UNIVERSITY OF CALIFORNIA
RIVERSIDE

The Electron-Temperature-Based Oxygen Abundance of Dwarf Galaxies at the Peak of Cosmic Star Formation

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Physics

by

Timothy Edwin Gburek

September 2022

Dissertation Committee:

Professor Brian Siana, Chairperson
Professor George Becker
Professor Naveen Reddy

The Dissertation of Timothy Edwin Gburek is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

The journey to this point has been a long and, at times, difficult one, but I have been fortunate enough to have had the love, help, and support of a plethora of people along the way: family, friends, teachers, classmates, colleagues – the list goes on. To all of you who have been there for me at some point in life, thank you.

I especially would like to thank my parents, Tim and Joan, and my sisters, Ellen and Angela. You have been my staunchest supporters, always there to guide me and encourage me regardless of what path I choose or what obstacle I encounter. This all would not have been possible without you.

I also especially would like to thank my doctoral advisor, Prof. Brian Siana, for his guidance, insights, kindness, and patience. I feel very fortunate to have had you as my mentor. To my undergraduate research advisors, Prof. Claudia Scarlata and Prof. Robert Gehrz, thank you as well for everything.

Thank you to the friends I have made during my time at UCR. Graduate school can be very difficult and daunting, but we've gotten through it together, and you all have made my time here very special.

Last, but certainly not least, I want to thank my dog, Ralph, who has brought so much joy and companionship into my life over the last couple of years.

I acknowledge partial financial support from the National Science Foundation under grant No. 1617013.

The text of this dissertation, in part, is a reprint of the material as it appears in *The Detection of [O III] λ4363 in a Lensed, Dwarf Galaxy at z = 2.59: Testing Metallicity*

Indicators and Scaling Relations at High Redshift and Low Mass. This article was published in *The Astrophysical Journal*, Volume 887, Issue 2, Article ID 168 on 2019 December 20. The co-authors of this publication, Brian Siana, Anahita Alavi, Najmeh Emami, Johan Richard, William R. Freeman, Daniel P. Stark, Christopher Snapp-Kolas, and Breanna Lucero provided commentary and feedback on the written work prior to its publication. Brian Siana directed and supervised the research which forms the basis for this dissertation.

ABSTRACT OF THE DISSERTATION

The Electron-Temperature-Based Oxygen Abundance of Dwarf Galaxies at the Peak of Cosmic Star Formation

by

Timothy Edwin Gburek

Doctor of Philosophy, Graduate Program in Physics
University of California, Riverside, September 2022
Professor Brian Siana, Chairperson

At $z \sim 2$, where cosmic star formation peaks, little is known about the galaxy stellar mass (M_*), star formation rate (SFR), and oxygen abundance (or gas-phase metallicity; O/H) parameter space in which typical dwarf galaxies lie ($M_* \leqslant 10^9$ $\mathrm{M}_\odot$). This is due to current technological limitations as well as uncertainty about the accuracy of indirectly-measured metallicities at high-z. Consequently, our models and understanding of galaxy formation and evolution are largely unconstrained in this mass regime.

Here we present a sample, one of the first of its kind, of 16 representative, star-forming, gravitationally-lensed dwarf galaxies at $1.7 \lesssim z \lesssim 2.6$ ($z_{\mathrm{mean}} = 2.30$) with a median M_* of $\log(M_*/\mathrm{M}_\odot)_{\mathrm{med}} = 8.29$ and a median SFR of $\mathrm{SFR}_{\mathrm{H}\alpha}^{\mathrm{med}} = 2.25$ $\mathrm{M}_\odot$ yr^{-1}. Sample spectroscopy and photometry were taken with the Keck Multi-Object Spectrometer For InfraRed Exploration and the *Hubble Space Telescope*, respectively. Study of these galaxies is done largely via a composite emission-line spectrum of the sample, from which we detect the faint, electron-temperature-sensitive, auroral emission line, [O III] λ4363, with which we

directly estimate metallicity. Within this sample, we also independently study the $z = 2.59$ galaxy, A1689-217, which yields a rare, individual detection of [O III] λ4363 at high-z.

In assessing indirect, locally-calibrated, strong-line metallicity relations at high-z, we find certain strong-line ratios are insensitive to metallicities around those we directly measure. Other indices show greater utility, but have metallicity calibrations that vary in applicability based on the high-z and calibration samples considered. We do show evidence supporting the creation of a redshift-invariant, empirical metallicity relation within the O_{32} vs. R_{23} excitation diagram. Our stacked sample shows excellent agreement with the direct-method $M_* - O/H$ relation (MZR) of Sanders et al. (2020) and disagrees with other high-z MZRs displaying significantly shallower slopes. We also find excellent agreement with the MZR from the Feedback In Realistic Environments simulations after recalculating our stellar masses, which likely underestimate contributions from older stellar populations under our fiducial assumptions. Finally, we find consistency with a redshift-invariant fundamental metallicity relation (FMR), though note a large scatter in dwarf galaxy metallicities.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The buildup of stellar mass in galaxies over cosmic time is a central component of galaxy evolution. However, the efficiency with which different galaxies form stars is not a constant value. It has been shown by numerous studies (e.g., Behroozi et al., 2019, and references therein) that galaxies with dark matter halo masses of $M_{\rm halo} \sim 10^{12}$ $\rm M_{\odot}$, like our Milky Way, are most efficient at converting their gas into stars while galaxies with more or less massive halos show a decrease in star formation efficiency of 1-2 orders of magnitude. This decrease in efficiency is believed to be due to feedback processes driven predominantly by active galactic nuclei (AGN) in the high-mass regime and supernovae (SNe) in the low-mass regime. At low masses, it has long been argued that SNe suppress star formation through galactic winds that heat the interstellar medium (ISM) and remove gas – the fuel for star formation – from the ISM or the galaxy altogether (Larson, 1974). However, the exact processes and physics that drive this feedback are currently not well-constrained.

In the quest to better understand these feedback mechanisms in low-mass galaxies, the fundamental physical properties of stellar mass (M_*) and gas-phase metallicity of the ISM ($Z_{\rm gas}$) are often considered due to their interconnected relations to several facets of galaxy evolution, including stellar mass build-up, chemical evolution, and gas flows. In more detail, galaxies accrete cold gas from the intergalactic medium (IGM) or circumgalactic medium (CGM) to the ISM, triggering star formation and increasing the stellar mass of the galaxy. The short-lived massive stars then die via SNe, which drive galactic outflows (feedback) that suppress star formation as described above. Chemically, metals are produced via stellar nucleosynthesis and are then injected into the ISM by stellar winds and SNe. Gas flows modulate this enrichment, whereby metal-poor gas inflows and metal-laden gas outflows dilute or remove metals from the ISM, respectively, with some of these ejected metals eventually returning to the ISM via gas recycling from the CGM (see illustrations in Lilly et al. 2013 and Tumlinson et al. 2017). It is clear that a connection exists between M_* and $Z_{\rm gas}$, and this relationship, and its evolution with time, prove to be a powerful tool in constraining feedback and outflow models necessary for understanding galaxy formation and evolution.

The connection between M_* and $Z_{\rm gas}$ of star-forming galaxies, first reported in the work of Lequeux et al. (1979), is a scaling relation, simply referred to as the mass-metallicity relation (MZR). In the modern MZR, the metallicity is represented by the oxygen abundance ($12 + \log(\rm O/H)$) of the ISM, this element so-chosen because it is the ISM's most abundant metal and has strong emission lines at easily-observable, rest-optical wavelengths. With the advent of large, fairly comprehensive, and statistically-powerful surveys, the MZR was first

studied for a vast sample of local star-forming galaxies by Tremonti et al. (2004), who used the Sloan Digital Sky Survey (SDSS) to show a clear, positive, linear correlation between M_* and O/H (typically represented by a power-law in the literature) over two decades in mass between $10^{8.5} \lesssim M_*/\mathrm{M}_\odot \lesssim 10^{10.5}$. At $M_* \gtrsim 10^{10.5}$ $\mathrm{M}_\odot$, a flattening was observed as metallicities approached an asymptotic value. The local MZR and its positive correlation between M_* and O/H has since been extended down to $M_* \approx 10^6$ $\mathrm{M}_\odot$ (e.g., Lee et al., 2006), yielding a powerful empirical metric over roughly five decades in stellar mass.

More recent statistical spectroscopic surveys at $1.0 \lesssim z \lesssim 3.5$, such as 3D-*HST* (Brammer et al., 2012a), FMOS-COSMOS (Kashino et al., 2019), KBSS (Steidel et al., 2014), and MOSDEF (Kriek et al., 2015), have allowed the study of the MZR at higher redshifts where the star formation rate density of the Universe peaks (Madau & Dickinson, 2014). At these redshifts, the positive correlation between M_* and O/H is also shown to exist, though with lower metallicities at higher redshifts and fixed stellar mass (e.g., Sanders et al., 2021). Unfortunately, unlike in the local Universe, at $z \sim 2.3$ the MZR has really only been studied down to $M_* \approx 10^9$ $\mathrm{M}_\odot$ with current facilities and instrumentation predating the *James Webb Space Telescope* (*JWST*). As the next generation of ground- and space-based telescopes come online, the parameters describing the MZR, and their evolution with redshift, remain important diagnostics to constrain for the modelling of galaxy evolution. The power-law slope of the low-mass end of the MZR can inform models of outflows and feedback while the scatter in O/H can reveal secondary dependencies of metallicity.

In the empirical study of the scatter of the MZR, Mannucci et al. (2010) and Lara-López et al. (2010) discovered that metallicity is dependent on both M_* and star formation

rate (SFR), such that the MZR derives from a more fundamental $M_* - \mathrm{SFR} - \mathrm{O/H}$ relation with less scatter in metallicity than the MZR. Mannucci et al. (2010) dubbed this relationship the fundamental metallicity relation, or FMR, which posits that metallicity is anti-correlated with SFR such that galaxies with above-average (below-average) SFR have below-average (above-average) O/H at fixed M_*. Mannucci et al. (2010) also suggested that the FMR exists outside of the local Universe and is redshift-invariant out to $z = 2.5$. Recent work with the spectroscopic surveys mentioned above has generally supported these claims, showing the FMR to have little ($\lesssim 0.1$ dex) to no evolution in metallicity at fixed M_* and SFR out to at least $z \sim 2.5$ and possibly out to $z \sim 3.3$ (e.g., Sanders et al., 2018, 2021). If the FMR is actually redshift-invariant, it suggests that the evolution seen in the MZR is actually the viewing of different parts of the FMR at different redshifts. In more physical terms, as gas fractions and SFRs rise with redshift at fixed M_* (e.g., Speagle et al., 2014), metallicities decline per the FMR, naturally explaining the observed evolution with redshift of the MZR.

Above, the importance and utility of the physical property of metallicity has been expounded upon. Of equal importance is how the metallicity is measured from observations as the accuracy and consistency of these estimations with redshift and between different methods can alter our results of, and conclusions from, scaling relations and their evolution. The best approach to empirically measuring metallicity involves the use of metal recombination lines. However, this method is limited to the very near Universe and high-resolution, high signal-to-noise spectra of bright H II regions owing to the fact that metal recombination lines are $10^3 - 10^4\times$ fainter than hydrogen Balmer recombination lines (Maiolino & Mannucci, 2019). More promising is the "direct," electron-temperature-based method,

which relies on collisionally-excited emission lines (CELs) from star-forming regions. This method estimates metallicity by first calculating other intrinsic properties of the ionized gas responsible for the strength of the CELs, the electron temperature (T_e) and electron density (n_e). With these measured properties and CELs in-hand, abundances are then determined from atomic physics (Osterbrock & Ferland, 2006; Maiolino & Mannucci, 2019). While this T_e-based method has much greater utility in the local Universe than the metal recombination line method, it is still limited, especially at high metallicity and high redshift, by its reliance on weak auroral emission lines which can be $\sim 30 - 100\times$ fainter than the strong emission lines (of the same ionic species) needed for T_e estimation (Jones et al., 2015). In fact, in the pre-*JWST* era, only ~ 20 total galaxies with significantly-detected auroral-line emission have been found at $z > 1$ (Patrício et al., 2018; Sanders et al., 2020), with many of these detections only made possible by the magnification effect of gravitational lensing.

The faintness of the emission lines necessary for the metal recombination line method and the direct, T_e-based method motivated the creation of indirect, strong-line ratio−metallicity calibrations (e.g., Pagel et al., 1979; Kewley & Dopita, 2002; Pettini & Pagel, 2004). These calibrations rely on the relationships between ratios of strong emission lines and metallicity determined empirically via the direct method or theoretically with photoionization models. While these strong-line methods have proven useful in measuring abundances of samples over a broader dynamic range in mass, metallicity, and redshift, they suffer from large systematic uncertainties (~ 0.7 dex) between different strong-line indices and calibrations (Kewley & Ellison, 2008). At high redshift, it is also unknown how accurate these strong-line methods are; due to the faintness of the auroral lines needed for

T_e-based metallicities in the calibrations samples, strong-line metallicity relations have been calibrated almost exclusively in the local Universe, where H II region conditions are likely different than at high-redshift (e.g., Kewley et al., 2013; Steidel et al., 2014; Shapley et al., 2015). Larger, statistical samples of high-z galaxies with auroral-line detections are needed in order to determine the accuracy of these locally-calibrated strong-line methods and, if necessary, to create new, high-z strong-line metallicity relations.

In the following chapters of this manuscript, we study the aforementioned methods and relations, namely the applicability at high-z of locally-calibrated strong-line metallicity relations, the MZR, and the FMR. We do so first with a rare detection of a high-z auroral-line-emitter, gravitationally-lensed by a foreground galaxy cluster. We then build upon this work by creating and studying a composite of representative, high-z, dwarf galaxy spectra, from which we also measure a T_e-sensitive auroral line. This auroral line, [O III] λ4363, allows both studies to be based on accurate direct metallicities despite the samples having redshifts of $z \gtrsim 2$.

In Chapter 2, we determine the physical properties of the $z = 2.59$ [O III] λ4363-emitter, A1689-217, a star-forming dwarf galaxy gravitationally-lensed by the Abell 1689 galaxy cluster. With this galaxy's T_e-based oxygen abundance in combination with other [O III] λ4363-based metallicities of galaxies in the literature at $0 \lesssim z \lesssim 3.1$, we study the applicability of the Jones et al. (2015) and Bian et al. (2018) strong-line metallicity relations at high-z. With these same samples, we also test the suggestion of Shapley et al. (2015) of the O_{32} vs. R_{23} excitation diagram being a powerful, redshift-invariant tool from which an empirical metallicity calibration can be derived on the direct-method abundance

scale. Considering the parameterizations of Mannucci et al. (2011) and Andrews & Martini (2013), we study the scatter and evolution with redshift of the FMR. Finally, we compare our galaxy against the predicted $z = 2.59$ MZR from the Feedback In Realistic Environments cosmological simulations (FIRE; Hopkins et al., 2014; Ma et al., 2016).

In Chapter 3, we present the composite spectrum of 16 gravitationally-lensed, star-forming dwarf galaxies at $\langle z \rangle = 2.3$, selected independent of emission-line strength. Through the [N II]-based BPT diagram (Baldwin et al., 1981) and the $z \sim 2.3$ $M_* -$ SFR relation, we show that this sample of galaxies is, on average, representative of typical dwarf galaxies at $z \sim 2.3$. With a significant [O III] λ4363 detection from the composite, we are able to directly calculate the composite oxygen abundance and compare it against several locally-calibrated strong-line metallicity relations. We also use this metallicity to better constrain the low-mass end of the empirical $z \sim 2.3$ MZR, in particular the recent direct-method MZR of Sanders et al. (2020). We further compare our composite against the predicted MZRs from The Next Generation Illustris (IllustrisTNG; Weinberger et al., 2017; Pillepich et al., 2018; Torrey et al., 2019) and FIRE cosmological simulations. Finally, we investigate the redshift evolution of the locally-defined, direct-method FMR as parameterized by Sanders et al. (2017). In Chapter 4, we summarize and discuss our results from both of the previous chapters and look forward to future high-z spectroscopic studies, with telescopes like *JWST*, that our findings will help inform.

Chapter 2

The Detection of [O III] λ4363 in a Lensed, Dwarf Galaxy at $z = 2.59$: Testing Metallicity Indicators and Scaling Relations at High Redshift and Low Mass

2.1 Introduction

Gas-phase metallicity, measured as nebular oxygen abundance, is a fundamental property of galaxies and is critical to understanding how they evolve across cosmic time. Metallicity traces the complex interplay between heavy element production via star forma-

tion/stellar nucleosynthesis and galactic gas flows, whereby infalling gas dilutes the interstellar medium (ISM) with metal-poor gas, and outflowing gas removes metals from the galaxy. These gas flows also relate to star formation and feedback, in which cold gas falls into the galaxy, triggering star formation that is later quenched by enriched outflows from supernovae that heat the ISM and remove the gas needed for star formation. As a tracer of the history of inflows and outflows, metallicity measurements at different redshifts constrain the timing and efficiency of processes responsible for galaxy growth.

This connection between metallicity and the build-up of stellar mass is encapsulated in the stellar mass (M_*) – gas-phase metallicity (Z) relation (MZR) of star-forming galaxies, seen both locally (e.g., Tremonti et al., 2004; Kewley & Ellison, 2008; Andrews & Martini, 2013) and at high redshift (e.g., Erb et al., 2006; Maiolino et al., 2008; Zahid et al., 2013; Henry et al., 2013b; Steidel et al., 2014; Sanders et al., 2015, 2019) where metallicities are lower at fixed stellar mass. The relation shows that low-mass galaxies are more metal-poor than their high-mass counterparts, possibly due to the increased effectiveness of galactic outflows (feedback) in shallower potential wells. Constraining the MZR and its redshift evolution is vital to constraining the processes ultimately responsible for galaxy formation and evolution.

The mass-metallicity relation has also been shown to derive from a more general relation between stellar mass, star formation rate (SFR), and oxygen abundance. This $M_* - \mathrm{SFR} - Z$ connection, the Fundamental Metallicity Relation (FMR), was first shown to exist by Mannucci et al. (2010) with ∼140,000 Sloan Digital Sky Survey (SDSS; Abazajian et al., 2009) galaxies, and independently by Lara-López et al. (2010) with ∼33,000 SDSS

galaxies. The FMR constitutes a 3D surface with these three properties, for which metallicity is tightly dependent on stellar mass and SFR with a residual scatter of ∼0.05 dex (Mannucci et al., 2010), a reduction in the scatter observed in the MZR. The FMR is also observed to be redshift-invariant out to $z = 2.5$ (Mannucci et al., 2010, see also sources within the review of Maiolino & Mannucci 2019), suggesting that the observed evolution of the MZR over this redshift range is the result of observing different parts of the locally-defined FMR at different redshifts. Above $z = 2.5$, galaxies have lower metallicities than predicted by the locally-defined FMR (Mannucci et al., 2010; Troncoso et al., 2014; Onodera et al., 2016). These studies analyze galaxies at $z \gtrsim 3$, where the strong optical emission lines used for metallicity determination are again observable in the H-band and K-band.

To accurately constrain the evolution of the MZR and FMR across redshift, metallicities must be estimated via a method that is consistent at all redshifts. Ideally, this is accomplished through first measuring other intrinsic nebular properties that dictate the strength of the collisionally-excited emission lines necessary for oxygen abundance determination. This "direct" method estimates the electron temperature (T_e) and density (n_e) of nebular gas, in conjunction with flux ratios of strong oxygen lines to Balmer lines, to determine the total oxygen abundance (e.g., Izotov et al., 2006). Electron temperature is calculated via a temperature-sensitive ratio of strong emission lines, commonly [O III] λ5007, to auroral emission lines, such as [O III] λ4363 or O III] $\lambda\lambda$1661,1666, from the same ionic species. The [O III] λ4363 line and flux ratio of [O III] $\lambda\lambda$4959,5007/[O III] λ4363 is preferred as all lines lie in the rest-optical part of the electromagnetic spectrum. However, the [O III] λ4363 line is faint, $\sim 40 - 100$ times weaker than [O III] λ5007 in low, sub-solar metallicity

galaxies, and still weaker in higher-metallicity sources where metal cooling is more efficient. This makes observing the line difficult locally, and especially difficult at high redshift. Only 11 galaxies at $z > 1$ have been detected (most via gravitational lensing) with significant [O III] λ4363 (Yuan & Kewley, 2009; Brammer et al., 2012b; Christensen et al., 2012; Stark et al., 2013; James et al., 2014; Maseda et al., 2014; Sanders et al., 2016a, 2019), and of those only 3 are at $z > 2$ (Sanders et al., 2016a, 2019).

In an effort to circumvent this problem and extend our ability to measure oxygen abundance to both high-metallicity and high-redshift galaxies, "strong-line" methods were developed to estimate abundances via flux ratios of strong, nebular emission lines (e.g., Jensen et al., 1976; Alloin et al., 1979; Pagel et al., 1979; Storchi-Bergmann et al., 1994). These indirect methods utilize calibrations of the correlations between these strong-line ratios and metallicities derived empirically with direct metallicity measurements of nearby H II regions and galaxies (e.g., Pettini & Pagel, 2004; Pilyugin & Thuan, 2005), theoretically with photoionization models (e.g., McGaugh, 1991; Kewley & Dopita, 2002; Dopita et al., 2013), or with a combination of both (e.g., Denicoló et al., 2002). However, as almost all of these calibrations have been done locally due to the inherent observational difficulties of the T_e-based, direct method (see Jones et al. 2015 for the first calibrations done at an appreciable redshift, $z \sim 0.8$), the question has naturally arisen as to whether these calibrations are accurate at high redshift.

With the statistical spectroscopic samples of high-redshift galaxies that now exist, there is evidence that physical properties of high-z, star-forming regions are different than what are observed locally. This is typically shown with the well-known offset of the locus of

star-forming, high-redshift galaxies relative to that of local, star-forming SDSS galaxies in the [O III] λ5007/Hβ vs. [N II] λ6583/Hα Baldwin−Phillips−Terlevich (N2-BPT; Baldwin et al., 1981) diagnostic diagram (Steidel et al., 2014; Shapley et al., 2015; Sanders et al., 2016b; Kashino et al., 2017; Strom et al., 2017). Numerous studies have tried to explain the primary cause of this evolution with various conclusions. It has been suggested that the offset derives from an elevated ionization parameter (Brinchmann et al., 2008; Cullen et al., 2016; Kashino et al., 2017; Hirschmann et al., 2017), elevated electron density (Shirazi et al., 2014), harder stellar ionizing radiation (Steidel et al., 2014; Strom et al., 2017, 2018), and/or an increased N/O abundance ratio in high-z galaxies (Masters et al., 2014; Shapley et al., 2015; Sanders et al., 2016b). It is also possible that there is no single primary cause, and the offset is due to a combination of the aforementioned property evolutions (Kewley et al., 2013; Maiolino & Mannucci, 2019). Nevertheless, there is considerable motivation to check the validity of locally-calibrated, strong-line metallicity methods at high redshift which utilize the emission lines in the N2-BPT plot and emission lines of other diagnostic diagrams, such as the S2-BPT variant ([O III] λ5007/Hβ vs. [S II] $\lambda\lambda$6716,6731/Hα) and the O_{32} vs. R_{23} (see Equations 2.2 and 2.3, respectively) excitation diagram.

In this paper, we present a detection of the auroral [O III] λ4363 emission line in a low-mass, lensed galaxy (A1689-217) at $z = 2.59$. We determine the direct metallicity of A1689-217 and combine it with other (re-calculated) direct metallicity estimates from the literature to examine the applicability of locally-calibrated, oxygen- and hydrogen-based, strong-line metallicity relations at high redshift. In Section 2.2 of this paper we give an overview of the spectroscopic and photometric observations of A1689-217 and their subse-

quent reduction. Section 2.3 discusses the emission-line spectrum of A1689-217, highlighting the detection of [O III] λ4363 and the method with which the spectrum was fit. Section 2.4 examines the physical properties of A1689-217 calculated from the photometry and spectroscopy. Section 2.5 discusses the results of the paper, focusing on the validity and evolution of strong-line metallicity relations with redshift, the evolution of ionization parameter with redshift, the position of A1689-217 in relation to the low-mass end of the FMR, and the position of A1689-217 relative to the predicted MZR from the FIRE hydrodynamical simulations. Section 2.6 gives a summary of our results. Section 2.7 revisits the [O III] λ4363 detection of Yuan & Kewley (2009) with a more sensitive spectrum of the galaxy, taken as part of our larger, dwarf galaxy survey. Throughout this paper, we assume a ΛCDM cosmology, with $H_0 = 70$ km s^{-1} Mpc^{-1}, $\Omega_\Lambda = 0.7$, and $\Omega_m = 0.3$.

2.2 Observations and Data Reduction

In this section, we discuss the spectroscopic and photometric observations and reduction for A1689-217, lensed by the foreground galaxy cluster Abell 1689. A1689-217 was initially detected via Lyman break dropout selection in the *Hubble Space Telescope* survey of Alavi et al. (2014, 2016). Based on its photometric redshift and high magnification (μ = 7.89), it was selected for spectroscopic observation of its rest-frame optical, nebular emission lines as part of a larger spectroscopic survey of star-forming, lensed, dwarf galaxies.

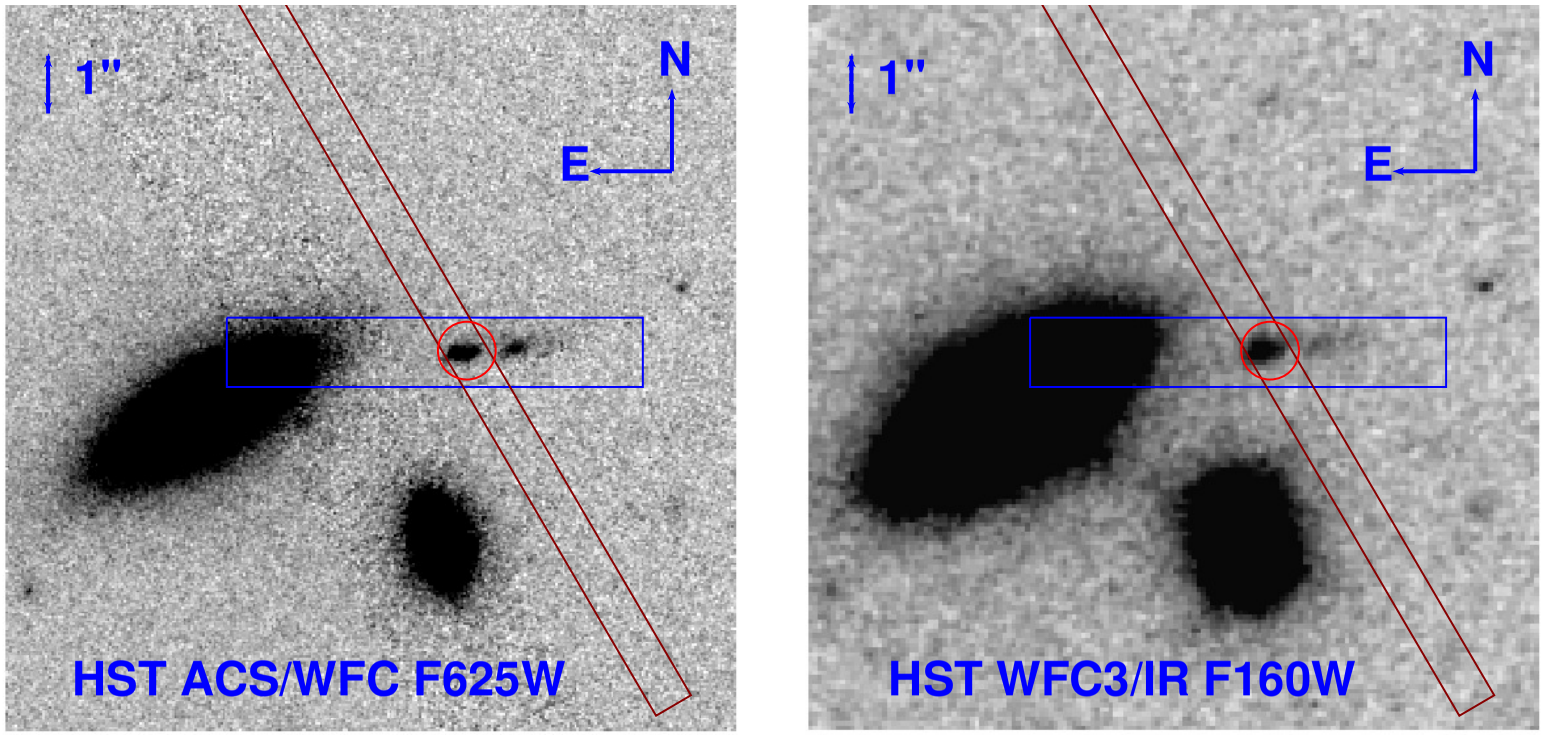

Figure 2.1: *HST* images of A1689-217 in the ACS/WFC F625W band and WFC3/IR F160W band. The 0.″7 MOSFIRE slit is shown in light brown, and the 1.″2 LRIS slit is shown in blue. A1689-217 is highlighted by the red circle. Foreground galaxies lie to the south and east of A1689-217. Both images are 12.″64 on each side.

2.2.1 Near-IR Spectroscopic Data

Near-IR (rest-optical) spectroscopic data for A1689-217 was taken on 2014 January 2 and 2015 January 17 with the Multi-Object Spectrometer for InfraRed Exploration (MOSFIRE; McLean et al., 2010, 2012) on the 10-m Keck I telescope. Spectroscopy was taken in the J, H, and K-bands with H-band and K-band data taken the first night (2014) and data in all three bands taken the second night (2015). J-band and H-band data consist of 120 second individual exposures while 180 second exposures were used in the K-band. In total, the integration time is 80 minutes in J-band, 104 minutes in H-band (56 minutes in 2014 and 48 minutes in 2015), and 84 minutes in K-band (60 minutes in 2014 and 24 minutes in 2015). The data were taken with a $0\farcs7$ wide slit (see orientation in Figure 2.1), giving spectral resolutions of $R \sim 3310$, 3660, and 3620 in the J, H, and K-bands, respectively. An ABBA dither pattern was utilized for all three filters with $1\farcs25$ nods for the J-band and $1\farcs2$ nods for the H and K-bands.

The spectroscopic data were reduced with the MOSFIRE Data Reduction Pipeline[1] (DRP). This DRP outputs 2D flat-fielded, wavelength-calibrated, background-subtracted, and rectified spectra combined at each nod position. Night sky lines are used to wavelength-calibrate the J and H-bands while a combination of sky lines and a neon arc lamp is used for the K-band. The 1D spectra were extracted using the IDL software, `BMEP`[2], from Freeman et al. (2019). The flux calibration of the spectra was first done with a standard star that was observed at an airmass similar to that of the A1689-217 observations, and then an absolute flux calibration was done using a star included in the observed slit mask.

[1] https://keck-datareductionpipelines.github.io/MosfireDRP/

[2] https://github.com/billfreeman44/bmep

2.2.2 Optical Spectroscopy

A deep optical (rest-frame UV) spectrum of A1689-217 was taken with the Low Resolution Imaging Spectrometer (LRIS; Oke et al., 1995; Steidel et al., 2004) on Keck I on 2012 February 24 with an exposure time of 210 minutes. The slit width was $1\farcs2$, and the slit was oriented E-W, as seen in Figure 2.1. We used the 400 lines/mm grism, blazed at 3400 Å, on the blue side. To reduce read-noise, the pixels were binned by a factor of two in the spectral direction. The resulting resolution is $R \sim 715$. The individual exposures were rectified, cleaned of cosmic rays, and stacked using the pipeline of Kelson (2003).

2.2.3 Near-UV, Optical, and Near-IR Photometry

Near-UV images of the Abell 1689 cluster, all of them covering A1689-217, were taken with the WFC3/UVIS channel on the *Hubble Space Telescope*. We obtained 30 orbits in the F275W filter and 4 orbits in F336W with program ID 12201, followed by 10 orbits in F225W and an additional 14 orbits in F336W (18 orbits total) with program ID 12931. The data were reduced and photometry was measured as described in Alavi et al. (2014, 2016).

In the optical, we used existing *HST* ACS/WFC images in the F475W, F625W, F775W, and F850LP filters (PID: 9289, PI: H. Ford) as well as in the F814W filter (PID: 11710, PI: J. Blakeslee), calibrated and reduced as detailed in Alavi et al. (2014). The number of orbits and the 5σ depths measured within a $0\farcs2$ radius aperture for all optical and near-UV filters are given in Alavi et al. (2016, Table 1). In the near-IR, we used existing WFC3/IR images in the F125W and F160W filters (PID: 11802, PI: H. Ford), both with 2,512 second exposure times.

Images of A1689-217 in the optical F625W filter and near-IR F160W filter are shown in Figure 2.1.

2.3 Emission-Line Spectrum of A1689-217

The MOSFIRE spectra yield several emission lines necessary for the direct measurement of intrinsic nebular properties of A1689-217, located at $z = 2.5918$ (see Section 2.3.2). Seen in both 1D and 2D in Figure 2.2, we strongly detect [O II] $\lambda\lambda$3726,3729, Hγ, Hβ, [O III] λ4959, and Hα. We also detect the auroral [O III] λ4363 line in the H-band (discussed in greater detail in Section 2.3.1). The [O III] λ5007 emission line, necessary for electron temperature (T_e) measurements, is not shown in Figure 2.2 because it sits at the edge of the H-band filter where transmission declines rapidly, and the flux calibration is uncertain. We instead scale up from the [O III] λ4959 line flux using the T_e-insensitive intrinsic flux ratio of the doublet, [O III] λ5007/[O III] λ4959 = 2.98 (Storey & Zeippen, 2000). We also note the lack of a significant detection of the [N II] $\lambda\lambda$6548,6583 doublet in this spectrum, placing A1689-217 in the upper-left corner of the N2-BPT diagnostic diagram as seen in Figure 2.3. We conclude that A1689-217 is not an AGN based on its very low [N II]/Hα ratio, lack of high-ionization emission lines like [Ne V], and narrow line widths ($\sigma_{\mathrm{H}\beta} \approx 53$ km s^{-1}). The optical spectrum shows strong Lyα emission (see Figure 2.4) with a rest-frame equivalent width of $\mathrm{EW}_{0,\mathrm{Ly}\alpha} = 138$ Å, redshifted by 282 km s^{-1}. The slit-loss-corrected, observed emission-line fluxes and uncertainties are given in Table 2.1 with the line-fitting technique described in Section 2.3.2.

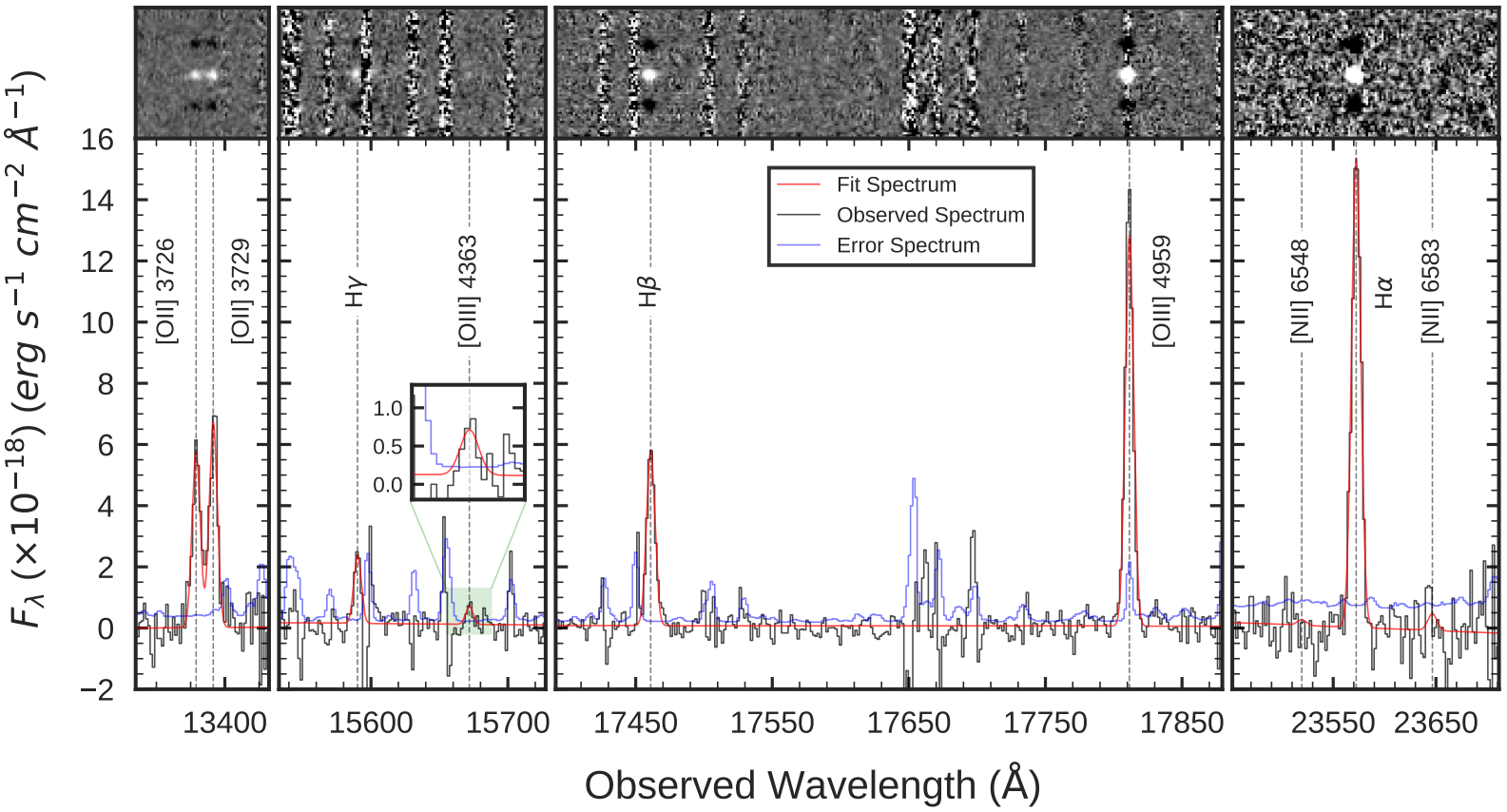

Figure 2.2: The $z = 2.5918$ observed spectrum of A1689-217 in the J, H, and K-bands of Keck/MOSFIRE. The top panel shows the two-dimensional spectrum while the bottom panel shows the observed (black), error (blue), and single-Gaussian fit (red) spectra in one dimension. The emission lines are labeled for reference. The portion of the spectrum containing [O III] λ4363 has been highlighted in green and magnified in the inset plot. A peak can be seen at the observed location of the line among 4 consecutive pixels with S/N > 1. We report a total significance in the detection of 4.2σ. Emission of [O III] λ4363 in the two-dimensional spectrum is also visible along with the expected symmetric negative images on either side resulting from nodding along the slit.

