HOW DO STAR-FORMING GALAXIES AT $z > 3$ ASSEMBLE THEIR MASSES?

Kyoung-Soo Lee[1], Henry C. Ferguson[2], Tommy Wiklind[3], Tomas Dahlen[2], Mark E. Dickinson[4], Mauro Giavalisco[5], Norman Grogin[2], Casey Papovich[6], Hugo Messias[7,8], Yicheng Guo[5], and Lihwai Lin[9]

[1] Department of Physics and Astronomy, Yale Center for Astronomy and Astrophysics, Yale University, New Haven, CT 06520, USA
[2] Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA
[3] ALMA/ESO, Santiago, Chile
[4] National Optical Astronomy Observatory, Tucson, AZ 85719, USA
[5] Department of Astronomy, University of Massachusetts, Amherst, MA 01003, USA
[6] Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843, USA
[7] Departamento de Astronomía, Av. Esteban Iturra 6to piso, Facultad de Ciencias Físicas y Matemáticas, Universidad de Concepción, Chile
[8] Centro de Astronomia e Astrofísica da Universidade de Lisboa, Observatório Astronómico de Lisboa, Tapada da Ajuda, 1349-018 Lisboa, Portugal
[9] Institute of Astronomy and Astrophysics, Academia Sinica, Taipei 106, Taiwan

ABSTRACT

We investigate how star-forming galaxies typically assemble their masses at high redshift. Taking advantage of the deep multi-wavelength coverage of the GOODS data set, we select two of the largest samples of high-redshift star-forming galaxies based on their UV colors and measure stellar mass of individual galaxies. We use template-fitting photometry to obtain optimal estimates of the fluxes in lower-resolution ground-based and *Spitzer* images using prior information about galaxy positions, shapes, and orientations. By combining the data and realistic simulations to understand measurement errors and biases, we make a statistically robust determination of stellar mass function (SMF) of the UV-selected star-forming galaxies at $z \sim 4$ and 5. We report a broad correlation between stellar mass and UV luminosity, such that more UV-luminous galaxies are, on average, also more massive. However, we show that the correlation has a substantial intrinsic scatter, particularly for UV-faint galaxies, evidenced by the fact there is a non-negligible number of UV faint but massive galaxies. Furthermore, we find that the low-mass end of the SMF does not rise as steeply as the UV luminosity function ($\alpha_{\rm UVLF} \approx -(1.7\text{--}1.8)$ while $\alpha_{\rm SMF} \approx -(1.3\text{--}1.4)$) of the same galaxies. In a smooth and continuous formation scenario where star formation rates (SFRs) are sustained at the observed rates for a long time, these galaxies would have accumulated more stellar mass (by a factor of ≈ 3) than observed and therefore the SMF would mirror more closely that of the UV luminosity function. The relatively shallow slope of the SMF is due to the fact that many of the UV-selected galaxies are not massive enough, and therefore are too faint in their rest-frame optical bands, to be detected in the current observations. Our results favor a more episodic formation history in which SFRs of low-mass galaxies vary significantly over cosmic time, a scenario currently favored by galaxy clustering. Our findings for the UV-faint galaxies at high redshift are in contrast with previous studies on more UV-luminous galaxies, which exhibit a tighter SFR–$M_{\rm star}$ correlation. The discrepancy may suggest that galaxies at different luminosities may have different evolutionary paths. Such a scenario presents a nontrivial test to theoretical models of galaxy formation.

Key words: galaxies: evolution – galaxies: formation – galaxies: high-redshift – galaxies: starburst – surveys

Online-only material: color figures

1. INTRODUCTION

Our knowledge of the distant universe has expanded substantially over the last decade thanks to deep multi-wavelength surveys. Currently, galaxies are routinely identified out to $z \sim 7$ and beyond, and their key quantities such as star formation rates (SFRs), stellar masses ($M_{\rm star}$), and extinction are measured via spectroscopy and multi-wavelength photometry (e.g., Papovich et al. 2001; Shapley et al. 2001; Stark et al. 2009; Mannucci et al. 2009). Large-area surveys have also enabled us to characterize the statistical properties of the high-redshift galaxy population as a whole. The UV luminosity function (UVLF) primarily reflects the distribution of SFRs within the population and cosmic SFR density (Steidel et al. 1999; Giavalisco et al. 2004b; Ouchi et al. 2004a; Sawicki & Thompson 2005; Bouwens et al. 2007; Reddy & Steidel 2009; McLure et al. 2009; Bouwens et al. 2011b). The stellar mass function (SMF) traces the distribution of stellar mass within the population and cosmic stellar mass density (Drory et al. 2005; Fontana et al. 2006; Marchesini et al. 2009, 2010; González et al. 2011). Morphological stud-

ies have quantified the relative spatial distribution and size of stellar components (Ferguson et al. 2004; Bouwens et al. 2004; Ravindranath et al. 2006; Lotz et al. 2006). Finally, the galaxy clustering measures the large-scale spatial correlation between galaxies and how it compares with that of underlying dark matter and dark matter halos (Giavalisco & Dickinson 2001; Ouchi et al. 2004b, 2005; Adelberger et al. 2005; Quadri et al. 2007; Lee et al. 2006, 2009a). Careful comparison of these statistical quantities at different cosmic epochs can provide powerful tools to gain insight into the evolutionary sequence of galaxy assembly in the high-redshift universe.

This tremendous observational progress offers the opportunity to establish whether galaxy growth is dominated by smooth accretion or more episodic events. While both modes certainly exist and are observed at high redshift, the relative importance of the two modes is currently unconstrained at $z > 3$ (but see Rodighiero et al. 2011; Elbaz et al. 2011 for measurements out to $z \sim 2$). Furthermore, there is observational evidence that the dominant mode may vary with galaxy or halo mass (e.g., van der Wel et al. 2011). A key measurement that directly

constrains the main mode of star formation (SF), or average star formation history (SFH), is the location of galaxies on the SFR–M_{star} plane. If galaxies assemble smoothly over a timescale comparable to the Hubble time, the majority of galaxies would form a tight sequence with a slope of close to unity (Noeske et al. 2007). On the other hand, in the case of bursty/episodic SF, one should observe significant scatter about the mean relation.

Unfortunately, at high redshift, different studies report conflicting results. Some find a strong correlation with a tight scatter (Daddi et al. 2007b; Pannella et al. 2009; Magdis et al. 2010; Lee et al. 2011; Sawicki 2012), while others find no or weak correlation with larger scatter (Shapley et al. 2005; Reddy et al. 2006; Mannucci et al. 2009). Some of this discrepancy may be real, if the main mode of galaxy growth changes with stellar mass or host halo mass (Renzini 2009; Lee et al. 2009a) or with different "types" of galaxies. However, it is likely that the discord between different measurements may be at least in part due to different and poorly understood selection effects suffered by different samples and/or different calibration methods employed by these studies to derive various SFRs and stellar mass.

Some of the existing samples of galaxies are based on a UV color selection, while some others use a combination of UV and optical colors. Some even require an additional detection in the mid-infrared (such as *Spitzer*), complicating the interpretations of their results in the global picture. Calibration is another issue at hand. For example, SFRs derived from polycyclic aromatic hydrocarbon (PAH) emission are known to depend strongly on galaxy metallicity and the internal distribution of H II regions and dust (Calzetti et al. 2007) even when the precise redshift is known. In addition, the presence of active galactic nuclei (AGN) in some high-redshift galaxies may contribute to the observed flux in the mid-infrared (e.g., Daddi et al. 2007a) and subsequently add to the observed scatter in the SFR–M_{star} relation unless such galaxies are identified or the AGN contribution is accounted for. These complications make it challenging to constrain the intrinsic SFR–M_{star} scaling law even for galaxies with the most comprehensive multi-wavelength and spectroscopic coverage. Furthermore, at $z > 3$, even these diagnostics become out of reach as the PAH emission redshifts to the longward of 24 μm (and Hα redshifts to >3 μm; but see Shim et al. 2011), and only galaxies with the most extreme starbursts such as submillimeter galaxies can be detected via other *Spitzer* MIPS bands or *Herschel*.

Here we take an alternative approach to constrain the main mode of SF by using two statistical measurements of galaxies, namely, the UVLF and the SMF of the *same* galaxies. The former constrains the distribution of *recent* SF in galaxies, while the latter measures the distribution of existing stellar mass in the same galaxies directly tracing the integral of all past SF activities. Hence, these two quantities represent the statistical distribution of the SFR (modulo dust) and stellar mass rather than those in individual galaxies. In essence, comparison of the two distributions will shed light on how one traces the other (when the effect of dust extinction is accounted for). If they trace each other closely, the overall shape of the two distributions will be similar. Alternatively, any difference between the two distributions can be attributed to the statistical mapping between the two as the two quantities are measured from the same galaxy population.

There are several advantages of this approach over the direct method discussed previously. First, as we will show later, the two quantities, UVLF and SMF, can be determined very robustly

by combining the observations and extensive simulations via rigorous statistical analyses. Second, our approach bypasses the need to measure SFRs very accurately for individual galaxies. As mentioned previously, the estimation of SFRs is hard even when precise photometry and redshift information are available, and it has been the primary challenge on which to make progress. Instead, we take advantage of the deepest surveys to statistically map the UV luminosity to SFRs (Bouwens et al. 2009; Reddy et al. 2010; Finkelstein et al. 2011; Castellano et al. 2012). In particular, at $z > 3$, this approach offers new possibilities to obtain useful insights into the typical mode of galaxy growth as deep imaging surveys are being conducted in the near- and mid-infrared from space (e.g., *Spitzer* Extended Deep Survey, Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey; Grogin et al. 2011; Koekemoer et al. 2011).

In this paper, we investigate how star-forming galaxies assemble their mass in the first two billion years. We take advantage of two of the deepest survey fields with the most comprehensive multi-wavelength coverage, namely, GOODS-N and GOODS-S. The combination of the large areal coverage ($\approx$0.1 deg^2 total) and depth achieved in these fields allows us to identify a large number of high-redshift star-forming galaxies and robustly measure their population properties from integrated galaxy light. We use the TFIT template-fitting photometry package to carry out the photometry and present the most extensive set of simulations to date to validate the technique and quantify the uncertainties. As we will demonstrate later, TFIT greatly improves the photometric accuracy for individual measurements over the conventional methods, as well as dramatically increases the number of sources with reliable multi-wavelength photometry.[10]

Throughout this work, we use $(\Omega_m, \Omega_\Lambda, \sigma_8, h_{100}) = (0.28, 0.72, 0.9, 0.72)$. Magnitudes are given in the AB system (Oke & Gunn 1983) unless noted otherwise.

2. DATA, SAMPLE SELECTION, AND MULTI-WAVELENGTH PHOTOMETRY

2.1. Data

The data set we use for our analyses consists of those obtained as part of the Great Observatories Origins Deep Survey (GOODS; Giavalisco et al. 2004b). The two fields, GOODS-N and GOODS-S, each cover roughly 160 arcmin^2 and include deep *Hubble Space Telescope* (*HST*)/Advanced Camera for Surveys (ACS) $B_{435}V_{606}i_{775}z_{850}$ (F435W, F606W, F775W, and F850LP) and deep *Spitzer* IRAC [3.6 μm], [4.5 μm], [5.8 μm],[11] and MIPS [24 μm] data. The GOODS-S field is covered by deep *J*-, *H*-, and K_S-band data taken with the ISAAC camera and the *U*-band data taken with the Visible Multiobject Spectrograph (VIMOS) on the Very Large Telescope (VLT; Retzlaff et al. 2010; Nonino et al. 2009, respectively). The GOODS-N field is covered by the *J*- and K_S-band data taken with the Wide-Field Infrared Camera (WIRCAM) on the Canada–France–Hawaii Telescope (CFHT; Wang et al. 2010, L. Lin et al. 2012, in preparation) and the *U*-band data taken with the MOSAIC prime focus camera (Capak et al. 2004) on the Kitt Peak National Observatory (KPNO) 4 m telescope.

---

[10] For example, Stark et al. (2009) noted that only 35% of the sources at $z \sim 4$ are sufficiently isolated to perform robust aperture photometry. On the other hand, this work using TFIT rejects less than 5% due to blending.

[11] We note that the majority of our samples are not detected in the [5.8 μm] and [8.0 μm] bands.

Table 1
The 5σ Limiting Magnitudes of the Data Set

Field	U	B_{435}[a]	V_{606}[a]	i_{775}[a]	z_{850}[a]	J	H	K_S	(3.6 μm)	(4.5 μm)	(5.8 μm)
GOODS–N	27.2	28.4	28.6	27.9	27.6	25.0	. . .	24.5	26.2	25.6	23.5
GOODS–S	28.1	28.4	28.6	27.9	27.6	25.7	25.5	25.1	26.2	25.6	23.5

Note. [a] For the ACS bands, the limiting magnitudes are computed within a 0″.2 diameter aperture. For the rest, the isophotal aperture defined in the detection band is used (see Section 2.4).

The 5σ limiting magnitudes of the data are tabulated in Table 1. Therein, the limiting magnitudes for the ACS bands are computed within a 0″.2 diameter aperture. For the rest of the photometric bands, we compute the photometric uncertainties (1σ errors) within the isophotal apertures defined in the detection band (z_{850}-band; see Section 2.4 for details).

2.2. "Dropout" Samples at z ~ 4 and 5

We adopt the Lyman-break technique (Steidel et al. 1996) to identify high-redshift star-forming galaxies ("Lyman-break galaxies" or LBGs hereafter) at $z \sim 4$ and ~5 with the color selection criteria discussed in Giavalisco et al. (2004a) and Lee et al. (2006, 2009a). Extensive spectroscopy campaigns have shown that these selection methods are very robust with minimal contamination by interlopers (Vanzella et al. 2006, 2009). While the adopted selection criteria differ slightly from others found in the literature (e.g., Bouwens et al. 2007), our main conclusions should remain identical to those derived using different selection as the proper account of the selection efficiency will make appropriate corrections to compensate for the difference. We used the UV color criteria used by Giavalisco et al. (2004a) to select galaxies at $z \sim 4$ and $z \sim 5$ as

$$(B_{435} - V_{606}) \geqslant 1.2 + 1.4 \times (V_{606} - z_{850})$$
$$\wedge (B_{435} - V_{606}) \geqslant 1.2 \wedge V_{606} - z_{850} \leqslant 1.2,$$

and

$$(V_{606} - i_{775}) > 1.5 + 0.9 \times (i_{775} - z_{850})$$
$$\vee (V_{606} - i_{775}) > 2.0 \wedge (V_{606} - i_{775}) \geqslant 1.2$$
$$\wedge (i_{775} - z_{850}) \leqslant 1.3$$
$$\wedge S/N(B_{435}) < 2,$$

where the symbols ∨ and ∧ are the logical "OR" and "AND" operators, respectively. In both samples, only sources with $S/N(z_{850}) \geqslant 6$ are considered. In addition to the color criteria, we also removed sources that have stellarity index (using the CLASS_STAR parameter in the SExtractor; Bertin & Arnouts 1996) greater than 0.8 from the sample when the source is brighter than 26.2 mag in the detection band (z_{850} band). For sources fainter than this limit, our simulation suggests that the stellarity measurement is not as reliable.

The number of photometric candidates satisfying these selection criteria is 3088 and 987 at $z \sim 4$ and 5, respectively. We refer to these samples as B_{435}-band and V_{606}-band dropouts hereafter as their targeted redshift range requires them to "drop out" in those passbands. Finally, we removed the most likely low-redshift interlopers based on photometric redshifts from Dahlen et al. (2010), which are calibrated against the largest compilation of spectroscopic redshifts available in the GOODS fields (Wirth et al. 2004; Cowie et al. 2004; Vanzella et al. 2006, 2009; D. Stern et al. 2012, in preparation). Using their estimates, we remove sources whose total integrated probability $P(z < 3)$

is greater than 70%. The number of sources removed using this criterion is 136 and 131 galaxies at $z \sim 4$ and 5, respectively. Figures 15 and 18 of Dahlen et al. (2010) show the distribution of photometric redshifts for the B_{435}- and V_{606}-band dropout samples.

2.3. Spitzer MIPS-detected Sources

We identify sources with the *Spitzer* MIPS 24 μm detection (with the formal signal-to-noise ratio S/N $\geqslant$ 3). A bona fide high-redshift galaxy at $z > 3$ with 24 μm detection (sampling the rest frame $\lambda \approx 4$–6 μm) implies the presence of unusually strong hot dust components produced by either an AGN or an extreme starburst (SFR $\gtrsim$ several hundreds $M_\odot$ yr^{-1}). Indeed, of the B_{435}- and V_{606}-band dropouts that are genuinely detected at 24 μm, several are detected in the deep X-ray data (Alexander et al. 2003; Luo et al. 2008), further supporting the possibility that they harbor an AGN. We cross-matched the GOODS MIPS catalogs with the dropout lists using a tolerance of 0″.5 to minimize blending problems. For the B_{435}-band dropouts in the GOODS-N, we find 24 sources with MIPS detections (four of which are detected in X-ray). Of those, 11 objects have spectroscopic redshifts (3 are foreground interlopers at $z < 3$ and the other 8 at $3.403 < z < 4.604$). In the GOODS-S, we find 24 sources with MIPS detections (5 of which are X-ray detected). Of the total of seven objects with spectroscopic redshifts, one lies at $z = 2.797$, while the remaining six lie at $3.055 < z < 3.891$ on the low side of the redshift distribution for the B_{435}-band dropouts. Four of the six sources at $3.055 < z < 3.891$ have X-ray detections. In summary, 44 out of 48 MIPS detections in the combined GOODS-North and South field have $f(24 \,\mu m) < 82 \,\mu Jy$. Three spectroscopically confirmed B_{435}-band dropouts have $100 \,\mu Jy < f(24 \,\mu m) < 160 \,\mu Jy$ and are also X-ray sources, thus likely AGNs.

