WHAT DO WE LEARN FROM IRAC OBSERVATIONS OF GALAXIES AT $2 < z < 3.5$?\(^1\)

**Stijn Wuyts,\(^2\)** Ivo Labbé,\(^3,4\)** Marin Franx,\(^5\)** Gregory Rudnick,\(^5\)** Peter G. van Dokkum,\(^6\)** Giovanni G. Fazio,\(^7\)** Natasha M. Förster Schreiber,\(^8\)** Jiasheng Huang,\(^7\)** Alan F. M. Moorwood,\(^9\)** Hans-Walter Rix,\(^10\)** Huub Röttgering,\(^2\)** and Paul van der Werf\(^5\)**

Received 2006 June 22; accepted 2006 September 15

**ABSTRACT**

We analyze very deep *HST*, VLT, and *Spitzer* photometry of galaxies at $2 < z < 3.5$ in the Hubble Deep Field–South. The sample is selected from the deepest public K-band imaging currently available. We show that the rest-frame $U - V$ versus $V - J$ color-color diagram is a powerful diagnostic of the stellar populations of distant galaxies. Galaxies with red rest-frame $U - V$ colors are generally red in rest-frame $V - J$ as well. However, at a given $U - V$ color a range in $V - J$ colors exists, and we show that this allows us to distinguish young, dusty galaxies from old, passively evolving galaxies. We quantify the effects of IRAC photometry on estimates of masses, ages, and the dust content of $z > 2$ galaxies. The estimated distributions of these properties do not change significantly when IRAC data are added to the $UBVJHK$ photometry. However, for individual galaxies the addition of IRAC can improve the constraints on the stellar populations, especially for red galaxies: uncertainties in stellar mass decrease by a factor of 2.7 for red $[(U - V)_{\text{rest}} > 1]$ galaxies, but only by a factor of 1.3 for blue $[(U - V)_{\text{rest}} < 1]$ galaxies. We find a similar color dependence of the improvement for estimates of age and dust extinction. In addition, the improvement from adding IRAC depends on the availability of full NIR $JHK$ coverage; if only $K$ band were available, the mass uncertainties of blue galaxies would decrease by a more substantial factor of 1.9. Finally, we find that a trend of galaxy color with stellar mass is already present at $z > 2$. The most massive galaxies at high redshift have red rest-frame $U - V$ colors compared to lower mass galaxies, even when allowing for complex star formation histories.

**Subject headings:** galaxies: evolution — galaxies: high-redshift — infrared: galaxies

**On-line material:** machine-readable tables

1. INTRODUCTION

Two of the major challenges in observational cosmology are understanding the history of star formation in galaxies and understanding the assembly of mass through cosmic time. In the local universe elaborate surveys mapped the diversity of nearby galaxies (e.g., Blanton et al. 2003) and characterized the dependence of their colors (Baldry et al. 2004) and star formation (Kauffmann et al. 2003). The study of their progenitors at $z \geq 2$ is important, since it is believed that at this epoch the most massive galaxies formed their stars (Glazebrook et al. 2004; van der Wel et al. 2005; Rudnick et al. 2006). The first method to efficiently identify distant galaxies was the Lyman-break technique (Steidel et al. 1999). Their stellar populations have been characterized by means of broadband photometry (e.g., Papovich et al. 2001; Shapley et al. 2005), optical spectroscopy (e.g., Shapley et al. 2003) and near-infrared (NIR) spectroscopy (Erb et al. 2003, 2006a, 2006b). Lyman-break galaxies (LBGs) have spectral energy distributions similar to nearby starburst galaxies.

In recent years new selection criteria provided evidence for a variety in color space among high-redshift galaxies as rich as in the local universe. Among the newly discovered populations are submillimeter galaxies (e.g., Smail et al. 2004), IRAC Extremely Red Objects (IEROs; Yan et al. 2004), BzK objects (Daddi et al. 2004), and distant red galaxies (DRGs; Franx et al. 2003). The latter are selected by the simple color criterion $(J - K)_{\text{Vega}} > 2.3$. Their rest-frame UV-to-optical spectral energy distributions (SEDs) resemble those of normal nearby galaxies of type Sbc–Sd (Förster Schreiber et al. 2004). NIR spectroscopy of DRGs (Kriek et al. 2006) and extension of the broadband photometry to mid-infrared wavelengths (Labbé et al. 2005) suggests that evolved stellar populations exist already at $2 < z < 3.5$. Rudnick et al. (2006) showed that DRGs contribute significantly to the mass density in rest-frame optically luminous galaxies. Van Dokkum et al. (2006) studied a stellar mass-limited sample of galaxies with $M > 10^{11} M_{\odot}$ and found that DRGs, rather than LBGs, are the dominant population at the high-mass end at $2 < z < 3$.

In this paper we exploit the 3–8 $\mu$m imaging of the Hubble Deep Field–South by *Spitzer’s* Infrared Array Camera (IRAC; Fazio et al. 2004) to extend the SED analysis of distant galaxies to the rest-frame NIR and constrain their stellar masses and stellar populations. Our sample is complete to $K_{\text{tot, AB}} = 25$. No color selection criteria are applied. The depth of our imaging allows us to probe down to stellar masses of a few $10^9 M_{\odot}$. We investigate whether IRAC helps to study the diversity of galaxies at high redshift and if the addition of IRAC improves the constraints on
stellar mass, age, and dust content. Finally, we investigate the dependence of galaxy color on stellar mass.

The paper is structured as follows. In § 2 we describe the data, IRAC photometry, and sample definition. Section 3 explains the modeling of the SEDs. The rest-frame optical to NIR color distribution of our K-selected sample is discussed in § 4. Section 5 provides an in-depth discussion of the constraints that IRAC places on estimates of age, dust extinction, and stellar mass. First wavelength and model dependence are discussed from a theoretical perspective. Next we discuss results from applying the models to our rest-frame optical to NIR colors as a function of stellar mass. Finally, the conclusions are summarized in § 7. Throughout the paper we adopt a cosmology with $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_m = 0.3$, and $\Omega_{\Lambda} = 0.7$.

2. DATA, PHOTOMETRY, AND SAMPLE SELECTION

2.1. Data

Observations of the HDF-S WFPC2 field were obtained with IRAC (Fazio et al. 2004) on the Spitzer Space Telescope (Werner et al. 2004) in 2004 June and 2005 June (GTO program 214). A $5' \times 5'$ field of view was covered by the four broadband filters at 3.6, 4.5, 5.8, and 8 $\mu$m. The data, reduction, and photometry will be described in detail by I. Labbé et al. (2006, in preparation). Briefly, we started with the basic calibrated data (BCD) as provided by the Spitzer Science Center pipeline. We applied a series of procedures to reject cosmic rays and remove artifacts such as column pulldown, muxbleed, and the “first-frame effect” (Hora et al. 2004). Finally, the frames were registered to and projected on a $2 \times 2$ blocked (0.2396 $\text{pixel}$ scale) version of an existing ISAAC K-band image (Labbé et al. 2003; hereafter L03), and average-combined. Characteristics such as exposure time, FWHM, limiting depth (5 $\sigma$, 3$''$ diameter aperture), and positional accuracy in each of the four IRAC bands are summarized in Table 1. A summary of the optical-to-NIR observations by L03 is provided in Table 2. All magnitudes quoted in this paper are in the AB system.

