JWST/NIRSpec measurements of extremely low metallicities in high equivalent width Lyα emitters

Maseda, M.V.; Lewis, Z.; Matthee, J.; Hennawi, J.F.; Boogaard, L.; Feltre, A.; ... ; Vitte, E.

JWST/NIRSpec Measurements of Extremely Low Metallicities in High Equivalent Width Lyα Emitters

Michael V. Maseda[1], Zach Lewis[1], Jorryt Matthee[2], Joseph F. Hennawi[3,4], Leindert Boogaard[5], Anna Feltre[6], Themiya Nanayakkara[7], Roland Bacon[8], Amy Barger[1,9,10], Jarle Brinchmann[11], Marijn Franx[4], Takuya Hashimoto[12,13], Hanae Inami[14], Haruka Kusakabe[15], Floriane Leclercq[16], Lucie Rowland[4], Anthony J. Taylor[16], Christy Tremonti[1], Tanya Urrutia[17], Joop Schaye[4], Charlotte Simmonds[18,19], and Eloïse Vitte[15,20]

[1] Department of Astronomy, University of Wisconsin-Madison, 475 N. Charter St., Madison, WI 53706, USA; maseda@astro.wisc.edu
[2] ETH Zürich, Department of Physics, Wolfgang-Pauli-Str. 27, 8093 Zürich, Switzerland
[3] Department of Physics, University of California, Broida Hall, Santa Barbara, Santa Barbara, CA 93106-9530, USA
[4] Leiden Observatory, Leiden University, P.O. Box 9513, 2300 RA, Leiden, The Netherlands
[5] Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany
[6] INAF—Osservatorio Astrofisico di Arcetri, Largo Enrico Fermi 5, I-50125 Firenze, Italy
[7] Centre for Astrophysics and Supercomputing, Swinburne University of Technology, P.O. Box 218, Hawthorn, 3122, VIC, Australia
[8] Univ Lyon, Univ Lyon1, ENS de Lyon, CNRS, Centre de Recherche Astrophysique de Lyon UMR5574, F-69230, Saint-Genis-Laval, France
[9] Department of Physics and Astronomy, University of Hawaii, 2505 Correa Rd., Honolulu, HI 96822, USA
[10] Institute for Astronomy, University of Hawaii, 2680 Woodlawn Dr., Honolulu, HI 96822, USA
[11] Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto, CAUP, Rua das Estrelas, PT4150-762 Porto, Portugal
[12] Division of Physics, Faculty of Pure and Applied Sciences, University of Tsukuba,Tsukuba, Ibaraki 305-8571, Japan
[13] Tomonaga Center for the History of the Universe (TCHoU), Faculty of Pure and Applied Sciences, University of Tsukuba, Tsukuba, Ibaraki 305-8571, Japan
[14] Hiroshima Astrophysical Science Center, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526, Japan
[15] Observatoire de Genève, Université de Genève, 51 Ch. des Maillettes, 1290 Versoix, Switzerland
[16] Department of Astronomy, University of Texas at Austin, 2515 Speedway Blvd. Stop C1400, Austin, TX 78712, USA
[17] Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, D-14482 Potsdam, Germany
[18] The Kavli Institute for Cosmology (KICC), University of Cambridge, Madingley Rd., Cambridge, CB3 0HA, UK
[19] Cavendish Laboratory, University of Cambridge, 19 JJ Thomson Ave., Cambridge, CB3 0HE, UK
[20] European Southern Observatory, Alonso de Cordova 3107, Vitacura, Santiago, Chile

Abstract

Deep Very Large Telescope/MUSE optical integral field spectroscopy has recently revealed an abundant population of ultra-faint galaxies ($M_{\rm UV} \approx -15$; $0.01\,L_\star$) at $z = 2.9-6.7$ due to their strong Lyα emission with no detectable continuum. The implied Lyα equivalent widths can be in excess of 100–200 Å, challenging existing models of normal star formation and indicating extremely young ages, small stellar masses, and a very low amount of metal enrichment. We use JWST/NIRSpec's microshutter array to follow up 45 of these galaxies (11 hr in G235M/F170LP and 7 hr in G395M/F290LP), as well as 45 lower-equivalent width Lyα emitters. Our spectroscopy covers the range $1.7-5.1$ micron in order to target strong optical emission lines: Hα, [O III], Hβ, and [N II]. Individual measurements as well as stacks reveal line ratios consistent with a metal-poor nature (2%–40% $Z_\odot$, depending on the calibration). The galaxies with the highest equivalent widths of Lyα, in excess of 90 Å, have lower [N II]/Hα (1.9σ) and [O III]/Hβ (2.2σ) ratios than those with lower equivalent widths, implying lower gas-phase metallicities at a combined significance of 2.4σ. This implies a selection based on Lyα equivalent width is an efficient technique for identifying younger, less chemically enriched systems.

Unified Astronomy Thesaurus concepts: Galaxy abundances (574); High-redshift galaxies (734); Galaxy evolution (594); Galaxy chemical evolution (580)

1. Introduction

Observing primordial galaxies has long been a goal in studies of the distant Universe. However, the identification and subsequent follow-up of these sources present unique challenges. Deep imaging surveys have efficiently identified galaxies at high-z based on their rest–UV continuum, yet practical considerations due to the finite depth of the imaging mean that the very youngest sources remain undetected until they have built up sufficient stellar mass (and hence also metals). Before this point, though, they would appear extremely luminous in Lyα $\lambda1216$ emission as their young,

metal-poor stellar populations would be efficient producers of ionizing photons (Partridge & Peebles 1967). Indeed, theoretical predictions for metal-poor or metal-free stellar populations suggest that the strength of Lyα with respect to the UV continuum (i.e., the equivalent width; EW) would be far in excess of what can be observed in metal-enriched systems, with EW $\approx 200-240$ Å demarcating these populations (Charlot & Fall 1991; Schaerer 2003). Identifying sources with Lyα EWs in excess of 200 Å is a direct way to isolate galaxies before they have experienced significant metal enrichment (Malhotra & Rhoads 2002; Raiter et al. 2010; Zackrisson et al. 2011).

The MUSE integral field spectrograph on the Very Large Telescope (VLT; Bacon et al. 2010) has opened a new parameter space for these searches: even in areas covered by the deepest optical/near-IR imaging in the Hubble Ultra Deep Field (UDF; $m_{\rm AB} \approx 30$; Illingworth et al. 2013; Rafelski et al. 2015), $10-30$ hr

MUSE spectroscopy has identified numerous emission line galaxies without a detected continuum counterpart (Bacon et al. 2017; Hashimoto et al. 2017). In particular, MUSE has uncovered an abundant population of $z = 2.9$–6.7 high-EW Lyα emitters (LAEs) that are the faintest unlensed spectroscopically confirmed galaxies ever detected at high redshift. Detailed stacking analyses reveal the average UV continua, which are as faint as $M_{\rm UV} = -15$ (or 0.01 $L_\star$ at $z = 4.5$; Maseda et al. 2018). These LAEs, which represent 20% of all detected LAEs in the field and 10% of all expected galaxies at this $M_{\rm UV}$ (Bouwens et al. 2014; Inami et al. 2017), must have high EWs with some estimates exceeding the canonical $200-240$ Å (rest frame).

Maseda et al. (2020) stack 200 hr deep Spitzer/IRAC photometry for 35 of these sources at $z = 4$–5, finding evidence for high-EW Hα emission, indicating rapid stellar mass buildup. They show that these sources are among the most efficient producers of ionizing photons ever discovered (see also Harikane et al. 2018; Simmonds et al. 2023), and they are likely to be younger than 3 Myr and have gas-phase metallicities of 3%–30% Z$_\odot$. However, these quantities are uncertain and constrained indirectly from the broadband Hubble Space Telescope (HST) and Spitzer imaging data.

With the launch of JWST, rest-frame optical emission lines such as [O III], Hβ, and Hα have become directly observable beyond $z \approx 2$–3, the limit for ground-based near-IR spectrographs. Here, we target some of the highest-EW LAEs with JWST/NIRSpec in order to detect these rest-frame optical emission lines. The observed strengths and ratios of emission lines can be used to determine numerous physical quantities in these galaxies such as their star formation rates (SFRs), gas-phase metallic abundances, and ionization properties, as has already been shown with cycle 1 data from JWST (e.g., Matthee et al. 2023; Roberts-Borsani et al. 2023; Schaerer et al. 2022; Taylor et al. 2022; Wang et al. 2022; Brinchmann 2023; Cameron et al. 2023; Fujimoto et al. 2023; Nakajima et al. 2023; Sanders et al. 2023a). In this first paper, we focus on the ratio of [O III] $\lambda5007$ to Hβ $\lambda4861$, with supplementary information provided by the ratio of [N II] $\lambda6584$ to Hα $\lambda6563$. These ratios use closely spaced lines, providing resilience against the effects of dust attenuation and/or poorly flux-calibrated data. Moreover, they can provide insight into the metal abundance of the gas in these galaxies, which are expected to be extremely low in the highest Lyα-EW systems.

