INSIGHTS ON THE FORMATION, EVOLUTION, AND ACTIVITY OF MASSIVE GALAXIES FROM ULTRA-COMPACT AND DISKY GALAXIES AT $Z = 2 - 3$

Tim Weinzirl[1], Shardha Jogee[1], Christopher J. Conselice[2], Casey Papovich[3],
Ranga-Ram Chary[4], Asa Bluck[5], Ruth Grützbauch[2], Fernando Buitrago[2], Bahram
Mobasher[6], Ray A. Lucas[7], Mark Dickinson[8], Amanda E. Bauer[9]
Draft version June 4, 2018

ABSTRACT

We present our results on the structure and activity of massive galaxies at $z = 1 - 3$ using one of the largest (166 with $M_\star \geq 5 \times 10^{10}\ M_\odot$) and most diverse samples of massive galaxies derived from the GOODS-NICMOS survey: (1) Sérsic fits to deep NIC3/F160W images indicate that the rest-frame optical structures of massive galaxies are very different at $z = 2 - 3$ compared to $z \sim 0$. Approximately 40% of massive galaxies are ultra-compact ($r_e \leq 2$ kpc), compared to less than 1% at $z \sim 0$. Furthermore, most ($\sim 65\%$) systems at $z = 2 - 3$ have a low Sérsic index $n \leq 2$, compared to $\sim 13\%$ at $z \sim 0$. We present evidence that the $n \leq 2$ systems at $z = 2 - 3$ likely contain prominent disks, unlike most massive $z \sim 0$ systems. (2) There is a correlation between structure and star formation rates (SFR). The majority ($\sim 85\%$) of non-AGN massive galaxies at $z = 2 - 3$, with SFR high enough to yield a 5σ (30μJy) 24 μm *Spitzer* detection have low $n \leq 2$. Such $n \leq 2$ systems host the highest SFR. (3) The frequency of AGN is $\sim 40\%$ at $z = 2 - 3$. Most ($\sim 65\%$) AGN hosts have disky ($n \leq 2$) morphologies. Ultra-compact galaxies appear quiescent in terms of both AGN activity and star formation. (4) Large stellar surface densities imply massive galaxies at $z = 2 - 3$ formed via rapid, highly dissipative events at $z > 2$. The large fraction of $n \leq 2$ disky systems suggests cold mode accretion complements gas-rich major mergers at $z > 2$. In order for massive galaxies at $z = 2 - 3$ to evolve into present-day massive E/S0s, they need to significantly increase (n, r_e). Dry minor and major mergers may play an important role in this process.

Subject headings: galaxies: bulges — galaxies: evolution — galaxies: formation — galaxies: fundamental parameters — galaxies: interactions — galaxies: structure

1. INTRODUCTION

Studies of high-redshift galaxies are essential for testing and constraining models of galaxy formation. Conventional wisdom suggests galaxies are assembled and shaped by a combination of mergers, smooth accretion, and internal secular evolution. Galaxies form inside cold dark matter halos that grow hierarchically through mergers with other halos and gas accretion (Somerville & Primack(1999); Cole et al. 2000; Steinmetz & Navarro 2002; Birnboim & Dekel 2003; Kereš et al. 2005; Dekel & Birnboim 2006; Dekel et al. 2009a; Dekel et al. 2009b; Kereš et al. 2005; Kereš et al. 2009; Brooks et al. 2009; Ceverino et al. 2010), while internal secular evolution (Kormendy & Kennicutt, 2004; Jogee et al. 2005) redistributes accreted material. Within the paradigm of hierarchical assembly, a number of issues remain. It is not known when and how the main baryonic components of modern galaxies (bulges, disks, and bars) formed, but the global stellar mass density rose substantially between $z \sim 1 - 3$, reaching $\sim 50\%$ of its present value by $z \sim 1$ (Dickinson et al. 2003b; Drory

et al. 2005; Conselice et al. 2007; Elsner et al. 2008; Pérez-González et al. 2008).

It is also not clear how high-redshift galaxies evolve into present-day galaxies. Complex baryonic physics such as mergers, gas dissipation, and feedback are all at work to an extent. There is also mounting evidence that cold-mode accretion (Birnboim & Dekel 2003; Kereš et al. 2005; Dekel & Birnboim 2006; Dekel et al. 2009a; Dekel et al. 2009b; Kereš et al. 2005; Kereš et al. 2009; Brooks et al. 2009; Ceverino et al. 2010) is important for building star-forming galaxies. This process is particularly effective in galaxies with halos of mass below $10^{12}\ M_\odot$ such that cold-mode accretion dominates the global growth of galaxies at high redshifts and the growth of lower mass objects at late times.

High-redshift galaxies are different from local galaxies. Within the framework of hierarchical assembly, early, high-redshift galaxies are expected to be smaller, at a given mass, than their present-day counterparts. The size difference is predicted to be a factor of a few at $z = 2 - 3$ (Loeb & Peebles 2003; Robertson et al. 2006; Khochfar &

[1] Department of Astronomy, University of Texas at Austin, Austin, TX

[2] University of Nottingham, School of Physics & Astronomy, Nottingham NG7 2RD, U.K.

[3] George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, Department of Physics and Astronomy, Texas A&M University, 4242 TAMU, College Station, TX 77843

[4] U.S. Planck Data Center, MS220-6 Caltech, Pasadena, CA 91125

[5] Gemini Observatory, Hilo, HI 96720, USA

[6] Department of Physics and Astronomy, University of California, Riverside, CA 92521

[7] Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD, 21218

[8] NOAO, 950 N. Cherry Avenue, Tucson, AZ 85719, USA

[9] Australian Astronomical Observatory, P.O. Box 296, Epping, NSW 1710, Australia

Silk 2006; Naab et al. 2007). Several recent studies using rest-frame optical data provide evidence for size evolution among massive galaxies (Guzman et al. 1997; Daddi et al. 2005; Trujillo et al. 2006, 2007; Zirm et al. 2007; Toft et al. 2007; Longhetti et al. 2007; Cimatti et al. 2008; Buitrago et al. 2008; van Dokkum et al. 2008, 2010; van der Wel et al. 2011). Aside from size evolution, there is some evidence that the nature of red galaxies changes at higher redshift. At $z \lesssim 1$, the red sequence primarily consists of old, passively evolving galaxies (Bell et al. 2004). Among extremely red galaxies (EROs) at $z = 1 - 2$, less than 40% are morphologically early types (Yan & Thompson 2003; Moustakas et al. 2004). It is well known that star formation rates were more intense at higher redshift (Daddi et al. 2007; Drory & Alvarez 2008), and a link has been found between star formation, size, and morphology at $z \sim 2.5$. Toft et al. (2007) and Zirm et al. (2007) find from NICMOS rest-frame optical imaging that blue star-forming galaxies are significantly more extended than red quiescent galaxies. Additionally, examples of rapidly star-forming galaxies (SFR $\sim 50 - 200 \ M_\odot \ \mathrm{yr}^{-1}$) at $z \sim 2 - 3$, whose ionized gas kinematics are consistent with turbulent rotating disks, are found in the SINS survey (Förster Schreiber et al. 2009; Genzel et al. 2008; Shapiro et al. 2008).

Progress on understanding the evolution of massive galaxies at high redshift has been hindered by significant observational challenges. The deep optical surveys carried out by $HST \ ACS$, such as the Hubble Ultra Deep Field (HUDF, Beckwith et al. 2006) and the Great Observatories Origins Deep Survey (GOODS, Giavalisco et al. 2004), trace rest-frame optical galaxy morphology only out to $z \sim 1$. At higher z, bandpass shifting effects cause filters to trace progressively bluer bands, and optical filters trace rest-frame UV at $z \gtrsim 2$. UV light traces massive young stars, but manages to set few constraints about the overall mass distribution, making it difficult to probe the structure and mass of galaxy components at early epochs.

Without high-resolution, deep, rest-frame optical imaging, it is not possible to robustly compare structural parameters in galaxies across redshift. NIR imaging is required to probe the rest-frame optical at $z \sim 1 - 3$. Unfortunately, deep NIR imaging with HST has been completed for a limited number of galaxies over relatively small fields and small volumes at $z > 1$, with most pointings being within the Hubble Deep Fields and the Hubble Ultra Deep Field due to the inefficiency of the NICMOS camera in covering large areas (e.g., Dickinson et al. 2004; Thompson et al. 2005; Zirm et al. 2007; van Dokkum et al. 2008). While ground-based NIR imaging surveys (e.g., Kajisawa et al. 2006; Retzlaff et al. 2010) efficiently cover wide fields at resolutions almost comparable to HST NICMOS, the depths reached are at least an order of magnitude shallower.

A large area, high-resolution, deep, space-based NIR survey would be bountiful for galaxy formation studies. The GOODS-NICMOS Survey (GNS; Conselice et al. 2011), covering 44 arcmin^2 of the GOODS fields with NIC3, is a strong first effort. The GOODS-North and GOODS-South are among the best-studied regions in the sky and are a natural choice for such a survey. The GOODS fields already have deep data from $HST \ ACS$

(Giavalisco et al. 2004), $Spitzer$ IRAC/MIPS (Dickinson et al. 2003a), and Chandra (Giacconi et al. 2002; Alexander et al. 2003; Lehmer et al. 2005; Luo et al. 2008), among others. GNS consists of 60 pointings centered on massive ($M_\star > 10^{11} \ M_\odot$) galaxies at $z > 2$, observed to a depth of $H = 26.8$ magnitudes. The value of GNS lies in the fact that the target fields were optimized to include massive galaxies selected by multiple methods in order to create an unbiased sample (see Conselice et al. 2011). There are additional massive galaxies in each field beyond the 60 main targets, so that there are 82 galaxies with $M_\star \geq 10^{11} \ M_\odot$ at $z = 1 - 3$ across all pointings. Thus, the GNS data contain one of the largest samples of very massive galaxies at high redshift with rest-frame optical imaging, and they robustly probe massive galaxies when the Universe was less than 1/3 of its current age, during the epoch of bulge and disk formation.

The goal of this work is to investigate the evolution of massive galaxies over $z = 1 - 3$ with this unique sample. We take advantage of the existing rich ancillary data to derive star formation rates (SFR) from 24 μm detections and look for AGN activity based on X-ray detections and mid-IR SEDs. We correlate rest-frame optical structural parameters with SFR to gain insight into how massive galaxies are expected to evolve.

The plan of this paper is as follows. We discuss the data and sample properties in §2. In §3 we describe the measurement of structural parameters, and in §3.2 we make a detailed comparison with $z \sim 0$ galaxies of similar stellar mass. A detailed artificial redshifting experiment is conducted in §3.3.1 to explore the impact of instrumental and redshift-dependent effects on structural parameters. In §4, we measure star formation properties based on $Spitzer$ MIPS 24 μm detections and discuss how they relate to structural properties. Estimates of the mass and fraction of cold gas in massive star-forming galaxies at $z = 2 - 3$ are presented in §5. In §6, we use a variety of techniques (X-ray properties, IR power-law, and IR-to-optical excess) to identify AGN and consider how galaxy activity relates to galaxy structure. Finally, in §7 and §8, we discuss and summarize our results. All calculations assume a flat ΛCDM cosmology with $\Omega_\Lambda = 0.7$ and $H_0 = 70$ km s^{-1} Mpc^{-1}.

2. DATA AND SAMPLE

2.1. Observations and Pointing Selections for GNS

Our data comes from the GOODS-NICMOS Survey (GNS; Conselice et al. 2011). GNS is a deep, 180-orbit survey with the HST NICMOS-3 camera in the F160W (H) band that probes optical light from galaxies between $z \sim 1 - 3$. The coverage extends over both ACS GOODS fields and is divided between 60 pointings centered on massive $M_\star \geq 10^{11} M_\odot$ galaxies at $z > 2$. Each pointing covers $51''.2$ x $51''.2$ and was observed to a depth of three orbits in nine exposures of ~ 900 seconds (~ 135 minutes per pointing). A total of ~ 8300 sources were detected across an effective area of ~ 44 arcmin^2. The 5σ limiting magnitude for an extended source with a $0''.7$ diameter is $H=26.8$ (AB). The NIC-3 images were drizzled with a pixfrac of 0.7 and output platescale of $0''.1$. The NIC3 camera is currently out of focus, and after detailed investigation (see §A), we find the point spread function (PSF) spans a full

width half maximum (FWHM) of $0\rlap{.}''26 - 0\rlap{.}''36$ with a mean value of $0\rlap{.}''3$.

The 60 GNS pointings were planned by identifying massive galaxies having a photometric redshift of $1.5 < z < 2.9$ and stellar mass $M_\star > 10^{11}~M_\odot$ via three color selection criteria. The target galaxies include Distant Red Galaxies (DRGs, Papovich et al. 2006), Extremely Red Objects (EROs, Yan et al. 2004), and BzK-selected galaxies (Daddi et al. 2004). All of these methods are designed to select red dusty or red passively evolving galaxies. DRGs have evolved stellar populations that are identified with $J - K > 2.3$ (Vega mag). EROs are selected based on $Spitzer$ and NIR data via $f_\nu(3.6\mu m)/f_\nu(z850) > 20$. This selection is sensitive to red populations that are either old or reddened, so EROs contain a mixture of young and old stellar populations. BzK galaxies are selected based on the quantity $BzK \equiv (z - K)_{AB} - (B - z)_{AB}$. Galaxies with $BzK > -0.2$ at $z > 1.4$ are identified as star-forming galaxies. Redder and possibly more evolved galaxies are identified with $BzK < -0.2$ and $(z - K)_{AB} > 2.5$. The final pointings were designed to include at least one red massive galaxy and to also maximize the total number of additional galaxies (e.g., Lyman-break galaxies and sub-mm galaxies) within each pointing.

2.2. *Our Sample of Massive Galaxies at z=1-3*

The sample of massive galaxies that we work with in this paper is not limited to the original color-selected massive galaxies at $z > 1.5$ defining the original 60 GNS pointings. Instead, our sample of massive galaxies at $z = 1 - 3$ is derived from the set of all galaxies mapped with NIC3/F160W across the 60 fields, and for which a reliable stellar mass and photometric redshift was estimated by Conselice et al. (2011), based on SED fits to the NIC3/F160W and optical imaging. A detailed description of how these quantities were estimated is in Conselice et al. (2011), and we only briefly summarize the methodology here.

The source extraction catalog for the NICMOS images across the 60 pointings of the GNS survey contains ~ 8300 sources with $H < 28$ and $V < 30$. For those galaxies detected in the $ACS~BViz$ and NICMOS H bands, we use the available photometric redshifts and stellar masses from Conselice et al. (2011). Photometric redshifts were determined by fitting template spectra to the $BVizH$ data. Stellar masses were measured by fitting the $BVizH$ magnitudes to a grid of SEDs generated from Bruzual & Charlot (2003) stellar population synthesis models, assuming a Salpeter IMF[10]. The grid includes different colors, ages of stellar populations, metallicities, dust content, and star formation histories as characterized by exponentially declining models. In general, the stellar masses derived depend on the SED used and the assumptions used in the SED modeling, such as the IMF, the metallicity, the extinction law, and star formation history (e.g., Borch et al. 2006; Marchesini et al. 2009; Conselice et al. 2011). The typical uncertainty in stellar mass across the sample is a factor of $\sim 2 - 3$.

In order to account for a small number (15) of additional massive ($M_\star \geq 5 \times 10^{10}~M_\odot$) red systems, which are un-

detected in the GOODS $ACS~BV$ and therefore do not have viable stellar masses from the above techniques, we use available masses and redshifts (Buitrago et al. 2008; Bluck et al. 2009) based on deep ground-based $RIJHK$ data along with ACS iz data, where available. Photometric redshifts are determined with a mixture of techniques (e.g., neural networks and Bayesian techniques) described more fully in Conselice et al. (2007). Stellar masses were measured from these data with uncertainties of a factor of $\sim 2 - 3$ with the multi-color stellar population fitting techniques from Conselice et al. (2007, 2008). As with the larger sample described above, a stellar mass is produced by fitting model SEDs to the observed SED for each galaxy. A Salpeter IMF is assumed, and the SED grids are constructed from Bruzual & Charlot (2003) stellar population synthesis models.

From the sample of galaxies with photometric redshifts and stellar masses determined as described above, we define the sample of massive galaxies used in this paper. We restrict our analysis to the redshift interval $z = 1 - 3$ over which our NIC3/F160W images probe the rest-frame optical light in order to avoid bandpass shifts into the rest-frame UV. This ensures that we measure all structural parameters in the rest-frame optical across $z = 1 - 3$, thereby reducing bandshift biases (see §3.1 for a quantitative estimate). Although the mass functions calculated for GNS by Mortlock et al. (2011) show that the mass completeness limit is $\sim 3 \times 10^9~M_\odot$ at $z \sim 3$, we apply a higher mass cut of $5 \times 10^{10}~M_\odot$ as our interest is specifically with the most massive galaxies.

Our final sample consists of the 166 (82) massive galaxies with $M_\star \geq 5 \times 10^{10}~M_\odot$ ($M_\star \geq 1 \times 10^{11}~M_\odot$) and $z = 1 - 3$. This is the largest HST-based dataset with rest-frame optical imaging of massive galaxies over $z = 1 - 3$. The galaxies with $M_\star \geq 10^{11}~M_\odot$ from Buitrago et al. (2008) are part of the sample. The other previous HST NICMOS studies (e.g., Toft et al. 2007; Zirm et al. 2007; van Dokkum et al. 2008) each contain, at most, $10 - 20$ systems with $M_\star \geq 10^{11}~M_\odot$. The full distributions of apparent H and V magnitude, stellar mass, and redshift for this sample are shown in Figure 1.

Figure 2 shows a comparison of the galaxy stellar mass function (SMF) of our GNS-based sample to the published SMF of other NIR-selected samples in the literature, such as the K-selected samples of Fontana et al. (2006), Kajisawa et al. (2009) and Marchesini et al. (2009), as well as the IRAC-selected sample of Pérez-Gonzalez et al. (2008). This figure essentially shows that for the mass range ($M_\star \geq 5 \times 10^{10}~M_\odot$) relevant for the GNS-based sample used in our paper, there is good agreement between the SMF of our sample and those from these four studies. In particular, at $M_\star \geq 5 \times 10^{10}~M_\odot$, the top panel shows that there is very good agreement with our sample, Fontana et al. (2006), and Pérez-Gonzalez et al. (2008) for three different redshift bins between $z = 1.5$ and $z = 3.0$. In the lower panel, at $M_\star \geq 5 \times 10^{10}~M_\odot$, the average SMF from Kajisawa et al. (2009) agrees with that of our sample within a factor of ~ 2 over $1.5 < z < 2.5$. The SMF from our GNS-based sample and Marchesini et al. (2009) show good agreement at $z = 2 - 3$, and are

---

[10] In §4 we use a Chabrier IMF for SFR estimates. Using a Chabrier IMF rather than a Salpeter IMF in estimating the stellar mass would lower the values by a factor of 0.25 dex or less.

slightly offset at z=1.3 to 2.0. The small offset may not be statistically significant if one includes all the sources of error. The error bars on the GNS mass functions include Poisson errors only. Marchesini et al. (2009) show that the dominant sources of error regarding stellar mass functions are cosmic variance and systematics from the assumptions used in the SED modeling. For a discussion of the SMF for lower mass ($M_\star \geq 5 \times 10^{10}$ $M_\odot$) galaxies, which are not included in the sample used in this paper, we refer the reader to Mortlock et al. (2011).

In our sample of 166, massive galaxies, spectroscopic redshifts are available for 44 galaxies ($26.5 \pm 3.4\%$ of the sample). These 44 galaxies are all bright with $V \leq 27$ and $H_{AB} < 23$. Among these 44 galaxies, the median photometric redshift error is $\delta z/(1+z) = 0.071$ (Grützbauch et al. 2010), $7/44$ ($15.9 \pm 5.5\%$) have $\delta z/(1+z) > 0.2$, and none have $\delta z/(1+z) > 0.5$.[11] For the remaining $122/166$ ($73.5 \pm 3.4\%$) of our sample galaxies without spectroscopic redshifts, photometric redshifts are used. Among these 122 galaxies, 60 ($49.2 \pm 4.5\%$) are fainter than $V > 27$, and the uncertainties in photometric redshifts may be larger than the median value of 0.071 cited above.

2.3. Properties and Selection Biases in the Sample

We estimate the number density of massive ($M_\star \geq 5 \times 10^{10}$ $M_\odot$) galaxies over $z = 2 - 3$ to be $\sim 5 \times 10^{-4}$ Mpc^{-3} (see Conselice et al. 2011 for a detailed discussion of the number density of massive galaxies in the GNS sample). The corresponding stellar mass density is $\sim 6 \times 10^7$ $M_\odot$ Mpc^{-3}. The massive GNS galaxies are collectively 10-100 times more abundant than SMGs, which have space densities of $10^{-5} - 10^{-6}$ Mpc^{-3} at $z \sim 2 - 3$ (Blain et al. 2002). Rather, the number density is in agreement with published values (Daddi et al. 2005; 2007) for other passively evolving and star-forming galaxies at $z \sim 2$.

How does our sample break down in terms of the typical color-selection methods, which are usually used to identify massive high redshift galaxies? About 63% (104/166) of this final sample is listed in existing catalogs for DRG (Papovich et al. 2006), BzK (Daddi et al. 2004), or ERO (Yan et al. 2004) galaxy populations. There are 8, 9, and 43 sources that are uniquely listed in one of the DRG, Bzk, or ERO galaxy catalogs, respectively. An additional 44 sources are listed in two or more of these catalogs. About 37% (62/166) sources were not previously identified as DRG, ERO, or BzK galaxies.

What are the selection biases impacting our sample? General biases in the selection of massive galaxies in the GNS survey have been discussed in Conselice et al. (2011), and we only discuss below the points relevant for our sample.

The 60 GNS pointings were selected to include massive galaxies identified via three color methods (DRG, BzK, and IERO). Combining all three color criteria, rather than using any single one, is already a step forward compared to many earlier studies because no single criterion would isolate a complete sample of massive galaxies (e.g., van Dokkum et al. 2006; Conselice et al. 2011). These three criteria all pick massive galaxies with red observed colors, but due to the range of criteria involved, they can pick both red dusty systems and red evolved stellar populations.

Another key step that makes our study less biased towards a specific type of massive galaxy is that our working sample at $z = 1 - 3$ is neither limited to nor defined by the original color-selected massive galaxies. Rather, it is derived from all galaxies within the survey area that are bright enough to be mapped with NIC3/F160W and for which a reliable stellar mass and photometric redshift could be determined by Conselice et al. (2011), as outlined in §2.2. The first potential bias in this final sample is introduced by excluding galaxies that are undetected by NIC3/F160W. The second potential bias is introduced by excluding detected galaxies for which no reliable stellar mass and photometric redshift could be determined. For instance, ultra-dusty galaxies, may not be detected in enough of the optical bands to allow a photometric redshift to be reliably estimated.

We assess the impact of the second bias by estimating how many massive galaxies we might miss due to the lack of available photometric redshift and stellar masses. Of the 8300 sources detected by GNS, 1076 have no photometric redshift and stellar mass measurements. Most (68%) of these 1076 sources are fainter ($H > 25$) than our sample of massive galaxies (Figure 1). Among GNS objects as bright ($H < 25$) as our sample of massive galaxies, only 8.5%, or 349/4083 have no redshift or stellar mass measurements. Furthermore, not all 8.5% of these bright ($H < 25$) sources will be massive, so that this fraction represents an upper limit on the sources we might not include in our sample due to the lack of a photometric redshift or stellar mass measurements.

We next discuss the impact of the first potential bias and the type of objects the GNS survey might not detect. It is relevant to ask whether we might miss galaxies with blue observed colors. We believe this is not the case for the following reasons. As discussed above, our working sample is not strongly biased against galaxies with blue observed colors because it is not limited to those massive galaxies selected by the three color methods (DRG, BzK, and IERO) that preferentially pick galaxies with red observed colors. Secondly, Conselice et al. (2011) explicitly show that many galaxies with blue observed ($z - H$) colors, which would have been undetected by these color selections, do get included in this final sample of massive galaxies for the GNS survey. Nearly all known Lyman Break Galaxies or BX/BM objects (Reddy et al. 2008) at $z = 2 - 3$ in the GNS fields are detected by the GNS NIC3/F160W imaging (Conselice et al. 2011).

In terms of rest-frame colors, rather than observed colors, it is also important to note that the galaxies detected by GNS at $z = 1 - 2$ or $z = 2 - 3$ include systems with both blue and red rest-frame $U - V$ colors. The rest-frame $U - V$ color ranges from about $-0.4 - 2.1$ for galaxies in the stellar mass range $M_\star \sim 10^9 - 10^{12}$ $M_\odot$ (Figure 3). The systems with blue rest-frame $U - V$ colors are preferentially at low masses, while GNS galaxies with $M_\star \geq 1 \times 10^{11}$

---

[11] While figure 6 of Conselice et al. (2011) shows that $\sim 15 - 20\%$ of bright ($20 < H_{AB} < 23$) galaxies with spectroscopic redshifts are catastrophic outliers in photometric redshift with $\delta z/(1+z) > 0.5$, it should be noted that there are no catastrophic outliers with such large $\delta z/(1+z) > 0.5$ among the 44 galaxies with spectroscopic redshifts in our sample of massive ($M_\star \geq 5 \times 10^{10} M_\odot$) galaxies at $z = 1 - 3$. The outliers with $\delta z/(1+z) > 0.5$ in the GNS survey have stellar masses below the cutoff value of our sample or/and lie outside its redshift range.

$M_\odot$ at $z = 2 - 3$ have preferentially red rest-frame $U - V$ colors, in the range of 1.0 to 1.7. These inherently red rest-frame $U - V$ colors of the massive galaxies at $z = 2 - 3$ could be due to a combination of old stellar populations and dusty young star-forming regions. We checked that the colors are consistent with stellar population synthesis models (based on Bruzual & Charlot (2003) and assuming a Chabrier IMF, an exponentially declining star formation history with a 100 Myr e-folding time). We find that even without dust extinction $U - V$ color rises rapidly. Assuming solar metallicity, $U - V$ is already ~ 1 at an age of 0.5 Gyr and reaches ~ 1.6 at 2 Gyr. For the case with dust extinction and an optical depth of 1, $U - V$ is ~ 1.1 after 0.5 Gyr and ~ 1.8 after 2 Gyr.

3. STRUCTURAL PROPERTIES OF MASSIVE GALAXIES

3.1. *Structural Decomposition*

We characterize the massive GNS galaxies with structural decomposition. Ideally, one would like to fit multiple components (bulge, disk, bar, nuclear point source, etc.) in the decomposition, but the $0\rlap{.}''3$ diameter (or full width half-maximum) of the PSF (corresponding to ~ 2.4 kpc at $z = 1 - 3$) prevents such detailed decompositions[12]. Instead, we choose to fit the 2D light distributions with only single Sérsic (1968) $r^{1/n}$ profiles, which have the form

$$I(r) = I_e \exp \left(-b_n \left[\left(\frac{r}{r_e} \right)^{1/n} - 1 \right] \right), \qquad (1)$$

where I_e is the surface brightness at the effective radius r_e and b_n is a constant that depends on Sérsic index n. Knowledge of the PSF is important for deriving structural parameters. We model the PSF (Appendix A) while taking into account both the variation in PSF with position on the NIC3 field and the dependence on the `drizzle` algorithm. We find a range in PSF FWHM of $\sim 0\rlap{.}''26 - 0\rlap{.}''36$.

It is clear that a single Sérsic profile is not a complete indicator of overall galaxy structure. For instance, in detailed images of nearby galaxies, the best-fit index n for a single Sérsic profile does not always correlate with the bulge Sérsic index obtained with 2D bulge-disk or bulge-disk-bar decomposition (Weinzirl et al. 2009). However, the single Sérsic index n is on average a good way to separate disk-dominated galaxies from the class of luminous spheroidal and bulged-dominated galaxies (see §3.3.1), and in studies of high-redshift galaxies the criterion $n \lesssim 2$ is often used to separate spirals or disk galaxies from ellipticals (e.g., Ravindranath et al. 2004; Bell et al. 2004; Jogee et al. 2004; Barden et al. 2005; Trujillo et al. 2007; Buitrago et al. 2008).

The NIC3/F160W images of the 166 sample galaxies were fit with a single Sérsic component using GALFIT (Peng et al. 2002). In each image, objects that were near, but not blended with, the primary source were masked out. For the fraction ($\sim 15\%$) of the primary galaxies that were blended or overlapping with another galaxy identified in the source extraction catalog, the blended sources were each fitted simultaneously with a separate Sérsic profile. Some fraction of primary galaxies appeared morphologically disturbed ($\sim 8\%$, see Figure 4 and §3.2), but these

were fitted with only a single Sérsic profile as they only counted as a single galaxy in the source extraction catalog.

Bandpass shifting causes the H-band central wavelength to move from 4000-8000 $\mathring{A}$ over $z = 1 - 3$. The $z = 1 - 2$ and $z = 2 - 3$ bins used in Figure 5, for example, correspond to 5333-8000 $\mathring{A}$ (I-band) and 4000-5333 $\mathring{A}$ (B-band), respectively. Even with the bandpass shifting, comparing the structural parameters (n, r_e) measured in these two bands to each other and to parameters of $z \sim 0$ galaxies measured in rest-frame B is a vast improvement over previous studies forced to compare the rest-frame UV at $z > 1$ to the rest-frame optical at $z < 1$. The systematic effects resulting from H-band changing from B to I-band over $z = 1 - 3$ are small, as can be inferred from studies of nearby galaxies. Graham (2001) presents bulge-disk decompositions of local $z \sim 0$ galaxies based on images in the B and I bands. The median ratio in B-band/I-band disk scalelength is 1.13, so that the disks are measured to be slightly larger in the B-band. If similar errors apply here, then the bias r_e due to bandpass shifting is on the order of 10%.

Another important consideration is the effect of potential AGN on the structural fits. When fitting high resolution images of nearby galaxies, it is well known that fitting a galaxy that hosts a point source with a single Sérsic component will lead to an artificially high Sérsic index n (typically $n > 4$; e.g., Weinzirl et al. 2009; Pierce et al. 2010). If a point source is added to the Sérsic model, the index n of the Sérsic component falls to more reasonable values. In the case of the massive GNS galaxies at $z = 1 - 3$, we expect that the low resolution ($0\rlap{.}''3$, corresponding to 2.5 kpc at $z \sim 2$) of the NIC3/F160W images will reduce the effect of potential point sources on the structural decomposition. However, for completeness, we have fitted all the galaxies at $z = 1 - 3$ in which a potential AGN was identified via a variety of techniques (§6) with both a Sérsic component and a point source. The fractional luminosity of the point source components, or PSF/Total light ratio, ranges from 1-46%, with a median of 10%. As expected, including the point source produces generally small changes in (n, r_e) and goes in the direction of lowering n and enlarging r_e. Overall, our results are not biased by the presence of AGN. In the rest of the paper, we therefore choose to use the structural parameters for a single Sérsic component fit.

3.2. *Derived Structural Properties at $z = 2 - 3$*

The results of the structural fits to the NIC3/F160W images of the 166 sample galaxies are shown in Table 1, Figure 4, Figure 5, and Figure 6.

