



# Metal Deficiency in Two Massive Dead Galaxies at $z \sim 2$

T. Morishita<sup>1</sup>, L. E. Abramson<sup>2</sup>, T. Treu<sup>2</sup>, X. Wang<sup>2</sup>, G. B. Brammer<sup>1</sup>, P. Kelly<sup>3,4</sup>, M. Stiavelli<sup>1</sup>, T. Jones<sup>5</sup>,  
K. B. Schmidt<sup>6</sup>, M. Trenti<sup>7</sup>, and B. Vulcani<sup>7</sup>

<sup>1</sup>Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA; [tmorishita@stsci.edu](mailto:tmorishita@stsci.edu)

<sup>2</sup>Department of Physics and Astronomy, UCLA, 430 Portola Plaza, Los Angeles, CA 90095-1547, USA

<sup>3</sup>Department of Astronomy, University of California, Berkeley, CA 94720-3411, USA

<sup>4</sup>School of Physics and Astronomy, University of Minnesota, 116 Church Street SE, Minneapolis, MN 55455, USA

<sup>5</sup>University of California Davis, 1 Shields Avenue, Davis, CA 95616, USA

<sup>6</sup>Leibniz-Institut für Astrophysik Potsdam, An der Sternwarte 16, D-14482 Potsdam, Germany

<sup>7</sup>School of Physics, Tin Alley, University of Melbourne VIC 3010, Australia

Received 2018 February 13; revised 2018 March 3; accepted 2018 March 5; published 2018 March 19

## Abstract

Local massive early-type galaxies are believed to have completed most of their star formation  $\sim 10$  Gyr ago and evolved without having substantial star formation since that time. If so, their progenitors should have roughly solar stellar metallicities ( $Z_*$ ), comparable to their values today. We report the discovery of two lensed massive ( $\log M_*/M_\odot \sim 11$ ),  $z \sim 2.2$  dead galaxies that appear markedly metal deficient given this scenario. Using 17-band *HST*+*K<sub>s</sub>*+*Spitzer* photometry and deep *Hubble Space Telescope* (*HST*) grism spectra from the Grism Lens Amplified Survey from Space (GLASS) and supernova (SN) Refsdal follow-up campaigns covering features near  $\lambda_{\text{rest}} \sim 4000$  Å, we find these systems to be dominated by A-type stars with  $\log Z_*/Z_\odot = -0.40 \pm 0.02$  and  $-0.49 \pm 0.03$  (30%–40% solar) under standard assumptions. The second system's lower metallicity is robust to isochrone changes, though this choice can drive the first system's from  $\log Z_*/Z_\odot = -0.6$  to 0.1. If these two galaxies are representative of larger samples, this finding suggests that evolutionary paths other than dry minor merging are required for these massive galaxies. Future analyses with direct metallicity measurements—e.g., by the *James Webb Space Telescope*—will provide critical insight into the nature of such phenomena.

**Key words:** galaxies: abundances – galaxies: evolution – galaxies: structure

## 1. Introduction

Stellar chemical compositions are a powerful means to link galaxies across cosmic time and therefore understand their evolution. Massive early-type galaxies in the local universe exhibit solar or super-solar stellar metallicities ( $Z_*$ ; e.g., Thomas et al. 2005). This holds beyond  $z \sim 0$ —previous observations report similar values in massive dead galaxies up to  $z \sim 1.6$  (e.g., Gallazzi et al. 2014; Lonoce et al. 2015; Onodera et al. 2015), indicating that either they formed from metal-enriched gas or retained a lot of their metals after forming from pristine gas.

These observations spanning  $\sim 10$  Gyr seem to fit in a recently popular evolution scenario for massive galaxies, where dry minor merging might grow galaxy sizes without adding much mass (e.g., Naab et al. 2009), explaining the observed size evolution by a factor of  $\gtrsim 3$  from  $z \sim 2$  to 0 (e.g., Szomoru et al. 2012; van der Wel et al. 2014). As these galaxies are already dead and near the mass of local ellipticals (e.g., Stiavelli et al. 1999; Daddi et al. 2005; Glazebrook et al. 2017), they are believed to evolve without further substantial episodes of in situ star formation over the rest of their lives. If so, the main  $z \sim 2$  progenitor must already have *at least* its  $z = 0$  descendant's metallicity: the low-mass satellites that it will accrete are probably metal poor (Gallazzi et al. 2005; Kirby et al. 2013), and thus can only dilute the final system's total metallicity (see also Choi et al. 2014).

The above scenario, however, may not be the complete picture. Other works (e.g., Newman et al. 2012; Nipoti et al. 2012; Sonnenfeld et al. 2014) show that minor mergers alone cannot accomplish the necessary size evolution. Furthermore, recent stellar metallicity estimates at  $z \gtrsim 2$  yield

values  $\sim 2.5\sigma$  below the local  $M_*$ – $Z_*$  locus (Kriek et al. 2016; Toft et al. 2017), adding to arguments for more-than-minor-merging growth. To clarify the picture, we must take steps toward a more complete galaxy census.

Here we report two additional massive ( $\log M_*/M_\odot \sim 11$ ) dead galaxies at  $z \sim 2.2$  that appear substantially under-enriched. Both galaxies have typical A-type spectra, but  $\log Z_*/Z_\odot = -0.40 \pm 0.02$  and  $-0.49 \pm 0.03$ , respectively. With the other spectrophotometric results at  $z \sim 2$ , these galaxies suggest something other than dry minor merging is required to explain the subsequent evolution of some fraction of  $z \gtrsim 2$  passive galaxies. While results based on a larger sample will be presented in a future study, this Letter reports the discovery of these two galaxies and briefly discusses its implications. Throughout, magnitudes are quoted in the AB system assuming  $\Omega_m = 0.3$ ,  $\Omega_\Lambda = 0.7$ ,  $H_0 = 70$  km s<sup>−1</sup> Mpc<sup>−1</sup>.

