CONSTRAINING THE MAJOR MERGING HISTORY OF MASSIVE GALAXIES: A COMPREHENSIVE ANALYSIS OF CLOSE PAIRS AND TIDAL FEATURES USING EMPIRICAL AND SIMULATED DATA

A Dissertation
IN
Physics
and
Electrical and Computer Engineering

Presented to the Faculty of the University
of Missouri–Kansas City in partial fulfillment of
the requirements for the degree

DOCTOR OF PHILOSOPHY

by
KAMESWARA BHARADWAJ MANTHA

B. Tech., Koneru Lakshmaiah University, Vijayawada, India, 2013

Kansas City, Missouri
2021

CONSTRAINING THE MAJOR MERGING HISTORY OF MASSIVE GALAXIES: A COMPREHENSIVE ANALYSIS OF CLOSE PAIRS AND TIDAL FEATURES USING EMPIRICAL AND SIMULATED DATA

Kameswara Bharadwaj Mantha, Candidate for the Doctor of Philosophy Degree

University of Missouri–Kansas City, 2021

ABSTRACT

Major galaxy merging is a fundamental aspect of the hierarchical structure-growth scenario of the universe, and it is theoretical expected to contribute to several key aspects of galaxy evolution. As such, empirically identifying major mergers is a key methodological step towards assessing the "merging – galaxy evolution" connection, and *close-pair* and *morphology-based* methods are established empirical merger identification techniques. Yet, the merger rate measurements from these methods vary up to a factor of five owing to their unique but analogous systematic biases, especially during the key epoch of galaxy growth ($\sim 7 - 11\,\mathrm{Gyr}$ ago), highlighting that the merger contribution to galaxy growth remains poorly constrained. As a step towards addressing key open questions pertaining to empirical merger identification methodologies, we carryout comprehensive

analysis of close pairs and merging induced tidal features (and in general galactic substructures) using forefront observational data from the *Hubble Space Telescope* (*HST*) and realistic mock observations from leading theoretical simulations. We analyze the incidence of major, similar-mass (mass ratio< 4) close pairs among a large sample of ~ 9800 massive galaxies ($\log M_{\star}/M_{\odot} \geq 10.3$) from the *HST*-CANDELS survey and quantify the major merger rate evolution over $\sim 11\,\mathrm{Gyr}$ in cosmic history (*published in Mantha et al., 2018*). Using the mock light cone data from the leading SantaCruz Semi-Analytical Model (SAM), we systematically analyze the impact of different observational effects on the measurement of close-pair frequency and provide detailed statistical corrections to account for them. We also developed a new public software tool to extract and quantify different kinds of faint morphological substructures hosted by massive galaxies in the *HST* imaging and demonstrated its applicability in extracting tidal features using mock observations of a galaxy merger from a cosmological simulation (*published in Mantha et al., 2019*). Finally, using supervised and unsupervised deep-learning models, we also investigate the automated characterization of different morphological substructures hosted within the parametric light-profile subtracted residual images of $\sim 10,000$ massive galaxies from the *HST* CANDELS survey.

APPROVAL PAGE

The faculty listed below, appointed by the Dean of the School of Graduate Studies, have examined a dissertation titled "Constraining the Major Merging History of Massive Galaxies: A Comprehensive Analysis of Close Pairs and Tidal Features Using Empirical and Simulated Data," presented by Kameswara Bharadwaj Mantha, candidate for the Doctor of Philosophy degree, and hereby certify that in their opinion it is worthy of acceptance.

Supervisory Committee

Daniel H. McIntosh, Ph.D., Committee Chair
Department of Physics & Astronomy

Mark Brodwin, Ph.D.
Department of Physics & Astronomy

Paul Rulis, Ph.D.
Department of Physics & Astronomy

Masud chowdhury, Ph.D., Inter-disciplinary Committee Chair
Department of Computer Science & Electrical Engineering

Reza Derakhshani, Ph.D.
Department of Computer Science & Electrical Engineering

CONTENTS

ILLUSTRATIONS

TABLES

ACKNOWLEDGEMENTS

This work is based on observations taken by the CANDELS Multi-Cycle Treasury Program with the NASA/ESA HST, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555. Support for Program number HST-GO-12060 was provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Incorporated, under NASA contract NAS5-26555. We acknowledge the use of the Sloan Digital Sky Survey (SDSS), NASA's Astrophysics Data System Bibliographic Services, `TOPCAT` [1], the core python package for the astronomy community `Astropy` [2]. This work has been partially supported by funding from the NASA Hubble Space Telescope Archival Research grant 15040, the Missouri Consortium of NASA's National Space Grant College and Fellowship Program, and from the University of Missouri Research Board.

CHAPTER 1

MAJOR MERGING HISTORY IN CANDELS. I. EVOLUTION OF THE INCIDENCE OF MASSIVE GALAXY-GALAXY PAIRS FROM $Z = 3$ TO $Z \sim 0$

Published in *Mantha et al., 2018, MNRAS, 475, 1549*

DOI: 10.1093/mnras/stx3260

Abstract

The rate of major galaxy-galaxy merging is theoretically predicted to steadily increase with redshift during the peak epoch of massive galaxy development ($1 \leq z \leq 3$). We use close-pair statistics to objectively study the incidence of massive galaxies (stellar $M_1 > 2 \times 10^{10} M_\odot$) hosting major companions ($1 \leq M_1/M_2 \leq 4$; i.e., <4:1) at six epochs spanning $0 < z < 3$. We select companions from a nearly complete, mass-limited ($\geq 5 \times 10^9 M_\odot$) sample of 23,696 galaxies in the five CANDELS fields and the SDSS. Using $5 - 50$ kpc projected separation and close redshift proximity criteria, we find that the major companion fraction $f_{\rm mc}(z)$ based on stellar mass-ratio (MR) selection increases from 6% ($z \sim 0$) to 16% ($z \sim 0.8$), then turns over at $z \sim 1$ and decreases to 7% ($z \sim 3$). Instead, if we use a major F160W flux ratio (FR) selection, we find that $f_{\rm mc}(z)$ increases steadily until $z = 3$ owing to increasing contamination from minor (MR>4:1) companions at $z > 1$. We show that these evolutionary trends are statistically robust to changes in companion proximity. We find disagreements between published results are resolved when selection criteria are closely matched. If we compute merger rates using constant fraction-to-rate conversion

factors ($C_{\rm merg,pair}$=0.6 and $T_{\rm obs,pair}$=0.65Gyr), we find that MR rates disagree with theoretical predictions at z>1.5. Instead, if we use an evolving $T_{\rm obs,pair}(z)\propto(1+z)^{-2}$ from Snyder et al., our MR-based rates agree with theory at $0<z<3$. Our analysis underscores the need for detailed calibration of $C_{\rm merg,pair}$ and $T_{\rm obs,pair}$ as a function of redshift, mass and companion selection criteria to better constrain the empirical major merger history.