2.2. Photometry

In this section we describe the steps taken to combine the IRAC data and optical-to-NIR data (L03) into one consistent K-band-selected photometric catalog. In this paper we limit ourselves to the $2.5' \times 2.5'$ field, for which very deep K-band data are available from L03. The main challenge in doing IRAC photometry is the proper treatment of source confusion and PSF matching of the data. Integrating for nearly 4 hr with IRAC at 3.6 and 4.5 $\mu$m reaches a depth only 1 mag shallower than 36 hr of ISAAC K-band imaging (10 $\sigma$ limit, $K_{\text{tot,AB}} = 25$), but the IRAC images have a PSF that is 4 times broader, causing many sources to be blended. Information on the position and extent of K-band-detected objects was used to fit and subtract the fluxes of neighboring sources. Each K-band-detected source was isolated using the SExtractor “segmentation map” and convolved individually to the considered IRAC PSF. Next, all convolved sources were fitted to the IRAC image, leaving only their fluxes as free parameters. We subsequently subtract the best-fit fluxes of all neighboring sources to remove the contamination. An illustration of this measurement technique is presented in Figure 1. The resulting cleaned IRAC images are matched to the broadest PSF (of the 8 $\mu$m image). We measured fluxes on the cleaned, PSF-matched images within a fixed 4.4$''$ diameter circular aperture. The aperture size is a compromise between quality of PSF matching (within 3% as derived from dividing growth curves) and adding too much noise. Finally, we applied for each source an aperture correction to scale the IRAC fluxes to the “color” apertures defined for the K-band catalog by L03. The correction factor is the ratio of the original K-band flux in the color aperture and the K-band flux in the 8 $\mu$m PSF-matched image within a 4.4$''$ diameter aperture. Photometric errors were calculated by taking the rms of fluxes in 4.4$''$ diameter apertures on empty places in the IRAC image. The end product is a photometric catalog with

---

## TABLE 1

| Instrument/Telescope | Filter   | Exposure Time (hr) | FWHM (arcsec) | Limiting Depth (5 $\sigma$, 3$''$ diameter aperture) | Positional Accuracy$^a$ (arcsec) |
|----------------------|----------|--------------------|---------------|--------------------------------------------------|---------------------------------|
| WFPC2/HST            | F300W    | 36.8               | 0.16          | 27.8                                            |                                 |
| WFPC2/HST            | F450W    | 28.3               | 0.14          | 28.6                                            |                                 |
| WFPC2/HST            | F606W    | 27.0               | 0.13          | 28.9                                            |                                 |
| WFPC2/HST            | F814W    | 31.2               | 0.14          | 28.3                                            |                                 |
| ISAAC/VLT            | J$_c$    | 33.6               | 0.45          | 26.9                                            |                                 |
| ISAAC/VLT            | H        | 32.3               | 0.48          | 26.4                                            |                                 |
| ISAAC/VLT            | K$_s$    | 35.6               | 0.46          | 26.4                                            |                                 |

---

$^a$ The rms difference between bright star positions in IRAC and K-band image.

## TABLE 2

| Instrument/Telescope | Filter | Exposure Time (hr) | FWHM (arcsec) | Limiting Depth (5 $\sigma$, 07 diameter aperture) |
|----------------------|--------|--------------------|---------------|--------------------------------------------------|
| WFPC2/HST            | F300W  | 36.8               | 0.16          | 27.8                                            |
| WFPC2/HST            | F450W  | 28.3               | 0.14          | 28.6                                            |
| WFPC2/HST            | F606W  | 27.0               | 0.13          | 28.9                                            |
| WFPC2/HST            | F814W  | 31.2               | 0.14          | 28.3                                            |
| ISAAC/VLT            | J$_c$  | 33.6               | 0.45          | 26.9                                            |
| ISAAC/VLT            | H      | 32.3               | 0.48          | 26.4                                            |
| ISAAC/VLT            | K$_s$  | 35.6               | 0.46          | 26.4                                            |

---

11 NIR data from the FIRS survey of the HDF-S is publicly available from http://www.strw.leidenuniv.nl/~firs.
consistent photometry from optical to MIR wavelengths with 11 filters (UBVRIJK+IRAC).

2.3. Sample Selection

From the catalog described in § 2.2 we selected all galaxies, well covered by all 11 filters, that have signal-to-noise ratio \((S/N) > 10\) in the \(K\) band. The sample reaches to a limiting total \(K\)-band magnitude \(K_{\text{tot,AB}} = 25\).

Since spectroscopic redshifts are only available for 63 out of 274 objects, we mostly rely on photometric redshift estimates to select high-redshift galaxies and compute rest-frame colors and luminosities. The photometric redshifts and derived rest-frame photometry were calculated as follows. We used an algorithm developed by Rudnick et al. (2001, 2003) to fit a nonnegative linear combination of galaxy templates to the SED of each galaxy. The template set consisted of empirical E, Sbc, Scd, and Im templates from Coleman et al. (1980), the two least reddened starburst templates from Kinney et al. (1996) and two Bruzual & Charlot (2003; hereafter BC03) single stellar populations (SSPs) with a Salpeter (1955) stellar initial mass function (IMF), aged 1 Gyr and 10 Myr, respectively. The empirical templates were extended into the IR using the BC03 stellar population synthesis code. The derived photometric redshifts show a good agreement with the available spectroscopic redshifts. The average value of \(\frac{z_{\text{spec}} - z_{\text{phot}}}{1 + z_{\text{spec}}}\) is 0.06, 0.09, and 0.08 for galaxies at \(0 < z < 1\), \(1 < z < 2\), and \(2 < z < 3.5\), respectively.

Once the redshift was derived, we calculated rest-frame luminosities and colors by interpolating between observed bands using the best-fit templates as a guide. For a detailed description, we refer to Rudnick et al. (2003).

The \(K\)-band–selected sample contains 121 sources at \(0 < z < 1\), 72 at \(1 < z < 2\), and 75 at \(2 < z < 3.5\). The \(K\)+IRAC photometry of the galaxies at \(2 < z < 3.5\) is provided in Table 3.

In § 4 we study the color-distribution of galaxies with \(L_{V} > 5 \times 10^{9} L_{\odot}\) over the whole redshift range. From that point on we focus on the high-redshift bin. Two commonly color-selected populations at \(z > 2\) are highlighted where they are of interest. LBGs are selected from the WFPC2 imaging using the criteria of Madau et al. (1996). DRGs are selected by the simple color criterion \((J - K)_{\text{AB}} > 1.34\) (Franx et al. 2003).

3. SED MODELING

To study physical characteristics of the galaxies such as stellar mass, stellar age, and amount of dust extinction, we make use of the evolutionary synthesis code developed by BC03. We fitted the synthetic spectra to our observed SEDs using the publicly available HYPERZ stellar population–fitting code, version 1.1 (Bolzonella et al. 2000). Redshifts were fixed to the \(z_{\text{phot}}\) measurement (see § 2.3; Rudnick et al. 2003) or \(z_{\text{spec}}\) when available. A minimum error of 0.08 mag was adopted to avoid the problem of data points with the largest errors being effectively ignored in the SED fits. We fitted three distinct star formation histories: a single stellar population (SSP) without dust, a constant star formation (CSF) history with dust (\(A_{V}\) varying from 0 to 4 in steps of 0.2), and an exponentially declining star formation history with an \(e\)-folding timescale of 300 Myr (\(\tau_{300}\)) and identical range of \(A_{V}\) values. The exponentially declining model allows for quiescent systems that underwent a period of enhanced star formation in their past. Förster Schreiber et al. (2004) showed that the estimated extinction values do not vary monotonically with the \(e\)-folding timescale \(\tau\), but reach a minimum around 300 Myr. Including the \(\tau_{300}\) model thus ensures that the allowed star formation histories encompass the whole region of parameter space that would be occupied when fitting models with different values of \(\tau\).

For each of the star-formation histories (SFHs), we constrained the time elapsed since the onset of star formation to a minimum of...
50 Myr, avoiding fit results with improbably young ages. The age of the universe at the observed redshift was set as an upper limit to the ages. Furthermore, we assume a Salpeter (1955) IMF with lower and upper mass cutoffs 0.1 and 100 $M_\odot$, and solar metallicity, and we adopt a Calzetti et al. (2000) extinction law. For each object the SFH resulting in the lowest $\chi^2$ of the fit was selected and corresponding model quantities such as age, mass, and dust extinction were adopted as the best-fit values. We calculated the mass-weighted age for each galaxy by integrating over the different ages of SSPs that build up the SFH, weighting with their mass fraction. We use this measure because it is more robust with respect to degeneracies in SFH than the time passed since the onset of star formation; it describes the age of the bulk of the stars. See Table 4 for a summary of the results of our SED modeling for the subsample of galaxies at $2 < z < 3.5$. In Figure 2 we show examples of $U$-to-8 $\mu$m SEDs with best-fit BC03 models of galaxies over the whole redshift range, illustrating that at all epochs a large variety of galaxy types is present.

