UC Davis
UC Davis Previously Published Works

Title

The Mass-Metallicity Relation at z ≃ 8: Direct-method Metallicity Constraints and Near-future Prospects

Permalink

https://escholarship.org/uc/item/4pb092bg

Journal

Astrophysical Journal, 903(2)

Authors

Jones, T
Sanders, R
Roberts-Borsani, G
et al.

Publication Date

2020-11-10

Peer reviewed

THE MASS-METALLICITY RELATION AT $Z \simeq 8$:
DIRECT-METHOD METALLICITY CONSTRAINTS AND NEAR-FUTURE PROSPECTS

TUCKER JONES[1], RYAN SANDERS[1,2], GUIDO ROBERTS-BORSANI[3], RICHARD S. ELLIS[4], NICOLAS LAPORTE[5,6],
TOMMASO TREU[3], AND YUICHI HARIKANE[4,7]

[1] Department of Physics, University of California Davis, 1 Shields Avenue, Davis, CA 95616, USA
[2] Hubble Fellow
[3] Department of Physics and Astronomy, University of California, Los Angeles, 430 Portola Plaza, Los Angeles, CA 90095, USA
[4] Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK
[5] Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
[6] Cavendish Laboratory, University of Cambridge, 19 JJ Thomson Avenue, Cambridge CB3 0HE, UK and
[7] National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan
Draft version September 22, 2020

ABSTRACT

Physical properties of galaxies at $z > 7$ are of interest for understanding both the early phases of star formation and the process of cosmic reionization. Chemical abundance measurements offer valuable information on the integrated star formation history, and hence ionizing photon production, as well as the rapid gas accretion expected at such high redshifts. We use reported measurements of [O III] 88μm emission and star formation rate to estimate gas-phase oxygen abundances in five galaxies at $z = 7.1 - 9.1$ using the direct T_e method. We find typical abundances $12 + \log{(O/H)} = 7.9$ ($\sim$0.2 times the solar value) and an evolution of 0.9 ± 0.5 dex in oxygen abundance at fixed stellar mass from $z \simeq 8$ to 0. These results are compatible with theoretical predictions, albeit with large (conservative) uncertainties in both mass and metallicity. We assess both statistical and systematic uncertainties to identify promising means of improvement with the Atacama Large Millimeter Array (ALMA) and the James Webb Space Telescope (JWST). In particular we highlight [O III] 52μm as a valuable feature for robust metallicity measurements. Precision of 0.1–0.2 dex in T_e-based O/H abundance can be reasonably achieved for galaxies at $z \approx 5$–8 by combining [O III] 52μm with rest-frame optical strong lines. It will also be possible to probe gas mixing and mergers via resolved T_e-based abundances on kpc scales. With ALMA and JWST, direct metallicity measurements will thus be remarkably accessible in the reionization epoch.

Subject headings: Galaxy chemical evolution (580) — High-redshift galaxies (734) — Galaxy evolution (594) — Reionization (1383)

1. INTRODUCTION

Deep spectroscopy of distant galaxies has recently begun to provide both redshift confirmations and a glimpse of their physical properties at $z > 7$. Multiple lines of evidence suggest that this period coincides with a phase transition of the universe from largely neutral at $z > 7$ to primarily ionized at $z < 6$, marking the "epoch of reionization" (e.g., Stark 2016; Mason et al. 2018). A key question is whether star formation in early galaxies produced enough ionizing photons to account for the reionization process. Current constraints rely largely on the rest-frame ultraviolet luminosity function of photometrically-selected galaxy samples, measured from Hubble Space Telescope imaging. This provides a good instantaneous snapshot of star formation rates reaching $z \approx 10$ (e.g., Ellis et al. 2013; Calvi et al. 2016; Oesch et al. 2018; Bouwens et al. 2019). However, information on the previous history of star formation and ionizing photon production remains ambiguous and challenging to obtain (e.g., Strait et al. 2020; Roberts-Borsani et al. 2020). Therefore relatively little is known about galaxy assembly in the first 500 Myr of the Universe at $z \gtrsim 10$.

Gas-phase metallicity[1] of galaxies is a valuable diag-

nostic property as it is sensitive to star formation history, as well as cosmological gas accretion and metal-enriched outflows which are expected to be prevalent at high redshifts. A strong correlation between galaxy stellar mass and metallicity has been established up to $z \simeq 3.5$, enabling interpretation of galaxy chemical evolution in terms of past star formation and gas flow rates (e.g., Maiolino & Mannucci 2019, and references therein). At higher redshifts, the widely used optical emission line diagnostics become inaccessible to ground-based observations, but will be observable with the upcoming James Webb Space Telescope (JWST). Currently only a few measurements of galaxy metallicities have been reported above $z > 4$ using a variety of methods (e.g., Shapley et al. 2017; Totani et al. 2006; Faisst et al. 2016; Cullen et al. 2019).

This paper is concerned with direct measurement of galaxy metallicities in the epoch of reionization. The standard "direct method" (or "T_e method") to calculate metallicity relies on nebular electron temperature T_e and density n_e. T_e is the more challenging to obtain and is typically determined from ratios of auroral and strong nebular emission lines (e.g., [O III]λ4363/[O III]λ5007). However, auroral line fluxes are well below the detection thresholds of typical high-redshift surveys. To date this method has been applied for tens of galaxies at $z \sim 1$ (e.g., Jones et al. 2015; Ly et al. 2014) and only a hand-

---

[1] Throughout this paper we use the term "metallicity" to refer to gas-phase abundance of oxygen relative to hydrogen, O/H.

ful at $z > 2$ (e.g., Sanders et al. 2020; Christensen et al. 2012; Gburek et al. 2019), with none at $z > 4$. Applying the T_e method with optical auroral lines at higher redshifts will be extremely challenging even with JWST. In contrast, several detections of far-IR [O III] 88μm emission have recently been achieved at $z > 7$ using the Atacama Large Millimeter Array (ALMA; Inoue et al. 2016; Laporte et al. 2017; Carniani et al. 2017; Hashimoto et al. 2018, 2019; Tamura et al. 2019). This breakthrough with ALMA provides a promising route toward direct metallicity measurements of bright $z > 7$ galaxies using [O III] 52μm,88μm combined with [O III]$\lambda\lambda$4959,5007 to determine T_e. In fact the T_e derived using far-IR lines is less sensitive to variations in temperature than auroral diagnostics, potentially enabling more robust results (e.g., Croxall et al. 2013; Esteban et al. 2009).

This paper addresses both available constraints and future prospects for galaxy metallicities at $z > 7$. We first describe our method and apply it to a sample of $z > 7$ galaxies (Section 2), followed by an assessment of systematic uncertainties (Section 3). We discuss implications of our results for chemical evolution in the reionization epoch in Section 4. In Section 5 we consider how to mitigate uncertainties with future observations using JWST and ALMA, feasibly reaching 0.1–0.2 dex precision in O/H. We summarize the main conclusions in Section 6. Throughout this paper we adopt a flat ΛCDM cosmology with $\Omega_M = 0.272$, $\Omega_\Lambda = 0.728$, and $H_0 = 70.4$ km s^{-1} Mpc^{-1} (Komatsu et al. 2011). We refer to emission lines by their rest-frame wavelengths, adopting common but separate conventions for optical (e.g., [O III]λ4363, implicitly in Angstroms) and infrared lines (e.g., [O III] 88μm).

2. METALLICITY CONSTRAINTS AT $Z > 7$ FROM CURRENT [O III] 88μm DATA

In this section we show that a combination of [O III] 88μm nebular emission and star formation rate (SFR) alone provides reasonable constraints on the metallicity, subject to assumptions about nebular physical properties. The basic methodology is as follows:

1. Estimate H I recombination line luminosity (e.g. Hβ) based on the photometrically-derived SFR

2. Calculate the abundance ratio O^{++}/H^+ from the ratio of [O III] 88μm/Hβ luminosities, with an assumed electron density (n_e) and temperature (T_e)

3. Assess systematic uncertainties in abundance based on a plausible range of n_e, T_e, and ionization correction factor (ICF)

A notable aspect is that T_e contributes relatively little uncertainty, whereas it typically dominates the error budget of [O III]λ4363-based measurements. This advantage results from the relative insensitivity of far-IR emission lines to temperature (as we discuss in Section 2.4).

2.1. $z > 7$ galaxy sample

We analyze the sample of six Lyman break galaxies with [O III] 88μm detections compiled by Harikane et al. (2020), spanning $z = 7.1$–9.1. The relevant measurements and original literature references are given

in Table 2. In the case of BDF-3299, the SFR and [O III] 88μm measurements correspond to different spatial regions which we consider separately, with appropriate upper and lower limits. The sample therefore contains 7 sources, of which 5 have measurements of both SFR and [O III] 88μm.

We estimate H I recombination line luminosities using the calibration of Kennicutt (1998) converted to a Chabrier stellar initial mass function (IMF):

$$L(H\alpha) \text{ [erg s}^{-1}] = 2.3 \times 10^{41} \times \text{SFR [M}_\odot \text{ yr}^{-1}] \quad (1)$$

as appropriate for the SFR values in Table 2 (for details see Harikane et al. 2020, and other tabulated references). This calibration is consistent (within 5%) with the mean relation for $z \sim 2$ galaxies from Shivaei et al. (2016). We report Hβ luminosity as a practical example (e.g., easily compared with [O III]$\lambda\lambda$4959,5007), with intrinsic ratio $\frac{H\alpha}{H\beta} = 2.79$ appropriate for Case B recombination and $T_e = 1.5 \times 10^4$ K. Both the SFR and luminosity values in Table 2 are corrected for dust attenuation (with a range $A_V \approx 0$–1 from original references), while observed Hβ fluxes are expected to be reduced.