Figure 4 shows examples of massive ($M_\star \geq 5 \times 10^{10}$ $M_\odot$) galaxies at $z = 2 - 3$ with different ranges of Sérsic index n and effective radius r_e. The majority ($\sim 82\%$; Table 1) of the massive GNS galaxies at $z = 2 - 3$ have $r_e \leq 4$ kpc. In such systems, structural features are generally hard to discern due to resolution effects, so that systems appear fairly featureless (top 4 rows of Figure 4). In the small fraction of massive galaxies at $z = 2 - 3$ with large $r_e > 4$ kpc, one can discern some structural

---

[12] For the more extended galaxies multiple components (e.g., bulge and disk) decomposition was attempted with limited success and this is discussed in §7.1.

features such as an elongated bar-like feature or a combination of a central condensation surrounded by a more extended lower surface brightness component, reminiscent of a bulge and disk (5th row). Row 6 contains morphologically disturbed systems. The fraction of such systems is small, only $\sim 8\%$, but this is a lower limit given redshift-dependent effects such as degraded physical resolution and surface brightness dimming.

The lower two rows of Figure 5 shows the rest-frame optical Sérsic index n and effective radius r_e for the samples of massive galaxies at $z = 1 - 2$ and $z = 2 - 3$. For comparison, the top row of Figure 5 also shows the rest-frame optical structural parameters for $z \sim 0$ galaxies of similar stellar mass taken from Allen et al. (2006), who performed a single component Sérsic fit to B-band images of galaxies in the Millennium Galaxy Catalogue (MGC), a large ground-based imaging and spectroscopic survey over 37.5 deg^2 (Liske et al. 2003; Driver et al. 2005). It is clear from Figure 5, Figure 6, and Table 1 that the massive galaxies at $z = 2 - 3$ are strikingly offset toward lower (n, r_e) compared to the massive ~ 0 galaxies.

Firstly, we find that *the majority (64.9 $\pm$ 5.4% for $M_\star \geq 5 \times 10^{10}$ $M_\odot$, and 58.5 $\pm$ 7.7% for $M_\star \geq 10^{11}$ $M_\odot$) of massive galaxies at $z = 2 - 3$ have low $n \leq 2$, while the fraction at $z \sim 0$ is five times lower.* We will present evidence in §7.1 that most of the massive systems with a low $n \leq 2$ harbor a massive disk component, so that our results point to the *predominance of disk-dominated systems among massive galaxies at $z = 2 - 3$.*

Secondly, we also find that massive galaxies at $z = 2-3$ typically have smaller r_e than massive galaxies at $z \sim 0$. *In particular, $\sim 40\%$ (39.0 $\pm$ 5.6% for $M_\star \geq 5 \times 10^{10}$ $M_\odot$ and 39.0 $\pm$ 7.6% for $M_\star \geq 1 \times 10^{11}$ $M_\odot$) of massive galaxies at $z = 2-3$ are ultra-compact ($r_e \leq 2$ kpc), compared to less than one percent at $z \sim 0$.* The massive ultra-compact ($r_e \leq 2$ kpc), galaxies at $z = 2 - 3$ have few counterparts among $z \sim 0$ massive galaxies.

The population of galaxies with low $n \leq 2$ and the population of ultra-compact ($r_e \leq 2$ kpc) galaxies show limited overlap. Only 28.0 $\pm$ 6.4% of the systems with low $n \leq 2$ are ultra-compact and the remaining majority (72.0$\pm$6.3% for $M_\star \geq 5 \times 10^{10}$ $M_\odot$, and 75.0$\pm$8.8% for $M_\star \geq 10^{11}$ $M_\odot$) are extended ($r_e > 2$ kpc). Conversely, among the ultra-compact ($r_e \leq 2$ kpc) systems, nearly half (46.7$\pm$9.1% for $M_\star \geq 5 \times 10^{10}$ $M_\odot$, and 37.5 $\pm$ 12.1% for $M_\star \geq 10^{11}$ $M_\odot$) have low $n \leq 2$.

Figure 7 further illustrates the striking difference between massive galaxies at $z = 2 - 3$ and $z \sim$ 0 by comparing their effective radius r_e and their mean rest-frame optical surface brightness $< \mu_e >$ within r_e. The value of $< \mu_e >$ was measured from the extinction-corrected rest-frame B-band light within r_e and is defined as:

$$\mu_e = B_{corr} + 2.5\log_{10}(2\pi r_e^2) - 10\log_{10}(1 + z) \qquad (2)$$

where B_{corr} is the extinction-corrected, rest-frame apparent B magnitude and $-10\log_{10}(1 + z)$ and is the correction for surface brightness dimming. The MGC galaxies at $z \sim 0$ are corrected only for Galactic extinction, while for the GNS galaxies the correction includes Galactic and internal extinction. *The mean rest-frame optical surface brightness can be 2.0 to 6.0 magnitudes brighter for the massive galaxies at $z = 2 - 3$ than for $z \sim 0$ massive galaxies.* This is due to their smaller sizes and likely

differences in the age of the stellar populations. The high mean rest-frame optical surface brightness of the massive galaxies at $z = 2 - 3$ translates into high mean stellar mass densities, and suggests that highly dissipative events played an important role in their formation (see §7).

It is worth noting that the use of deeper images for the $z \sim 0$ galaxies could make the large offset in (n, r_e) at $z = 2-3$ versus $z \sim 0$ even stronger. The MGC B-band images have a median sky background of 22 mag/arcsec^2. Low surface brightness halos may be detected around some of the $z \sim 0$ galaxies in deeper exposures. This is true for some massive elliptical and cD galaxies, and in these cases the (n, r_e) are significantly boosted if the halo is region is also fitted (Kormendy et al. 2009).

How do these results compare with earlier studies? While many of the earlier studies focused on small samples, this work is a step forward because of the improved number statistics that come with an unbiased and complete sample of massive galaxies. The observed *apparent* size evolution in our data generally agrees with results reported in other studies of massive galaxies (e.g., Daddi et al. 2005; Trujillo et al. 2007; Zirm et al. 2007; Toft et al. 2007; Buitrago et al. 2008; van Dokkum et al. 2008; 2010; Williams et al. 2010).

The ratio in r_e of high-redshift galaxies with respect to $z \sim 0$ galaxies, or $r_e/r_{e,z\sim0}$, can be modeled as a power law in redshift of the form $\alpha(1 + z)^\beta$, where α and β are constants. Using the $z \sim 0$ massive ($M_\star \geq 5 \times 10^{10}$ $M_\odot$) MGC galaxies as the normalization, we measure α and β for different subsamples of the massive galaxies and summarize the results in Table 2. For all galaxies the slope β is -1.30 for a fit over $z = 0 - 3$. For disk-like $n \leq 2$ galaxies β is also -1.30, and for $n > 2$ galaxies β is -1.52. For non-AGN host galaxies with SFR_{IR} detected above the 5σ detection limit (see §4), β is -1.21, while for non-AGN host galaxies not detected by *Spitzer* the slope is substantially steeper (-1.67).

These results are comparable to the findings of earlier studies. Buitrago et al. (2008) show for massive ($M_\star \geq 10^{11}$ $M_\odot$) galaxies over $z = 0-3$ that β varies from -0.8 for $n < 2$ disk-like galaxies to -1.5 for $n > 2$ spheroidal galaxies. Williams et al. (2010) find β is -0.88 for all massive ($M_\star \geq 6.3 \times 10^{10} M_\odot$) galaxies over $z = 0.5 - 2$. van Dokkum et al. (2010) find a slope of -1.27 for massive ($M_\star \geq 10^{11}$ $M_\odot$) galaxies over z=0-2, which is a good match to our slope (-1.30) for massive ($M_\star \geq 5 \times 10^{10}$ $M_\odot$) galaxies of all n over $z = 0 - 3$. Compared to massive $z \sim 0$ galaxies, the implied mean size evolution is a factor of ~ 4 from $z = 2 - 3$ and a factor of ~ 3 from $z = 1 - 2$. In order to determine whether this apparent size evolution is real, one needs to address a number of systematic effects, as outlined in the next section.

3.3. *Impact of Systematic Effects on Structural Properties*

In the previous section we found that the massive galaxies at $z = 2 - 3$ are strikingly offset toward lower (n, r_e) compared to the massive ~ 0 galaxies. It is relevant to ask whether the large fraction of low (n, r_e) systems we observe among massive galaxies at $z = 2 - 3$, compared to massive galaxies at $z \sim 0$ is real or due to a number of systematic effects. We address the most important effects

in the main text and include the others in Appendix B. We consider the issues listed below:

1. Is it possible that the distribution of (n, r_e) for massive galaxies at $z \sim 0$ and at $z = 2 - 3$ is intrinsically similar, but that some selection effects at $z = 2.5$ is making us preferentially detect the compact low n systems, thereby causing an artificial excess of the latter? We argue that this is very unlikely because even if we take all the massive compact low n systems at $z \sim 0$, and appropriately scale them for the difference in number density between $z \sim 0$ and $z = 2.5$, we still would fall way short of reproducing the observed number densities of compact low n systems. The number density of massive $(M_\star \geq 1 \times 10^{11} \, M_\odot)$ galaxies at $z = 2.5$ is approximately 30% that at $z \sim 0$. If we take the most compact $(r_e \leq 2 \text{ kpc})$ and low $n \leq 2$ systems at $z \sim 0$, and scale this number by 30%, we find a much lower number density $(2.8 \times 10^{-6} \text{ gal Mpc}^{-3})$, than the observed no density $(5.0 \times 10^{-5} \text{ gal Mpc}^{-3})$ at $z = 2.5$ for such compact systems.

2. Can redshift-dependent systematic effects cause structural parameters, such as the high Sérsic index n of massive galaxies at $z \sim 0$, to 'degrade' into the regime of low $n \leq 2$ values, measured in the $z = 2 - 3$ systems? We address this issue in §3.3.1.

3. How robust are our fits to the NIC3/F610W images of the $z = 2 - 3$ galaxies? Could some of the galaxies with a best-fit Sérsic index $n \leq 2$ have similarly good fits with much higher n? We show in Appendix B.1 and Appendix B.2 that this is unlikely. We are confident that the fraction of $n \leq 2$ systems is not being overestimated.

4. Can the offset in (n, r_e) between the $z = 2 - 3$ galaxies and the $z \sim 0$ galaxies be caused by systematic differences between the fitting techniques applied by us to the NIC3/F610W images of $z = 2 - 3$ galaxies and the fitting techniques used by Allen et al. (2006) on the B-band images of the massive $z \sim 0$ galaxies in MGC? We conduct additional tests (see Appendix B.3) and conclude that this is also not the case.

3.3.1. *Artificial Redshifting*

We next investigate whether redshift-dependent systematic effects could potentially cause the offset in (n, r_e) shown in Figure 5 between massive galaxies at $z \sim 0$ and $z = 2 - 3$, by causing the (n, r_e) of massive $z \sim 0$ galaxies to 'degrade' into the regime of low $n \leq 2$ and low r_e exhibited by the $z = 2 - 3$ systems.

Ideally one would investigate this question by artificially redshifting the entire MGC subsample of 385 massive $z \sim 0$ galaxies shown in Figure 5 out to $z \sim 2.5$, and re-decomposing the redshifted galaxies. However, this is extremely time consuming, and, furthermore, many of the galaxies do not have high quality SDSS $ugriz$ images which are needed for redshifting software (FERENGI; Barden et al. 2008) to work. We therefore decide to artificially

redshift a smaller, but *representative* sample S1 of 255 galaxies. S1 consists of 42 massive $(M_\star \geq 5 \times 10^{10} \, M_\odot)$ MGC galaxies combined with 213 nearby $(z < 0.05)$ massive galaxies having high quality and well-resolved SDSS imaging. We ensure the (n, r_e) of the 255 galaxies in S1 match those of the entire subsample of MGC galaxies shown in Table 1, Figure 8, and Figure 9. We also ensure that the distribution of Hubble types of sample S1 matches those of the MGC subsample; the MGC subsample contains $\sim 66\%$ E/S0 galaxies versus $\sim 34\%$ Spirals, while sample S1 is $\sim 64\%$ E/S0 galaxies and $\sim 36\%$ Spirals. Many of the galaxies in S1 are well studied and include E, S0, and Sabc galaxies from Barden et al. (2008), E galaxies in Kormendy et al. (2009), as well as S0s and bulge-dominated spirals from Eskridge et al. (2002).

We used FERENGI (Barden et al. 2008) to artificially redshift the SDSS $ugriz$ images (tracing rest-frame UV-to-optical light) of $z \sim 0$ galaxies, out to $z = 2.5$, and re-observe them with the NIC3 F160W filter to the same depth as the GNS survey. During this process, FERENGI mimics the effects of surface brightness dimming, instrumental resolution, transmission efficiency, and PSF effects. It also corrects for other geometrical effects of cosmological redshift by appropriately re-binning input images for the desired redshift and platescale.

Specifically, during artificial redshifting, as is standard convention, FERENGI assumes surface brightness dimming at the rate of $(1 + z)^{-4}$ for the bolometric luminosity of the full redshifted rest-frame optical SED. For galaxies where only part of this redshifted rest-frame optical SED falls within the NIC3/F160W filter bandwidth, the observed flux per unit wavelength f_λ relates to the rest-frame luminosity per unit wavelength at redshift z via a $(1 + z)^{-3}$ dependence (e.g., Weedman 1986). The exact surface brightness dimming in such a case will be set by the integral of f_λ over the filter-detector response function and depends on the detailed shape of the SED (e.g., Hogg 1999; Hogg et al. 2002). In practice, when using the FERENGI software, the relevant degree of surface brightness dimming is automatically applied when FERENGI convolves the redshifted images with the NIC3 F160W PSF and then re-observes the redshifted SED with the NIC3 F160W (H) filter-detector, while taking into account the filter-detector characteristics, such as bandwidth and transmission efficiency. An exposure time of three-orbits (8063 seconds) and a resolution of $0\rlap{.}''2$/pixel is assumed to mimic the GNS survey. A sky background equal to the mean sky background of the GNS NIC3 images (0.1 counts/second) was added to the redshifted images. Poisson noise, sky noise, and read noise (29 e^- for NIC3) were then added to the redshifted images.

During artificial redshifting of local galaxies, it is standard procedure to incorporate surface brightness evolution (Barden et al. 2008) because galaxies at higher redshifts have been observed to have higher mean surface brightness after applying the standard correction for the geometrical effect of cosmological surface brightness dimming. For instance, Lilly et al. (1998) find that surface brightness for disk-dominated galaxies of similar properties increases on average by 0.8 magnitudes by $z = 0.7$. Barden et al. (2005) find from the GEMS ACS survey that galaxies with $M_V \lesssim -20$ show a brightening of ~ 1 magnitude in rest-

frame V-band by $z \sim 1$. Labbé et al. (2003) find a disk-like galaxy with spectroscopic redshift $z = 2.03$ to have a rest-frame B-band surface brightness ~ 2 magnitudes brighter than nearby galaxies. Finally in our own study, the mean surface brightness within r_e of massive galaxies at $z = 2 - 3$ is 2 to 6 magnitudes higher than that of massive galaxies at $z \sim 0$, with a mean offset of ~ 4.5 magnitudes (Figure 7).

In our experiment of artificially redshifting massive galaxies from $z \sim 0$ to $z = 2.5$, we applied a conservative value of 2.5 magnitudes of surface brightness evolution. This value is motivated by several considerations: a) 2.5 magnitudes of surface brightness evolution is on the conservative side as many of the massive galaxies at $z = 2.5$ show even more evolution (Figure 7). Thus, using this value will not lead to overoptimistic recovery of faint features during the experiment; b) The adopted 2.5 magnitudes of evolution out to $z = 2.5$ corresponds to one magnitude of brightening per unit redshift. This rate of brightening is comparable to those seen in studies out to $z \sim 2$ (Lilly et al. 1998; Barden et al. 2005; Labbé et al. 2003); c) Using the Bruzual & Charlot (2003) models, one can show that the passive evolution of a single stellar population from $z = 2.5$ to $z = 0$, assuming an exponentially declining star formation history associated with an e-folding time of 100 Myr, will lead the rest-frame B luminosity to decline by 2.5 to 3 magnitudes, depending on the chosen metallicity.

While we believe that 2.5 magnitudes of surface brightness evolution is a conservative and reasonable value to use during the artificial redshfiting experiment, for the sake of completeness, we have also tested the effect of applying a surface brightness evolution (brightening) of 0, 1.25, 2.5, and 3.75 magnitudes between $z \sim 0$ and $z = 2.5$. There is a discernible difference in the recovered morphology and structural parameters between 0 and 1.25 magnitudes of brightening, but less difference between 1.25, 2.5, or 3.75 magnitudes of brightening. More details on the use of zero surface brightness evolution are given in point 4 at the end of this section.

After artificially redshifting S1 from $z \sim 0$ to $z = 2.5$, we fit both the original galaxy images and their redshifted counterparts with single Sérsic profiles. We compare the rest-frame optical structural parameters in the original and redshifted images in order to assess the influence of redshift-dependent systematic effects (e.g., surface brightness dimming, loss of spatial resolution) and see how well the structural parameters are recovered. We also compare the redshifted distribution of (n, r_e) to the one actually observed in the GNS massive galaxies to assess whether they are similar. Note that the structural parameters are measured at $z \sim 0$ from g-band images, while at $z = 2.5$ they are measured from the artificially redshifted images in the NIC3/F160W band so that all parameters are measured in the rest-frame blue optical light, thereby avoiding bandpass shifting problems. Our main results are outlined below.

1. Figure 8 shows the (n, r_e) distribution obtained by redshifting the sample S1 (magenta points in row 1) of 255 $z \sim 0$ massive galaxies to $z \sim 2.5$ (blue points in row 2). This redshifted distribution of $(n,$

$r_e)$ is still significantly offset from those observed in the massive GNS galaxies at $z = 2 - 3$ (red points in row 2).

This difference is shown more quantitatively in Figure 9 where results in discrete bins of n and r_e are compared. The massive galaxies at $z = 2 - 3$ (red line) includes $64.9 \pm 5.4\%$ of systems with low $n \leq 2$, while the corresponding fraction for the redshifted sample (blue line) is $10.6 \pm 1.9\%$. Similarly, for the r_e distribution of the massive galaxies at $z = 2 - 3$, $39.0 \pm 5.6\%$ have $r_e \leq 2$ kpc, while the redshifted sample has $1.2 \pm 0.7\%$. We therefore conclude that cosmological and instrumental effects are not able to account for the large offset shown in Figure 8 and Figure 9 between the (n, r_e) distributions of the massive galaxies at $z = 2 - 3$ and those at $z \sim 0$.

2. It is very interesting to look at how the structural parameters of galaxies of different morphological types change during the redshifting. Figure 10 compares the rest-frame optical structural parameters in massive E, S0, and spirals at $z \sim 0$ to the structural parameters recovered after these galaxies were artificially redshifted.

From Figure 10, one can see that r_e is recovered to better than a factor of 1.5 for the vast majority of redshifted E/S0 and spirals of early-to-late Hubble types. In the case of a small fraction of $z \sim 0$ galaxies with highly extended halos or disks and associated large r_e, the recovered r_e at $z = 2.5$ can be nearly a factor of two lower than the original r_e at $z \sim 0$. Inspection of the surface brightness profiles shows that this effect primarily happens because surface brightness dimming prevents the outer lower surface brightness components of the galaxies from being adequately recovered after redshifting.

It is striking that even after redshifting out to $z = 2.5$, practically none of the massive $z \sim 0$ galaxies fall into the regime of $r_e \leq 2$ kpc (shown as shaded areas) inhabited by the ultra-compact systems, which make up $\sim 40\%$ ($39.0 \pm 5.6\%$ for $M_\star \geq 5 \times 10^{10} \, M_\odot$ and $39.0 \pm 7.6\%$ for $M_\star \geq 1 \times 10^{11} \, M_\odot$) of the massive galaxies at $z = 2 - 3$ (see §3.2). *Thus, these massive ultra-compact ($r_e \leq 2$ kpc) systems at $z = 2 - 3$ appear to truly have no analogs among $z \sim 0$ massive galaxies, in terms of their size, structure, and optical surface brightness.*

The top row of Figure 10 shows the distribution of Sérsic index n before and after redshifting out to $z = 2.5$. The recovered Sérsic index n can be lower or higher than the original n at $z \sim 0$, but is recovered to better than a factor of two in all cases. The shaded area in the plots represents the regime of $n \leq 2$ where the majority ($64.9 \pm 5.4\%$ for $M_\star \geq 5 \times 10^{10} \, M_\odot$ and $58.5 \pm 7.7\%$ for $M_\star \geq 1 \times 10^{11} \, M_\odot$) of massive GNS galaxies at $z = 2 - 3$ lie (Table 1). It is interesting to note that massive E and S0s, which are spheroid-dominated and bulge-dominated systems, do not typically lie in the $n \leq 2$ regime, before or after redshifting. In contrast, a

---

[13] The Hubble types are based on the bulge-to-total light ratio (B/T), which we measured with bulge-disk and bulge-disk-bar decomposition

large fraction of $z \sim 0$ spirals with intermediate-to-late Hubble types[13] populate the $n \leq 2$ regime, both before and after redshifting. Disk features on large and small scales (e.g., outer disk or disky pseudobulge) lead to an overall single Sérsic index $n \leq 2$ for the entire galaxy. *It is possible that similar disk features are responsible at least in part, for the low $n \leq 2$ values shown by the majority ($\sim 65\%$) of the massive GNS galaxies at $z = 2 - 3$.* We discuss this point further in §7.

3. One important question is whether the use of deeper images of the $z \sim 0$ galaxies would change the conclusion of the redshifting experiment. In the present experiment, we used SDSS g-band images, which have an exposure time of 54 seconds and a typical sky background of 22 mag/arcsec^2. Deeper exposures of nearby galaxies may potentially detect an outer low surface brightness halo (if such a halo exists), which is missed in the SDSS images, and in that case lead us to measure *larger* (n, r_e) at $z \sim 0$ with a Sérsic fit. Such halos can be found in very local massive elliptical and cD galaxies, where the measured (n, r_e) can increase significantly if the halo is included in the fit (Kormendy et al. 2009). However, such low surface brightness halos will be dimmed out and not recovered during the artificial redshifting of these deep images, so that the (n, r_e) parameters recovered at $z = 2.5$ will be similar to those we presently obtain from the SDSS images. The net effect will be that using deeper images of local massive galaxies during the artificial redshifting will at most raise the (n, r_e) at $z \sim 0$, but not at $z = 2.5$. Thus the difference in the (n, r_e) at $z \sim 0$ compared $z = 2.5$ will be unchanged (for systems without halos) or amplified (for systems with such halos). Our overall conclusion from the redshifting experiment regarding degradation of the profiles to $n \leq 2$ and $r_e \leq 2$ kpc would remain unchanged or be even stronger.

4. Finally, as one additional test, we repeated the redshifting experiment assuming zero surface brightness evolution, rather than 2.5 magnitudes of brightening, out to $z = 2.5$. Even in this case there is still a large offset in the (n, r_e) distributions of the redshifted sample S1 compared to the massive GNS galaxies. Specifically, the fraction of systems with low $n \leq 2$ ($22.0 \pm 2.6\%$) is still significantly less than that for massive GNS galaxies at $z = 2 - 3$ ($64.9 \pm 5.4\%$). Likewise, there are still few systems with $r_e \leq 2$ kpc ($1.6 \pm 0.8\%$) compared to the high fraction ($39.0 \pm 5.6\%$) found at $z = 2 - 3$. Thus, even without surface brightness evolution it is still true that cosmological and instrumental effects are not able to account for the large offset between massive galaxies at $z = 2 - 3$ versus $z \sim 0$.

4. STAR FORMATION ACTIVITY

4.1. *Matching GNS Galaxies to MIPS 24 μm Counterparts*

The *Spitzer* GOODS Legacy Program (Dickinson et al. 2003a; Dickinson et al. in preparation) provides deep *Spitzer* MIPS 24 μm observations of the GOODS fields. In the discussion below, we only consider MIPS 24 μm counterparts with $f_{24\mu m} \geq 30$ μJy, the 5σ flux limit. The MIPS images have a PSF diameter of 6$''$ (~ 42 kpc at $z = 1 - 3$), versus the NIC3/F160W PSF of 0$.''$3. MIPS 24 μm counterparts of the massive GNS galaxies were identified by selecting the closest MIPS 24 μm source within a maximum matching radius of 1$.''$5. We initially find 84/166 massive GNS galaxies with MIPS 24 μm counterparts with $f_{24\mu m} \geq 30$ μJy and further refine these matches below.

There are several potential problems with the above procedure. Firstly, it allows for the situation where a given MIPS 24 μm source could be matched to several massive GNS galaxies. This would happen if some massive GNS galaxies were crowded within a radius of a few arcseconds so that the MIPS source would be within 1$.''$5 of all of them. This situation occurs for 2/84 ($\sim 2\%$) of massive galaxies with a MIPS counterpart. We reject these two cases, reducing the number of unique and secure matches from 84 to 82.

A second possible caveat is that within the large MIPS 24 μm PSF of 6$''$ diameter, there may be several other NIC3/F160W sources, in addition to the main massive GNS galaxy to which the MIPS source is matched. These extra NIC3/F160W sources may even be lower mass galaxies not in our sample of massive ($M_\star \geq 5 \times 10^{10}$ $M_\odot$) galaxies. In such a scenario, all the extra NIC3 sources could potentially contribute to the MIPS 24 μm flux, and assigning all the 24 μm flux of the MIPS counterpart to the nearest massive GNS galaxy would overestimate the 24 μm flux of this galaxy. In order to assess the extent of this potential problem, we proceed as follows. For the MIPS 24 μm counterpart assigned previously to each massive GNS galaxy, we determine how many extra NIC3/F160W sources with $M_\star \geq 10^9$ $M_\odot$, in addition to the massive GNS galaxy, lie within a circle of diameter 6$''$ (i.e., the PSF diameter) centered on the MIPS source. Of the 82 massive GNS galaxies with a secure MIPS 24 μm counterpart, 30 involve cases where there are extra NIC3 sources, along with the massive GNS galaxy, inside the MIPS PSF diameter.

Next, we estimate the relative expected contributions of the massive GNS galaxy and the extra NIC3/F160W sources to the overall 24 μm flux by using the stellar mass ratio of the main massive GNS galaxy (e.g., $M_{\star 1}$) and of the contaminating source (e.g., $M_{\star 2}$), scaled by a function that takes into account the different redshifts of the two sources. Specifically, for the two sources with stellar mass $M_{\star 1}$ and $M_{\star 2}$, having redshifts z_1 and z_2 and luminosity distances D_{L1} and D_{L2}, the stellar mass ratio $M_{\star 1}/M_{\star 2}$ is scaled by $((1 + z_2) \times D_{L2}^2)/((1 + z_1) \times D_{L1}^2)$. In 8 of 30 cases, the contribution of the extra NIC3 contaminating sources to the overall 24 μm flux is $> 20\%$ that of the main GNS galaxy, and spans $\sim 40\%$ to $\sim 126\%$. We reject these latter 8 cases rather than try to correct for the contamination, which in all cases is distributed across two or more nearby galaxies. For the remaining 22 cases, the contamination by extra NIC3/F160W sources is $< 20\%$ and we deem that our afore-described procedure of assigning

of $z \sim 0$ g-band images.

all the 24 μm flux of the MIPS counterpart to the massive GNS galaxy is reasonable.

Therefore, in summary, 74/166 (44.6 $\pm$ 3.9%) massive ($M_\star \geq 5 \times 10^{10}$ $M_\odot$) GNS galaxies have a reliable MIPS 24 μm counterpart (with $f_{24\mu m} \geq 30$ μJy) whose entire flux is assigned to the massive GNS galaxy. In contrast, 82/166 (49.4 $\pm$ 3.9%), massive GNS galaxies do not have a reliable MIPS counterpart with $f_{24\mu m} \geq 30$ μJy and in these cases we can only measure upper limits on their SFR. Table 3 lists the fraction of massive GNS galaxies with a MIPS 24 μm counterpart as a function of redshift.

4.2. *Star Formation Rates*

In order to estimate the SFR, the total IR luminosity ($L_{\rm IR}$) over 8–1000 μm is first estimated from the observed 24 μm flux (corresponding to rest-frame wavelengths of 6–12 μm over $z = 1-3$) by using SED templates from Chary & Elbaz (2001). Using solely 24 μm flux density to measure $L_{\rm IR}$ works well for inferred $L_{\rm IR} \lesssim 10^{12}$ $L_\odot$ galaxies at $z \sim 2$, but $L_{\rm IR}$ is overestimated by a factor of ~ 3 in more luminous galaxies (e.g., Papovich et al. 2007). Early results from Herschel (e.g., Elbaz et al. 2010; Nordon et al. 2010; D. Lutz, private communication) suggest that at $z > 1.5$, the SFRs extrapolated from 24 μm fluxes may overestimate the true SFR, typically by a factor of 2 to 4 and possibly as much as a factor of 10. This overestimate could be due to a rise in the strength of PAH features, changes in the SEDs, or AGN contamination at $z > 1.5$. Murphy et al. (2009) find that estimates of $L_{\rm IR}$ from 24 μm flux density alone are incorrect because the templates used are based on local galaxies with smaller PAH equivalent widths than galaxies of similar luminosity at high-redshift. We account for this discrepancy by making a correction for galaxies with inferred $L_{\rm IR} > 6 \times 10^{11}$ $L_\odot$ using

$$\log_{10}(L_{\rm IR}) = 0.59 \times \log_{10}(L_{\rm IR}^{24}) + 4.8, \qquad (3)$$

where $L_{\rm IR}^{24}$ is the infrared luminosity inferred solely from 24 μm flux density (R. Chary, private communication). The upper-left and upper-right panels of Figure 11 show the distribution of 24 μm flux and the inferred $L_{\rm IR}$.

The obscured star formation rate can be calculated using the expression

$$\mathrm{SFR_{IR}} = 9.8 \times 10^{-11} L_{\rm IR} \qquad (4)$$

from Bell et al. (2007). This calculation is based on a Chabrier IMF (Chabrier 2003) and assumes that the infrared emission is radiated by dust that is heated primarily by massive young stars. Uncertainties in the SFR estimates are a factor of ~ 2 or higher for individual galaxies.

If an AGN is present, then $\mathrm{SFR_{IR}}$ only gives an upper limit to the true SFR. In §6, we adopt several techniques to identify AGN candidates in the sample and estimate the mean SFR for galaxies with and without a candidate AGN (see Table 3). The upper-right panel of Figure 11 shows $L_{\rm IR}$ for AGN and non-AGN, and the bottom panels show $\mathrm{SFR_{IR}}$. The AGN candidates dominate the tail of highest $L_{\rm IR}$ and $\mathrm{SFR_{IR}}$. Among the HyLIRGs[14], 9/11 ($\sim 82\%$) turn out to be AGN. After excluding the AGN candidates, the mean $L_{\rm IR}$ is a factor of ~ 8 times lower, while the mean $\mathrm{SFR_{IR}}$ is reduced a factor of ~ 1.5 to ~ 2.7, and the difference rises with redshift (Table 3).

