmer break strength ([Matthee et al. 2026](#)), suggesting that both are driven by the same dense gaseous envelope through scattering and collisional (de-)excitation ([Torrallba et al. 2026a](#); [de Graaff et al. 2025b](#); [Chang et al. 2025](#)). The blue UV slopes ($\beta_{UV} \leq -1.5$) in our sample already suggest that the line of sight to the central source is relatively unobscured, though this alone cannot distinguish between dust attenuation and the dense gas scattering effects that characterize LRDs. Measuring the Balmer decrement provides a more direct handle on the dust and gas conditions along the line of sight, while acknowledging that elevated decrements can reflect both dust reddening and collisional effects in a dense medium, a degeneracy that the Balmer break strength and photoionization modeling of Sect. 5 help to break. We therefore measure both quantities from the PRISM spectra as follows.

The $H\alpha$ emission line of each source was decomposed into a narrow and a broad Gaussian component following the criteria described in Sect. 3.1. In addition, we fit the $H\beta$ emission line with a single Gaussian plus a power-law continuum, using the same Levenberg–Marquardt minimization and LSF convolution described in Sect. 3.1. A single-component model is adopted for $H\beta$ because the broad component is not independently detected

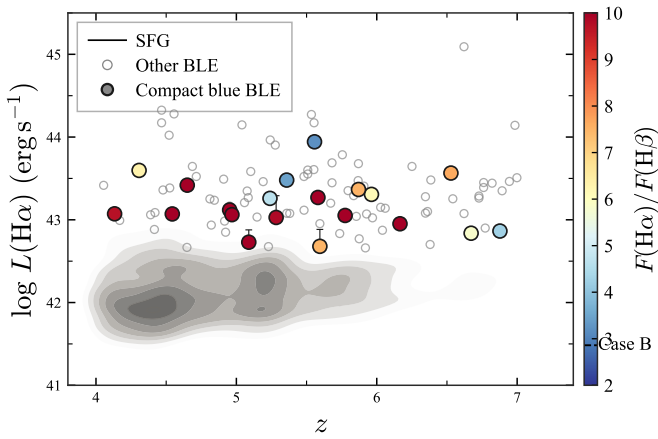


Fig. 4: $H\alpha$ luminosity as a function of redshift. Filled circles show the 20 compact BLEs, color-coded by the observed Balmer decrement $F_{H\alpha}/F_{H\beta}$; the color scale goes from the Case B value of 2.86 (blue) to 10 (red). Open circles show the other BLE sample; gray contours the SFG sample from the DJA compilation.

at sufficient S/N in the PRISM spectra; a two-component fit would introduce poorly constrained degeneracies between narrow and broad amplitudes at this spectral resolution. Following de Graaff et al. (2025b), the Balmer decrement $F(H\alpha)/F(H\beta)$ is computed using the total $H\alpha$ and $H\beta$ fluxes, as the spectral resolution of NIRSpec/PRISM does not allow a reliable decomposition into broad and narrow components.

Fig. 4 shows the total $H\alpha$ luminosity $L_{H\alpha}$ as a function of redshift for our 20 compact, blue BLEs (filled circles, color-coded by the observed Balmer decrement), compared with the other BLE sample (open circles) and the SFG sample from the DJA compilation (gray contours). The compact blue BLEs span $\log(L_{H\alpha}/\text{erg s}^{-1}) \approx 42.7\text{--}43.9$, with a median of 43.2, consistent with the broader BLE distribution (median $\log L_{H\alpha} = 43.3$). Both samples lie systematically above the typical SFG luminosities at comparable redshifts (e.g., Covelo-Paz et al. 2024).

Strikingly, every source in the sample has a Balmer decrement above the case B recombination value of 2.86, spanning $F(H\alpha)/F(H\beta) \approx 3.2\text{--}18.8$ with a median of 8.7, far above the value of 3.06 measured for low-redshift blue AGN by Dong et al. (2008). The two sources with decrements closest to case B – GS_3073 ($F(H\alpha)/F(H\beta) \sim 3.21$) and GN-16813 (~ 3.46) – are consistent with a relatively clear nebular sight line, in agreement with the strong $\text{Ly}\alpha$ detection and low H I column inferred for GN-16813 (Sect. 4.3 and 4.4). The majority of the sample shows substantially elevated values: 16 out of 20 sources have $F(H\alpha)/F(H\beta) > 5$, and 8 exceed $F(H\alpha)/F(H\beta) > 10$. These high decrements point to a dense, partially covering gas envelope that preferentially drives the Balmer decrements away from case B, consistent with the geometry proposed for LRDs (e.g., Sneppen et al. 2026b; Matthee et al. 2026). This is discussed in the context of our photo-ionization modeling in Sect. 4.5.

4.2. Optical emission line diagnostics

Classical BPT diagrams (Baldwin et al. 1981) lose discriminating power at high redshift: lower metallicities and harder ionizing fields push both star-forming galaxies and AGN narrow-line regions toward overlapping loci, making AGN selection based on rest-frame optical diagnostics ambiguous (e.g.,

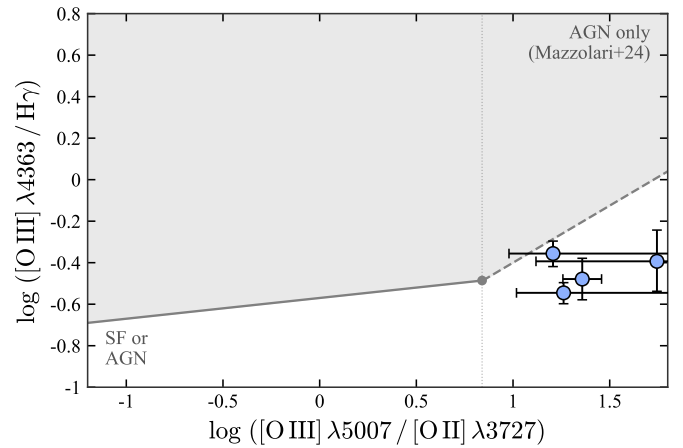


Fig. 5: Optical AGN diagnostic diagram following Mazzolari et al. (2024). y-axis: auroral-to-Balmer ratio $\log([\text{O III}] \lambda 4363 / \text{H}\gamma)$; x-axis: ionization-sensitive ratio $\log([\text{O III}] \lambda 5007 / [\text{O II}] \lambda 3727)$. Filled circles show the compact BLEs with grating coverage for all the four lines.

Harikane et al. 2023; Übler et al. 2023; Scholtz et al. 2025). We instead exploit the medium- and high-resolution grating spectra, which resolve emission lines blended at PRISM resolution, searching for $[\text{O II}] \lambda\lambda 3727, 3729$, $[\text{O III}] \lambda\lambda 4959, 5007$, $\text{H}\gamma$, and the auroral line $[\text{O III}] \lambda 4363$. Of particular interest is the latter, recently proposed as an AGN diagnostic at high redshift: Mazzolari et al. (2024) showed that $[\text{O III}] \lambda 4363 / \text{H}\gamma$ is significantly elevated in AGN due to the harder ionizing spectrum driving the gas to higher temperatures. At PRISM resolution $[\text{O II}] \lambda 3727$ and $\text{H}\gamma$ are blended with neighboring lines, restricting this analysis to the 4/20 sources with complete grating coverage (see Table A.1). Line fluxes and equivalent widths are measured with *unite* (Hvinding 2025).

Fig. 5 shows the $[\text{O III}] \lambda 4363 / \text{H}\gamma$ versus $[\text{O III}] \lambda 5007 / [\text{O II}] \lambda 3727$ diagnostic diagram (Mazzolari et al. 2024) for compact BLEs with grating coverage and significant ($\text{S/N} \geq 3$) detections in all four lines. The demarcation line from Mazzolari et al. (2024), based on the photoionization models of Gutkin et al. (2016) and Feltre et al. (2016), separates the AGN-only region (upper left) from the region accessible to both star-forming galaxies and AGN. Our sources predominantly occupy the latter region, consistent with star-forming galaxies or composite systems rather than AGN.

The fact that our sources are not identified as AGN by this optical diagnostic does not contradict their broad-line nature. Rather, it suggests that the $[\text{O III}] \lambda 4363$ emission in these objects is not strongly enhanced relative to $\text{H}\gamma$, which may reflect low gas-phase metallicity (although this is probably only relevant at metallicities below $\approx 5\%$ solar that are rare at these redshifts, e.g. Kotiwal et al. 2026) or a softer ionizing spectra than typical AGN models, yielding more modest temperature increases, or some contribution from star-formation-powered photoionization to the narrow-line emission. Regardless, none of the rest-frame optical emission-line diagnostics convincingly shows that the broad-line sources are powered by a classical AGN spectrum.

4.3. $\text{Ly}\alpha$ detection

$\text{Ly}\alpha$ is a powerful probe of the neutral gas covering fraction in the ISM. Its emerging equivalent width, escape fraction and ve-

locity offset are sensitive to the H I column density and geometry along the line of sight (e.g., Tang et al. 2026) and therefore also to the ionizing photon escape fraction (e.g., Verhamme et al. 2017). At the resolution of NIRSpec/PRISM, however, the Ly α line profile is severely blurred by the instrumental LSF (FWHM $\sim 10^4$ km s $^{-1}$ at $z \sim 4$ –7, where $R \sim 30$ –50), making it impossible to characterize the line profile, measure velocity offsets, or detect lines with $EW_0 \lesssim 20$ Å (Jones et al. 2024). PRISM-based Ly α measurements are therefore restricted to integrated line flux and equivalent width: they are sufficient to identify Ly α emitters and quantify the LAE fraction across the sample, but insufficient to constrain the escape kinematics or the H I column in individual sources. With these caveats in mind, we fit the Ly α emission line of all analyzed sources (both SFGs and BLEs) following the method of Napolitano et al. (2024).

The line profile is modeled as a Gaussian convolved with the NIRSpec/PRISM instrumental line spread function, superimposed on a step-function continuum that accounts for the sharp break in Ly α opacity blueward of the line. The blue-side continuum level is fixed to zero, consistent with full IGM absorption at $z \gtrsim 4$, while the red-side continuum is modeled as a linear function fitted over the rest-frame window 1260–1900 Å. The continuum level at the Ly α wavelength, $C_{Ly\alpha}$, is evaluated by averaging the best-fit linear model over 1216–1240 Å. The LSF is approximated as a Gaussian with σ_{LSF} derived from the point-source resolution curve evaluated at the observed Ly α wavelength. The free parameters (line amplitude and FWHM) are sampled with the affine-invariant MCMC ensemble sampler EMCEE (Foreman-Mackey et al. 2013), using $N_{walkers} = 64$ walkers, $N_{steps} = 2000$ steps and a burn-in of 500 steps. Parameter uncertainties correspond to the 16th–84th percentile range of the marginalized posteriors.

A Ly α detection is claimed at $S/N \geq 3$, with the peak position searched within a velocity window of -300 to $+1000$ km s $^{-1}$ relative to the systemic Ly α wavelength. The rest-frame equivalent width is computed as $EW_0(Ly\alpha) = F_{Ly\alpha}/[(1+z)C_{Ly\alpha}]$, where $C_{Ly\alpha}$ is the continuum flux density evaluated redward of the line from the best-fit linear model. For non-detections, a 3σ upper limit on the EW is derived from the local noise at the expected Ly α peak position and the LSF width following $EW_{0,lim} = \sqrt{2\pi} \sigma_{LSF} \sigma_{noise}/[(1+z)C_{Ly\alpha}]$.

To ensure the LAE fractions are not biased by sources whose continuum is too faint to constrain the EW, we apply an injection-recovery completeness test to all non-detections. We inject a synthetic Ly α line with $EW_0 = 25$ Å – the LAE threshold – into each spectrum at $+500$ km s $^{-1}$ redward of the systemic wavelength, perturb the spectrum with $N_{trials} = 100$ independent noise realizations, and attempt to recover the injected line with the full MCMC fitter. Non-detections for which the median recovered $S/N \geq 3$ are classified as genuine non-LAEs and retained in the analysis; those for which the injection is not recovered are classified as incomplete and excluded from the denominator when computing $X_{Ly\alpha}$.

Among the 20 compact blue BLEs, 12 are detected in Ly α with $S/N \geq 3$, of which 11 satisfy $EW_0 > 25$ Å, yielding an overall Ly α emitter (LAE) fraction of 55%. The detected equivalent widths span $EW_0 \approx 24$ –235 Å, with a median of ≈ 65 Å.

Fig. 6 shows EW_0 as a function of M_{UV} for all three samples. At fixed M_{UV} , the compact blue BLEs show systematically elevated EW_0 relative to the SFG population across the full redshift range $4 \lesssim z \lesssim 7$ spanned by the sample (median $z \approx 5.5$), consistent with a stronger ionizing continuum boosting intrinsic Ly α production and/or a higher Ly α escape fraction.

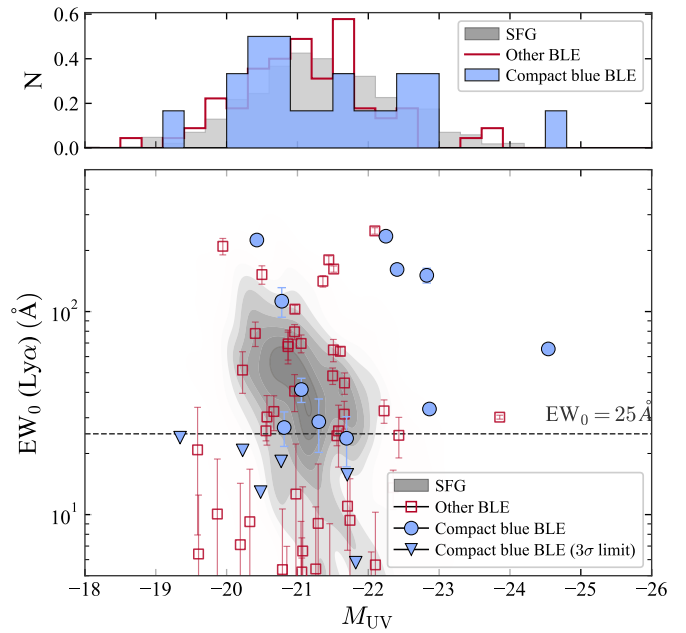


Fig. 6: Ly α rest-frame equivalent width EW_0 as a function of absolute UV magnitude M_{UV} . Filled circles show compact blue BLEs with significant detection ($S/N \geq 3$), with 1σ error bars; downward triangles are 3σ upper limits. Open squares show the other BLE sample with 1σ error bars. Grey filled contours show the SFG comparison sample. The M_{UV} distribution for the three populations is presented in the upper panel.