2.2. Oxygen abundance

The combination of [O III] 88μm and Hβ luminosities now allows an estimate of the ion abundance ratio O^{++}/H^+, which we compute using the PYNEB package (Luridiana et al. 2012). We adopt fiducial values of $T_e = 1.5 \times 10^4$ K and $n_e = 250$ cm^{-3}, which gives a simple relation for doubly ionized oxygen abundance:

$$12 + \log(O^{++}/H^+) = 7.735 + \log \frac{L_{88\mu m}}{10^7 \, L_\odot} - \log \frac{\text{SFR}}{M_\odot \, \text{yr}^{-1}} \quad (2)$$

expressed here in terms of SFR. These fiducial values are chosen on the basis of available measurements for samples at $z \gtrsim 2$ (e.g., Sanders et al. 2016, 2020; Strom et al. 2017); both the fiducial T_e and n_e are considerably higher than for typical galaxies at $z \simeq 0$ (e.g., Andrews & Martini 2013). The $z \gtrsim 2$ sample with temperature measurements available is in fact representative of even higher-z galaxy demographics in terms of specific SFR, young inferred ages, emission line equivalent widths, and other properties. Resulting ion abundance ratios are given in Table 2, along with total abundance $12 + \log(O/H)$ for an estimated ionization correction of 0.17 dex (Section 3.3).

This simple exercise demonstrates the method of determining oxygen abundances, with fiducial assumptions suggesting approximately 0.05–0.4 times the solar value for this $z > 7$ sample (adopting solar $12 + \log(O/H) = 8.69$; Asplund et al. 2009). However we must first verify the method (Section 2.3) and critically assess the uncertainties (Section 3), before discussing implications of the results.

2.3. Validation of the method at $z = 0$

To verify the reliability of using [O III] 88μm and SFR to calculate metallicity, we apply the method described above to nearby galaxies with Herschel Space Observatory data from the Dwarf Galaxy Survey (DGS; Madden et al. 2013). The DGS sample is chosen as being
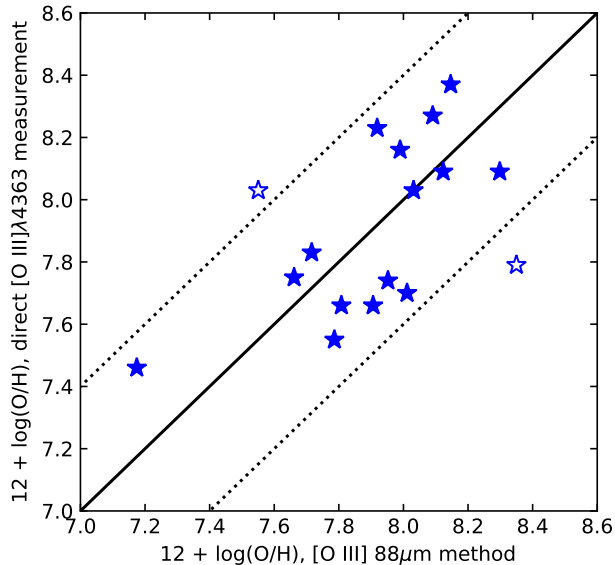
FIG. 1.— Comparison of $12 + \log{(O/H)}$ determined from [O III] 88μm and SFR, compared to measurements based on standard optical emission line methods (using [O III]λ4363 as a T_e diagnostic). Blue stars show all galaxies with suitable [O III] 88μm measurements from Herschel taken as part of the Dwarf Galaxy Survey (DGS). The DGS sample probes sub-solar metallicities relevant for comparison to high redshift galaxies (here we adopt a solar abundance $12 + \log{(O/H)} = 8.69$). The solid black line shows the one-to-one relation, with dashed lines at ± 0.4 dex (corresponding to 2σ scatter of the solid symbols), demonstrating reasonably good agreement. We note that the dashed lines serve to illustrate scatter about the one-to-one relation and are not intended to represent confidence intervals. The galaxies which differ by >0.4 dex (open symbols) can be reconciled with direct measurements as discussed in the text.

likely the most appropriate comparison based on properties such as nebular excitation and metallicity. Table 1 lists relevant physical properties for all 19 DGS targets for which infrared-based star formation rates are available (Madden et al. 2013), and for which Herschel spectroscopy covers the entire far-IR emission region. Figure 1 shows the resulting metallicities calculated using [O III] 88μm fluxes from Cormier et al. (2015), with distances and SFRs from Madden et al. (2013), as given in Table 1. These [O III] 88μm-based metallicities are compared with optical [O III]λ4363-based direct measurements, as reported by Madden et al. (2013) for 17 of the 19 galaxies (using the method of Izotov et al. 2006; Table 1). For comparison purposes we subtract 0.2 dex from the [O III] 88μm-based metallicities in Figure 1, to account for systematic differences in the [O III]λ4363 method attributed to temperature fluctuations (i.e., the well-known "abundance discrepancy factor" or ADF; Esteban et al. 2009; López-Sánchez et al. 2012; Blanc et al. 2015). [O III] 88μm is not significantly affected by such temperature fluctuations, although it may be subject to other biases. We have otherwise adopted the same fiducial values as for the $z > 7$ sample described above.

Figure 1 shows generally good agreement between our method and previously published O/H values (Table 1), with RMS scatter of 0.26 dex. The average offset is <0.01 dex, although a systematic offset can easily be introduced (or removed) by varying the fiducial assumptions such as

density, ionization correction, or the ADF. It is notable that the scatter in Figure 1 is <0.3 dex given that a single T_e, n_e, and ICF has been assumed for the entire DGS sample. Only two of the 17 galaxies in Figure 1 are discrepant by >0.35 dex, and we consider these cases specifically:

- *Mrk 1450*: Metallicity from the fiducial [O III] 88μm method is lower by $\Delta(\log{O/H}) = -0.48$ than the [O III]λ4363-based value reported in Madden et al. (2013). This discrepancy appears to arise largely from the SFR, which is used only as a proxy for H I line luminosity. The extinction-corrected H I Balmer line flux (Izotov et al. 1994) corresponds to SFR $= 0.09\,M_\odot\,{\rm yr}^{-1}$ (via Equation 1), comparable to the SFR $= 0.11\,M_\odot\,{\rm yr}^{-1}$ reported by Sargsyan & Weedman (2009) based on 1.4 GHz measurements, and lower than the 0.43 $M_\odot\,{\rm yr}^{-1}$ from Madden et al. (2013). Adopting this direct measurement of H I line luminosity along with T_e, n_e, and ICF (Izotov et al. 2014, 1994), we find good agreement with $\Delta(\log{O/H}) = 0.02$.

- *UM 461*: The difference $\Delta(\log{O/H}) = 0.56$ cannot be explained by n_e and T_e, which are both similar to the fiducial values (Lagos et al. 2018). However, Lagos et al. (2018) report an Hα-based SFR $= 0.077\,M_\odot\,{\rm yr}^{-1}$ (indicating higher H I luminosity c.f. 0.01 $M_\odot\,{\rm yr}^{-1}$ from Madden et al. 2013) and a relatively low ICF of 0.03 dex. Adopting these measurements, the difference is $\Delta(\log{O/H}) = -0.14$.

In sum, the largest outliers from Figure 1 are in fact consistent (< 0.2 dex) with the [O III] 88μm method when direct measurements are used instead of fiducial values. The RMS scatter in Figure 1 is only 0.2 dex if these two galaxies are removed or corrected; these cases serve to illustrate potential causes and rates of outliers. The data show a positive correlation at 97% confidence level (Pearson correlation coefficient $\rho = 0.53$), or 99.7% confidence ($\rho = 0.70$) if these two galaxies are excluded.

We have also performed the comparison shown in Figure 1 with metallicities derived from the Pilyugin & Thuan (2005) method (Table 1; for a detailed comparison see Appendix A of Madden et al. 2013). This gives nearly identical results with an RMS scatter of 0.30 dex (0.20 dex with outliers removed), and positive correlation at 95% confidence level (99.9% with outliers removed), although the metallicities from this method are on average 0.09 dex higher. The three cases which differ by >0.35 dex are all reconciled (within <0.2 dex) when direct measurements of T_e, n_e, and ICF are used instead of fiducial values. The main conclusions of this comparison are therefore unchanged regardless of which metallicity values in Table 1 (via Madden et al. 2013) are adopted.