[14] HyLIRGs are defined to have $L_{IR} \geq 10^{13}$ $L_\odot$

How do our measurements of $\mathrm{SFR_{IR}}$ compare with UV-based SFR derived in other studies of high-redshift galaxies? The left panel of Figure 12 plots $\mathrm{SFR_{IR}}$ versus $M_\star$ for the massive ($M_\star \geq 5 \times 10^{10}$ $M_\odot$) GNS galaxies at $z = 2-3$ with 24 μm flux above the 5σ limit (30 μJy). We demonstrate that the SFR derived at $z = 2-3$ for non-AGN are in approximate agreement with the UV-based SFR from Daddi et al. (2007). Drory & Alvarez (2008) parameterize SFR as a function of mass and redshift for a wide range in stellar mass ($M_\star \sim 10^9 - 10^{12}$ $M_\odot$). In the right panel of Figure 12, the black line shows average SFR versus redshift for a 5×10^{10} $M_\odot$ galaxy as calculated by Drory & Alvarez (2008). The mean $\mathrm{SFR_{IR}}$ for massive non-AGN GNS galaxies, with $\mathrm{SFR_{IR}}$ above the 5σ limit, are higher by a factor of $\sim 1.5 - 4$ over $z = 1-3$, with the offset worsening with redshift. This disagreement with mean $\mathrm{SFR_{IR}}$ is not just a bias caused by the requirement that $\mathrm{SFR_{IR}}$ exceed the 5σ limit, which selects the most intense star-forming systems at each redshift. Even if the upper limits on $\mathrm{SFR_{IR}}$ are included, our $\mathrm{SFR_{IR}}$ do not show the same break and flattening seen at $z \sim 2$ by Drory & Alvarez (2008). Finally, Bauer et al. (2011) measure dust-corrected UV-based SFR ($\mathrm{SFR_{UV,corr}}$) for galaxies in GNS over $1.5 < z < 3$. Among massive ($M_\star \geq 5 \times 10^{10}$ $M_\odot$) galaxies, $\mathrm{SFR_{UV,corr}}$ can differ by as much as a factor of 10, but for higher $\mathrm{SFR_{IR}}$ the difference is typically a factor of $\sim 2-3$.

4.3. *Relation Between Star Formation and Structure*

Figure 13 shows the distribution of $\mathrm{SFR_{IR}}$ among systems of different n. On the LHS panel, galaxies with $\mathrm{SFR_{IR}}$ below the 5σ detection limit are shown as downward pointing arrows. The potential AGN candidates identified in §6 are coded separately as $\Sigma_{\mathrm{SFR_{IR}}}$ is likely overestimating the true SFR in the galaxy. For the histograms on the RHS panel, the y-axis shows the fraction of massive GNS galaxies in each redshift bin, while on the x-axis, we plot the actual value of $\mathrm{SFR_{IR}}$ for systems with $\mathrm{SFR_{IR}}$ above the 5σ detection limit (indicated by the vertical line), and the upper limit for the other systems.

The massive galaxies at $z = 1-3$ display several interesting relations between their star formation activity and structure, as characterized by the Sérsic index n. Firstly, among the non-AGN massive ($M_\star \geq 5 \times 10^{10}$ $M_\odot$) galaxies at $z = 2-3$, the fraction of galaxies with low $n \leq 2$ having $\mathrm{SFR_{IR}}$ high enough to produce a 24 μm flux above the 5σ detection limit is (53.4$\pm$10.9%), which is significantly higher than the corresponding fraction (15.4 $\pm$ 10.0%) for systems with $n > 2$. Secondly, among the non-AGN massive ($M_\star \geq 5 \times 10^{10}$ $M_\odot$) galaxies at $z = 2-3$ with $\mathrm{SFR_{IR}}$ above the 5σ detection limit, *the majority (84.6 $\pm$10.0%) have low $n \leq 2$, while none have $n > 4$.* The corresponding numbers for the redshift bin $z = 1-2$ are 67.7 $\pm$ 8.0% and 11.8 $\pm$ 5.5%, respectively. Thirdly, the RHS panel of Figure 13 shows that the high SFR tail in each redshift bin is populated primarily by $n \leq 2$ systems. While the $n \leq 2$ disky systems have a wide range of $\mathrm{SFR_{IR}}$ (21 to 626 $M_\odot$ yr^{-1} at $z = 1-2$, 53 to 1466 $M_\odot$ yr^{-1} at $z = 2-3$), *they include the systems of the highest SFR at both $z = 1-2$ and $z = 2-3$.* Thus, the systems with low $n \leq 2$ seem to be more actively star-forming than the systems of high

$n > 3$.

Most ($72.0 \pm 6.3\%$ of systems with low $n \leq 2$ are extended ($r_e > 2$ kpc) so that a relation is also expected between SF activity and size. We thus investigate next the relationship between SFR and effective radius r_e. The distribution of $\mathrm{SFR_{IR}}$ for different r_e ranges is shown in Figure 14. The same convention as for Figure 13 is adopted, with upper limits being plotted for galaxies with $\mathrm{SFR_{IR}}$ below the 5σ detection limit, and only non-AGN systems being plotted on the RHS panel. We find that among the non-AGN massive ($M_\star \geq 5 \times 10^{10}\ M_\odot$) galaxies at $z = 2 - 3$, the fraction of ultra-compact ($r_e \leq 2$ kpc) objects with $\mathrm{SFR_{IR}}$ above the 5σ detection limit is only $15.0 \pm 8.0\%$ compared to the fraction ($32.4 \pm 8.0\%$) for the whole sample. Thus, among non-AGN massive galaxies over $z = 2 - 3$, *the ultra-compact ($r_e \leq 2$ kpc) galaxies show a deficiency by a factor of ~ 2.2 of systems with $\mathrm{SFR_{IR}}$ above the detection limit, compared to the whole sample.* At $z = 1 - 2$, the deficiency is a factor of ~ 3.5. Furthermore, as illustrated by the RHS panel of Figure 14, although there are some ultra-compact ($r_e \leq 2$ kpc) galaxies with high $\mathrm{SFR_{IR}}$, on average, the mean $\mathrm{SFR_{IR}}$ of the $z = 2 - 3$ and $z = 1 - 2$ is significantly lower than that of more extended galaxies.

5. CONSTRAINTS ON COLD GAS CONTENT

The high estimated $\mathrm{SFR_{IR}}$ found in §4 suggest that copious cold gas reservoirs are present to fuel the star formation. For the massive GNS galaxies with $\mathrm{SFR_{IR}}$ measurements above the 5σ detection limit, we assume half of $\mathrm{SFR_{IR}}$ lies within the circularized rest-frame optical half-light radius $\left(r_c = r_e \times \sqrt{b/a}\right)$ from single component Sérsic fits, and thereby estimate that the deprojected SFR per unit area as

$$\Sigma_{\mathrm{SFR_{IR}}} = \frac{0.5 \times \mathrm{SFR_{IR}}}{\pi \times r_c^2}. \tag{5}$$

In galaxies that AGN host candidates, $\Sigma_{\mathrm{SFR_{IR}}}$ is likely overestimating the true SFR in the galaxy (see §4). If potential AGN candidates are included, $\Sigma_{\mathrm{SFR_{IR}}}$ ranges from $\sim 0.10 - 360.8\ M_\odot\ \mathrm{yr}^{-1}\ \mathrm{kpc}^{-2}$, with a mean value of $\sim 19.4\ M_\odot\ \mathrm{yr}^{-1}\ \mathrm{kpc}^{-2}$ over $z = 1 - 3$. After excluding the potential AGN candidates $\Sigma_{\mathrm{SFR_{IR}}}$ ranges from $\sim 0.24 - 360.8\ M_\odot\ \mathrm{yr}^{-1}\ \mathrm{kpc}^{-2}$, with a mean value of $\sim 14.8\ M_\odot\ \mathrm{yr}^{-1}\ \mathrm{kpc}^{-2}$. This range is comparable to that seen in BzK/normal galaxies, ULIRGS, and submillimeter galaxies (e.g., see Daddi et al. 2010b).

We use a standard Schmidt-Kennicutt law (Kennicutt 1998), with a power-law index of 1.4 and a normalization factor of 2.5×10^{-4}, to estimate the cold gas surface density from $\Sigma_{\mathrm{SFR_{IR}}}$. The results are uncertain by at least a factor of ~ 2.5 because different relations between molecular gas surface density and SFR surface density have been suggested for various types of star-forming systems over a broad range of redshifts (Kennicutt 2008; Gnedin & Kravtsov 2010; Daddi et al. 2010b; Genzel et al. 2010; Tacconi et al. 2010). If potential AGN candidates are included, the resulting implied cold gas surface density

$$\Sigma_{\mathrm{gas}} = \left[\frac{10^4 \times \Sigma_{\mathrm{SFR_{IR}}}}{2.5} \right]^{1/1.4} \tag{6}$$

ranges from $\sim 73 - 25,091\ M_\odot\ \mathrm{pc}^{-2}$, with a median value of $\sim 907\ M_\odot\ \mathrm{pc}^{-2}$ over $z = 1 - 3$ (Figure 15). The corresponding values after excluding AGN candidates are $\sim 136 - 25091\ M_\odot\ \mathrm{pc}^{-2}$, with a median value of $\sim 607\ M_\odot\ \mathrm{pc}^{-2}$ (Figure 15). These values are again comparable to those observed in BzK/normal galaxies, ULIRGS, and submillimeter galaxies (e.g., see Daddi et al. 2010b).

In the subsequent discussion, we only cite values obtained after excluding AGN candidates, but Figure 15 also shows the values for the full sample of galaxies. Next we estimate the cold gas fraction relative to the baryonic mass within r_c. For each galaxy, we use the above cold gas surface density to estimate the total cold gas mass within the circularized rest-frame optical half-light radius,

$$M_{\mathrm{gas}}(r_c) = \Sigma_{\mathrm{gas}} \times \pi \times r_c^2. \tag{7}$$

M_{gas} ranges from $3.4 \times 10^9 - 1.0 \times 10^{11}\ M_\odot$, with a mean value of $1.9 \times 10^{10}\ M_\odot$ (Figure 15). The baryonic mass (M_{Baryon}) within r_c is taken to be the sum of cold gas mass and stellar mass within r_c, and we assume that the latter term is half of the total stellar mass of the galaxy.

The cold gas fraction ($f_{\mathrm{gas}}(r_c)$) within the circularized rest-frame optical half-light radius r_c is defined as

$$f_{\mathrm{gas}}(r_c) \equiv M_{\mathrm{gas}} / \left[M_{\mathrm{gas}} + M_\star \right]. \tag{8}$$

The cold gas fraction ($f_{\mathrm{gas}}(r_c)$) ranges from $6.5 - 65.4\%$, with a mean of $\sim 23\%$ over $z = 1 - 3$ (Figure 15). Figure 16 shows how $f_{\mathrm{gas}}(r_c)$ varies as a function of stellar mass and redshift, both with and without the AGN candidates. For galaxies with $5 \times 10^{10}\ M_\odot \leq M_\star < 10^{11}\ M_\odot$ above the 5σ detection limit, the mean $f_{\mathrm{gas}}(r_c)$ (without AGN candidates) rises from $\sim 19\%$ to $\sim 25\%$ to $\sim 41\%$ across the three redshift bins. In comparison, for $M_\star \geq 10^{11}\ M_\odot$ galaxies, the mean cold gas fraction is $\sim 14\%$ to $\sim 23\%$. The 1σ error bars are large and there is considerable overlap between the two mass ranges. *Still, the highest cold gas fractions within the circularized rest-frame optical half-light radius at a given redshift are found among the less massive galaxies,* consistent with downsizing.

Our inferred cold gas fractions ($f_{\mathrm{gas}}(r_c)$) within the circularized rest-frame optical half-light radius r_c can be higher or lower than the total cold gas fraction of the galaxy, depending on whether the molecular gas is centrally concentrated or extended, respectively. While bearing this caveat in mind, we note that our inferred values for $f_{\mathrm{gas}}(r_c)$ are consistent with previous direct measurements of the total cold gas fraction at high-redshift. Daddi et al. (2008, 2010a) report gas fractions of 50-65% in massive ($M_\star \sim 4 \times 10^{10} - 1 \times 10^{11}\ M_\odot$) IR-selected BzK galaxies at $z \sim 1.5$. Tacconi et al. (2010) also measure cold gas fraction from CO observations of high-redshift galaxies at $z = 1.1 - 2.4$. For stellar masses spanning $M_\star \sim 3 \times 10^{10} - 3.4 \times 10^{11}\ M_\odot$, they find cold gas fractions in the range of $\sim 14 - 78\%$.

6. AGN IN MASSIVE GALAXIES AT $z = 1 - 3$

6.1. *Frequency of AGN*

We use a variety of techniques (X-ray properties, IR power-law, IR-to-optical excess, and mid-IR colors) to identify Active Galactic Nuclei (AGN) among the massive GNS galaxies because selection based on X-ray emission alone may fail at high redshift in the case of Compton-thick AGN where much of the soft X-ray emission is Compton scattered or absorbed by thick columns of gas

($N_H \gg 10^{24}$ cm^{-2}; Brandt et al. 2006). We briefly summarize here and in Table 4 the number of AGN identified by each of the selection methods[15].

1. X-ray counterparts to the massive GNS sources were searched for in the CDF-N and CDF-S catalogs of Alexander et al. (2003) and Luo et al. (2008), as well as the ECDF-S catalogs of Lehmer et al. (2005). A total of 33/166 massive GNS galaxies had counterparts within 1″.5 across all catalogs.

2. Following Alonso-Herrero et al. (2006) and Donley et al. (2008), we look for AGN power-law emission over $z = 1 - 3$ using SEDs from the IRAC bands at 3.6, 4.5, 5.8, and 8.0 μm. The IRAC SEDs were fit with a power-law SED ($f_\nu \propto \nu^\alpha$). There are only 3/166 sources with power-law index $\alpha \leq -0.5$ that are considered power-law galaxies (PLGs) and obscured AGN candidates.

3. Heavily obscured AGN may be present in highly reddened, IR-excess galaxies. Fiore et al. (2008) identify obscured AGN candidates in IR-bright, optically faint, red galaxies over $z = 1.2 - 2.6$ using the criteria $f_{24\mu m}/f_R \geq 1000$ and $R - K > 4.5$. We search for such IR-bright, optically faint systems with $f_{24\mu m}/f_R > 1000$ in our sample of massive galaxies. R-band flux is determined by linear interpolation between the ACS V and i-band fluxes. We find 25 sources meeting this criteria, of which 16 are new AGN candidates not identified via the above two methods.

Among the 166 massive GNS galaxies at $z = 1 - 3$, the AGN fraction is 49/166 or $29.5 \pm 3.5\%$. When the results are broken down in terms of redshift, the AGN fraction rises with redshift, increasing from $17.9 \pm 6.1\%$ at $z = 1 - 1.5$ to $40.3 \pm 8.8\%$ at $z = 2 - 3$. The percentage of AGN among all massive GNS galaxies is higher than at $z \sim 1$, where it is reported that less than 15% of the total 24 μm emission at $z < 1$ is in X-ray luminous AGN (e.g., Silva et al. 2004; Bell et al. 2005; Franceschini et al. 2005; Brand et al. 2006).

6.2. Relation Between AGN Activity and Structure

We summarize the properties of the AGN host candidates and discuss their implications in terms host galaxy structure.

Figure 17 shows the single Sérsic n versus r_e. Most ($80.6 \pm 7.9\%$) of the AGN hosts at $z = 2 - 3$ have $r_e > 2$ kpc and are not ultra-compact. AGN appear to be found preferentially in the more extended galaxies. Indeed, at $z = 2 - 3$, the AGN fraction in ultra-compact galaxies is ~ 2.7 times lower than in extended galaxies ($20.0 \pm 16.3\%$ versus $53.2 \pm 10.0\%$). At $z = 1 - 2$ the deficiency is a factor of 5.6. Thus, the ultra-compact galaxies are more quiescent in terms of both AGN activity and SFR activity (see §4).

Furthermore, a significant fraction of these AGN ($64.6 \pm 10.7\%$) have disky ($n \leq 2$) morphologies. Over half ($58.2 \pm 11.6\%$) of the AGN candidates are both disky and not ultra-compact. Similar statistics apply over $z = 1 - 2$. The disky nature of AGN hosts at $1.5 < z < 3$ has been measured previously by Schawinski et al. (2011). From decomposition of the rest-frame optical light for 20 AGN imaged with HST WFC3, they measure a mean Sérsic index of 2.54 and a mean effective radius of 3.16 kpc. Their results for (n, r_e) are consistent with our results for $z = 2 - 3$ in Table 4 and Figure 17. Furthermore, Kocevski et al. (2011, in prep.) find from visual classification of rest-frame optical morphologies that $51.4^{+5.8}_{-5.9}$ of X-ray selected AGN ($L_X \sim 10^{42-44}$ erg s^{-1}) at $1.5 < z < 2.5$ reside in galaxies with visible disks; only $27.4^{+5.8}_{-4.6}$ reside in pure spheroids.

If the disky AGN host candidates host massive black holes, then massive black holes are present in galaxies that are not dominated by a massive spheroid. In the local Universe, nearly all massive galaxies are believed to host a central supermassive black hole (Kormendy 1993; Magorrian et al. 1998; Ferrarese & Merritt 2000; Gebhardt et al. 2000; Marconi & Hunt 2003), and the black hole mass is tightly related to the bulge stellar velocity dispersion (Ferrarese & Merritt 2000; Gebhardt et al. 2000). This has led to the suggestion that the black hole and bulge or spheroid probably grew in tandem (e.g., Cattaneo & Bernardi 2003; Hopkins et al. 2006). The presence at $z = 2 - 3$ of luminous and potentially massive black holes in high mass galaxies that do not seem to have a prominent bulge or spheroid may be at odds with this picture.

7. DISCUSSION

7.1. Do Massive Galaxies With $n \leq 2$ at $z = 2 - 3$ Host Disks?

We have shown in §3.2 that the majority ($64.9\% \pm 5.4\%$ for $M_\star \geq 5 \times 10^{10}$ $M_\odot$, and $58.5\% \pm 7.7\%$ for $M_\star \geq 10^{11}$ $M_\odot$) of massive galaxies at $z = 2 - 3$ have low $n \leq 2$, while the fraction at $z \sim 0$ is five times lower. We also demonstrated via artificial redshifting experiments and extensive tests (§3.3 and the Appendix) that this difference between $z = 2 - 3$ and $z \sim 0$ is real and not driven primarily by systematic effects. Furthermore, most ($\sim 72\%$) of these with low $n \leq 2$ massive galaxies at $z = 2 - 3$ are extended with $r_e > 2$ kpc, rather than being ultra-compact.

What is the nature of the large population of galaxies with low $n \leq 2$ at $z = 2 - 3$? We present below different lines of evidence which suggest that many of these massive galaxies at $z = 2 - 3$ with $n \leq 2$, particularly the extended ($r_e > 2$ kpc) systems, likely host a significant disk component.

1. Some insight into the interpretation of $n \leq 2$ values can be gleaned by considering massive galaxies at $z \sim 0$. As discussed in §3.3.1 and illustrated in Figure 10, massive E and S0s, which are spheroid-dominated and bulge-dominated systems, are predominantly associated with $n > 2$, both at $z \sim 0$ and after artificially redshifting to $z = 2.5$. In contrast, spiral galaxies of intermediate to low bulge-to-total ratios, often have an overall low Sérsic index

---

[15] The mid-IR selection criteria of Lacy et al. (2004) and Stern et al. (2005) were investigated but considered unreliable. Contamination from high-redshift star-forming galaxies drastically reduces their accuracy (e.g., Donley et al. 2008). Applying these methods at $z = 1 - 3$ would add more false-positives than true AGN.

$n \leq 2$ (Figure 10) because they have a disk component, such as an outer disk or a central disky pseudobulge (e.g., Kormendy & Kennicutt 2004; Jogee 1999; Jogee et al 2005), which contributes significantly to the total blue light of the galaxy. An extension of these arguments to $z = 2-3$ suggests the large fraction $\sim 65\%$ of massive galaxies at $z = 2-3$ with low $n \leq 2$ is driven, at least partially, by the presence of an outer disk or central disky pseudobulge.

2. We next consider the relationship between disk structure and projected ellipticity e. The top panels of Figure 18 show the deconvolved ellipticity $e = 1 - b/a$ determined by GALFIT for the massive ($M_\star \geq 5 \times 10^{10}~M_\odot$) galaxies at $z = 2 - 3$ with $n \leq 2$ and $n > 2$. The lower left and right panels of Figure 18 show the distributions of deconvolved ellipticity determined by GIM2D of similarly massive spiral (Sabc and Sd/Irr) and E/S0[16] galaxies in the MGC catalog.

   The projected ellipticity distribution of massive galaxies at $z = 2 - 3$ with $n \leq 2$ is quite different from that of $z \sim 0$ massive E/S0 galaxies. For local E/S0s, the distribution of e drops sharply at $e > 0.35$ and there are few systems at $e > 0.5$. In contrast, for the massive galaxies at $z = 2 - 3$ with $n \leq 2$, the e distribution continues to rise out to $e \sim 0.5$. There is also a significant fraction ($\sim 58\%$) of systems with $n \leq 2$ having e above 0.5, specifically in the range of 0.5 to 0.75. In effect, a Kolmogorov-Smirnov (KS) test (Table 5) shows that the galaxies at $z = 2 - 3$ with $n \leq 2$ have a 0% KS-test probability of coming from the same distribution as local massive E/S0s in MGC. These comparisons suggest that *the massive galaxies at $z = 2 - 3$ with $n \leq 2$ are very different from local bulge-dominated and spheroid-dominated E/S0s.*

   Among the massive systems with $n \leq 2$ at $z = 2-3$, $28.0 \pm 6.4\%$ are ultra-compact ($r_e \leq 2$ kpc). Thus, our conclusion complements the results of van der Wel et al. (2011) who analyze WFC3 images of a small sample of 14 massive ($M_\star \geq 6 \times 10^{10}~M_\odot$), quiescent, and compact ($r_e \leq 2$ kpc) galaxies at $1.5 < z < 2.5$ and report that most ($65 \pm 15\%$) are disk-dominated systems They find that 5 of 14 galaxies are flat in projection and have an ellipticity ≥ 0.45.

   What is the nature of the massive galaxies at $z = 2 - 3$ with $n \leq 2$? Figure 18 and the KS tests in Table 5 show that the massive galaxies at $z = 2-3$ with $n \leq 2$ are more similar to $z \sim 0$ massive Sd/Irr (KS probability of 23.5% and $D = 0.317$) and to $z \sim 0$ massive Sabc spirals (KS probability of 4.8% and $D = 0.221$) than to $z \sim 0$ massive E/S0s. However, the similarity to massive late-type spirals at $z \sim 0$ is clearly limited, since most massive galaxies at $z = 2-3$ with $n \leq 2$ have smaller half-light radii (r_e primarily below 7 kpc; Figure 5) than any of the $z \sim 0$ massive systems. It is possible that they host less extended and thicker disks than present-day massive spirals.

   Another possibility is that the massive galaxies at $z = 2 - 3$ with $n \leq 2$ might be related to clump-cluster and chain galaxies (Cowie et al. 1995; van den Bergh et al. 1996; Elmegreen et al. 2005, 2009a, 2009b). Such galaxies very often host disk structures (Elmegreen et al. 2009a), and many of them appear to represent a population of highly clumped disk galaxies viewed at different orientations (Elmegreen et al. 2008; Elmegreen et al. 2005). While clumpy disks may be among the massive GNS galaxies with low $n \leq 2$, we cannot identify them due to resolution effects. Finally, we note that in principle a low Sérsic index could be the result of a merger that has not fully coalesced. However, as noted in §3.2 most massive GNS galaxies do not visually appear to be made of multiple distorted systems in early phases of mergers. Artificial redshifting of present-day interacting systems show that our GNS images should be able to resolve systems in early phases of merging, such as NGC4568 and NGC 3396, but would be unlikely to resolve late merger phases, such as Arp 220 into two separate systems.

3. Another line of evidence for massive galaxies at $z \sim 2$ with potentially thick disks comes from the SINS survey (Genzel et al. 2008; Shapiro et al. 2008; Förster Schreiber et al. 2009), which provides ionized gas kinematics of $z \sim 2$ star-forming galaxies and finds examples of clumpy, turbulent, and geometrically thick systems having high velocity dispersions ($\sigma \sim 30-120$ km/s). About $\sim 1/3$ of such systems show rotating disks kinematics. Furthermore, Förster Schreiber et al. (2011) find from *HST* NIC2 imaging that five star-forming galaxies with rotating disk kinematics are well-characterized with shallow $n \leq 1$ Sérsic profiles. Compared to these SINS galaxies, the massive GNS galaxies at $z = 2 - 3$ are more massive on average.

4. In this work (§3.1), we fitted the NIC3/F160W images of the massive galaxies at $z = 2 - 3$ with single Sérsic components, rather than separate bulge and disk components because the low resolution (PSF FWHM of $0\farcs3$ corresponding to ~ 2.4 kpc at $z = 1 - 3$) of the images prevent reliable multiple component decomposition for all the galaxies, particularly the fairly compact ones. However, for the galaxies with large $r_e \geq 4$ kpc we attempted a bulge-plus-disk decomposition following the techniques outlined in Weinzirl et al. (2009). The decomposition was reliable only for the more extended systems within this group and yielded bulge-to-total light ratios below 0.4, indeed suggesting the presence of a significant disk component among massive galaxies at $z = 2 - 3$ with $n \leq 2$.

5. It is also interesting to note that most ($\sim 72\%$ for $M_\star \geq 5 \times 10^{10}~M_\odot$) of these massive galaxies at $z = 2-3$ with low $n \leq 2$ are extended ($r_e > 2$ kpc)

---

[16] The MGC catalog assigns the 'E/S0' Hubble type and unfortunately does not allow us to identify Es separately.

rather than ultra-compact systems. This is in itself does not prove that disk components exist in low $n \leq 2$ systems, but it is suggestive of such a picture. Furthermore, we found in §4.3 that at $z = 2 - 3$, the $n \leq 2$ disky systems have a wide range of SFR_{IR} and include systems of the highest SFR_{IR}. This result is generally consistent with the idea that the systems with $n \leq 2$ are actively star-forming and host copious amounts of gas (§5), which tends to settle in disk-like configurations.

6. For completeness, we note that in principle the presence of a massive disk component is not the only way to produce a low Sérsic index $n \leq 2$ in massive galaxies at $z = 2 - 3$. For the ultra-compact ($r_e \leq 2$ kpc) massive galaxies with $n \leq 2$, it has been argued that such systems could be somewhat like a massive elliptical, which has a bright high surface brightness central component surrounded by a very extended low surface brightness envelope. If the low surface brightness envelope is somehow not detected by the NIC3/F160W images, then the latter could yield a lower $n \leq 2$, as the wings of the surface brightness profile would be effectively clipped. However, this scenario does not seem likely since our artificial redshifting experiments (§3.3.1) show that $z \sim 0$ massive Es are not degraded into ultra-compact systems. Furthermore, Szomoru et al. (2010) confirm the absence of a low surface brightness halo in an ultra-compact, massive galaxy at $z = 1.9$ from extremely deep ($H \sim 28$ mag arcsec^{-2}) WFC3 imaging.

In summary, based on all the above tests and arguments, we conclude that the massive galaxies at $z = 2 - 3$ with $n \leq 2$, particularly the more extended systems with $r_e > 2$ kpc, likely host a massive disk component, which contributes significantly to the rest-frame blue light of the galaxies.

7.2. *Formation of Massive Galaxies By $z = 2 - 3$*

How do the massive galaxies with ultra-compact ($r_e \leq 2$ kpc) and low $n \leq 2$ disky structures form by $z = 2 - 3$? The surface brightness in the rest-frame B-band of the massive galaxies at $z = 2 - 3$ is on average 4.5 magnitudes brighter than massive $z \sim 0$ galaxies (Figure 7). This implies a large mass surface density of young-to-intermediate-age stars had to built up in less than a few Gyr. Implied stellar mass surface densities exceed several $10^{10} M_\odot$ pc^{-2} even for conservative mass-to-light ratios. This implies that *rapid and highly dissipative gas-rich events must have led to the formation of these massive galaxies by $z = 2 - 3$.* Both gas accretion and wet major mergers at $z > 2$ are likely to have played an important role because at such high redshifts, the short dynamical timescales associated with mergers, and the short cooling time associated with gas accretion imply that both mechanisms would lead to a rapid buildup of cold gas. The latter can in turn lead to rapid star formation and dense stellar remnants (e.g., Wuyts et al. 2009; Wuyts et al. 2010; Khochfar & Silk 2011; Bournaud et al. 2011).

A further constraint on the formation pathway is provided by the structure of the massive galaxies at $z = 2 - 3$.

We have shown in §3.2 that as much as $\sim 65\%$ of the massive galaxies at $z = 2 - 3$ have a low $n \leq 2$, and we further argued in §7.1 that most of these systems with $n \leq 2$ at $z = 2 - 3$ likely host a massive disk component. Major mergers of low-to-moderate gas fraction (e.g., $\leq 30\%$) will typically produce merger remnants with a de Vaucouleurs type profile and a Sérsic index $n > 3$ (Naab, Khochfar, & Burkert 2006; Naab & Trujillo 2006). Mergers with moderate-to-high gas fractions are expected to produce lower Sérsic n that are still in general > 2. For instance, Figure 14 of Hopkins et al. (2009) show the Sérsic index of major merger remnants for a range of orbits and a range of progenitors with gas fractions spanning from 10-100%. Although some massive ($M_\star \geq 10^{11} M_\odot$) remnants with $n \sim 1$ arise in mergers with $f_{gas} \geq 80\%$, most remnants of gas-rich ($f_{gas} \geq 40\%$) mergers have a Sérsic index $n > 2$. Furthermore, Rothberg & Joseph (2004) find from K-band imaging of 52 merger remnants that $\sim 51\%$ (26/51) have $n > 3$, $\sim 37\%$ (19/51) have $n \sim 2 - 3$, and only a small fraction ($\sim 12\%$, 6/51) have $n \sim 1 - 2$. Thus, when considering isolated gas-rich major mergers, namely those not fed by cold streams, it is challenging to produce a population of merger remnants where $\sim 65\%$ of the systems have $n \leq 2$.

The challenge of producing a large population of disky ($n \leq 2$) systems with high SFRs from isolated gas-rich major mergers may be an indication that the accretion of cold gas along cosmological filaments ((Birnboim & Dekel 2003; Kereš et al. 2005; Dekel & Birnboim 2006; Dekel et al. 2009a; Dekel et al. 2009b; Kereš et al. 2005; Kereš et al. 2009; Brooks et al. 2009; Ceverino et al. 2010) may be particularly important in the build-up of massive galaxies by $z = 2 - 3$. As merger remnants at $z > 2$ acquire gas via cold-mode accretion, a gas disk is expected to form (Khochfar & Silk 2009a; Burkert et al. 2010). Depending on the angular momentum of the accreted gas, it can settle into a compact disk component or into an outer extended disk. Burkert et al. (2010) discuss a scenario where turbulent rotating disks can form, segregating into compact ($r_e \sim 1 - 3$ kpc) dispersion-dominated ($1 \leq v/\sigma \leq 3$) systems and more extended ($r_e \sim 4 - 8$ kpc), rotation-dominated ($v/\sigma > 3$) disks. The formation of a gas disk via cold-mode accretion and its subsequent conversion into a stellar disk, would lower the overall Sérsic index of the massive galaxies at $z = 2 - 3$, making them more in line with the observed values.