Fig. 7 quantifies this enhancement via the Ly α emitter fraction $X_{Ly\alpha}$, i.e. the fraction of sources with $EW_0 > 25$ Å, as a function of M_{UV} in bins of $\Delta M_{UV} = 1.5$ mag. Poisson errors are used for the SFG sample; small-number statistics (Gehrels 1986) for the BLEs; bins with fewer than 3 sources are omitted. The SFG sample spans $-18 \lesssim M_{UV} \lesssim -24$, with $X_{Ly\alpha}$ ranging from $\approx 7\%$ at the bright end to a peak of $\approx 53\%$ near $M_{UV} \approx -21$, in broad agreement with published LAE fractions for UV-selected galaxies at these redshifts (e.g., Stark et al. 2011; De Barros et al. 2017; Napolitano et al. 2024). The other BLE sample spans $-19 \lesssim M_{UV} \lesssim -24$ and shows systematically elevated fractions of ≈ 11 –39% across this range. The compact blue BLEs stand out most strikingly: both bins in which the subsample is represented reach $X_{Ly\alpha} \gtrsim 50\%$, consistently above both the SFG and other BLE levels at matched luminosity. The redshift dependence of $X_{Ly\alpha}$ is shown in Fig. 8, where both the SFG and other BLE populations show little evolution between $z \approx 4.6$ and $z \approx 6.1$, with fractions of ≈ 35 –39% in both redshift bins (e.g., Stark et al. 2011; De Barros et al. 2017; Jones et al. 2024). The compact blue BLEs show higher fractions in both bins, rising from $\approx 50\%$ at $z \approx 4.8$ to $\approx 67\%$ at $z \approx 6.2$, though the large uncertainties preclude a firm conclusion about redshift evolution within this subsample.

4.4. UV emission lines

Rest-frame UV emission lines provide a direct probe of the ionization conditions and hardness of the ionizing continuum, complementing the Ly α and H α diagnostics discussed above. We search for the brightest UV metal and helium lines in the NIRSpec/PRISM spectra of the 20 compact, blue BLEs: N IV $\lambda\lambda 1483, 1487$, C IV $\lambda\lambda 1548, 1551$, He II $\lambda 1640$,

Table 3: Ly α measurements for the 20 compact blue broad-line emitters, sorted by M_{UV} .

Source	M_{UV} [AB mag]	$F_{Ly\alpha}$ [10^{-18} erg s $^{-1}$ cm $^{-2}$]	EW_0 [Å]	S/N
GS_3073	-24.54 ± 0.01	105.84 ± 0.27	65.4 ± 0.2	391.5
capers-cos07-v4_6368_35805	-22.86 ± 0.03	8.98 ± 0.51	33.1 ± 1.9	17.7
GN-16813	-22.83 ± 0.16	57.37 ± 4.87	151.6 ± 12.9	11.8
uncover-v4_2561_11254	-22.50 ± 0.03	< 0.48	< 1.4	1.2
gto-wide-uds12-v4_1215_1259	-22.41 ± 0.09	32.63 ± 1.48	161.0 ± 7.3	22.1
rubies-egs53-v4_4233_50052	-22.25 ± 0.04	51.03 ± 1.14	235.1 ± 5.2	44.9
abell2744-greene-v4_8204_69688	-21.82 ± 0.03	< 1.49	< 5.8	1.0
rubies-uds32-v4_4233_966096	-21.70 ± 0.16	< 1.28	< 15.7	1.0
jades-gdn198-v4_1181_38147	-21.69 ± 0.10	2.74 ± 0.74	23.8 ± 6.5	3.7
jades-gdn2-v4_1181_38509	-21.30 ± 0.11	1.51 ± 0.44	28.9 ± 8.3	3.5
capers-egs61-v4_6368_20952	-21.05 ± 0.11	4.81 ± 0.65	41.5 ± 5.6	7.4
uncover-v4_2561_38108	-20.81 ± 0.08	2.46 ± 0.47	27.0 ± 5.2	5.2
rubies-uds23-v4_4233_172350	-20.78 ± 0.27	6.05 ± 1.00	111.8 ± 18.5	6.0
macs1423-v4_1208_4103248	-20.77 ± 0.24	< 2.62	< 18.3	2.9
capers-egs47-v4_6368_19300	-20.48 ± 0.07	< 1.60	< 12.9	2.8
jades-gdn09-v4_1181_73488	-20.43 ± 0.07	30.95 ± 1.21	225.3 ± 8.8	25.5
rubies-egs62-v4_4233_42232	-20.38 ± 0.34	< 2.13	...	1.1
nexus-obs3-v4_5105_10835	-20.24 ± 0.41	10.48 ± 2.80	...	3.7
jades-gdn11-v4_1181_1093	-20.23 ± 0.25	< 1.05	< 20.7	1.2
ceers-ddt-v4_2750_1768	-19.34 ± 0.26	< 0.71	< 23.9	1.2

Notes. Upper limits (3σ) are given for sources with $S/N < 3$. For nexus-obs3 v4_5105_10835 and rubies-egs62-v4_4233_42232, the continuum redward of Ly α is unresolved in the PRISM spectrum, preventing a reliable EW measurement (...).

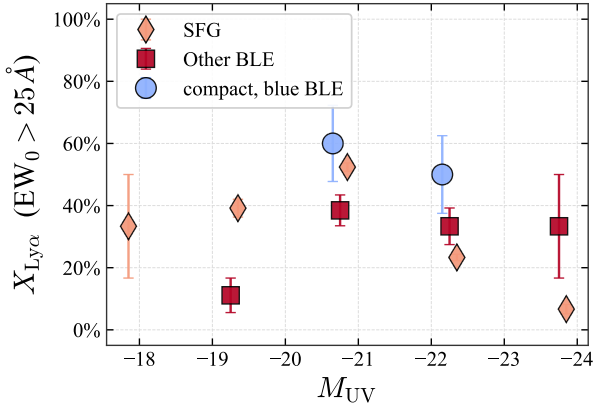


Fig. 7: Ly α emitter fraction $X_{Ly\alpha}$ (fraction of sources with $EW_0 > 25 \text{ \AA}$) as a function of M_{UV} , in bins of $\Delta M_{UV} = 1.5$ mag. Bins containing fewer than 3 sources are omitted. Error bars represent Poisson uncertainties for SFGs and small-number statistics for the other two samples. Orange diamonds: SFG sample. Dark red squares: other BLE sample (excluding compact blue BLEs). Blue circles: compact blue BLE subsample.

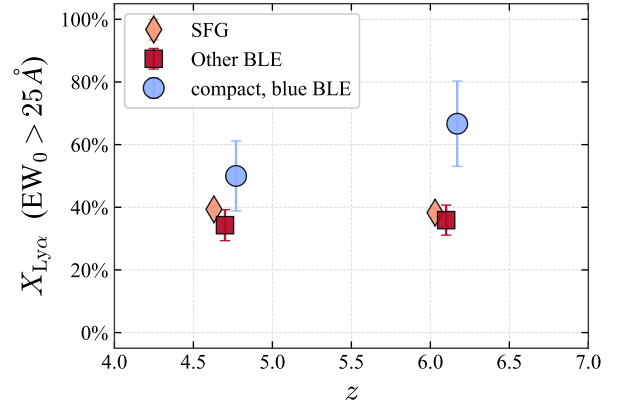


Fig. 8: Ly α emitter fraction $X_{Ly\alpha}$ (fraction of sources with $EW_0 > 25 \text{ \AA}$) as a function of redshift, in bins of $\Delta z \sim 1.5$. Symbols are as in Fig. 7.

O III] $\lambda\lambda 1661, 1666$, and C III] $\lambda\lambda 1907, 1909$. Each line is fitted with `unite` (Hviding 2025). A detection is claimed at $S/N \geq 3$.

O III] $\lambda\lambda 1661, 1666$ is the most commonly detected UV line in our sample, recovered in 4 out of 20 sources; we note however that at the PRISM resolution this line is blended with He II $\lambda 1640$. N IV] $\lambda\lambda 1483, 1487$, which requires ionizing photons above 47.4 eV, is detected in 3 sources. C IV] $\lambda\lambda 1548, 1551$ and C III] $\lambda\lambda 1907, 1909$ are each detected in 2 sources. The overall detection rates are low, as expected given the PRISM resolution and the modest per-pixel S/N ratio of these faint sources; confirmed detections are found exclusively among the sources with the highest continuum S/N in the sample. However, we note

that PRISM spectra of *typical* star-forming galaxies do not show such UV lines (e.g., Hayes et al. 2025).

The UV line properties collectively reinforce the picture established by Ly α and broad H α . The simultaneous detection of N IV] and C IV] in rubies-egs53-v4_4233_50052 – both requiring $> 47 \text{ eV}$ photons – and of N IV] at large equivalent width ($EW_0 > 20 \text{ \AA}$) in abell2744-greene-v4_8204_69688 places these sources in the AGN locus of the C IV] EW diagnostic (Nakajima et al. 2018). Notably, N IV] detections at high equivalent width are rare in the general galaxy population and suggest nitrogen enhancement (e.g., Cameron et al. 2023; Zhu et al. 2026), as also reported for GN-z11 (Maiolino et al. 2024b) and other extreme high- z systems (Berg et al. 2026). However, these high-ionization features are absent in the stacked spectrum (Sect. 4.5), suggesting that the UV emission in the sample as a whole reflects a superposition of host galaxy and central engine contributions,

with the AGN UV signature emerging only in the highest-S/N individual sources.

At the resolution of NIRSpec/PRISM ($R \sim 50\text{--}100$), UV doublets are unresolved, O III is blended with He II , and lines with $\text{EW}_0 \lesssim 20 \text{ \AA}$ fall below the effective detection threshold even at moderate continuum S/N ratio (e.g., Jones et al. 2024). Medium- and high-resolution spectroscopy is therefore required to resolve blended features, detect weaker UV lines, identify ISM absorption, and disentangle AGN, nebular, and stellar-wind contributions (e.g., Keerthi Vasani et al. 2026).

4.5. Overall spectral shape

To illustrate the population-averaged spectral properties of the compact, blue BLEs, we construct a median rest-frame stack of all 20 sources. For each source, the observed-frame PRISM spectrum is shifted to rest-frame wavelengths using the spectroscopic redshift and normalized over the rest-frame window 5200–5500 Å. All spectra are then interpolated onto a common rest-frame wavelength grid with spacing $\Delta\lambda_{\text{rest}} = \Delta\lambda_{\text{inst}}/(1 + z_{\text{median}})$, where $\Delta\lambda_{\text{inst}}$ is the instrumental pixel scale. The stacked spectrum is computed as a weighted median, with weights derived from the per-pixel S/N ratio estimated via a rolling median, in order to minimize the impact of outliers and spectral artifacts. A weighted mean stack and its standard error are computed in parallel for comparison, and the results are consistent.

The resulting stack is shown in Fig. 9, together with equivalent stacks of the LRDs from the “other BLE” sample and the SFG comparison sample constructed with the same procedure, and a QSO template at $1 < z < 2.1$ (Selsing et al. 2016) for reference. The rest-frame UV continuum exhibits a consistently blue slope with $\beta_{\text{UV}} = -1.72 \pm 0.04$ for the median stack. For comparison, the median stack of the LRDs has a $\beta_{\text{UV}} = -1.46 \pm 0.04$ (in agreement with Ando et al. 2026).

Despite sharing several spectral features with broad-line quasars – including broad Balmer emission and strong high-ionization lines – compact blue BLEs are fundamentally distinguished by their significantly lower luminosities, and differences in the optical β slopes. Moreover, we do not identify prominent *broad* lines in the UV such as $\text{Ly}\alpha$, C IV , C III or Mg II , which are hallmarks of classical broad-line quasar spectra. In the rest-frame optical stack, the compact blue BLEs exhibit $\text{EW}_0(\text{H}\alpha) \approx 900 \text{ \AA}$, significantly larger than the $\sim 200 \text{ \AA}$ typical of luminous quasars (e.g., Vanden Berk et al. 2001). Compared to the SFG population, both BLE samples are distinguished by the absence of $[\text{S II}]\lambda\lambda 6718, 6733$ emission (3σ upper limits of $\text{EW}_0 < 7.4$ and $4.6 \text{ \AA}$ for the compact blue BLE and LRD stacks respectively, versus detections of $\text{EW}_0 \approx 21 \text{ \AA}$ in the SFG stack), consistent with a harder ionizing continuum suppressing low-ionization forbidden line emission. The compact blue BLE stack also shows strong $\text{He I } \lambda 5875$ ($\text{EW}_0 = 15.9 \pm 4.3 \text{ \AA}$) and $\text{He I } \lambda 7065$ ($27.3 \pm 6.5 \text{ \AA}$) emission – notably stronger than in the SFG stack (12.5 ± 0.7 and $5.1 \pm 2.0 \text{ \AA}$ respectively).

The rest-frame optical part of the compact blue BLE and LRD stacks are strikingly similar, sharing comparable $\text{H}\alpha$ equivalent widths, Balmer line ratios, optical β slopes, and $\text{He I } \lambda 5875$ and $\text{He I } \lambda 7065$ line strengths. This resemblance suggests that the compact blue BLEs and LRDs may share a common powering source, with the observed differences in UV properties reflecting variations in covering gas that may be due to differences in the viewing angle or in evolutionary phases of the dense envelope (Matthee et al. 2026; Sneppen et al. 2026a), rather than a funda-

mentally distinct physical nature. We explore this connection in detail in Sect. 5.

5. Photoionization modeling with SIROCCO

Considering the spectral similarities between compact blue BLEs and LRDs discussed in Sect. 4.5, together with their similar X-Ray faintness and lack of variability (Brazzini et al. 2026), compact blue BLEs have been proposed to represent the counterparts of LRDs viewed from sight lines in the envelope with lower column densities (Matthee et al. 2026; Sneppen et al. 2026b), where the UV continuum and associated high-ionization lines can escape. This is supported by their elevated LAE fractions relative to both the BLE and SFG baselines at matched luminosity (Fig. 7), and their large $\text{Ly}\alpha$ equivalent widths (Fig. 6), all pointing to low H I columns along the line of sight.