1.1 Introduction

In an hierarchical universe, collisions between similar-mass galaxies (major mergers) are expected to occur, and many theoretical studies predict such merging plays an important role in the formation and evolution of massive galaxies. A key measurement for quantifying the role of major merging in galaxy development is the merger rate and its evolution during cosmic history. A host of studies have measured major merger rates at redshifts $z \leq 1.5$, primarily based either on close-pair statistics [3–6], clustering statistics [7, 8], and morphological disturbances and asymmetries [9, 10]. These studies have all found higher incidences of major merging at earlier look-back times and a strong to moderate decrease to the present epoch, in broad agreement with many theoretical predictions [11–13]. Despite these successes, large scatter (factor of 10) exists between even the most stringent individual constraints, owing to systematic uncertainties in different methodologies and merger timescales. These issues are compounded for empirical estimates at the epoch of peak galaxy development ($z \sim 2-3$; 'cosmic high-noon'). Some early empirical estimates based on both methodologies found increasing major merger incidence at $z > 1.5$ [14], but recent studies find a possible flattening or turnover in merger

rates between $1 < z < 3$ [15–17]. These new empirical trends are in strong disagreement with recent theoretical models predicting that merger rates continue to rise from $z = 1$ to $z = 3$ and beyond [13, 18, 19]. These discrepancies and the large variance between past measurements highlight the need for improved major merger constraints, especially during the critical high-noon epoch.

A host of selection-effect issues has plagued many previous attempts to constrain major merger statistics at high redshift, from low-number statistics and significant sample variance due to small-volume pencil-beam surveys, and rest-frame UV selections of both disturbed morphologies and close pairs. While the identification of close pairs is less prone to some systematics, the lack of statistically useful samples of spectroscopic redshifts or even moderately small-uncertainty photometric redshifts at $z > 1$ until very recently have limited the usefulness of this method. Moreover, the wildly varying close-companion selection criteria among previous studies is a plausible explanation for tensions between empirical merger rates and theoretical predictions [18]. In this study, we will address many of these shortcomings and systematically explore the impact of major close-companion selection criteria by analyzing major companion fractions in a sample of 10,000 massive host galaxies (stellar mass $M_1 \geq 2 \times 10^{10} M_{\odot}$) from the *five* Hubble Space Telescope (*HST*) legacy fields in CANDELS [20, 21] and the SDSS survey. This comprehensive sample provides statistically useful major companion counts, down to a mass limit of $M_2 = 5 \times 10^9 M_{\odot}$, from rest-frame optical images over a large volume out to $z = 3$.

The hierarchical major merging of similar mass halos via gravitational accretion is the underlying physical driver of galaxy-galaxy major merging. Cosmological simulations predict that the major halo-halo merger rate rises steeply with redshift as $R \propto (1+z)^{2-3}$ [22–24], which is in agreement with simple analytical predictions based on Extended Press-Schechter (EPS) theory of $R \propto (1+z)^{2.5}$ [25, 26]. Cosmologically-motivated simulations of galaxy formation and evolution predict major galaxy-galaxy merger rates that follow $R \propto (1+z)^{1-2}$ over a wide redshift range (e.g., $z < 6$) [19]. While there is some debate on the increasing merger rate evolution among theoretical studies due to model-dependencies [13], some works claim flattening of merger rates with increasing redshift [27], most agree with an increasing incidence (within a factor-of-two uncertainty). Not only are merger rates expected to be higher at early cosmic times, but major galaxy merging is predicted to play a crucial role in nearly all aspects of the formation and evolution of massive galaxies including buildup of spheroidal bulges and massive elliptical galaxies [28–32], triggering and enhancement of star formation (SF) including nuclear starbursts [33–36], and the fueling of active galactic nuclei (AGN) [13, 37–39] and subsequent SF quenching [40, 41].

Many empirical studies support the predictions that major merging may explain the documented build-up of massive and quenched (non-star-forming and red) galaxy number densities and their stellar content growth at $z < 1$ [42–44], enhancement of SF activity [45, 46], and elevation of AGN activity [47–49]. Despite this agreement, some studies find a weak major merging-SF connection and suggest mergers may not be the dominant contributor to in-situ galactic SF [50–52]. Moreover, other studies find a lack

of a merger-AGN connection [53–56]. These conflicting observations lend support to theories that predict violent disk instabilities (VDI) due to the rapid hierarchical accretion of cold gas may be responsible for key processes like bulge formation and AGN triggering [57, 58]. Indeed, a recent CANDELS study by [59] found the observed evolution of massive quenched spheroids at $z < 3$ is better matched to SAM predictions that include both mergers and disk-instability prescriptions. Therefore, the role of major merging in galaxy evolution remains a critical open question. Hence, measuring the frequency and rate at which major mergers occur at different cosmic times using large, uniformly selected close-pair samples is a key step towards answering the role played by them in massive galaxy development.

Theoretical simulations predict that galaxies involved in major close pairs will interact gravitationally and coalesce over time into one larger galaxy, and thereby make them effective probes of ongoing or future merging. Many studies in the past have employed the close-pair method to estimate the frequency of major merging as a function of cosmic time. This typically involves searching for galaxies that host a nearby companion meeting a number of key criteria: (i) 2-dimensional projected distance, (ii) close redshift-space proximity, and satisfies a nearly-equal mass ratio M_1/M_2 between the host (1) and companion (2) galaxies. For each criterion, a wide range of choices is used in the literature. For projected separation $R_{\rm proj}$, a search annulus is often employed with minimum and maximum radii. Common choices vary between $R_{\rm max} \sim 30 - 140\,$kpc [60, 61] and $R_{\rm min} \sim 0 - 14$ kpc [14, 16]. Depending on available redshift information, the

choice of physical proximity criterion ranges from stringent spectroscopic velocity differences (commonly $\Delta v_{12} \leq 500\,\mathrm{km\,s^{-1}}$) [62] to a variety of photometric redshift $z_{\rm phot}$ error overlaps [6,63] To study similar-mass galaxy-galaxy mergers, previous studies have adopted stellar-mass-ratio selections ranging from $2 > M_1/M_2 > 1$ (or 2:1) [64] to $5 > M_1/M_2 > 1$ (5:1) [65], with 4:1 being by far the most common mass ratio criterion. In the absence of stellar-mass estimates, flux ratio F_1/F_2 is often used as a proxy for M_1/M_2 [66]. The wide range of adopted close-companion selection criteria lead to a large scatter in two decades of published pair-derived merger rates with redshift evolution spanning $R \propto (1+z)^{0.5-3}$ at $0 < z < 1.5$, and sometimes even indicating a flat or turnover in merger rates at $z > 1.5$ [16, 17, 63, 67]. This large scatter in pair-derived merger rate constraints highlights the strong need for tighter constraints at cosmic high-noon, and motivates a careful analysis of selection effects.