The outline of this paper is as follows. In Section 2, we present our sample and our JWST/NIRSpec observations. In Section 3, we present the results of our spectroscopy, including emission line ratios from individual objects as well as stacks based on the EW of Lyα. In Section 4, we discuss the implications of these results, make concluding remarks, and present prospects for future work with NIRSpec.

2. Data

2.1. Target Selection

Our targets are selected from the MUSE HUDF survey (Bacon et al. 2023) and the MUSE Wide survey (Herenz et al. 2017; Urrutia et al. 2019). We preferentially target LAEs with plausibly high Lyα EWs, via bright Lyα emission and/or faint measurements or limits on the rest–UV continuum magnitudes from the HST imaging (e.g., Hashimoto et al. 2017; Maseda et al. 2018, 2020; Kerutt et al. 2022).

We prioritize sources based on their Lyα EW and the likelihood that their rest-optical emission lines (Hα, [O III], and/or Hβ) are measurable with NIRSpec in the combination of G235M and G395M (see Section 2.2). Specifically, we perform a first mask design with the NIRSpec mask planning tool (MPT; Karakla et al. 2014) at our assigned roll angle with the highest-EW sources only. We then calculate the predicted wavelengths covered for each object and downweight cases where the detector gap and/or the red edge of the NRS2 detector restricts the observability of these emission lines, based on the MUSE Lyα-based redshifts.

We allow for spectra to overlap on the detector to maximize the multiplexing, with the understanding that this can lead to potential confusion in the 2D frames. As our sources are emission line dominated (Maseda et al. 2020), are spatially compact (Maseda et al. 2018), and have known redshifts from MUSE spectroscopy, we can always disentangle the spectra on the detector provided the sources themselves are not over-lapping in the cross-dispersion direction. We impose a limit of two additional spectra overlapping with the spectra of any of our primary targets at any one position on the detector to minimize unnecessarily high background levels (a detailed description for how spectral overlap can be calculated as a function of shutter position will be given by N. Bonaventura et al. 2023, in preparation).

Our final microshutter array (MSA) design consists of 1×3 microshutters targeting 125 unique galaxies. Of the 125, 64 are selected from the MUSE UDF Data Release 2 catalog, 26 are from the MUSE Wide Survey, and 35 are based on grism data from the 3D-HST survey (Momcheva et al. 2016). For the following analysis, we only consider the galaxies that were selected based on the presence of Lyα in their spectrum, i.e., the 90 galaxies selected from MUSE. These galaxies have redshifts between 2.9 and 6.5, with a median redshift of 4.5.

2.2. NIRSpec Observations

Our MSA configuration was observed on 11 September 2022 (GO 1671). Target acquisition was performed with MSATA using eight reference objects in three quadrants, the CLEAR filter, and NRSRAPID readout. The HUDF does not contain a sufficient number of stars to use for the target acquisition process, so we supplement the reference target list with compact galaxies. We utilize the "stellarity" measurements from the 3D-HST photometric catalog (Skelton et al. 2014) to pre-select compact objects. We then estimate their magnitudes in the NIRSpec imaging filters (CLEAR, F110W, F140X) by fitting the 3D-HST photometry, including Spitzer/IRAC 3.6 and 4.5 micron data, using MAGPHYS (da Cunha et al. 2008) at the best-fit grism redshift (Momcheva et al. 2016). The MPT selects the eight objects from this larger list for the target acquisition.

The MSA configuration was observed using a three-point nodding pattern. Each nodded exposure consisted of three integrations: 20 groups per integration with NRSIRS2 in G235M/F170LP and 100 groups per integration with NRSIRS2RAPID in G395M/F290LP.[21] With three total nodding sequences in G235M/F170LP, we obtain a total of 39,783.9 s of exposure time; with two total nodding sequences in G395M/F290LP, we obtain a total of 26,522.6 s of exposure

---

[21] The different readout patterns result in different data rates, the limitations of which prevent us from observing both in the preferred NRSIRS2RAPID.

time. Only one of our primary MUSE-selected targets was affected by an inoperable microshutter (see Appendix A).

The NIRSpec data was reduced using level 1 and level 2 of the version 1.10.2 NIRSpec pipeline and the `jwst_1097.pmap` context (Bushouse et al. 2022). We perform an additional $1/f$ noise correction and a median bias correction using `msaexp`.[22] The calibrated, unrectified 2D frames are combined using `PypeIt`[23] (Prochaska et al. 2020a, 2020b). The sky background, bar shadows, and detector-level artifacts are removed with a pixel-level background subtraction based on adjacent exposures in the nodding sequence.

We note that, in this version of the NIRSpec pipeline, the errors are systematically underestimated compared to the measured pixel-to-pixel variation in the fluxes (as seen in, e.g., Brinchmann 2023; Trump et al. 2023). We observe this already in the level 1 rate files, where the observed pixel-to-pixel flux variations in two consecutive exposures are larger than the quoted uncertainties: in the case of G235M (NRSIRS2 readout), this is a factor of 1.91, and in the case of G395M (NRSIRS2RAPID readout), this is a factor of 1.70. These discrepancies can be caused by a number of issues that are currently not being accounted for in the pipeline, such as thermal instabilities or inter-pixel capacitance. For what follows, we include these empirical factors in all of our uncertainties; a more detailed discussion can be found in Appendix B. While this has a small impact on what we consider to be a detection, there is no evidence that this would have a systematic impact on the stacked line ratios that are the main result of this paper.

2.3. Spectral Extractions and Redshift Determinations

Our observed sample consists of 90 LAEs selected from MUSE spectroscopy. Although the peak of Lyα emission is typically known to be offset from the systemic redshift of the galaxy by several hundred kilometers per second (e.g., Shapley et al. 2003; Trainor et al. 2015; Muzahid et al. 2020; Matthee et al. 2021), the redshifts from the peak of the Lyα emission in MUSE nonetheless give us an idea of where to look for rest-optical emission in the NIRSpec data. In future work, we will systematically assess the incidence rate of emission as a function of source properties.

We begin by extracting a cutout of the reduced, unrectified 2D spectrum at the expected position of $\lambda_{H\alpha} \pm 50$ Å, and between $\lambda_{H\beta} - 50$ Å and $\lambda_{\text{[OIII]}\,5007} + 50$ Å ($\Delta\lambda \approx 1000$ km s^{-1}), where λ represents the predicted line centers based on the Lyα redshifts. Within these 2D cutouts, we visually inspect for the presence of strong emission lines (including negative residuals from the nodded background subtraction) and remove spurious cases where, e.g., a strong cosmic ray residual with underestimated uncertainties can mimic a true emission line. The widths of the search windows are narrow enough that we would not confuse, e.g., Hβ and [O III] in the case that only a single line is present in the spectrum.

1D spectra are extracted using a boxcar aperture with a size corresponding to the nodding length (0″5) that is constant with wavelength. In cases where a line is visible in the 2D spectrum, we extract the boxcar at the spatial position of the strongest line, taking into account the curvature of the slits. Otherwise, we center the boxcar on the catalog source position for each

object (i.e., its R.A. and decl. as projected onto the MSA shutter), meaning the position that corresponds to the flux-weighted Lyα centroid from MUSE.

When a line is observed in the NIRSpec data, we use the wavelength corresponding to the peak line flux as the initial guess to the redshift fitting on the 1D spectrum with `pyPlatefit`. Otherwise, we start the `pyPlatefit` fit at the MUSE-based Lyα redshift, but allowing for alternative redshift solutions to be found (see Bacon et al. 2023 for a description of the `pyPlatefit` redshift fitting process; and Appendix C for a description of the accuracy of the NIRSpec wavelength calibration). Our final sample of MUSE LAEs with redshifts from our NIRSpec observations consists of 46 objects (i.e., 46/90 objects from our initial sample have at least one emission line detected).

Figures 1 and 2 show the NIRSpec spectra for all objects in the sample with at least one detected optical emission line. The objects for which we did not detect any rest-optical emission lines with NIRSpec either are too faint, are affected by artifacts in the data, and/or have an incorrect MUSE-based redshift. Stacked NIRSpec spectra for undetected objects based on the MUSE redshifts, which may shed light on the relative contributions of each failure mode and will be presented in future work.