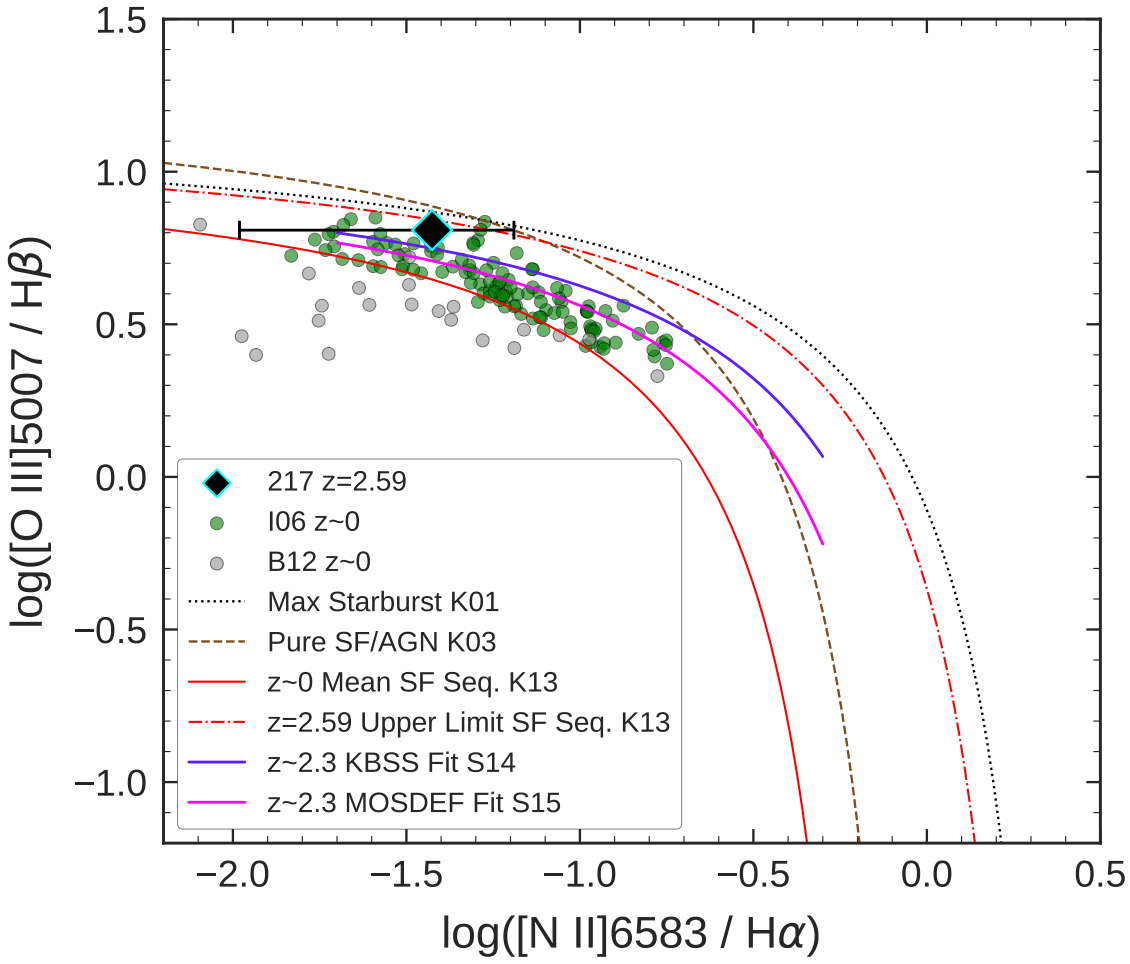

Figure 2.3: The [O III] λ5007/Hβ vs. [N II] λ6583/Hα N2-BPT diagram. A1689-217 is denoted by the black diamond with cyan border and lies offset from the $z \sim 0$ mean star-forming sequence of Kewley et al. (2013, K13) (solid red line). The galaxy displays high excitation and a very low [N II]/Hα ratio, with the large error bars resulting from the lack of a significant [N II] λ6583 detection. The green and gray points represent the $z \sim 0$ comparison samples (see Section 2.5) of Izotov et al. (2006, I06) and Berg et al. (2012, B12), respectively. The dotted black line is the "maximum starburst" curve from Kewley et al. (2001, K01). The dashed brown line is the demarcation between star-forming galaxies and AGN from Kauffmann et al. (2003, K03). The purple line is the best fit to the $z \sim 2.3$ star-forming galaxies in Steidel et al. (2014, S14) while the magenta line is the best fit to the $z \sim 2.3$ star-forming galaxies in Shapley et al. (2015, S15). The red, dot-dashed line represents the theoretical $z = 2.59$ upper-limit, star-forming abundance sequence as given by Kewley et al. (2013, K13).

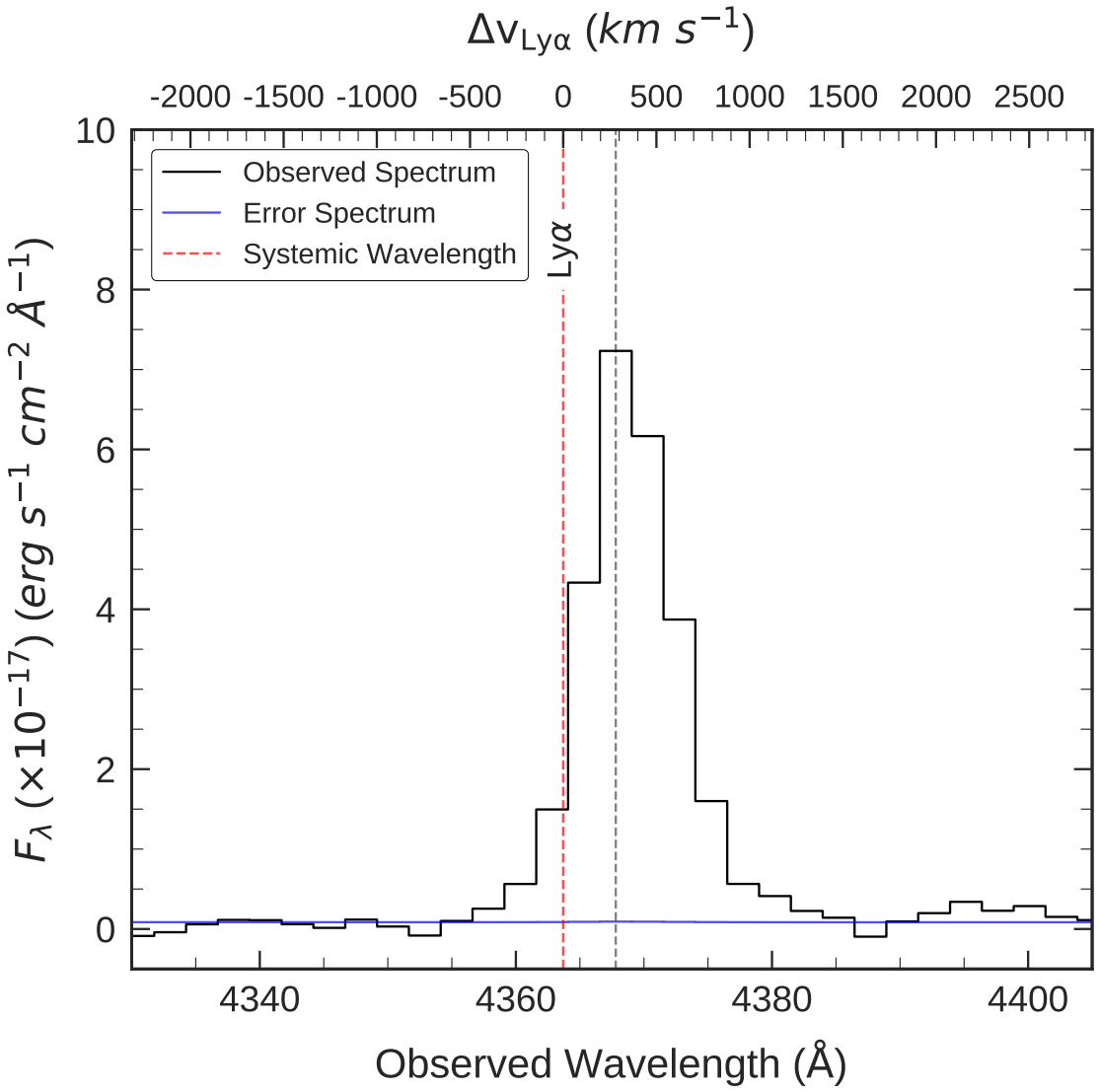

Figure 2.4: The Lyα emission line of A1689-217, observed with Keck/LRIS. The observed and error spectra are shown in black and blue, respectively. The systemic wavelength of Lyα is denoted by the dashed red line. The observed peak of the Lyα line, marked by the dashed gray line, displays a velocity offset (labeled on the upper x-axis) from the systemic redshift of $\Delta v_{\mathrm{Ly}\alpha} = 282$ km s^{-1}.

Table 2.1: Emission-line fluxes and rest-frame equivalent widths for A1689-217.

Line	$\lambda^a_{\rm rest}$	$\lambda_{\rm obs}$	$f^b_{\rm obs}$	$f^{b,c}_{\rm corr}$
[O II]	3726.03	13 383.21	40.8 ± 1.7	222 ± 9
[O II]	3728.82	13 393.21	47.3 ± 2.2	257 ± 12
Hγ^d	4340.46	15 590.12	18.3 ± 1.4	81 ± 6
[O III]	4363.21	15 671.84	4.8 ± 1.1	21 ± 5
Hβ^d	4861.32	17 460.96	53.2 ± 1.4	192 ± 5
[O III]	4958.91	17 811.48	118.7 ± 4.9	414 ± 17
Hα^d	6562.79	23 572.34	206.0 ± 6.9	507 ± 17
[N II]	6583.45	23 646.52	7.8 ± 5.6	19 ± 14
EW$_0$(Lyα)^e				$137.9^{+8.3}_{-8.5}$
EW$_0$([O III]λ5007)				860.4 ± 52.2
EW$_0$(Hα)				520.7 ± 28.7

Notes: The [O III]λ5007 line lies at the edge of the H-band filter, so the flux for this line is found via the intrinsic flux ratio of the doublet: [O III]λ5007/[O III]λ4959 = 2.98.

[a]Rest-frame wavelengths in air (Å)

[b]Fluxes are in units of 10^{-18} erg s^{-1} cm^{-2} and are uncorrected for lens magnification. $f_{\rm obs}$ and $f_{\rm corr}$ refer to the observed and dust-corrected fluxes, respectively. Both $f_{\rm obs}$ and $f_{\rm corr}$ are slit-loss-corrected.

[c]The intrinsic flux uncertainties do not include other systematic errors associated with inter-filter calibrations and dust correction, though these additional errors are propagated throughout all of our calculations.

[d]Emission-line fluxes not corrected for underlying stellar absorption as these corrections are small and uncertain (see Section 2.4.2)

[e]Rest-frame equivalent widths in Å

2.3.1 Detection of [O III] λ4363

We report a 4.2σ detection of the T_e-sensitive, auroral [O III] λ4363 line. In Figure 2.2, there is visible emission in the 2D spectrum at the observed wavelength and spatial coordinates expected for the emission line (as well as the expected symmetric negative images on either side resulting from nodding along the slit). In the magnified inset plot of the highlighted region of the 1D spectrum, there is a clear peak centered at the observed wavelength expected for [O III] λ4363 at $z = 2.5918$. We note that this peak is part of 4 consecutive pixels that have a S/N > 1. We also note that at A1689-217's redshift, the [O III] λ4363 line is not subject to sky line contamination and thus conclude that this detection is robust.

2.3.2 Fitting the Spectrum

The spectrum of A1689-217 was fit using the Markov Chain Monte Carlo (MCMC) Ensemble sampler `emcee`[3] (Foreman-Mackey et al., 2013). In each filter we fit single-Gaussian profiles to the emission lines and a line to the continuum. In the H-band, due to the large wavelength separation between Hβ and [O III] λ4363, Hβ and [O III] λ4959 were fit separately from Hγ and [O III] λ4363. While the width and redshift were free parameters in the H and K-bands, in the H-band they were only fit with the much higher S/N lines of Hβ and [O III] λ4959 and then adopted for Hγ and [O III] λ4363. In the J-band, due to the small wavelength separation of the [O II] doublet, and thus the partial blending of the lines (seen in Figure 2.2), the redshift and width were taken to be the values fit to the highest S/N line

[3] https://emcee.readthedocs.io/en/v2.2.1/

in the spectrum (Hβ). The redshift of A1689-217 reported in this paper (see Table 2.2) is the weighted average of the redshifts fit to the H and K-bands.

2.4 Properties of A1689-217

Estimates of various physical properties of A1689-217 are summarized in Table 2.2, with select properties discussed in greater detail in the sections below.

2.4.1 Stellar Mass and Age

The stellar mass is estimated by fitting stellar population synthesis models to the *HST* optical and near-IR photometry. Because some of the emission lines have high equivalent widths (see Table 2.1), we have corrected the photometry by subtracting the contribution from the emission lines (e.g., Lyα, [O II] $\lambda\lambda$3726,3729, Hγ, [O III] λ4363). We have also added in quadrature an additional 3% flux error in all bands to account for systematic errors in the photometry (Alavi et al., 2016). We use the stellar population fitting code `FAST`[4] (Kriek et al., 2009) with the Bruzual & Charlot (2003) stellar population synthesis models, and a constant star formation rate with a Chabrier initial mass function (IMF; Chabrier, 2003). As suggested by Reddy et al. (2018a) for high-redshift, low-mass galaxies, we use the SMC dust extinction curve (Gordon et al., 2003) with A_V values varying between $0.0 - 2.0$. We fix the metallicity at 0.2 $Z_\odot$ and the redshift at the spectroscopic value. The stellar age can vary between $7.0 < \log(t)$ [yr] < 10.0. The 1σ confidence intervals are derived from a Monte Carlo method of perturbing the broadband photometry within the corresponding photometric uncertainties and refitting the spectral energy distribution (SED)

[4] `http://w.astro.berkeley.edu/~mariska/FAST.html`

300 times. The best-fit parameters for A1689-217, corrected for the lensing magnification factor, $\mu = 7.89$, when necessary, are $A_V = 0.25$, $\log(M_*/M_\odot) = 8.07$, SFR $= 2.75$ $M_\odot$ yr^{-1}, and $t_{\rm age} \sim 50$ Myr, with the best-fit, de-magnified SED model shown in Figure 2.5.

The young age of the stellar population is perhaps not surprising as the large Hα equivalent width (EW$_{0,\rm H\alpha} = 521$ Å) strongly suggests that A1689-217 is undergoing an intense burst of star formation, as seen in a subset of galaxies at high redshift (Atek et al., 2011; van der Wel et al., 2011; Straughn et al., 2011; Atek et al., 2014; Tang et al., 2018). Because the stellar population associated with this recent burst is young, it has a low mass-to-light ratio and can easily be hiding a significant mass in older stars. To understand how much stellar mass we might be missing, we investigated adding a maximally old stellar population, formed in a single burst at $z = 6$ (1.6 Gyr old at $z = 2.5918$). We found that the stellar mass could be increased by a factor of 3.3 before the reduced χ^2 is increased by a factor of two (seen in Figure 2.5). Thus, we use 3.3$\times$ the mass from the SED fit, or $\log(M_*/M_\odot) < 8.59$, as the upper-limit of the stellar mass.

We note that many of the high-redshift galaxies with [O III] λ4363 detections have high equivalent width Balmer lines and may selectively be in a burst relative to the typical galaxy at these redshifts (Ly et al., 2015). Thus, a simple star formation history fit to the photometry might be dominated by the recent burst and will significantly underestimate the stellar mass. This is important to consider when ultimately trying to measure the MZR with these galaxies.

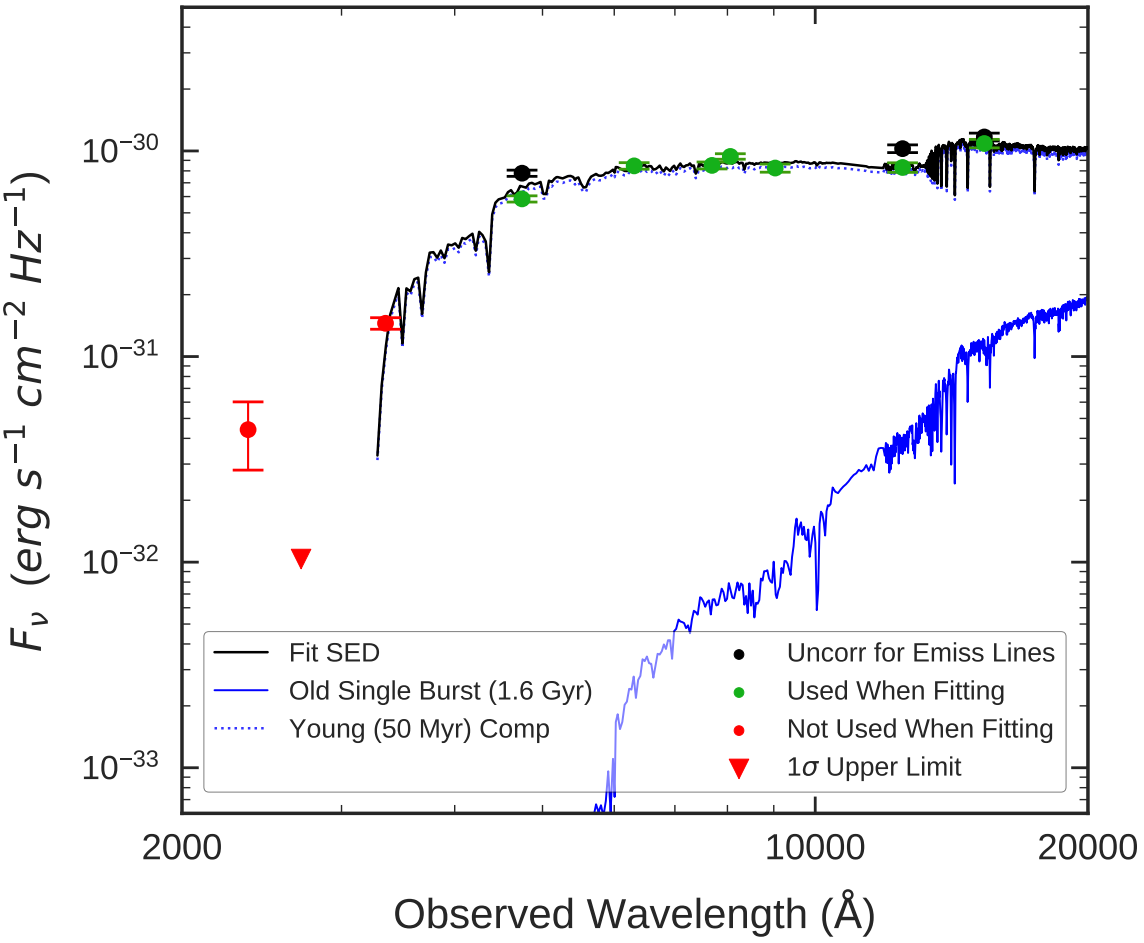

Figure 2.5: The de-magnified, observed photometry and best-fit SED model (black line) for A1689-217. The green data points represent the emission-line-subtracted photometry used for the SED fitting. The black data points represent the photometry before correction for emission lines. The red points signify WFC3/UVIS photometry not used in the fitting because of Lyα-forest absorption. An additional 3% flux error, used to account for systematic errors in the photometry, has been added in quadrature to the flux errors in each of the bands prior to SED fitting and is reflected in the error bars of all (green, black, and red) photometric data points. The SED redshift is fixed to the spectroscopic value of $z_{\rm spec} = 2.5918$. The best-fit model indicates a young stellar population ($\sim$ 50 Myrs). Also plotted is a maximally-old (1.6 Gyr) stellar population (blue solid line) that can be added to the fit while slightly scaling down the best-fit, constant-SFR SED (blue dotted line). Adding this older component can increase the stellar mass by a factor of 3.3 at a doubling of the reduced χ^2, so it is treated as an upper limit to the stellar mass.

2.4.2 Nebular Extinction and Star Formation Rate

To properly estimate galactic properties and conditions within the interstellar medium (ISM), several of which rely on flux ratios, the wavelength-dependent extinction from dust must be accounted for. This extinction can be quantified with Balmer line ratios calculated from observed hydrogen emission-line fluxes. With the strong detections of Hγ, Hβ, and Hα in the spectrum of A1689-217, we estimate the extinction due to dust by assuming Case B intrinsic ratios of Hα/Hβ = 2.79 and Hα/Hγ = 5.90 for T_e = 15,000 K and n_e = 100 cm^{-3} (Dopita & Sutherland, 2003), approximately the electron temperature and density of A1689-217 (see Section 2.4.3).[5] We note the presence of underlying stellar absorption of the Balmer lines in Figure 2.5 but do not make any corrections to the emission-line fluxes of Hγ, Hβ, or Hα here as these corrections amount to small percent differences in the fluxes of $\sim$ 3.5%, $\sim$ 1.1%, and $\sim$ 0.1%, respectively, and are also based on an uncertain star formation history. Assuming the extinction curve of Cardelli et al. (1989) with an R_V = 3.1, we find the color excess to be $E(B-V)_{\rm gas} = A_V/R_V = 0.39 \pm 0.05$. We use this result to correct the observed emission-line fluxes for extinction due to dust and list the corrected values in Table 2.1. We note that the nebular extinction is significantly higher than the best-fit extinction of the stellar continuum derived from the SED fit ($A_V = 0.25$) and indicated by the flat (in f_ν) SED seen in Figure 2.5. This difference in nebular vs. stellar extinction is likely due to the young age of the burst, indicating that the nebular regions are still enshrouded within their birth cloud (Charlot & Fall, 2000). We also note here that some T_e-derived metallicities at high redshift are calculated with dust corrections

[5]The variation in the intrinsic Balmer line ratios with temperature is small over the temperature range typical of H II regions. We obtain $T_e \sim$ 15,000 K after correcting for dust regardless of using the Balmer ratios corresponding to 15,000 K or the commonly assumed 10,000 K.

based on the stellar SEDs. If many of these galaxies are in a burst of recent star formation, the stellar attenuation may not be a reliable indicator of the nebular extinction. This is especially concerning for galaxies with O III] $\lambda\lambda$1661,1666 detections (rest-UV auroral lines used to estimate T_e) instead of [O III] λ4363, as the attenuation at these wavelengths is much larger.

The star formation rate (SFR) of A1689-217 is calculated with the galaxy's dust-corrected Hα luminosity (L(Hα)) and the relation between SFR and L(Hα) from Kennicutt (1998). The conversion factor of the relation is re-calculated assuming a Chabrier (2003) IMF with 0.2 $Z_\odot$, roughly the oxygen abundance of A1689-217 (see Section 2.4.4). The resulting SFR is divided by the magnification factor (μ = 7.89) from the lensing model. We estimate that A1689-217 has a SFR = 16.2 $\pm$ 1.8 $M_\odot$ yr^{-1}. The uncertainty in this measurement does not include the uncertainty in the magnification as the magnification and its error are dependent on the assumptions inherent to the lensing model. We also note here that the Hα-derived SFR is nearly six times larger than the SED-derived SFR. Much of this discrepancy can be explained if the stellar population has a harder ionizing spectrum due to low Fe abundance (Steidel et al., 2014) and/or binary stellar evolution (Eldridge & Stanway, 2009). A harder ionizing spectrum produces more ionizing photons, seen in the Hα recombination line, relative to the non-ionizing UV and thus should yield Hα-based SFRs that are larger than those derived via fitting to rest-UV photometry.

2.4.3 Electron Temperature and Density

The electron temperature (T_e) and electron density (n_e) are intrinsic nebular properties that are responsible for the strength of collisionally-excited lines that allow for a

direct measurement of the gas-phase metallicity of H II regions. We calculate the electron temperature in the O^{++} region, T_e([O III]), using the temperature-sensitive line ratio [O III] $\lambda\lambda$4959,5007/[O III] λ4363 and the IRAF task NEBULAR.TEMDEN (Shaw & Dufour, 1994). This temperature-sensitive ratio is dependent on electron density, though below $n_e \approx 10^3$ cm^{-3} – the low-density regime within which A1689-217 and this paper's literature comparison sample reside – T_e([O III]) is insensitive to the density (Osterbrock & Ferland, 2006). We therefore calculate T_e([O III]) non-iteratively, assuming a fiducial electron density of $n_e = 150$ cm^{-3}, appropriate for H II regions (Sanders et al., 2016b). This yields a result of T_e([O III]) = 14,300 ± 1,500 K.[6] To calculate the electron temperature in the O^+ region, T_e([O II]), the auroral doublet [O II] $\lambda\lambda$7320,7330 is needed. These lines are not within our wavelength coverage, so we utilize the T_e([O III]) – T_e([O II]) relation of Campbell et al. (1986) to obtain an electron temperature in the O^+ region of T_e([O II]) = 13,000 ± 1,100 K.

The electron density is estimated with the doublet ratio [O II] λ3729/[O II] λ3726 and the IRAF task NEBULAR.TEMDEN. The aforementioned T_e([O II]) = 13,000 K is used in the calculation. We obtain an electron density for A1689-217 of $n_e = 220^{+70}_{-60}$ cm^{-3}. This measurement is consistent with the typical electron density found by Sanders et al. (2016b) for $z \sim 2.3$ star-forming galaxies, ~ 250 cm^{-3}, a factor of ~ 10 higher than densities in local star-forming regions. It should be noted, however, that while our measurement agrees with Sanders et al. (2016b) and others (e.g., Steidel et al., 2014; Kashino et al., 2017), our galaxy is $\sim 0.9 - 1.4$ dex lower in stellar mass (see Section 2.4.1 and Figure 2.5) than the

[6] Assuming any $n_e < 1{,}000$ cm^{-3} results in variations of our calculated T_e of $< 0.5\%$.

mass ($\sim 10^{9.5}$ $M_\odot$) above which Sanders et al. (2016b) is confident their density estimate holds true.

2.4.4 Oxygen Abundance

The oxygen abundance, or gas-phase metallicity, is calculated using the analytic ionic abundance expressions of Izotov et al. (2006). These equations make use of the values found for T_e([O II]), T_e([O III]), and n_e from the previous section. We assume that the oxygen abundance comprises contributions from the populations of the O^+ and O^{++} zones of an H II region with negligible contributions from higher oxygen ionization states.

$$\frac{O}{H} \approx \frac{O^+}{H^+} + \frac{O^{++}}{H^+} \tag{2.1}$$

We calculate an oxygen abundance for A1689-217 of 12+log(O/H) = 8.06 ± 0.12 (0.24 $Z_\odot$; Asplund et al., 2009).

2.4.5 Uncertainties

To calculate the 1σ uncertainties of the intrinsic emission-line fluxes, flux ratios, and other properties of A1689-217, we utilize a Monte Carlo approach in which a given value is sampled $N = 10^5$ times. The uncertainties in the intrinsic emission-line fluxes are found by first sampling the probability distribution of A1689-217's extinction in the visual band (A_V), needed for the extinction at a given wavelength (A_λ), and the probability distribution of each emission line's observed flux. The final probability distribution of A_V is the result of multiplying the probability distributions of A_V found for each of the Balmer decrements

Table 2.2: Properties of A1689-217.

Property	Value
R.A. (J2000)	$13^{\rm h}\,11^{\rm m}\,27\overset{\rm s}{.}62$
Dec. (J2000)	$-01^\circ\,21'\,35\overset{''}{.}62$
z	2.591 81
	$\pm$ 0.000 01
μ	7.89 ± 0.40
$\log(M_*/M_\odot)^{a,b}$	$8.07 - 8.59$
$M^a_{\rm UV,1700}$	-18.67 ± 0.04
$E(B-V)_{\rm gas}$	0.39 ± 0.05
SFR^a [M$_\odot$ yr^{-1}]	16.2 ± 1.8
n_e [cm^{-3}]	220^{+70}_{-60}
T_e([O II]) [K]	$13\,000 \pm 1100$
T_e([O III]) [K]	$14\,300 \pm 1500$
12+log(O$^+$/H$^+$)	7.56 ± 0.12
12+log(O^{++}/H$^+$)	7.90 ± 0.12
12+log(O/H)	8.06 ± 0.12
Z [Z$_\odot$]	$0.24^{+0.08}_{-0.06}$

[a]Most probable value corrected for the listed magnification factor, μ. The uncertainty does not include the uncertainty in the magnification.

[b]The lower and upper bounds of the stellar mass estimate. The lower bound corresponds to our best-fit SED model ($t \sim 50$ Myr), and the upper bound corresponds to a young stellar component ($t = 50$ Myr) in combination with a 1.6 Gyr old burst component. See Section 2.4.1 and Figure 2.5 for further details.

considered for A1689-217, Hα/Hβ and Hα/Hγ, the uncertainty for each ratio coming from its observed statistical error added in quadrature with a 5% inter-filter systematic error. The visual-band extinction and the emission lines are each sampled N times from a normal distribution centered on the most probable A_V or observed flux, respectively, with a standard deviation given by the 1σ error of the value being sampled. The A_V values are then used to calculate N extinction magnitudes for each emission line, with which each iteration of each emission-line sample is dust-corrected, giving a sample of N intrinsic fluxes for each line. A posterior histogram is then generated for the intrinsic flux of each line, and a 68% confidence interval is fit, allowing a 1σ uncertainty to be determined for each line's intrinsic flux.

In the calculation of the flux-ratio uncertainties, we take the samples of intrinsic emission-line fluxes and calculate N-length samples of the desired flux ratios, for which posterior histograms are created and 1σ errors estimated as for the intrinsic emission-line fluxes. The properties of A1689-217 have their uncertainties estimated in the same manner.

2.5 Discussion

2.5.1 Strong-Line Ratio – Metallicity Diagnostics

Having calculated the intrinsic emission-line fluxes and direct-metallicity estimate of A1689-217, we study the evolution of both nebular physical properties and the relationships between strong-line ratios and T_e-based metallicities.

Jones et al. (2015) presented the first calibrations between strong-line ratios and direct metallicities at significant redshift, utilizing a sample of 32 star-forming galaxies at

$z \sim 0.8$ from the DEEP2 Galaxy Redshift Survey (Davis et al., 2003; Newman et al., 2013). Because the flux ratio of [O III]λ4363/[O III]λ5007 is generally $\lesssim 3\%$, random noise creates a large scatter in the measurement of this temperature-sensitive ratio. To combat this effect, all 32 galaxies in the Jones et al. sample were selected because they have high S/N in [O III]λ5007 and low noise in the location of [O III]λ4363. More specifically, the galaxies in the sample have a ratio of [O III]λ5007 flux to uncertainty in the [O III]λ4363 flux (f_{5007}/σ_{4363}) of $\geqslant 300$. This ratio, which they call the "sensitivity" (this term used hereafter to denote this ratio), not only reduces the effects of random noise but also the bias toward very low metallicity ($12+\log(\mathrm{O/H}) \lesssim 8.3-8.4$ or $Z \lesssim 0.4-0.5$ $\mathrm{Z}_{\odot}$) galaxies that comes with selecting a sample via [O III]λ4363 significance instead (see their Figure 1).

Jones et al. (2015) found that the relations between direct metallicity and ratios of neon, oxygen, and hydrogen emission lines derived from their sample are consistent (albeit with larger uncertainties) with the relations derived from a subset (subject to the same sensitivity requirement) of the $z \sim 0$ star-forming galaxies from Izotov et al. (2006) – a subsample itself from Data Release 3 of the Sloan Digital Sky Survey (Abazajian et al., 2005). Jones et al. showed that these relations do not evolve from $z = 0$ to $z \sim 0.8$.

2.5.1.1 Comparison Samples Across Cosmic Time

In a similar manner to Jones et al. (2015) and Sanders et al. (2016a) with their object COSMOS-1908, we will use the measurements of A1689-217, compared to other [O III]λ4363 sources at various redshifts, to further study the evolution of the calibrations in Jones et al. (2015), particularly at higher redshift. We note that unlike in Jones et al.

(2015) and Sanders et al. (2016a), the relations involving [Ne III]λ3869 are not studied here because this line falls out of our spectroscopic coverage of A1689-217.

In addition to the 32, $z \sim 0.8$ galaxies from Jones et al. (2015), we also consider two local, $z \sim 0$ comparison samples: 113 star-forming galaxies with spectral coverage of the optical [O II] doublet from Izotov et al. (2006) – the same $z \sim 0$ sample used in Jones et al. (2015) – and 28 H II regions (21 total galaxies) from Berg et al. (2012). The galaxies from Berg et al. (2012) comprise a low-luminosity subsample of the *Spitzer* Local Volume Legacy (LVL) catalog (Dale et al., 2009) and have high-resolution MMT spectroscopy for [O III]λ4363 detection. This particular sample was chosen because of its low-luminosity and the volume-limited – as opposed to flux-limited – nature of its parent LVL sample, the combination of which allows for the statistical study of local dwarf galaxies ($5.90 \leqslant \log(M_*/M_\odot) \leqslant 9.43$ here). These Berg et al. sample qualities are similar to those of our high-z parent survey, to which A1689-217 belongs, in the sense that we are looking at very low-mass objects (via lensing) in a small volume as opposed to less-typical, more luminous objects in a larger volume.

Both of the local comparison samples adhere to the sensitivity cut placed on the Jones et al. (2015) sample. Additionally, as in Izotov et al. (2006), we arrived at our stated comparison sample sizes by removing all galaxies (or H II regions) with both [O III]λ4959/Hβ < 0.7 and [O II]λ3727/Hβ > 1.0, ensuring high-excitation samples that do not discriminate against very metal-deficient sources with high excitation. Global oxygen abundance and strong-line ratio values for galaxies in the Berg et al. (2012) sample with multiple H II regions meeting these cuts are taken as the average of the individual H II region values,

weighted by the uncertainties calculated for the abundances and ratios, respectively, as detailed in Section 2.4.5.

At low-to-intermediate redshifts, we also include 9 of the 20, $z < 0.9$, high-sSFR galaxies with [O III]λ4363 detections from Ly et al. (2014) and the Subaru Deep Field (Kashikawa et al., 2004), excluding the rest of the sample due to the inability to determine dust corrections, unreliable T_e estimates, missing Hβ or stellar mass (necessary for our study of the FMR in Section 2.5.4), and the presence of a LINER. Due to this sample being so small, we do not apply the sensitivity cut of Jones et al. (2015), which would remove 5 of the 9 objects, but note that all galaxies pass the cut of Izotov et al. (2006).

In addition to the low- and intermediate-redshift samples, we also compare A1689-217 to the galaxies of James et al. (2014) at $z = 1.43$, Stark et al. (2013) at $z = 1.43$, Christensen et al. (2012) at $z = 1.83$, and Sanders et al. (2016a) at $z = 3.08$. Each of these galaxies has an [O III]λ4363 detection and corresponding, re-calculated, direct metallicity estimate. We do not compare to the galaxy reported in Yuan & Kewley (2009) as our deeper spectrum of this galaxy shows that the claimed [O III]λ4363 detection is not correct. See Section 2.7 for more details. All comparison samples in this paper, at $z \sim 0-3.1$, are dust-corrected using the Cardelli et al. (1989) extinction curve, with an $R_V = 3.1$ (except for Jones et al. 2015, who use an $R_V = 4.05$ though show that their results are insensitive to this value), and have had their physical properties re-calculated using the methods detailed in Sections 2.4.3 and 2.4.4.

We do not include any O III]$\lambda\lambda$1661,1666 sources in our comparison samples as do some other similar studies (e.g., Patrício et al., 2018; Sanders et al., 2019) due to added

complications when considering both the optical and ultraviolet. These complications lie primarily in the very uncertain extinction law in the UV and the large wavelength separation between these auroral lines and [O III]λ5007, as well as in issues arising from observing in these different regimes (e.g., different instruments, slit widths, seeing).

