The Astrophysical Journal, 772:136 (13pp), 2013 August 1 doi:10.1088/0004-637X/772/2/136

A REST-FRAME OPTICAL VIEW ON $z \sim 4$ GALAXIES. I. COLOR AND AGE DISTRIBUTIONS FROM DEEP IRAC PHOTOMETRY OF THE IUDF10 AND GOODS SURVEYS*

P. A. Oesch[1,6], I. Labbé[2], R. J. Bouwens[2], G. D. Illingworth[1], V. Gonzalez[1,3], M. Franx[2], M. Trenti[4], B. P. Holden[1], P. G. van Dokkum[5], and D. Magee[1]

[1] UCO/Lick Observatory, University of California, Santa Cruz, CA 95064, USA; poesch@ucolick.org
[2] Leiden Observatory, Leiden University, NL-2300 RA Leiden, Netherlands
[3] University of California, Riverside, CA, USA
[4] Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
[5] Department of Astronomy, Yale University, New Haven, CT 06520, USA

Received 2012 November 2; accepted 2013 June 13; published 2013 July 17

ABSTRACT

We present a study of rest-frame UV-to-optical color distributions for $z \sim 4$ galaxies based on the combination of deep *HST*/ACS+WFC3/IR data with *Spitzer*/IRAC imaging. In particular, we use new, ultra-deep data from the IRAC Ultradeep Field program (IUDF10), together with previous, public IRAC data over the GOODS fields. Our sample contains a total of ∼2600 galaxies selected as *B*-dropout Lyman-break Galaxies in the HUDF and its deep parallel field HUDF09-2, as well as GOODS-North/South. This sample is used to investigate the UV continuum slopes β and Balmer break colors ($J_{125} - [4.5]$) as a function of rest-frame optical luminosity (using [4.5] to avoid optical emission lines). We find that galaxies at $M_z < -21.5$ (roughly corresponding to $L^*_{z\sim4}$) are significantly redder than their lower luminosity counterparts. The UV continuum slopes and the $J_{125} - [4.5]$ colors are well correlated, indicating that the dust reddening at these redshifts is better described by an SMC-like extinction curve, rather than the typically assumed Calzetti reddening. After dust correction, we find that the galaxy population shows mean stellar population ages in the range $10^{8.5}$ to 10^9 yr, with a dispersion of ∼0.5 dex, and only weak trends as a function of luminosity. Only a small fraction of galaxies shows Balmer break colors consistent with extremely young ages, younger than 100 Myr. Under the assumption of smooth star-formation histories, this fraction is 12%–19% for galaxies at $M_z < -19.75$. Our results are consistent with a gradual build-up of stars and dust in galaxies at $z > 4$ with only a small fraction of stars being formed in short, intense bursts of star-formation.

Key words: galaxies: evolution – galaxies: formation – galaxies: high-redshift – galaxies: luminosity function, mass function

Online-only material: color figures

1. INTRODUCTION

Great progress has been made over the last decade in exploring galaxies at very early times. Large optical and near-IR (NIR) surveys with *Hubble Space Telescope* (*HST*), such as the HUDF/HUDF09 (Beckwith et al. 2006; Oesch et al. 2007; Bouwens et al. 2011), GOODS/CANDELS (Giavalisco et al. 2004a; Grogin et al. 2011; Koekemoer et al. 2011), or BORG/HIPPIES (Trenti et al. 2011; Bradley et al. 2012; Yan et al. 2011) allowed us to identify large numbers of galaxies at $z \sim 4$–8, which are used to study the galaxy build-up in the first 1.5 Gyr of cosmic time.

One of the main shortcomings of *HST* surveys, however, is that they only probe the rest-frame UV light of high-redshift galaxies. Even the recently installed NIR camera WFC3/IR only probes to $\lesssim$2600 Å at $z > 4$ and the crucial rest-frame optical wavelengths are out of reach of *HST*. The observed light of high-redshift galaxies is thus strongly susceptible to dust extinction.

The only way to efficiently probe the rest-frame optical light of faint $z > 4$ galaxies is with the *Spitzer Space Telescope*, which samples up to ∼9000 Å at $z > 4$ in its 4.5 μm channel. Several surveys have obtained deep imaging in the [3.6] and [4.5] bands over a number of fields to cover the rest-frame optical of the large samples of high-redshift galaxies that were identified with *HST*. Such data are crucial for an unbiased study of the galaxy stellar mass build-up at $z \gtrsim 4$. They were already used to measure stellar population properties and galaxy masses at $z \sim 4$–7 (e.g., Yan et al. 2006; Eyles et al. 2007; Verma et al. 2007; Stark et al. 2009; Yabe et al. 2009; Labbé et al. 2010b, 2010a; González et al. 2010, 2012; Lee et al. 2012b), and recently even out to $z \sim 8$ (Labbé et al. 2012; Yan et al. 2011).

Despite some debate in the early literature, the rest-frame UV data alone indicate that UV bright galaxies at $z > 2$, on average, show redder UV spectral slopes β than their lower luminosity counterparts (e.g., Reddy et al. 2008; Bouwens et al. 2009, 2012; Castellano et al. 2012; Dunlop et al. 2012; Finkelstein et al. 2012). It is now clear that the galaxies at the bright end of the UV luminosity function (LF) show significantly redder colors than the average galaxy already as early as $z \sim 4$–6 (e.g., Lee et al. 2011; González et al. 2012; Willott et al. 2013). Furthermore, a clear trend to redder UV-to-optical colors is found in brighter, more massive galaxies at $z \sim 3$–4 (Papovich et al. 2004). These trends are likely caused by an increase in dust extinction at bright luminosities, since small amounts of dust have strong effects on the rest-frame UV colors (e.g., Wilkins et al. 2011; Bouwens et al. 2012).

When including *Spitzer*/IRAC photometry to measure rest-frame UV-to-optical colors as a function of UV luminosity, González et al. (2012) and Lee et al. (2011) find that these colors

* Based on data obtained with the *Hubble Space Telescope* operated by AURA, Inc. for NASA under contract NAS5-26555. Based on observations with the *Spitzer Space Telescope*, which is operated by the Jet Propulsion Laboratory, California Institute of Technology under NASA contract 1407.

[6] Hubble Fellow.

Table 1
Data and Sample Used in This Paper

Field	Area (arcmin^2)	3σ Depth [3.6]	3σ Depth [4.5]	Galaxy Sample (Total/$>3\sigma$ in [4.5])
HUDF	4.7	27.1	26.8	230/59
HUDF09-2	4.7	26.8	26.4	118/23
GOODS-S	160	26.2	25.7	1635/813
GOODS-N	120	26.2	25.7	660/378

Note. Depths are total AB magnitudes.

are broadly consistent with being just due to dust extinction assuming a similar stellar population age at different UV luminosities. Such a measurement can be significantly biased, however, if the galaxy population shows large scatter in dust extinction at a fixed stellar mass. An obvious next step is therefore to study galaxy colors as a function of rest-frame optical luminosity, which is significantly less affected by dust extinction and is a much more reliable proxy for galaxy stellar mass.

The main reason that such studies are not readily done is the difficulty of measuring accurate IRAC photometry for faint high-redshift sources, given the wide IRAC point-spread function (PSF) of $\sim$1.″6 ($\sim$10$\times$ the *HST* PSF). Without a neighbor-subtraction scheme, only about 30%–40% of galaxies are isolated enough for reliable aperture photometry (e.g., Stark et al. 2009), resulting in small sample sizes and low number statistics. Several groups have therefore developed sophisticated techniques to obtain photometry from crowded IRAC data over the last few years (see, e.g., Labbé et al. 2006; Grazian et al. 2006; Laidler et al. 2007).

Another complication of rest-frame optical studies from broad-band IRAC photometry at $z > 4$ are strong emission lines. There is now ample evidence that rest-frame optical lines can provide a very significant contribution to the observed *Spitzer*/IRAC fluxes of up to 0.5 mag (e.g., Chary et al. 2005; Schaerer & de Barros 2009; Shim et al. 2011; Stark et al. 2013; González et al. 2012; Labbé et al. 2012). Such lines have to be taken into account when interpreting galaxy colors or when fitting spectral energy distributions (SEDs). For instance, using synthetic models with emission lines Schaerer & de Barros (2009) and de Barros et al. (2012) estimate that the majority of galaxies at $z \sim 4$–7 are best fit with extremely young galaxy ages[7] of <100 Myr. This would have very important consequences on the total inferred cosmic star-formation rate density. However, it is a concern whether the many high-quality photometric constraints in the rest-frame UV compared to the few data points in the rest-frame optical could be biasing the ages determined from standard SED fits.

A simple way to test such predictions is to study galaxy rest-frame UV-to-optical color distributions. In this paper, we use deep *Spitzer*/IRAC imaging for a detailed investigation of galaxy color distributions of a large sample of $z \sim 4$ galaxies as a function of rest-frame optical luminosities. With these data we study the origin of the UV continuum reddening of galaxies, and we constrain the fraction of galaxies that are consistent with extremely young ages (<100 Myr) based on dust-corrected UV-to-optical colors.

Throughout this paper, we adopt $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$, $H_0 = 70$ km s^{-1} Mpc^{-1}, i.e., $h = 0.7$. Magnitudes are given in the AB system (Oke & Gunn 1983).

[7] If not mentioned otherwise, throughout this paper, we define age as the time since the onset of star-formation.

2. DATA AND GALAXY SAMPLE

2.1. Observations

Here we analyze *HST* and *Spitzer* data over four different fields: the two wide-area GOODS/CANDELS fields, the ultra-deep HUDF, and its deep parallel, the HUDF09-2.[8] All these fields are covered by Advanced Camera for Surveys (ACS) imaging from B_{435} to z_{850} from the UDF05 (PI: Stiavelli Oesch et al. 2007, 2009), the HUDF (Beckwith et al. 2006) and GOODS surveys (Giavalisco et al. 2004b). WFC3/IR data in three filters including J_{125} and H_{160}, was obtained as part of the HUDF09 (PI: Illingworth; Bouwens et al. 2011), the ERS (PI: O'Conell; Windhorst et al. 2011), and the CANDELS MCT programs (PI: Faber/Ferguson; Grogin et al. 2011, Koekemoer et al. 2011). For GOODS-North, we only include the first six visits obtained until July 19, 2012. The data reach to limits of $H_{160} = 27.0$–28.5 mag over the $\sim$300 arcmin^2 GOODS fields, and $H_{160} = 29.3$–29.9 mag over the two HUDF09 fields (covering a total of $\sim$10 arcmin^2).

Most importantly, all fields are covered by deep *Spitzer*/IRAC (Fazio et al. 2004) imaging in the [3.6] and [4.5] channels. In particular, we use new ultra-deep data from the 262 hr IRAC Ultradeep Field program (IUDF10; PI: Labbé; see also Labbé et al. 2012). This survey pushes the IRAC [3.6] and [4.5] imaging to exposure times of $\sim$120 hr and $\sim$80–120 hr over the HUDF and the HUDF09-2, respectively. The two wider-area GOODS fields were previously covered by 23 hr of public IRAC imaging (M. Dickinson et al., in preparation).