## 2. Data and Sample

### 2.1. Hubble Space Telescope (*HST*) Data

We require high-quality spectral coverage over the rest-frame 3700–4200 Å window and broadband photometry over a much wider wavelength range to pin down details of the youngest and oldest stellar populations and break the age–metallicity degeneracy. Such data are difficult to obtain from the ground, but lensing magnification and deep near-infrared (NIR) *HST* spectrophotometric campaigns provide an ideal way to study passive galaxies at such early epochs.

The spectral range of the Wide Field Camera 3 (WFC3)/G141 grism covers age-sensitive features—the Balmer and 4000 Å breaks—for galaxies at  $z \sim 2$ , such that high-quality *HST* photometry covering 0.2–1.7 μm from Cluster Lensing

and Supernova Survey (CLASH; Postman et al. 2012) and Hubble Frontier Fields (HFF; Lotz et al. 2017) can constrain stellar metallicities at sufficient confidence. Among the survey lines of sight, the  $z = 0.544$  galaxy cluster MACSJ1149.5+2223 is an optimal target: it is covered by 17-band photometry spanning F225W–F160W to typical  $5\sigma$  depths of  $\sim 29$  (for the HFF data) to  $\sim 27$  mag (for CLASH). Moreover, spectroscopy from the Grism Lens Amplified Survey from Space (GLASS; Schmidt et al. 2014; Treu et al. 2015) and the follow-up *HST* General Observer/Director’s Discretionary Time (HSTGO/DDT) campaign of supernova (SN) Refsdal (Proposal ID 14041; Kelly et al. 2015, 2016) provide 34 orbit-deep G141 spectra at four position angles (PAs; 81,000 s total integration). A collection of the spectra is available at MAST (doi:10.17909/T9ZT2B).

We match all imaging to F160W resolution using unsaturated stars as point-spread functions (PSFs) before conducting photometry with SExtractor (Bertin & Arnouts 1996). Apertures are set to radii of  $0''.72$ , but measurements are scaled based on the aperture-to-total flux ratio in F160W.

As for spectra, we retrieve the WFC3/G141 data from the Mikulski Archive for Space Telescopes archive, and run the Grizli software package for one-dimensional optimal extraction. We then stack the extracted spectrum from each orbit using inverse-variance weighting based on the pipeline-output uncertainties. We use jackknife resampling to estimate the error on the stack’s mean (e.g., Onodera et al. 2015) and add that to the  $\sigma_{\text{rms}}$  formal error to obtain our final spectral uncertainties. We reach a signal-to-noise ratio (S/N)  $\sim 20$  per  $\sim 45 \text{ \AA}$  pixel at rest-frame  $4000 \text{ \AA}$ . This is slightly finer than the Nyquist sampling of G141 thanks to sufficient dithering across the orbits, each of which produces a shift in the dispersion direction.

The line spread function (LSF) is estimated using each galaxy’s morphology at each PA, to which the fitting templates are convolved (Section 3). While the spectra and templates are matched to the broadest one among all PAs, LSF variations at different PAs show negligible variation given the sources’ point-like/symmetric morphology.

## 2.2. $K_s$ and Infrared Array Camera (IRAC) Photometry

In addition to precise spectroscopic constraints near  $\lambda_{\text{RF}} \sim 4000 \text{ \AA}$ , well-characterized spectral energy distribution (SED) features at rest-frame  $\gtrsim 1 \mu\text{m}$  are essential for estimating dust and metallicity content. This regime is redward of *HST*’s bandpasses for  $z \gtrsim 1$  galaxies, but accessible via ground-based  $K_s$  and *Spitzer*/IRAC photometry. We obtain  $K_s$  and IRAC1 ( $3.6 \mu\text{m}$ )/IRAC2 ( $4.5 \mu\text{m}$ ) photometry from the publicly available catalog by Castellano et al. (2016), based on Keck/Multi-Object Spectrometer For Infrared Exploration (MOS-FIRE; Brammer et al. 2016) and *Spitzer* Frontier Fields observations (PI: T. Soifer, P. Capak).

To account for aperture-loss differences compared to the *HST* photometry, as well as zeropoint uncertainties in the ground-based data, we renormalize the public data based on corrections derived from the best-fit *HST*-only SED templates for  $z < 1$  galaxies, where *HST* photometry captures the full wavelength range to  $\lambda_{\text{RF}} \sim 1 \mu\text{m}$ . The medians of this correction are  $-2.8\%$ ,  $-3.0\%$ , and  $3.4\%$  for  $K_s$ , IRAC1, and IRAC2, respectively. We incorporate these small corrections into the photometric error budget.

## 2.3. Sample Galaxies

To cover the  $4000 \text{ \AA}$  break at sufficient S/N, we limit our initial pre-selection to galaxies with  $H_{160} < 24$  and  $1.6 < z_{\text{phot}} < 3.3$ . A rest-frame color cut at  $U-V > 1.0$  is then applied to bias the sample toward passive galaxies. Combined, these criteria leave 19 suitable galaxies in the MACS1149 field. Unfortunately, contamination from nearby spectra affects most of these, leaving a final sample of two galaxies with sufficiently clean spectra. The spectra have sufficient S/N of  $\sim 20$  per pixel at rest-frame  $4000 \text{ \AA}$  ( $B \sim 21.7$  and  $22.4$  mag, respectively). The rest of the sample will not support robust results in the following SED analysis.