If compact blue BLEs and LRDs are primarily distinguished by envelope geometry and line-of-sight opacity, standard diagnostics of ξ_{ion} and $f_{\text{esc}}^{\text{LyC}}$ may not be directly applicable. They are calibrated on typical star-forming galaxies and rest on assumptions most critically case B recombination and an optically thin Lyman series – that are unlikely to hold when a dense gas envelope surrounds the central ionizing source. In this geometry, ionizing photons are reprocessed before escaping, and the emergent line and continuum fluxes reflect the combined effects of column density, covering factor, and viewing angle rather than the intrinsic photon budget (Chang et al. 2026).

For this reason, we use 3D² Monte Carlo radiative transfer simulations with SIROCCO (Matthews et al. 2025) to model the ionizing photon production and escape directly. While CLOUDY remains the standard tool for photoionization modeling, it assumes a static medium and cannot self-consistently predict resolved emission-line profiles, since it lacks a treatment of the gas kinematics. SIROCCO simultaneously reproduces the continuum and line-profile shapes for a given gas configuration, retaining the flexibility to explore different geometries, density structures, and clumping factors.

The constraints on the ionizing spectrum from the SED shape (Sect. 4.5) and on the gas geometry from the Balmer line profiles together provide sufficient information to set up the simulations (Sect. 5.1). SIROCCO employs a Monte Carlo radiative transfer framework specifically designed to treat optically thick, multi-dimensional flows. The code self-consistently solves the ionization structure, radiation field, electron column density, recombination line strengths and widths, extent of partially ionized material, and the strength of Balmer absorption lines and breaks throughout the gas cocoon (Parkinson et al. 2025; Matthews et al. 2025). However, SIROCCO has a number of limitations (Matthews et al. 2025). Line transfer is treated in the Sobolev approximation which assumes that the wind has a sufficiently large velocity gradient that a photon traversing the wind sweeps through the resonance wavelength of any given transition on a timescale much shorter than the transition lifetime. As a result, only permitted transitions can be excited: the photon passes through resonance before a forbidden or semi-forbidden transition, which requires a comparatively long interaction time, can occur. This naturally suppresses forbidden line emission such as $[\text{O III}]\lambda\lambda 4959, 5007$, which are therefore not considered in the SIROCCO wind model. The approximation neglects thermal and microturbulent broadening of bound-bound transitions,

² SIROCCO is a 2.5-dimensional code, where the coordinate grid is 2D and symmetric about the z -axis. However, photon transport occurs in 3D, allowing rotational motion around the y -axis.

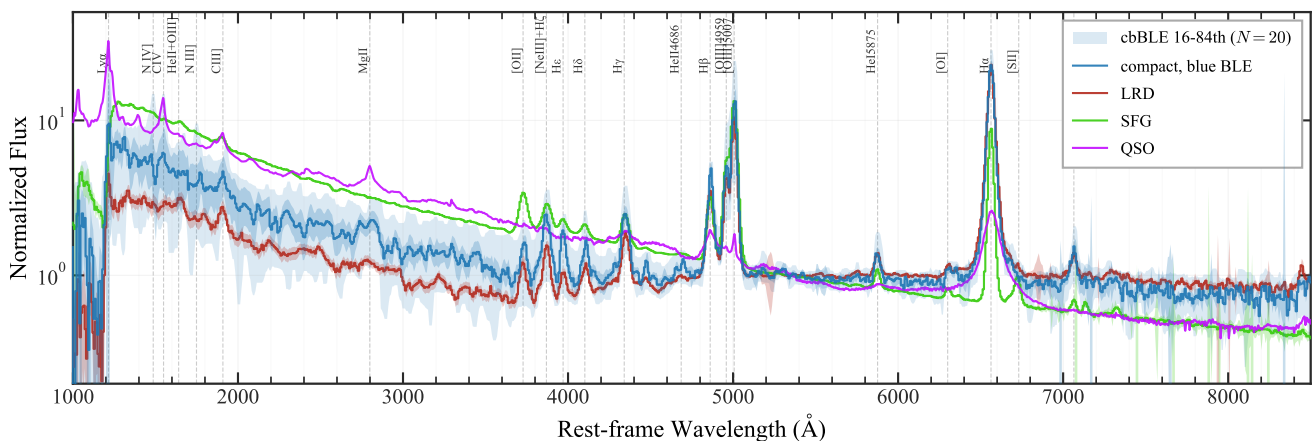


Fig. 9: Rest-frame stacked spectra of high-redshift galaxy populations. The compact, blue BLEs’ median stack is shown in blue, the green line shows the median spectra of star-forming galaxies at $z = 4 - 7$ from DJA. The shaded region indicating $\pm 1\sigma$. The red line shows the median stack of the LRDs from the “Other BLE” sample. The purple line shows the median stack of QSO at $z \sim 1.5$ from Selsing et al. (2016). Vertical dashed gray lines mark prominent emission lines. All spectra are normalized between 5200–5500 Å to facilitate comparison of spectral shapes. The compact, blue BLEs’ stack exhibits distinct spectral features compared to the star-forming population, particularly in the strength and profile of emission lines such as H α , H β , and He I, suggesting different physical conditions and/or dominant ionization mechanisms.

but may only introduce minor uncertainties in the cores of optically thick lines. Finally, the code assumes thermal and ionization equilibrium throughout the wind, which may not hold in the fastest-moving regions of the outflow where the local dynamical timescale is shorter than the recombination timescale.

5.1. Model setup and physical configuration

To identify a physical configuration capable of reproducing the observed spectral properties of the compact blue BLE sample, we explore a grid of SIROCCO radiative transfer models varying the wind geometry, column density, and ionizing source properties (the full grid is described in Sect. 6.1.3 and Appendix B). Here we present the single best-fit model from this exploration, selected by minimizing the reduced χ^2 between the model and the stacked compact blue BLE spectrum, jointly across the four Balmer-break bins presented in Matthee et al. (2026) and all viewing angles (Sect. 5.2).

The SIROCCO model has an axisymmetric two-component geometry consisting of a polar outflow and an equatorial inflow, first proposed by Sneppen et al. (2026b), who showed that a biconical outflow embedded in an equatorial gas distribution can simultaneously reproduce the broad Balmer profiles and Balmer absorption of the LRDs. Physically, a central ionizing source is surrounded by a gas envelope whose H I column density decreases from the equatorial plane towards the poles: the highest equatorial columns produce the strongest Balmer breaks, while the lowest inclinations intercept little H I and resemble the compact blue BLEs studied here (whose broad Balmer wings nonetheless remain shaped by the high column of ionized gas in the envelope). This viewing-angle sequence naturally reproduces the full spectral diversity of the population, from the reddest, most absorbed LRDs to the unobscured compact blue BLEs. However, we note that a suite of spherically symmetric models with (evolutionary-)variations in the column densities would yield qualitatively similar results, as long as a similar density and ionization structure along the line of sight can be captured.

In our best models, the two zones are separated by the boundary angle $\theta_b \approx 70.5^\circ$ (measured from the polar axis), corresponding to a solid-angle ratio $\Omega = \cos \theta_b / (1 - \cos \theta_b) = 0.5$ between the components; the polar outflow thus subtends approximately one third of the total solid angle, broadly consistent with the fraction of broad-line emitters showing Balmer absorption – a signature of the equatorial inflow – among LRD samples ($\approx 30 - 65\%$; Matthee et al. 2026; Juodžbalis et al. 2026). The transition between the two zones is smoothed over a 3° taper to avoid numerical discontinuities at the boundary.

The gas moves radially at $v(\theta) = v_{\text{out}} = 700 \text{ km s}^{-1}$ in the polar region ($\theta < \theta_b$) and inflows at $v(\theta) = \eta v_{\text{out}}$ with $\eta = 0.01$ in the equatorial region ($\theta > \theta_b$). For a steady-state wind of mass-loss rate $\dot{M}$ spanning inner and outer radii R_{in} and R_{out} , mass conservation fixes the density,

$$\rho(r, \theta) = \frac{\dot{M}}{4\pi r^2 v(\theta)}, \quad v(\theta) = \begin{cases} v_{\text{out}}, & \theta < \theta_b \\ \eta v_{\text{out}}, & \theta > \theta_b, \end{cases} \quad (1)$$

which yields the $\rho \propto r^{-2}$ radial dependence of a steady wind and an equatorial-to-polar density enhancement $\rho_{\text{in}}/\rho_{\text{out}} = 1/\eta = 100$. We normalize $\dot{M}$ so that the polar sight lines intercept a column density of $\log N_{\text{H}} \approx 23.5 \text{ cm}^{-2}$, consistent with the values inferred for the compact blue BLEs (Sneppen et al. 2026a); the equatorial sight lines then reach $\log N_{\text{H}} \approx 25.5 \text{ cm}^{-2}$, characteristic of the heavily absorbed red LRDs (e.g., Matthee et al. 2026). Crucially, the emergent broad line widths are not set by this bulk motion but arise primarily from electron scattering in the dense envelope, with profiles dependent on the electron-scattering optical depth and temperature, the former determined by the density structure of Eq. 1 (Sneppen et al. 2026b).

Keplerian rotation $v_\phi = \sqrt{GM_{\text{BH}}/r} \sin \theta$, where $M_{\text{BH}} = 10^7 M_\odot$ is the mass of the central source, is included to provide angular-momentum support in the equatorial gas, though the emergent spectra along polar sight lines are only weakly sensitive to this term. Through Eq. 1 the radial density slope also governs the ionization structure of the envelope: steeper profiles concentrate the ionized region near the inner boundary, while shallower ones let the ionization front extend outward. Finally,

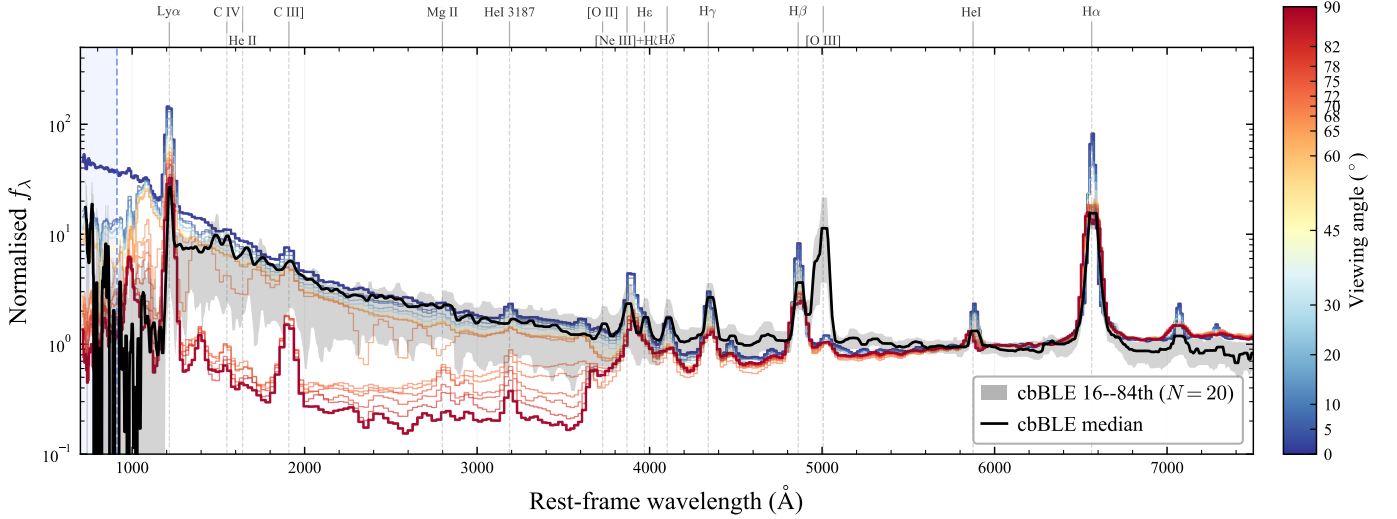


Fig. 10: Best Sirocco radiative transfer model spectra degraded to NIRSpec/PRISM resolution, shown for all viewing angles from pole-on (blue, $\theta = 0^\circ$) to edge-on (red, $\theta = 90^\circ$), compared to the mean stacked spectrum of our sample of compact blue broad-line emitters (black), along with the 16th and 84th percentiles. All spectra are normalized to the median flux in the 6200–6500 Å rest-frame window. The shaded blue region marks the LyC ($\lambda < 912$ Å) region. Key features are labeled at the top.

the medium is clumpy, with a volume filling factor $f = 0.01$; this raises the local recombination rate by $1/f = 100$ relative to a smooth medium of the same mean density, enhancing recombination-line emission while allowing high column densities at lower volume-averaged density.

The central ionizing source is modeled as a blackbody with temperature $T_{\text{BB}} = 5.0 \times 10^4$ K and bolometric luminosity $L = 10^{44}$ erg s $^{-1}$, typical of LRDs (de Graaff et al. 2025b). We adopt this simple parameterization rather than an AGN power-law spectrum: as discussed in Sects. 4.2 and 4.4, the emission-line diagnostics do not unambiguously identify the central engine, and a blackbody provides a physically agnostic but sufficiently hard ionizing continuum to reproduce the observed high-ionization lines without presupposing an accreting black hole. The wind metallicity is set to $Z = 0.01 Z_\odot$ (consistent with Gentile et al. 2026).

The choice of viewing angle for comparison with the compact blue BLE stack follows naturally from this geometric picture. Their lack of Balmer absorption and blue UV continua suggest sight lines dominated by the polar direction, i.e. $\theta \ll \theta_b$. We therefore compare the stacked spectrum primarily with model spectra extracted at $\theta = 0$ – 30° , while using the full range of viewing angles ($\theta = 0^\circ$ – 90°) to illustrate the diversity of emergent spectral morphologies and the connection between compact blue BLEs and LRDs within a unified framework. The fixed model parameters are summarized in Table B.1.

5.2. Emergent spectra and viewing-angle dependence

To compare directly with the observations, we degrade the angle-specific model spectra to the NIRSpec/PRISM resolution ($R \sim 100$) and rescale the flux to the luminosity distance at the median redshift of our sample ($z = 5.3$). The degraded spectra are shown in Fig. 10, overlaid on the stacked PRISM spectrum of the compact blue BLE sample.