Numerical simulations of the gravitation interactions between merging galaxies can produce disturbed morphological features due to strong tidal forces [68–70]. As such, morphological selections have also been used to empirically identify mergers. These selections are broadly divided into visual classifications [71, 72], analysis of image−model residuals [43, 73], and automated measures of quantitative morphology such as Gini-M20 [74] and CAS [75]. Although morphology-based studies broadly find merger rates to be rising strongly with redshift as $(1+z)^{2-5}$ [76, 77], sometimes finding as high as $25\% - 50\%$ of their sample as mergers [78], there are significant study-to-study discrepancies where some studies find no merger rate evolution [9,79]. Morphology-based selections depend on identifying relatively fainter disturbances than the galaxy, which makes

this method prone to systematics. The cosmological surface-brightness of galaxies falls off as $(1+z)^{-4}$, which can lead to a biased identification of faint merger-specific features as a function of redshift. In addition, most of these morphology-based merger rates are based on small-volume, pencil-beam surveys probing the rest-frame UV part of the spectrum, especially at $z > 1$. This can lead to over-estimation of merger rates due to contamination from non-merging, high star-forming systems with significant substructure that can be confused as two merging galaxies. Recent theoretical developments suggest that VDI can also cause disturbances in the host galaxy morphology and mimic merger-like features [80–82], which in principle may complicate the measurement of morphology-based merger rates. Thus, to robustly identify plausible merging systems out to high redshifts ($z \sim 3$) without having to rely on imaging-related systematics strongly, we resort to the close-pair method in this study. We acknowledge that the close-pair method has its limitations at high redshift where galaxies have large photometric redshift uncertainties, which may lead to incorrect merger statistics. In this study, we initially exclude the galaxies with unreliable redshifts from our analysis, but later add back a certain fraction of them by employing a statistical correction.

In this work, we analyze galaxy-galaxy close pairs in a large sample of 5698 *massive* galaxies ($M_{\rm stellar} \geq 2 \times 10^{10} M_{\odot}$) from the state-of-the-art Cosmic Assembly Near-Infrared Deep Extragalactic Legacy Survey (CANDELS) [20, 21] of five highly-studied extragalactic fields at six epochs spanning $z \sim 0.5 - 3.0$ (with a width $\Delta z = 0.5$). To simultaneously anchor our findings to $z = 0$, we take advantage of 4098 massive galaxies from the Sloan Digital Sky Survey (SDSS) [83], Data release 4 (DR4) [84] at

$z \sim 0.03 - 0.05$, which is matched in resolution to CANDELS and probes an average of $\sim 1.3 \times 10^6$ Mpc^3 per redshift bin. With the available data, we also perform rigorous analyses to understand the impact of different close-companion selection criteria on the derived results.

We structure this chapter as follows: In § 1.2, we provide a brief description of the CANDELS and SDSS data products (redshifts and stellar masses) and describe the selection of massive galaxies hosting major companions based on the stellar mass complete massive galaxy sample. In § 1.3, we describe the calculation of major companion fraction and its redshift evolution including necessary statistical corrections. In § 1.4, we discuss the impact of close-companion selection choices on the derived major companion fractions. In § 1.5, we calculate the major merger rates based on the major companion fractions. We synthesize detailed comparisons of the companion fractions and merger rates to other empirical studies and theoretical model predictions, and also discuss plausible reasons and implications of disagreement between the observed and theoretical merger rates. We present our conclusions in § 1.6. Throughout this paper, we adopt a cosmology of H_0 = 70 km s^{-1}Mpc^{-1} ($h = 0.7$), $\Omega_{\rm M}$ = 0.3 and Ω_Λ = 0.7, and use the AB magnitude system [85].

1.2 Data and Galaxy Sample

In this study, we analyze close galaxy-galaxy pairs selected from a large sample of massive galaxies from the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey [20, 21], spanning redshifts $0.5 \leq z \leq 3$, and subdivided into five epochs probing

a volume of $\sim 1.3 \times 10^6$ Mpc^3 each. We anchor our findings to $z \sim 0$ using a sample from the SDSS that is matched in volume and resolution to the CANDELS sample. To reliably track the major merging history since $z = 3$ using close-pair method, we start with a mass-limited sample of galaxies that will allow a complete selection of massive $> 2 \times 10^{10}\, M_\odot$ galaxies with major companions meeting our chosen stellar mass ratio: $1 \leq M_1/M_2 \leq 4$ (1, 2 represent host and companion galaxies, respectively). In this section, we describe the relevant details of the data necessary to achieve this sample selection.

1.2.1 CANDELS : 5 Legacy Fields

1.2.1.1 Photometric Source Catalogs

The five CANDELS *HST* legacy fields – UDS, GOODS-S, GOODS-N, COSMOS, and EGS – have a wealth of multi-wavelength data and cover a total area of $\sim 800\,\mathrm{arcmin}^2$ ($\sim 0.22\,\mathrm{deg}^2$). The CANDELS survey observations and image processing are described in [20] and [21], respectively. We use the photometric source catalogs from UDS [86], GOODS-S [87], [88] (GOODS-N), COSMOS [89], and EGS [90]. Each catalog was generated with a consistent source detection algorithm using SExtractor applied to the F160W (H-band) 2-orbit depth CANDELS mosaic image produced for each field. These authors used profile template fitting [91] to provide uniform photometry and spectral energy distributions (SEDs) for each galaxy at wavelengths spanning 0.4μm to 1.6μm, supplemented by ground-based data [87] and *spitzer*/IRAC photometry (3.6μm to 8.0μm) from the S-CANDELS survey [92] . Each photometric object was assigned a flag (PhotFlag) to identify plausible issues using a robust automated routine described in [86]. We

use PhotFlag = 0 to remove objects with contaminated photometry due to nearby stars, image artifacts or proximity to the F160W coverage edges. This cut removes $\sim 3-5\%$ of raw photometric sources depending on the field. We also use the stellarity index from SExtractor (Class_star ≥ 0.95) to eliminate bright star-like sources. We estimate that this additional cut removes active compact galaxies that makeup $\sim 1.3\%$ of our total desired mass-limited sample. We note that including these galaxies has no significant impact on our conclusions. We tabulate the total raw and good photometric source counts for the five CANDELS fields in Table 1.