All the data and samples used in this paper are summarized in Table 1.

2.2. Measurements of the $z \sim 4$ Galaxy Sample

The galaxy sample analyzed here is identified using a standard $z \sim 4$ Lyman-break Galaxy (LBG) B_{435}-dropout selection based on the *HST*/ACS data (see, e.g., Giavalisco et al. 2004a; Bouwens et al. 2007). Sources were identified in a χ^2 image based on the V_{606}, i_{775}, and z_{850} ACS data, before applying the following color selection criteria:

$$(B_{435} - V_{606} > 1.1) \wedge B_{435} - V_{606} > (V_{606} - z_{850}) + 1.1 \wedge (V_{606} - z_{850} < 1.6).$$

This selects galaxies at $z \sim 3.2$–4.4, with a mean redshift $\bar{z} = 3.8$ and results in a total sample of 2643 galaxies down to $M_{\rm UV} = -16$ mag over all four fields. *HST* photometry is obtained in small elliptical Kron apertures from PSF matched images (see Bouwens et al. 2012, for details). The IRAC photometry is derived following the procedure of Labbé et al. (2010a, 2010b). In particular, we subtract neighboring foreground sources based on their profiles in the *HST* J_{125} and H_{160} images, after convolution to the IRAC PSF. We then perform aperture photometry on the cleaned images in 2″ diameter apertures, and correct to total fluxes using the growth curves of nearby stars in the field. The typical corrections are a factor 2.2, consistent with expectations from the IRAC photometry handbook. A comparison of our photometry against the GOODSMUSIC catalog over GOODS-South (Santini et al. 2009) shows no bias and a scatter of 0.20–0.25 mag.

The automated IRAC cleaning procedure does not always work. In particular, if a source is too close to an extremely

[8] The second HUDF09 parallel field to the northeast is not used due to its lack of B_{435} imaging, which is needed for the $z \sim 4$ LBG selection.

IRAC-bright neighbor, reliable photometry cannot be extracted. We therefore inspect all sources by eye for failures in the neighbor subtraction and flag those. These are excluded from the final analysis. With our procedure we are able to obtain clean IRAC photometry for ∼75% of all sources. This is compared to a ∼30%–40% success rate for simple aperture photometry on isolated sources only. We therefore more than double the galaxy sample compared to previous analyses which omit such a neighbor subtraction. Details on the IRAC cleaning and photometry will be presented in I. Labbé et al. (in preparation).

Since the IRAC flux in [3.6] can be significantly affected by strong Hα emission for up to 50% of our $z \sim 3.2$–4.4 sample (namely the sources lying at $z \geqslant 3.8$; e.g., Shim et al. 2011; Stark et al. 2013), we restrict our analysis to IRAC [4.5] band detections. This samples rest-frame 9000 Å, which is devoid of strong line emission. Our galaxy sample consists of 1273 sources with IRAC [4.5] detections more significant than 3σ. 82 of these come from the ultra-deep IUDF10 data.

Throughout the paper, we use this sample of ∼1300 galaxies with S/N([4.5]) > 3 as the primary sample. However, in Appendix A, we also show the effect of limiting the sample at S/N([4.5]) > 5, and we demonstrate that the choice of [4.5] detection significance has no impact on our results.

Using the observed [4.5] fluxes, we compute the rest-frame absolute magnitude at 9000 Å, assuming a fixed redshift for all sources at the mean of our sample $\bar{z} = 3.8$. In the following, we will denote this as M_z.

One of the main goals of our analysis is to identify the origin of the UV continuum reddening in these galaxies. To this end, we compute the UV continuum slopes, β ($f_\lambda \propto \lambda^\beta$), as in Bouwens et al. (2012) from a power-law fit to the observed fluxes in the *HST* bands V_{606}, i_{775}, z_{850}, $Y_{098/105}$ (where available), and J_{125}.

2.3. Low-redshift Contamination and Spectroscopic Sample

One important consideration in any analysis using Lyman-break selected galaxy samples is low redshift contamination. Large spectroscopic surveys based on rest-frame UV spectroscopy have found contamination fractions in $z \sim 4$–6 samples of ∼10%–20% (e.g., Vanzella et al. 2009; Rhoads et al. 2009).

We have matched our primary galaxy sample with the ESO Master spectroscopic catalog,[9] which combines the spectroscopic observations over the GOODS-South field from several surveys. The most important ones for our redshifts of interest are the ESO FORS2 and VIMOS surveys (Vanzella et al. 2009; Balestra et al. 2010).

In total, we only find 66 galaxies with reliable spectroscopic redshift measurements in our primary sample of ∼1300 galaxies. Only six of these (i.e., 9%) are catastrophic outliers with $z_{\rm spec} < 2.5$. The vast majority (55 sources, i.e., 83%) have $z_{\rm spec} > 3.2$.

This is therefore consistent with expected contamination fractions, and shows that the vast majority of galaxies are indeed at the correct redshift of our selection.

In Appendix B, we show that even at the very bright end, our sample is not dominated by contamination (see Figure 12). Rather the low redshift contaminants are randomly distributed throughout our sample both in terms of z-band luminosity and rest-frame UV-to-optical colors.

Furthermore, Figure 12 shows that an analysis of UV-to-optical color distributions can currently not be done with existing spectroscopic catalogs due to low completeness at rest-frame UV magnitudes fainter than 25.5 mag.

In a second check, we use the photometric redshift analysis of the GOODSMUSIC team (Santini et al. 2009). This catalog is primarily limited at $z_{850} = 26$ mag, containing 424 of our primary *B*-dropout galaxy sample over the GOODS-South field. The photometric redshift distribution of these sources is strongly peaked at $z \sim 3.2$–4.4 as expected for our selection. However, the distribution shows a tail to lower redshift. Similar to the spectroscopic sample, the majority of the photometric redshifts from GOODSMUSIC lie at $z > 3.2$ (82% of the sample, or 348 sources). On the other hand, only 8% (33 sources) appear to be catastrophic outliers having $z_{\rm phot} < 2.5$. Again, these catastrophic outliers are randomly distributed both in color and rest-frame optical luminosity, and excluding these sources does not affect our analysis significantly.

In summary, our analysis of both photometric and spectroscopic redshifts indicates that our sample shows a very low fraction of catastrophic outliers (<10%), having redshifts at $z < 2.5$, and the vast majority of the sources lies at the redshift of interest at $z = 3.3$–4.5. Low-redshift contamination is not expected to bias our results as these sources appear to be randomly distributed in color and magnitudes (see, e.g., Figure 12).

2.4. Possible AGN Contamination

Since one of the main goals in this paper is to infer the star-formation histories of $z \sim 4$ galaxies, it is important to make sure that the light in these galaxies is in fact dominated by stellar emission rather than by an active galactic nucleus (AGN). All our fields are covered by extremely deep X-ray and *Spitzer*/MIPS 24 μm data, which can be used to infer the presence of an AGN in these systems.

For galaxies in the CDFS, we cross-correlated our source list with the X-ray catalog of Xue et al. (2011) based on the 4 Ms *Chandra* imaging, and we visually checked all sources with $M_z < -21.5$ for lower significance detections. In the CDFN, we adopted the 2 Ms *Chandra* catalogs and images of Alexander et al. (2003). The sensitivity of these data is sufficient to detect Compton-thin AGNs with luminosities down to ~ 7–10×10^{43} erg s^{-1}. Out of our primary sample of ∼1300 galaxies, only 5 show weak X-ray detections, consistent with AGN activity, 3 of which are also present in the Fiore et al. (2012) X-ray catalog of the CDFS. We exclude all five sources from our analysis.

Compton-thick AGNs can be identified with MIPS 24 μm data, which is available over the GOODS fields as part of a large campaign (PI: Dickinson). Such obscured AGNs typically show a strong mid-IR excess and can thus efficiently be identified with MIPS imaging (see, e.g., Fiore et al. 2008).

We used the most recent 24 μm catalogs over the GOODS fields of Magnelli et al. (2011) to check for detections of our primary sample. None of our sources was present individually in these catalogs. While a large fraction of galaxies in our sample are blended with brighter foreground sources (given the very broad PSF of MIPS), a proper neighbor subtraction as we have done for the IRAC data is beyond the scope of this paper. However, we performed again a visual inspection of the brightest sources in our primary sample (with $M_z < -21.5$), through which we have identified three possible individual detections. We flagged these as potential contaminants, and removed these from our analysis, but note that this choice does not impact our result.

[9] See http://www.eso.org/sci/activities/garching/projects/goods/MasterSpectroscopy.html.

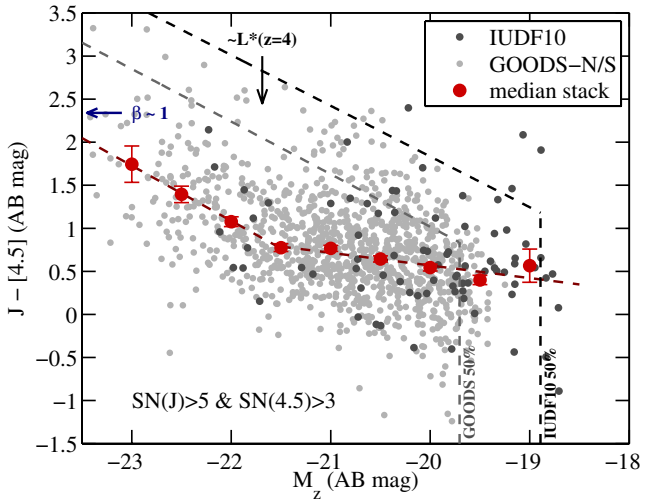

Figure 1. The observed Balmer break color $J_{125} - [4.5]$ (sampling rest-frame 2500 Å–9000 Å) vs. rest-frame optical magnitude for $z \sim 4$ galaxies. The gray dots represent galaxies with $>3\sigma$ detections in [4.5] and $>5\sigma$ in J_{125} (dark gray: IUDF10, light gray: GOODS). The median colors in bins of M_z are indicated by red dots with errorbars computed from bootstrap resampling. A clear luminosity dependence is evident, with a significant steepening at the bright end, where the Balmer break colors become increasingly redder. The increase in reddening sets in at around $M_z = -21.5$, which roughly corresponds to $L_z^*(z = 4)$, as estimated based on the characteristic magnitude of the UV LF and on the average UV-to-optical colors. This is indicated by the vertical black arrow. The dashed red line represents two linear fits to the $J_{125} - [4.5]$ vs. M_z relation at $M_z < -21.5$ and $M_z > -21.5$. Shown as gray and black dashed lines are the expected 50% completeness limits of the GOODS and IUDF10 data sets, respectively. The horizontal blue arrow shows the approximate location of galaxies with UV continuum slopes $\beta = 1$ according to the relation we find later in Figure 4. The LBG selection volume is significantly reduced for sources redder than that.