Our stack samples only the weak-break, near-pole-on end of the sequence. To test the model across the full range of Balmer break strengths, we additionally compare it to the LRD stacks of

Matthee et al. (2026), binned by Balmer break strength into four bins of increasing break values. Because the Balmer break is set by the neutral-hydrogen column along the line of sight, which in our geometry increases with inclination, these bins probe progressively more equatorial sight lines. We assign to each bin the model angle that minimizes the reduced χ^2 between the binned stack and the degraded model spectra, computed jointly with our compact blue stack: this yields best-matching angles of $\theta \approx [0, 5, 68, 82, 90]^\circ$ and a total reduced chi-square of $\chi^2_\nu = 12$. The recovered trend between the viewing-angle and the Balmer break strength, measured as discussed in Sect. 3.2, is shown in Fig. 11, top panel.

5.3. Constraints on the ionizing photon production and escape

We use the Sirocco model spectra to derive ξ_{ion} and $f_{\text{esc}}^{\text{LyC}}$ as a function of viewing angle. The ionizing photon rate, expressed in units of photons per second, is computed as

$$Q(\theta) = \int_0^{912 \text{ Å}} \frac{L_\lambda(\theta) \lambda}{hc} d\lambda,$$

where $L_\lambda(\theta) = 4\pi D_{\text{model}}^2 F_\lambda(\theta)$ is the angle-specific luminosity at the model distance $D_{\text{model}} = 100$ pc. The total escaping ionizing budget is $Q_{\text{tot}} = \sum_\theta Q(\theta)$, and the viewing-angle-dependent escape fraction is defined as $f_{\text{esc}}^{\text{LyC}}(\theta) = Q(\theta)/Q_{\text{int}}$, where Q_{int} is the intrinsic ionizing photon production rate of the central source prior to radiative transfer. Fig. 11, bottom panel, shows the viewing-angle-dependent escape fraction, which ranges from ~ 0.01 to ~ 0.35 : the polar viewing angles have the highest values due to the lower column density along the outflow cone. We assess the robustness of $f_{\text{esc}}^{\text{LyC}}(\theta)$ by sweeping the upper integration cutoff over $\lambda_{\text{cut}} \in [912 \pm 20]$ Å in 41 steps, propagating the resulting variance as a systematic uncertainty.

To estimate ξ_{ion} , we use the intrinsic spectrum to compute both Q_{int} and the monochromatic UV luminosity $L_\nu(1500 \text{ Å})$, giving $\xi_{\text{ion}} = Q_{\text{int}}/L_\nu(1500 \text{ Å})[\text{Hz erg}^{-1}]$. The predicted $\log \xi_{\text{ion}}$

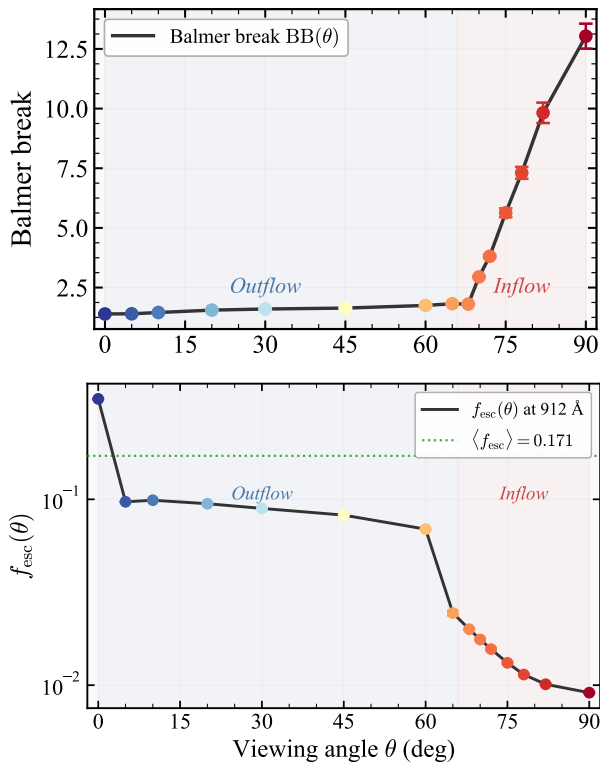


Fig. 11: Viewing-angle dependence of the Balmer break strength (top) and LyC escape fraction (bottom) for the best-fit Sirocco model. The blue and red shaded regions denote lines of sight through the polar outflow ($\theta \lesssim 70.5^\circ$) and equatorial inflow ($\theta \gtrsim 70.5^\circ$), respectively. The Balmer break strength rises sharply across the outflow–inflow transition, coincident with the increase in neutral hydrogen column density along equatorial sightlines. This enhanced neutral column renders the gas optically thick to ionizing photons, producing the corresponding decline in the LyC escape fraction, $f_{\text{esc}}(\theta)$, at $\theta \gtrsim 70^\circ$.

and the viewing-angle distribution of $f_{\text{esc}}^{\text{LyC}}$ are discussed in Sect. 6.2.

6. Discussion

6.1. The nature of compact broad-line emitters: a coherent physical picture

The spectral analysis of Sect. 4 support a physical picture in which compact blue BLEs and LRDs are drawn from the same parent population, with their observed differences reflecting the orientation of the line of sight relative to the axisymmetric gas cocoon as we modeled with Sirocco (Sect. 5), rather than an intrinsically distinct physical nature. In line with this picture, Brazzini et al. (2026) report that one of the compact blue BLEs in our sample (GS_3073) is extremely X-ray weak, and we find the same lack of X-ray emission for the remaining sources in our sample with archival X-ray coverage, mirroring the X-ray weakness reported for LRDs (e.g., Ananna et al. 2024; Setton et al. 2025; Yue et al. 2024). Both classes also lack significant photometric variability on timescales of months to years (e.g., Ji et al. 2025; Zhang et al. 2025; Brazzini et al. 2026; Liu et al. 2026). We now bring these threads together and discuss their quantitative implications.

6.1.1. Viewing-angle classification of the BLE population

Of the 99 confirmed broad-line emitters in the parent sample, 20 satisfy the compact blue BLE selection criteria ($\beta_{\text{UV}} \leq -1.5$, $\beta_{\text{opt}} < 0.5$, and $C_{\text{UV}}, C_{\text{opt}} \leq 1.7$), i.e. $\sim 20\%$ of the PRISM-selected BLE sample. The remaining $\sim 80\%$ are predominantly LRDs (see Fig. 2), characterized by red optical slopes and elevated Balmer breaks. This apparent dominance of LRDs must, however, be interpreted with care, because the PRISM broad-line recovery rate is a strong function of $\text{H}\alpha$ luminosity and line width (Sect. 3.1). Within the Sirocco geometry, we set the wind boundary to $\theta_b \approx 70.5^\circ$, to reproduce the blue-to-LRD ratio measured directly from the PRISM sample. Correcting for the luminosity- and line-width-dependent incompleteness of the broad-line selection (Sect. 3.1), which preferentially removes the fainter blue emitters, raises the intrinsic blue fraction only modestly, from $\sim 26\%$ to $\sim 32\%$, corresponding to $\theta_b \approx 67^\circ$. This $\sim 3^\circ$ shift confirms that our adopted geometry is robust against the selection incompleteness: the completeness correction moves the boundary angle by less than the width of the transition taper and leaves the polar-to-equatorial partition essentially unchanged. In this physical configuration, the pole-on spectrum ($\theta = 0^\circ$) reproduces 8 sources; the intermediate polar spectra at $\theta \approx 5^\circ$, 20° , and 30° match 1, 4, and 3 sources respectively, covering the full range of compact blue BLE spectral types. The remaining 4 sources occupy the transition region at $\theta \approx 45^\circ$ – 65° (with a mix of polar and equatorial gas along the sight line), with 1 source at $\theta \approx 60^\circ$ and 3 at $\theta \approx 65^\circ$, consistent with the borderline cases discussed in Sect. 3.4.

Based on a small sample of LRDs and LBDs at $2 \lesssim z \lesssim 8$ identified by JWST in the GOODS fields, Geris et al. (2026) find that LRD models require either clumpy or equatorial geometries, consistent with our model. While they favor a “dust-obscured LBDs” interpretation, our Sirocco modeling shows that the Balmer break and spectral sequence are primarily governed by gas column density and electron-scattering geometry rather than variations in dust attenuation (Sneppen et al. 2026a; Matthee et al. 2026).

6.1.2. The compact blue BLEs as an unobscured window onto the central engine

So far, we have remained agnostic about the nature of the central ionizing source, which could range from an accreting black hole to a hot stellar populations in star clusters. The envelope geometry can naturally explain the broad Balmer lines through scattering and radiative transfer effects, meaning that line widths alone do not uniquely diagnose the underlying engine. Instead, we turn to excitation-sensitive UV emission lines, whose ionization potentials and relative strengths provide a more direct probe of the hardness of the ionizing radiation field.

Rest-frame UV medium-resolution grating spectroscopy ($R \sim 1000$) is available for one source in the sample, GN-16813, from the DIVER program (GO-8018, PI Lin). At this resolution, $\text{C IV } \lambda\lambda 1548, 1551$ and $\text{N IV } \lambda\lambda 1483, 1486$ are confirmed as the dominant UV emission features, with rest-frame equivalent widths of $\text{EW}_0 \approx 43$ and $31 \text{ \AA}$ respectively. The grating data also recover $\text{He II } \lambda 1640$, $\text{O III } \lambda\lambda 1661, 1666$, $\text{C III } \lambda\lambda 1907, 1909$, and $\text{N III } \lambda 1748$ at $\text{EW}_0 \approx 7^3, 14, 25$, and $5 \text{ \AA}$ respectively. These lines fall entirely below the PRISM detection threshold (Fig. C.1, Table C.1, in Appendix C).

³ broad+narrow components

The `unite` fit to the $\text{He II} + \text{O III}$ window reveals that $\text{He II } \lambda 1640$ requires a broad component in addition to the narrow nebular emission traced by the O III lines, inconsistent with a purely nebular origin. Broad $\text{He II } \lambda 1640$ with $\text{FWHM} \gtrsim 1000 \text{ km s}^{-1}$ is a well-established signature of very massive stars (VMS; Gräfenr & Vink 2015; Marques-Chaves et al. 2024), whose Wolf-Rayet-type stellar winds drive He II into broad emission at low metallicity. This interpretation is supported by recent JWST detections of VMS signatures – strong $\text{N V } \lambda 1240$ and $\text{C IV } \lambda 1550$ P-Cygni profiles alongside broad $\text{He II } \lambda 1640$ – in UV-bright galaxies at $z \sim 8.7$ (Marques-Chaves et al. 2026b). The $\text{N IV}]$ detection further indicates nitrogen enhancement (Cameron et al. 2023; Zhang et al. 2026), consistent with the CNO-processed material exposed by these winds. Taken together, the UV spectrum of GN-16813 is not classically AGN-dominated but suggests a significant contribution from an extremely young ($\lesssim 2 \text{ Myr}$), massive stellar population. An additional AGN contribution cannot be excluded, however, given the broad $\text{H}\alpha$ emission and the line ratios placing GN-16813 in or near the AGN locus of multiple BPT diagnostics (Fig. 5); this inconsistent classification across diagnostics prevents firm conclusions on the dominant ionizing mechanism and underscores the need for higher-resolution spectroscopy to disentangle star formation, shocks, and AGN activity.

$\text{Ly}\alpha$ emission is detected in the grating spectrum with a redshifted peak at $v_{\text{red}} \approx +260 \text{ km s}^{-1}$. This relatively small velocity offset points to a low H I column density along the line of sight, consistent with efficient $\text{Ly}\alpha$, and potentially LyC , escape (Verhamme et al. 2015). The absence of low-ionization interstellar absorption lines in the spectrum suggests a low H I covering fraction, which would favor a high escape fraction of ionizing photons (e.g., Saldana-Lopez et al. 2022). Mapping the complete rest-UV diagnostics (C IV , He II , O III , C III , $\text{N IV}]$ and $\text{N III}]$) at $R \sim 2700$ is therefore the next step, both to settle the nature of GN-16813 and to extend this analysis to the remaining 19 compact blue BLEs lacking deep grating data.

The $\text{He II } \lambda 4686/\text{H}\beta$ ratio provides a diagnostic of the hardness of the ionizing source that is accessible even at PRISM resolution in the brightest sources. An elevated $\text{He II}/\text{H}\beta$, above the value predicted by standard photoionization models for star-forming regions, constitutes direct evidence for a photon field capable of doubly ionizing helium, a signature of AGN accretion discs or extreme stellar populations (Shirazi & Brinchmann 2012; Pérez-González et al. 2026). In the stacked spectrum of the compact blue BLEs, $\text{He II } \lambda 4686$ is detected with $\log(\text{He II}/\text{H}\beta) \approx -1.0$. This value exceeds that of typical star-forming galaxies and their high-redshift analogs ($\log(\text{He II } \lambda 4687/\text{H}\beta) \sim -1.8$, Bian et al. 2020), and is consistent with – or slightly harder than – the stacked value of $\log(\text{He II } \lambda 4687/\text{H}\beta) = -1.45 \pm 0.09$ reported for DESI LRDs (Lin et al. 2026), which itself is interpreted as intermediate between the star-forming and AGN loci (Shirazi & Brinchmann 2012). The systematically harder $\text{He II}/\text{H}\beta$ ratio in the compact blue BLEs relative to LRDs is qualitatively expected in the viewing-angle scenario: at pole-on sight lines, the observer intercepts a smaller column of the dense, partially ionized envelope that reprocesses and softens the emergent spectrum at equatorial viewing angles. The stacked $\text{He II}/\text{H}\beta$ ratio therefore supports the interpretation that compact blue BLEs provide a less-reprocessed view of the same underlying engine responsible for LRDs.

