The Astrophysical Journal, 924:25 (24pp), 2022 January 1 https://doi.org/10.3847/1538-4357/ac312a
© 2022. The Author(s). Published by the American Astronomical Society.
OPEN ACCESS

The Number Densities and Stellar Populations of Massive Galaxies at $3 < z < 6$: A Diverse, Rapidly Forming Population in the Early Universe

Z. Cemile Marsan[1,8], Adam Muzzin[1], Danilo Marchesini[2], Mauro Stefanon[3], Nicholas Martis[2], Marianna Annunziatella[2,4], Jeffrey C. C. Chan[5], Michael C. Cooper[6], Ben Forrest[5], Percy Gomez[7], Ian McConachie[5], and Gillian Wilson[5]
[1] Department of Physics and Astronomy, York University, 4700, Keele Street, Toronto, ON MJ3 1P3, Canada; cmarsan@yorku.ca
[2] Department of Physics and Astronomy, Tufts University, 574 Boston Avenue Suites 304, Medford, MA 02155, USA
[3] Leiden Observatory, Leiden University, 2300 RA Leiden, The Netherlands
[4] Centro de Astrobiología (CSIC-INTA), Ctra de Torrejón a Ajalvir, km 4, E-28850 Torrejón de Ardoz, Madrid, Spain
[5] Department of Physics and Astronomy, University of California, Riverside, 900 University Avenue, Riverside, CA 92521, USA
[6] Center for Cosmology, Department of Physics and Astronomy, University of California, Irvine, 4129 Frederick Reines Hall, Irvine, CA, USA
[7] W.M. Keck Observatory, 65-1120 Mamalahoa Highway, Kamuela, HI 96743, USA
Received 2020 October 2; revised 2021 October 15; accepted 2021 October 15; published 2022 January 6

Abstract

We present the census of massive ($\log(M_*/M_\odot) > 11$) galaxies at $3 < z < 6$ identified over the COSMOS/UltraVISTA Ultra-Deep field stripes: consisting of ≈100 and ≈20 high-confidence candidates at $3 < z < 4$ and $4 < z < 6$, respectively. The $3 < z < 4$ population is comprised of post-starburst, UV-star-forming, and dusty star-forming galaxies in roughly equal fractions, while UV-star-forming galaxies dominate at $4 < z < 6$. We account for various sources of biases in the spectral energy distribution (SED) modeling, finding that the treatment of emission line contamination is essential for understanding the number densities and mass growth histories of massive galaxies at $z > 3$. The significant increase in observed number densities at $z \sim 4$ ($> \times 5$ in $\lesssim$600 Myr) implies that this is the epoch at which $\log(M_*/M_\odot) > 11$ galaxies emerge in significant numbers, with stellar ages (≈500–900 Myr) indicating rapid formation epochs as early as $z \sim 7$. Leveraging ancillary multiwavelength data sets, we perform panchromatic SED modeling to constrain the total star formation activity of the sample. The star formation activity of the sample is generally consistent with being on the star formation main sequence at the considered redshifts, with ≈15%–25% of the population showing evidence of suppressed star formation rates, indicating that quenching mechanisms are already at play by $z \sim 4$. We stack the available Hubble Space Telescope imaging, confirming their compact nature ($r_e \lesssim 2.2$ kpc), consistent with expected sizes of high-z star-forming galaxies. Finally, we discuss how our results are in-line with the early formation epochs and short formation timescales inferred from the fossil records of the most massive galaxies in the universe.

Unified Astronomy Thesaurus concepts: Galaxy evolution (594)

1. Introduction

In the nearby universe, the most massive galaxies are a relatively homogenous population: with early-type morphologies, little dust attenuation, and quiescent, metal-rich old (>10 Gyr) stellar populations (Gallazzi et al. 2005, 2006; Thomas et al. 2005; McDermid et al. 2015; Choi et al. 2014; Bellstedt et al. 2020). This implies that the majority of their stars were formed in the first few Gyr of cosmic history (i.e., $z > 2$) through short and intense bursts of star formation (e.g., Renzini 2006; van Dokkum & van der Marel 2007; Thomas et al. 2010). Indeed, numerous works have demonstrated that old, passively evolving galaxies exist at $1.5 < z < 2.5$ (Cimatti et al. 2004, 2008; Daddi et al. 2005; Kriek et al. 2009; Whitaker et al. 2013; Belli et al. 2017b, 2019; Toft et al. 2017; Newman et al. 2018; Carnall et al. 2019), while recent ambitious spectroscopic campaigns have confirmed the existence of massive, evolved galaxies at $3 < z < 4$ (Marsan et al. 2015, 2017; Glazebrook et al. 2017; Schreiber et al. 2018b; Tanaka et al. 2019; Forrest et al. 2020, 2021; Valentino et al. 2020; Saracco et al. 2020). Related to this, the number density of the most massive galaxies shows no evolution in the ≈3.5 Gyr between $z = 0.4$ and $z \sim 1$–1.5 (Kawinwanichakij et al. 2020), and evolves very little in the prior ∼3–4 Gyr to $z \approx 4$, providing additional evidence that they must have assembled important amounts of their stellar content rapidly beyond $z > 3$ (Franx et al. 2003; Cimatti et al. 2004; Daddi et al. 2004; Pérez-González et al. 2008; Wiklind et al. 2008; Marchesini et al. 2010; Brammer et al. 2011; Caputi et al. 2012; Ilbert et al. 2013; Muzzin et al. 2013a; Duncan et al. 2014; Nayyeri et al. 2014; Straatman et al. 2014; Tomczak et al. 2014; Grazian et al. 2015; Davidzon et al. 2017).

Massive galaxies display more diverse stellar populations with increasing cosmic look-back time: at $z \sim 2$, half of the massive galaxies are already devoid of star formation and have old stellar ages suggesting that they quenched their star formation at even earlier times (e.g., Kriek et al. 2006; Franx et al. 2008; Toft et al. 2009; McCracken et al. 2010; Wuyts et al. 2011; Brammer et al. 2011; Whitaker et al. 2011; Kado-Fong et al. 2017; Morishita et al. 2018; D'Eugenio et al. 2020), while at $z \sim 3$ the population is dominated by dusty, star-forming galaxies (Marchesini et al. 2010; Spitler et al. 2014; Martis et al. 2016, 2019; Deshmukh et al. 2018). Massive galaxies in the early universe are also more compact than their

[8] York Science Fellow.

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

local counterparts (Daddi et al. 2005; van der Wel et al. 2014; Straatman et al. 2015; Mowla et al. 2019), with a more dramatic size evolution observed for the quiescent population, e.g., van Dokkum et al. (2008); Williams et al. (2010); Newman et al. (2012). The resolved light distributions for the most massive galaxies at $z \sim 1.5$–3 show a range in morphologies with significant fractions (∼30%–40%) exhibiting features indicative of interactions and /or companions in close proximity (Marsan et al. 2019; Stockmann et al. 2020).

Deeper extragalactic surveys constructed in recent years have uncovered a substantial population of massive, evolved galaxies existing beyond $z > 3$, although statistical samples are still missing. The discovery of massive, evolved galaxy candidates at $z \gtrsim 4$ has been claimed by several works (Stefanon et al. 2015; Caputi et al. 2015; Mawatari et al. 2016, 2020; Wang et al. 2016, 2019; Merlin et al. 2018, 2019; Alcalde Pampliega et al. 2019; Girelli et al. 2019; Santini et al. 2019; Guarnieri et al. 2019; Carnall et al. 2020); however, these largely rely on detections in relatively few photometric data points. Thus, the task of determining when the first massive galaxies emerge and understanding the baryonic physics responsible for their accelerated evolution in the first billion years of the universe is still incomplete.

In this study, we aim to study the most massive galaxies present in the first ∼2.5 Gyr of the universe using a statistically large sample, and extend the investigation of the most massive galaxies even further at $z > 4$, the epoch at which these extreme systems are expected to emerge. While this has been examined over smaller fields by previous works (e.g., Mobasher et al. 2005; Wiklind et al. 2008; Fontana et al. 2009; Nayyeri et al. 2014), this work takes advantage of the unique combination of deep near- and mid-infrared observations over a wide area (≈0.84 deg^2 in the COSMOS field) provided by the UltraVISTA survey and the Spitzer Matching survey of the UltraVISTA ultra-deep Stripes (SMUVS), crucial to selecting statistical samples of evolved galaxies in the early universe (also see Caputi et al. 2015). The COSMOS field benefits from a multitude of (deep) ancillary data sets collected across the electromagnetic spectrum. Particularly relevant for understanding the massive high-z galaxy population, we utilize the available far-infrared (FIR), radio, and X-ray data sets, as well as imaging data from the Hubble Space Telescope (HST), to identify multiwavelength counterparts and explore the prevalence of active galactic nucleus (AGN), the level of star formation activity obscured by dust, as well as size estimates.

This paper is structured as follows: Section 2 describes the parent DR3 galaxy catalog and the quality cuts employed to identify the *bona fide* mass-complete ($\log(M_*/M_\odot) > 11$) sample of galaxies at $3 < z < 6$, as well as the ancillary multiwavelength data sets incorporated in our analyses. We dissect and characterize the stellar populations and AGN fraction of the sample in Section 3. Tests for the robustness of stellar-mass estimates and the high-confidence number density measurements of massive galaxies at $3 < z < 6$ are presented in Section 4. Section 5 explores the star formation activity of the sample via the modeling of their extended UV-FIR/radio spectral energy distributions (SEDs). The sizes of high-z massive galaxies are investigated by the stacking analysis of the available HST imaging in Section 6. In Section 7, we discuss the implications of our results in the context of formation mechanisms and evolutionary pathways inferred for the most massive galaxies in the universe. The summary is presented in Section 8. The standard ΛCDM cosmological parameters $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$ with $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, and a Chabrier (2003) initial mass function (IMF) are assumed throughout this paper. All magnitudes are in the AB system.

2. Data and Sample

This analysis is based on the K_S-selected catalog constructed over the UltraVISTA near-infrared imaging survey (McCracken et al. 2012) in the COSMOS (Scoville et al. 2007) field, covering an area totalling ≈0.84 deg^2 in the four strips of *ultra-deep* UltraVISTA imaging (Data Release 3), where both IRAC and optical coverage is available (covering the three main *deep* strips and half of the final deep strip). Briefly, the DR3 photometric catalog used in this work was constructed using the same procedure as in Muzzin et al. (2013b), but the NIR depths are ≈1.2 mag deeper compared to the DR1 release, reaching $Y = 25.8$, $J = 25.6$, $H = 25.4$, and $K_S = 25.2$ mag (5σ, $2\farcs1$ diameter aperture).

In Section 2.1, we summarize the galaxy catalog used to obtain the mass-complete galaxy sample at $3 < z < 6$ ($M_* \geqslant 10^{11} M_\odot$), and describe the quality cuts employed to arrive at the *bona fide* sample of massive galaxies in Section 2.2 (105 and 23 galaxies at $z < 4$ and $z > 4$, respectively). We take advantage of the wealth of photometry across a range of wavelengths in the COSMOS field in order to reliably constrain the properties of massive galaxies at $3 < z < 6$ (e.g., determining the AGN hosts, constraining the dust-obscured SFRs). The ancillary multiwavelength source catalogs used in this analysis and counterpart identification methods are outlined in Section 2.3.

2.1. UltraVISTA DR3 Photometric Catalog

In addition to the 30 photometric bands used in the construction of the UltraVISTA DR1 K_S-selected catalog presented in (Muzzin et al. 2013b), we include the ultra-deep optical imaging (≈27.0 mag AB) from the Subaru Hyper Suprime-Cam (DR1; Aihara et al. 2018). Particularly crucial for identifying evolved galaxies at early cosmic epochs, this catalog is complemented by the full-depth Spitzer/IRAC mosaics built combining S-COSMOS (Sanders et al. 2007), the Spitzer Large Area Survey with Hyper-Suprime-Cam (Mehta et al. 2018), and SMUVS (Caputi et al. 2017; Ashby et al. 2018), pushing the available IRAC imaging ∼1 mag deeper at 3.6 μm and 4.5 μm over the ultra-deep stripes (compared to DR1). The catalog also includes UVISTA NB118, five CFHTLS deep, and NMBS filters (Whitaker et al. 2011), totalling 49 bands where there is overlap. The total K_S-band 90% completeness limit adopted in this work is $K_S = 24.5$ mag, corresponding to a stellar-mass-completeness limit of $M_* = 10^{11} M_\odot$ at $z = 6$ (Figure 1).

Photometric redshifts (z_{phot}) are calculated for each entry in the K_S-selected catalog using the photometric redshift code EAZY (Brammer et al. 2008). Briefly, EAZY fits the SEDs using linear combinations of template SEDS. The template set employed here is comprised of those derived from the PEGASE models (Fioc & Rocca-Volmerange 1999), a red template from the models of Maraston (2005), a post-starburst Bruzual & Charlot (2003) model, as well as a template to account for galaxies that are both old and dusty. Photometric redshifts were determined with EAZY allowing solutions in the

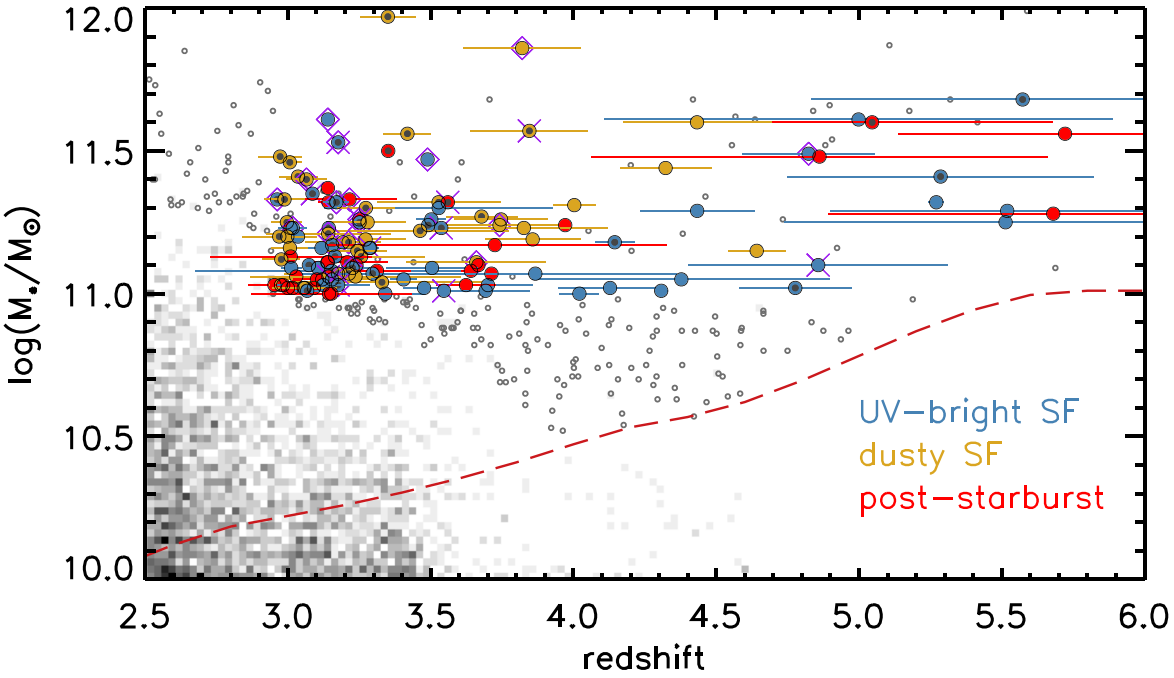

Figure 1. Galaxy stellar masses as a function of redshift. The gray scale represents galaxies down to the 90% K_S-band completeness limit of the UltraVISTA DR3 catalog. Empty circles are used to highlight objects in low-density regions above the adopted 95% mass-completeness limit, which is shown by the red dashed curve. Large symbols indicate the *bona fide* mass-complete ($\log(M_*/M_\odot) > 11.0$) galaxy sample (Section 2.2), colored according to the classification scheme in Section 3.1: *red, blue*, and *gold* for post-starburst, UV-star-forming, and dusty star-forming subsamples, respectively. Galaxies in our sample with X-ray and radio identified counterparts are indicated with X and diamond symbols, respectively. Small (central) black points indicate those with $>3\sigma$ Spitzer 24 μm band detections.

range $0 < z < 6$ using the `v1.0` template error function and $K_{S,tot}$ magnitude prior.

Similar to Muzzin et al. (2013b), stellar population parameters for catalog sources were estimated using the FAST code (Kriek et al. 2009) to fit galaxy SEDs with Bruzual & Charlot (2003) stellar population synthesis models. We assume solar metallicity, a Chabrier (2003) IMF, and the Calzetti et al. (2000) dust-extinction law. The set of template SEDs was constructed adopting exponentially declining star formation histories (SFHs) of the form SFR $\propto exp(-t/\tau)$, where τ is the *e*-folding star formation timescale. We adopted a grid for $\log(\tau)$ between 7.0 and 10.0 in steps of 0.1 dex and for dust attenuation (A_V) between 0 and 5 mag in steps 0.1 mag. The time since the onset of star formation (t) was allowed to range between 10 Myr and the maximum age of the universe at the redshift of sources in steps of 0.1 dex. All galaxies were fit assuming the z_{phot} EAZY solutions, or spectroscopic redshifts when available. The stellar masses (M_*) are determined by multiplying the mass-to-light ratio of the best-fit SED by its corresponding normalization factor.

Figure 1 shows the stellar mass of galaxies down to the 90% $K_{S,\ tot}$-band completeness of the survey ($K_S = 24.5$, in gray scale). We estimate the redshift-dependent mass-completeness limit of the survey following the approach outlined in Marchesini et al. (2009) and also adopted in Tomczak et al. (2014). Briefly, we first scaled the stellar mass of each galaxy down to the K_S-band 90% completeness limit. Then, in each redshift bin, the upper envelope below which 95% of the galaxies are positioned is estimated to be the 95% mass-completeness limit. This empirical 95% mass-completeness limit is indicated as the red dashed curve in Figure 1.

The resulting stellar-mass-completeness limit we adopt in this work is $M_* = 10^{11} M_\odot$ over the targeted redshift range $3.0 \leqslant z < 6.0$. Large colored symbols in Figure 1 indicate the *bona fide* sample of high-redshift massive galaxy candidates that have passed our initial quality assessments (summarized in Section 2.2).

2.2. Sample Selection

From the DR3 galaxy catalog outlined in 2.1, we selected all targets with stellar masses ($M_* \geqslant 10^{11} M_\odot$) and z_{phot} consistent with $3 < z < 6$ (within uncertainties), yielding ≈180 objects. We visually inspected the NIR image stamps, and excluded targets near bright stars with potential contaminated photometry (∼30 targets). Targets with poor fits to the observed photometry were removed after checking the photometric redshift distribution estimated by EAZY and the best-fit SEDs obtained using FAST (eliminating another ∼20 objects, the majority of which were quasars or diffraction spikes). This yields a sample of 105 and 23 galaxies selected at $3 < z < 4$ and $4 < z < 6$, respectively. We find that some objects in this sample have photometric redshift solutions that hit the $z_{max} = 6$ boundary; therefore we rerun EAZY and FAST extending $z_{max} = 7$.

2.2.1. Spectroscopic Redshifts

The high-z massive galaxy sample includes twelve objects with spectroscopic redshift measurements in the literature. The majority of redshifts are obtained using deep NIR spectroscopy: "C1-23152" in Marsan et al. (2015) at $z_{spec} = 3.351$, "C1-15182" in Marsan et al. (2017) at $z_{spec} = 3.370$, and "ZF-COSMOS-20115" presented in Glazebrook et al. (2017) at $z_{spec} = 3.717$. Eight additional objects from the sample of massive galaxies at $3 < z < 4$ were targeted with follow-up Keck-MOSFIRE spectroscopy as part of the MAGAZ3NE Survey, yielding seven $z_{spec} > 3$ redshift confirmations, while one was found to be $z_{spec} = 2.80$ (Forrest et al. 2021). Two

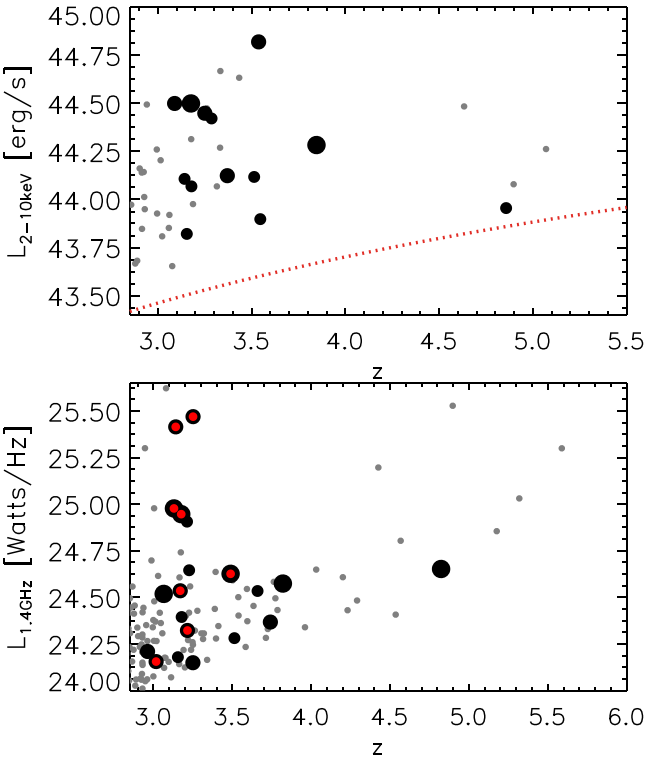

Figure 2. Results of cross-matching the K_S-selected UltraVISTA DR3 sample with the Chandra COSMOS-Legacy (Civano et al. 2016) and VLA-COSMOS 3 GHz Large Project (Smolčić et al. 2017) source catalogs. Targets among the *bona fide* massive galaxy sample with X-ray or Radio counterparts are shown with larger filled circles, with the symbol size correlating with stellar mass. Top: rest-frame X-ray luminosity ($L_{2-10keV}$) of DR3 sources with X-ray counterparts identified $r < 2''$, as a function of redshift. Bottom: rest-frame 1.4 GHz luminosity ($L_{1.4GHz}$) of DR3 sources with 3 GHz counterparts within $r < 0.''75$. Red points indicate objects where the integrated radio emission is likely dominated by AGN activity.

galaxies have high-quality redshifts from the DEIMOS 10k spectroscopy catalog (Hasinger et al. 2018; at $z_{spec} = 3.086$ and 3.176; six additional galaxies have $z_{spec} > 3$, although with lower-quality flags). The redshifts for these twelve objects are fixed to the z_{spec} when performing SED fits.