## 3. Spectrophotometric SED Fitting

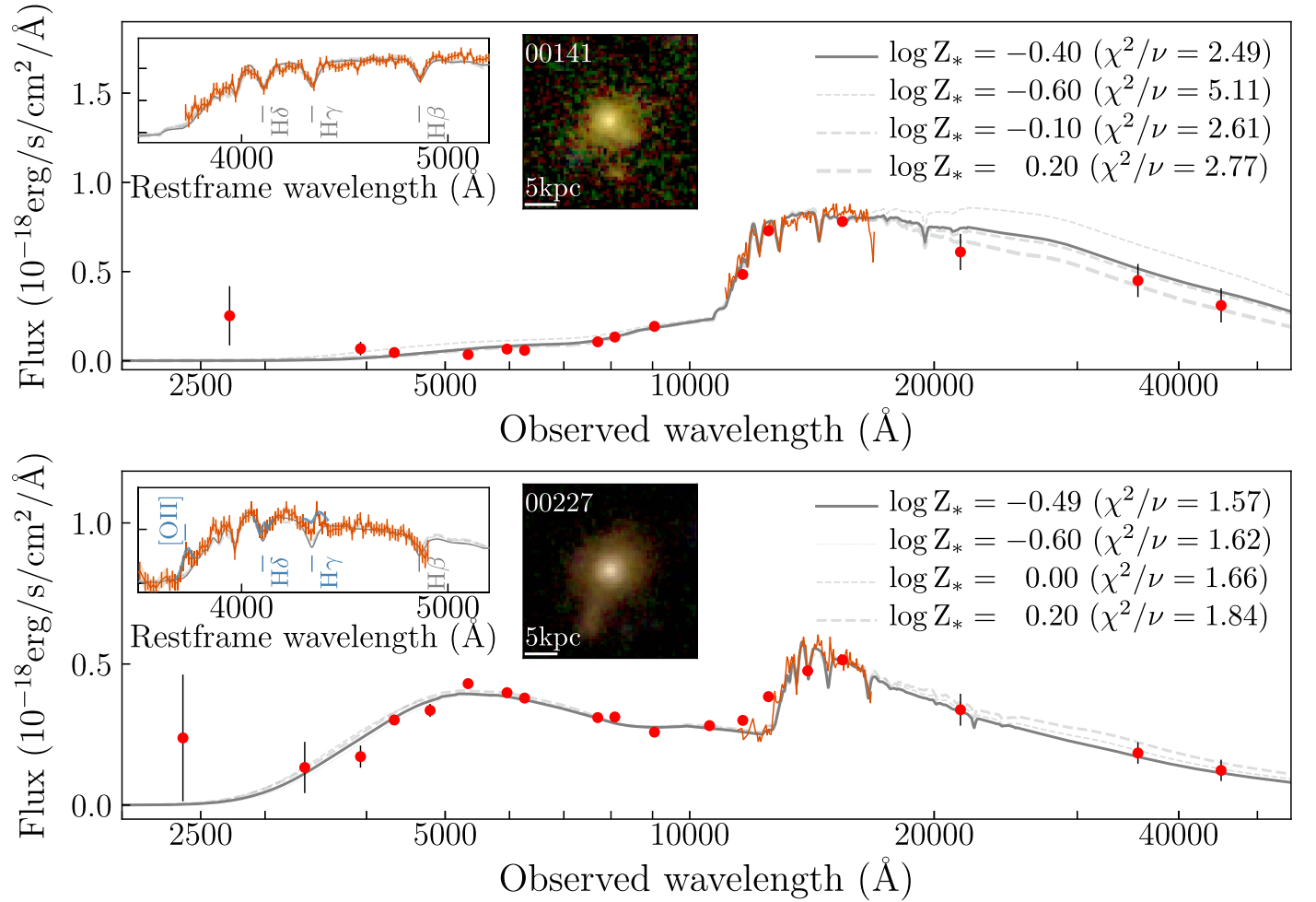
With extensive multi-band photometry and deep NIR spectra, we deconstruct the stellar populations of the two high- $z$  passive galaxies. A forthcoming paper presents full details of the SED fitting, but our scheme was established and tested by Kelson et al. (2014) using a similar data set at  $z \lesssim 1$ . The core of the method is to find the best combination of amplitudes,  $\{a_i=[0,1,2,3,4]\}$ , for a set of stellar templates of different ages,  $\{t_i=[0,1,2,3,4]\}$ , that matches the data. This flexibility avoids systematic biases in, e.g., stellar mass that can arise when using functional forms for star formation histories.

Our baseline results adopt the Flexible Stellar Population Synthesis code (FSPS; Conroy et al. 2009; Conroy & Gunn 2010) to generate four simple stellar population (SSP) templates with ages of [0.01, 0.1, 0.5, 1.0, 3.0] Gyr (roughly O, B, early/late-A, and F stars), based on Padova isochrones and the MILES stellar library. We assume a Salpeter IMF and Calzetti dust law. Our main results are robust to those assumptions except for the choice of isochrones, which systematically affects one system (see below). The number and choice of templates was set to fully exploit the data based on result reproducibility tests using simulations. Assuming non-zero durations of star formation for each template—e.g., exponentially declining star formation histories—pushes systems to older ages, and thus lower metallicities to compensate for the subsequently redder SEDs. Thus, SSPs produce conservative results for our purposes.

The interpolation scheme in the FSPS python module generates each template at metallicities spanning  $\log Z_*/Z_\odot \in [-1.0, 0.5]$  in increments of 0.05 dex. Metallicity and extinction are fixed across all five SSP templates; i.e., the fits have  $5 + 1 + 1$  free parameters.