2.4. Line Flux Determinations

Once the redshift for each spectrum is fixed, we perform a Markov Chain Monte Carlo analysis of the spectra using `pyPlatefit`. Spectra are perturbed according to their uncertainties and refit until convergence, with the resulting mean and standard deviation of the line flux measurements being quoted throughout. We fit the G235M and G395M data separately. In cases when a line is covered in both, we quote the G235M result as those observations have longer exposure times and hence a higher expected signal-to-noise ratio (S/N). To minimize compounding assumptions about the absolute error values as well as the flux calibrations (see Appendix D), we do not attempt to combine the spectra in these overlapping regions.

As an additional verification of our uncertainties, we compare the measurements for emission lines from each of the individual three-nodded sequences described in Section 2.2. As discussed in Brinchmann (2023), these "duplicate" observations are useful in understanding systematic issues with the uncertainties such as the impact of correlated noise (which, e.g., would not be apparent when looking at pixel-to-pixel variations alone as in Appendix B). We use our line fitting methodology on each of the individual exposures and compare the measurements and the uncertainties for [O III], Hβ, and Hα. The standard deviation of the resulting flux differences is 0.996, justifying our increase of the pixel-level errors in Section 2.2.

3. Results

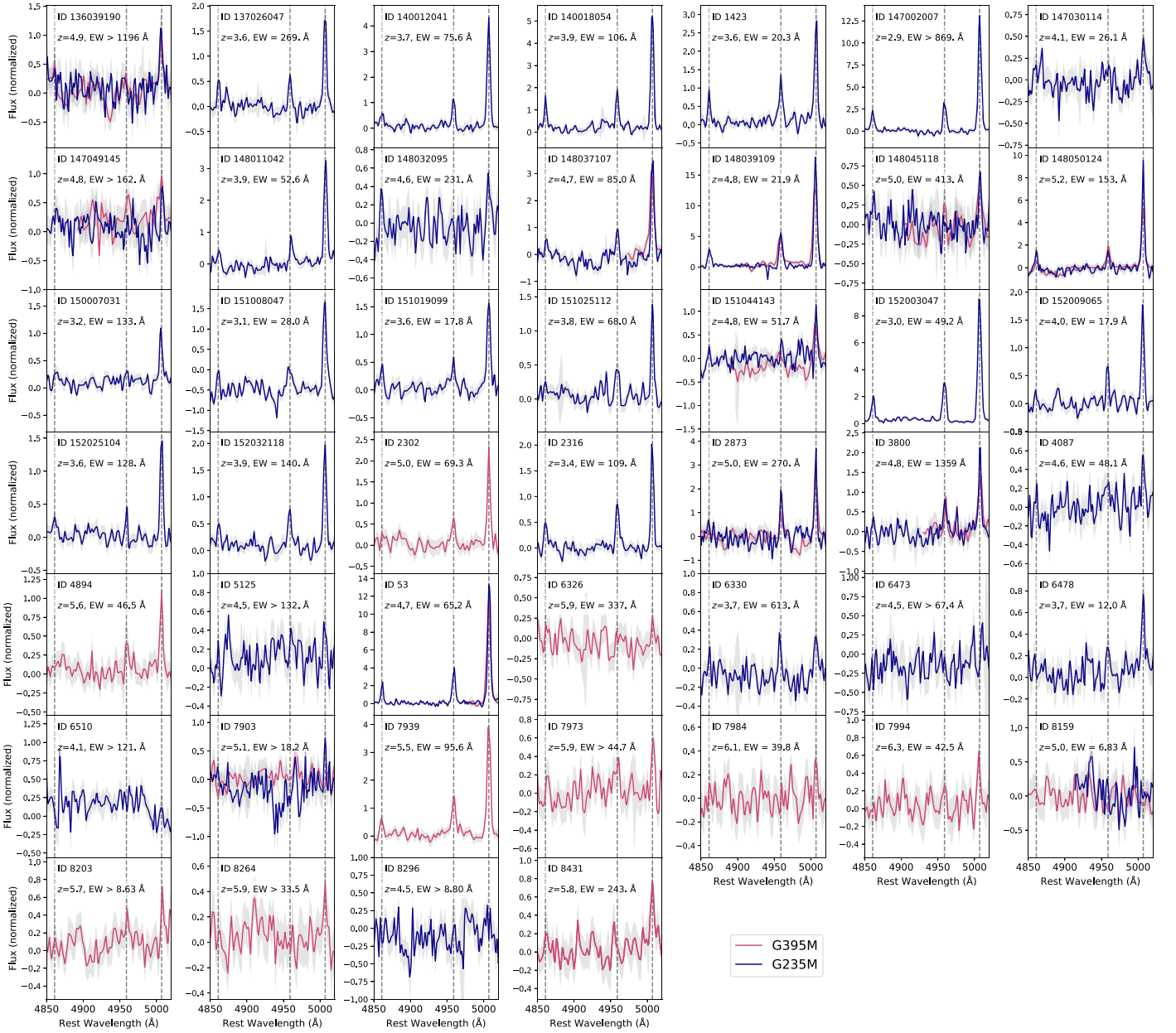For brevity, we define the following emission line ratios:

$$N2 = \log\left(\text{[NII]}\,\lambda 6584/H\alpha\right);$$
$$R3 = \log\left(\text{[O III]}\,\lambda 5007/H\beta\right).$$

Note that the ratios of [O III] $\lambda 5007$ to $\lambda 4959$ and [N II] $\lambda 6584$ to $\lambda 6548$ are fixed in our fitting procedure to the values in Storey & Zeippen (2000).


Figure 1. NIRSpec spectra of LAEs showing the region around [O III] and Hβ for objects in our sample. The darker (blue) line shows G235M data, and the lighter (pink) line shows G395M data. Shaded regions show the $\pm 1\sigma$ errors on the spectra.

Considering the close wavelength spacing between the lines in both ratios, they are relatively insensitive to dust attenuation and/or wavelength-dependent flux calibration issues (see Appendix D). For each galaxy, we calculate the ratio according to the measured fluxes and uncertainties of each emission line as detailed in Section 2.3. In the case where one of the emission lines is not detected above a signal-to-noise of 1, we instead use the 2σ upper limit to the flux when calculating the limit on the line ratio.

For the 44 galaxies with spectral coverage of both N2 and R3 ratios, we can construct line ratio diagnostic diagrams, which can be used to differentiate between star formation and active galactic nucleus ionization in the galaxy (e.g., Baldwin et al. 1981; Veilleux & Osterbrock 1987; Kewley et al. 2001). All of our objects with measurements of both ratios are consistent with being in the star-forming region using the demarcation line of Kewley et al. (2013) extrapolated to the

median redshift of our sample $z = 4.5$ within 1σ. Beyond this simple demarcation, though, these line ratios encode information about the interstellar medium (ISM) conditions inside the galaxies. In particular, galaxies with different ionizing radiation fields, metallicities, electron densities, and ionization parameters will lie in different parts of line ratio space (Kewley et al. 2013).