2.3. Identifying Counterparts in Ancillary Datasets

We complement the DR3 photometry with the rich set of ancillary data in the COSMOS field assembled across the electromagnetic spectrum. These will be used to constrain the dust-obscured star formation rates and investigate the prevalence of AGN. Appendix A provides the details of the matching between the DR3 photometric catalog and the X-ray catalog from Civano et al. (2016) and the radio catalog from Smolčić et al. (2017), as well as the analysis to estimate X-ray and radio luminosities. The top panel of Figure 2 shows the calculated $L_{2-10keV}$ as a function of redshift for all objects in the DR3 stellar-mass-complete galaxy sample as a function of redshift. The red curve indicates the corresponding $L_{2-10keV}$ limit assuming the quoted X-ray source catalog depth. The bottom panel of Figure 2 shows the calculated $L_{1.4GHz}$ as a function of redshift for all objects in the DR3 stellar-mass-complete galaxy sample as a function of redshift. Red points mark objects whose integrated radio emission exceeds predictions from being powered by star formation alone (see Section 3.4).

Given the increasing importance of the dusty star-forming (dSF) population among the most massive galaxies at early epochs (Martis et al. 2016), observations at FIR or submillimeter wavelengths are crucial to understanding the processes that drive the rapid formation and assembly of these *monsters*. In order to get a better handle on the amount of star formation occurring behind dust, we supplement the UltraVISTA DR3 (UV-IRAC) photometry of high-z massive galaxies (Section 2.1) with cross-matched observations at mid-infrared wavelengths and longer. In Appendix A, we describe in detail the IR data sets used to model the full UV-to-FIR SEDs assuming energy balance (Section 5). Briefly, Spitzer-MIPS 24 μm fluxes were extracted, and we supplemented the UltraVISTA DR3 photometry with Herschel-PACS (at 100 and 160 μm) and SPIRE (at 250, 350, and 500 μm) fluxes, SCUBA-2 850 μm fluxes, and the Atacama Large Millimeter Array (ALMA) fluxes at 870 μm (Band 7) and 1.2 mm (Band 6).

We identify ($>3\sigma$) MIPS 24 μm counterparts for 54 objects in the *bona fide* high-z massive galaxy sample. Close to half of the matched sample have a low signal-to-noise ratio (S/N) 24 μm fluxes (21 with $3\sigma < S/N_{24\mu m} < 5\sigma$). While this band is useful to estimate SFR obscured by dust in galaxies, the observed 24 μm band samples increasingly shorter wavelengths at higher-z, probing the $\sim$6 μm by $z \sim 3$ (thermal emission from hot dust), arguably contaminated by emission from obscured AGN (Marchesini et al. 2010; Cowley et al. 2016; Alcalde Pampliega et al. 2019), leading to overestimated SFRs (Marsan et al. 2017; Martis et al. 2019).

None of the *bona fide* high-z massive galaxy candidates are robustly detected ($>3\sigma$) at Herschel-PACS 100 μm, while only two are detected at 160 μm (with $S/N \approx 4\sigma$–6σ). 20 candidates have $>3\sigma$ counterparts identified in *at least one* of the SPIRE bands, whereas only 7 are robustly detected in *all three* (250 μm, 350 μm, and 500 μm).

We find 850 μm counterparts for 26 objects in the *bona fide* massive galaxy sample (4 at $z_{peak} > 4$) from the SCUBA-2 COSMOS survey (S2COSMOS; Simpson et al. 2019). Only 10 out of these 26 targets are also detected in the MIPS 24 μm band (3 at $<5\sigma$ significance).

We find ALMA counterparts for 17 galaxies in the *bona fide* sample by cross-matching with the A^{3}COSMOS data set (Liu et al. 2019). Three of these objects comprise multiple components: two submillimeter/AzTEC sources ("J1000 +0234" and "AzTEC-5"; see Toft et al. 2014; Gómez-Guijarro et al. 2018) and the curious "Jekyll and Hyde" system at $z_{spec} = 3.717$ (see Simpson et al. 2017; Schreiber et al. 2018b).

3. Characterizing the $3 < z < 6$ Massive Galaxy Sample

3.1. Classification Scheme

Given the exquisite sampling of SEDs provided by UVISTA-DR3, we decide to come up with our own method to classify the K_S-selected $3 < z < 6$ massive galaxy population. We attempt to broadly dissect this population based on the shapes of their rest-frame UV-optical SEDs, primarily probing emission from their stellar component (see also Spitler et al. 2014; Forrest et al. 2018).

Below we describe the basis on which we classified the UMGs:

1. *Dusty star-forming*: SEDs with little to no observed UV fluxes, where primarily the rest-frame optical break is

observed (redshifted to observed NIR wavelengths) with a plateau observed in the IRAC bands due to dust-heated radiation.
2. *Old and quiescent:* SEDs with little to no observed UV fluxes, where primarily the rest-frame optical break is observed. Although a more *rounded bump* is present in the IRAC bands, *we do not find strong evidence for this type of SED*.

If *both* breaks are apparent with a *sharper* rest-frame optical break (i.e., *concave*-like SEDs in the IRAC bands):

1. *UV-bright star-forming*: SEDs with more prominent rest-frame UV fluxes, i.e., $peak(f_{\lambda,\mathrm{UV}}) \gtrsim peak(f_{\lambda,\mathrm{opt}})$.
2. *Post-starburst*: SEDs dominated by O/A stars and $peak(f_{\lambda,\mathrm{UV}}) < peak(f_{\lambda,\mathrm{opt}})$.

Figure 3 shows examples of high-z massive galaxies to illustrate the galaxy classification scheme outlined above (also see Figure 6 in Forrest et al. 2018). This classification is consistent with the positions of galaxies on the $U-V$ versus $V-J$ diagram (see Section 3.3). In the left panel of Figure 4, we display the z_{peak} histogram for the *bona fide* high-z massive galaxy sample. The dSF, UV-star-forming (UV-SF), and post-starburst (pSB) subsamples are shown in gold (outlined), blue (filled), and red (hatched), respectively. Given the faintness of the sample, we tested for the robustness of redshift estimates using a Monte Carlo approach. 100 photometric catalogs were simulated by allowing the photometry at each observed filter to vary given the formal errors. The simulated catalogs were each fit with EAZY separately, and we calculated the standard deviation for the z_{peak} (σ_z) distribution for each galaxy.

The right top (bottom) panel in Figure 4 shows the $\sigma(z_{\mathrm{peak}})/(1+z)$ histogram for the *bona fide* $3<z<4$ ($4<z<6$) massive galaxy sample. At $3<z<4$, the $\sigma(z_{\mathrm{peak}})$ distribution aligns with the $1\,\sigma$ z_{peak} errors obtained for the fiducial EAZY runs (majority $\delta(z)<0.2$). The redshift probability distributions appear well constrained, especially for the UV-SF and pSB population, as most of these objects show both the Lyman and the optical breaks (given the young ages of the population, the latter is mostly due to the Balmer features rather than 4000 Å break). At $z>4$, however, there exists a significant fraction of massive galaxies with $\sigma(z_{\mathrm{peak}})/(1+z)>0.15$. These objects are largely undetected outside of the (broad and widely separated) H, K_S, and IRAC bands, which prohibits noteworthy breaks in the observed SEDs to be identified. This, combined with the rather ambiguous SED shapes of dusty galaxies, naturally leads to the larger uncertainties in redshift estimates (i.e., dust and redshift being degenerate parameters in SED fitting). We find that, although the massive galaxy z_{phot} are increasingly uncertain, they are not significant enough to affect the number of selected objects.

3.2. Stellar Populations

Table 1 displays the range of key stellar population parameters obtained for the *bona fide* mass-complete galaxy sample, separated according to their UV-NIR SEDs in redshift bins of $3<z<4$ and $4<z<6$. The reported values correspond to the median and scatter calculated in each subsample. We will further explore the star formation activity for the sample in Section 5 by incorporating constraints at longer wavelengths with MAGPHYS. We refer the reader to Appendix C.1 for a more detailed comparison of the best-fit stellar population parameters obtained using these two commonly employed SED modeling methods. Briefly, the distributions of A_V and stellar age estimates are in relatively good agreement using either; however, the dSF galaxies have stellar masses offset to larger values ($\sim$0.2 dex) by MAGPHYS.

Although more detailed follow-up data is required for a precise analysis of SFHs, given the sample size, we specifically explore trends in dust content, stellar ages, and SFH timescales. The subsamples at $3<z<4$ are characterized by similar distributions in stellar-mass ($\log(M_*/M_\odot)\approx 11.1$) and SFR-weighted ages ($\sim$0.5–0.9 Gyr, albeit with large scatter), implying that a large fraction of their stellar masses had formed by $z\sim 5$. The dust-extinctions and e-folding timescales (inferred using the *fiducial* exponentially declining SFH), however, show a range across the sample. As expected, the dust obscuration is largest for the dSF galaxies ($A_V\sim$ 2–4 mag), while the UV-SF and pSB population is characterized with $A_V<2$ mag. The pSB population is distinguished by their very short inferred star formation timescales ($\tau << 150$ Myr), while the actively star-forming galaxies (both UV-bright and dust) appear to have more extended actively star-forming epochs ($\tau > 400$ Myr).

At $4<z<6$, the median SFR-weighted stellar ages are $\gtrsim$0.5 Gyr, suggesting formation redshifts as early as $z\sim$ 7–8. Similar trends are observed for dust-extinction and star formation timescales as at $3<z<4$, although with much more scatter: dust extinction increases across the population (pSB to UV-SF to dSF), and the pSB galaxies have very short star formation timescales ($\tau << 50$ Myr).

It is worth noting that, based on their UV-NIR SEDs, and $U-V$ versus $V-J$ colors, our *bona fide* sample lacks a population of old and quiescent galaxies, which seems to rule out quenching to not occur much before $z\approx$ 3–5 in the most massive galaxies at these epochs (also see Carnall et al. 2020).

3.3. Rest-frame U – V versus V – J Colors

When large samples with deep spectroscopy are lacking, a common approach is to use rest-frame $U-V$ versus $V-J$ colors (Williams et al. 2009) to separate galaxy populations, especially at high-z (Muzzin et al. 2013a; Whitaker et al. 2013; Straatman et al. 2014). Rest-frame colors were measured using the best-fit EAZY templates, as described in Brammer et al. (2009; Muzzin et al. 2013a). Briefly, rest-frame colors are calculated from the best-fit EAZY template. As one might expect, the robustness of calculated rest-frame colors depends on (1) how realistic of a description the chosen template set is for the specific galaxy population and (2) how well sampled the galaxy SED is, at the corresponding wavelengths for the filters of color calculations.

The top panels in Figure 5 display the rest-frame $U-V$ versus $V-J$ colors at $3.0<z<4.0$ (left) and $4.0<z<6.0$ (right). The gray scale represents galaxies above the 95% stellar-mass-completeness limits in the DR3 catalog (gray scale); filled points mark the *bona fide* sample of massive galaxies, colored according to classification established in Section 3.1 (*gold:* dusty-SF, *blue:* UV-SF, and *red:* pSB). The rest-frame color cuts from Whitaker et al. (2015) are shown with the thick solid (diagonal) and light gray dashed (at $U-V=1.3$) demarcations. The thick dashed lines show the color cuts we introduce to broadly separate the population

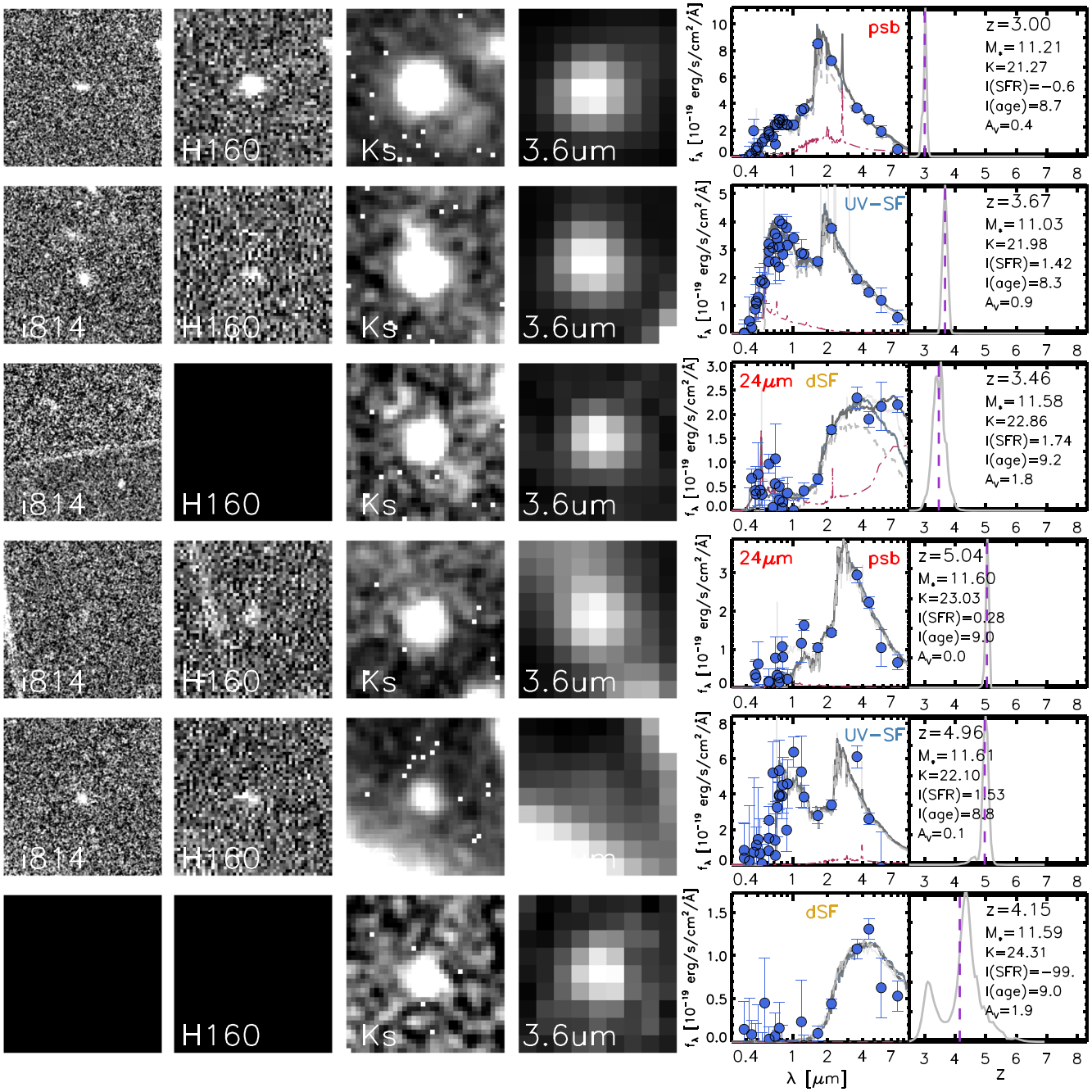

Figure 3. Examples of massive galaxies at $3 < z < 4$ (first three rows) and $4 < z < 6$ (last three rows) sorted according to the classification scheme outlined in Section 3.1. First four columns correspond to $5'' \times 5''$ cutouts from available HST (i_{814} and H_{160}) imaging, VISTA K_S, and Spitzer IRAC 3.6 μm imaging. The panels in column five display the observed UV-IRAC photometry, along with best-fit FAST SEDs for each object (*gray curves*). The best-fit AGN templates (see Section 4.1.2) are indicated with red dotted–dashed curves. The galaxy *type* and Spitzer 24 μm $> 5\sigma$ detections are listed in these panels. The last column shows the photometric redshift probability distribution, with the z_{peak} indicated by the vertical purple dashed lines. Best-fit (fiducial) FAST stellar population parameters are also listed.

based on their SED classification: the *quiescent wedge* is expanded to include pSB galaxies by removing the $U - V > 1.3$ requirement (extension shown); a cut at $U - J = 3.2$ is used to further distinguish between dusty and relatively unobscured star-forming galaxies. The dust vector is shown for an extinction of $A_V = 1$ mag assuming Calzetti et al. (2000) extinction curve.

The errors on the rest-frame color estimates were calculated employing a Monte Carlo approach (hence the uncertainty in redshift estimates are folded in). The standard deviation of rest-frame colors obtained in this manner for the sample are shown with the error bars in the top panels of Figure 5. The objects in the massive galaxy sample with MIPS 24 μm ($S/N \geqslant 3$) detections are indicated by small black filled circles, while purple X and diamond symbols indicate those with X-ray and radio identified counterparts, respectively.

The bottom panels in Figure 5 also show the color–color evolution tracks for different SFHs with varying degrees of dust extinction. Solid tracks represent the color evolution for SFHs assuming no dust ($A_V = 0$), while dashed curves indicate

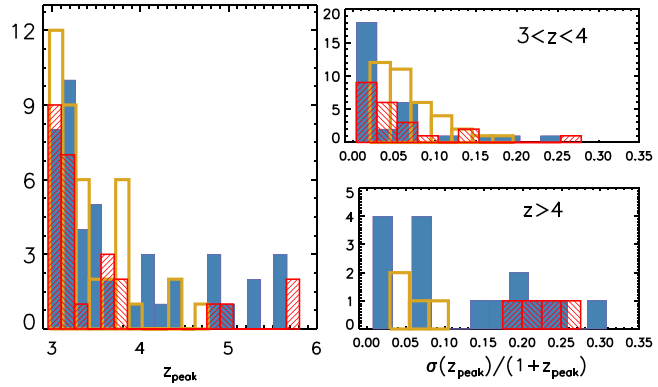

Figure 4. Left panel: z_{peak} distribution for the *bona fide* high-z massive galaxy sample, colored according to our SED classification scheme: dSF—gold (outlined), SF—blue (filled), and pSB—red (hatched) histograms. Right top (bottom) panel shows the distribution for the fractional error in the photometric redshifts ($\sigma_z/(1+z_{peak})$), measured by perturbing and refitting the UV-IRAC photometry of each object) for the *bona fide* $3<z<4$ ($4<z<6$) massive galaxy sample.

Table 1
Stellar Population Parameters for the *Bona Fide* Sample of Massive Galaxies at $3<z<4$ and $4<z<6$

	pSB	UV-SF	dSF
$3<z<4$	$N=29$ (28%)	$N=39$ (37%)	$N=37$ (35%)
$\log(M_*/M_\odot)$	11.09 (0.2)	11.1 (0.14)	11.2 (0.22)
A_V	0.6 (0.44)	1.1 (0.37)	2.2 (0.60)
$\log(age)$[a]	8.72 (0.27)	8.86 (0.33)	8.94 (0.45)
$\log(\tau)$[b]	7.9 (0.51)	8.8 (0.73)	8.70 (1.08)
$4<z<6$	$N=4$ (17%)	$N=15$ (65%)	$N=4$ (17%)
$\log(M_*/M_\odot)$	11.52 (0.12)	11.25 (0.22)	11.38 (0.17)
A_V	0.15 (0.15)	0.80 (0.46)	1.55 (0.54)
$\log(age)$[a]	8.79 (0.08)	8.82 (0.23)	8.95 (0.08)
$\log(\tau)$[b]	7.45 (0.50)	8.50 (0.65)	8.60 (0.76)

Notes. Listed values are medians and standard deviation of the subsample in each redshift bin computed by model UV-IRAC SEDs using FAST.
[a] Denotes SFR-weighted mean stellar ages calculated for an exponentially declining SFH, where $[age]=$ yr; see Förster Schreiber et al. (2004).
[b] Indicates the e-folding star formation timescale ($[\tau]=$ yr) for an exponentially declining star formation history.

the dust-attenuated tracks (A_V values labeled in panels). Both bottom panels include the tracks for an exponentially declining SFH with $\tau=100$ Myr and $A_V=0$ (solid purple curves), a constant SFH ($A_V=0$ and 2.5 mag: orange solid and dashed curves). We highlight that the $\tau=100$ Myr color evolution is quite similar to the track of a SSP (solid gray track in right panel), with the latter reaching the same $U-V$ and $V-J$ colors as the $\tau=100$ Myr model at slightly younger ages (by a few hundred Myr). The bottom left panel ($3<z<4$) displays tracks for exponentially declining SFHs with $\tau=100$ Myr SFH with $A_V=0.5$ (dashed purple) and $\tau=500$ Myr ($A_V=0$ and 1.0 mag: solid and dotted–dashed green curves), while the right panel ($4<z<6$) shows tracks for $\tau=300$ Myr ($A_V=0$ and 1.5 mag: solid and dotted–dashed teal curves). Small open circles along the color evolution tracks correspond to age steps since the onset of star formation. The color tracks are plotted up to the maximum allowed age of the universe corresponding to the lower limit of each redshift bin, and empty circles mark model colors at the specified ages (in Gyr).