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

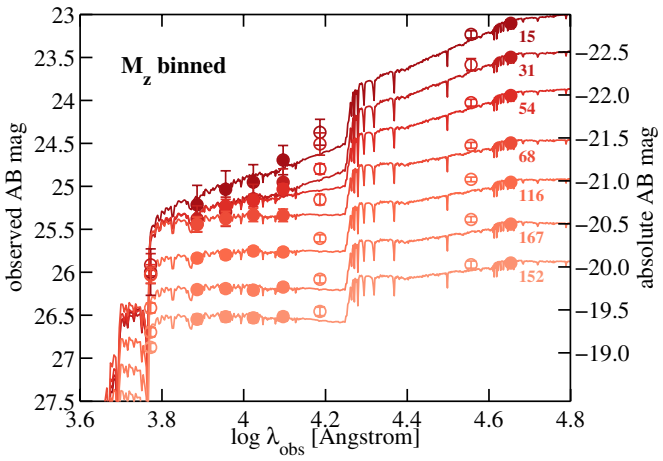

Figure 2. Median stacked SEDs of galaxies in bins of z-band luminosities. The numbers indicate how many galaxies contribute to each stack. Dramatic trends toward redder colors are seen as a function of rest-frame optical luminosities, in particular at $M_z < -21.5$. Galaxies are clearly not self-similar as is inferred from studies based on binning galaxies in UV luminosities only (e.g., González et al. 2012). The open symbols indicate filter fluxes which are not used when deriving synthetic SED fits to the stacks, since they are expected to change significantly across the extent of the LBG redshift selection. The H_{160}-band probes longward of 4000 Å for the lower redshift tail of the selection, while strong Hα emission may contaminate the [3.6] measurement for 50% of the sample (at $z = 3.8$–5.0). As can be seen, both these filter fluxes are indeed significantly higher than the best-fit SED for all stacks. The synthetic SEDs shown are based on Bruzual & Charlot (2003) models that are fit to the stacked fluxes.

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

Note also that the faint X-ray fluxes detected through stacking of high-z galaxy samples are consistent with being powered solely by star-formation (see, e.g., Cowie et al. 2012; Basu-Zych et al. 2013). Together with the tests above, it is clear that AGN contamination is very small in our $z \sim 4$ LBG sample and the light of these sources is expected to be dominated by stars.

3. RESULTS

3.1. Galaxy Colors as a Function of Rest-frame Optical Luminosity

Our large sample of galaxies with robust IRAC photometry allows us to investigate galaxy colors as a function of rest-frame optical luminosity. Of particular interest is the $J_{125} - [4.5]$ color, which straddles the Balmer break. At $z \sim 4$, J_{125} samples 2500 Å, and [4.5] samples 9000 Å. Note that J_{125} is chosen over H_{160} because the latter is contaminated by light longward of 4000 Å for the lower redshift tail of the B-dropout distribution, which complicates its interpretation.

The $J_{125} - [4.5]$ color–magnitude diagram is shown in Figure 1. This plot reveals that more luminous galaxies are significantly redder than their lower luminosity counterparts. Galaxies at $M_z = -22.5$ show colors of $J_{125} - [4.5] = 1.4$, while galaxies at $M_z \sim -20$ have $J_{125} - [4.5] = 0.6$. In particular, at $M_z > -21.5$ the color distribution is well described by a Gaussian with an observed dispersion of 0.4–0.5 mag (see also later Figure 6).

Interestingly, a luminosity dependence in the color–magnitude diagram is seen at all luminosities. However, the dependence is significantly steeper at the brightest luminosities. The color–magnitude relation is well reproduced by a two component linear relation, with a break roughly at $L_*(z = 4)$.[10] The slope of the relation changes from -0.64 ± 0.02 at $L \gtrsim L*$ to -0.15 ± 0.05 for $L \lesssim L*$.

Our finding of redder colors at brighter rest-frame optical luminosities is completely consistent with previous results from Papovich et al. (2004), who studied the $i_{775} - K_s$ colors of GOODS-S B-dropout galaxies as a function of K_s magnitude. The latter samples the rest-frame B-band. These authors already found a clear trend in the color–magnitude relation, with a well-defined blue lower envelope at $z \sim 4$ that reddens to brighter magnitudes, similarly to what we see in Figure 1 using a longer wavelength baseline and a much larger galaxy sample.

Most previous studies that were based on binning galaxies as a function of rest-frame UV luminosities suggested that the average UV-to-optical SED of galaxies changes only very little with luminosity. For instance, González et al. (2012) find rest-frame $U - V$ colors that get redder by only 0.026 per mag in UV luminosity (see also Figure 3). Such weak trends were in good agreement with the suggestion from smoothed particle hydrodynamics simulations that found star-formation histories of galaxies to be essentially self-similar (Finlator et al. 2011).

Our findings above suggest that using rest-frame optical luminosities to bin and stack galaxy SEDs, much more significant trends should become apparent. This is indeed the case as shown in Figure 2, where we present average SEDs in bins of M_z. These are simply computed based on averaging the fluxes of all sources

[10] $L*$ is crudely estimated based on the average $i_{775} - [4.5]$ color and the characteristic luminosity of the UV LF at $z = 4$ of $M^*_{\rm UV} = -21.06$ (Bouwens et al. 2007). This results in an estimated $M_z^* = -21.7$ mag, which is significantly fainter than the characteristic luminosity of the rest-frame V-band LF as measured by Marchesini et al. (2012) who find $M_V^* = -22.76^{+0.40}_{-0.63}$ mag.

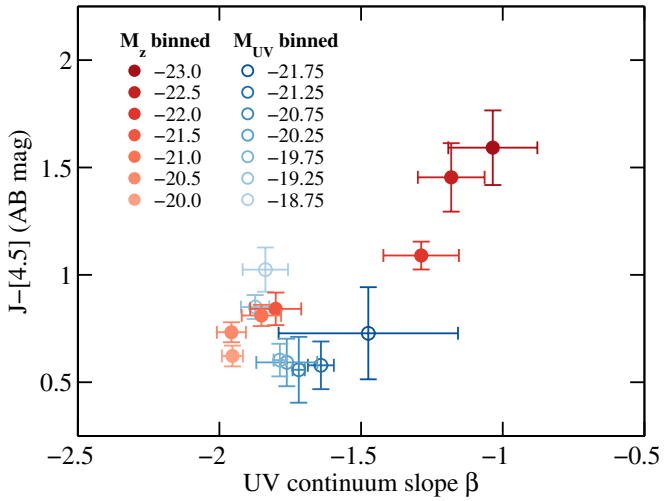

Figure 3. The relation between rest-frame UV-to-optical color and UV continuum slope of the median stacked SEDs of galaxies in bins of z-band luminosities (red filled) and UV luminosities (blue open). While the average galaxy SED changes very little as a function of UV luminosity (blue circles), a clear trend to redder colors is seen toward brighter rest-frame optical luminosities (red dots), both in the UV continuum slope and in the rest-frame UV-to-optical colors.

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

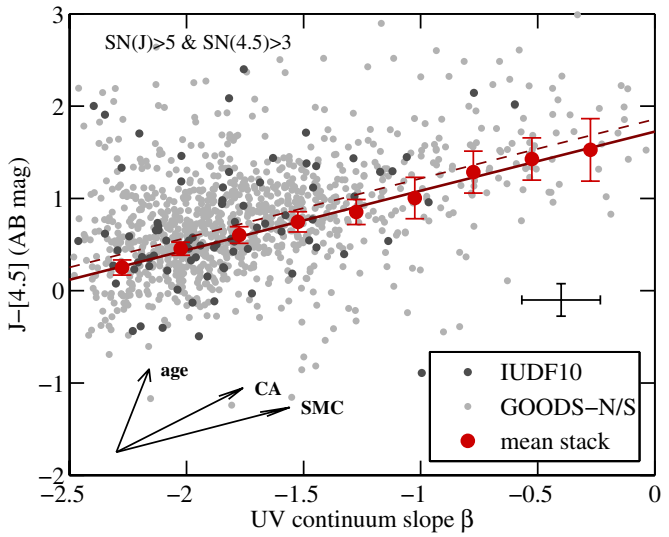

Figure 4. The relation between the UV continuum slope β and the observed $J-[4.5]$ color for $z\sim4$ galaxies. As in Figure 1, the filled dots correspond to sources with $>5\sigma$ detections in J_{125} and S/N([4.5]) >3. Dark and light gray dots denote galaxies in the IUDF10 and GOODS-North/South respectively. The mean stacked color (including the measured fluxes of IRAC undetected galaxies) in bins of β are indicated as red dots with errorbars corresponding to the uncertainty on the mean. The red solid line indicates a linear fit to these stacked colors. The correlation between β and the $J-[4.5]$ color is primarily driven by dust extinction, as shown by the vectors in the lower left. CA indicates the Calzetti (2001) star-burst dust reddening vector with $A_V=0.5$ mag, and SMC represents the Pei 1992 reddening vector with $A_V=0.25$ mag. Also shown is an age vector which is derived for an SED with constant star-formation and an age from 0.1 to 1 Gyr. The dashed line shows the effect of including only galaxies that are detected at 3σ in [4.5]. This results only in a zeropoint offset to marginally redder $J-[4.5]$ colors. However, the slope is essentially identical to the case where all galaxies are included showing the robustness of this relation (see text).

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

in the specific M_z bin, without any weighting scheme. A significant increase in reddening toward brighter z-band luminosities is obvious in these stacks. While these SEDs span 3 mag in M_z, they only vary by $\sim$1.5 mag in the UV luminosity. Furthermore, the stacked SEDs show a clear correlation between the UV continuum slope and the UV-to-optical color, with brighter galaxies in M_z showing redder SEDs overall (see Figure 3). This indicates that galaxies that are luminous in the rest-frame optical are significantly more dust obscured than fainter systems, which is likely the origin for the steeper trend in the $J_{125}-[4.5]$ color we found in Figure 1.

Finally, from these stacks it is clear that Hα flux contamination is indeed a potential problem for our B-dropout sample. All stacked SEDs show a significant excess at 3.6 μm compared to the best-fits SED. This excess ranges from 0.06–0.14 mag, with a tentative trend to a larger excess at fainter luminosities. Since Hα is only present in the [3.6] filter at $z=3.8$–5.0 (i.e., in $\sim$50% of our redshift selection), these numbers are roughly consistent with the mean excess of 0.27 mag as inferred by Stark et al. (2013), in particular for fainter sources.

A potential problem with the above measurements are the combination of different selection effects that can result in biased distributions. In Figure 1, we show the estimated completeness limits of our data as dashed lines. The vertical limit is due to the IRAC S/N cut, while the slanted line is mostly caused by the J_{125}-band cut, but its slope is influenced by the correlation between $J_{125}-[4.5]$ versus β (see Figure 4) due to the LBG selection. The completeness lines are estimated based on simulating a large number (10^5) of galaxies with synthetic SEDs with $M_z=-24$ to -18, with UV slopes in the range $\beta=-3$ to 1, and with redshifts according to our redshift selection function (see, e.g., Bouwens et al. 2007). The fluxes of these SEDs are then perturbed with Gaussian dispersions and all the selection criteria are applied, from which we compute the completeness as a function of M_z and $J_{125}-[4.5]$.