However, many key questions remain unanswered. Can theoretical models account for the observed fractions of massive galaxies with low $n \leq 2$, as well as the fraction of galaxies with ultra-compact ($r_e > 2$ kpc) sizes? Can the relation between structure, SFR, and AGN activity discussed in §4.3 and §6.2, as well as the range in SFR at a given stellar mass, be accounted for? We will address these questions in a future paper (Jogee et al., in preparation) where we perform detailed comparisons to different theoretical scenarios.

7.3. *Transformation of Massive Galaxies at $z = 2 - 3$ Into Present-Day E and S0s*

Next we discuss the transformation of massive galaxies at $z = 2 - 3$ into their more massive present-day descendants, which are primarily E and S0s. During this trans-

formation, the massive galaxies will need to significantly increase n since the majority ($\sim 65\%$) of massive galaxies at $z = 2-3$ have low $n \leq 2$, while the corresponding fraction among massive systems at $z \sim 0$ is five times lower (Table 1 and Figure 5). Similarly, the galaxies will also need to significantly raise r_e, since approximately 40% of massive galaxies at $z = 2-3$ are in the form of ultra-compact ($r_e \leq 2$ kpc) galaxies compared to less than 1% at $z \sim 0$ (Table 1 and Figure 5). In general, the massive $z = 2-3$ galaxies must experience a substantial growth in r_e by up to a factor of ~ 6, a dimming in rest-frame optical surface brightness within r_e by up to 6 magnitudes (Figure 7), and their n must increase to $n > 2$. An increase in (n, r_e) and a dimming in μ_e can be achieved via several pathways.

A natural pathway to produce large changes in (n, r_e, μ_e) is a dry major merger of two disk systems. This produces a remnant with $n \sim 4$, a higher r_e, and a lower surface brightness within r_e than the progenitors (Naab, Khochfar, & Burkert 2006; Naab & Trujillo 2006; Naab et al. 2009). In this case, the change in n is produced by the transformation of galaxies with disks into systems dominated by spheroids or bulges. This type of transformation must take place from $z = 2-3$ to $z \sim 0$ in many of the massive galaxies because $\sim 65\%$ of them at $z = 2-3$ have $n \leq 2$, which we argued is indicative of a massive disk in many cases (§7.1). In contrast the E/S0s at $z \sim 0$ are dominated by spheroids or bulges.

Other lines of evidence support the idea that dry major mergers play a role in making the most massive $z \sim 0$ ellipticals. The most massive local ellipticals are found to harbor cores (missing light), which are believed to be scoured by binary black holes that form in dry major mergers (Kormendy et al. 2009). From a study of the tidal features associated with bulge-dominated early-type galaxies, van Dokkum (2005) concludes that today's most luminous ellipticals form through mergers of gas-poor, bulge-dominated systems. Kriek et al. (2008) focus on massive red-sequence galaxies at $z \sim 2.3$ with little or no ongoing star formation, finding that the changes in color and number density of galaxies on the high-mass end ($M_\star \geq 1 \times 10^{11}$ $M_\odot$) of the red sequence from $z \sim 2.3$ to the present are better explained by a combination of passive evolution and red mergers that induce little star formation, rather than by passive evolution alone.

While dry major mergers play a role in the evolution of massive galaxies, it remains debated whether they can account for the full size and mass evolution of massive galaxies. From a theoretical standpoint, the predicted dry major merger rate appears to be too low. From simulations, Khochfar & Silk (2009b) find that only between $10\% - 20\%$ of massive ($M_\star > 6.3 \times 10^{10}$ $M_\odot$) galaxies have had a dry major merger in the last Gyr at any redshift $z < 1$. Hopkins et al. (2010) find from semi-empirical models that the importance of major mergers in bulge formation scales with galaxy stellar mass. Namely, an L_* galaxy with $M_\star \sim 10^{11} M_\odot$ at $z = 0$ will experience only one dry major merger at $z < 2$. Shankar et al. (2010) calculate that the frequency of dry mergers increases with final stellar mass, and they find that by $z = 0$ massive ($M_\star > 10^{11} M_\odot$) early-type galaxies undergo on average < 1 dry major merger since their formation.

From an observational standpoint, direct measurements of the dry major merger rate at $z < 1$ are highly uncertain. Bell et al. (2006) suggest that present-day spheroidal galaxies with $M_V < -20.5$ on average have undergone anywhere between 0.5 and 2 dry major mergers since $z \sim 0.7$. The analysis carries large uncertainties as it is based on a small number (~ 6) of observed dry major mergers. Several observational studies report that between 16% to 35% of massive ($M_\star > 2.5 \times 10^{10}$ $M_\odot$) galaxies have undergone a major merger since $z \sim 0.8$ (e.g., Jogee et al. 2009; Lotz et al. 2008; Conselice et al. 2009), but it should be noted that most of the major mergers in the above studies are star-forming systems, and there are very few dry major mergers. Robaina et al. (2010) find that galaxies with $M_\star > 1 \times 11^{10}$ $M_\odot$ have undergone, on average, only 0.5 mergers since $z \sim 0.7$ involving progenitor galaxies that are both more massive than $M_\star > 5 \times 10^{10}$ $M_\odot$. Hammer et al. (2009) focus on starbursts with disturbed ionized gas morphologies and kinematics at $z \sim 0.65$, and they argue based on modeling that ~ 6 Gyr ago 46% of the galaxy population was involved in major mergers, most of which were gas-rich. Kaviraj et al. (2011) find that theoretically and empirically determined major merger rates at $z < 1$ are too low by factors of a few to account for the fraction of disturbed systems they find among morphologically classified early-type massive ($M_\star > 1 \times 10^{10}$ $M_\odot$) galaxies at $0.5 < z < 0.7$. They suggest that the overall evolution of massive early type galaxies, particularly the low-level star formation activity, may be heavily influenced by minor merging at late epochs. At higher redshifts $1 < z < 2$, higher major merger rates are reported than at $z < 1$ (e.g., Conselice et al. 2003), but the frequency of dry major mergers is claimed to be low (Williams et al. 2011).

An alternate pathway that could be at least as important as major mergers consists of consecutive dry minor mergers or accretion of externally formed stars such that stellar mass is cumulatively added to the outskirts of a compact galaxy (e.g., Naab et al. 2009; Feldman et al. 2010). Naab & Trujillo (2006) show that successive minor mergers can, on average, raise the Sérsic index of the merger remnant about as effectively as major mergers. Furthermore, it is claimed from simulations and analytical arguments that dry minor mergers produce a much larger increase in size (r_e) and a larger fall in average stellar mass densities within r_e than do dry major mergers (Naab et al. 2009; Bezanson et al. 2009). Shankar et al. (2011) find in simulations that massive ($M_\star \geq 10^{11}$ $M_\odot$) $z \sim 0$ galaxies grow primarily by dry minor mergers, especially at $z < 1$. Oser et al. (2010; 2011) use cosmological simulations to study 40 individual massive galaxies with present-day stellar masses of $M_\star > 6.3 \times 10^{10}$ $M_\odot$. They find that massive galaxies at $z > 2$ are dominated by "in situ" star formation fueled by in-falling cold gas within the galaxy. As cold-mode accretion becomes inefficient at $z \approx 2$, accretion of externally created stars (i.e., stellar satellites) dominates at $z < 2$. For galaxies of present-day stellar mass $M_\star > 6.3 \times 10^{10}$ $M_\odot$, the average number-weighted merger mass-ratio is $\sim 1:16$, while the average mass-weighted merger mass-ratio is $\sim 1:5$. In other words, the mass growth since $z \sim 2$ is dominated by minor mergers with a mass ratio of 1:5. The importance

of stellar accretion increases with galaxy mass and toward lower redshift, and it substantially raises the galaxy stellar mass and size. For systems with present-day stellar mass $M_\star > 6.3 \times 10^{10}\ M_\odot$, a size evolution of up to a factor of $\sim 5 - 6$ occurs from $z = 2$ to $z \sim 0$. However, one strong caveat of these simulations is that all their massive ($M_\star > 1 \times 10^{11}\ M_\odot$) galaxies at $z = 2$ are ultra-compact ($r_e \leq 2$ kpc), while observations (see Fig. 5) show a large fraction of such massive galaxies at $z = 2$ are extended ($r_e = 3 - 10$ kpc), with a wide range in star formation rate. The increase of size and mass induced by minor mergers in these simulations is qualitatively in agreement with our results on size evolution for the ultra-compact systems and also with the inside-out growth reported by van Dokkum et al. (2010) from stacking deep rest-frame R-band images of massive galaxies over the redshift range of 0.6 to 2.0.

However, many questions remain unresolved. While dry minor mergers appear to be effective at inducing significant evolution in mass and size from $z \sim 2$ to $z \sim 0$ in the simulations of Oser et al. (2010; 2011), it is unclear if they can really drive the large change in Sersic index n required by the observations. Furthermore, these simulations focus only on ultra-compact ($r_e \leq 2$ kpc) galaxies, and are not representative of the large dominant population of more extended galaxies at $z = 2 - 3$. Finally, it is not clear whether minor mergers can account for the changes in effective surface brightness between $z = 2 - 3$ and $z \sim 0$. We will evaluate these issues more thoroughly with a detailed comparison to models in a subsequent paper (Jogee et al. in preparation).

8. SUMMARY

We present a study of the structure, activity, and evolution of massive galaxies at $z = 1 - 3$ using deep (5σ limiting magnitude of H=26.8 AB for an extended source of diameter $0\rlap{.}''7$), high resolution (PSF$\sim 0\rlap{.}''3$) NIC3/F160W images from the GOODS-NICMOS Survey (GNS), along with complementary ACS, *Spitzer* IRAC and MIPS, and Chandra X-ray data. One of the strengths of our study is that the NIC3/F160W data provide rest-frame optical imaging over $z = 1 - 3$ for one of the largest (166 galaxies with $M_\star \geq 5 \times 10^{10}\ M_\odot$ and 82 with $M_\star \geq 10^{11}\ M_\odot$), most diverse, and relatively unbiased samples of massive galaxies at $z = 1 - 3$ studied to date. Our main results are summarized below.

1. Structure of massive galaxies at rest-frame optical wavelengths: We analyze the rest-frame optical structure of the massive galaxies by fitting single Sérsic profiles to the 2D light distribution in the NIC3/F160W images. We find that the rest-frame optical structures of the massive galaxies are very different at $z = 2 - 3$ compared to $z \sim 0$, with their Sérsic index n and half-light radius r_e being strikingly offset toward lower values compared to $z \sim 0$. (Table 1 and Figure 5). Through extensive tests and artificial redshifting experiments we conclude that the offset in (n, r_e) between massive galaxies at $z = 2 - 3$ and $z \sim 0$ is real and not primarily driven by systematic effects related to the fitting techniques instrumental effects, or redshift-dependent effects (e.g., cosmological surface brightness dimming and the loss of spatial resolution). In effect, we find *a large population of ultra-compact*

($r_e \leq 2$ kpc) systems, as well as *a dominant population of systems with low $n \leq 2$ disky morphologies* at $z = 2 - 3$. We further describe these populations below.

We find that approximately 40% ($39.0 \pm 5.6\%$ for $M_\star \geq 5 \times 10^{10}\ M_\odot$ and $39.0 \pm 7.6\%$ for $M_\star \geq 1 \times 10^{11}\ M_\odot$) of the massive galaxies at $z = 2 - 3$ are in the form of ultra-compact ($r_e \leq 2$ kpc) galaxies compared to less than 1% at $z \sim 0$ (Table 1 and Figure 5). These ultra-compact galaxies are practically unmatched among $z \sim 0$ massive galaxies, and their surface brightness in the rest-frame optical can be 4-6 magnitudes brighter (Figure 7).

Secondly, we find that the majority ($64.9\% \pm 5.4\%$ for $M_\star \geq 5 \times 10^{10}\ M_\odot$, and $58.5\% \pm 7.7\%$ for $M_\star \geq 10^{11}\ M_\odot$) of massive galaxies at $z = 2 - 3$ have low $n \leq 2$, while the corresponding fraction among massive systems at $z \sim 0$ is five times lower. Most ($\sim 72\%$) of these massive galaxies at $z = 2 - 3$ with low $n \leq 2$ have $r_e > 2$ kpc, and therefore complement the ultra-compact galaxies. We further explore the meaning of a Sérsic index $n \leq 2$ at $z = 2 - 3$, and present evidence that *most of the massive galaxies with $n \leq 2$ at $z = 2 - 3$, particularly the extended ($r_e > 2$ kpc) ones, likely host a prominent disk*, unlike the majority of massive galaxies at $z \sim 0$. Our evidence is based on rest-frame optical morphologies, ellipticities, artificial redshifting experiments, as well as bulge-to-total ratios from bulge-plus-disk decompositions of extended systems.

2. Star formation rates: We estimate star formation rates using IR luminosities (8-1000 μm) derived from the *Spitzer* 24 μm flux for massive GNS galaxies having a secure MIPS 24 μm counterpart and a 24 μm flux exceeding the 5σ detection limit of 30 μJy. AGN host candidates are excluded because the inferred IR luminosities overestimate the true star formation rate.

We find a strong link between galaxy structure and SFR. Among the non-AGN massive ($M_\star \geq 5 \times 10^{10}\ M_\odot$) galaxies at $z = 2 - 3$ with SFR$_{\rm IR}$ high enough to yield a 5σ (30 μJy) *Spitzer* 24 μm detection, the majority ($84.6 \pm 10.0\%$) have low $n \leq 2$. While the $n \leq 2$ disky systems have a wide range of SFR$_{\rm IR}$ (53 to 1466 $M_\odot$ yr^{-1} at $z = 2 - 3$), they include the systems of the highest SFR$_{\rm IR}$ at both $z = 1 - 2$ and $z = 2 - 3$. In contrast, the massive ultra-compact objects at $z = 2 - 3$ are less likely by a factor of ~ 2.2 to have SFR$_{\rm IR}$ above the detection limit, compared to the whole sample of non-AGN massive galaxies.

3. AGN activity: Using a variety of techniques (X-ray properties, IR power-law, and IR-to-optical excess) to identify AGN, we find that 49/166 ($29.5 \pm 3.5\%$) of the massive galaxies at $z = 1 - 3$ are AGN candidates. The AGN fraction rises with redshift, increasing from $17.9 \pm 6.1\%$ at $z = 1 - 1.5$ to $40.3 \pm 8.8\%$ at $z = 2 - 3$ (Table 4).

We find a relationship between host galaxy structure and AGN activity that complements the relationship between SFR and structure. Among massive galaxies at $z = 2 - 3$, AGN appear to be found preferentially in galaxies that are not ultra-compact, as evidenced by the fact that most ($80.6 \pm 7.9\%$) AGN hosts have $r_e > 2$ kpc. In fact, at $z = 2 - 3$, the AGN fraction in ultra-compact galaxies is ~ 2.7 times lower than in extended galaxies ($20.0 \pm 16.3\%$ versus $53.2 \pm 10.0\%$). *Thus, ultra-compact galaxies appear quiescent in terms of both SFR and AGN activity.* In terms of their Sérsic index n, a large fraction

(64.6±10.7%) of AGN hosts at $z = 2-3$ have disky ($n \leq 2$) morphologies.

4. Cold gas content: We apply a standard Schmidt-Kennicutt law (Kennicutt 1998) to the $\mathrm{SFR_{IR}}$ of the non-AGN host candidates. The high estimated $\mathrm{SFR_{IR}}$ suggest that copious cold gas reservoirs are present. We estimate that the average cold gas surface density in non-AGN hosts ranges from ~ 136 to $\sim 25,091$ $M_\odot$ pc^{-2} at $z = 1-3$, with a median value of ~ 607 $M_\odot$ pc^{-2} (Figure 15). The implied cold gas fraction within the rest-frame optical half-light radius ranges from 6.5% to 65.4%, with a mean of $\sim 41\%$ at $z = 2-3$ (Figure 15). The highest gas fractions at a given redshift are found among the less massive galaxies, consistent with downsizing.

5. Formation of massive galaxies by $z = 2-3$: The massive galaxies at $z = 2-3$ already have an average rest-frame optical surface brightness within r_e that can be up to 3-6 magnitudes brighter than $z \sim 0$ massive galaxies. The associated high stellar mass densities imply that massive galaxies at $z = 2-3$ must have formed via rapid, highly dissipative events at $z > 2$. Both gas-rich major mergers and gas accretion at $z > 2$ are viable as their associated short dynamical timescales and short gas cooling times at $z > 2$ would lead to a rapid buildup of mass. However, the large fraction ($\sim 65\%$) of massive galaxies at $z = 2-3$ with $n \leq 2$ and disky morphologies suggest that cold-mode accretion at $z > 2$ must have played an important role in the build-up of massive galaxies by $z = 2-3$, since it may be challenging to have such a large fraction of of merger remnants with low $n \leq 2$ from isolated gas-rich major mergers.

6. Transformation of massive galaxies at $z = 2-3$ into present-day E and S0s: In order for massive galaxies at $z = 2-3$ to evolve into $z \sim 0$ massive systems (which are primarily E and S0s), they need to radically change their rest-frame optical structure and distributions of (n, r_e). In particular they need to raise n well above 2, increase r_e by an average factor of 3-4, and dim the average rest-frame optical surface brightness. Dry major mergers can induce changes in galaxy size, Sérsic index, and stellar surface density, but they may be too rare to account for all the needed evolution. Successive dry minor mergers have been shown to influence galaxy size, Sérsic index, and stellar surface density in a similar direction. We suggest the transformation of massive $z = 2-3$ galaxies into $z \sim 0$ galaxies will occur through a combination of dry major mergers, minor mergers. We will investigate in the relative importance and efficiency of these mechanisms in a future paper.

REFERENCES

Alexander, D. M., Bauer, F. E., Chapman, S. C., Smail, I., Blain, A. W., Brandt, W. N., & Ivison, R. J. 2005, ApJ, 632, 736
Alexander, D. M., et al. 2003, AJ, 126, 539
Allen, P. D., Driver, S. P., Graham, A. W., Cameron, E., Liske, J., & de Propris, R. 2006, MNRAS, 371, 2
Alonso-Herrero, A., et al. 2006, ApJ, 640, 167
Barden, M., Jahnke, K., Haüßler, B. 2008, ApJS, 175, 105
Barden, M., et al. 2005, ApJ, 635, 959
Bauer, A. E., Conselice, C. J., Pérez-González, P. G., Grützbauch, R., Bluck, A. F. L., Buitrago, F., Mortlock, A., 2011, MNRAS, in press, arXiv:1106.2656
Beckwith, S. V. W., et al. 2006, AJ, 132, 1729
Bell, E. F., Phleps, S., Somerville, R. S., Wolf, C., Borch, A., & Meisenheimer, K. 2006, ApJ, 652, 270
Bell, E. F., Zheng, X. Z., Papovich, C., Borch, A., Wolf, C., & Meisenheimer, K. 2007, ApJ, 663, 834
Bell, E. F., et al. 2004, ApJL, 600, L11
Bell, E. F., et al. 2005, ApJ, 625, 23
Bell, E. F., et al. 2006, ApJ, 640, 241
Bezanson, R., van Dokkum, P. G., Tal, T., Marchesini, D., Kriek, M., Franx, M., & Coppi, P. 2009, ApJ, 697, 1290
Binney, J. & Merrifield, M. 1998, Galactic Astronomy (Princeton: Princeton Univ. Press)
Birnboim, Y., & Dekel, A. 2003, MNRAS, 345, 349
Blain, A. W., Smail, I., Ivison, R. J., Kneib, J.-P., & Frayer, D. T. 2002, Phys. Rep., 369, 111
Bluck, A. F. L., Conselice, C. J., Bouwens, R. J., Daddi, E., Dickinson, M., Papovich, C., & Yan, H. 2009, MNRAS, 394, L51
Borch, A., Meisenheimer, K., Bell, E. F., et al. 2006, A&A, 453, 869
Bournaud, F., Dekel, A., Teyssier, R., Cacciato, M., Daddi, E., Juneau, S., & Shankar, F. 2011, arXiv:1107.1483
Brand, K., et al. 2006, ApJ, 644, 143
Brandt et al. 2006, in Physics of Active Galactic Nuclei at all Scales, ed. D. Alloin, R. Johnson, & P. Lira (Berlin: Springer), 185
Brooks, A. M., Governato, F., Quinn, T., Brook, C. B., & Wadsley, J. 2009, ApJ, 694, 396
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Buitrago, F., Trujillo, I., Conselice, C. J., Bouwens, R. J., Dickinson, M., & Yan, H. 2008, ApJL, 687, L61
Burkert, A., et al. 2010, ApJ, 725, 2324
Cattaneo, A., & Bernardi, M. 2003, MNRAS, 344, 45
Ceverino, D., Dekel, A., & Bournaud, F. 2010, MNRAS, 404, 2151
Chabrier, G. 2003, ApJL, 586, L133
Chary, R., & Elbaz, D. 2001, ApJ, 556, 562
Cimatti, A., et al. 2008, A&A, 482, 21
Cole, S., Lacey, C. G., Baugh, C. M., & Frenk, C. S. 2000, MNRAS, 319, 168
Conselice, C. J. 2009, MNRAS, 399, L16
Conselice, C. J., Bershady, M. A., Dickinson, M., & Papovich, C. 2003, AJ, 126, 1183
Conselice, C. J., Rajgor, S., & Myers, R. 2008, MNRAS, 386, 909
Conselice, C. J., et al. 2011, MNRAS, 226
Conselice, C. J., et al. 2007, MNRAS, 381, 962
Cowie, L. L., Hu, E. M., & Songaila, A. 1995, AJ, 110, 1576
Daddi, E., Cimatti, A., Renzini, A., Fontana, A., Mignoli, M., Pozzetti, L., Tozzi, P., & Zamorani, G. 2004, ApJ, 617, 746
Daddi, E., Dannerbauer, H., Elbaz, D., Dickinson, M., Morrison, G., Stern, D., & Ravindranath, S. 2008, ApJL, 673, L21
Daddi, E., et al. 2005, ApJ, 626, 680
Daddi, E., et al. 2007, ApJ, 670, 156
Daddi, E., et al. 2010a, ApJ, 713, 686
Daddi, E., et al. 2010b, ApJL, 714, L118
Dekel, A., & Birnboim, Y. 2006, MNRAS, 368, 2
Dekel, A., Sari, R., & Ceverino, D. 2009a, ApJ, 703, 785
Dekel, A., et al. 2009b, Nature, 457, 451
Dickinson, M., Giavalisco, M., & GOODS Team 2003a, in The Mass of Galaxies at Low and High Redshift, ed. R. Bender & A. Renzini (Heidelberg: Springer-Verlag), 324
Dickinson, M., Papovich, C., Ferguson, H. C., & Budavári, T. 2003b, ApJ, 587, 25
Dickinson, M., et al. 2004, ApJL, 600, L99
Donley, J. L., Rieke, G. H., Pérez-González, P. G., & Barro, G. 2008, ApJ, 687, 111
Driver, S. P., Liske, J., Cross, N. J. G., De Propris, R., & Allen, P. D. 2005, MNRAS, 360, 81
Drory, N., & Alvarez, M. 2008, ApJ, 680, 41
Drory, N., Salvato, M., Gabasch, A., Bender, R., Hopp, U., Feulner, G., & Pannella, M. 2005, ApJL, 619, L131
Elbaz, D., et al. 2010, A&A, 518, L29
Elmegreen, B. G., Bournaud, F., & Elmegreen, D. M. 2008, ApJ, 688, 67
Elmegreen, B. G., Elmegreen, D. M., Fernandez, M. X., & Lemonias, J. J. 2009a, ApJ, 692, 12
Elmegreen, D. M., Elmegreen, B. G., Marcus, M. T., et al. 2009b, ApJ, 701, 306
Elmegreen, D. M., Elmegreen, B. G., Rubin, D. S., & Schaffer, M. A. 2005, ApJ, 631, 85
Elsner, F., Feulner, G., & Hopp, U. 2008, A&A, 477, 503
Eskridge, P. B., et al. 2002, ApJS, 143, 73
Feldmann, R., Carollo, C. M., Mayer, L., Renzini, A., Lake, G., Quinn, T., Stinson, G. S., & Yepes, G. 2010, ApJ, 709, 218
Ferrarese, L., & Merritt, D. 2000, ApJL, 539, L9
Fiore, F., et al. 2008, ApJ, 672, 94
Fontana, A., Salimbeni, S., Grazian, A., et al. 2006, A&A, 459, 745
Förster Schreiber, N. M., et al. 2009, ApJ, 706, 1364
Förster Schreiber, N. M., Shapley, A. E., Erb, D. K., Genzel, R., Steidel, C. C., Bouché, N., Cresci, G., & Davies, R. 2011, ApJ, 731, 65
Franceschini, A., et al. 2005, AJ, 129, 2074
Gebhardt, K., et al. 2000, ApJL, 539, L13
Genzel, R., et al. 2008, ApJ, 687, 59
Genzel, R., et al. 2010, MNRAS, 407, 2091
Giacconi, R., et al. 2002, ApJS, 139, 369
Giavalisco, M., et al. 2004, ApJL, 600, L93
Gnedin, N. Y., & Kravtsov, A. V. 2010, ApJ, 714, 287
Graham, A. W. 2001, AJ, 121, 820
Grützbauch, R., Chuter, R. W., Conselice, C. J., Bauer, A. E., Bluck, A. F. L., Buitrago, F., & Mortlock, A. 2010, MNRAS, in press, arXiv:1011.4846
Guzman, R., Gallego, J., Koo, D. C., Phillips, A. C., Lowenthal, J. D., Faber, S. M., Illingworth, G. D., & Vogt, N. P. 1997, ApJ, 489, 559
Hammer, F., Flores, H., Puech, M., et al. 2009, A&A, 507, 1313
Häussler, B., et al. 2007, ApJS, 172, 615
Hogg, D. W. 1999, arXiv:astro-ph/9905116
Hogg, D. W., Baldry, I. K., Blanton, M. R., & Eisenstein, D. J. 2002, arXiv:astro-ph/0210394
Hopkins, P. F., Cox, T. J., Dutta, S. N., Hernquist, L., Kormendy, J., & Lauer, T. R. 2009, ApJS, 181, 135
Hopkins, P. F., Hernquist, L., Cox, T. J., Robertson, B., & Springel, V. 2006, ApJS, 163, 50
Hopkins, P. F., et al. 2010, ApJ, 715, 202
Jogee, S. 1999, Ph.D. Thesis,
Jogee, S., Scoville, N., & Kenney, J. D. P. 2005, ApJ, 630, 837
Jogee, S., et al. 2009, ApJ, 697, 1971
Jogee, S., et al. 2004, ApJ, 615, L105
Kajisawa, M., Ichikawa, T., Tanaka, I., et al. 2009, ApJ, 702, 1393
Kajisawa, M., et al. 2006, PASJ, 58, 951
Kaviraj, S., Tan, K.-M., Ellis, R. S., & Silk, J. 2011, MNRAS, 411, 2148
Kennicutt, R. C., Jr. 1998, ApJ, 498, 541
Kennicutt, R. C., Jr. 2008, in Astronomical Society of the Pacific Conference Series, Vol. 390, Pathways Through an Eclectic Universe, ed. J. H. Knapen, T. J. Mahoney, & A. Vazdekis, 149
Kereš, D., Katz, N., Fardal, M., Davé, R., & Weinberg, D. H. 2009, MNRAS, 395, 160
Kereš, D., Katz, N., Weinberg, D. H., & Davé, R. 2005, MNRAS, 363, 2
Khochfar, S., & Silk, J. 2006, ApJL, 648, L21
Khochfar, S., & Silk, J. 2009a, ApJL, 700, L21
Khochfar, S., & Silk, J. 2009b, MNRAS, 397, 506
Khochfar, S., & Silk, J. 2011, MNRAS, 410, L42
Kocevski, D. D., et al. 2011, ApJ submitted, arXiv:1109.2588
Kormendy, J., Fisher, D. B., Cornell, M. E., & Bender, R. 2009, ApJS, 182, 216
Kormendy, J. in the Nearest Active Galaxies (eds Beckman, J., Colina, L., & Netzer, H.) 197-218 (Consejo Superior de Investigaciones Cientificas, Madrid, 1993)
Kormendy, J., & Kennicutt, R. C., Jr. 2004, ARAA, 42, 603
Kriek, M., van der Wel, A., van Dokkum, P. G., Franx, M., & Illingworth, G. D. 2008, ApJ, 682, 896
Krist, J. 1995, Astronomical Data Analysis Software and Systems IV, 77, 349
Labbé, I., et al. 2003, ApJL, 591, L95
Lacy, M., et al. 2004, ApJS, 154, 166
Lehmer, B. D., et al. 2005, ApJS, 161, 21
Lilly, S., et al. 1998, ApJ, 500, 75
Liske, J., Lemon, D. J., Driver, S. P., Cross, N. J. G., & Couch, W. J. 2003, MNRAS, 344, 307
Loeb, A., & Peebles, P. J. E. 2003, ApJ, 589, 29
Longhetti, M., et al. 2007, MNRAS, 374, 614
Lotz, J. M., Davis, M., Faber, S. M., et al. 2008, ApJ, 672, 177
Luo, B., et al. 2008, ApJS, 179, 19
Magorrian, J., et al. 1998, AJ, 115, 2285
Marchesini, D., van Dokkum, P. G., Förster Schreiber, N. M., et al. 2009, ApJ, 701, 1765
Marconi, A., & Hunt, L. K. 2003, ApJL, 589, L21

Mortlock, A., Conselice, C. J., Bluck, A. F. L., Bauer, A. E., Grützbauch, R., Buitrago, F., & Ownsworth, J. 2011, MNRAS, 413, 2845

Moustakas, L. A., et al. 2004, ApJL, 600, L131

Murphy, E. J., Chary, R.-R., Alexander, D. M., Dickinson, M., Magnelli, B., Morrison, G., Pope, A., & Teplitz, H. I. 2009, ApJ, 698, 1380

Naab, T., Khochfar, S., & Burkert, A. 2006, ApJL, 636, L81

Naab, T., Johansson, P. H., Ostriker, J. P., & Efstathiou, G. 2007, ApJ, 658, 710

Naab, T., Johansson, P. H., & Ostriker, J. P. 2009, ApJL, 699, L178

Naab, T., & Trujillo, I. 2006, MNRAS, 369, 625

Nordon, R., et al. 2010, A&A, 518, L24

Oser, L., Ostriker, J. P., Naab, T., Johansson, P. H., & Burkert, A. 2010, ApJ, 725, 2312