At $3<z<4$ (Figure 5, left panels), the rest-frame colors of massive galaxies span a wide range in *UVJ* space, similar to the large spread (>1.5 mag) found for a mass-complete ($\log(M_*/M_\odot)>10.6$) sample of galaxies at $3<z<4$ (Spitler et al. 2014). The massive galaxy sample at $3<z<4$ appears to be equally distributed (in numbers) in each subgroup. The classification of massive galaxies based on their observed SEDs (outlined in 3.1) aligns with expectations from the estimated rest-frame colors, and the separation between the dSF (gold) and UV-SF (blue) massive galaxies is reassuringly perpendicular to the dust vector (A_V) and quiescent wedge (see also Alcalde Pampliega et al. 2019). The dSF massive galaxies are red in both $U-V$ and $V-J$ colors, while the pSB massive galaxy population (red points) is located primarily toward bottom-left of the quiescent wedge, indicating young stellar ages (Belli et al. 2019; Carnall et al. 2020).

It is not surprising that a substantial fraction of the *pSB*-tagged massive galaxies at $3<z<4$ are just outside of the traditional UVJ boundary, as this criteria is known to be a relatively pure (but likely incomplete) method for selecting quiescent galaxies at high-z (although see e.g., Martis et al. 2019; Roebuck et al. 2019). Studies have found that the horizontal cut at $U-V=1.3$ may be increasingly irrelevant at identifying passive galaxies at $z>3$, as it is possible for the colors of *dust-free* galaxies where star formation has ceased abruptly $\sim$few $\times$ 100 Myr prior to observation to fall below this cut (Marsan et al. 2015; Merlin et al. 2018; Schreiber et al. 2018c; Forrest et al. 2021), a common occurrence, especially at high redshifts where galaxies had little available time to evolve passively after quenching. The colors of the pSB massive galaxies suggest that at $3<z<4$, observations are converging toward the epoch of quenching for UMGs. This is consistent with the picture that the typical passive massive galaxy at $z\approx3$–4 is a pSB, as also suggested by the analysis of HST WFC3/IR G141 stacked grism spectrum for passive massive ($\log(M_*/M_\odot)\sim11.2$) galaxies at $z\sim2.8$ (D'Eugenio et al. 2020).

Focusing on the higher-redshift bin, $4<z<6$ (Figure 5, right panels), estimates of rest-frame colors are less reliable (evidenced by the larger uncertainties), as the rest-frame V and J bands are redshifted farther into the NIR wavelengths. At $z>4$, the rest-frame V band falls between K_S and IRAC 3.6μm bands, while the rest-frame J band is only probed by ≈IRAC 5.8 μm and 8.0 μm, where the available photometry is significantly shallower. In fact, all of the galaxies in our $z>4$ sample lack significant detections in IRAC 5.8 μm and 8.0 μm photometry ($S/N\lesssim3\sigma$). Although the full SED is taken into consideration when inferring rest-frame colors for galaxies with EAZY, at $z>4$ the photometry especially critical to estimating the rest-frame $V-J$ colors at $z>4$ are limited. Therefore we caution not to overinterpret the implied stellar populations of the red points located in the quiescent wedge.

Although we elected to choose a sample above the 90% mass-completeness limit for UltraVISTA, we note that at $z>4$ our sample may be missing some highly obscured star-forming galaxies by the nature of K_S selection (Caputi et al. 2015; Alcalde Pampliega et al. 2019)—although the redshift distribution and abundance of such objects are still uncertain. We find that the $U-V$ versus $V-J$ diagram is inadequate to constrain the stellar populations of the $4<z<6$ massive

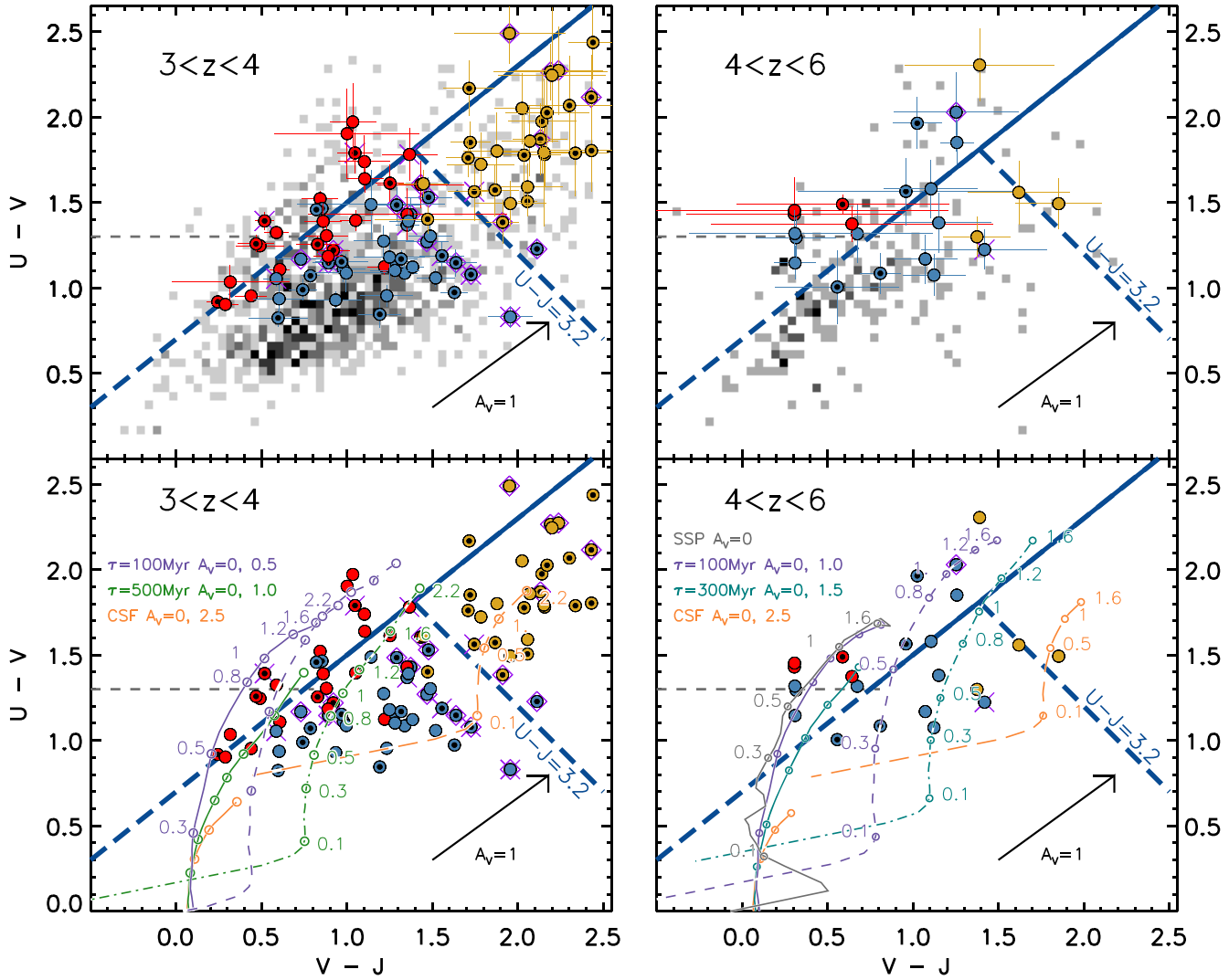

Figure 5. Rest-frame $U-V$ vs. $V-J$ color diagram at $3.0 < z < 4.0$ (left panels) and $4.0 < z < 6.0$ (right panels). The gray scale in the top panels represents the distribution of galaxies above the 95% completeness limits from the K_S selected UltraVISTA DR3 catalog at the corresponding redshifts. *Red*, *blue*, and *gold* filled circles indicate post-starburst, UV-star-forming, and dusty star-forming massive ($\log(M_*/M_\odot) \gtrsim 11$) galaxies, respectively. The thick demarcations show the extended diagonal color–color cut from Whitaker et al. (2015) and a slice at $U-J=3.2$ to broadly separate these three populations. Massive galaxies with MIPS 24 μm detections ($>3\sigma$) are marked with small black filled circles, while purple X and diamond symbols indicate massive galaxies with X-ray and radio counterparts, respectively. Errorbars (top panels) are the 1σ distribution of the colors estimated using an MC approach. Bottom panels show the color evolution tracks of Bruzual & Charlot (2003) models with various SFHs: an SSP (single burst of star formation) with no dust (gray solid curve, right panel); exponentially declining SFHs with e-folding timescales $\tau = 100$ Myr (purple, both panels), 300 Myr (teal, right panel), and 500 Myr (green, left panel) with no dust (solid lines) and $A_V = 0.5$, 1.5, 1.0 mag (dashed curves, same colors for corresponding SFHs), respectively; a constant SFH with no dust (CSF; orange solid curve in both panels) and $A_V = 2.5$ mag (dashed orange curves). The empty circles represent the model colors at the specified ages (in Gyr). The dust vector indicates an extinction of $A_V = 1$ mag for a Calzetti et al. (2000) extinction curve.

galaxy sample owing to the presence of (unconstrained) strong/high-equivalent width optical nebular emission biasing the derived rest-frame V band magnitudes, along with the associated uncertainties in the derived rest-frame J band magnitudes due to the low number of filters probing this spectral region at these redshifts.

3.4. Incidence of Active Galactic Nuclei

We estimate the bolometric AGN luminosity ($L_{\rm bol,AGN}$) for the 13 X-ray counterpart identified objects in our sample (appendix A.1) assuming a bolometric correction factor of $c_{\rm bol} = 10$. Following Marconi et al. (2004), Rigby et al. (2009), and Lusso et al. (2012), this is a conservative average for $c_{\rm bol}$ given the parameter space in luminosity probed by the X-ray matched massive galaxies. Examining the top panel of Figure 2, the resulting $L_{\rm bol,AGN}$ for the 13 X-ray matched galaxies corresponds to $L_{\rm bol}({\rm AGN}) \gtrsim 10^{44}$ ergs^{-1}. Only two have a hardness ratio (HR) < -0.2, indicating that the majority of X-ray detected massive galaxies are obscured AGN.

To identify massive galaxies whose integrated radio emission is likely dominated by AGN activity (as opposed to star formation), we use the redshift-dependent radio excess threshold defined in (Delvecchio et al. 2017); $\log\left(\frac{L_{1.4\,\rm GHz}}{\rm SFR_{IR}}\right) = 21.984 \times (1+z)^{0.013}$. The SFRs we use in this are those obtained with UV-FIR SED modeling with MAGPHYS (Section 5, which does *not* account for AGN contribution). Combining this with the fact that the majority of FIR fluxes used in the MAGPHYS modeling are merely (Herschel) upper limits or low-significance detections ($<3\sigma$), we argue that these values are *upper* limits for realistic SFRs,

and therefore this radio-AGN identification scheme is a conservative one. There are likely more galaxies that harbor radio-AGN than the ones we are selecting. We find that the radio power for 8 out of the 20 radio-counterpart identified massive galaxies (appendix A.2) is consistent with harboring radio-AGN. These objects are indicated with red points in the bottom panel of Figure 2.

In summary, 19 (1) out of our sample of 105 (23) $3<z<4$ ($4<z<6$) massive galaxies show evidence of harboring powerful AGN by the detection of luminous radio or X-ray counterparts. We measure the AGN fraction to be $f_{\rm AGN}=0.15^{+0.05}_{-0.04}$ for the full sample, while splitting the sample into two redshift bins yields $0.18^{+0.05}_{-0.04}$ and $0.04^{+0.04}_{-0.04}$ at $3<z<4$ and $4<z<6$, respectively. Considering instead that *all* the galaxies in our sample with radio counterparts are the AGN hosts would result in $f_{\rm AGN}\approx 20\%$. We note that this estimate should be viewed as conservative lower limits since, effectively, the AGN selection methods we employed for the high-z universe are biased toward brighter objects. These fractions are much lower than what was found in Marsan et al. (2017), possibly because the $3<z<4$ sample here extends to much fainter objects.

To illustrate this, we use the extended sample of galaxies with NIR spectroscopy at $3<z_{\rm spec}<4$ ($N=10$; whereas 5 were in Marsan et al. 2017) and repeat the analysis. Three objects in this sample are identified as AGN via X-ray or radio detections, while 5 are observed to have [OIII]/Hβ ratios and limits consistent with an ionizing source as AGN (Schreiber et al. 2018c; Forrest et al. 2021). Only 2 of the latter 5 have X-ray or radio counterparts, hence leading to 3 additional sources to be classified as AGN by [OIII]/Hβ line ratios alone (one radio-AGN identified target was found to have line ratio consistent instead with star formation; Forrest et al. 2021). In short, considering the rest-frame, the optical line diagnostics increases the inferred AGN fraction in the sample by a factor of 2 ($\sim$30% to $\sim$50%–60%).

4. Number Densities

4.1. Robustness of Stellar-mass Estimates

Here we investigate three potential components in SED modeling which may be biasing the estimated physical parameters of our sample, in order to arrive at a *robust* high-z massive galaxy sample with well-constrained stellar masses. Specifically, we vary some assumptions made either in SED modeling using FAST (different SFH forms, inclusion of AGN) or strength of emission line contamination to photometry.

4.1.1. Impact of Assumed SFHs

One of the main sources of uncertainty in stellar-mass estimates is the lack of knowledge of the SFHs, and it has been shown that the choice of SFH may vary stellar-mass inferences by $\sim$0.1–0.3 dex (Michałowski et al. 2014; Mobasher et al. 2015; Carnall et al. 2019). The galaxy catalog from which we selected the *bona fide* sample of the high-z massive galaxies assumes an exponentially declining form for the SFH, which has been shown to be decent at describing the massive end of the galaxy population (provided they are not starbursts). However, with increasing look-back times, a larger variety of SFHs is needed to model observed SEDs. In this section, we test the influence of the assumed SFH on stellar-mass estimates.

We investigated the impact of the different SFHs on the estimated stellar population parameters by using two additional parametric forms for the SFH (in addition to the regular exponentially declining model): delayed exponentially declining and constant truncated SFH. We also consider a scenario in which a recent burst of star formation is triggered through gas infall. A recent episode of star formation may outshine the underlying older stellar population, causing inaccurate stellar-mass estimates. To investigate this, we created stellar libraries with GALAXEV (Bruzual & Charlot 2003) assuming a two-component SFH. We use exponentially declining SFHs (with $\log(\tau[\rm yr])=7.1,\ 7.3,\ 7.5,\ 7.7,\ 7.9$) to model the old stellar population accounting for the majority of a galaxy's stellar mass, and introduce a burst contributing a fraction of the stellar mass (5%, 10%, 20%, 30%, or 50%) when the older stellar population reaches an age of 0.2, 0.3, 0.7, 1.0, 1.2, or 1.4 Gyr. Using FAST, we fit the observed photometry of our sample with these different SFHs using the same grid in dust content, stellar age, and star formation timescale as in the SFH=exponentially declining case.

Panels in Figure 6 show the resulting stellar-mass estimates obtained assuming three different forms for the SFH versus the stellar masses obtained using the fiducial exponentially declining SFH. For the two-component SFH fits, given the large range in burst/old population component permutations, we impose a χ^2 criterion to compare stellar masses that are (similarly) well fit. The stellar masses corresponding to the two-component SFH fits are indicated with small gray circles in the right panel of Figure 6. We find that applying these three SFHs does not cause a statistically significant offset in stellar-mass estimates; stellar masses are on average 0.01 ± 0.03, 0.01 ± 0.05, and 0.03 ± 0.07 less than the fiducial (exponentially declining) SFH, for the delayed declining, the truncated, and the two-component SFH cases, respectively.

4.1.2. AGN Contamination to Observed UV-IRAC SED

AGN continuum emission can potentially contribute to the rest-frame UV-optical SED, biasing the derived stellar population parameters of the sample (see e.g., Florez et al. 2020). We investigated this by using a version of the FAST code (Kriek et al. 2009) modified to perform a two-component (galaxy +AGN) SED fit, presented in Aird et al. (2018). This two-component fitting approach uses the same grid of galaxy templates as the standard FAST, and a library of eight empirical AGN templates to determine the best-fit SED by marginalizing over the full grid of possible galaxy and AGN combinations. We refer the reader to Aird et al. (2018) for a more detailed description of this approach. Briefly, the AGN templates are comprised of five AGN-dominated SWIRE templates from Polletta et al. (2007) and three composite SEDs of X-ray selected AGNs from Silva et al. (2004). We remodel the observed UV-IRAC photometry of the high-z massive galaxy sample in the same manner as for the fiducial FAST runs (i.e., SFH fixed to exponentially declining parameterization and identical EAZY redshift solutions).

Figure 7 displays the resulting stellar-mass estimates when including AGN templates to the UV-IRAC SED modeling. Although it appears that the stellar masses of some massive galaxies may be biased by >0.2 dex, the effect on the overall population is negligible: including AGN to SED modeling in this manner decreases the stellar-mass estimates on average by 0.04 ± 0.08. We stress that the derived $\log(M_*/M_\odot)$ for

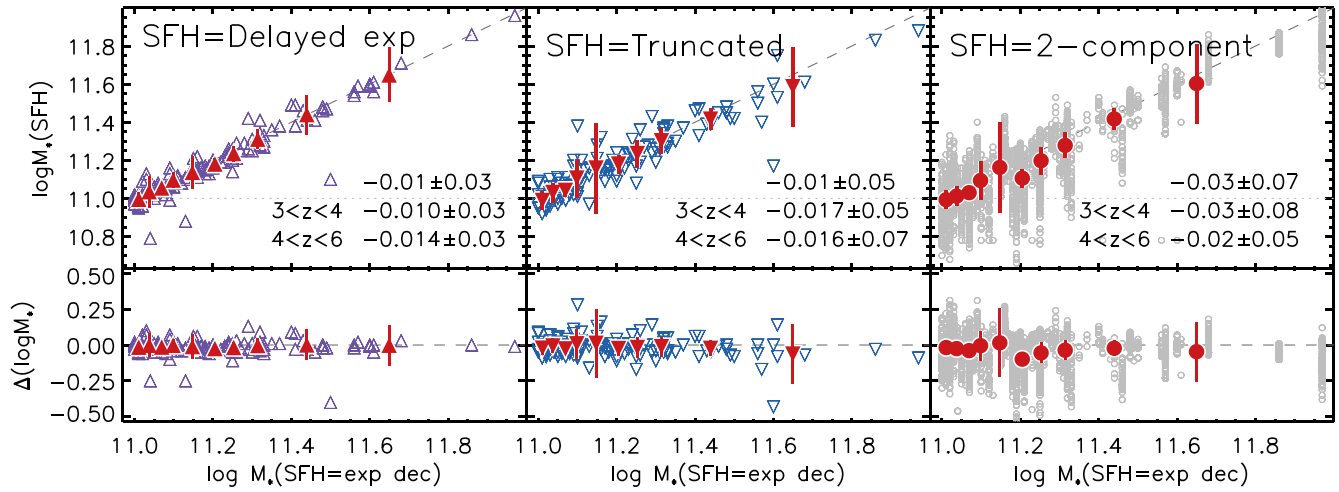

Figure 6. The change in estimated stellar masses when varying the SFHs used to model UV-IRAC SEDs with FAST. *Left*, *central*, and *right* panels display results when assuming the *delayed exponentially declining*, the *constant truncated*, and the *two-component* SFH parameterization vs. the *fiducial* stellar masses used (SFH = exponentially declining). The open symbols indicate individual objects in the *bona fide* $3 < z < 6$ massive galaxy sample. Filled red symbols and error bars mark the running means and their associated standard deviation. The robust mean difference in the obtained stellar masses ($\Delta \log M_* = \log M_{*,\mathrm{SFH}} - \log M_{*,\mathrm{SFH=exp}}$) for the assumed SFHs and their associated standard deviations are listed in the top panels (for the full *bona fide* $3 < z < 6$ massive galaxy sample, and separated in two redshift bins).