Clearly, at faint luminosities, we are potentially biased against red sources in the GOODS data. However, the $J_{125}-[4.5]$ color distribution peaks at sufficiently bluer colors than the completeness limit. Using the simulations above, we estimate that the observed mean color at $M_z=-20$ is not significantly biased due to selection effects ($\lesssim$0.05 mag).

Finally, photometric errors in [4.5] might bias the measurement of the color–magnitude relation of Figure 1, since both axes depend on the [4.5] flux measurement. A positive error in [4.5] would result in a brighter M_z with a redder $J_{125}-[4.5]$ color. We therefore checked and confirmed that an analogous steepening in the color–magnitude relation is still seen if the absolute magnitude derived from the [3.6] photometry is used instead of M_z. This relation is not shown here due to the possible flux contamination from strong Hα lines, as explained earlier. Nevertheless, this test suggests that the steepening in the color–magnitude relation is not significantly affected by scatter in the [4.5] flux measurements.

3.2. UV Continuum Slopes and Balmer Breaks

In order to investigate the origin of red UV continuum slopes β in brighter galaxies, we study the relation between β and the $J_{125}-[4.5]$ color of $z\sim4$ galaxies, which effectively compares the reddening in the UV with the Balmer break.

As shown in Figure 4, there is a clear and relatively well-defined relation between the UV slope β and the $J_{125}-[4.5]$ color, with UV redder galaxies showing redder colors across their Balmer break. The relation is well fit by the linear relation:

$$J_{125}-[4.5]=(0.64\pm0.03)\times\beta+(1.72\pm0.04) \quad (1)$$

where we have included all galaxies with J_{125} detections $>5\sigma$, irrespective of their [4.5] detection significance. Excluding

sources with non-detections in the IRAC [4.5] band results in a relation with an identical slope, but with a slightly redder zeropoint (1.86 ± 0.04; see dashed line in Figure 4). This is to be expected since IRAC detected sources are expected to be biased to higher IRAC fluxes, i.e., redder $J_{125} - [4.5]$ colors.

Similarly, excluding the reddest sources with $\beta > -1$ does not change the relation significantly, resulting in $J_{125} - [4.5] = (0.58 \pm 0.04) \times \beta + (1.60 \pm 0.06)$.

For sources with $\beta < -1.5$, the color dispersion perpendicular to the dust reddening is only 0.4 mag, which increases to 0.5–0.6 mag for redder sources. These dispersions are most likely explained by a variation in stellar population ages as indicated by the age vector in Figure 4 (see also Section 3.4).

Interestingly, a Calzetti (2001) dust curve would lead to a steeper relation than what is seen in the data. This can be seen from the dust reddening vector shown in the lower left of Figure 4. The change in the UV slope for the reddening vectors are computed based on a power-law fit to an SED with $\beta = -2$ over the same wavelength range as probed by our *HST* data (1600–2600 Å), while the $J_{125} - [4.5]$ is derived from the dust-curve directly using the extinction at 2500 Å and at 9000 Å.

As can be seen in Figure 4, Calzetti dust results in a slope of 1.29 and is clearly inconsistent with the relation given above. This means that the correlation between β and $J_{125} - [4.5]$ cannot simply be explained by Calzetti dust reddening only, but would require that dustier galaxies show younger stellar population ages in a correlated manner which results exactly in the observed slope.

A simpler explanation of the $J_{125} - [4.5]$ versus β correlation, however, is that galaxies are instead reddened by SMC-like dust. Using the parameterization of Pei (1992) SMC reddening results in a slope of 0.66, which is in excellent agreement with the observed relation (see also the reddening vector in Figure 4). Our data therefore provides tentative evidence that SMC reddening might be more appropriate for galaxies at $z \sim 4$. At least for brighter galaxies, this is in agreement with recent measurements of the infrared-excess for $z \sim 4$ sources with *Herschel* (Lee et al. 2012a). At fainter luminosities, the results from Lee et al. are consistent with the standard Meurer et al. (1999) and Calzetti (2001) extinction instead.

We therefore test whether we find a luminosity dependence in the $J_{125} - [4.5]$ versus β relation in our data. In Figure 5, we show this relation for three bins of z-band luminosity, down to $M_z < -19.75$, which is the completeness limit of the GOODS IRAC data.

As is evident, the brightest galaxies with $M_z < -21.75$ show a very large range of UV continuum slopes from $\beta \sim -2$ to $\beta \sim 0$. Across this full range, their $J_{125} - [4.5]$ colors follow the observed slope for the full sample extremely well, consistent with the *Herschel* results and with a reddening of SMC-like dust. However, the colors of these bright sources are systematically shifted by about 0.3 mag to the red, likely due to systematically older stellar population ages compared to the average galaxy in our sample. The fact that there are almost no bright galaxies with $\beta < -2$ indicates that the most luminous (and thus the most massive) galaxies at $z \sim 4$ must have formed both their stars and their dust over an extended period of time.

As we move to lower luminosities, we find that the mean UV slope clearly shifts to bluer, more dust-free colors. Interestingly, in the faintest bin ($M_z \sim -20.25$) the $J_{125} - [4.5]$ color appears to be independent of β. However, this is potentially due to incompleteness, which starts affecting galaxies as blue as $J_{125} - [4.5] = 0.9$ in that bin (see Figure 1).

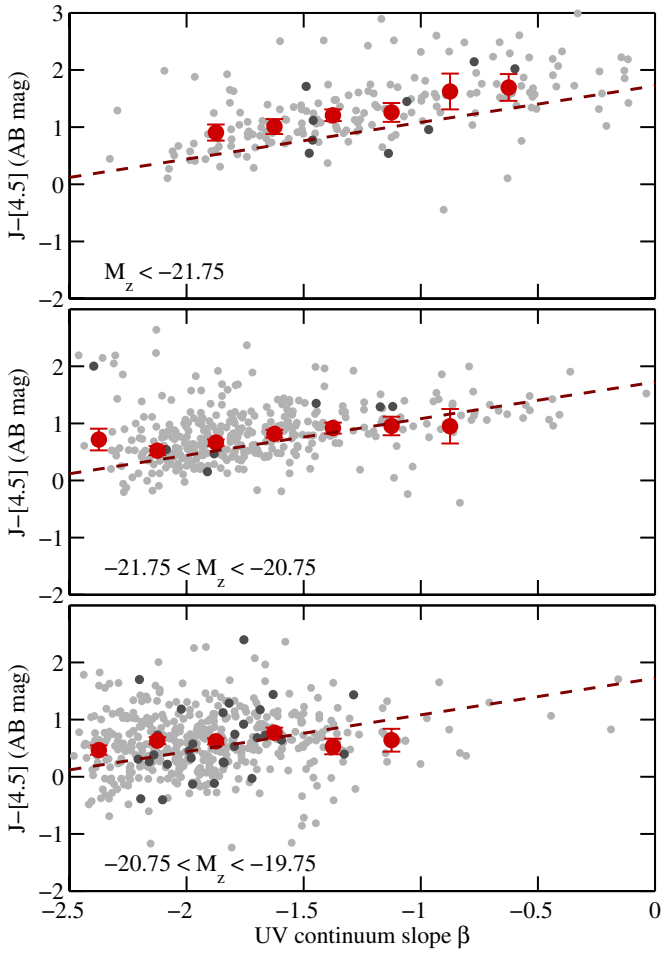

Figure 5. The relation between the UV continuum slope β and the observed $J - [4.5]$ color for $z \sim 4$ galaxies as a function of 9000 Å luminosity. The dashed red line is the mean relation derived for all galaxies (Equation (1)) and is repeated in every panel. As in Figure 4, the filled dots correspond to sources with $>5\sigma$ detections in J_{125} and S/N([4.5]) > 3. Dark and light gray dots denote galaxies in the IUDF10 and GOODS-North/South respectively. The mean stacked color (including the measured fluxes of IRAC undetected galaxies) in bins of β are indicated as red dots with errorbars corresponding to the uncertainties on the means. Only bins with more than 15 sources to stack are shown.

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

3.3. Dust-corrected UV-to-Optical Colors

The goal of our study is to constrain the intrinsic distribution of Balmer break colors in the $z \sim 4$ LBG population. Having found the correlation in Figure 4, we have a means of empirically correcting the $J_{125} - [4.5]$ colors for dust reddening. Our fiducial dust correction procedure is as follows: starting from the observed $J_{125} - [4.5]$ and β values for galaxies, we find the intersection of the reddening vector with the track of a synthetic SED model with constant star-formation and with ages in the range 10^6–3×10^9 yr. For the reddening vector, we use the empirical slope of Equation (1), which is essentially equivalent to SMC reddening.

This procedure is chosen to more accurately represent the likely dust extinction in young galaxies, which typically have bluer UV continuum slopes than the value of -2.23 for which the Meurer et al. (M99; 1999) correction results in zero dust extinction. The intrinsic, dust-corrected β values of our fiducial approach lie in the range -2.8 to -2.3. In Appendix C, we show a schematic figure which explains our dust correction (Figure 14,

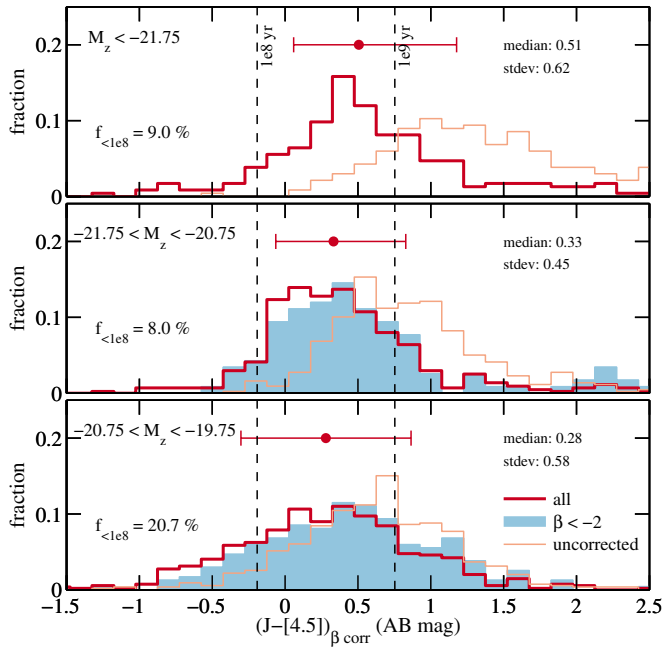

Figure 6. Histograms of dust-corrected colors $(J_{125} - [4.5])_{\beta\mathrm{corr}}$ in three bins of rest-frame optical absolute magnitude at 9000 Å (thick red line). The underlying, shaded blue histograms represent only the blue galaxy population with UV-continuum slopes $\beta < -2$ for which the dust correction is small. Also shown as thin, orange lines are the un-corrected color histograms for the full population, indicating that the dust-correction is much more significant for brighter galaxies. The dots with horizontal errorbars represent the median and 1σ range of the distribution. These indicate that unreddened galaxy colors show weak trends with rest-frame optical luminosity, but that the width of the distribution does not change dramatically. Vertical dashed lines indicate the colors of SEDs with constant star-formation rates and stellar population ages of 0.1 and 1 Gyr, which enclose the majority of the observed colors. The fraction of galaxies that are consistent with very young ages (<0.1 Gyr) is indicated on the left in each panel. This amounts to only 12% when corrected for photometric scatter and averaged over all galaxies with $M_z < -19.75$ mag. Note, however, that the exact number depends on the assumed star-formation history.