Oser et al. 2011, ApJ, in press, arXiv:1106.5490

Papovich, C., et al. 2006, ApJ, 640, 92

Papovich, C., et al. 2007, ApJ, 668, 45

Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, AJ, 124, 266

Pérez-González, P. G., Rieke, G. H., Villar, V., et al. 2008, ApJ, 675, 234

Pérez-González, P. G., et al. 2008, ApJ, 675, 234

Pierce, C. M., et al. 2010, MNRAS, 405, 718

Ravindranath, S., et al. 2004, ApJL, 604, L9

Reddy, N. A., Steidel, C. C., Pettini, M., Adelberger, K. L., Shapley, A. E., Erb, D. K., & Dickinson, M. 2008, ApJS, 175, 48

Retzlaff, J., Rosati, P., Dickinson, M., Vandame, B., Rité, C., Nonino, M., Cesarsky, C., & GOODS Team 2010, A&A, 511, A50

Robaina, A. R., Bell, E. F., van der Wel, A., Somerville, R. S., Skelton, R. E., McIntosh, D. H., Meisenheimer, K., & Wolf, C. 2010, ApJ, 719, 844

Robertson, B., Cox, T. J., Hernquist, L., Franx, M., Hopkins, P. F., Martini, P., & Springel, V. 2006, ApJ, 641, 21

Rothberg, B., & Joseph, R. D. 2004, AJ, 128, 2098

Schawinski, K., Treister, E., Urry, C. M., Cardamone, C. N., Simmons, B., & Yi, S. K. 2011, ApJL, 727, L31

Sersic, J. L. 1968, Cordoba, Argentina: Observatorio Astronomico, 1968

Shankar, F., Marulli, F., Bernardi, M., Boylan-Kolchin, M., Dai, X., & Khochfar, S. 2010, MNRAS, 405, 948

Shankar, F., Marulli, F., Bernardi, M., Mei, S., Meert, A., & Vikram, V. 2011, MNRAS submitted, arXiv:1105.6043

Shapiro, K. L., et al. 2008, ApJ, 682, 231

Silva, L., Maiolino, R., & Granato, G. L. 2004, MNRAS, 355, 973

Simard, L., et al. 2002, ApJS, 142, 1

Somerville, R. S., & Primack, J. R. 1999, MNRAS, 310, 1087

Steinmetz, M., & Navarro, J. F. 2002, NewA, 7, 155

Stern, D., et al. 2005, ApJ, 631, 163

Szomoru, D., et al. 2010, ApJL, 714, L244

Tacconi, L. J., et al. 2010, Nature, 463, 781

Thompson, R. I., et al. 2005, AJ, 130, 1

Toft, S., et al. 2007, ApJ, 671, 285

Trujillo, I., Conselice, C. J., Bundy, K., Cooper, M. C., Eisenhardt, P., & Ellis, R. S. 2007, MNRAS, 382, 109

Trujillo, I., et al. 2006, ApJ, 650, 18

van den Bergh, S., Abraham, R. G., Ellis, R. S., Tanvir, N. R., Santiago, B. X., & Glazebrook, K. G. 1996, AJ, 112, 359

van der Wel, A., et al. 2011, ApJ, 730, 38

van Dokkum, P. G. 2005, AJ, 130, 2647

van Dokkum, P. G., et al. 2006, ApJL, 638, L59

van Dokkum, P. G., et al. 2008, ApJL, 677, L5

van Dokkum, P. G., et al. 2010, ApJ, 709, 1018

Weedman D. W. 1986. Quasar Astronomy, pp. 60-62. Cambridge: Cambridge University Press

Weinzirl, T., Jogee, S., Khochfar, S., Burkert, A., & Kormendy, J. 2009, ApJ, 696, 411

Williams, R. J., Quadri, R. F., & Franx, M. 2011, ApJL, 738, L25

Williams, R. J., Quadri, R. F., Franx, M., et al. 2010, ApJ, 713, 738

Wuyts, S., Cox, T. J., Hayward, C. C., Franx, M., Hernquist, L., Hopkins, P. F., Jonsson, P., & van Dokkum, P. G. 2010, ApJ, 722, 1666

Wuyts, S., et al. 2009, ApJ, 700, 799

Yan, H., et al. 2004, ApJ, 616, 63

Yan, L., & Thompson, D. 2003, ApJ, 586, 765

Zirm, A. W., et al. 2007, ApJ, 656, 66

v

TABLE 1

REST-FRAME OPTICAL SÉRSIC INDEX n AND r_e IN MASSIVE ($M_\star \geq 5 \times 10^{10}\ M_\odot$) GALAXIES

z	Morphology	$n \leq 2$	$n > 2$	$n > 3$
		$M_\star \geq 5 \times 10^{10}\ M_\odot$		
$z = 2 - 3$ ($N = 77$)	All	**64.9** $\pm 5.44\%$	35.1 $\pm 5.44\%$	18.2 $\pm 4.40\%$
$z = 1 - 2$ ($N = 89$)	All	**49.4** $\pm 5.30\%$	50.6 $\pm 5.30\%$	30.3 $\pm 4.87\%$
$z \sim 0$ ($N = 385$)	All	**13.0** $\pm 1.71\%$	87.0 $\pm 1.71\%$	74.3 $\pm 2.23\%$
	E/S0	0.8 $\pm 0.45\%$	64.9 $\pm 2.43\%$	58.7 $\pm 2.51\%$
	Sabc	10.4 $\pm 1.56\%$	20.8 $\pm 2.07\%$	14.8 $\pm 1.81\%$
	Sd/Irr	1.82 $\pm 0.68\%$	1.30 $\pm 0.58\%$	0.78 $\pm 0.45\%$
		$M_\star \geq 1 \times 10^{11}\ M_\odot$		
$z = 2 - 3$ ($N = 41$)	All	**58.5** $\pm 7.69\%$	41.5 $\pm 7.69\%$	17.1 $\pm 5.88\%$
$z = 1 - 2$ ($N = 41$)	All	**34.1** $\pm 7.41\%$	65.9 $\pm 7.41\%$	43.9 $\pm 7.45\%$
$z \sim 0$ ($N = 115$)	All	**10.4** $\pm 2.85\%$	89.6 $\pm 2.85\%$	80.9 $\pm 3.67\%$
	E/S0	1.7 $\pm 1.22\%$	72.2 $\pm 4.18\%$	67.0 $\pm 4.39\%$
	Sabc	6.09 $\pm 2.23\%$	13.9 $\pm 3.23\%$	12.2 $\pm 3.05\%$
	Sd/Irr	2.61 $\pm 1.49\%$	3.48 $\pm 1.71\%$	1.74 $\pm 1.22\%$
z	Morphology	$r_e \leq 2$ kpc	$2 < r_e \leq 4$ kpc	$r_e > 4$ kpc
		$M_\star \geq 5 \times 10^{10}\ M_\odot$		
$z = 2 - 3$ ($N = 77$)	All	**39.0** $\pm 5.56\%$	42.9 $\pm 5.64\%$	18.2 $\pm 4.40\%$
$z = 1 - 2$ ($N = 89$)	All	**24.7** $\pm 4.57\%$	48.3 $\pm 5.30\%$	27.0 $\pm 4.70\%$
$z \sim 0$ ($N = 385$)	All	**0.52** $\pm 0.37\%$	1.8 $\pm 0.68\%$	97.7 $\pm 0.77\%$
	E/S0	0.26 $\pm 0.26\%$	1.8 $\pm 0.68\%$	63.6 $\pm 2.45\%$
	Sabc	0.00 $\pm 0.00\%$	0.0 $\pm 0.00\%$	31.2 $\pm 2.36\%$
	Sd/Irr	0.26 $\pm 0.26\%$	0.00 $\pm 0.00\%$	2.86 $\pm 0.85\%$
		$M_\star \geq 1 \times 10^{11}\ M_\odot$		
$z = 2 - 3$ ($N = 41$)	All	**39.0** $\pm 7.62\%$	41.5 $\pm 7.69\%$	19.5 $\pm 6.19\%$
$z = 1 - 2$ ($N = 41$)	All	**22.0** $\pm 6.46\%$	56.1 $\pm 7.75\%$	22.0 $\pm 6.46\%$
$z \sim 0$ ($N = 115$)	All	**0.87** $\pm 0.87\%$	1.74 $\pm 1.22\%$	97.39 $\pm 1.49\%$
	E/S0	0.00 $\pm 0.00\%$	1.7 $\pm 1.22\%$	72.2 $\pm 4.18\%$
	Sabc	0.00 $\pm 0.00\%$	0.00 $\pm 0.00\%$	20.0 $\pm 3.73\%$
	Sd/Irr	0.87 $\pm 3.73\%$	0.00 $\pm 0.00\%$	5.22 $\pm 2.07\%$

Note. — Rows 1-12: For a given redshift (Column 1), morphology (Column 2), and stellar mass range, Columns 3, 4, and 5 list the fraction of galaxies in three separate bins of Sérsic index n. Rows 13-24: Same as the above except that Columns 3, 4, and 5 reflect bins of half-light radius r_e.

TABLE 2
FIT OF MASSIVE GALAXIES TO $r_e/r_{e,z\sim0} = \alpha(1+z)^\beta$ OVER $z = 0 - 3$

Sample (1)	$\alpha(\pm1\sigma)$ (2)	$\beta(\pm1\sigma)$ (3)
All n	1.15(0.30)	-1.30(0.24)
$n \leq 2$	1.11(0.32)	-1.30(0.29)
$n > 2$	1.20(0.31)	-1.52(0.26)
Non-AGN hosts with high SFR_{IR}[1]	1.15(0.33)	-1.22(0.30)
Non-AGN hosts with low SFR_{IR}[2]	1.67(0.33)	-1.67(0.28)

Note. — [1] Non-AGN hosts with 24 μm flux above the *Spitzer* 5σ limit (30 μJy). [2] Non-AGN hosts with 24 μm flux below the *Spitzer* 5σ limit (30 μJy).

TABLE 3
FRACTION OF MASSIVE ($M_\star \geq 5 \times 10^{10}\ M_\odot$) GALAXIES WITH 24 μM DETECTIONS

z	SFR$_{min}$ ($M_\odot$ yr^{-1})	Fraction with $f_{24\mu m} \geq 30\mu$Jy (%)	Mean SFR (AGN + non-AGN) ($M_\odot$ yr^{-1})	Mean SFR (Non-AGN) ($M_\odot$ yr^{-1})
(1)	(2)	(3)	(4)	(5)
$z = 1 - 1.5$	4.29	43.6 $\pm$7.9%	63.8 $\pm$ 12.9	44.0 $\pm$7.3
$z = 1.5 - 2$	12.4	48.0 $\pm$7.1%	222.8 $\pm$ 58.5	128.9 $\pm$37.6
$z = 2 - 3$	47.2	42.9 $\pm$5.6%	1145.6 $\pm$ 274.5	418.8 $\pm$142.9

Note. — Column 2 estimates the detection limit on SFR given the 5σ limit on $f_{24\mu m}$ of 30 μJy. The expected SFR$_{IR}$ at 30 μJy is determined by linear regression of the distribution of $f_{24\mu m}$ versus SFR$_{IR}$ in each redshift bin. Column 3 lists the percentage of massive GNS galaxies with $f_{24\mu m} > 30\mu$Jy. Column 4 shows the mean SFR among all galaxies having $f_{24\mu m} > 30\ \mu$Jy. Column 5 shows the mean SFR among all galaxies without any evidence for AGN activity (see §6.1). The error bars in Column 4 and Column 5 represent the standard error on the mean.

Table 4

Summary of AGN Detection and Properties

z	Total Number	X-ray AGN	PLG	IR Excess AGN	AGN Fraction	Median n	Median r_e (kpc)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$z = 1 - 1.5$	7	7	0	0	17.9 ±6.1%	2.12	4.48
$z = 1.5 - 2$	11	6	0	5	22.0 ±5.9%	1.85	3.73
$z = 2 - 3$	31	20	3	11	40.3 ±8.8%	1.42	2.83

TABLE 5

SUMMARY OF KOLMOGOROV-SMIRNOV TEST ON ELLIPTICITY

Sample 1	Sample 2	Probability (%)	KS Test D
(1)	(2)	(3)	(4)
$n \leq 2\ z = 2 - 3$	MGC E/S0	0	0.489
$n \leq 2\ z = 2 - 3$	MGC Spiral (Sabc + Sd/Irr)	4.78	0.221
$n \leq 2\ z = 2 - 3$	MGC Sd/Irr	23.5	0.317
$n > 2\ z = 2 - 3$	MGC E/S0	34.3	0.184
$n > 2\ z = 2 - 3$	MGC Spiral (Sabc + Sd/Irr)	14.0	0.237
$n > 2\ z = 2 - 3$	MGC Sd/Irr	15.8	0.370

Note. — Columns 1 and 2 list the two samples for which ellipticity was compared in each KS test. Column 3 lists the probability that Sample 1 and Sample 2 are drawn from the same distribution. Column 4 lists the Kolmogorov-Smirnov statistic specifying the maximum separation between the cumulative ellipticity distribution functions for Sample 1 and Sample 2.

Appendices

Knowledge of the PSF is important to assess data quality and for deriving structural parameters. NIC3 is out of focus, so the PSF can deviate from the theoretically expected one. PSF convolution with GALFIT is commonly performed with a user-provided bright, unsaturated star. Not all of the GNS tiles contain suitably bright, unsaturated stars. It is not advisable to adopt a set of PSF stars from a subset of pointings because the NIC3 PSF depends on position within the NIC3 field and is also subject to interpolation artifacts introduced by `drizzle` that are dependent on the adopted dither pattern (J. Krist, private communication).

As a result, the best available option for handling PSF convolution is to make synthetic NIC3 PSFs with Tiny Tim (Krist 1995). For each galaxy, Tiny Tim PSFs were generated for all the galaxy's positions in the individual, undrizzled exposures. Telescope breathing was accounted for with each PSF by refining the Tiny Tim parameters to match the Pupil Alignment Mechanism (PAM) value recorded in the headers of the undrizzled frames. Blank, zero-valued frames retaining the WCS information of the undrizzled frames were made. The synthetic PSFs were inserted into the blank frames precisely where each galaxy would be in the individual frames. The blank frames were drizzled together in the same way as the data with a pixfrac of 0.7 and a final output platescale of $0\rlap{.}{''}1$/pixel. This process was repeated for all 166 massive ($M_\star \geq 5 \times 10^{10}$) galaxies in our sample.

This approach accounts both for variation in PSF with position on the NIC3 field and for the dependence on the `drizzle` algorithm. The range of FWHM in the final drizzled synthetic PSFs is $\sim 0\rlap{.}{''}26 - 0\rlap{.}{''}36$,[17] with a mean value of $0\rlap{.}{''}3$. The mean PSF diameter of the science images ($0\rlap{.}{''}3$) is 2.5 kpc at $z = 2$, under the adopted cosmology.

B.1. *Tests on Robustness of Fits and Parameter Coupling*

How robust are the results that a dominant fraction of the massive galaxies at $z = 2 - 3$ have a low $n \leq 2$ and that a large fraction are ultra-compact? In particular, how non-degenerate are the fits? Could some of the galaxies with an $n \leq 2$ have similarly good fits with higher n?

First, one should note that the errors quoted by GALFIT on the structural parameters cannot be used to assess the robustness of the fits because the errors quoted by GALFIT underestimate the true parameter errors (Häussler et al. 2007), which are dominated by the systematics of galaxy structure, and in particular, by parameter coupling and degeneracy.

The task of assessing the coupling between model parameters is complicated when models have a large number of free parameters. The single Sérsic profiles fit to the NICMOS galaxy images have 7 free parameters (centroid, luminosity, r_e, n, axis ratio, and position angle). While GALFIT selects a best-fit by minimizing χ^2 for a given set of input guesses, it is not clear whether the minimized χ^2 is an absolute minimum or local minimum. Investigating the χ^2 values for all combinations of fit parameters over the full multi-dimensional parameter space is prohibitively time consuming and computationally expensive. Instead, we will adopt a simpler approach of focusing on strong coupling between r_e and n, and exploring how χ^2 varies as these parameters are moved away from the initial solution picked by GALFIT.

One important point should be noted when using changes in χ^2, or $\Delta\chi^2$, for fits to different models. When errors are normally distributed, the multi-dimensional ellipsoids for a given $\Delta\chi^2$ contour can be associated with a statistical confidence level (e.g., $\Delta\chi^2 \sim 1$ corresponds to a 68% confidence level). However, since the errors in the GALFIT models are not normally distributed, but are instead dominated by the systematics of galaxy structure, this means that we cannot a priori assign a confidence level to a given $\Delta\chi^2$. As outlined in the test below, we can still use the shape of $\Delta\chi^2$ as a function of n or r_e as a guide to the quality of fit in the sense that sharp rises in χ^2 as n is varied away from the best-fit value are taken as indicative of poorer fits. But, we cannot a priori say how much poorer the fits are in a statistical sense. This is a well-known and hard problem in structural fitting. We will return to this point in section §B.2.

We carry out the test below for all galaxies in our sample We denote as $\chi^2_{\min,0}$, the value of χ^2 obtained when GALFIT fits the galaxy with n and r_e as free parameters. The associated best-fit parameters are $n_{\min,0}$ and $r_{e,\min,0}$. We then fit single Sérsic profiles with n fixed at discrete values (0.5-10), while allowing all other parameters to freely vary. The initial inputs to these fits were the same as those used to generate the model in which n is a free parameter. We let GALFIT find the best-fit for each of these fixed n models by minimizing χ^2, and we record for each such best-fit the following quantities: the fixed value of n, the best-fit value of r_e, and the associated minimum in χ^2 called $\chi^2_{\min}$. We then evaluate how the difference $\chi^2_{\min}$ - $\chi^2_{\min,0}$ varies as a function of r_e and n, as we move to values away from $n_{\min,0}$ and $r_{e,\min,0}$.

The test was carried out for all galaxies. Figure B1 shows the results of the test for four representative galaxies with $n \sim 1 - 4$. The first column of Figure B1 shows how ($\chi^2_{\min}$-$\chi^2_{\min,0}$) changes when n is varied away from $n_{\min,0}$ at discrete values (0.5-10) and GALFIT is allowed to vary all other parameters to get a best-fit that yields $\chi^2_{\min}$. The second column shows the corresponding best-fit r_e for that $\chi^2_{\min}$. Red stars in the plots denote $n_{\min,0}$ and $r_{e,\min,0}$, which are associated with $\chi^2_{\min,0}$. The shape of $\chi^2_{\min}$ - $\chi^2_{\min,0}$ is asymmetric for n and r_e. The coupling between n and r_e means ($\chi^2_{\min}$ - $\chi^2_{\min,0}$) varies in a similar way with both n and r_e.

We can see that in Figure B1, the absolute minimum χ^2 values occur at the $n_{\min,0}$ and $r_{e,\min,0}$ values, which GALFIT picked when it was allowed to freely fit the galaxies without fixing n. Shifting n away from $n_{\min,0}$ (denoted by the red stars) by ± 1 can increase $\chi^2_{\min}$ by several 10s or 100s of χ^2 units. While only 4 representative galaxies are shown in Figure

---

[17] The range in PSF FWHM comes from differing positions in the NIC3 field and the PAM values used to create the synthetic PSFs.

B1, we show results for the whole sample in Figure B2. This figure illustrates that the distribution of $(\chi^2_{\min} - \chi^2_{\min,0})$ for $n_{\min,0} - 1$ (top panel) and $n_{\min,0} + 1$ (bottom panel), and demonstrates for the whole sample, $\chi^2_{\min}$ generally changes substantially when n is shifted away from $n_{\min,0}$. We draw two primary conclusions from Figure B1:

1. For galaxies with $n_{\min,0} > 2$ (rows 3 and 4), $\chi^2_{\min}$ - $\chi^2_{\min,0}$ rises sharply at lower $n < n_{\min,0}$, suggesting that lower n values are unlikely to yield a good fit for such systems. At $n > n_{\min,0}$, $\chi^2_{\min}$ - $\chi^2_{\min,0}$ rises less sharply, but the rise is still substantial as demonstrated in by the high-magnification inset plots in rows 3 and 4 of column 1.

2. The most important point to take from Figure B1 is that for galaxies with $n_{\min,0} < 2$ (as in rows 1 and 2), $\chi^2_{\min}$ - $\chi^2_{\min,0}$ rises rapidly at higher $n > n_{\min,0}$, thereby making it unlikely that a higher $n > 2$ would provide a similarly good fit. *Thus, we have a great degree of confidence that we are not highly overestimating the number of $n \leq 2$ galaxies in the sample.*

B.2. *Recovery of Parameters From Simulated Images*

Section §B.1 tests how well parameters are recovered in real galaxies, but we cannot a priori assign a confidence level to a given $\Delta\chi^2$ because the errors in the GALFIT models are not normally distributed. However, we can run an extra complementary test where we use simulated *idealized* galaxies whose (n, r_e) are a priori known. The drawback of using idealized galaxies as opposed to the real galaxies fitted in §3.1 is that the former lack the complexity of real galaxies, since they are simply generated from GALFIT models and exactly described by a functional form, such as a Sersic model with a specified (n, r_e). However, the advantage is that we do know the (n, r_e) values a priori and can therefore compare these values to those obtained once these idealized galaxies are inserted into frames with noise properties corresponding to the NIC3 GNS images of our sample galaxies at $z = 1 - 3$.

This test is performed by simulating 1000 galaxy images, each with a unique set of Sérsic parameters: surface brightness at the effective radius μ_e, effective radius r_e, Sérsic index n, axis ratio b/a, and position angle PA. The parameters are chosen randomly from uniform distributions spanning the parameter space of the observed galaxies. The ranges in μ_e, r_e, n, b/a, and PA are 16 to 32 mag/arcsec^2, $0''.05$ to $1''.0$, 0.5 to 10, 0.3 to 1.0, and -90 to 90 degrees, respectively. The chosen range in input μ_e mimics the effect of surface brightness dimming, and the range in r_e ensures the simulated objects span the angular size of the real GNS galaxies. The simulated galaxies were created with GALFIT and convolved with a drizzled PSF image. They were set within a sky background equivalent to the mean NIC3 sky background within GNS (0.1 counts/sec). Source noise, sky noise, and read noise (29 e^-) were added to the frames.

The simulated images were then re-fit with GALFIT to derive (n, r_e). Initial guess parameters for $(\mu_e, r_e, n, b/a, PA)$ were generated randomly from uniform distributions spanning ± 1.5 mag/arcsec^2 in μ_e, $\pm 0''.3$ in r_e, ± 2 indices in n, 0.3 to 1 in b/a, and -90 to 90 degrees in PA. Figure B3 shows the recovery in (n, r_e) plotted against surface brightness. The dashed vertical lines represent the minimum, median, and maximum μ_e for the observed massive galaxies. Figure B3 shows (n, r_e) are well recovered across the full range in observed μ_e. The recovery as a function of μ_e severely degrades only at several mag/arcsec^2 fainter than observed μ_e. In $\sim 95\%$ of cases, n and r_e are recovered to within 10% of their input values for the range of observed μ_e among the massive galaxies in our sample.

B.3. *Tests on MGC Fits*

The structural parameters for the massive galaxies at $z \sim 0$, are derived by Allen et al. (2006) by using the GIM2D code (Simard et al. 2002) to fit single Sérsic component to the MGC B-band images. We derived the structural parameters for the massive galaxies at $z = 1 - 3$, by using the GALFIT code (Peng et al. 2002) on the NIC3/F610W images (§3.1). One might wonder whether the dramatic shift in Figure 5 of the $z = 2 - 3$ galaxies toward lower (n, r_e) compared to the the $z \sim 0$ MGC galaxies may be caused by systematic differences between the fitting techniques used by us versus those by Allen et al. (2006). This would be the case only if the fits by Allen et al. (2006) give systematically higher (n, r_e) than ours for the same galaxies. As we show below this is not the case.

In order to address this issue, we have applied GALFIT to a subset of B-band MGC images and compared our resulting structural parameters to the GIM2D-based results given in the MGC catalogue. The comparison (top row of Figure B4) shows that the GIM2D-based fits of Allen et al. (2006) are not biased to higher (n, r_e) compared to our GALFIT-based fits for the $z \sim 0$ MGC galaxies. In fact, for large r_e, the GIM2D-based values may even be lower. in many cases.

These results are consistent with extensive comparisons of single component Sérsic fits from GALFIT and GIM2D conducted by Häussler et al.(2007) on both simulated and real galaxy data. They concluded that both codes provide reliable fits with little systematic error for galaxies with effective surface brightnesses brighter than that of the sky, as long as one is not dealing with highly crowded fields.

Another possible source of difference between the structural parameters of the $z \sim 0$ and $z = 2 - 3$ massive galaxies might be the fact that Allen et al. (2006) fitted the $z \sim 0$ massive galaxies with only a single Sérsic component, and did not include an extra point source component in galaxies with evident nuclear sources. It seems unlikely that the much larger fraction of higher (n, r_e) systems at $z \sim 0$ in Figure 5 is mainly driven by this effect. To illustrate this, we have fitted the $z \sim 0$ MGC galaxies in the top row of Figure B4 with a combination of a single Sérsic component and a point source model using GALFIT. The bottom row of Figure B4 shows the results. The values of r_e are not changed systematically. The Sérsic index is lowered by the addition of the point source, but only 20% of the sources with $n > 2$ in the single Sérsic fit have $n \leq 2$ after including the point source. Since not all $z \sim 0$ MGC galaxies in Figure 5 will have