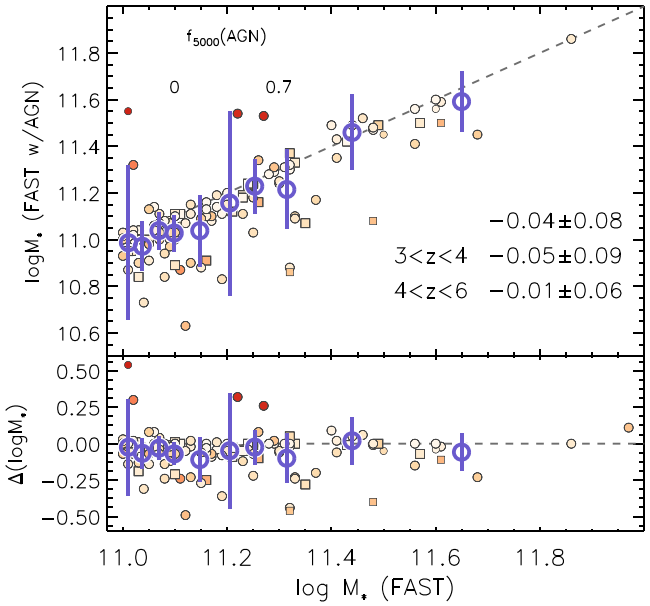

Figure 7. The effect of including AGN templates in the UV-IRAC SED modeling for the *bona fide* $3 < z < 6$ massive galaxy sample (filled symbols). Large open circles mark the running means and their associated standard deviation. The robust mean difference in the obtained stellar masses ($\Delta \log M_* = \log M_{*,\mathrm{FAST}} - \log M_{*,\mathrm{FAST\,w/AGN}}$) and their associated standard deviations (for the full *bona fide* $3 < z < 6$ massive galaxy sample, and separated in two redshift bins) are quoted in the top panel. Square symbols indicate objects that are identified as harboring AGN in Section 3.4.

individual galaxies may not be entirely accurate, as there does not seem to be an obvious offset in stellar-mass estimates for the AGN-tagged massive galaxies (see Section 3.4, square filled symbols in Figure 7). In fact, the objects with the largest estimated AGN contribution to rest-frame optical wavelengths (5000 Å) appear to be massive galaxies that do not have other indications of harboring an AGN. Nevertheless, this should be a sufficient sanity check for the overall sample.

4.1.3. Nebular Emission Line Contamination to Photometry

Obtaining reliable stellar-mass estimates for high-z galaxies requires probing the light arising from older stellar populations (rest-frame optical or longer wavelength light). At $z > 3$, this corresponds to observed near- and mid-IR wavelengths, where the available photometric filters are limited and may be significantly contaminated by emission features (due to increasing line equivalent widths with redshift), yielding overestimated stellar masses (Labbé et al. 2013; Stark et al. 2013). For example, a prominent/strong [OIII] emission feature at $3 < z < 4$ can mimic the NIR colors associated with more massive, evolved galaxies (or the Balmer break of recently quenched galaxies; see Merlin et al. 2018).

At $z > 2.5$, Hα, [NII], and [SII] features fall out of spectral coverage range from ground, and the next most prominent optical spectral feature, [OIII], is redshifted beyond the observed K_S band at $z > 3.7$. Hence, current spectroscopic surveys with direct measurements of the full suite of emission lines and their equivalent widths (EW) are generally limited to the lower-redshift ($z < 3$) universe (i.e., Reddy et al. 2018).

At higher redshifts ($z > 3$), Spitzer-IRAC photometry/colors have been employed to infer the emission line EWs of galaxies (Smit et al. 2014, 2016; Faisst et al. 2016). Additionally, the strength and contribution of emission lines to IR broadband photometry are highly uncertain at $z > 4$, and will have to wait until the launch of JWST.

The galaxy template library used to model the observed SEDs with FAST do not include prescriptions for nebular emission lines, and not accounting for these strong features can bias, in particular, the estimated stellar masses. We investigate the systematic effects that nebular line contamination to photometry may have on the derived parameters implementing the approach similar to that explored in Stefanon et al. (2015); namely, this is accomplished by correcting each galaxy's observed photometry for potential line contamination using independent prescriptions for equivalent width estimates and recomputing the stellar population parameters in adopting the same configuration as in Section 2.1. The three approaches are outlined below:

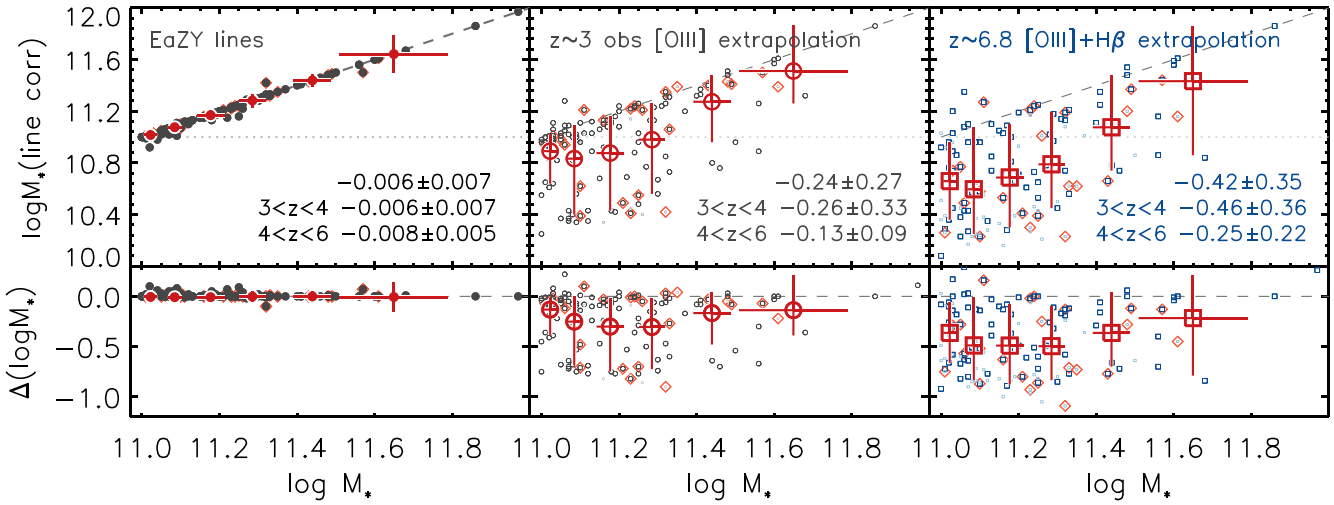

Figure 8. Stellar masses obtained after correcting for nebular line contamination to photometry, using the three prescriptions for estimating equivalent widths in Section 4.1.3. Left: using the EWs computed from EAZY templates. Center: Extrapolating nebular lines based on the range EWs of observed [OIII] emission in spectroscopic studies of $z\sim3$ massive galaxies (K-band spectroscopy). This is the more *realistic* upper limit scenario for emission line contribution. Right: the *extreme* upper limit to the emission line contamination prescription based on extrapolating EW ([OIII]+Hβ) estimated for galaxies at $\sim z=6.8$ by Smit et al. (2014). Larger dark symbols indicate fits where the emission line corrected photometry yields similarly well-fit SED models as the original photometry, while small light symbols mark instances where emission line corrected photometry is not modeled as well. Galaxies tagged as harboring AGN are indicated with orange diamonds. The mean difference in the obtained stellar masses ($\Delta\log M_* = \log M_{*,\rm orig} - \log M_{*,\rm line\ corr}$) and the associated standard deviations are quoted in the top panels.

1. The *minimal* correction: EW recovered from best-fit EAZY templates
2. The *empirical* correction: extrapolating the measured EW([OIII]) for massive galaxies at $z=3$–4
3. The *extreme* correction: extrapolating the estimated EW([OIII]+ Hβ) upper limits at $z\sim6.8$

Figure 8 displays the effect of nebular emission contamination on the derived stellar masses assuming the three different prescriptions for equivalent width estimates outlined above.

In our first approach, we use the best-fit EAZY template, which is constructed using a combination of stellar models from the lower-z universe. This can be thought of as the *minimal* correction to the observed photometry due to emission features, because the EWs used in this approach bias the observed photometry the least. This can be understood as the combination of (1) the rising star formation activity (on average) for galaxies with redshift (as evidenced by the evolution of the star formation main sequence with redshift; e.g., Whitaker et al. 2014) and (2) the lower metallicities expected for galaxies at early cosmic epochs (due to fewer cycles of chemical enrichment and ongoing star formation being fuelled by more pristine gas). We identified bands that could be contaminated by the (redshifted) main nebular emission lines (Lyα, [OII], Hβ, [OIII], Hα, [NII], and [SII]). The potential contribution for each feature was computed from the line EW recovered from the best-fit EAZY template, and the corresponding flux was then rescaled by the factor $R(\lambda_{\rm obs})/(\lambda_{\rm obs}\int R(\lambda)/\lambda d\lambda)$, where $R(\lambda)$ is the filter efficiency (see Equation (1) in Smit et al. 2014).

For a *realistic* emission line correction, we adopt the maximum EW([OIII]) measured from K_S-band spectroscopy of $z\approx3$–3.5 massive galaxies, corresponding to $\rm EW_{\it rest}\approx280\pm20$Å (Marsan et al. 2015, 2017; Schreiber et al. 2018c). This value is consistent with the $\rm EW_{rest}$ directly measured for Hβ + [OIII] and [OII] emitters at $z\sim3.5$ (Khostovan et al. 2015, 2016) in the mass range considered here. Lastly, for the *maximal* emission line correction to photometry, we used the EW([OIII]+ Hβ) ($\rm EW_{\it rest}\approx1550$Å) estimates at $z\sim6.8$ (Smit et al. 2014, 2015). We interpret this last approach as an upper limit for nebular emission line contamination, as it is based on a lower-mass sample with significant excesses observed in the Spitzer-IRAC bands (which by construction require substantial EWs).

In the latter two cases, we assume a simple $(1+z)^\beta$ redshift evolution for $\rm EW_{rest}$([OIII]), where $\beta=1.8$, which has been shown to describe the $\rm EW_{rest}$ evolution for Hα and Hβ+[OIII] (e.g., Sobral et al. 2014; Mármol-Queraltó et al. 2016; Rasappu et al. 2016) out to $z\sim2$, although there is evidence that a shallower ($\beta=1.3$) evolution may be present beyond $z>3$ (e.g., Khostovan et al. 2016; Faisst et al. 2016). To constrain the EW of all the (remaining) nebular emission lines we adopt the line intensity ratios corresponding to $Z=0.004$ (subsolar metallicity, $0.02\times Z_\odot$), see Table 1 of Anders & Fritze-v. Alvensleben (2003). For each galaxy, we computed the (estimated) observed-frame equivalent widths for each emission line considered. The necessary correction factor for each emission line was determined by comparing the estimated $\rm EW_{obs}$ to the bandwidths of the corresponding filters that the features will fall in at the redshift of galaxies.

The left, middle, and right panels of Figure 8 display the stellar masses obtained after correcting the observed photometry assuming *minimal, empirical,* and *maximal* emission line contamination to observed photometry. We impose a $\Delta\chi^2$ criterion to only keep values obtained by similarly good fits. We indicate the fits that satisfy $\chi^2_{\rm line\ corr}\leqslant\chi^2_{\rm orig}+1$ by the larger and darker symbols in each panel, respectively. The average decrease in estimated stellar masses caused by the assumed *minimal*, *realistic,* and *extreme* emission line strength prescriptions are 0.006 ± 0.007, 0.24 ± 0.27 and 0.42 ± 0.35, respectively.

4.2. Massive End of the Stellar-mass Function at $3<z<6$

In the previous section, in order to infer high-confidence measurements for the abundance of massive galaxies in the

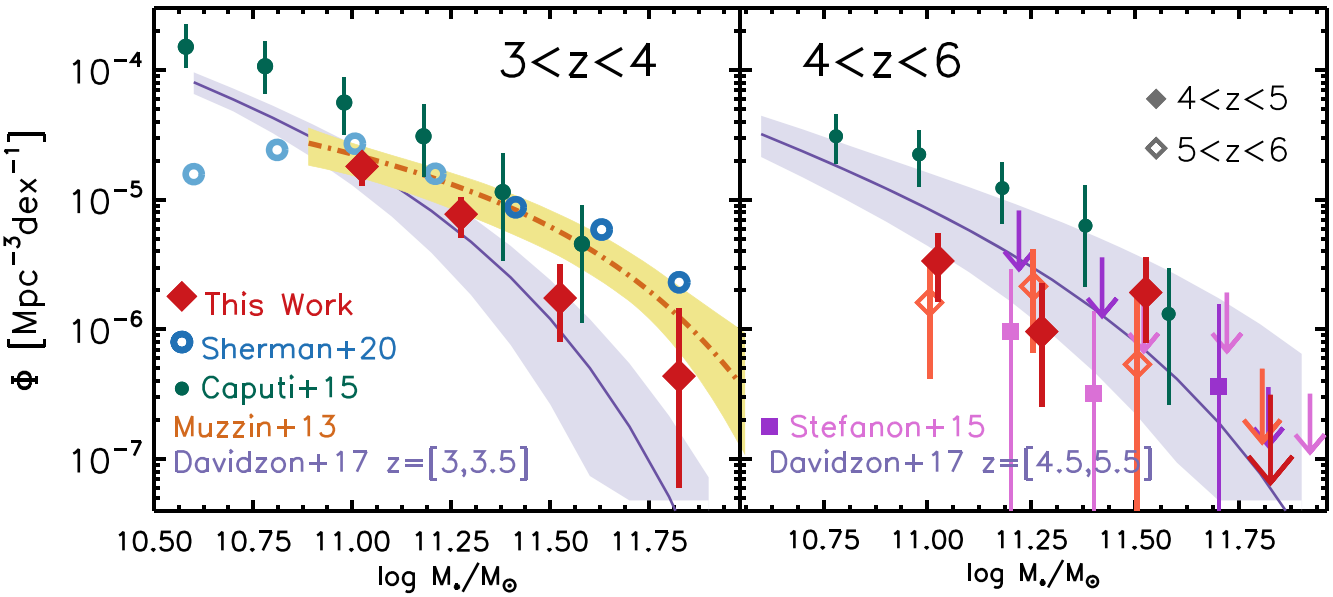

Figure 9. Stellar-mass functions at $3 < z < 4$ (left panel) and $4 < z < 6$ (right panel). Diamonds indicate calculations using the $1/V_{\rm max}$ method for the *robust* sample of high-z massive galaxies. The right panel marks measurements at $4 < z < 5$ and $5 < z < 6$ with filled and open diamonds, respectively. Orange dotted–dashed and purple solid curves correspond to high-z SMFs (1σ errors indicated by shaded regions) previously calculated for galaxies in the COSMOS field from Muzzin et al. (2013a) ($3 < z < 4$, left panel) and Davidzon et al. (2017) ($3 < z < 3.5$, $4.5 < z < 5.5$). Green filled circles show measurements from Caputi et al. (2015) at $3 < z < 4$ (left panel) and $4 < z < 5$ (right panel). Small blue open circles in the left panel are measurements from Sherman et al. (2020) for star-forming massive galaxies ($3 < z < 3.5$, darker symbols above their 80% mass-completeness limit). Pink and purple filled squares in the right panel show the SMF from Stefanon et al. (2015) computed at $4 < z < 5$ and $5 < z < 6$, respectively.

early universe (4.1.), we demonstrated that varying the assumed SFH form (Figure 6) or including AGN (Figure 7) in SED fits does not bias the stellar-mass estimates for galaxies in our sample, while the contribution of emission lines to photometry (Figure 8) may significantly alter their stellar-mass distribution. Thus, we use the results from different emission line correction prescriptions to constrain the number density of very massive galaxies at $z > 3$. Specifically, the empirical emission line contamination scenario is used to infer the robust number densities for the most massive galaxies at $3 < z < 6$. The *bona fide* (Section 2.2) and *maximally* emission line corrected (4.1.3) samples are used to serve as reliable upper and lower limits for the number density calculations (see Figure B1).

We estimate the abundance of massive ($\log(M_*/M_\odot) > 11$) high-z galaxies using the $1/V_{\rm max}$ formalism (Avni & Bahcall 1980), dividing the sample into two redshift bins: $3 < z < 4$ and $4 < z < 6$, corresponding to approximately equal slices in look-back time (≈620 Myr). Upper and lower limits to number counts were calculated using the small sample recipe of Gehrels (1986). We estimate the contribution of cosmic variance to the total error budget using the prescription of Moster et al. (2011). The total area covered by UltraVISTA/DR3 is ≈0.84 deg^2, corresponding to ∼24%, ∼35%, and 50% fractional uncertainty at $3 < z < 4$, $4 < z < 5$, and $5 < z < 6$, respectively for $\log(M_*/M_\odot) \approx 11.25$ galaxies. We assumed an additional 5% fractional error on the cosmic volumes probed in the redshifts considered. The final uncertainties on density estimates were calculated by summing the above uncertainties in quadrature.

Panels in Figure 9 display the calculated SMF at $3 < z < 4$ and $4 < z < 6$, diamonds indicate SMF densities computed for the *robust* high-z massive galaxy sample, and include observed SMFs from literature for comparison. Orange dotted–dashed and purple solid curves correspond to the high-z SMFs previously calculated for galaxies in the COSMOS field from Muzzin et al. (2013a) ($3 < z < 4$) and Davidzon et al. (2017) ($3 < z < 3.5$ and $4.5 < z < 5.5$), respectively. Shaded regions indicate 1σ errors. Green filled circles in the left (right) panel show SMF measurements from Caputi et al. (2015) at $3 < z < 4$ ($4 < z < 5$) corrected for a Chabrier (2003) IMF. In the right panel, the SMFs from Stefanon et al. (2015) calculated at $4 < z < 5$ and $5 < z < 6$ are plotted with pink and purple filled squares, respectively. We specifically show their "default" (i.e., no accounting for nebular emission lines) and *luminosity prior +old/dusty template* measurements. We note that Stefanon et al. (2015) computed SMFs assuming several different prescriptions for emission line corrections, and the resulting SMFs are essentially all identical, consisting mostly of upper limits. Measurements from Sherman et al. (2020) reported for the star-forming massive galaxies identified at $3 < z < 3.5$ across 17.2 deg^2 are indicated with small open circles.

At $3 < z < 4$, our measurements are consistent with the SMF from Davidzon et al. (2017), while lower than all other literature points. The excess seen in the SMFs from Muzzin et al. (2013a) and Caputi et al. (2015) can be explained by the different survey depths used in these works, as well as the fact that the stellar masses inferred in these works do not account for emission line contamination (while emission lines are included while modeling SEDs in Davidzon et al. 2017). The latter case is especially obvious at $\log(M_*/M_\odot) < 11.5$, where the SMF calculated using our *bona fide* (not corrected for emission lines) massive galaxy sample is entirely consistent with these works (see Figure B1). Interpreting the observed discrepancy with the SMF from Sherman et al. (2020) is less trivial, but can be attributed to the differences in the galaxy selection methods and photometry used to construct SEDs. While Sherman et al. (2020) do implement nebular emission in their SED modeling, their calculations are based on measurements using (less deep, by >2 mag) observations with 9

Table 2
Cumulative Number Density Measurements for $\log(M_*/M_\odot) \geqslant 11.0$ Galaxies Using the *Bona Fide* Sample and After Correcting for Nebular Line Contamination to the Photometry

Redshift	n [10^{-6} Mpc^{-3}] *bona fide*	Empirical Lines	Extreme Lines
$3 < z < 4$	10.6 ± 2.7	5.6 ± 1.6	2.4 ± 0.8
$4 < z < 5$	1.7 ± 0.8	1.5 ± 0.7	$1.2^{+0.6}_{-0.5}$
$5 < z < 6$	$1.0^{+0.7}_{-0.6}$	$1.0^{+0.7}_{-0.6}$	$0.7^{+0.6}_{-0.5}$

Note. Quoted uncertainties include Poisson errors and the effect of cosmic variance.

photometric bands and do not sample beyond the IRAC channel 2 band, i.e., $\lambda \gtrsim 5\ \mu$m.

At $4 < z < 6$, our measurements are again below the SMF measurement from Caputi et al. (2015) at $4 < z < 5$, and in agreement with the SMF by Davidzon et al. (2017) in this redshift range. Our measurements are also consistent with the SMF reported in Stefanon et al. (2015)—specifically, comparing the open (closed) diamonds indicating our measurements at $4 < z < 5$ ($5 < z < 6$) with pink (purple) squares correspond to identical redshift intervals from Stefanon et al. (2015).

The apparently flatter SMF at $4 < z < 6$ is likely due to a combination of the following: low-number statistics, the extremely large stellar-mass estimates (notice how in Figure 1, the distribution in $\log(M_*/M_\odot)$ is increasingly skewed toward larger values at $z > 4$), and the fact that, on average, stellar masses at $z > 4$ appear to be less affected by contamination from nebular emission lines compared to the lower-redshift bin considered (see inset text in Figure 8). This means that, although the uncertainties in emission line contamination prescriptions may contribute significantly to the total error budget in the number density estimates at $3 < z < 4$, inferences at $4 < z < 6$ still appear to be dominated by low-number statistics.

Table 2 lists the number density calculated for $\log(M_*/M_\odot) \geqslant 11$ galaxies for the *bona fide* massive galaxy sample (Section 2.2) and assuming the empirical and extreme prescriptions used to account for emission line contamination to observed photometry. As expected, number densities computed with the *maximal* emission line contributions are the lowest.