As for the V_{606}-band dropouts in the GOODS-N, we find four MIPS detections, none of which have spectroscopic redshifts available. However, two of them are very bright in 24 μm (113 μJy and 347 μJy) and thus are very likely to be foreground interlopers. In the GOODS-S, we find 11 MIPS detections, 3 of which are confirmed to be foreground interlopers, while another galaxy lies at $z = 4.762$ on the low side of the redshift distribution for the V_{606}-band dropouts. The galaxy at $z = 4.762$ is also a known submillimeter galaxy (Coppin et al. 2009). For the remaining seven, without spectroscopic redshifts, one is unusually bright (109 μJy) to be at $z > 4.5$. In summary, for the combined GOODS-N and GOODS-S V_{606}-band dropouts, 12 out of 15 MIPS detections have $f(24 \,\mu m) < 60 \,\mu Jy$. Two with no spectroscopic confirmation have $f(24 \,\mu m) \gtrsim 110 \,\mu Jy$ (no X-ray detection), and another has $f(24 \,\mu m) = 347 \,\mu Jy$ (also without spec-z or X-ray).

Based on these statistics, we conclude that rather few bona fide B_{435}-band dropouts and almost no V_{606}-band dropouts are detected at 24 μm. Thus, we exclude all the sources with MIPS detection from our sample. However, we note that inclusion or
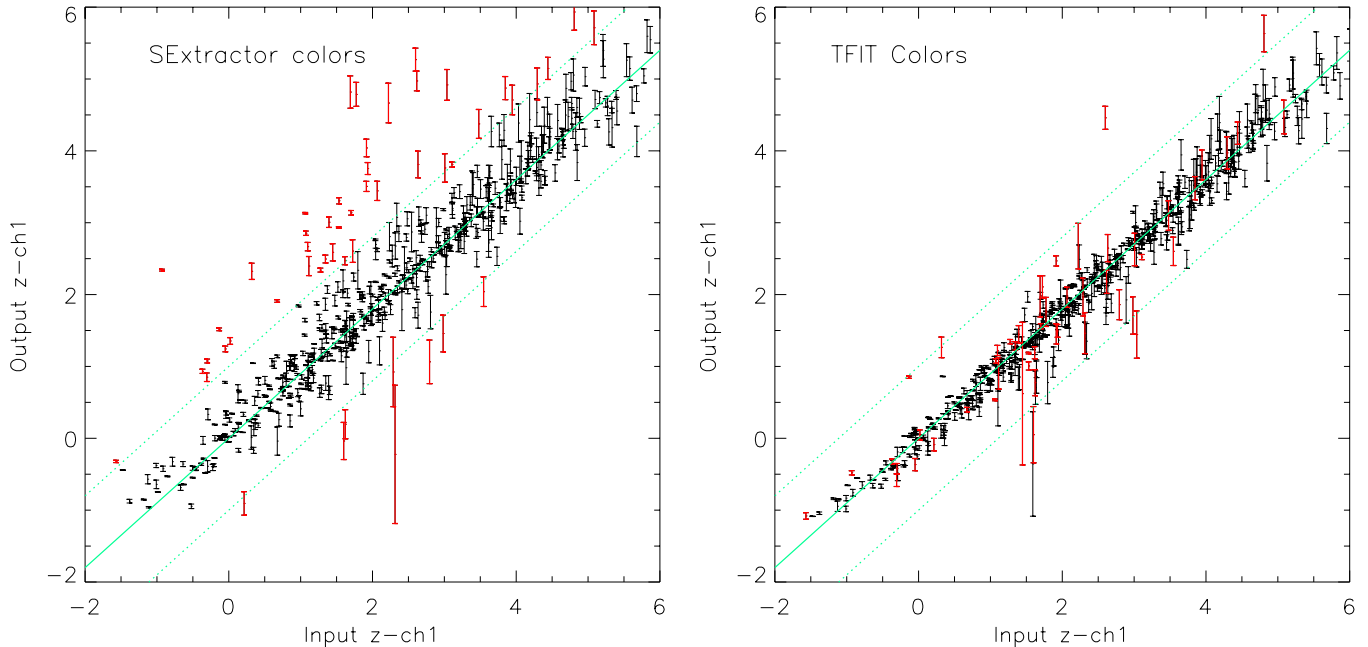
Figure 1. Optimal multi-wavelength photometry by TFIT. Extended sources with a range of colors are simulated to compare the performance of conventional aperture photometry and TFIT. Left: the $z - ch1$ (ACS z_{850} band $-$ *Spitzer* [3.6 μm]) colors measured by SExtractor are compared with the input colors for 600 simulated galaxies. Large scatter is expected due to source blending and confusion. Right: the colors of the same sources measured by TFIT. The sources with the "aperture color" bias >1 mag are shown in red on both panels. Note that most of the same sources are successfully recovered in TFIT-derived colors.

exclusion of these sources makes little difference in our main conclusions as their number is few.

2.4. Template-fitting Photometry (TFIT)

Reliable estimates of physical parameters begin with reliable photometry. This is a challenge for mixed-resolution data sets like GOODS, where the point-spread function (PSF) width varies by an order of magnitude across the full wavelength range. Galaxies that are cleanly identified in the *HST* images may be horribly blended with their neighbors in the *Spitzer* images.

Template-fitting (TFIT) photometry is designed to overcome these challenges and provide accurate flux estimates and colors for galaxies in multi-wavelength mixed-resolution data sets such as the data used in this work. While we refer interested readers to Papovich et al. (2001) and Laidler et al. (2007) for more details on the TFIT algorithm, we also note that similar algorithms have been implemented by others to obtain optimal photometry (e.g., Fernández-Soto et al. 1999; Labbé et al. 2006; Grazian et al. 2006; de Santis et al. 2007). Here, we introduce basic concepts of TFIT to highlight the main difference between this work and previous ones in the literature. We further demonstrate the effectiveness of TFIT for faint high-redshift galaxies by directly comparing the TFIT results with the conventional methods, namely, aperture photometry.

The basic premise of TFIT is that the best flux of a source can be estimated by simultaneous fitting of all the adjacent sources rather than just performing photometry within a fixed aperture. This strategy is particularly effective for crowded fields where sources are often substantially blended with one another and PSFs are large enough to not just affect their immediate neighbors but possibly those farther out from any given source. Even though deep fields such as GOODS are by design mostly devoid of bright sources in the optical bands, the same field is crowded in the deep IRAC data.

In order to achieve simultaneous fitting of multiple sources, TFIT constructs a realistic model of how an object would appear in low-resolution data based on the observed morphology of the object in a deep high-resolution image. The cutout image of each object is created from the high-resolution data and then convolved with the kernel designed to reproduce the PSF of the low-resolution image and block-averaged to the desired pixel scale. The image created with this procedure is called a *template*, which provides the best approximation of how the same object appears in the low-resolution image. This assumes no significant dependence of galaxy morphology on wavelength—i.e., no morphological *k*-correction. While such an assumption may not be valid in some cases, this can be in part overcome by using the high-resolution image observed in the wavelength range closest to the low-resolution one. For the GOODS TFIT photometry, we use the z_{850} band to create templates for IRAC photometry and the B_{435} band for the *U*-band photometry.

Once the templates of all galaxies are created, TFIT performs chi-square minimizations on subsections of the data to find the best-fit fluxes for each object. Furthermore, TFIT at this stage corrects for any second-order astrometric misregistration that may be present for data sets taken with different instruments. Such misregistration can bias the measured flux and colors not only for TFIT but also for aperture photometry as well.

We carried out simulations to compare the performance of TFIT with aperture photometry returned by SExtractor (Bertin & Arnouts 1996). Artificial galaxies (with known input fluxes) are inserted into the real data (ACS and IRAC), and fluxes are recovered by both aperture photometry and TFIT. As for aperture photometry, we adopt MAG_AUTO for both ACS and IRAC as the best estimate. Figure 1 shows the $z - [3.6\,\mu m]$ colors measured from SExtractor (left) and TFIT (right) in comparison with the true input colors of the same sources. It is evident that the colors measured by TFIT are much closer to the true values than those by aperture photometry, and also that photometric scatter is considerably smaller. Furthermore,

catastrophic outliers identified in aperture photometry (shown in red in both panels) are minimized in TFIT photometry as they are located much closer to the one-to-one line (green) in TFIT colors. This is not surprising because such errors typically result from significant contamination due to crowding. Such catastrophic failures are expected to increase in deeper images where the source density is higher. We also note that such cases are expected to be frequent for faint high-redshift sources whose flux can be significantly affected by usually brighter neighbors when the conventional method is adopted.

Throughout this work, we utilize the GOODS Grand Unified TFIT catalog[12] to measure the spectral properties of LBGs. The catalog includes the SExtractor photometry of the HST/ACS band (B_{435}, V_{606}, i_{775}, z_{850}) and TFIT photometry for the remainder of our photometric bands ($UJHK_S$, [3.6 μm], [4.5 μm], [5.8 μm]). The ACS z_{850} band (F850LP) was used as the detection band for the source extraction and to define isophotal area for other ACS bands and also as high-resolution templates for the TFIT photometry. This ensures that the measured flux by either SExtractor or TFIT will return accurate colors within the same isophotal apertures, and that aperture correction is straightforward to handle. Galaxy colors are computed using the MAG_ISO for the ACS bands and TFIT fluxes for the others, while MAG_AUTO in the z_{850} band is taken as the total magnitude. The limiting magnitude for each band is estimated from the distribution of TFIT photometric errors (1σ) for all galaxies in the catalog, which typically exhibits a well-defined peak. The peak value of the error distribution represents the typical uncertainty given the sensitivity of the data. Whenever TFIT fluxes are consistent with zero within errors, we take the 1σ errors as upper flux limits.

Finally, we have compared the GOODS TFIT catalog with the public GOODS-MUSIC catalog (v2.0) discussed in Santini et al. (2009), which uses a similar algorithm (CONPHOT) to deal with source blending (de Santis et al. 2007). The photometries from the two catalogs are found to be consistent with each other, with the median offset less than 0.1 mag down to 26.0 mag (for IRAC [3.6 μm]). The scatter, however, is quite large ($\approx$0.27 mag) toward the faintest bin. While the origin of the relatively large scatter is beyond the scope of this work as it is difficult to quantify without detailed analyses coupled with simulations, no systematic bias found between the two catalogs implies that our main results will remain robust against the particular choice of algorithms adopted for optical photometry discussed above.

3. MEASURING STELLAR MASSES OF HIGH-REDSHIFT GALAXIES

3.1. SED Fitting

Stellar population synthesis models are used to derive stellar masses of galaxies in our sample. We use the updated version of the stellar population synthesis models of Bruzual & Charlot (2003, known as the BC03 model), which we refer to as the CB07 code hereafter. We use the CB07 models to compute the broadband color evolution of galaxies with different SFHs and ages. We use a Chabrier (2003) initial mass function (IMF) with a lower and an upper mass cutoff at 0.1 $M_{\odot}$ and 100 $M_{\odot}$, respectively, while the metallicity is fixed to solar. While the CB07 model includes enhanced contribution from the TP-AGB stars (see, e.g., Maraston 2005), it returns stellar masses very

---

[12] The catalog is soon to be made available at http://www.stsci.edu/science/goods/DataProducts/.

similar to the BC03 because the light from UV-selected galaxies is dominated by younger OB stars even at the rest-frame optical wavelengths sampled by the IRAC channels at $z > 3$. The TP-AGB stars would contribute more at longer wavelengths (rest-frame near-infrared) and for older stellar populations (>1 Gyr; also see Figure 5 in Stark et al. 2009).

The model galaxy spectral energy distributions (SEDs) are redshifted in the range of $z = 0.2$–8 with $\Delta z = 0.1$. Internal dust extinction is computed according to the prescription of the Calzetti et al. (2000) law in the range of $E(B - V) = 0.0$–0.95 with $\Delta E(B - V) = 0.025$. We also attenuated the resulting SEDs from the neutral hydrogen absorption in the intergalactic medium using the Madau (1995) prescription. The age of our model SEDs extends from 5 Myr to the age of the universe at the given redshift. The SFH is parameterized as an exponentially decreasing SFR, where τ represents the e-folding decay time. We use $\tau = 0.1, 0.2, 0.3, 0.4, 0.6, 0.8$, and 1.0 Gyr and constant star formation (CSF; $\tau \rightarrow \infty$). We precompute a large number of model SEDs spanning the parameter space described above and then integrate them through the filter response functions of all the observed bands (HST/ACS: $B_{435} V_{606} i_{775} z_{850}$; $Spitzer$/IRAC: [3.6 μm], [4.5 μm], [5.8 μm]; VLT/ISAAC: JHK_S; CFHT/WIRCAM: JK_S, KPNO/MOSAIC U band, and VIMOS/U band) to compute the galaxy colors for each point in our model grid.

Using the precomputed model grid, we find the best-fit parameters for each galaxy in our sample. To evaluate the goodness of fit, we first determine the overall normalization to match the observed photometry and then compute the chi-square using all relevant data points. Because model SEDs are normalized to have total mass of 1 $M_{\odot}$, the normalization factor directly determines the stellar mass. For each galaxy, we also determine the absolute UV magnitude at 1700 Å, M_{1700}, by interpolating the best-fit model spectra. The evaluation of the chi-square value is done in flux units rather than magnitudes, which helps in properly handling the data points with non-detection. Whenever the photometry in a given passband is consistent with zero within the 1σ error, we assign as a flux upper limit a value that is representative of the sensitivity limit for the band. This treatment is often useful to avoid a large stellar mass being assigned as a solution when TFIT returns very large photometric uncertainties. Such cases can occur when a source is hopelessly blended with much brighter neighbors. However, in most cases, using TFIT-returned flux and flux errors (instead of flux upper limits) for the fitting should yield similar results. Data points with upper limits are included in the chi-square evaluation only when the model flux is higher than the upper limit.

The flux upper limits in these bands are listed in Table 1, which are similar but not identical to those given in Stark et al. (2009) for the same data. The main reasons for this slight difference are twofold. First, we used the source-weighted rms map rather than the exposure map, which helps make a more realistic estimate of the noise near bright sources. Second, TFIT includes the flux uncertainties arising from source blending, which, in particular, for faint sources like most high-redshift galaxies, is an important factor to account for as the majority of them are blended with brighter neighbors.

Instead of using the conventional approach of finding a single value as the best-fit stellar mass, we determine the full probability distribution function for each galaxy. In practice, this means that we retain all models that are within $\Delta \chi_r^2 = 1$ from the minimum χ^2 model as acceptable fits, and thus the stellar
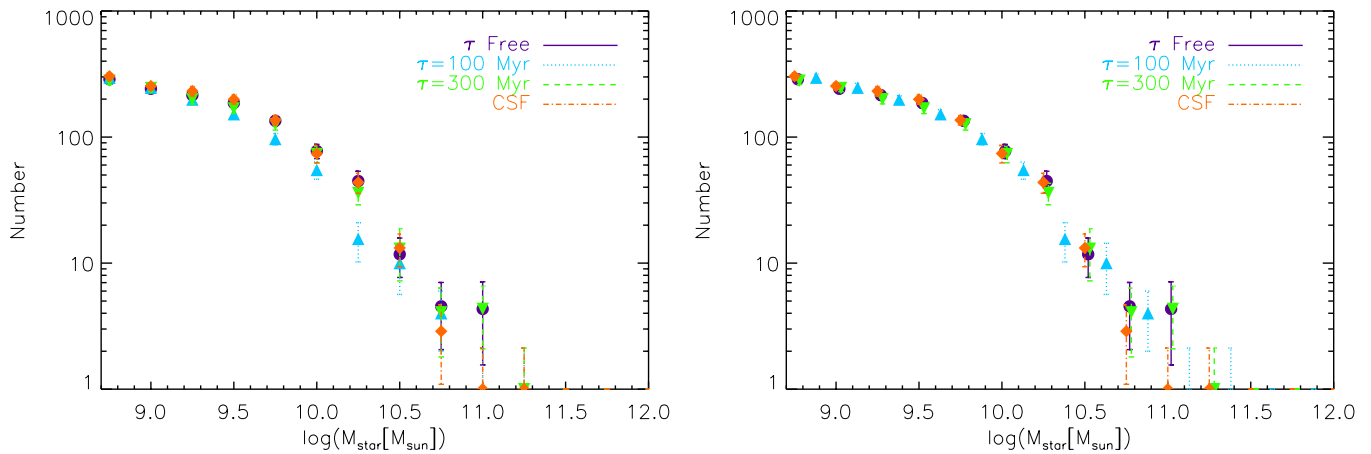
Figure 2. Left: the observed number counts of LBGs at $z \sim 4$ in stellar mass bins when different SF histories are assumed. Larger τ values generally return higher stellar-mass-to-UV-light ratios and thus slightly larger masses for given rest-frame optical fluxes. Right: when a small offset is applied to the masses to correct for this effect, the number counts using different SFHs agree with one another very well. This means that even if mass measurements are biased toward lower or higher masses because of our assumption in SF histories, the overall shape of the number counts is likely robust.

mass of the galaxy is described by a histogram of stellar masses spanned by all acceptable models. While this approach returns results that are fully consistent with the conventional method when stellar mass of a given galaxy is in question, it has several advantages from a statistical standpoint. For example, there can be multiple models with χ^2 values very close to one another but over a range of stellar mass. These situations become very real when one is dealing with galaxies at the low-mass end. Picking one model over the others may bias the outcome in an unpredictable fashion unless the systematics that affect the fitting are well understood. Choosing to construct the number counts using the full probability distribution makes it easier to correct the "observed" number counts for the measurement biases and systematics, which can be quantified using extensive simulations of galaxies. In Section 4, we present our simulation procedures in detail and further discussions on how best to measure the galaxy counts in stellar mass bins given the photometric uncertainties.