1.2.1.2 Redshifts & Stellar Masses

We use the CANDELS team photometric redshift and stellar mass catalogs available for each field. For the CANDELS UDS and GOODS-S fields, the redshifts are published in [93], and the masses are found in [94]. For the remaining fields, we use the catalogs: GOODS-N [88], COSMOS [89], and EGS [90]. As discussed extensively in [93], photometric redshift probability distribution functions $P(z)$ were computed for each galaxy by fitting the SED data. This exercise was repeated by six participants (#ID 4, 6, 9, 11, 12, and 13 in [93]) who performed SED fitting using different codes (EAZY, HyperZ) and template sets (BC03, PEGASE, EAZY). Additional detailed discussion on individual code functionality and their respective fitting priors can be found in [93]. A team photometric redshift ($z_{\rm phot}$) was computed for each source equal to the median of the six $P(z)$ peak redshifts. When compared to a known spectroscopic sample, these

photometric redshifts have an outlier removed RMS scatter $\sigma_z \sim 0.029$ (see [93] for definition). Additionally, spectroscopic redshifts ($z_{\rm spec}$) are also available for small subsets of galaxies in each field. The best available redshift $z_{\rm best}$ is cataloged as either the team $z_{\rm phot}$ or the good quality $z_{\rm spec}$ measurements when available, which are defined by the flag q_zspec = 1 [93]. Note that the compilation of redshifts included in our analysis sample does not include grism redshifts. We limit our selection of massive galaxies to $0.5 \leq z_{\rm best} \leq 3.0$, and we employ a redshift bin size $\Delta z = 0.5$ to probe evolution between 5 and 11 Gyr ago using five roughly equal co-moving volumes ranging between 7×10^5 Mpc^3 $- 1.3 \times 10^6$ Mpc^3. We exclude redshifts $z_{\rm best} < 0.5$ since this volume is ~ 10 times smaller ($\sim 1.3 \times 10^5$ Mpc^3).

The stellar masses ($M_{\rm stellar}$) were estimated for each source by fitting the multiband photometric data to SED templates with different stellar population model assumptions[1] fixed to the object's $z_{\rm best}$. The team stellar mass (see [94, 95]) for each source is chosen as the median of the estimates based on the same assumptions of IMF [96] and stellar population templates [97]. Using the median mass estimate, we select a mass-limited ($M_{\rm stellar} \geq 5 \times 10^9 M_\odot$) sample of 14,513 potential companion galaxies in a redshift range $0.5 \leq z_{\rm best} \leq 3$ (for breakdown, see Table 1). As described in the next section, this provides a sample with high completeness.

[1]Each model is defined by a set of stellar population templates, Initial Mass Function (IMF), Star Formation History (SFH), metallicity and extinction law assumptions; see [95]

1.2.1.3 $0.5 \leq z \leq 3.0$ Sample Completeness

We demonstrate the completeness of massive CANDELS galaxies with redshifts $0.5 \leq z \leq 3.0$ by adopting the method introduced in [89, 98]. Briefly, [98] computes a stellar-mass limit as a function of redshift, above which nearly all the galaxies are observable and complete. They do so by estimating the limiting stellar-mass ($M_{\rm stellar,lim}$) distributions for the 20% faintest sample population[2], where $M_{\rm stellar,lim}$ of a galaxy is the mass it would have if the apparent magnitude ($H_{\rm mag}$) is equal to the limiting H-band magnitude ($H_{\rm lim}$). We estimate the $M_{\rm stellar,lim}$ by following Nayyeri17 relation between the observed galaxy $M_{\rm stellar}$ and its $M_{\rm stellar,lim}$ as $\log_{10}(M_{\rm stellar,lim}) = \log_{10}(M_{\rm stellar}) + 0.4(H_{\rm mag} - H_{\rm lim})$ (see [89]) and use the published H-band 5σ limiting magnitudes [20, 21, 86, 89, 90].

In Figure 29, we show the normalized cumulative distributions of $M_{\rm stellar,lim}$ for the 20% faintest CANDELS $\log_{10}(M_{\rm stellar}/M_{\odot}) \geq 9.7$ galaxy samples[3] in narrow ($\Delta z = 0.25$) redshift slices at $z > 1$. At all redshift bins up to $z = 2.25$, we find that all the galaxies in our mass-limited sample have $M_{\rm stellar} > M_{\rm stellar,lim}$, which implies 100% completeness. At redshifts $2.25 < z < 2.75$ and $2.75 < z < 3$, we find that the desired sample selection is $> 95\%$ complete and 90% complete, respectively. Additionally, we test the impact of surface brightness on the measured stellar-mass completeness by analyzing the *effective* H-band surface brightness ($\rm SB_H$) distributions of our desired mass-limited $\log_{10}(M_{\rm stellar}/M_{\odot}) \geq 9.7$ galaxy sample at five redshift bins between $0.5 < z < 3$. We

[2]By considering the 20% faintest galaxy sample of the apparent magnitude distribution at each redshift bin, only those galaxies with representative mass-to-light ratios close to the $H_{\rm lim}$ are used towards estimating the $M_{\rm stellar,lim}$ (see for additional details [98]).

[3]We compute the distributions independently for the five CANDELS fields and present the mean of them at each redshift slice. We find that the behavior of individual field distributions is not significantly different from each other and with the mean distribution.

use a H-band surface brightness limit $\mathrm{SB}_H = 26.45\,\mathrm{mag/arcsec^2}$ based on the model-galaxy recovery simulations by [16] and find that 100% and $> 95\%$ of our desired galaxies have $\mathrm{SB}_H < 26.45\,\mathrm{mag/arcsec^2}$ at redshifts $0.5 < z < 2$ and $2 < z < 3$, respectively. This implies that even the population that constitutes *lowest* 10% of the SB_H distribution (low surface brightness galaxies; hereby LSB galaxies) in our desired sample can be robustly detected up to $z = 3$. As the LSB galaxies only make up a small fraction (less than 10%) of our desired mass-limited sample, we expect that a smaller completeness among these LSB galaxies will not have an significant impact on the close-pair statistics presented in this study. These tests permit us to robustly search for major companions associated with $\log_{10}(M_{\mathrm{stellar}}/M_{\odot}) \geq 10.3$ galaxies unaffected by significant incompleteness. We include the breakdown of $N_{\mathrm{m}} = 5698$ massive galaxies per CANDELS field in Table 1.

1.2.2 SDSS

1.2.2.1 Redshifts & Stellar Masses

To anchor evolutionary trends to $z \sim 0$, we employ redshifts and stellar masses from Sample III of the SDSS Group Catalog described in [99]. Briefly, this catalog contains $\sim 400,000$ galaxies spanning a redshift range $0.01 < z < 0.2$ from the ~ 4500 square degree sky coverage of the SDSS Data Release 4 (DR4) [84]. [99] computed $(g - r)$ color-based M_{stellar} estimates using the [100] SED fitting based mass-to-light ratio calibrations and K-corrections from the NYU-VAGC [101]. For consistency, these masses were corrected by -0.1 dex to convert from a 'diet' Salpeter IMF to a [96] IMF