These results demonstrate that the combination of [O III] 88μm and SFR alone can empirically reproduce standard T_e-method metallicity measurements to within $\sim$0.3 dex for individual galaxies in the DGS sample, and < 0.2 dex using prior information (such as ICF and n_e). We conclude that the basic method presented here is sound.
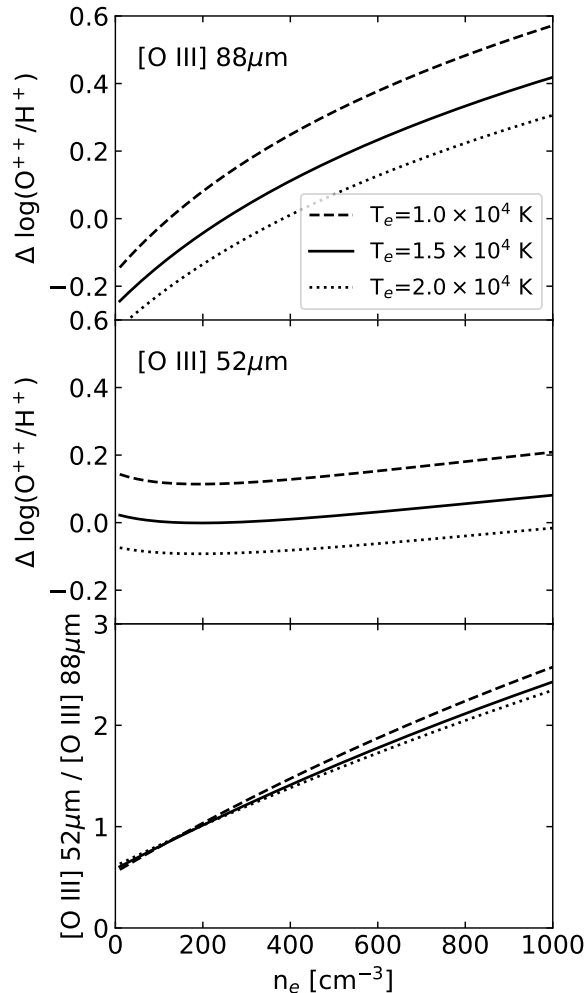
FIG. 2.— Illustration of how assumed density and temperature values affect the derived metallicity. The top panel shows how derived O^{++} abundance varies relative to the fiducial values (Section 2) at fixed [O III] 88μm flux. A strong dependence on n_e is apparent due to suppression of [O III] 88μm emission by collisional de-excitation. In contrast, the middle panel shows that [O III] 52μm-based abundance is nearly independent of density for $n_e \lesssim 1000$ cm^{-3}, with only modest temperature dependence. [O III] 52μm is therefore a promising metallicity diagnostic for future work. The bottom panel demonstrates how the [O III] doublet ratio is a robust density diagnostic (e.g., Palay et al. 2012), with [O III] 52μm becoming the stronger line at $n_e > 200$ cm^{-3}.

2.4. The advantage of [O III] 52μm, 88μm over [O III]λ4363

A principal motivation for the [O III] 88μm-based approach is the demonstrated success in detecting far-IR lines at $z > 7$. It is not a very pragmatic method at low redshift, where [O III]λ4363 is instead the most common T_e diagnostic. This is in part due to challenges of infrared observations and the relative ease of detecting optical auroral lines in nearby galaxies. However, the situation is reversed at high redshifts where far-IR lines become accessible to ALMA. A key advantage of far-IR lines is their relative brightness compared to auroral lines. Another major advantage noted previously is that far-IR lines are relatively insensitive to temperature (e.g., Croxall et al. 2013). Over a range of 9,000–

20,000 K, [O III] 88μm varies by only 15% in emissivity, whereas [O III]λ4363 varies by a factor 33$\times$. This causes [O III]λ4363 and other auroral line measurements to be strongly biased toward high-temperature regions. As a result, current $z \gtrsim 1$ auroral line samples are a biased subset of the overall galaxy population (e.g., Sanders et al. 2020). In contrast, far-IR lines depend primarily on ion abundances and thus can provide more robust metallicities across a broader sample.

While this work is necessarily limited to available measurements of the 88 μm line, we note that [O III] 52μm is a superior diagnostic. Figure 2 demonstrates that both lines have similarly weak temperature dependence, but [O III] 52μm is far less sensitive to collisional de-excitation at densities $n_e \lesssim 1000$ cm^{-3}, and thus enables considerably more robust metallicity measurements. Moreover the [O III] 52μm line is comparable or brighter (with minimum flux ratio $\frac{f_{52}}{f_{88}} \geq 0.6$; Figure 2). Therefore we consider observations of the 52 μm line to be a highly desirable priority for future work. Unfortunately it is not widely available even for nearby galaxies, and hence our team is pursuing observations for a benchmark sample with SOFIA (the Stratospheric Observatory For Infrared Astronomy).

3. SOURCES OF UNCERTAINTY

The scatter in Figure 1 and case studies discussed in Section 2.3 underscore potential systematic errors arising from the limited data currently available for $z > 7$ galaxies. This section focuses on assessing the main sources of systematic uncertainty: nebular temperature T_e, density n_e, ionization correction factor (ICF), and H I line luminosity. Each of these contributes at the level of $\sim$0.2 dex. Our approach is to consider the plausible range of values for each parameter in the $z > 7$ sample, which we list in Table 3 along with the corresponding range of metallicity. We also tabulate the partial derivative of O/H with respect to each parameter, illustrating the sensitivity to moderate changes in the assumed values. This reveals the dominant sources of error and consequently *how to minimize uncertainty* for future studies. The effects of T_e and n_e are illustrated as an example in Figure 2. Below we discuss each parameter separately.

3.1. Nebular temperature T_e

Higher T_e results in lower metallicity, although with modest dependence. We consider a maximal range of $T_e(O^{++}) = 9{,}000$–20,000 K. This spans the full range of nebular T_e values seen in galaxy surveys (e.g., Izotov et al. 2006; Andrews & Martini 2013), hence we view this as highly conservative. Allowing values as low as 5,000 K would increase the range by only 0.07 dex above the nominal bound. Restricting to 11,000–20,000 K, the range in O/H would decrease to $\pm$0.1 dex. Thus the O^{++} abundance can be reasonably constrained even with completely unknown T_e.

3.2. Nebular density n_e

Higher n_e results in higher metallicity, due to collisional de-excitation suppressing the [O III] 88μm emission (Figure 2). Our lower bound of 10 cm^{-3} represents a strict limit; lower densities would change O/H by only $\sim$1%.

The upper bound of 600 cm^{-3} is the approximate maximum range found at $z \sim 2.3$ from the MOSDEF survey (from [O II] and [S II]; Sanders et al. 2016). Higher densities are inferred for some objects but not at a statistically significant level. Nonetheless, higher densities are possible (e.g., $n_e > 1000$ cm^{-3} in extreme $z \sim 3$ starbursts; Zhang et al. 2018) in which case true metallicities would be above the nominal bounds.

3.3. Ionization correction factor

Only the O^{++} ion is directly constrained with present data, representing a lower limit on total oxygen abundance. To estimate the ICF we first adopt a fiducial reddening-corrected $O_{32} = \frac{[\text{O III}]\lambda\lambda 4959,5007}{[\text{O II}]\lambda\lambda 3727,3729}$ flux ratio which is readily compared with observations. High [O III] 88μm fluxes and [O III]/[C II] ratios generally suggest high O_{32} for the $z > 7$ sample (e.g., Harikane et al. 2020). This is supported by observed mid-IR colors of $z > 6$ galaxies indicating extremely large [O III] equivalent widths (typically W$_{[\text{OIII}]4959,5007+\text{H}\beta} \approx 500-1500$ Å; e.g., Labbé et al. 2013; Smit et al. 2014, 2015; Laporte et al. 2014; Huang et al. 2016; Roberts-Borsani et al. 2016), confirmed in lower-redshift analogs (e.g., Mainali et al. 2020). Empirically such high equivalent widths suggest $O_{32} \simeq 2-10$ (accounting for scatter in the relation; Sanders et al. 2020). However this could be overestimated if there are significant Balmer breaks in the stellar spectra. Theoretical modeling likewise suggests large ionization parameters and high O_{32}. For example, Katz et al. (2019) predict $O_{32} \simeq 3-8$ for simulated massive galaxies at $z \simeq 10$.

Motivated by these expectations we adopt a fiducial $O_{32} = 3$, combined with $n_e = 250$ cm^{-3} and $T_e(\text{O}^+)$ based on an H II region relation:

$$T_e(\text{O}^+) = 0.7 \times T_e(\text{O}^{++}) + 3000 \text{ K}$$

(the "$T_2 - T_3$ relation"; Campbell et al. 1986; Stasińska 1982. An empirical alternative $T_e(\text{O}^+) = T_e(\text{O}^{++}) - 1300$ K from Andrews & Martini 2013 produces very similar results.). This gives 33% of oxygen in the O$^+$ state, or an ICF of $+0.17$ dex from O^{++} to total O abundance. As an approximate upper bound we take $O_{32} = 1$, $n_e = 600$ cm^{-3}, and $T_e(\text{O}^{++}) = 9,000$ K, giving ICF $= 0.39$ dex. As a lower bound we take $O_{32} = 10$, $n_e = 10$ cm^{-3}, and $T_e(\text{O}^{++}) = T_e(\text{O}^+) = 20,000$ K, giving ICF $= 0.04$ dex. These cases span the maximal ranges considered for T_e and n_e. However, lower O_{32} or $T_e(\text{O}^+)$ values are possible and would lead to higher O/H (e.g., for He 2-10 discussed in Section 2.3).

The above estimates consider only singly and doubly ionized oxygen. Neutral ions can be safely ignored for H II regions. The O^{3+} state may contribute in cases where O^{++}/O$^+$ is very large, but the expected effect is of order $\sim 1\%$ for the conditions described above (e.g., Izotov et al. 2006; Guseva et al. 2012). If anything, O^{3+} would contribute in more extreme cases of small ICF where O$^+$ has been underestimated, offsetting the overall effect. Hence we consider the ICF from O^{3+} and higher ionization states to be negligible.

3.4. Hydrogen recombination line luminosity

Higher H I recombination line luminosity implies higher H$^+$ abundance and lower O/H. We estimate uncertainty using the comparison of H I Balmer emission with photometrically-derived SFR by Shivaei et al. (2016), who find a scatter of $\sigma = 0.17-0.26$ dex for various data sets in their $z \sim 2$ sample. Therefore we adopt a somewhat conservative 0.25 dex systematic uncertainty in H I line luminosity (e.g., Hβ). This is comparable in many cases to statistical uncertainty in SFR of the $z > 7$ sample, whose spectral energy distributions (SEDs) are less well sampled than galaxies at lower redshift.