The redshift distributions and the accuracy of photometric redshifts for our samples are well understood (Vanzella et al. 2009; Dahlen et al. 2010). Thus, we rely on photometric redshift estimation to derive physical parameters of each galaxy. During the SED fitting, redshift is varied within the best-fit photometric redshift by $\pm\Delta z = 0.125$, which accounts for a typical uncertainty of the photometric redshift estimation (see Figure 9 in Dahlen et al. 2010), except when the accurate spectroscopic redshift is known. Varying the redshifts is important in obtaining realistic errors in the derived SED parameters for these galaxies as the majority (95% of the full B_{435}-band dropout sample) does not have a secure spectroscopic redshift. Typically, the errors in stellar mass when the redshift is varied are 20%–25% larger than those when the redshift is fixed to the best-fit photometric redshift.
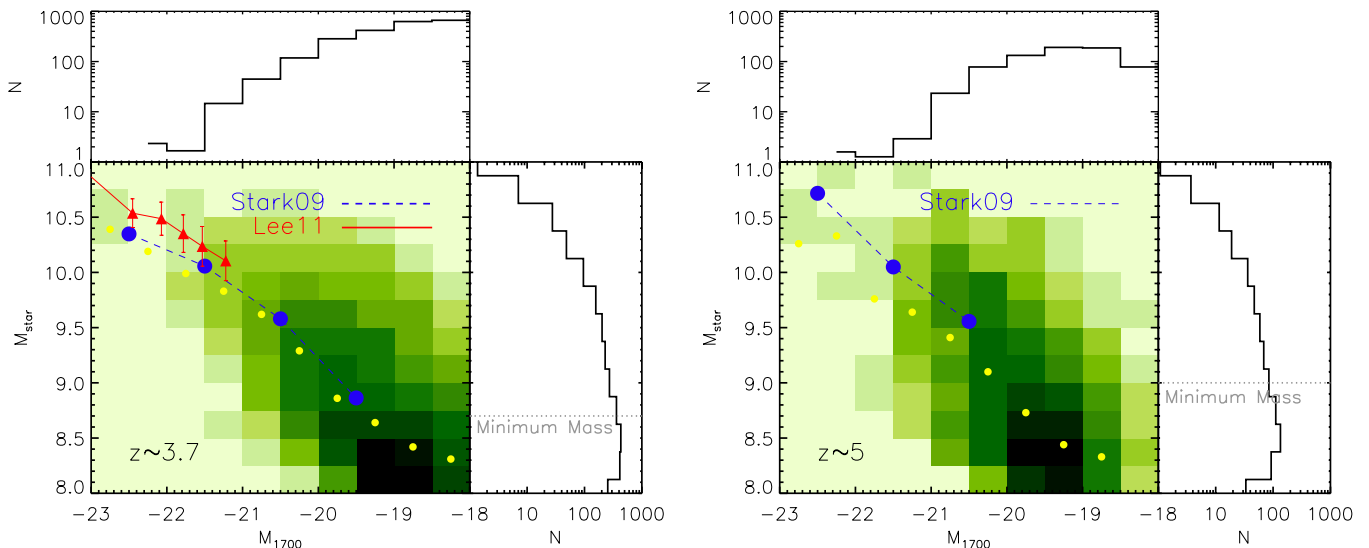
3.2. The Observed Number Counts of Galaxies

Following the procedure described above, we measured the stellar masses of the B_{435}- and V_{606}-band dropouts and constructed the number counts. In order to assess robustness of our results, we investigate how the final number counts depend on the assumed SFH. In Figure 2, we compare the number counts in the stellar mass bins for the B_{435}-band dropouts when different SFHs are assumed. While we only include CSF history

and exponentially declining τ models in our fitting procedure, comparing the results using a short τ model (100 Myr, in this case) and CSF model gives an idea of how the results would differ for scenarios in which SF is assumed to take place on short, "bursty" timescales, versus longer, more continuous modes.

We find that the observed number of galaxies in a given mass bin is slightly higher when a larger τ value is assumed as shown in the left panel of Figure 2. This is because models with a larger τ typically result in a higher stellar-mass-to-UV-light ratio (mass-to-light ratio, hereafter) and thus predict slightly larger stellar mass for given rest-frame optical fluxes. However, different mass-to-light ratios seem to result in a solid shift rather than change in the overall shape of the number counts. They agree well with one another when a small offset is applied to correct for different mass-to-light ratios. In the right panel of Figure 2, we show the same number counts with a small offset $\Delta \log[M_{star}/M_{\odot}] = 0, 0.13, 0.03, 0.0$ for the free-τ model, $\tau = 100, 300$, and CSF model, respectively. This shows that the overall shape of the number counts can be measured robustly regardless of the assumed SFH. One possibility is that a typical τ value depends systematically on galaxy UV luminosity or stellar mass. For example, if low-mass galaxies have, on average, shorter SF timescales, such a dependency would manifest as a slightly steeper slope on the low-mass end. While it is difficult to model such an effect without more observational or theoretical input, Figure 2 suggests that such an effect should have a relatively small impact on the galaxy counts. Throughout this work, we use the number counts measured with "free-τ" models for our analyses unless stated otherwise.

There are several factors that may contribute, to a varying degree, to the estimation of the number counts. First, a hidden population of old stars formed from a previous generation of SF may contribute to the stellar mass. The upper limit on this contribution can be set by assuming a population formed from an instantaneous burst at a very high redshift, which maximizes the mass-to-light ratio. Based on the multi-wavelength data set (out to K band, which corresponds to the rest-frame V-band) for $z \sim 3$ galaxies selected from the HDF-N, Papovich et al. (2001) estimated that galaxies, on average, can be hiding stellar mass up to a factor of three. While this can certainly be the case for some galaxies in our sample, the relative contribution from an old population is likely smaller because the universe is younger at $z \sim 4$ and ~ 5, and thus

Figure 3. Number counts of LBGs in UV luminosity and stellar mass bins at $z \sim 4$ (left) and $z \sim 5$ (right). We show in the main panels the location of all galaxies in our sample. Each cell is color-coded with the observed number of galaxies. The average stellar mass at a fixed UV magnitude is also shown as yellow filled symbols. Our results indicate an overall correlation between UV luminosity and stellar mass albeit with a significant scatter. Our estimates are in good agreement with those of Stark et al. (2009), as marked by dashed line and filled symbols, but extend further down to lower masses using TFIT photometry. For $z \sim 4$, we also show similar measurements by Lee et al. (2011) for very UV-luminous ($M_{1700} < -21$) galaxies. In upper panels and right panels, we show the galaxy number counts in UV magnitudes and stellar mass bins, respectively. The minimum mass we include in our stellar mass function analyses, which we determine based on our simulations (Section 4), is also indicated on right panels. Our number counts in UV magnitudes are in excellent agreement with Bouwens et al. (2007), confirming that the UV luminosity function has a steep faint-end slope. A large concentration of UV-faint galaxies in low masses is due largely to non-detection in IRAC channels and thus represents upper mass limits in some galaxies. This suggests that at $M_{1700} > -19$ the median stellar mass for galaxies falls more steeply than that for more luminous ones.

the galaxies have had less time to have formed a significant stellar population unrelated to the current SF. Furthermore, the *Spitzer* data sample out to, at least, the rest-frame z (I) band, ≈ 9000 (7500) Å, at $z = 4$ (5) as the majority of the galaxies with stellar mass $M_{star} \geqslant 10^{8.7}\ M_{\odot}$ are detected in the [3.6 μm] and [4.5 μm] bands. For galaxies that are not formally detected in the *Spitzer* bands, it is difficult to assess how much mass may be hidden in old stellar populations. However, the average SED of *Spitzer*-faint galaxies constructed by stacking (see Figure 8 of Stark et al. 2009) suggests that it is unlikely that such a contribution is more substantial than their more luminous counterparts. Furthermore, a large population of maximally old galaxies would be inconsistent with the rapid redshift evolution of the UV SFR density and UVLF observed at $z \gtrsim 4$ (Bouwens et al. 2007, 2008).

Another interesting possibility concerns a recent result by Shim et al. (2011), who reported a strong Hα line emission in a large fraction (>70%) of galaxies detected in the *Spitzer* [3.6 μm] and [4.5 μm] with high S/Ns. While the study is based on a relatively small sample (70 galaxies) of spectroscopically confirmed sources at $3.8 < z < 5.0$, they convincingly showed that the flux excess in the *Spitzer* [3.6 μm] band is indeed due to strong Hα emission from SF and not from hidden AGN activities. From the viewpoint of measuring stellar mass using broadband photometry, the presence of strong emission lines can skew the mass estimation because emission lines are not included in our SED modeling. Furthermore, the observed equivalent widths of the Hα emission, they claimed, suggest that it requires extremely young ages, top-heavy IMF, or sub-solar metallicity (or a combination of these). In practice, IRAC [3.6 μm] is one of the main drivers in the SED fitting procedure that determines the stellar mass as it samples the rest-frame I band. We visually inspected the observed SEDs of the B_{435}-band dropout sample to check whether such an excess is common also in the photometric sample. While the effect of line emission in

the photometric data is fairly subtle to identify visually and may be particularly hard to see for the fainter galaxies without spectroscopic redshifts, we determine that the number of sources with a noticeably large [3.6 μm] excess is indeed small, and thus unlikely to result in change in the overall shape of the galaxy number counts. Because the strength of Hα emission line should be lower for UV-faint (lower SFR) galaxies (a typical z_{850}-band magnitude for the Shim et al. (2011) sample is 25.0, while the majority (>95%) of the galaxies in our sample are fainter than 25.0 mag), the lack of Hα-excess sources seems to be in line with a possibility that the line emission is indeed due to SF. Regardless of the nature of the flux excess reported by Shim et al. (2011), we conclude that it is unlikely that it will change our results significantly.

3.3. The Observed M_{UV}–M_{star} Scaling Law

Having measured both UV luminosity and stellar mass of the dropout samples, we present how these galaxies populate the plane of the observed luminosity and stellar mass in the main panels of Figure 3. The figure is color-coded by the total number of galaxies in each given cell and thus illustrates the overall trend of the relative location of galaxies on the plane, as well as the number of galaxies at different mass and luminosity bins. The average stellar mass at a fixed UV luminosity is also computed from our measurements as shown as yellow filled circles. Top and right panels indicate the sum of all columns and rows (of the main panels), i.e., galaxy counts as a function of UV luminosity and stellar mass, respectively. We also note that a large concentration of UV-faint galaxies at very low masses ($10^{8.0}$–$10^{8.5}\ M_{\odot}$) is partly an artifact produced by a combination of flux upper limits and our adopted fitting method. When a source is not formally detected in the *Spitzer* IRAC channels, the minimum χ^2 is achieved when the model flux is below all the flux upper limits set by the data. As a result, the best-fit solution is returned with a range of low stellar masses, which is primarily

determined by a range of stellar-mass-to-UV-light ratios allowed within our model grid (including the models with the lowest mass-to-light ratios as the data do not provide any constraints), and thus does not reflect a true value. To avoid any systematic bias in our results by this effect, we confine our analyses only to galaxies with stellar mass $M_{star} \geqslant 10^{8.7} M_\odot$ for the B_{435}-band dropouts and $M_{star} \geqslant 10^9 M_\odot$ for the V_{606}-band dropouts. These limits, the minimum stellar mass considered in our analyses, are also indicated in Figure 3 on right panels as "Minimum Mass." Our simulations (Section 4) indicate that the current data allow statistically robust measurements down to this limit. A majority of simulated galaxies in this mass range are formally undetected in the IRAC bands or have large photometric errors.

Our measures are largely consistent with those made by Stark et al. (2009), marked by dashed lines, but are generally slightly lower (by ∼0.1 dex) than their estimate. The discrepancy may be due to the fact that they adopted a single value for each stellar mass rather than taking into account the full distribution, and as a result, the median is slightly skewed toward higher masses. As we will show in Section 4, the uncertainty in the mass determination at $10^9 M_\odot$ is somewhat large and thus requires a statistical approach to construct the distribution. Furthermore, the discrepancy may reflect the difference in the two samples that ours includes a much larger number of galaxies (≈90% of the photometric sample in comparison to their 35%) because TFIT photometry does not require that sources be well isolated.

In Figure 3, we show that there is an overall correlation between stellar mass and UV luminosity down to $M_{star} \approx 10^{8.7} M_\odot$ in that more UV-luminous galaxies are on average more massive. However, it is equally evident that there is a substantial scatter to this correlation. In particular, there exist galaxies that are quite massive but faint in the UV. Roughly 36% and 20% of the galaxies in the B_{435}- and V_{606}-band dropout sample that are more massive than $M_{star} = 10^{9.5} M_\odot$ are fainter than $M_{1700} = -20$. The implication is that there is a non-negligible population of galaxies whose SFRs were considerably higher in the past but are currently fading in the UV. It is possible that some of them may be faint simply due to more dust reddening, but as we will show later (Section 6.1), dust alone is not enough to explain the observed UV luminosity distribution of these low-mass galaxies. Interestingly, the opposite is not true in that the region of low stellar mass and high UV luminosity is largely devoid of galaxies. This suggests that the majority of the UV-luminous galaxies have been forming stars for quite some time (at least a few hundred million years). This interpretation is consistent with a recent study of the most UV-luminous galaxies at the same redshift by Lee et al. (2011, also shown in Figure 3), who concluded that they undergo smooth SFHs (also see Shim et al. 2011).

4. UNDERSTANDING MEASUREMENT UNCERTAINTIES/BIASES

SED fitting is a procedure that attempts to fit a number of galaxy parameters based on multiple data points taken at different depths and resolutions. Thus, its outcome may depend strongly on the range of galaxy parameters (IMF, SFHs, age) and the quality of the photometric data. Therefore, it is worth paying close attention to how well each galaxy parameter can be measured given the set of observations. The most relevant issues to our analyses are: How robust are the derived galaxy properties (e.g., extinction, stellar mass) when the SFH and redshift are unknown? What are their uncertainties, and how do they scale with the quality of photometry? How well can we

measure the galaxy light within the same effective aperture from the multi-resolution data set, using a photometric technique of choice (e.g., SExtractor, TFIT)? Creating a mock data set with realistic properties (noise, seeing, galaxy colors, etc.) is key to properly accounting for all factors that may contribute to the measurement errors and systematics. To address these questions, we carried out a large set of simulations. The procedure is designed to simulate galaxies as realistically as possible every step of the way by mimicking the noise properties, size distribution, photometric extraction, and derivation of physical parameters via SED fitting of the real high-redshift galaxies.

4.1. Galaxy Simulations on Mixed-resolution Data Set

The simulation procedure can be summarized as follows: (1) create galaxy model SEDs spanning a wide range of physical parameters (age, reddening, SFHs, redshift, and stellar mass); (2) compute input photometry of these SEDs in the observed passbands; (3) insert mock galaxies into the images in the different photometric bands; (4) measure photometry using TFIT software; and (5) carry out SED fitting to "measure" output physical parameters. By design, the output of our simulations is identical to that of the real galaxies except that input parameters are known for simulations. By comparing the input and output parameters, we can quantitatively assess the robustness of the measurements at various levels and use that information to correct for the systematics. As for the range of physical parameters used to create mock galaxies, we use the same ranges that we explore in the SED fitting procedure for real galaxies. By doing so, we ensure that we can quantify all the measurement errors and biases within the assumptions that went into deriving the physical parameters of real galaxies. Any non-standard physical characteristics of real galaxies, such as exotic IMFs or sub- or super-solar metallicities, that may have contributed to our measurements, however, cannot be accounted for in our simulations and thus are beyond the scope of this work. The descriptions of steps 1 and 2 are given in Section 3.1. The details of inserting mock galaxies in the real data (step 3) are found in Ferguson et al. (2004).

The main improvement of our simulations over the previous version described in Ferguson et al. (2004) is that we extend the technique to create multi-resolution data sets. This revision is straightforward as all of the data have pixel scales that are integer multiples of one another ($0\farcs03$ pixel^{-1} for the HST/ACS data, $0\farcs5$ pixel^{-1} or $0\farcs30$ pixel^{-1} for the ground-based data, and $0\farcs60$ pixel^{-1} for the $Spitzer$/IRAC data). All images are initially created in the highest resolution, convolved with the appropriate PSFs, and then block-summed to the desired pixel scale before being added to the real science images (step 3). Only 30 galaxies are inserted at a time over a 4 arcmin^2 area (4096×4096 in the pixel resolution of the ACS bands) to avoid unrealistic crowding. They are placed at random locations within the images with random orientation angle. The light profiles of galaxies are chosen to be either of de Vaucouleurs profile or of exponential disk. The size distribution is set to follow a lognormal distribution as observed at these redshifts (Ferguson et al. 2004).