We use a Monte Carlo Markov Chain to find the best combination of  $a_i$ ,  $A_V$ ,  $Z_*$ , and their covariances in the parameter ranges ( $a_i \in [0, 200]$  and  $A_V \in [0, 4]$  mag) with flat priors. We set the number of walkers to 200 and use  $N_{\text{mc}} = 10000$  realizations. Redshifts determined beforehand via  $\chi^2$ -minimization (and visual inspection of absorption lines) are fixed during parameter estimation. Final results are drawn from the last  $N_{\text{mc}}/2$  steps to avoid biases from the initial exploration. Emission lines detected at  $> 1.5\sigma$  across 3 pixels ( $\sim 135 \text{ \AA}$ ) are masked from the best-fit template of each iteration and later fit with gaussians.

One caveat is that the metallicity determined above is not a direct measurement, as metallicity sensitive indicators, such as



**Figure 1.** SEDs and fits for 00141 (top) and 00227 (bottom). Photometry and spectra (red points and lines) are shown with errors (black and red bars, respectively) overlaid on the best-fit model (gray solid line). Gray dashed lines show fit models with different metallicities at local  $\chi^2/\nu$ -minima ( $\log Z_*/Z_\odot \sim [-0.6, 0, 0.2]$ ; thin to thick). Insets show the region around  $\lambda_{\text{rest}} \sim 4000 \text{ \AA}$  and  $3''6 \times 3''6$  F814/125/160W pseudo-color images of the sources. Weak emission lines are detected in 00227 and modeled with Gaussian (blue lines).

**Table 1**  
Summary of the SED Fitting Parameters

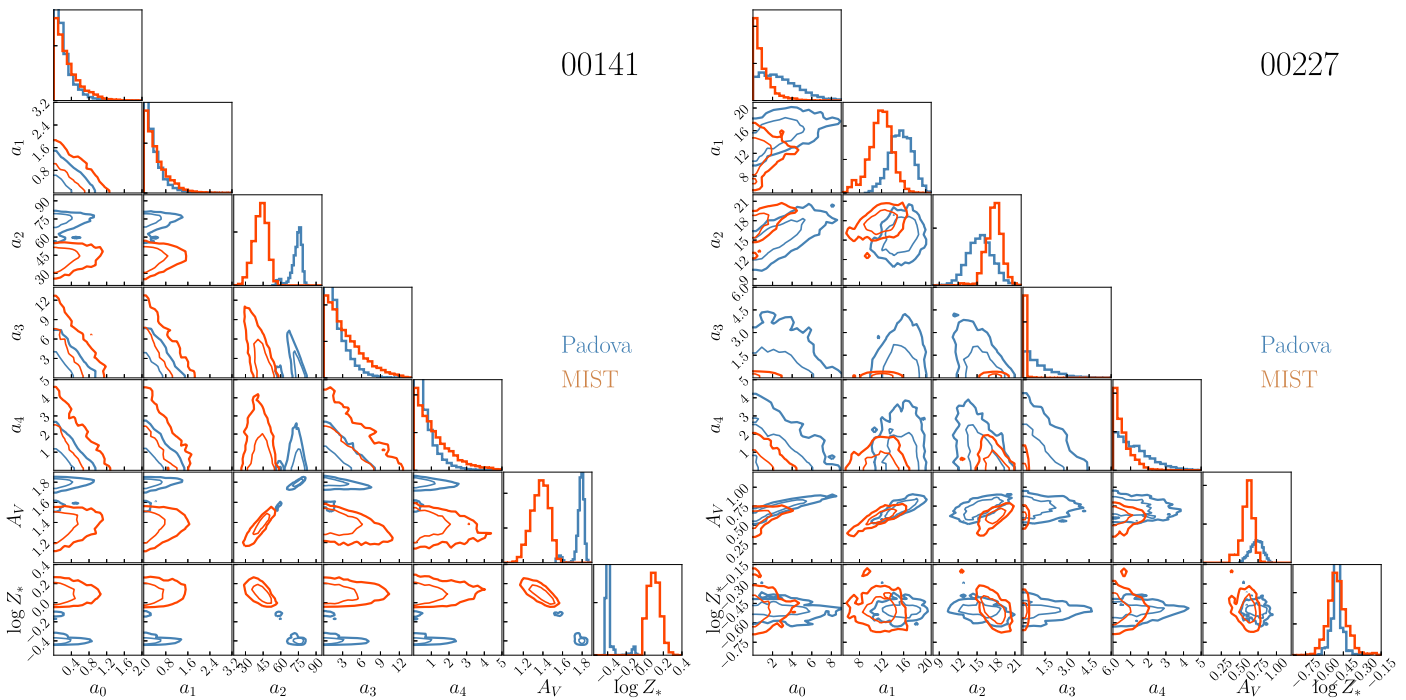
ID	R.A. (degree)	Decl. (degree)	$z_{\text{spec.}}$	$\mu$	$\log Z_*$ ( $Z_\odot$ )	$a_0$ (10 Myr)	$a_1$ (100 Myr)	$a_2$ (500 Myr)	$a_3$ (1 Gyr)	$a_4$ (3 Gyr)	$A_V$ (mag)	$M_*$ ( $10^{10} M_\odot$ )
Padova Isochrone												
00141	177.4034	22.4185	$1.967^{+0.002}_{-0.001}$	$1.85^{+0.83}_{-0.11}$	$-0.40^{+0.02}_{-0.02}$	$0.18^{+0.26}_{-0.13}$	$0.22^{+0.40}_{-0.16}$	$74.06^{+2.86}_{-4.85}$	$1.47^{+2.07}_{-1.09}$	$0.52^{+0.73}_{-0.38}$	$1.79^{+0.02}_{-0.03}$	$17.86^{+0.43}_{-0.54}$
00227	177.4068	22.41624	$2.412^{+0.004}_{-0.004}$	$1.68^{+0.37}_{-0.09}$	$-0.49^{+0.04}_{-0.03}$	$0.71^{+1.77}_{-0.62}$	$14.72^{+1.81}_{-1.89}$	$14.10^{+1.59}_{-1.52}$	$0.88^{+1.31}_{-0.67}$	$1.37^{+1.22}_{-0.94}$	$0.66^{+0.06}_{-0.07}$	$5.78^{+0.68}_{-0.56}$
MESA Isochrones and Stellar Tracks (MIST) Isochrone												
00141	177.403	22.419	$1.967^{+0.001}_{-0.002}$	$1.85^{+0.83}_{-0.11}$	$0.10^{+0.07}_{-0.07}$	$0.23^{+0.34}_{-0.17}$	$0.31^{+0.52}_{-0.23}$	$42.78^{+5.36}_{-5.59}$	$2.79^{+3.28}_{-2.00}$	$0.94^{+1.19}_{-0.70}$	$1.37^{+0.08}_{-0.08}$	$14.72^{+1.22}_{-1.14}$
00227	177.407	22.416	$2.411^{+0.004}_{-0.002}$	$1.68^{+0.37}_{-0.09}$	$-0.50^{+0.07}_{-0.06}$	$0.67^{+1.10}_{-0.51}$	$11.98^{+1.80}_{-2.13}$	$17.75^{+1.09}_{-1.31}$	$0.01^{+0.05}_{-0.01}$	$0.40^{+0.62}_{-0.31}$	$0.62^{+0.06}_{-0.08}$	$5.39^{+0.46}_{-0.40}$