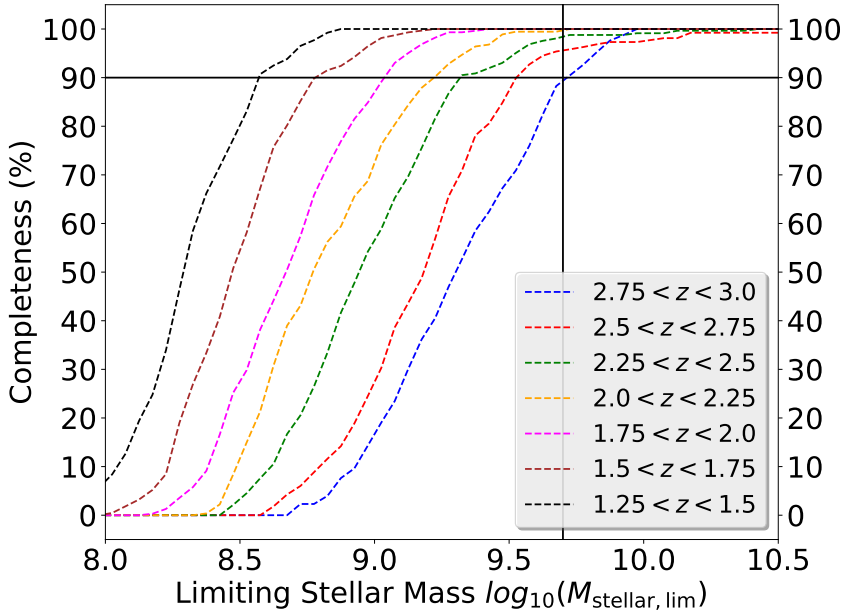

Figure 1: Stellar-mass completeness of the $\log_{10}(M_{\rm stellar}/M_{\odot}) \geq 9.7$ CANDELS galaxy sample. We show the normalized cumulative distributions of the limiting stellar masses ($M_{\rm stellar,lim}$) for the 20% faintest galaxies of the desired mass-limited sample (dashed lines), color-coded according to their respective redshift slices ($\Delta z = 0.25$) at $z \geq 1$ (see § 1.2.1.3 text for details). We show our desired mass limit in solid vertical black line and we mark the 90% completeness in solid horizontal black line. At the highest redshift slice ($2.75 < z < 3$), we find that 90% of the galaxies with $\log_{10}(M_{\rm stellar}/M_{\odot}) \geq 9.7$ have their limiting stellar masses smaller than the desired major companion mass limit, implying that the CANDELS mass-limited sample of $\log_{10}(M_{\rm stellar}/M_{\odot}) \geq 9.7$ galaxies is at least 90% in the desired redshift range of this study $0.5 \leq z \leq 3.0$.

basis as in CANDELS. Besides the IMF, [100] assumed similar exponentially declining star formation histories as the CANDELS team $M_{\rm stellar}$ participants, but used PÉGASE stellar population models [102] in contrast to [97], respectively (for details, see [95]). However, [103] explored the impact of these model assumptions and found that both PÉGASE and [97] yield similar results in terms of [100] color and mass-to-light ratio calibrations. In addition for a sample of galaxies with SDSS+GALEX photometry, [104] found good agreement between SED fitting-derived stellar masses and independent SDSS photometry-based estimates. Hence, we conclude that the CANDELS and SDSS stellar mass estimates are not systematically different.

We select Sample III galaxies within a redshift range $0.03 \leq z \leq 0.05$ and sky area 1790 sq.deg ($\rm RA = 100\,deg - 210\,deg$ & $\rm DEC = 17\,deg - 69\,deg$) to match the CANDELS sample in volume and resolution. Using these cuts, we find 9183 galaxies with $\log_{10}(M_{\rm stellar}/M_{\odot}) \geq 9.7$. We present the SDSS selection information in Table 1. We are aware of more recent datasets than the SDSS-DR4; e.g.,the SDSS-DR7 [105] has an improvement in photometric calibration from 2% (DR4) to 1% (DR7). However, owing to the contribution from $\sim 20\%$ random and $\sim 25\%$ model dependent systematic uncertainties for [100] $M_{\rm stellar}$ estimates, we argue that these small photometric improvements have no significant impact on our results. Hence, we use the SDSS-DR4 because it is readily available and it meets our volume and resolution requirements.

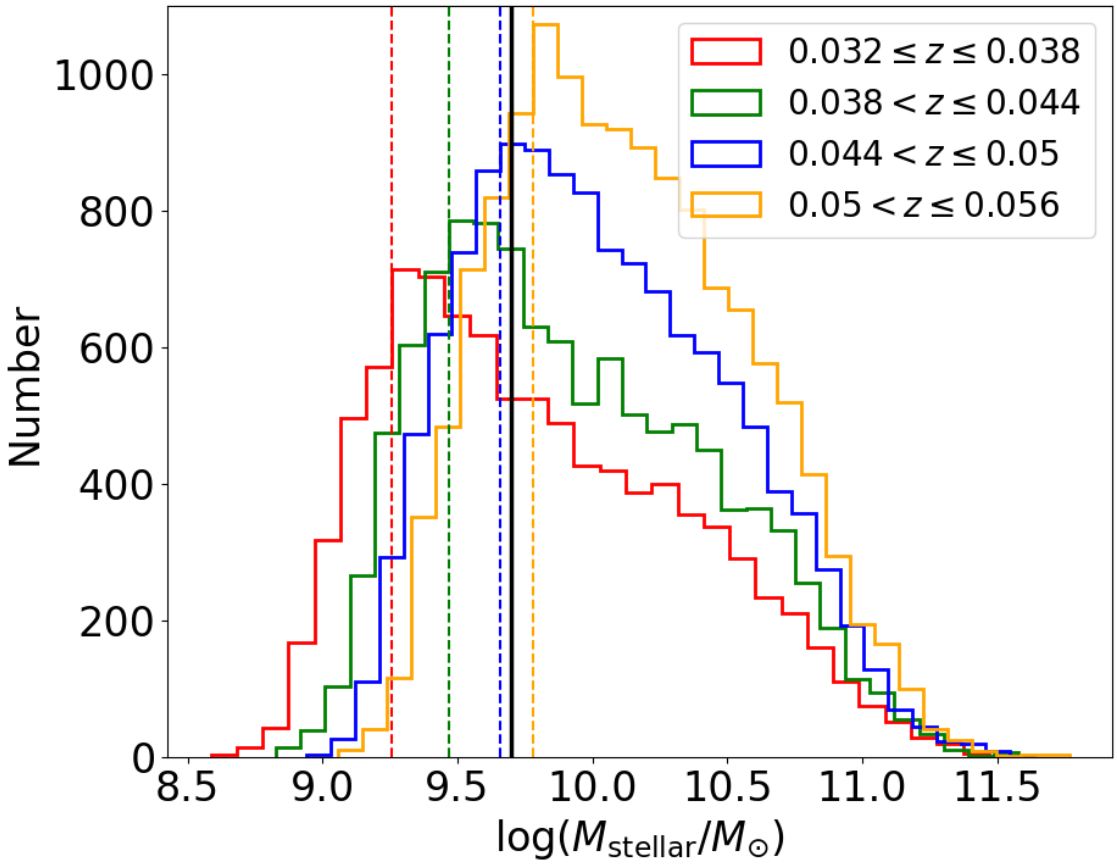

Figure 2: Stellar mass distributions for galaxies in the SDSS group catalog [99] for four narrow redshift slices (see legend). The stellar mass at which the mass distribution turns over owing to the $r < 17.77$ mag criteria for the SDSS spectroscopic target selection [106] is given by the vertical dashed lines. The solid black line indicates our desired stellar mass limit of the companion galaxies ($M_{\rm stellar} = 5 \times 10^9 M_\odot$). We find that companion galaxies are highly complete at $z \leq 0.05$.