2.5.1.2 The Evolution of the Strong-Line Ratio – Metallicity Calibrations

In our effort to further quantify the evolution at high redshift of the locally-calibrated, strong-line metallicity relations, as well as other physical properties, we consider the position of A1689-217, and the other high-redshift galaxies, in relation to the Jones et al. (2015) calibrations and other lower-redshift comparison samples in the four panels of Figure 2.6. We find that A1689-217 is consistent with the local best-fit relations of Jones et al. (2015) in the top two and bottom-left panels, given A1689-217's uncertainties and the relations' intrinsic scatter. We observe A1689-217 to be $\sim$1.6σ above the best-fit R_{23} (see Equation 2.3 for R_{23} ratio) relation at its metallicity of $Z = 8.06$, though we do not claim it to be inconsistent with the relation based on A1689-217's uncertainties in both parameters, especially oxygen abundance, combined with the scatter around the relation. A1689-217's elevated R_{23} value is a consequence of A1689-217 being above the local relation in the [O III]λ5007/Hβ ratio and especially in the [O II]$\lambda\lambda$3726,3729/Hβ ratio, though both ratios are consistent with the local calibrations. When also considering the other $z > 1$ sources in addition to A1689-217, we do not observe any significant systematic offsets in line ratio or metallicity for any of the relations. We therefore suggest that there is no evidence of evolution from $z \sim 0$ to $z \sim 3.1$ in the relations between direct metallicity and emission-

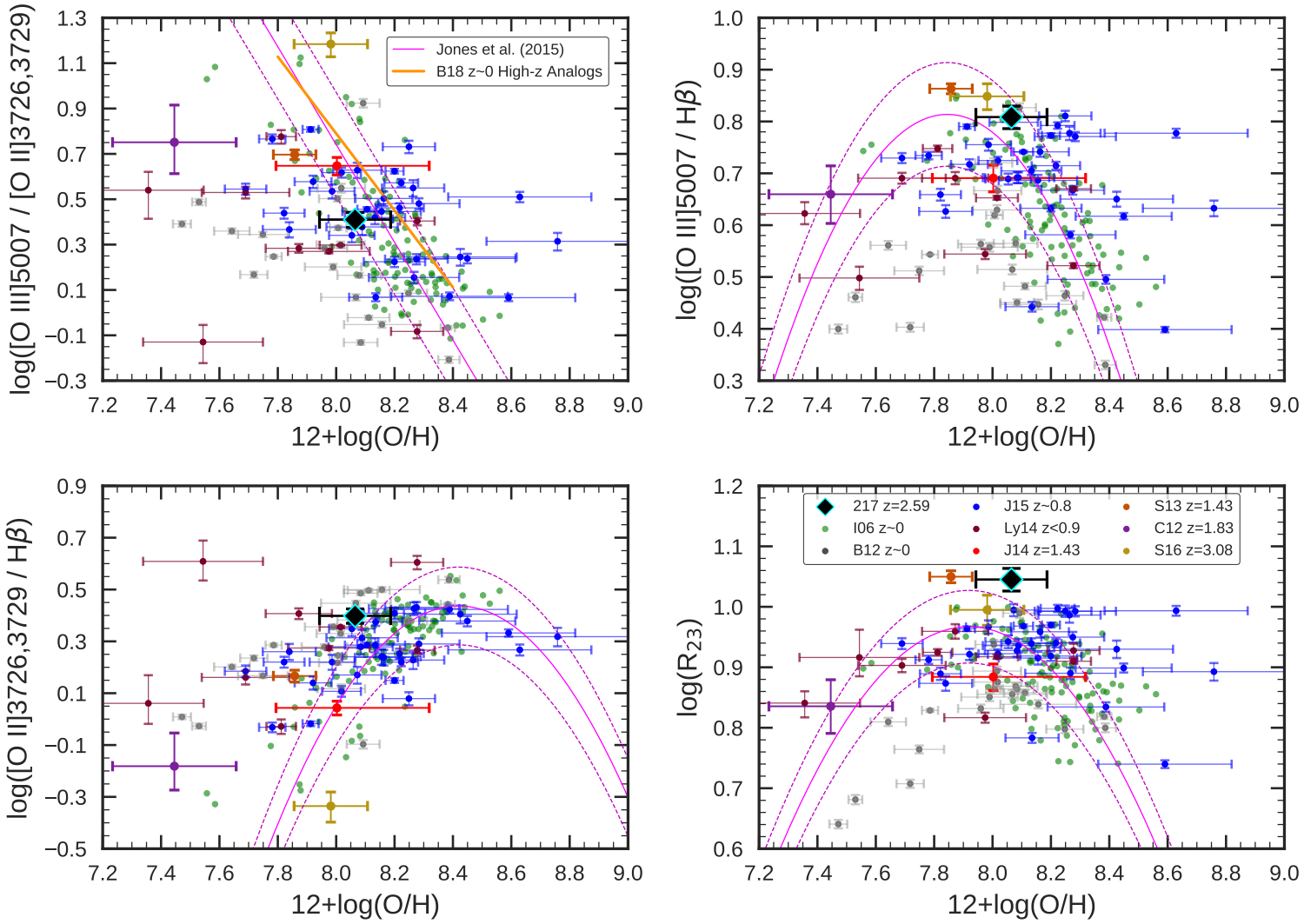

Figure 2.6: The full caption follows on the next page.

Figure 2.6: (Figure on previous page.) Strong emission-line ratios vs. direct-method oxygen abundance for A1689-217 and comparison samples ranging in redshift from $z \sim 0$ to $z \sim 3.1$. A1689-217 is denoted by the black diamond with cyan border. The $z \sim 0$ sample of Izotov et al. (2006, I06) is given by the green points. The $z \sim 0.8$ sample of Jones et al. (2015, J15) is given by the blue data points. The $z < 0.9$ sample of Ly et al. (2014, Ly14) is given by the dark red data points. The red, dark orange, purple, and gold points correspond to the $z = 1.43$ galaxy of James et al. (2014, J14), the $z = 1.43$ galaxy of Stark et al. (2013, S13), the $z = 1.83$ galaxy of Christensen et al. (2012, C12), and the $z = 3.08$ galaxy of Sanders et al. (2016a, S16), respectively. The solid magenta lines show the best-fit relations between the strong-line ratios and metallicity as determined by Jones et al. (2015) with the $z \sim 0$ Izotov et al. (2006) sample. The accompanying dashed magenta lines represent the 1σ intrinsic scatter around the best-fit relations. The orange line in the upper-left panel is the best-fit relation, based on stacked spectra of $z \sim 0$ high-z analogs, of Bian et al. (2018, B18). The sample of Berg et al. (2012, B12) $z \sim 0$ LVL galaxies is represented by the gray points and included to show the disparity between this low-excitation (see also Figure 2.3), low-sSFR (median sSFR ~ 0.2 Gyr^{-1} for the objects used here) sample and the other comparison samples when investigating these strong-line ratio – metallicity relations.

line ratios involving only oxygen and hydrogen. However, larger samples of [O III]λ4363 detections are needed in order to significantly constrain the evolution out to high redshift.

We do caution, however, that 4 out of the 5 $z > 1$ galaxies lie at or very near the turnover portion of the [O III]λ5007/Hβ and R_{23} relations, where variation in the strong-line ratio is small over the corresponding oxygen abundance range, limiting the constraining power of the relations when determining the metallicity at fixed line-ratio. This is seen as well in the recent work of Sanders et al. (2019), who study the relationships between strong-line ratios and direct metallicity using a sample of 18 galaxies at $1.4 \lesssim z \lesssim 3.6$ with [O III]λ4363 or O III]$\lambda\lambda$1661,1666 auroral-line detections, including 3 new [O III]λ4363 detections from the MOSFIRE Deep Evolution Field survey (MOSDEF; Kriek et al., 2015). They show an abundance of objects with $7.7 < 12+\log(\mathrm{O/H}) < 8.1$ lying at these turnovers and caution against the use of these line ratios at high-z for galaxies within this metallicity regime.

In addition to the strong-line metallicity relations of Jones et al. (2015), we plot the [O III]/[O II] direct metallicity calibration of Bian et al. (2018) (top-left panel of Figure 2.6), who utilized stacked spectra with [O III]λ4363 of $z \sim 0$ high-z analogs that lie at the same location on the N2-BPT diagram as $z \sim 2.3$ star-forming galaxies. This calibration is favored in Sanders et al. (2019) for its linear relation between the strong-line ratio and metallicity, its ability to closely reproduce (~ 0.1 dex) the average metallicity of their $z > 1$ sample, and its derivation from an analog sample selected via strong-line ratios rather than global galaxy properties. Within the range of applicability, $12+\log(\mathrm{O/H}) = 7.8-8.4$, there is generally good agreement between the relation, our various samples (including A1689-

217), and the relation of Jones et al. (2015) as the relation of Bian et al. (2018) lies within the intrinsic scatter around that of Jones et al. (2015).

We note that the majority of the Berg et al. (2012) line ratios do not follow the local relations with direct metallicity. While there is good agreement between the local Jones et al. relations and the few H II regions in the Berg et al. sample with $8.2 \lesssim 12 + \log(\mathrm{O/H}) \lesssim 8.4$, the bulk of the H II region sample, having $12 + \log(\mathrm{O/H}) \lesssim 8.1$, lies removed from these relations. This is seen as well in the strong-line ratio direct metallicity plots of Sanders et al. (2019, Figure 3), who find agreement at $12 + \log(\mathrm{O/H}) \sim 8.3$ between the median relations of individual $z = 0$ H II regions and their $z \sim 0$ and $z > 1$ galaxy samples, but similar divergences below an oxygen abundance of ~ 8.0. As Sanders et al. suggests, this may be due to an incomplete sample of local, high-excitation, low-metallicity H II regions, possibly a result of the short-lived nature of individual star-forming regions and their rapidly changing ionizing spectra.

2.5.2 O_{32} vs. R_{23} Excitation Diagram and its Use as a Metallicity Indicator

The O_{32} vs. R_{23} excitation diagram relates optical emission-line ratios given by the following equations:

$$O_{32} = \frac{[\mathrm{O\ III}]\lambda\lambda 4959, 5007}{[\mathrm{O\ II}]\lambda\lambda 3726, 3729} \tag{2.2}$$

$$R_{23} = \frac{[\mathrm{O\ II}]\lambda\lambda 3726, 3729 + [\mathrm{O\ III}]\lambda\lambda 4959, 5007}{\mathrm{H}\beta} \tag{2.3}$$

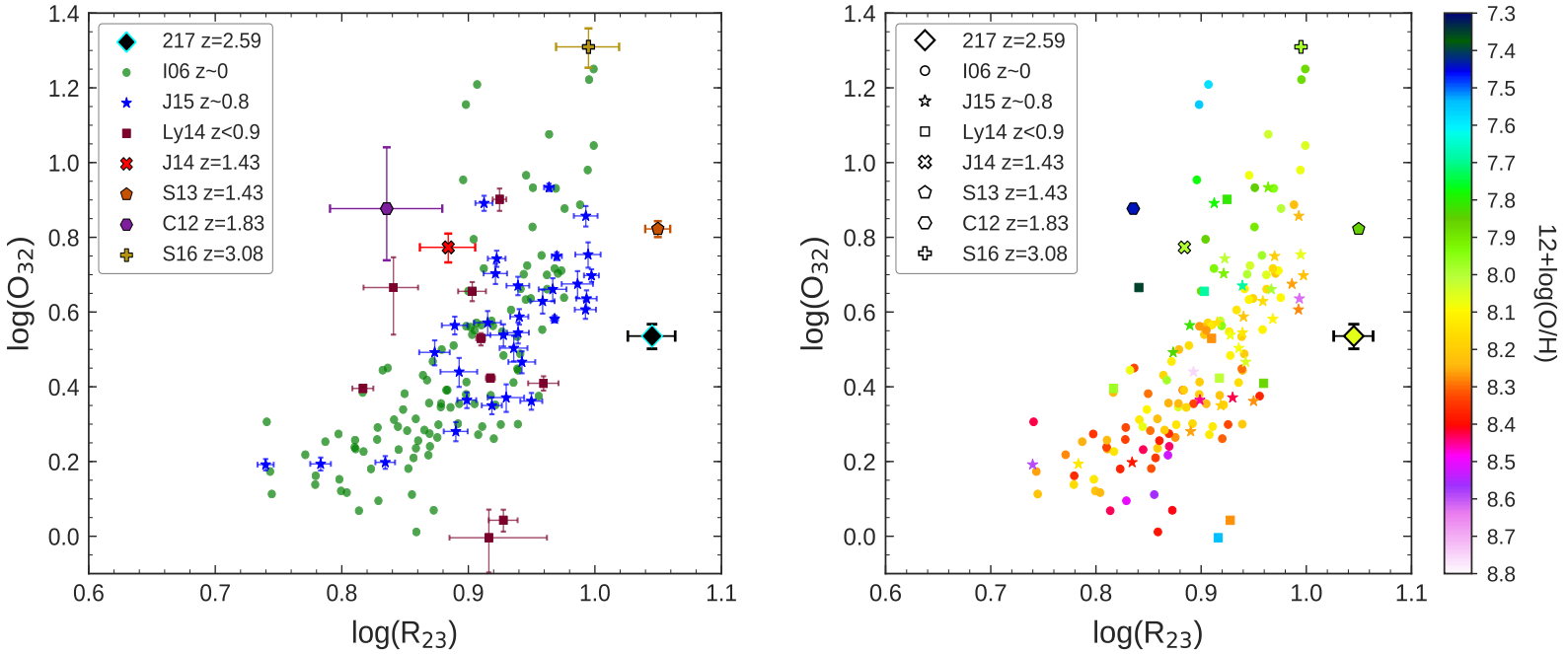

Figure 2.7: High-excitation tail of the O_{32} vs. R_{23} excitation diagram. (Left) A1689-217 and the comparison samples, with error bars, following the same color scheme as in Figure 2.6. A1689-217 is represented by a diamond with a cyan border. (Right) A1689-217 and the comparison samples color-coded by their direct metallicity estimates. A1689-217 is again represented by a diamond, now with a black border. In both panels, the Izotov et al. (2006, I06) sample is represented by circles. The Jones et al. (2015, J15) sample is represented by stars. The Ly et al. (2014, Ly14) sample is denoted by squares. The James et al. (2014, J14) galaxy, Stark et al. (2013, S13) galaxy, Christensen et al. (2012, C12) galaxy, and Sanders et al. (2016a, S16) galaxy are given by an $\times$, pentagon, hexagon, and plus sign, respectively. The color mapping of this plot demonstrates the roughly monotonic and redshift-independent decrease in oxygen abundance from low-to-high O_{32} and R_{23} as first demonstrated empirically by Shapley et al. (2015).

As seen in the high-excitation tail of O_{32} vs. R_{23} displayed in Figure 2.7 for A1689-217 and the comparison samples, as well as in full in the literature (e.g., Nakajima et al., 2013; Nakajima & Ouchi, 2014; Shapley et al., 2015; Sanders et al., 2016b; Strom et al., 2017), the excitation diagram characteristically has a strong correlation between higher O_{32} and R_{23} values. It has also been shown by Nakajima & Ouchi (2014) with a sample of $z = 2 - 3$ Lyman Break Galaxies (LBGs), Shapley et al. (2015); Sanders et al. (2016b) with $z \sim 2.3$ galaxies from the MOSDEF survey, and Strom et al. (2017) with $z \sim 2.3$ galaxies from the KBSS survey that high-redshift, star-forming galaxies follow the same distribution as local SDSS galaxies toward higher O_{32} and R_{23} values. Indeed, when looking at the galaxies in the left panel of Figure 2.7, we see no evidence for significant evolution at any of the redshifts considered by our samples.

Individually, the O_{32} ratio serves as a commonly used diagnostic of the ionization parameter of a star-forming region (see Kewley & Dopita, 2002; Sanders et al., 2016b) while the R_{23} ratio is a commonly used diagnostic for the gas-phase oxygen abundance of a star-forming region (Pagel et al., 1979). However, as detailed in Kewley & Dopita (2002), O_{32} is dependent on metallicity, and R_{23} is dependent on the ionization parameter. Furthermore, as seen in Figure 2.6, the R_{23} diagnostic is double-valued (Kewley & Dopita, 2002) and not very sensitive to the majority of the sub-solar oxygen abundances studied in this work. The variation of ~ 0.3 dex in $\log(R_{23})$ seen here in Figures 2.6 and 2.7 supports the findings of Steidel et al. (2014, see Figure 11), who show, via photoionization models, that $\log(R_{23})$ is nearly independent of input oxygen abundance in high-redshift galaxies with gas-phase metallicities ranging from $0.2-1.0$ $Z_{\odot}$.

If instead these two ratios are considered simultaneously in the O_{32} vs. R_{23} excitation diagram, the double-valued nature of the R_{23} diagnostic is removed, and a combination of ionization parameter and metallicity can be obtained. Kewley & Dopita (2002), Nakajima et al. (2013), Nakajima & Ouchi (2014), and Strom et al. (2018) have all utilized this excitation diagram in combination with photoionization models to calculate oxygen abundances, out to $z \sim 2$ in the latter three studies. Shapley et al. (2015) took an empirical approach to suggesting this excitation diagram's value as an abundance indicator, using the direct metallicity estimates from stacked SDSS spectra of Andrews & Martini (2013) to show a nearly monotonic decrease in metallicity from low-to-high O_{32} and R_{23}. They showed that while R_{23} considered alone does not vary greatly with metallicity, the position within the 2D space defined by these two line ratios correlates strongly with metallicity. They further argued that due to the apparent lack of evolution in high-redshift galaxies along the high-excitation end of the diagram, a redshift-independent (out to $z \sim 2.3$, at least) metallicity calibration deriving from direct abundance estimates could be devised based on the location of a galaxy along the O_{32} vs. R_{23} sequence.

We investigate this claim further with A1689-217 and the comparison samples in the right panel of Figure 2.7. Here we have again plotted A1689-217 and the other samples on the high-excitation tail of the O_{32} vs. R_{23} diagram with each galaxy now color-coded by its direct metallicity estimate. Unlike in the left panel of Figure 2.7, we do not plot the error bars for the galaxies (except for A1689-217) so as to more clearly illustrate any present trends. We see that there is indeed a nearly monotonic decrease in metallicity as one moves from the lower $\log(O_{32}) \sim 0.1$ and $\log(R_{23}) \sim 0.8$ along the sequence to higher values in

both ratios. We also note that with redshift, there does not appear to be any significant evolution of the samples in either O_{32} or R_{23} as well as in metallicity. The $z \sim 0$ sample from Izotov et al. (2006) and the $z \sim 0.8$ sample from Jones et al. (2015) track the excitation sequence very similarly with comparable metallicity values as a function of position along the sequence. The intermediate- and high-redshift galaxies also do not collectively display any systematic offsets in their line-ratio values and do not show any evidence of evolution in their metallicities as a function of location on the sequence. These galaxies follow the same metallicity distribution seen by the lower-redshift samples.

We do take note of the large scatter, particularly in $\log(R_{23})$, of the $z > 1$ galaxy sample. At fixed $\log(O_{32})$, the galaxies of Christensen et al. (2012, C12) and James et al. (2014, J14) lie furthest to the left in $\log(R_{23})$ compared to the lower-redshift samples while the galaxy of Stark et al. (2013, S13) and A1689-217 lie furthest to the right, having significantly higher R_{23} than the comparison samples. This observed scatter may be the consequence of underestimated uncertainties that do not account for systematic errors in the measurement and dust-correction of the emission lines, or it may hint at a larger intrinsic scatter in this line ratio at high redshift when compared to the relatively narrow high-excitation tail defined locally. In either case, our conclusions should not be significantly affected as R_{23}, taken by itself, is not very sensitive to metallicity in the moderately sub-solar regime we are studying. A proper analysis of this scatter will require larger statistical samples with well constrained R_{23} and accurate metallicities that span a broad dynamic range.

The conclusions made from Figure 2.7 support the findings of Shapley et al. (2015) of the O_{32} vs. R_{23} excitation diagram being a useful, redshift-invariant oxygen abundance indicator, based on the direct metallicity abundance scale, out to at least $z \sim 2.3$ and perhaps $z \sim 3.1$ with the inclusion here of COSMOS-1908 (Sanders et al., 2016a). While much larger samples of intermediate- and high-redshift galaxies with direct metallicity estimates are required to confirm or refute the observed lack of evolution in this excitation diagram, its potential as an abundance indicator is important for several reasons (see Jones et al., 2015; Shapley et al., 2015; Sanders et al., 2016b). If this excitation sequence and its relation to metallicity are redshift-independent, then a local relation based on the much richer SDSS sample can be developed and applied accurately at high redshift. This sequence and a corresponding abundance calibration are based on line ratios solely involving strong oxygen and hydrogen emission lines, avoiding biases in nitrogen-based abundance indicators resulting from systematically higher N/O abundance ratios at high redshift (Masters et al., 2014; Shapley et al., 2015; Sanders et al., 2016b). Finally, an indicator using this excitation sequence would be based on the direct metallicity abundance scale, with direct metallicities most closely reflecting the physical conditions present in star-forming regions due to their relation to electron temperature and density.

2.5.3 The Evolution of the Ionization Parameter

The ionization parameter, defined as the ratio of the number density of hydrogen-ionizing photons to the number density of hydrogen atoms in the gas, characterizes the ionization state of the gas in a star-forming region and is often determined via the O_{32} (see Equation 2.2) line ratio. It has been suggested that at high redshift, galaxies have system-

atically higher ionization parameters than are usually found in local galaxies (Brinchmann et al., 2008; Nakajima et al., 2013; Nakajima & Ouchi, 2014; Steidel et al., 2014; Kewley et al., 2015; Cullen et al., 2016; Kashino et al., 2017). These studies have shown this largely based on comparisons at fixed stellar mass (e.g., Kewley et al., 2015; Sanders et al., 2016b), comparison to the average ionization parameter of the entire SDSS (e.g., Nakajima & Ouchi, 2014), and comparisons at fixed metallicity (e.g., Cullen et al., 2016; Kashino et al., 2017).

However, studying the [O III]λ5007/[O II]$\lambda\lambda$3726,3729 and [O III]λ5007/Hβ ratios at fixed metallicity in Figure 2.6, we do not see any systematic offset of the high-redshift galaxies toward higher ionization parameter proxy (the former ratio) or higher excitation (the latter ratio) at fixed O/H. This is in agreement with Sanders et al. (2016a), who studied the same high-z comparison galaxies, as well as Sanders et al. (2019), who enlarged their high-z sample with 3 new [O III]λ4363 detections from the MOSDEF survey and O III]$\lambda\lambda$1661,1666 sources from the literature. In regard to the former ratio, A1689-217 (z = 2.59) and the z = 1.43 galaxy of James et al. (2014) lie very close to the locally-calibrated, best-fit relation, within the 1σ intrinsic scatter around the relation. The z = 3.08 galaxy of Sanders et al. (2016a) lies above the best-fit relation and scatter, but the z = 1.43 galaxy of Stark et al. (2013) and the z = 1.83 galaxy of Christensen et al. (2012) lie below them. When considering the latter ratio, all four high-redshift galaxies lie near the best-fit relation within the intrinsic scatter. These results from Figure 2.6 are corroborated in the O_{32} vs. R_{23} excitation diagram of Figure 2.7. We see no collective systematic offset of these galaxies in O_{32} at fixed R_{23} (a diagnostic for oxygen abundance).

The conclusions drawn from Figures 2.6 and 2.7 contrast with studies such as Cullen et al. (2016) and Kashino et al. (2017), who argue for increased ionization parameter at fixed O/H in high-redshift galaxies. Instead, our results support the suggestions of Sanders et al. (2016a,b, 2019), who argue for an absence of evolution in the ionization parameter at fixed metallicity. Sanders et al. (2016b) used ~ 100 star-forming galaxies at $z \sim 2.3$ from the MOSDEF survey to suggest that while high-redshift galaxies do in fact have systematically higher O_{32} values at fixed stellar mass relative to local galaxies, they have similar O_{32} values at fixed R_{23}. They argue that, with the high-redshift MOSDEF sample following the same distribution as local galaxies along the higher O_{32} and R_{23} end of the excitation sequence, and this end corresponding to lower metallicities (Shapley et al., 2015), the ionization state of high-redshift, star-forming galaxies must be similar to metal-poor local galaxies. This is corroborated by Sanders et al. (2019), who show that, on average, their $z > 1$ auroral-line-emitting sample lies on local relations between ionization parameter and direct-method oxygen abundance, positioned in the same location as metal-poor, $z \sim 0$ SDSS stacks and local H II regions. Sanders et al. (2016b) further argue that the difference in offset when comparing to constant stellar mass as opposed to constant metallicity is due to the evolution of the mass-metallicity relation, where high-redshift galaxies have systematically lower metallicities than local galaxies at fixed stellar mass (Sanders et al., 2015).

It is important to note that the results of this paper support the notion of a lack of evolution in ionization parameter at fixed metallicity without the use of nitrogen in the metallicity estimates. As stated earlier, using direct metallicities and diagnostics (R_{23})

not involving nitrogen avoids possible systematic offsets in the abundance estimates due to higher N/O abundance ratios at high redshift.

2.5.4 Low-Mass End of the Fundamental Metallicity Relation

The Fundamental Metallicity Relation (Mannucci et al., 2010) is a 3D surface defined by a tight dependence of gas-phase metallicity on stellar mass and SFR and is suggested to exist from $z = 0$ out to $z = 2.5$ without evolution (e.g., Mannucci et al., 2010; Henry et al., 2013b; Maiolino & Mannucci, 2019). From this surface, Mannucci et al. (2010) define a projection, μ_α vs. 12+log(O/H), where μ_α is a linear combination of stellar mass and SFR relying on the observed correlation and anti-correlation of metallicity with stellar mass and SFR, respectively.

$$\mu_\alpha \;=\; \log(M_*) - \alpha \log(\mathrm{SFR}) \tag{2.4}$$

Mannucci et al. (2010) suggest that if $\alpha = 0.32$ in this relation, the scatter in metallicity at fixed μ_α is minimized, all galaxies out to $z = 2.5$ show the same dependence of metallicity on $\mu_{0.32}$, and all galaxies out to this redshift occupy the same range of $\mu_{0.32}$ values.

Unfortunately, the FMR of Mannucci et al. (2010) is defined by low-redshift SDSS galaxies with stellar masses down to $\log(M_*/M_\odot) = 9.2$, $\sim$1.1 (0.6) dex above the lower- (upper-) limit stellar mass of A1689-217 (see Section 2.4.1 and Figure 2.5). In SFR, this FMR only probes galaxies with $-1.45 \leqslant \log(\mathrm{SFR}) \leqslant 0.8$, whereas A1689-217 has a log(SFR) = 1.2. Furthermore, the redshift-invariant nature of the FMR and $\mu_{0.32}$ − metallicity projection only applies out to $z = 2.5$, with A1689-217 lying just beyond this redshift at

$z = 2.59$. Perhaps most importantly, the Mannucci et al. (2010) FMR is defined with metallicities calculated via locally-calibrated, strong-line diagnostics (Maiolino et al., 2008), the applicability of such indirect methods at high redshift being a primary focus of this paper.

Addressing the limited stellar mass range, Mannucci et al. (2011) extended the FMR, or more specifically the $\mu_{0.32}$ − metallicity projection, down to a stellar mass of $\sim 10^{8.3}$ $M_\odot$ using $\sim$ 1300 galaxies from the Mannucci et al. (2010) sample with $8.3 < \mu_{0.32} < 9.4$. They found that these low-mass galaxies extend the FMR with a smooth, linear relation between gas-phase metallicity and $\mu_{0.32}$ given, for $\mu_{0.32} < 9.5$, by:

$$12 + \log(\mathrm{O/H}) \;=\; 8.93 + 0.51(\mu_{0.32} - 10) \tag{2.5}$$

Recognizing that metallicity estimates based on different methods can differ drastically for the same galaxies (Kewley & Ellison, 2008), Andrews & Martini (2013) investigated the μ_α (Equation 2.4) FMR projection using the T_e-based metallicities they calculated with their stacked SDSS spectra. Using galaxies with $7.5 \lesssim \log(M_*/M_\odot) \lesssim 10.6$ and $-1.0 \leqslant \log(\mathrm{SFR}) \leqslant 2.0$ binned by M_* and SFR, they found that $\alpha = 0.66$ minimized the scatter in their metallicities at fixed μ_α. While this calibration of the μ_α − metallicity projection utilizes direct-method oxygen abundances, it still suffers from both a lack of high-redshift data due to the faintness of T_e-sensitive auroral lines and a poor sampling of low-mass, high-SFR galaxies like A1689-217 (see Figure 1 of Andrews & Martini 2013 for the distribution in M_* and SFR of their sample).

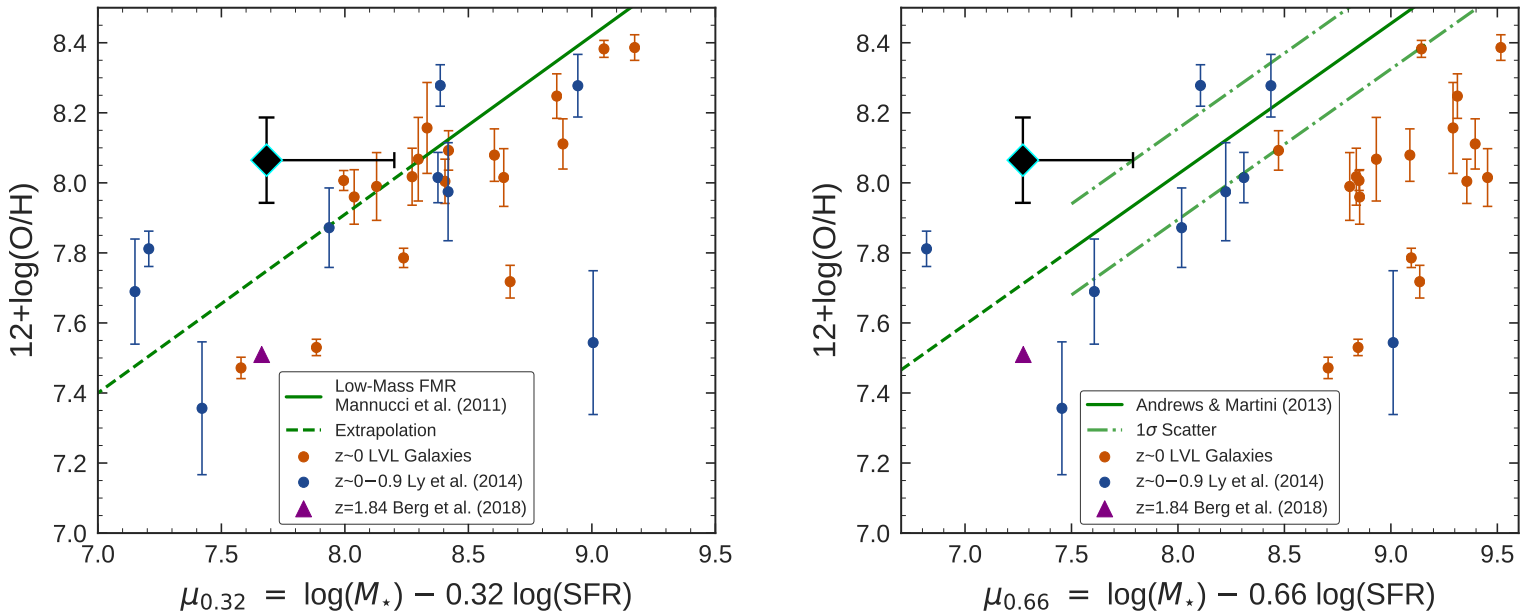

Figure 2.8: (Left) The low-mass extension of the Fundamental Metallicity Relation (FMR; Mannucci et al., 2010) as given by the projection of metallicity vs. μ_α (in solar units; see Equation 2.4) for $\alpha = 0.32$. This extension (Equation 2.5) was calculated by Mannucci et al. (2011) down to $\mu_{0.32} \sim 8.3$ (solid line), so an extrapolation in $\mu_{0.32}$ is shown here for lower values (dashed line). (Right) The FMR, metallicity vs. μ_α projection as calculated by Andrews & Martini (2013). These authors report a minimization in the scatter of metallicity at fixed μ_α for $\alpha = 0.66$. This linear relation, with slope $m = 0.43$, is only calibrated down to $\mu_{0.66} \sim 7.5$ (solid line), so an extrapolation in $\mu_{0.66}$ is given here (dashed line). The scatter in the projection (based on stacks instead of individual galaxies) is given to be $\sigma = 0.13$ dex and is shown by the dot-dashed lines. Metallicities used in Mannucci et al. (2010, 2011) are based on strong-line methods whereas metallicities in Andrews & Martini (2013) are T_e-based from stacks of SDSS spectra. In both panels, A1689-217 is given by the black diamond with cyan border. Its stellar mass of $\log(M_*/M_\odot) = 8.07$ is likely a lower limit not accounting for an undetected older stellar population (see Section 2.4.1 and Figure 2.5), so we show the increase A1689-217 would experience in μ_α for a factor of ~ 3.3 increase in stellar mass. A similar galaxy (in M_* and SFR) to A1689-217 from Brammer et al. (2012b) and Berg et al. (2018) has its lower-limit metallicity (see text for details) plotted as the purple triangle. A sample of $z < 0.9$ galaxies from Ly et al. (2014) is shown in blue, and a low-mass, $z \sim 0$, Local Volume Legacy (LVL) sample (see text for details) is shown by the dark orange data points. All galaxy samples have their direct metallicities plotted in both panels.

We test the validity of the FMRs of Mannucci et al. (2011) and Andrews & Martini (2013) in the poorly-sampled $M_* - \mathrm{SFR}$ parameter space occupied by A1689-217. In Figure 2.8, we plot A1689-217 against the low-mass FMR extension (left) given by Equation 2.5, extrapolated down by ~ 0.6 dex in $\mu_{0.32}$, and against the T_e-based FMR (right), extrapolated down by ~ 0.2 dex in $\mu_{0.66}$. We also plot the $z = 1.84$ highly-ionized, lensed galaxy (SL2SJ02176-0513) of Brammer et al. (2012b) and Berg et al. (2018), which, when adjusted for a Chabrier (2003) IMF with 0.2 $Z_\odot$, has a very similar stellar mass ($\log(M_*/M_\odot) = 8.03$) and SFR (14 $M_\odot$ yr^{-1}) as A1689-217. Despite these similar properties, SL2SJ02176-0513 has a much lower metallicity ($12+\log(\mathrm{O/H}) \geqslant 7.51$) than A1689-217, however. We note that its metallicity is reported as a lower limit due to both the lack of spectroscopic coverage of the [O II]$\lambda\lambda$3726,3729 emission lines needed for the determination of $\mathrm{O}^+/\mathrm{H}^+$ (see Equation 2.1) and the possibility of a contribution from O^{+3} to O/H. Nevertheless, as detailed in Berg et al. (2018), this lower limit should be close to the actual value as the highly-ionized nature of the galaxy makes the O^+ contribution to the oxygen abundance very small (estimated at 2% of the total oxygen abundance; included in our stated lower-limit metallicity), and the ionization correction factor (ICF) for contribution of O^{+3} is also estimated to be small (ICF = 1.055; not included in our stated lower-limit metallicity).

For further comparison of A1689-217 and the FMRs to other low-mass galaxies spanning a broad range of star formation activity, we also include in Figure 2.8 the partial Ly et al. (2014) sample used in this work (median $\log(M_*/M_\odot) \sim 8.4$ and median specific star formation rate (sSFR) ~ 9.3 Gyr^{-1}) and a $z \sim 0$ LVL subsample (median $\log(M_*/M_\odot)$ ~ 7.7 and median sSFR ~ 0.2 Gyr^{-1}). The Ly et al. (2014) sample, in addition to using

the metallicities re-derived in this work, uses SFRs re-calculated assuming a Cardelli et al. (1989) extinction law. Stellar masses for this sample are the values given in Ly et al. (2014) for a Chabrier (2003) IMF with 0.2 $Z_{\odot}$. The LVL objects used here comprise the subset of the Berg et al. (2012) sample used in Figures 2.3 and 2.6 of which the objects are a part of both the sample used in Berg et al. (2012) and the sample in Weisz et al. (2012). Metallicities used here are those re-calculated in this paper with the emission-line fluxes from Berg et al. (2012). Stellar masses for these galaxies are taken from Weisz et al. (2012) while the SFRs are calculated from Hα measurements taken by Kennicutt et al. (2008) and Lee et al. (2009) as part of the 11HUGS survey. All SFRs for A1689-217 and the comparison samples are calculated via Balmer recombination lines, assuming a Chabrier (2003) IMF with 0.2 $Z_{\odot}$, and all metallicities are calculated via the "direct" method.

With the lower-limit stellar mass estimated by our SED fitting ($\log(M_*/M_{\odot}) = 8.07$), A1689-217 lies $\sim 2.6\sigma$ (2.9σ) above the extrapolation of the low-mass FMR extension of Mannucci et al. (2011) (T_e-based FMR of Andrews & Martini 2013). However, as mentioned in Section 2.4.1 and seen in Figure 2.5, an unseen, older stellar population component can exist in A1689-217 without significantly altering the observed SED, raising the stellar mass estimate of A1689-217 by as much as a factor of 3.3 (up to $\log(M_*/M_{\odot}) = 8.59$). An increase in stellar mass will correspondingly increase the measured value of μ_{α} (Equation 2.4) and bring A1689-217 into better agreement with both FMRs. This is seen in Figure 2.8, where the horizontal bar extending from A1689-217 represents the range of μ_{α} values corresponding to our estimated range of stellar masses for A1689-217. If the mass estimate is even $\sim 2\times$ what we state as the lower bound, A1689-217 is consistent with the FMR of

Andrews & Martini (2013) within the 1σ scatter around the relation and the uncertainty in A1689-217's oxygen abundance. Without this mass increase, A1689-217 is very likely already consistent with the extrapolation of the low-mass end of the FMR as given by Mannucci et al. (2011) considering the 1σ dispersions in metallicity seen at fixed $\mu_{0.32}$ in their work (see right panel of their Figure 1). We therefore suggest that A1689-217 is consistent with both FMRs within the observed scatter around each relation.

An important takeaway from Figure 2.8 is the large scatter seen around both μ_α − metallicity projections. This is well illustrated when comparing A1689-217 and SL2SJ02176-0513 from Berg et al. (2018). Despite having similar sSFRs ~ 135 Gyr^{-1}, these galaxies differ in oxygen abundance by ~ 0.55 dex, lying on either side of both FMRs. Large scatter is also seen in the Ly et al. (2014) comparison sample, despite the sample being generally consistent with both FMRs. This scatter observed in Figure 2.8 around the FMRs is likely due to the increased variation in star formation histories and current star formation activity in dwarf galaxies (Mannucci et al., 2011; Emami et al., 2018) and suggests that physical processes of gas flows, enrichment, and star formation have not yet reached equilibrium (Ly et al., 2015). Physical timescale effects in dwarf galaxies with bursty star formation may lead to large dispersions in the metallicities of galaxies with similar properties, like we see with A1689-217 and SL2SJ02176-0513, whereby we may be observing more metal-rich galaxies at a time when recent star formation has enriched the gas, but not yet removed metals from the galaxy via supernovae and other stellar feedback (Ly et al., 2015).

In consideration of the LVL sample here, we note the systematic offsets of the galaxies (median $\log(\mathrm{SFR}) \sim -1.9$ and median $\log(M_*/M_\odot) \sim 7.7$) particularly from the

relation of Andrews & Martini (2013), but also slightly below the relation of Mannucci et al. (2011) on average. While an in-depth study of these offsets is beyond the scope of this work, they may arise from a lack of examination of the $M_* -$ SFR parameter space occupied by the LVL galaxies. Mannucci et al. (2011) only probe down to $M_* \sim 10^{8.3}$ $M_\odot$ and log(SFR) ~ -1.45 while Andrews & Martini (2013) study a sample with the vast majority of objects having log(SFR) > -1 and $\log(M_*/M_\odot) > 8$. The extreme offset of the LVL galaxies from the Andrews & Martini (2013) relation may also result from the stronger dependence of μ_α on SFR in this calibration ($\alpha = 0.66$) compared to that in Mannucci et al. (2010) ($\alpha = 0.32$).