This gradient is directly visible when the sample is split by the viewing-angle assignments of Sect. 5.2 (Fig. 12). The median stack of the 9 sources best matching the pole-on sight lines ($\theta \leq 20^\circ$) displays a systematically bluer and brighter UV con-

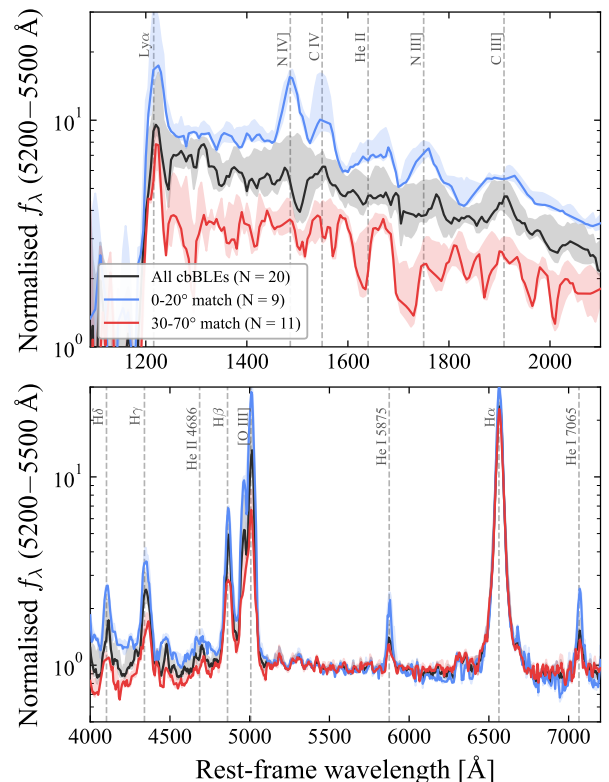


Fig. 12: Median rest-frame stacked spectra of compact blue BLEs split by assigned viewing angle. The pole-on subsample ($\theta \leq 20^\circ$, $N = 9$, blue) and the edge-on subsample ($30 \leq \theta \leq 70^\circ$, $N = 11$, red) are compared to the full sample of 20 compact blue BLEs (black). All the stacks are normalized to the median flux in the 5200–5500 Å rest-frame window; shaded regions show the 1σ uncertainty on the median. *Top*: rest-frame UV (1100–2100 Å). *Bottom*: rest-frame optical (4000–7200 Å). Key emission lines are indicated by vertical dashed lines.

tinuum, with the detection of $\text{N IV}]$, $\text{C IV}]$ and $\text{N III}]$ ($EW_0 = 25.7, 10, 12.5 \text{ Å}$ respectively, versus $EW_0 < 3.7, < 3.6, < 4.1$ for the full sample). These sources also have the largest $\text{Ly}\alpha$ EWs: among the $\text{Ly}\alpha$ emitters in the sample (Sect. 4.3), those assigned to $\theta \leq 20^\circ$ have a median $EW_0(\text{Ly}\alpha) \approx 152 \text{ Å}$ compared to $\approx 42 \text{ Å}$ for the $\text{Ly}\alpha$ emitters at larger viewing angles. This is precisely the result expected when pole-on sight lines intercept the lowest H I column densities: both the ionizing continuum and $\text{Ly}\alpha$ photons escape more freely, and these are therefore the sources with the highest $f_{\text{esc}}^{\text{LyC}}$ in the Sirocco model (Fig. 11).

The UV emission lines discussed above are sensitive to hard photons, but can also be powered by hot stellar spectra (Marques-Chaves et al. 2026b). A more decisive test of the AGN nature can be provided by $[\text{Ne v}] \lambda 3426$, which requires photons above 97.1 eV and is an unambiguous tracer of a very hard radiation field, typically associated with the AGN accretion disc or corona (e.g., Gilli et al. 2010; Mignoli et al. 2013). We searched for $[\text{Ne v}] \lambda 3426$ both in the individual PRISM spectra and in stacked one, finding no significant detection in either. From the stack we place a 3σ upper limit on the line ratio $\log(\text{Ne53}) \equiv \log([\text{Ne v}]/[\text{Ne III}] \lambda 3869) < -0.5$; combined with the measured $\log([\text{O III}] \lambda 5007/\text{H}\beta) \approx 0.6$, this places the compact blue BLEs in the “composite” region of the Cleri et al. (2023) diagnostic, and excludes both the traditional supermassive BH accretion-

disk AGN locus and the Population III/intermediate-mass BH region, which require $\log(\text{Ne53}) > -0.3$. This non-detection is physically consistent with the picture emerging from the $\text{He II}/\text{H}\beta$ analysis: the ionizing spectrum is harder than that of typical star-forming galaxies but softer than that of luminous type-1 AGN or quasars (Wang et al. 2025; Ji et al. 2025). Indeed, $[\text{Ne V}]$ has not been reported in any LRD to date (Lin et al. 2026; Park et al. 2026), and its absence is now understood as a generic property of this class of sources.

6.1.3. Average escape fraction across the population

If the BLE population as a whole – compact blue BLEs and LRDs alike, excluding the rare quasars with much lower number densities – is drawn from the same underlying class of objects, then the viewing angle simultaneously determines both the observed spectral type and the $f_{\text{esc}}^{\text{LyC}}$. In this framework, the observed distribution of spectral types provides an empirical proxy for the population-averaged $\langle f_{\text{esc}}^{\text{LyC}} \rangle$, obtained by weighting the escape fraction at each viewing angle by the corresponding fraction of sight lines. The Sirocco model suggests a steep decline of $f_{\text{esc}}^{\text{LyC}}$ with viewing angle (Fig. 11): polar sight lines ($\theta \lesssim 20^\circ$) reach $f_{\text{esc}}^{\text{LyC}} \approx 0.34$, intermediate angles ($\theta \sim 45^\circ$) ~ 0.02 , and equatorial sight lines ($\theta \gtrsim 70^\circ$) fall below 10^{-2} . The solid-angle-weighted average is $\langle f_{\text{esc}}^{\text{LyC}} \rangle = 0.171$, increasing to ~ 0.20 when weighted by the observed source distribution across angular bins. This same geometry explains the spectral diversity of the population: high-column equatorial sight lines suppress LyC escape and produce strong Balmer breaks, while polar sight lines show weaker breaks and higher escape fractions (Fig. 11). Since both observables are controlled by the same line-of-sight column density, the Balmer break provides a direct proxy for $f_{\text{esc}}^{\text{LyC}}$ without requiring individual radiative-transfer modeling.

To quantify this connection, we exploit the full Sirocco model grid rather than the single best-fit model discussed above: we ran 1229 models spanning a range of column densities, wind geometries, and ionizing source properties (Appendix B). For each model we degrade every viewing-angle spectrum to NIR-Spec/PRISM resolution and compare it against the four Balmer-break-binned stacks of Matthee et al. (2026), which span the dynamic range in the BLE population, assigning to each bin the viewing angle that jointly minimizes the spectral reduced χ^2 and the discrepancy in the $\text{H}\alpha$ FWHM. The resulting reduced- χ^2 distribution is strongly bimodal, with a well-separated population of good-fit models at low χ^2_{ν} and a tail of poor fits at $\chi^2_{\nu} \gg 100$; we therefore retain the 150 models below the natural break in the distribution ($\chi^2_{\nu} \lesssim 30$). The elevated absolute χ^2_{ν} reflects the very small formal uncertainties of the stacked spectra rather than a poor match to the data (Sect. 5.2), so we use the reduced χ^2 to rank and select models rather than as an absolute goodness-of-fit. While this threshold is not unique, varying it within the low- χ^2_{ν} population does not qualitatively affect our conclusions.

Fig. 13 shows $f_{\text{esc}}^{\text{LyC}}$ as a function of the Balmer break for this selected sample, with each model contributing one point per bin at its assigned viewing angle. Despite spanning a wide range of input physical parameters, the models trace a remarkably consistent trend: $f_{\text{esc}}^{\text{LyC}}$ declines steeply with increasing Balmer break before flattening at a residual level of a few $\times 10^{-3}$ for $\text{BB} \gtrsim 10$. Our best model lies within the 16th–84th percentile of the full population at every bin, indicating that it is representative of the broader grid rather than a fine-tuned outlier.

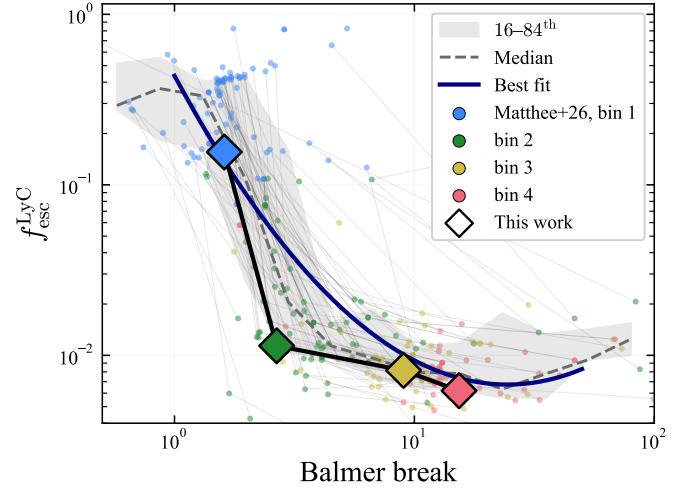


Fig. 13: LyC escape fraction versus Balmer break strength for the 150 best-fitting Sirocco models, selected from the full grid by minimizing the cumulative reduced- χ^2 across the four Balmer-break-binned stacks of Matthee et al. (2026) and our stack. Each colored point shows one model evaluated at the viewing angle assigned to a given bin (blue = bin 1, bluest; green = bin 2; yellow = bin 3; red = bin 4, reddest), with lines connecting the four bins for each model. The gray shaded region marks the 16th–84th percentile range of the population in bins of Balmer break, and the gray dashed line shows the median. The blue curve gives the best-fit second-degree polynomial calibration in log–log space (Eq. 2). Our best-fit model from Sect. 5.2 is highlighted with black-edged diamonds, falling consistently within the population envelope at every bin.

Fitting a second-degree polynomial in log–log space,

$$\log_{10} f_{\text{esc}}^{\text{LyC}} = p_0 + p_1 \log_{10} \text{BB} + p_2 (\log_{10} \text{BB})^2, \quad (2)$$

we obtain $p_0 = -0.42 \pm 0.07$, $p_1 = -2.2 \pm 0.2$, and $p_2 = 0.66 \pm 0.14$. We use this calibration in Sect. 6.2 to estimate $f_{\text{esc}}^{\text{LyC}}$ for the LRD population from their observed Balmer break strengths.

6.2. Implications for the contribution to reionization

Having established the physical nature of the compact blue BLEs and constrained their ionizing properties through Sirocco modeling, we now quantify their contribution to the reionization photon budget. The total ionizing photon production rate of a source population is given by the luminosity-weighted integral

$$\dot{N}_{\text{ion}} = \int \xi_{\text{ion}}(L_{\nu}) f_{\text{esc}}(L_{\nu}) L_{\nu} \phi(L_{\nu}) dL_{\nu}, \quad (3)$$

where $\phi(L_{\nu})$ is the UV luminosity function (LF) of the population, and $\xi_{\text{ion}}(L_{\nu})$ and $f_{\text{esc}}(L_{\nu})$ are allowed to vary with UV luminosity rather than being fixed to a single representative value. This generalization is important because, in our framework, both quantities are governed by the same line-of-sight column density that sets the Balmer break (Sect. 6.1.3): the most UV-luminous, least obscured sight lines preferentially have the lowest Balmer breaks and the highest f_{esc} , while progressively more obscured sight lines have higher Balmer breaks and lower f_{esc} .

We use the calibration in Eq. (13) to assign each BLE population a representative f_{esc} based on its characteristic Balmer break, rather than adopting a single value for all the sources.

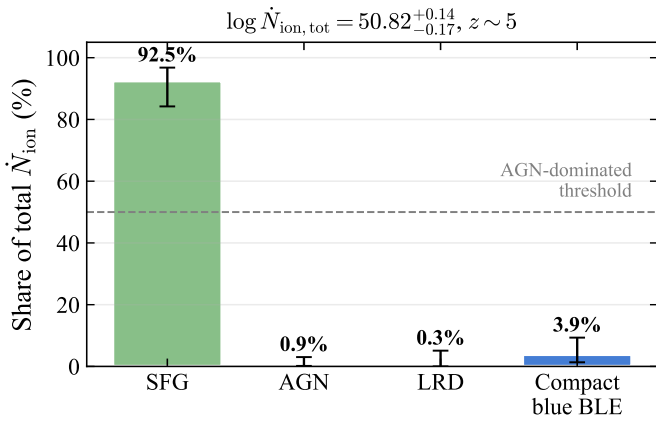


Fig. 14: Distribution of the total ionizing photon production rate $\dot{N}_{\text{ion}}$ for the SFG, AGN, LRDs and compact, blue BLEs, following our “fiducial” model. Error bars are the 68% confidence interval from a Monte Carlo propagation of the input uncertainties (f_{esc} , ξ_{ion} , $L_{\text{UV}}/L_{\text{H}\alpha}$, LF, and population split).

Rather than comparing the (incomplete) distributions of the different populations in our total spectroscopic sample (SFG, BLEs – including classical AGN and LRDs, and the compact blue BLEs), we estimate their contribution to the reionization photon budget from the known luminosity functions. For the star-forming galaxies, we integrate the UV LF of Bouwens et al. (2015) at $z \sim 5$ from $M_{\text{UV}} = -24$ down to $M_{\text{UV}} = -12$, adopting a canonical $\log \xi_{\text{ion}} = 25.2$ (e.g., Llerena et al. 2025; Simmonds et al. 2024; Mascia et al. 2024; Prieto-Lyon et al. 2023) and average $f_{\text{esc}} = 0.1$ (e.g., Mascia et al. 2024; Simmonds et al. 2024; Giovinazzo et al. 2026).

For the BLEs, we integrate the broad- $\text{H}\alpha$ LF at $z \sim 4 - 5.5$ from Matthee et al. (2024) and split it in luminosity: guided by our analysis, we assign emitters brighter than $\log(L_{\text{H}\alpha}/\text{erg s}^{-1}) = 43.5$ to the classical AGN population and fainter ones to LRDs (66%) and compact blue BLEs (33%). We convert $\text{H}\alpha$ to an ionizing-photon rate differently for each population: for the AGN, whose broad $\text{H}\alpha$ arises in a standard broad-line region, we adopt case-B recombination, $Q = L_{\text{H}\alpha}/[1.37 \times 10^{-12} (1 - f_{\text{esc}})]$; for the LRDs and compact blue BLEs, where the $\text{H}\alpha$ is affected by electron scattering in the dense gas, we instead bridge $\text{H}\alpha$ to the UV continuum via the measured $L_{\text{UV}}/L_{\text{H}\alpha}$ ratio, and use $\log \xi_{\text{ion}} = 25.4$ estimated from our best model, avoiding a case-B assumption. We adopt escape fractions of $f_{\text{esc}} = 0.30$ for the AGN (a viewing-angle-weighted average over unobscured and obscured sight lines, D’Amato et al. 2020), 0.2 for the compact blue BLEs, and 0.01 for the LRDs.