We fitted all objects in our sample twice, once with and once without IRAC photometry. We repeated the SED modeling with the same parameter settings using the models by Maraston (2005; hereafter M05). The results are discussed in § 5.2.2. Variations in modeled parameters due to a different metallicity are addressed in § 5.2.3. The effects of adopting a different extinction law are discussed in § 5.2.4. Unless noted otherwise, we refer to stellar mass, mass-weighted age, and dust extinction values derived from the $U$-to-8 $\mu$m SEDs with BC03 models.

### 4. REST-FRAME OPTICAL TO NEAR-INFRARED COLOR DISTRIBUTION

At redshifts above 1 all rest-frame NIR bands have shifted redward of observed $K$, and mid-infrared photometry is needed to compute rest-frame NIR fluxes from interpolation between observed bands. It has only been with the advent of IRAC on the Spitzer Space Telescope that the rest-frame NIR opened up for the study of high-redshift galaxies. As the 3.6 and 4.5 $\mu$m images are much deeper than the 5.8 and 8.0 $\mu$m images (see Table 1), we focus on the rest-frame $J$ band ($J_{\text{rest}}$).

Several studies have focussed on the optical to NIR colors and inferred stellar populations of particular color-selected samples (e.g., Shapley et al. 2005; Labbé et al. 2005). In this section we take advantage of the multiband data and the very deep $K$-band selection to study the rest-frame optical to NIR colors of all galaxies up to $z = 3.5$ without color bias. For the first time we can therefore investigate what range in optical to NIR colors high-redshift galaxies occupy, how their optical to NIR colors relate to pure optical colors, and what this tells us about the nature of their stellar populations.

In Figure 3 we present a color-color diagram of ($U$ – $V$)$_{\text{rest}}$ versus ($V$ – $J$)$_{\text{rest}}$ for the redshift bins $0 < z < 1$, $1 < z < 2$, and $2 < z < 3.5$. A clear correlation of ($U$ – $V$)$_{\text{rest}}$ with ($V$ – $J$)$_{\text{rest}}$ is observed at all redshifts. The ($U$ – $V$)$_{\text{rest}}$ color samples the Balmer/4000 Å break. The large wavelength range spanned by ($U$ – $V$)$_{\text{rest}}$ and ($V$ – $J$)$_{\text{rest}}$ together is useful to probe reddening by dust.

To study how the color distribution compares to that in the local universe, we indicate the colors of galaxies in the low-redshift New York University Value-Added Galaxy Catalog (NYU_VAGC; Blanton et al. 2005) with small gray dots. The low-z NYU_VAGC is a sample of nearly 50,000 galaxies at 0.033 < $z$ < 0.05 extracted from the Sloan Digital Sky Survey (SDSS data release 4; Adelman-McCarthy et al. 2006). It is designed to serve as a reliable reference for the local galaxy population and contains matches to the Two Micron All Sky Survey Point Source Catalog and Extended Source Catalog (2MASS; Cutri et al. 2000). Only the subsample of 20,180 sources that are detected in the 2MASS $J$-band are plotted in Figure 3. This results effectively in a reduction of the blue peak of the bimodal $U$ – $V$ distribution. We only show those galaxies (both for the local sample
and for our sample of HDF-S galaxies) with a rest-frame $V$-band luminosity $L_V > 5 \times 10^9 L_\odot$. At this luminosity the distribution of low-$z$ NYU_VAGC galaxies with SDSS and 2MASS detections starts falling off. From the much deeper HDF-S imaging the luminosity cut weeds out low- to intermediate-redshift dwarf galaxies.

The same trend of optically red galaxies being red in optical to NIR wavelengths that we found for galaxies up to $z = 3$: is observed in the local universe. However, there are two notable differences in the color distribution between distant and local galaxies. First, a population of luminous high-redshift galaxies with very blue $U$-$V$ colors ($U$-$V$ rest) exists without an abundant counterpart in the local universe. The 2MASS observations are not deep enough to probe very blue $V$-$J$ colors, but we can ascertain that $95\%$ of all low-$z$ NYU_VAGC sources with $L_V > 5 \times 10^9 L_\odot$ lie in the range $0.73 < U - V < 2.24$. About half of the blue galaxies at $z > 2$ with $(U - V)_{\text{rest}} < 0.73$ and $L_V > 5 \times 10^9 L_\odot$ satisfy the Lyman-break criterion. Their stellar populations have been extensively studied (e.g., Papovich et al. 2001; Shapley et al. 2001; among many others) and their blue SEDs (see, e.g., objects 242 and 807 in Fig. 2) are found to be well described by relatively unobscured star formation. The rest-frame optical bluing with increasing redshift of galaxies down to a fixed $L_V$ is thoroughly discussed by Rudnick et al. (2003).

A second notable difference with respect to the color distribution of nearby galaxies is present at $(U - V)_{\text{rest}} > 1$, where most local galaxies reside. Our sample of HDF-S galaxies has a median offset with respect to the SDSS+2MASS galaxies of $0.22 \pm 0.04$ mag toward redder $(V - J)_{\text{rest}}$ at a given $(U - V)_{\text{rest}}$.  

Fig. 2.—Rest-frame $U$-$V$ vs. $V$-$J$ color-color diagram of all galaxies with $L_V > 5 \times 10^9 L_\odot$. SDSS+2MASS galaxies (small gray dots) are plotted as a local reference. Color coding refers to the redshift bin. Galaxies with red $U$-$V$ colors are also red in $V$-$J$. Compared to the local SDSS galaxies the high-redshift color distribution extends to bluer $U$-$V$ colors (where Lyman-break galaxies are located) and for the same $U$-$V$ color to redder $V$-$J$ colors. 

Fig. 3.—Rest-frame $U$-$V$ vs. $V$-$J$ color-color diagram of all galaxies with $L_V > 5 \times 10^9 L_\odot$. SDSS+2MASS galaxies (small gray dots) are plotted as a local reference. Color coding refers to the redshift bin. Galaxies with red $U$-$V$ colors are also red in $V$-$J$. Compared to the local SDSS galaxies the high-redshift color distribution extends to bluer $U$-$V$ colors (where Lyman-break galaxies are located) and for the same $U$-$V$ color to redder $V$-$J$ colors.
axes that are red in and slightly redder best-fit mass-weighted age. The tracks show an increase toward redder optical colors are indeed found to be young, the median with age. The spread in colors similar to that of local galaxies to colors up to a magnitude redder. The larger spread in colors at a given (U - V)rest is not caused by photometric uncertainties. After subtraction in quadrature of the scatter expected from measurement errors (0.05 mag), we obtain an intrinsic scatter of 0.3 mag, significantly larger than that for SDSS+2MASS galaxies (0.19 mag) at a 4.5 σ level.

In order to understand the nature of galaxies with similar or redder (V - J)rest than the bulk of nearby galaxies, we make use of stellar population synthesis models by BC03. In Figure 4 we draw age tracks for three different dust-free star formation histories in the (U - V)rest versus (V - J)rest color-color diagram. The solid line represents a single stellar population (SSP), the dashed line a continuous star formation model (CSF), and the dotted line an exponentially declining star formation model with an e-folding timescale of 300 Myr (τ300). All star formation histories show an evolution to redder (U - V)rest and (V - J)rest with age. The τ300 model first has similar colors as a CSF model and eventually moves to the same region in color space as an evolved SSP, namely, to where the red peak of the SDSS bimodal U - V distribution is located. In the absence of dust a population with a constant star formation history only reaches U - V = 1 in a Hubble time.

We now investigate how the location in this color plane is related to stellar populations. Using the best-fit model parameters (see § 3) we plot the mass-weighted ages for the galaxies with L_U > 5 x 10^9 L⊙ with color-coding on Figure 4. Galaxies with blue optical colors are indeed found to be young, the median mass-weighted age for galaxies at (U - V)rest < 1 being 250 Myr. At (U - V)rest > 1 galaxies with a wide range of stellar ages are found. The oldest stellar populations show the blueest (V - J)rest colors at a given (U - V)rest. Over the whole redshift range galaxies are present that have red optical colors and whose SEDs are consistent with evolved stellar populations and low dust content. According to their best-fit model, three of them started forming stars less than 0.5 Gyr after the big bang and already at z > 2.5 have star formation rates less than a percent of the past-averaged value. We note that in the Chandra Deep Field-South, Papovich et al. (2006) find a number density of passively evolving galaxies at high redshift that is nearly an order of magnitude lower than in the HDF-S, possibly owing to the fact that the HDF-S observations probe to fainter K-band magnitudes. The red (V - J)rest side of the color distribution is made up of galaxies that are best fitted by young stellar populations. Since the age tracks alone cannot explain the presence of galaxies with such red SEDs from the optical throughout the NIR, we investigate the role of dust in shaping the galaxy color distribution.