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

Appendix C), and we show the effect of using the standard M99 dust correction for comparison.

Histograms of both corrected and observed colors are shown in Figure 6, again for three bins of z-band luminosity down to $M_z = -19.75$. Our fiducial dust-correction results in average offsets of 0.8 mag at $M_z < -21.75$, and only 0.25 mag at $M_z = -20.25$. These offsets in $J_{125} - [4.5]$ color correspond to $A_V = 0.41$ mag and 0.13 mag for the Pei (1992) SMC extinction law, respectively.

While the original colors were strongly luminosity dependent, the corrected $J_{125} - [4.5]$ colors show only a weak trend with luminosity. The brightest sources ($M_z < -21.75$) are only ~0.2 mag redder on average than galaxies at $M_z > -21.75$. Note that these histograms are well sampled. From bright to faint, the three luminosity bins in Figure 6 contain 234, 438, and 545 galaxies, respectively.

For the two lower luminosity bins, we also show the color histograms for blue sources with $\beta < -2$, for which the dust corrections are smallest. This provides a good consistency check. Overall, sources with $\beta < -2$ comprise 30% of the sample at $M_z < -19.75$ (326 galaxies). However, at $M_z < -21.75$ this $\beta < -2$ sample contains only 16 sources, preventing a reliable comparison (which is why these are omitted from Figure 4). In both fainter bins, the $\beta < -2$ sample shows de-reddened color distributions in excellent agreement with the full sample, indicating that our dust-correction procedure is reliable and unbiased.

3.4. Limits on the Predominance of Extremely Young Galaxy Ages

Several papers have recently claimed that a dominant fraction of $z > 4$ galaxies may exhibit very young stellar population ages (<100 Myr; e.g., Schaerer & de Barros 2010; de Barros et al. 2012). This would have very important consequences, since it would imply extremely short duty cycles for LBGs, and large corrections to the cosmic star-formation rate density (e.g., Jaacks et al. 2012). One major difficulty in inferring stellar population ages from SED fitting is the degeneracy between dust extinction, age, and metallicity in reproducing the reddening of the UV slope. These problems are exacerbated at $z > 4$, since *HST* only probes rest-frame UV wavelengths, and the only rest-frame optical information typically comes from only two *Spitzer* bands.

Given these complications, it is thus instructive to use assumptions as simple and model independent as possible to test what range of ages is consistent with the data. This can be done using our simple dust-corrected Balmer break colors $(J_{125} - [4.5])_{\beta\mathrm{corr}}$. In Figure 6, we indicate the colors of SEDs with constant star-formation and ages of 10^8 yr and 10^9 yr. These are based on Bruzual & Charlot (2003) models, with Salpeter initial mass function and sub-solar metallicities ($Z = 0.2\,Z_\odot$). As can be seen, these lines nicely bracket the peak of the color distribution, indicating that the average $z \sim 4$ LBG shows stellar population ages in this range.

The fraction of sources with colors consistent with very young ages below 10^8 yr (i.e., with $(J_{125} - [4.5])_{\beta\mathrm{corr}} < -0.19$) is very small; less than 10% at $M_z < -20.75$ and only 21% at $M_z = -20.25$. Integrated over all galaxies with $M_z < -19.75$ for which we are complete, the number of galaxies with $(J_{125} - [4.5])_{\beta\mathrm{corr}} < -0.19$ is 14%. A simple correction of this number for photometric scatter suggests that this fraction intrinsically is even smaller, i.e., 12%. The formal uncertainty on this number is only 1%. However, we note that this fraction is highly model dependent. In particular, using an SED model with rising star-formation rates, the young fraction is significantly reduced, due to bluer galaxy colors at a given age. Additionally, in Appendix C we estimate that using the standard M99 and Calzetti (2001) dust relations and SEDs with constant star-formation would result in a young fraction of 19%. Despite of these uncertainties, the observed color distributions indicate that the galaxy population may only contain a relatively small number of extremely young sources, in contrast to earlier findings in the literature based on full SED fitting.

3.5. Fits to the Color Distribution at $z \sim 4$

Using our corrected color distribution, we can set interesting constraints on the ages and star-formation histories (SFHs) of $z \sim 4$ galaxies. In particular, we can fit for the distribution of galaxy ages under the assumption of smooth SFHs. This is done as follows: we sample galaxy ages from a normal distribution in log(age) with mean μ_A and width σ_A. We then compute the expected galaxy $J_{125} - [4.5]$ color for the given age, assuming either a constant star-formation rate (CSF) or a power-law rising star-formation history (RSF; $t^{1.7}$). The latter is the population average SFH consistent with the evolution of the UV LF (Papovich et al. 2011). The photometry is then perturbed using the real flux uncertainties of our galaxy sample, and the

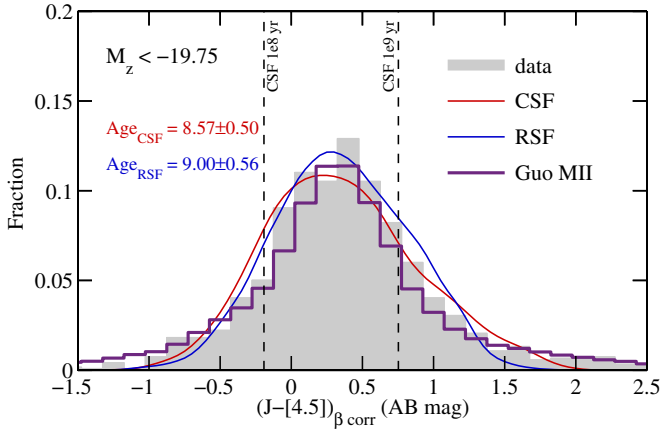

Figure 7. The color distribution for all galaxies with $M_z < -19.75$, together with three different model distributions. The filled histograms show the dust-corrected $J_{125} - [4.5]$ colors from the previous figure. The two thin lines show the best-fit color distributions when assuming a Gaussian distribution in log(age) and two different SFHs (CSF: constant star-formation, RSF: a power law rising SFR $t^{1.7}$). Both reproduce the color distributions fairly well. However, they fall somewhat short at the bluest and reddest tails of the distribution. The colors of the Guo et al. (2011) SAM are shown as thick purple line. These result in a color distribution in excellent agreement with the observed corrected colors, indicating that the SAMs predict SFHs that represent reality very well. As in the previous figure, the vertical dashed lines indicate the colors expected for a CSF SED model at an age of 0.1 and 1 Gyr.

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

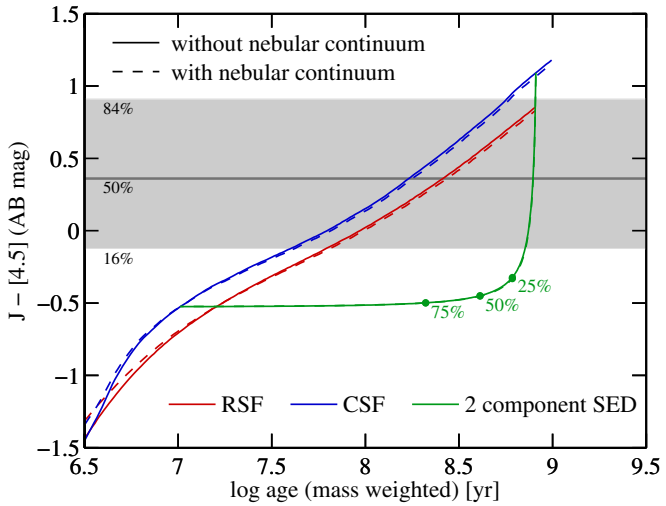

Figure 8. The $J_{125} - [4.5]$ color evolution of different SFHs as a function of mass-weighted age. The star-formation histories are: a power-law $\propto t^{1.7}$ (RSF), constant star-formation (CSF), and a two-component SFH with a varying mass-fraction of an extremely young stellar population (of age 20 Myr) on top of an older population (of age 1.6 Gyr). The location of 25%, 50%, and 75% young mass fractions are indicated by small dots. Solid lines are for pure stellar continuum, while dashed lines include the UV reddening due to nebular continuum emission from hydrogen and helium recombination in the inter-stellar medium of a galaxy (assuming ionization equilibrium and an escape fraction of 50% for ionizing photons). The shaded gray area spans 68% ($\sim 1\sigma$) of the observed galaxy color distribution at $M_z < -19.75$ mag (see Figure 6). The majority of galaxies clearly has less than a 25% contribution from extremely recent bursts, and sources are consistent with a mass weighted stellar population age between 50 and 500 Myr.

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

resulting histogram is compared to the data using a Poissonian likelihood for the number counts.

The best-fit color distributions are shown in Figure 7. These correspond to Gaussian age distributions with log(age) = 8.57 and $\sigma_{\log \mathrm{age}} = 0.50$ for constant star-formation and log(age) = 9.00 and $\sigma_{\log \mathrm{age}} = 0.56$ for the rising star-formation histories. The somewhat older mean stellar population age for the latter SFH is a result of bluer galaxy colors at a fixed stellar population age, given the dominance of younger stars. With these ages, the mean formation redshift of our sample is inferred to be $z_f = 5.0$ or $z_f = 9.3$ assuming CSF or RSF, respectively, and a mean observed redshift of $z = 3.8$ for the sample.

Given the mean μ_A and width σ_A of these best-fit Gaussian age distributions, the fraction of sources with very young ages is small. For the CSF case, the tail of sources with ages less than 100 Myr is 13%, while it is only 4% when assuming RSF. Interestingly, however, both age distributions include a significant fraction of galaxies with ages younger than 400 Myr (50% for CSF and 24% for RSF), which is the time between $z \sim 5$ and $z \sim 4$. These sources would therefore not have been present in the higher redshift LBG sample.

While the Gaussian age distributions for both assumptions on the SFHs do reproduce the observed color histograms rather well, they fall somewhat short at the bluest and reddest tails of the distribution. This indicates that real star-formation histories contain a modest component of bursty star-formation.