3.5. Total systematic uncertainty

We now assess the total systematic uncertainty, noting that the various contributions are not fully independent. Higher T_e decreases both O^{++}/H$^+$ and the ICF toward lower O/H. Higher n_e increases O^{++}/H$^+$ and slightly increases the ICF. T_e also correlates with n_e (i.e., higher overall pressure), such that their combined effects may lead to somewhat lower systematic uncertainty in O/H. Overall we estimate the combined uncertainty from T_e, n_e, and ICF as $\simeq 0.3$ dex considering their covariances. Scatter in the SFR–$L(\text{H}\beta)$ relation has no clear connection to the other parameters and we treat it as independent. In combination we arrive at an estimate of $\sigma_{sys}(\log(\text{O/H})) = 0.4$ dex for the *total systematic error* in O/H for the $z > 7$ sample. We reiterate that this estimate conservatively spans a broad range of physical parameters (e.g., $T_e = 9,000-20,000$ K). For comparison, current measurements based on [O III]$\lambda 4363$ at $z \gtrsim 2$ have typical statistical uncertainties of ~ 0.2 dex in O/H (e.g., Sanders et al. 2020), comparable to the scatter from applying our method to the DGS sample (Figure 1). In our view it is remarkable that such precision can be achieved at $z > 7$ solely from [O III] 88μm flux and a photometry-based estimate of the SFR.

4. CHEMICAL ENRICHMENT AT $Z \simeq 8$

Having verified our methodology and assessed the uncertainties, we now discuss our results for metallicity of the $z > 7$ sample. First we consider timescales required for enrichment and implications for extended star formation histories, followed by a discussion of the mass-metallicity relation and its evolution to the present day. We then comment on spatially resolved structure in the sample.

4.1. Enrichment and star formation timescales

The metallicity estimates in Table 2 range from $\sim \frac{1}{20}$ to $\frac{1}{3}$ the solar value, albeit with uncertainty spanning ~ 1 dex. The average is $12 + \log(\text{O/H}) = 7.94$ or ~ 0.2 solar. Notably in two cases the 1-σ lower bounds are $\gtrsim 0.1$ solar, indicating a significant degree of chemical enrichment only ~ 650 Myr after the big bang. We can place an approximate lower bound on the enrichment timescale by considering a closed-box chemical evolution model, with an estimated gas depletion time $\tau \approx 300$ Myr at these redshifts (e.g., Scoville et al. 2017). For a yield of approximately solar abundance (e.g., Woosley & Weaver 1995), a metallicity of 0.1 solar requires that $\sim 10\%$ of gas in these systems has already been processed into stars, requiring ~ 30 Myr of constant star formation. The enrichment timescale could be reduced if oxygen yields are

higher, as may be expected at low metallicity due to stellar evolution effects and possibly a top-heavy IMF (e.g., Jeřábková et al. 2018). With yields from low-metallicity stars $y_O = 0.007$–0.039, as tabulated by Vincenzo et al. (2016, for a Chabrier IMF of different mass cutoffs), an abundance of 0.1 solar is reached in only 6–33 Myr for a closed-box model. However, gas inflows and outflows would increase the time required to enrich interstellar gas. Both effects are expected to be prevalent given the sample's high SFRs. As an example, Langan et al. (2020) report that gas flows reduce effective yields by a factor of 10 in simulated $z \approx 8$ galaxies. This drives the timescale to ∼100–300 Myr to reach enrichment of ∼0.1 solar.

We thus view 100 Myr as a reasonable order-of-magnitude estimate of the timescale of past star formation in the $z > 7$ sample, in order to explain the average metallicities. Ages of ∼10 Myr are plausible only in the absence of gas flows, which we view as an unlikely scenario, or for the lower-metallicity galaxies. This contrasts with ages $\lesssim 10$ Myr inferred for most of the sample from single stellar population SED fits. Our chemical evolution results thus indicate that the true stellar mass is likely dominated by an extended period of star formation which is not captured by single-component photometric models (e.g., Hashimoto et al. 2018; Roberts-Borsani et al. 2020), especially given limited available data beyond the rest-frame ultraviolet (UV). These results are generally consistent with previous metallicity estimates, for example of SXDF-NB1006-2 (0.05–1 solar; Inoue et al. 2016) and MACS0416-Y1 (∼0.2 solar; Tamura et al. 2019), where in both cases the authors conclude that dust and metal enrichment requires the presence of an evolved underlying stellar population.

A key diagnostic in determining the extent of previous star formation is the strength of the stellar Balmer break. However, this feature is challenging to distinguish from strong nebular emission in photometry of the $z > 7$ sample. Red mid-IR 3.6μm - 4.5μm "IRAC excess" colors measured with the Spitzer Space Telescope can be explained by very young stellar populations with strong nebular emission lines, or by more extended star formation histories with weaker emission lines and a prominent Balmer break (permitting ∼30 times higher stellar mass in the latter case; Roberts-Borsani et al. 2020).

The strongest indications of extended star formation based on chemical evolution are for MACS1149-JD1 and B14-65666, where our method indicates relatively high metallicity. The lower bound on metallicity is $\gtrsim 0.1$ solar in both cases, considering both systematic and measurement uncertainties. Such enrichment suggests previous star formation lasting several tens of Myr or more as discussed above. This result bolsters previous interpretation of MACS1149-JD1 where the IRAC excess cannot be explained by line emission, clearly favoring a Balmer break (with stellar population age $\gtrsim 100$ Myr; Hashimoto et al. 2018).

In the case of B14-65666, the photometric data are consistent with a young starburst of age $\lesssim$10 Myr and strong nebular emission (Hashimoto et al. 2019), but the data also permit extended star formation with a moderate Balmer break (Roberts-Borsani et al. 2020). Such a young single-starburst age would imply extremely high effective yields to reach the metallicity we derive. A more likely scenario in our view is that the gas may have been pre-enriched by previous generations of star formation, which do not necessarily dominate the observed SED nor the IRAC excess, but can nonetheless dominate the total stellar mass and chemical enrichment. This is demonstrated by the 2-component SED fit of Roberts-Borsani et al. (2020): a young (3 Myr) starburst contributes most of the broadband photometric flux, while an extended star formation component constitutes 92% of the stellar mass (and we note that this model provides the best fit to the data based solely on log-likelihood).

Thus our metallicity estimates provide additional constraints with which to determine the star formation histories of high-z galaxies, despite some ambiguity in photometric analyses. This supports the general results of Roberts-Borsani et al. (2020) that current near- and mid-IR data permit a broad range of star formation histories and stellar masses, which are not captured by single-component models. While JWST spectroscopy will directly resolve this issue, we conclude that chemical enrichment constraints from ALMA are valuable in support of interpreting stellar population synthesis models.

4.2. The $z \simeq 8$ mass-metallicity relation

We show the mass-metallicity relation (MZR) in Figure 3 for two different sets of stellar masses reported in the literature (summarized in Table 4). Masses in the left panel are based on fits which suggest ages $\lesssim 10$ Myr for most sources. It is perhaps better to treat these as lower limits following the discussion above (e.g., Inoue et al. 2016). The right panel shows the mean and range of masses for two limiting cases from Roberts-Borsani et al. (2020), where allowing for extended star formation histories can increase the stellar mass by a factor of $\sim 30\times$. (We note that SXDF-NB1006-2 was not included in the sample of Roberts-Borsani et al. 2020 due to insufficient mid-IR photometry, and is not shown in the right panel of Figure 3.) Here we consider the implications for metallicity. As discussed above, the young and low-mass case generally requires very large [O III]$\lambda\lambda 4959,5007$ fluxes to reproduce observed mid-IR colors. Such strong line emission implies high T_e and O_{32} (hence lower ICF) compared to our fiducial values (Sections 2 and 3). As a result our metallicity estimates would be revised downward by ∼0.2 dex. The high-mass extreme corresponds to an opposite effect with much weaker $W_{[OIII]4959,5007+H\beta} \lesssim 200$ Å, such that metallicity estimates would be revised upward. Since the low/high mass cases would imply lower/higher metallicity, even ±0.8 dex in mass would shift the sample by $\lesssim 0.2$ dex in O/H relative to the theoretical mass-metallicity relation of Ma et al. (2016). In this sense the uncertainty in MZR evolution is plausibly smaller than uncertainty in metallicity alone. Otherwise this mass range would propagate to ±0.3 dex in MZR evolution for the conservative case of treating mass and metallicity independently. Our interpretation of MZR evolution from the right panel of Figure 3 is therefore not strongly affected by the large uncertainties in stellar mass.