2.5.5 A Comparison Against the MZR Predictions of FIRE

The FIRE[7] (Feedback In Realistic Environments) simulations (Hopkins et al., 2014) are cosmological zoom-in simulations that contain realistic physical models and resolution of the multi-phase structure of the ISM, star formation, and stellar feedback. Ma et al. (2016) utilize these simulations to study the evolution of the stellar mass $-$ gas-phase metallicity relation from $z = 0 - 6$ for galaxies spanning the stellar mass range $M_* = 10^4 - 10^{11}$ $M_\odot$ at $z = 0$. They predict an MZR that has a slope which does not vary appreciably with redshift. They fix the slope to the mean value with redshift, $m = 0.35$ (which almost perfectly agrees with the best-fit slope between $z = 1.4$ and $z = 3.0$ $-$ see their Figure 3), and report an MZR that evolves with z as:

[7] https://fire.northwestern.edu/

$$12 + \log(\mathrm{O/H}) \;=\; 0.35[\log(\mathrm{M_*/M_\odot}) - 10] + 0.93\exp(-0.43\mathrm{z}) + 7.95 \qquad (2.6)$$

Comparing A1689-217 against this prediction, at $z = 2.5918$, with A1689-217's lower- (upper-) limit stellar mass of $\log(M_*/M_\odot) = 8.07$ (8.59; see Section 2.4.1 and Figure 2.5), we find that the metallicity of A1689-217 ($12 + \log(\mathrm{O/H}) = 8.06 \pm 0.12$) is $\sim 4.0\sigma$ (2.5σ) above the predicted oxygen abundance of $12 + \log(\mathrm{O/H}) = 7.58$ (7.76). Comparing the prediction in Equation 2.6 also against the galaxy, SL2SJ02176-0513, of Berg et al. (2018) at $z = 1.8444$ and $\log(M_*/M_\odot) = 8.03$, we find that the lower-limit metallicity of the galaxy (7.51; see Berg et al. 2018 and Section 2.5.4 for details on the lower limit) lies 0.17 dex below the prediction of $12 + \log(\mathrm{O/H}) = 7.68$. Further comparing the position of both of these galaxies to the scatter around the MZR in Figure 3 of Ma et al. (2016), we see that A1689-217 lies above all simulated galaxies at its lower-limit stellar mass, but likely among the objects scattered high in oxygen abundance at its upper-limit stellar mass. SL2SJ02176-0513 lies below the best-fit relation, but is consistent within the scatter.

Considered together, despite being at different redshifts, these results at least show that there is significant scatter of dwarf galaxies around the MZR at roughly fixed stellar mass. This is likely due to time variations in the metallicities of dwarf galaxies resulting from the bursty nature of their star formation and its connection to gas inflows/outflows (Ma et al., 2016). Due to the extremely metal-poor nature of SL2SJ02176-0513 (~ 0.07 $\mathrm{Z_\odot}$) and its general agreement with the predicted MZR, as well as the discrepancy of A1689-217 from the MZR, particularly when considering the lower-end of A1689-217's mass range, these

results may also suggest that the slope ($m = 0.35$) in Equation 2.6 is too steep. However, larger observational samples are needed to verify this suggestion.

2.6 Summary

In this paper, we present a 4.2σ detection of the temperature-sensitive, auroral [O III]λ4363 emission line in a lensed, star-forming, dwarf galaxy at $z = 2.59$, A1689-217. With the extinction-corrected fluxes of the rest-optical, nebular emission lines, we estimate the electron temperature and density of this galaxy and calculate, directly, an oxygen abundance of 12+log(O/H) = 8.06 $\pm$ 0.12 (0.24 $Z_{\odot}$). With this measurement, and intrinsic strong-line ratios calculated for A1689-217, we report the following:

1. We study the evolution with redshift of strong-line ratio $-$ direct metallicity relations calibrated and suggested to be redshift-invariant out to $z \sim 0.8$ by Jones et al. (2015). With a $z \sim 0$ comparison sample from Izotov et al. (2006), the 32 $z \sim 0.8$ galaxies from Jones et al. (2015), 9 $z < 0.9$ galaxies from Ly et al. (2014), and 4 high-redshift galaxies (z = 1.43, 1.43, 1.83, 3.08) with [O III]λ4363 detections in addition to A1689-217, we find no evidence for evolution of the Jones et al. strong-line ratio $-$ metallicity calibrations. We also study the [O III]/[O II] metallicity calibration of Bian et al. (2018), the preferred metallicity diagnostic in the strong-line metallicity study of Sanders et al. (2019). We find general agreement between this relation and our samples as well as with the relation of Jones et al. (2015). We note divergences from the Jones et al. relations of our $z \sim 0$ LVL H II region sample below $12 + \log(\mathrm{O/H}) \sim 8.1$, similar to H II region divergences seen in Sanders et al. (2019).

2. Using the same comparison samples, we find no significant evolution with redshift in the high-excitation tail of the O_{32} vs. R_{23} excitation diagram. The different galaxy samples do not display any relative offsets in either O_{32} or R_{23}, with intermediate- and high-redshift galaxies following the same distribution as local galaxies, albeit with larger scatter of the $z > 1$ sample in $\log(R_{23})$. We also observe the nearly monotonic decrease in direct metallicity with increasing O_{32} and R_{23} seen in Shapley et al. (2015). As with the strong-line ratios, we find no evidence for evolution with redshift of the metallicity as a function of position along the excitation sequence. The combination of these results supports the conclusions of Shapley et al. (2015) that the O_{32} vs. R_{23} excitation diagram can be a useful, direct-metallicity-based, redshift-invariant, empirical oxygen abundance indicator.

3. Through our study of both the strong-line ratio − metallicity relations and the O_{32} vs. R_{23} excitation diagram, we find no evolution with redshift of the ionization parameter at fixed O/H. This result is in agreement with Sanders et al. (2016a,b, 2019), who report the same finding and suggest that the ionization state of high-z, star-forming galaxies is similar to local, metal-poor galaxies.

4. We plot A1689-217 against both the $\mu_{0.32}$ − metallicity projection of the Fundamental Metallicity Relation (FMR) as extended to low stellar mass by Mannucci et al. (2011) and the $\mu_{0.66}$ − metallicity projection of Andrews & Martini (2013), wherein the metallicities are T_e-based as opposed to the strong-line basis of Mannucci et al. (2011). The stated stellar mass range ($\log(M_*/M_\odot) = 8.07 - 8.59$) and SFR (16.2 $M_\odot$ yr^{-1}) of A1689-217 yield a range in $\mu_{0.32}$ ($\mu_{0.66}$) of $\sim 7.7 - 8.2$ ($\sim 7.3 - 7.8$) and thus require

slight extrapolations of both FMRs in μ_α (~ 0.6 dex in $\mu_{0.32}$ and ~ 0.2 dex in $\mu_{0.66}$). We also compare A1689-217 and the FMRs to other low-mass galaxy samples at low-to-high redshift with a large range in current star formation activity. Together, these samples show a large scatter around the FMR, likely due to large variations in star formation history and current star formation activity in dwarf galaxies. With this observed scatter, and the uncertain mass estimate of A1689-217 resulting from the possibility of the presence of an unseen, older stellar population within the galaxy, we conclude that A1689-217 is consistent with both FMRs studied.

5. We compare the locations in $M_* - Z$ parameter space of A1689-217 and the galaxy from Berg et al. (2018) to the predicted MZR from the FIRE hydrodynamical simulations (Ma et al., 2016). A1689-217 lies $\sim 0.3 - 0.5$ dex above the predicted relation while the object from Berg et al. (2018) lies ~ 0.2 dex below the relation, suggesting a large scatter in the relation at low-mass and/or a slightly shallower MZR slope than predicted.

This study adds another crucial data point at high redshift in terms of direct oxygen abundance estimates and dwarf galaxy properties. With the measurements of A1689-217 and their comparisons to measurements of other auroral-line-emitting galaxies at various redshifts, we are able to further constrain the validity of several diagnostics at high redshift and low stellar mass, such as locally-calibrated strong-line ratio − direct metallicity relations and the FMR. However, large statistical samples of high-redshift [O III]λ4363 sources and very low mass dwarf galaxies are needed to properly constrain these diagnostics. Regardless,

this and other similar studies help to prepare us for those large surveys that will be conducted with the next generation of ground and space-based telescopes.

2.7 Yuan 2009 Detection

This paper includes a re-analysis of previously reported high-redshift ($z > 1$) detections of [O III]λ4363. Yuan & Kewley (2009) reported a $\sim 3\sigma$ detection of [O III]λ4363 in a $z = 1.7$ galaxy behind Abell 1689, referred to as "Lens22.3" in their paper and first reported as a multiply-imaged galaxy in Broadhurst et al. (2005). As part of our larger campaign to obtain near-IR spectra of lensed, high-redshift galaxies, we obtained a MOSFIRE J-band spectrum of Lens22.3 as well as of another image of the same galaxy (referred to as Lens22.1 in Broadhurst et al., 2005). Both images were observed in the same slit mask for 1,440 seconds on 2015 January 20 and 4,320 seconds on 2016 February 1 in $\sim 0\farcs6$ seeing on both nights. Though our exposure times are somewhat shorter than the Yuan & Kewley (2009) observations (5,760 seconds vs. 6,800 seconds), the much higher spectral resolution ($R \sim 3300$ vs. $R \sim 500$) and narrower slit width ($0\farcs7$ vs. $1\farcs0$) of the MOSFIRE observations result in a superior sensitivity to narrow emission lines. For a specific comparison in the J-band, our detections of Hβ are 35σ and 28σ for Lens22.3 and Lens22.1, respectively, compared to 10σ for the Yuan & Kewley (2009) detection. For additional sensitivity to faint lines, we normalized the two spectra (by the [O III]λ4959 flux) and created a weighted-average spectrum, resulting in an Hβ detection of 48σ.

The 2D spectra of Lens22.3 and Lens22.1 and the stacked 1D spectrum can be seen in Figure 2.9. Strong [O III]λ4959, Hβ, and a 23σ detection of $H\gamma$ can be seen. However,

there is no evidence of an [O III]λ4363 line. Given the reported $H\beta$/[O III]λ4363 $\sim$ 3.7, we should have detected the line at $\sim 9.2\sigma$. Given the much lower spectral resolution of the Subaru/MOIRCS spectrum of Yuan & Kewley (2009), we believe that the line detected in the MOIRCS spectrum was likely the Hγ line. That would also help explain why the line center reported in that spectrum was at a somewhat lower redshift than the other lines ($z = 1.696$ vs. $z = 1.705$).

A more detailed analysis of this spectrum and the rest of our sample will be reported in future works.

Table 2.3: Emission-Line Fluxes of Lens22.3 and Lens22.1

Line	Relative Flux[a]	S/N
Hγ	0.49	23
[O III]λ4363	< 0.03	...
Hβ	1.0	48
[O III]λ4959	2.00	67

[a]Fluxes relative to Hβ flux

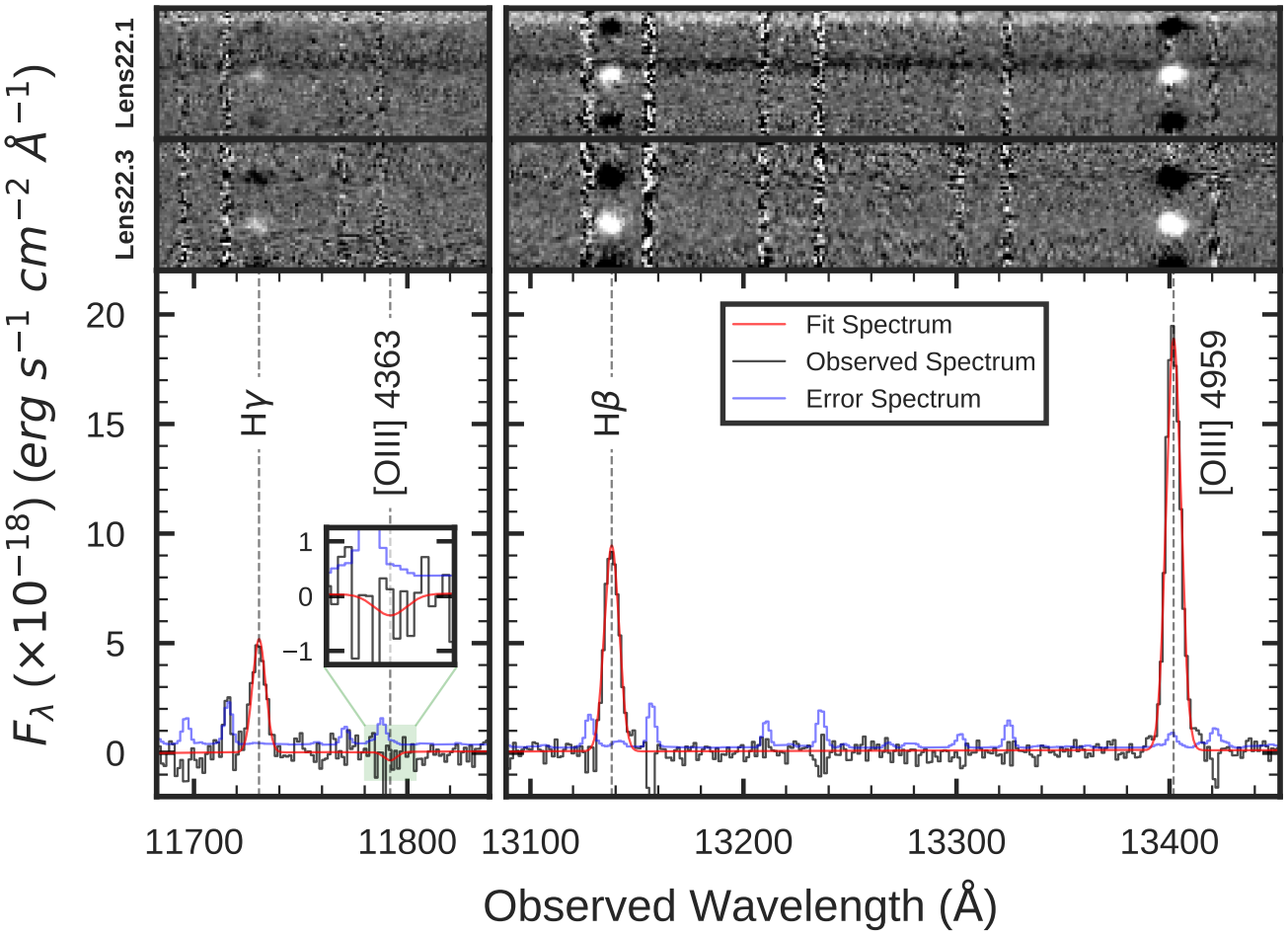

Figure 2.9: The 2D spectra of Lens22.1 (top) and Lens22.3 (bottom) (images referred to in Broadhurst et al. 2005), two images of the same galaxy at $z = 1.7026$. Below is plotted the combined 1D spectrum of both images (black), the error spectrum (blue), and the best-fit continuum and emission lines (red). Strong emission lines are seen in Hγ, Hβ, and [O III]λ4959, but no detection is seen in [O III]λ4363 (either in 1D or 2D), in disagreement with the claimed detection in Yuan & Kewley (2009). The portion of the 1D spectrum containing [O III]λ4363 has been highlighted in green and magnified in the inset plot.

Chapter 3

The Direct-Method Oxygen Abundance of Typical Dwarf Galaxies at Cosmic High-Noon

3.1 Introduction

The gas-phase metallicity, or gas-phase oxygen abundance ($12 + \log(\mathrm{O/H})$) of the interstellar medium (ISM) of galaxies, is a cornerstone in the study of galaxy formation and evolution. The metallicity traces the stellar mass buildup of galaxies through the enrichment over time of the ISM by heavy elements produced via stellar nucleosynthesis. Galaxies, however, are not closed boxes and have inflows of metal-poor gas – the fuel for star formation – from the circumgalactic medium (CGM) and intergalactic medium (IGM) as well as outflows of metal-laden gas triggered by feedback from supernovae and/or active galactic nuclei

(AGN). This modulation of the enrichment of the ISM, via gas flows, shows metallicity to also be an important physical property in the study of the cycle of baryons into, out of, and within (baryon recycling) galaxies.

The combination of these processes is reflected in the scaling relation of the gas-phase metallicity (Z) with the stellar mass (M_*) of star-forming galaxies, known more succinctly as the mass−metallicity relation or MZR. This relation demonstrates that stellar mass and metallicity are positively and tightly correlated, whereby in the local Universe, below a characteristic mass of $M_* \approx 10^{10.0-10.5}$ $M_\odot$, the MZR is described by a power-law. Above this characteristic mass, or "turnover" mass, the MZR flattens and asymptotically approaches an upper-limit oxygen abundance. Locally, the MZR has been shown to exist over five decades in M_* from $10^6 \lesssim M_*/M_\odot \lesssim 10^{11}$ (e.g., Tremonti et al., 2004; Lee et al., 2006; Kewley & Ellison, 2008; Andrews & Martini, 2013; Maiolino & Mannucci, 2019; Curti et al., 2020; Sanders et al., 2021). The MZR has also been shown in numerous studies to exist at high-z out to $z > 3$, though with an evolution such that galaxies at fixed M_* have lower metallicities at higher redshifts (e.g., Erb et al., 2006; Maiolino et al., 2008; Henry et al., 2013a,b; Zahid et al., 2013, 2014a,b; Steidel et al., 2014; Maiolino & Mannucci, 2019; Sanders et al., 2015, 2020, 2021; Strom et al., 2022). Additionally, at higher redshifts, the turnover mass is found to be larger than seen locally (Zahid et al., 2013, 2014a,b), and at $z > 2$, it is unknown whether the turnover mass exists at all. At $z > 2$ and $M_* \gtrsim 10^9$ $M_\odot$, the MZR has been described by a single power-law (e.g., Sanders et al., 2021).

In constraining the shape, scatter, and evolution in the MZR, insight is gained into the physics of how star formation processes and baryon flows are connected and how

galaxy growth is structured and regulated (e.g., Finlator & Davé, 2008; Davé et al., 2012; Ma et al., 2016; Torrey et al., 2014, 2019). For instance, the slope of the low-mass end of the MZR can relate galactic metal retention to how efficient outflows (which remove gas and metals from the ISM/galaxy) and stellar feedback are in regulating star formation and stellar mass growth (e.g., Torrey et al., 2014). Correlated scatter in the MZR can inform of secondary dependencies of the metallicity to properties such as gas-mass and SFR, giving further insight into current conditions of a galaxy as well as elucidating more fundamental relationships between mass, metallicity, and other properties (e.g., Ma et al., 2016; Torrey et al., 2019).

Through the empirical study of the scatter in the MZR, Mannucci et al. (2010) and Lara-López et al. (2010) found that metallicities of galaxies in the Sloan Digital Sky Survey (SDSS) do in fact have a secondary dependence on SFR, a dependence reaffirmed in more recent work, albeit to varying degrees of the strength of that dependence (e.g., Yates et al., 2012; Andrews & Martini, 2013; Maiolino & Mannucci, 2019; Curti et al., 2020; Sanders et al., 2021). This $M_* - \mathrm{SFR} - \mathrm{O/H}$ relation is referred to as the fundamental metallicity relation (FMR) and displays a reduced scatter in metallicity of ~ 0.05 dex (compared to ~ 0.1 dex in the MZR; Tremonti et al., 2004). In effect, the FMR is a 3D surface that posits that metallicity is anti-correlated with SFR such that, at fixed M_*, galaxies with above-average (below-average) SFRs will have below-average (above-average) O/H. In addition to a reduction in the intrinsic scatter of metallicity, Mannucci et al. (2010) also suggested that the FMR is redshift-invariant out to $z \sim 2.5$. If true, this naturally explains the evolution in the normalization of the MZR to be observations at various redshifts of different regions

of the locally-defined FMR; this is physically motivated by the anti-correlation of O/H with SFR and the observed increase of SFR with redshift at fixed M_* (e.g., Speagle et al., 2014; Whitaker et al., 2014; Sanders et al., 2021). As galaxy samples at high-z have increased in size, evidence has grown that the FMR is indeed redshift-invariant, at least to within ~ 0.1 dex, out to $z \sim 2.5$ and even possibly $z \sim 3.3$ (e.g., Henry et al., 2013a,b; Cresci et al., 2019; Maiolino & Mannucci, 2019; Sanders et al., 2018, 2020, 2021). However, this evidence is largely based on metallicities indirectly-calculated via prescriptions calibrated in the local Universe, and it is unknown how accurate these methods are at high-z.

In order to properly assess the evolution of the MZR and FMR, metallicities must be estimated accurately at low- and high-z. This requires an accurate understanding of the nebular physical conditions of star-forming galaxies at different redshifts. Fortunately, a procedure that addresses both of these requirements exists and is applicable at various redshifts, this procedure being the "direct" method of oxygen abundance determination. This method relies on first estimating the electron temperature (T_e) and electron density (n_e) of the ionized nebular gas as these properties are responsible for the strength of the collisionally-excited oxygen emission lines needed for this procedure ([O II] $\lambda\lambda$3726, 3729 and [O III] $\lambda\lambda$4959, 5007). These properties are then considered together with flux ratios of the collisionally-excited lines to hydrogen Balmer recombination lines in order to estimate the total oxygen abundance (e.g., Izotov et al., 2006; Osterbrock & Ferland, 2006). Unfortunately, this direct method relies on weak auroral emission lines to calculate T_e, which is determined from the flux ratio of strong emission lines (e.g., [O III] $\lambda\lambda$4959, 5007) to auroral emission lines (e.g., O III] $\lambda\lambda$1661, 1666 or [O III] λ4363) of the same ionic species. While

use of [O III] λ4363 is common in this methodology as it lies in the rest-optical with [O III] $\lambda\lambda$4959, 5007, this line is $\sim 30-100\times$ fainter than [O III] λ5007 (e.g., Jones et al., 2015, Figure 1), typically decreasing in strength with increasing galactic metallicity. As such, large, representative samples of [O III] λ4363-emitters (and thus direct metallicities) have been difficult to acquire with current facilities and instrumentation, especially at high O/H (and M_* by the MZR) and at $z > 1$ where only a handful of [O III] λ4363 detections exist, mostly thanks to gravitational lensing (Brammer et al., 2012b; Christensen et al., 2012; Stark et al., 2013; James et al., 2014; Patrício et al., 2018; Gburek et al., 2019; Sanders et al., 2016a, 2020). Moreover, due to the faintness of the auroral lines, both in the UV and optical, the currently-detected auroral-line-emitters at $z > 1$ are clearly biased and are more representative of extreme emission line galaxies (EELGs) than of "typical" star-forming galaxies seen at these redshifts (Sanders et al., 2020). These high-z auroral-line-emitters tend to fall well above the mean $M_* - SFR$ relation defined by typical galaxies at a given redshift and have flux ratios indicative of higher ionization parameters and lower metallicities than average.

To overcome the current limitations of the direct-metallicity method, and therein study more representative samples of galaxies across a wider dynamic range of metallicities and redshifts, indirect "strong-line" methods of determining oxygen abundance were developed (Jensen et al., 1976; Alloin et al., 1979; Pagel et al., 1979). These methods allow for metallicity estimation when [O III] λ4363 cannot be detected. Instead, strong-line methods rely on locally-calibrated empirically (e.g., Pettini & Pagel, 2004; Jones et al., 2015; Bian et al., 2018; Curti et al., 2020) or theoretically-determined (e.g., McGaugh, 1991; Kewley & Dopita, 2002; Dopita et al., 2013) relations between metallicity and flux ratios of strong,

rest-optical, nebular emission lines. However, while these strong-line methods have proven very useful in understanding the enrichment of local galaxies, they have several drawbacks of their own. For example, depending on the strong-line index and calibration used, metallicity estimates can vary by up to 0.7 dex (Kewley & Ellison, 2008). In part, this is due to how the strong-line methods are calibrated. Calibrations based on photoionization models tend to produce higher metallicity estimates than empirical, T_e-based calibrations (Curti et al., 2020, Figure 3). Empirical calibrations can also suffer from sample selection effects whereby individually-detected [O III] λ4363-emitters yield metallicities of more extreme star-forming regions whereas metallicities from galaxy samples stacked in order to detect [O III] λ4363 may be more representative of "typical" galaxies (Curti et al., 2017; Sanders et al., 2020) that fall on the $M_* - \mathrm{SFR}$ relation.

When considering high-z galaxies, it is unknown if these locally-calibrated strong-line relations, reflective of H II-region conditions in the local Universe, are applicable for estimating metallicity. Excitation diagrams have shown that star-forming region conditions likely evolve with redshift; this is most notably seen in the [O III] λ5007/Hβ vs. [N II] λ6583/Hα Baldwin – Phillips – Terlevich (N2-BPT; Baldwin et al., 1981) diagram, where the locus of star-forming, high-z galaxies is offset from the locus of local, star-forming, SDSS galaxies (e.g., Steidel et al., 2014; Shapley et al., 2015; Strom et al., 2017, 2018; Kashino et al., 2017, 2019; Runco et al., 2022). While it is a current matter of debate as to what is driving this evolution in the locus and thus the H II region physical conditions (see Kewley et al. (2013) for an analysis of several possibilities such as the ionization parameter, electron density, hardness of the ionizing spectrum, and N/O abundance ratio), it is clear

that caution must be taken when applying strong-line metallicity methods at high redshift. Calibrations are needed that are derived from objects with analogous physical conditions to typical, star-forming, high-z galaxies.

In this paper, we analyze a $\langle z \rangle = 2.3$ composite spectrum of 16 gravitationally-lensed, typical, star-forming dwarf galaxies selected independent of emission-line strength. In particular, we study the direct-method metallicity from this composite, derived from a detection of the T_e-sensitive [O III] λ4363 auroral-line. The paper is organized as follows: In Section 3.2, we discuss our observations, data reduction, and sample selection. In Section 3.3, we discuss our spectral-fitting and stacking methodologies, introduce our composite spectrum, and calculate physical properties of our stacking sample and composite. In Section 3.4, we present our analysis and discussion in regard to how representative our sample is of typical, $z \sim 2.3$, star-forming dwarf galaxies, the applicability of locally-calibrated strong-line metallicity diagnostics at high-z, the slope and normalization of the $z \sim 2.3$ MZR, and the redshift evolution of the FMR. In Section 3.5, we summarize our results. Finally, in Section 3.6, we briefly describe the reasoning and methods behind our refitting of the Bian et al. (2018) strong-line metallicity relations. Throughout this paper, uncertainties reflect our bootstrapped error spectrum for the composite unless stated otherwise. We assume a ΛCDM cosmology with $H_0 = 70$ km s^{-1} Mpc^{-1}, $\Omega_\Lambda = 0.7$, and $\Omega_m = 0.3$.

3.2 Observations, Data Reduction, and Sample Selection

The focus of this paper is the careful analysis of a stack – from which the oxygen abundance is directly measured – of 16 gravitationally-lensed, star-forming, dwarf galaxies

at the peak of cosmic star formation. These galaxies at $1.7 < z < 2.6$ have stellar masses of $\log(M_*/M_\odot) < 9.0$ and probe typical dwarf galaxies in this epoch, complimenting the recent large statistical studies of more massive galaxies at these redshifts, such as the Keck Baryonic Structure Survey (KBSS-MOSFIRE; Steidel et al., 2014) and the MOSFIRE Deep Evolution Field survey (MOSDEF; Kriek et al., 2015). In this section, we detail the photometric and spectroscopic observations and data reduction of these galaxies and the larger parent surveys from which the galaxies are drawn. We also discuss the selection strategy of these 16 objects chosen for stacking.

3.2.1 Photometric Data and Reduction

The galaxy stacking sample is drawn from a spectroscopic follow-up survey of the photometric *Hubble Space Telescope* (*HST*) survey of Alavi et al. (2014, 2016), which was conducted to study faint, low-mass, star-forming galaxies gravitationally-lensed by the foreground galaxy clusters Abell 1689, MACS J0717.5+3745, and MACS J1149.5+2223, among others (hereafter A1689, MACS J0717, and MACS J1149, respectively). This *HST* survey compliments the *Hubble* Frontier Fields (HFF; Lotz et al., 2017) survey of lensing clusters by both adding deep near-ultraviolet (UV) images of the HFF clusters (of which MACS J0717 and MACS J1149 are members) to the deep HFF optical and near-infrared (IR) datasets as well as by adding or including deep near-UV to near-IR photometry of another lensing cluster, A1689.

For galaxies lensed by A1689, near-UV images were taken over two programs in the F225W, F275W, and F336W bandpasses with the Wide Field Camera 3 (WFC3)/UVIS channel on the *HST*. As part of Program ID 12201 (PI: B. Siana), F275W was observed

for 30 orbits, and F336W was observed for 4 orbits. As part of Program ID 12931 (PI: B. Siana), F336W was observed for an additional 14 orbits (18 orbits total), and F225W was observed for 10 orbits. In the optical, we used existing *HST* photometry, taken with the Advanced Camera for Surveys (ACS)/WFC channel, in the F475W, F625W, F775W, and F850LP bandpasses (PID: 9289; PI: H. Ford) as well as in the F814W bandpass (PID: 11710; PI: J. Blakeslee). A summary of the number of orbits for each near-UV and optical filter, as well as the 5σ depths for a $0\farcs2$ radius aperture, can be found in Alavi et al. (2016, Table 1). In the near-IR, existing images taken over 1-2 orbits with the F125W and F160W filters and the *HST* WFC3/IR channel (PID: 11802; PI: H. Ford) were used. We note that the near-IR footprint for A1689 is smaller than the near-UV and optical footprints, covering 10 of the 13 stacking sample galaxies (see sample selection in Section 3.2.3) lensed by A1689.

Galaxies behind the lensing clusters MACS J0717 and MACS J1149 were observed with the WFC3/UVIS channel for 8 orbits in both the F275W and F336W bandpasses as part of the Alavi et al. (2016) *HST* survey under Program ID 13389 (PI: B. Siana). In the optical and near-IR, these clusters were observed with *HST* Director's Discretionary time as part of the Hubble Frontier Fields survey (Lotz et al., 2017). As with all clusters in this survey (6 total), MACS J0717 and MACS J1149 were observed for 70 orbits with each ACS/WFC and WFC3/IR (140 orbits total). These HFF clusters are observed in the F435W, F606W, and F814W filters with ACS/WFC and the F105W, F125W, F140W, and F160W filters with WFC3/IR (PID: 13498 for MACS J0717; PID: 13504 for MACS J1149; PI: J. Lotz). As for the optical and near-UV photometry of A1689, the depths and orbits (both from the HFF survey and other projects) for each filter are listed for MACS J0717

in Alavi et al. (2016, Table 1). This information can be found for MACS J1149 via the Mikulski Archive for Space Telescopes (MAST) website for the HFF survey.[1]

The data reduction, calibration, and photometric measurements for MACS J0717 and MACS J1149 are detailed in Alavi et al. (2016), as is the UV data reduction and calibration for A1689. The reduction and calibration of the optical data from A1689, as well as the photometric measurements for this cluster, are discussed in Alavi et al. (2014). The near-IR photometry of A1689 was reduced in the same way as the UV and optical data with the exception that a larger pixel scale of $0\farcs08$ was used in the final drizzled images. As described in Alavi et al. (2014, 2016), our main photometric catalog for A1689 is built on the UV and optical images with a pixel scale of $0\farcs04$. For the areas of A1689 with near-IR coverage, the multi-band photometry (from UV to near-IR) was remeasured on images with larger pixel scales and that are PSF-matched to the F160W data. The estimations of photometric redshifts, which were used to select the spectroscopic follow-up survey sample detailed in Section 3.2.2, are described in Alavi et al. (2016).

3.2.1.1 Lens Models

When working with objects gravitationally-lensed by foreground galaxy clusters, accurate lens models are imperative for correcting observed photometry and spectroscopy for the lensing magnification. This correction is necessary for the determination of an object's intrinsic properties (e.g., stellar mass, SFR, etc.). Alavi et al. (2016) detail the lens models considered and used for the HFF clusters and A1689, all of which, while constructed with different assumptions and methodologies, are constrained by the location and redshift of

[1] https://archive.stsci.edu/prepds/frontier/macs1149.html

known multiply-imaged systems. As stated in Alavi et al. (2016), for the HFF clusters we use the lens models derived by the Clusters As TelescopeS (CATS) collaboration,[2] specifically the models of Limousin et al. (2016) and Jauzac et al. (2016) for MACS J0717 and MACS J1149, respectively. For A1689, we use the lens model of Limousin et al. (2007). These parametric models are all derived via mass reconstruction done with the `LENSTOOL`[3] software (Jullo et al., 2007).

3.2.2 Spectroscopic Data and Reduction

As a follow-up to the photometric *HST* survey of Alavi et al. (2014, 2016), a spectroscopic survey was conducted between 2014 January and 2017 March to obtain near-IR (rest-optical) spectroscopy of select galaxies with the Multi-Object Spectrometer For InfraRed Exploration (MOSFIRE; McLean et al., 2010, 2012) on the 10 m Keck I telescope. Galaxies for this survey were selected to have high magnifications, observed optical magnitudes (F606W or F625W) less than 26.0 (AB), and photometric redshifts in three redshift ranges, $1.37 \leqslant z \leqslant 1.70$, $2.09 \leqslant z \leqslant 2.61$, and $2.95 \leqslant z \leqslant 3.80$, so that the strong, rest-optical, nebular emission lines of the galaxies lie in the near-IR atmospheric transmission windows. Early selection of galaxies in MACS J0717 and MACS J1149 used photometric redshifts from the CLASH survey (Postman et al., 2012). In all, 151 sources were observed across 9 masks. For galaxies that fall into the two lowest redshift ranges, the strong, nebular emission lines targeted are [O II] $\lambda\lambda$3726, 3729, Hβ, [O III] $\lambda\lambda$4959, 5007, Hα, and [N II] $\lambda\lambda$6548, 6583. To this end, observations of galaxies in the lowest redshift range were

[2] https://archive.stsci.edu/prepds/frontier/lensmodels/

[3] https://projets.lam.fr/projects/lenstool/wiki

conducted using the Y-, J-, and H-band filters, whereas the J-, H-, and K-band filters were used for the two highest redshift ranges. We note that while we targeted the strong, nebular emission lines in the highest redshift range as well, Hα and the [N II] doublet were not observed as they fall outside of the K-band's wavelength coverage.

Observations used an ABBA dither pattern with a $2\farcs5$ dither spacing. The individual exposure time for J-band and H-band data was 120 s and was 180 s for Y-band and K-band data. In total, across the 9 masks, the J-band was observed between 48 m and 112 m, the H-band between 56 m and 112 m, and the K-band between 60 m and 120 m, for average total exposure times of 81 m, 85 m, and 82 m, respectively. Data in the Y-band were taken for one mask in A1689 for a total of 96 m. In each mask, we used $0\farcs7$-wide slits, yielding spectral resolutions of $R = 3388,\ 3318,\ 3660,$ and 3610 for the Y-, J-, H-, and K-bands, respectively.[4] Our typical FWHM seeing for a given mask/filter combination was $0\farcs71$.

The spectroscopic data obtained with MOSFIRE were reduced with the MOSFIRE Data Reduction Pipeline[5] (DRP). This DRP returns a 2D flat-fielded, wavelength-calibrated, background-subtracted, and rectified spectrum for each slit in a given mask. This output spectrum is a composite of the multiple spectra taken at each nod position. For Y-, J-, and H-band spectra, wavelength calibration is performed using the night-sky lines, whereas a combination of night-sky lines and a neon arc lamp is used for K-band spectra owing to the faintness of the sky lines and the dominance of thermal noise at the red end of the band. Once the 2D spectra were produced, the 1D spectra were extracted using the custom IDL

[4] https://www2.keck.hawaii.edu/inst/mosfire/grating.html

[5] https://keck-datareductionpipelines.github.io/MosfireDRP/

software BMEP[6] from Freeman et al. (2019). Each spectrum is flux-calibrated with two stars. A standard star of spectral type B9 V to A2 V is first used to apply a wavelength-dependent calibration. It is ensured that this standard star was observed at an air mass similar to that of the mask under consideration. Following this step, an absolute flux calibration is conducted using a star that was included in the corresponding mask.

3.2.3 Sample Selection for Dwarf Galaxy Stack

The galaxies that comprise the stack mentioned in the opening of this section are drawn from the photometric and spectroscopic surveys detailed above. These galaxies are required to have a robust spectroscopic redshift and spectroscopic coverage of the strong, rest-optical, nebular emission lines: [O II] $\lambda\lambda$3726, 3729, Hβ, [O III] λ4959, Hα, and [N II] $\lambda\lambda$6548, 6583. Additionally, these galaxies must have spectroscopic coverage of Hγ and the faint [O III] λ4363 auroral emission line. The auroral line is essential for determining gas-phase metallicity directly as it is a component of the emission-line ratio used to estimate electron temperature (T_e; see Section 3.3.5 for more details). These redshift and coverage requirements yield a sample of 18 galaxies and 24 total spectra when accounting for multiply-imaged systems, of which we have four in our sample. A final cut is made on stellar mass (see Section 3.3.2 on mass estimation) to ensure that our sample lies in the dwarf galaxy regime ($\log(M_*/M_\odot) < 9.0$). With this cut, two galaxies are removed from our sample, yielding a final count of 16 galaxies (22 total spectra) ranging in redshift from $z = 1.70$ to $z = 2.59$ ($z_{\rm mean} = 2.30$).