We can obtain a first constraint on the contribution of extremely short bursts to the total star-formation of galaxies from the expected $J_{125} - [4.5]$ color for a two-component SED with a very young population added to an older underlying one. This is shown in Figure 8. The Balmer break color saturates very quickly close to the color of the young component, even for a small mass fraction (<10%) produced in a recent burst. Bursty star-formation would therefore significantly tilt the color distribution to young galaxy ages, even if only a small mass fraction is built in bursts. The exact numbers depend on the detailed assumptions on the underlying old population and burst ages, but this suggests that *the majority of galaxies is consistent with only a very small fraction of stars being formed in very recent bursts.*

A further constraint on bursty star-formation can be obtained from the equivalent width distribution of strong rest-frame optical lines. Based on galaxies with spectroscopic redshifts in the range $z = 3.8$–5.0, Stark et al. (2013) estimated the combined Hα+[S II]+[N II] equivalent widths from the IRAC [3.6] flux excess. They find that this distribution peaks at $\log(W/\text{Å}) = 2.57$ and has a dispersion of 0.25 dex. We can estimate the predicted equivalent widths from our age distribution using a simple model for hydrogen recombination. Assuming a 10% escape fraction of ionizing photons (consistent with the most recent estimates for LBGs from Nestor et al. 2013) and photoionization equilibrium, we can derive the Hα emission line strengths from the tables in Osterbrock & Ferland (2006) and from the total number of ionizing photons that are produced in the SED model. This calculation shows that our age distributions result in equivalent widths that are lower than the observed values of Stark et al. (2013) by only ~0.15 dex, indicating that such smooth SFHs can produce most, but not all of the rest-frame optical line strengths. Bursty star-formation is thus certainly present to produce the most extreme tail of Hα emission, but it is likely not dominant.

Furthermore, we can use these calculations to check what fraction of the equivalent width distribution measured by Stark et al. (2013) requires young galaxy ages. For an SED with an age of 100 Myr, we derive a combined Hα+[S II]+[N II] equivalent width of 480 Å and 790 Å for our CSF and RSF SFH models, respectively. The Stark et al. (2009) distribution contains ~30% and ~10% of galaxies with larger values. This provides further

evidence for relatively small fractions of very young galaxies in LBG samples.

In order to compare our observations with more realistic SFHs, we use the results of the Guo et al. (2011) semi-analytical model (SAM) based on the Millennium-II dark matter simulation. In order to extract the $J_{125} - [4.5]$ colors of the SAM, we fit SEDs to the dust-free absolute magnitudes of their simulated galaxies at $z = 3.8$ with Bruzual & Charlot (2003) templates, and extract the J_{125} and [4.5] magnitudes of the best-fit SEDs. These intrinsic colors are then perturbed by the flux errors from our observational data. The result of this is shown as thick purple line in Figure 7. Obviously, the predicted color histogram is in very good agreement with the observed one, indicating that this SAM predicts SFHs that are a very good match to real galaxies. Note, however, that the Guo et al. (2011) model does not do as well in reproducing the dust-reddened (i.e., directly observed) color distributions. Furthermore, their model falls short of bright galaxies in the LFs at blue wavelengths at $z > 2$. As argued in, e.g., Henriques et al. (2012) both these effects are likely connected to uncertainties in the dust modeling. Analyses similar to the ones in our paper will help to further test and refine such models.

4. DISCUSSION

The large sample of galaxies with reliable IRAC photometry analyzed in this paper allowed us for the first time to study the correlations between UV and optical colors at $z \sim 4$, as a function of optical luminosity. The relatively well-defined correlation we find between the $J_{125} - [4.5]$ color and the UV slope β in Figure 4 provides direct evidence that the main driver for UV reddening in LBG samples is dust extinction.

The shallow slope of this relation is particularly intriguing, since Calzetti (2001) dust reddening would result in significantly steeper trends. This could in principle be the result of correlated changes in galaxy ages with dust extinction, such that younger galaxies are more heavily obscured. However, a more simple explanation is that dust extinction at $z \sim 4$ is more accurately described by an SMC-like dust curve.

Certainly, our measurements are not ideal for determining the exact shape of the reddening law. However, the fact that the brightest galaxies do follow a $J_{125} - [4.5]$ versus β trend that is parallel to the mean trend but offset by a constant value strongly suggests that the correlation is driven by SMC dust, and that the offset is due to older stellar population ages compared to the mean. If this is confirmed in future analyses, the use of SMC-like dust would result in small corrections of the reported SFRs for $z > 4$ galaxies in the literature.

While both local Lyman-break Analogs and the majority of bright LBGs at $z \sim 2$–3 are more consistent with Calzetti dust extinction (e.g., Overzier et al. 2011; Reddy et al. 2012, 2010), SMC-like extinction has been found to better reproduce the infrared excess of young, low-luminosity galaxies at $z \sim 2$–3 (e.g., Siana et al. 2008, 2009; Reddy et al. 2012). On the other hand, direct measurements of the IR excess with *Herschel* of $z \sim 4$ galaxies suggests an inverted luminosity dependence, with brighter galaxies having SMC-like dust, while fainter galaxies are found to follow the M99 relation (Lee et al. 2012a).

Clearly, the appropriate dust curve is still an open question. Solving this problem will require a multi-wavelength approach using complete samples of galaxies with unbiased spectroscopic redshifts, and combining deep *HST* and *Spitzer* data with *Herschel* and ALMA observations. Further progress should thus be possible in the very near future.

For now, we use the $J_{125} - [4.5]$ versus β relation to empirically correct the Balmer break color for dust extinction. While the observed $J_{125} - [4.5]$ colors are significantly redder for the brightest sources, the dust-corrected colors only show a relatively weak dependence on rest-frame optical luminosity. A residual 0.2 mag difference between galaxies at $M_z < -21.75$ and fainter sources is easily explained by a small increase in the stellar population age by only ∼0.2 dex.

Dust extinction is clearly the main contributor to red galaxy colors even at the brightest luminosities, and the steepening of the color–magnitude relation seen in Figure 1 is the result of a significant increase in the dust reddening in the most massive galaxies. This could have important consequences for the interpretation of the UV LFs, since increased dust extinction can effectively saturate the UV luminosities (see, e.g., Reddy et al. 2008; Bouwens et al. 2009).

The steeper color–magnitude relation at $L > L*$ is an indication that the most massive galaxies produce a significantly higher dust mass relative to their stellar mass than sub-$L*$ galaxies, or that they can more easily hold on to the dust produced in their supernovae compared to lower mass systems (see also Bouwens et al. 2012; Finkelstein et al. 2012). Although Finkelstein et al. (2012) only rely on rest-frame UV data to infer stellar masses, their results are consistent with our analysis, finding that the most massive galaxies show surprisingly red UV continuum slopes, which would require significant amounts of dust to be present as early as $z \sim 7$.

A more robust exploration of this effect is now possible based on the large samples with IRAC photometry (V. Gonzalez et al., in preparation).

In contrast to the strong trends of dust extinction, the galaxy ages only vary weakly with luminosity. The inferred mean galaxy ages increase by only 35% to 60% across 2 mag in z-band luminosity, depending on the assumed SFH (see Figure 9). This is derived using the same simulations as described in Section 3.5. The best-fit distributions show that there are essentially no bright galaxies with colors consistent with being dominated by very young bursts. However, this fraction appears to be increasing toward fainter luminosities (lower panel Figure 9), which can further be explored in the near future with the deeper S-CANDELS data.

Unfortunately, current simulations of high-redshift galaxies mostly focus on the sub-$L*$ population, due to the very large volumes necessary to sample the brightest galaxies. However, the very steep increase in dust extinction in $>L*$ galaxies that we find in our data presents an interesting test for galaxy evolution models. For example, we show that the corrected $J_{125} - [4.5]$ color distribution is almost perfectly reproduced by the Guo et al. (2011) SAM models, indicating that this SAM is reliably predicting galaxy SFHs. However, these same models fail in accurately capturing the observed steepening in the color magnitude relation and in predicting the UV LF and its evolution. Both these effects are likely connected to the uncertain treatment of dust extinction in such models (see, e.g., Gonzalez-Perez et al. 2013; Henriques et al. 2012).

5. SUMMARY

In this paper, we combined ultra-deep *HST* and *Spitzer* imaging to study rest-frame UV-to-optical colors of galaxies at $z \sim 4$. In particular, we perform a sophisticated neighbor subtraction in the IRAC images to obtain a sample of 1273 B-dropout galaxies with clean IRAC [4.5] detections more significant than 3σ from two fields of the HUDF09 survey as

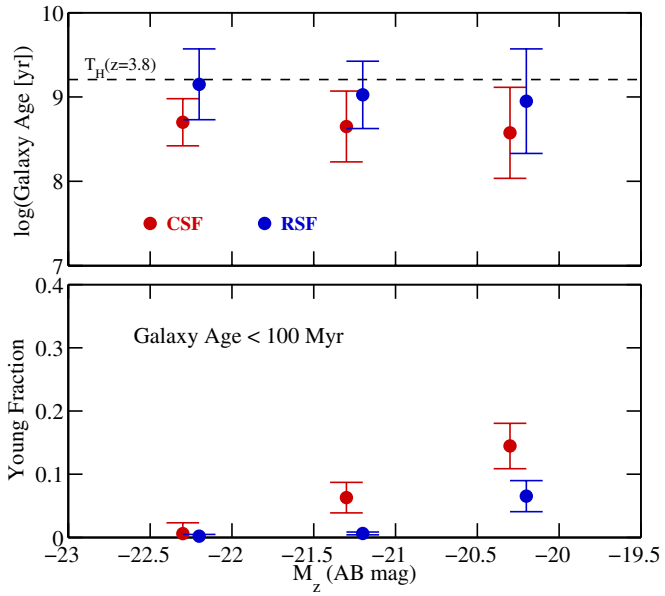

Figure 9. (Top) The parameters of the best-fit age distributions as a function of rest-frame optical luminosities. Mean galaxy ages are increasing only by 35% (60%) from $M_z = -20.25$ to $M_z = -22.25$, while the width of the age distribution decreases slightly from $\sigma \simeq 0.55$ at faint luminosities to $\sigma \simeq 0.3$–0.4, depending on the assumed SFH. (Bottom) The fraction of galaxies with ages younger than 100 Myr based on the inferred age distributions. Clearly, the brightest galaxies are not dominated by short bursts with <5% showing galaxy ages younger than 100 Myr. For constant star-formation histories, this increases to 15% at lower luminosity, while it stays rather small for rising SFHs.

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

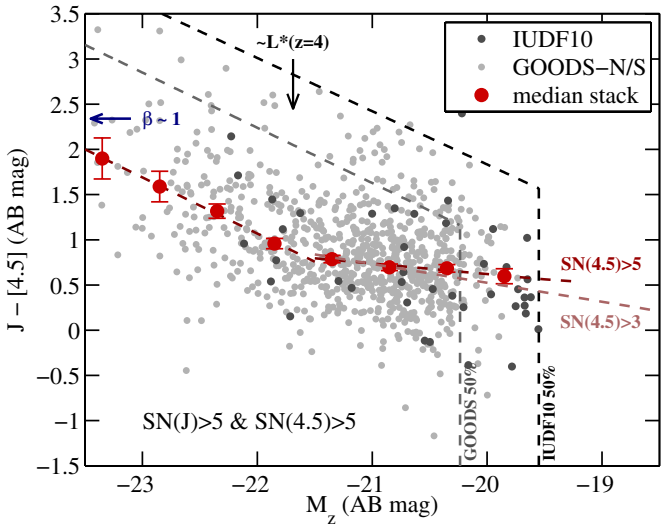

Figure 10. Same as Figure 1 but cutting the galaxy sample to $>5\sigma$ detections in [4.5], instead of $>3\sigma$. The relation of the observed Balmer break color $J_{125} - [4.5]$ vs. rest-frame optical magnitude remains unchanged at the bright end (above $L*$). Only a small change is seen at fainter luminosities (compare the dark red and light red dashed lines). Nevertheless, at all luminosities, a clear luminosity dependence is still evident.