Although the current sample is insufficient to measure the MZR slope at $z \simeq 8$, it provides a useful basis to examine redshift evolution and compare with theoretical work. In this sense the relevant quantity is the (unweighted) mean $12 + \log(O/H) = 7.9$, at stellar mass $M_* \sim 10^9 - 10^{9.5} M_\odot$. The comparison with T_e-based
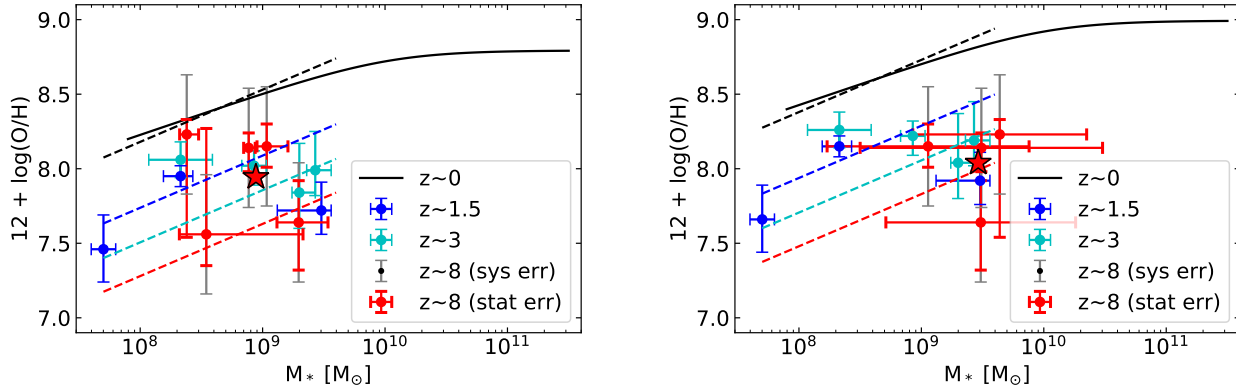
FIG. 3.— Mass-metallicity relation of $z > 7$ galaxies, compared with $z \simeq 1.5$–3.5 galaxies with [O III]$\lambda4363$-based measurements listed in Table 4, and the $z \simeq 0$ relation from stacked spectra (black line; Curti et al. 2020). Theoretical simulation-based results from Ma et al. (2016) are shown as dashed lines with the same redshift color-coding ($z = 0$, 1.5, 3, and 8).
Left: Mass-metallicity relation with original stellar mass values reported in the literature. *Right:* Stellar masses for the $z > 7$ sample from Roberts-Borsani et al. (2020, summarized in Table 4): error bars in mass span the 1- and 2-component SED fit results, while data points represent the midpoint of these two cases. Metallicities are from this work (Table 2) and are the same in both panels. The galaxy SXDF-NB1006-2 is not shown in the right panel as it was not included in the sample of Roberts-Borsani et al. (2020), due to insufficient mid-IR photometry. Metallicities from the literature are increased by 0.2 dex in the right panel, showing the approximate effect of temperature fluctuations which can systematically underestimate [O III]$\lambda4363$-based values. The abundance scale is such that the most massive $z \simeq 0$ galaxies have moderately super-solar metallicities (by ∼0.1 and 0.3 dex in the left and right panels, respectively, where we adopt solar $12 + \log(\text{O/H}) = 8.69$ from Asplund et al. 2009). In both panels, red error bars represent statistical measurement uncertainty for the $z > 7$ sample, grey error bars show the estimated 0.4 dex systematic uncertainty in O/H, and red stars are the $z > 7$ unweighted mean.

metallicities at lower redshift is also subject to systematic uncertainty arising from the [O III]$\lambda4363$ diagnostic, which exhibits an ADF of ∼0.2 dex compared to other methods (as noted in Section 2, e.g., Esteban et al. 2009; Blanc et al. 2015). Hence in the right panel of Figure 3 we increase all lower-redshift values by 0.2 dex in O/H. This increase brings the local MZR to $12 + \log(\text{O/H}) \approx 9.0$ for massive galaxies ($M_* \gtrsim 10^{10.5}$ $M_\odot$), closer to most results in the literature (e.g., Tremonti et al. 2004; Kewley & Ellison 2008; Mannucci et al. 2010). We note that joint analyses of optical and far-IR lines in $z \simeq 0$ galaxies are needed to fully address this discrepancy. At fixed M_*, the resulting evolution in O/H from $z = 8 \rightarrow 0$ is ∼0.6 and 0.9 dex for the two panels in Figure 3. We consider the latter to be a more appropriate relative comparison, and thus estimate the evolution as 0.9 ± 0.5 dex in O/H accounting conservatively for systematic uncertainty in both metallicity and M_*.

Several groups have reported theoretical predictions for the mass-metallicity relation at high redshifts based on cosmological simulations. Bright $z \simeq 8$ galaxies near the mass range of our sample are predicted to have metallicities near ∼0.1 solar (Moriwaki et al. 2018; Katz et al. 2019; Langan et al. 2020), with evolution in the MZR to $z = 8$ of ∼0.5–0.9 dex (Figure 3; Ma et al. 2016; Torrey et al. 2019). These predictions are compatible and in fact in good agreement with our results. A potentially interesting point of divergence is the MZR evolution from $z \simeq 4-8$, where predictions range from ∼0–0.3 dex (Langan et al. 2020), reflecting differences in galaxy gas content and metal-enriched outflows. Precise metallicities from future JWST and ALMA data (Section 5) can usefully test these models. Chemical evolution studies can therefore address the formative processes of gas accretion and stellar feedback in reionization-era galaxies.

4.3. *Spatially resolved metallicity*

The sensitivity and angular resolution of ALMA (and JWST) permits resolved metallicity measurements on ∼1 kpc scales at $z > 7$, using the approach outlined herein. Resolved structure is already evident with ALMA in many cases (e.g., Hashimoto et al. 2019; Smit et al. 2018; Matthee et al. 2017). Such information can provide considerable insight into the earliest phases of galaxy assembly, for example regarding in-situ growth versus hierarchical merging. Distinguishing mergers from disk galaxies can be challenging even with high-quality kinematic data (e.g., Simons et al. 2019), while resolved metallicity can be a powerful discriminant (e.g., Wang et al. 2017, 2020). For in-situ galaxy growth, spatial variations in metallicity provide insight into gas mixing and feedback processes (e.g., Ma et al. 2017; Jones et al. 2013; Wang et al. 2019).

BDF-3299 is a compelling example of resolved structure in our sample. Our fiducial case gives limits $12 + \log(\text{O/H}) < 7.9$ and > 8.1 for the UV- and IR-bright components respectively. Likewise B14-65666 is resolved into two distinct clumps, where luminosities reported by Hashimoto et al. (2019) suggest that clump "A" is more metal-rich by a few tenths of a dex than clump "B" for our fiducial case. These nominal differences could indicate mergers in both systems, where the more enriched component likely dominates the stellar mass. Although the metallicities of each component are not distinguishable given the uncertainties, it illustrates the prospects of resolving chemical enrichment and metal mixing at $z > 7$. This is a truly remarkable possibility in our view, especially as no such resolved T_e measurements have yet been achieved beyond the relatively nearby universe.

5. REMOVING SYSTEMATIC UNCERTAINTIES: <0.2 DEX PRECISION WITH JWST AND ALMA

Having discussed implications of our measurements for chemical enrichment in the reionization epoch, we now

consider prospects for improved precision with future JWST and ALMA data. While prospects for stellar mass and strong-line abundance methods have been widely recognized (e.g., Roberts-Borsani et al. 2020), here we focus on T_e-based metallicity using the far-IR emission lines and [O III] 52μm in particular.

The results of Section 3 provide both an assessment of the dominant uncertainties and guidance on how to minimize them. Each source of systematic uncertainty can be reduced or eliminated with a modest number of emission line measurements, summarized in Table 3. We frame the discussion below in approximate order of the largest to smallest sources of uncertainty, leaving ICF for last as it depends on T_e and n_e.

1. $L(H\beta)$: JWST can provide both Hβ flux and the correction for dust attenuation (from ratios such as Hα/Hβ or Paα/Hβ). Attenuation correction is also critical for T_e (via [O III]$\lambda\lambda$4959,5007) and ICF (via [O II]$\lambda\lambda$3727,3729). We note that Paschen series lines are accessible to JWST/MIRI at $z > 7$ and are more robust than Balmer lines to the dust attenuation curve. The Paα feature is especially promising, having approximately 10$\times$ lower attenuation than Hβ.

2. n_e: The most promising approach in our view is to measure [O III] 52μm, accessible in the 450 μm atmospheric window with ALMA (Band 9) for $z \simeq 7.1-8.6$. With $n_{crit} = 3470\,\mathrm{cm}^{-3}$, [O III] 52μm is nearly independent of density for normal H II region conditions, while the [O III] 52μm/[O III] 88μm ratio provides the density $n_e(O^{++})$. Figure 2 illustrates the clear advantage of [O III] 52μm for robust O/H measurements. Density diagnostics such as the optical [S II] and [O II] doublets are also accessible to JWST, although they do not directly probe the O^{++} ionization state.

3. T_e: [O III]λ5007 and/or λ4959 luminosity provides T_e in combination with the far-IR lines, ideally [O III] 52μm. We view the [O III]λ4363 temperature diagnostic as far less pragmatic than [O III] 52μm,88μm given its low expected flux (see Section 5.3).

4. ICF (O^+ and O^{3+} abundance): [O II]$\lambda\lambda$3727,3729 is the most accessible feature to address the ICF, although we consider some lingering uncertainty from unknown $T_e(O^+)$. A range $\pm$1000 K ($\pm$3000 K) in $T_e(O^+)$ corresponds to only $\pm$0.03 dex ($\pm$0.1 dex) in $\log(O/H)$ for the fiducial conditions. The magnitude of ICF uncertainty is related to [O II] flux, such that higher O_{32} translates to smaller uncertainty. O^{3+} is expected to contribute minimally ($\sim$1%; Section 3.3), although it should be reconsidered in cases where O^+/O^{++} is very low ($\lesssim 0.05$).

The four points above outline a clear path toward removing systematic uncertainties using a modest set of strong emission lines: [O III] 52μm, [O III]λ5007, Hβ, other H I lines such as Paα and/or Hα, and [O II]$\lambda\lambda$3727,3729 (Table 3). To avoid aperture correction uncertainties, rest-optical fluxes are best determined with grism (WFSS: wide field slitless spectroscopy) or integral field spectroscopy (IFS) modes of JWST. With these features, systematic uncertainty is limited by the ICF and estimated to be $\lesssim$0.05-0.1 dex in O/H.