[6] https://github.com/billfreeman44/bmep

We note here that Hγ coverage is included as a requirement so as to provide another Balmer decrement with which to estimate the dust extinction from the stack. Due to the close proximity of Hγ (4340 Å) to [O III] λ4363, this inclusion does not affect our sample size. We also note here that we do not require spectroscopic coverage of the [O III] λ5007 line of the [O III] $\lambda\lambda$4959, 5007 doublet so as to maximize our galaxy count by including those sources for which [O III] λ5007 falls just redward of a given filter. Instead, when necessary, we make use of the T_e-insensitive intrinsic intensity ratio of the doublet: [O III] λ5007/[O III] λ4959 = 2.98 (Storey & Zeippen, 2000). Lastly, while we do not select galaxies based on the strength of any given emission line, we do note that each spectrum has a signal-to-noise ratio (S/N) for [O III] λ5007 of S/N $>$ 5, ensuring accurate normalization of each spectrum (by [O III] λ5007) during the stacking process (see Section 3.3.3). A summary of our sample, and some of the galaxies' physical properties, are listed in Table 3.1.

3.3 Measurements and Stacking Methodology

In this section, we detail our methodologies for fitting the spectroscopy and photometry of the dwarf galaxies in our stacking sample. We also discuss how various physical properties are estimated either for the individual galaxies or for the "sample-average" dwarf galaxy, represented by a composite spectrum of these dwarfs. We begin by discussing the measurements made for individual galaxies and then proceed to the construction and analysis of the composite spectrum.

Table 3.1: Summary and Properties of $\langle z \rangle = 2.3$ Dwarf Galaxy Stacking Sample

Galaxy	Spec. ID	z^a	R.A.^{b}	Dec.^{b}	$\log(\frac{M_*}{M_\odot})^c_{\rm fid}$	$\log(\frac{M_*}{M_\odot})^d$	$\frac{\rm SFR}{M_\odot\ {\rm yr}^{-1}}^e$	$\log(\frac{L_{H\alpha}}{\rm erg\ s^{-1}})^f$
1	A1689-1037	1.70089	13:11:35.197	-01:20:25.040	$7.71^{+0.20}_{-0.37}$	$8.12^{+0.22}_{-0.23}$	0.151 ± 0.004	$40.511^{+0.013}_{-0.013}$
2	A1689-1197	1.70261	13:11:29.689	-01:20:08.769	$8.45^{+0.05}_{-0.05}$	$8.87^{+0.13}_{-0.10}$	$\_^h$	$\_^h$
	A1689-370	1.70257	13:11:32.406	-01:21:16.027	$8.31^{+0.05}_{-0.05}$	$8.71^{+0.12}_{-0.11}$	1.779 ± 0.020	$41.583^{+0.005}_{-0.005}$
	Composite^g	1.70259	–	–	$8.36^{+0.03}_{-0.04}$	$8.79^{+0.08}_{-0.08}$	1.779 ± 0.020	$41.583^{+0.005}_{-0.005}$
3	A1689-280	1.70316	13:11:31.886	-01:21:26.014	$7.91^{+0.04}_{-0.05}$	$8.70^{+0.23}_{-0.21}$	0.692 ± 0.011	$41.173^{+0.007}_{-0.007}$
4	A1689-257	1.70355	13:11:26.426	-01:21:31.277	$7.81^{+0.08}_{-0.10}$	$8.47^{+0.18}_{-0.14}$	1.809 ± 0.027	$41.590^{+0.006}_{-0.007}$
5	A1689-1751	2.38159	13:11:31.333	-01:19:18.559	$8.82^{+0.07}_{-0.09}$	$9.23^{+0.17}_{-0.12}$	3.071 ± 0.145	$41.820^{+0.020}_{-0.021}$
6	A1689-232	2.38709	13:11:32.794	-01:21:27.893	$7.06^{+1.10}_{-7.06}$	$8.70^{+0.20}_{-0.19}$	2.368 ± 0.185	$41.707^{+0.033}_{-0.035}$
7	M0717-3958	2.39329	07:17:27.442	+37:45:25.475	$8.69^{+0.11}_{-0.16}$	$9.42^{+0.09}_{-0.12}$	2.072 ± 0.139	$41.649^{+0.028}_{-0.030}$
	M0717-4517	2.39330	07:17:27.050	+37:45:09.695	$8.99^{+0.07}_{-0.09}$	$9.46^{+0.09}_{-0.09}$	2.350 ± 0.128	$41.704^{+0.023}_{-0.024}$
	Composite^g	2.39329	–	–	$8.84^{+0.07}_{-0.08}$	$9.44^{+0.07}_{-0.07}$	2.223 ± 0.094	$41.680^{+0.018}_{-0.019}$
8	A1689-1059	2.41141	13:11:25.228	-01:20:19.309	$8.40^{+0.03}_{-0.03}$	$9.03^{+0.17}_{-0.21}$	5.041 ± 0.464	$42.035^{+0.038}_{-0.042}$
9	A1689-1467	2.51903	13:11:26.118	-01:19:42.837	$8.21^{+0.10}_{-0.13}$	$8.76^{+0.18}_{-0.17}$	2.287 ± 0.441	$41.692^{+0.077}_{-0.093}$
10	A1689-1216	2.54082	13:11:31.981	-01:20:07.173	$8.70^{+0.06}_{-0.07}$	$9.07^{+0.15}_{-0.15}$	0.822 ± 0.144	$41.248^{+0.070}_{-0.084}$
	A1689-1292	2.54064	13:11:26.528	-01:19:55.146	$8.90^{+0.06}_{-0.07}$	$9.07^{+0.11}_{-0.13}$	0.667 ± 0.097	$41.157^{+0.059}_{-0.068}$
	A1689-537	2.54046	13:11:29.795	-01:21:05.969	$8.60^{+0.08}_{-0.10}$	$9.05^{+0.17}_{-0.19}$	1.725 ± 0.397	$41.570^{+0.090}_{-0.114}$
	Composite^g	2.54065	–	–	$8.70^{+0.04}_{-0.04}$	$9.07^{+0.08}_{-0.08}$	0.755 ± 0.079	$41.211^{+0.043}_{-0.048}$
11	A1689-470	2.54112	13:11:26.213	-01:21:09.695	$7.28^{+0.31}_{-7.28}$	$8.49^{+0.22}_{-0.21}$	1.235 ± 0.160	$41.425^{+0.053}_{-0.060}$
12	A1689-1451	2.54201	13:11:28.682	-01:19:42.849	$8.22^{+0.08}_{-0.09}$	$9.10^{+0.17}_{-0.16}$	0.756 ± 0.128	$41.211^{+0.068}_{-0.080}$
13	A1689-722	2.54247	13:11:33.915	-01:20:52.526	$8.78^{+0.20}_{-0.36}$	$8.81^{+0.22}_{-0.24}$	3.771 ± 0.291	$41.910^{+0.032}_{-0.035}$

Notes: The table continues on the next page.

Table 3.1: Continued

Galaxy	Spec. ID	z^a	R.A.^{b}	Dec.^{b}	$\log(\frac{M_*}{M_\odot})^c_{\rm fid}$	$\log(\frac{M_*}{M_\odot})^d$	$\frac{\rm SFR}{M_\odot\ yr^{-1}}^e$	$\log(\frac{L_{H\alpha}}{erg\ s^{-1}})^f$
14	M0717-1531	2.55185	07:17:32.547	+37:45:02.348	$9.08^{+0.11}_{-0.14}$	$9.94^{+0.25}_{-0.06}$	5.558 ± 0.441	$42.078^{+0.033}_{-0.036}$
	M0717-3187	2.55159	07:17:35.089	+37:45:48.120	$8.90^{+0.10}_{-0.14}$	$9.67^{+0.17}_{-0.15}$	5.965 ± 0.268	$42.109^{+0.019}_{-0.020}$
	M0717-5970	2.55167	07:17:30.613	+37:44:22.798	$8.82^{+0.17}_{-0.29}$	$9.66^{+0.08}_{-0.10}$	3.969 ± 0.283	$41.932^{+0.030}_{-0.032}$
	Composite^g	2.55168	–	–	$8.93^{+0.07}_{-0.09}$	$9.72^{+0.07}_{-0.07}$	5.109 ± 0.178	$42.041^{+0.015}_{-0.015}$
15	A1689-217	2.59181	13:11:27.623	-01:21:35.622	$8.23^{+0.04}_{-0.04}$	$9.22^{+0.17}_{-0.18}$	9.194 ± 0.313	$42.297^{+0.015}_{-0.015}$
16	M1149-2185	2.59366	11:49:40.162	+22:25:07.571	$8.83^{+0.05}_{-0.06}$	$9.33^{+0.04}_{-0.04}$	33.602 ± 3.037	$42.859^{+0.038}_{-0.041}$

[a]Spectroscopic redshift – all uncertainties are $\sigma_z \lesssim 6 \times 10^{-5}$.

[b]Right Ascension: hh:mm:ss.sss; Declination: dd:mm:ss.sss; Equinox: J2000

[c]Our fiducial de-magnified stellar mass estimates assuming constant star formation histories (SFH). See Section 3.3.2.

[d]De-magnified stellar masses estimated assuming non-parametric SFHs. These estimates are considered in Section 3.4.3.2.

[e]Star formation rates calculated from Hα luminosities assuming a Chabrier (2003) IMF. See Section 3.3.4 and Equation 3.1.

[f]All Hα luminosities are corrected for slit-loss (Section 3.3.1.1), magnification (Section 3.2.1.1), and dust extinction (Section 3.3.4). The dust extinction correction applied to each luminosity is the same and is derived from the composite spectrum of the total sample. The luminosities are not corrected for stellar absorption, which on average would result in an increase of $< 1\%$.

[g]The redshift, stellar masses, and Hα luminosity are weighted-averages. The SFR is calculated from this luminosity.

[h]Hα was not used here or in the composite of A1689-1197 and A1689-370 because it is at the edge of our Y-band spectrum of this image (with [N II] λ6583 falling outside of our coverage).

3.3.1 Fitting the Individual Emission-Line Spectra

Each emission-line spectrum in our stacking sample (22 total), corresponding to either the single image of a galaxy or one of a multiply-imaged galaxy, is fit using the Markov Chain Monte Carlo Ensemble sampler `emcee`[7] (Foreman-Mackey et al., 2013). Lines are fit to each spectrum's continuum, and single-Gaussian profiles are fit to the emission lines. To minimize the impact on the spectral-fitting from pixels contaminated by sky lines, we removed, prior to fitting, any pixels with a corresponding error spectrum value $> 3\times$ the median error value over the range of the fit.

When fitting, each spectroscopic band (Y, J, H, K) was considered separately. For each spectrum, the slope and intercept of the continuum were free parameters. In the band containing Hα and the [N II] doublet (the H- or the K-band), the free parameters also included the redshift of the spectrum, the width of the emission lines (each line having the same width), and the amplitudes of the Hα and [N II] λ6583 lines, with the amplitude of [N II] λ6583 constrained such that [N II] λ6583/[N II] λ6548 = 2.95 (Acker et al., 1989).

In the band containing Hγ, [O III] λ4363, Hβ, and [O III] λ4959 (the J- or H-band), two fits were conducted due to the large wavelength separation between [O III] λ4363 and Hβ. The portion of the spectrum containing Hβ and [O III] λ4959 (and [O III] λ5007 if covered) was fit first, having the free parameters of line-width for the filter, redshift, and emission-line amplitudes. If [O III] λ5007 is within the spectrum's wavelength coverage, its amplitude was fit with the constraint that [O III] λ5007/[O III] λ4959 = 2.98 (Storey & Zeippen, 2000). Otherwise, the amplitude of [O III] λ4959 was fit, and the line's flux was

[7] https://emcee.readthedocs.io/en/v2.2.1/

multiplied by the aforementioned intensity ratio in order to estimate the [O III] λ5007 flux. With a best-fit width and redshift in-hand from the first fit to the filter, the fainter Hγ and [O III] λ4363 lines were then fit with these two parameters fixed.

Finally, in the band containing the [O II] $\lambda\lambda$3726, 3729 doublet (the Y- or J-band), the redshift and line-width were fixed to the values fit to the complete spectrum's highest S/N line in order to avoid complications resulting from the doublet lines' small wavelength separation. In addition to the continuum parameters, only the [O II] lines' amplitudes were free parameters in these fits.

Ultimately, the final redshift given to the full spectrum is the weighted average of the redshifts fit to the J (H)- and H (K)-bands.

3.3.1.1 Slit-Loss Correction

When measuring the emission-line fluxes from spectra observed through slit masks, care must be taken to account for loss of flux outside of the slits in order to recover the full integrated flux values. To this end, our emission-line fluxes were slit-loss-corrected on a galaxy-by-galaxy basis using the methodology of Emami et al. (2020).

3.3.2 SED-Fitting and Stellar Mass Estimation

To determine the stellar masses of the galaxies in our stacking sample, we fit spectral energy distributions (SEDs) to our *HST* near-UV to near-IR photometry (we note that three galaxies lensed by A1689 lack near-IR photometry; see Section 3.2.1). At high redshift, observations suggest that high equivalent width emission lines are fairly common, particularly in lower-mass galaxies like those in our sample (Reddy et al., 2018b). Therefore,

prior to SED-fitting, we subtracted off any contribution to the photometry from the slit-loss-corrected, nebular emission lines. We also added an additional 3% flux error, in quadrature, to all bands in order to account for systematic errors in the photometry (Alavi et al., 2016). To this emission-line-corrected photometry, we then fit Bruzual & Charlot (2003) stellar population synthesis models using the SED-fitting code `FAST`[8] (Kriek et al., 2009). We assume constant star formation histories (SFH), a Chabrier (2003) initial mass function (IMF), stellar metallicities of 0.2 $Z_\odot$ or 0.4 $Z_\odot$, and a Calzetti et al. (2000) dust attenuation curve. The redshifts of the galaxies are fixed to their fit spectroscopic values.

Uncertainties on the properties estimated by `FAST` (e.g., stellar mass, SFR, A_V, etc.) are derived using a Monte Carlo approach where the photometry being fit is perturbed within its uncertainties and is then refit, this process being repeated 300 times. From these 300 realizations of the SED, 68% confidence intervals are determined for each estimated property. In Table 3.1, we list the best-fit stellar mass, and its uncertainty, of each galaxy in our stacking sample. The stellar mass associated with our full-sample composite, detailed below, is taken to be the median of these individual masses, $\log(M_*/\mathrm{M}_\odot)_{\mathrm{med}} = 8.29$. We note that the best-fit SEDs, stellar masses, and all other affected properties are de-magnified based on the lensing models discussed in Section 3.2.1.1.

3.3.3 The Composite Spectrum

While the individual galaxies in our sample display several nebular emission lines at high-S/N (e.g., [O III], Hα), the galaxies are still inherently faint even with high magnification via gravitational-lensing. As a result, many other useful, fainter lines are undetected

[8] https://w.astro.berkeley.edu/~mariska/FAST.html

or marginally-detected in our individual spectra. Such lines can include Hβ and Hγ for estimating extinction from dust, the [O II] doublet for calculating electron density, and especially [O III] λ4363 for estimating electron temperature and metallicity directly. A composite spectrum, or stacked spectrum, of all of our sample galaxies offers a solution to this problem by including in our study both galaxies for which we have individual line detections and galaxies for which we only have upper limits. This composite gives the advantages of both increasing the S/N of faint spectral features and displaying the average spectrum and properties of dwarf galaxies like those in our sample. Additionally, we use composites of the individual spectra of galaxies multiply-imaged by lensing in order to increase the effective exposure times and S/N of those galaxies' spectroscopy.

Our methodology for creating composite spectra is similar when stacking multiple images of sources (4 multiply-imaged galaxies; see Table 3.1) or all of the galaxies in our sample (16 total). We first create the composites for our multiply-imaged galaxies as these composites represent their corresponding galaxies in the full-sample stack. We begin by shifting the slit-loss-corrected, observed spectra to the rest-frame and converting the flux densities into luminosity densities. Each spectrum is then normalized by its slit-loss-corrected, [O III] λ5007 emission-line luminosity. This normalization serves two purposes. It de-magnifies each spectrum implicitly by dividing the magnified spectrum by its magnified [O III] λ5007 luminosity. It also, in the case of stacking our full sample, prevents our composite electron temperature (see Section 3.3.5) from being biased by the brightest [O III] λ5007 source (Sanders et al., 2020). We note here that, prior to normalizing, the spectra and [O III] λ5007 luminosities are not corrected for dust extinction both due to the faintness

of Hβ and Hγ and the sky line contamination of these lines in several of our individual sources. Following normalization, each spectroscopic band's science spectrum and corresponding error spectrum are resampled with the Python tool, `SpectRes`[9,10] (Carnall, 2017), onto a common wavelength grid with a rest-frame dispersion – for the full-sample stack – of 0.38 Å pix^{-1} in the band (Y or J) containing [O II], 0.47 Å pix^{-1} in the band (J or H) containing Hγ through [O III] λ4959, and 0.63 Å pix^{-1} in the band (H or K) containing Hα and [N II]. (Hereafter, these bands will be referred to as the YJ-band, JH-band, and HK-band, respectively.) These rest-frame dispersions are computed by shifting the MOSFIRE J-, H-, and K-band observed-frame dispersions[4] to the median redshift of the stacking sample, $z_{\rm med} \approx 2.465$. Once resampled, the spectra from the full galaxy sample are combined at each wavelength element by taking the median value of all luminosity densities at that point. When stacking the spectra of a multiply-imaged galaxy, the average at each pixel is taken instead, with luminosity densities weighted by their associated uncertainty values. Finally, the composite spectrum of the full stacking sample is multiplied by the median [O III] λ5007 luminosity of the sample, whereas the composite for a multiply-imaged galaxy is multiplied by the lowest-luminosity [O III] λ5007 measurement (a proxy for the least magnified measurement). We note that, in this work, only ratios of emission lines are used from the composite of the full stacking sample since individual luminosity measurements rely on a normalization dependent on our stacking methodology.

The uncertainty spectrum of the stack of all of our sample galaxies is derived via a Monte Carlo approach with bootstrapping. We first create a bootstrapped sample of

[9] https://spectres.readthedocs.io/en/latest/

[10] https://github.com/ACCarnall/SpectRes

number count $N_{\rm boot} = 16$, the number of galaxies in our full stack, by randomly drawing galaxies for the sample with replacement. For each galaxy in this bootstrapped sample, its science spectrum (already shifted and converted to the rest-frame and luminosity densities, respectively) is perturbed according to its error spectrum. This perturbed spectrum is then normalized by its corresponding [O III] λ5007 emission-line luminosity, which has been perturbed based on its own uncertainty value. The normalized, perturbed spectra are then resampled and stacked according to the procedure detailed above. This process is repeated 500 times in order to create an array of composite luminosity densities at each wavelength element. Our uncertainty spectrum is then comprised of the standard deviations of the values in each of these arrays. By constructing our error spectrum via bootstrapping, our uncertainties represent both our measurement errors and sample variance. The composite spectrum of our full stacking sample, as well as its bootstrapped uncertainty spectrum (the light-gray-shaded region), are shown in Figure 3.1. In Figure 3.1, we also show the statistical-only uncertainty spectrum (the dark-gray-shaded region), which was created as described above, but without bootstrap-resampling each iteration. We note that, unless specified otherwise, the stated uncertainties on measurements derived from the composite spectrum reflect the use of the bootstrapped uncertainty spectrum and its consideration of sample variance.

3.3.3.1 Fitting the Composite Emission-Line Spectrum

Prior to their inclusion in the full-sample stack, the composite spectra of the multiply-imaged galaxies are fit in the same manner as the spectra of individual galaxy

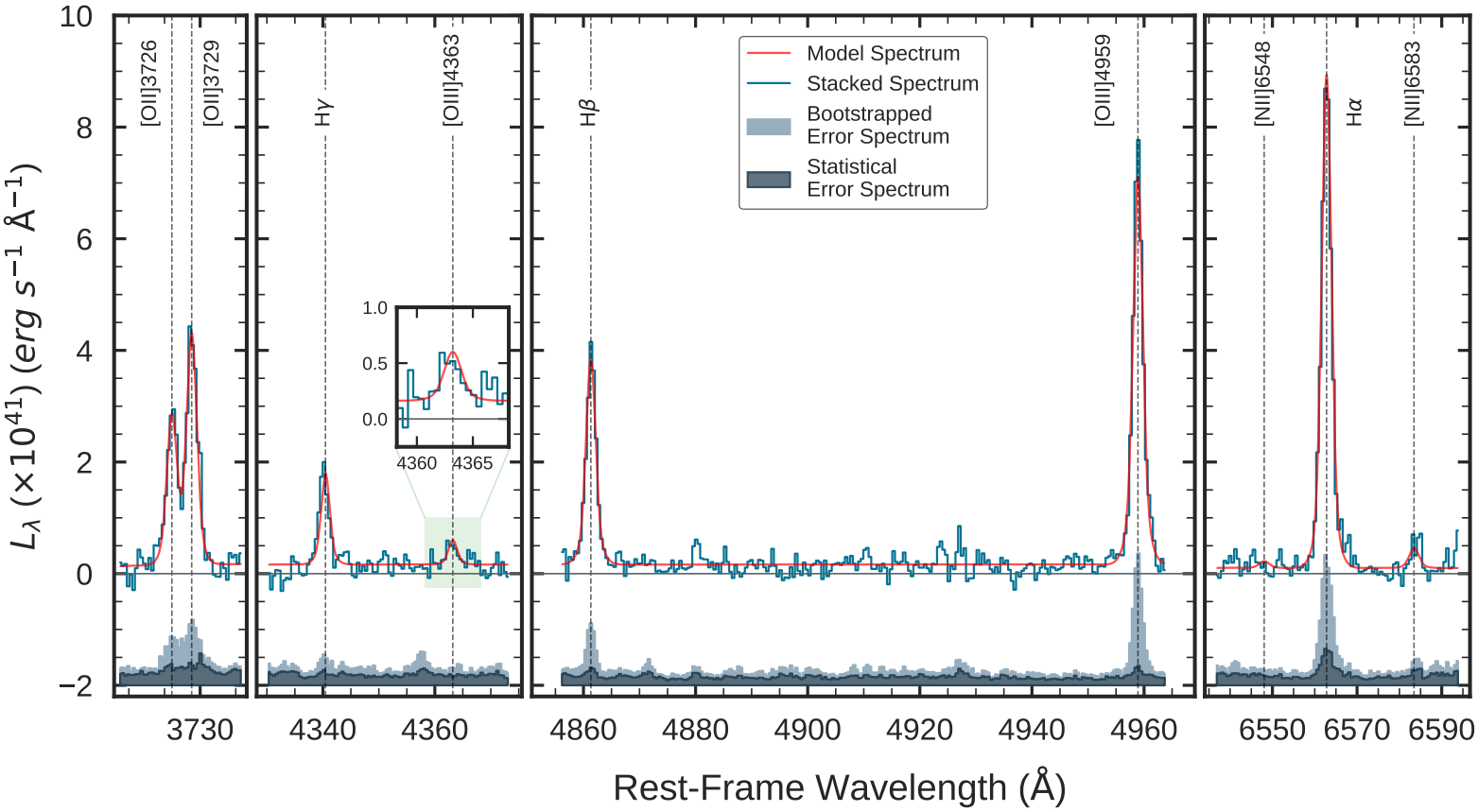

Figure 3.1: The rest-optical composite spectrum of the 16 typical, star-forming, $\langle z \rangle = 2.30$ dwarf galaxies in our stacking sample. The stacked spectrum is shown as the blue unfilled histogram, and its best-fit model is displayed in red. Here we show two offset uncertainty spectra (see Section 3.3.3), one estimated via bootstrapping in order to account for cosmic variance (the light blue-gray-filled histogram), and the other a statistical-only error spectrum (the dark blue-gray-filled histogram). We note that the bootstrapped uncertainty spectrum was used when fitting the model shown here. In the inset box corresponding to the green-shaded spectral region, we show a zoom-in of the weak [O III] λ4363 auroral line, measured at 2.5σ (4.1σ) significance when considering the bootstrapped (statistical-only) uncertainty spectrum. We note that, as a result of our stacking procedure (see Section 3.3.3), the luminosity density values (L_λ) of the composite have a constant but arbitrary magnification. While this makes individual emission-line luminosities unreliable, it does not affect strong-line ratios.

images (see Section 3.3.1). For the full-sample composite, we adopt a slightly different fitting methodology. While spectral-fitting is still facilitated with `emcee`, and the continuum is still fit with a line, the emission lines are fit with a combination of two Gaussian profiles due to the non-Gaussian shape of the high-S/N lines (Steidel et al., 2016). The resultant profiles fit to the emission lines consist of a "broad" Gaussian component and a "narrow" Gaussian component, both of which are centered on the rest-frame wavelength of the emission line in question. In order to maintain consistent line profiles for all emission lines in a given spectral band (*YJ*, *JH*, or *HK*), the fitting of these profiles, and the spectrum, is done in two rounds.

In the first round, the two Gaussian components are fit to the brighter, higher-S/N lines in the composite: Hβ and [O III] λ4959 in the *JH*-band and, separately, Hα in the *HK*-band. For each spectral band, the broad Gaussian has its 1σ-width fixed at 100 km s^{-1} while the narrow-Gaussian 1σ-width is left as a free parameter.[11] Additionally, the free parameters for each fit include the slope and intercept of the continuum, the amplitude(s) of the narrow-Gaussian profile(s), and a narrow-to-broad amplitude ratio for the spectral band.

In the second round, each spectral band is fit in its entirety. The broad Gaussian keeps its 1σ-width fixed to 100 km s^{-1} in all bands. In the *JH*- and *HK*-bands, the narrow 1σ-widths and amplitude ratios fit in the first round are held fixed and applied to all lines

[11] We tested the validity of fixing the broad-Gaussian 1σ-width by comparing luminosities fit to Hβ, [O III] λ4959, and Hα with either a fixed broad-width ($\sigma_{\rm b}$ = 100 km s^{-1}) or the broad-width left as a free parameter. When comparing the luminosities assuming a fixed-width to the weighted-average luminosities of three runs where the broad-width was left free, we found an average percentage difference of $\sim$ 0.4% and $\sim$ 1.0% when assuming the bootstrapped or statistical-only uncertainty spectrum, respectively. The weighted-average luminosities fell well within the uncertainties of the fixed-width luminosities. We therefore find our choice of fixed broad-Gaussian 1σ-width to be robust.

in the corresponding bands: Hγ, [O III] λ4363, Hβ, [O III] λ4959 in the JH-band and Hα and [N II] $\lambda\lambda$6548, 6583 in the HK-band. Free parameters in these bands are the narrow-Gaussian amplitudes of each line and the linear continuum parameters. (The [N II] λ6583 narrow-Gaussian amplitude is constrained in the same manner as in the individual spectra; see Section 3.3.1.) In the YJ-band containing [O II] $\lambda\lambda$3726, 3729, the amplitude ratio from the JH-band is adopted, but the narrow 1σ-width is left as a free parameter. Like with the other bands, the narrow-Gaussian amplitudes and linear continuum parameters are also fit.

The spectral model from these two rounds of fitting can be seen as the red line in Figure 3.1. The total luminosity of each emission line, representing the addition of the broad- and narrow-Gaussian component luminosities, is given in Table 3.2 relative to the total Hβ luminosity. Of particular interest for this study is the 2.5σ (4.1σ) detection of [O III] λ4363 in our composite with the bootstrapped (statistical-only) error spectrum, which will be used in Sections 3.3.5 and 3.3.6 to estimate the electron temperature and gas-phase metallicity directly.

3.3.4 Dust Extinction and SFRs of the Sample

In order to estimate a galaxy's intrinsic emission-line luminosities, from which its galactic properties and interstellar medium (ISM) conditions are derived, a wavelength-dependent correction to the observed luminosities must be made to account for extinction from nebular dust. This correction is typically quantified via Balmer recombination-line ratios of observed hydrogen emission-line luminosities. Ideally, dust extinction would have been compensated for on a galaxy-by-galaxy basis prior to stacking our sample. Unfortunately, many of our individual galaxy spectra have Balmer lines that are too faint or too

impacted by sky lines for this approach to be used. Instead, the nebular dust extinction "typical" of star-forming, dwarf galaxies like those in our sample is estimated via our full-sample composite spectrum.

Prior to using Balmer decrements to estimate dust extinction, corrections to Balmer emission-line luminosities must be made to account for Balmer absorption in the atmospheres of (primarily A-type) stars. Using the Balmer emission lines from our individual spectra and the model continua from our SED-fitting, we estimated the sample-median stellar absorption corrections for Hγ, Hβ, and Hα and applied them to our composite Balmer emission-line luminosities, increasing the measured values by $\sim 5.0\%$, $\sim 1.7\%$, and $\sim 0.4\%$, respectively.

To calculate the nebular dust extinction with our composite spectrum, we assumed Case B intrinsic Balmer ratios of Hα/Hβ = 2.79 and Hα/Hγ = 5.90 for an electron temperature and electron density of T_e([O III]) = 15,000 K and n_e = 100 cm^{-3}, respectively[12] (Dopita & Sutherland, 2003). Further assuming the extinction curve of Cardelli et al. (1989) with $R_V = 3.1$, we find a "typical" color excess of $E(B-V)_{\rm gas} = A_V/R_V = 0.14^{+0.11}_{-0.09}$. This result allows us to correct our composite emission-line luminosities for extinction due to dust, yielding the typical intrinsic values of these lines in star-forming, dwarf galaxies at high redshift. The dust-corrected (and stellar-absorption-corrected in the case of the Balmer lines) emission-line luminosities of the composite spectrum, relative to Hβ, are listed in Table 3.2.

[12]We note that while an electron temperature is assumed when selecting Balmer ratio values, the dependence of those ratios on T_e is weak for typical temperatures in H II regions. To confirm that our assumption of T_e = 15,000 K is valid, we dust-corrected the composite spectrum then calculated T_e (see Section 3.3.5) assuming intrinsic Balmer ratios corresponding to T_e = 10,000 K, 12,500 K, 15,000 K, and 20,000 K (Dopita & Sutherland, 2003). With each variation, we consistently calculated from our dust-corrected composite spectrum a $T_e \sim$ 15,000 K.

Table 3.2: Emission-Line Luminosities of the Composite Spectrum

Line	$\lambda^a_{\rm rest}$	$L^b_{\rm meas}$	$L^{b,c}_{\rm corr}$	A^d_λ
[O II]	3726.032	0.61 ± 0.12	0.70 ± 0.13	0.67
[O II]	3728.815	0.95 ± 0.16	1.09 ± 0.19	0.67
Hγ	4340.459	0.40 ± 0.08	0.44 ± 0.09	0.59
[O III]	4363.209	0.11 ± 0.05	0.11 ± 0.05	0.58
Hβ	4861.321	1.00 ± 0.16	1.00 ± 0.16	0.51
[O III]	4958.910	1.95 ± 0.30	1.89 ± 0.29	0.49
Hα	6562.794	3.14 ± 0.48	2.70 ± 0.42	0.36
[N II]	6583.448	0.13 ± 0.07	0.11 ± 0.06	0.35

Notes: The luminosity of [O III] λ5007 can be calculated via the intrinsic ratio [O III] λ5007/[O III] λ4959 = 2.98.

[a]Rest-frame wavelengths in air (Å).

[b]Luminosities relative to $L_{\rm H\beta}$. We only use ratios of the luminosities from the composite spectrum since the spectrum's normalization is dependent on our stacking methodology.

[c]All luminosities are dust-corrected using the corresponding extinction magnitude (A_λ) given in the last column. Balmer lines are corrected for stellar absorption. See Section 3.3.4. The listed uncertainties do not include systematic errors associated with the dust correction, though these errors are propagated throughout all of our calculations.

[d]Dust extinction magnitudes at $\lambda_{\rm rest}$.

When considering SFRs, similar to how our stellar masses are being reported, we calculate the SFR for each individual galaxy and report the composite SFR as the median value of the sample. These SFRs are calculated with slit-loss-corrected, dust-corrected, Hα luminosities ($L_{\mathrm{H}\alpha}$), de-magnified according to the lensing models in Section 3.2.1.1. The dust-extinction correction of each $L_{\mathrm{H}\alpha}$ value is conducted with the "typical" extinction estimate for the sample found via the composite spectrum. The $L_{\mathrm{H}\alpha}$ values are not corrected for stellar absorption, which on average would result in an increase of $< 1\%$. The $L_{\mathrm{H}\alpha}$ values are converted to SFRs using Equation 3.1 below:

$$\mathrm{SFR}\ (\mathrm{M}_\odot\ \mathrm{yr}^{-1}) = 4.645 \times 10^{-42}\ L_{\mathrm{H}\alpha}\ (\mathrm{ergs\ s}^{-1}) \tag{3.1}$$

This equation is of the same form as the relation in Kennicutt (1998) for calculating SFRs from recombination lines. However, the conversion factor here has been recalculated assuming a metallicity of 0.2 $\mathrm{Z}_\odot$ and a Chabrier (2003) IMF. Our estimates for the SFRs of the individual galaxies are given in Table 3.1, and the SFR associated with the composite spectrum is taken to be the sample-median value of $\mathrm{SFR}_{\mathrm{med}} = 2.25\ \mathrm{M}_\odot\,\mathrm{yr}^{-1}$.

3.3.5 Electron Temperature and Electron Density

The "direct" calculation of metallicity relies on collisionally-excited oxygen emission lines and the nebular properties of electron temperature (T_e) and electron density (n_e), which are responsible for the strength of the collisionally-excited lines. Electron temperature is calculated in two ionization zones of the star-forming, H II regions. In the O^{++} zone, T_e([O III]) is calculated using the electron-temperature-sensitive emission-line ratio [O III]

$\lambda\lambda$ 4959, 5007/[O III] λ4363 and the IRAF routine `NEBULAR.TEMDEN` (Shaw & Dufour, 1994). We note that while T_e([O III]) does have a dependence on electron density, n_e, below $n_e \approx 10^3$ cm^{-3}, T_e([O III]) is insensitive to n_e (Osterbrock & Ferland, 2006; Izotov et al., 2006) and can be calculated assuming the typical $z \sim 2$ H II region electron density of a few hundred per cubic centimeter (Sanders et al., 2016b). For our calculation, we assume $n_e = 150$ cm^{-3} and obtain an electron temperature in the O^{++} region of T_e([O III]) = 15,800 $\pm$ 3,100 K. We note in regard to the assumed n_e that Gburek et al. (2019) studied a galaxy in our present stacking sample, A1689-217, that had a similar electron temperature of T_e([O III]) = 14,300 K. When calculating this temperature, they found that assuming any $n_e < 10^3$ cm^{-3} changed their result by $< 0.5\%$, suggesting our current assumption is robust.

Ideally, the electron temperature in the O^+ ionization region is calculated using measurements of the [O II] $\lambda\lambda$7320, 7330 auroral emission-line doublet. Unfortunately, for the galaxies in our stacking sample, we do not have spectroscopic coverage of these lines. Instead, we calculate T_e([O II]) via the T_e([O III])$-T_e$([O II]) relation of Campbell et al. (1986), reprinted here in Equation 3.2:

$$T_e([\text{O II}]) = 0.7\, T_e([\text{O III}]) + 3000 \text{ K} \tag{3.2}$$

Use of this equation gives us an electron temperature in the O^+ region of T_e([O II]) = 14,100 $\pm$ 2,200 K.

The electron density, n_e, can be derived with the doublet ratio [O II] λ3729/[O II] λ3726, the T_e([O II]) electron temperature, and the IRAF routine `NEBULAR.TEMDEN`. For our composite spectrum, we calculate [O II] λ3729/[O II] λ3726 = 1.56 $\pm$ 0.32 (1.51 $\pm$ 0.12 when

using the statistical-only uncertainty spectrum and associated fits). This corresponds to an electron density of $n_e = 1^{+152}_{-0}$ cm^{-3} ($n_e = 1^{+49}_{-0}$ cm^{-3}), where the "best-fit" n_e value is set to the low-density limit of $n_e = 1$ cm^{-3} as a result of the best-fit [O II] ratio exceeding the maximum theoretical bound of [O II] λ3729/[O II] λ3726 $\lesssim 1.5$ (Osterbrock & Ferland, 2006; Sanders et al., 2016b).

We note that we have significantly detected the component emission lines of the [O II] $\lambda\lambda$3726, 3729 doublet and resolved their individual peaks (S/N(3726, 3729) = (6.4, 7.8); see Figure 3.1). The electron density associated with our dwarf galaxy sample and with the best-fit ratio of these lines is significantly lower than the densities found in more massive galaxies at $1.5 \lesssim z \lesssim 2.5$, which typically lie in the range of $n_e \approx 100 - 300$ cm^{-3} (Steidel et al., 2014; Sanders et al., 2016b; Kashino et al., 2017; Kaasinen et al., 2017; Davies et al., 2021).

3.3.6 Direct Oxygen Abundance

We directly calculate the oxygen abundance, or gas-phase metallicity, of our composite spectrum using the ionic abundance equations of Izotov et al. (2006). These equations utilize the values of T_e([O II]), T_e([O III]), and n_e given in the preceding section as well as the dust-corrected emission-line ratios of [O II] $\lambda\lambda$3726, 3729/Hβ and [O III] $\lambda\lambda$4959, 5007/Hβ. We assume that the total oxygen abundance is the summation of the ionic abundances in the H II region O^+ and O^{++} ionization zones as seen in Equation 3.3. Any higher ionization states of oxygen are deemed to have a negligible contribution to the metallicity.

$$\frac{\mathrm{O}}{\mathrm{H}} \approx \frac{\mathrm{O}^{+}}{\mathrm{H}^{+}} + \frac{\mathrm{O}^{++}}{\mathrm{H}^{+}} \tag{3.3}$$

From our composite spectrum, we report a typical gas-phase metallicity for high-redshift, star-forming, dwarf galaxies of $12 + \log(\mathrm{O/H}) = 7.87^{+0.24}_{-0.22}$ ($0.15^{+0.11}_{-0.06}$ $\mathrm{Z}_{\odot}$; Asplund et al., 2021). This metallicity estimate, as well as the calculations from our composite spectrum of the other physical properties detailed in Section 3.3, are summarized in Table 3.3.

3.4 Results and Discussion

This section of the manuscript will take the measurements derived in the previous section from our dwarf galaxy sample and composite spectrum and analyze them in the context of strong-line abundance diagnostics and global galaxy scaling relations. Prior to this, however, it is crucial to look at our sample and stack compared to the broader galaxy population at $1.7 \lesssim z \lesssim 2.6$ in order to assess how representative our sample is of typical, dwarf galaxies at this epoch.

