TO STACK OR NOT TO STACK: SPECTRAL ENERGY DISTRIBUTION PROPERTIES OF Lyα-EMITTING GALAXIES AT z = 2.1

Carlos J. Vargas1,2, Hannah Bish1, Viviana Acquaviva3, Eric Gawiser1, Steven L. Finkelstein1, Robin Ciardullo5,6, Matthew L. N. Ashby7, John Feldmeier8, Henry Ferguson9, Caryl Gronwall5,6, Lucia Guaita10, Alex Hagen5,6, Anton Koekemoer9, Peter Kurczynski1, Jeffrey A. Newman11, and Nelson Padilla12

1 Department of Physics and Astronomy, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA
2 Department of Astronomy, New Mexico State University, Las Cruces, NM 88001, USA
3 Physics Department, New York City College of Technology, City University of New York, 30 Jay Street, Brooklyn, NY 11201, USA
4 Department of Astronomy, The University of Texas at Austin, Austin, TX 78712, USA
5 Department of Astronomy & Astrophysics, The Pennsylvania State University, University Park, PA 16802, USA
6 Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802, USA
7 Harvard-Smithsonian Center for Astrophysics, 60 Garden St., Cambridge, MA 02138, USA
8 Department of Physics and Astronomy, Youngstown State University, Youngstown, OH 44555, USA
9 Space Telescope Science Institute, Baltimore, MD, USA
10 Department of Astronomy, Oskar Klein Center, Stockholm University, Roslagstullsbacken 21, Stockholm, Sweden
11 Department of Physics and Astronomy, University of Pittsburgh, 3409 O’Hara St., Pittsburgh, PA 15260, USA
12 Departamento de Astronomía y Astrofísica, Universidad Católica de Chile, Santiago, Chile

Received 2013 September 22; accepted 2013 December 27; published 2014 February 10

ABSTRACT

We use the Cosmic Assembly Near-Infrared Deep Extragalactic Legacy Survey (CANDELS) GOODS-S multi-wavelength catalog to identify counterparts for 20 Lyα emitting (LAE) galaxies at z = 2.1. We build several types of stacked spectral energy distributions (SEDs) of these objects. We combine photometry to form average and median flux-stacked SEDs, and postage-stamp images to form average and median image-stacked SEDs. We also introduce scaled flux stacks that eliminate the influence of variation in overall brightness. We use the SED fitting code SpeedyMC to constrain the physical properties of individual objects and stacks. Our LAEs at z = 2.1 have stellar masses ranging from 2 × 10^7 M_☉ to 8 × 10^9 M_☉ (median = 3 × 10^8 M_☉), ages ranging from 4 Myr to 500 Myr (median = 100 Myr), and E(B − V) between 0.02 and 0.24 (median = 0.12). Although still low, this represents significantly more dust reddening than has been reported for LAEs at higher redshifts. We do not observe strong correlations between Lyα equivalent width (EW) and age or E(B − V). The Lyα radiative transfer (q) factors of our sample are predominantly close to one and do not correlate strongly with EW or E(B − V). The absence of strong correlations with EW or q implies that Lyα radiative transfer is highly anisotropic and/or prevents Lyα photons from scattering in dusty regions. The SED parameters of the flux stacks match the average and median values of the individual objects, with the flux-scaled median SED performing best with uncertainties reduced by a factor of two. Median image-stacked SEDs provide a poor representation of the median individual object, and none of the stacking methods capture the large dispersion of LAE properties.

Key words: galaxies: evolution – galaxies: high-redshift

Online-only material: color figures

1. INTRODUCTION

The nature of radiation emitted from astrophysical sources offers insights into their physical properties. Lyα photons are emitted when a neutral hydrogen atom undergoes a transition from the first excited state to the ground state. An immense number of Lyα photons are emitted at 1216 Å rest frame from H II regions surrounding young, massive stars. The large luminosity of the Lyα emission line makes it easily identified even over vast distances, and thus it is widely used in research in extragalactic astronomy. Lyα emitting (LAE) galaxies are typically discovered using the narrowband (NB) filter technique (Cowie & Hu 1998; Rhoads et al. 2000), in which a detection of an LAE is made by finding greater signal-to-noise ratios (S/Ns) in an NB filter image than in images obtained with adjacent broadband filters.

LAE galaxies at 2 < z < 3 are thought to be building blocks of present-day galaxies like our own Milky Way (Gawiser et al. 2007; Guaita et al. 2011). These galaxies are easy to detect thanks to the strength of the Lyα line, but they usually have low stellar masses and are therefore dim in the continuum. As a result, it has been a common procedure to stack multiple photometric data points of LAEs in order to enhance their S/N (Gawiser et al. 2006a, 2007; Nilsson et al. 2007; Pirzkal et al. 2007; Finkelstein et al. 2007; Lai et al. 2008; Pentericci et al. 2009; Ono et al. 2010a, 2010b; Ouchi et al. 2010; Nilsson et al. 2011; Guaita et al. 2011; Finkelstein et al. 2011a; Acquaviva et al. 2011, 2012b).

Inconsistencies in the literature, particularly in estimating ages and sample diversity, motivate a study of the validity of the stacking method for analyzing the spectral energy distributions (SEDs) of dim LAEs. To date, most SED studies of high-redshift LAEs have been done by stacking the catalog fluxes or ground-based postage-stamp images of galaxies too faint to study individually. By stacking numerous faint objects (via average or median), we increase the S/N significantly. It is believed that a study of these images will represent the characteristics of a sample of galaxies on the average. However, this interpretation of median as typical relies on the assumption that the best-fit SED parameters of the stacked galaxies have similar physical properties. An additional key assumption is that the median (average) SEDs will match the median (average) physical...
properties of individual galaxy SEDs. This assumption appears reasonable for stellar mass if the mass-to-light ratio of the stellar populations of different LAEs is similar, but it is not guaranteed for other parameters like age or $E(B - V)$.