The resulting escaping ionizing-photon production rate density at $z = 5$ is $\log(\dot{N}_{\text{ion}}/\text{s}^{-1} \text{Mpc}^{-3}) = 50.82^{+0.14}_{-0.17}$, consistent with the ionizing emissivity inferred from the $\text{Ly}\alpha$ forest at these redshifts (Becker & Bolton 2013). Uncertainties are the 68% confidence intervals from a Monte Carlo propagation of the uncertainties in the adopted f_{esc} , ξ_{ion} , $L_{\text{UV}}/L_{\text{H}\alpha}$ conversions, LF normalizations, and the AGN/non-AGN split. Star-forming galaxies supply the bulk of this budget ($\approx 92.5^{+4.8}_{-8.8}\%$), while the full broad-line population contributes $\approx 5\%$. Within the latter, the compact blue BLEs dominate ($\approx 3.9^{+6.2}_{-2.2}\%$ of the total), whereas the AGN and the LRDs contribute only $\approx 0.9^{+2.2}_{-0.7}\%$ and $\approx 0.3^{+5.5}_{-0.3}\%$ respectively. As expected by the correlation between Balmer break and f_{esc} , although the LRDs contribute the most to the broad- $\text{H}\alpha$ LF, they are negligible reionizers. The broad-line contribution

to reionization is therefore carried almost entirely by the minority of low-column sources. We consider this the “fiducial” model. Its principal uncertainty is precisely the composition of the broad-line population below our completeness limit, which the luminosity split fixes by construction. Adopting instead an equal split between AGN and non-AGN broad-line emitters, the latter divided between LRDs (33.5%) and compact blue BLEs (16.5%), raises the broad-line share of the ionizing budget to $11.8^{+10.8}_{-6.2}\%$, with the AGN term alone growing from 0.9% to $\sim 8\%$. The two scenarios differ by a factor of ~ 2.5 in broad-line emissivity while remaining statistically consistent with one another. Both scenarios fall well short of the $\geq 50\%$ required for AGN-dominated reionization (Madau et al. 2024), but a $\sim 5\%$ contribution is non-negligible, particularly toward the end of reionization, where the ionizing budget is most sensitive to small changes in emissivity (Robertson et al. 2015; Finkelstein et al. 2019). Discriminating between them – and thus placing the broad-line contribution on a firmer footing – requires the systematic analysis of higher-resolution grating spectra and deeper surveys, which recover the fainter emitters that the PRISM misses.

A natural consequence of the elevated ionizing output of compact blue BLEs and LRDs is that they may reside within self-generated ionized bubbles, large enough to render the surrounding IGM transparent to their own $\text{Ly}\alpha$ radiation along the line of sight, even at $z \gtrsim 7$ (Kokorev et al. 2023; Tang et al. 2026), qualitatively similar to the proximity effect seen around quasars (Eilers et al. 2025). This would also help explain the elevated $\text{Ly}\alpha$ emitter fractions reported for this population (Sect. 4.3), since efficient $\text{Ly}\alpha$ escape through the IGM typically requires a sufficiently large ionized region around the source (e.g., Matthee et al. 2018; Mason & Gronke 2020). Following the latter, we estimate the size of the bubbles that could be powered by these sources: at $z = 6$, the bluest sources in our sample ionize bubbles with proper radii of $R_{\text{ion}} = 2.15$ pMpc for $f_{\text{esc}} = 0.4$ and $R_{\text{ion}} = 1.88$ pMpc for $f_{\text{esc}} = 0.2$, without the contribution from any neighboring sources. At $z = 7$, the same range of escape fractions yields $R_{\text{ion}} = 1.61$ pMpc ($f_{\text{esc}} = 0.4$) and 1.41 pMpc ($f_{\text{esc}} = 0.2$), corresponding to co-moving radii of ~ 15 and 11 cMpc, respectively. Interestingly, these sizes are similar to the scale at which the correlation between the $\text{Ly}\alpha$ transmission excess and galaxy distance is largest during the late stages of reionization (Kashino et al. 2026; Jin et al. 2026), i.e. these sizes correspond to the typical regions with excess ionization over the background.

7. Conclusions

We have presented a systematic census of compact, blue broad-line emitters (BLEs) at $4 \leq z \leq 7$, selected from DJA to isolate sources whose UV continuum and broad emission lines arise from the same compact physical region. By combining rest-frame UV and optical spectroscopy with SIROCCO radiative transfer modeling, we address the four questions posed in Sect. 1.

1. *What fraction of sources exhibit bright, blue UV emission that is predominantly powered by the same region as the broad lines?* Starting from a parent sample of 4145 galaxies with simultaneous $\text{Ly}\alpha$ -to- $\text{H}\alpha$ coverage, we identify 99 broad-line emitters, of which only 20 ($\approx 20\%$) satisfy our combined blue-slope and UV/optical compactness criteria (Sect. 3). Although the PRISM broad-line selection is incomplete and biased toward luminous, red sources – so this fraction is a lower limit (Sect. 3.1) – compact blue BLEs represent a genuine minority of the BLE population, con-

sistent with the geometric expectation from our viewing-angle picture (Sect. 6.1). We emphasize, however, that the adopted thresholds ($\beta_{\text{UV}} \leq -1.5$ and $\beta_{\text{opt}} < 0.5$) are somewhat arbitrary: in the viewing-angle framework, the transition from compact blue BLE to LRD spectral types is continuous rather than discrete, with the observed β_{UV} and β_{opt} varying smoothly as a function of the line-of-sight column density (Sect. 5.2).

2. *What are the UV continuum and emission line properties of the blue, compact broad-line emitters as a population?* The compact blue BLEs display blue UV and optical continua, elevated H α equivalent widths, Balmer decrements systematically above the case B value, strong He I emission, and a non-detection of [S II] $\lambda\lambda 6718, 6733$ relative to the SFG population, consistent with a harder ionizing continuum in dense gas (Sect. 4.5). Their Ly α emitter fractions ($X_{\text{Ly}\alpha} \approx 50\text{--}67\%$) are systematically higher than both the SFG and other BLE populations at matched luminosity, pointing to low H I columns along these sight lines (Sect. 4.3).
3. *What spectral diagnostics provide the most effective constraints on the nature and hardness of the ionizing continuum, and can AGN accretion and stellar processes be distinguished using the available data?* High-ionization UV lines (C IV, N IV), He II and an elevated He II/H β ratio indicate a harder ionizing continuum than typical star-forming galaxies (Sect. 4.4). However, neither the rest-frame optical diagnostics (Sect. 4.2) nor the UV spectrum of GN-16813 (Sect. 6.1) uniquely distinguishes AGN accretion from an extremely young, massive stellar population. The absence of [Ne V] $\lambda 3426$ throughout the sample further supports an ionizing spectrum harder than normal star formation but softer than luminous Type-1 AGN (Sect. 6.1). We therefore adopt a geometry-based interpretation that remains agnostic to the nature of the central engine.
4. *What is the contribution of this population to the reionization photon budget and how sensitive is this estimate to the assumed powering mechanism, surrounding gas geometry, and structure?* SIROCCO modeling of an axisymmetric polar-outflow, equatorial-inflow geometry reproduces the spectral diversity from unobscured compact blue BLEs to heavily absorbed LRDs as a function of viewing angle alone (Sect. 5). From the full grid of 1229 models, we derive an empirical calibration between the LyC escape fraction and the observable Balmer break strength (Sect. 6.1.3, Eq. 2). Combining this calibration with the elevated ξ_{ion} predicted by our best-fit model, we find that compact blue BLEs contribute $\approx 4\%$ of the total ionizing photon budget at $z \sim 5$ (Sect. 6.2). While this falls well short of the $\geq 50\%$ contribution required for AGN-dominated reionization, it is non-negligible, particularly toward the tail end of reionization where the photon budget is most sensitive to small changes in the emissivity.

Several caveats apply to these results. Our $f_{\text{esc}}^{\text{LyC}}$ and ξ_{ion} estimates rely on a single physically motivated wind geometry, calibrated to the stacked spectra; while this geometry self-consistently reproduces the observed spectral diversity across viewing angle, alternative gas distributions or evolutionary histories cannot be excluded with the current data (see Begelman & Dexter 2025; Matthee et al. 2026). The ionizing budget calculation is also sensitive to the faint-end slope and completeness of the BLE UV luminosity function, which remains poorly constrained at $\log L_{\nu} \lesssim 28 \text{ erg s}^{-1} \text{ Hz}^{-1}$ (Sect. 6.2). Finally, the nature of the central ionizing source – whether AGN accretion, extreme massive stars, or some combination of the

two – remains uncertain, and resolving this question will require higher-resolution UV spectroscopy of the kind currently available only for GN-16813 (Sect. 4.4).

Our analysis is also limited to $4 \leq z \leq 7$ by the requirement of simultaneous Ly α -to-H α coverage with NIRSpec/PRISM (Sect. 3). This is unlikely to represent a physical upper redshift limit for the population: CANUCS-LRD-z8.6 (Tripodi et al. 2025), independently identified in the literature as a classical LRD, in fact satisfies our compact blue BLE selection criteria, and SPURS gdn_4762_33609 at $z = 7.18$ (Baccus & Xu 2025) shows the same absence of Ly α alongside a rich set of UV emission lines that characterizes our sample (Sect. 7). The existence of these sources beyond $z = 7$ suggests that the compact blue BLE population extends to the earliest stages of reionization, and that our results are likely to be representative of the population’s ionizing properties across the full EoR.

Finally, at these redshifts LyC is unobservable due to IGM absorption, so the $f_{\text{esc}}^{\text{LyC}}$ –Balmer break calibration of Sect. 6.1.3 is necessarily model-based rather than empirically calibrated: $f_{\text{esc}}^{\text{LyC}}$ cannot be directly measured for any source in our sample, compact blue BLE or LRD alike. Testing this calibration therefore requires LRD-like broad-line emitters at $z \lesssim 3$, where the LyC remains observable against the residual IGM opacity. Identifying such a sample, i.e. broad-line emitters spanning the same range of Balmer break strengths as our high- z population, with deep enough UV spectroscopy to either detect or place stringent limits on their LyC emission, would provide a critical empirical check on whether the Balmer break is, as our models predict, a reliable proxy for the ionizing escape fraction at the EoR itself.

Acknowledgements. SM thanks Nicola Mascia for useful discussions on efficient sample selection. Funded by the European Union (ERC, AGENTS, 101076224). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. This work is based on observations made with the NASA/ESA/CSA James Webb Space Telescope. The raw data were obtained from the Mikulski Archive for Space Telescopes 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. These observations are associated with programs 1180, 1181, 1208, 1212, 1213, 1215, 1216, 1286, 1345, 1433, 2198, 2561, 2750, 2767, 4106, 4233, 5105, 5224, 5664, 6368, 6585, 8018, and 8204. The authors acknowledge the observing teams that adopted a zero-exclusive-access period for several of the programs listed above, making those data public immediately. The data products presented herein were retrieved from the Dawn JWST Archive (DJA). DJA is an initiative of the Cosmic Dawn Center (DAWN), which is funded by the Danish National Research Foundation under grant DNRF140. This work was supported by the International Space Science Institute (ISSI) in Bern, through ISSI International Team project #25-659 ‘Little Red Dots, Big Open Questions’. The authors used Claude (Anthropic) to assist with code development for the data analysis pipelines. All AI-assisted content was reviewed, verified, and edited by the authors, who take full responsibility for the accuracy and integrity of this work.

References

- Ananna, T. T., Bogdán, Á., Kovács, O. E., Natarajan, P., & Hickox, R. C. 2024, *ApJ*, 969, L18
- Ando, M., Harikane, Y., Katz, H., Inayoshi, K., & Tanaka, T. S. 2026, *arXiv e-prints*, arXiv:2606.03522
- Atek, H., Chemerynska, I., Furtak, L. J., et al. 2026, *arXiv e-prints*, arXiv:2604.23823
- Baccus, C. & Xu, X. 2025, *arXiv e-prints*, arXiv:2512.03281
- Baggen, J. F. W., Scoggins, M. T., van Dokkum, P., et al. 2026, *ApJ*, 1002, L4
- Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, *PASP*, 93, 5
- Becker, G. D. & Bolton, J. S. 2013, *MNRAS*, 436, 1023
- Begelman, M. C. & Dexter, J. 2025, *arXiv e-prints*, arXiv:2507.09085
- Berg, D. A., Naidu, R. P., Chisholm, J., et al. 2026, *ApJ*, 1003, 112
- Bezanson, R., Labbe, I., Whitaker, K. E., et al. 2024, *ApJ*, 974, 92
- Bian, F., Kewley, L. J., Groves, B., & Dopita, M. A. 2020, *MNRAS*, 493, 580