In Figure 3, we present our measurements as well as limits on the line ratios of R3 and N2 for our sample of MUSE-selected LAEs. 36 of our 46 targets have detections or upper or lower limits on R3 and N2. Similar to the $z > 4$ star-forming galaxies from Cameron et al. (2023), we see relatively few detections (19) of [N II] in our individual spectra. In particular, at the typical [O III]/Hβ ratios observed here, the lack of [N II] limits our ability to interpret the physical conditions of the ISM in individual galaxies. A more detailed investigation of these conditions utilizing additional line ratios will be presented in
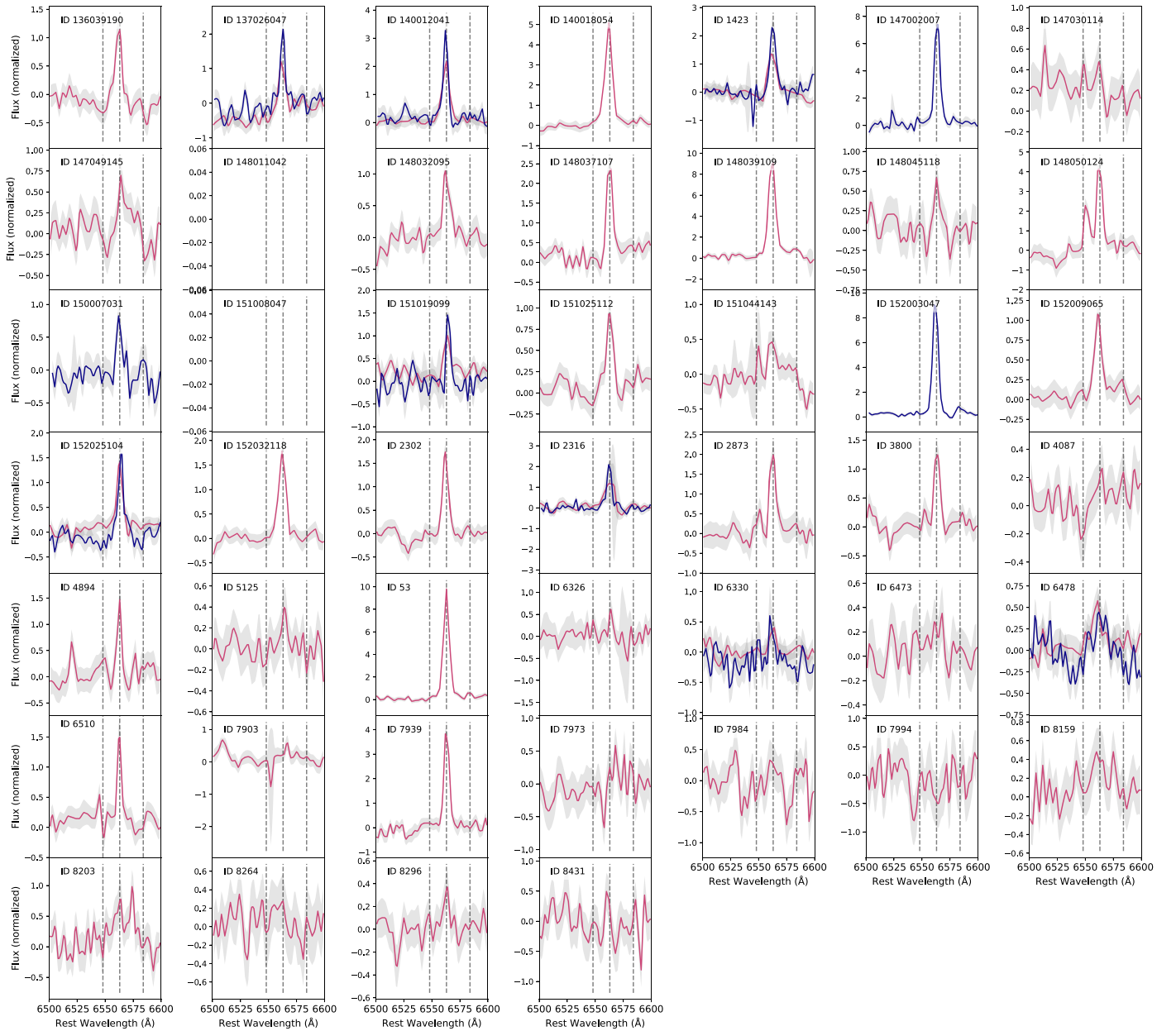
**Figure 2.** Same as Figure 1, but for the region around [N II] and Hα.

forthcoming work (Z. Lewis et al. 2023, in preparation), but here, we focus on stacked spectra where precise measurements of the (average) line ratios are possible.

3.1. Line Ratios from Stacked Spectra

In order to achieve a higher S/N in each of the lines, we create median stacks of our NIRSpec data. For the 46 MUSE LAEs with confirmed redshifts from NIRSpec, we split our sample into two parts based on their Lyα EWs. Note that the stacks can include objects that are missing spectral coverage of either R3 or N2 and otherwise do not appear in Figure 3. We select 23 objects that (a) have a measured Lyα rest-frame EW in excess of 90 Å or (b) are undetected in the HST imaging, implying we can only derive a lower-limit on the Lyα EW (Maseda et al. 2020; Kerutt et al. 2022). This constitutes our *high EW* stack, with a median EW of 154 Å. We create a second stack including the 23 objects with lower EWs (median

value of 43 Å; see the mean Lyα EW from MUSE-detected sources at $z = 4.9$ of 88 Å; Hashimoto et al. 2017). The final stack consists of all 46 MUSE LAEs with redshift detections in NIRSpec, noting that all of these sources are well-detected in Lyα from MUSE. The median redshifts for the stacks are $z = 4.8$, 4.8, and 4.6, respectively. Spectra in each bin are interpolated onto the same rest-frame wavelength grid, and the median flux at each wavelength is measured. While the individual measurements of emission line ratios can be quite noisy, typically from very faint [N II] or Hβ, the stacked measurements result in significant ($>2\sigma$) detections in all emission lines. The error bars are determined by taking 50,000 random samples of objects in each of the categories and reperforming the median stacking including random perturbations to each 1D spectrum according to its noise vector.

Figure 3 shows the resulting stacks (lower panels) as well as the values for R3 and N2 from each of the individual Monte
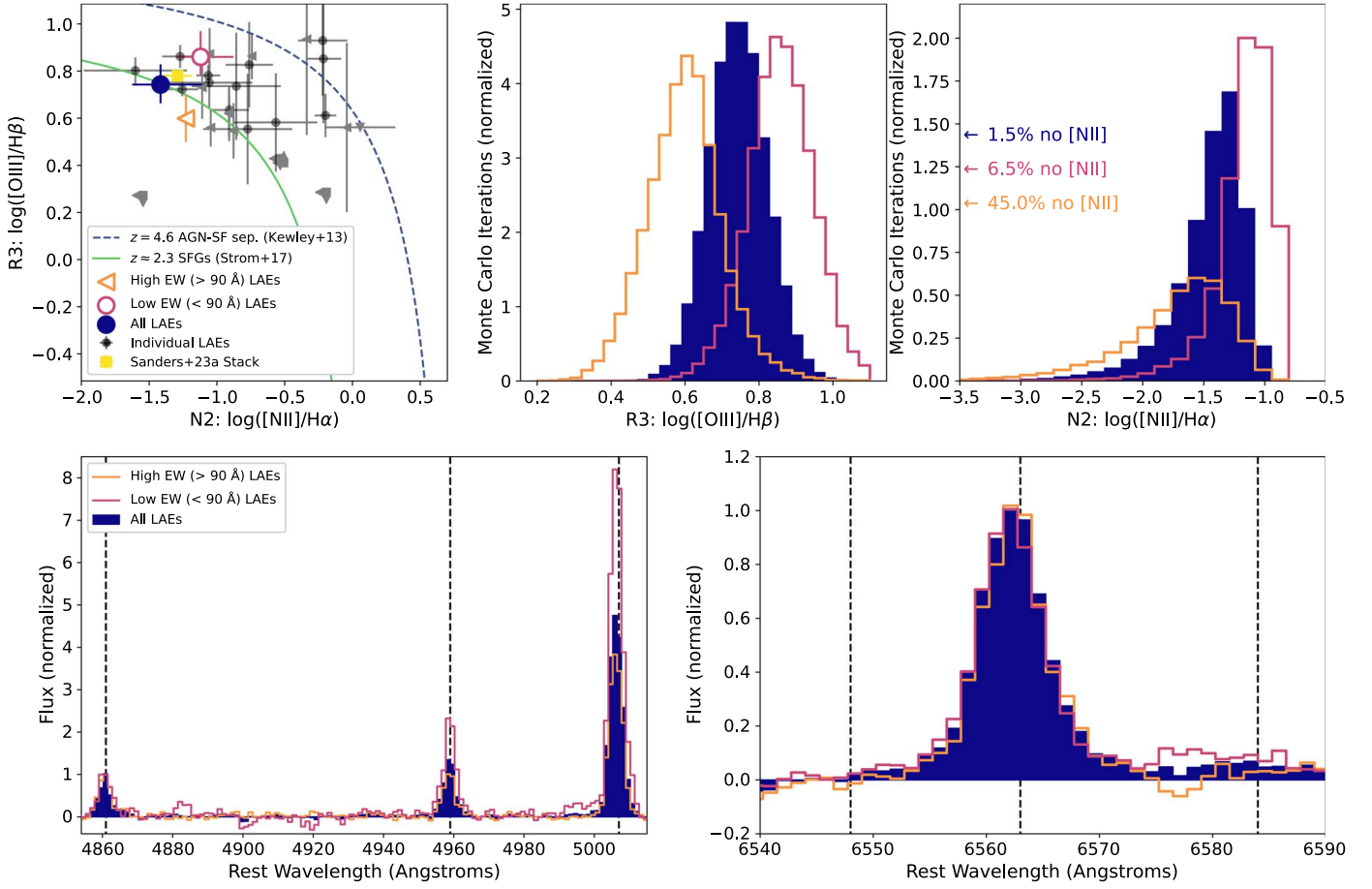
Figure 3. Upper left: R3 vs. N2 ionization diagnostic plot for our sample of LAEs at $z \approx 4.6$. For objects with S/N < 1 in any of the individual lines, we use gray triangles to represent limits (2σ) on the ratio(s); objects without a S/N > 1 detection in both components of R3 or N2 are omitted, although they are still included in the stacks. The stacked values from our full sample of LAEs lie on the sequence of $z \approx 2.3$ star-forming galaxies (SFGs) from Strom et al. (2017), and below the $z = 4.6$ extrapolated demarcation between active galactic nuclei (AGNs) and star formation from Kewley et al. (2013; although we caution that this extrapolation is not well-calibrated at these redshifts). Compared to the Sanders et al. (2023a) stack of $z \approx 4.5$ SFGs, we observe a trend of lower R3 and N2 values when restricting the stack to higher-EW Lyα. Upper center and right: observed R3 and N2 ratios for our sample from each Monte Carlo iteration (Section 3.1). Lower panels: stacked spectra for the three EW subsamples showing Hβ, [O III], Hα, and [N II] (vertical dashed lines). Each plot is normalized to the flux of Hβ or Hα.