Results from the analysis of stacked samples of LAEs indicate a large spread in the properties of these galaxies and introduce further concern about the validity of the stacking technique. A recent study of median image-stacked LAEs at $z = 3.1$ found them to be older than median flux-stacked LAEs at $z = 2.1$ (Acquaviva et al. 2012a); this puzzling result would imply that LAEs grow younger as cosmic time progresses. However, it might instead result from the failure of image and/or flux stacking to accurately measure the SED characteristics of individual galaxies, and from systematic differences between image stacking and flux stacking. Additionally, a previous study of individual LAEs at $z = 4.5$ (Finkelstein et al. 2009) found a wide range of physical properties, including many LAEs that appeared dusty, something that would not have been expected given the earlier results at $z = 3.1$ from Gawiser et al. (2006b), which implied, via a stacking analysis, that LAEs were dust free.

In this paper, we perform SED fitting on 20 LAEs at $z = 2.1$, as well as on their flux- and image-stacked SEFs, in order to clarify whether these discrepancies are attributable to the failure of one or all stacking methods. The use of deep Hubble Space Telescope (HST) and Spitzer Infrared Array Camera (IRAC) data gives us the opportunity, for the first time, to study rest-frame ultraviolet (UV) and optical properties of individual objects, which is crucial for accurate constraints on age and stellar mass. We use the Markov Chain Monte Carlo (MCMC) algorithm SpeedyMC (Acquaviva et al. 2012a) to analyze the SED properties of our sample. SpeedyMC is a streamlined implementation of GaMC (Acquaviva et al. 2011) designed to handle large samples of SEDs. This code utilizes Bayesian statistics to determine the expectation values of SED parameters and provide estimates of the uncertainties associated with their measurements. All magnitudes referenced in the text hereafter are in the AB magnitude system. All SED fitting calculations assume $H_0 = 73$ km s$^{-1}$ Mpc$^{-1}$, $\Omega_M = 0.258$, and $\Omega_\Lambda = 0.742$.

2. DATA

The starting point for our analysis was the sample of 216 LAE candidates at $z = 2.1$ discovered in deep 3727 Å NB images of the $30' \times 30'$ Extended Chandra Deep Field-South (ECDF-S) by Guaita et al. (2010) as refined by Bond et al. (2012), who excluded 34 additional LAE candidates that appear to be low-redshift contaminants due to extended morphology in HST V-band optical images. These objects typically have low S/Ns in ground-based broadband images, with a median magnitude of $R \sim 25.5$, and are unresolved at $0''.1$. Thus, the addition of deep HST imaging is crucial.

The Cosmic Assembly Near-Infrared Deep Extragalactic Legacy Survey (CANDELS; Grogin et al. 2011; Koekemoer et al. 2011) has produced a deep $H$-band-selected multiwavelength catalog of the GOODS-S field, which encompasses the central $16' \times 10'$ of ECDF-S (Guo et al. 2013). The observations are 10 epochs deep in $H$ band (limiting magnitude of 27.2), which corresponds to roughly six $HST$ orbits. The catalog contains photometry from $U$-band images from CTIO and VIMOS, Advanced Camera for Surveys (BViz) images from the GOODS survey (Giavalisco et al. 2004), CANDELS WFC3 images (F098M, F105W, F125W, and F160W), $K$-band images from Hawk-I and ISAAC, and Spitzer Space Telescope IRAC (Fazio et al. 2004) 3.6, 4.5 $\mu$m images from SEDs (Ashby et al. 2013) and 5.8, 8.0 $\mu$m images from GOODS (Dickinson et al. 2003). For SED fitting, we excluded both $U$ bands due to expected Ly$\alpha$ contamination at this redshift, F098M due to having coverage in a minority of our sample that might have different median physical properties than the rest, and both $K$ bands due to incomplete coverage and shallower photometry. All of our analysis made use of imaging and photometry from the remaining 11 space-based bands.

We match the positions of our catalog of known LAEs at $z = 2.1$ (Guaita et al. 2010) to the positions of objects contained within the version of the CANDELS Multi-wavelength catalog presented in Guo et al. (2013). The positions within the $z = 2.1$ LAE catalog are accurate to within $0''.1$. Objects within a distance of $0''.5$ were recorded as a match, and their CANDELS SEDs were used in our analysis. Only two objects showed multiple possible counterparts within this tolerance, and they were discarded due to the strong likelihood of neighboring object contamination. In total, we found 20 LAE counterparts in the CANDELS Multi-wavelength catalog. We note that only 30 of the original 216 LAEs are within the GOODS-S field, and the 10 “missing” objects were probably too dim to have been detected at the $5\sigma$ level in $H$ band (F160W) and were thus excluded from the CANDELS catalog. In our sample, 9 of 20 LAEs were observed with VLT+FORS (Berry et al. 2012), yielding five confirmed redshifts over $2.05 < z < 2.08$, consistent with the expected scatter in redshifts from NB selection. The other four objects showed insufficient S/N to yield a spectroscopic redshift.

The main objective of this paper is the comparison of the SED fitting properties of individual objects and stacks. The SED of the stacks can be obtained via two different procedures: by combining the cataloged fluxes of the individual objects, or by stacking the postage-stamp images of the 20 individual objects at each wavelength and then performing photometry on the combined images. We refer to the SEDs obtained via these two methods as “flux-stacked” and “image-stacked,” respectively. We took the median and average of the flux densities in each band of all of the $z = 2.1$ LAEs with CANDELS photometry in that band and will henceforth refer to these as the median and average flux stacks, respectively.