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

well as from GOODS-North and South. The [4.5] channel data is preferred over the shorter wavelength [3.6] photometry in order to avoid any flux contamination by strong Hα emission (e.g., Shim et al. 2011; Stark et al. 2013).

Using this sample we find:

1. The observed $J_{125} - [4.5]$ color is strongly dependent on rest-frame optical luminosity. In particular, the relation steepens significantly for galaxies with $M_z < -21.5$ compared to the lower luminosity trend (see Figure 1), indicating that the most massive galaxies produce and retain more dust relative to their stellar mass than lower mass galaxies.
2. A relatively well-defined correlation is found between $J_{125} - [4.5]$ and the UV continuum slope β (see Figure 4). The most simple explanation for this slope is that the UV-to-optical SEDs of $z \sim 4$ galaxies are better reproduced by an SMC-like dust extinction rather than the standard Calzetti dust curve. This result will have to be confirmed in the future by combining large samples of galaxies with spectroscopic redshifts with additional longer-wavelength data to constrain the amount of dust extinction.
3. The dust-corrected colors of galaxies are only weakly dependent on luminosity, and their distribution is well reproduced by a Gaussian in log(age) of galaxies, with a mean log(age) = 8.57 or 9.00, depending on the assumed star-formation history (either constant or rising with time).
4. Only a very small fraction of 12%–19% of galaxies shows colors that are consistent with extremely young galaxy ages, below 100 Myr. The exact number of this fraction is dependent on the assumptions about the star-formation histories of galaxies. However, the vast majority of galaxies is consistent with a smooth build-up and only a negligible fraction of their stars being formed in very recent burst.

This paper gives a brief preview of the power of combining *HST* and *Spitzer* observations for studying galaxies at $z > 4$. After completion of the still on-going S-CANDELS program, these types of analyses can be extended to cover more than twice the area of our current sample to a depth of 50 hr (about 0.4 mag deeper than the GOODS IRAC data used here). These data sets will allow us to constrain SFH and age distributions of large samples of galaxies at $z > 3$. In particular when combined with additional constraints on dust extinction, e.g., from *Herschel* or ALMA, and with spectroscopic redshifts from the new and upcoming multi-object NIR spectrographs. This should allow for the next big steps forward over the next few years in working toward a self-consistent picture of star-formation and mass build-up before the peak of cosmic star-formation.

We are very thankful for useful comments and suggestions by the anonymous referee. Support for this work was provided by NASA through Hubble Fellowship grant HF-51278.01 awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS 5- 26555. This work was additionally supported by NASA grant NAG5-7697 and NASA grant HST-GO-11563.01.

Facilities: HST (ACS/WFC3), *Spitzer* (IRAC)

APPENDIX A

THE EFFECTS OF RESTRICTING THE ANALYSIS TO S/N([4.5]) > 5 SOURCES

In this section, we test the effect of using an S/N([4.5]) > 5 cut in our analysis instead of S/N([4.5]) > 3 as adopted in the main text. As expected, the effect of a higher S/N-cut is only seen at fainter z-band luminosities, where fewer sources are included. However, as shown in Figures 10 and 11, this only has a negligibly small impact on the average relations between

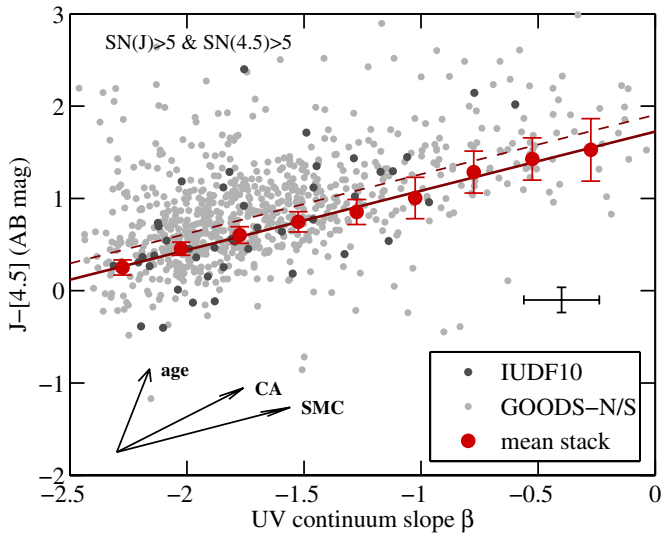

Figure 11. Same as Figure 4, but using only galaxies with S/N([4.5]) > 5. As can be seen in comparison to the figure in the main text, the higher cut in S/N does not result in a much cleaner β-color relation, indicating that the observed scatter is mostly intrinsic. As in Figure 4, the red dots correspond to a mean stack of *all* galaxies. These are therefore unchanged compared to the lower S/N cut. The dashed line shows the effect of using only detected sources. The slope of the β-color relation is again unchanged compared to the case where all galaxies are considered.

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

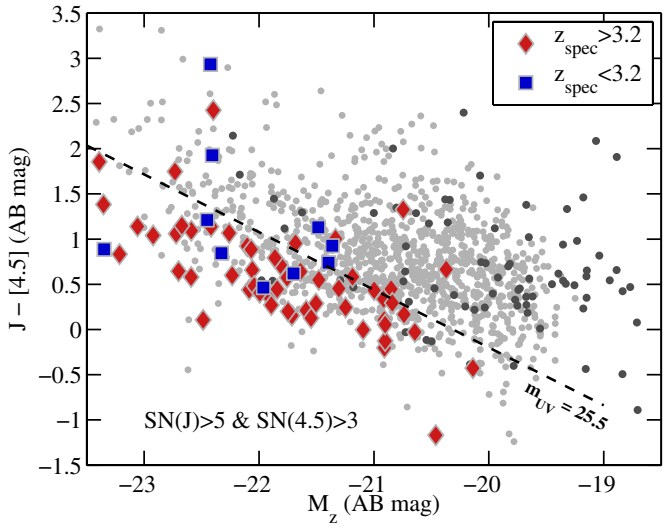

Figure 12. Same as Figure 1, additionally indicating the sample with spectroscopic redshifts. Only 66 galaxies in our sample have a spectroscopic redshift. 11 of these galaxies (17%) have a redshift measurement at $z_{spec} < 3.25$ (blue filled squares), while the remaining 55 (83%) lie at $z_{spec} > 3.25$ (red filled diamonds). The low redshift contaminants are randomly distributed throughout the sample and it is therefore unlikely that the color trends are purely driven by contamination. The existing redshift measurements are essentially all obtained from observed-frame optical spectroscopy, e.g., in the *i*-band. The dashed line indicates the approximate typical magnitude limit of such surveys of $i_{775} = 25.5$ mag. As can be seen, the vast majority of our sample is significantly fainter than this limit and redshifts can typically only be obtained if such sources show Lyα in emission. This clearly demonstrates that such an analysis cannot be done with current spectroscopic samples. However, rest-frame optical spectroscopy with new multi-object NIR spectrographs promises to obtain highly complete samples at much fainter rest-frame UV luminosities in the near future based on the detection of strong O III and Hβ line emission.

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

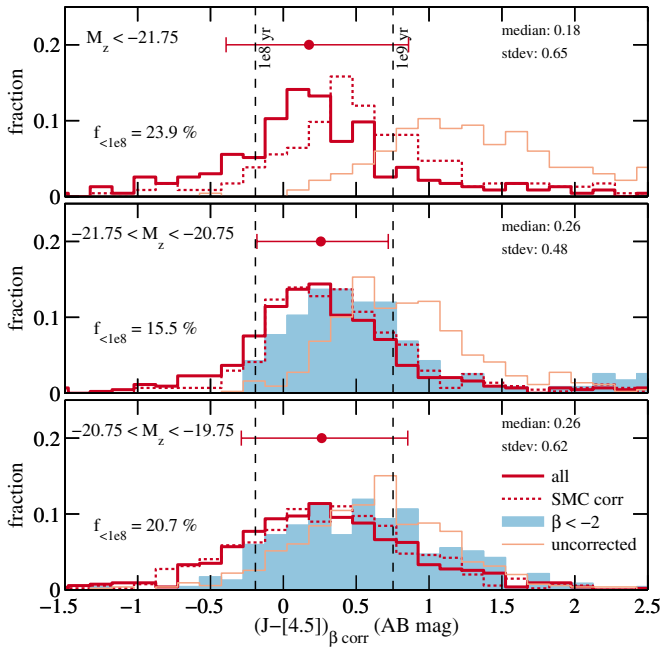

Figure 13. Same as Figure 6, but showing the dust-corrected colors when using the Meurer et al. (1999) A_{UV}–β relation and the Calzetti dust curve. For comparison, the dashed line shows the histogram for our fiducial dust correction as described in the main text. As can be seen, using the Calzetti dust correction results in dust corrected colors that are bluest at the brightest luminosities. Furthermore, the corrected colors for the bluest galaxies with $\beta < -2$ are significantly redder than for the full population. If this were true, this would indicate that the bluest galaxies are the oldest in the sample, which is not expected. Therefore, the combination of Meurer and Calzetti dust likely over-corrects the $J_{125} - [4.5]$ colors. Nevertheless, the integrated fraction of galaxies with corrected colors bluer than what is expected for a CSF population at 100 Myr is only 19%.

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

luminosities and colors, and does not affect the results of our paper. We therefore chose to keep the analysis at S/N([4.5]) > 3 throughout the main text of this paper.

APPENDIX B

THE SPECTROSCOPIC REDSHIFT SAMPLE

In order to check our sample of photometrically selected LBGs at $z \sim 4$, we matched all galaxies against the ESO Master spectroscopic catalog (see http://www.eso.org/sci/activities/garching/projects/goods/MasterSpectroscopy.html). This combines the spectroscopic observations over the GOODS-South field from several surveys, including in particular the ESO FORS2 and VIMOS surveys (Vanzella et al. 2009; Balestra et al. 2010), which are most important for our redshift of interest.

Only 66 galaxies with reliable spectroscopic redshift measurements are found out of our sample of ∼1300 galaxies. As can be seen in Figure 12, only a very small fraction of these are found to be catastrophic outliers (only six sources), with no particular trend with luminosity. The figure also shows that, given the low spectroscopic completeness at rest-frame UV magnitudes fainter than ∼25.5 mag, we could not restrict our analysis to sources with spectroscopic redshifts and draw conclusions which would have the same statistical weight or be relevant over the same dynamic range. However, this may change in the near future thanks to the advent of efficient NIR multi-object spectrographs, which will allow for detecting strong rest-frame optical lines such as O III with *K*-band spectroscopy up to $z \sim 3.8$ (see B. P. Holden et al., in preparation).