5.1. *Feasibility of JWST and ALMA observations*

In this section we comment on expected emission line fluxes and detection feasibility. H I lines will likely present the greatest challenge for JWST. For the brighter sources B14-65666 and SXDF-NB1006-2, with a conservative $A_V = 1$ magnitude of dust extinction, the luminosity estimates from Table 2 translate to expected fluxes $f_{H\beta} \simeq 9 \times 10^{-18}$ ergs s^{-1} cm^{-2}, $f_{H\alpha} \simeq 3 \times 10^{-17}$, and $f_{Pa\alpha} \simeq 7\times10^{-18}$ ergs s^{-1} cm^{-2}. Hβ and Hα would be $\sim$2 times stronger for estimates of $A_V \lesssim 0.3$ (Hashimoto et al. 2019; Inoue et al. 2016), while Paα is relatively insensitive to A_V. Thus the brighter objects are expected to have H I line fluxes $\gtrsim 10^{-17}$ ergs s^{-1} cm^{-2}. [O III]λ5007 is likely to be stronger by a factor of a few, based on both empirical and theoretical estimates (e.g., Roberts-Borsani et al. 2020; Katz et al. 2019). [O II]$\lambda\lambda$3727,3729 may be weaker than Hβ, but uncertainty in [O II] flux does not necessarily limit the precision in O/H (Section 5.2; Belli et al. 2013). JWST IFS and WFSS sensitivity estimates suggest that all of these key features can be detected at $\gtrsim 10\sigma$ significance within a few hours or less, for the brighter sources. We note that grism spectroscopy offers comparable or superior spatial resolution and sensitivity over a far larger field of view than IFS, which may even be useful for multiplexing of clustered $z > 7$ sources. JWST's grism modes thus provide a powerful method to probe the relevant nebular emission lines (e.g. [O II], Hβ, and [O III]λ5007 at $z > 7$ with NIRCam WFSS).

Finally, the key [O III] 52μm line is comparable or brighter than [O III] 88μm (depending on n_e; Figure 2). The minimum flux ratio is $\frac{f_{52}}{f_{88}} = 0.6$ for low densities, increasing to 1.8 for $n_e = 600$ cm^{-3} (and 1.1 for the fiducial case $n_e = 250$ cm^{-3}). The main challenge is that [O III] 52μm falls in ALMA's high-frequency bands, requiring excellent conditions and of order $\sim$10-hour integrations for the brightest objects. From the sample in Table 2, those at $z < 8$ have [O III] 52μm at relatively accessible frequencies. We note that compact ALMA array configurations are optimal for this work, but are typically available only during poor-weather months, with little time suitable for high-frequency observations. Scheduling compact configurations during months with reliable high-frequency weather (e.g. May through August) would greatly improve the feasibility of observing [O III] 52μm at high redshifts with ALMA.

5.2. *Forecasted precision in oxygen abundance*

To illustrate the potential of combining rest-optical and far-IR emission lines, we perform a calculation of T_e, n_e, and $12 + \log(O/H)$ with full error propagation. We adopt 10% flux uncertainty for Hβ, Paα, and [O III] 88μm; 20% flux uncertainty for [O III] 52μm and [O II]$\lambda\lambda$3727,3729; and $\pm$1000 K uncertainty in $T_e(O^+)$ from the $T_2 - T_3$ relation (Section 3.3). Line fluxes are chosen to give the fiducial $T_e = 1.5 \times 10^4$ K, $n_e = 250$ cm^{-3}, ICF = 0.17 dex, with 0.15 solar abundance and $A_V = 1$ magnitude (although varying A_V does not affect the results). The total result is $12 + \log(O/H) =$

7.88 ± 0.17, with [O III] 52μm dominating the error budget (±0.15 dex). The other substantial error source is Paα (±0.07; the same precision is obtained for Hα with 3% flux uncertainty). Hβ, [O II]$\lambda\lambda$3727,3729, and T$_e$(O$^+$) each contribute ±0.03 while [O III] 88μm contributes negligibly (< 0.01 dex, illustrating mild effects of n$_e$). [O III]λ5007 likewise contributes negligible error, assuming a flux sensitivity comparable to Hβ. Increasing the flux uncertainty to 20% for all lines results in ±0.23 dex uncertainty in O/H.

We conclude that signal-to-noise ratios (SNR) $\gtrsim$ 5 with good flux calibration are sufficient to achieve $\lesssim$ 0.2 dex precision in O/H. The limiting factors above are [O III] 52μm and Paα, which anchor the O^{++} and H$^+$ abundances. The latter can be reasonably improved with higher sensitivity and multiple H I lines (e.g. simultaneously measured with JWST/MIRI-MRS). [O III] 52μm is likely to be limited by flux calibration accuracy of ∼10%, corresponding to a best-case precision of ±0.08 dex achievable with deep ALMA observations.

5.3. *Comparison of [O III] 52μm and [O III]λ4363*

In Section 2.4 we argue that the far-IR offers considerable advantages over optical auroral lines at high redshifts. For comparison, we repeat the analysis of Section 5.2 using only rest-optical emission lines, with T$_e$ based on [O III]λ4363. For the same physical properties, a flux uncertainty of 25% in [O III]λ4363 (i.e., 4σ detection) propagates to $12 + \log{(O/H)} = 7.88^{+0.17}_{-0.12}$ assuming infinite SNR in all other lines. This precision is comparable to the case outlined in Section 5.2. However, [O III]λ4363 is 12 times weaker than Hβ and 60 times weaker than [O III]λ5007 for our fiducial conditions (e.g., T$_e$ = 1.5 × 10^4 K). Hence achieving SNR=4 in [O III]λ4363 requires $\gtrsim$10 times better sensitivity at rest-optical wavelengths (∼100× longer integration times with JWST!) compared to achieving SNR=5 in Hβ, and may be even more challenging if T$_e$ is lower. We therefore find the [O III] 52μm approach with ALMA to be far more pragmatic.

6. DISCUSSION

We present a simple method to determine gas-phase oxygen abundance from far-IR emission lines and SFR measurements, and apply it to a sample of galaxies at $z = 7.1 - 9.1$.[2] We find typical $12 + \log{(O/H)} = 7.9$ or ∼0.2 times the solar abundance. We determine evolution in the MZR of 0.9 ± 0.5 dex in O/H from $z = 8 \to 0$ at fixed stellar mass, accounting conservatively for systematic uncertainties in both mass and metallicity. The results reveal a substantial degree of chemical enrichment in reionization-era galaxies, indicating that star formation has likely been ongoing for ∼100 Myr in much of the sample (supporting multi-component studies of stellar population properties; e.g., Roberts-Borsani et al. 2020). The underlying evolved stellar populations are not cap-

tured by single-component photometric fits, yet can easily dominate the stellar mass and integrated number of UV photons produced by these galaxies. This in turn is important for understanding the cosmic reionization process at $z > 7$.

Notably, Roberts-Borsani et al. (2020) constrain the extent of older stellar populations largely using [O III] 88μm and far-IR dust continuum. These ALMA measurements enable modeling of how nebular emission and stellar continuum breaks contribute to observed mid-IR photometry. Additional higher-frequency ALMA data can help to constrain both the dust mass and nebular emission (as [O III] 52μm would eliminate the factor of 3× uncertainty from the range of densities considered herein, compared to [O III] 88μm). While JWST spectroscopy presents a clear path forward, further ALMA observations are a promising intermediate step to refine estimates of stellar population properties.

We have discussed prospects for improving metallicity measurements with future JWST and ALMA data, concluding that 0.1–0.2 dex precision is feasible with modest integration times for the best $z > 7$ targets. The [O III] 52μm emission line is highlighted as an important feature for securing robust T$_e$ and abundance measurements. While challenging high-frequency observations are required, we nonetheless argue that [O III] 52μm is far more accessible than the [O III]λ4363 T$_e$ diagnostic. While our discussion focuses on $z > 7$ galaxies, the same methods can be applied at lower $z = 5.2$–6.3 where [O III] 52μm is redshifted into ALMA's Band 10. ALMA's sensitivity to fixed [O III] 52μm luminosity is approximately equal in Band 9 ($z \approx 7$–8) as in Band 10, as the ∼ 2× higher noise is compensated by lower luminosity distance. Intrinsically luminous (or gravitationally lensed) $z \simeq 6$ galaxies are thus detectable within a few hours with ALMA in high-frequency weather. As noted in Section 5, these measurements would ideally make use of future ALMA cycles with compact arrays scheduled during seasons with reliable high-frequency weather.

Future abundance measurements can provide remarkable insight into the early assembly history of galaxies. Knowledge of metallicity will improve photoionization models used to interpret high-redshift emission lines, including inferences on the stellar ionizing spectrum. Metallicity also serves as a constraint on integrated enrichment from past star formation history, especially in combination with improved stellar mass determinations. Star formation histories and the stellar ionizing spectrum in turn have important implications for the timeline of cosmic reionization and the role of star forming galaxies. Additionally, measuring evolution in the mass-metallicity relation from $z \simeq 5$–8 will test different theoretical models and provide insight into the gas content, accretion rates, and stellar feedback which drive early galaxy evolution. Current theoretical models vary by ∼0.3 dex of evolution in O/H, which can be distinguished with modest samples using the techniques described herein. Finally, spatially resolved metallicity measurements may be feasible for bright and moderately extended sources, offering a useful probe of gas mixing processes and hierarchical merger assembly. JWST is undoubtedly poised to revolutionize our knowledge of reionization-era galaxies, and we view ALMA observations of far-IR [O III] lines as an extremely valuable com-

---

[2] Following the submission of this work, Yang & Lidz (2020) presented an analytic model for [O III] fine structure emission based on similar physical arguments. They verify their model with CLOUDY calculations, and reproduce our metallicity estimates from Equation 2 to within 0.03 dex, deriving consistent metallicities for the sample discussed herein. This provides further validation of our methodology with an independent framework.

ponent for chemical evolution studies.