To create the image-stacked SED, we performed both average and median stacking on cutout images in each band centered around each of our LAEs with CANDELS imaging in that band. The extracted HST cutouts measured 211 pixels on a side, while the IRAC cutouts measured 21 pixels. As the processed HST image pixels measure $0''.06$ and the processed IRAC image pixels measure $0''.6$, the cutsouts in all bands had approximately the same size of $12''6$ on a side. We note that of the 20 objects, 6 of them were located in the northern $\sim25\%$ of the GOODS-S field, which has WFC3 data from the WFC3 Early Release Science program (Windhorst et al. 2011). While these images were obtained prior to the CANDELS data, they were reprocessed alongside the CANDELS data to make a single mosaic which was used to extract the cutouts. Along with the science images, we also extracted cutouts from the rms images for use with Source Extractor (Bertin & Arnouts 1996; see Koekemoer et al. 2011 for details on the science and rms mosaic construction).

To create our image stacks, we created median and average images by taking the median/average value for each pixel from all of the cutout images for a given band, excluding pixels with rms values $>10$—a very high value, indicative of a problem with the given pixel in that image. We then measured photometry on
the image stacks using Source Extractor in dual-image mode, with the F160W stack as the detection image for the \textit{HST} bands and the 3.6 $\mu$m image as the detection band for the \textit{HST} images. We measured photometry using Kron-style elliptical apertures, denoted by MAG\_AUTO within Source Extractor. For the \textit{HST} bands, we measured colors in small Kron apertures (with the parameter PHOT\_AUTOPARAMS set to 1.2, 1.7), with the fluxes corrected to total using an aperture correction derived by the ratio of the default MAG\_AUTO flux value to our smaller Kron flux (this procedure has been optimized for faint, high-redshift sources; see Finkelstein et al. 2010 for details). For the IRAC stacks, we measured photometry directly from MAG\_AUTO, which is designed to approximate a total flux. The photometry for the individual objects was also performed in the same way. For reference, typical S/N values in \textit{HST} band of the individual objects are $\sim 8$, while that of the average flux stack is $\sim 16$.

To further our analysis, we also introduce an additional set of image stacks centered on the object positions found in the MUSYC NB catalog of LAEs. Comparison of the primary \textit{HST}-centered image stacks with these NB-centered stacked images will illustrate any improvement that results from eliminating the $\sim 0.1$ NB centroiding errors (Guaita et al. 2010) and any astrometric offsets between ground-based and space-based coordinates. Figure 1 shows F160W filter postage-stamp images of each galaxy in our sample and those of the \textit{HST} and NB-centered image stacks. Typically, LAEs have half-light radii of 0.5–2 kpc in \textit{HST} images (Bond et al. 2012; Gronwall et al. 2007; Malhotra & Rhoads 2002). While some of the galaxies seen in Figure 1 are mildly resolved, their light distribution is accounted for by the TFFIT (Laidler et al. 2007) process used to make the CANDELS catalogs, and the majority of their light will fall inside the apertures used to obtain photometry from image stacks. In order to explore the effects of potential IRAC contamination, we introduce IRAC-clean image stacks. These stacks were composed of 11 individual objects determined by visual inspection to be free of significant flux contamination from neighboring objects in the IRAC 3.6 $\mu$m band. The IRAC-clean stack postage-stamp images are compared to those of the entire sample in Figure 2. A slight loss of S/N in the clean sample is apparent due to the reduced sample size.

2.1. Error Estimation

The uncertainty in the fluxes for individual objects’ SEDs includes an effective zero-point calibration error, which is added in quadrature to the photometric uncertainty. We assume that these zero-point errors are 3% of the flux in \textit{HST} bands and 8% in the IRAC bands (CANDELS team 2012, private communication).

Errors on the flux-stacked photometry are determined by a bootstrap procedure, as in Guaita et al. (2011) and Acquaviva et al. (2011). This accounts for both the photometric error and the sample variance resulting from the spread in the SEDs of different galaxies; for our sample, the latter is the dominant source of uncertainty. The same calibration uncertainty used for the individual objects’ SED is added in quadrature to the bootstrap error in both the average and median flux stacked SED. For image-stacked SEDs, we adopt a conservative error
However, centroiding errors are inevitable; for CANDELS objects detected at $S/N > 5$, these errors should be at most a single pixel (Koekemoer et al. 2011). When we simulate 1 pixel rms centroiding errors but no additional sources of error, the average flux and average $HST$-centered image stacks lose 24% of the flux, the median flux stack loses 29%, and the median $HST$-centered image stack loses 43% of the flux. Adding realistic levels of photometric and background subtraction errors makes no significant difference in these results. The larger flux loss in median image stacks may explain why these are the dimmest stacked SEDs at wavelengths imaged by $HST$, as seen in Figure 3.

The overall stacking behavior is similar at the much larger pixel scale of IRAC bands (e.g., 1.2 pixels sampling a 1.8 PSF at 3.6 μm), where we continued to assume that the centroiding errors were no larger than 0.06 given the usage of CANDELS $H$-band catalog positions for the $HST$-centered image stacks and for TFIT photometry on IRAC images used to make the multiwavelength catalog inputs for flux stacking. These centroiding errors cause no measurable loss in flux at IRAC resolution. However, this could create a mild color bias in the stacks since, as noted above, 24%–43% of the flux is being lost to centroiding errors in $HST$ bands. The worst effect is expected in the median $HST$-centered image stack, which does exhibit a red $H - [3.6]$ color in Figure 3. It should be noted that our simulations do not include the non-Gaussian PSF wings typical in IRAC bands or any attempt to account for the contamination by neighboring objects that results from them.