**Note.** 50th percentile is quoted as the best-fit parameter with 16th/84th percentiles as statistical uncertainties. Systematic uncertainties derived from simulations (Section 4) are  $\sim 0.25$  dex,  $0.25$  dex, and  $0.17$  dex for  $A_V$ ,  $\log Z_*/Z_\odot$ , and  $\log M_*/M_\odot$ , respectively.  $a_i$ : Amplitudes of the templates, with  $i = [0, 1, 2, 3, 4]$  representing  $t_i = [0.01, 0.1, 0.5, 1, 3]/\text{Gyr}$ .  $\mu$ : Median of magnification by the foreground cluster. Magnification errors are not included in the parameter errors as all are  $\mu$ -independent except for  $M_*$ .

Fe and Mg, fall outside the G141 spectral range. The broadband-defined “ $Z_*$ ” here is thus different from the total metallicity derived from absorption line measurements (see Section 5).

#### 4. Results

Figure 1 shows the data and fits and Table 1 summarizes the best-fit parameters. Both galaxies prefer sub-solar metallicities under the standard assumption of Padova isochrones:



**Figure 2.** Covariance matrices of the SED fitting parameters, for 00141 (left) and 00227 (right), based on Padova (blue) and MIST (red) isochrones. Contours enclose the top 68%, 95%, and 99.7% of the distribution, respectively.

$\log Z_*/Z_\odot = -0.40^{+0.02}_{-0.02}$  and  $-0.49^{+0.04}_{-0.03}$  for 00141 and 00227, respectively, with estimates and uncertainties reflecting the 16th/50th/84th percentile values.<sup>8</sup> Also, both best-fit templates are dominated by  $\sim 0.5$ -Gyr-old stellar populations:  $\sim 99\%$  of 00141’s light is in the  $a_2$ -template, while 00227’s stellar population is younger, with  $\gtrsim 44\%$  in  $a_1$  and  $a_2$ . Stellar masses are  $M_* = 1.8 \times 10^{11} M_\odot$  and  $5.8 \times 10^{10} M_\odot$ , respectively. 00141 exhibits no significant emission, while 00227 shows very weak Balmer infilling (equivalent widths of  $H\delta \sim 0.8 \pm 0.2 \text{ \AA}$  and  $H\gamma \sim 1.8 \pm 0.2 \text{ \AA}$ ) and [O II] emission ( $\sim 3.4 \pm 0.7 \text{ \AA}$ ), suggesting the latter was indeed quenched more recently. Both galaxies have compact/centrally concentrated morphologies with effective radii  $r_e = 1.8\text{--}2 \text{ kpc}$  and Sérsic indices of  $n = 8$ . These values are typical for passive galaxies at similar redshifts (e.g., Szomoru et al. 2012), and confirm that significant dry minor merging is needed if they are to grow to match the size of  $z = 0$  early-types.

Figure 2 shows the fit parameter covariances. While some covary strongly (e.g.,  $a_2$  and  $A_V$ ), none significantly impact the metallicity assessment. Simulations of our SEDs (discussed momentarily) confirm these results.