3.4.1 How Representative is our Sample?

Here we will consider two main diagnostics, the [N II] Baldwin-Phillips-Terlevich diagram (N2-BPT; Baldwin et al., 1981) and the M_{*} – SFR relation, or "star-forming main sequence."

Table 3.3: Properties of the Dwarf Galaxy Composite

Property	Value
$z^{a}_{\rm mean}$	2.30
$\log(M_*/{\rm M}_\odot)^{a,b}_{\rm med,\,fiducial}$	8.29
$\log(M_*/{\rm M}_\odot)^{a,c}_{\rm med}$	8.92
${\rm SFR}^{\rm med}_{\rm H\alpha}$ $({\rm M}_\odot\ {\rm yr}^{-1})^{a}$	2.25
$E(B-V)_{\rm gas}$	$0.14^{+0.11}_{-0.09}$
$n_e^{\rm boot}$ $({\rm cm}^{-3})^{d}$	1^{+152}_{-0}
$n_e^{\rm stat}$ $({\rm cm}^{-3})^{d}$	1^{+49}_{-0}
T_e([O II]) (K)	$14{,}100 \pm 2{,}200$
T_e([O III]) (K)	$15{,}800 \pm 3{,}100$
$12+\log({\rm O}^+/{\rm H}^+)$	$7.27^{+0.28}_{-0.19}$
$12+\log({\rm O}^{++}/{\rm H}^+)$	$7.71^{+0.27}_{-0.20}$
$12+\log({\rm O/H})_{\rm direct}$	$7.87^{+0.24}_{-0.22}$
Z $({\rm Z}_\odot)$	$0.15^{+0.11}_{-0.06}$

Notes: All uncertainties here (except for $n_e^{\rm stat}$) derive from the composite bootstrapped error spectrum. See Figure 3.1.

[a]Mean and median values of the individual galaxies in the stacking sample. See Table 3.1. All other values derive from the composite spectrum.

[b]Our fiducial median stellar mass assuming constant SFHs.

[c]The median stellar mass assuming non-parametric SFHs. See Section 3.4.3.2.

[d]$n_e^{\rm boot}$ and $n_e^{\rm stat}$ assume the bootstrapped and statistical-only error spectrum (and associated fits), respectively. Both "best-fit" values are set as the low-density limit of $n_e = 1$ cm^{-3}. See Section 3.3.5.

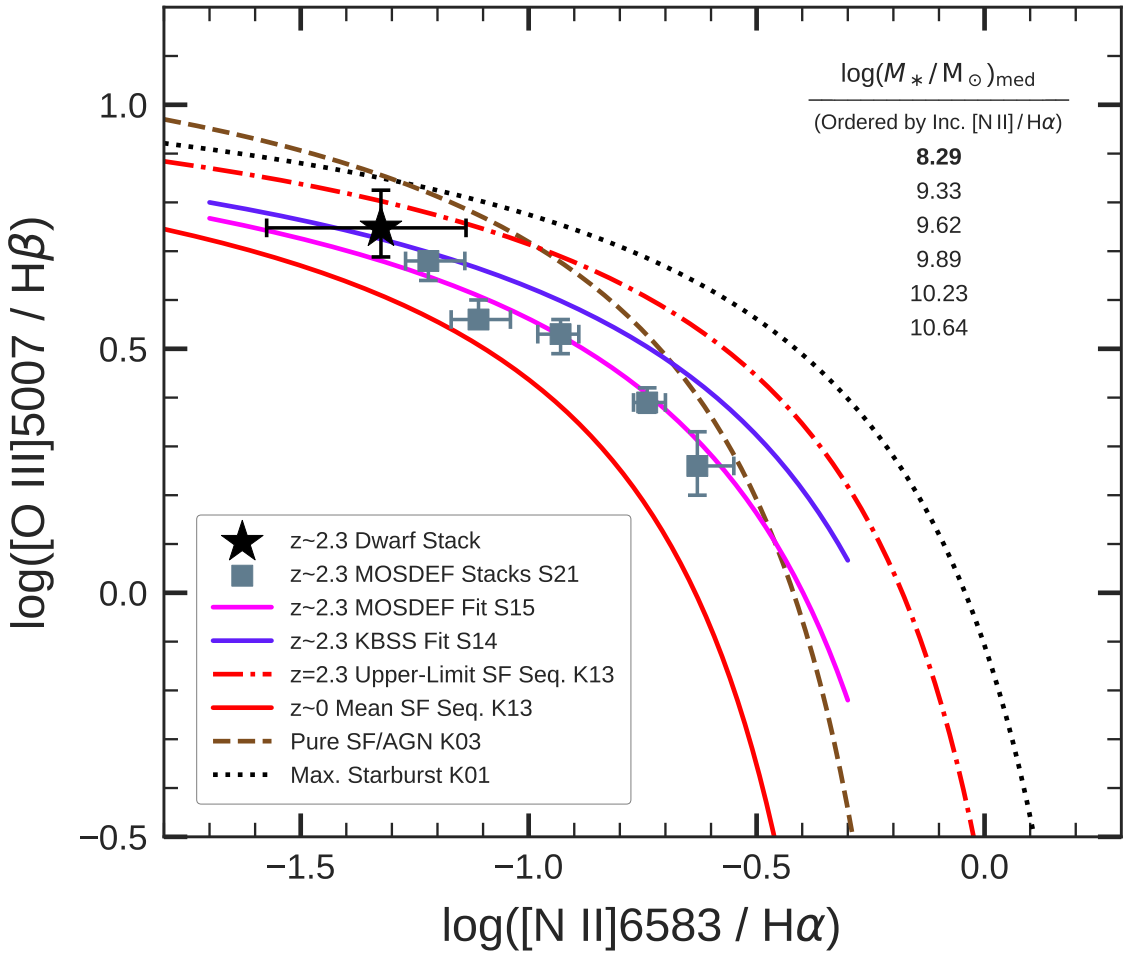

Figure 3.2: The [O III] λ5007/Hβ vs. [N II] λ6583/Hα BPT diagnostic diagram. Our $\langle z \rangle = 2.30$ dwarf galaxy composite is shown as the black star. The $z \sim 2.3$ M_*-binned stacks of star-forming (SF) MOSDEF galaxies from Sanders et al. (2021, S21) are shown as gray squares. The median M_* of each of these stacks, and of our composite (in bold), is listed in the upper right-hand corner of the plot by order of increasing [N II]/Hα of the stacks, highlighting that our stack is an extension to lower M_* (and O/H via the MZR) of the MOSDEF survey. The $z \sim 2.3$ SF sequences of the MOSDEF (Shapley et al., 2015, S15) and KBSS-MOSFIRE (Steidel et al., 2014, S14) surveys are shown by the magenta and purple lines, respectively. Like these sequences and the MOSDEF stacks, our composite also lies offset from the $z \sim 0$ mean SF sequence given by the red line and parameterized by Kewley et al. (2013, K13). We plot the $z = 2.30$ SF sequence upper-limit from K13 as the red dot-dashed line. The demarcation between SF galaxies and AGN of Kauffmann et al. (2003, K03) is given by the dashed brown line, and the "maximum starburst" curve of Kewley et al. (2001, K01) is given by the dotted black line.

3.4.1.1 N2-BPT Diagnostic Diagram

In Figure 3.2, we show the location of our $\langle z \rangle = 2.30$ stack of star-forming, dwarf galaxies on the [O III] λ5007/Hβ vs. [N II] λ6583/Hα BPT diagnostic diagram. This diagram is a useful tool for distinguishing between star-forming galaxies (SFGs) and AGN through optical strong-line ratios (Kewley et al., 2001; Kauffmann et al., 2003; Kewley et al., 2013, K01, K03, K13, respectively), all without needing to apply a dust correction to the line fluxes. The SFG locus of the BPT is also a probe of changing physical conditions in star-forming regions with redshift (e.g., Kewley et al., 2013, Figure 2). This has been an active area of research in numerous high-redshift statistical studies (e.g., Steidel et al., 2014; Shapley et al., 2015, S14, S15, respectively) which have shown that high-z SFGs cluster around a locus offset toward higher [O III] λ5007/Hβ and/or [N II] λ6583/Hα when compared to the star-forming locus of $z \sim 0$ SDSS galaxies. We show this in Figure 3.2, where the SFG locus of $z \sim 0$ SDSS galaxies is given by the solid red line (Kewley et al., 2013), and the offset SFG loci of the $z \sim 2.3$ KBSS-MOSFIRE and MOSDEF surveys are displayed by the purple (Steidel et al., 2014) and magenta (Shapley et al., 2015) lines, respectively. Recent work by Runco et al. (2022) has shown that these high-z SFG loci actually converge with consistent emission-line-fitting applied to each sample and reaffirm the existence of an offset in the BPT between local and high-z SFGs.

We show, via the black star in Figure 3.2, that our stack of $z \sim 2.3$ star-forming galaxies is also offset from the SDSS SFG locus, lying in parameter space consistent with KBSS-MOSFIRE and MOSDEF. While we cannot place each individual galaxy in our sample on this plot due to skyline contamination, particularly of Hβ, we note that our stack lies

below the "maximum starburst" demarcation (dotted black line) of Kewley et al. (2001), the empirical demarcation (dashed brown line) between SFGs and AGN of Kauffmann et al. (2003), and the theoretical, $z = 2.30$, upper-limit SFG locus (red dotted-dashed line) of Kewley et al. (2013). We also note that the large uncertainty of our stack in log([N II] / Hα) is the result of [N II] λ6583 only being detected in the composite spectrum with 2σ significance.

While the [N II] λ6583 measurement in our composite is fairly uncertain, the location of our stack along the x-axis of the N2-BPT is interesting when compared to the M_*-binned stacks of SFGs (the blue-gray squares) from the MOSDEF survey and Sanders et al. (2021, S21). This is because of the monotonic relationship that exists between log([N II] λ6583/Hα) and metallicity; as this strong-line ratio increases, metallicity increases (Pettini & Pagel, 2004; Maiolino et al., 2008; Curti et al., 2017; Bian et al., 2018; Sanders et al., 2021). By the mass-metallicity relation (e.g., Tremonti et al., 2004; Sanders et al., 2021), as this ratio increases, the stellar mass of galaxies should then also increase on average. We see this with the MOSDEF stacks, where in the upper right-hand section of the plot, we list the median stellar masses of the stacks in order of increasing [N II]/Hα. Here in this list we have also included, in bold, the median stellar mass of our dwarf galaxy stack, which, based on its positioning in the BPT, predictably has the lowest listed median stellar mass. In this, we show that our sample is a complementary extension in stellar mass to the MOSDEF (and KBSS-MOSFIRE) survey, extending its mass range into the dwarf galaxy regime ($M_* < 10^9$ M$_\odot$).

3.4.1.2 The Star-Forming Main Sequence

One of the primary goals of this study is to analyze our dwarf galaxy composite relative to the $z \sim 2.3$ mass-metallicity relation (MZR) of star-forming galaxies (see Section 3.4.3), which is a scaling relation between galaxy stellar mass (M_*) and gas-phase oxygen abundance, or metallicity (O/H). However, in order to properly contextualize our findings in relation to the broader $z \sim 2.3$ dwarf galaxy population, we must consider the SFRs associated with the stack and the sample that comprises it. This is due to the existence of the fundamental metallicity relation (FMR) between M_*, SFR, and O/H which has been demonstrated locally at $z \sim 0$ (e.g., Mannucci et al., 2010, 2011; Lara-López et al., 2010; Andrews & Martini, 2013; Curti et al., 2020) and at high redshift out to $z \sim 3.3$ (e.g., Henry et al., 2013a,b; Cresci et al., 2019; Sanders et al., 2018, 2021). The FMR demonstrates that, at a fixed M_*, a galaxy with an above-average (below-average) SFR will typically have a below-average (above-average) metallicity. Therefore, if our stacking sample (and thus composite spectrum) is biased in SFR, it will not have an average metallicity representative of typical dwarf galaxies at $z \sim 2.3$. This would be problematic when comparing the metallicity of our composite to the low-mass end of the MZR.

To investigate whether our stacking sample has a bias in SFR, we plot our sample and its median values against the $M_* - \mathrm{SFR}$ star-forming main sequence (SFMS) in Figure 3.3. Galaxies that lie on this relation, which is redshift-dependent, are considered to be representative of the typical galaxy at that corresponding stellar mass and redshift. In Figure 3.3, we compare to the $z \sim 2.3$ SFMS parameterizations of Sanders et al. (2021) and Whitaker et al. (2014).

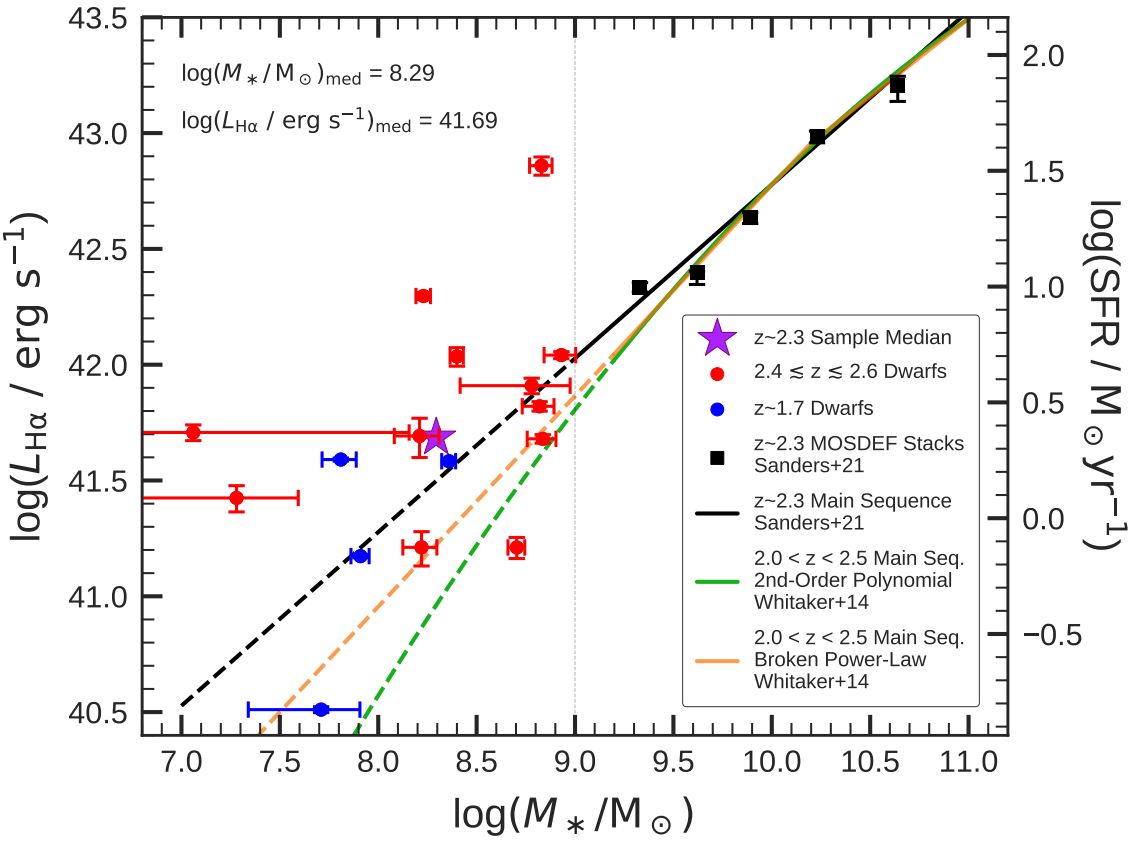

Figure 3.3: M_* vs. dust-corrected $L_{\rm H\alpha}$ on the left-hand axis and SFR on the right-hand axis. Our stacking sample is shown by the colored circles, with blue circles representing our four $z \sim 1.7$ galaxies and red circles our twelve $2.4 \lesssim z \lesssim 2.6$ galaxies. The purple star lies at the median M_* and $L_{\rm H\alpha}$ (or SFR) of our $z \sim 2.3$ sample. The black squares show the $z \sim 2.3$ M_*-binned stacks of MOSDEF galaxies from Sanders et al. (2021). We compare our sample median values against the M_* – SFR relations, or "star-forming main sequences" (SFMS), of Sanders et al. (2021, black line) and Whitaker et al. (2014, orange and green lines) to determine how representative our sample is of typical, $z \sim 2.3$ dwarf galaxies. The Sanders et al. (2021) relation is a power-law derived from Hα SFRs. The orange and green relations of Whitaker et al. (2014) are parameterized by a broken power-law and second-order polynomial, respectively; both relations derive from UV+FIR SFRs. Dashed portions of the SFMSs are extrapolations.

In comparing to the $z \sim 2.3$ SFMS of Sanders et al. (2021), we do so with stellar masses and SFRs calculated in a manner highly consistent with the methodologies adopted in Sanders et al. The stellar masses of our stacking sample and the MOSDEF galaxies used to calibrate the SFMS both rely on emission-line-corrected photometry and the SED-fitting (with `FAST`; Kriek et al., 2009) assumptions of constant star formation histories, the Calzetti et al. (2000) attenuation curve, and the Chabrier (2003) IMF. The SFRs in both studies are calculated from dust-corrected (via the Cardelli et al. 1989 extinction curve) Hα luminosities ($L_{\mathrm{H}\alpha}$). We note that the conversion factor between $L_{\mathrm{H}\alpha}$ and SFR (see, for example, Equation 3.1) is dependent on assumptions such as the stellar metallicity and can vary between different authors. Therefore, we plot dust-corrected $L_{\mathrm{H}\alpha}$ on the left-hand axis of Figure 3.3 so that our stacking sample (red and blue circles) and the Sanders et al. (2021) MOSDEF stacks (black squares) can be directly compared without the additional SFR conversion. Ultimately, however, we find that the SFR conversion factor used by both studies is very similar, and we continue our analysis of how representative our stacking sample is via SFR, given on the right-hand axis of Figure 3.3.

The $z \sim 2.3$ SFMS of Sanders et al. (2021) is parameterized as a power-law over the stellar mass range $9.0 < \log(M_*/\mathrm{M}_\odot) < 11.0$. In Figure 3.3, we plot their best-fit relation[13] over this range as a solid black line and extrapolate into the dwarf galaxy regime as seen by the dashed black line. In plotting our stacking sample in the $M_* - \mathrm{SFR}$ parameter space, we differentiate the $z \sim 1.7$ galaxies from the $2.4 \lesssim z \lesssim 2.6$ galaxies by blue and red points, respectively. We note that three of the four $z \sim 1.7$ galaxies lie along or above the

[13]The fitting of the SFMS was done with the four lowest-mass bins. Additionally, there is evidence that the lowest-mass ($M_* < 10^{9.5}$ $\mathrm{M}_\odot$) bin may be biased high in SFR (Shivaei et al., 2015; Sanders et al., 2021).

extrapolation of the $z \sim 2.3$ SFMS. These objects are biased high in SFR relative to typical $z \sim 1.7$ galaxies that have a lower SFR at fixed M_* due to the redshift evolution of the SFMS (Speagle et al., 2014; Whitaker et al., 2014; Sanders et al., 2021). The other group of $z > 2.38$ galaxies are found to scatter around the Sanders et al. extrapolation. Considered together, at the median mass of the complete $\langle z \rangle = 2.3$ stacking sample, $\log(M_*/\mathrm{M}_\odot)_{\mathrm{med}} = 8.29$, the median SFR of the sample, $\log(\mathrm{SFR}/\mathrm{M}_\odot\,\mathrm{yr}^{-1})_{\mathrm{med}} = 0.35$, lies $\Delta\log(\mathrm{SFR}) \sim 0.19$ dex above the extrapolation of the $z \sim 2.3$ SFMS of Sanders et al. (2021). This median point is shown as the purple star.

With the relative offset in SFR of our stacking sample in hand, we estimate the bias in O/H, resulting from the FMR, of our composite spectrum. Considering the strength of the SFR dependence of direct-method O/H at fixed M_* from Sanders et al. (2020),

$$\Delta\log(\mathrm{O/H}) \approx -0.29 \times \Delta\log(\mathrm{SFR}\,/\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}) \tag{3.4}$$

our sample stack is biased by $\Delta\log(\mathrm{O/H}) \sim -0.06$ dex, a value half the statistical uncertainty of our composite direct-method metallicity estimate ($\sigma_{\mathrm{stat}} \approx 0.12$ dex). We therefore conclude that, when comparing to the SFMS of Sanders et al. (2021), our stacking sample of dwarf galaxies does not have a major bias in SFR or O/H and, on average, is representative of typical dwarf galaxies at $z \sim 2.3$ with $M_* \gtrsim 10^8$ $\mathrm{M}_\odot$.

In Figure 3.3, we also plot the $2.0 < z < 2.5$ SFMS parameterizations of Whitaker et al. (2014), which were fit to M_*-binned stacks above a mass-completeness limit of $10^{9.2}$ $\mathrm{M}_\odot$. Whereas Sanders et al. (2021) fit the SFMS with a power-law, Whitaker et al. (2014) fit the SFMS with both a second-order polynomial (green line) and a broken power-law (or-

ange line) for which a separate slope was fit above and below a characteristic mass of $\log(M_*/\mathrm{M}_\odot) = 10.2$. We note that in our recreation of the second-order polynomial fit, we use the more precise polynomial coefficients given in the erratum (Whitaker et al., 2020) to Whitaker et al. (2014, Table 1), in order to more accurately portray the curve. Similar to Sanders et al. (2021) and the MOSDEF survey, the sample of Whitaker et al. (2014) is composed of star-forming galaxies from the CANDELS fields (Grogin et al., 2011; Koekemoer et al., 2011), though has a larger galaxy count and different sample selection. Like with the MOSDEF galaxies and our stacking sample, the stellar masses are determined with `FAST` assuming a Calzetti et al. (2000) attenuation curve and Chabrier (2003) IMF; however, the star-formation histories are taken to be exponentially declining. Unlike in Sanders et al. (2021) and our stacking sample though, the SFRs are estimated from the combination of rest-frame ultraviolet (UV) light and light re-radiated by dust in the far-infrared (FIR).

Figure 3.3 shows that, at least qualitatively, the parameterizations from Whitaker et al. (2014) generally agree with the power-law fit ($\beta = 0.75$) and MOSDEF stacks of Sanders et al. (2021), though begin to diverge as a result of steeper slopes near unity below $\log(M_*/\mathrm{M}_\odot) \lesssim 10$. This divergence, in part due to the choices of parametric form made by each author, is particularly pronounced in the dwarf galaxy mass regime where our sample lies. Unfortunately, there does not currently exist a high-redshift statistical sample in this regime that would allow us to confidently use a given parameterization. Therefore, we determine the most applicable SFMS based on consistency with our study in physical property estimation as well as the selection approach of our stacking sample. As detailed above, our methodologies for estimating M_* and SFR are most analogous to those of Sanders

et al. (2021), mitigating systematic uncertainties between estimations of physical properties calculated with different techniques. Additionally, when selecting our stacking sample, we did not require a detection of [O III] λ5007, Hα, or [O III] λ4363 which typically bias a sample toward higher SFR at fixed M_*. Our sample selection instead suggests a more representative sample like we see when comparing to the SFMS of Sanders et al. (2021). We therefore conclude that the SFMS of Sanders et al. (2021) is the most applicable comparison and that our stacking sample is representative of typical, star-forming, dwarf galaxies at $z \sim 2.3$.

3.4.2 Strong-Line Metallicity Calibrations at High-z

A major outstanding issue and active area of research in high-z astronomy is how to accurately calculate the gas-phase metallicities of the star-forming galaxies in the various large, statistical, spectroscopic surveys at $z > 1$ and $M_* \gtrsim 10^9\,\mathrm{M}_\odot$ (e.g., 3D-*HST*; Brammer et al. 2012a, KBSS-MOSFIRE; Steidel et al. 2014, MOSDEF; Kriek et al. 2015, FMOS-COSMOS; Kashino et al. 2019). The cause of this problem is two-fold. For one, auroral lines such as O III] $\lambda\lambda$1661, 1666 or [O III] λ4363, needed for direct, T_e-based metallicity estimation, are exceedingly faint, especially at high-redshift and with increasing galaxy stellar mass. Additionally, it is not fully understood how changing physical conditions with redshift in star-forming regions affect locally-calibrated, strong-line ratio metallicity diagnostics for indirect metallicity estimation. In other words, the accuracy and applicability of these strong-line metallicity calibrations at high-redshift is an open question which several

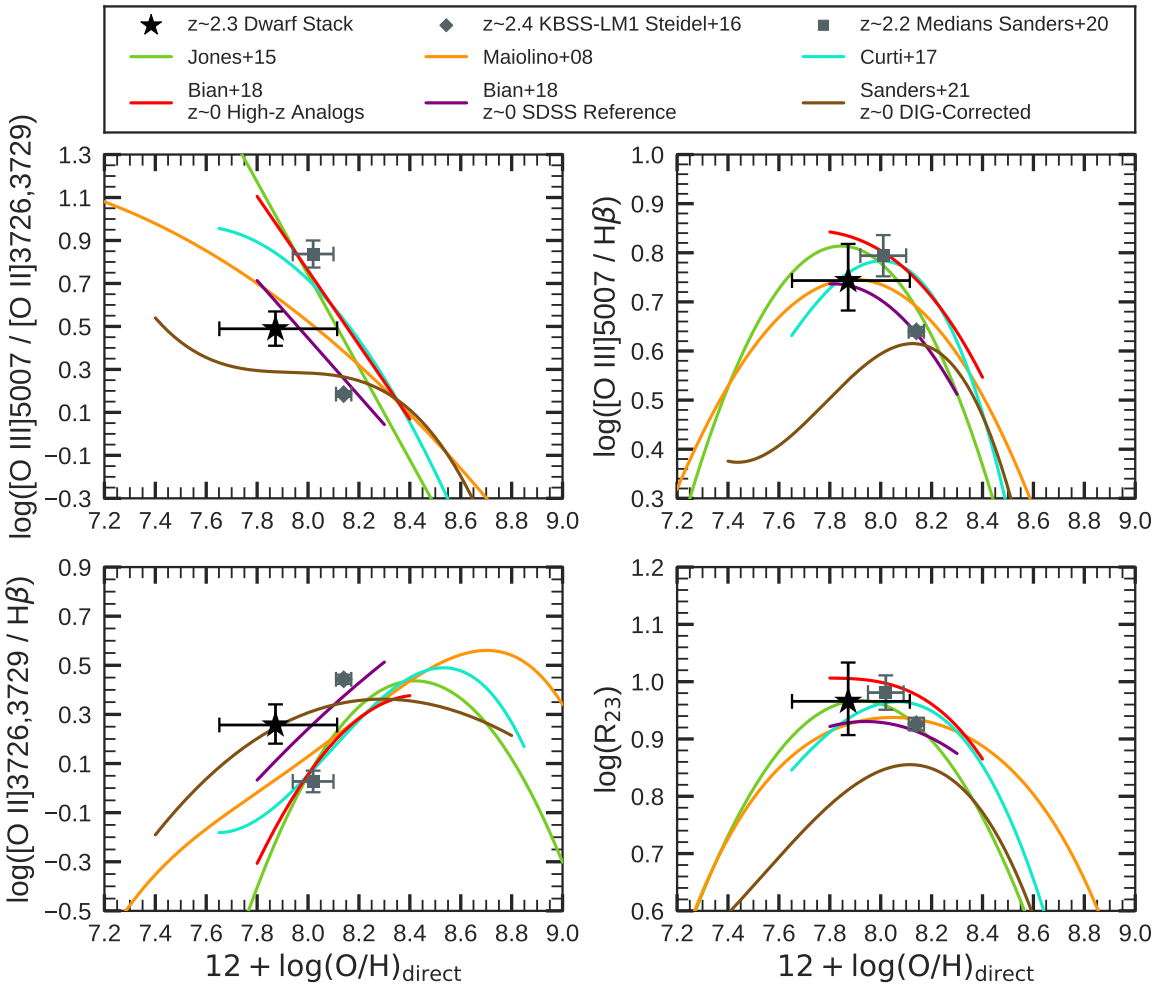

Figure 3.4: Strong emission-line ratios as a function of direct-method metallicity. From left-to-right and top-to-bottom we consider the oxygen-based strong-line ratios of O_{32}, O_3, O_2 and R_{23}. Our $z \sim 2.3$ composite is displayed via the black star in each panel. For reference, we also include the $z \sim 2.4$ KBSS-LM1 composite of 30 star-forming galaxies also selected independent of emission-line strength (gray diamond; Steidel et al., 2016), as well as the median values of the $z \sim 2.2$ auroral-line sample of Sanders et al. (2020, gray square). We compare these points to the locally-calibrated, strong-line metallicity relations of Maiolino et al. (2008, orange line), Jones et al. (2015, green line), Curti et al. (2017, cyan line), Bian et al. (2018, both the local reference – purple line – and high-z analog – red line – relations), and Sanders et al. (2021, DIG-corrected; brown line). We note the metallicity-insensitivity of the O_3 and R_{23} indices at the oxygen abundances considered here. We find that our composite and KBSS-LM1 have their metallicities best reproduced by the local reference relations of Bian et al., in contrast to the auroral-line sample of Sanders et al. that favors the high-z analog relations.

studies have tried to address (Jones et al., 2015; Sanders et al., 2016a; Patrício et al., 2018; Gburek et al., 2019; Sanders et al., 2020).

In Figure 3.4, we revisit this issue with our $z \sim 2.3$ dwarf galaxy composite. Here we plot our stack (the black stars) in the parameter space of T_e-based oxygen abundance versus various commonly-used, oxygen-based, strong emission-line ratios. The dust-corrected emission-line ratios and direct metallicity ($12 + \log(\mathrm{O/H}) = 7.87^{+0.24}_{-0.22}$) of our stack are measured (see Section 3.3) from the composite spectrum in Figure 3.1. Our stack in these plots is compared to several locally-calibrated strong-line metallicity diagnostics from the literature (Maiolino et al., 2008; Jones et al., 2015; Curti et al., 2017; Bian et al., 2018; Sanders et al., 2021) in an attempt to shed light on which strong-line ratios are serviceable at the typical metallicity of our stacking sample as well as which calibrations most favorably reproduce our composite metallicity at fixed strong-line ratio. While we cannot comment on the shape or slope of the various calibrations, we can get a sense of the appropriate normalization of these relations when considering high-z dwarf galaxies. In Figure 3.4, we consider four strong-line ratios: $\mathrm{O_{32}} = \log([\mathrm{O\ III}]\ \lambda5007/[\mathrm{O\ II}]\ \lambda\lambda3726, 3729)$, $\mathrm{O_3} = \log([\mathrm{O\ III}]\ \lambda5007/\mathrm{H}\beta)$, $\mathrm{O_2} = \log([\mathrm{O\ II}]\ \lambda\lambda3726, 3729/\mathrm{H}\beta)$, and $\mathrm{R_{23}} = \log(([\mathrm{O\ III}]\ \lambda\lambda4959, 5007 + [\mathrm{O\ II}]\ \lambda\lambda3726, 3729)/\mathrm{H}\beta)$. We do not consider strong-line metallicity calibrations based on [N II] in this work due to our low detection significance ($\sim 2\sigma$) of [N II] $\lambda6583$ in the composite spectrum as well as concerns in the literature (e.g., Masters et al., 2014, 2016) of elevated N/O abundance ratios at high-redshift.

Of immediate note when considering the location of our stack relative to the $\mathrm{O_3}$- and $\mathrm{R_{23}}$-based calibrations in the right-hand panels of Figure 3.4 is that our stack lies

at or near the apex of these relations in what is called the “turnover" between the high and low metallicity branches of these calibrations. In these turnover regimes, the strong-line ratio is insensitive to the metallicity of a galaxy, giving these relations little value as useful metallicity indicators over the oxygen abundance range spanned by the turnover region (somewhere roughly between $7.7 \lesssim 12 + \log(\mathrm{O/H}) \lesssim 8.3$ depending on the strong-line index $-$ $\mathrm{O_3}$ or $\mathrm{R_{23}}-$ and calibration used). We therefore do not recommend the use of these strong-line indices for $z \sim 2$ dwarf galaxies similar to our those in our stacking sample. These results and conclusion are not particularly surprising as [O III] λ4363-emitter studies (Sanders et al., 2016a; Gburek et al., 2019; Sanders et al., 2020) and photoionization modeling (e.g., Steidel et al., 2014) have shown that it is quite common for $z \sim 2$ star-forming galaxies, over a couple orders of magnitude in stellar mass, to have metallicities that lie within these insensitive turnover regions.

A more interesting result is revealed when looking at the dwarf galaxy stack relative to the $\mathrm{O_{32}}$- and $\mathrm{O_2}$-based calibrations in the left-hand panels of Figure 3.4. In particular, we focus on the comparisons with the strong-line metallicity relations of Bian et al. (2018), who used stacked SDSS spectra to create T_e-based, empirical metallicity calibrations over the metallicity range $7.8 < 12 + \log(\mathrm{O/H}) < 8.4$. These calibrations were created from two distinct SDSS samples, a reference sample of galaxies lying within $\pm$0.05 dex of the $z \sim 0$ star-forming sequence of the N2-BPT diagram (parameterized by Kewley et al. 2013) and a high-redshift analog sample of SDSS galaxies lying within $\pm$0.04 dex of the offset $z \sim 2.3$ star-forming sequence of the BPT diagram (Steidel et al., 2014). With these selection criteria based on nebular emission-line ratios, the calibrations of Bian et al. (2018) should represent

the conditions of star-forming regions in low- and high-redshift galaxies, respectively. We note that the Bian et al. calibrations presented in Figure 3.4 were re-fit for this study for the reasons, and via the methods, described in Section 3.6.

When comparing our stack (black stars) to the strong-line metallicity relations of Bian et al. (2018), we find that our stack favors the local reference calibrations (purple curves), such that when considering the O_{32} and O_2 indices, the local reference relations reproduce our composite metallicity to within $\lesssim 0.15$ dex at fixed strong-line ratio. The high-z analog relations of Bian et al. (red curves) and the calibrations of Maiolino et al. (2008, orange curves), Jones et al. (2015, green curves), and Curti et al. (2017, cyan curves) are all $\gtrsim 1\sigma$ inconsistent with our stack in the O_{32} and O_2 panels, the lone exception being the O_{32} relation of Maiolino et al. This result of inconsistency with these relations has added significance in that we are analyzing a composite of 16 galaxies rather than an individual source. Also, our uncertainties here take into account sample variance in addition to statistical error (see Section 3.3.3). We note that, in all panels other than that of the O_2 index, our composite is also highly inconsistent with the calibrations of Sanders et al. (2021, brown curves). At metallicities below $12 + \log(O/H) < 8.4$, the Sanders et al. (2021) relations are calibrated with the H II region spectra of dwarf galaxies from Berg et al. (2012) and the *Spitzer* Local Volume Legacy survey (Dale et al., 2009). These dwarf galaxies have been shown in previous studies to not follow the other strong-line metallicity relations considered in this work, possibly as a result of biases from the selection methods of the various calibration samples or an incompleteness in low-metallicity, high-excitation H II regions (Gburek et al., 2019; Sanders et al., 2020, 2021).

It is interesting that our $z \sim 2.3$ stack generally best agrees with the local reference calibrations of Bian et al. (2018), particularly when compared to the findings of Sanders et al. (2020), who compiled O III] $\lambda\lambda$1661, 1666 and [O III] λ4363-detected sources at $z > 1$ from the literature and the MOSDEF survey and conducted a similar study of strong-line metallicity diagnostics at the median redshift of the compiled sample, $z_{\rm med} \sim 2.2$. These authors found that at the median metallicities and median line ratios of the galaxies in their $z > 1$ sample (gray squares in Figure 3.4; each median limited to galaxies with detections of the respective line ratio), the high-redshift analog calibrations of Bian et al. (2018) best reproduced their metallicities at fixed line-ratio. Moreover, Sanders et al. (2020) found general agreement between their median points and the Curti et al. (2017) calibrations as well as general agreement with the calibration sample ($z \sim 0$ SDSS galaxies from Izotov et al. 2006) of the Jones et al. (2015) relations. Consequently, when considering the results of our work and those of Sanders et al. (2020), both studies at similar redshift and metallicity, a tension exists in regard to the evolution with redshift of strong-line metallicity diagnostics and therefore which calibrations are reliable at high-z.

A likely source of the discrepancy and tension seen between the results of this work and that of Sanders et al. (2020) lies in how each galaxy sample was selected. For our dwarf galaxy stacking sample, we did not select galaxies based on the strength of any particular rest-optical emission-line (see Section 3.2.3). This is important particularly when considering [O III] λ4363, [O III] λ5007 ([O III] λ4959 in our case), and Hα. By avoiding selection based on line-strength, we mitigate biases in our sample such as high sSFRs (SFR / M_*) and high excitation, resulting in a sample which very nearly falls on an extrapolation of the mean

$z \sim 2.3$ M_* – SFR relation (see Section 3.4.1.2 and Figure 3.3). In contrast, the $z \sim 2.2$ sample of Sanders et al. (2020) comprises galaxies selected on auroral-line detection (either O III] $\lambda\lambda$1661, 1666 or [O III] λ4363, detected at $(\mathrm{S/N})_{\mathrm{med}} = 6.0^{+1.5}_{-2.5}$) for T_e-based metallicity estimation. Due to the faintness of these auroral lines – [O III] λ5007 is $\sim 30-100\times$ brighter than [O III] λ4363 (Jones et al., 2015) – this detection requirement preferentially selects younger galaxies with high-excitation, highly-ionized star-forming regions. These galaxies lie well above the M_* – SFR relation. In fact, the sample of Sanders et al. (2020) lies an average ~ 0.6 dex above the best-fit $z \sim 2.3$ M_* – SFR relation of Sanders et al. (2018) (which is very similar to the fit of the same relation in Sanders et al. 2021) and has O_{32} values ~ 0.5 dex higher on average than typical $z \sim 2.3$ star-forming MOSDEF galaxies at fixed M_* (Sanders et al., 2020, Figure 16). Sanders et al. shows that this auroral-line sample is not representative of typical $z \sim 2.3$ star-forming galaxies, but rather coincident with $z \sim 2$ extreme emission-line galaxies (EELGs). In effect, at roughly fixed O/H in Figure 3.4, we are comparing two galaxy samples that differ significantly in the ionization state and extremity of their star-forming regions. As such, both samples are not well-represented by one single strong-line metallicity calibration.