We also simulated the NB-centered image stacks where the ground-based MUSYC NB-detected LAE positions were adopted, causing 0.1 centroid errors. If we had used ~FWM diameter apertures, which are formally optimal for point sources, in the $HST$ images, the NB-centered average (median) stacks would underestimate $HST$-band fluxes by a factor of two (6). However, the larger apertures used in $H$ band to determine aperture corrections are big enough to correct this bias for the average NB-centered image stack, which is well worth the loss of $S/N$.

It is important to note that no matter what aperture size is used, our simulations show that median image stack photometry is biased low by a large factor whenever centroiding errors are significant compared to the PSF. This occurs because the various objects exhibit only partial overlap, leading median image values to be dominated by image background rather than objects. This provides a significant note of caution about the usage of median image stacks, although the effect, as seen in Figure 3, appears to be modest for our NB-centered median image stack and should be negligible for our $HST$-centered median image stack due to the small expected centroiding errors.

### 3. SED FITTING METHODOLOGY

Information about the physical properties of galaxies, including redshift, stellar mass, age, dust content, and metallicity, can be determined by fitting their SEDs. The algorithm that was used for this analysis, SpeedyMC, is a faster version of GaLMC introduced in Acquaviva et al. (2012a). Rather than using GALAXEV to generate the model SED at each point, a template library on a grid of locations encompassing the entire parameter space is generated beforehand. The exploration of parameter space is then carried out using the same MCMC algorithm, but at each location, multi-linear interpolation between the pre-computed spectra is used to calculate the model SED and its corresponding $\chi^2$ value. The use of the SpeedyMC method

![Figure 3. SEDs of each individual object (gray) plotted with all of the stacked SEDs in our analysis. Red is the average $HST$ image stack, blue is the median $HST$ image stack, magenta is the average NB-centered image stack, cyan is the median NB-centered image stack, green is the median stack flux, orange is the average flux stack, brown is the average scaled flux stack, and dark green is the median scaled flux stack. One can observe the median flux stack’s robustness to outlying sources, especially in the Spitzer IRAC bands. There are three objects that are particularly bright in the Spitzer IRAC bands, implying larger stellar masses. These objects have CANDELS catalog numbers 17254, 25902, and 205 and are the three largest mass objects in the sample.](image-url)
Figure 4. Left panel: a plot of the SED of individual object 3826 overlaid with its MCMC best-fit SED template. These templates correspond to the model SED that produces the lowest χ² value by SpeedyMC. Note that the U-band data are affected by Lyα emission and that the emission-line contribution is, in this case, predicted extremely well by the best-fit model to the 11 space-based bands used for SED fitting. Right panel: a similar plot for the median flux stack.

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

allows us to fit the SEDs of galaxies at a rate 20,000 times faster than GalMC, and corresponding to about 1 s per galaxy on a 2.2 GHz MacBook Pro laptop.

In our analysis, the parameter space consisted of stellar mass, age, and dust reddening defined by excess color E(B − V). We assumed a constant star formation history and fixed the redshifts at z = 2.1 and the metallicity, Z, at the value of 0.2 Z☉ as indicated by both spectroscopic analysis and SED fitting results (Acquaviva et al. 2011; Finkelstein et al. 2011). We used the stellar population synthesis models of Bruzual & Charlot (2003, hereafter BC03) and S. Charlot & G. Bruzual (2007, private communication, hereafter CB07) and included nebular emission according to the procedure described in Acquaviva et al. (2011). Galactic absorption was taken into account using the Calzetti law (Calzetti et al. 1994), with a value RV of 4.05, and starlight absorption by the intergalactic medium using the prescription from Madau (1995). We used the Salpeter initial mass function (IMF) for consistency with the previous literature on this subject. Our reference grid contained 100 values of both age and E(B − V) (the stellar mass is a free normalization parameter that can be varied without the need to build extra models in the grid). The number of values in the grid was chosen by refining the resolution of the grid until the resulting probability distributions obtained from GalMC and SpeedyMC were virtually identical, a sign that the linear interpolation between reference SEDs was working correctly. We limit the grid to range 10^4–10^15 M☉, and in E(B − V) we range 0–1.

We ran six chains on each object, using 100,000 steps for each and starting from six different locations to ensure sampling of the entirety of parameter space and minimize “chain locking” in local minima. The “GetDist” software from Lewis & Bridle (2002) was used to analyze the chains and to make sure that the distribution inferred from the chains had converged to the true one. In the left panel of Figure 4, we plot the observed SED of CANDELS object 3826 overlaid with its best-fitting model produced by SpeedyMC. A similar plot for the median flux stacked SED is included as the right panel of Figure 4. We plot the observed U-band data points in this figure as an illustration of possible Lyα contamination in that band.

4. RESULTS

The BC03 stellar population synthesis models performed better than the CB07 ones, resulting in smaller χ² values on average. This behavior is in agreement with results in the recent literature that seemed to favor a low contribution of thermally pulsating asymptotic giant branch stars (e.g., Kriek et al. 2010; Meidt et al. 2012; Zibetti et al. 2013). For these reasons, we show the parameters obtained by using the BC03 models in Table 1 and the figures. Individual LAEs at z = 2.1 are found to have stellar masses ranging from 2.3 × 10^7 M☉ to 8.5 × 10^9 M☉, ages ranging from 0.004 Gyr to 0.47 Gyr, and E(B − V) between 0.02 and 0.24.