ACKNOWLEDGEMENTS

REFERENCES

Andrews, B. H., & Martini, P. 2013, ApJ, 765, 140
Asplund, M., Grevesse, N., Sauval, A. J., et al. 2009, ARA&A, 47, 481
Belli, S., Jones, T., Ellis, R. S., & Richard, J. 2013, ApJ, 772, 141
Blanc, G. A., Kewley, L., Vogt, F. P. A., et al. 2015, ApJ, 798, 99
Bouwens, R. J., Stefanon, M., Oesch, P. A., et al. 2019, ApJ, 880, 25
Calvi, V., Trenti, M., Stiavelli, M., et al. 2016, ApJ, 817, 120
Campbell, A., Terlevich, R., & Melnick, J. 1986, MNRAS, 223, 811
Carniani, S., Maiolino, R., Pallottini, A., et al. 2017, A&A, 605, A42
Christensen, L., Laursen, P., Richard, J., et al. 2012, MNRAS, 427, 1973
Cormier, D., Madden, S. C., Lebouteiller, V., et al. 2015, A&A, 578, A53
Croxall, K. V., Smith, J. D., Brandl, B. R., et al. 2013, ApJ, 777, 96
Cullen, F., McLure, R. J., Dunlop, J. S., et al. 2019, MNRAS, 487, 2038
Curti, M., Mannucci, F., Cresci, G., et al. 2020, MNRAS, 491, 944
Ellis, R. S., McLure, R. J., Dunlop, J. S., et al. 2013, ApJ, 763, L7
Esteban, C., Bresolin, F., Peimbert, M., et al. 2009, ApJ, 700, 654
Faisst, A. L., Capak, P. L., Davidzon, I., et al. 2016, ApJ, 822, 29
Furusawa, H., Kashikawa, N., Kobayashi, M. A. R., et al. 2016, ApJ, 822, 46
Gburek, T., Siana, B., Alavi, A., et al. 2019, ApJ, 887, 168
Guseva, N. G., Izotov, Y. I., Fricke, K. J., et al. 2012, A&A, 541, A115
Harikane, Y., Ouchi, M., Inoue, A. K., et al. 2020, ApJ, 896, 93
Hashimoto, T., Laporte, N., Mawatari, K., et al. 2018, Nature, 557, 392
Hashimoto, T., Inoue, A. K., Mawatari, K., et al. 2019, PASJ, 71, 71
Huang, K.-H., Bradač, M., Lemaux, B. C., et al. 2016, ApJ, 817, 11
Inoue, A. K., Tamura, Y., Matsuo, H., et al. 2016, Science, 352, 1559
Izotov, Y. I., Thuan, T. X., & Lipovetsky, V. A. 1994, ApJ, 435, 647
Izotov, Y. I., Stasińska, G., Meynet, G., Guseva, N. G., & Thuan, T. X. 2006, A&A, 448, 955
Izotov, Y. I., Thuan, T. X., & Guseva, N. G. 2014, MNRAS, 445, 778
Jeřábková, T., Hasani Zonoozi, A., Kroupa, P., et al. 2018, A&A, 620, A39
Jones, T., Ellis, R. S., Richard, J., & Jullo, E. 2013, ApJ, 765, 48
Jones, T., Martin, C., & Cooper, M. C. 2015, ApJ, 813, 126
Katz, H., Galligan, T. P., Kimm, T., et al. 2019, MNRAS, 487, 5902
Kennicutt, R. C. 1998, ARA&A, 36, 189
Kewley, L. J., & Ellison, S. L. 2008, ApJ, 681, 1183
Komatsu, E., Smith, K. M., Dunkley, J., et al. 2011, ApJS, 192, 18
Labbé, I., Oesch, P. A., Bouwens, R. J., et al. 2013, ApJ, 777, L19
Lagos, P., Scott, T. C., Nigoche-Netro, A., et al. 2018, MNRAS, 477, 392
Langan, I., Ceverino, D., & Finlator, K. 2020, MNRAS, 494, 1988
Laporte, N., Streblyanska, A., Clement, B., et al. 2014, A&A, 562, L8
Laporte, N., Ellis, R. S., Boone, F., et al. 2017, ApJ, 837, L21
Laporte, N., Katz, H., Ellis, R. S., et al. 2019, MNRAS, 487, L81
López-Sánchez, Á. R., Dopita, M. A., Kewley, L. J., et al. 2012, MNRAS, 426, 2630

Luridiana, V., Morisset, C., & Shaw, R. A. 2012, IAU Symposium, 422
Ly, C., Malkan, M. A., Nagao, T., et al. 2014, ApJ, 780, 122
Ma, X., Hopkins, P. F., Faucher-Giguère, C.-A., et al. 2016, MNRAS, 456, 2140
Ma, X., Hopkins, P. F., Feldmann, R., et al. 2017, MNRAS, 466, 4780
Madden, S. C., Rémy-Ruyer, A., Galametz, M., et al. 2013, PASP, 125, 600
Mainali, R., Stark, D. P., Tang, M., et al. 2020, MNRAS, 494, 719
Maiolino, R., Carniani, S., Fontana, A., et al. 2015, MNRAS, 452, 54
Maiolino, R., & Mannucci, F. 2019, A&A Rev., 27, 3
Mannucci, F., Cresci, G., Maiolino, R., et al. 2010, MNRAS, 408, 2115
Mason, C. A., Treu, T., Dijkstra, M., et al. 2018, ApJ, 856, 2
Matthee, J., Sobral, D., Boone, F., et al. 2017, ApJ, 851, 145
Moriwaki, K., Yoshida, N., Shimizu, I., et al. 2018, MNRAS, 481, L84
Oesch, P. A., Bouwens, R. J., Illingworth, G. D., et al. 2018, ApJ, 855, 105
Palay, E., Nahar, S. N., Pradhan, A. K., et al. 2012, MNRAS, 423, L35
Pilyugin, L. S. & Thuan, T. X. 2005, ApJ, 631, 231
Roberts-Borsani, G. W., Bouwens, R. J., Oesch, P. A., et al. 2016, ApJ, 823, 143
Roberts-Borsani, G. W., Ellis, R. S., & Laporte, N. 2020, MNRAS, 497, 3440
Sanders, R. L., Shapley, A. E., Kriek, M., et al. 2016, ApJ, 816, 23
Sanders, R. L., Shapley, A. E., Reddy, N. A., et al. 2020, MNRAS, 491, 1427
Sargsyan, L. A. & Weedman, D. W. 2009, ApJ, 701, 1398
Scoville, N., Lee, N., Vanden Bout, P., et al. 2017, ApJ, 837, 150
Shapley, A. E., Sanders, R. L., Reddy, N. A., et al. 2017, ApJ, 846, L30
Shivaei, I., Kriek, M., Reddy, N. A., et al. 2016, ApJ, 820, L23
Simons, R. C., Kassin, S. A., Snyder, G. F., et al. 2019, ApJ, 874, 59
Smit, R., Bouwens, R. J., Labbé, I., et al. 2014, ApJ, 784, 58
Smit, R., Bouwens, R. J., Franx, M., et al. 2015, ApJ, 801, 122
Smit, R., Bouwens, R. J., Carniani, S., et al. 2018, Nature, 553, 178
Stark, D. P. 2016, ARA&A, 54, 761
Stasińska, G. 1982, A&AS, 48, 299
Strait, V., Bradač, M., Coe, D., et al. 2020, ApJ, 888, 124
Strom, A. L., Steidel, C. C., Rudie, G. C., et al. 2017, ApJ, 836, 164
Tamura, Y., Mawatari, K., Hashimoto, T., et al. 2019, ApJ, 874, 27
Torrey, P., Vogelsberger, M., Marinacci, F., et al. 2019, MNRAS, 484, 5587
Totani, T., Kawai, N., Kosugi, G., et al. 2006, PASJ, 58, 485
Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, ApJ, 613, 898
Vincenzo, F., Matteucci, F., Belfiore, F., et al. 2016, MNRAS, 455, 4183
Wang, X., Jones, T. A., Treu, T., et al. 2017, ApJ, 837, 89
Wang, X., Jones, T. A., Treu, T., et al. 2019, ApJ, 882, 94
Wang, X., Jones, T. A., Treu, T., et al. 2020, arXiv e-prints, arXiv:1911.09841
Woosley, S. E., & Weaver, T. A. 1995, ApJS, 101, 181
Yang, S. & Lidz, A. 2020, arXiv:2007.14439
Zhang, Z.-Y., Ivison, R. J., George, R. D., et al. 2018, MNRAS, 481, 59