Carlo iterations (upper center and upper right panels) and the overall R3 and N2 values for each stack (upper left panel). In the case of N2 for the high-EW stack, we only obtain a reliable upper limit to the value as [N II] is not always detected in the stacks (45.0% of the time the [N II] flux is either zero or negative). The stack of all LAEs has comparable line ratios to the Sanders et al. (2023a) stack of star-forming galaxies at similar redshifts (not selected by Lyα), which itself closely resembles the values for $z \approx 2.3$ star-forming galaxies from Strom et al. (2017).

Within our sample, we see lower R3 and N2 values in the stack of sources with the highest Lyα EW, implying a lower gas-phase metallicity. Using the Monte Carlo iterations, we see that N2 is lower in the high-EW stack at a significance of 1.9σ compared to the low-EW stack. R3 is lower at a significance of 2.2σ. Figure 4 shows the relationship between R3 and Lyα EW, providing further evidence that higher EWs correspond to lower R3 values. This strongly suggests a lower gas-phase oxygen abundance in the high-EW LAEs, as we discuss in the following section. Furthermore, since the lines are closely spaced in wavelength, it is a robust result that is not strongly affected by systematics (see Appendix D).

4. Discussion and Conclusions

As shown in Trainor et al. (2016), Maseda et al. (2018), Maseda et al. (2020), and Kerutt et al. (2022), an elevated Lyα EW typically signifies a low-metallicity, young, vigorously star-forming galaxy. This trend of higher Lyα EWs with lower gas-phase metallicities typically manifests only as a lower N2 value; the R3 values are typically *elevated* in the higher-EW systems (Trainor et al. 2016; Cullen et al. 2020; Du et al. 2021). This, however, is expected for metallicities in excess of $\approx 10\%$ $Z_\odot$ (seen in the right panel of Figure 4) where R3 increases as the metallicity decreases. At the extremely low metallicities of the MUSE-selected sources, we expect metallicity to be directly correlated to the value of R3. This difference is due to the double-valued nature of R3.

Although nitrogen-based calibrations such as R3N2 (i.e., the ratio of R3 to N2) have not been calibrated at these redshifts using JWST data (Laseter et al. 2023), if we were to adopt the R3N2 calibration of Bian et al. (2018), we would observe a lower metallicity in the high-EW subset compared to the low-EW subset at a significance of 2.39σ. If we use the empirical calibration between R3 and oxygen abundance from Bian et al. (2018), which has been shown to match high-z observations
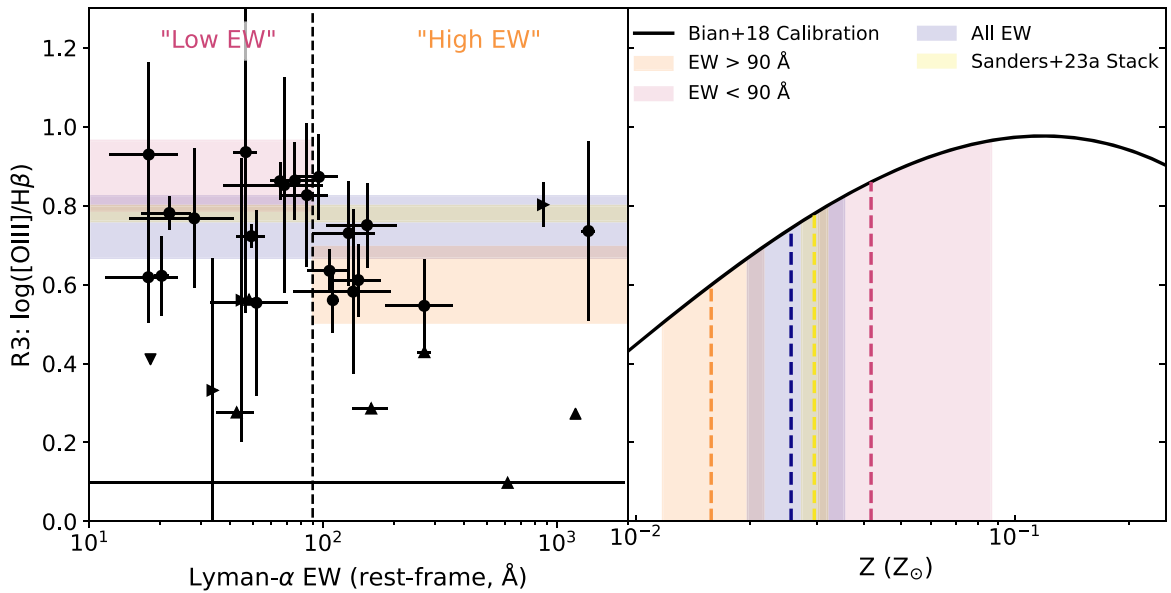
Figure 4. Left: NIRSpec-derived R3 ratio vs. MUSE-derived Lyα EW. In cases where the Lyα EW or R3 value are limits, triangles are used. The shaded regions are the $\pm 1\sigma$ ranges for R3 for each of the stacks based on Lyα EW. We observe systematically lower R3 values as a function of Lyα EW, indicative of lower gas-phase metallicities. Right: the relationship between R3 and gas-phase metallicity using the calibration from Bian et al. (2018). The shaded regions show the 1σ stacked R3 ranges, including those from Sanders et al. (2023a), and their corresponding metallicities. Our high-EW LAEs have lower R3 values than the "typical" SFGs from Sanders et al. (2023a) at a significance of 2.0σ, indicating lower metallicities (vertical dashed lines).

with NIRSpec (Laseter et al. 2023), we observe that the highest-EW objects have a metallicity of $1.5 \pm 0.50\%$ $Z_\odot$, compared to $4.1^{+4.5}_{-1.1}\%$ $Z_\odot$ for the lower-EW subset and $2.5^{+0.98}_{-0.60}\%$ $Z_\odot$ for all LAEs (Figure 4). If we use the nebular photoionization models from Gutkin et al. (2016; which assume a constant SFR for the last 100 Myr), to match the R3 and N2 values of our low- and high-EW samples, we determine oxygen abundances of $60^{+10}_{-22}\%$ and $<37\%$ $Z_\odot$ (2σ), respectively. The stack of all LAEs would have a metallicity of $20^{+11}_{-19}\%$ $Z_\odot$.

The discrepancy between the absolute metallicities derived using the two different methods partially highlights the need for strong-line diagnostics calibrated from JWST data that include nitrogen lines (Laseter et al. 2023; Sanders et al. 2023b). We also need a detailed characterization of the star formation histories and emission line properties of individual galaxies in the sample (Z. Lewis, in preparation). However, all of these values are in the range of metallicities derived by Maseda et al. (2020), based on the Lyα EWs, the ionizing photon production efficiencies, and the UV continuum slopes.

If higher Lyα EWs indeed correspond to lower metallicities, it would be expected that this MUSE-derived sample, with half of the sources plausibly having EWs in excess of 90 Å, would show a lower metallicity than the Trainor et al. (2016) sample with a median EW of 56 Å. This is what we observe, where the stack of all of our LAEs has a metallicity of 2.5% $Z_\odot$ (or 20% $Z_\odot$ using the Gutkin et al. 2016 calibrations) compared to 22% $Z_\odot$ for the Trainor et al. (2016) sample. Moreover, Trainor et al. (2016) observe that the continuum-faintest sources in their sample (i.e., those with plausibly the highest EWs) may actually be in the low-metallicity regime with a depressed value of R3. This has a dependence on the ionization parameter, where Trainor et al. (2016) observe a positive correlation between Lyα EW and the ionization parameter. In fact, a larger ionization parameter may actually drive the elevated Lyα EWs observed in our sample (Erb et al. 2016). While we cannot

directly constrain the ionization parameter with the optical lines presented in this work, future stacking experiments with these data can potentially yield constraints.