Figure 5 shows the relationships between parameters of age, stellar mass, and E(B − V). In our analyses of parameters that may have statistically significant correlations with other parameters, we use both the Pearson product-moment correlation coefficient and the Spearman rank correlation coefficient. The Pearson and Spearman correlation coefficients (ρ) describe the strength and direction of the correlation and range from −1 to 1. We define 1 − ρ as the confidence with which one can reject the null hypothesis of no correlation. In order for two parameters to be correlated at 95% confidence, their ρSpearman value must be less than 0.05. Because the intrinsic distribution of parameters is not well known, we adopt the Spearman p values. We see a correlation of age with stellar mass, as expected. The assumption of a constant star formation history forces older galaxies to be more massive, and thus we see a correlation between age and stellar mass. No correlation is found between age and E(B − V), implying that older LAEs generally host more dust.

4.1. Correlations with Equivalent Width

Figure 6(a) shows stellar mass versus the rest-frame Lyα equivalent width (EW). There is a rough upper envelope on
the points such that no objects exhibit rest-frame EW > 60 Å and stellar mass > 10^7 M_☉. Because EW was measured via NB 3727 Å and broadband U, B photometry (Guaita et al. 2010), there is no direct selection effect on stellar mass. Hence, the lack of high-stellar-mass, high-EW objects implies that galaxies with high stellar mass also have high rest-frame UV luminosity. This should be reflected via a related dearth of objects with high stellar mass and low star formation rate (SFR), i.e., low specific SFRs. Also, the correlation between stellar mass and EW is likely due to larger mass galaxies hosting more O- and B-type stars, producing the UV radiation that leads to Lyα emission.

Figure 6(b) shows stellar population ages versus rest-frame EW. Here, one might expect to see a strong correlation, as the Lyα EW of a starburst is a strongly decreasing function of age, and for continuous SFR the correlation remains strong (Shapley et al. 2003). The observed lack of correlation implies that Lyα radiative transfer is not trivial; however, it could also be produced by a disconnect between our measured ages and the true age of the starbursting population caused by our assumption of a single stellar population with constant SFR.

Figure 6(c) shows dust reddening, E(B − V), versus rest-frame EW. Though statistically correlated, the correlation is weak. One could expect that dustier galaxies would more effectively quench Lyα photons to produce lower EWs. The lack of strong correlation therefore implies that Lyα radiative transfer is working to prevent Lyα photons from repeated resonant

| CANDELS ID      | M* (10^9 M_☉) | Age (Gyr) | E (B − V) | Best-fit \( \chi^2 / \text{dof} \) |
|-----------------|---------------|-----------|-----------|-----------------|
| 205 (Clean)     | 2.474 ± 0.328 | 0.119 ± 0.032 | 0.169 ± 0.014 | 81.9/8          |
| 991 (Clean)     | 0.407 ± 0.053 | 0.095 ± 0.029 | 0.137 ± 0.018 | 18.2/8          |
| 7097 (Clean)    | 0.971 ± 0.250 | 0.116 ± 0.070 | 0.156 ± 0.013 | 111.3/8         |
| 9125 (Clean)    | 0.280 ± 0.052 | 0.473 ± 0.138 | 0.029 ± 0.059 | 20.1/8          |
| 10220 (Clean)   | 0.717 ± 0.154 | 0.211 ± 0.100 | 0.094 ± 0.020 | 151.0/8         |
| 11824 (Clean)   | 0.023 ± 0.005 | 0.013 ± 0.052 | 0.031 ± 0.022 | 14.0/8          |
| 12369 (Clean)   | 0.064 ± 0.030 | 0.053 ± 0.051 | 0.058 ± 0.011 | 13.4/8          |
| 14889 (Clean)   | 0.101 ± 0.078 | 0.015 ± 0.017 | 0.156 ± 0.010 | 13.8/8          |
| 17254 (Clean)   | 8.518 ± 0.205 | 0.280 ± 0.183 | 0.240 ± 0.016 | 189.3/8         |
| 22689 (Clean)   | 0.286 ± 0.118 | 0.092 ± 0.093 | 0.064 ± 0.015 | 5.0/8           |
| 24168 (Clean)   | 1.116 ± 0.219 | 0.147 ± 0.070 | 0.121 ± 0.019 | 61.0/8          |
| 2210           | 1.533 ± 0.281 | 0.160 ± 0.069 | 0.168 ± 0.018 | 28.0/8          |
| 3826           | 0.024 ± 0.014 | 0.004 ± 0.004 | 0.083 ± 0.034 | 4.2/8           |
| 6696           | 0.231 ± 0.415 | 0.076 ± 0.581 | 0.192 ± 0.025 | 11.3/8          |
| 7408           | 0.053 ± 0.006 | 0.002 ± 0.011 | 0.017 ± 0.010 | 4.0/8           |
| 1192           | 0.154 ± 0.038 | 0.031 ± 0.010 | 0.113 ± 0.014 | 14.1/8          |
| 18836          | 0.124 ± 0.089 | 0.013 ± 0.013 | 0.155 ± 0.017 | 24.0/8          |
| 21026          | 0.060 ± 0.086 | 0.017 ± 0.041 | 0.072 ± 0.016 | 4.9/8           |
| 21553          | 0.572 ± 0.184 | 0.258 ± 0.203 | 0.056 ± 0.035 | 6.1/8           |
| 25902          | 2.966 ± 0.847 | 0.290 ± 0.312 | 0.150 ± 0.020 | 94.0/8          |

| Average of individual results | 1.03 | 0.123 | 0.113 |
| Median of individual results | 0.283 | 0.094 | 0.117 |
| Scatter of individual results | 1.896 | 0.121 | 0.059 |