Figure 5 shows again the (U - V)rest versus (V - J)rest color-color diagram, now color-coded by best-fit dust extinction, expressed in A_V. The arrow indicates an A_V of 1 mag using a Calzetti et al. (2000) extinction law. It is immediately apparent that the optical to NIR color-color diagram is a useful diagnostic for distinguishing stellar populations with various amounts of dust extinction. At the bluest (U - V)rest colors there is little evidence for dust obscuration. The degree of dust extinction increases as we move along the dust vector to redder colors.

Independent constraints on dust-enshrouded activity in distant galaxies can be derived from MIPS 24 μm imaging (Webb et al. 2006; Papovich et al. 2006). The mid-infrared emission is usually thought to be powered by a dusty starburst in which polycyclic aromatic hydrocarbon (PAH) features are produced or by an active galactic nucleus (AGN). Of the area with very deep U-to-8 μm in the HDF-S 95% is covered by a 1 hr MIPS pointing. We performed the same photometric procedure to reduce confusion as for the IRAC photometry (see § 2.2). Fluxes were measured...
Within a 6′ diameter aperture and then scaled to total using the growth curve of the 24 μm PSF.

In Figure 6 we plot the $(U - V)_{\text{rest}}$ versus $(V - J)_{\text{rest}}$ color-color diagram of all objects in the redshift interval $1.5 < z < 3$ with $L_V > 5 \times 10^9 L_\odot$ that are covered by MIPS (empty circles). At these redshifts, strong PAH features, if present, move through the MIPS 24 μm passband. Six sources have a MIPS 24 μm detection above 28 μJy (3σ). Their 24 μm flux is indicated by the colored filled circles. Object 767 (red circle) is well detected with $F_{24 \mu m} = 95$ μJy. As noted by Labbé et al. (2005) its SED shows an 8 μm excess with respect to the best-fitting template. The combination of 8 μm excess and 24 μm detection suggests that this galaxy hosts an AGN whose power-law SED dominates throughout the mid-infrared. All other 24 μm detections are located in the part of the diagram where our U-to-8 μm SED modeling found dusty young populations. None of the blue relatively unobscured star-forming galaxies or red evolved galaxies show evidence of PAH emission from the observed 24 μm flux. There are various reasons why not all star-forming dusty galaxies have a 24 μm detection. The density of the UV radiation field exciting the PAHs may vary among galaxies. Furthermore, the narrow PAH features with respect to the width of the 24 μm passband make the 24 μm flux very sensitive to redshift. Overall, MIPS observations agree well with SED modeling and rest-frame optical-NIR color characterization.

We conclude that over the whole redshift range from $z = 0$ to 3.5 a trend is visible of galaxies with redder optical colors showing redder optical to NIR colors. However, at a given optical color, a spread in optical to NIR colors is observed that is larger than for nearby galaxies. At $(U - V)_{\text{rest}} > 1$ evolved galaxies with little dust extinction are found at the bluest $(V - J)_{\text{rest}}$. Dusty young star-forming galaxies occupy the reddest $(V - J)_{\text{rest}}$ colors. This is once more illustrated by the SEDs of galaxies with $(U - V)_{\text{rest}} > 1$ presented in Figure 7. The top row shows SEDs of objects at the blue side of the $(V - J)_{\text{rest}}$ color distribution. The bottom panels show SEDs of galaxies matched in $(U - V)_{\text{rest}}$ but with comparatively redder $(V - J)_{\text{rest}}$ colors. The latter galaxies have comparatively younger ages and a larger dust content. Since this distinction could not be made on the basis of $(U - V)_{\text{rest}}$ color alone, the addition of IRAC 3.6–8 μm photometry to our $U$-to-$K$ SEDs proves very valuable for the understanding of stellar populations at high redshift.

We verified that no substantial changes occur to the rest-frame optical-to-NIR color distribution and its interpretation in terms of age and dust content of the galaxies when we derive photometric redshifts by running HYPERZ with redshift as free parameter instead of using the algorithm developed by Rudnick et al. (2003; see § 2.3).

5. Constraints on Stellar Population Properties at $2 < z < 3.5$: Age, Dust, and Mass

We now proceed to analyze in more detail the constraints that IRAC places on the stellar populations of the subsample of galaxies at $2 < z < 3.5$ (75 galaxies). In particular we will focus on stellar mass, which likely plays a key role in galaxy evolution at all redshifts (e.g., Kaufmann et al. 2003; Bundy et al. 2005; Drory et al. 2005; Rudnick et al. 2006). Fortunately, estimates of stellar mass from modeling the broadband SEDs are generally more robust than estimates of dust content and stellar age (Bell & de Jong 2001; Shapley et al. 2001; Papovich et al. 2001; Förster Schreiber et al. 2004). Nevertheless, translating colors to mass-to-light ratios and subsequently stellar masses requires a good understanding of the effects of age and dust.

5.1. Predictions from Stellar Population Synthesis Models

5.1.1. Wavelength Dependence: Optical versus Near-Infrared

In its simplest form the stellar mass of a galaxy can be estimated from one color (see, e.g., Bell & de Jong 2001). To illustrate this process, we present the evolutionary track of a dustfree BC03 model in a $M/L_V$ versus $U - V$ diagram (Fig. 8). Up to 2.5 Gyr after the onset of star formation (Fig. 8, top right corner) the track represents a one-component population with a star formation history that is exponentially declining, with an e-folding timescale of 300 Myr. For most of the galaxies in our sample this was the best-fitting star formation history. For more extreme star formation histories, such as an SSP or CSF, the process of estimating $M/L$ values follows similar arguments. The filled circles on Figure 8 represent age steps of 100 Myr. As the stellar population ages, its $V$-band luminosity fades with only a small decrease in stellar mass from mass loss, moving the galaxy up in $M/L_V$. Simultaneously, the $U - V$ color reddens as the hot early-type stars with short lifetimes die. The dust vector indicating a reddening of $A_V = 1$ mag runs parallel to the age track of the one-component model. Ironically, the mass estimate benefits greatly from this degeneracy between age and dust in the optical. Under the assumption of a monotonic star formation history $(U - V)_{\text{rest}}$ can uniquely be translated to $M/L_v$, regardless of the precise role of dust or age. Only a normalization with $L_V$ is needed to derive the stellar mass. A similar relation was used by Rudnick et al. (2003) to translate the integrated $(U - V)_{\text{rest}}$ color of high-redshift galaxies into a global $M/L_V$ and stellar mass density $\rho_s$. They found that the conversion to mass-to-light ratio is more robust from the $(U - V)_{\text{rest}}$ color than from the $(U - B)_{\text{rest}}$ or $(B - V)_{\text{rest}}$ color.

What if the actual star formation history is more complex? What effect does it have on the derived stellar mass? There is
ample evidence from local fossil records (e.g., Trager et al. 2000; Lancon et al. 2001; Freeman & Bland-Hawthorn 2002; Förster Schreiber et al. 2003; Angeretti et al. 2005) and high-redshift studies (e.g., Papovich et al. 2001, 2005; Ferguson et al. 2002) that galaxies of various types have complex and diverse star formation histories, often with multiple or recurrent episodes of intense star formation. Such a scenario is also predicted by cold dark matter models (e.g., Somerville et al. 2001; Nagamine et al. 2005; De Lucia et al. 2006). In order to address this question qualitatively, we consider the case of a two-component population. At $t = 2.5$ Gyr we added a burst of star formation to the $C28$ model, lasting 100 Myr and contributing 20% to the mass. To follow the evolution of the two-component population closely, we mark 10 Myr time steps with open circles. Over a time span of only 10 Myr the galaxy color shifts by 1.6 mag toward the blue, while the $M/L_V$ value stays well above the one-component corresponding to that color. As the newly formed stars grow older, the galaxy moves toward the top right corner of the diagram again. The offset of $M/L_V$ with respect to the one-component model is a decreasing function of $U/V$. This means that if a bursty star formation is mistakenly fitted with a one-component model, the mass and mass-to-light ratio are underestimated more for blue than for red galaxies, confirming what Shapley et al. (2005) found for a sample of star-forming galaxies at $z > 2$.