Table 1: Galaxy Sample Information. Columns: (1) name of the field/survey; (2) the total number of sources in photometry catalog before (after) applying good-source cuts described in Section 1.2.1.1; (3) the redshift range of interest in our study, used to select the mass-limited sample counts in (4,5), where the subsets with spectroscopic redshift information are given in parenthesis.

Name (1)	Phot Sources (2)	Redshift Range (3)	$\log_{10}(M_{\rm stellar}/M_{\odot}) \geq 9.7$ (4)	$\log_{10}(M_{\rm stellar}/M_{\odot}) \geq 10.3$ (5)
UDS	35932 (33998)	$z \in [0.5, 3.0]$	3019 (260)	1223 (141)
GOODS-S	34930 (34115)	$z \in [0.5, 3.0]$	2491 (892)	942 (403)
GOODS-N	35445 (34693)	$z \in [0.5, 3.0]$	2946 (494)	1133 (209)
COSMOS	38671 (36753)	$z \in [0.5, 3.0]$	3232 (11)	1307 (9)
EGS	41457 (37602)	$z \in [0.5, 3.0]$	2825 (199)	1093 (72)
CANDELS (Total)	186,435 (177,161)	$z \in [0.5, 3.0]$	14,513 (1856)	5698 (834)
SDSS-DR4 (1790 sq. deg)	141,564	$z \in [0.03, 0.05]$	9183 (8524)	4098 (3859)

1.2.2.2 $z \sim 0$ Sample Completeness

The [99] sample is magnitude-limited ($r < 17.77$ mag) due to the SDSS spectroscopic target selection; this provides a $\sim 90\%$ $z_{\rm spec}$ completeness [106]. Yang et al. included additional redshifts from supplementary surveys to improve the incompleteness due to spectroscopic fiber collisions [107]. As such, our $z \sim 0$ sample selection has 92.8% $z_{\rm spec}$ completeness (Table 1). Nevertheless, several merger studies demonstrate that the SDSS spectroscopic incompleteness grows with decreasing galaxy-galaxy separation [43, 48]. We account for this issue and provide detailed corrections in § 1.3.2. In addition, we demonstrate the stellar mass completeness of $\log_{10}(M_{\rm stellar}/M_{\odot}) \geq 9.7$ galaxies by employing the method by [108]. In Figure 2, we show the stellar mass distributions for narrow redshift intervals ($\Delta z = 0.006$) at $z \leq 0.05$. We find that the $r < 17.77$ mag limit produces a turnover in counts at different masses as a function of redshift. At $z \leq 0.05$, the mass at which the distributions turn over (become incomplete) is well below our limit of $\log_{10}(M_{\rm stellar}/M_{\odot}) = 9.7$. This indicates that our mass-limited sample is highly complete for selecting possible major companions in a complete sample of $N_{\rm m} = 4098$ massive galaxies with $\log_{10}(M_{\rm stellar}/M_{\odot}) \geq 10.3$ and $0.03 \leq z \leq 0.05$ (see Table 1).

1.2.3 Selection of Massive Galaxies Hosting Major Companions

1.2.3.1 Projected Separation

With our well-defined mass-limited samples for CANDELS and SDSS in hand, we start by identifying the massive ($M_{\rm stellar} \geq 2 \times 10^{10} M_{\odot}$) galaxies hosting a major

projected companion satisfying $1 \leq M_1/M_2 \leq 4$ and a projected physical separation of $5\ \mathrm{kpc} \leq R_{\mathrm{proj}} \leq 50\ \mathrm{kpc}$. The choice of $R_{\mathrm{proj}} \leq 50$ kpc is common in close-pair studies [18, 60, 109] which is supported by the numerical simulation results showing that major bound companions with this separation will merge within $\lesssim 1\,\mathrm{Gyr}$. Additionally, source blending from smaller separations ($\lesssim 1.2\ \mathrm{kpc}$) can cause incompleteness at $z \gtrsim 0.04$ for SDSS and at $z \gtrsim 2.5$ for CANDELS. Thus, we adopt a lower limit of $R_{\mathrm{proj}} = 5$ kpc ($\sim 4\times$ the resolution), which also corresponds to the typical sizes of $\log_{10}(M_{\mathrm{stellar}}/M_{\odot}) \geq 9.7$ galaxies at $2.0 \leq z \leq 2.5$. In summary, we find $N_{\mathrm{proj}} = 318$ and $N_{\mathrm{proj}} = 2451$ unique (duplicate resolved) massive galaxies hosting major projected companions in SDSS ($0.03 \leq z \leq 0.05$) and CANDELS (total of all five fields at $0.5 \leq z \leq 3.0$), respectively. We tabulate the breakdown of N_{proj} by redshift per each CANDELS field in Table 2.

1.2.3.2 Plausible Physical Proximity (SDSS)

We note that projected proximity does not guarantee true physical proximity as foreground and background galaxies can be projected interlopers. A common and effective method to define physical proximity is to isolate systems with a small velocity separation, which indicate that the host and companion galaxies are plausibly gravitationally bound. For the SDSS sample, we employ the common criteria $\Delta v_{12} = |v_1 - v_2| \leq 500\ \mathrm{km\,s^{-1}}$ [5,60,62], where v_1 and v_2 are the velocities of the host and companion galaxies, respectively. Merger simulations find that systems that satisfy $\Delta v_{12} \leq 500\ \mathrm{km\,s^{-1}}$ typically merge within $0.5-1$ Gyr [110]. Other studies show that close-pair systems with

$\Delta v_{12} > 500\ \mathrm{km\,s^{-1}}$ are not likely to be gravitationally bound [64, 111]. However, owing to spectroscopic redshift incompleteness (see § 1.2.2.2), we are only able to apply this velocity selection to a subset of galaxies from § 1.2.3.1 which have spectroscopic redshifts. In doing so, we find $N_{\mathrm{phy}} = 106$ massive galaxies hosting a major projected companion (in § 1.2.3.1) meeting $\Delta v_{12} \leq 500\ \mathrm{km\,s^{-1}}$ criteria in the SDSS ($0.03 \leq z \leq 0.05$) sample. We describe the statistical correction for missing major companions due to spectroscopic incompleteness in § 1.3.2.