As we lack detailed spectra, the metallicities reported here might depend on the isochrones adopted during SED fitting. We tested all isochrones available in FSPS, and found that, while 00227’s result is robust,  $Z_*$  changes significantly for 00141, reaching a low of  $-0.6$  (PAдова & TRIESTE Stellar Evolution Code (PARSEC); Bressan et al. 2012) and a high of  $\log Z_*/Z_\odot \sim 0.10 \pm 0.07$  (MIST; Choi et al. 2016). This is mostly caused by the different treatments of the asymptotic giant branch, which contributes significantly to the galaxy’s luminosity at ages near 00141’s of  $\sim 500 \text{ Myr}$ . Because we

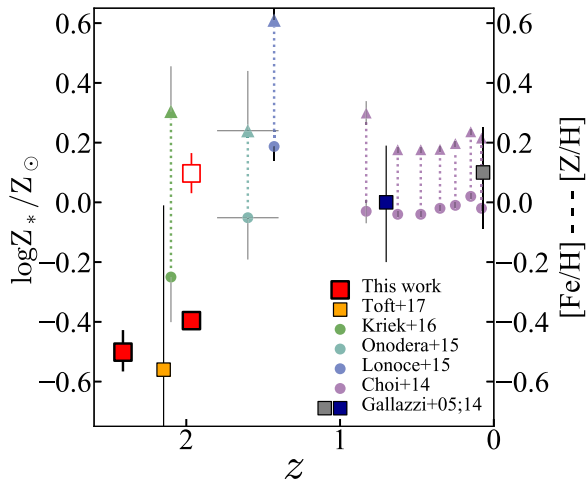
cannot determine the best isochrone with the current data set, we base our discussion on the result with the standard Padova isochrone to facilitate contextualizing our results with those in the literature. However, we present the MIST results in Table 1 and related figures to illustrate the effect of this choice.

In addition to isochrone systematics, there is a chance that, due to dust or statistical fluctuations, an observed sub-solar SED could be measured from a galaxy that is truly enriched to solar levels. To estimate this contamination rate, we tested our fitting method on 1000 simulated SEDs covering the same spectrophotometric bandpasses at comparable S/N with parameters randomly chosen over the fitting range. While full results will be discussed in our forthcoming study, the simulation shows that the input and extracted estimates are unbiased up to  $A_V = 1.3 \text{ mag}$ . Above that, our inferences underestimate  $A_V$  by up to 0.2 mag. If so, to return a sufficiently reddened SED, the fitter would tend to *increase*  $Z_*$ , returning an overestimate of the true metallicity, making our results conservative. Finally, we find  $\sim 20\%$  and  $\sim 5\%$  chances that 00141’s  $\log Z_*/Z_\odot = -0.4$  and 00227’s  $\log Z_*/Z_\odot = -0.49$ , respectively, could have come from a galaxy with truly solar metallicity.

## 5. Implications of Sub-solar Metallicities in Massive Dead Galaxies

The sub-solar abundances found in this study, suggest that these galaxies are in the early stages of *chemical* evolution, if already quite mature in terms of *mass*. While perhaps not surprising given the age of the universe (e.g., De Lucia & Blaizot 2007), these results do differ from the solar to super-solar values reported previously for massive quiescent galaxies at lower redshifts (Gallazzi et al. 2005, 2014; Choi et al. 2014; Onodera et al. 2015) as summarized in Figure 3. (Gallazzi et al. 2014 and Leethochawalit et al. 2018, however, show  $Z_*$  evolution at  $z < 0.7$ .)

<sup>8</sup> While it is known that metallicities are effectively differentiated in the rest-frame NIR (e.g., Lee et al. 2007), our  $K_S/\text{IRAC}$  data do not provide the tightest constraints. Instead, the grism spectrum and broadband photometry at rest-frame  $\sim 2000$  to  $5000 \text{ \AA}$  do by breaking the age/metallicity degeneracy.



**Figure 3.** Stellar metallicities over time comparing our results using Padova (red filled squares) and MIST isochrones (open squares) to previous literature  $\log Z_*/Z_\odot$  estimates (Gallazzi et al. 2005, 2014; Toft et al. 2017, small squares) of massive galaxies ( $\log M_*/M_\odot \gtrsim 11$ ). We show  $[\text{Fe}/\text{H}]$  and  $[Z/\text{H}]$  (circles and triangles, respectively) estimated from direct iron/ $\alpha$ -abundances (Choi et al. 2014; Lonoce et al. 2015; Onodera et al. 2015; Kriek et al. 2016), though the direct comparison should be made with care (see the main text).

Now,  $Z_*$  here derives from broad SED features. Given the complexities of stellar evolution, it is therefore perhaps not easily interpreted, e.g., as reflecting iron or total (iron +  $\alpha$ ) abundances. This fact complicates comparisons to other measurements based on highly detailed spectroscopy. As such, in those cases (Choi et al. 2014; Lonoce et al. 2015; Onodera et al. 2015; Kriek et al. 2016) we show in Figure 3 both  $[\text{Fe}/\text{H}]$  and the total abundance,  $[Z/\text{H}]$ , where  $[Z/\text{H}] = [\text{Fe}/\text{H}] + 0.94$   $[\alpha/\text{Fe}]$  (Thomas et al. 2003). Other results reported only as “ $\log Z_*/Z_\odot$ ” are reproduced as published (Gallazzi et al. 2005, 2014; Toft et al. 2017).

Assuming that our two galaxies do not represent an outlier population, the differences illustrated in Figure 3 raise interesting questions about the diversity of evolutionary paths of these objects to  $z \sim 0$ . What can be securely said is that the popular scenario for high- $z$  passive galaxy evolution does not hold for these objects: while it may grow their sizes sufficiently, dry minor merging cannot correspondingly raise their metallicities to those of their  $z = 0$  counterparts. Given their low mass, any satellites are almost certain to have lower  $Z_*$  than these centrals (Pasquali et al. 2010; Kirby et al. 2013), so such merging would only further dilute these systems’ metal content. Furthermore, most metallicity gradients in local early-types are negative—or at least flat—with higher metallicity in their centers (e.g., Martin-Navarro et al. 2018). If the objects studied here were to somehow merge with more metal-rich satellites, their  $z = 0$  gradients would go in the opposite direction (Nipoti et al. 2012).

Given this boundary condition, there appear to be three plausible routes for our galaxies to take.