The histogram at the bottom of Figure 8 indicates the $(U-V)_{\text{rest}}$ color distribution of galaxies in the HDF-S at $z > 2$. The population of LBGs is marked in blue, DRGs in red. The possible underestimate in mass-to-light ratio, and thus mass, is largest for blue galaxies, up to a factor of 3 for $(U-V)_{\text{rest}} = 0.2$, the bluest color reached by this two-component model. For

**Fig. 7.—** Comparison of galaxies with similar $(U-V)_{\text{rest}}$ color but different $(V-J)_{\text{rest}}$ color. The top row shows the galaxies with blue $(V-J)_{\text{rest}}$ colors, and the bottom row shows galaxies with matching $(U-V)_{\text{rest}}$ color but much redder $(V-J)_{\text{rest}}$ color. The systematic difference in the SEDs of the two rows is striking. Fits indicate old bursts of star formation with little dust in the top row, and dusty young galaxies in the bottom row. This demonstrates the power of $(V-J)_{\text{rest}}$ in separating these classes. Note that the $U$-band photometry for objects 224 and 249 deviates by more than $2\sigma$ from the predicted $U$-band flux of the best-fit template.

**Fig. 8.—** Evolutionary track of a two-component stellar population in the $M/L_V$ vs. $U-V$ plane. Filled circles mark age steps of 100 Myr. Open circles represent 10 Myr age steps. The dust vector, indicating an extinction of $A_V = 1$ mag, lies parallel to the age track. The histogram represents the color distribution of galaxies at $2 < z < 3.5$, with DRGs highlighted in red, and LBGs in blue. The age track starts as an exponentially declining star formation model ($\tau = 300$ Myr; BC03). At 2.5 Gyr a new burst of star formation is introduced, lasting 100 Myr and contributing 20% to the mass. Translating $U-V$ into $M/L_V$ assuming one-component models can lead to underestimates of $M/L_V$ and thus stellar mass. The possible underestimate is largest for blue galaxies.
DRGs only a modest amount of mass can be hidden under the glare of a young burst of star formation. The exact error that bursts cause depends on the form of the bursty star formation history. The loop toward bluer colors is highly uncertain for the blue galaxies in our sample, and the additional IRAC observations do not improve the constraints on stellar mass-to-light ratio. The situation is further complicated by the effect of dust. $V - J$ is a lot more sensitive to dust than $M/L_j$, illustrated by the dust vector of $AV = 1$ mag. The effects of dust and age no longer conspire to give robust mass estimates at a given $V - J$ color. At redder $V - J$ the situation improves as the slope of the age track flattens. Here the inclusion of a rest-frame NIR color clearly reduces the uncertainty in stellar $M/L$ that stems from the poor knowledge of the star formation history. The loop toward bluer colors is a magnitude smaller in size and we see no large offsets in $M/L$ between the one- and two-component modeling.

We have discussed the different behavior of dust and age in simplified one- and two-component models, and have investigated the improvements expected from the inclusion of the rest-frame NIR with respect to the rest-frame optical. While additional rest-frame NIR data can lead to better $M/L$ estimates, in particular for redder galaxies ($U - V > 1; V - J > 0.4$), it is clear that we need to take advantage of the full $U$ to $8 \mu m$ SED information to derive reliable estimates of stellar mass, stellar age, and dust content.

5.1.2. Model Dependence: Bruzual & Charlot versus Maraston

It is important to note that different stellar population synthesis models do not paint a consistent picture of evolution in the rest-frame NIR. To illustrate, we compare BC03 models to M05 models under the same assumption of Salpeter initial mass function and solar metallicity.

Whereas the age track in a $M/L_V$ versus $U - V$ diagram behaves similarly for M05 and BC03, the NIR evolution of a $\tau_{300}$ model looks very different (see Fig. 10). The gray dashed line represents the age track of a BC03 $\tau_{300}$ model with superposed burst at 2.5 Gyr, as described in § 5.1.1. In black we overplot the age track of a two-component model with identical parameters by M05. In the 0.2–2 Gyr age range the two models look strikingly different. At the same $V - J$ color the M05 model predicts $M/L_j$ values that are up to a factor 2.5 smaller than those of the BC03 model. The offset between $M/L_j$ as predicted from...
one- and two-component modeling is also larger by a similar factor. The BC03 and M05 models differ in several aspects: the stellar evolutionary tracks adopted to construct the isochrones, the synthesis technique and the treatment of the thermally pulsating asymptotic giant branch (TP-AGB) phase. The Padova stellar tracks (Fagotto et al. 1994) used by BC03 include a certain amount of convective-core overshooting, whereas the Frascati tracks (Cassisi et al. 1997) do not. The two stellar evolutionary models also differ for the temperature distribution of the red giant branch phase. The higher NIR luminosity originates mainly from a different implementation of the TP-AGB phase (M05). Following the fuel consumption approach, M05 finds that this phase in stellar evolution has a substantial impact on the NIR luminosity at ages between 0.2 and 2 Gyr. BC03 follow the isochrone synthesis approach, characterizing properties of the stellar population per mass bin. The latter method leads to smaller luminosity contributions by TP-AGB stars. We refer to recent studies from M05, van der Wel et al. (2006), and Maraston et al. (2006) for discussions of the model differences in greater detail.

For our purpose it is sufficient to state that a given $V - J$ color corresponds to younger ages, lower mass-to-light ratios and thus lower masses for the M05 model than for the BC03 model. Most importantly, we note that for M05 models inclusion of NIR data does not reduce stellar mass uncertainties caused by the unknown star formation history.

5.2. Constraints on Mass, Dust and Age from Modeling Our Observed Galaxies

5.2.1. Wavelength Dependence: Optical versus Near-Infrared

Having investigated the qualitative relationship between $M/L$ and the rest-frame optical-to-NIR color in § 5.1.1, we now quantify the effect of inclusion of IRAC MIR photometry on the stellar population constraints of galaxies at $2 < z < 3.5$. Our goal is to investigate whether and how the addition of IRAC imaging changes our best estimate of the stellar population properties and their confidence intervals.

First we compare the distribution of stellar mass, dust content and mass-weighted stellar age as fit with or without IRAC. The top row of Figure 11 shows a direct comparison of the inferred model parameters with or without IRAC photometry for all galaxies at $2 < z < 3.5$. The filled histograms in the bottom row of Figure 11 show the distribution of mass, dust extinction, and age derived from the full $U$-to-$K$ photometry. Both the median and the width of the distribution stays the same for all three parameters. Defining the difference between mass, mass-weighted age, and $A_V$ as $\Delta \log(M) = \log(M_{\text{withIRAC}}) - \log(M_{\text{noIRAC}})$, $\Delta \log(a_{w}) = \log(a_{w,\text{withIRAC}}) - \log(a_{w,\text{noIRAC}})$, and $\Delta \log(age_{w}) = \log(age_{w,\text{withIRAC}}) - \log(age_{w,\text{noIRAC}})$, we find a median and normalized median absolute deviation (equal to the rms for a Gaussian distribution) $\hat{x}$, $\frac{\text{NMAD}(x)}{\sigma}$ of $(-0.007 \pm 0.009, 0.07)$, $(0.00 \pm 0.03, 0.30)$, and $(0.00 \pm 0.02, 0.16)$, respectively. The average and standard deviation $\langle x \rangle$ of $\Delta \log(M)$, $\Delta \log(a_{w})$, and $\Delta \log(age_{w})$ are $(-0.04 \pm 0.02; 0.13)$, $(-0.08 \pm 0.04; 0.36)$, and $(-0.02 \pm 0.03; 0.28)$, respectively. Thus, the differences for the galaxy sample as a whole after including IRAC are very small. The results for stellar mass are similar to what Shapley et al. (2005) found for a more specific sample of optically selected star-forming galaxies at $z \sim 2$.