When considering the sample selection methodologies for our stacked sample and the $z > 1$ auroral-line sample of Sanders et al. (2020) against those for the calibration samples used to parameterize the relations shown in Figure 3.4, it is perhaps not surprising that several relations appear to agree with the Sanders et al. auroral-line sample while being $> 1\sigma$ inconsistent with our stacked sample. In the low-metallicity regime ($12 + \log(\mathrm{O/H}) \lesssim 8.4$), the locally-calibrated strong-line metallicity relations are generally defined

by *individual* galaxies with [O III] λ4363 detections. This is the case for the relations of Maiolino et al. (2008), Jones et al. (2015), and Curti et al. (2017). This requirement of an [O III] λ4363 detection in individual galaxies, as well as the BPT-related requirement of the Bian et al. (2018) high-z analog calibration (detailed above), selectively probe galaxies with ISM conditions more extreme than in typical $z \sim 0$ star-forming galaxies. Rather, ISM conditions more akin to those in the $z > 1$ auroral-line sample are probed, leading to low-metallicity strong-line calibrations that are likely biased high in O_3 and O_{32} (and low in O_2) and that closely predict the Sanders et al. (2020) $z > 1$ sample metallicity at fixed strong-line ratio. Meanwhile, the selection criterion for the Bian et al. (2018) local reference calibrations (detailed above) selectively probes galaxies with less extreme star-forming conditions and better predicts the metallicity at fixed strong-line ratio of our dwarf galaxy stack, which is not reliant on emission-line detections.[14]

3.4.2.1 The KBSS-LM1 Composite of Steidel et al. (2016)

In this subsection, we briefly discuss the $z \sim 2.4$ composite spectrum of Steidel et al. (2016) in the context of the strong-line ratio – direct metallicity relations displayed in Figure 3.4. This composite, referred to in Steidel et al. (2016) as "KBSS-LM1," is derived from the rest-frame far-UV and optical spectra of 30 star-forming galaxies from the KBSS-MOSFIRE spectroscopic survey. The galaxies comprising KBSS-LM1 were notably selected to have emission-line measurements or limits (not detections) of, among other lines, [O II] $\lambda\lambda$3726, 3729, Hβ, [O III] $\lambda\lambda$4959, 5007, and Hα. The sample of galaxies was also selected

[14] A more detailed study of the biases in certain locally-calibrated strong-line metallicity relations (particularly of Curti et al. 2017), and how those biases factor into the observed evolution of strong-line metallicity relations with redshift, can be found in Sanders et al. (2020).

such that it broadly represents the full KBSS-MOSFIRE sample in SFR, M_*, and O/H, the latter of which was calculated via the T_e-sensitive UV emission-line doublet, O III] $\lambda\lambda$1661, 1666. The median M_* and derived T_e-based metallicity of KBSS-LM1 are, as reported by Sanders et al. (2020), $\log(M_*/\mathrm{M}_\odot) = 9.8 \pm 0.3$ and $12+\log(\mathrm{O/H}) = 8.14 \pm 0.03$, respectively. With these selection criteria, KBSS-LM1 is therefore similar to our $z \sim 2.3$ dwarf galaxy composite in the sense that both samples should be representative of typical $z \sim 2.3$ galaxies at their respective stellar masses. Indeed, Sanders et al. (2020, Figure 16) find KBSS-LM1 to lie just above the $z \sim 2.3$ star-forming main sequence as well as amongst the typical $z \sim 2.3$ MOSDEF galaxies in O_{32} at fixed M_*.

In Figure 3.4, along with our $z \sim 2.3$ dwarf galaxy stack (black stars) and $z \sim 2.2$ auroral-line sample of Sanders et al. (2020, gray squares), we plot KBSS-LM1 as a gray diamond in each panel. As with our dwarf galaxy stack, albeit at higher O/H owing to its higher M_*, we see that KBSS-LM1 is best represented by the local reference calibrations of Bian et al. (2018) instead of the high-redshift analog calibrations, though note the small statistical error bars (not bootstrapped) of KBSS-LM1. This result suggests that $z \sim 2.3$ star-forming galaxies with $8.4 \lesssim \log(M_*/\mathrm{M}_\odot) \lesssim 9.8$ that lie on the M_*-SFR relation will, on average, have their metallicities most accurately predicted at fixed strong-line ratio via the local reference strong-line metallicity calibrations of Bian et al. (2018). In contrast, $z \sim 2.3$ EELGs and galaxies with more extreme ISM conditions, like the $z \sim 2.2$ auroral-line sample, may require independently-calibrated strong-line metallicity relations more akin to the high-z analog calibrations of Bian et al. (2018).

3.4.3 The Stellar Mass – Gas-Phase Metallicity Relation

This section explores our $z \sim 2.3$ dwarf galaxy sample in relation to the low-mass end of the stellar mass – gas-phase metallicity relation (MZR). The MZR, shown to exist both locally and at high redshift, displays a positive correlation, parameterized as a power-law, between galaxy stellar mass (M_*) and gas-phase oxygen abundance (O/H) at lower stellar masses before flattening asymptotically at higher masses ($M_* \gtrsim 10^{10.5}$ $M_\odot$ locally; Tremonti et al., 2004; Erb et al., 2006; Maiolino et al., 2008; Andrews & Martini, 2013; Steidel et al., 2014; Curti et al., 2020; Sanders et al., 2020, 2021; Strom et al., 2022). The MZR also evolves with time, such that at higher redshifts, the average metallicity of star-forming galaxies is lower at fixed M_* than it is locally (Erb et al., 2006; Maiolino et al., 2008; Zahid et al., 2013, 2014a,b; Steidel et al., 2014; Sanders et al., 2020, 2021). Here we compare our dwarf galaxy composite against recent empirical and theoretical parameterizations of the MZR in an effort to better constrain the low-mass slope and normalization of the relation at $z \sim 2.3$.

3.4.3.1 Comparison to MZRs Derived from Observations

In the left-hand panel of Figure 3.5, we plot our $z \sim 2.3$ dwarf galaxy composite against the $z \sim 2.2$ direct-method MZR of Sanders et al. (2020), the $z \sim 2.3$ strong-line MZR of Sanders et al. (2021), and the $z \sim 2.3$ MZR of Strom et al. (2022), who calculated their metallicities via photoionization modelling. The direct-method MZR shown here (black line) is derived from the Sanders et al. (2020) $z > 1$ ($z_{\rm med} \sim 2.2$) auroral-line sample considered in Section 3.4.2 above. However, as previously mentioned, this sample of

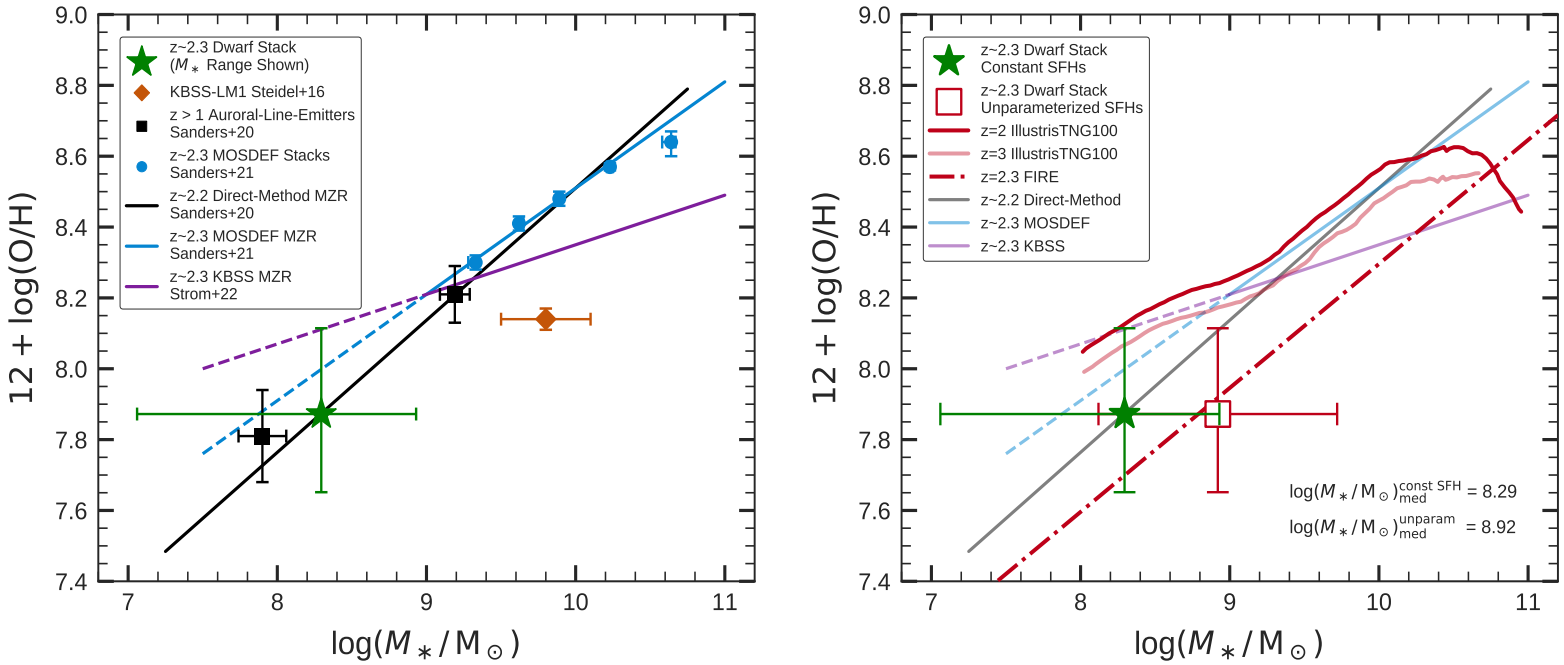

Figure 3.5: The stellar mass – gas-phase metallicity relation (MZR). *Left*: We compare our $z \sim 2.3$ dwarf galaxy stack (green star) at our sample's median mass, $\log(M_*/\mathrm{M}_\odot)_{\mathrm{med}} = 8.29$, against the $z \sim 2.2$ direct-method MZR of Sanders et al. (2020, black line) and the extrapolations of the $z \sim 2.3$ strong-line MZR of Sanders et al. (2021, blue solid/dashed line) and the $z \sim 2.3$ photoionization-model-based MZR of Strom et al. (2022, purple solid/dashed line). We find excellent agreement between our stack and the direct-method MZR, the most comparable of the relations. For reference, we also include the mean values of the $z > 1$ auroral-line sample split into two M_* bins (black squares; Sanders et al., 2020), the MOSDEF galaxy stacks from Sanders et al. (2021, blue squares), and the $z \sim 2.4$ KBSS-LM1 composite of Steidel et al. (2016, orange diamond). *Right*: We compare our $z \sim 2.3$ stack (green star; same as in left panel) to the predicted $z = 2$ and $z = 3$ MZRs (dark-red and light-red lines, respectively) of the IllustrisTNG100 simulations (Torrey et al., 2019) and the $z = 2.3$ MZR of the FIRE simulations (red dot-dashed line; Ma et al., 2016), finding consistency, within errors, with both sets of MZRs. When comparing our stack after recalculating the sample stellar masses assuming more realistic non-parametric SFHs (red unfilled square; $\log(M_*/\mathrm{M}_\odot)_{\mathrm{med}} = 8.92$), we find excellent agreement with the MZR from FIRE. The MZRs from the left-hand panel are recreated here for reference. We note that for each display of our stack, the x-axis error bar corresponds to the M_* range of our sample.

galaxies lies $\langle \Delta \log(\mathrm{SFR}) \rangle \sim 0.6$ dex above the $z \sim 2.3$ $M_* - \mathrm{SFR}$ relation, which has the effect, via the FMR, of yielding a MZR biased low in O/H relative to the MZR expected for typical galaxies that fall on the star-forming main sequence at this redshift. As such, the direct-method MZR displayed in Figure 3.5 was parameterized by Sanders et al. (2020) *after* applying SFR-corrections to the $z > 1$ sample metallicities via Equation 3.4. Additionally, Sanders et al. (2020) adjusted this MZR to account for a low-redshift bias in the low-mass $z > 1$ galaxies. These galaxies (low-mass black square; 5 galaxies) were found to have a median redshift $\Delta z_{\mathrm{med}} \sim 0.5$ lower than the high-mass sample galaxies (high-mass black square; 9 galaxies), and were estimated, on average, to have a metallicity biased 0.1 dex high due to the redshift evolution of O/H at fixed M_* ($d\log(\mathrm{O/H})/dz \approx -0.2$; Sanders et al., 2020). With these adjustments, the direct-method MZR reproduced in Figure 3.5 is an estimation of the MZR – on the T_e-based abundance scale – of typical galaxies lying on the star-forming main sequence at a redshift of $z \sim 2.2$.

The blue line (and its extrapolation) in the left-hand panel of Figure 3.5 is a recreation of the $z \sim 2.3$ strong-line MZR of Sanders et al. (2021), fit as a power-law to the four lowest-mass bins of MOSDEF galaxies displayed here as blue squares (The highest-mass bin suffers from incompleteness and was excluded from the fit (Sanders et al., 2021, Section 2.4)). These MOSDEF stacks were also used by Sanders et al. to parameterize their $M_* - \mathrm{SFR}$ relation (see Section 3.4.1.2 and Figure 3.3). The metallicities of the MOSDEF stacks were calculated via the high-z analog calibrations of Bian et al. (2018) and the α-element-based strong-line ratios of O_{32}, O_3, and $\log([\mathrm{Ne\ III}]\ \lambda 3869/[\mathrm{O\ II}]\ \lambda\lambda 3726, 3729)$. These relations were chosen so as to use calibrations that most closely reproduce the

excitation sequences, and thus likely ISM conditions, of the MOSDEF $z \sim 2.3$ star-forming sample.

The $z \sim 2.3$ MZR of Strom et al. (2022), and its extrapolation, are shown as the purple line in the left-hand panel of Figure 3.5. This relation was fit to 195 individual star-forming galaxies from the KBSS survey. The metallicities in this study were estimated with photoionization models, described in Strom et al. (2018, 2022), that are able to reproduce the rest-UV and rest-optical spectroscopic properties of high-z, star-forming galaxies.

When comparing our $z \sim 2.3$ dwarf galaxy stack (green star; x-axis error bar represents the M_* range of the stacking sample) against the direct-method and strong-line MZRs of Sanders et al. (2020) and Sanders et al. (2021), respectively, we find that at the median stellar mass of our stacking sample, $\log(M_*/\mathrm{M}_\odot)_{\mathrm{med}} = 8.29$, and the direct-method metallicity calculated from the composite spectrum, $12 + \log(\mathrm{O/H}) = 7.87^{+0.24}_{-0.22}$, the stack lies virtually on top of the $z \sim 2.2$ direct-method MZR. Within uncertainties, our stack (and the mean values of the low- and high-mass bins of the $z > 1$ auroral-line sample; Sanders et al., 2020, black squares) is also consistent with the extrapolation of the strong-line MZR, lying below this extrapolation by ~ 0.13 dex. Correcting the metallicity of our composite for the slightly-high SFR bias of the stacking sample (increasing $\log(\mathrm{O/H})$ by ~ 0.06 dex; see Section 3.4.1.2) reduces the offset of the stack from the extrapolation of the strong-line MZR while moving the stack ~ 0.06 dex above the direct-method MZR. This consistency with both MZRs supports their fit power-law parameters, in particular the slope, which was fit by Sanders et al. (2020) to be $\beta = 0.37$ and by Sanders et al. (2021) to be $\beta = 0.30$. However, while we are comparing to the strong-line MZR of Sanders et al. (2021) due to

the authors' careful consideration in selecting applicable strong-line calibrations for high-z star-forming regions, we note that systematic uncertainties still exist between metallicities calculated directly versus with strong-line proxies. Therefore, in constraining the slope of the $z \sim 2.3$ MZR, our results and direct-method of metallicity estimation most favorably suggest the slope fit to the direct-method MZR of Sanders et al. (2020), $\beta = 0.37$.

When considering the $z \sim 2.3$ photoionization-model-based MZR of Strom et al. (2022), we find that our stack lies $\approx 1\sigma$ below the extrapolation of this relation, in disagreement with the shallow slope ($\beta = 0.14$) proposed by Strom et al. However, we must again take into account the difference in methods of abundance estimation between the two studies. As discussed in the literature (e.g., Kewley & Ellison, 2008; Maiolino & Mannucci, 2019), photoionization models typically overestimate metallicities by ~ 0.2 dex or more compared to direct metallicities. In general, this is due to a poorly-constrained combination of factors, such as photoionization models accounting for dust depletion or T_e-based metallicities potentially being biased low due to temperature fluctuations or gradients in star-forming regions leading to nebular spectra dominated by brighter auroral lines from high-T_e zones. While the discrepancy between direct and theoretical metallicity estimates is stronger at higher metallicities (e.g., Kewley & Ellison, 2008; Maiolino & Mannucci, 2019; Curti et al., 2020), Steidel et al. (2016) found that photoionization models predict an oxygen abundance 0.25 dex higher than their T_e-based estimate of 12+log(O/H) = 8.14 (0.29 $Z_\odot$) for KBSS-LM1, a composite of 30 star-forming KBSS galaxies that we briefly discussed in Section 3.4.2.1. If we apply this same offset to the T_e-based metallicity of our dwarf galaxy composite, our stack will lie very near the extrapolation of the Strom et al. (2022) MZR.

This said, there are too many systematic uncertainties involved to accurately compare our composite against this MZR at present, and we conclude that the best estimate for the slope of the $z \sim 2.3$ MZR is $\beta = 0.37$, given by the direct-method MZR of Sanders et al. (2020).

We note that the stellar masses of the samples and studies considered here are calculated via SED-fitting with consistent assumptions in star-formation history, IMF, and extinction law. We also note that the KBSS-LM1 composite of Steidel et al. (2016) lies significantly below the displayed MZRs, including the direct-method MZR, when plotted at its reported direct metallicity of $12+\log(\mathrm{O/H}) = 8.14 \pm 0.03$. However, we do acknowledge the very small *statistical-only* uncertainty of this metallicity.

3.4.3.2 Comparison to MZRs from Cosmological Simulations

In the right-hand panel of Figure 3.5, we now compare our $z \sim 2.3$ dwarf galaxy stack against predicted MZRs from well-known cosmological simulations, the Feedback in Realistic Environments simulations[15] (FIRE; Hopkins et al., 2014) and The Next Generation Illustris simulations[16] (IllustrisTNG; Weinberger et al., 2017; Pillepich et al., 2018), which are the successor to the Illustris simulation (Vogelsberger et al., 2014a,b; Genel et al., 2014). Our stack, with the median mass ($\log(M_*/\mathrm{M}_\odot) = 8.29$) and direct metallicity ($12 + \log(\mathrm{O/H}) = 7.87^{+0.24}_{-0.22}$) as reported throughout this paper, is shown by the green star, with the x-axis error bar representing the M_* range of our stacking sample. The predicted MZR from FIRE, given by the redshift-dependent, fixed-slope ($\beta = 0.35$), gas-phase MZR fitting function in Ma et al. (2016), is evaluated at $z = 2.30$ (the mean redshift of our

[15]FIRE: https://fire.northwestern.edu/

[16]IllustrisTNG: https://www.tng-project.org/

stacking sample) and is represented by the red dot-dashed line. The $z = 2$ and $z = 3$ MZRs from IllustrisTNG (Torrey et al., 2019), specifically from the TNG100 simulation volume, are displayed by the solid dark-red and light-red lines, respectively. For reference to MZRs derived from observations, in the right-hand panel we reproduce, from the left-hand panel, the $z \sim 2.2$ direct-method MZR of Sanders et al. (2020, black line), the $z \sim 2.3$ strong-line MZR of Sanders et al. (2021, blue line), and the $z \sim 2.3$ photoionization-model-based MZR of Strom et al. (2022, purple line).

Properly assessing the predicted MZRs considered in this section requires a reliable and accurate empirical metallicity estimation method such as the direct-method, which estimates oxygen abundances through directly probing physical properties (T_e and n_e) of star-forming regions. We find that at fixed M_*, the direct metallicity of our stack (green star) is consistent within uncertainties with both the $z = 2.3$ FIRE MZR and $2 \leqslant z \leqslant 3$ MZR of IllustrisTNG. We also observe that the $z \sim 2.2$ direct-method MZR agrees favorably in slope and normalization with the IllustrisTNG MZR above $M_* \approx 10^9$ $M_\odot$. Below this mass, the IllustrisTNG MZR deviates away from this slope and the direct-method MZR toward higher metallicity values, displaying a “bump" in the low-mass regime. This bump is the result of the minimum wind velocity ($v_{\rm min} = 350$ km s^{-1}) enforced in the stellar feedback models of IllustrisTNG, with $v_{\rm min}$ put in place so that the simulations match the low-end of the galaxy stellar mass function (Pillepich et al., 2018). As Torrey et al. (2019) explain, while $v_{\rm min}$ is not directly a function of M_*, it is generally set as the wind velocity in galaxies with $M_* \lesssim 10^9$ $M_\odot$ due to the low dark matter velocity dispersions in their halos. The higher metallicities at these masses that we see as the bump in the

MZR then arise because these fixed-velocity winds eject less gas, and therefore fewer metals, than would be the case if the wind velocity were allowed to be $v < 350$ km s^{-1}. While further observations are needed to either confirm or deny this bump in the MZR, as well as evaluate the applicability of the minimum wind velocity assumption, the direct-method MZR of Sanders et al. (2020) suggests that such a bump likely does not exist and that wind velocities in low-mass galaxies can extend lower than $v < 350$ km s^{-1}. This suggestion is reinforced below when revisiting our dwarf galaxy stacking sample with stellar masses recalculated under more realistic assumptions.

In this work, in order to facilitate fair comparisons of our empirical results to those in the literature, when estimating the stellar masses of our stacking sample via SED-fitting, we made assumptions consistent with those generally found in the literature; in particular, we assumed constant star formation histories (SFH). However, the SFHs of galaxies, particularly of galaxies at high-z owing to their higher-EW emission lines, are likely not well-described by such simple parameterizations. Instead, stellar masses are likely more accurate if calculated assuming non-parametric SFHs which can better reveal the presence of older stellar populations that are hidden in the rest-UV and rest-optical by brighter, younger stars (Gburek et al., 2019; Tang et al., 2022; Whitler et al., 2022; Topping et al., 2022). Having these more realistic, typically larger stellar masses is important when comparing to simulation results. Therefore, we recalculated the stellar masses of our stacking sample, via the `PROSPECTOR`[17,18] SED-fitting code (Johnson et al., 2021), assuming non-parametric SFHs. The resultant median stellar mass of our sample under these assumptions

[17] https://prospect.readthedocs.io/en/latest/

[18] https://github.com/bd-j/prospector

becomes $\log(M_*/\mathrm{M}_\odot) = 8.92$, an increase to our fiducial median stellar mass of 0.63 dex. Our dwarf galaxy stack, shifted to this recalculated median M_*, is shown in Figure 3.5 as the red unfilled square. As with the x-axis error bar of our fiducial point (the green star), the x-axis error bar of the red square represents the recalculated range of M_* in our stacking sample.

When comparing our $z \sim 2.3$ dwarf galaxy stack to the simulated MZRs following recalculating the sample stellar masses under the assumption of more realistic non-parametric SFHs, we find excellent agreement between our composite and the $z = 2.3$ FIRE MZR of Ma et al. (2016), especially if we apply the metallicity correction ($\Delta \log(\mathrm{O/H}) \sim 0.06$ dex) to our stack to account for our bias high in SFR. In comparison to the $z = 2$ and $z = 3$ IllustrisTNG MZRs of Torrey et al. (2019), we find that our composite lies $\sim 1.6\sigma$ below the $z = 2$ MZR and $\sim 1.3\sigma$ below the $z = 3$ MZR, suggesting that a stronger O/H evolution at fixed M_* with redshift (like that seen in Figure 7 of Ma et al. 2016 relative to other simulations of the time – including the original Illustris simulation Torrey et al. 2014) and/or tuning of the $z = 0$ MZR normalization is needed. Like the direct-method MZR of Sanders et al. (2020), our composite also disagrees with the existence of a metallicity bump in the MZR at $M_* \lesssim 10^9$ $\mathrm{M}_\odot$, suggesting that the minimum wind velocity assumption in IllustrisTNG should be revisited.

3.4.4 $M_* - \mathrm{SFR} - \mathrm{O/H}$ Relation at $z \sim 2.3$

The $M_* - \mathrm{SFR} - \mathrm{O/H}$ relation, or fundamental metallicity relation (FMR), posits that the MZR has a secondary dependence on SFR such that, when considering all three properties, the scatter in metallicity at fixed M_* and SFR is reduced compared to the scatter

in metallicity at fixed M_* alone (Mannucci et al., 2010; Lara-López et al., 2010; Andrews & Martini, 2013; Sanders et al., 2018; Curti et al., 2020; Sanders et al., 2021). Also as a result of this secondary dependence, at fixed M_*, a higher (lower) than average SFR yields a lower (higher) than average O/H. Of further interest is that numerous studies have shown that the FMR is redshift-invariant to within ~ 0.1 dex in metallicity out to $z \sim 3.3$ (e.g., Mannucci et al., 2010; Henry et al., 2013a,b; Sanders et al., 2018, 2021). In effect, the observation of the evolution of the MZR over this redshift range, whereby O/H decreases with increasing redshift at fixed M_*, is actually the observation at different redshifts of different portions of the locally-defined FMR since SFR increases with redshift at fixed M_* (Whitaker et al., 2014; Speagle et al., 2014; Sanders et al., 2021). Unfortunately, however, this redshift-invariance of the FMR is still a matter of debate owing to the uncertainties regarding the applicability of locally-calibrated, strong-line metallicity calibrations at high-redshift (see Section 3.4.2), widely-used in lieu of hard-to-measure direct metallicities.

Fortunately, in this work we have a direct, T_e-based oxygen abundance from our dwarf galaxy composite spectrum with which we can evaluate the redshift evolution of the FMR. Additionally, given the properties of our sample, we are able to probe $M_* - \mathrm{SFR} - \mathrm{O/H}$ parameter space that to-date has been poorly-sampled at $z \sim 2.3$. We do so via the commonly-used 2D planar projection of the 3D FMR, established by Mannucci et al. (2010). In this projection, the metallicity is a function of the linear combination of M_* and SFR, denoted by μ_α and described by the equation

$$\mu_\alpha = \log(M_*/\mathrm{M}_\odot) - \alpha \log(\mathrm{SFR}/\mathrm{M}_\odot\,\mathrm{yr}^{-1}) \tag{3.5}$$

where α is the parameter which denotes the strength of the SFR-dependence of the FMR as well as the value which minimizes the scatter in O/H at fixed μ_α. This parameter, α, is generally found to be lower (a weaker SFR dependence) when determined with strong-line metallicities (e.g., $\alpha = 0.32$; Mannucci et al., 2010) and higher (a stronger SFR dependence) when determined with direct-method metallicities (e.g., $\alpha = 0.66$; Andrews & Martini, 2013), though recent estimations via strong-line metallicities by Curti et al. (2020, $\alpha = 0.55$) and Sanders et al. (2021, $\alpha = 0.60$) have brought these α-estimates into better agreement. For this work, we use the value of $\alpha = 0.63$, which derives from the SDSS $M_* -$ SFR stacks of Andrews & Martini (2013), with their direct-method metallicities corrected for diffuse ionized gas (DIG) contamination by Sanders et al. (2017).

In Figure 3.6, we show our $z \sim 2.3$ dwarf galaxy stack (green star; x-axis error bar represents the $\mu_{0.63}$ range of the stacking sample) in the direct-method O/H $- \mu_{0.63}$ parameter space, plotted at the $\mu_{0.63}$ value ($\mu_{0.63} = 8.07$) given by the stacking sample's median stellar mass ($\log(M_*/\mathrm{M}_\odot) = 8.29$) and median SFR ($\log(\mathrm{SFR}) = 0.353$). We compare our composite to the Sanders et al. (2017) best-fit linear representation of the FMR (black line) tracing the $z \sim 0$, DIG-corrected, Andrews & Martini (2013) $M_* -$ SFR stacks. We also include in Figure 3.6 the two mass bins of the Sanders et al. (2020) $z > 1$ auroral-line sample (black squares), which these authors found to be consistent, within ~ 0.1 dex at fixed $\mu_{0.63}$, with lower redshift samples in the $\mu_{0.63}$ direct-method FMR projection. Adding to these results, we find our dwarf galaxy composite to have a T_e-based metallicity that lies $\sim 0.5\sigma$ or ~ 0.11 dex below the linear relation representing the direct-method FMR. We therefore join numerous other authors in suggesting, via direct-method metallicities, that

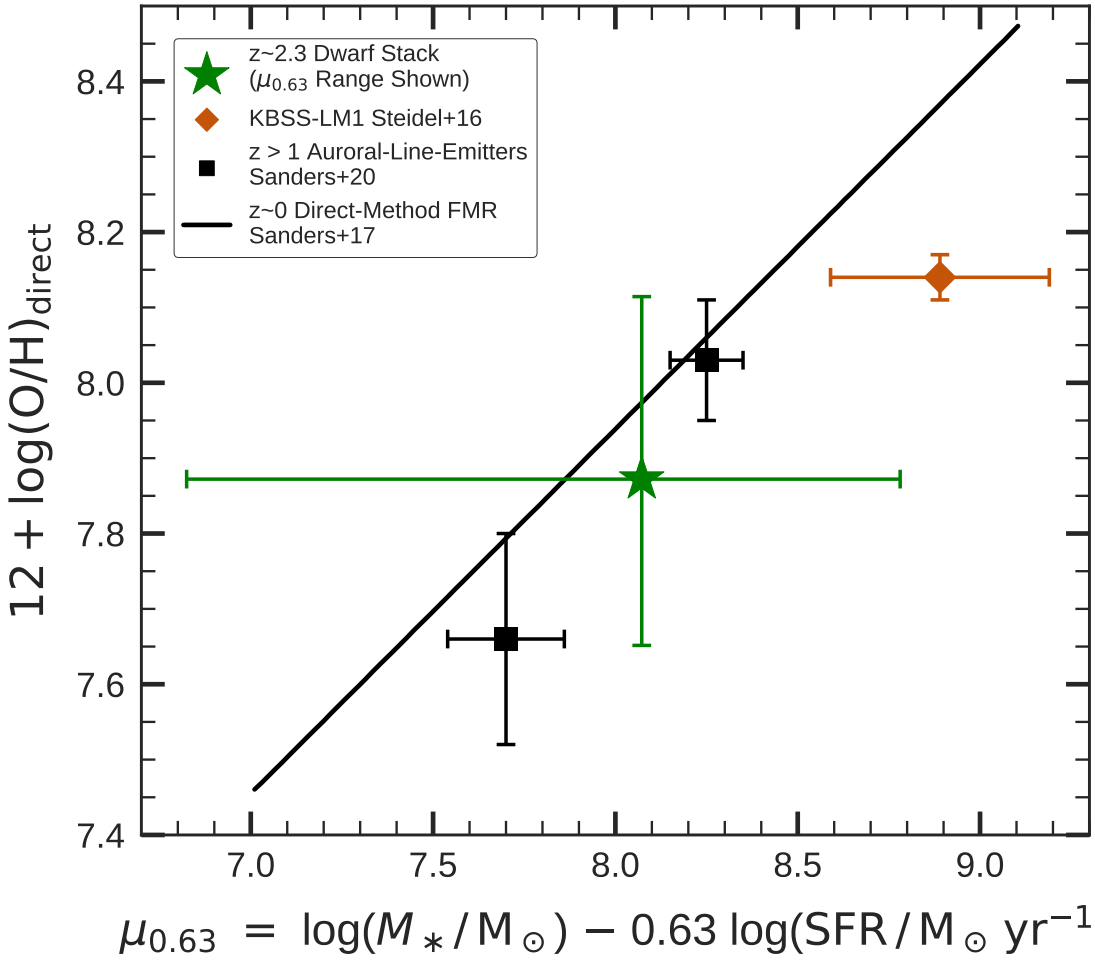

Figure 3.6: The direct-method $M_* - \mathrm{SFR} - \mathrm{O/H}$ fundamental metallicity relation (FMR), represented here by the $\mathrm{O/H} - \mu_\alpha$ planar projection (Mannucci et al., 2010). Referencing Equation 3.5 for μ_α, we take $\alpha = 0.63$, the value determined by Sanders et al. (2017) using the DIG-corrected, T_e-based metallicities of the Andrews & Martini (2013) SDSS $M_* - \mathrm{SFR}$ stacks. The best-fit linear relation by Sanders et al. (2017) to these $z \sim 0$ stacks is shown here by the black line. Our $z \sim 2.3$ composite is displayed as the green star with $\mu_{0.63}$ calculated from our sample's median M_* ($\log(M_*/\mathrm{M}_\odot) = 8.29$) and SFR ($\log(\mathrm{SFR}) = 0.353$). The x-axis error bar represents our sample's range in $\mu_{0.63}$. Our stack lies ~ 0.11 dex or $\sim 0.5\sigma$ below the direct-method FMR, though is consistent with a redshift-invariant FMR within uncertainties. For reference, we also show the M_*-binned, $z > 1$ auroral-line sample of Sanders et al. (2020, black squares) and the $z \sim 2.4$ KBSS-LM1 composite of Steidel et al. (2016, orange diamond).

the FMR evolves at most by ~ 0.1 dex in O/H at fixed M_* and SFR from $z = 0$ to at least $z \sim 2.3$, though note that within the uncertainties on the metallicity, our stack is consistent with a redshift-invariant FMR.

3.5 Summary

In this study, we analyze the median composite spectrum of 16 typical, star-forming, dwarf galaxies ($7.06 \leqslant \log(M_*/\mathrm{M}_\odot) \leqslant 8.93$; $\log(M_*/\mathrm{M}_\odot)_{\mathrm{median}} = 8.29$) at redshifts $1.7 < z < 2.6$ ($z_{\mathrm{mean}} = 2.30$) selected independent of the strength of any particular emission line. These galaxies are gravitationally-lensed by the foreground clusters Abell 1689, MACS J0717.5+3745, and MACS J1149.5+2223. In our composite spectrum, we find a 2.5σ (4.1σ) detection of the faint, T_e-sensitive, [O III] λ4363 auroral line when considering our bootstrapped (statistical-only) error spectrum, allowing us to directly calculate an oxygen abundance from the composite of $12 + \log(\mathrm{O/H})_{\mathrm{direct}} = 7.87^{+0.24}_{-0.22}$ ($0.15^{+0.11}_{-0.06}$ $\mathrm{Z}_\odot$). We summarize the results using this T_e-based metallicity, and other conclusions, in this final section.

1. To determine how representative our dwarf galaxy sample is of typical, star-forming, $z \sim 2.3$ dwarf galaxies, we first considered our composite in the context of the N2-BPT diagram, where we found that our stack lies offset from the $z \sim 0$ SDSS star-forming sequence in the same parameter space as the $z \sim 2.3$ star-forming galaxies of larger statistical surveys (e.g., MOSDEF; Shapley et al., 2015). We also show that our composite lies at higher [O III] λ5007/Hβ and lower [N II] λ6583/Hα than any of the M_*-binned MOSDEF stacks of Sanders et al. (2021); our composite extends the

trend seen with these MOSDEF stacks of lower stellar mass and metallicity at higher [O III]/Hβ and lower [N II]/Hα.

2. We also considered our stacking sample against an extrapolation of the $z \sim 2.3$ M_* – SFR SFMS of Sanders et al. (2021), finding the sample to scatter on either side of this relation. Our stacking sample has a median SFR = 2.25 $M_\odot\,yr^{-1}$, which lies $\Delta\log(SFR) \approx 0.19$ dex above this SFMS at fixed M_* ($\log(M_*/M_\odot)_{med} = 8.29$), corresponding to a bias in O/H of $\Delta\log(O/H) \approx -0.06$ dex via the FMR, well within even our statistical metallicity uncertainty ($\sigma \sim 0.12$ dex). We conclude that our stacking sample is not largely biased in SFR or O/H and is thus, on average, representative of typical, star-forming, $z \sim 2.3$ dwarf galaxies with stellar masses between $10^8 \lesssim M_*/M_\odot \lesssim 10^9$. Our sample serves as an initial extension of representative, statistical, spectroscopic surveys at $z \sim 2.3$ into the dwarf galaxy mass regime.

3. We analyzed the applicability at $z \sim 2.3$ of several locally-calibrated, oxygen-based, strong-line metallicity relations from the literature. We find that at $12+\log(O/H)_{dir} = 7.87^{+0.24}_{-0.22}$, our stack lies in the metallicity-insensitive "turnover" region of the O_3 and R_{23} calibrations, signalling their ineffectiveness for metallicity estimation of typical, $z \sim 2.3$, dwarf galaxies. When considering the O_{32} and O_2 indices together, our stack's metallicity is most accurately reproduced (within $\lesssim 0.15$ dex) at fixed strong-line ratio by the local reference calibrations of Bian et al. (2018), in agreement with that seen for the $z \sim 2.4$ KBSS-LM1 composite of Steidel et al. (2016), who also selected their sample independent of line-strength. We generally disagree with the conclusions of Sanders et al. (2020), who argue that their $z \sim 2.2$ auroral-line sample

favors the high-z analog calibrations of Bian et al. (2018). While both samples are at similar redshift and metallicity, we argue that our discrepancy in conclusion is due to sample selection effects as well as biases in the low-metallicity strong-line calibration samples. Indeed, by being selected for having a detection of a T_e-sensitive auroral-line, the sample of Sanders et al. (2020) is strongly biased in SFR, O_{32}, and EW_0([O III] λ5007) relative to typical, $z \sim 2.3$, star-forming galaxies.