For each run, once the mock images are created, we repeat the identical procedures for measuring their photometry (step 4) and SED parameters (step 5) as for real galaxies. This means that we run SExtractor on the ACS images with the same setup for source detection and photometry and then use the ACS z_{850} band as templates to run TFIT on other passbands. The U band (below the Lyman limit for $z > 3$ galaxies) was not included
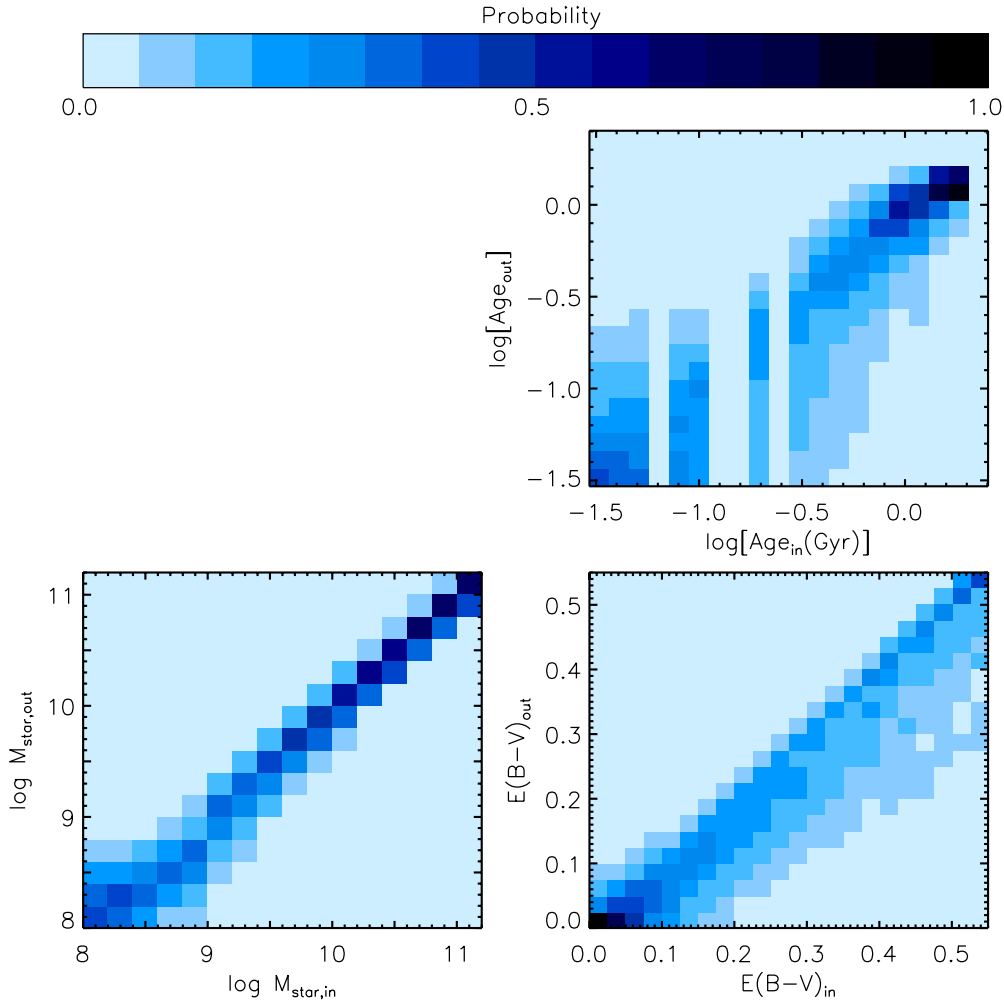
Figure 4. Probability distribution function for the stellar mass (lower left), internal reddening (lower right), and age parameters (top right) recovered from our SED fitting procedure as a function of input parameters. The color bar (top) indicates the probability value for each (input, output) cell. The discreteness of input ages reflects non-equal steps of ages adopted for the simulated galaxies. These results are based on 34,501 B_{435}-band dropouts planted in the GOODS-S and GOODS-N field according to the simulations procedure we described in the text (Section 4). Based on these results, we determined that stellar mass is the most robust quantity that can be recovered, at least based on the broadband photometry via SED fitting. The internal reddening can be recovered to a somewhat lesser degree, indicated by a larger error. The population age is the least robust parameter, which is likely driven by the current sensitivity limit of the near-infrared data sampling the Balmer/4000 Å region (H and K_S bands for the B_{435}-band dropouts).

(A color version of this figure is available in the online journal.)

in the simulations, as mock galaxies do not suffer from low-redshift interlopers, for which the U band is mainly used to check non-detection for real galaxies.

4.2. How Well Can We Measure Galaxy Parameters?

Using the procedures described above, we have simulated about 200,000 galaxies, of which ≈35,000 galaxies are selected as B_{435}-band dropouts and ≈26,000 as V_{606}-band dropouts using the color selection criteria given in Equations (1) and (2). These galaxies span a wide range of UV luminosities, SFHs, age, extinction, and stellar mass. In Figure 4, we show the input and recovered value of population age, dust reddening (parameterized as $E(B-V)$), and stellar mass, for simulated galaxies. Stellar mass is the most robust quantity that we can recover based on the available broadband photometry at the current depth, while internal dust reddening can be recovered to a somewhat lesser degree (bottom right). As for population ages, our SED fitting fares very poorly with a large scatter (top right) unless the age of input galaxies is much older (>500 Myr) than the values derived by SED fitting for the majority of the

observed LBGs. The poor recovery of population age is likely attributed to the current (shallow) depth of the near-infrared data from ground, which samples near the Balmer/4000 Å break (H and K_S bands for the B_{435}-band dropouts at $z \sim 4$). These results are in qualitative agreement with Lee et al. (2009b), who, based on the SED fitting of mock galaxies created by a semianalytic model, reported that stellar mass can be recovered most reliably (within 0.1 dex of the true values) even when the assumed SFH is clearly different from the intrinsic one.

In Figure 5, we show typical SEDs of real (left) and simulated (right) galaxies at different stellar masses. For the simulated galaxies, we also indicate the input photometry as open circles, illustrating the increasing degree of photometric scatter and errors toward the low-mass end. On right panels, we show the probability distribution of the *recovered* stellar mass at various input mass values. As expected, it is evident that the accuracy of stellar mass estimation decreases rapidly toward low masses as a result of larger photometric errors in relevant passbands (in this case, mainly driven by the IRAC [3.6 μm] and [4.5 μm]). Increasingly large spread of recovered mass with respect to
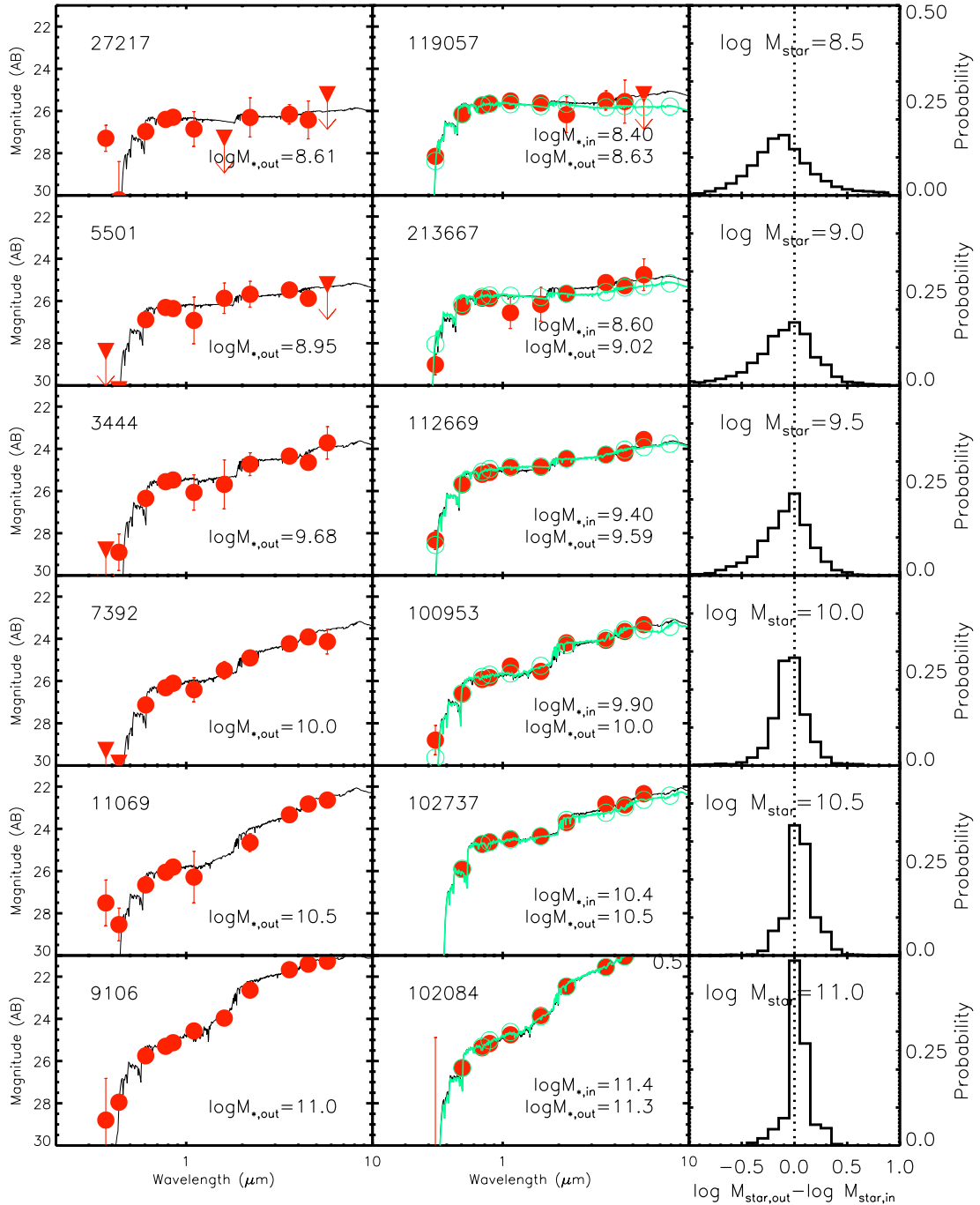
Figure 5. We show the spectral energy distributions of typical galaxies at different stellar mass bins for both *real* (left panels) and *simulated* galaxies (middle panels). Filled circles represent flux in each photometric band with TFIT errors, while solid lines indicate the best-fit population models from which the stellar mass is determined. All stellar masses are given in units of solar mass. In the histograms on right panels, we show the distribution of *recovered stellar mass* with respect to the *input* mass. The typical accuracy in the determination of stellar mass can be assessed based on the mean and dispersion of the output mass distribution. We conservatively set the mass limit to $\log[M_{star}/M_{\odot}] = 8.7$ and 9.0 at $z \sim 4$ and 5, respectively. Below these limits, output masses are increasingly uncertain as evidenced by an offset with large scatter (top right panel).

(A color version of this figure is available in the online journal.)

the true value further supports the statistical approach we take in constructing galaxy counts in mass bins. In any SMF that rises relatively steeply toward the low-mass end, such scatter will result in artificial steepening of the observed slope as more low-mass galaxies will scatter into higher mass bins than the opposite.

In our analyses throughout this paper, we use the mass probability distribution (right panels of Figure 5) for two purposes:

first to assess and thereby determine how much we can push the limit of the available data without introducing systematic biases into our measurements, and second to statistically correct the observed number count for the "mass spread" due to photometric errors. As for the former, we conclude that stellar mass can be measured reliably down to $\log[M_{star}/M_{\odot}] = 8.7$ and 9 at $z \sim 4$ and 5, respectively. As evident in the right panels of Figure 5, the probability distribution becomes skewed toward

the masses below the limit. This is mainly due to the fact that sources become very faint (generally, in all bands). This leads to two independent consequences that affect the SED fitting procedure. First, SExtractor tends to underestimate the isophotal area from the detection image, which is later used to define the "effective aperture" in other bands to measure flux using TFIT. More quantitative discussions on this issue will be discussed in an upcoming paper (A. Galametz et al. 2012, in preparation). Second, because the same sources are (usually) intrinsically fainter in the *Spitzer* bands, the SED fitting procedure is often driven by the flux upper limits. Because there is no meaningful constraint on the rest-frame optical wavelengths, the fitting returns a wide distribution of stellar masses as acceptable (see discussions in Section 3.2).

5. THE UV LUMINOSITY FUNCTION AND STELLAR MASS FUNCTION OF LBGs

5.1. The UV Luminosity Function

We measure the UVLF of LBGs in our samples using a methodology similar to that presented by Bouwens et al. (2007). The procedure consists of (1) constructing the number counts of the real galaxies in apparent magnitude bins, (2) quantifying the photometric and selection completeness of the observed galaxies by using simulations (Section 4), and (3) determining the range of UVLF parameters that reproduces the observed number counts. In quantifying the completeness, we treat internal dust reddening as an explicit variable.

We define the selection efficiency, $p_S(m', z, \epsilon)$, as the probability that a galaxy at redshift z of the (true) apparent magnitude m' with the reddening parameter $E(B - V) = \epsilon$ is detected and selected as a dropout in our selection. In addition to the selection efficiency, we further quantify the photometric error that affects the number counts of the dropouts as the probability that a galaxy (m', z, ϵ) is observed at the apparent magnitude m, hereafter referred to as $p_M(m|m', z, \epsilon)$. To relate the intrinsic UVLF to the observed number density, we define the effective volume as

$$
\begin{aligned}
V_{m,k} &= \int_z \int_{m'} W(M(m', z, \epsilon) - M_k) \mathcal{P}(m|m', z, \epsilon) \frac{dV}{dz} dm' dz \\
&\approx \sum_{m'} \sum_z W(M(m', z, \epsilon) - M_k) \mathcal{P}(m|m', z, \epsilon) \frac{dV}{dz}(z) \\
&\quad \times \Delta m' \Delta z,
\end{aligned}
\tag{1}
$$

where W is a window function, $\mathcal{P}(m|m', z, \epsilon) \equiv p_S(m, z, \epsilon) \times p_M(m|m', z, \epsilon)$, dV/dz is the cosmic volume covered by the survey area, m is the *observed* apparent magnitude, m' is the *true* apparent magnitude, and M_k is the absolute UV magnitude at 1700 Å. The effective volume averaged over the intrinsic distribution of reddening ϵ is

$$
\langle V_{m,k} \rangle_\epsilon = \frac{\Sigma_n V_{m,k} \mathcal{D}(\epsilon_n, M_k) \Delta\epsilon}{\Sigma_n \mathcal{D}(\epsilon_n, M_k) \Delta\epsilon},
\tag{2}
$$

where $\mathcal{D}$ characterizes the distribution of reddening that can be luminosity dependent (M_k). The observed number count for a given luminosity function can be computed as

$$
N_m = \sum_k \phi_L(M_k) \langle V_{m,k} \rangle_\epsilon,
\tag{3}
$$

where $\phi_L(M_k)$ is the UVLF evaluated at the absolute magnitude bin $M_{1700} = M_k$, where the UVLF is parameterized as a Schechter function:

$$
\begin{aligned}
\phi_L(M_k) &= (0.4 \ln 10) \phi^* 10^{0.4(M^*_{\rm UV} - M_k)(1+\alpha)} \\
&\quad \times \exp[-10^{0.4(M^*_{\rm UV} - M_k)}].
\end{aligned}
\tag{4}
$$

As for the reddening distribution $\mathcal{D}$, we assumed a normal distribution with a mean of $\epsilon = 0.14$ if $M_k \leqslant -21.06(L^*)$, $\epsilon = 0$ if $M_k \geqslant -18.56(0.1L^*)$, and linearly declining in between. The 1σ scatter of the distribution remains $\sigma_\epsilon = 0.14$ at all luminosities. Our prescription is essentially identical to that used by Bouwens et al. (2007).

In Figure 6, we show the observed number counts together with those published in Stark et al. (2009) based on the GOODS v1.0 catalog. Both measurements agree well except in a few faintest magnitude bins, where our deeper catalog should be more complete. Thick dashed lines indicate the expected number counts computed according to Equation (3) when the best-fit UVLF parameters given by Bouwens et al. (2007) are assumed. While we are unable to independently determine the UVLF parameters based on the GOODS data alone, the good agreement between the two suggests that the faint-end slope is steep in the range of $-(1.7–1.8)$. To illustrate the fact that our data are indeed more consistent with a steep slope (< -1.7), we also show in Figure 6 as a set of dotted lines the model number counts with the slope $-(1.6, 1.4, 1.2)$ when the characteristic luminosity $M^*_{\rm UV}$ is fixed. All functions are normalized at $z_{850} = 24.97$ and 25.71 at $z \sim 4$ and 5, roughly corresponding to the respective characteristic luminosity at the median redshift of each sample (shown as filled arrows).