Having determined that the overall distribution of best-fit age, dust content, and stellar mass does not change after including IRAC, the question remains whether IRAC helps improve the constraints on the stellar population characteristics for individual galaxies. We address this question using the measure $\sigma_{\text{noIRAC}}/\sigma_{\text{withIRAC}}$, defined as the ratio of confidence intervals without and with IRAC. The 1 $\sigma$ confidence intervals, representing random uncertainties propagating from photometric errors, are derived...
from Monte Carlo simulations. For each galaxy SED we create 100 mock SEDs for which the flux-point in each band is randomly drawn from a Gaussian with the measured flux as the mean and its error as the standard deviation. Next, each SED was fitted with the same fitting procedure as the observed version. As we want to isolate the effect of including IRAC observations on the confidence intervals, we fix the redshift to want to isolate the effect of including IRAC observations on the with the same fitting procedure as the observed version. As we simulate this effect by omitting J and H and repeating the Monte Carlo simulations with and without IRAC. The median reduction of the 1σ mass confidence interval now increases to a factor 1.9 when including IRAC.

We conclude that in the presence of very deep observed J, H, and K photometry, inclusion of mid-infrared data places few extra constraints on the stellar populations of blue galaxies. However, for galaxies redder than (U − V)_{rest} = 1, IRAC reduces the confidence interval by a substantial factor of 2.5–3.

5.2.2. Model Dependence: Bruzual & Charlot versus Maraston

In § 5.1.2 we pointed out strong differences in the rest-frame optical-to-NIR colors between the BC03 and M05 models. In this paragraph we quantify how our results change if we use M05 models. The median and normalized median absolute deviation, average, and standard deviation of the differences between BC03 fits and M05 fits (Δ log(M_{BC03M05}), Δ A_{V,BC03M05}, and Δ log(age_{w,BC03M05}) are summarized in Table 5. As expected, the BC03 models predict older ages and thus higher stellar masses than the M05 models for our z = 2–3.5 galaxies. The estimated mass for the M05 models is systematically lower by a factor of 1.4. Maraston et al. (2006) found a similar discrepancy for a sample of 7 galaxies in the Hubble Ultra Deep Field that satisfy the BzK criterion (Daddi et al. 2004) for z > 1.4 passively evolving galaxies. Apart from a systematic shift a scatter of 0.1 in dex is found in Δ log(M_{BC03M05}), meaning the choice of stellar population synthesis model introduces a considerable systematic uncertainty.

It is of great importance to test whether Δ log(M_{BC03M05}) = log(M_{BC03}) − log(M_{M05}) correlates with redshift, color, or stellar mass, since such dependencies, if present, could bias studies of galaxy evolution or trends with mass. In Figure 13 we plot Δ log(M_{BC03M05} versus redshift, (U − V)_{rest} color, and stellar mass (the latter derived from BC03 models). We show galaxies

![Image of a page from a scientific paper]

**Fig. 12.—Tightening of the confidence interval around best-fit stellar mass, age, and dust extinction as a function of rest-frame U − V color for galaxies at 2 < z < 3.5. The median improvement after including the IRAC photometry is a factor 2.7 for red galaxies (dashed red line), significantly larger than the factor 1.3 for blue galaxies (dashed blue line). A similar color-dependence is found for constraints on age and dust extinction.**

**TABLE 5**

| Parameter        | medianΔ_{BC03M05} | σ_{MAMS}Δ_{BC03M05} | meanΔ_{BC03M05} | σΔ_{BC03M05} |
|------------------|-------------------|---------------------|-----------------|--------------|
| log (M_{*})....... | 0.14 ± 0.01       | 0.06                | 0.15 ± 0.01     | 0.08         |
| A_{V}.............. | −0.20 ± 0.00      | 0.00                | −0.18 ± 0.02    | 0.19         |
| log (Age_{w})..... | 0.29 ± 0.02       | 0.17                | 0.34 ± 0.02     | 0.18         |
with \( L_V > 5 \times 10^9 L_\odot \) at \( 1 < z < 2 \) (open symbols) and at \( 2 < z < 3.5 \) (filled symbols); no evidence for a redshift dependence is found. For the \( (U - V)_{\text{rest}} \) (middle panel) and stellar mass (right), the \( p \)-values for statistical significance from the Spearman rank-order correlation test are also larger than 0.05, meaning no significant correlation is found. Fitting a line to the points in the \( \Delta \log (M_{BC03}/M_{M05}) \) versus \( (U - V)_{\text{rest}} \) diagram, a difference of 0.06 dex in \( \Delta \log (M_{BC03}/M_{M05}) \) is found over the 2 mag range in \( (U - V)_{\text{rest}} \) color spanned by the galaxies in our sample. Even if a trend of increasing \( \Delta \log (M_{BC03}/M_{M05}) \) with redder \( (U - V)_{\text{rest}} \) color is real, it only introduces a small bias of the order of 15%. A similar conclusion can be drawn for the dependence on stellar mass.

### 5.2.3. Metallicity Dependence

We test how variations from solar metallicity affect the estimates of stellar mass, mass-weighted age, and dust extinction. We study the effect of a different metal abundance by fitting BC03 templates with metallicity \( Z = 0.2 Z_\odot \) to the observed SEDs, leaving the extinction law to Calzetti et al. (2000). NIR spectroscopy of DRGs (van Dokkum et al. 2004) and LBGs (Erb et al. 2006a) indicates that a range of \( Z = 0.2 - 1 Z_\odot \) is appropriate for galaxies at \( 2 < z < 3.5 \). Furthermore, at metallicities below \( Z = 0.2 Z_\odot \), the tracks and spectral libraries used to build the BC03 templates become more uncertain by lack of observational constraints. Decreasing the metallicity from \( Z = Z_\odot \) to \( Z = 0.2 Z_\odot \) lowers the estimated stellar masses of galaxies at \( 2 < z < 3.5 \) by 0.1 dex, leads to a mass-weighted age that is typically lower by 0.2 dex, and is compensated by an average increase in \( A_V \) of 0.2 mag. The fact that age estimates are more strongly affected than estimates of stellar mass when changing the assumed metallicity was demonstrated in detail by Worthey (1994). While absolute values of ages and dust extinctions may be biased as just described, the relative age and dust trends within the galaxy population as discussed in \$4 based on the standard SED modeling (see \$3) are robust.

### 5.2.4. Dependence on Extinction Law

The Calzetti et al. (2000) extinction law was empirically derived from observations of local starburst galaxies. We quantify the variations in stellar population properties due to the adopted extinction law by comparing our modeling results with a Calzetti et al. (2000) law to those obtained with reddening laws from Fitzpatrick (1986) for the Large Magellanic Cloud (LMC) and Prévot et al. (1984) for the Small Magellanic Cloud (SMC), leaving the metallicity at solar. Stellar masses, mass-weighted ages, and \( A_V \) values of galaxies at \( 2 < z < 3.5 \) derived with the LMC law models are similar to those obtained with the Calzetti et al. (2000) law. The SMC law, which rises more steeply toward shorter wavelengths in the near-UV, gives similar mass estimates, \( A_V \) values that are on average smaller by 0.3 mag, and mass-weighted stellar ages that are older by 0.23 dex, with the ages of the oldest galaxies being limited by the age of the universe constraint. As for metallicity, we conclude that using a different extinction law has a larger impact on the age estimates than on estimates of stellar mass.

### 6. STELLAR MASS–OPTICAL COLOR RELATION

In this section we study the relation between the rest-frame optical color of high-redshift galaxies and their stellar mass. We start with a model-independent approach in Figure 14, plotting the metallicity at solar. Stellar masses, mass-weighted ages, and \( A_V \) values of galaxies at \( 2 < z < 3.5 \) derived with the LMC law models are similar to those obtained with the Calzetti et al. (2000) law. The SMC law, which rises more steeply toward shorter wavelengths in the near-UV, gives similar mass estimates, \( A_V \) values that are on average smaller by 0.3 mag, and mass-weighted stellar ages that are older by 0.23 dex, with the ages of the oldest galaxies being limited by the age of the universe constraint. As for metallicity, we conclude that using a different extinction law has a larger impact on the age estimates than on estimates of stellar mass.