4. At the median stellar mass of our stacking sample, $\log(M_*/M_\odot) = 8.29$, we compared our composite direct metallicity, $12+\log(O/H) = 7.87^{+0.24}_{-0.22}$, against the $z \sim 2.2$ direct-method MZR of Sanders et al. (2020), the $z \sim 2.3$ strong-line MZR of Sanders et al. (2021), and the $z \sim 2.3$ photoionization-model-based MZR of Strom et al. (2022). After correcting for the slight SFR bias of the stacking sample, we find that our $z \sim 2.3$ stack lies ~ 0.06 dex above the direct-method MZR and ~ 0.07 dex below the strong-line MZR at fixed M_*, well within our uncertainties. Our stack lies $\approx 1\sigma$ below the Strom et al. MZR. In constraining the slope of the MZR, we defer to the direct-method MZR, with which we show excellent agreement, as metallicities for this relation and our composite were calculated consistently. Therefore, we suggest that the slope of the $z \sim 2.3$ MZR is that given by this direct-method relation of Sanders et al. (2020), $\beta = 0.37$.

5. We also compared our composite, and the MZRs derived from observations, to the $z = 2.30$ MZR from the FIRE simulations (Ma et al., 2016) as well as to the $z = 2$ and $z = 3$ MZRs from the IllustrisTNG100 simulations (Torrey et al., 2019). At the stack's fiducial median stellar mass, $\log(M_*/M_\odot) = 8.29$, our composite is consistent within

uncertainties with both sets of simulations. However, when recalculating our sample stellar masses assuming more realistic non-parametric SFHs, the median stellar mass is increased to $\log(M_*/\mathrm{M}_\odot) = 8.92$, moving the stack into excellent agreement with the FIRE MZR and $\sim 1.5\sigma$ below the $2 \leqslant z \leqslant 3$ IllustrisTNG MZR. This tension with IllustrisTNG is in part caused by a "bump" in its MZR deriving from a constant minimum wind velocity ($v_{\mathrm{min}} = 350$ km s^{-1}) applied to galaxies with $M_* \lesssim 10^9$ M$_\odot$. Between our stack with recalculated masses and the direct-method MZR of Sanders et al. (2020), we suggest that the low-mass end of the MZR does not contain this bump.

6. Our $z \sim 2.3$ dwarf galaxy composite was compared to the locally-defined, direct-method FMR in order to test the relation's redshift-invariance. We made this comparison via the FMR projection proposed by Mannucci et al. (2010) and given in Equation 3.5, with $\alpha = 0.63$ (Sanders et al., 2017). At $\mu_{0.63} = 8.07$, calculated with the stacking sample's median stellar mass, $\log(M_*/\mathrm{M}_\odot) = 8.29$, and median SFR, $\log(\mathrm{SFR}) = 0.353$, we find our composite to lie $\sim 0.5\sigma$ or ~ 0.11 dex below the Sanders et al. (2017) best-fit linear relation in direct-method $\mathrm{O/H} - \mu_{0.63}$ space; this relation is fit to the $z \sim 0$ DIG-corrected stacks of Andrews & Martini (2013). We therefore agree with many in the literature who suggest that the FMR is redshift-invariant within ~ 0.1 dex at fixed M_* and SFR from $z \sim 0 - 2.3$.

This study compliments other larger spectroscopic surveys of representative, star-forming galaxies at $z \sim 2$ and $M_* \gtrsim 10^9$ M$_\odot$ by serving as an initial extension into the dwarf galaxy mass regime. In analyzing our sample of representative dwarf galaxies, we are able

to use a direct oxygen abundance to provide initial constraints on the low-mass slope of the high-z MZR and probe scarcely-studied parameter space of the FMR. While our sample size is small, our work provides a reference point for future statistical studies of high-z dwarf galaxies with the newly-operational *James Webb Space Telescope*, which will greatly increase our understanding of the processes responsible for galaxy formation and evolution.

3.6 Consistent Fits of the Strong-Line Metallicity Calibrations of Bian+18

We note that while Bian et al. (2018) provide fits to their high-z analog stacks for the strong-line indices of R_{23}, O_{32}, and O_3 (with [O III] λ4959 added to [O III] λ5007 in the latter two indices unlike in this work; see Bian et al. (2018) and Sanders et al. (2021), Footnote 17), they do not provide fits for the O_2 index or to the local reference stacks. Therefore, we used orthogonal distance regression and the stated flux and T_e-based metallicity values given in Bian et al. (2018) to consistently fit both samples for each line index considered in Figure 3.4 and Section 3.4.2. As in Bian et al. (2018), third-order polynomials were assumed for the R_{23} and O_3 high-z analog fits; we adopted the same functional form for the O_2 high-z analog fit as well. For the fits of the local reference stacks and these same strong-line indices, we instead assumed second-order polynomial functions due to there only being 4 bins with estimated direct metallicities compared to 6 bins for the high-z analog sample. For the O_{32} index, as in Bian et al. (2018), we fit a linear functional form to both the local reference and high-z analog stacks. The coefficients of our fits are given in Table 3.4 below.

Table 3.4: Coefficients of the Refit Bian+18 Relations

Ratio	c_0	c_1	c_2	c_3
Local Reference Relations				
O_{32}	11.1767	-1.3414		
O_3	-57.7632	14.9708	-0.9578	
O_2	-25.2699	5.3889	-0.2750	
R_{23}	-26.4726	6.9012	-0.4345	
High-z Analog Relations				
O_{32}	14.5895	-1.7287		
O_3	117.8668	-48.8336	6.7607	-0.3107
O_2	-80.9802	14.1526	0.0755	-0.0723
R_{23}	263.2135	-101.2448	13.0313	-0.5591

Notes: Ratio definitions given in Section 3.4.2.

Chapter 4

Conclusion

In the effort to better constrain models of galaxy formation and evolution, particularly with regard to understanding the processes behind the inefficiency of star formation in low-mass galaxies, the research and findings presented here focus on three main areas: the accuracy of high-z, gas-phase oxygen abundance estimates derived from locally-calibrated, strong-line metallicity relations, the low-mass slope and normalization of the high-z stellar mass – gas-phase metallicity relation (MZR), and the evolution with redshift of the fundamental metallicity relation (FMR). Of particular interest in each of these three areas is that our sample is comprised of representative, $z \sim 2.3$, star-forming dwarf galaxies with $M_* < 10^9$ M$_\odot$. This sample represents a first-look into the typical, high-redshift, dwarf galaxy population and is an extension of the statistical, spectroscopic, $z \sim 2$ surveys of typical star-forming galaxies with $M_* \gtrsim 10^9$ M$_\odot$ (e.g., MOSDEF; Kriek et al., 2015). Through a combination of the magnification effect from gravitational lensing and the ability to better detect faint emission lines by stacking individual galaxy spectra, the findings presented

here are based on metallicities calculated via the direct, T_e-based method which directly measures properties of star-forming regions and is applicable at all redshifts (Osterbrock & Ferland, 2006; Maiolino & Mannucci, 2019). Here we summarize our results from our studies of the $z = 2.59$ [O III] λ4363-emitter, A1689-217, and the composite of 16 dwarf galaxies at $\langle z \rangle = 2.3$ selected independent of emission-line strength. We do so in the context of existing and future high-z spectroscopic studies of star-forming galaxies.

When studying the MZR or FMR at high-redshift, it is imperative that metallicities are accurately determined. However, due to the limitations of the direct method, particularly the faintness of auroral emission lines, indirect, locally-calibrated relations between strong emission-line ratios and direct metallicities are used. Unfortunately, at high-z, the accuracy of these strong-line relations is not well-known due to evolving physical conditions with redshift in star-forming regions (Kewley et al., 2013; Steidel et al., 2014; Shapley et al., 2015). With a T_e-based metallicity measured for A1689-217 and for our dwarf galaxy composite, we address this issue and assess which, if any, oxygen-based strong-line metallicity calibrations (Maiolino et al., 2008; Jones et al., 2015; Curti et al., 2017; Bian et al., 2018; Sanders et al., 2021) are applicable at high-z. In both studies, we show that at high-z the strong-line indices of [O III] λ5007/Hβ and R_{23} are insensitive to metallicities similar to those of our composite or A1689-217, reaffirming the same conclusion put forth by other authors (e.g., Steidel et al., 2014; Sanders et al., 2020). Rather, we see greater sensitivity and utility at these metallicities with the strong-line indices [O III] λ5007/[O II] $\lambda\lambda$3726, 3729 and [O II] $\lambda\lambda$3726, 3729/Hβ. However, we show that when considering $z \sim 2.3$ samples selected independent of emission-line strength (such as our dwarf sample or the KBSS-LM1 sample

of Steidel et al. 2016) against a $z \sim 2.3$ sample selected by auroral-line detection (Sanders et al., 2020), there is a clear distinction in which calibrations most accurately reproduce the metallicities of the differently-selected samples. We suggest that the applicability at high-z of different strong-line metallicity calibrations may depend on the selection and properties of the high-z galaxies being studied as well as the selection of the local galaxies used to calibrate the relations.

In addition to the aforementioned strong-line indices and their direct comparison to T_e-based metallicities, in our study of A1689-217 we also considered the utility of the O_{32} (here O_{32} = [O III] $\lambda\lambda$4959, 5007/[O II] $\lambda\lambda$3726, 3729) vs. R_{23} excitation diagram as an empirical metallicity indicator. Shapley et al. (2015) were the first authors to advocate for an empirical calibration of this excitation diagram, in which they showed with the $z \sim 0$ SDSS composites of Andrews & Martini (2013) that, on average, there is a monotonic decrease in T_e-based metallicity with increasing O_{32} and R_{23}. These authors also suggested that this calibration should be applicable out to at least $z \sim 2$ due to there being no evidence of evolution with redshift of the high-excitation tail of the diagram. We tested this suggestion with the auroral-line-detected samples considered in our work, ranging in redshift from $z \sim 0$ to $z \sim 3.1$; we showed that our samples support the conclusions and suggestions of Shapley et al. (2015) that an empirical calibration of the O_{32} vs. R_{23} diagram would be a powerful, direct-metallicity-based, redshift-invariant metallicity indicator. While not discussed in Chapter 3, our dwarf galaxy composite shows some disagreement with these conclusions, though the KBSS-LM1 composite of Steidel et al. (2016) and the auroral-line sample of Sanders et al. (2020) do show general agreement. We argue that our results and

those of Shapley et al. (2015) warrant further study and possible creation of a metallicity calibration of the O_{32} vs. R_{23} excitation diagram, particularly at high-redshift as statistical samples of auroral-line-emitters become available via *JWST*.

In the study of galaxy formation and evolution, the mass-metallicity scaling relation is a vital tool as it probes the interconnected processes of galactic stellar mass buildup, chemical enrichment, and the flow of baryons into and out of galaxies. An area of focus in our studies, the slope of the low-mass end of the MZR helps to constrain models of outflows and stellar feedback, processes that are responsible for the observed efficiency with which low-mass galaxies form stars. With our representative dwarf galaxy composite, we are able to further constrain the recently-published, $z \sim 2.3$, direct-method MZR of Sanders et al. (2020), the first parameterization of the direct-method MZR at $z \sim 2$. When comparing to this relation, for which biases in SFR and redshift were carefully removed by Sanders et al., we find excellent agreement, a result that is robust due to the consistency with which metallicities (and stellar masses) were calculated between studies. We disagree with significantly shallower $z \sim 2.3$ MZR slopes proposed in the literature (e.g., Strom et al., 2022), though note that comparisons to these other studies are subject to significant systematic uncertainties that exist between metallicities calculated via different methods (Maiolino & Mannucci, 2019). When comparing our dwarf galaxy composite to MZRs from cosmological simulations, we find excellent agreement with the predicted MZR at $z \sim 2.3$ from the FIRE simulations (Ma et al., 2016). However, we show this agreement *after* recalculating our dwarf galaxy stellar masses assuming more realistic, non-parametric SFHs rather than our fiducial constant SFH assumption. While fair comparisons of different studies require

consistently-calculated stellar masses, as expanded upon below, we echo other recent works in suggesting that the stellar masses of high-z star-forming galaxies like ours may be underestimated when using simple SFH parameterizations, like a constant SFH, and when only rest-UV and rest-optical photometry is available for SED-fitting (Tang et al., 2022; Whitler et al., 2022; Topping et al., 2022).

In order to accurately determine a star-forming galaxy's stellar mass, the presence of both younger and older stellar populations must be accounted for. As we and others have shown, old, redder stellar populations can easily be hidden in the rest-UV and rest-optical by bright, young stars. We show this to be a possibility in the SED of A1689-217, whereby adding a maximally-old stellar population increases the fiducial stellar mass by 3.3× (0.52 dex) without significantly affecting the quality of the rest-UV to rest-optical SED fit (see Chapter 2.4.1 and Figure 2.5). When assuming a non-parametric SFH, which can better reveal older stellar populations than simpler SFH parameterizations, we find that the stellar mass of A1689-217 increases by an order of magnitude compared to when a constant SFH is assumed (see Table 3.1). When considering our dwarf galaxy stacking sample, we find that the median stellar mass is 0.63 dex larger when assuming non-parametric SFHs. These findings demonstrate that care must be taken when estimating stellar mass via SED-fitting of bluer photometric bands.

Finally, we explored the scatter and evolution with redshift of the FMR, a relation between galaxy stellar mass, SFR, and gas-phase metallicity which has been proposed to be redshift-invariant out to at least $z \sim 2.5$ (Mannucci et al., 2010). The FMR, which posits an anti-correlation between O/H and SFR, is thought to be more fundamental than the MZR

and can naturally explain the evolution of the MZR as gas-fractions and SFRs rise with redshift (e.g., Speagle et al., 2014). Proposed by Mannucci et al. (2010) and commonly used in the literature, in both of our studies we considered the 2D O/H $-$ μ_α planar projection of the 3D FMR with α values determined from studies using direct-method metallicities as opposed to strong-line metallicities. In our study of A1689-217 and the low-mass comparison samples therein, we show a large scatter around the low-μ_α end of the FMR where dwarf galaxies lie. We agree with Mannucci et al. (2011), who also find a large scatter around the FMR with dwarf SDSS galaxies, in suggesting this scatter is due to a large variation in dwarf galaxy SFHs and current star formation activity. In our study of our dwarf galaxy composite, we find that the composite lies $\sim 0.5\sigma$ or ~ 0.11 dex below the locally-defined, direct-method FMR. This result is in agreement with numerous other studies in the literature that suggest the FMR evolves at most by ~ 0.1 dex in metallicity at fixed M_* and SFR out to $z \sim 2.5$ and even possibly $z \sim 3.3$ (e.g., Sanders et al., 2018, 2021); our best-fit composite metallicity lies at this evolutionary limit below the FMR, but is consistent within uncertainties with a redshift-invariant FMR out to $z \sim 2.3$.

As mentioned above, the galaxies considered here comprise an initial sample of representative, star-forming dwarf galaxies at the peak of cosmic star formation. These galaxies probe scarcely-studied M_*, SFR, and O/H parameter space at $z \sim 2$. As part of this sample, A1689-217 represents one of only a handful of detected auroral-line-emitters at $z > 1$. While the study of high-z galaxies like those in our sample is important for many reasons, our current sample-size is unfortunately very limited due to the sample's inherent faintness and current technological limitations. The sample that we do have, as

well as many of the detected high-z auroral-line-emitters in the literature, have only been found due to gravitational lensing. Fortunately, with the increased sensitivity of upcoming 30 m ground-based telescopes and the newly-operational *James Webb Space Telescope*, we will be able to detect these high-z dwarf galaxies and auroral-line-emitters in much more statistically-significant numbers at $z \gtrsim 2$ without gravitational lensing. Moreover, with *JWST*'s wavelength coverage out to ~ 5.3 μm, rest-optical emission lines necessary for T_e-based metallicities can be observed out to $z \sim 9.5$. This combination of sensitivity and wavelength coverage will allow for strong-line metallicity relations to be calibrated outside of the local Universe, will much more effectively sample the low-mass end of the high-z MZR, and will allow for studies of metallicity evolution with redshift (and thus evolution of the MZR and FMR) into the epoch of reionization. As these different projects come to fruition, our sample of typical dwarf galaxies will serve as an important collection and reference point at cosmic high-noon.

Bibliography

Abazajian, K., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2005, AJ, 129, 1755, doi: 10.1086/427544

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, The Astrophysical Journal Supplement Series, 182, 543, doi: 10.1088/0067-0049/182/2/543

Acker, A., Köppen, J., Samland, M., & Stenholm, B. 1989, The Messenger, 58, 44

Alavi, A., Siana, B., Richard, J., et al. 2014, ApJ, 780, 143, doi: 10.1088/0004-637X/780/2/143

—. 2016, ApJ, 832, 56, doi: 10.3847/0004-637X/832/1/56

Alloin, D., Collin-Souffrin, S., Joly, M., & Vigroux, L. 1979, A&A, 78, 200

Andrews, B. H., & Martini, P. 2013, ApJ, 765, 140, doi: 10.1088/0004-637X/765/2/140

Asplund, M., Amarsi, A. M., & Grevesse, N. 2021, A&A, 653, A141, doi: 10.1051/0004-6361/202140445

Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, Annual Review of Astronomy and Astrophysics, 47, 481, doi: 10.1146/annurev.astro.46.060407.145222

Atek, H., Siana, B., Scarlata, C., et al. 2011, ApJ, 743, 121, doi: 10.1088/0004-637X/743/2/121

Atek, H., Kneib, J.-P., Pacifici, C., et al. 2014, ApJ, 789, 96, doi: 10.1088/0004-637X/789/2/96

Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, Publications of the Astronomical Society of the Pacific, 93, 5, doi: 10.1086/130766

Behroozi, P., Wechsler, R. H., Hearin, A. P., & Conroy, C. 2019, MNRAS, 488, 3143, doi: 10.1093/mnras/stz1182

Berg, D. A., Erb, D. K., Auger, M. W., Pettini, M., & Brammer, G. B. 2018, ApJ, 859, 164, doi: 10.3847/1538-4357/aab7fa

Berg, D. A., Skillman, E. D., Marble, A. R., et al. 2012, ApJ, 754, 98, doi: 10.1088/0004-637X/754/2/98

Bian, F., Kewley, L. J., & Dopita, M. A. 2018, ApJ, 859, 175, doi: 10.3847/1538-4357/aabd74

Brammer, G. B., van Dokkum, P. G., Franx, M., et al. 2012a, ApJS, 200, 13, doi: 10.1088/0067-0049/200/2/13

Brammer, G. B., Sánchez-Janssen, R., Labbé, I., et al. 2012b, ApJ, 758, L17, doi: 10.1088/2041-8205/758/1/L17

Brinchmann, J., Pettini, M., & Charlot, S. 2008, MNRAS, 385, 769, doi: 10.1111/j.1365-2966.2008.12914.x

Broadhurst, T., Benítez, N., Coe, D., et al. 2005, ApJ, 621, 53, doi: 10.1086/426494

Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000, doi: 10.1046/j.1365-8711.2003.06897.x

Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682, doi: 10.1086/308692

Campbell, A., Terlevich, R., & Melnick, J. 1986, MNRAS, 223, 811, doi: 10.1093/mnras/223.4.811

Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245, doi: 10.1086/167900

Carnall, A. C. 2017, arXiv e-prints, arXiv:1705.05165. https://arxiv.org/abs/1705.05165

Chabrier, G. 2003, Publications of the Astronomical Society of the Pacific, 115, 763, doi: 10.1086/376392

Charlot, S., & Fall, S. M. 2000, ApJ, 539, 718, doi: 10.1086/309250

Christensen, L., Laursen, P., Richard, J., et al. 2012, MNRAS, 427, 1973, doi: 10.1111/j.1365-2966.2012.22007.x

Cresci, G., Mannucci, F., & Curti, M. 2019, A&A, 627, A42, doi: 10.1051/0004-6361/201834637

Cullen, F., Cirasuolo, M., Kewley, L. J., et al. 2016, MNRAS, 460, 3002, doi: 10.1093/mnras/stw1181

Curti, M., Cresci, G., Mannucci, F., et al. 2017, MNRAS, 465, 1384, doi: 10.1093/mnras/stw2766

Curti, M., Mannucci, F., Cresci, G., & Maiolino, R. 2020, MNRAS, 491, 944, doi: 10.1093/mnras/stz2910

Dale, D. A., Cohen, S. A., Johnson, L. C., et al. 2009, ApJ, 703, 517, doi: 10.1088/0004-637X/703/1/517

Davé, R., Finlator, K., & Oppenheimer, B. D. 2012, MNRAS, 421, 98, doi: 10.1111/j.1365-2966.2011.20148.x

Davies, R. L., Förster Schreiber, N. M., Genzel, R., et al. 2021, ApJ, 909, 78, doi: 10.3847/1538-4357/abd551

Davis, M., Faber, S. M., Newman, J., et al. 2003, in Discoveries and Research Prospects from 6- to 10-Meter-Class Telescopes II, ed. P. Guhathakurta, Vol. 4834, 161–172

Denicoló, G., Terlevich, R., & Terlevich, E. 2002, MNRAS, 330, 69, doi: 10.1046/j.1365-8711.2002.05041.x

Dopita, M. A., & Sutherland, R. S. 2003, Astrophysics of the Diffuse Universe (Berlin: Springer)

Dopita, M. A., Sutherland, R. S., Nicholls, D. C., Kewley, L. J., & Vogt, F. P. A. 2013, The Astrophysical Journal Supplement Series, 208, 10, doi: 10.1088/0067-0049/208/1/10

Eldridge, J. J., & Stanway, E. R. 2009, MNRAS, 400, 1019, doi: 10.1111/j.1365-2966.2009.15514.x

Emami, N., Siana, B., Alavi, A., et al. 2020, ApJ, 895, 116, doi: 10.3847/1538-4357/ab8f97

Emami, N., Siana, B., Weisz, D. R., et al. 2018, arXiv e-prints, arXiv:1809.06380. https://arxiv.org/abs/1809.06380

Erb, D. K., Shapley, A. E., Pettini, M., et al. 2006, ApJ, 644, 813, doi: 10.1086/503623

Finlator, K., & Davé, R. 2008, MNRAS, 385, 2181, doi: 10.1111/j.1365-2966.2008.12991.x

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306, doi: 10.1086/670067

Freeman, W. R., Siana, B., Kriek, M., et al. 2019, ApJ, 873, 102, doi: 10.3847/1538-4357/ab0655

Gburek, T., Siana, B., Alavi, A., et al. 2019, ApJ, 887, 168, doi: 10.3847/1538-4357/ab5713

Genel, S., Vogelsberger, M., Springel, V., et al. 2014, MNRAS, 445, 175, doi: 10.1093/mnras/stu1654

Gordon, K. D., Clayton, G. C., Misselt, K. A., Land olt, A. U., & Wolff, M. J. 2003, ApJ, 594, 279, doi: 10.1086/376774

Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35, doi: 10.1088/0067-0049/197/2/35

Henry, A., Martin, C. L., Finlator, K., & Dressler, A. 2013a, ApJ, 769, 148, doi: 10.1088/0004-637X/769/2/148

Henry, A., Scarlata, C., Domínguez, A., et al. 2013b, ApJL, 776, L27, doi: 10.1088/2041-8205/776/2/L27

Hirschmann, M., Charlot, S., Feltre, A., et al. 2017, MNRAS, 472, 2468, doi: 10.1093/mnras/stx2180

Hopkins, P. F., Kereš, D., Oñorbe, J., et al. 2014, MNRAS, 445, 581, doi: 10.1093/mnras/stu1738

Izotov, Y. I., Stasińska, G., Meynet, G., Guseva, N. G., & Thuan, T. X. 2006, A&A, 448, 955, doi: 10.1051/0004-6361:20053763

James, B. L., Pettini, M., Christensen, L., et al. 2014, MNRAS, 440, 1794, doi: 10.1093/mnras/stu287

Jauzac, M., Richard, J., Limousin, M., et al. 2016, MNRAS, 457, 2029, doi: 10.1093/mnras/stw069

Jensen, E. B., Strom, K. M., & Strom, S. E. 1976, ApJ, 209, 748, doi: 10.1086/154773

Johnson, B. D., Leja, J., Conroy, C., & Speagle, J. S. 2021, ApJS, 254, 22, doi: 10.3847/1538-4365/abef67

Jones, T., Martin, C., & Cooper, M. C. 2015, ApJ, 813, 126, doi: 10.1088/0004-637X/813/2/126

Jullo, E., Kneib, J. P., Limousin, M., et al. 2007, New Journal of Physics, 9, 447, doi: 10.1088/1367-2630/9/12/447

Kaasinen, M., Bian, F., Groves, B., Kewley, L. J., & Gupta, A. 2017, MNRAS, 465, 3220, doi: 10.1093/mnras/stw2827

Kashikawa, N., Shimasaku, K., Yasuda, N., et al. 2004, PASJ, 56, 1011, doi: 10.1093/pasj/56.6.1011

Kashino, D., Silverman, J. D., Sanders, D., et al. 2017, ApJ, 835, 88, doi: 10.3847/1538-4357/835/1/88

—. 2019, ApJS, 241, 10, doi: 10.3847/1538-4365/ab06c4

Kauffmann, G., Heckman, T. M., Tremonti, C., et al. 2003, MNRAS, 346, 1055, doi: 10.1111/j.1365-2966.2003.07154.x

Kelson, D. D. 2003, PASP, 115, 688, doi: 10.1086/375502

Kennicutt, Robert C., J. 1998, Annual Review of Astronomy and Astrophysics, 36, 189, doi: 10.1146/annurev.astro.36.1.189

Kennicutt, Robert C., J., Lee, J. C., Funes, J. G., et al. 2008, ApJS, 178, 247, doi: 10.1086/590058

Kewley, L. J., & Dopita, M. A. 2002, The Astrophysical Journal Supplement Series, 142, 35, doi: 10.1086/341326

Kewley, L. J., Dopita, M. A., Leitherer, C., et al. 2013, ApJ, 774, 100, doi: 10.1088/0004-637X/774/2/100

Kewley, L. J., Dopita, M. A., Sutherland, R. S., Heisler, C. A., & Trevena, J. 2001, ApJ, 556, 121, doi: 10.1086/321545

Kewley, L. J., & Ellison, S. L. 2008, ApJ, 681, 1183, doi: 10.1086/587500

Kewley, L. J., Zahid, H. J., Geller, M. J., et al. 2015, ApJ, 812, L20, doi: 10.1088/2041-8205/812/2/L20

Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36, doi: 10.1088/0067-0049/197/2/36

Kriek, M., van Dokkum, P. G., Labbé, I., et al. 2009, ApJ, 700, 221, doi: 10.1088/0004-637X/700/1/221

Kriek, M., Shapley, A. E., Reddy, N. A., et al. 2015, The Astrophysical Journal Supplement Series, 218, 15, doi: 10.1088/0067-0049/218/2/15

Lara-López, M. A., Cepa, J., Bongiovanni, A., et al. 2010, A&A, 521, L53, doi: 10.1051/0004-6361/201014803

Larson, R. B. 1974, MNRAS, 169, 229, doi: 10.1093/mnras/169.2.229

Lee, H., Skillman, E. D., Cannon, J. M., et al. 2006, ApJ, 647, 970, doi: 10.1086/505573

Lee, J. C., Gil de Paz, A., Tremonti, C., et al. 2009, ApJ, 706, 599, doi: 10.1088/0004-637X/706/1/599

Lequeux, J., Peimbert, M., Rayo, J. F., Serrano, A., & Torres-Peimbert, S. 1979, A&A, 80, 155

Lilly, S. J., Carollo, C. M., Pipino, A., Renzini, A., & Peng, Y. 2013, ApJ, 772, 119, doi: 10.1088/0004-637X/772/2/119

Limousin, M., Richard, J., Jullo, E., et al. 2007, ApJ, 668, 643, doi: 10.1086/521293

—. 2016, A&A, 588, A99, doi: 10.1051/0004-6361/201527638

Lotz, J. M., Koekemoer, A., Coe, D., et al. 2017, ApJ, 837, 97, doi: 10.3847/1538-4357/837/1/97

Ly, C., Malkan, M. A., Nagao, T., et al. 2014, ApJ, 780, 122, doi: 10.1088/0004-637X/780/2/122

Ly, C., Rigby, J. R., Cooper, M., & Yan, R. 2015, ApJ, 805, 45, doi: 10.1088/0004-637X/805/1/45

Ma, X., Hopkins, P. F., Faucher-Giguère, C.-A., et al. 2016, MNRAS, 456, 2140, doi: 10.1093/mnras/stv2659

Madau, P., & Dickinson, M. 2014, ARA&A, 52, 415, doi: 10.1146/annurev-astro-081811-125615

Maiolino, R., & Mannucci, F. 2019, A&A Rv, 27, 3, doi: 10.1007/s00159-018-0112-2

Maiolino, R., Nagao, T., Grazian, A., et al. 2008, A&A, 488, 463, doi: 10.1051/0004-6361:200809678

Mannucci, F., Cresci, G., Maiolino, R., Marconi, A., & Gnerucci, A. 2010, MNRAS, 408, 2115, doi: 10.1111/j.1365-2966.2010.17291.x

Mannucci, F., Salvaterra, R., & Campisi, M. A. 2011, MNRAS, 414, 1263, doi: 10.1111/j.1365-2966.2011.18459.x

Maseda, M. V., van der Wel, A., Rix, H.-W., et al. 2014, ApJ, 791, 17, doi: 10.1088/0004-637X/791/1/17

Masters, D., Faisst, A., & Capak, P. 2016, ApJ, 828, 18, doi: 10.3847/0004-637X/828/1/18

Masters, D., McCarthy, P., Siana, B., et al. 2014, ApJ, 785, 153, doi: 10.1088/0004-637X/785/2/153

McGaugh, S. S. 1991, ApJ, 380, 140, doi: 10.1086/170569

McLean, I. S., Steidel, C. C., Epps, H., et al. 2010, in Ground-based and Airborne Instrumentation for Astronomy III, Vol. 7735, 77351E

McLean, I. S., Steidel, C. C., Epps, H. W., et al. 2012, in Ground-based and Airborne Instrumentation for Astronomy IV, Vol. 8446, 84460J

Nakajima, K., & Ouchi, M. 2014, MNRAS, 442, 900, doi: 10.1093/mnras/stu902

Nakajima, K., Ouchi, M., Shimasaku, K., et al. 2013, ApJ, 769, 3, doi: 10.1088/0004-637X/769/1/3

Newman, J. A., Cooper, M. C., Davis, M., et al. 2013, The Astrophysical Journal Supplement Series, 208, 5, doi: 10.1088/0067-0049/208/1/5

Oke, J. B., Cohen, J. G., Carr, M., et al. 1995, PASP, 107, 375, doi: 10.1086/133562

Onodera, M., Carollo, C. M., Lilly, S., et al. 2016, ApJ, 822, 42, doi: 10.3847/0004-637X/822/1/42

Osterbrock, D. E., & Ferland, G. J. 2006, Astrophysics of Gaseous Nebulae and Active Galactic Nuclei (Sausalito, CA: University Science Books)

Pagel, B. E. J., Edmunds, M. G., Blackwell, D. E., Chun, M. S., & Smith, G. 1979, MNRAS, 189, 95, doi: 10.1093/mnras/189.1.95

Patrício, V., Christensen, L., Rhodin, H., Cañameras, R., & Lara-López, M. A. 2018, MNRAS, 481, 3520, doi: 10.1093/mnras/sty2508

Pettini, M., & Pagel, B. E. J. 2004, MNRAS, 348, L59, doi: 10.1111/j.1365-2966.2004.07591.x

Pillepich, A., Springel, V., Nelson, D., et al. 2018, MNRAS, 473, 4077, doi: 10.1093/mnras/stx2656

Pilyugin, L. S., & Thuan, T. X. 2005, ApJ, 631, 231, doi: 10.1086/432408

Postman, M., Coe, D., Benítez, N., et al. 2012, ApJS, 199, 25, doi: 10.1088/0067-0049/199/2/25

Reddy, N. A., Oesch, P. A., Bouwens, R. J., et al. 2018a, ApJ, 853, 56, doi: 10.3847/1538-4357/aaa3e7

Reddy, N. A., Shapley, A. E., Sanders, R. L., et al. 2018b, ApJ, 869, 92, doi: 10.3847/1538-4357/aaed1e

Runco, J. N., Reddy, N. A., Shapley, A. E., et al. 2022, MNRAS, 513, 3871, doi: 10.1093/mnras/stac1115

Sanders, R. L., Shapley, A. E., Zhang, K., & Yan, R. 2017, ApJ, 850, 136, doi: 10.3847/1538-4357/aa93e4

Sanders, R. L., Shapley, A. E., Kriek, M., et al. 2015, ApJ, 799, 138, doi: 10.1088/0004-637X/799/2/138

—. 2016a, ApJ, 825, L23, doi: 10.3847/2041-8205/825/2/L23

—. 2016b, ApJ, 816, 23, doi: 10.3847/0004-637X/816/1/23

—. 2018, ApJ, 858, 99, doi: 10.3847/1538-4357/aabcbd

Sanders, R. L., Shapley, A. E., Reddy, N. A., et al. 2019, arXiv e-prints, arXiv:1907.00013. https://arxiv.org/abs/1907.00013

—. 2020, MNRAS, 491, 1427, doi: 10.1093/mnras/stz3032

Sanders, R. L., Shapley, A. E., Jones, T., et al. 2021, ApJ, 914, 19, doi: 10.3847/1538-4357/abf4c1

Shapley, A. E., Reddy, N. A., Kriek, M., et al. 2015, ApJ, 801, 88, doi: 10.1088/0004-637X/801/2/88

Shaw, R. A., & Dufour, R. J. 1994, in Astronomical Data Analysis Software and Systems III, ed. D. R. Crabtree, R. J. Hanisch, & J. Barnes, Vol. 61, 327

Shirazi, M., Brinchmann, J., & Rahmati, A. 2014, ApJ, 787, 120, doi: 10.1088/0004-637X/787/2/120

Shivaei, I., Reddy, N. A., Shapley, A. E., et al. 2015, ApJ, 815, 98, doi: 10.1088/0004-637X/815/2/98

Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, ApJS, 214, 15, doi: 10.1088/0067-0049/214/2/15

Stark, D. P., Auger, M., Belokurov, V., et al. 2013, MNRAS, 436, 1040, doi: 10.1093/mnras/stt1624

Steidel, C. C., Shapley, A. E., Pettini, M., et al. 2004, ApJ, 604, 534, doi: 10.1086/381960

Steidel, C. C., Strom, A. L., Pettini, M., et al. 2016, ApJ, 826, 159, doi: 10.3847/0004-637X/826/2/159

Steidel, C. C., Rudie, G. C., Strom, A. L., et al. 2014, ApJ, 795, 165, doi: 10.1088/0004-637X/795/2/165

Storchi-Bergmann, T., Calzetti, D., & Kinney, A. L. 1994, ApJ, 429, 572, doi: 10.1086/174345

Storey, P. J., & Zeippen, C. J. 2000, MNRAS, 312, 813, doi: 10.1046/j.1365-8711.2000.03184.x

Straughn, A. N., Kuntschner, H., Kümmel, M., et al. 2011, AJ, 141, 14, doi: 10.1088/0004-6256/141/1/14

Strom, A. L., Rudie, G. C., Steidel, C. C., & Trainor, R. F. 2022, ApJ, 925, 116, doi: 10.3847/1538-4357/ac38a3

Strom, A. L., Steidel, C. C., Rudie, G. C., Trainor, R. F., & Pettini, M. 2018, ApJ, 868, 117, doi: 10.3847/1538-4357/aae1a5

Strom, A. L., Steidel, C. C., Rudie, G. C., et al. 2017, ApJ, 836, 164, doi: 10.3847/1538-4357/836/2/164

Tang, M., Stark, D., Chevallard, J., & Charlot, S. 2018, arXiv e-prints, arXiv:1809.09637. https://arxiv.org/abs/1809.09637

Tang, M., Stark, D. P., & Ellis, R. S. 2022, MNRAS, 513, 5211, doi: 10.1093/mnras/stac1280

Topping, M. W., Stark, D. P., Endsley, R., et al. 2022, arXiv e-prints, arXiv:2203.07392. https://arxiv.org/abs/2203.07392

Torrey, P., Vogelsberger, M., Genel, S., et al. 2014, MNRAS, 438, 1985, doi: 10.1093/mnras/stt2295

Torrey, P., Vogelsberger, M., Marinacci, F., et al. 2019, MNRAS, 484, 5587, doi: 10.1093/mnras/stz243

Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, ApJ, 613, 898, doi: 10.1086/423264

Troncoso, P., Maiolino, R., Sommariva, V., et al. 2014, A&A, 563, A58, doi: 10.1051/0004-6361/201322099

Tumlinson, J., Peeples, M. S., & Werk, J. K. 2017, ARA&A, 55, 389, doi: 10.1146/annurev-astro-091916-055240

van der Wel, A., Straughn, A. N., Rix, H. W., et al. 2011, ApJ, 742, 111, doi: 10.1088/0004-637X/742/2/111

Vogelsberger, M., Genel, S., Springel, V., et al. 2014a, MNRAS, 444, 1518, doi: 10.1093/mnras/stu1536

—. 2014b, Nature, 509, 177, doi: 10.1038/nature13316

Weinberger, R., Springel, V., Hernquist, L., et al. 2017, MNRAS, 465, 3291, doi: 10.1093/mnras/stw2944

Weisz, D. R., Johnson, B. D., Johnson, L. C., et al. 2012, ApJ, 744, 44, doi: 10.1088/0004-637X/744/1/44

Whitaker, K. E., Franx, M., Leja, J., et al. 2014, ApJ, 795, 104, doi: 10.1088/0004-637X/795/2/104

—. 2020, ApJ, 896, 175, doi: 10.3847/1538-4357/ab98fc

Whitler, L., Stark, D. P., Endsley, R., et al. 2022, arXiv e-prints, arXiv:2206.05315. https://arxiv.org/abs/2206.05315

Yates, R. M., Kauffmann, G., & Guo, Q. 2012, MNRAS, 422, 215, doi: 10.1111/j.1365-2966.2012.20595.x

Yuan, T. T., & Kewley, L. J. 2009, ApJ, 699, L161, doi: 10.1088/0004-637X/699/2/L161

Zahid, H. J., Dima, G. I., Kudritzki, R.-P., et al. 2014a, ApJ, 791, 130, doi: 10.1088/0004-637X/791/2/130

Zahid, H. J., Geller, M. J., Kewley, L. J., et al. 2013, ApJ, 771, L19, doi: 10.1088/2041-8205/771/2/L19

Zahid, H. J., Kashino, D., Silverman, J. D., et al. 2014b, ApJ, 792, 75, doi: 10.1088/0004-637X/792/1/75