In comparison with the results by Bouwens et al. (2007), we find a slight discrepancy at the bright end, where their estimate is about a factor of two higher than what we and Stark et al. (2009) found. This is likely due to the fact that we discarded some number of bright sources after inspecting their SEDs (as well as MIPS detection; see Section 2.3). Many of the bright sources have UV colors consistent with high-redshift LBGs (and thus are included in their sample), but longer-wavelength data suggest that they are unlikely to be at high redshift. Clearly, both our estimates and those of Bouwens et al. (2007) suffer from small-number statistics at the bright end, and the discrepancy is driven by a mere 10–20 sources and their photometric redshifts; more complete spectroscopy of these sources and larger-area surveys will be needed to better quantify the UVLF at the bright end.

5.2. The Stellar Mass Function of SFGs

Using robust measurements (and error estimates) of SED properties and UV luminosities for individual galaxies, we next determine the SMF of LBGs at $z \sim 4$ and 5. We caution readers that these measurements do not in any way represent the *total* SMF of *all* galaxies, but rather provide information on how much of the cosmic stellar mass budget is contributed by actively star-forming and relatively UV-bright galaxies. Furthermore, as we will see later, by comparing the shape of the SMF with that of the UVLF measured for the same galaxies, one can obtain crucial clues to how (UV–visible) SF has proceeded with cosmic time.

We compute the number density of LBGs at each stellar mass bin as follows. First, the number density of galaxies at a fixed apparent magnitude m_j and stellar mass $M_{*,m}$ is obtained by correcting the observed number $N_{\rm obs}(m_j, M_{*,m})$ for the selection incompleteness by dividing it by the effective volume $V_{\rm eff}(m_j)$.
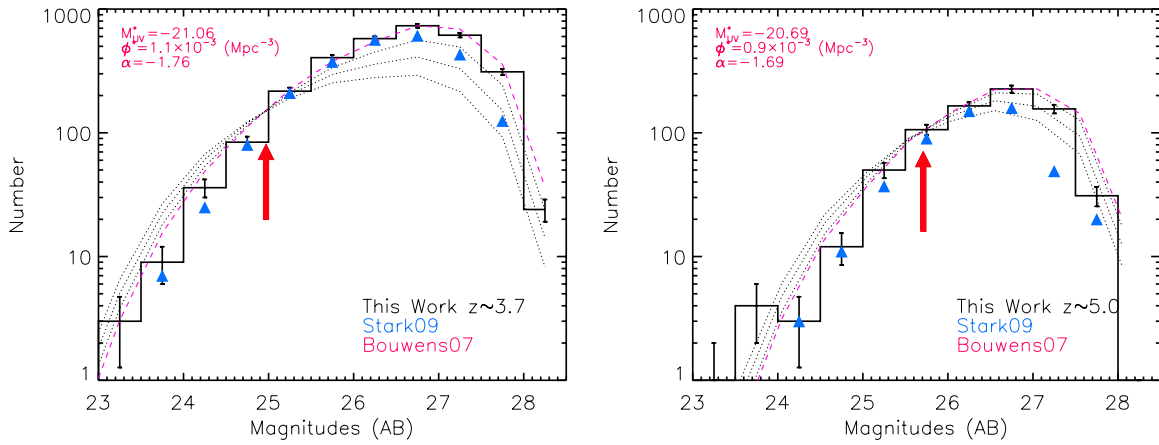
Figure 6. Total *observed* number of galaxies in UV magnitude bins compared to those in the literature at $z \sim 4$ (left) and 5 (right). Our number counts are in excellent agreement with those by Stark et al. (2009) except the very faint end. This discrepancy is likely explained by the fact that the data set (v2.0 release) used in this work has a slightly higher completeness than theirs (v1.0 release). Assuming the best-fit parameters of the Bouwens et al. (2007) determination of the UV luminosity function, we compute the expected galaxy number counts using Equation (3). We find good agreement between the observed and expected number counts, which confirms that the faint-end slope of the UV luminosity function is indeed steep. Also shown as dotted lines in each panel are the expected number counts when a set of shallower slopes ($\alpha = -1.6, -1.4, -1.2$ from top to bottom) are assumed while the characteristic luminosity M^*_{1700} is fixed to the same value as shown in the top left corner. All models are normalized at $z_{850} = 24.97$ and 25.71 at $z \sim 4$ and 5, roughly corresponding to the respective characteristic luminosity at the median redshift of each sample (shown as filled arrows).

(A color version of this figure is available in the online journal.)

Using Equations (1) and (2), the total effective volume at a fixed apparent magnitude m_j is

$$V_{\text{eff}}(m_j) = \Sigma_k \langle V_{j,k} \rangle_\epsilon \quad (5)$$

$$\phi_M(m_j, M_{*,m}) = \frac{N_{\text{obs}}(m_j, M_{*,m})}{V_{\text{eff}}(m_j)}. \quad (6)$$

The *total* number density of LBGs at a stellar mass bin is obtained by adding the contributions from all the apparent magnitude bins:

$$\phi_M(M_{*,m}) = \Sigma_j \phi_M(m_j, M_{*,m}). \quad (7)$$

To estimate realistic errors at each mass bin, we account for both random Poisson errors (based on the observed number at each bin) and sample variance (based on the difference between the observed number at two fields). The Poisson error at each bin is computed as $\Delta\phi_M(m_j, M_{*,k}) = [\sqrt{N_{\text{obs}}(m_j, M_{*,k}) + 0.75} + 1]/V_{\text{eff}}(m_j)$ using the approximation given by Gehrels (1986). The total Poisson error at each mass bin is the quadratic sum of all magnitude bins:

$$\Delta\phi_M^{\mathcal{P}}(M_{*,k}) = \sqrt{\Sigma_j [\Delta\phi_M(m_j, M_{*,k})]^2}. \quad (8)$$

The errors associated with sample variance are estimated by accounting for the difference in the observed number between the two fields as

$$\Delta\phi_M^{SV}(M_{*,k}) = \left| \phi_M^S(M_{*,m}) - \phi_M^N(M_{*,m}) \right| / 2. \quad (9)$$

The total error budget is obtained as a quadratic sum of the random error and sample variance at each mass bin and is given as $\Delta\phi_M = \sqrt{[\Delta\phi_M^{\mathcal{P}}]^2 + [\Delta\phi_M^{SV}]^2}$.

Using the prescription described above, we computed the SMF of LBGs at $z \sim 4$ and 5. The results are shown in Figure 7 together with other measurements in the literature (Marchesini et al. 2009, 2010; Stark et al. 2009; Caputi et al. 2011; González et al. 2011; Santini et al. 2012). We also indicate

the number density measured in the GOODS-S and GOODS-N fields as small open symbols at each mass bin (circles and triangles). Our measurements at $z \sim 4$ are in broad agreement with those of Marchesini et al. (2009, 2010, blue open circles using set 7 in their Table 4), who measured the *total* SMF of *all* galaxies at $3 < z < 4$. The fact that, at the massive end, their estimated number density is a factor of 2–3 higher than ours is consistent with their observation that only $\approx$30% of their sample at the very massive end would be selected as LBGs, while the rest is either passive or heavily obscured by dust. The measurements of Santini et al. (2012, green downward triangles) at $3.5 < z < 4.5$, mostly consistent with those of Marchesini et al. (2009), are also in line with our data in their intermediate-mass bins. In their least massive bin $\approx$10^{10} $M_\odot$, their point is higher by a factor of two, which may be due to cosmic variance as the bin includes only a handful of sources. We also compare our results with those of Caputi et al. (2011, cyan open and filled diamonds for $z \sim 3.2$ and $z \sim 3.9$, filled diamonds for $z \sim 5$) after we corrected their masses by 0.16 dex. The correction is made to convert their masses (assuming Salpeter IMF) to ours (Chabrier IMF). While at $\log[M_{\text{star}}/M_\odot] > 10.5$ their data are consistent with our estimates, they are systematically lower by more than a factor of two than our estimates and those of Marchesini et al. (2009, 2010) and Santini et al. (2012) at lower masses at both redshift bins. While the origin of this discrepancy is unclear, possible causes include that (1) some of real high-redshift sources may have been missed by their photometric redshift selection or (2) their incompleteness correction may have been underestimated. Relatively large discrepancies present among similar studies based on the rest-frame optical selection (e.g., Fontana et al. 2006; Pérez-González et al. 2008; Marchesini et al. 2009, 2010; Caputi et al. 2011; Santini et al. 2012) may also be explained at least in part by non-negligible cosmic variance. Even the lowest-mass galaxies selected by these studies are among the most massive galaxies at high redshift, which are observed to be strongly clustered (e.g., Lin et al. 2011).

We also compare our results with those that measured the SMF for star-forming galaxies. The results of Stark et al.
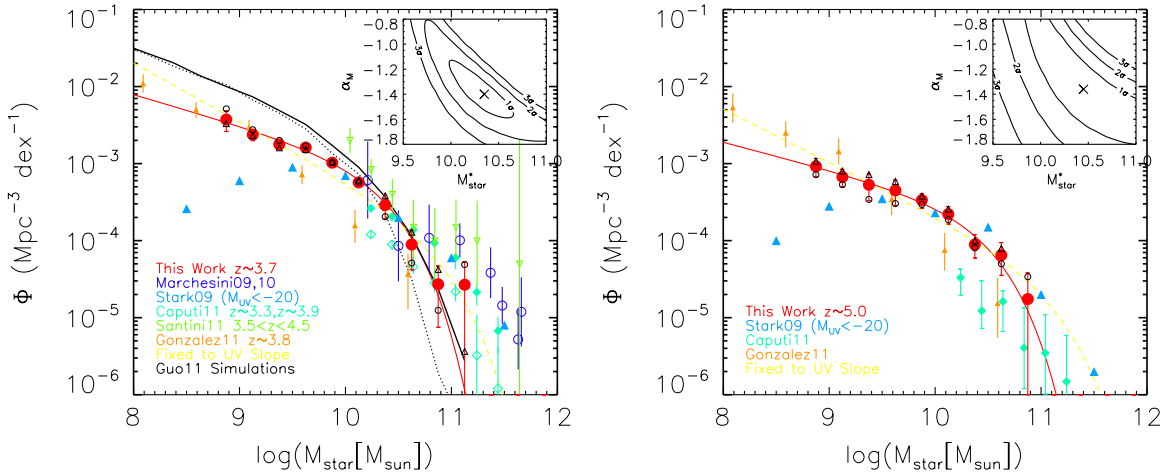
Figure 7. Stellar mass functions of LBGs at $z \sim 4$ and 5 are shown. The number density of galaxies at each stellar mass bin is obtained by first correcting the observed number for the selection incompleteness (for the UV selection), then summing up the contributions from all magnitude bins. The error bars fully account for random Poisson errors, as well as those arising from sample variance. Filled circles represent our best estimates combining the statistics of both GOODS fields, while small open symbols (circle and triangle) at each mass bin indicate the number density computed in each field. Also shown are similar measures in the literature (Stark et al. 2009; Marchesini et al. 2009, 2010; Caputi et al. 2011; González et al. 2011; Santini et al. 2012, in blue filled triangles, blue open circles, cyan diamonds, orange upward triangles, and green downward triangles, respectively). For $z \sim 4$, we show the model predictions from Guo et al. (2011, solid and dotted yellow lines) based on Millennium and Millennium II simulations. The best-fit Schechter parameters (shown in solid red lines) indicate that the low-mass end slope of the stellar mass function is ≈ -1.4 and thus is considerably shallower than the faint-end slope of the UV luminosity function derived for the same galaxies, $\approx -(1.7–1.8)$. Top right insets illustrate the confidence intervals at the 68%, 95%, and 99% level.

(2009, blue triangles) are in god agreement with ours when we confine our analyses only to galaxies at $M_{1700} < -20$ as they did. A slight discrepancy at the massive end is likely due to our stringent selection, which resulted in removal of a few galaxies at the massive end, and possibly due to slightly different criteria applied to remove likely low-redshift interlopers. These "cleaning" processes affect the massive end the most, and clearly, the two measurements are comparable in lower-mass bins at $\log[M_{star}/M_\odot] < 10.5$. Finally, González et al. (2011, orange upward triangles) found the low-mass end slope of the SMF to be $\alpha_M \approx -1.4$, in excellent agreement with our analyses. At $z \sim 3.7$, their measurements are consistent with ours in their low-mass bins ($10^8 M_\odot < M_{star} < 10^{9.5} M_\odot$) even though they went further down to below $10^8 M_\odot$, whereas we chose not to go below $10^{8.7} M_\odot$. At $z \sim 5$, their overall normalization is about 50% higher than ours even though the estimated low-mass slope is still $\alpha = -1.39$, very close to ours, $\alpha_M = -1.36$ (see discussions later). At the massive end, however, their data points are lower by more than a factor of two compared to our own results and those of Stark et al. (2009). Because their sample is based on the WFC3 Early Release Science (Windhorst et al. 2011) data, which cover about one-third of the GOODS-S field (i.e., $\approx 15\%$ of the total area covered by our study and that of Stark et al. 2009, who used both GOODS-N and GOODS-S), it is possible that their massive bins suffer more from cosmic variance than others.

Based on our measurements, we determined the best-fit parameters when a Schechter function is assumed for the SMF:

$$\phi_M(M_{star}) = (\ln 10)\phi_M^* [10^{(M_{star} - M_{star}^*)(1+\alpha_{SMF})}$$
$$\times \exp[-10^{(M_{star} - M_{star}^*)}]. \qquad (10)$$

Note that all parameters have subscripts "M" to distinguish from similar Schechter parameters for the UVLF discussed previously. To properly account for the measurement errors in mass, we use the probability distribution discussed in Section 4.2 (also shown in Figures 4 and 5) to convert the *intrinsic* SMF to

the *observed* one:

$$\phi_M(M_{*,m}) = \Sigma_k \phi_M(M_{*,k}) p(M_{*,m}|M_{*,k}), \qquad (11)$$

where $M_{*,k}$ is the true stellar mass of a galaxy and $M_{*,m}$ is the measured one. We use the chi-square minimization method to search for the best parameter set (α_{SMF}, M_{star}^*). For a given parameter set (α_{SMF}, M_{star}^*), we determine the normalization parameter ϕ_M^* such that the total number density of galaxies at $M_{star} \geqslant M_{min}$ equals the observed number density. The minimum stellar mass M_{min} is set to $10^{8.7} M_\odot$ and $10^9 M_\odot$ at $z \sim 4$ and 5, respectively, for the reasons discussed in Section 4.2. However, our results are insensitive to the minimum mass as long as it is set within the mass range over which the determination of stellar mass is relatively robust. On the top right inset in Figure 7, we show the confidence levels for the best-fit Schechter parameters, namely, the characteristic stellar mass M_{star}^* and low-mass end slope α_{SMF}.

At $z \sim 4$, the best-fit value for the low-mass end slope α_{SMF} is ≈ -1.4. Also at $z \sim 5$, the best-fit slope is -1.4 even though a large degeneracy between M_{star}^* and α_M (due to lack of constraints on the massive end) makes a relatively steep slope also acceptable. The slope is much shallower than the faint-end slope of the UVLF for the same galaxies but is similar to that found from stellar-mass-selected studies (Fontana et al. 2006; Marchesini et al. 2009) and a recent study of LBGs based on a smaller sample by González et al. (2011). To visually illustrate this difference, we show in Figure 7 (dashed lines) the best-fit Schechter function when the slope α_{SMF} is fixed to the faint-end slope of the UVLF from Bouwens et al. (2007), the values of which are indicated in Figure 7. Clearly, a very steep low-mass end slope $\alpha_{SMF} = -(1.7–1.8)$ does not provide a good fit to the observed SMF even for the $z \sim 5$ case. The shallow slope on the low-mass end seems to be a robust feature even considering somewhat large uncertainties for the normalization parameter as evidenced by a relatively large cosmic variance particularly at $z \sim 5$. In Table 2, we summarize the best-fit Schechter parameters for both samples. For the

Table 2
The Schechter Parameters for the Stellar Mass Function of LBGs at $z \sim 4$ and ~ 5

SFH	B_{435}-dropouts ($z \sim 4$)			V_{606}-dropouts ($z \sim 5$)		
	$\log M_{\mathrm{star}}^{*}$	ϕ_{M}^{*} (10^{-4} Mpc^{-3})	α_{SMF}	$\log M_{\mathrm{star}}^{*}$	ϕ_{M}^{*} (10^{-4} Mpc^{-3})	α_{SMF}
All SFHs[a]	$10.35^{+0.30}_{-0.35}$	$3.9^{+6.3}_{-2.4}$	$-1.40^{+0.28}_{-0.20}$	$10.45^{+1.05}_{-0.65}$	$1.1^{+4.5}_{-1.0}$	$-1.36^{+0.76}_{-0.48}$
All SFHs[b]	...	...	...	$10.35_{-0.55}$	$1.4^{+4.1}_{-0.6}$	$-1.28^{+0.68}_{-0.28}$
CSF[a]	$10.35^{+0.25}_{-0.30}$	$3.7^{+5.6}_{-2.4}$	$-1.46^{+0.26}_{-0.22}$	$10.50^{+1.00}_{-0.55}$	$0.9^{+3.1}_{-0.9}$	$-1.42^{+0.52}_{-0.44}$
CSF[b]	...	...	...	$10.35_{-0.40}$	$1.4^{+2.6}_{-0.5}$	$-1.32^{+0.42}_{-0.24}$
300 Myr[a]	$10.35^{+0.50}_{-0.40}$	$3.5^{+7.7}_{-2.8}$	$-1.40^{+0.38}_{-0.28}$	$11.20^{+0.30}_{-1.10}$	$0.2^{+2.3}_{-0.1}$	$-1.60^{+0.50}_{-0.26}$
300 Myr[b]	...	...	...	$10.35_{-0.25}$	$1.2^{+1.2}_{-0.4}$	$-1.34^{+0.24}_{-0.20}$
100 Myr[a]	$10.15^{+0.55}_{-0.35}$	$4.0^{+6.8}_{-3.2}$	$-1.40^{+0.34}_{-0.28}$	$10.50^{+1.00}_{-0.55}$	$0.6^{+2.4}_{-0.5}$	$-1.50^{+0.50}_{-0.46}$
100 Myr[b]	...	...	...	$10.35_{-0.40}$	$1.0^{+2.0}_{-0.5}$	$-1.40^{+0.40}_{-0.32}$

Notes.
[a] The characteristic mass $\log M_{\mathrm{star}}^{*}$ is allowed to vary during the fit in the mass range $9 \leqslant \log M_{\mathrm{star}}^{*} \leqslant 11.5$.
[b] The characteristic mass $\log M_{\mathrm{star}}^{*}$ is allowed to vary only below $M_{\mathrm{star}}^{*} = 10.35$, the best-fit value for the B_{435}-band dropouts.