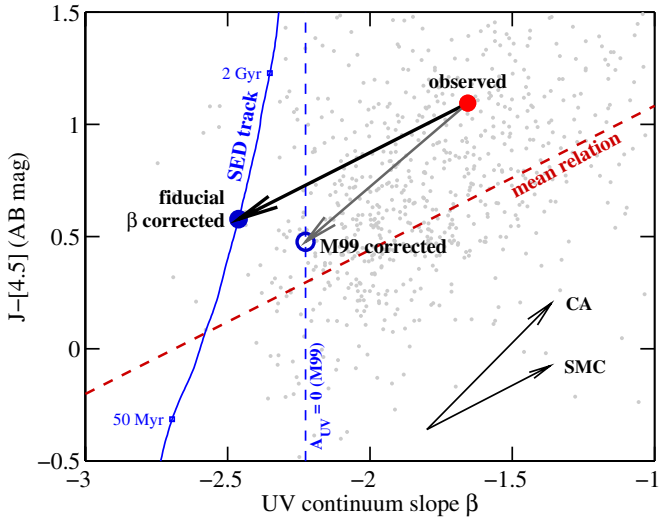

Figure 14. Schematic figure showing our fiducial dust correction procedure. The solid line shows the track of a Bruzual & Charlot (2003) SED model with constant star-formation at 0.2 $Z_{\odot}$ as the SED ages from 10^6 yr to 3×10^9 yr. The locations at ages of 50 Myr and 2 Gyr are labeled and marked with small squares. For every galaxy we start from the observed location in the β vs. $J_{125} - [4.5]$ plot (red dot), and we find the intersection of the negative reddening vector with the SED track (filled blue dot) to determine the dust-corrected $J_{125} - [4.5]$ color. For comparison, the standard Meurer et al. (1999) correction together with a Calzetti (2001) dust curve is also shown by the dark gray vector, which would result in the open blue dot. In particular for very UV-red galaxies the difference between the two dust corrections can be significant. The M99 relation assumes a dust-free UV slope of $\beta = -2.23$ (vertical dashed line), which is substantially redder than the values measured for star-forming SEDs with ages $\lesssim$1 Gyr. For the few galaxies that lie to the left of the SED track, we do not apply any dust correction to the color. The red dashed line represents the mean relation between β and $J_{125} - [4.5]$ of Equation (1), shown in Figure 4.

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

APPENDIX C

THE EFFECTS OF USING THE MEURER ET AL. (1999) DUST CORRECTION

As we show in the main text of this paper, an SMC like dust curve appears to be more consistent with the $J_{125} - [4.5]$ versus β relation (see Figure 4). This motivated our use of an SMC slope to correct galaxy colors for dust. However, for consistency with much of the previous literature, we also show the effects of using the combination of the Meurer et al. (1999) relation between $A_{\rm UV}$ and the UV slope β together with a Calzetti (2001) dust curve. The Meurer corrected color histograms are shown in Figure 13.

As can be seen, this approach results in dust corrected colors that are bluest at the brightest luminosities. Furthermore, the corrected colors for the bluest galaxies with $\beta < -2$ are significantly redder than for the full population. If this were true, this would indicate that the bluest galaxies are the oldest in the sample, which is not expected. Therefore, the combination of Meurer and Calzetti dust likely over-corrects the $J_{125} - [4.5]$ colors.

Despite of the larger correction, the integrated fraction of galaxies with corrected colors bluer than what is expected for a CSF population at 100 Myr is still only 19%.

REFERENCES

Alexander, D. M., Bauer, F. E., Brandt, W. N., et al. 2003, AJ, 126, 539
Balestra, I., Mainieri, V., Popesso, P., et al. 2010, A&A, 512, A12
Basu-Zych, A. R., Lehmer, B. D., Hornschemeier, A. E., et al. 2013, ApJ, 762, 45
Beckwith, S. V. W., Stiavelli, M., Koekemoer, A. M., et al. 2006, AJ, 132, 1729
Bouwens, R. J., Illingworth, G. D., Franx, M., & Ford, H. 2007, ApJ, 670, 928
Bouwens, R. J., Illingworth, G. D., Franx, M., et al. 2009, ApJ, 705, 936
Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2011, ApJ, 737, 90
Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2012, ApJ, 754, 83
Bradley, L. D., et al. 2012, ApJ, 760, 108
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Calzetti, D. 2001, PASP, 113, 1449
Castellano, M., Fontana, A., Grazian, A., et al. 2012, A&A, 540, A39
Chary, R.-R., Stern, D., & Eisenhardt, P. 2005, ApJL, 635, L5
Cowie, L. L., Barger, A. J., & Hasinger, G. 2012, ApJ, 748, 50
de Barros, S., Schaerer, D., & Stark, D. P. 2012, arXiv:1207.3663
Dunlop, J. S., McLure, R. J., Robertson, B. E., et al. 2012, MNRAS, 420, 901
Eyles, L. P., Bunker, A. J., Ellis, R. S., et al. 2007, MNRAS, 374, 910
Fazio, G. G., Hora, J. L., Allen, L. E., et al. 2004, ApJS, 154, 10
Finkelstein, S. L., Papovich, C., Salmon, B., et al. 2012, ApJ, 756, 164
Finlator, K., Oppenheimer, B. D., & Davé, R. 2011, MNRAS, 410, 1703
Fiore, F., Grazian, A., Santini, P., et al. 2008, ApJ, 672, 94
Fiore, F., Puccetti, S., Grazian, A., et al. 2012, A&A, 537, A16
Giavalisco, M., Dickinson, M., Ferguson, H. C., et al. 2004a, ApJL, 600, L103
Giavalisco, M., Ferguson, H. C., Koekemoer, A. M., et al. 2004b, ApJL, 600, L93
González, V., Bouwens, R. J., Labbé, I., et al. 2012, ApJ, 755, 148
González, V., Labbé, I., Bouwens, R. J., et al. 2010, ApJ, 713, 115
Gonzalez-Perez, V., Lacey, C. G., Baugh, C. M., Frenk, C. S., & Wilkins, S. M. 2013, MNRAS, 429, 1609
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
Henriques, B. M. B., White, S. D. M., Lemson, G., et al. 2012, MNRAS, 421, 2904
Jaacks, J., Nagamine, K., & Choi, J.-H. 2012, MNRAS, 427, 403
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, ApJL, 649, L67
Labbé, I., González, V., Bouwens, R. J., et al. 2010a, ApJL, 716, L103
Labbé, I., González, V., Bouwens, R. J., et al. 2010b, ApJL, 708, L26
Labbé, I., et al. 2012, arXiv:1209.3037
Laidler, V. G., Papovich, C., Grogin, N. A., et al. 2007, PASP, 119, 1325
Lee, K.-S., Alberts, S., Atlee, D., et al. 2012a, ApJL, 758, L31
Lee, K.-S., Dey, A., Reddy, N., et al. 2011, ApJ, 733, 99
Lee, K.-S., Ferguson, H. C., Wiklind, T., et al. 2012b, ApJ, 752, 66
Magnelli, B., Elbaz, D., Chary, R. R., et al. 2011, A&A, 528, A35
Marchesini, D., Stefanon, M., Brammer, G. B., & Whitaker, K. E. 2012, ApJ, 748, 126
Meurer, G. R., Heckman, T. M., & Calzetti, D. 1999, ApJ, 521, 64
Nestor, D. B., Shapley, A. E., Kornei, K. A., Steidel, C. C., & Siana, B. 2013, ApJ, 765, 47
Oesch, P. A., Carollo, C. M., Stiavelli, M., et al. 2009, ApJ, 690, 1350
Oesch, P. A., Stiavelli, M., Carollo, C. M., et al. 2007, ApJ, 671, 1212
Oke, J. B., & Gunn, J. E. 1983, ApJ, 266, 713
Osterbrock, D. E., & Ferland, G. J. 2006, Astrophysics of Gaseous Nebulae and Active Galactic Nuclei (Sausalito, CA: University Science Books)
Overzier, R. A., Heckman, T. M., Wang, J., et al. 2011, ApJL, 726, L7
Papovich, C., Dickinson, M., Ferguson, H. C., et al. 2004, ApJL, 600, L111
Papovich, C., Finkelstein, S. L., Ferguson, H. C., Lotz, J. M., & Giavalisco, M. 2011, MNRAS, 412, 1123
Pei, Y. C. 1992, ApJ, 395, 130
Reddy, N., Dickinson, M., Elbaz, D., et al. 2012, ApJ, 744, 154
Reddy, N. A., Erb, D. K., Pettini, M., Steidel, C. C., & Shapley, A. E. 2010, ApJ, 712, 1070
Reddy, N. A., Steidel, C. C., Pettini, M., et al. 2008, ApJS, 175, 48
Rhoads, J. E., Malhotra, S., Pirzkal, N., et al. 2009, ApJ, 697, 942
Santini, P., Fontana, A., Grazian, A., et al. 2009, A&A, 504, 751
Schaerer, D., & de Barros, S. 2009, A&A, 502, 423
Schaerer, D., & de Barros, S. 2010, A&A, 515, A73
Shim, H., Chary, R.-R., Dickinson, M., et al. 2011, ApJ, 738, 69
Siana, B., Smail, I., Swinbank, A. M., et al. 2009, ApJ, 698, 1273
Siana, B., Teplitz, H. I., Chary, R.-R., Colbert, J., & Frayer, D. T. 2008, ApJ, 689, 59
Stark, D. P., Ellis, R. S., Bunker, A., et al. 2009, ApJ, 697, 1493
Stark, D. P., Schenker, M. A., Ellis, R. S., et al. 2013, ApJ, 763, 129

Trenti, M., Bradley, L. D., Stiavelli, M., et al. 2011, ApJL, 727, L39
Vanzella, E., Giavalisco, M., Dickinson, M., et al. 2009, ApJ, 695, 1163
Verma, A., Lehnert, M. D., Förster Schreiber, N. M., Bremer, M. N., & Douglas, L. 2007, MNRAS, 377, 1024
Wilkins, S. M., Bunker, A. J., Stanway, E., Lorenzoni, S., & Caruana, J. 2011, MNRAS, 417, 717
Willott, C. J., McLure, R. J., Hibon, P., et al. 2013, AJ, 145, 4
Windhorst, R. A., Cohen, S., Hathi, N. P., et al. 2011, ApJS, 193, 27
Xue, Y. Q., Luo, B., Brandt, W. N., et al. 2011, ApJS, 195, 10
Yabe, K., Ohta, K., Iwata, I., et al. 2009, ApJ, 693, 507
Yan, H., Dickinson, M., Giavalisco, M., et al. 2006, ApJ, 651, 24
Yan, H., Yan, L., Zamojski, M. A., et al. 2011, ApJL, 728, L22