At $3 < z < 4$, the mass-complete ($\log(M_*/M_\odot) > 11$) number density we calculated for the *bona fide* (emission lines not included in SED modeling) sample is consistent with the measurements reported in Muzzin et al. (2013a) ($n = 9^{+6}_{-3} \times 10^{-6}$ Mpc^{-3}), as well as the mass-complete ($\log(M_*/M_\odot) > 11$) sample of galaxies identified in the CANDELS/GOODS fields (Alcalde Pampliega et al. 2019; $14.4 \pm 3.7 \times 10^{-6}$ Mpc^{-3}) over the same redshift range. At $z > 4$, the number densities calculated considering the *bona fide* SED modeling scenario are lower (by $2\sigma - 3\sigma$) than measurements from Alcalde Pampliega et al. (2019) ($n = 8.5 \pm 3.0 \times 10^{-6}$ Mpc^{-3} and $5.9 \pm 2.6 \times 10^{-6}$ Mpc^{-3}, at $4 < z < 5$ and $5 < z < 6$, respectively), with increasing discrepancy when nebular emission lines are considered to contaminate observed photometry. We note that the reported values from Alcalde Pampliega et al. (2019) do not include the effect of cosmic variance in the error budget, which we estimate to be as high as 50% and 110% for $\log(M_*/M_\odot) \approx 11.25$ at $3 < z < 6$.

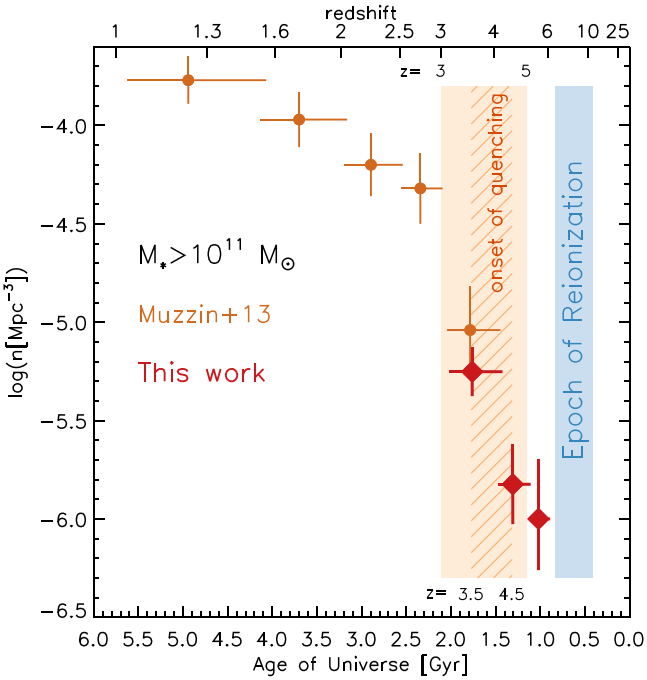

Figure 10. Number density evolution for $\log(M_*/M_\odot) > 11$ galaxies as a function of redshift (top axis) and cosmic time (bottom axis). Red diamonds show our calculation assuming the *empirical* prescription for emission line contamination to photometry. Orange filled circles show measurements from Muzzin et al. (2013a). The number density increases by a factor of $\gtrsim$5 and $\approx$3–4 in the $\approx$950 Myr between $z = 3$–5 (orange shaded region) and $\approx$460 Myr between $z = 3.5$–4.5 (hatched region), respectively. Gray shaded region represents the epoch of reionization ($z \approx 7$–11).

In summary, accounting for nebular emission contamination to photometry causes the calculated abundances to decrease by a factor of $\times \approx 2$–3 on average, and a factor of $\sim$5 in the most extreme case. This highlights that ignoring the contribution of nebular emission lines in SED modeling may significantly bias the inferred properties of the high-z galaxies (although the exact magnitude of this effect is relatively unconstrained with currently available observations).

We show Figure 10 to highlight the number density evolution for $\log(M_*/M_\odot) > 11$ galaxies as a function of cosmic time (bottom axis) and redshift (top axis). Red diamonds indicate our measurements for the *robust* sample of massive galaxies, while orange filled symbols show measurements from (Muzzin et al. 2013a) to the lower redshifts. The orange shaded region indicates the cosmic time between $z = 3$–5 ($\approx$950 Myr): the epoch of early formation and assembly for massive galaxies, where the number density increases by $\gtrsim$1.2 dex. The hatched area marks $z = 3.5$–4.5 ($\approx$460 Myr), the epoch at which quenching is beginning to take place (explored further in Section 5). Finally, the blue shaded region shows the epoch of reionization at $z \approx 7$–11 (Robertson et al. 2015; Planck Collaboration et al. 2020; Hoag et al. 2019; Mason et al. 2019).

5. Star Formation Activity

We use the counterparts identified in ancillary radio and FIR data sets (Sections A.2 and A.3) to better constrain the amount of dust-obscured star formation activity occurring in the *bona fide* sample of the massive galaxies identified in Section 2.2. Specifically, we incorporate the cross-matched Spitzer 24 μm, Herschel- PACS (100 μm, 160 μm)/SPIRE (250 μm, 300 μm,

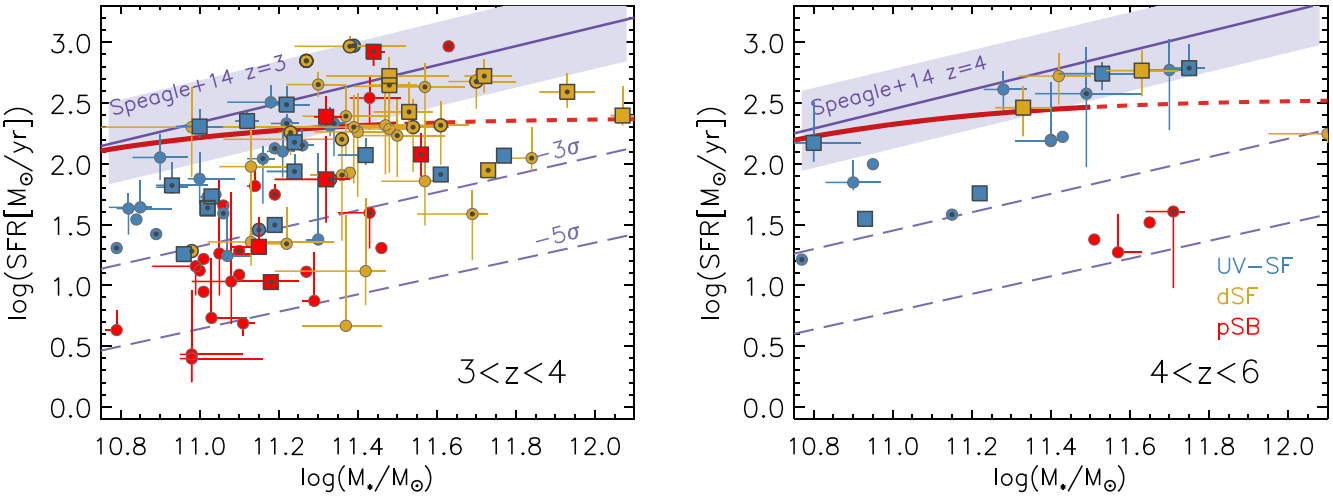

Figure 11. SFRs vs. stellar mass at $3 < z < 4$ (left panel) and $4 < z < 6$ (right panel) obtained using MAGPHYS. *Square* symbols indicate objects with FIR/radio counterparts (hence well-constrained IR SEDs), while *filled circles* are targets lacking strong FIR constraints. Color-coding is identical to previous figures. Galaxies with $>3\sigma$ MIPS 24 μm are marked with small *bullseye* points. Purple region shows the star formation MS fit at $z = 3$ and $z = 4$ from Speagle et al. (2014), with the 3σ and 5σ negative offsets indicated with dashed lines. Red curve is the fit to the star-forming main sequence from Tomczak et al. (2016) at $z = 3$ and $z = 4$ (dashed line is extrapolation to the high-mass end).

500 μm), SCUBA-2 (850 μm), ALMA (Band 7/870 μm, Band 6/1.2 mm), and VLA (3GHz) observations to model the observed UV-radio photometry using the high-z extension of MAGPHYS (da Cunha et al. 2008, 2015). This code incorporates energy balance to model the dust-radiated IR SEDs, ensuring stronger constraints on the SFRs compared to estimates from UV-IRAC photometry alone.

Similar to FAST, MAGPHYS employs the Chabrier (2003) IMF and the Bruzual & Charlot (2003) library to model stellar emission, but the two SED fitting techniques differ in several assumptions used to infer galaxy properties: the SFH is parametrized by a continuous delayed exponentially declining form (SFH $\propto (t\, e^{-t/\tau})/\tau^2$, where $\tau \equiv$ star formation timescale), superimposed with stochastic bursts of star formation; the two-component dust prescription of Charlot & Fall (2000) is adopted to attenuate stellar emission; the emission in the IR is modeled as described in da Cunha et al. (2008); metallicity is allowed to vary from $0.2Z_{\odot}$–$2Z_{\odot}$. The redshifts of targets are fixed ($z_{\rm peak}$ from EAZY or $z_{\rm spec}$, when available) when running MAGPHYS.

In Section 2.3, we identified FIR/submillimeter/radio counterparts for 38 (6) out of the 105 (23) sources in the mass-complete sample of K_S-selected galaxies at $3 < z < 4$ ($4 < z < 6$), resulting in the extended FIR SEDs to be studied for $\sim$1/3 of the overall sample. For the DR3 sources lacking FIR/radio counterparts, we incorporate the upper limits on the Spitzer and Herschel fluxes using the depths of the corresponding surveys (3σ upper limits of 4.5, 9.8, 17.4, 18.9, and 20.4 mJy for the 100, 160, 250, 350, and 500 μm bands, respectively). We choose to include the limits in these bands as they probe the redshifted peak of IR dust emission for $z > 3$ galaxies (see for example, Schreiber et al. 2018a; Simpson et al. 2019), and yield more realistic constraints for SFRs compared to using more restricted UV-IRAC SED models. We refer the reader to Appendix C.1 for more details regarding modeling with MAGPHYS and the comparison with FAST best-fit parameters. As the 24 μm data points may be sensitive to AGN contamination at the redshifts considered, we additionally ran MAGPHYS ignoring the 24 μm data points, finding the derived stellar populations to be similar within the statistical uncertainties.

Panels in Figure 11 display the median stellar mass versus SFRs for the *bona fide* sample of high-z massive galaxies obtained using MAGPHYS, where the SFR values correspond to the average calculated over the last 10 Myr. Error bars correspond to the 16th and 84th percentiles calculated from likelihood distributions. The colors of points mark the object classification scheme outlined in Section 3.1. Square symbols indicate objects in our sample with FIR/radio counterparts and well-constrained FIR SEDs, while circles represent those that lack strong constraints in the FIR. We compare our results with the star formation main sequence at $z = 3$ and $z = 4$ using the best-fit relations from Speagle et al. (2014; purple curve, with shaded region indicating the measured scatter) and Tomczak et al. (2016; red curve, dashed line indicates extrapolations to the high-mass end).

Given that the majority of the high-z massive galaxy sample is relatively unconstrained in their IR SEDs, and that SFR estimates may vary widely depending on the indicator(s) used, we refrain from analyzing the precise SFR value of individual objects, but rather investigate trends in the ensemble. The range in SFRs is consistent with the scatter measured in the other high-z populations, although we note that extreme-starbursts appear to be missing in our sample. Reassuringly, galaxies that were identified as pSB, based on their UV-IRAC SEDs, are generally distributed toward lower SFRs. We conclude that the K_S-selected massive galaxy sample at $3 < z < 6$ is largely consistent with being on the star-forming main sequence, with a tail extending to $\log({\rm sSFR/yr^{-1}}) < -10$—providing evidence that suppression of star formation has already begun by $z \sim 4$ for the most massive galaxies.

We additionally investigated the dependence of inferred SFR properties when considering different FIR SED templates. Specifically, we reestimated the star formation rates for the subsample of massive galaxies with FIR detections using the library of the model infrared templates introduced in Schreiber et al. (2018a). The SFRs estimated obtained using the Schreiber et al. (2018a) templates are, on average, larger by a factor of $\sim$2–3 than those estimated using MAGPHYS, albeit with large

scatter (see Appendix C.2). We note, however, that this offset arises mainly because the peak of the dust/IR SEDs is largely unconstrained (as it corresponds to Herschel wavelengths), and hence this translates to the rather limited discriminating power for the Schreiber et al. (2018a) templates to infer IR properties for the sample.

6. Sizes

We take advantage of the publicly available HST imaging surveys in COSMOS to investigate the structures of high-z UMGs. Specifically, we use the optical F814W (i_{814}, covering 1.7 deg^2; Koekemoer et al. 2007, 5σ point source limit $i_{814} = 27.2$) and NIR F160W (H_{160}; over 0.66 deg^{2}of the COSMOS field from the DASH Survey; Momcheva et al. 2017; Mowla et al. 2019; 5σ point source limit $H_{160} = 25.1$) mosaics available over the COSMOS field. The ACS i_{814} filter probes the rest-frame UV (≈1600–2000 Å) wavelengths of galaxies at $3 < z < 4$, enabling a high-resolution view of the star-forming regions unobscured by dust. The reddest filter on HST, WFC3 H_{160}, is better than bluer filters for tracing the distribution of stellar mass in high-z galaxies, although we caution that H_{160} is far from ideal for this, as it begins to probe rest-frame wavelengths blue-ward of the optical/Balmer break for galaxies at $z > 3$.

These mentioned HST imaging surveys are not deep enough to detect the majority of high-z massive galaxies with sufficient S/N to model their individual 2D light profiles. Therefore we opt to perform a stacking analysis to obtain constraints for the *average* size of massive galaxies at $3 < z < 6$. Two of the spec-z confirmed galaxies in our massive galaxy sample have HST/H_{160} band size measurements from the literature (Marsan et al. 2015; Straatman et al. 2015)—we removed these objects from the stacking analysis in the following section.

6.1. HST Stacking

High-resolution stacks were constructed for the UMG sample separated in two redshift bins ($3 < z < 4$ and $4 < z < 6$) for both HST imaging filters considered here (i_{814} and H_{160}) in a similar manner to that presented in van Dokkum et al. (2010) and Hill et al. (2017). As a first step, we created the i_{814} and H_{160} cutouts of size $20'' \times 20''$ centered using the coordinates for high-z massive galaxies taken from the UltraVISTA DR3 catalog (corresponding to K_S-band centroids). These individual cutouts were inspected to remove the diffraction spike features, along with those that were contaminated by or blended with nearby bright objects.

We constructed segmentation maps for the individual HST cutouts using SExtractor (Bertin & Arnouts 1996) in a low S/N detection threshold setting. These segmentation maps were used to identify which stamps to incorporate into our stacking (i.e., we excluded/discarded stamps that essentially had no signal associated with the location of main object in order to mitigate the effects of uncertain centroiding), as well as create bad pixel masks that flag neighboring objects. The resulting segmentation maps reveal a slight offset between the K_S-band and HST centroids of UMGs, and we choose to correct the individual HST cutouts prior to the final stacking procedure (otherwise, we will be artificially smearing the stacks). These are subarcsec offsets, typically $<0\farcs25$, and can be explained by the difference in the PSFs (VISTA K_S FWHM $\sim 0\farcs8$ versus HST FWHM $< 0\farcs2$). This could also be due to (inhomogeneous) dust distribution since i_{814} and H_{160} probes ever closer to rest-frame UV wavelengths at $z > 3$.

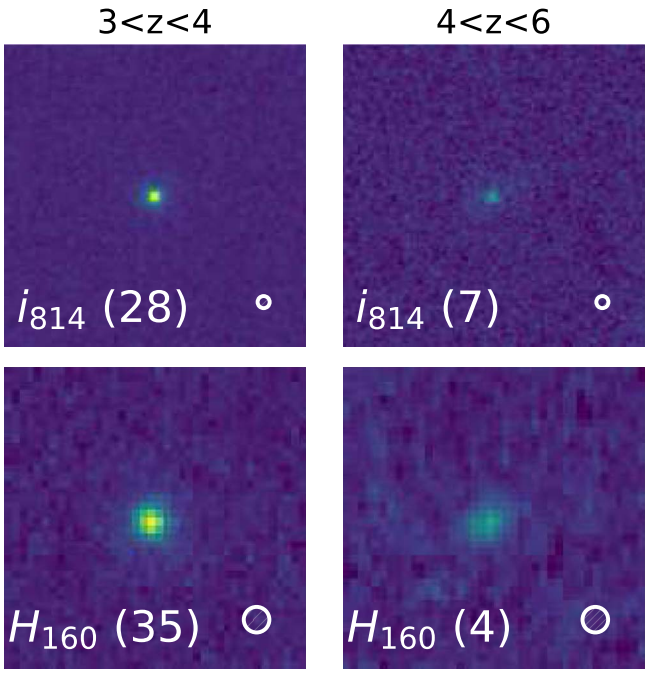

Figure 12. Stacked HST i_{814} (top row) and H_{160} (bottom row) images of massive galaxy sample at $3 < z < 4$ and $4 < z < 6$ (left and right column, respectively). The number of objects used in each stack are quoted in parenthesis. Stamps are $4'' \times 4''$ in size, with PSFs indicated in the bottom right corner (FWHM $\approx 0\farcs17$ and $\approx 0\farcs08$ for H_{160} and i_{814} bands).

After shifting and centering the individual HST stamps (and their corresponding bad pixel maps), the cutouts were normalized to the observed (aperture corrected) H band fluxes in the UltraVISTA DR3 catalog—this is to ensure that the final stacks are not biased by the brightest massive galaxies among the stacking sample. (van Dokkum et al. 2010 find similar results when normalizing using either fixed aperture or aperture corrected fluxes.) A *final* weight map was created by summing the bad pixel masks of associated galaxies; this is used to obtain the final weighted stacked images. The final weighted stacks were created by summing the normalized and masked cutouts of individual galaxies and dividing by the corresponding *final* weight map calculated in each redshift bin, for each HST imaging filter considered here. Figure 12 displays the $4'' \times 4''$ cutouts of the resulting HST i_{814} and H_{160} image stacks for the massive galaxy sample at $z = 3$–4 and $z = 4$–6. Each panel lists the number of objects used to obtain the displayed stack. As expected, the stacked HST images reveal the compact morphology of high-z UMGs, as well as the observed dimming with increasing look-back time.

6.2. Modeling 2D Light Profiles

We fit the 2D light profile distributions of the high-z massive galaxy HST stacks using a single-component Sérsic model. Empirical point-spread function (PSF) images, to be included in light profile modeling, were constructed using bright, unsaturated, and isolated stars located within the H_{160} and i_{814} HST mosaics. Given the compact nature of high-z galaxies, measurements may be sensitive to particular PSFs. For this, and to account for PSF variations across the field, we used PSFs from 10 individual stars in each mosaic (versus stacking to

Table 3
Size Estimates for HST Stacks of Massive Galaxies

	$3 < z < 4$			$4 < z < 6$		
Band	n	r_e [″]	r_e [kpc]	n	r_e [″]	r_e [kpc]
i_{814}	3.7 ± 0.2	0.089 ± 0.007	0.68 ± 0.06	…	…	…
H_{160}	2.5, 4	$0.26 \pm 0.01/0.458 \pm 0.01$	2.35 ± 0.09	1, 2.5, 4	0.31 ± 0.02	2.02 ± 0.15

Note. Listed values are weighted averages and 1σ uncertainty of the best-fit Sérsic indices (or fixed n, when used in GALFIT modeling) and effect radii.

obtain a single PSF). The measured PSF FWHMs are $r_e \approx 0\farcs17$ and $\approx 0\farcs08$ for H_{160} and i_{814} bands.

We used GALFIT (Peng et al. 2002) to determine the best-fit total magnitude, the half-light radius (r_e), the Sérsic index (n), the axis ratio (b/a), the position angle, and the centroid for a single object. The stacked HST images were modeled using the 10 empirical PSFs obtained from the mosaics, and the range of best-fit r_e values were used to constrain sizes. Table 3 lists the size estimates obtained using the stacked HST images of the massive galaxy sample. We convert the observed angular sizes to physical scales using the median redshift of galaxies in the created stacks (i_{814}: $\langle z \rangle = 3.15 \pm 0.24,\ \ 4.37 \pm 0.35$; H_{160}: $\langle z \rangle = 3.15 \pm 0.21,\ \ 4.88 \pm 0.35$). More details regarding the 2D light profile modeling with GALFIT can be found in Appendix D.

Figure 13 shows the observed stellar-mass versus size measurements for massive galaxies at $z \gtrsim 3$, with our H_{160} band size measurements at $3 < z < 4$ and $4 < z < 6$ shown by the green and yellow filled diamonds (error bars correspond to the range of sizes obtained using fixed $n = 2.5,\ 4.0$ fits). We compare with size measurements of high-z galaxies from the literature obtained with the subarcsec resolution imaging. Blue and pink dashed lines indicate the mass-size relation at $2.5 < z < 3$ for star-forming and quiescent galaxies, respectively (Mowla et al. 2019). Dash-triple dotted lines show the extrapolation for star-forming galaxies in ZFOURGE at $z \sim 3.5,\ 4.0,\ 5.0$ (Allen et al. 2017). Maroon and blue open circles are (UVJ-selected) quiescent and star-forming massive galaxies in ZFOURGE at $z \sim 4$ (Straatman et al. 2015), while open squares represent the size measurements for quiescent massive galaxies at $z \sim 4$ obtained using AO-assisted K_S-band imaging (Kubo et al. 2018). Our H_{160}-band size measurements at both redshift bins are consistent with the extrapolation of size-mass trends for star-forming galaxies at $z \sim 4$ and ~ 5 (Allen et al. 2017). Note that they are also just as compact as the massive quiescent galaxies observed at $z \sim 2.75$ (Mowla et al. 2019).