V_{606}-band dropouts, we also provide the best-fit parameters when the characteristic mass M_{star}^{*} is forced not to exceed that at $z \sim 4$. Such a treatment generally helps constrain the low-mass end slope better by reducing the large degeneracy between the two parameters. While it remains to be seen if it is indeed the case, the brightening of the characteristic luminosity L_{UV}^{*} observed at high redshift (Bouwens et al. 2007) likely supports such a scenario unless the mass-to-light ratios change significantly with redshift to reverse the effect. We also note that removing the least robust mass bin also makes the slope even shallower (albeit leaving larger overall uncertainties). Finally, we compare our measurements against recent predictions by Guo et al. (2011) using semi-analytic models implemented on the Millennium and Millennium II simulations at $z \sim 4$ (shown as black lines on the left panel of Figure 7). The dotted line indicates their predictions for the intrinsic SMF, while the solid line represents the values accounting for measurement errors (by convolving the intrinsic SMF assuming 0.3 dex scatter in stellar mass). On the massive end, our estimates are in excellent agreement with theirs. However, at the mass regime below the characteristic mass $\sim 10^{10} M_{\odot}$, their model clearly overpredicts the abundance by a factor of 2–3 compared to all the existing studies. Furthermore, the predicted slope for the low-mass end of the SMF is more in line with that of the UVLF shown by a dashed line, much steeper than our measurements.

Finally, we note that not all studies find the low-mass end slopes to be as shallow as our results and those of González et al. (2011). For example, Kajisawa et al. (2009) studied the galaxy SMF out to $z \sim 3$ and reported that the low-mass end slopes become steeper with redshift, from $\alpha_{M} = -(1.2\text{–}1.3)$ at $z \sim 0.7$ to $\alpha_{M} = -1.6$ at $z \sim 3$. Santini et al. (2012) reported a similar steepening of the slope with redshift, from $\alpha_{M} = -1.44 \pm 0.03$ at $z \sim 0.8$ to -1.86 ± 0.16 at $z \sim 3$. For galaxies at $z \sim 3.2$, 3.9, and 4.7, Caputi et al. (2011) found the low-mass end slope to be ≈ -1.8 at all probed redshift bins. With the current observational constraints, these results are not necessarily contradictory to shallower slopes reported for the SMFs of star-forming galaxies. Most studies based on mass-selected samples agree that the characteristic mass M_{star}^{*} lies well above $10^{11} M_{\odot}$ within a range of population synthesis models and assumptions (see all the references discussed above). On the other hand, the number density of star-forming galaxies at around $10^{11} M_{\odot}$ is observed to be very low, and consequently, the SMF measured only for star-forming galaxies tends to have

a much lower characteristic mass (10^{10}–$10^{10.6} M_{\odot}$) in most studies. Such low characteristic mass is found independent of whether the masses are fit for individual galaxies (such as this work and González et al. 2011) or an average mass-to-light ratio was assumed (e.g., Reddy & Steidel 2009; McLure et al. 2009). Hypothetically, if the fraction of star-forming (therefore, UV–visible) galaxies increases toward lower masses, then it is conceivable that the *true* mass function will have a high characteristic mass above $10^{11} M_{\odot}$ but will converge toward that of SMF at low masses where galaxies are dominated by star-forming galaxies. To test this possibility, we tried to fit our data to a Schechter function after forcing the characteristic mass to be greater than $10^{11} M_{\odot}$. When we only include our data points down to $\approx 3 \times 10^{9} M_{\odot}$, which is the lowest mass bin currently probed by deep surveys (Kajisawa et al. 2009), we find that a slope as steep as $-(1.8\text{–}1.9)$ can provide a good fit. However, once we include the lower-mass bins, such a steep slope is no longer acceptable. In other words, the steep slope found by mass-selected studies may be explained by a combination of the M_{star}^{*}–α_{M} degeneracy (see inset of Figure 7 for the elongated shape of the contours) and the limited mass range currently probed by mass-selected samples, *if* the low-mass end is primarily dominated by star-forming galaxies. Deeper data are critically needed to extend the mass range and the fraction of star-forming galaxies at different masses to help better understand the observed discrepancies between different studies.

6. DISCUSSIONS

We have measured the SED properties of LBGs at $z \sim 4$ and 5, two of the largest photometric samples ever assembled, consisting of 3088 and 987 galaxies, respectively. The use of TFIT, which can very robustly measure galaxy light within a fixed physical aperture over data with a wide range of PSF sizes, and a large set of galaxy simulations to quantify measurement errors and systematics of TFIT photometry at different flux levels has allowed us to make the most robust measurements to date of galaxy properties, such as stellar mass, age, and UV luminosity, for a large sample and carry out a large statistical study of the high-redshift star-forming galaxy population as a whole. Based on these measurements, we concluded that (1) there is a significant *intrinsic* scatter to the correlation between the observed UV luminosity and stellar mass; (2) the number counts of galaxies in the bins of UV luminosity strongly

support the previous studies (Bouwens et al. 2007; Reddy & Steidel 2009) that the UVLF at these redshifts increases steeply, $\alpha_{\rm UVLF} \approx -(1.7-1.8)$, toward the faint end; and finally, (3) the number density of the same galaxies in the bins of stellar mass increases rather mildly, $\alpha_{\rm SMF} \approx -(1.3-1.4)$, compared to the UVLF of the same galaxies. In what follows, we discuss these findings in further detail and their possible implications on how these UV-selected galaxies assemble their masses over cosmic time.

6.1. The $M_{1700}-M_{\rm star}$ Scaling Relation

The locations of galaxies on the SFR–$M_{\rm star}$ plane indicate typical SFHs of the galaxy population in question. If most galaxies have been assembling their mass smoothly over a long period of time, one would expect them to form a tight "main sequence" on the SFR–$M_{\rm star}$ plane as stellar mass is, to the zeroth order, the time integral of the past SFH. Indeed, a relatively tight main sequence has been observed for star-forming galaxies in the local universe (Noeske et al. 2007) and for massive star-forming galaxies at $z = 1$ and 2 (Elbaz et al. 2007, 2011; Daddi et al. 2007b; Rodighiero et al. 2011). The observed scatter around the main sequence is 0.25 dex in all cases. For a galaxy population at $z > 2$, however, the trend is still unclear. While a similar (positive) correlation between SFRs and stellar mass exists (Pannella et al. 2009; Lee et al. 2011, from stacking), many observations reported much larger scatter ($\sigma \gtrsim 0.46$ dex; Reddy et al. 2006; Mannucci et al. 2009; Magdis et al. 2010).

As shown in Figure 3, our results suggest that there is a significant scatter in the correlation between UV luminosity (M_{1700}) and stellar mass. It is evident at both redshifts that at a fixed stellar mass, there is a non-negligible fraction of galaxies that extend down to nearly 2 mag fainter than the "main sequence." It is difficult to use these observations to estimate the true extent of this faint tail because our samples are UV selected and thus are not complete in mass. It is worth noting that the UV selection misses a large number of very faint galaxies ($M_{1700} > -19.5$) as the completeness of the data decreases at the faint end. If we correct for the missing galaxies, the faint-end tail is even more significant than that observed here. Furthermore, as the UV selection systematically misses red galaxies with heavily obscured SF, inclusion of such a population would further increase the scatter.

The main limitation of interpreting the observed $M_{1700}-M_{\rm star}$ scaling relation is the fact that UV light from these galaxies is systematically dimmed by internal dust (Calzetti et al. 1997, 2000). At a fixed UV luminosity, the distribution of UV extinction can vary significantly (e.g., Bouwens et al. 2009; Reddy et al. 2010). For example, the 1σ scatter in the color excess $E(B - V)$ of 0.1 (or extinction at 1700 Å, A_{1700}, of 1.0 mag) means that galaxies at a fixed UV luminosity can have SFRs different up to a factor of six (from -1σ to 1σ). In turn, the intrinsic distribution of UV extinction can cause a significant scatter in the observed $M_{1700}-M_{\rm star}$ scaling relation, even if the intrinsic relation between SFR and $M_{\rm star}$ is tight. To account for this effect, we use the UV-slope measurements of Bouwens et al. (2009). The main reason for using Bouwens et al. (2009) measurements instead of our own is their inclusion of the Hubble Ultra Deep Field (HUDF) and HUDF Parallel Fields in the analyses. These data not only provide higher-S/N measurements of colors for sources detected in the GOODS data but also allow us to define a larger sample of UV-faint sources form which the color distribution is computed. Assuming the Calzetti et al. (2000) extinction curve, the intrinsic scatter in

the UV continuum slope β (their Table 4) translates into the intrinsic scatter of 0.7–0.9 mag in the extinction at 1700 Å or A_{1700}. A more recent result by Castellano et al. (2012) also found a similar scatter. As for the observed $M_{1700}-M_{\rm star}$ relation, we correct it for the selection completeness we estimated from our simulations. For example, at $M_{1700} = -21$ ($-20, -19$), the completeness is 89 (72, 34)%. Because our sample is selected based on the UV colors and flux limit, it is fair to assume that at a fixed stellar mass correcting for the UV completeness is appropriate. In other words, the distribution of stellar mass at a fixed UV luminosity has no bearing on altering the intrinsic distribution we attempt to recover here.

In Figure 8, we show the observed UV luminosity distribution of galaxies at $z \sim 4$ at three different stellar mass bins, $M_{\rm star} = 10^9, 10^{9.5}, 10^{10} M_\odot$. Histograms represent the distribution of absolute UV magnitudes for all galaxies (solid) and in individual fields (dashed and dotted for GOODS-S and GOODS-N, respectively). The smooth function in each panel approximates the emergent distribution of M_{1700} when there is a perfect correlation between SFR intrinsic UV luminosity, L_{1700}, and stellar mass and dust extinction is solely responsible for the spread in the observed UV luminosity. All functions have a standard deviation of $\sigma = 0.8$ mag, as discussed above, and are renormalized to match the peak of the observed distribution. We find that there is an excess of UV-faint galaxies at least in the lowest-mass bin. The observed distribution (top left) already suggests the excess although marginally. However, when the distribution is corrected for the selection incompleteness, the existence of such excess becomes more apparent. At $M_{1700} \sim -19.5$, the number of galaxies is higher by at least a factor of two when corrected for the selection effect. The selection completeness at this luminosity is still relatively high ($\approx$50%), and thus our estimate should be robust. Toward the massive end, the discrepancy between the observed distribution and that expected from UV extinction alone appears to lessen in both observed and corrected distribution. In particular, at the highest-mass bin, the observed distribution is in excellent agreement with the smooth function (except at the faintest end, where the completeness level is very low and probably not very robust). The implication is that the dust-corrected UV luminosity (directly proportional to SFR) and stellar mass are very tightly correlated at the massive end, while there is a significant scatter in the relation at the low-mass end.

6.2. Inferences from the Shape of the UV Luminosity Function and Stellar Mass Function

Now we turn our attention to the SMF of LBGs. In Section 5.1, we showed that the number counts of galaxies in the UV luminosity bin are consistent with a steep faint-end slope ($\alpha_{\rm UVLF} \approx -(1.7-1.8)$), in excellent agreement with other LF measures of high-redshift star-forming galaxies in the literature (Bouwens et al. 2007; Reddy & Steidel 2009). Interestingly, when we estimated the SMF of the *same* galaxies (Section 5.2), we found that the low-mass end slope of the mass function is relatively shallow, $\alpha_{\rm SMF} = -(1.3-1.4)$, compared to the faint-end slope of the UVLF (see Figure 7). The overall $M_{1700} - M_\star$ correlation discussed in Section 6.1 means that the majority of the galaxies on the low-mass end of the SMF mainly populate the faint end of the UVLF. The implication is that a substantial fraction of the galaxies on the faint end of the UVLF (which is very steep) have stellar mass lower than what the sensitivities of our data allow (i.e., $M_{\rm star} \sim 10^{8.7} M_\odot$ and $10^9 M_\odot$ at $z \sim 4$ and 5, respectively; see Section 4.2).
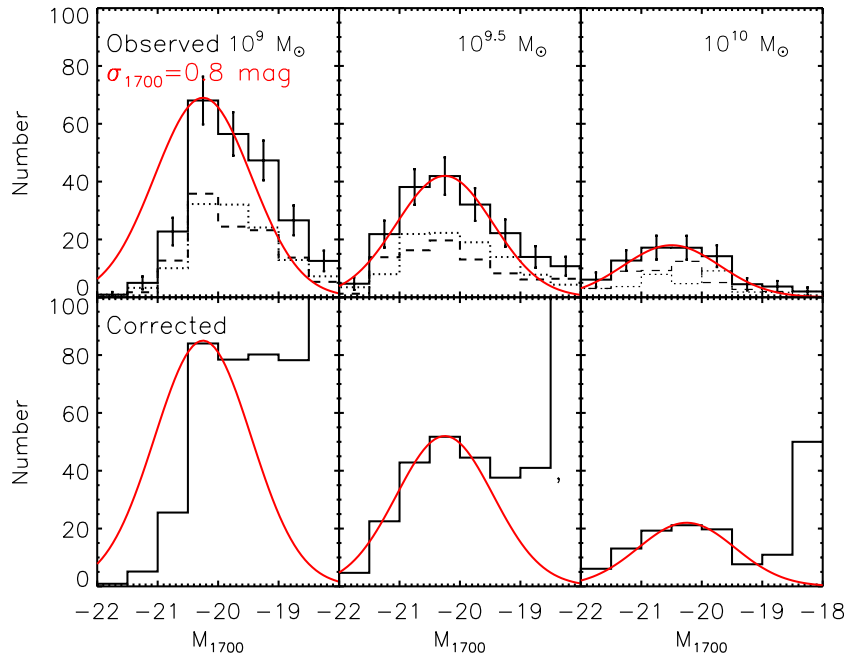
Figure 8. Distribution of absolute UV magnitudes in stellar mass bins of $M_{star} = 10^9$, $10^{9.5}$, 10^{10} $M_\odot$ (from left). In upper panels, we show the observed distribution of UV magnitudes for all galaxies (solid histogram) and for galaxies in each field (dashed and dotted histogram) in the respective bins. The error bars indicate Poisson errors. The smooth function represents a normal distribution with standard deviation of $\sigma = 0.8$ mag. The scatter is chosen to match the UV slope measurements by Bouwens et al. (2009) and its correlation with extinction (Meurer et al. 1999). If dust extinction alone is responsible for the observed spread of the observed UV luminosity, then the function would closely reflect the observations. In lower panels, we show the same distribution but corrected for the selection incompleteness. (A color version of this figure is available in the online journal.)

The relatively shallow low-mass end slope of the SMF has important implications for typical SFH of galaxies in our sample. If most galaxies have been forming stars in the past at comparable rates as when observed, then, to the zeroth order, the SMF of LBGs should have a similar shape as the UVLF as $M_{star} \propto$ SFR. Introducing scatter to this relation increases the overall normalization of the function but does not alter the overall shape or the low-mass end slope as long as the scatter is roughly constant. What is needed to effectively map galaxies in the UVLF to that in the SMF is to assume luminosity-dependent stellar-mass-to-UV-light ratios. This may mean that either the median ratio M_{star}/L_{1700} is systematically lower for galaxies on the faint end of the UVLF than more UV-luminous ones or the M_{star}/L_{1700} distribution has a long tail toward the low M_{star}/L_{1700} end. Both scenarios would effectively result in a large number of UV-faint galaxies having stellar masses lower than the minimum stellar mass ($\approx 10^{8.7}$ $M_\odot$) that we deem reliable.