nuclear sources, the fraction of sources impacted will be even less. We thus conclude that the presence of a point source in some of the $z \sim 0$ MGC galaxies and the inclusion of such a point source in the model fits, are not sufficient to shift the $z \sim 0$ MGC galaxies into the parameter space occupied by the $z = 2 - 3$ massive galaxies in Figure 5.
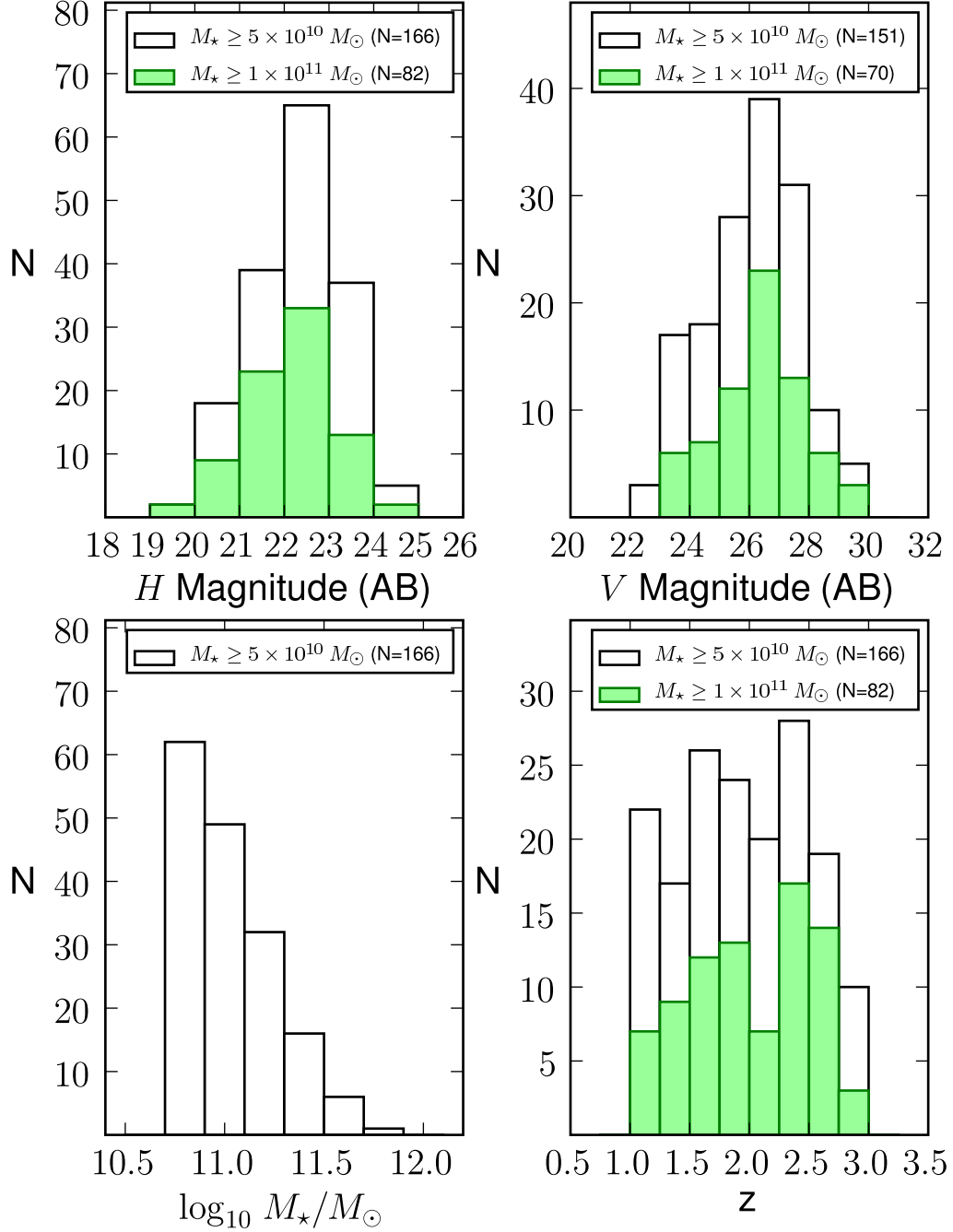
FIG. 1.— The distribution of apparent H (F160W), V apparent magnitude, stellar mass, and redshift for the final, complete sample of 166 galaxies with $M_\star \geq 5 \times 10^{10}$ $M_\odot$ and redshift $z = 1 - 3$.
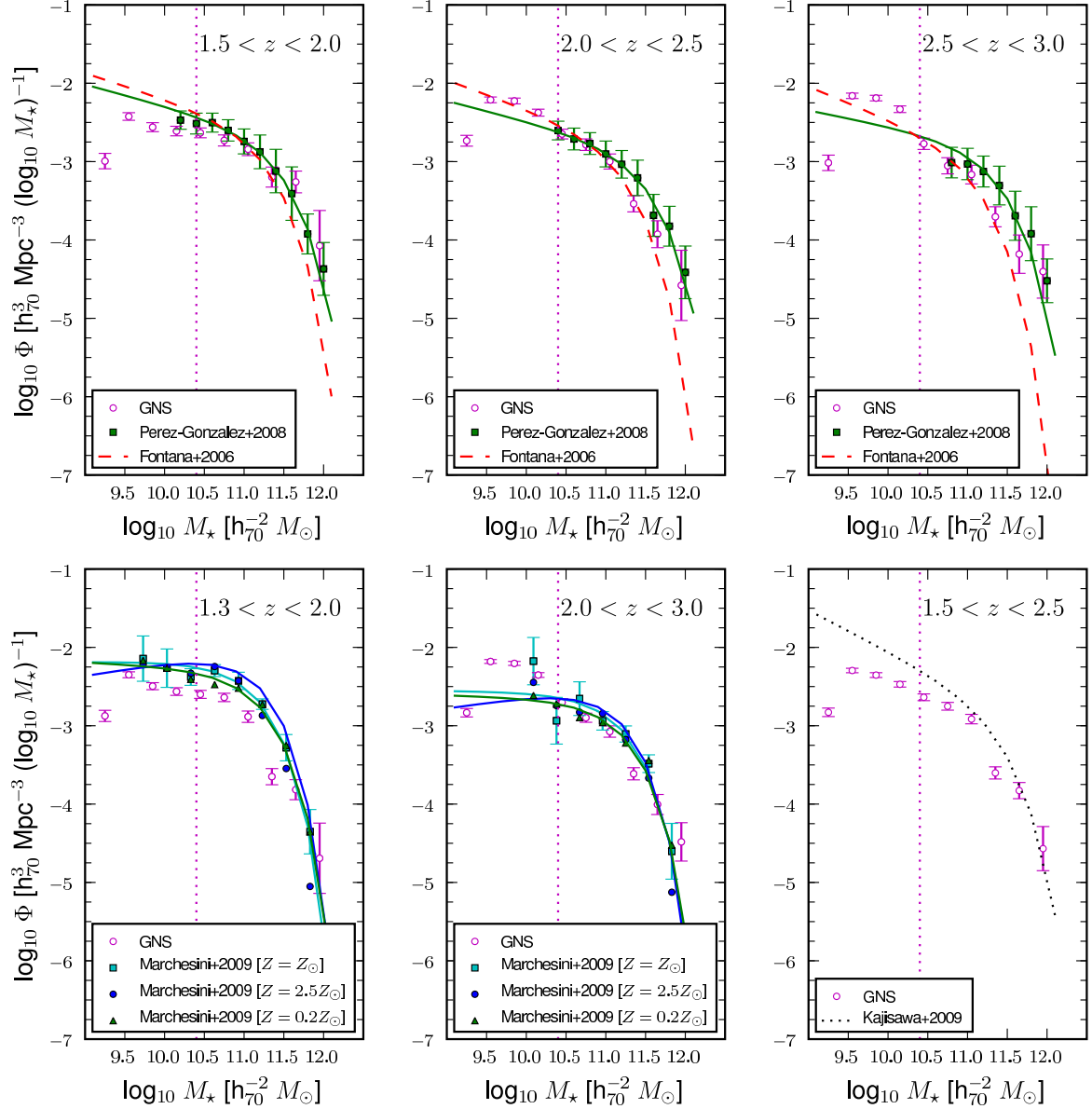
FIG. 2.— We compare the galaxy stellar mass functions for GNS over $z = 1 - 3$ with those from other studies that are based on K or IRAC-selected samples (Kajisawa et al. 2009; Marchesini et al. 2009; Pérez-González et al. 2008; Fontana et al. 2006). The vertical line in each plot marks the mass cut ($M_\star \geq 5 \times 10^{10}\ M_\odot$) for the GNS-based sample used in this paper. We include the data points with error bars from the other studies, where available, along with each Schechter function fit. Some studies (Kajisawa et al. 2009; Marchesini et al. 2009) present results for multiple sets of SED-modeling assumptions, and in these cases we show the results for the assumptions that most closely match those used for GNS by Conselice et al. (2011). For Kajisawa et al. (2009), we show the mass function calculated with Bruzual & Charlot (2003) stellar templates. For Marchesini et al. (2009), we show the stellar mass functions calculated with Bruzual & Charlot 2003 templates, metallicities of 0.2, 1, and 2.5 $Z_\odot$, a Kroupa IMF, and a Calzetti extinction law, but in the above plot, we scale their mass functions by +0.2 dex along the x-axis to convert their Kroupa IMF to a Salpeter IMF. For the GNS mass functions, in comparison, the best metallicity is determined on a galaxy-by-galaxy basis from a set of discrete values spanning 0.005 to 2.5 $Z_\odot$. The error bars for Marchesini et al. (2009) take into account the uncertainties due to cosmic variance, Poisson error, photometric redshifts, and stellar SED templates. The error bars from Pérez-González et al. (2008) account for Poisson error and uncertainty in photometric redshifts. In comparison, the error bars on the GNS mass functions show only Poisson error.
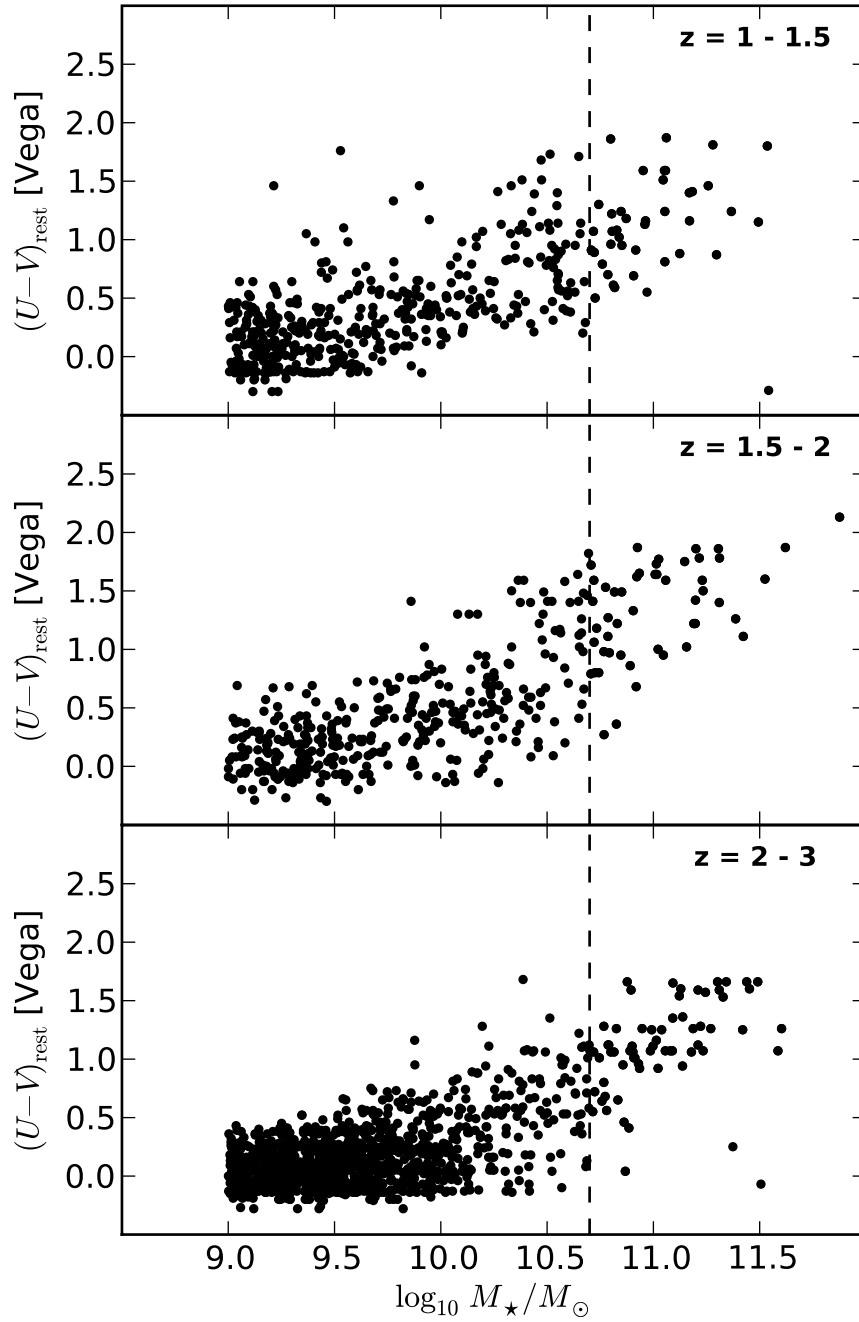
FIG. 3.— For all galaxies detected in the GOODS-NICMOS Survey (GNS) over $z = 1 - 3$, the rest-frame $U - V$ color is plotted against $M_\star$ for different redshift bins. Blue systems are preferentially at low masses, while the most massive ($M_\star \geq 1 \times 10^{11}$ $M_\odot$) galaxies are preferentially red. The vertical line denotes $M_\star = 5 \times 10^{10}$ $M_\odot$, the mass cut we adopt for our final sample of 166 galaxies.