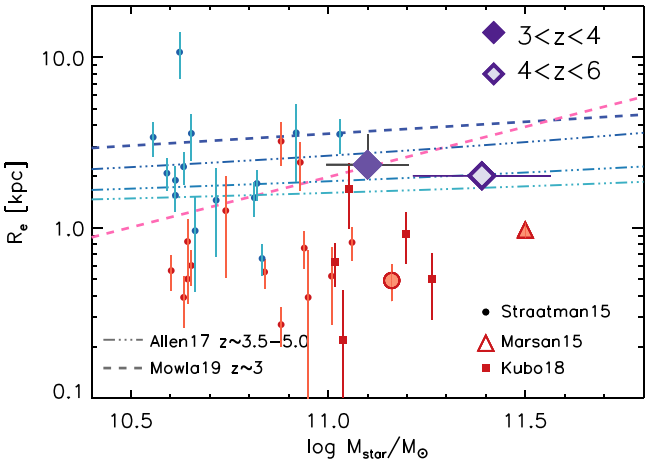

Figure 13. Stellar-mass and size measurements at $z \gtrsim 3$. Filled diamonds represent effective radii obtained from the H_{160} band stacks of the DR3 massive galaxy sample at $3 < z < 4$ and $4 < z < 6$. Filled circle and triangle symbols correspond to galaxies in our sample with available size measurements from the literature (Straatman et al. 2015; Marsan et al. 2015). Other measurements from the literature are shown (see text for details).

7. Discussion

7.1. Evolutionary Pathway to z ∼ 0

What are these massive galaxies at $z > 3$ expected to look like in the $z \sim 0$ universe? Answering this requires linking galaxy populations observed across various snapshots in cosmic time. Recent works investigating the progenitors of the most massive ($\log(M_*/M_\odot) > 11.5$) galaxies in the local universe using abundance matching techniques (Marchesini et al. 2014; Hill et al. 2017) infer stellar-mass growth by ~ 0.5 dex since $z \sim 3$ and 1.5 dex since $z \sim 5$. At $z = 0.35$, galaxies with $M_* \approx 5 \times 10^{11}\,M_\odot$ have a cumulative number density of $3 \times 10^{-6}\,\mathrm{Mpc}^{-3}$ (Marchesini et al. 2014), comparable with the cumulative number density estimate for $M_* \gtrsim 10^{11} M_\odot$ galaxies at $z \approx 4$ (after accounting for *empirical* emission line contamination to observed photometry). A comprehensive progenitor/descendant analysis is beyond the scope of this work, but one can argue that most (if not all) $M_* > 10^{11}\,M_\odot$ galaxies at $z > 4$ must evolve to become local *ultra-massive* galaxies by simply extrapolating from previous investigations. These objects are likely the dense cores that accumulate stellar mass through minor mergers to become the most massive early type/brightest cluster galaxies observed at $z \sim 0$.

7.2. Implications for Formation Epochs and Mechanisms

We find that the number density of massive galaxies increases by at least fivefold in the ~ 1 Gyr of cosmic time between $z \sim 3$ and $z \sim 4$–5. Stacking available HST imaging reveals that they are quite compact in size (<2.5 kpc), hinting that dissipative processes are responsible for the very early and rapid stellar-mass assembly in these systems. Contrary to the local universe, we find that the massive end of the galaxy population at $3 < z < 6$ exhibits a variety of stellar populations, with evidence that at least some candidates display signatures of suppressed levels of star formation (either through SED shapes or inferred SFR/sSFRs)—although the latter is less robust beyond $z > 4$. This suggests that the processes responsible for quenching galaxies are already at play by $z \sim 4$.

Although the analysis presented in this work is based on merely photometry, our results corroborate previous findings (e.g., stellar archaeological studies; Thomas et al. 2005; abundance patterns of massive quenched galaxies at $z \sim 2$; Kriek et al. 2016; Belli et al. 2017a; Morishita et al. 2019), converging upon $z > 5$ (and possibly as early as $z \sim 7$) as the

epoch at which the most massive galaxies in the universe begin to form rapidly (Estrada-Carpenter et al. 2020; Carnall et al. 2020; Forrest et al. 2021). We find a handful of galaxies on the way to becoming quenched at $3 < z < 4$, while the number of such viable candidates above $z > 4$ diminishes rapidly. This suggests that $z \approx 4$ is the epoch of the first quenching of galaxies, albeit with small number statistics. Indeed, follow-up spectroscopy of the brightest massive galaxy candidates identified in NIR extragalactic catalogs have confirmed the existence of a range of massive galaxies beyond $z > 3$ (Marsan et al. 2015, 2017; Schreiber et al. 2018c; Forrest et al. 2021), including those with no significant ongoing star formation, as evidenced by the absence of emission features in high S/N stellar continuum spectra out to $z \sim 4$ (Glazebrook et al. 2017; Tanaka et al. 2019; Forrest et al. 2020; Valentino et al. 2020).

7.3. Limitations and Outlook

It is essential to improve $z_{\rm phot}$ precision and accuracy (especially at $z > 4$) to better constrain the observed abundances and properties of the most massive galaxies at high-z. This will also allow for targets to be efficiently selected for detailed follow-up studies. Increasing the sampling of SEDs is crucial for identifying evolved stellar populations (in particular, the rest-frame optical break feature) and improving photometric redshift estimates. NIR medium-band surveys such as NMBS (Whitaker et al. 2011) and ZFOURGE (Straatman et al. 2016) have demonstrated that observations *splitting* the J and H broadband filters significantly enhances the ability to identify $z < 2$ dusty-forming galaxies that can masquerade as quiescent objects at high-z. The F2 Extragalactic Near-IR K-split (FENIKS; PI: Papovich) Survey uses split $K_{\rm S}$-band filters in a similar manner to enable analogous science at $z > 4$ (Esdaile et al. 2021).

Despite the encouraging results obtained with follow-up studies, our results show that the massive end of the galaxy population at $z > 3$ extends to $K_{\rm S} = 24$, revealing that the brightest objects ($K_{\rm S} < 22$) currently observed are merely the *tip of the iceberg*. This highlights that there is much to be learned regarding the formation of the most massive galaxies (see also Forrest et al. 2021). Extending the detailed spectral analysis (to obtain robust stellar populations and constrain SFHs) to beyond just the brightest candidates will have to wait until the launch of JWST. JWST will be crucial especially at $z > 4$, where most rest-frame optical features sensitive to SFH are redshifted to wavelengths not accessible with ground-based telescopes.

An obvious caveat to the $K_{\rm S}$-selection method employed (with current survey depths) is that it is effectively *blind* to the most dust enshrouded star-forming high-z galaxies. This population is beginning to be uncovered largely thanks to ALMA, such as *optically-dark* galaxies (Schreiber et al. 2018b; Alcalde Pampliega et al. 2019; Wang et al. 2019) and serendipitous *ALMA-only* sources (Williams et al. 2019; Romano et al. 2020; Loiacono et al. 2021). In Appendix E, we explored how the K-band selection employed in this work is dependent on the SFH, stellar age, and dust extinction for massive galaxies observed at $z > 3$. At $3 < z < 4$, the current sample is expected to be complete for all considered SFHs with stellar ages <0.9 Gyr with A_V up to $\sim$1.8–2.5 mag. At $z \sim 5$–6, however, the survey is expected to miss pSB galaxies older than $\sim$0.6 Gyr (corresponding to the onset of star formation at $z \gtrsim 9$–10) and $A_V > 1.3$ mag, while the star-forming galaxies with stellar ages <0.6 Gyr are detectable with dust attenuation up to $A_V \sim 2.3$ mag. Searches for massive star-forming galaxies have been made at $z > 7$ (e.g., Bowler et al. 2014, 2015, 2017; Stefanon et al. 2017, 2019, 2021); however, none with $M_* \gtrsim 10^{11} M_\odot$ have yet been located, suggesting that, although we are formally incomplete for quiescent pSB galaxies at $5 < z < 6$ (the likely descendants of such galaxies), such objects must be extremely rare, and therefore, our incompleteness for such extreme objects is unlikely to significantly affect our stellar-mass functions or stellar-mass density measurements.

Finally, extending detailed follow-up studies to observe galaxy components beyond stars is key to understanding the formation in these remarkably early systems. Obtaining resolved maps of their multiphase interstellar medium can help place meaningful limits on, for example, star formation efficiencies and baryonic budgets.

8. Summary

We use the UltraVISTA DR3 $K_{\rm S}$-selected photometric catalog (A. Muzzin et al. 2022, in preparation) to assemble a mass-complete ($M_* \gtrsim 10^{11} M_\odot$) sample of galaxies at $3 < z < 6$, corresponding to the most massive galaxies in the early universe. The NIR depths and exquisite sampling of SEDs provided by the 49 band UV-IRAC photometry allow us to construct a statistical sample of high-confidence candidates ($\approx$100 and 25 objects at $3 < z < 4$ and $4 < z < 6$, respectively).

The rapid number density evolution observed at $z \sim 4$ (increase by a factor of >5 in $\lesssim$500 Myr) points to the $z \sim 4$–6 universe as the epoch at which the high-mass end of the galaxy population is built up. The considerable fraction of the mass-complete sample showing evidence of suppressed star formation activity ($\sim$10%–25%, depending on the preferred photometry-based diagnostic) implies that we are approaching the epoch of first quenching at $z \sim 4$–5. The estimated stellar populations, as well as the observed trends in the rest-frame UVJ colors, provide additional support for this scenario:

1. At $3 < z < 4$, galaxies identified as pSB, UV-bright star-forming, and dSF based on their UV-IRAC SEDs contribute equally to the massive galaxy population ($\sim$30% for pSB and $\sim$35% for both UV-SF and dSF). At $4 < z < 6$, the fraction of UV-SF identified galaxies increases by a factor of 2 (to $\sim$65%), and the fraction of both pSB and dSF decreases by 2 (both to $\sim$17%). SEDs of the ensemble ($3 < z < 6$) reveal median mass-weighted stellar ages $\sim$500–900 Myr, but a range in the amount of dust-obscuration and inferred SFH timescales is consistent with the spread in rest-frame $U - V$ and $V - J$ colors.
2. We tested the potential bias of stellar-mass estimates by considering different SFHs, the inclusion of AGN templates, and various prescriptions for emission line contamination to the observed photometry in UV-IRAC SED modeling. We measure robust abundances for the most massive galaxies ($\log (M_*/M_\odot) > 11$) at $3 < z < 6$ and present the updated SMFs. Our findings are largely in agreement with previous measurements in literature; however, the improved statistics made available by deeper NIR measurements allow us to highlight the rapid emergence of the most massive galaxies at $z \approx 4$.

3. We incorporate ancillary data at FIR-radio wavelengths to model the extended UV-FIR SEDs of the mass-complete sample using MAGPHYS with the aim of obtaining reliable constraints for their IR luminosities and dust-obscured star formation rates. Despite their extreme stellar masses, the (UV-bright and dusty) actively star-forming population is consistent with being on the star formation main sequence at $3 < z < 6$. There is a large range in inferred SFRs and they are in general agreement with our classification based on UV-IRAC SEDs: ∼10%–15% (∼25%) of the selected sample at $3 < z < 4$ ($4 < z < 6$) have SFR $\gtrsim 300\ M_{\odot}/yr$; at both redshift bins, ∼15%–20% of the population show evidence of suppressed star formation activity ($<3\sigma$, i.e., ≈1 dex below MS, or $sSFR \lesssim 0.1$ Gyr^{-1}).
4. We perform the stacking analysis using available HST imaging across the COSMOS field (i_{814} and H_{160} bands) to obtain size estimates for the sample at $3 < z < 4$ and $4 < z < 6$. Consistent with previous works, we find the massive galaxy population at high-z is remarkably compact ($r_e(H_{160}) \lesssim 2.3$ kpc and $r_e(i_{814}) \lesssim 0.6$ kpc).

Z.C.M. gratefully acknowledges support from the Faculty of Science at York University as a York Science Fellow. This work was supported by the National Science Foundation through grants AST-1513473, AST-1517863, AST-1518257, and AST-1815475 by HST program number GO-15294, and by grant numbers NNX16AN49G and 80NSSC17K0019 issued through the NASA Astrophysics Data Analysis Program. Support for program number GO-15294 was provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Incorporated, under NASA contract NAS5-26555.

Based on data products from observations made with the ESO Telescopes at the La Silla Paranal Observatory under ESO program ID 179.A-2005 and on data products produced by CALET and the Cambridge Astronomy Survey Unit on behalf of the UltraVISTA consortium.

Appendix A
Multiwavelength Counterpart Identification

A.1. X-Ray Counterparts

We use the publicly available Chandra COSMOS-Legacy source catalog (Civano et al. 2016; limiting depth 8.9×10^{-16} erg cm^{-2} s^{-1} in the full 0.5-10 keV band) to search for X-ray counterparts (with DET_LM ⩾ 10.8) around the positions of the DR3 K_S-band source catalog using a search radius of $r = 2''$. We convert the observed full-band X-ray fluxes to rest-frame 2–10 keV luminosities ($L_{2-10\mathrm{keV}}$) for sources according to Equation (4) in Marchesi et al. (2016) assuming a fixed photon index of $\Gamma = 1.4$, and fixing their redshifts to best-fit z_{phot} solutions, or spectroscopic redshifts, when available.

A.2. Radio Counterparts

We utilize the publicly available source catalog from the VLA-COSMOS 3 GHz Large Project (Smolčić et al. 2017; reaching median rms ∼2.3 μJy/beam at an angular resolution of 0″.75) to search for radio counterparts of UltraVISTA DR3 sources. We identify radio-counterparts for 22/128 of objects in the *bona fide* high-z massive galaxy sample, using a search radius of $r = 0''.7$ and a S/N > 5.5 cut (to account for spurious sources as recommended in Smolčić et al. 2017). Assuming a radio spectral index, $\alpha = -0.7$ (i.e., $S_\nu \propto \nu^\alpha$) and the best-fit EAZY (or spec-z when available) redshifts, we convert the observed-frame 3 GHz flux densities to rest-frame 1.4 GHz luminosities.

A.3. Infrared Counterparts

1. **S***pitzer-MIPS 24 μm Counterparts.* The 24 μm photometry for objects in the DR3 catalog were measured on deblended images constructed using the positions of K_S band detections. Briefly, all sources in the K_S band were convolved with a kernel derived from the bright PSF stars in the K_S and Spitzer-MIPS 24 μm images (Le Floc'h et al. 2009). An individual *cleaned* 24 μm image was produced for each object after subtracting off the total flux of all nearby sources, and the 24 μm flux for each object was obtained by performing aperture photometry on these images (5″ diameter aperture). We refer the reader to Wuyts et al. (2008) for a more detailed discussion and demonstration of this method.
2. *Herschel-PACS/SPIRE Counterparts.* We supplement the UltraVISTA DR3 photometry using the publicly available Herschel data from the Herschel-PACS Evolutionary Probe (Lutz et al. 2011) and Herschel Multi-Tiered Extragalactic Survey (Oliver et al. 2012; Hurley et al. 2017, with SPIRE 250 μm, 350 μm, and 500 μm filters). The 3σ depths for PACS imaging at 100 μm and 160 μm are 4.5 and 9.8 mJy (FWHM = 7″–11″), while for the SPIRE 250 μm, 350 μm, and 500 μm maps they are 9.5, 8.1, and 11.4 mJy (FWHM = 18″−35″). For both surveys, we use the catalogs of fluxes extracted using Spitzer 24μm prior positions. We match sources with at least 3σ detection in any of the Herschel bands to MIPS-detected sources in the UltraVISTA DR3 catalog using a matching radius of 1″.5. We refer the reader to Martis et al. (2019) for a more detailed discussion of the matching and flux extraction technique.
3. *SCUBA-2 850 μm Counterparts.* We cross-match our sample of K_S-selected $3 < z < 6$ massive galaxy sample with publicly available 850 μm source catalog provided across the COSMOS field by S2COSMOS (Simpson et al. 2019). The source catalog consists of 850 μm sources detected at a significance of ⩾4σ across ∼2.6 square degrees in the COSMOS field (median noise $\sigma = 1.2$–1.7 mJy beam^{-1}, FWHM = 14″.9). We use a search radius of $r \lesssim 3''$ for counterpart identification. The $S_{850\mu m}$ flux densities are in the range ≈2.0–13.2 mJy with a median measured value ∼5.5 mJy.
4. *ALMA Counterparts.* Wide-field FIR/submillimeter surveys conducted using single-dish instruments generally lack the sensitivity to place meaningful constraints on the star formation in sources beyond the most IR-luminous high-z galaxies. Their large beam sizes (and hence, low angular resolution) also mean that their analysis is riddled with source-blending issues (e.g., Karim et al. 2013; Trakhtenbrot et al. 2017; Wardlow et al. 2018; Simpson et al. 2020). To capture the FIR SEDs for selected objects with higher angular resolution and deeper flux limits, we augment the photometry of massive high-z galaxies taking advantage of the A^{3}COSMOS (Liu et al. 2019) data set constructed using publicly available ALMA

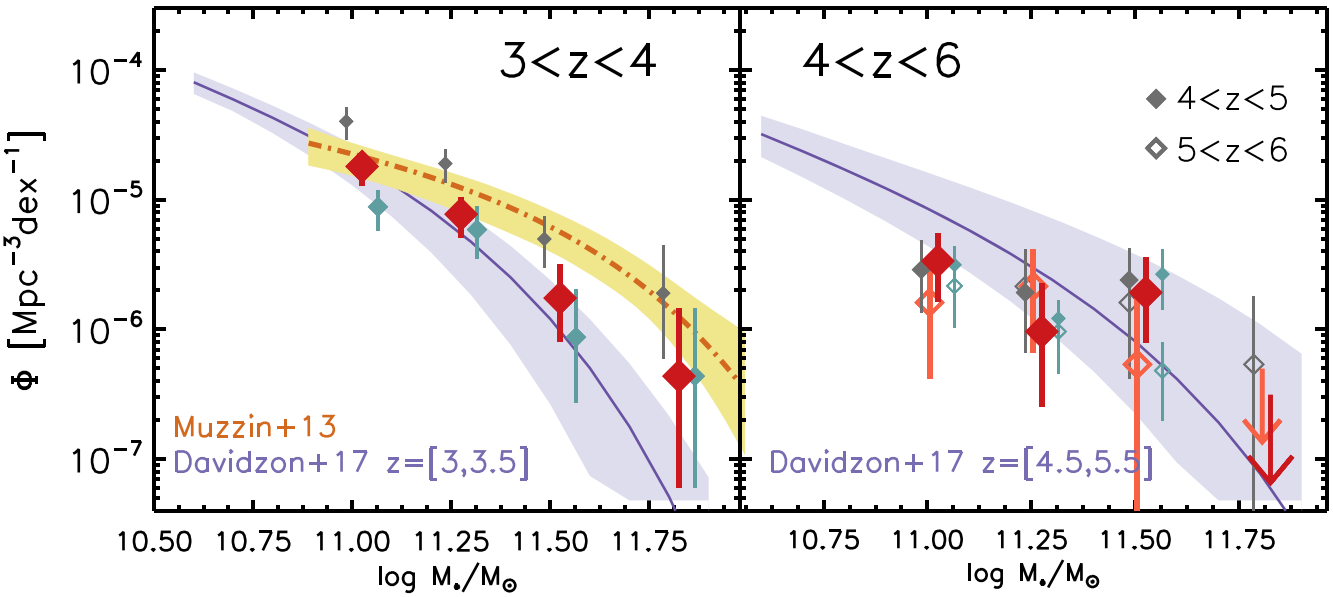

Figure B1. Stellar-mass functions at $3 < z < 4$ (left panel) and $4 < z < 6$ (right panel). Diamonds indicate calculations using the $1/V_{\rm max}$ method for the robust (large, red), *bona fide* (medium, gray), and *maximally* emission line corrected (small, light blue) sample of high-z massive galaxies. Shaded regions show, for comparison, the SMFs from Muzzin et al. (2013a) and Davidzon et al. (2017), same as plotted in Figure 9.

archival data in the COSMOS field. Specifically, we use the blind-extracted source catalog ($S/N_{\rm peak} > 5.4$ in Band 6 and/or 7 continuum data) using a search radius of $r = 1''$, finding matches for 17 of the high-z massive galaxies (2 at $z_{\rm phot} > 4$). We note that, although this data set overcomes source-blending issues, it is based on a compilation of various pointed surveys, leading to mixed coverage and depths across the field. There are 5 targets with Band 7 (870 μm) flux densities in the range $S_\nu \sim 1.5-9$ mJy (median $S_{870} \approx 4.7$ mJy), while 16 have Band 6 ($\sim$1.2 mm) measurements in the range $S_\nu \sim 0.3-3.6$ mJy (median $S_{1.2} \approx 1.8$ mJy).