JWST/NIRSpec has given us the unique opportunity to study the rest-frame-optical spectral region of galaxies at $z \approx 4$–6 for the first time. With 7 and 11 hr of exposure time in G395M/F290LP and G235M/F170LP, respectively, we detect [O III], Hβ, Hα, and/or [N II] in 46 out of 90 MUSE-selected LAEs down to a UV magnitude of $M_{\rm UV} \approx -15$. We measure the [N II]/Hα ratio (limits) for 19 (25) galaxies and the [O III]/Hβ ratio for 25 (21) galaxies. 13 galaxies have detections of all four emission lines studied here. After correcting the error estimates propagated through the data reduction pipeline (Appendix B), we focus on these same line ratios in the stacked spectra. We observe that, when split into "high" and "low" Lyα EW stacks at 90 Å, the "high" stack consists of galaxies with systematically lower gas-phase metallicities, potentially as low as 1.5% $Z_\odot$ when using the Bian et al. (2018) R3 calibration. The "low" EW stack is still plausibly very low metallicity (4.1% $Z_\odot$), with R3 and N2 values that are lower than similar results from other intermediate- to high-z stacks. We stress that stacking is important to take into account a number of galaxies that lack individual detections of [N II] and/or Hβ, where otherwise requiring detections of those lines would bias the sample mean to higher [O III]/Hβ and [N II]/Hα ratios.

In galaxies with subsolar metallicity, one of the primary sources of radiative cooling is via collisionally excited ions like [O III]. However, at oxygen abundances less than $\approx 10\%$ $Z_\odot$, there is so little oxygen in the ISM that the ratio of [O III] to Hβ would start to decrease (see Figure 4). Although this has been observed in some of the most metal-poor galaxies locally (e.g., Berg et al. 2016; Laseter et al. 2022; Umeda et al. 2022), this is the first evidence so far of a systematically suppressed [O III]/Hβ ratio at high z from JWST spectroscopy (e.g., Cameron et al. 2023; Laseter et al. 2023; Nakajima et al. 2023; Sanders

et al. 2023a, 2023b). We interpret this as evidence for extremely low metallicities in the highest-EW LAEs. This, however, is still a significant level of metal enrichment compared to truly "Population III" stellar populations, which are expected to show R3 ratios below −1 (Inoue 2011). In future work, we will utilize improved data reduction and analysis tools to derive metallicities and excitation diagnostics for individual galaxies. With absolute flux calibrations, we will determine SFRs from Hβ to compare to the results of Maseda et al. (2020), who determined a nominal Hα-based SFR of 1.2 $M_\odot$ yr^{-1} and a UV-based SFR of 0.1 $M_\odot$ yr^{-1} in the highest-EW LAEs (the core of this sample), where the discrepancy could be due to episodic star formation histories. Nevertheless, our initial results with JWST/NIRSpec already highlight the huge advances that can be made with this facility.

et al. 2016); pypeit (Prochaska et al. 2020b); pyplatefit (Bacon et al. 2023).

Appendix A
NIRSpec Microshutter Operability

As noted in Rawle et al. (2022) as well as papers on ERO data (e.g., Taylor et al. 2022; Curti et al. 2023), there can be intermittent failures of otherwise functional MSA shutters. We diagnose the incidence rate of this affect by comparing our planned MSA configuration with our post-acquisition confirmation image, which ultimately reveals which shutters are allowing flux transmission during the science exposures (nb. we cannot fully diagnose if any of these shutters are *partially* open, but this would not affect our primary science as we are only concerned with line ratios and not any absolute flux calibration).

Of the 386 shutters that were commanded open in our observations, three did not open (0.8%). Two of the failures were shutters in quadrant 2: (51, 7) and (151, 159). One was in quadrant 4: (91, 151). As of APT version 2022.7.1, none of these shutters are determined to be inoperable. This is significantly smaller than the ≈4% quoted by Rawle et al. (2022). However, this could be due to the relatively few commanded-open shutters placed in quadrants 1 and 2 (90 and 48, respectively) compared to those in quadrants 3 and 4 (87 and 161), as shutters in these two quadrants are approximately twice as likely to experience an intermittent failure (Rawle et al. 2022).

Appendix B
NIRSpec Noise Estimates

Using the STScI NIRSpec pipeline as described in Section 2.2, the quoted uncertainties in the 2D rate files do not accurately explain the pixel-to-pixel variations present in the data: the errors are systematically underestimated. In Figure 5, we show data from both the standard level 1 and 2 products ("rate" and "cal" files) in the left and center panels, respectively. We plot the background-subtracted flux (differenced using our nodding pattern) compared to the quoted uncertainties on each of the pixels. If the noise is properly accounted for, this would be a Gaussian distribution with a σ of 1. However, this is not observed to be the case (nb. without a median bias level subtraction performed by `msaexp`, the NRS2 values are offset toward negative values instead of being centered on zero, visible in Figure 5). To account for this, we empirically determine a multiplicative factor that would turn these distributions into normal Gaussians. We therefore scale all of our error estimates by a factor of 1.91 (G235M) or 1.70 (G395M), leading to the distributions seen in the right panel of Figure 5.
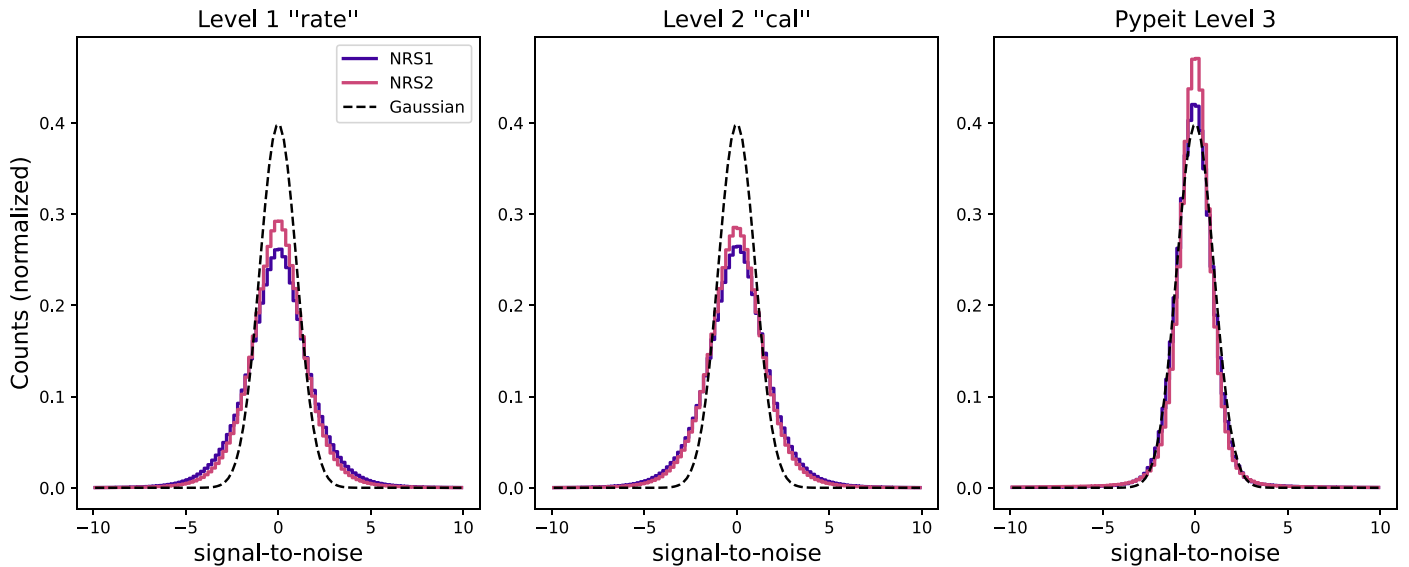
Figure 5. Comparison of the background-subtracted signal-to-noise values for (left) all pixels in the level 1 processed "rate" files, (center) pixels illuminated by open MSA shutters in the level 2 "cal" files, and (right) pixels in the `Pypeit`-calibrated 2D frames, including the extra factors of 1.91 and 1.70 in the noise estimates as described in the text. Each of the two detectors are plotted separately. The discrepancy in the distributions of the level 1 and 2 products compared to Gaussian statistics (dashed line) motivates the extra noise factors used in `Pypeit`.