$n > 2$ **Systems**

$r_e \leq 2$ kpc

$2 < r_e \leq 4$ kpc

$n \leq 2$ **Systems**

$r_e \leq 2$ kpc

$2 < r_e \leq 4$ kpc

$4 < r_e \leq 8$ kpc

Irregular/Disturbed Systems

$2.7 \leq r_e \leq 3.6$ kpc
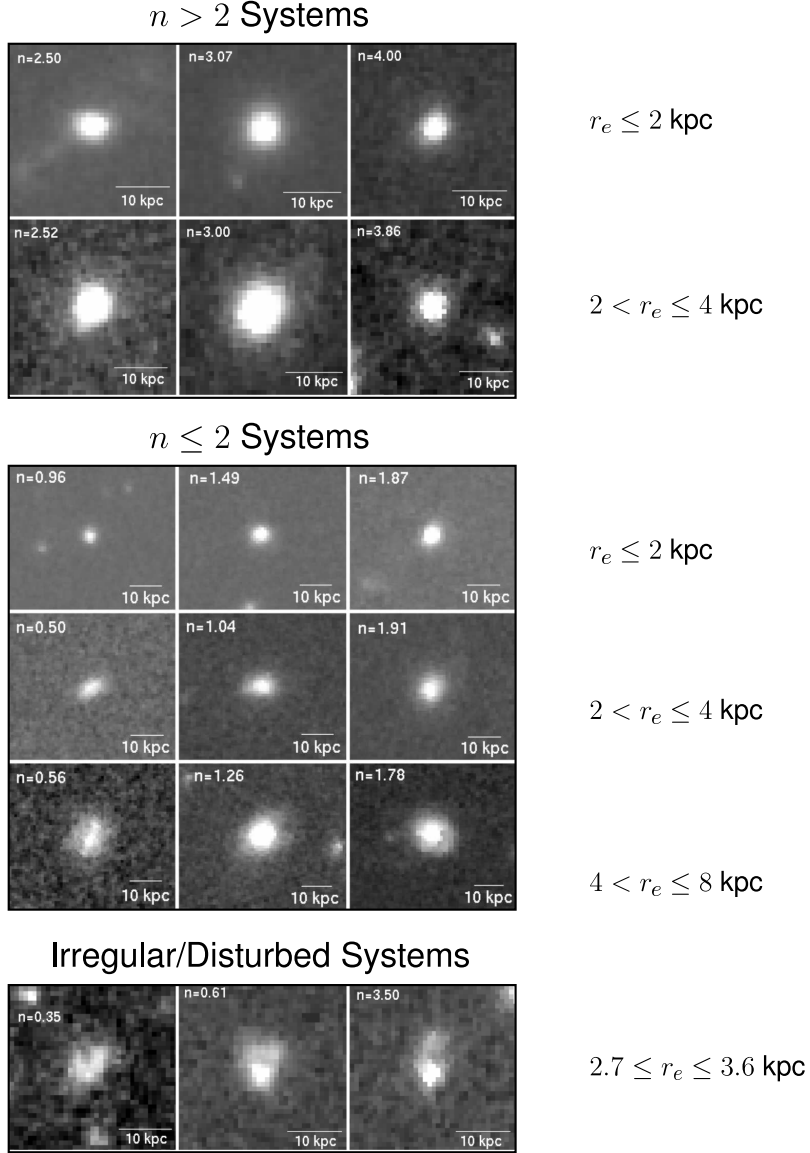
FIG. 4.— NICMOS F160W images for representative GNS galaxies with $M_\star \geq 5 \times 10^{10}$ $M_\odot$ at $z = 2 - 3$. The Sérsic index n and effective radius r_e referenced here are based on fitting single Sérsic components to the NICMOS images, as described in §3. The top panel shows example systems with Sérsic index $n > 2$ and half-light radii $r_e \leq 4$ kpc. The middle panel shows examples with $n \leq 2$ and $r_e \leq 8$ kpc. The majority ($\sim 82\%$; Table 1) of the massive GNS galaxies have $r_e \leq 4$ kpc. In such systems, structural features are generally hard to discern due to resolution effects, so that systems appear fairly featureless (top 4 rows). In the small fraction of massive galaxies at $z = 2 - 3$ with large $r_e > 4$ kpc, one can discern some structural features such as an elongated bar-like feature or a combination of a central condensation surrounded by a more extended lower surface brightness component, reminiscent of a bulge and disk (row 5). The bottom panel (row 6) contains systems that appear morphologically disturbed.
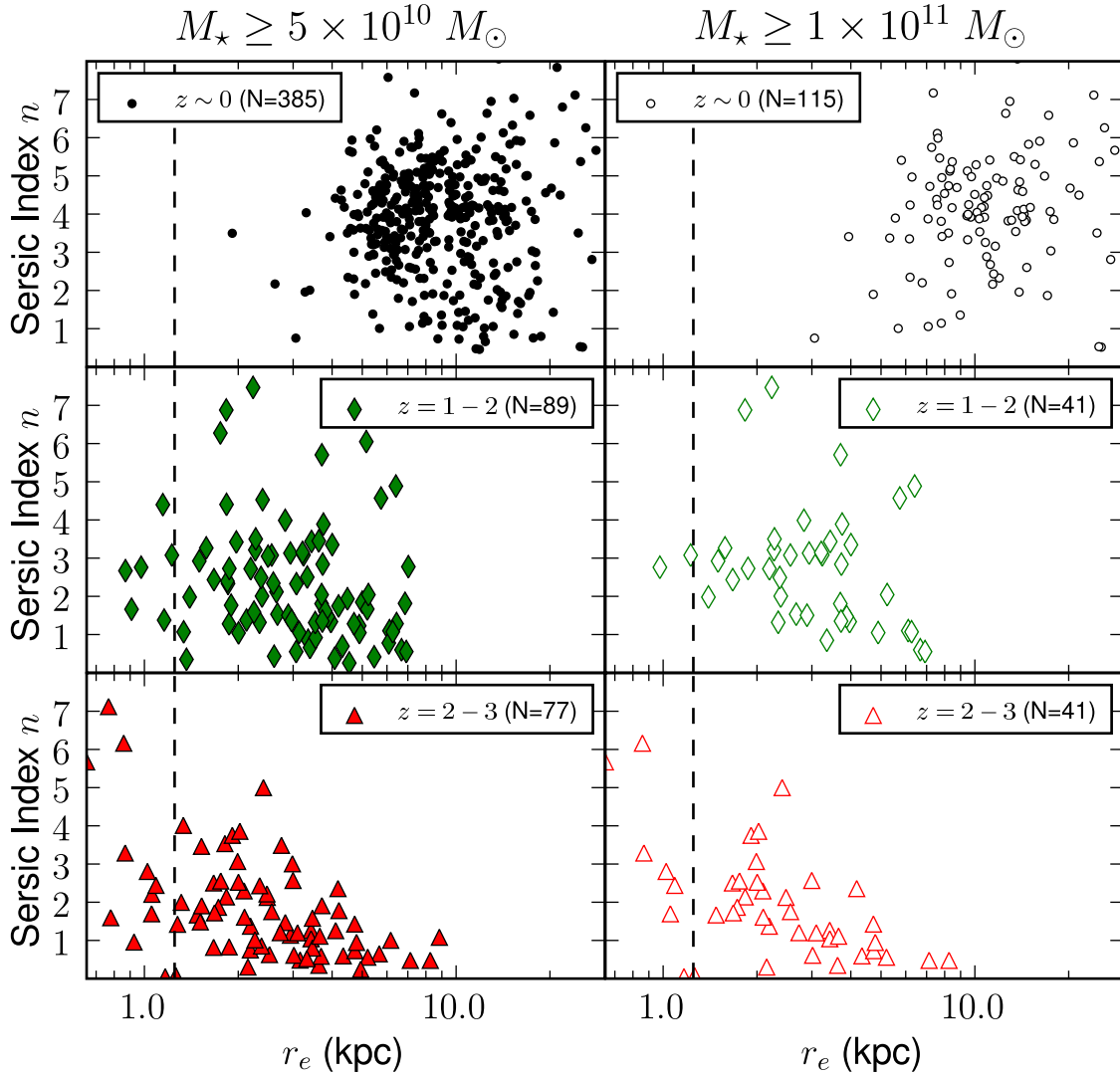
FIG. 5.— The B-band Sérsic index n and effective radius r_e derived from single Sérsic profile fits to massive ($M_\star \geq 5 \times 10^{10}\ M_\odot$) galaxies are plotted for the three redshift bins listed in Table 1. In the top row, the black points represent fits to $z \sim 0$ galaxies by Allen et al. (2006) on B-band images of galaxies from the Millennium Galaxy Catalog (Liske et al. 2003). The lower two rows are based on our fits to the NIC3 F160W images of massive GNS galaxies at $z = 1 - 2$ and $z = 2 - 3$. Note that the massive galaxies at $z = 2 - 3$ are strikingly offset toward lower (n, r_e) compared to the massive $z \sim 0$ galaxies, and have five times more low $n \leq 2$ disky systems (see also Figure 6). The black dashed line represents the typical half-width half maximum of the NICMOS3 PSF at $z = 1 - 3$ of ~ 1.2 kpc.
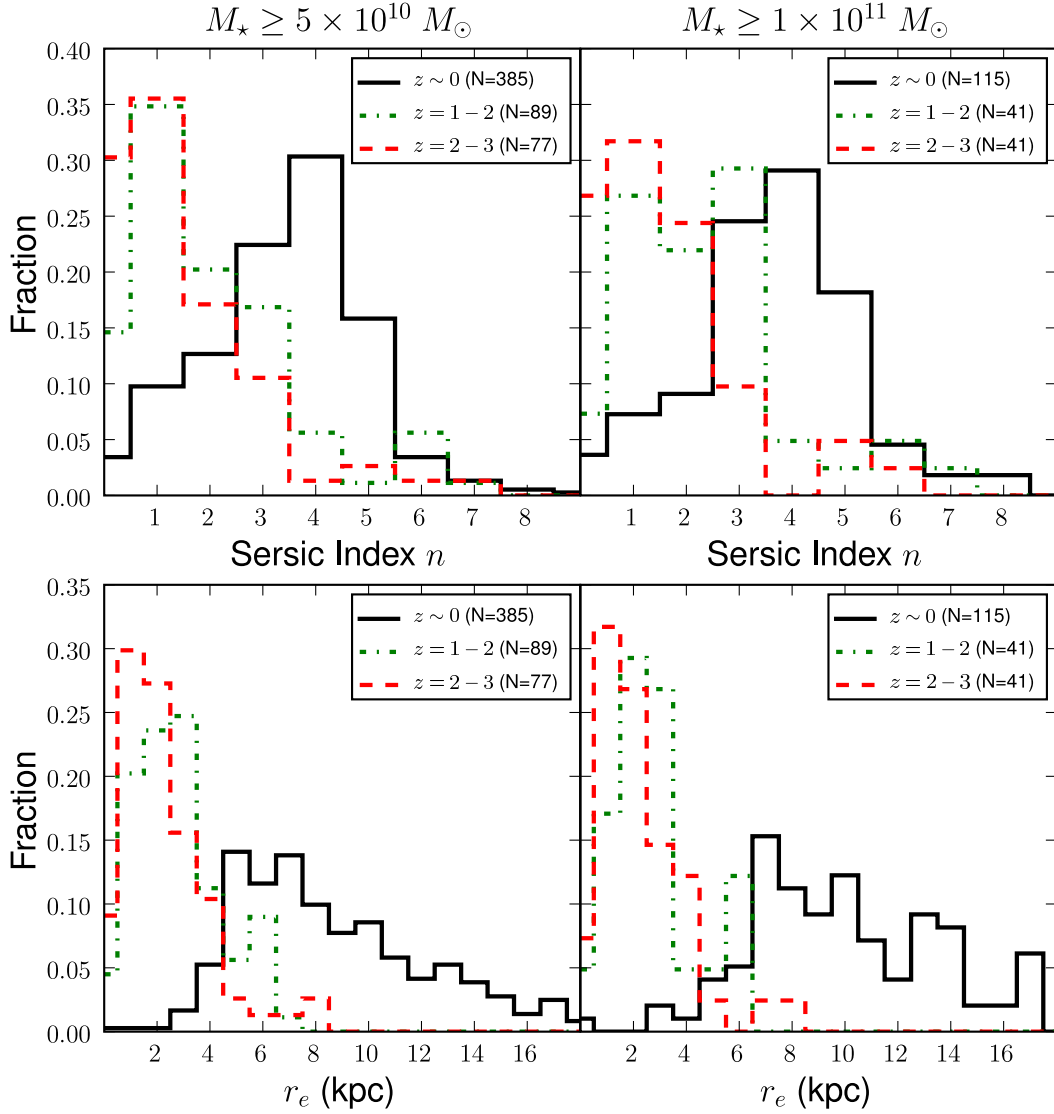
FIG. 6.— Left column: The distributions of rest-frame optical Sérsic index and effective radius r_e based on single Sérsic profile fits to massive ($M_\star \geq 5 \times 10^{10}~M_\odot$) galaxies are plotted for the three redshift bins listed in Table 1: at $z \sim 0$ (solid line), based on the fits of Allen et al. (2006) on B-band images of galaxies from the Millennium Galaxy Catalog (Liske et al. 2003), and at $z = 1 - 2$ (dash-dotted line) and $z = 2 - 3$ (dashed line), based on our fits to the NIC3 F160W images of massive GNS galaxies. Note that a significant fraction ($39.0 \pm 5.56\%$) of massive ($M_\star \geq 5 \times 10^{10}~M_\odot$) galaxies at $z = 2 - 3$ have $r_e \leq 2$ kpc, compared to only $0.52 \pm 0.37\%$ at $z \sim 0$. Note also that most ($64.9 \pm 5.4\%$) of massive galaxies at $z = 2 - 3$ have low $n \leq 2$ (disky) structures compared to only $13.0 \pm 1.7\%$ at $z \sim 0$. Right column: Same as left column but for the mass range $M_\star \geq 1 \times 10^{11}~M_\odot$.
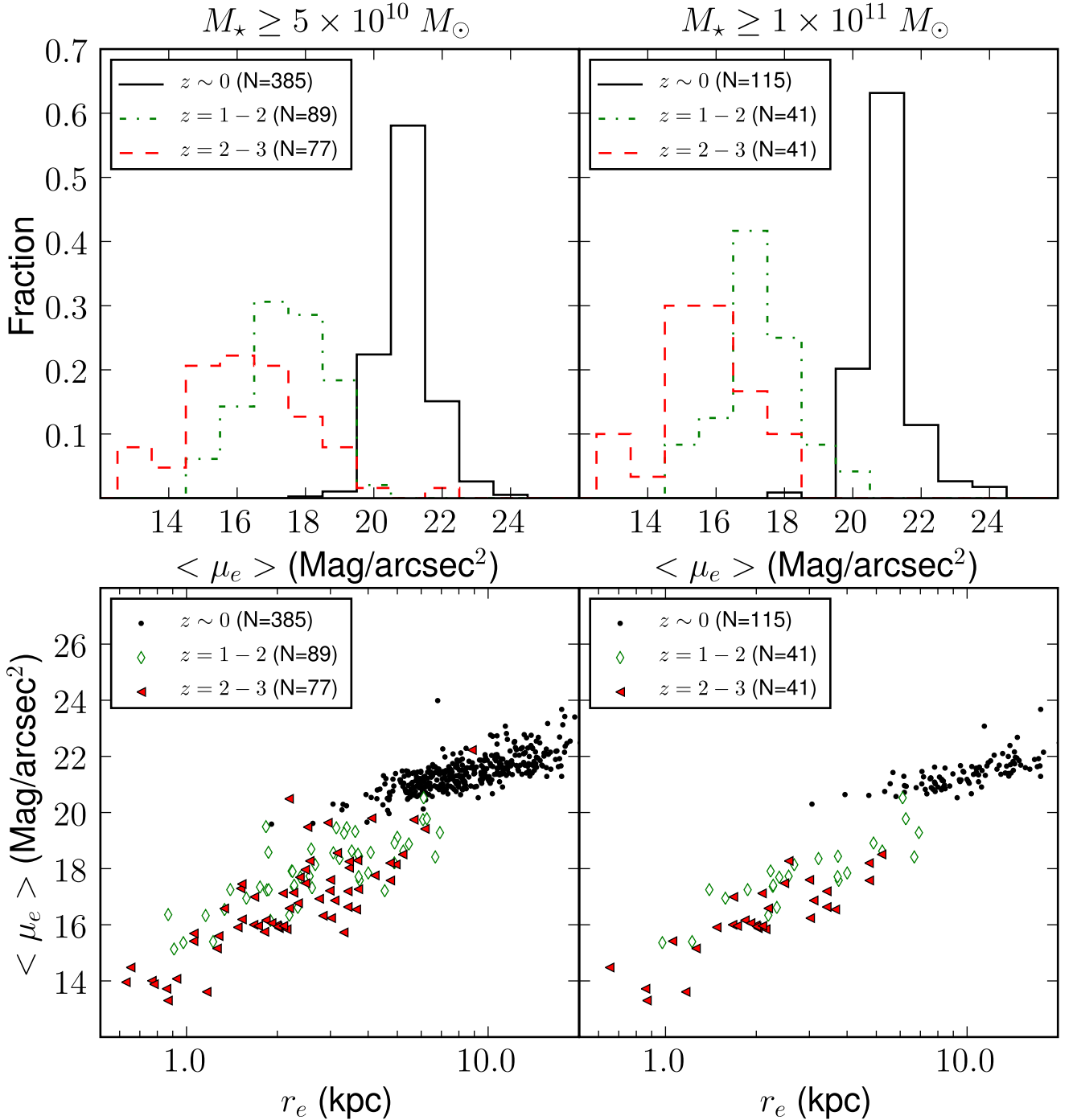
FIG. 7.— Left column: The top panel shows mean extinction-corrected rest-frame B-band surface brightness within the effective radius ($< \mu_e >$) for massive ($M_\star \geq 5 \times 10^{10}\ M_\odot$) galaxies for the three redshift bins listed in Table 1. The solid line is for $z \sim 0$ MGC galaxies. The dash-dotted line ($z = 1 - 2$) and the dashed line ($z = 2 - 3$) are based on our fits to the NIC3 F160W images of massive GNS galaxies. The GNS galaxies at $z = 2 - 3$ have a mean surface brightness of 16.8 mag/arcsec^2 and are systematically brighter than the $z \sim 0$ MGC galaxies, which have a mean surface brightness of 21.3 mag/arcsec^2. In the bottom panel, surface brightness within the effective radius is plotted against effective radius r_e for the same redshift bins. Right column: The same plots are repeated for galaxies with $M_\star \geq 1 \times 10^{11}\ M_\odot$. Surface brightness is calculated with the extinction-corrected rest-frame B-band light and is defined as $< \mu_e > = B_{\rm corr} + 2.5\log_{10}(2\pi r_e^2) - 10\log_{10}(1+z)$, where $B_{\rm corr}$ is the extinction-corrected, rest-frame B apparent magnitude.
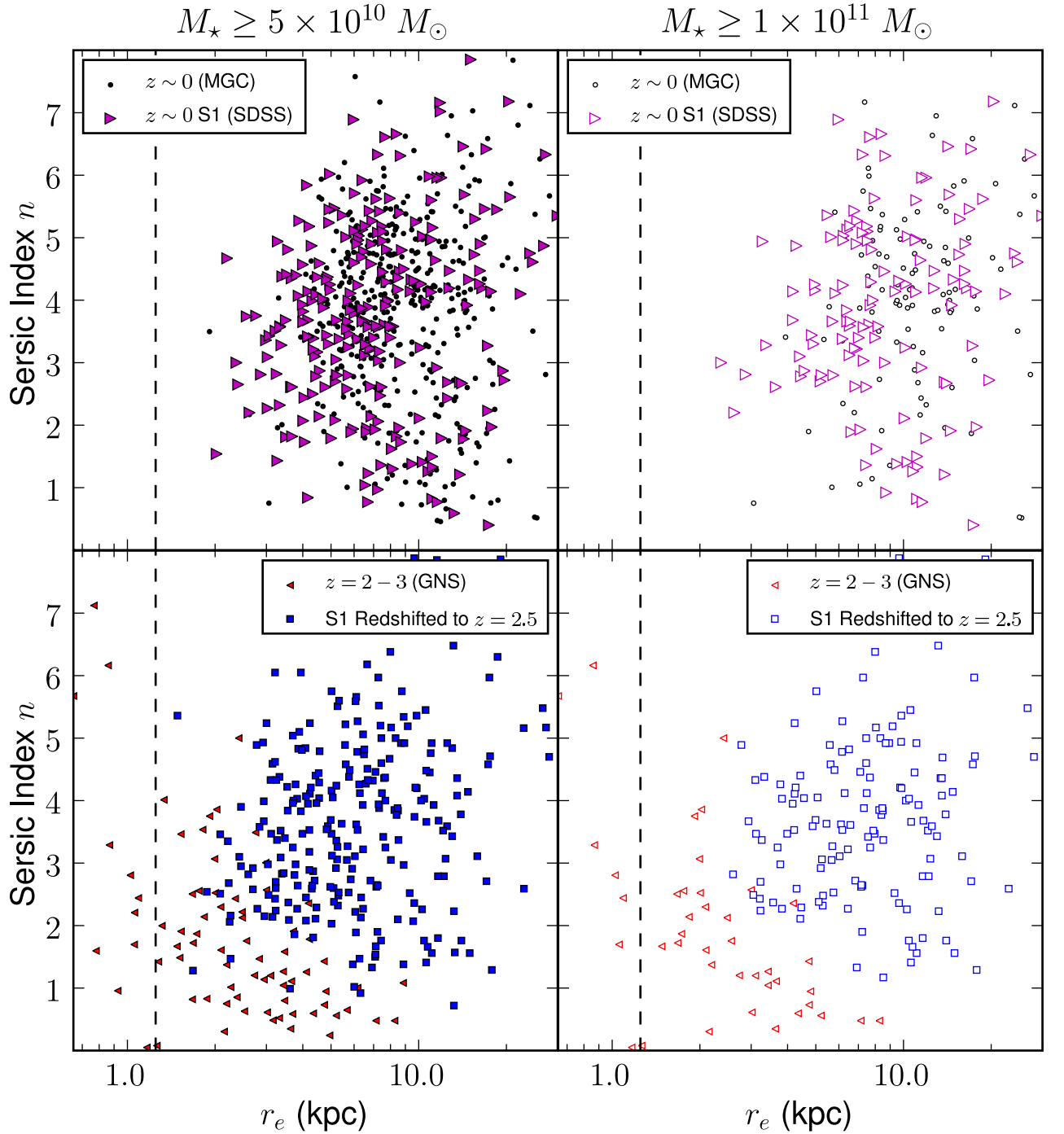
FIG. 8.— Top row: The black points show the massive ($M_\star \geq 5 \times 10^{10}\ M_\odot$) $z \sim 0$ galaxies from MGC described earlier in Figure 5. The magenta points denote the SDSS-based sample S1 of 255 representative massive ($M_\star \geq 5 \times 10^{10}\ M_\odot$) $z \sim 0$ galaxies used in the redshifting experiment. Note the (n, r_e) distribution of S1 covers the same parameter space as that of the MGC sample. This is also shown quantitatively in Figure 9. Row 2: We show as blue squares the (n, r_e) distribution obtained after redshifting S1 to $z = 2.5$ and 're-observing' it with NIC3/F16W as in the GNS survey. We assume a surface brightness evolution of 2.5 magnitudes and brighten each redshifted galaxy by this amount. The actual observed (n, r_e) distributions of the massive galaxies at $z = 2 - 3$ in the GNS survey are significantly offset toward lower values compared to the redshifted galaxies. The black dashed line represents the typical half-width half max of the NICMOS3 PSF at $z = 1 - 3$ of ~ 1.2 kpc.
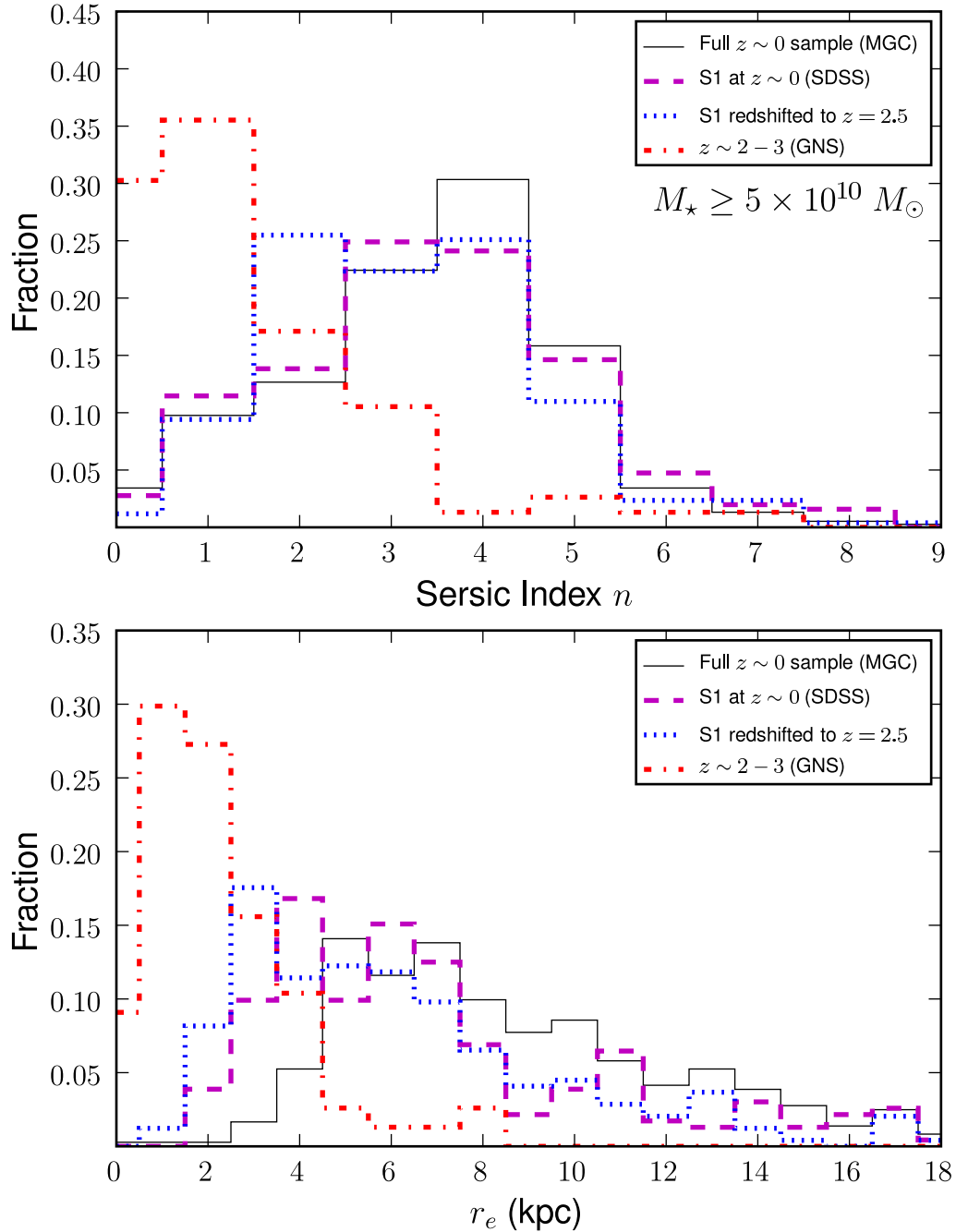
FIG. 9.— This figure illustrates the same information as in Figure 8 but in more quantitative terms. It shows the n and r_e distributions for the full MGC sample of massive $z \sim 0$ galaxies (black line) and the representative sample S1 of 255 galaxies used in the redshifting experiment (magenta line). Sample S1 does a good job of matching the full MGC sample and is typically within $\pm 10\%$ for a given bin. We also contrast the (n, r_e) values after redshfiting S1 to $z = 2.5$ (blue line) with the actual distribution observed in the massive the GNS galaxies at $z = 2 - 3$ (red line). While $64.9 \pm 5.4\%$ and $39.0 \pm 5.6\%$ of the massive $z = 2 - 3$ galaxies have $n \leq 2$ and $r_e \leq 2$ kpc, respectively, the corresponding fractions for the redshifted sample are $10.6 \pm 1.9\%$ and $1.2 \pm 0.7\%$. The results shown here are for galaxies with $M_\star \geq 5 \times 10^{10} \ M_\odot$, but a similar result is obtained for $M_\star \geq 1 \times 10^{11} \ M_\odot$.
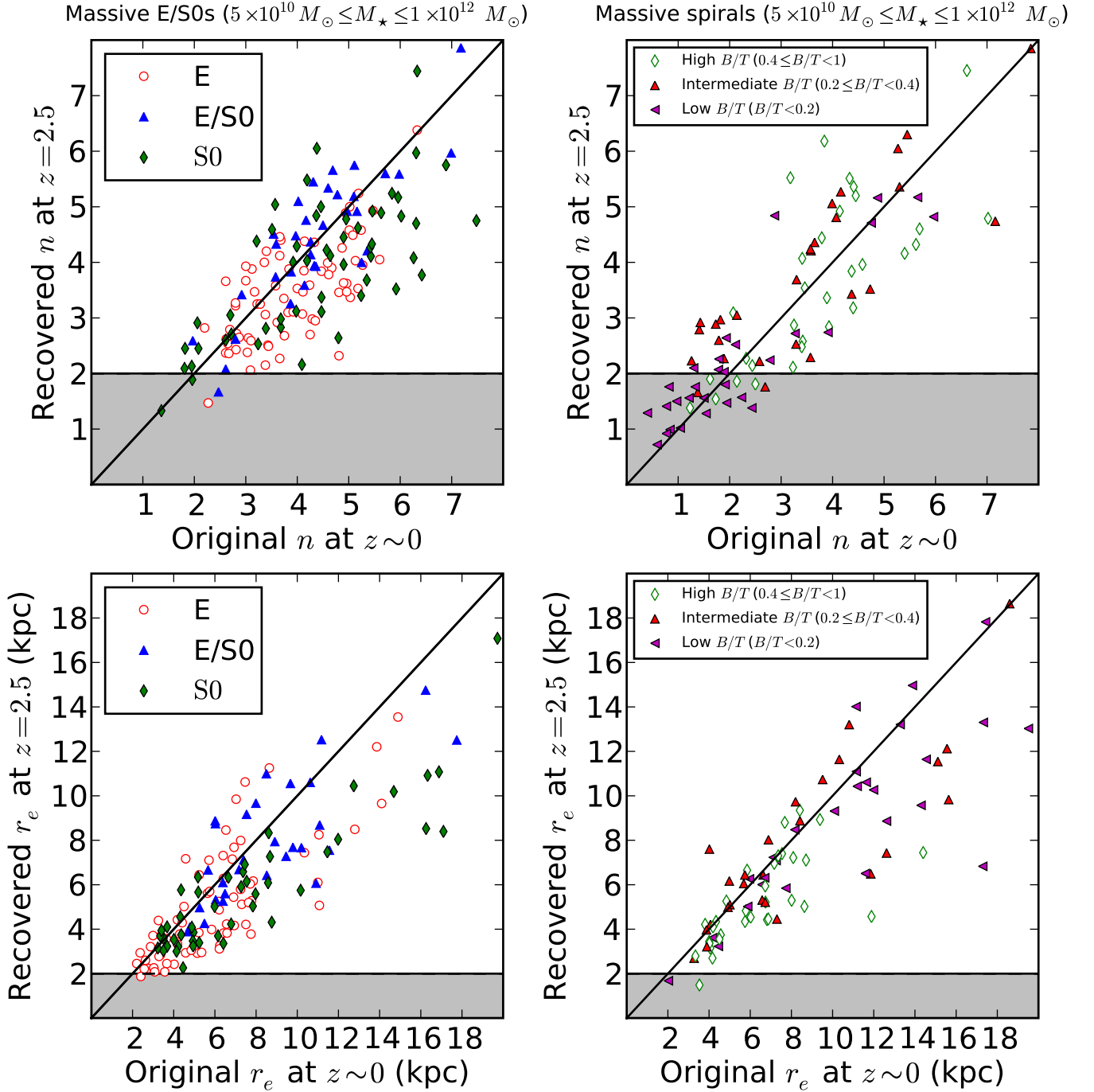
FIG. 10.— Left column: The panels compare the rest-frame optical structural parameters (Sérsic index n and effective radius r_e) of massive ($M_\star \geq 5 \times 10^{10}\ M_\odot$) elliptical and S0 galaxies at $z \sim 0$ to the structural parameters recovered after these galaxies were artificially redshifted to $z = 2.5$, brightened by 2.5 magnitudes in surface brightness, and re-observed with NIC3/F160W. At $z \sim 0$, the structural parameters were measured from g-band images, while at $z = 2.5$ they are measured from the artificially redshifted images in the NIC3/F160W band, so that all parameters are measured in the rest-frame blue optical light. The black lines represent equality, while the shaded area represents the regime of $n \leq 2$ and $r_e \leq 2$ kpc, where $64.9 \pm 5.4\%$ and $39.0 \pm 5.6\%$, respectively, of massive GNS galaxies at $z = 2-3$ lie (Table 1 and Figure 8). *The plots show that the Sérsic index n and effective radius r_e of the massive $z \sim 0$ E and S0s may be lower or higher after redshifting out to $z = 2.5$, but they do not, in general, drop to values as low as $n \leq 2$ and $r_e \leq 2$ kpc, and avoid the shaded area.* Right column: Same as left column, but this time for massive ($M_\star \geq 5 \times 10^{10}\ M_\odot$) $z \sim 0$ spiral galaxies. The galaxies are coded by bulge-to-total light ratio (B/T). B/T was measured with bulge-disk and bulge-disk-bar decomposition of the $z \sim 0$ g-band images. The top plot shows that it is mainly massive $z \sim 0$ late-type spirals of low B/T that yield Sérsic index n as low as $n \leq 2$ after redshifting, and populate the shaded area where $64.9 \pm 5.4\%$ of massive GNS galaxies at $z = 2-3$ lie. However, as shown by this lower plot, the local massive spirals have much larger r_e ($r_e \gg 2$ kpc) and after artificial redshifting avoid the shaded area where $39.0 \pm 5.6\%$ of the massive GNS galaxies at $z = 2-3$ lie.