Appendix B
Calculated SMFs with Different Emission Line Contamination Prescriptions

Panels in Figure B1 display the SMFs calculated at $3 < z < 4$ and $4 < z < 6$ using the three approaches to estimate nebular emission line contamination to the observed UV-NIR photometry (Section 4.1.3). Red (large) diamonds in Figure B1 indicate the densities computed for the robust high-z massive galaxy sample (same as in Figure 9). The same calculations, repeated using the *bona fide* sample and the sample corrected assuming the extreme emission line prescription, are plotted with gray and light blue diamonds, respectively. It is evident from Figure B1 that the calculated SMFs for massive galaxies at $3 < z < 4$ depend heavily on the assumed emission line prescriptions, while low-sample statistics still dominates the total error budget at $4 < z < 6$.

Appendix C
Comparison of Inferred Stellar Populations

C.1. Stellar Populations Estimated by FAST and MAGPHYS

Here we compare the best-fit stellar population parameters obtained for the *bona fide* massive galaxy sample using FAST and EAZY. As the massive galaxy population becomes increasingly more dominated by the dust-obscured star formation, we tested how the estimated stellar population properties varied when FIR constraints are considered in SED fitting. We specifically considered the key stellar population parameters investigated in this work: $\log(M_*)$, SFR, A_V and stellar age. Here we remind the reader that (more restricted) UV-NIR SEDs are used with FAST (Sections 2.1 and 4.1) versus the extended UV-FIR/radio SEDs with MAGPHYS (Section 5). Martis et al. (2019) report that M_*, A_V, and SFR estimates obtained using MAGPHYS are fairly robust to Herschel photometry.

Table C1 lists the average $\log(M_*)$, log(SFR), A_V, and log (age) MAGPHYS estimates for the sample, separated according to the classification in Section 3.1. Panels in Figure C1 show the comparison of M_*, A_V, and SFR and stellar age estimates for the high-z massive galaxy sample (defined in Section 3) using FAST and MAGPHYS. We use the median (50th) and the corresponding 16th and 84th percentiles for the parameter distribution from MAGPHYS. The distributions for

Table C1
Stellar Population Parameters for Massive Galaxies at $3 < z < 4$ and $4 < z < 6$

	pSB	UV-SF	dSF
$3 < z < 4$	$N = 29$	$N = 39$	$N = 37$
$\log(M_*/M_\odot)$	11.15 (0.20)	11.17 (0.24)	11.46 (0.27)
A_V	0.73 (0.58)	0.84 (0.59)	2.90 (0.75)
log(*age*)	8.85 (0.17)	8.96 (0.21)	8.91 (0.23)
log(*SFR*)[a]	1.27 (0.61)	2.01 (0.38)	2.23 (0.43)
$4 < z < 6$	$N = 4$	$N = 15$	$N = 4$
$\log(M_*/M_\odot)$	11.61 (0.08)	11.28 (0.33)	11.5 (0.19)
A_V	0.94 (0.27)	0.76 (0.70)	2.41 (0.30)
log(age)	8.76 (0.07)	8.80 (0.13)	8.76 (0.08)
log(SFR)[a]	1.45 (0.13)	2.19 (0.51)	2.52 (0.28)

Notes. Listed values are the medians and standard deviation of the subsample in each redshift bin computed by model UV-FIR SEDs using MAGPHYS.
[a] SFR in units of $M_\odot/yr$.

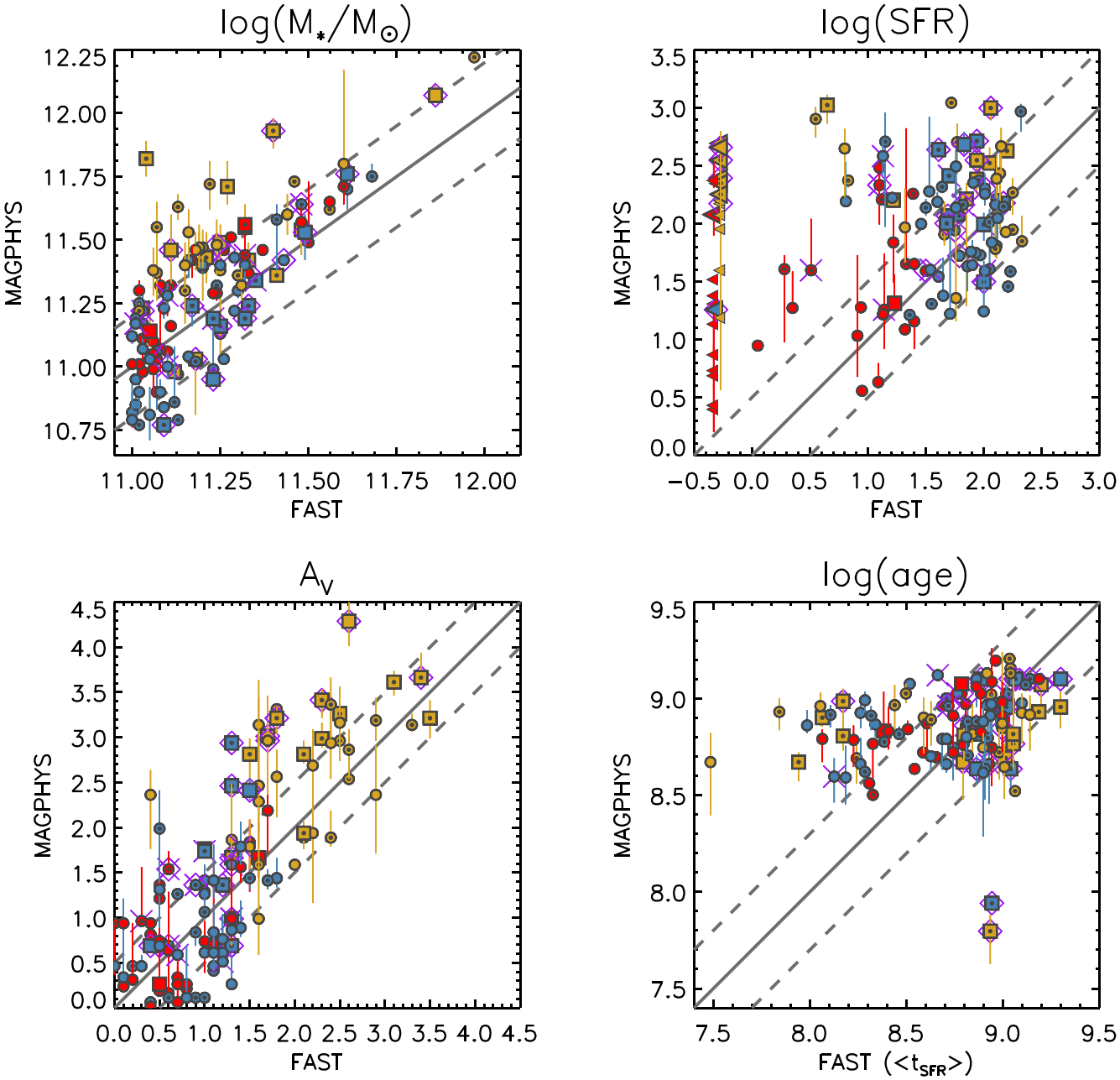

Figure C1. Comparison of best-fit stellar-population parameters obtained by modeling UV-NIR SEDs with FAST versus the modeling with MAGPHYS, which incorporates Spitzer 24 μm, Herschel (PACS and SPIRE), SCUBA-2 850 μm, and ALMA (Band 6 and 7) detections and upper limits. Symbols are color-coded according to the dissection of the massive galaxy sample in Section 3.1. Solid diagonal lines show 1:1 relation, and dashed lines indicate the following: $\pm$0.25 dex in log(M_*), $\pm$0.5 dex in log(SFR), $\pm$0.5 mag in A_V, and $\pm$0.3 dex in log(t_{age}).

best-fit parameters obtained with these two methods agree relatively well, albeit with some inconsistencies, which we outline below.

The estimated stellar masses for the pSB (red symbols) and UV-SF (blue symbols) population are consistent within $\sim\pm$0.2 dex, comparable with the variation reported in many previous works, e.g., Michałowski et al. (2014), Mobasher et al. (2015), and Carnall et al. (2019). The MAGPHYS stellar-mass estimates of the dusty-SF galaxies, however, are offset by $\sim$+0.25 dex compared to those calculated by FAST (see also Liu et al. 2019). The largest discrepancy between the investigated stellar population parameters is seen in the *SFR* estimates (top right panel in Figure C1). This is expected, as the implementation of FIR constraints to account for dust reemission, fundamentally alters how the two methods calculate SFRs. FAST underestimates the total SFR as by construction it does not know anything about fluxes beyond $\lambda_{obs} > 10\,\mu$m—where dust reemitted radiation is significant.

C.2. Star Formation Rates Using Schreiber et al. (2018a) IR Templates

We used the library of templates presented in Schreiber et al. (2018a) to model the SEDs of the FIR-detected galaxies in our sample. The templates in this library are characterized by three parameters: the dust mass (M_{dust}), the dust temperature (T_{dust}), and the mid-to-total infrared color ($IR8 = LIR/L8$). The templates show trends of increasing T_{dust} and IR8 with redshift

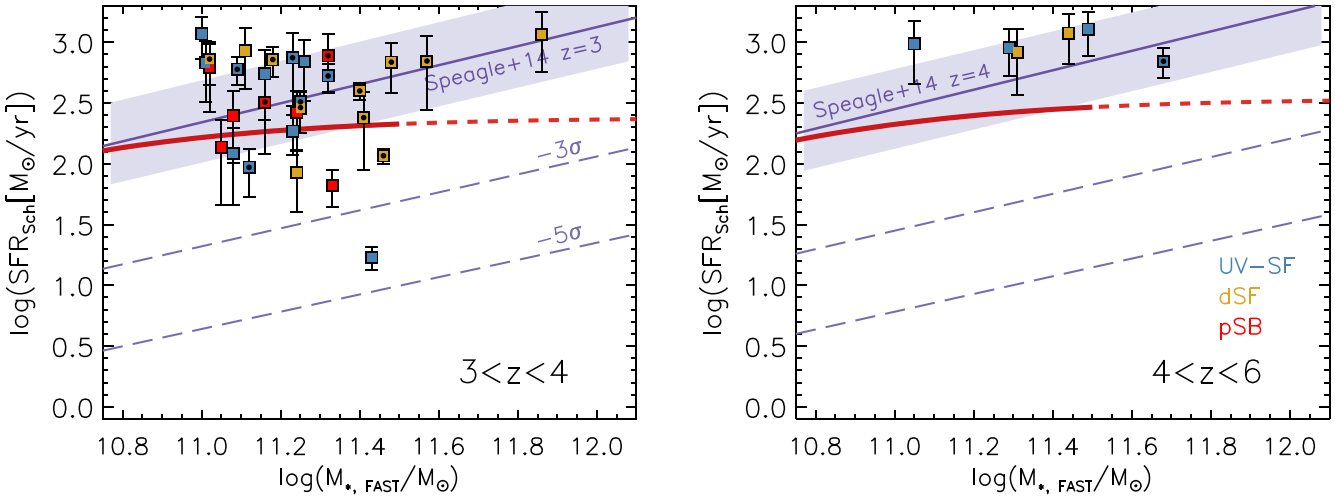

Figure C2. SFRs vs. stellar mass for the subsample of massive galaxies with far-infrared (FIR) detections (square symbols as in Figure 11; those with $>3\sigma$ MIPS 24 μm detections are marked with *bullseye* points). These SFR estimates are obtained using Schreiber et al. (2018a) FIR templates, while stellar masses correspond to those derived using FAST.

and distance from the SFR $- M_*$ main sequence. Essentially, one can obtain an IR template from the Schreiber et al. (2018a) library by specifying a redshift and offset from the main sequence. The offset from the main sequence is characterized by the *rate of starburstines* parameter, *rsb*, i.e., the ratio between the SFR of the object and the SFR it would have on the main sequence (a galaxy on the main sequence would have $rsb = 1$). For each FIR-detected galaxy in our sample, we extracted Schreiber et al. (2018a) templates using the (photometric) redshift of the object assuming a range in values for $rsb = 0.1,\ 0.5,\ 1.0,\ 5.0$. At $3 < z < 4$, these templates correspond to $T_{\rm dust}$ values ranging from ∼30K to ∼45K; while at $4 < z < 6$, they range from ∼35–40 K to ∼50 K.

We estimated SFRs using the Schreiber et al. (2018a) templates as follows. For each FIR detected object, we extracted the IR SED templates corresponding to $rsb = 0.1$, 0.5, 1.0, 5.0, at the desired redshift of the galaxy. For each template (4 for each galaxy considered), we calculated the $L_{IR}(8 - 1000\mu m)$ that minimizes the χ^2 to FIR detections (excluding 24 μm photometry as in Section 5, as it may be contaminated by AGN) and are consistent with upper limits to non-detections (essentially, constraining the normalization of the templates). We then estimated SFRs using the Kennicutt (1998) relation, corrected for a Chabrier (2003) IMF.

In Figure 11 of the main text, we plotted the star formation main sequence derived from MAGPHYS stellar masses and MAGPHYS IR SFRs. In Figure C2, we reproduce this plot, this time using the Schreiber et al. (2018a) templates to determine the SFRs and the FAST SED fits to determine the stellar masses. As discussed above and shown in Figure C1, FAST stellar masses are largely consisted with those estimated by MAGPHYS, except for the dusty-star-forming population where an offset of ∼0.25 dex is observed. On average, the Schreiber et al. (2018a) templates imply SFRs that are on average a factor of ∼2–3 greater than MAGPHYS estimates, although there is large scatter ($\Delta \log(\mathrm{SFR}) = 0.43 \pm 0.53$ dex). Overall this leads to a higher normalization of the star-forming sequence for our sample, although we note that inferred SFRs are largely consistent within the scatter of the main sequence at the considered redshifts.

We note that the discrepancy between SFRs from the two methods is driven exclusively by galaxies that have only upper limits in PACS/SPIRE. When galaxies are detected in PACS/SPIRE, the two methods provide SEDs and SFRs that are in good agreement. For galaxies with only upper limits, we note that the inferred Schreiber et al. (2018a) templates are much brighter in the submillimeter compared to the inferred MAGPHYS templates; however, they still lie below the upper limits. In many cases the inferred IR SED from the Schreiber et al. (2018a) templates is only slightly below the PACS/SPIRE upper limits; however, formally there is no disagreement. Taken at face value, the two methods imply significantly different IR SEDs and dust temperatures for PACS/SPIRE undetected galaxies. Additional submillimeter observations would be extremely valuable to test if there truly is a wide range of dust temperatures for massive high-z galaxies as implied by MAGPHYS, or if the population is more homogenous as implied by the Schreiber et al. (2018a) templates.

Appendix D
2D Light Profile Modeling of HST Stacks

For the i_{814} $3 < z < 4$ stacks, the best-fit Sérsic indices and half-light radii ranged between $n \sim 3 - 6$ and $r_e \sim 2.3$–3.4 pixels (corresponding to $\sim 0\farcs07$–$0\farcs1$, indicating that it is just barely resolved). We were not able to obtain reliable size measurements for the high-redshift ($4 < z < 6$) i_{814} stacks given the low S/N. Regardless, Figure 12 shows that they appear to be as compact as their $3 < z < 4$ counterparts. Interpretation of size estimates based on i_{814} band images are complicated due to dust-obscuration considerations.

Modeling the light profile of the H_{160} image stacks did not converge when all the Sérsic profile parameters were allowed to vary. Therefore, the size estimates were obtained using fixed $n = 1.0,\ 2.5,\ 4.0$. For the $3 < z < 4$ H_{160} stack modeling, the r_e best-fit estimates were correlated with the specific n Sérsic profiles used: $\langle r_e \rangle = 0\farcs175 \pm 0\farcs003$, $0\farcs260 \pm 0\farcs004$, $0\farcs458 \pm 0\farcs008$, for $n = 1,\ 2.5,\ 4.0$, respectively (weighted averages of fits using the 10 different PSFs). Visually inspecting the residuals from the $n = 1$ model's fits showed them to be not good fits (over-subtracting a compact center);

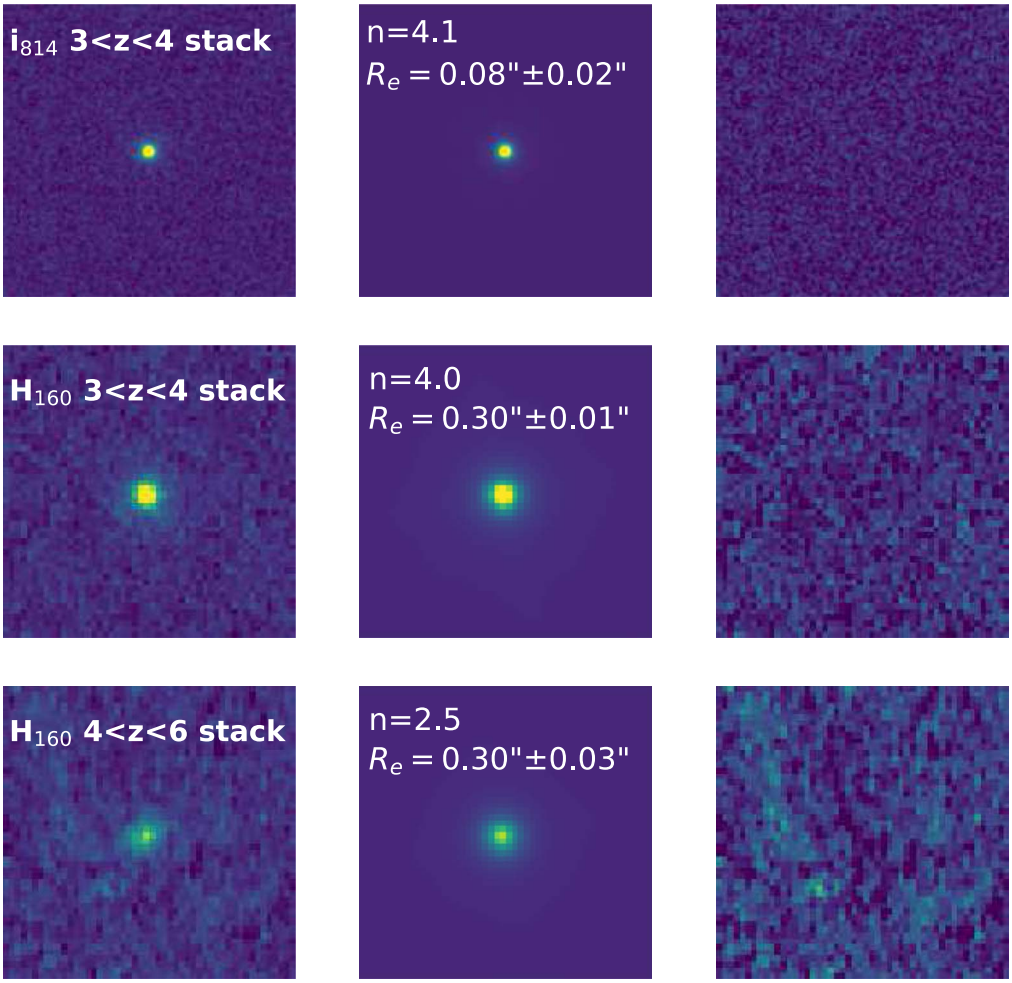

Figure D1. GALFIT modeling results. Stamps are $5'' \times 5''$. Columns show the following: (left) input images, (middle) single-component Sérsic model, and (right) residual of fit model.

we, therefore, do not consider them further. The $4 < z < 6$ H_{160} r_e estimates obtained for all the fixed n profiles were consistent with each other (within 1σ uncertainties). We calculate the weighted averages sizes for the $3 < z < 4$ and $4 < z < 6$ stacks of H_{160} imaging to be $\langle r_e \rangle = 0\farcs302 \pm 0\farcs012$ and $0\farcs310 \pm 0\farcs023$.

Figure D1 displays examples of single-component Sérsic profile models for the HST stacks analyzed in Section 6.2. Columns are as follows: the HST stack image cutouts, the best-fit 2D light profile obtained with GALFIT (n and r_e indicated in the row), and the residual image.

Appendix E
Completeness of the Sample

Given that the rest-frame optical break of galaxies to be will be redshifting through the K-band at $z \gtrsim 4$, we investigated the sensitivity of the adopted survey limit and attempted to quantify which types of galaxies (in terms of stellar age, star formation activity, and dust content) would be absent from the current sample.

We used GALAXEV to compute the spectral evolution for composite stellar populations (motivated by the SED fits for our sample; Section 3.2) using Bruzual & Charlot (2003) stellar population synthesis models, assuming a Chabrier (2003) IMF and fixed solar metallicity. We produced stellar population evolution models for a range of parametric SFHs (exponentially declining: with $\tau = 50$ Myr, 100 Myr, and 300 Myr, and delayed exponentially declining $\tau = 50$ Myr, 150 Myr, and 250 Myr models), and with varying amounts of dust attenuation. The two-component dust prescription of Charlot & Fall (2000) is employed in GALAXEV to model dust attenuation according to two input parameters: the total effective V-band optical depth in birth clouds affecting young stars (τ_V), and the fraction of it (μ) that is contributed by the diffuse ISM (affecting light from older stars). We simulated dust-extinction models with μ fixed at 0.4, and with varying $\tau_V = 1, 4, 5.5, 7.5$, yielding V-band dust attenuation values (A_V) ranging from $\approx$0.3 mag to $\approx$3 mag.