Average flux stack | 1.050 ± 0.079 | 0.184 ± 0.471 | 0.146 ± 0.042 | 3.6/8 |
Median flux stack | 0.511 ± 0.347 | 0.148 ± 0.335 | 0.108 ± 0.048 | 1.7/8 |
Average HST image stack | 1.192 ± 0.262 | 0.309 ± 0.873 | 0.124 ± 0.066 | 22.5/8 |
Median HST image stack | 0.810 ± 0.451 | 0.305 ± 0.509 | 0.108 ± 0.075 | 11.8/8 |
Average NB image stack | 0.222 ± 0.180 | 0.019 ± 0.044 | 0.165 ± 0.006 | 43.5/8 |
Median NB image stack | 0.828 ± 0.396 | 0.462 ± 0.462 | 0.101 ± 0.042 | 16.9/8 |

Scaled average flux stack | 0.525 ± 0.280 | 0.084 ± 0.093 | 0.105 ± 0.025 | 48.2/8 |
Scaled median flux stack | 0.382 ± 0.165 | 0.091 ± 0.096 | 0.131 ± 0.013 | 3.31/8 |

Clean average flux stack | 1.314 ± 0.124 | 0.206 ± 0.635 | 0.160 ± 0.057 | 4.6/8 |
Clean median flux stack | 0.639 ± 0.182 | 0.348 ± 0.098 | 0.048 ± 0.016 | 3.7/8 |
Clean average HST image stack | 0.088 ± 0.031 | 0.008 ± 0.034 | 0.140 ± 0.028 | 20.9/8 |
Clean median HST image stack | 0.049 ± 0.059 | 0.009 ± 0.048 | 0.089 ± 0.062 | 23.2/8 |
Clean average NB image stack | 0.095 ± 0.126 | 0.007 ± 0.029 | 0.150 ± 0.023 | 15.8/8 |
Clean median NB image stack | 0.079 ± 0.214 | 0.016 ± 0.170 | 0.131 ± 0.042 | 10.4/8 |

Notes. Below the results for each individual object, we display the individual object sample average, mean, and standard deviation in each parameter, for comparison to stacking results.
Figure 5. These plots show the relationships between (a) stellar mass and age, (b) stellar mass and $E(B-V)$, and (c) age and $E(B-V)$ of the sample. The black points represent the best-fitting parameter values for the individual objects, and the remaining color scheme is retained from Figure 3. The error bars plotted are those output by SpeedyMC and are asymmetrical. The best-fitting correlation derived from the Spearman correlation analysis is plotted as a diagonal line. The Pearson and Spearman coefficients, $(\rho_{\text{Pearson}}, \rho_{\text{Spearman}})$, are listed. We also list the $p_{\text{Pearson}}$ and $p_{\text{Spearman}}$ values, which must be less than 0.05 to rule out the null hypothesis of no correlation at 95% confidence.

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

Figure 6. These plots show the relationships between (a) stellar mass and Ly$\alpha$ EW, (b) age and Ly$\alpha$ EW, and (c) $E(B-V)$ and Ly$\alpha$ EW of the individual objects. The best-fitting correlation derived from the Spearman correlation analysis is plotted as a diagonal line, with $\rho$ and $p$ values listed as in Figure 5. The vertical red dashed line represents the 20 Å EW selection criterion within the MUSYC NB catalog.

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

4.2. The Ly$\alpha$ q Factor

As mentioned above, we excluded the U-band data from SED fitting due to this likelihood of Ly$\alpha$ emission making a significant contribution to this broadband photometry and the impossibility of properly accounting for Ly$\alpha$ radiative transfer in our SED templates. However, this makes the U-band data ideal for an estimation of Ly$\alpha$ q factors, which represent the extent to which Ly$\alpha$ emission from a galaxy has been enhanced ($q > 1$) or quenched ($q < 1$) along the line of sight to Earth. As was first done by Finkelstein et al. (2008), we define $q = \tau_{\text{Ly} \alpha}/\tau_{\lambda=1216}$, where $\tau_\lambda = k_\lambda E(B-V)/1.086$ and $k_\lambda$ follows a Calzetti et al. (2000) dust attenuation law. The q factor has typically been measured from the same NB imaging used to select LAEs, leading to a selection against low values of q. The deep U-band data in the CANDELS catalog allow us to make a measurement of q from broadband imaging sensitive to Ly$\alpha$ that is uncorrelated with the original selection of the LAE sample. A q factor of unity implies that Ly$\alpha$ photons see the same dust column as UV continuum photons, whereas a q factor of zero points to Ly$\alpha$ photons seeing no dust extinction whatsoever. A negative q factor would imply Ly$\alpha$ enhancement by a top-heavy IMF, clumpy dust, or anisotropic radiative transfer, and $q > 1$ implies that Ly$\alpha$ photons see a greater dust column than UV continuum photons, as expected for resonant scattering.

Panel (a) of Figure 7 shows that the mode in q factors lies between 0 and 1. In a few of our objects, noise in the U-band photometry results in very low, even negative, values of q. There is no statistically significant correlation with $E(B-V)$ in panel (b), but panel (c) shows a correlation with EW in the expected direction.

We can compare this study’s q factors with that of a previous study by Finkelstein et al. (2007). In that study, a sample of four LAEs at $z = 4.4$ were studied. The q factors for that sample were all between 0 and 1, which is the most common range for q factors in our study, although we find a broader distribution.
4.3. Are LAEs on the SFR–$M_*$ “Main Sequence”?  