![Fig. 14.—Rest-frame \( U - V \) color vs. absolute \( J \) magnitude for galaxies at \( 2 < z < 3.5 \). Lyman-break galaxies are plotted in blue. DRGs (red symbols) populate the red side of the \( U - V \) color distribution. Black symbols denote those objects that do not meet either criteria. The solid line marks the \( K \)-band selection of our sample. The dust vector indicates an extinction of \( A_V = 1 \) mag. The most luminous galaxies in the rest-frame NIR have redder rest-frame optical colors than fainter galaxies.](image-url)
rest-frame $(U-V)_{\text{rest}}$ versus rest-frame $J_{\text{rest}}$ magnitude for all galaxies at $2 < z < 3.5$. The emission of low-mass long-lived stars that make up the bulk of the mass in a galaxy peaks in the rest-frame NIR. $J_{\text{rest}}$ is therefore expected to be a reasonably good tracer of stellar mass. The galaxies that satisfy the DRG selection criterion (red symbols) are found at redder $(U-V)_{\text{rest}}$ than the Lyman-break galaxies (blue symbols). The reddest $(U-V)_{\text{rest}}$ colors are found at the brightest $J_{\text{rest}}$ magnitudes. Note, however, that the observed trend is partially driven by the $K$-band selection of our sample. The line on Figure 14 indicates the magnitude at which a galaxy with identical colors to our observed galaxies would fall out of the sample. Even if we only consider galaxies brighter than the limiting $J_{\text{rest}} = -21.5$, to which we are complete over the whole $(U-V)_{\text{rest}}$ color range, we find that galaxies redder than $(U-V)_{\text{rest}} < 1$ are $1$ mag brighter than galaxies with $(U-V)_{\text{rest}} < 1$, significant at the $3 \sigma$ level. Studying a sample without color bias (as advocated by van Dokkum et al. 2006) proves crucial to picking up the trend of $(U-V)_{\text{rest}}$ with $J_{\text{rest}}$. We note that Meurer et al. (1999) found that LBGs with higher rest-frame UV luminosities tend to have redder rest-frame UV colors, illustrating that, while trends of color with luminosity are most notable in samples without color bias, they are still present in at least some color selected samples.

If $J_{\text{rest}}$ is a reasonable tracer of stellar mass, we expect to see a similar or stronger trend of $(U-V)_{\text{rest}}$ with the stellar mass. This is shown in Figure 15. The plotted mass is derived from one-component SED modeling of the $U$-to-$8 \mu$m SED, as described in §3. The typical error bar is indicated in the bottom left corner. The depth of our $K$ detection band allows us to probe stellar masses from $3 \times 10^{11} M_\odot$ down to $2 \times 10^9 M_\odot$. A correlation of $(U-V)_{\text{rest}}$ with stellar mass is clearly visible. The most massive galaxies have a red optical color. LBGs and other blue galaxies at high redshift contain typically 5 times less stellar mass than the DRGs in our sample. Again the $K$-band selection of our sample (solid line) limits our ability to detect faint red galaxies. Therefore, we cannot exclude the presence of low-mass red galaxies. The lack of massive blue galaxies seems to be real. Rigopoulou et al. (2006) find a comoving density of $\Phi = (1.6 \pm 0.5) \times 10^{-3}$ Mpc$^{-3}$ for LBGs with $M > 10^{11} M_\odot$ at an average redshift $(z) \simeq 3$, consistent with the absence of such massive but rare LBGs in our sample. However, the lack of massive blue galaxies could be an artifact of our choice of simple star formation histories. As demonstrated in §5.1.1 a severe underestimate of the stellar mass is possible when the true star formation history is more complex than that of the modeled one-component stellar population. When a young burst of star formation is superposed on a maximally old population, its blue light will dominate the $(U-V)_{\text{rest}}$ color and the mass from the underlying population will be hidden. In order to constrain the possible underestimate in mass, we fit two-component models to our SEDs. Erb et al. (2006b) describe a procedure to achieve this in two steps, where first a maximally old population is fit to the $K(+IRAC)$ data and subsequently a young population is fit to the (primarily UV) residual. However, this procedure does not guarantee a good fit in the $\chi^2$ sense. Instead, we decided to perform a simultaneous fit of both old and young components. We constructed template SEDs consisting of a maximally old single stellar population with a recent burst of star formation that started $100$ Myr ago and lasted until the moment of observation superposed. We made templates in which the mass fraction created in the burst is $2^x$ with $x$ going from $-6$ to $2$ in steps of $1$. We assume that the same reddening by dust applies to the old and the young population, with $A_V$ ranging from $0$ to $3$ in steps of $0.2$. Without this assumption, one could in principle hide an infinite amount of mass in an old population as long as an optically thick medium is shielding it from our sight. However, such a scenario is
physically implausible. Since we are interested in an upper limit on the mass, as opposed to the most likely value, we do not search for the least-squares solution over all $x$. Instead we perform the fit for every burst fraction and select the highest mass that still has $Δχ^2 = χ^2_{\text{red, two-component}} - χ^2_{\text{red, min, one-component}} < 2$.

Fitting the two-component models to the $U$-to-$8 \, \mu m$ SEDs of our galaxy sample at $2 < z < 3.5$, we indeed see that a higher stellar mass is allowed when more complex star formation histories are adopted (Fig. 16). The upper bound on stellar mass that we derive from this particular two-burst model is in the median a factor 1.7 higher than the one-component estimate for galaxies redder than $(U - V)_{\text{rest}} = 1$. For blue galaxies the median increase is a factor 2.1. Despite the fact that more mass can be hidden in blue galaxies, a trend of optical color with stellar mass remains visible. We performed a Mann-Whitney U-test to compare the $(U - V)_{\text{rest}}$ colors of galaxies with different stellar mass. We conservatively adopted the one-component stellar mass for galaxies with $(U - V)_{\text{rest}} > 1$ and the two-component upper limit for objects with $(U - V)_{\text{rest}} < 1$. To avoid selection effects we only consider galaxies more massive than $M = 10^{10} \, M_\odot$. Dividing them in two mass bins with an equal number of objects the Mann-Whitney U-test (Walpole & Myers 1985) confirms at a 99% significance level that the mean of the $(U - V)_{\text{rest}}$ distributions differs. Applying the same two-component models to the $U$-to-$K$ SEDs (omitting IRAC), the median upper mass estimate increases to a factor 2.3 above the one-component estimate for red objects and a factor 3.7 for blue objects. We conclude that, as expected from §5.1.1, more mass can be hidden in blue than in red galaxies, but this effect is insufficient to remove the trend of stellar mass with color. Furthermore, the amount of mass that can be hidden is constrained by addition of IRAC photometry.

The color dependence that we derive for the amount of mass that can be hidden in an underlying old population confirms findings from Shapley et al. (2005) based on a sample of star-forming galaxies at $z \sim 2$. The predominance of distant red galaxies at the high-redshift end was illustrated recently by van Dokkum et al. (2006) using a mass-selected sample of galaxies at $2 < z < 3$ with $M > 10^{11} \, M_\odot$. Only with very deep imaging, such as that of the HDF-S analyzed in this paper, it is possible to probe down to lower masses and prove that the most massive galaxies have red $(U - V)_{\text{rest}}$ colors compared to lower mass galaxies.

7. SUMMARY

We investigated the rest-frame optical to NIR color distribution of galaxies up to $2 < z < 3.5$ in the Hubble Deep Field-South. At all redshifts, galaxies with redder $(U - V)_{\text{rest}}$ tend to have redder $(V - J)_{\text{rest}}$ as is the case in the local universe. At $(U - V)_{\text{rest}}$ colors comparable to that of local galaxies, the color distribution of distant galaxies extends to redder $(V - J)_{\text{rest}}$. At $(U - V)_{\text{rest}} > 1$ the population of galaxies at the red $(V - J)_{\text{rest}}$ end is well described by dust-enshrouded star-forming models, whereas galaxies with $(V - J)_{\text{rest}}$ similar to that of local galaxies are consistent with old passively evolving systems. We conclude that $(U - V)_{\text{rest}}$ alone allows us to isolate blue relatively unobscured star-forming galaxies, but the addition of $(V - J)_{\text{rest}}$ is necessary to distinguish young dusty from old passively evolved systems. At redshifts above $z = 1$, this means IRAC observations are crucial in understanding the wide variety in stellar populations. We note that our analysis is not subject to uncertainties due to field-to-field variations, but surveys over much larger areas are needed to study the relative contributions of galaxies with different stellar populations.