We chose a range of redshifts for the onset of star formation ranging from $z_{\rm form} = 3.5$ to $z_{\rm form} = 10$, and computed the observed K-band magnitude for the full suite of SEDs (all combinations of SFHs, A_V, and z_{form}) such that they match a galaxy with stellar mass $\log(M_*/M_\odot) = 11$ at the observed redshifts of $z_{\rm obs} = 3, 3.5, 4, 4.5, 5, 6$. We used this analysis to inform the *detectability* for each SED type at the considered epochs (i.e., if $K \leqslant 24.5$ for a specific SED at some $z_{\rm obs}$, then such an object should be detected). Our analysis recovers the expected trend: at fixed observed epochs ($z_{\rm obs}$) and dust attenuation, galaxies with younger stellar ages are always brighter, and hence more likely to be selected.

At $3 < z < 4$, the current Ks sample is expected to be complete for all considered SFHs with stellar ages <0.9 Gyr with A_V up to ≈1.8–2.5 mag. The adopted $K = 24.5$ AB limit implies that quenched/pSB galaxies that began forming in a short burst ($\tau < 100$ Myr) earlier than at $z_{\rm form} > 10$ (corresponding to <0.5 Gyr after the Big Bang), with dust attenuation more than $A_V > 1.5$ mag, would be missing from the current sample. When considering more extended SFHs, the survey is expected to capture star-forming galaxies with $A_V \lesssim 2.5$ mag that began forming at $z_{\rm form} < 10$.

These simulations reproduce the rapid drop in observed K-band magnitudes that is expected for galaxies at $4 < z < 5$ (as the rest-optical break is redshifted through the filter), and show that the detectability of SEDs is more sensitive to the assumed SFH, redshift for the onset of star formation (hence, stellar ages), and dust attenuation magnitude. At $z \approx 5$–6, the survey is expected to miss the oldest pSB galaxies (∼0.6 Gyr; corresponding to $z_{\rm form} > 9$–10) with $A_V > 1.3$ mag, while star-forming galaxies with <0.6 Gyr are expected to be captured with dust attenuation up to $A_V \sim 2.3$ mag.

ORCID iDs

Z. Cemile Marsan https://orcid.org/0000-0002-7248-1566
Adam Muzzin https://orcid.org/0000-0002-9330-9108
Danilo Marchesini https://orcid.org/0000-0001-9002-3502
Mauro Stefanon https://orcid.org/0000-0001-7768-5309
Nicholas Martis https://orcid.org/0000-0003-3243-9969
Marianna Annunziatella https://orcid.org/0000-0002-8053-8040
Jeffrey C. C. Chan https://orcid.org/0000-0001-6251-3125
Michael C. Cooper https://orcid.org/0000-0003-1371-6019
Ben Forrest https://orcid.org/0000-0001-6003-0541
Percy Gomez https://orcid.org/0000-0003-0408-9850
Ian McConachie https://orcid.org/0000-0002-2446-8770
Gillian Wilson https://orcid.org/0000-0002-6572-7089

References

Aihara, H., Armstrong, R., Bickerton, S., et al. 2018, PASJ, 70, S8
Aird, J., Coil, A. L., & Georgakakis, A. 2018, MNRAS, 474, 1225
Alcalde Pampliega, B., Pérez-González, P. G., Barro, G., et al. 2019, ApJ, 876, 135
Allen, R. J., Kacprzak, G. G., Glazebrook, K., et al. 2017, ApJL, 834, L11
Anders, P., & Fritze-v. Alvensleben, U. 2003, A&A, 401, 1063
Ashby, M. L. N., Caputi, K. I., Cowley, W., et al. 2018, ApJS, 237, 39
Avni, Y., & Bahcall, J. N. 1980, ApJ, 235, 694
Belli, S., Newman, A. B., & Ellis, R. S. 2017a, ApJ, 834, 18
Belli, S., Newman, A. B., & Ellis, R. S. 2019, ApJ, 874, 17
Belli, S., Genzel, R., Förster Schreiber, N. M., et al. 2017b, ApJL, 841, L6
Bellstedt, S., Robotham, A. S. G., Driver, S. P., et al. 2020, MNRAS, 498, 5581
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
Bowler, R. A. A., Dunlop, J. S., McLure, R. J., & McLeod, D. J. 2017, MNRAS, 466, 3612
Bowler, R. A. A., Dunlop, J. S., McLure, R. J., et al. 2014, MNRAS, 440, 2810
Bowler, R. A. A., Dunlop, J. S., McLure, R. J., et al. 2015, MNRAS, 452, 1817
Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, ApJ, 686, 1503
Brammer, G. B., Whitaker, K. E., van Dokkum, P. G., et al. 2009, ApJL, 706, L173
Brammer, G. B., Whitaker, K. E., van Dokkum, P. G., et al. 2011, ApJ, 739, 24
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682
Caputi, K. I., Dunlop, J. S., McLure, R. J., et al. 2012, ApJL, 750, L20
Caputi, K. I., Ilbert, O., Laigle, C., et al. 2015, ApJ, 810, 73
Caputi, K. I., Deshmukh, S., Ashby, M. L. N., et al. 2017, ApJ, 849, 45
Carnall, A. C., Leja, J., Johnson, B. D., et al. 2019, ApJ, 873, 44
Carnall, A. C., Walker, S., McLure, R. J., et al. 2020, MNRAS
Chabrier, G. 2003, ApJL, 586, L133
Charlot, S., & Fall, S. M. 2000, ApJ, 539, 718
Choi, J., Conroy, C., Moustakas, J., et al. 2014, ApJ, 792, 95
Cimatti, A., Daddi, E., Renzini, A., et al. 2004, Natur, 430, 184
Cimatti, A., Cassata, P., Pozzetti, L., et al. 2008, A&A, 482, 21
Civano, F., Marchesi, S., Comastri, A., et al. 2016, ApJ, 819, 62
Cowley, M. J., Spitler, L. R., Tran, K.-V. H., et al. 2016, MNRAS, 457, 629
da Cunha, E., Charlot, S., & Elbaz, D. 2008, MNRAS, 388, 1595
da Cunha, E., Walter, F., Smail, I. R., et al. 2015, ApJ, 806, 110
Daddi, E., Cimatti, A., Renzini, A., et al. 2004, ApJL, 600, L127
Daddi, E., Renzini, A., Pirzkal, N., et al. 2005, ApJ, 626, 680
Davidzon, I., Ilbert, O., Laigle, C., et al. 2017, A&A, 605, A70
Delvecchio, I., Smolčić, V., Zamorani, G., et al. 2017, A&A, 602, A3
Deshmukh, S., Caputi, K. I., Ashby, M. L. N., et al. 2018, ApJ, 864, 166
D'Eugenio, C., Daddi, E., Gobat, R., et al. 2020, ApJL, 892, L2
Duncan, K., Conselice, C. J., Mortlock, A., et al. 2014, MNRAS, 444, 2960
Esdaile, J., Labbé, I., Glazebrook, K., et al. 2021, AJ, 162, 225
Estrada-Carpenter, V., Papovich, C., Momcheva, I., et al. 2020, ApJ, 898, 171
Faisst, A. L., Capak, P., Hsieh, B. C., et al. 2016, ApJ, 821, 122
Fioc, M., & Rocca-Volmerange, B. 1999, arXiv:astro-ph/9912179
Florez, J., Jogee, S., Sherman, S., et al. 2020, MNRAS, 497, 3273
Fontana, A., Santini, P., Grazian, A., et al. 2009, A&A, 501, 15
Forrest, B., Tran, K.-V. H., Broussard, A., et al. 2018, ApJ, 863, 131
Forrest, B., Annunziatella, M., Wilson, G., et al. 2020, ApJL, 890, L1
Forrest, F., Marsan, Z. C., Annunziatella, A., & Wilson, G. 2021, ApJ, 903, 47
Förster Schreiber, N. M., van Dokkum, P. G., Franx, M., et al. 2004, ApJ, 616, 40
Franx, M., van Dokkum, P. G., Förster Schreiber, N. M., et al. 2008, ApJ, 688, 770
Franx, M., Labbé, I., Rudnick, G., et al. 2003, ApJL, 587, L79
Gallazzi, A., Charlot, S., Brinchmann, J., & White, S. D. M. 2006, MNRAS, 370, 1106
Gallazzi, A., Charlot, S., Brinchmann, J., White, S. D. M., & Tremonti, C. A. 2005, MNRAS, 362, 41
Gehrels, N. 1986, ApJ, 303, 336
Girelli, G., Bolzonella, M., & Cimatti, A. 2019, A&A, 632, A80
Glazebrook, K., Schreiber, C., Labbé, I., et al. 2017, Natur, 544, 71
Gómez-Guijarro, C., Toft, S., Karim, A., et al. 2018, ApJ, 856, 121
Grazian, A., Fontana, A., Santini, P., et al. 2015, A&A, 575, A96
Guarnieri, P., Maraston, C., Thomas, D., et al. 2019, MNRAS, 483, 3060
Hasinger, G., Capak, P., Salvato, M., et al. 2018, ApJ, 858, 77
Hill, A. R., Muzzin, A., Franx, M., et al. 2017, ApJ, 837, 147
Hoag, A., Bradač, M., Huang, K., et al. 2019, ApJ, 878, 12
Hurley, P. D., Oliver, S., Betancourt, M., et al. 2017, MNRAS, 464, 885
Ilbert, O., McCracken, H. J., Le Fèvre, O., et al. 2013, A&A, 556, A55
Kado-Fong, E., Marchesini, D., Marsan, Z. C., et al. 2017, ApJ, 838, 57
Karim, A., Swinbank, A. M., Hodge, J. A., et al. 2013, MNRAS, 432, 2
Kawinwanichakij, L., Papovich, C., Ciardullo, R., et al. 2020, ApJ, 892, 7
Kennicutt, R. C., Jr. 1998, ARA&A, 36, 189
Khostovan, A. A., Sobral, D., Mobasher, B., et al. 2015, MNRAS, 452, 3948
Khostovan, A. A., Sobral, D., Mobasher, B., et al. 2016, MNRAS, 463, 2363
Koekemoer, A. M., Aussel, H., Calzetti, D., et al. 2007, ApJS, 172, 196
Kriek, M., van Dokkum, P. G., Labbé, I., et al. 2009, ApJ, 700, 221
Kriek, M., van Dokkum, P. G., Franx, M., et al. 2006, ApJL, 649, L71
Kriek, M., Conroy, C., van Dokkum, P. G., et al. 2016, Natur, 540, 248
Kubo, M., Tanaka, M., Yabe, K., et al. 2018, ApJ, 867, 1
Labbé, I., Oesch, P. A., Bouwens, R. J., et al. 2013, ApJL, 777, L19
Le Floc'h, E., Aussel, H., Ilbert, O., et al. 2009, ApJ, 703, 222
Liu, D., Lang, P., Magnelli, B., et al. 2019, ApJS, 244, 40
Loiacono, F., Decarli, R., Gruppioni, C., et al. 2021, A&A, 646, A76
Lusso, E., Comastri, A., Simmons, B. D., et al. 2012, MNRAS, 425, 623
Lutz, D., Poglitsch, A., Altieri, B., et al. 2011, A&A, 532, A90
Maraston, C. 2005, MNRAS, 362, 799
Marchesi, S., Civano, F., Elvis, M., et al. 2016, ApJ, 817, 34
Marchesini, D., van Dokkum, P. G., Förster Schreiber, N. M., et al. 2009, ApJ, 701, 1765
Marchesini, D., Whitaker, K. E., Brammer, G., et al. 2010, ApJ, 725, 1277
Marchesini, D., Muzzin, A., Stefanon, M., et al. 2014, ApJ, 794, 65
Marconi, A., Risaliti, G., Gilli, R., et al. 2004, MNRAS, 351, 169
Mármol-Queraltó, E., McLure, R. J., Cullen, F., et al. 2016, MNRAS, 460, 3587
Marsan, Z. C., Marchesini, D., Brammer, G. B., et al. 2017, ApJ, 842, 21
Marsan, Z. C., Marchesini, D., Brammer, G. B., et al. 2015, ApJ, 801, 133
Marsan, Z. C., Marchesini, D., Muzzin, A., et al. 2019, ApJ, 871, 201
Martis, N. S., Marchesini, D. M., Muzzin, A., et al. 2019, ApJ, 882, 65
Martis, N. S., Marchesini, D., Brammer, G. B., et al. 2016, ApJL, 827, L25
Mason, C. A., Naidu, R. P., Tacchella, S., & Leja, J. 2019, MNRAS, 489, 2669

Mawatari, K., Yamada, T., Fazio, G. G., Huang, J.-S., & Ashby, M. L. N. 2016, PASJ, 68, 46
Mawatari, K., Inoue, A. K., Hashimoto, T., et al. 2020, ApJ, 889, 137
McCracken, H. J., Capak, P., Salvato, M., et al. 2010, ApJ, 708, 202
McCracken, H. J., Milvang-Jensen, B., Dunlop, J., et al. 2012, A&A, 544, A156
McDermid, R. M., Alatalo, K., Blitz, L., et al. 2015, MNRAS, 448, 3484
Mehta, V., Scarlata, C., Capak, P., et al. 2018, ApJS, 235, 36
Merlin, E., Fontana, A., Castellano, M., et al. 2018, MNRAS, 473, 2098
Merlin, E., Fortuni, F., Torelli, M., et al. 2019, MNRAS, 490, 3309
Michałowski, M. J., Hayward, C. C., Dunlop, J. S., et al. 2014, A&A, 571, A75
Mobasher, B., Dickinson, M., Ferguson, H. C., et al. 2005, ApJ, 635, 832
Mobasher, B., Dahlen, T., Ferguson, H. C., et al. 2015, ApJ, 808, 101
Momcheva, I. G., van Dokkum, P. G., van der Wel, A., et al. 2017, PASP, 129, 015004
Morishita, T., Abramson, L. E., Treu, T., et al. 2018, ApJL, 856, L4
Morishita, T., Abramson, L. E., Treu, T., et al. 2019, ApJ, 877, 141
Moster, B. P., Somerville, R. S., Newman, J. A., & Rix, H.-W. 2011, ApJ, 731, 113
Mowla, L. A., van Dokkum, P., Brammer, G. B., et al. 2019, ApJ, 880, 57
Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013a, ApJS, 206, 8
Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013b, ApJ, 777, 18
Nayyeri, H., Mobasher, B., Hemmati, S., et al. 2014, ApJ, 794, 68
Newman, A. B., Belli, S., Ellis, R. S., & Patel, S. G. 2018, ApJ, 862, 126
Newman, A. B., Ellis, R. S., Bundy, K., & Treu, T. 2012, ApJ, 746, 162
Oliver, S. J., Bock, J., Altieri, B., et al. 2012, MNRAS, 424, 1614
Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, AJ, 124, 266
Pérez-González, P. G., Trujillo, I., Barro, G., et al. 2008, ApJ, 687, 50
Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, A&A, 641, A6
Polletta, M., Tajer, M., Maraschi, L., et al. 2007, ApJ, 663, 81
Rasappu, N., Smit, R., Labbé, I., et al. 2016, MNRAS, 461, 3886
Reddy, N. A., Shapley, A. E., Sanders, R. L., et al. 2018, ApJ, 869, 92
Renzini, A. 2006, ARA&A, 44, 141
Rigby, J. R., Diamond-Stanic, A. M., & Aniano, G. 2009, ApJ, 700, 1878
Robertson, B. E., Ellis, R. S., Furlanetto, S. R., & Dunlop, J. S. 2015, ApJL, 802, L19
Roebuck, E., Sajina, A., Hayward, C. C., et al. 2019, ApJ, 881, 18
Romano, M., Cassata, P., Morselli, L., et al. 2020, MNRAS, 496, 875
Sanders, D. B., Salvato, M., Aussel, H., et al. 2007, ApJS, 172, 86
Santini, P., Merlin, E., Fontana, A., et al. 2019, MNRAS, 486, 560
Saracco, P., Marchesini, D., La Berbera, F., et al. 2020, ApJ, 905, 40
Schreiber, C., Elbaz, D., Pannella, M., et al. 2018a, A&A, 609, A30
Schreiber, C., Labbé, I., Glazebrook, K., et al. 2018b, A&A, 611, A22
Schreiber, C., Glazebrook, K., Nanayakkara, T., et al. 2018c, A&A, 618, A85
Scoville, N., Aussel, H., Brusa, M., et al. 2007, ApJS, 172, 1
Sherman, S., Jogee, S., Florez, J., et al. 2020, MNRAS, 491, 3318
Silva, L., Maiolino, R., & Granato, G. L. 2004, MNRAS, 355, 973
Simpson, J. M., Smail, I., Wang, W.-H., et al. 2017, ApJL, 844, L10
Simpson, J. M., Smail, I., Swinbank, A. M., et al. 2019, ApJ, 880, 43
Simpson, J. M., Smail, I., Dudzevičiūtė, U., et al. 2020, MNRAS, 495, 3409
Smit, R., Bouwens, R. J., Labbé, I., et al. 2016, ApJ, 833, 254
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
Smolčić, V., Novak, M., Bondi, M., et al. 2017, A&A, 602, A1
Sobral, D., Best, P. N., Smail, I., et al. 2014, MNRAS, 437, 3516
Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, ApJS, 214, 15
Spitler, L. R., Straatman, C. M. S., Labbé, I., et al. 2014, ApJL, 787, L36
Stark, D. P., Schenker, M. A., Ellis, R., et al. 2013, ApJ, 763, 129
Stefanon, M., Bouwens, R. J., Labbé, I., et al. 2021, ApJ, 922, 29
Stefanon, M., Bouwens, R. J., Labbé, I., et al. 2017, ApJ, 843, 36
Stefanon, M., Marchesini, D., Muzzin, A., et al. 2015, ApJ, 803, 11
Stefanon, M., Labbé, I., Bouwens, R. J., et al. 2019, ApJ, 883, 99
Stockmann, M., Toft, S., Gallazzi, A., et al. 2020, ApJ, 888, 4
Straatman, C. M. S., Labbé, I., Spitler, L. R., et al. 2014, ApJL, 783, L14
Straatman, C. M. S., Labbé, I., Spitler, L. R., et al. 2015, ApJL, 808, L29
Straatman, C. M. S., Spitler, L. R., Quadri, R. F., et al. 2016, ApJ, 830, 51
Tanaka, M., Valentino, F., Toft, S., et al. 2019, ApJL, 885, L34
Thomas, D., Maraston, C., Bender, R., & Mendes de Oliveira, C. 2005, ApJ, 621, 673
Thomas, D., Maraston, C., Schawinski, K., Sarzi, M., & Silk, J. 2010, MNRAS, 404, 1775
Toft, S., Franx, M., van Dokkum, P., et al. 2009, ApJ, 705, 255
Toft, S., Smolčić, V., Magnelli, B., et al. 2014, ApJ, 782, 68
Toft, S., Zabl, J., Richard, J., et al. 2017, Natur, 546, 510
Tomczak, A. R., Quadri, R. F., Tran, K.-V. H., et al. 2014, ApJ, 783, 85
Tomczak, A. R., Quadri, R. F., Tran, K.-V. H., et al. 2016, ApJ, 817, 118
Trakhtenbrot, B., Lira, P., Netzer, H., et al. 2017, ApJ, 836, 8
Valentino, F., Tanaka, M., Davidzon, I., et al. 2020, ApJ, 889, 93
van der Wel, A., Franx, M., van Dokkum, P. G., et al. 2014, ApJ, 788, 28
van Dokkum, P. G., & van der Marel, R. P. 2007, ApJ, 655, 30
van Dokkum, P. G., Franx, M., Kriek, M., et al. 2008, ApJL, 677, L5
van Dokkum, P. G., Whitaker, K. E., Brammer, G., et al. 2010, ApJ, 709, 1018
Wang, T., Elbaz, D., Schreiber, C., et al. 2016, ApJ, 816, 84
Wang, T., Schreiber, C., Elbaz, D., et al. 2019, Natur, 572, 211
Wardlow, J. L., Simpson, J. M., Smail, I., et al. 2018, MNRAS, 479, 3879
Whitaker, K. E., Labbé, I., van Dokkum, P. G., et al. 2011, ApJ, 735, 86
Whitaker, K. E., van Dokkum, P. G., Brammer, G., et al. 2013, ApJL, 770, L39
Whitaker, K. E., Franx, M., Leja, J., et al. 2014, ApJ, 795, 104
Whitaker, K. E., Franx, M., Bezanson, R., et al. 2015, ApJL, 811, L12
Wiklind, T., Dickinson, M., Ferguson, H. C., et al. 2008, ApJ, 676, 781
Williams, C. C., Labbe, I., Spilker, J., et al. 2019, ApJ, 884, 154
Williams, R. J., Quadri, R. F., Franx, M., van Dokkum, P., & Labbé, I. 2009, ApJ, 691, 1879
Williams, R. J., Quadri, R. F., Franx, M., et al. 2010, ApJ, 713, 738
Wuyts, S., Labbé, I., Förster Schreiber, N. M., et al. 2008, ApJ, 682, 985
Wuyts, S., Förster Schreiber, N. M., Lutz, D., et al. 2011, ApJ, 738, 106