Lastly, in Figure 8, we explore the behavior of the individual object LAEs in the SFR–stellar mass relation. Our sample of LAEs are subject to a selection effect in both mass and SFR. The CANDELS catalog is mass selected, and the MUSYC catalog is SFR selected. SFRs below $1 \, M_\odot \, \text{yr}^{-1}$ will not be detectable in the NB survey. Galaxies with stellar mass $<3 \times 10^7 \, M_\odot$ would not be detectable in the CANDELS $H$-band catalog, even with the youngest ages observed in our sample. At stellar masses $>10^8 \, M_\odot$, where the SFR selection effect is unimportant, our LAEs appear to lie systematically above the galaxy main sequence (MS), implying larger SFRs than expected for galaxies of their mass.

Interestingly, the scatter seems to increase toward low masses. This could happen because a star formation episode of a few $M_\odot \, \text{yr}^{-1}$ will disturb a lower mass galaxy much more than a higher mass galaxy. In the analysis done in Ibar et al. (2013) and seen in their Figure 8, $H\alpha$ emitters at $z = 1.47$ also fall above the MS, though that study probes much higher masses. One should note that all other studies shown on this plot used the same (Salpeter) IMF, BC03 models, and constant star formation histories as used in this study, ruling out a systematic offset in SFR or $M_*$ due to varying assumptions.

4.4. Stacking Results

In order to better discuss how well the individual object SED fitting parameters agree with that of the various stacking techniques, three rows were added to Table 1, containing the average, median, and $1\sigma$ scatter of the individual object parameters. For simplicity, we define the failure of a stacking technique in a particular parameter to be when the average (median) of the individual results resides outside of the stacked parameter’s error region. Using this definition, the median HST image stack fails in stellar mass, the average NB image stack fails in all three parameters, and the median NB image stack fails in stellar mass. The average flux stack, median flux stack, and average HST image stack match the individual object results in all three SED fitting parameters. It should be stressed that we do not make direct comparisons between median stacks and average parameter values (or vice versa). The spread of individual object and stacked parameters can be seen in Figure 5.

While it is not entirely surprising that the NB-centered image stacks perform less well than the HST-centered image stacks, failure of any HST-centered stacking results is striking, since stacking using the CANDELS established positions for the individual objects is a best-case scenario for stacking accuracy. Lower precision position estimates would provide a substantial amount of error in any stacking analysis. It is possible that this discrepancy in the median image stack could be due to a majority of objects receiving contaminating flux from nearby objects in the IRAC infrared bands.

The average NB image stack is poor at estimating the average parameters for the sample, but other stacking techniques are relatively reasonable. While the error bars for the stacks shown in...
Figure 3 include sample variance, the large dispersion of properties is still not captured by any stacking method. As can be seen in Table 1 and Figure 5, there is a large dispersion in the values of mass, age, and $E(B - V)$ measured from individual objects, and this dispersion is far larger than the measurement uncertainties. A similarly large dispersion is seen in the SED properties of individual LAEs identified by the Hobby–Eberly Telescope Dark Energy Experiment Pilot Survey (A. Hagen et al., in preparation). As typically implemented, the stacking approach misses this dispersion, and this is a significant weakness versus fitting individual object SEDs. In circumstances where stacking must be used, however, there are statistical methods that can be used to infer the scatter in properties within the population. An upper limit to the uncertainty in the parameters inferred for the stacked population can be obtained by assuming that the dispersion in properties of the underlying population is much larger than the photometric errors (although one should then interpret a median as a typical object). In this case, the scatter in the parameters for a population of $N$ objects could be quantified as $\sqrt{N}$ times the uncertainty found for their average-stacked SED; this effectively turns the standard deviation of the average back into the standard deviation of the population. We do find that the sample variance found during our bootstrap resampling dominates the flux uncertainties on the stacked SED in most wavebands. However, applying this method to the average flux stack values in Figure 1 would significantly overestimate the observed scatter in individual object properties. A more sophisticated approach would be to use jackknife techniques (Lupton 1993) where stacks are made from each set of $N - 1$ galaxies and the resulting variance in best-fit parameters is turned into an estimate of the scatter of the population.

4.5. Scaled Flux Stacks

While straightforward calculations of average (median) stacks have been standard in the literature, it is also possible to scale all input fluxes (or images) to a common brightness e.g., in $H$ band, before stacking. This eliminates the influence of variation in overall brightness and focuses the stacking on determining an average (median) SED shape. To test the efficacy of this approach, we introduce scaled average and scaled median flux stacks. The scaled average (median) flux stack was created by determining the average (median) $H$-band flux and then multiplying each individual input SED and its uncertainties by the factor needed to match that value. As with other stacks, we include uncertainties due to sample variance by following the identical procedure in bootstrap simulations. By their definition, these new scaled average (median) stacked SEDs have the same average (median) as the corresponding flux-stacked SEDs. Their shapes turn out to be similar. Nonetheless, there is significantly less variance among the bootstrap samples due to having removed the impact of brightness variations, and the scaled flux stacks have uncertainties reduced by a factor of $\sim 2$ as a result. In Figure 5, we see that these scaled flux stacks perform quite well in tracing the properties of the sample, at least as well as the simple flux-stacked SEDs but with significantly reduced parameter uncertainties due to the reduced uncertainties in the SED. We conclude that scaling is a superior method for producing stacked SEDs and recommend its use in future studies.