Appendix C
NIRSpec Wavelength Calibration

Unlike ground-based spectrographs, there are typically no dedicated arc calibrations taken with NIRSpec after MSA observations, nor are there OH skylines that can be used. As such, it is valuable to provide additional verification that the wavelength calibration step is performed correctly. We have

the opportunity to check this with our R1000 spectra in cases where the G235M and G395M gratings overlap (i.e., $\approx$28500–30000 Å).

The results for individual objects are given in Table 1. The mean velocity offset between the two gratings is 22.38 ± 16.75 km s^{-1}, i.e., it is consistent with no systematic offset between the two gratings.

Table 1
Comparison of Emission Line Redshifts in G235M and G395M

MUSE ID	Line	z_{G235M}	z_{G395M}	Δv (km s^{-1})
53	[O III]	$4.7763 \pm 5.9841\text{e-}5$	$4.7757 \pm 1.0410\text{e-}4$	31.030 ± 6.2364
2316	Hα	$3.4713 \pm 4.1071\text{e-}4$	$3.4702 \pm 1.6707\text{e-}3$	72.638 ± 115.43
2873	[O III]	$5.0499 \pm 1.9912\text{e-}4$	$5.0500 \pm 3.2134\text{e-}4$	-8.4822 ± 18.746
3800	[O III]	$4.8208 \pm 1.7813\text{e-}4$	$4.8214 \pm 6.0100\text{e-}4$	-32.666 ± 36.234
136039190	[O III]	$4.9438 \pm 9.9560\text{e-}4$	$4.9440 \pm 7.2192\text{e-}4$	-9.6310 ± 62.070
140012041	Hα	$3.7931 \pm 2.0336\text{e-}4$	$3.7932 \pm 2.7577\text{e-}4$	-2.7580 ± 21.446
148039109	[O III]	$4.8682 \pm 5.3343\text{e-}5$	$4.8674 \pm 7.9943\text{e-}5$	38.600 ± 4.9132
151044143	[O III]	$4.8677 \pm 3.3596\text{e-}4$	$4.8665 \pm 9.4421\text{e-}4$	61.655 ± 51.240
152025104	Hα	$3.6585 \pm 1.9973\text{e-}4$	$3.6577 \pm 3.7174\text{e-}4$	50.993 ± 27.176

Note. Redshift determinations are all from `pyPlatefit`, using a Monte Carlo algorithm to determine the uncertainties. When multiple lines are available in both gratings (i.e., [O III] and Hβ), we utilize the strongest line in the group for the redshift determination.

Appendix D
Systematic Uncertainties on [O III]/Hβ and [N II]/Hα

Due to the fixed nature of the NIRSpec MSA shutters and the wavelength-dependent point-spread function, any emission lines that are separated in wavelength will suffer from different amounts of geometrical slit losses. For our input catalog, the NIRSpec pipeline currently assumes that all sources are all spatially extended and treats them as uniformly filling the open MSA area. Since the geometrical slit losses will increase with wavelength, the pipeline correction could systematically bias our R3 and N2 values low.

Although these pairs of lines are close together in wavelength, we may nevertheless be systematically underestimating the line ratios due to this effect. In order to assess the magnitude of this, we utilize a set of simulations using v2.0 of the `Pandeia` (Pontoppidan et al. 2016) simulation engine. We simulate extended sources with $n = 1$ Sérsic indices across a grid of half-light radii, centered on various positions across an MSA shutter (see also Maseda et al. 2019). We then calculate the differential slit losses at the positions of [O III] and Hβ as well [N II] and Hα (and [O III] and [O II], for reference) compared to the pipeline correction factor, which is calculated assuming the sources are uniform across the open shutter area.

The result is shown in Figure 6 as a function of half-light radius, assuming a fiducial redshift of 5. As an example, at this redshift, the pipeline introduces a 0.4% correction to the [O III]/Hβ ratio for the slightly larger losses at the wavelength of [O III]. We calculate that the true correction for sources with an $n = 1$ Sérsic light profile (i.e., not uniform), and a half-light radius of $0\rlap{.}{''}1$ is 0.9%–1.5%, depending on the size and centroid of the source within the open area of an MSA shutter. Hence, the necessary 'correction' from the measured line ratio to the true one is $\approx 0.5\%$–1.1% (see the y-axis of Figure 6). The effect is even smaller for the more closely spaced [N II]/Hα ratio. These systematic offsets are not large enough to cause the discrepancy between our two stacks even if, e.g., the high-EW sources are systematically smaller, which would cause an artificial suppression of their R3 and N2 ratios, or if there are systematic spatial offsets between the Lyα emission centroid and the rest-optical emission lines (e.g., Matthee et al. 2016; Lemaux et al. 2021; M. Maseda et al. 2023, in preparation). One additional verification comes from observations of both the $\lambda 5007$ and $\lambda 4959$ components of [O III]: for sources with detections of both lines at $>2\sigma$, the average ratio is 2.96 ± 0.0937, comparable to the value of 2.98 expected from Storey & Zeippen (2000). The only case where both lines are detected with a signal-to-noise value greater than 2 and the deviation from the expected 2.98 is larger than 2σ is 147002007 (central in Figure 1) where there is a negative flux residual from the cosmic ray subtraction near the 4959 emission line in the 2D frame, biasing the boxcar-extracted flux low without properly contributing to an elevated noise level. We expect improvements to the cosmic ray subtraction in future versions of the pipeline.

A similar systematic effect in our ability to measure line ratios could occur due to the 1D spectral extractions, where the assumption of a fixed extraction window (e.g., a boxcar) could result in the preferential loss of flux at long wavelengths compared to short wavelengths. As we adopt a boxcar width equal to the full open area of the shutter, our 1D extractions are not by themselves introducing a systematic effect to our measurement of the emission line ratios compared to the
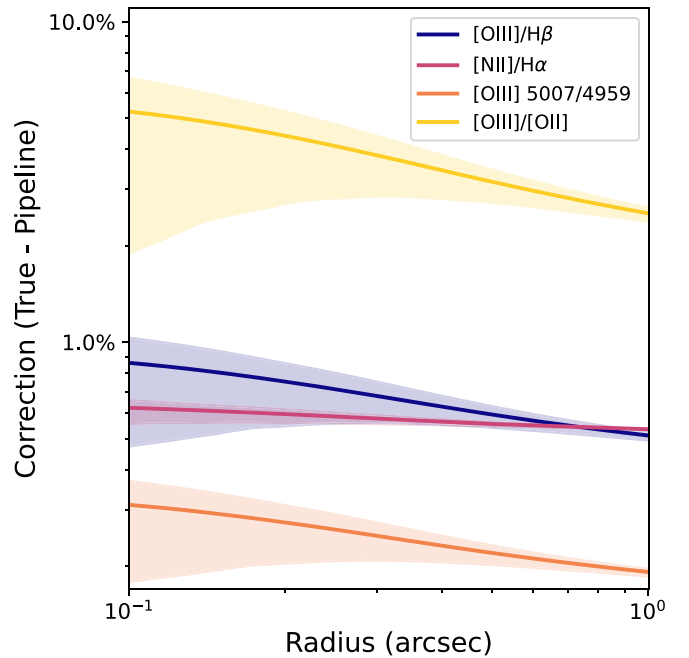
Figure 6. Difference between the *true* correction factor for differential slit losses (from `Pandeia` simulations, assuming an $n = 1$ Sérsic profile for a $z = 5$ galaxy) and the assumption used by the pipeline that sources uniformly illuminate the open shutter area. The shaded regions denote the $+/-2\sigma$ distribution in these values for each line ratio, which encapsulates the variety of spatial offsets possible for a source in the open shutter area. The NIRSpec pipeline, by default, systematically lowers the observed line ratios as there are larger flux losses at longer wavelengths. However, the magnitude of this systematic offset cannot explain the offsets we observe between our two stacks in Figure 3.

geometrical effects described above and the nature of the nodded background subtraction, which removes flux that extends to large radii. A more careful "optimal" spectral extraction (Horne 1986) combined with a "master sky" background subtraction will be necessary for the precise flux calibration needed for accurate measurement of ratios for distantly separated emission lines.

Finally, dust attenuation could have the opposite effect whereby the R3 and N2 ratios would be systematically biased toward higher values. However, as discussed in Maseda et al. (2020), there is little to no dust in galaxies selected based on high Lyα EWs. Moreover, this effect would also be small compared to the observed difference in the two stacks: an A_V of 1 would cause a bias in R3 of only 3% or 0.01 dex.