In order to make more quantitative statements on these interpretations, we consider toy models in which the M_{star}–M_{1700} scaling law is described by a mean and scatter. We begin by taking the UVLF, assign stellar masses to the UV-selected galaxies according to the scaling law, and thereby compute the SMF of the same galaxies. This can be done by a simple integral:

$$\phi_M(\log M_{star}) = \int_0^\infty \phi(M_{1700})\mathcal{P}(\log M_{star}|M_{1700})dM_{1700},$$

(12)

where M_{star} is in units of $M_\odot$ and $\mathcal{P}(\log M_{star}|M_{1700})$ is a probability distribution of stellar mass at a fixed UV luminosity, which we assume to be a normal distribution. At the bright end ($M_{1700} \lesssim -(20$–$21)$), we fix the mean scaling law (where the probability distribution peaks) to what we observe in
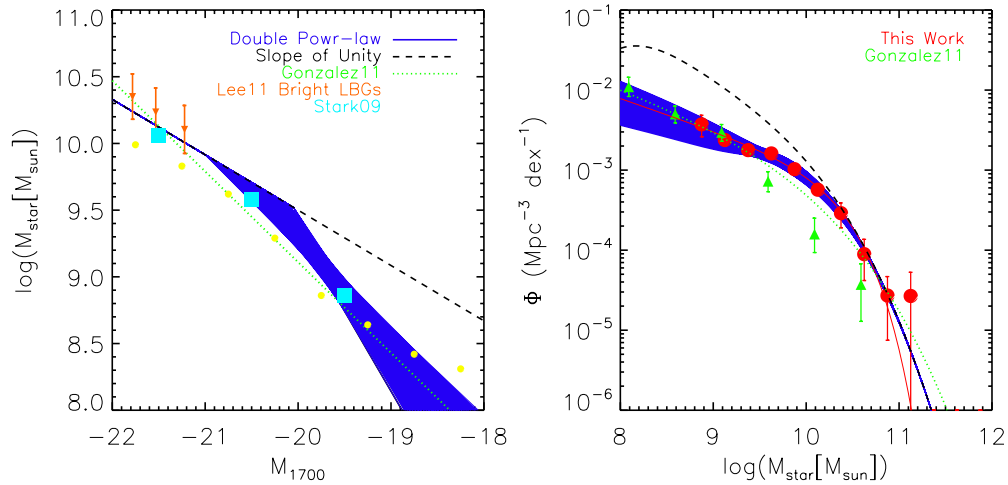
Figure 3:

$$\log M_{star} = \gamma_{bright} M_{1700} + 1.2.$$

(13)

We assume $\gamma_{bright} = -0.415$, the value that is consistent with the observed scaling relation at the bright end (Stark et al. 2009; Lee et al. 2011; also see Figure 3). A slope close to -0.4 means that the UV luminosity, L_{1700}, scales linearly with stellar mass. Furthermore, we assume a constant scatter of 0.3 dex around the mean scaling law. While the scatter has never been measured robustly for these galaxies, Lee et al. (2011) concluded that the M_{1700}–M_{star} correlation must be rather tight on the bright end based on the stacking analysis of a large sample of $L \gtrsim L^*$ ($M_{1700} < -21.5$) galaxies at $z \sim 4$. In any case, adding more scatter does little other than increasing the overall normalization of the SMF. This is because a larger number of galaxies on the faint end of the UVLF scatters into the massive end of the SMF than UV-luminous galaxies scatter into the low-mass end (similar to the "Malmquist bias"). As scatter does not affect the low-mass end slope, and we have no means to robustly measure this scatter observationally for the UV-faint galaxies, we set the scatter to be 0.3 dex throughout this discussion.

First, we compute the SMF using the scaling law observed for the relatively UV-luminous galaxies (Equation (13)). The results are shown in Figure 9 (dashed line). Clearly, the low-mass end slope in this case is much steeper than the observed one. This is expected because the assumed scaling law implies that the UV luminosity and stellar mass scale linearly ($M_{star} \propto L_{1700}$), and thus the low-mass end slope of the SMF should have the identical functional shape as the UVLF. Based on the fact that the assumed scaling law reproduces the bright end of the SMF and that it is determined from the observations where the measurements have the least uncertainties, we determine that the single power-law mapping between the observed UV luminosity and stellar mass does not provide a good description of our observations.

Figure 9. Inferred $M_{\mathrm{star}}-M_{1700}$ scaling law of LBGs at $z \sim 4$ shown together with the observational measures (Stark et al. 2009; Lee et al. 2011, cyan squares and downward triangles, respectively) (see Section 6.2). We also show the average stellar mass at a fixed UV luminosity directly determined from our measurements (yellow circles; see Figure 3). The $M_{\mathrm{star}}-M_{1700}$ scaling law is modeled as a double power law, and the acceptable models that reproduce the observed stellar mass function (right) are shown in blue. A single power-law model with the slope of unity well describes the observed $M_{\mathrm{star}}-M_{1700}$ relation on the bright end (Lee et al. 2011, downward triangles) but produces too many low-mass galaxies as the low-mass end slope is steep, $\alpha_{\mathrm{SMF}} = -(1.7-1.8)$, as shown in dashed lines on both panels. A model in which mass-to-light ratios decrease significantly toward the faint end provides a better description of the observed number counts in stellar mass bins, such as a double power law (blue swaths) or a steeper slope (shown in green as measured by González et al. 2011). A steeper decline of stellar mass toward the faint end of the UVLF implies that these galaxies have not sustained their SF at the rates seen at the time of observations.

(A color version of this figure is available in the online journal.)

Next, we try a double power-law model in which the scaling law is fixed to Equation (13) at the bright end, but the slope changes to a different value at the faint end. This approach is a two-parameter model (the luminosity M_{break} at which the break takes place, and the power-law slope at the faint end γ_{faint}). Over a wide range of M_{break} and γ_{faint}, we computed the SMFs to find the best scaling law that describes the observed SMF. In Figure 9, we illustrate the range of the scaling laws and their corresponding SMFs, which return the reduced chi-square of $\chi_r^2 \leqslant 1.2$ (68% confidence level). The best-fit model is $M_{\mathrm{break}} = -20.45$ and $\gamma_{\mathrm{faint}} = -0.84$. Clearly, all the acceptable scaling laws have the power-law slope $\gamma_{\mathrm{faint}} \leqslant -0.66$, much steeper than that on the bright end ($\gamma_{\mathrm{bright}} \approx -0.4$).

Alternatively, it is possible that the scaling law has a steeper but single slope as suggested by González et al. (2011). In Figure 9, we also show their best-fit scaling law ($\log M_{\mathrm{star}} = -0.68 M_{1700} - 4.49$ with 0.5 dex scatter around the mean) and corresponding SMF prediction in dotted lines. Indeed, their scaling law is consistent with the current observations in most luminosities (Stark et al. 2009; Lee et al. 2011, filled squares and downward triangles, respectively) and is in good agreement with our estimation (blue swaths) assuming double power laws. It also reproduces well the shallow slope of the galaxy SMF as measured independently by this work and theirs. The disagreement between the model SMF (dotted line) and their data (filled triangles in the right panel of Figure 9) can likely be explained by the fact that they constructed the SMF using individual measurements from their data and not scaling by the mean mass-to-light ratios as we did here. Their scaling law slightly underpredicts the galaxy space density at the intermediate masses ($\log[M_{\mathrm{star}}/M_\odot] = 9.5-10.5$). More data are likely needed to improve the accuracies of measurements in the relevant mass ranges. Despite minor discrepancies, the two studies clearly agree in that the observed mass-to-light ratios are luminosity dependent such that they is considerably lower by a factor of 2–4 for sub-L^* galaxies compared to UV-bright galaxies.

6.3. The Impact of Luminosity-dependent Extinction

As discussed in Section 6.2, the difference in the observed slopes between the UVLF and SMF of the same galaxies seems robust. As shown in Figure 9, if galaxies obey a single power law (dashed lines) on the observed $M_{1700} - M_{\mathrm{star}}$ plane, the abundance of low-mass ($\lesssim 10^{9.5}\,M_\odot$) galaxies would be overpredicted by a factor of several. However, the observed difference may not entirely be attributed to changing mass-to-light ratios *if* internal dust extinction strongly depends on UV luminosity. If the extinction systematically increases with UV luminosity, the space density of UV-bright galaxies would be boosted more substantially than that of UV-faint galaxies after correcting for dust. As a result, the "dust-corrected" UV luminosity function (intrinsic UVLF, hereafter) would have a shallower faint-end slope than the observed UVLF. Hence, it is important to quantify how much of this effect would be needed to explain the observed $M_{1700} - M_{\mathrm{star}}$ scaling law without invoking changing mass-to-light ratios.

Unfortunately, existing works are somewhat conflicting on this topic at present, ranging from a relatively strong dependence (Bouwens et al. 2009, 2011a; Lee et al. 2011; Castellano et al. 2012) to little or no dependence (Ouchi et al. 2004a; Finkelstein et al. 2011; Dunlop et al. 2012) at $z \gtrsim 4$. The discrepancies among different works likely lie in the limited knowledge of the precise redshifts for the samples, measurement errors arising from the use of a limited number of photometric bands (often one or two colors), and accounting for the selection effects of high-redshift samples that may mimic the trend, or quite possibly a combination of these effects. Obviously, the strongest luminosity-dependent trend for extinction will have the most dramatic impact on the interpretations of our results. Therefore, we explore this possibility by using the most recent results of Bouwens et al. (2011a), who reported a clear trend of luminosity-dependent extinction.

We use their updated best-fit formula given by Bouwens et al. (2011a), $\beta = -0.10(M_{\mathrm{UV}} + 19.5) - 1.98$ and the 1σ

scatter of $\sigma_\beta = 0.3$ (Bouwens et al. 2009). Assuming the correlation between the extinction (A_{1600}) and UV slope (β) observed for local starburst galaxies (Calzetti et al. 2000, $A_{1600} = 2.31\beta + 4.85$), we estimate the effect of luminosity-dependent extinction on the shape of the UVLF. We start with a Schechter function for the intrinsic UVLF and then apply luminosity-dependent dust extinction to repopulate galaxies on the observed UV luminosity space and compute the observed UVLF. The Schechter parameters for the intrinsic UVLF were varied until we reproduced the observed UVLF at $z \sim 4$ (Bouwens et al. 2007). The use of a Schechter function is not essential to our results as long as the faint end is described as a power law. The best-fit faint-end slope for the intrinsic UVLF is -1.62, somewhat shallower than that of the observed UVLF of -1.76. Assuming a slightly different correlation from Meurer et al. (1999)—$A_{1600} = 1.99\beta + 4.43$—does not change the slope by more than $\Delta\alpha = 0.01$.

As expected, the resultant faint-end slope of the intrinsic UVLF is shallower than that observed if there is a positive correlation between extinction and UV luminosity. However, it seems that the "flattening" that occurs as a result of luminosity-dependent extinction is relatively mild even for the strongest trend reported so far and thus is insufficient to explain the observed shallow slope $-(1.3-1.4)$ for the SMF at the 1.5σ level (the discrepancy is at the $\approx 2\sigma$ level in the absence of such a trend). Hence, we conclude that changing mass-to-light ratios are indeed needed to explain the observed shape of the SMF even if extinction is luminosity dependent. Careful analyses on this topic will shed further light on more proper interpretations of our results.

6.4. The Origin of the Changing Mass-to-light Ratios

In what follows, we consider mainly three physical scenarios that might have caused the observed scaling law to deviate from a single power law. The first scenario we explore is that more UV-luminous galaxies have, on average, higher formation redshifts than UV-fainter ones, while *all* galaxies have sustained their SF at relatively constant levels as observed. This allows us to reproduce the observed SMF while maintaining that most galaxies have formed over a long smooth SFH. Such a picture is qualitatively in line with theoretical expectations that more massive galaxies formed earlier in more massive halos ("hierarchical formation"). Furthermore, smooth formation histories for average galaxies are favored by numerical simulations as cold gas is accreted efficiently via filamentary structures without shock heating (Kereš et al. 2005, 2009; Dekel et al. 2009). However, this picture appears to directly contradict two observations. First, ages of galaxies would be luminosity dependent in this scenario. Reading off the left panel of Figure 9, galaxies with UV luminosity $M_{1700} = -19.5$ have stellar mass of roughly one-third (blue swath) of what they would have amassed if the SF varies smoothly over time[13] (dashed line). In other words, those galaxies should be on average three times younger than those at $M_{1700} = -21$. Instead, we find that all galaxies have similar ages ranging over 200–400 Myr when a CSF history is assumed, in agreement with Stark et al. (2009). However, we note that it would be hard to constrain population ages younger than 200 Myr as the Balmer break would be weak at the 0.1–0.2 mag level. With the current depth of the available data (mainly limited

by the near-infrared data from the ground), large photometric errors can easily wash away the trend, even if it existed. Second, a perhaps more convincing observational constraint comes from the clustering studies. Lee et al. (2009a) found that, based on the measurement of the correlation function of identical samples, the observed small-scale clustering is inconsistent with halo clustering if the same number density is imposed (i.e., all halos above a mass threshold host a visible galaxy). They argued that if only a fraction (15%–40%) of halos host galaxies at a given time, the shape of the correlation function for galaxies and halos would be well matched. Because our first scenario posits that galaxies have long continuous formation histories, it directly contradicts the short duty cycle implied by the clustering studies by Lee et al. (2009a).

The second possibility is that the formation redshifts of galaxies depend on luminosity, but the average SFRs of galaxies also rise with time. Rising SFH has been suggested based on several high-redshift observations. These include the arguments that galaxies at high redshift appear to be very young while forming a relatively tight sequence on the SFR–M_{star} plane, which apparently contradict one another unless rising SFH is invoked (Renzini 2009; Maraston et al. 2010; Lee et al. 2011). Another compelling argument is that the median SFR at a fixed comoving density rises with redshift at least for the most UV-luminous galaxies (Papovich et al. 2011). Recent numerical simulations also support this view (e.g., Finlator et al. 2011). While a rising SFH would make galaxies appear young regardless of their formation redshifts, the second scenario also predicts a long duty cycle and thus contradicts the clustering constraints.

Finally, we consider a scenario in which the average SFH is more episodic. One immediate feature would be that galaxies are not UV-luminous at all times, and therefore only a fraction of galaxies would be observed at a given time. The exact fraction would be determined by how long the UV–visible SF phase is sustained during the observed sampling of cosmic time. Furthermore, one would observe a substantial scatter on the SFR–M_{star} plane and the M_{star}–M_{1700} plane (independent of any variations caused by a dispersion in extinction), as some galaxies would be observed at the rise or the decline of their SF episode and thus lie far outside of the "SF main sequence." The episodic SFH scenario, however, does not preclude the existence of the main sequence albeit with larger scatter than that expected from continuous SF scenarios. In this case, the main sequence would represent how the average strength of each episode depends on halo mass. In fact, the halo-mass-dependent UV SF scenario is supported by clustering studies that the correlation length (i.e., the clustering strength) increases with UV luminosity, implying that more UV-luminous galaxies are, on average, hosted by more massive halos (Giavalisco & Dickinson 2001; Ouchi et al. 2004b; Adelberger et al. 2005; Kashikawa et al. 2006; Lee et al. 2006, 2009a; Hildebrandt et al. 2007, 2009). Such requirements can be met if more massive halos accrete, on average, a larger amount of cold gas to fuel SF episodes. In such a picture, a rising SFH is a direct result of the fact that the host halo masses and thus possibly cold gas accretion therein increase with cosmic time. In addition, the episodic SF scenario would be able to explain young ages observed at high redshift as the galaxy light is dominated by a recently formed stellar population, even if there exist older generations of stars formed from earlier episodes.

It is worth noting that, in this interpretation, the observed M_{star}–M_{1700} relation can be used to estimate the median SF duty

---

[13] We note that *any* SFH that varies smoothly over time would produce a slope of unity, including a constant SFH, exponentially or linearly increasing/declining models unless galaxy ages (or formation redshifts) change with stellar mass.

cycle (or the duration of a typical SF episode). Because the unit power-law slope of the M_{star}–M_{1700} relation is achieved when SF varies smoothly over time, the deviation from it can immediately be interpreted as due to a non-unity duty cycle. In this regard, the left panel of Figure 9 can provide a useful diagnostic to estimate the SF duty cycle. To the zeroth order, the duty cycle should be a ratio of the stellar mass that galaxies should have amassed if SF was continuous at the observed level (dashed line indicating the unit power-law slope) to the observed stellar mass (blue swath). Interestingly, the duty cycle of our $z \sim 4$ sample computed using this method is $\approx 20\%$–40% (evaluated at $M_{1700} \approx -19.4$ AB, roughly corresponding to the survey depth), in good agreement with that estimated from our clustering study (Lee et al. 2009a).

One interesting implication of this interpretation is that the SF duty cycle changes with luminosity or halo mass. At the UV-luminous regime ($M_{1700} \lesssim -21$), SF proceeds rather smoothly with long duty cycles and possibly at increasing rates. This is in line with other studies of UV-luminous galaxies (Papovich et al. 2011; Lee et al. 2011) at these redshifts. On the other hand, at the UV faint end ($M_{1700} \gtrsim -20$), the dominant mode of SF becomes progressively shorter and thus more "episodic." This may be due to more sporadic replenishment of cold gas in relatively low-mass halos (Lee et al. 2009a), which likely host these UV-faint galaxies, or due to an increasingly more efficient feedback mechanism, which can easily shut down SF by driving away available cold gas (Finlator et al. 2011). Whatever the main driver may be, the differential evolution of SF in dark matter halos suggested by this scenario presents a nontrivial test to theoretical models of galaxy formation.