TABLE 1
PHYSICAL PROPERTIES OF THE HERSCHEL DGS SAMPLE

Object name	Distance[a] [Mpc]	12+log(O/H)[a] (PT05)	12+log(O/H)[b] (I06)	M_*[a] [10^8 M$_\odot$]	SFR[a] [M$_\odot$ yr^{-1}]	[O III] 88μm flux[c] [10^{-18} W m^{-2}]	L([O III] 88μm)[d] [10^7 L$_\odot$]	12+log(O/H)[e] (Eq. 2)
Haro 11	92.1	8.36	8.23	339.0	28.6	1720.0	45.62	8.11
Haro 2	21.7	8.23	—	20.8	0.9	972.0	1.432	8.11
Haro 3	19.3	8.28	8.37	18.8	0.8	1850.0	2.156	8.34
He 2-10	8.7	8.43	—	18.2	0.79	3380.0	0.8003	7.91
HS 1222+3741	181.7	7.79	7.83	14.6	1.507	14.6	1.508	7.91
HS 1304+3529	78.7	7.93	7.66	4.2	0.5	31.9	0.618	8.00
HS 1330+3651	79.7	7.98	7.66	6.3	0.382	29.8	0.5923	8.09
II Zw 40	12.1	8.23	8.09	4.7	0.43	3590.0	1.645	8.48
Mrk 1450	19.8	7.84	8.03	0.86	0.47	262.0	0.3213	7.74
Mrk 930	77.8	8.03	8.09	44.2	3.12	421.0	7.973	8.31
NGC 1140	20.0	8.38	8.27	23.4	0.57	1080.0	1.352	8.28
NGC 5253	4.0	8.25	8.16	6.7	0.24	9010.0	0.451	8.18
Pox 186	18.3	7.70	7.75	0.1	0.04	33.7	0.0353	7.85
SBS 1159+545	57.0	7.44	7.46	0.5	0.17	4.81	0.0489	7.36
UGC 4483	3.2	7.46	7.55	0.03	0.0007	25.7	0.000823	7.98
UM 133	22.7	7.82	7.70	0.81	0.02	24.5	0.0395	8.20
UM 448	87.8	8.32	8.03	241.0	12.36	1060.0	25.56	8.22
UM 461	13.2	7.73	7.79	0.26	0.01	79.0	0.04306	8.54
VII Zw 403	4.5	7.66	7.74	0.1	0.003	81.6	0.00517	8.14

[a] As reported in Table 2 of Madden et al. (2013); references for distance and metallicity are given therein. Metallicity is calculated using the method of Pilyugin & Thuan (2005, abbreviated as PT05).
[b] As reported in Table 6 of Madden et al. (2013). Metallicity is calculated using the method of Izotov et al. (2006, abbreviated as I06). Missing values indicate cases where [O III]λ4363 was not available, and therefore a direct-method metallicity measurement was not possible.
[c] As reported in Table 4 of Cormier et al. (2015).
[d] Calculated using the listed values of distance and [O III] 88μm flux.
[e] Calculated using the listed values of [O III] 88μm luminosity and SFR via Equation 2, with +0.17 dex ionization correction.

TABLE 2
OXYGEN ABUNDANCE ESTIMATES OF $z > 7$ GALAXIES

Object name[a]	z_{spec}[a]	L([O III] 88μm)[a] [L$_\odot$]	SFR[a] [M$_\odot$ yr^{-1}]	L(Hβ)[b] [erg s^{-1}]	$12 + \log(\mathrm{O}^{++}/\mathrm{H}^+)$ $\pm\sigma_{stat}$	$12 + \log(\mathrm{O/H})$[c] $\pm\sigma_{stat} \pm \sigma_{sys}$	References[d]
MACS1149-JD1[e]	9.110	$(7.4 \pm 1.6) \times 10^7$	$4.2^{+0.8}_{-1.1}$	3.43×10^{41}	$7.98^{+0.15}_{-0.14}$	$8.15^{+0.15}_{-0.14} \pm 0.4$	H18, L19
A2744-YD4[e]	8.382	$(7.0 \pm 1.7) \times 10^7$	$12.9^{+11.1}_{-6.0}$	1.05×10^{42}	$7.47^{+0.28}_{-0.32}$	$7.64^{+0.28}_{-0.32} \pm 0.4$	L17, L19
MACS0416-Y1[e]	8.312	$(1.2 \pm 0.3) \times 10^9$	$57^{+175.0}_{-0.2}$	4.65×10^{42}	$8.06^{+0.10}_{-0.69}$	$8.23^{+0.10}_{-0.69} \pm 0.4$	T19
SXDF-NB1006-2	7.215	$(9.9 \pm 2.1) \times 10^8$	219^{+105}_{-176}	1.79×10^{43}	$7.39^{+0.71}_{-0.21}$	$7.56^{+0.71}_{-0.21} \pm 0.4$	I16
B14-65666	7.168	$(3.4 \pm 0.4) \times 10^9$	200^{+82}_{-38}	1.63×10^{43}	$7.97^{+0.10}_{-0.16}$	$8.14^{+0.10}_{-0.16} \pm 0.4$	H19, F16
BDF-3299 (UV)[f]	7.109	$< 5.5 \times 10^7$	>5.7	$> 4.65 \times 10^{41}$	7.72	7.89	C17, M15
BDF-3299 (IR)[f]	7.109	$(1.8 \pm 0.2) \times 10^8$	<12	$< 9.80 \times 10^{41}$	7.91	8.08	C17, M15

[a] As compiled in Table 2 of Harikane et al. (2020).
[b] Hβ luminosity calculated from SFR via Equation 1.
[c] $12 + \log(\mathrm{O/H})$ includes a fiducial correction of +0.17 dex from singly-ionized oxygen.
[d] References for redshift, [O III] 88μm luminosity, and SFR. C17: Carniani et al. (2017), F16: Furusawa et al. (2016), H18: Hashimoto et al. (2018), H19: Hashimoto et al. (2019), I16: Inoue et al. (2016), L17: Laporte et al. (2017), L19: Laporte et al. (2019), M15: Maiolino et al. (2015), T19: Tamura et al. (2019)
[e] Luminosity and SFR values are corrected for lensing magnification. All other quantities are independent of magnification.
[f] Far-IR line emission is spatially offset from the rest-UV continuum, and we give results separately for these two spatial locations. For the rest-UV component we list the reported upper limit on L([O III] 88μm) (which assumes a line width of 100 km s^{-1}), and the reported SFR assuming no extinction (which we express as a lower limit). For the IR [O III] 88μm-emitting component we list the reported 2-σ upper limit on SFR (although this limit may be underestimated, as noted by Carniani et al. 2017). We refer readers to Carniani et al. (2017) for further details of the measurements, limits, and spatial structure of this system. Oxygen abundances correspond to the listed values in both cases and should be treated as an upper (lower) limit for the UV (IR) region. Given the complexities of upper and lower limits, we do not attempt to quantify the uncertainty range.

TABLE 3
SOURCES OF SYSTEMATIC UNCERTAINTY IN OXYGEN ABUNDANCE

Parameter	Fiducial value	Range	$\sigma_{\log(\mathrm{O/H})}$[a]	$\frac{\partial \log(\mathrm{O/H})}{\partial \log X}$[b]	Prospects for eliminating systematic error at $z > 7$
T_e	15,000 K	9,000–20,000 K	$^{+0.17}_{-0.09}$	-0.72	[O III]$\lambda\lambda$4959,5007 with JWST, & [O III] 52μm with ALMA
n_e	250 cm^{-3}	10–600 cm^{-3}	$^{+0.23}_{-0.24}$	0.47	[O III] 52μm with ALMA
L(Hβ)	(Equation 1)	RMS $\simeq$ 0.25 dex	± 0.25	-1.0	Hβ and other H I lines with JWST
$\frac{\mathrm{O}^+ + \mathrm{O}^{3+}}{\mathrm{O}_{tot}}$	0.33	0.08–0.59	$^{+0.22}_{-0.13}$	0.25	[O II]$\lambda\lambda$3727,3729 with JWST

[a] Here σ corresponds to the difference in $\log(\mathrm{O/H})$ between the fiducial value and extrema of the range.
[b] Partial derivative of $\log(\mathrm{O/H})$ with respect to each parameter (X: T_e, n_e, etc.), evaluated at the fiducial values.

TABLE 4
Literature mass and metallicity measurements

Object name	z	$12 + \log{(\mathrm{O/H})}$ ([O III]λ4363-based)	$\log \frac{\mathrm{M_*}}{\mathrm{M_\odot}}$ [a]	$\log \frac{\mathrm{M_*}}{\mathrm{M_\odot}}$ [b] (R-B20, 1-comp)	$\log \frac{\mathrm{M_*}}{\mathrm{M_\odot}}$ [b] (R-B20, 2-comp)	References
			$z > 7$ sample			
MACS1149-JD1	9.110		$9.03^{+0.17}_{-0.08}$	8.23 ± 0.01	9.88 ± 0.01	H18
A2744-YD4	8.382		$9.29^{+0.24}_{-0.18}$	8.71 ± 0.02	10.26 ± 0.06	L17
MACS0416-Y1	8.312		$8.38^{+0.10}_{-0.06}$	8.93 ± 0.01	10.35 ± 0.03	T19
SXDF-NB1006-2	7.215		$8.54^{+0.79}_{-0.22}$			I16
B14-65666	7.168		$8.89^{+0.05}_{-0.04}$	9.25 ± 0.04	9.49 ± 0.99	H19
			$z = 1 - 4$ sample			
S13[c]	1.425	$7.95^{+0.07}_{-0.07}$	$8.33^{+0.1}_{-0.14}$			
AEGIS-11452[c]	1.6715	$7.72^{+0.19}_{-0.16}$	$9.48^{+0.08}_{-0.36}$			
C12a[c]	1.8339	$7.46^{+0.23}_{-0.22}$	$7.7^{+0.1}_{-0.1}$			
GOODS-S-41547[c]	2.5451	$7.84^{+0.33}_{-0.24}$	$9.3^{+0.13}_{-0.06}$			
A1689-217[d]	2.5918	8.06 ± 0.12	$8.07 - 8.59$			
COSMOS-1908[c]	3.0767	$8.02^{+0.10}_{-0.13}$	$8.93^{+0.1}_{-0.1}$			
COSMOS-23895[c]	3.6372	$7.99^{+0.26}_{-0.17}$	$9.43^{+0.13}_{-0.07}$			

[a] References for stellar mass are given in the final column, with same codes as Table 2.
[b] 1- and 2-component masses from R-B20: Roberts-Borsani et al. (2020). Masses are corrected for lensing magnification, where relevant.
[c] As compiled by Sanders et al. (2020).
[d] From Gburek et al. (2019).