- Bosman, S. E. I., Davies, F. B., Becker, G. D., et al. 2022, *MNRAS*, 514, 55
- Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2015, *ApJ*, 803, 34
- Bradley, L., Sipőcz, B., Robitaille, T., et al. 2024, *astropy/photutils*: 2.0.2
- Brammer, G. 2023a, *grizli*
- Brammer, G. 2023b, *msaexp: NIRS spec analysis tools*
- Brazzini, M., D'Eugenio, F., Maiolino, R., et al. 2026, *arXiv e-prints*, arXiv:2601.22214
- Bunker, A. J., Saxena, A., Cameron, A. J., et al. 2023, *A&A*, 677, A88
- Cameron, A. J., Saxena, A., Bunker, A. J., et al. 2023, *A&A*, 677, A115
- Chang, S.-J., Gronke, M., Matthee, J., & Mason, C. 2025, *arXiv e-prints*, arXiv:2508.08768
- Chang, S.-J., Gronke, M., Matthee, J., & Mason, C. 2026, *MNRAS*, 545, staf2131
- Chen, C.-H., Ho, L. C., Li, R., & Zhuang, M.-Y. 2025, *ApJ*, 983, 60
- Cleri, N. J., Olivier, G. M., Hutchison, T. A., et al. 2023, *ApJ*, 953, 10
- Covelo-Paz, A., Giovanazzo, E., Oesch, P. A., et al. 2024, *arXiv e-prints*, arXiv:2409.17241
- Cristiani, S., Serrano, L. M., Fontanot, F., Vanzella, E., & Monaco, P. 2016, *MNRAS*, 462, 2478
- Curtis-Lake, E., Cameron, A. J., Bunker, A. J., et al. 2025, *arXiv e-prints*, arXiv:2510.01033
- D'Amato, Q., Gilli, R., Vignali, C., et al. 2020, *A&A*, 636, A37
- De Barros, S., Pentericci, L., Vanzella, E., et al. 2017, *A&A*, 608, A123
- de Graaff, A., Brammer, G., Weibel, A., et al. 2025a, *A&A*, 697, A189
- de Graaff, A., Hviding, R. E., Naidu, R. P., et al. 2025b, *arXiv e-prints*, arXiv:2511.21820
- de Graaff, A., Rix, H.-W., Naidu, R. P., et al. 2025c, *A&A*, 701, A168
- Dong, X., Wang, T., Wang, J., et al. 2008, *MNRAS*, 383, 581
- Eilers, A.-C., Yue, M., Matthee, J., et al. 2025, *ApJ*, 991, L40
- Eisenstein, D. J., Willott, C., Alberts, S., et al. 2023, *arXiv e-prints*, arXiv:2306.02465
- Feltre, A., Charlot, S., & Gutkin, J. 2016, *MNRAS*, 456, 3354
- Finkelstein, S. L., Bagley, M. B., Arrabal Haro, P., et al. 2025, *ApJ*, 983, L4
- Finkelstein, S. L., D'Aloisio, A., Paardekoooper, J.-P., et al. 2019, *ApJ*, 879, 36
- Flury, S. R., Jaskot, A. E., Ferguson, H. C., et al. 2022, *ApJS*, 260, 1
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, 125, 306
- Furtak, L. J., Zitrin, A., Weaver, J. R., et al. 2023, *MNRAS*, 523, 4568
- Gehrels, N. 1986, *ApJ*, 303, 336
- Gentile, F., Giavalisco, M., Daddi, E., et al. 2026, *arXiv e-prints*, arXiv:2606.06575
- Geris, S., Maiolino, R., Ji, X., et al. 2026, *arXiv e-prints*, arXiv:2606.21614
- Giallongo, E., Grazian, A., Fiore, F., et al. 2019, *ApJ*, 884, 19
- Gilli, R., Vignali, C., Mignoli, M., et al. 2010, *A&A*, 519, A92
- Giovinazzo, E., Oesch, P. A., Weibel, A., et al. 2026, *A&A*, 707, A352
- Gräfenr, G. & Vink, J. S. 2015, *A&A*, 578, L2
- Grazian, A., Giallongo, E., Boutsia, K., et al. 2018, *A&A*, 613, A44
- Grazian, A., Giallongo, E., Fiore, F., et al. 2020, *ApJ*, 897, 94
- Greene, J. E., Labbe, I., Goulding, A. D., et al. 2024, *ApJ*, 964, 39
- Gupta, A., Trott, C. M., Jaiswar, R., et al. 2024, *ApJ*, 973, 169
- Gutkin, J., Charlot, S., & Bruzual, G. 2016, *Monthly Notices of the Royal Astronomical Society*, 462, 1757
- Harikane, Y., Zhang, Y., Nakajima, K., et al. 2023, *ApJ*, 959, 39
- Hayes, M. J., Saldana-Lopez, A., Citro, A., et al. 2025, *ApJ*, 982, 14
- Heintz, K. E., Watson, D., Brammer, G., et al. 2024, *Science*, 384, 890
- Hopkins, P. F., Richards, G. T., & Hernquist, L. 2007, *ApJ*, 654, 731
- Horne, K. 1986, *PASP*, 98, 609
- Hviding, R. E. 2025, *TheSkyentist/Unite: Version 0*, Zenodo
- Hviding, R. E., de Graaff, A., Miller, T. B., et al. 2025, *A&A*, 702, A57
- Inayoshi, K., Shangguan, J., Chen, X., Ho, L. C., & Haiman, Z. 2025, *arXiv e-prints*, arXiv:2505.05322
- Inoue, A. K., Shimizu, I., Iwata, I., & Tanaka, M. 2014, *MNRAS*, 442, 1805
- Ishikawa, Y., Eilers, A.-C., Naidu, R. P., et al. 2026, *arXiv e-prints*, arXiv:2607.09647
- Izotov, Y. I., Schaerer, D., Worseck, G., et al. 2018, *MNRAS*, 474, 4514
- Ji, X., Maiolino, R., Übler, H., et al. 2025, *MNRAS*, 544, 3900
- Ji, X., Pezzulli, G., D'Eugenio, F., et al. 2026, *arXiv e-prints*, arXiv:2604.03370
- Jin, X., Yang, J., Wang, F., et al. 2026, *arXiv e-prints*, arXiv:2606.14959
- Jones, G. C., Bunker, A. J., Saxena, A., et al. 2024, *A&A*, 683, A238
- Juodžbalis, I., Ji, X., D'Eugenio, F., et al. 2026, *arXiv e-prints*, arXiv:2607.19057
- Juodžbalis, I., Maiolino, R., Baker, W. M., et al. 2025, *arXiv e-prints*, arXiv:2504.03551
- Kakiichi, K., Jin, X., Wang, F., et al. 2025, *arXiv e-prints*, arXiv:2503.07074
- Kashino, D., Lilly, S. J., Matthee, J., et al. 2023, *ApJ*, 950, 66
- Kashino, D., Lilly, S. J., Matthee, J., et al. 2026, *ApJ*, 997, 280
- Keerthi Vasan, G., C., Senchyna, P., Mason, C. A., et al. 2026, *arXiv e-prints*, arXiv:2606.06609
- Kocevski, D. D., Finkelstein, S. L., Barro, G., et al. 2025, *ApJ*, 986, 126
- Kocevski, D. D., Onoue, M., Inayoshi, K., et al. 2023, *ApJ*, 954, L4
- Kokorev, V., Caputi, K. I., Greene, J. E., et al. 2024, *ApJ*, 968, 38
- Kokorev, V., Fujimoto, S., Labbe, I., et al. 2023, *ApJ*, 957, L7
- Kotiwale, G., Matthee, J., Kashino, D., et al. 2026, *A&A*, 706, A165
- Kulkarni, G., Worseck, G., & Hennawi, J. F. 2019, *MNRAS*, 488, 1035
- Labbe, I., Greene, J. E., Bezanson, R., et al. 2025, *ApJ*, 978, 92
- Lin, X., Fan, X., Cai, Z., et al. 2026, *ApJ*, 997, 364
- Liu, Z., Naidu, R. P., Secunda, A., et al. 2026, *arXiv e-prints*, arXiv:2604.13000
- Llerena, M., Pentericci, L., Napolitano, L., et al. 2025, *A&A*, 698, A302
- Madau, P., Giallongo, E., Grazian, A., & Haardt, F. 2024, *ApJ*, 971, 75
- Madau, P. & Haardt, F. 2015, *ApJ*, 813, L8
- Maiolino, R., Scholtz, J., Curtis-Lake, E., et al. 2024a, *A&A*, 691, A145
- Maiolino, R., Scholtz, J., Curtis-Lake, E., et al. 2024b, *A&A*, 691, A145
- Marques-Chaves, R., Álvarez-Márquez, J., Colina, L., et al. 2026a, *arXiv e-prints*, arXiv:2602.02322
- Marques-Chaves, R., Martins, F., Schaerer, D., Dessauges-Zavadsky, M., & Palacios, A. 2026b, *arXiv e-prints*, arXiv:2604.21493
- Marques-Chaves, R., Schaerer, D., Kuruvanthodi, A., et al. 2024, *A&A*, 681, A30
- Mascia, S., Pentericci, L., Calabrò, A., et al. 2024, *A&A*, 685, A3
- Maseda, M. V., de Graaff, A., Franx, M., et al. 2024, *A&A*, 689, A73
- Mason, C. A. & Gronke, M. 2020, *MNRAS*, 499, 1395
- Matthee, J., Naidu, R. P., Brammer, G., et al. 2024, *ApJ*, 963, 129
- Matthee, J., Naidu, R. P., Kotiwale, G., et al. 2025, *ApJ*, 988, 246
- Matthee, J., Naidu, R. P., Pezzulli, G., et al. 2022, *MNRAS*, 512, 5960
- Matthee, J., Sobral, D., Gronke, M., et al. 2018, *A&A*, 619, A136
- Matthee, J., Torralba, A., Pezzulli, G., et al. 2026, *arXiv e-prints*, arXiv:2603.17667
- Matthews, J. H., Long, K. S., Knigge, C., et al. 2025, *MNRAS*, 536, 879
- Mazzolari, G., Gilli, R., Maiolino, R., et al. 2024, *arXiv e-prints*, arXiv:2412.04224
- Mignoli, M., Vignali, C., Gilli, R., et al. 2013, *A&A*, 556, A29
- Naidu, R. P., Forrest, B., Oesch, P. A., Tran, K.-V. H., & Holden, B. P. 2018, *MNRAS*, 478, 791
- Naidu, R. P., Matthee, J., Katz, H., et al. 2025, *arXiv e-prints*, arXiv:2503.16596
- Naidu, R. P., Matthee, J., Kramarenko, I., et al. 2024, *arXiv e-prints*, arXiv:2410.01874
- Nakajima, K., Schaerer, D., Le Fèvre, O., et al. 2018, *A&A*, 612, A94
- Napolitano, L., Pentericci, L., Santini, P., et al. 2024, *A&A*, 688, A106
- Park, K., Torralba, A., Matthee, J., et al. 2026, *arXiv e-prints*, arXiv:2605.14233
- Parkinson, E. J., Knigge, C., Dai, L., et al. 2025, *MNRAS*, 540, 3069
- Pérez-González, P. G., Barro, G., Carniani, S., et al. 2026, *arXiv e-prints*, arXiv:2602.20247
- Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, *A&A*, 641, A6
- Pollock, C. L., Gottumukkala, R., Heintz, K. E., et al. 2025, *arXiv e-prints*, arXiv:2506.15779
- Price, S. H., Bezanson, R., Labbe, I., et al. 2025, *ApJ*, 982, 51
- Prieto-Lyon, G., Strait, V., Mason, C. A., et al. 2023, *A&A*, 672, A186
- Rinaldi, P., Bonaventura, N., Rieke, G. H., et al. 2025, *ApJ*, 992, 71
- Rivera-Thorsen, T. E., Dahle, H., Gronke, M., et al. 2017, *Astronomy and Astrophysics*, 608, L4
- Robertson, B. E., Ellis, R. S., Furlanetto, S. R., & Dunlop, J. S. 2015, *ApJ*, 802, L19
- Romano, M., Grazian, A., Giallongo, E., et al. 2019, *A&A*, 632, A45
- Saldana-Lopez, A., Schaerer, D., Chisholm, J., et al. 2022, *A&A*, 663, A59
- Sarrouh, G. T. E., Asada, Y., Martis, N. S., et al. 2025, *arXiv e-prints*, arXiv:2506.21685
- Scarlata, C., Hayes, M., Panagia, N., et al. 2024, *arXiv e-prints*, arXiv:2404.09015
- Scholtz, J., Maiolino, R., D'Eugenio, F., et al. 2025, *A&A*, 697, A175
- Selsing, J., Fynbo, J. P. U., Christensen, L., & Krogager, J.-K. 2016, *A&A*, 585, A87
- Setton, D. J., Greene, J. E., de Graaff, A., et al. 2025, *ApJ*, 995, 118
- Sharma, M., Theuns, T., Frenk, C., et al. 2017, *MNRAS*, 468, 2176
- Shen, Y., Zhuang, M.-Y., Li, J., et al. 2024, *arXiv e-prints*, arXiv:2408.12713
- Shirazi, M. & Brinchmann, J. 2012, *MNRAS*, 421, 1043
- Simmonds, C., Tacchella, S., Hainline, K., et al. 2024, *MNRAS*, 535, 2998
- Sneppen, A., Watson, D., Matthews, J. H., & Nikopoulos, G. 2026a, *arXiv e-prints*, arXiv:2606.12509
- Sneppen, A., Watson, D., Matthews, J. H., et al. 2026b, *arXiv e-prints*, arXiv:2601.18864
- Stark, D. P., Ellis, R. S., & Ouchi, M. 2011, *ApJ*, 728, L2
- Stark, D. P., Topping, M. W., Endsley, R., & Tang, M. 2026, in *Encyclopedia of Astrophysics*, Volume 4, Vol. 4, 453–499
- Steidel, C. C., Bogosavljević, M., Shapley, A. E., et al. 2018, *ApJ*, 869, 123
- Tang, M., Stark, D. P., Mason, C. A., et al. 2026, *arXiv e-prints*, arXiv:2604.03563
- Torralba, A., Matthee, J., Pezzulli, G., et al. 2026a, *A&A*, 707, A75
- Torralba, A., Matthee, J., Pezzulli, G., et al. 2026b, *A&A*, 705, A147
- Tripodi, R., Martis, N., Markov, V., et al. 2025, *Nature Communications*, 16, 9830
- Übler, H., Maiolino, R., Curtis-Lake, E., et al. 2023, *A&A*, 677, A145
- Valentino, F., Brammer, G., Gould, K. M. L., et al. 2023, *ApJ*, 947, 20
- Valentino, F., Heintz, K. E., Brammer, G., et al. 2025, *A&A*, 699, A358
- Vanden Berk, D. E., Richards, G. T., Bauer, A., et al. 2001, *AJ*, 122, 549
- Vanzella, E., Grazian, A., Hayes, M., et al. 2010, *A&A*, 513, A20
- 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
- Wang, B., de Graaff, A., Davies, R. L., et al. 2025, *ApJ*, 984, 121
- Witsok, J., Jakobsen, P., Maiolino, R., et al. 2025, *Nature*, 639, 897
- Yue, M., Eilers, A.-C., Ananna, T. T., et al. 2024, *ApJ*, 974, L26
- Zhang, Y., Morishita, T., & Stuvelli, M. 2026, *ApJ*, 998, 141
- Zhang, Z., Jiang, L., Liu, W., & Ho, L. C. 2025, *ApJ*, 985, 119
- Zhu, P., Trussler, J., & Kewley, L. J. 2026, *ApJ*, 998, 5

Appendix A: Total sample

Table A.1 summarizes the JWST/NIRSpec observing programs and medium- and high-resolution grating coverage available for each of the 20 compact blue BLEs in our sample. Only 12 of the 20 sources have additional coverage with at least one medium- or high-resolution grating, and no source in the sample has the complete combination of all five grating configurations listed. Coverage of the bluest grating, G140M/F100LP, which probes the rest-frame UV at the redshifts of our sample, is available for only 5 sources; of these, GN-16813 is the only source with sufficient depth and wavelength coverage to enable the detailed UV line analysis presented in Sect. 4.4. The remaining 8 sources in the sample have PRISM-only coverage, and their spectral properties are therefore constrained exclusively by the low-resolution analysis of Sect. 4. Fig. A.1 shows the NIRCам stamps and NIRSpec/PRISM spectra for the 14 sources not already shown in Fig. 3.