6.5. SF Duty Cycle versus Invisible Halos Scenario from Finlator et al. (2011)

In light of our findings discussed above, it is interesting to consider an alternate interpretation to the short SF duty cycle scenario that explains the observations at high redshift. The two main observations include uniformly young population ages (Stark et al. 2009) and a relatively low halo occupancy implied by the clustering study (Lee et al. 2009a). The former implies that galaxies observed at higher redshift (e.g., $z \sim 6$) cannot be progenitors of those observed at lower redshift (e.g., $z \sim 4$) because their population ages are similar. On the other hand, the clustering study implies that only a fraction (15%–40%) of all halos host a visible star-forming galaxy at a given point in time. Both can be explained naturally if UV–visible SF does not continue for the Hubble time, but rather has a typical timescale of ≈ 300 Myr.

Another interpretation put forward by Finlator et al. (2011) is that there is a significant scatter in the baryonic mass and halo mass relation in low-mass halos. Motivated by hydrodynamic simulations, they predict that galaxies in a significant fraction of low-mass halos will be affected more severely by strong outflows ("momentum-driven winds;" Davé et al. 2006), which suppress SF therein. In turn, these galaxies would have much lower stellar mass than others hosted in halos of similar masses. As a result, galaxies hosted by low-mass halos would span a broad range of SFRs (and stellar masses), and the observed galaxies only represent the top end of this distribution (i.e., galaxies whose SF is least suppressed by feedback). This scenario can also naturally explain a low halo occupancy number similar to that measured by Lee et al. (2009a, 0.15–0.40) for LBGs, as a large fraction of halos will remain "dark" out of observational reach. They further predicted that the observed UV-selected galaxies will go on to have a smoothly rising SFH

with a duty cycle of unity, while others will remain undetected by observations, which explains the young ages (their stellar population is always dominated by recently formed stars).

One caveat in their scenario is that the observed galaxies would form a tight sequence on the SFR–M_{star} plane (see their Figure 10) as their SFH is smoothly rising with time. One direct result is that the SMF of LBGs should still closely mirror the UVLF and thus the low-mass end slope of the SMF is close to $\alpha_{UVLF} \sim -2.0$, a typical value for the faint-end slope predicted by simulations (e.g., Finlator et al. 2006; Choi & Nagamine 2010). A very steep SMF slope is in direct conflict with our findings and those of González et al. (2011) that the SMF rises considerably more shallowly than the UVLF of the same galaxies and that the stellar-mass-to-UV-light ratio indeed changes with luminosity. More careful comparison of the observed $\mathcal{M}_{star}/\mathcal{L}_{UV}$ with the predictions on a larger sample will provide a useful test for the validity of such a scenario.

7. SUMMARY

Using one of the deepest multi-wavelength data sets on two independent fields, we have investigated the statistical properties of star-forming galaxies (LBGs) at $z \sim 4$ and 5, namely, their number counts as a function of UV luminosity and stellar mass, as well as how these two quantities are related to each other. These statistics have important implications for the average SFH of these galaxies as they represent the ongoing SF and the time integral of the past SF, respectively. Based on our analyses, we conclude the following:

1. The galaxy counts in the rest-frame UV magnitudes (at 1700 Å) are consistent with a steep faint-end slope, $\alpha \approx -(1.7$–$1.8)$, of the UVLF, in good agreement with several existing studies at high redshift (Reddy & Steidel 2009; Bouwens et al. 2007, see Section 5.1 and Figure 6).

2. Based on the locations on the M_{1700}–M_{star} plane populated by galaxies in our samples, we find a broad correlation between stellar mass and UV luminosity, such that more UV-luminous galaxies are, on average, also more massive (Section 6.1; see Figure 3). However, the correlation has a substantial intrinsic scatter, in particular, for UV-faint galaxies. This is evidenced by the fact that there is a non-negligible number of UV faint but massive galaxies that are present in our samples. Roughly 35 (20)% of the massive galaxies ($M_{star} \geqslant 10^{9.5} M_\odot$) are UV faint ($M_{UV} > -20$) at $z \sim 4$ (5). In contrast to our findings for UV-faint galaxies, we also report that the region of the low stellar mass and high UV luminosity is largely devoid of galaxies, suggesting that the majority of the UV-luminous galaxies may have somewhat more extended SFH (for at least several hundred million years) than their UV-fainter counterparts.

3. While the current data do not allow us to directly quantify the *intrinsic* scatter of the SFR–M_{star} scaling law, the distribution of the UV magnitudes at a fixed stellar mass strongly suggests that there exists a significant intrinsic scatter at least on the low-mass end (Figure 8). Considering the adopted UV color selection, which systematically misses highly dust-obscured systems, the true distribution is likely even wider than that estimated in our analyses. Better sensitivities in the near-infrared (sampling the UV slope 2000–3500 Å) and mid-infrared (sampling longward of the Balmer/4000 Å break) are needed to make more precise measurements of the intrinsic SFR–M_{star} scaling law.

4. We make the statistically robust estimates of the SMFs for LBGs at $z \sim 4$ and 5 (Section 5.2). Our measurements suggest that the low-mass end slope of the SMF is $\alpha_{SMF} \approx -(1.3-1.4)$ and thus is not as steep as that of the UVLF of the same galaxies at both redshifts (Figure 7), in agreement with a recent study by González et al. (2011) based on a smaller sample. The direct implication is that a large fraction of the UV-selected galaxies are not massive enough, and therefore are too faint in their rest-frame optical bands, to be detected in the current IRAC data. The broad $M_{1700}-M_{star}$ correlation observed for these galaxies implies that those "missing" galaxies with stellar mass $M_{star} \lesssim 10^{8.7} M_{\odot}$ are mostly UV faint. In a scenario in which most galaxies have a relatively continuous SFH, these galaxies would have accumulated more stellar mass (by a factor of several) than observed, and as a result, the SMF would more closely mirror that of the UVLF of the same galaxies. Hence, our results favor a more episodic formation history in which SFRs of galaxies largely fluctuate over cosmic time. The duty cycle inferred from our SMF measurements is 20%–40% (Section 6.4), in good agreement with that implied from a clustering study by Lee et al. (2009a).

5. Using a simple toy model, we demonstrate that the stellar-mass-to-UV-light ratio should decrease systematically for galaxies with $M_{1700} \gtrsim -20.5$ (Section 6.2; also see Figure 9) in order to reproduce the observed low-mass end slope of the SMF. Assuming roughly constant mass-to-light ratios for all galaxies would result in significant overestimation for the abundance of low-mass galaxies, in direct violation of the current SMF measurements. We explore a possible impact of the luminosity-dependent dust extinction on the interpretations of our results using the strongest trend reported to date and conclude that it is likely insufficient to explain the observed shape of the SMF, although such a trend would imply a milder decline of the mass-to-light ratios toward faint UV luminosities than that expected in the absence of the luminosity-dependent extinction. We discuss several possible scenarios as the physical origins of the implied SFR–M_{star} scaling law (Section 6.4). We conclude that while we cannot completely rule out a possibility that UV-fainter galaxies have, on average, lower formation redshifts based on the observations, it is most likely that the "episodic SF" scenario is in best agreement with the available set of observations including clustering, abundance of low-mass galaxies, and population age constraints of high-redshift galaxies. One interesting implication of such a scenario is that the SF duty cycle increases with luminosity or halo mass. A direct result would be that galaxies with high SFR will continue to form stars with duty cycles close to unity and assemble their mass rapidly, while galaxies with low SFR will lag behind as their mass assembly is punctuated by relatively dormant phases. The differential evolution of galaxies at different luminosities presents a nontrivial test to theoretical models of galaxy formation.

REFERENCES

Adelberger, K. L., Steidel, C. C., Pettini, M., et al. 2005, ApJ, 619, 697
Alexander, D. M., Bauer, F. E., Brandt, W. N., et al. 2003, AJ, 126, 539
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
Bouwens, R. J., Illingworth, G. D., Blakeslee, J. P., Broadhurst, T. J., & Franx, M. 2004, ApJ, 611, L1
Bouwens, R. J., Illingworth, G. D., Franx, M., et al. 2009, ApJ, 705, 936
Bouwens, R. J., Illingworth, G. D., Franx, M., & Ford, H. 2007, ApJ, 670, 928
Bouwens, R. J., Illingworth, G. D., Franx, M., & Ford, H. 2008, ApJ, 686, 230
Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2011a, arXiv:1109.0994
Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2011b, ApJ, 737, 90
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682
Calzetti, D., Kennicutt, R. C., Engelbracht, C. W., et al. 2007, ApJ, 666, 870
Calzetti, D., Meurer, G. R., Bohlin, R. C., et al. 1997, AJ, 114, 1834
Capak, P., Cowie, L. L., Hu, E. M., et al. 2004, AJ, 127, 180
Caputi, K. I., Cirasuolo, M., Dunlop, J. S., et al. 2011, MNRAS, 413, 162
Castellano, M., Fontana, A., Grazian, A., et al. 2012, A&A, 540, A39
Chabrier, G. 2003, PASP, 115, 763
Choi, J.-H., & Nagamine, K. 2010, MNRAS, 407, 1464
Coppin, K. E. K., Smail, I., Alexander, D. M., et al. 2009, MNRAS, 395, 1905
Cowie, L. L., Barger, A. J., Hu, E. M., Capak, P., & Songaila, A. 2004, AJ, 127, 3137
Daddi, E., Alexander, D. M., Dickinson, M., et al. 2007a, ApJ, 670, 173
Daddi, E., Dickinson, M., Morrison, G., et al. 2007b, ApJ, 670, 156
Dahlen, T., Mobasher, B., Dickinson, M., et al. 2010, ApJ, 724, 425
Davé, R., Finlator, K., & Oppenheimer, B. D. 2006, MNRAS, 370, 273
de Santis, C., Grazian, A., Fontana, A., & Santini, P. 2007, New Astron., 12, 271
Dekel, A., Sari, R., & Ceverino, D. 2009, ApJ, 703, 785
Drory, N., Salvato, M., Gabasch, A., et al. 2005, ApJ, 619, L131
Dunlop, J. S., McLure, R. J., Robertson, B. E., et al. 2012, MNRAS, 420, 901
Elbaz, D., Daddi, E., Le Borgne, D., et al. 2007, A&A, 468, 33
Elbaz, D., Dickinson, M., Hwang, H. S., et al. 2011, A&A, 533, A119
Ferguson, H. C., Dickinson, M., Giavalisco, M., et al. 2004, ApJ, 600, L107
Fernández-Soto, A., Lanzetta, K. M., & Yahil, A. 1999, ApJ, 513, 34
Finkelstein, S. L., Papovich, C., Salmon, B., et al. 2011, arXiv:1110.3785
Finlator, K., Davé, R., Papovich, C., & Hernquist, L. 2006, ApJ, 639, 672
Finlator, K., Oppenheimer, B. D., & Davé, R. 2011, MNRAS, 410, 1703
Fontana, A., Salimbeni, S., Grazian, A., et al. 2006, A&A, 459, 745
Gehrels, N. 1986, ApJ, 303, 336
Giavalisco, M., & Dickinson, M. 2001, ApJ, 550, 177
Giavalisco, M., Dickinson, M., Ferguson, H. C., et al. 2004a, ApJ, 600, L103
Giavalisco, M., Ferguson, H. C., Koekemoer, A. M., et al. 2004b, ApJ, 600, L93
González, V., Labbé, I., Bouwens, R. J., et al. 2011, ApJ, 735, L34
Grazian, A., Fontana, A., de Santis, C., et al. 2006, A&A, 449, 951
Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35
Guo, Q., White, S., Boylan-Kolchin, M., et al. 2011, MNRAS, 413, 101
Hildebrandt, H., Pielorz, J., Erben, T., et al. 2007, A&A, 462, 865
Hildebrandt, H., Pielorz, J., Erben, T., et al. 2009, A&A, 498, 725
Kajisawa, M., Ichikawa, T., Tanaka, I., et al. 2009, ApJ, 702, 1393
Kashikawa, N., Yoshida, M., Shimasaku, K., et al. 2006, ApJ, 637, 631
Kereš, D., Katz, N., Fardal, M., Davé, R., & Weinberg, D. H. 2009, MNRAS, 395, 160
Kereš, D., Katz, N., Weinberg, D. H., & Davé, R. 2005, MNRAS, 363, 2
Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36
Labbé, I., Bouwens, R., Illingworth, G. D., & Franx, M. 2006, ApJ, 649, L67
Laidler, V. G., Papovich, C., Grogin, N. A., et al. 2007, PASP, 119, 1325
Lee, K., Giavalisco, M., Conroy, C., et al. 2009a, ApJ, 695, 368
Lee, K.-S., Dey, A., Reddy, N., et al. 2011, ApJ, 733, 99
Lee, K.-S., Giavalisco, M., Gnedin, O. Y., et al. 2006, ApJ, 642, 63
Lee, S.-K., Idzi, R., Ferguson, H. C., et al. 2009b, ApJS, 184, 100
Lin, L., Dickinson, M., Jian, H.-Y., et al. 2011, arXiv:1111.2135
Lotz, J. M., Madau, P., Giavalisco, M., Primack, J., & Ferguson, H. C. 2006, ApJ, 636, 592
Luo, B., Bauer, F. E., Brandt, W. N., et al. 2008, ApJS, 179, 19
Madau, P. 1995, ApJ, 441, 18
Magdis, G. E., Rigopoulou, D., Huang, J., & Fazio, G. G. 2010, MNRAS, 401, 1521

Mannucci, F., Cresci, G., Maiolino, R., et al. 2009, MNRAS, 398, 1915
Maraston, C. 2005, MNRAS, 362, 799
Maraston, C., Pforr, J., Renzini, A., et al. 2010, MNRAS, 407, 830
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
McLure, R. J., Cirasuolo, M., Dunlop, J. S., Foucaud, S., & Almaini, O. 2009, MNRAS, 395, 2196
Meurer, G. R., Heckman, T. M., & Calzetti, D. 1999, ApJ, 521, 64
Noeske, K. G., Weiner, B. J., Faber, S. M., et al. 2007, ApJ, 660, L43
Nonino, M., Dickinson, M., Rosati, P., et al. 2009, ApJS, 183, 244
Oke, J. B., & Gunn, J. E. 1983, ApJ, 266, 713
Ouchi, M., Hamana, T., Shimasaku, K., et al. 2005, ApJ, 635, L117
Ouchi, M., Shimasaku, K., Okamura, S., et al. 2004a, ApJ, 611, 660
Ouchi, M., Shimasaku, K., Okamura, S., et al. 2004b, ApJ, 611, 685
Pannella, M., Carilli, C. L., Daddi, E., et al. 2009, ApJ, 698, L116
Papovich, C., Dickinson, M., & Ferguson, H. C. 2001, ApJ, 559, 620
Papovich, C., Finkelstein, S. L., Ferguson, H. C., Lotz, J. M., & Giavalisco, M. 2011, MNRAS, 412, 1123
Pérez-González, P. G., Rieke, G. H., Villar, V., et al. 2008, ApJ, 675, 234
Quadri, R., van Dokkum, P., Gawiser, E., et al. 2007, ApJ, 654, 138
Ravindranath, S., Giavalisco, M., Ferguson, H. C., et al. 2006, ApJ, 652, 963
Reddy, N. A., Erb, D. K., Pettini, M., Steidel, C. C., & Shapley, A. E. 2010, ApJ, 712, 1070
Reddy, N. A., & Steidel, C. C. 2009, ApJ, 692, 778
Reddy, N. A., Steidel, C. C., Fadda, D., et al. 2006, ApJ, 644, 792
Renzini, A. 2009, MNRAS, 398, L58
Retzlaff, J., Rosati, P., Dickinson, M., et al. 2010, A&A, 511, A50
Rodighiero, G., Daddi, E., Baronchelli, I., et al. 2011, ApJ, 739, L40
Santini, P., Fontana, A., Grazian, A., et al. 2009, A&A, 504, 751
Santini, P., Fontana, A., Grazian, A., et al. 2012, A&A, 538, A33
Sawicki, M. 2012, MNRAS, 421, 2187
Sawicki, M., & Thompson, D. 2005, ApJ, 635, 100
Shapley, A. E., Steidel, C. C., Adelberger, K. L., et al. 2001, ApJ, 562, 95
Shapley, A. E., Steidel, C. C., Erb, D. K., et al. 2005, ApJ, 626, 698
Shim, H., Chary, R.-R., Dickinson, M., et al. 2011, ApJ, 738, 69
Stark, D. P., Ellis, R. S., Bunker, A., et al. 2009, ApJ, 697, 1493
Steidel, C. C., Adelberger, K. L., Giavalisco, M., Dickinson, M., & Pettini, M. 1999, ApJ, 519, 1
Steidel, C. C., Giavalisco, M., Pettini, M., Dickinson, M., & Adelberger, K. L. 1996, ApJ, 462, L17
van der Wel, A., Straughn, A. N., Rix, H.-W., et al. 2011, ApJ, 742, 111
Vanzella, E., Cristiani, S., Dickinson, M., et al. 2006, A&A, 454, 423
Vanzella, E., Giavalisco, M., Dickinson, M., et al. 2009, ApJ, 695, 1163
Wang, W.-H., Cowie, L. L., Barger, A. J., Keenan, R. C., & Ting, H.-C. 2010, ApJS, 187, 251
Windhorst, R. A., Cohen, S. H., Hathi, N. P., et al. 2011, ApJS, 193, 27
Wirth, G. D., Willmer, C. N. A., Amico, P., et al. 2004, AJ, 127, 3121