1.2.3.3 Plausible Physical Proximity (CANDELS)

Most galaxies in the CANDELS catalogs do not have a spectroscopic redshift. Hence, we use a proximity method based on photometric redshifts and their uncertainties (σ_{z}) to select plausible, physically close companions [6,16,112]. As described in § 1.2.1.2, each galaxy's z_{phot} value is the median of the peak values (z_{peak}) of multiple photometric $P(z)$ distributions computed by the CANDELS team. However, the $P(z)$ data was not thoroughly analyzed to derive z_{phot} errors for all of the CANDELS fields. Thus, we compute σ_{z} values from a single participant $P(z)$ dataset that produces z_{peak} values that are consistent with the published team z_{best} values. This is necessary to achieve z_{phot} errors that are consistent with the z_{best} and stellar masses (calculated with z_{best}) that we use in this study. We find that the S. Wuyts[4] photometric redshifts produced the best match to z_{best} after testing all participant $P(z)$ data. The Wuyts $P(z)$ distributions for each CANDELS galaxy were computed using the photometric redshift code EAZY [113] and

[4]Method 13 as specified in [93].

PÉGASE [102] stellar synthesis template models. We optimize[5] the $P(z)$ for each galaxy and use this distribution to compute the uncertainty ($\sigma_{\rm z}$) defined as the 68% confidence interval of the photometric redshift $z_{\rm phot}$ (see [114] for details).

In Figure 3, we show the photometric redshift uncertainties ($\sigma_{\rm z}$) as a function of $z_{\rm best}$ for each galaxy in our sample ($M_{\rm stellar} \geq 5 \times 10^9 M_{\odot}$). We find that the $\sigma_{\rm z}$ distributions in the CANDELS fields are qualitatively similar to each other. We find the $\sigma_{\rm z}$ distributions have small scatter up to $z \sim 1.5$ with their medians typically ranging between $0.02 \leq \sigma_{\rm z,med} \leq 0.05$, and much larger scatter at $z \gtrsim 1.5$ with the medians ranging between $0.06 \leq \sigma_{\rm z,med} \leq 0.08$. This large scatter is because the observed filters no longer span the 4000Å break, which leads to larger uncertainties during template SED-fitting (for additional details, see [114]). For each CANDELS field, we show the 80% and 95% outlier limits of the redshift normalized error [$\sigma_{\rm z}/(1 + z_{\rm best})$] distribution and present their values in Table 3. While the 95% clipping limit rejects extreme outliers typically with $z_{\rm best} > 1.5$, the 80% limit does a reasonable job representing the upper envelope of the $\sigma_{\rm z}$ distribution at all redshifts. Therefore, to exclude galaxies with large $z_{\rm phot}$ errors, we elect to exclude those $\sigma_{\rm z}$ above the 80% clipping limit. Hereafter, we define the large-error $z_{\rm phot}$ as *unreliable*.

For each galaxy in our CANDELS sample, we adopt the [6] (hereafter, B09) redshift proximity criteria given by :

$$\Delta z_{12}^2 \leq \sigma_{z,1}^2 + \sigma_{z,2}^2 \, , \tag{1.1}$$

[5]We shift the $P(z)$ distributions and raise them to a power such that when compared to the test set of spectroscopic redshifts ($z_{\rm spec}$), the 68% confidence interval of the $P(z)$ should include $z_{\rm spec}$ 68% of the time. A detailed description is given in [114]

where $\Delta z_{12} = (z_{\rm best,1} - z_{\rm best,2})$ is the redshift difference of the host and companion galaxies, and $\sigma_{z,1}$ and $\sigma_{z,2}$ are their photometric redshift errors, respectively. It is important to note that projected pairs containing widely separated galaxies in redshift space that have large $z_{\rm phot}$ errors can satisfy Equation 1.1. Hence, we apply the redshift proximity criteria only to those galaxies with reliable photometric redshifts. In summary, we select $N_{\rm phy} = 504$ massive galaxies hosting major companions satisfying $5\ {\rm kpc} \leq R_{\rm proj} \leq 50$ kpc, $1 \leq M_1/M_2 \leq 4$, and Equation 1.1. We present the breakdown of $N_{\rm phy}$ in each redshift bin per CANDELS field in Table 2.

In § 1.3.1, we describe a statistical correction to add back a subset of galaxies excluded because of an unreliable $\sigma_{\rm z}$ that could be statistically satisfying the redshift proximity criteria. Additionally, owing to the possibility that some companion galaxies may satisfy the close redshift proximity criterion by random chance, we discuss the statistical correction for random chance pairing in § 1.3.3.3. We acknowledge the mismatch between the redshift proximity methods that we employ for the SDSS and CANDELS. In § 1.4, we test the impact of this mismatch and find that it does not significantly impact our results and conclusions.

1.3 Frequency of Major Merging

To track the history of major merging, we start by analyzing the fraction of massive ($M_{\rm stellar} \geq 2 \times 10^{10} M_{\odot}$) galaxies hosting a major companion selected in § 1.2.3. The

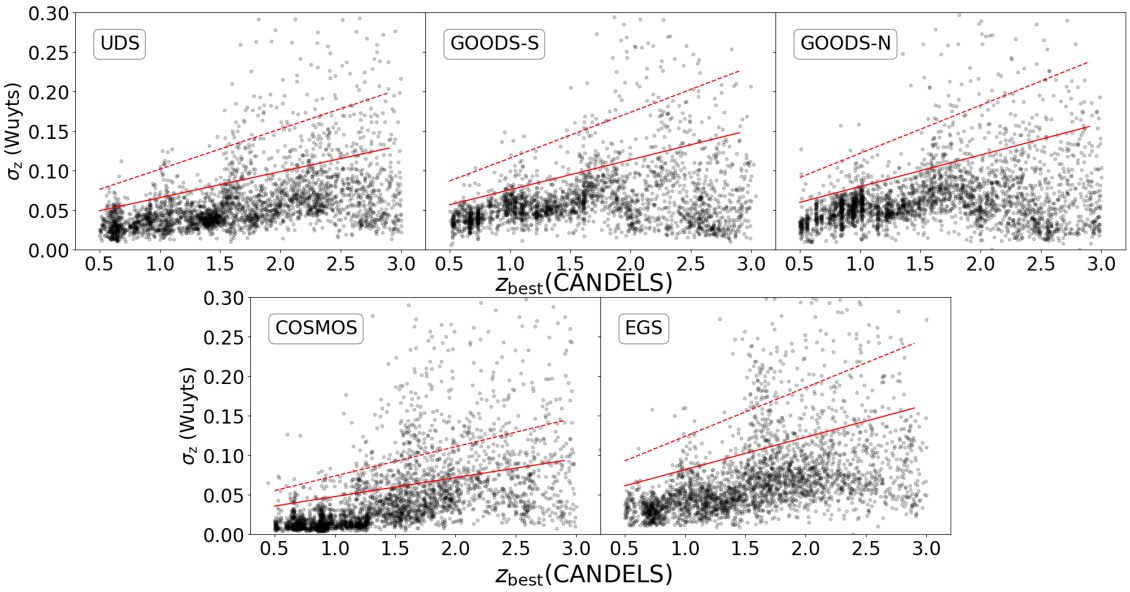

Figure 3: Photometric redshift uncertainties (σ_z) as a function of $z_{\rm best}$ for galaxies with $M_{\rm stellar} \geq 5 \times 10^9 M_\odot$ in each CANDELS field. The σ_z values are the 1σ photometric redshift errors from the optimized $P(z)$ distributions originally derived by S. Wuyts (see text for details). In each panel, we show the 80% and 95% outlier clipped limits of the redshift normalized uncertainty $\sigma_z/(1+z_{\rm best})$ distribution in solid and dashed lines, respectively. We present these limits in Table 3.