- (1) Major mergers: Merging with another  $\sim$ equal-mass galaxy of  $Z_* \gtrsim Z_\odot$  would bring them toward/onto the local mass–metallicity relation. While major mergers are rare, globally, and probably too rare to happen to *all* massive galaxies (e.g., Bundy et al. 2009), given the amount of time from the observed redshift to today and the increasing merger rate increases toward higher- $z$  (e.g., Lotz et al. 2011), this route cannot be excluded here. If

this is taken as the preferred solution, measurements like this may provide independent tests of the merger rate once sufficient samples exist.

- (2) A second burst of star formation. Newly accreted gas delivered either by *wet* satellite galaxies or cosmic accretion can raise metallicities at later times via in situ star formation. Dissipation efficiently supplies the gas to the central part of the system and induces starbursts. Despite uncertainties in the metallicity of the inflowing gas and the size of these bursts—perhaps  $\sim 10\%$ – $15\%$  of the systems’ extant stellar mass (e.g., MacArthur et al. 2008)—the large numbers of minor mergers in massive systems since  $z \sim 2$  (e.g., Rodriguez-Gomez et al. 2015) make this scenario attractive, as do observations of blue core early-types (e.g., Treu et al. 2005), and other metrics of minor merger-triggered star formation at intermediate redshift (e.g., Kaviraj et al. 2011). The centrally concentrated nature of star formation in this scenario would also produce negative metallicity gradients, qualitatively consistent with those in the local universe. However, the increase of the central density would result in smaller galaxy size, contradicting the observed size evolution.
- (3) The galaxies remain as they are. If this is the case, their descendant systems must be outliers today, suggesting most of the size *and* metallicity evolution of the passive population is due purely to progenitor bias. The situation is unclear at present: Poggianti et al. (2013) found local compact galaxies (size outliers) to have younger light-weighted ages, which could also reflect low metallicities. Other studies, however, find about solar values for compact local relics (e.g., Martin-Navarro et al. 2018). More spectroscopic studies of local compact galaxies will help resolve this picture further.

## 6. Summary











We identify two galaxies with SEDs that imply substantially sub-solar stellar metallicities under standard assumptions, adding to recent, similar results. Taken together, these findings suggest that some—if not many— $z \gtrsim 2$  passive galaxies are unlikely to evolve exclusively through dry minor merging (a currently popular scenario). As Kriek et al. (2016) also argued, this might imply that  $z \gtrsim 2$  galaxies are a distinct population from those even at  $z \sim 1.5$ , which already exhibit solar metallicities (see also Newman et al. 2012). Rather, the observed metallicity evolution of massive galaxies could be driven by newly quenched galaxies (a.k.a. progenitor bias; van Dokkum & Franx 1996; Carollo et al. 2013), as may also partially explain their size growth.

Ground-based  $K_s$  and IRAC photometry covers the rest-NIR in this study, a regime key to robust metallicity measurements (e.g., Lee et al. 2007). Despite the depth of such data, our lack of high-resolution spectroscopy does leave systematic uncertainties that must wait to be resolved by the *James Webb Space Telescope* (*JWST*). Most notable among these are isochrone effects, which can have a substantial impact at ages  $\sim 500$  Myr. Once in hand, large samples of direct *JWST* iron and  $\alpha$ -element abundances at high- $z$  would bear perhaps on our understanding of the cosmic merger rate, and the duty cycle of late-time star formation in once-dead galaxies in addition to their early chemical evolution.

We thank the anonymous referee for insightful suggestions, and Camilla Pacifici and Mia Bovill for fruitful discussion. Support for GLASS (*HST*-GO-13459, 14041) was provided by NASA through a grant from the Space Telescope Science Institute, operated by the Association of Universities for Research in Astronomy, Inc., under contract NAS 5-26555. This work uses gravitational lensing models by PIs Bradač Natarajan & Kneib (CATS v4.1), Merten & Zitrin, Sharon, Williams, Keeton, Bernstein & Diego, and the GLAFIC group. B.V. acknowledges support from an Australian Research Council Discovery Early Career Researcher Award (PD0028506).

*Software:* Astropy (Muna et al. 2016), emcee (Foreman-Mackey et al. 2013), FSPS (Conroy et al. 2009; Conroy & Gunn 2010; Foreman-Mackey et al. 2014), Grizli (Brammer 2018), Imfit (Newville et al. 2017).

### ORCID iDs

T. Morishita  <https://orcid.org/0000-0002-8512-1404>  
 L. E. Abramson  <https://orcid.org/0000-0002-8860-1032>  
 T. Treu  <https://orcid.org/0000-0002-8460-0390>  
 X. Wang  <https://orcid.org/0000-0002-9373-3865>  
 G. B. Brammer  <https://orcid.org/0000-0003-2680-005X>  
 P. Kelly  <https://orcid.org/0000-0003-3142-997X>  
 T. Jones  <https://orcid.org/0000-0001-5860-3419>  
 K. B. Schmidt  <https://orcid.org/0000-0002-3418-7251>  
 M. Trenti  <https://orcid.org/0000-0001-9391-305X>  
 B. Vulcani  <https://orcid.org/0000-0003-0980-1499>