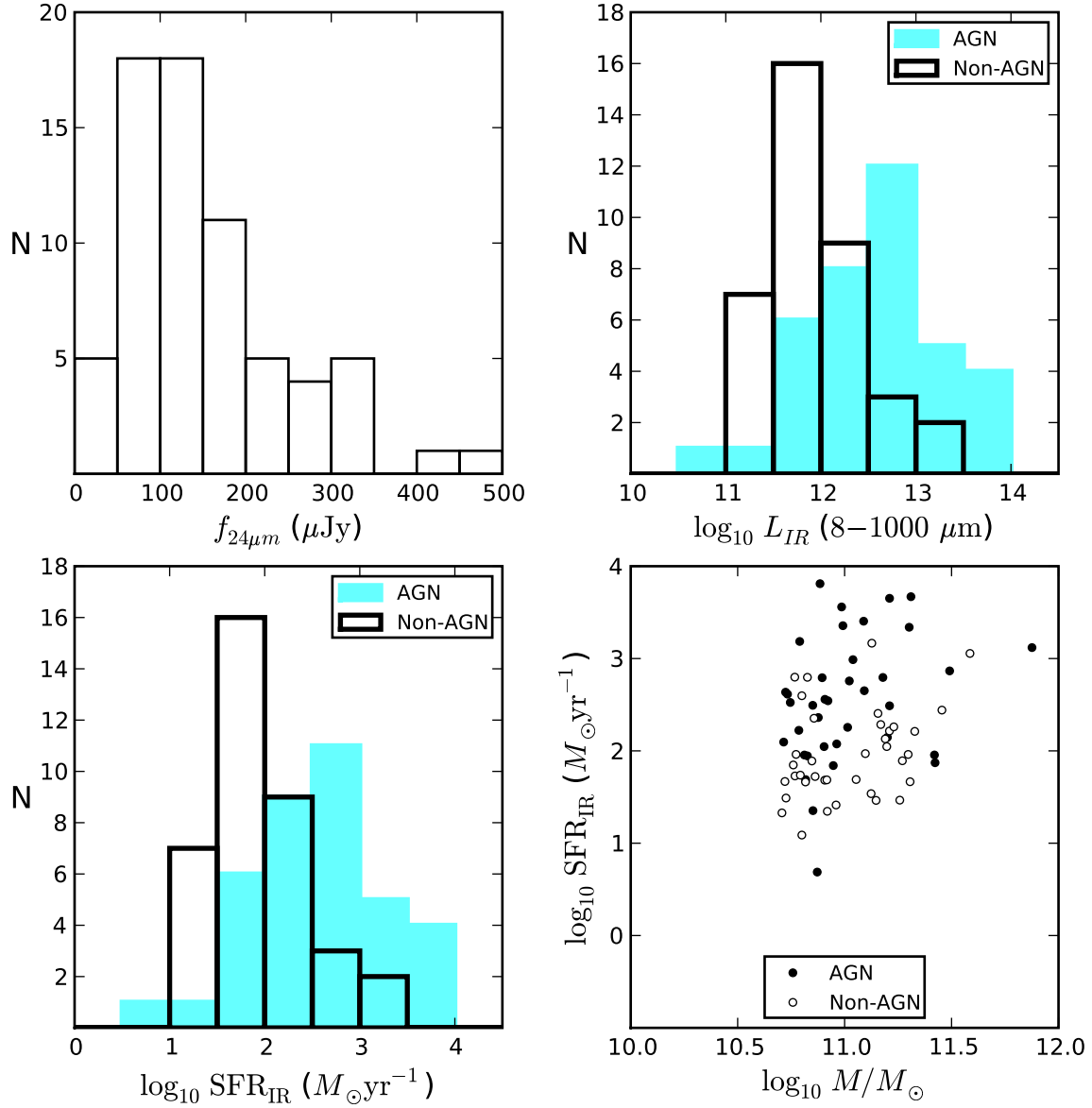
FIG. 11.— Top left: The $f_{24\mu m}$ distribution for the massive ($M_\star \geq 5 \times 10^{10}~M_\odot$) GNS galaxies with reliable MIPS 24 μm counterpart. Upper right: The inferred L_{IR} distribution over 8–1000 μm. Lower left: The inferred SFR$_{IR}$ distribution based on L_{IR}, which is estimated using the Chary & Elbaz (2001) templates, with a correction at L$_{IR}$ > 6 × 10^{11} $L_\odot$. Lower right: SFR$_{IR}$ versus $M_\star$. For sources containing an AGN, the measured L_{IR} and SFR$_{IR}$ are upper limits. The upper right and bottom panels use different coding for sources identified in §6 as hosting an AGN.
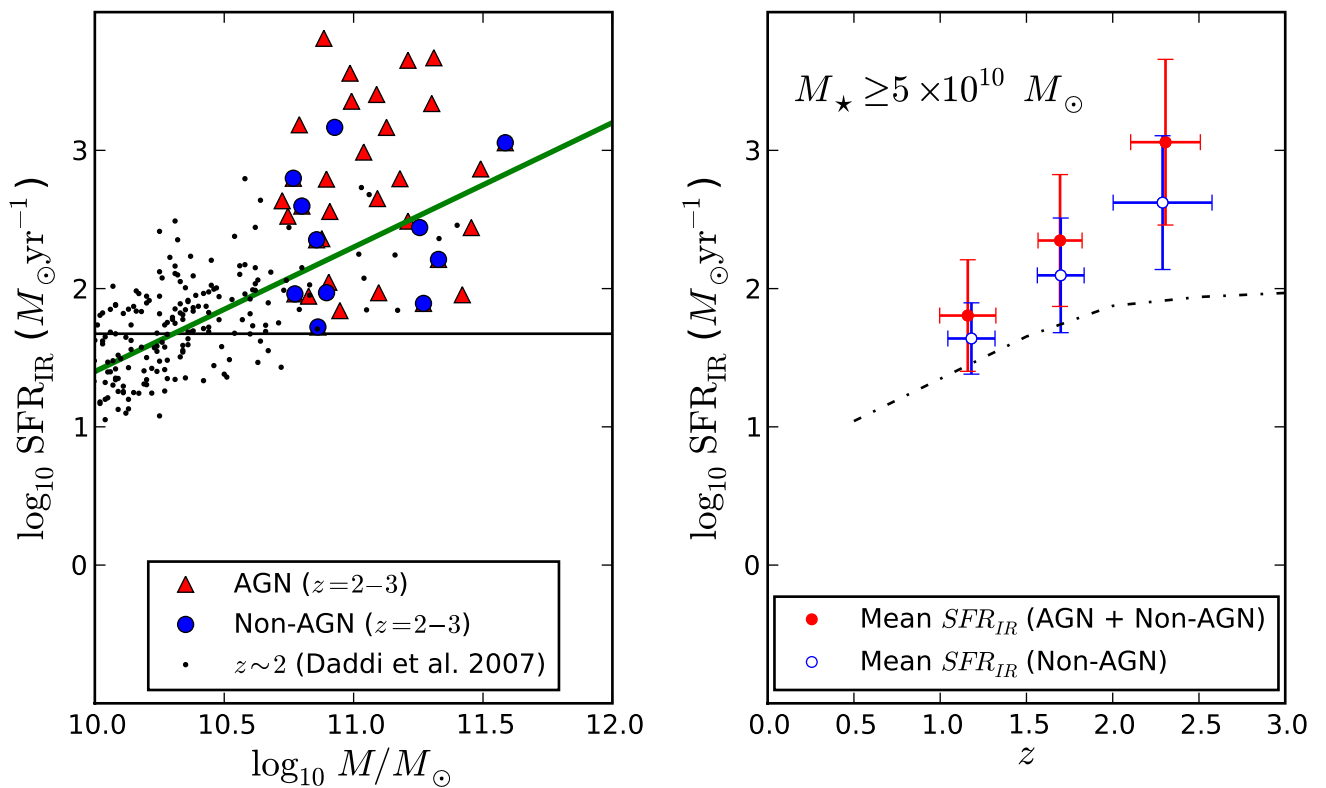
FIG. 12.— The left-hand panel shows SFR$_{\rm IR}$ versus $M_\star$ at $z = 2 - 3$. The AGN candidates are coded as triangle symbols, and their SFR$_{\rm IR}$ likely overestimate their true SFR. The completeness limit in SFR$_{\rm IR}$ (corresponding to the limiting 24 μm flux of $\sim$30 μJy) is shown as a black solid line. The black dots represent SFR from UV measurements by Daddi et al. (2007) for $z \sim 2$; the diagonal green line is their corresponding SFR-mass correlation at $z \sim 2$. The right-hand panel shows mean SFR$_{\rm IR}$ in the different redshift bins for sources with SFR$_{\rm IR}$ above the detection limit. The error bars are the 1σ standard deviation around the mean. The black line shows average UV-based SFR versus redshift for a galaxy with 5×10^{10} $M_\odot$, as calculated by Drory & Alvarez (2008).
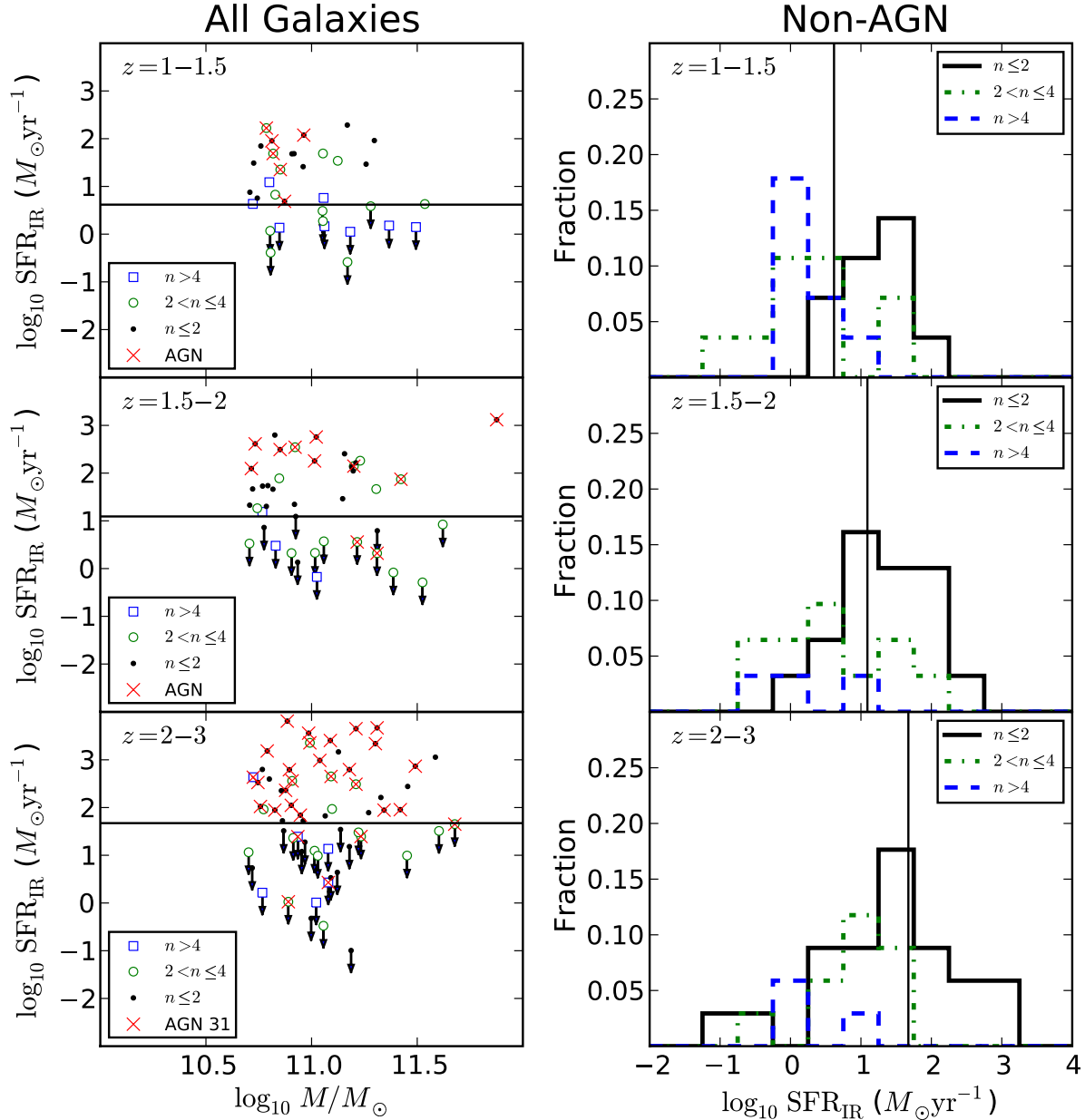
FIG. 13.— Left column: SFR_{IR} is plotted versus $M_\star$, for all galaxies with $M_\star \geq 5 \times 10^{10}\ M_\odot$, in different redshift bins spanning 1-1.4 Gyr in cosmic time. Data are sorted according to the Sérsic index n calculated in §3.1. AGN candidates (see §6) are labeled with red x's. Galaxies with SFR_{IR} below the detection limit (shown as a horizontal line) are shown with downward pointing arrows because they are upper limits. *At $z = 2 - 3$ the majority $(84.6 \pm 10.0\%)$ of massive non-AGN galaxies with SFR_{IR} above the detection limit have $n \leq 2$ (disky) structures.* Right column: For non-AGN sources, histograms show *the fraction of massive galaxies* in each redshift bin with a given SFR_{IR} for separate ranges of n. The vertical black lines mark the SFR_{IR} detection limit. For sources to the left of the line, we plot upper limits for SFR_{IR}. *The high SFR tail in each redshift bin is populated primarily by systems with low $n \leq 2$ (disky) structures.*
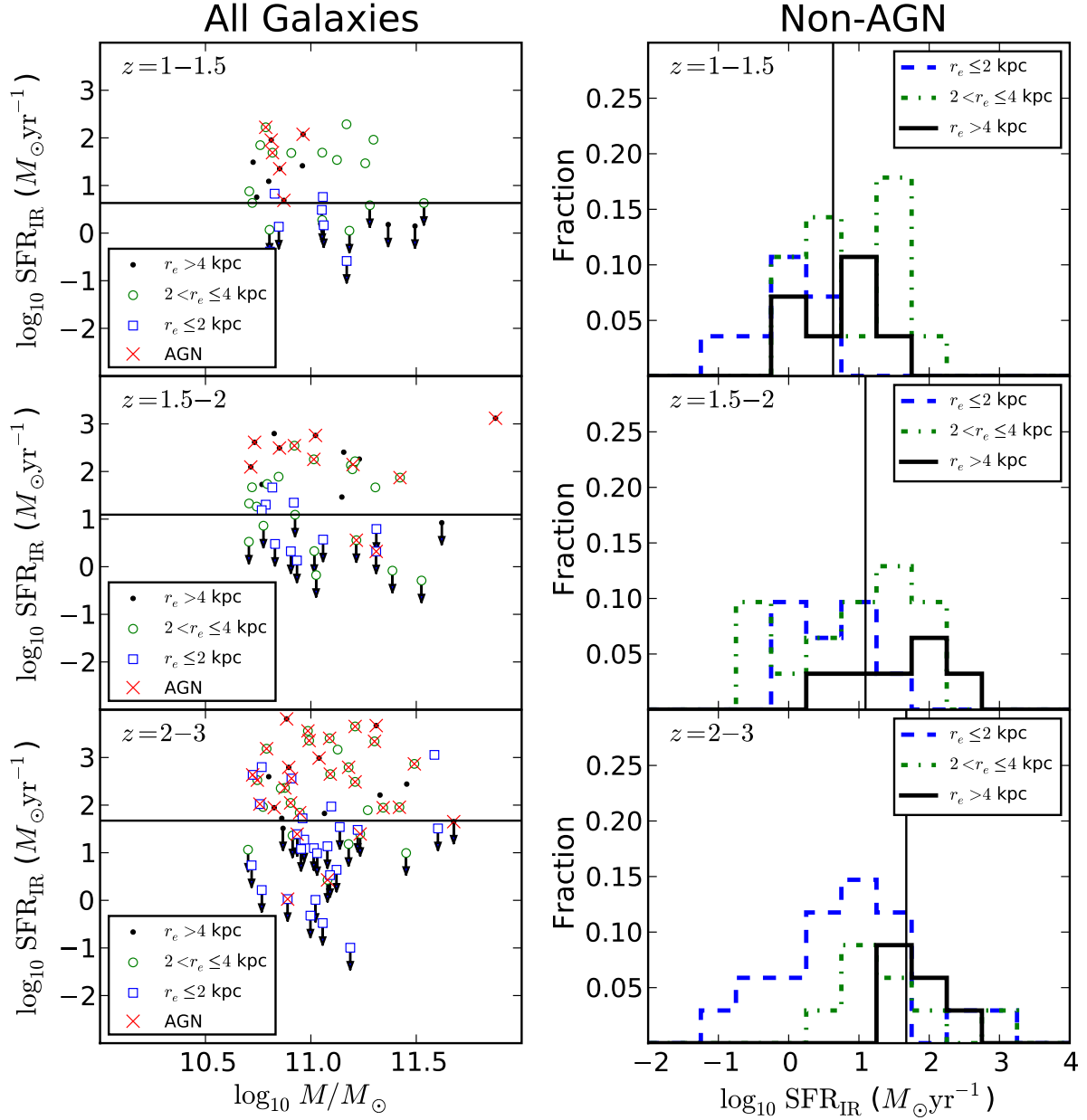
FIG. 14.— Same as Figure 13, but now the data are sorted by half-light radius r_e. Note that only a small fraction of the ultra-compact ($r_e \leq 2$ kpc) galaxies have SFR$_{\rm IR}$ above the 5σ detection limit. Some ultra-compact galaxies have high SFR$_{\rm IR}$, but, on average, their mean SFR$_{\rm IR}$ are lower than in more extended systems.
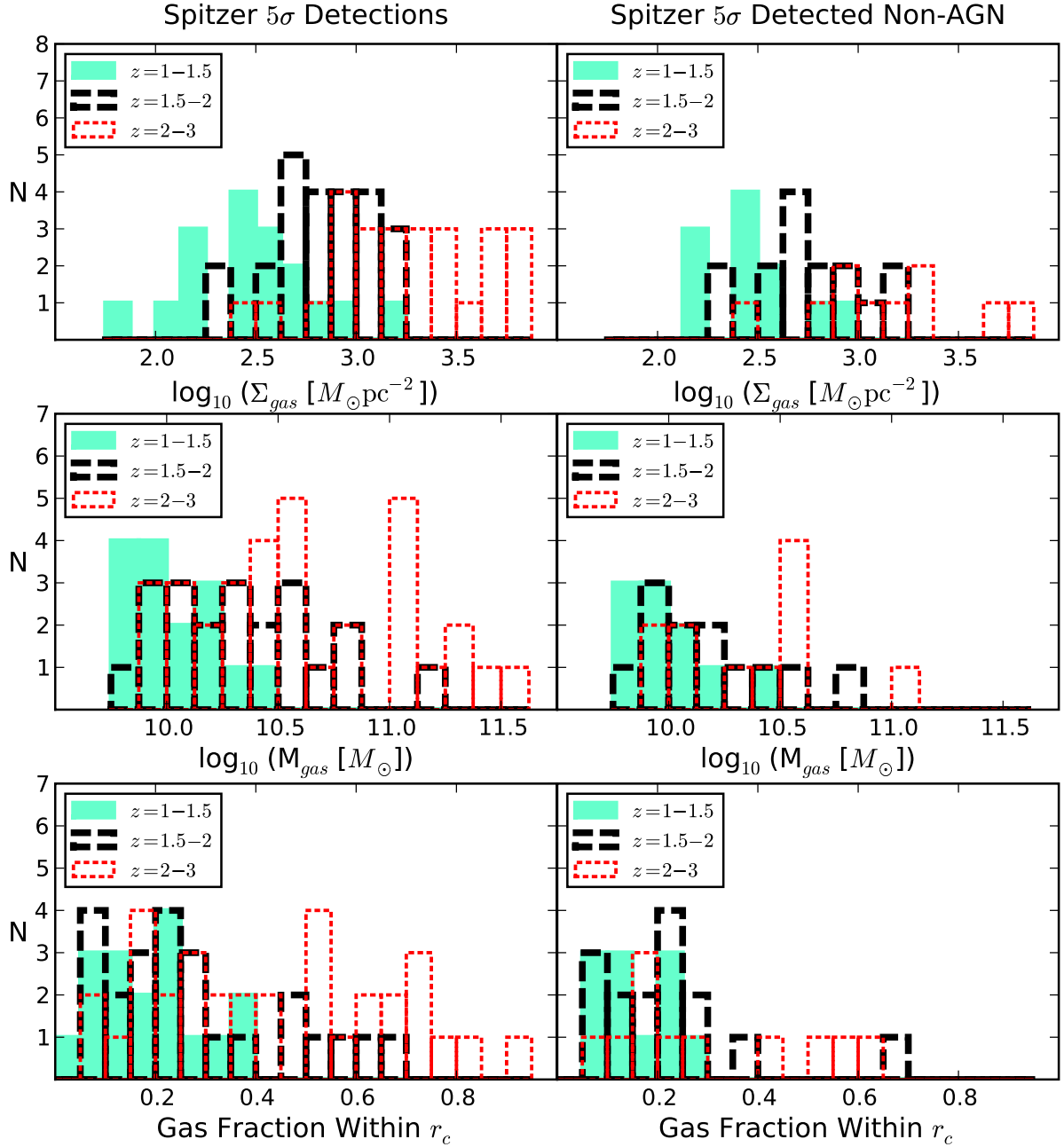
FIG. 15.— Left column: For galaxies with SFR$_{\rm IR}$ above the 5σ detection limit, the distributions of cold gas surface density ($\Sigma_{\rm gas}$), cold gas mass $M_{\rm gas}$, and cold gas fraction ($f_{\rm gas}(r_c)$) within the circularized optical half-light radius r_c are shown for different redshift ranges. $\Sigma_{\rm gas}$ is calculated using a Schmidt-Kennicutt law (Kennicutt 1998) with power-law index 1.4 a normalization factor of 2.5×10^{-4}. The cold gas fraction ($f_{\rm gas}(r_c) \equiv M_{\rm gas}/(M_{\rm gas} + M_\star)$) is calculated relative to the total baryonic mass within r_c. Right column: Same as left column except that only non-AGN sources are shown.
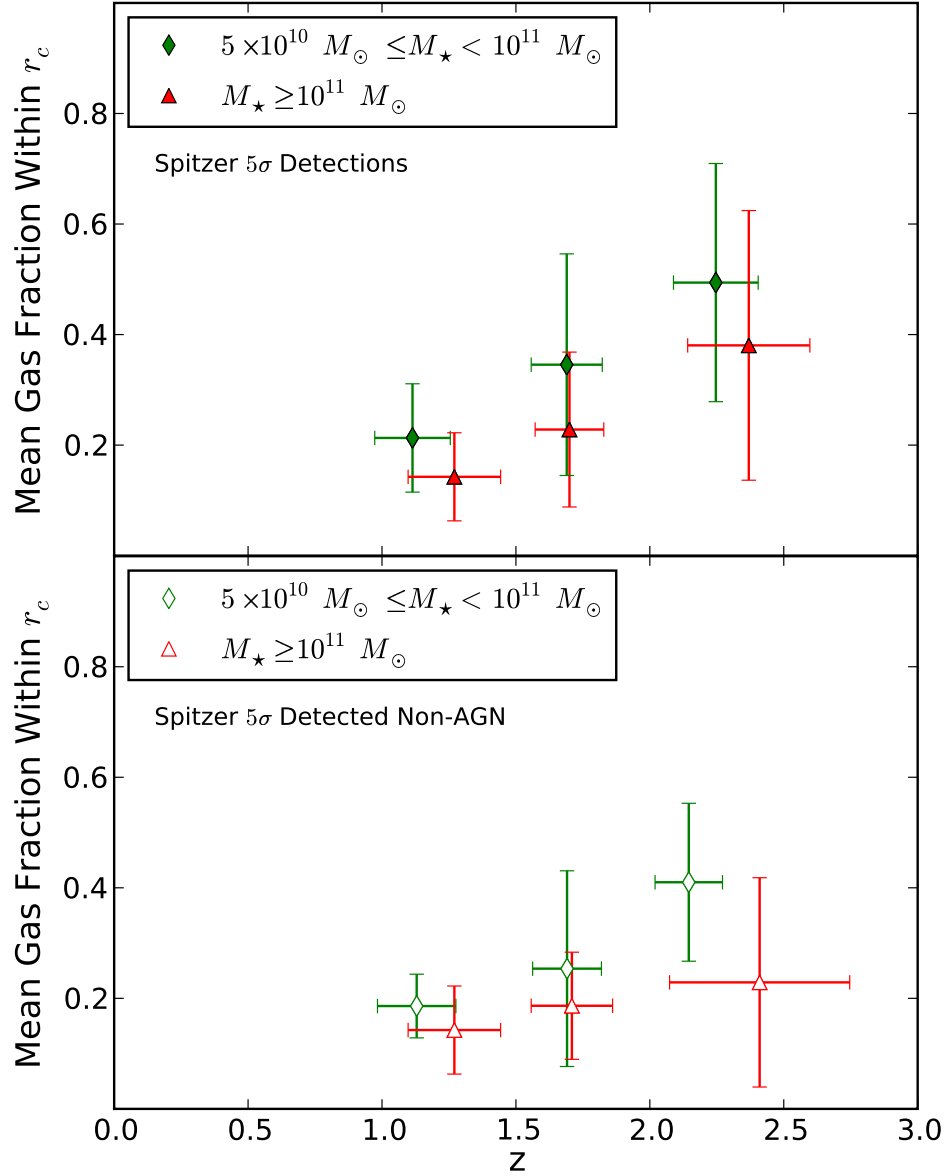
FIG. 16.— Top: For galaxies with SFR$_{\mathrm{IR}}$ above the 5σ detection limit, the mean cold gas fraction ($f_{\mathrm{gas}}(r_c) \equiv M_{\mathrm{gas}}/(M_{\mathrm{gas}} + M_\star)$) within the circularized optical half-light radius r_c is shown in three redshift bins for all galaxies with $5 \times 10^{10}~M_\odot \leq M_\star < 10^{11}~M_\odot$ and $M_\star \geq 10^{11}~M_\odot$. The error bars indicate the 1σ scatter in gas fraction and redshift. Bottom: Same as the top except that only non-AGN sources are shown.
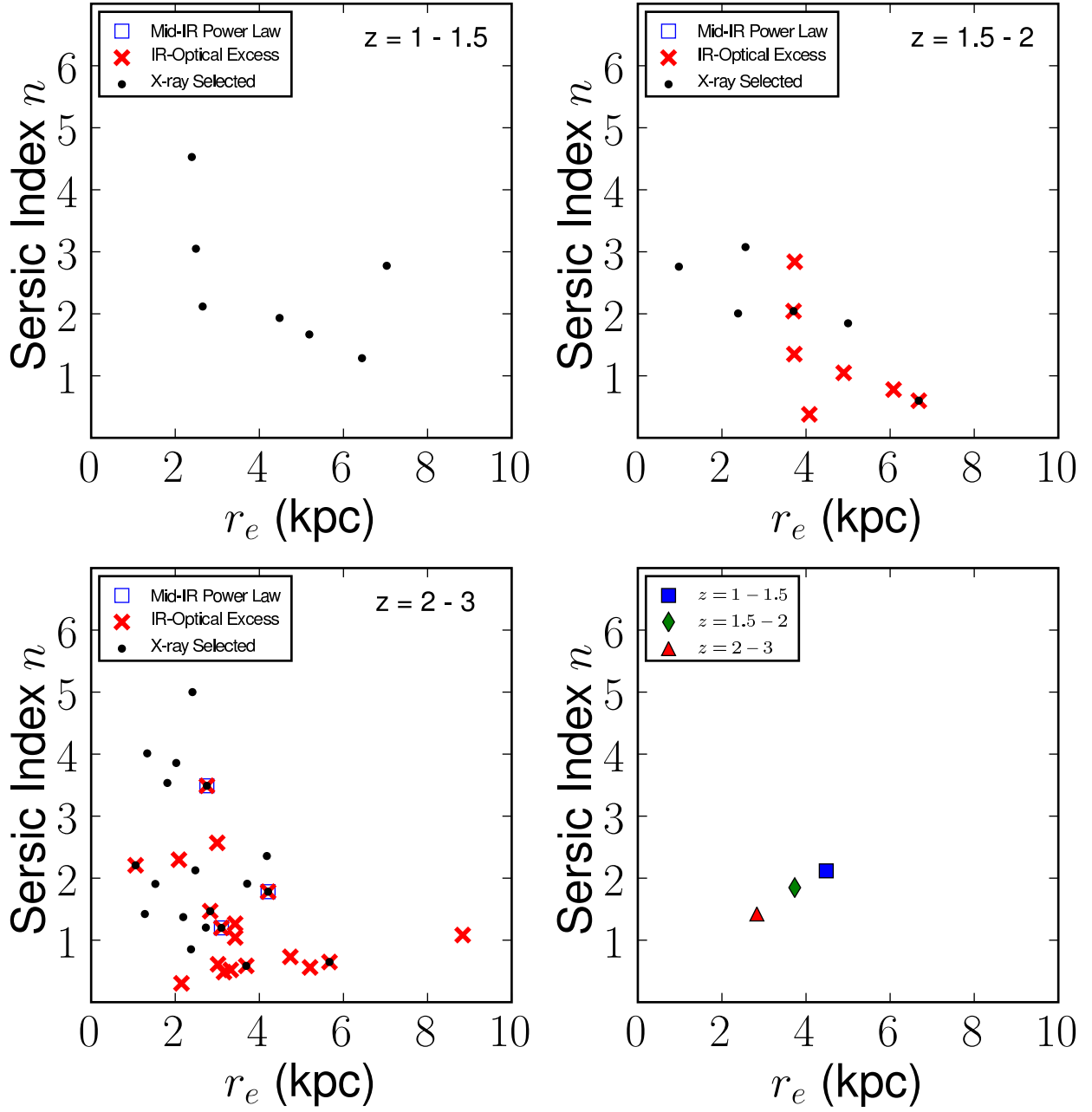
FIG. 17.— The upper and lower-left panels show single Sérsic index n versus effective radius r_e for the 49 AGN candidates selected either based on X-ray properties, mid-IR power-law, or IR-to-optical excess. The lower-right panel shows the median Sérsic index and r_e in each redshift bin.
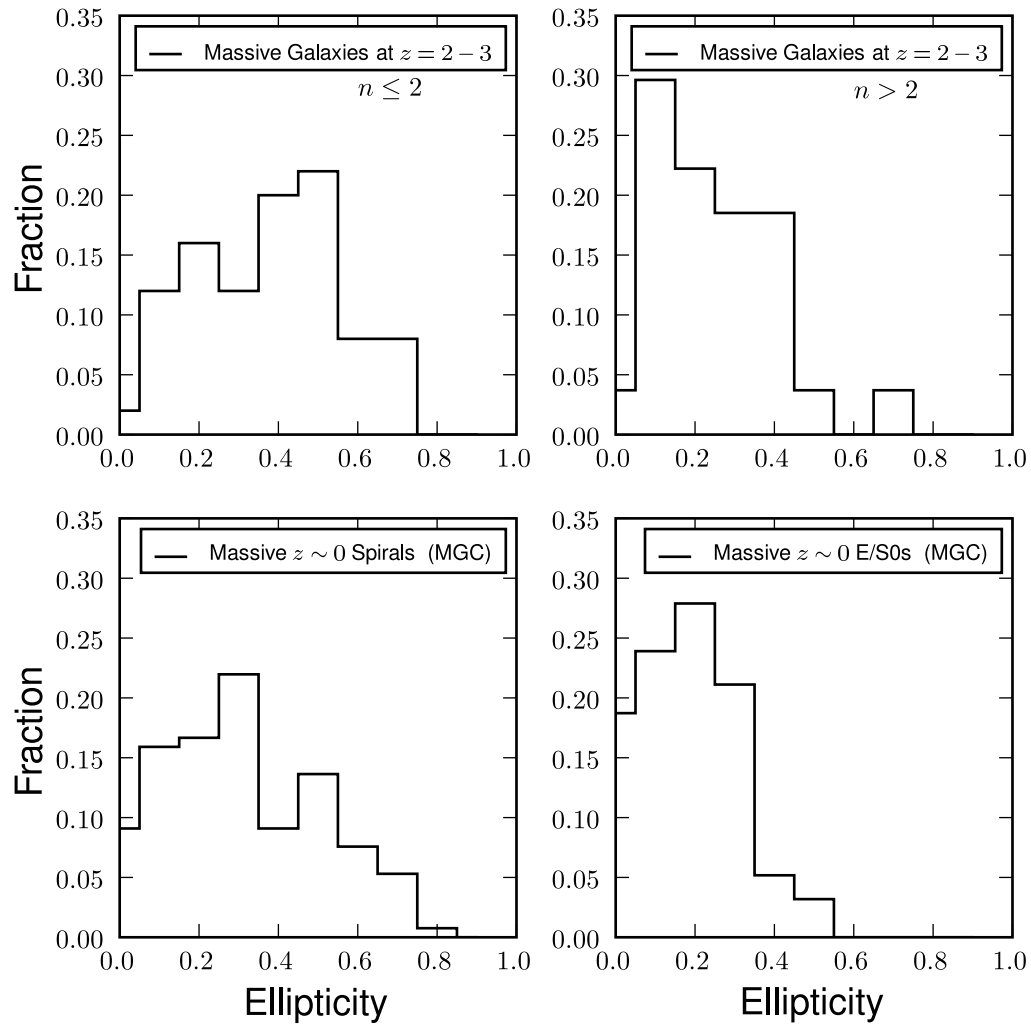
FIG. 18.— In the top panels, the deconvolved ellipticity $(1 - b/a)$ measured by GALFIT is shown for massive $(M_\star \geq 5 \times 10^{10}\ M_\odot)$ GNS galaxies at $z = 2 - 3$ with $n \leq 2$ and $n > 2$. The bottom panels show the deconvolved ellipticity for similarly massive E/S0 and Spiral galaxies as measured with GIM2D by Allen et al. (2006).
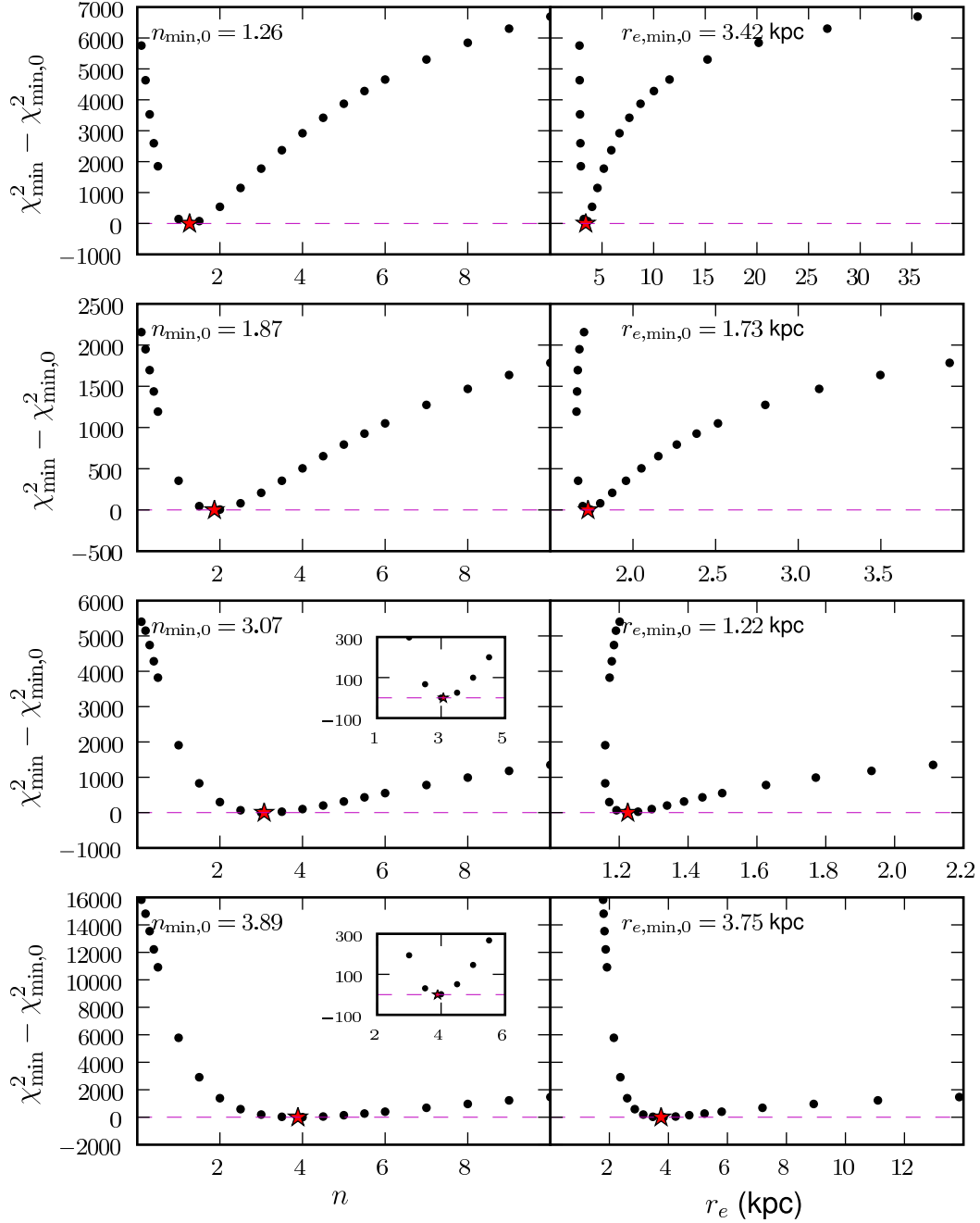
FIG. B1.— For four representative galaxies with $n \sim 1 - 4$, the first and second columns show the difference $\chi^2_{\mathrm{min}} - \chi^2_{\mathrm{min},0}$ versus n and r_e, respectively. $\chi^2_{\mathrm{min},0}$ is the minimum χ^2 obtained when all parameters are freely fit, and χ^2_{min} is the minimum χ^2 when n is held at discrete values (0.5-10). The r_e in the second column are the best-fit results for a given n and χ^2_{min}. The red stars mark the best-fit $n_{\mathrm{min},0}$ and $r_{e,\mathrm{min},0}$ corresponding to $\chi^2_{\mathrm{min},0}$. The insets in rows 3 and 4 of column 1 show a magnified view around the minimum in $\chi^2_{\mathrm{min}} - \chi^2_{\mathrm{min},0}$. Note that for galaxies with $n_{\mathrm{min},0} < 2$ (rows 1 and 2), χ^2_{min} - $\chi^2_{\mathrm{min},0}$ rises sharply at higher $n > n_{\mathrm{min},0}$, thereby making it unlikely that a higher $n > 2$ would provide a similarly good fit.
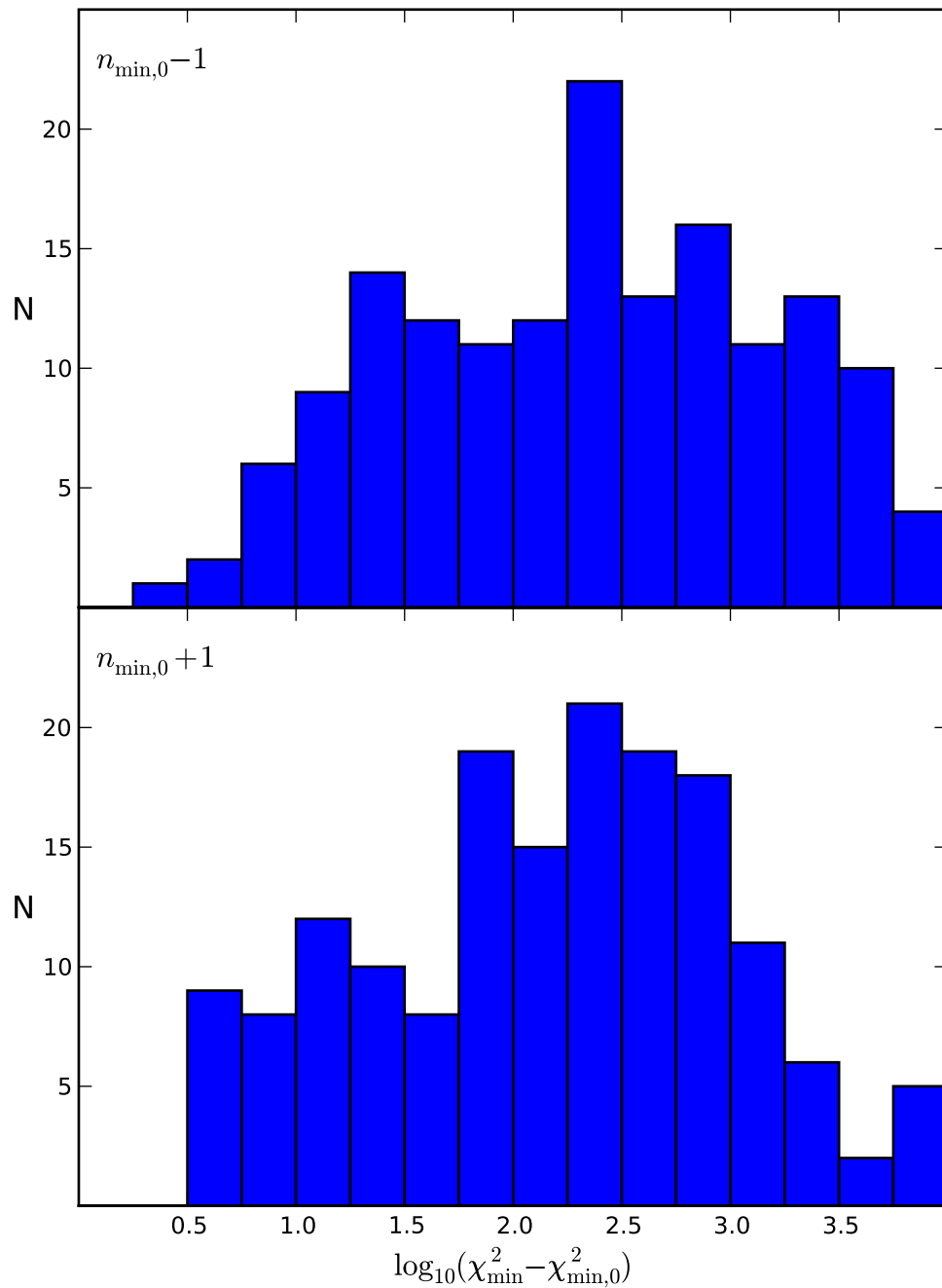
FIG. B2.— The quantity $\chi^2_{\min} - \chi^2_{\min,0}$ from Figure B1 is shown for all massive GNS galaxies well fitted with a single Sérsic profile. The top panel evaluates $\chi^2_{\min} - \chi^2_{\min,0}$ at $n_{\min,0} - 1$, and the bottom panel evaluates $\chi^2_{\min} - \chi^2_{\min,0}$ at $n_{\min,0} + 1$, where $n_{\min,0}$ is the best-fit Sérsic index corresponding to $\chi^2_{\min,0}$.
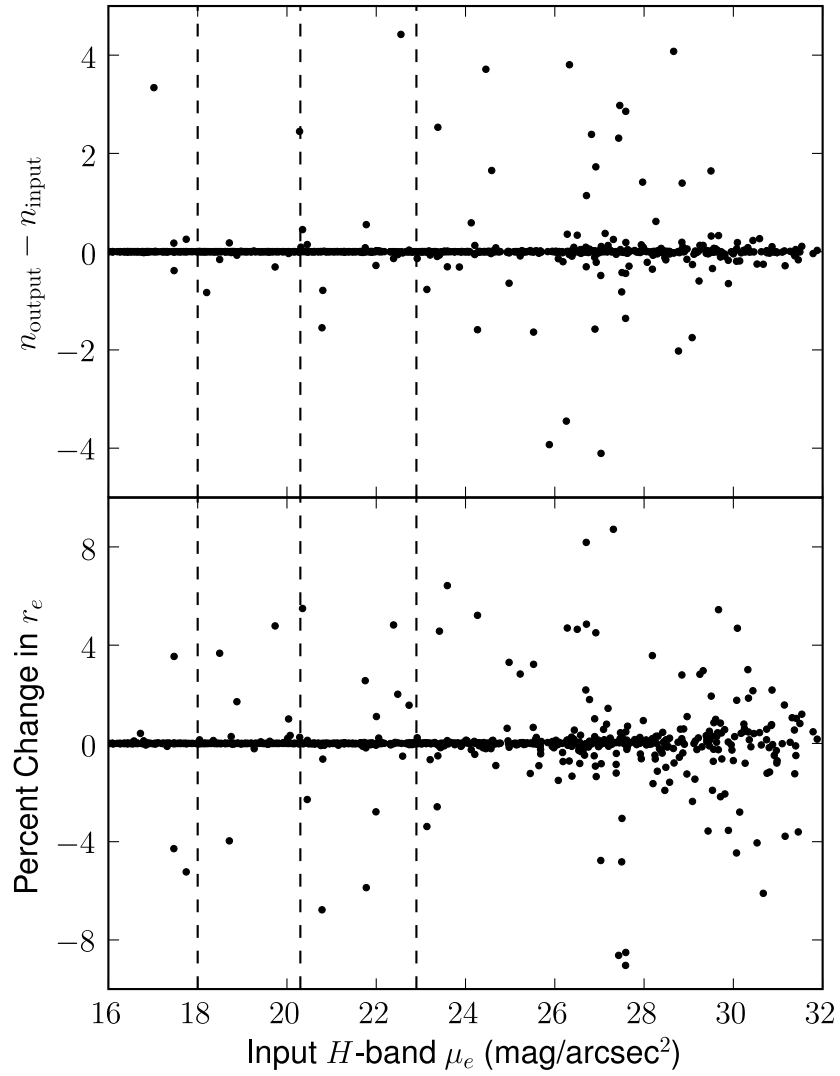
Fig. B3.— For the simulations described in §B.2, the difference between input and output Sérsic index n and effective radius r_e are plotted against effective surface brightness μ_e, the surface brightness at r_e. The vertical lines correspond to the range in μ_e in the NIC3/F160W band for the massive galaxies at $z = 1 - 3$ in our sample.
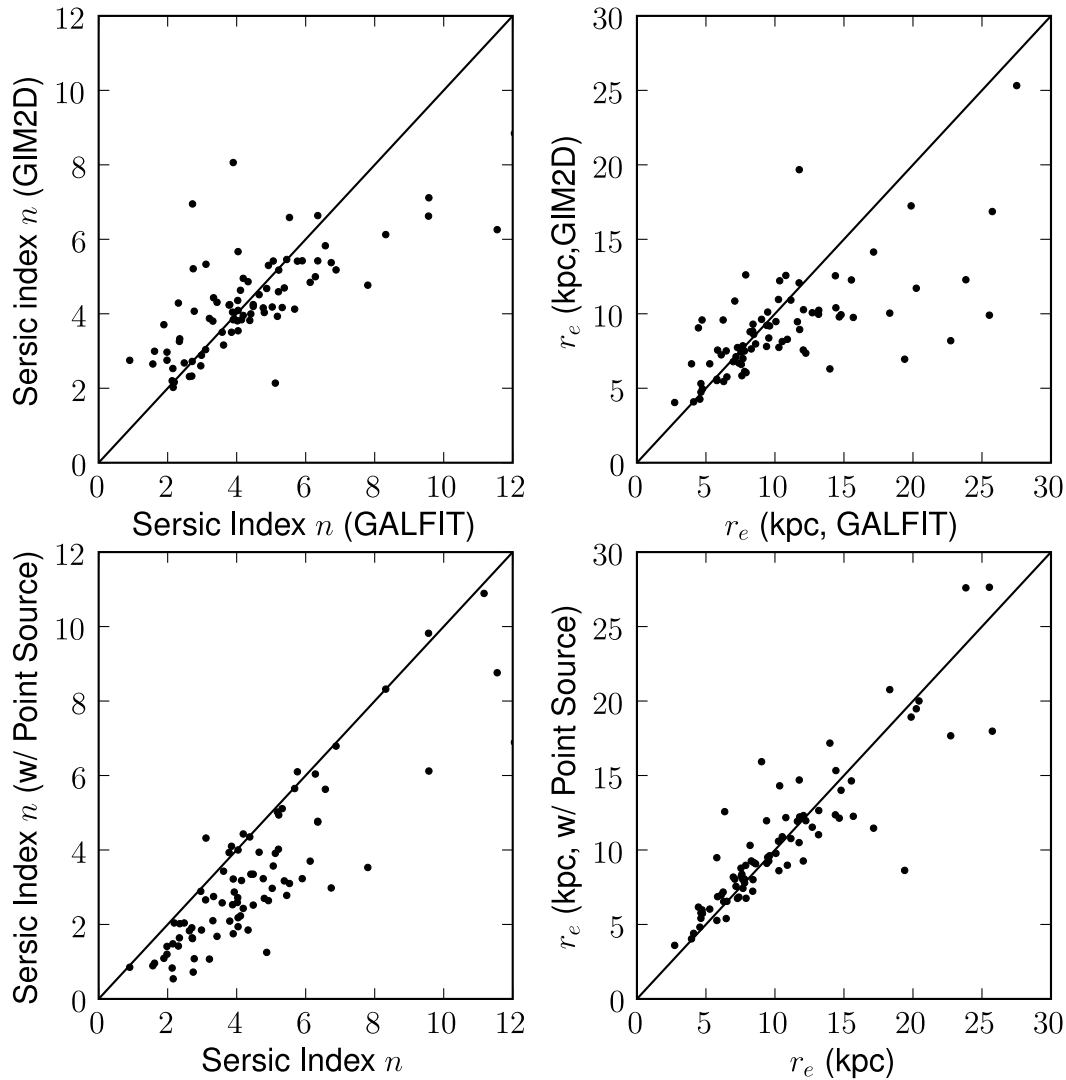
FIG. B4.— Top row: We demonstrate for a subset of $z \sim 0$ galaxies in the MGC catalog that the GIM2D-based (n, r_e) values from Allen et al. (2006) are not biased to higher values compared to our GALFIT-based fits for the same galaxies. All fits are performed on the B-band images from MGC. Bottom row: We show the effects of adding a point source in the GALFIT models fitted to the $z \sim 0$ MGC galaxies. The values obtained using a model made of a Sérsic component plus a point source are plotted along the y-axis, while the x-axis shows the values obtained with a single Sérsic component. The values of r_e are not changed systematically. The Sérsic index is lowered by the addition of the point source, but only 20% of sources with $n > 2$ in the single Sérsic fit have $n \leq 2$ after including the point source.