ORCID iDs

Michael V. Maseda ⓘ https://orcid.org/0000-0003-0695-4414
Jorryt Matthee ⓘ https://orcid.org/0000-0003-2871-127X
Joseph F. Hennawi ⓘ https://orcid.org/0000-0002-7054-4332
Leindert Boogaard ⓘ https://orcid.org/0000-0002-3952-8588
Themiya Nanayakkara ⓘ https://orcid.org/0000-0003-2804-0648
Amy Barger ⓘ https://orcid.org/0000-0002-3306-1606
Jarle Brinchmann ⓘ https://orcid.org/0000-0003-4359-8797
Marijn Franx ⓘ https://orcid.org/0000-0002-8871-3026
Takuya Hashimoto ⓘ https://orcid.org/0000-0002-0898-4038
Hanae Inami ⓘ https://orcid.org/0000-0003-4268-0393
Haruka Kusakabe ⓘ https://orcid.org/0000-0002-3801-434X
Anthony J. Taylor ⓘ https://orcid.org/0000-0003-1282-7454
Tanya Urrutia ⓘ https://orcid.org/0000-0001-6746-9936

Joop Schaye ● https://orcid.org/0000-0002-0668-5560
Eloïse Vitte ● https://orcid.org/0000-0001-5121-1260

References

Astropy Collaboration, Price-Whelan, & Lim, A. M. 2022, ApJ, 935, 167
Bacon, R., Accardo, M., Adjali, L., et al. 2010, Proc. SPIE, 7735, 773508
Bacon, R., Brinchmann, J., Conseil, S., et al. 2023, A&A, 670, A4
Bacon, R., Conseil, S., Mary, D., et al. 2017, A&A, 608, A1
Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, PASP, 93, 5
Berg, D. A., Skillman, E. D., Henry, R. B. C., Erb, D. K., & Carigi, L. 2016, ApJ, 827, 126
Bian, F., Kewley, L. J., & Dopita, M. A. 2018, ApJ, 859, 175
Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2014, ApJ, 793, 115
Brammer, G. 2023, grizli, v1.8.2, Zenodo, doi:10.5281/zenodo.7712834
Brinchmann, J. 2023, MNRAS, 525, 2087
Bushouse, H., Eisenhamer, J., Dencheva, N., et al. 2022, JWST Calibration Pipeline, v1.8.2, Zenodo, doi:10.5281/zenodo.7229890
Cameron, A. J., Saxena, A., Bunker, A. J., et al. 2023, A&A, 677, A115
Charlot, S., & Fall, S. M. 1991, ApJ, 378, 471
Cullen, F., McLure, R. J., Dunlop, J. S., et al. 2020, MNRAS, 495, 1501
Curti, M., D'Eugenio, F., Carniani, S., et al. 2023, MNRAS, 518, 425
da Cunha, E., Charlot, S., & Elbaz, D. 2008, MNRAS, 388, 1595
Du, X., Shapley, A. E., Topping, M. W., et al. 2021, ApJ, 920, 95
Erb, D. K., Pettini, M., Steidel, C. C., et al. 2016, ApJ, 830, 52
Fujimoto, S., Arrabal Haro, P., Dickinson, M., et al. 2023, ApJL, 949, L25
Gutkin, J., Charlot, S., & Bruzual, G. 2016, MNRAS, 462, 1757
Harikane, Y., Ouchi, M., Shibuya, T., et al. 2018, ApJ, 859, 84
Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Natur, 585, 357
Hashimoto, T., Garel, T., Guiderdoni, B., et al. 2017, A&A, 608, A10
Herenz, E. C., Urrutia, T., Wisotzki, L., et al. 2017, A&A, 606, A12
Horne, K. 1986, PASP, 98, 609
Hunter, J. D. 2007, CSE, 9, 90
Illingworth, G. D., Magee, D., Oesch, P. A., et al. 2013, ApJS, 209, 6
Inami, H., Bacon, R., Brinchmann, J., et al. 2017, A&A, submitted
Inoue, A. K. 2011, MNRAS, 415, 2920
Karakla, D., Shyrokov, A., Pontoppidan, K., et al. 2014, Proc. SPIE, 9149, 91491Z
Kerutt, J., Wisotzki, L., Verhamme, A., et al. 2022, A&A, 659, A183
Kewley, L. J., Dopita, M. A., Leitherer, C., et al. 2013, ApJ, 774, 100
Kewley, L. J., Dopita, M. A., Sutherland, R. S., Heisler, C. A., & Trevena, J. 2001, ApJ, 556, 121
Laseter, I. H., Barger, A. J., Cowie, L. L., & Taylor, A. J. 2022, ApJ, 935, 150
Laseter, I. H., Maseda, M. V., Curti, M., et al. 2023, arXiv:2306.03120
Lemaux, B. C., Fuller, S., Bradač, M., et al. 2021, MNRAS, 504, 3662
Malhotra, S., & Rhoads, J. E. 2002, ApJL, 565, L71

Maseda, M. V., Bacon, R., Franx, M., et al. 2018, ApJL, 865, L1
Maseda, M. V., Bacon, R., Lam, D., et al. 2020, MNRAS, 493, 5120
Maseda, M. V., Franx, M., Chevallard, J., & Curtis-Lake, E. 2019, MNRAS, 486, 3290
Matthee, J., Mackenzie, R., Simcoe, R. A., et al. 2023, ApJ, 950, 67
Matthee, J., Sobral, D., Hayes, M., et al. 2021, MNRAS, 505, 1382
Matthee, J., Sobral, D., Oteo, I., et al. 2016, MNRAS, 458, 449
Momcheva, I. G., Brammer, G. B., van Dokkum, P. G., et al. 2016, ApJS, 225, 27
Muzahid, S., Schaye, J., Marino, R. A., et al. 2020, MNRAS, 496, 1013
Nakajima, K., Ouchi, M., Isobe, Y., et al. 2023, arXiv:2301.12825
Partridge, R. B., & Peebles, P. J. E. 1967, ApJ, 148, 377
Pontoppidan, K. M., Pickering, T. E., Laidler, V. G., et al. 2016, Proc. SPIE, 9910, 991016
Prochaska, J. X., Hennawi, J., Cooke, R., et al. 2020a, pypeit/PypeIt: Release 1.0.0, v1.0.0, Zenodo, doi:10.5281/zenodo.3743493
Prochaska, J. X., Hennawi, J. F., Westfall, K. B., et al. 2020b, JOSS, 5, 2308
Rafelski, M., Teplitz, H. I., Gardner, J. P., et al. 2015, AJ, 150, 31
Raiter, A., Schaerer, D., & Fosbury, R. A. E. 2010, A&A, 523, A64
Rawle, T. D., Giardino, G., Franz, D. E., et al. 2022, Proc. SPIE, 12180, 121803R
Roberts-Borsani, G., Treu, T., Chen, W., et al. 2023, Natur, 618, 480
Sanders, R. L., Shapley, A. E., Topping, M. W., Reddy, N. A., & Brammer, G. B. 2023a, ApJ, 955, 54
Sanders, R. L., Shapley, A. E., Topping, M. W., Reddy, N. A., & Brammer, G. B. 2023b, arXiv:2303.08149
Schaerer, D. 2003, A&A, 397, 527
Schaerer, D., Marques-Chaves, R., Barrufet, L., et al. 2022, A&A, 665, L4
Shapley, A. E., Steidel, C. C., Pettini, M., & Adelberger, K. L. 2003, ApJ, 588, 65
Simmonds, C., Tacchella, S., Maseda, M. V., et al. 2023, MNRAS, 523, 5468
Skelton, R. E., Whitaker, K. E., Momcheva, I. G., et al. 2014, ApJS, 214, 24
Storey, P. J., & Zeippen, C. J. 2000, MNRAS, 312, 813
Strom, A. L., Steidel, C. C., Rudie, G. C., et al. 2017, ApJ, 836, 164
Taylor, A. J., Barger, A. J., & Cowie, L. L. 2022, ApJL, 939, L3
Trainor, R. F., Steidel, C. C., Strom, A. L., & Rudie, G. C. 2015, ApJ, 809, 89
Trainor, R. F., Strom, A. L., Steidel, C. C., & Rudie, G. C. 2016, ApJ, 832, 171
Trump, J. R., Arrabal Haro, P., Simons, R. C., et al. 2023, ApJ, 945, 35
Umeda, H., Ouchi, M., Nakajima, K., et al. 2022, ApJ, 930, 37
Urrutia, T., Wisotzki, L., Kerutt, J., et al. 2019, A&A, 624, A141
Veilleux, S., & Osterbrock, D. E. 1987, ApJS, 63, 295
Wang, X., Cheng, C., Ge, J., et al. 2022, arXiv:2212.04476
Zackrisson, E., Inoue, A. K., Rydberg, C. E., & Duval, F. 2011, MNRAS, 418, L104