4.6. Clean Stacks

As briefly discussed above, in an attempt to test whether or not IRAC contamination could be a strong source of error, we compiled an additional set of stacks using only the 11 individual objects whose CANDELS images were free of nearby sources of flux. This was determined by a visual inspection of the photometric images of each object. If an object’s image contained nearby extraneous sources, it was not included in these “clean” stacks. Our visual inspection ignored the apparent IRAC flux (or lack thereof) from the LAE itself, so that our “clean” sample should not be biased toward IRAC-bright LAEs. While one could design an automated criterion for this, our goal was to mimic the analysis performed in Lai et al. (2008), which used the same type of visual inspection. Our parallel analysis with these clean stacks produced results similar to the original stacks. Hence, there is no evidence that IRAC contamination causes the discrepancies in our flux-stacked versus image-stacked SEDs seen in Figure 3. However, the clean stack analysis lowers the sample size of an already sparse set of objects, so our clean stacking analysis might suffer from significant information loss. Clean stack results are included in Table 1.

The amount gained in a clean versus crowded analysis is questionable. Figure 2 shows the average and median stacked images for the entire sample and the clean sample. Inspection of the rightmost panel shows that a clean sample of 11 objects produces a less reliable median. Rather than excluding crowded regions, the CANDELS catalog photometry used to create our flux stacks utilizes TFIT (Laidler et al. 2007), which seems to perform better than careful qualitative crowdedness assessments by eye. This could cause the flux stacks to be more accurate than image stacks.

5. DISCUSSION AND CONCLUSIONS

Stacking techniques have been employed for studying low-N objects. However, some stacking methods may be more accurate at modeling the parameters of a sample than others. Through fitting individual SEDs of 20 LAEs at redshift 2.1 individually, and then fitting six types of stacked SEDs composed of the same 20 objects, we have tested the validity of these methods. Our choice of SED fitting under the assumption of a constant star formation history was made for two major reasons. First, our initial motivation for this paper came from inconsistencies in stacked LAE studies. Most of these studies used constant star formation histories in their SED fitting. Using other star formation histories in this study would make it more difficult to compare our results to other studies. Second, other functional forms of star formation histories, such as exponentially increasing or decreasing star formation histories and starburst star formation histories, do not significantly alter the SED fitting result for LAEs at $z = 2.1$ (Acquaviva et al. 2012b; Guaita et al. 2011).

We find that median flux stacking and average $HST$-centered image-stacking techniques correctly represent the age and stellar mass of a sample, while the median $HST$-centered image stack performs less adequately. This is a stark result, since our analysis serves as a best-case scenario for stacking. Stacking objects with CANDELS-defined coordinates maximizes the accuracy of our stacking sample. We further investigated this result by establishing a set of clean stacks, which are composed of objects free of nearby sources of flux, and with simulations. The clean stacking analysis sheds little light on the underlying reason for the discrepancies in image-stacked parameter estimates of a sample of individual objects. Future stacking analyses should be wary of disparities in analyses due to errors possibly brought on by the stacking method itself. The simulations show that median image stack photometry is biased considerably low
when centroiding errors are significant compared to the PSF. Though this effect is not, in practice, catastrophic for either the NB- or HST-centered image stacks, it provides a note of caution. We caution the reader of inconsistencies any stacking analysis may bring forth, and we recommend the use of a scaled stacking method where stacking is absolutely necessary.

We can compare our results to other LAE stacking studies that also used SED fitting to derive galaxy parameters. There are three such studies at \(z \approx 2.1\): Guaita et al. (2011), Nilsson et al. (2011), and Acquaviva et al. (2012b). The SED fitting results from these studies are summarized in Table 2. Each of these studies at \(z \approx 2.1\) is in agreement within the uncertainties with the typical age, stellar mass, and \(E(B - V)\) found in this study with the exception of Acquaviva et al. (2012b), who report stellar masses higher by a factor of two, and Nilsson et al. (2011), who report stellar masses higher by a factor of 50 and ages higher by a factor of five. The latter discrepancy is likely caused by their higher \(\text{Ly} \alpha\) flux detection limit and the sensitivity of the median or average flux stack, which reduces uncertainties significantly.

Our main conclusions are the following:

1. A lack of correlation between \(\text{Ly} \alpha\) EW and age implies either complicated radiative transfer mechanisms or an inappropriate assumption of a constant SFR in a starbursting sample.
2. The narrow distribution of \(q\) values peaking near 1 and lack of correlation between \(q\) and \(E(B - V)\) imply that radiative transfer mechanisms seem to be working to prevent Ly\(\alpha\) photons from resonantly scattering in dusty regions.
3. Our sample of LAEs lies systematically above the SFR–stellar mass relation galaxy “MS” and shows an increase in scatter above this relation at low mass. This may be caused by ongoing starbursts in these galaxies causing a greater excursion in their specific SFRs.
4. Though some types of stacking represent the average and median properties of a sample well, the large dispersion of individual object properties is obscured by stacking. We recommend a new approach using an \(H\)-band scaled median or average flux stack, which reduces uncertainties significantly.

We would like to thank both the CANDELS and MUSYC collaborations for making this work possible. We would also like to thank the anonymous referee for helpful comments that greatly improved the quality of the paper. This work is based on observations taken by the CANDELS Multi-Cycle Treasury Program with the NASA/ESA HST, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555. These observations are associated with programs GO-9352, GO-9425, GO-9583, GO-9728, GO-10189, GO-10339, GO-10340, GO-11359, GO-12060, and GO-12061. Observations have been carried out using the Very Large Telescope at the ESO Paranal Observatory under Program ID(s): LP168.A-0485. This work is also based in part on observations made with the Spitzer Space Telescope, which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA. Support for Program numbers 12060.57 and 12445.56 was provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Incorporated, under NASA contract NAS5-26555. The Institute for Gravitation and the Cosmos is supported by the Eberly College of Science and the Office of the Senior Vice President for Research at the Pennsylvania State University. This material is based on work supported by the National Science Foundation under CAREER grant AST-1055919 awarded to Eric Gawiser. This research has made use of NASA’s Astrophysics Data System.