Appendix B: SIROCCO

In addition to the single best-fit model discussed in Sect. 5, we explored a broad grid of 1229 SIROCCO models to test the robustness of the viewing-angle interpretation and to derive the empirical $f_{\text{esc}}^{\text{LyC}}$ –Balmer break calibration of Sect. 6.1.3. All models share the same underlying two-component geometry described in Sect. 5.1 but span a wide range of central source properties, wind kinematics, geometry, and computational setup. Table B.1 lists the parameters set for the best model presented in Sect. 5. Table B.2 lists the varied parameters across the entire grid. The grid was constructed to broadly cover the parameter space expected for compact, dusty/gas-enriched AGN-like sources at sub-parsec to parsec scales, with central black hole masses spanning 10^6 – $10^9 M_{\odot}$, bolometric luminosities spanning 10^{43} – $10^{45} \text{ erg s}^{-1}$, and wind mass-loss rates spanning nearly four orders of magnitude. For each model, spectra were extracted at up to 20 discrete viewing angles spanning $\theta = 0$ – 90° (pole-on to edge-on), enabling us to map the angle dependence of both the Balmer break and $f_{\text{esc}}^{\text{LyC}}$ self-consistently within each model, as described in Sect. 6.1.3.

Appendix C: UV lines in GN-16813

GN-16813 is the only source in our sample with deep medium-resolution NIRSpec grating spectroscopy in the rest-frame UV. We fit N IV] + C IV, He II + O III], N III], and C III] (Fig. C.1) with *unite* (see Sec. 4.4, Hviding 2025). The two panels on the left additionally show the first and second spectral orders of Ly α . The resulting line fluxes, rest-frame equivalent widths, and diagnostic ratios are listed in Table C.1. Uncertainties correspond to the 16th–84th percentile range of the MCMC posteriors, and the fluxes are not corrected for extinction; doublet equivalent widths are the summed contribution of both components.

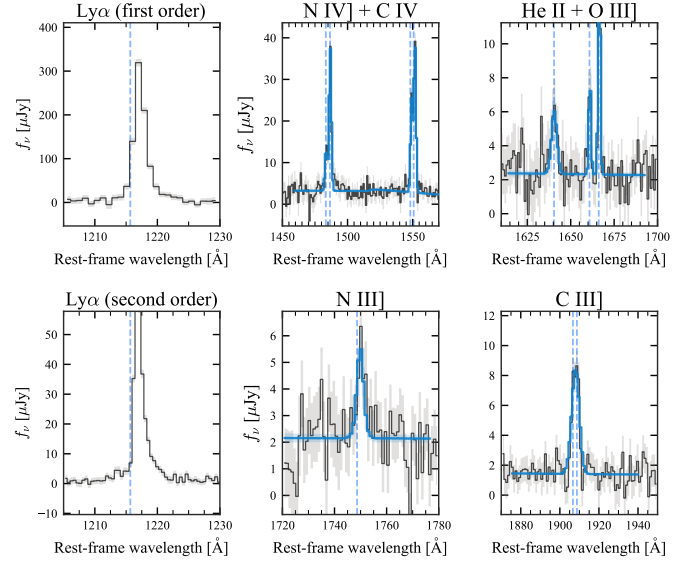


Fig. C.1: Rest-UV spectroscopy of GN-16813. *unite* best-fit line profiles (solid curves) for the four spectral windows covering N IV] + C IV, He II + O III], N III], and C III], overlaid on the continuum-subtracted grating spectrum. On the left panels, first and second order of Ly α are presented.

Table C.1: UV emission-line measurements and diagnostic ratios for GN-16813.

Line	F [$10^{-18} \text{ erg s}^{-1} \text{ cm}^{-2}$]	EW_0 [Å]
C IV] $\lambda\lambda 1548, 1551$	$6.34^{+0.29}_{-0.31}$	43^{+5}_{-4}
He II $\lambda 1640$	$0.98^{+0.23}_{-0.22}$	7^{+1}_{-2}
N IV] $\lambda\lambda 1483, 1486$	$6.30^{+0.32}_{-0.33}$	31^{+3}_{-3}
N III] $\lambda 1748$	$0.73^{+0.19}_{-0.19}$	5^{+2}_{-1}
O III] $\lambda\lambda 1661, 1666$	$2.06^{+0.23}_{-0.20}$	14^{+2}_{-2}
C III] $\lambda\lambda 1907, 1909$	$2.24^{+0.16}_{-0.16}$	25^{+2}_{-2}

Ratio	Value
C IV]/He II	$6.47^{+1.48}_{-1.55}$
C III]/He II	$2.29^{+0.54}_{-0.56}$
N IV]/He II	$6.43^{+1.48}_{-1.55}$
O III]/He II	$2.10^{+0.53}_{-0.53}$
C IV]/C III]	$2.83^{+0.24}_{-0.25}$
N IV]/C III]	$2.81^{+0.25}_{-0.25}$
N III]/C III]	$0.33^{+0.09}_{-0.09}$
N IV]/C IV]	$0.99^{+0.07}_{-0.07}$

Notes. Fluxes are not corrected for extinction. Doublet EWs are the sum of the two components. Asymmetric uncertainties correspond to the 16th–84th percentile range of the MCMC posteriors.

Table A.1: Summary of the data available for the 20 compact blue broad-line emitters in our sample.

Source	Program(s)	G140M/F100LP	G235M/F170LP	G395M/F290LP	G235H/F170LP	G395H/F290LP
jades-gdn09_1181_73488	1181	✓	✓	✓	✗	✓
abell2744-green_8204_69688	8204	✗	✓	✓	✗	✗
capers-egs47_6368_19300	6368	✗	✗	✗	✗	✗
nexus-obs3_5105_10835	5105	✗	✗	✗	✗	✗
rubies-egs62_4233_42232	4233	✗	✗	✓	✗	✗
uncover_2561_38108	2561	✗	✗	✗	✗	✗
ceers-ddt_2750_1768	2750	✗	✗	✗	✗	✗
rubies-egs53_4233_50052	4233	✗	✗	✓	✗	✗
capers-egs61_6368_20952	6368	✗	✗	✗	✗	✗
GN-16813	5664, 8018	✓	✗	✗	✗	✓
GS_3073	1216	✗	✗	✗	✗	✓
rubies-uds23_4233_172350	4233	✗	✗	✓	✗	✗
jades-gdn11_1181_1093	1181	✓	✓	✓	✗	✓
macs1423_1208_4103248	1208	✗	✗	✗	✗	✗
jades-gdn198_1181_38147	1181	✓	✓	✓	✗	✓
gto-wide-uds12_1215_1259	1213	✗	✗	✗	✓	✓
rubies-uds32_4233_966096	4233	✗	✗	✓	✗	✗
capers-cos07_6368_35805	6368	✗	✗	✗	✗	✗
jades-gdn2_1181_38509	1181	✓	✓	✓	✗	✓
uncover_2561_11254	2561	✗	✗	✗	✗	✗

Table B.1: Parameters of the best Sirocco model presented in this work.

Category	Parameter	Value
<i>Central source</i>	Mass	$M_{\text{BH}} = 10^7 M_{\odot}$
	Bolometric luminosity	$L = 10^{44} \text{ erg s}^{-1}$
	Ionizing spectrum	Blackbody, $T_{\text{BB}} = 5.0 \times 10^4 \text{ K}$
	Schwarzschild radius	$R_{\text{Schw}} = 2.95 \times 10^{13} \text{ cm}$
	Disk	None
<i>Gas envelope</i>	Coordinate system	Polar (axisymmetric)
	Boundary angle	$\theta_b = 70.5^\circ$
	Boundary taper	3°
	Density profile	$\rho \propto r^{-2} [-1, -3]$
	Initial temperature	$T_{\text{init}} = 10^3 \text{ K}$
	Metallicity	$Z = 0.01 Z_{\odot}$
	Filling factor	$f = 0.01$ (clumped)
<i>Kinematics</i>	Polar outflow velocity	$v_{\text{out}} = 700 \text{ km s}^{-1}$
	VRATIO (η)	0.01
	Density contrast	$\rho_{\text{in}}/\rho_{\text{out}} = 100$
	Rotation	Keplerian ($v_{\phi} = \sqrt{GM/r} \sin \theta$)

Table B.2: Parameters varied across the grid of 1229 Sirocco models.

Category	Parameter	Range
<i>Central source</i>	Blackbody temperature	$2.0 \times 10^4 - 1.585 \times 10^5 \text{ K}$
	Bolometric luminosity	$10^{43} - 10^{45} \text{ erg s}^{-1}$
	Black hole mass	$10^6 - 10^9 M_{\odot}$
	Radius	$2.95 \times 10^{12} - 8.857 \times 10^{14} \text{ cm}$
<i>Stellar wind component</i>	Acceleration exponent	1–2
	$\dot{M}$	$0.011 - 0.702 M_{\odot} \text{ yr}^{-1}$
	v_{∞}	$10^8 - 5 \times 10^8 \text{ cm s}^{-1}$
	v_{base}	$10^6 - 5 \times 10^7 \text{ cm s}^{-1}$
<i>Wind grid and material</i>	Filling factor	0.001–1
	$\dot{M}_{\text{wind}}$	$8.67 \times 10^{-4} - 9.584 M_{\odot} \text{ yr}^{-1}$
	Metallicity	0.01–1 $Z_{\odot}$
	R_{max}	$10^{16} - 10^{19} \text{ cm}$
	R_{min}	$10^{15} - 3 \times 10^{16} \text{ cm}$
	Initial temperature	$10^3 - 10^5 \text{ K}$

Notes. Number of unique discrete values sampled for each parameter is given in parentheses. Each grid point was run with the full set of viewing angles listed in Table B.1, subject to the active viewing-angle configuration for that run.

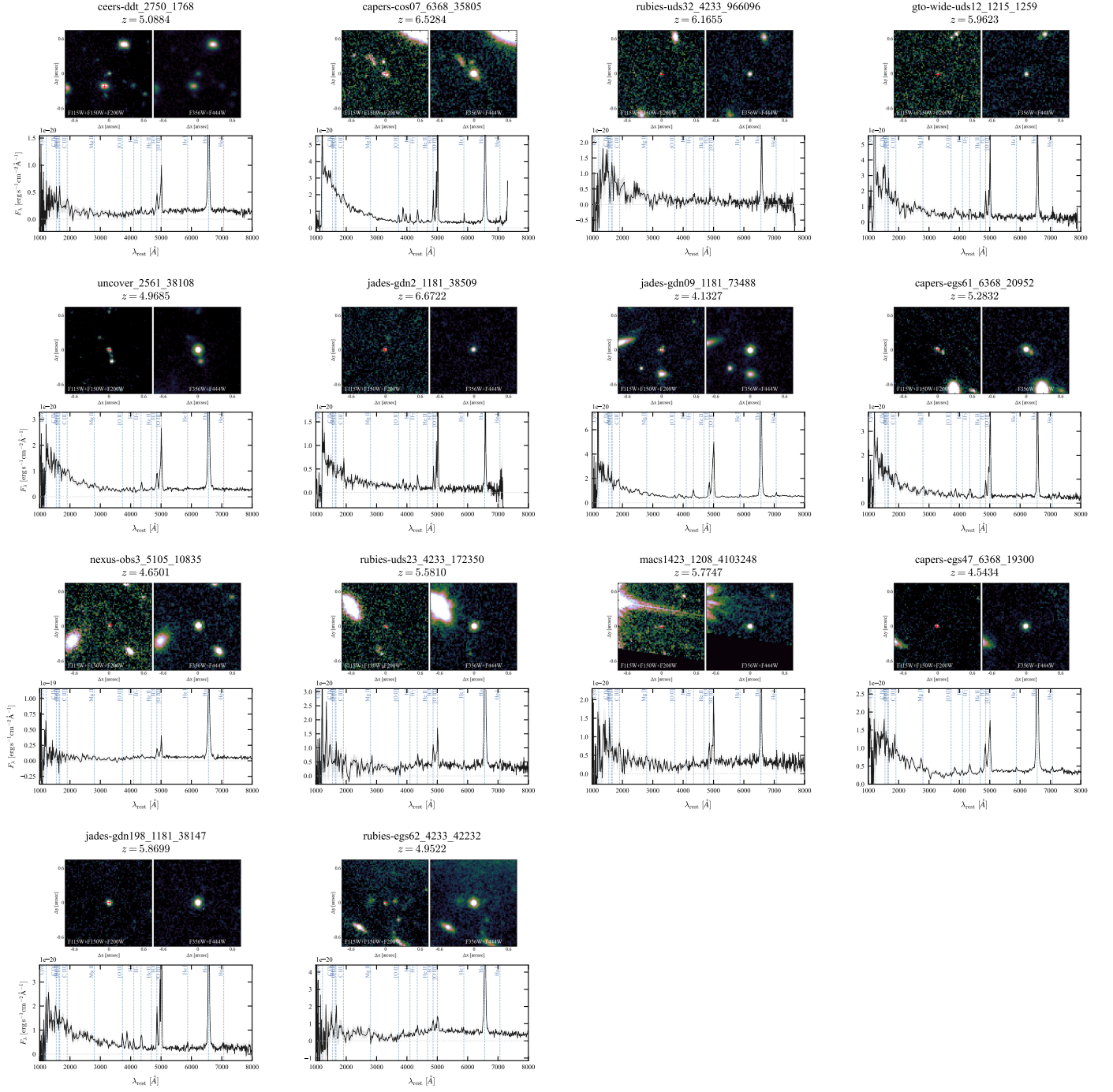


Fig. A.1: NIRCcam stamps and NIRSpec/PRISM spectra for the remaining 14 compact blue BLs. For each source, the upper panel shows NIRCcam composites in the rest-frame UV (F115W+F150W+F200W, left) and optical (F356W+F444W, right), with the optical centroid marked by a red cross on the UV stamp. The lower panel shows the PRISM spectrum (black) with the 1 σ uncertainty band (grey). Sources are ordered by increasing B_{UV} , with the redshift indicated above each stamp.