### References

- Bertin, E., & Arnouts, S. 1996, *A&AS*, 117, 393  
 Brammer, G. 2018, gbrammer/grizli: Preliminary release, v0.4.0, Zenodo, doi:10.5281/zenodo.1146905  
 Brammer, G. B., Marchesini, D., Labbé, I., et al. 2016, *ApJS*, 226, 6  
 Bressan, A., Marigo, P., Girardi, L., et al. 2012, *MNRAS*, 427, 127  
 Bundy, K., Fukugita, M., Ellis, R. S., et al. 2009, *ApJ*, 697, 1369  
 Carollo, C. M., Bschorr, T. J., Renzini, A., et al. 2013, *ApJ*, 773, 112  
 Castellano, M., Amorín, R., Merlin, E., et al. 2016, *A&A*, 590, A31  
 Choi, J., Conroy, C., Moustakas, J., et al. 2014, *ApJ*, 792, 95  
 Choi, J., Dotter, A., Conroy, C., et al. 2016, *ApJ*, 823, 102  
 Conroy, C., & Gunn, J. E. 2010, *ApJ*, 712, 833  
 Conroy, C., Gunn, J. E., & White, M. 2009, *ApJ*, 699, 486  
 Daddi, E., Renzini, A., Pirzkal, N., et al. 2005, *ApJ*, 626, 680  
 De Lucia, G., & Blaizot, J. 2007, *MNRAS*, 375, 2  
 Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, 125, 306  
 Foreman-Mackey, D., Sick, J., & Johnson, B. 2014, python-fsps: Python bindings to FSPS, v0.1.1, Zenodo, doi:10.5281/zenodo.12157  
 Gallazzi, A., Bell, E. F., Zibetti, S., Brinchmann, J., & Kelson, D. D. 2014, *ApJ*, 788, 72  
 Gallazzi, A., Charlot, S., Brinchmann, J., White, S. D. M., & Tremonti, C. A. 2005, *MNRAS*, 362, 41  
 Glazebrook, K., Schreiber, C., Labbé, I., et al. 2017, *Natur*, 544, 71  
 Kaviraj, S., Tan, K.-M., Ellis, R. S., & Silk, J. 2011, *MNRAS*, 411, 2148  
 Kelly, P. L., Brammer, G., Selsing, J., et al. 2016, *ApJ*, 831, 205  
 Kelly, P. L., Rodney, S. A., Treu, T., et al. 2015, *Sci*, 347, 1123  
 Kelson, D. D., Williams, R. J., Dressler, A., et al. 2014, *ApJ*, 783, 110  
 Kirby, E. N., Cohen, J. G., Guhathakurta, P., et al. 2013, *ApJ*, 779, 102  
 Kriek, M., Conroy, C., van Dokkum, P. G., et al. 2016, *Natur*, 540, 248  
 Lee, H.-C., Worthey, G., Trager, S. C., & Faber, S. M. 2007, *ApJ*, 664, 215  
 Leethochawalit, N., Kirby, E. N., Moran, S. M., et al. 2018, *ApJ*, in press (arXiv:1802.09560)  
 Lonoce, I., Longhetti, M., Maraston, C., et al. 2015, *MNRAS*, 454, 3912  
 Lotz, J. M., Jonsson, P., Cox, T. J., et al. 2011, *ApJ*, 742, 103  
 Lotz, J. M., Koekemoer, A., Coe, D., et al. 2017, *ApJ*, 837, 97  
 MacArthur, L. A., Ellis, R. S., Treu, T., et al. 2008, *ApJ*, 680, 70  
 Martin-Navarro, I., Vazdekis, A., Falcon-Barroso, J., et al. 2018, *MNRAS*, 475, 3700  
 Muna, D., Alexander, M., Allen, A., et al. 2016, arXiv:1610.03159  
 Naab, T., Johansson, P. H., & Ostriker, J. P. 2009, *ApJL*, 699, L178  
 Newman, A. B., Ellis, R. S., Bundy, K., & Treu, T. 2012, *ApJ*, 746, 162  
 Newville, M., Nelson, A., Ingargiola, A., et al. 2017, Imfit/Imfit-py 0.9.7, Zenodo, doi:10.5281/zenodo.802298  
 Nipoti, C., Treu, T., Leauthaud, A., et al. 2012, *MNRAS*, 422, 1714  
 Onodera, M., Carollo, C. M., Renzini, A., et al. 2015, *ApJ*, 808, 161  
 Pasquali, A., Gallazzi, A., Fontanot, F., et al. 2010, *MNRAS*, 407, 937  
 Poggianti, B. M., Calvi, R., Bindoni, D., et al. 2013, *ApJ*, 762, 77  
 Postman, M., Coe, D., Benítez, N., et al. 2012, *ApJS*, 199, 25  
 Rodríguez-Gomez, V., Genel, S., Vogelsberger, M., et al. 2015, *MNRAS*, 449, 49  
 Schmidt, K. B., Treu, T., Brammer, G. B., et al. 2014, *ApJL*, 782, L36  
 Sonnenfeld, A., Nipoti, C., & Treu, T. 2014, *ApJ*, 786, 89  
 Stiavelli, M., Treu, T., Carollo, C. M., et al. 1999, *A&A*, 343, L25  
 Szomoru, D., Franx, M., & van Dokkum, P. G. 2012, *ApJ*, 749, 121  
 Thomas, D., Maraston, C., & Bender, R. 2003, *MNRAS*, 339, 897  
 Thomas, D., Maraston, C., Bender, R., & Mendes de Oliveira, C. 2005, *ApJ*, 621, 673  
 Toft, S., Zabl, J., Richard, J., et al. 2017, *Natur*, 546, 510  
 Treu, T., Ellis, R. S., Liao, T. X., et al. 2005, *ApJ*, 633, 174  
 Treu, T., Schmidt, K. B., Brammer, G. B., et al. 2015, *ApJ*, 812, 114  
 van der Wel, A., Franx, M., van Dokkum, P. G., et al. 2014, *ApJ*, 788, 28  
 van Dokkum, P. G., & Franx, M. 1996, *MNRAS*, 281, 985