We analyzed the constraints that IRAC places on stellar mass, stellar age, and dust content of galaxies at $2 < z < 3.5$. No evidence is found for systematic offsets when determining the stellar population characteristics with or without IRAC. However, the ratio of confidence intervals on stellar mass, mass-weighted age, and dust extinction is typically reduced by a factor 2.7, 2.9 and 1.7, respectively, for red $(U - V)_{\text{rest}} > 1$ galaxies. In general, IRAC does not provide stronger constraints for blue galaxies $(U - V)_{\text{rest}} < 1$ when very deep NIR imaging is available (as is the case for the HDF-S).

We caution that in characterizing the stellar populations using M05 models, we find stellar masses that are typically a factor 1.4 lower than for BC03 models with a scatter of 0.1 in dex.

A trend of brighter $J_{\text{rest}}$ with redder $(U - V)_{\text{rest}}$ is observed for galaxies at $2 < z < 3.5$, where the NIR luminosity serves as a (imperfect but model independent) tracer for stellar mass. Plotting $(U - V)_{\text{rest}}$ versus modeled stellar mass, we arrive at a similar conclusion: the most massive galaxies in our sample have red rest-frame optical colors. A possible concern is that this trend with mass is caused by our simplistic choice of star formation histories. When we allow for more complex star formation histories, more mass can be hidden in the case of a one-component stellar population and the amount depends on the color of the galaxy. We used two-component stellar populations, consisting of a maximally old population with a young population superposed, to set an upper bound on the stellar mass present. Even though relatively more mass can be hidden in blue galaxies compared to red galaxies, under the assumption of an equal dust reddening of the young and old component, a trend of $(U - V)_{\text{rest}}$ increasing to redder colors with stellar mass remains visible.

Support from NASA LTSA grant NNG04GE12G is gratefully acknowledged.

REFERENCES

Adelman-McCarthy, J. K., et al. 2006, ApJS, 162, 38
Angert, L., Tosi, M., Greggio, L., Sabha, E., Aloisi, A., & Leitherer, C. 2005, AJ, 129, 2203
Baldry, I. K., Glazebrook, K., Brinkmann, J., Ivezić, Z., Lupton, R. H., Nichol, R. C., & Szalay, A. S. 2004, ApJ, 600, 681
Bell, E. F., & de Jong, R. S. 2001, ApJ, 550, 212
Blanton, M. R., et al. 2003, ApJ, 594, 186
———, 2005, AJ, 129, 2562
Bolzonella, M., Miralles, J.-M., & Pelló, R. 2000, A&A, 363, 476
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000 (BC03)
Bundy, K., Ellis, R. S., & Conselice, C. J. 2005, ApJ, 625, 621
Calzetti, D., et al. 2000, ApJ, 533, 682
Cassisi, S., Castelli, M., & Castellani, V. 1997, A&A, 317, 108
Coleman, G. D., Wu, C.-C., & Weedman, D. W. 1980, ApJS, 43, 393
Cutri, R. M., et al. 2000, The 2MASS Explanatory Supplement (Pasadena: IPAC), http://www.ipac.caltech.edu/2mass/releases/allsky/doc/expsup.html

Daddi, E., Cimatti, A., Renzini, A., Fontana, A., Mignoli, M., Pozzetti, L., Tozzi, P., & Zamorani, G. 2004, ApJ, 617, 746
De Lucia, G., Springel, V., White, S. D. M., Croton, D., & Kauffmann, G. K. 2006, MNRAS, 366, 499
Doré, N., Salvato, M., Gabasch, A., Bender, R., Hopp, U., Feulner, G., & Pannella, M. 2005, ApJ, 619, L131
Erb, D. K., Shapley, A. E., Pettini, M., Steidel, C. C., Reddy, N. A., & Adelberger, K. L. 2006a, ApJ, 644, 813
Erb, D. K., Steidel, C. C., Shapley, A. E., Pettini, M., Reddy, N. A., & Adelberger, K. L. 2006b, ApJ, 646, 107
Fazio, G. G., et al. 2004, ApJS, 154, 10
Ferguson, H. C., Dickinson, M., & Papovich, C. 2002, ApJ, 569, L65
 Fitzpatrick, E. L. 1986, AJ, 92, 1068
Förster Schreiber, N. M., et al. 2003, ApJ, 599, 193
Forster Schreiber, N. M., et al. 2004, ApJ, 616, 40
Franx, M., et al. 2003, ApJ, 587, L79
Freeman, K., & Bland-Hawthorn, J. 2002, ARA&A, 40, 487
Glazebrook, K., et al. 2004, Nature, 430, 181
Hopp, J. L., et al. 2004, Proc. SPIE, 5487, 77
Kauffmann, G., et al. 2003, MNRAS, 341, 54
Kinney, A. L., et al. 1996, ApJ, 467, 38
Kriek, M., et al. 2006, ApJ, 645, 44
Labbé, I., et al. 2003, AJ, 125, 1107 (L03)
———. 2005, ApJ, 624, L81
Lançon, A., et al. 2001, ApJ, 552, 150
Madau, P., et al. 1996, MNRAS, 283, 1388
Maraston, C. 2005, MNRAS, 362, 799 (M05)
Maraston, C., et al. 2006, ApJ, 652, 85
Meurer, G. R., Heckman, T. M., & Calzetti, D. 1999, ApJ, 521, 64
Nagamine, K., Cen, R., Hernquist, L., Ostriker, J. P., & Springel, V. 2005, ApJ, 627, 608
Papovich, C., Dickinson, M., & Ferguson, H. C. 2001, ApJ, 559, 620
Papovich, C., Dickinson, M., Giavalisco, M., Conselice, C. J., & Ferguson, H. C. 2005, ApJ, 631, 101
Papovich, C., et al. 2006, ApJ, 640, 92
Prévot, M. L., Lequeux, J., Prévot, L., Maurice, E., & Rocca-Volmerange, B. 1984, A&A, 132, 389
Rigopoulou, D., et al. 2006, ApJ, 648, 81
Rudnick, G., et al. 2001, AJ, 122, 2205
———. 2003, ApJ, 599, 847
———. 2006, ApJ, 650, 624
Salpeter, E. E. 1955, ApJ, 121, 161
Shapley, A. E., et al. 2001, ApJ, 562, 95
———. 2003, ApJ, 588, 65
———. 2005, ApJ, 626, 698
Smail, I., Chapman, S. C., Blain, A. W., & Ivison, R. J. 2004, ApJ, 616, 71
Somerville, R. S., Primack, J. R., & Faber, S. M. 2001, MNRAS, 320, 504
Steidel, C. C., et al. 1996, ApJ, 462, L17
———. 1999, ApJ, 519, 1
Trager, S. C., Faber, S. M., Worthey, G., & González, J. J. 2000, AJ, 120, 165
van der Wel, A., Franx, M., van Dokkum, P. G., Rix, H.-W., Illingworth, G. D., & Rosati, P. 2005, ApJ, 631, 145
van der Wel, A., Franx, M., Wuyts, S., van Dokkum, P. G., Huang, J., Rix, H.-W., & Illingworth, G. D. 2006, ApJ, 652, 97
van Dokkum, P. G., et al. 2004, ApJ, 611, 703
———. 2006, ApJ, 638, L59
Walpole, R. E., & Myers, R. H. 1985, Probability and Statistics for Engineers and Scientists (3rd. ed.; New York: Macmillan)
Webb, T. M., et al. 2006, ApJ, 636, L17
Werner, M. W., et al. 2004, ApJS, 154, 1
Worthey, G. 1994, ApJS, 95, 107
Yan, H., et al. 2004, ApJ, 616, 63