major companion fraction[6] is

$$f_{\rm mc}(z) = \frac{N_{\rm mc}(z)}{N_{\rm m}(z)} \, , \tag{1.2}$$

where at each redshift bin, $N_{\rm mc}$ is the number of massive galaxies hosting a major companion after statistically correcting the $N_{\rm phy}$ counts (see § 1.3.1 and 1.3.2), and $N_{\rm m}$ is the number of massive galaxies. The companion fraction $f_{\rm mc}$ is commonly used in the literature, and is the same as $f_{\rm merg}$ used by [16]. We use the samples described in § 1.2.3 to derive $N_{\rm mc}(z)$ separately for the five redshift bins in CANDELS ($0.5 \leq z \leq 3.0$) and for the SDSS $z \sim 0$ anchor. We then discuss the application of a correction to account for galaxies satisfying the companion selection criteria by random chance. Finally, we characterize the companion fraction and use an analytical function to quantify the redshift evolution of $f_{\rm mc}$ during $0 < z < 3$.

1.3.1 Deriving $N_{\rm mc}(z)$ for CANDELS

For redshifts $0.5 \leq z \leq 3$, we compute $N_{\rm mc}$ corrected for incompleteness owing to unreliable photometric redshifts as

$$N_{\rm mc} = N_{\rm phy} + C_1 N_{\rm proj,unreliable} \, , \tag{1.3}$$

where $N_{\rm phy}$ is the number of massive galaxies with reliable $z_{\rm phot}$ values that host a major companion (§ 1.2.3.3), $N_{\rm proj,unreliable}$ is the number of galaxies hosting major projected companions that are excluded because of unreliable $z_{\rm phot}$ values (§ 1.2.3.1), and C_1 is the correction factor used to statistically add back a subset of excluded galaxies that are

[6]While the companion fraction is related to the pair fraction, it is important to be clear that it is not the same (see § 1.5.1).

expected to satisfy the redshift proximity criteria we employ. We estimate C_1 as

$$C_1 = \frac{N_{\rm phy}}{N_{\rm proj} - N_{\rm proj,unreliable}} \, . \tag{1.4}$$

To study $f_{\rm mc}(z)$ from the overall sample and also its field-to-field variations, we calculate $N_{\rm mc}$ for five $\Delta z = 0.5$ bins separately for each CANDELS field and for the total sample using counts tabulated in Table 2. For example, in the $0.5 \leq z \leq 1$ bin, CANDELS contains 671 massive galaxies with reliable $z_{\rm phot}$ values hosting a major projected companion. This results in $C_1 = 193/671 = 0.29$ and a corrected count of $N_{\rm mc} = 230$; i.e., we add back 29% of the previously excluded galaxies at these redshifts. We tabulate C_1 and $N_{\rm mc}$ values in Table 2. We also note no significant difference in the computed $N_{\rm mc}$ values whether we use an 80% or 95% σ_z clipping limit (§ 1.2.3) to remove unreliable photometric redshifts.

1.3.2 Deriving $N_{\rm mc}(z)$ for SDSS

To achieve an accurate low-redshift anchor for the fraction of massive galaxies hosting a major companion meeting our $\Delta v_{12} \leq 500\,{\rm km\ s^{-1}}$ velocity separation criterion, we calculate $N_{\rm mc}$ corrected for the SDSS spectroscopic incompleteness that varies with projected separation using $\Delta R_{\rm proj} = 5\,{\rm kpc}$ bins as follows:

$$N_{\rm mc} = N_{\rm phy} + \sum_{i=1}^{9} \left(C_2 \, N_{\rm proj,nospec} \right)_i \, . \tag{1.5}$$

For each of nine bins between $R_{\rm proj} = 5 - 50\,{\rm kpc}$, we compute a correction factor $C_{2,i}$ necessary to add back a statistical subset of the $N_{\rm proj,nospec}$ galaxies in the bin that lack spectroscopic redshifts but that we expect to satisfy the $\Delta v_{12} \leq 500{\rm km\ s^{-1}}$ criterion. Following the same logic as in Equation 1.4, we estimate this correction at each $\Delta R_{\rm proj}$ bin

based on the counts of spectroscopic galaxies hosting a major projected companion and the fraction that satisfy $\Delta v_{12} \leq 500\,\mathrm{km\,s^{-1}}$. Owing to our well-defined sample volume ($0.03 < z < 0.05$), the total sample of spectroscopic hosts with plausible physical companions (§ 1.2.3.2) is limited to $N_{\rm phy} = 109$ over the nine separation bins. To reduce random errors from small number statistics, we use a larger redshift range ($0.01 \leq z \leq 0.05$) and SDSS footprint ($\sim 4000\,\mathrm{deg^2}$), to calculate the correction factor at each $\Delta R_{\rm proj}$ bin:

$$C_{2,i} = \left(\frac{N'_{\rm phy}}{N'_{\rm proj} - N'_{\rm proj,nospec}} \right)_i . \tag{1.6}$$

In Figure 4, we plot C_2 and the factors in Equation 1.5 as a function of $R_{\rm proj}$ using this larger sample (emphasized with a *prime*). For example, in the $20 \leq R_{\rm proj} \leq 25$ kpc separation bin, we find 119 massive galaxies hosting a major projected companion, of which 35 have small velocity separations and 44 lack spectroscopic redshifts. This results in a 47% correction ($C_2 = 35/75 = 0.47$) at this separation. We find the probability for a small-separation pair ($R_{\rm proj} = 5$ kpc) to satisfy $\Delta v_{12} \leq 500$ km s^{-1} (despite lacking the redshift information) is 85% ($C_2 = 0.85$) and this rapidly decreases to $\sim 30\%$ ($C_2 \sim 0.3$) at $R_{\rm proj} = 30$ kpc, and remains statistically constant between $R_{\rm proj} = 30 - 50$ kpc. This correction is important since the spectroscopic incompleteness $N'_{\rm proj,nospec}/N'_{\rm proj}$ ranges from >0.6 ($\sim 5 - 10$ kpc) to 0.2 ($\sim 45 - 50$ kpc) over the separations we probe, which is in agreement with trends published in Figure 2 from [48].

Table 2: Detailed breakdown of variables involved in estimating the $f_{\rm mc}$ and $f_{\rm mc,c}$ at five redshift bins between $0.5 \leq z \leq 3$. Columns: (1) name of the CANDELS field; (2) the CANDELS team $z_{\rm best}$ bin; (3) number count of massive ($\log M_{\rm stellar}/M_{\odot} \geq 10.3$) galaxies; (4) number of massive galaxies hosting a major projected companion (§ 1.2.3.1), those of which that have unreliable photometric redshift values are shown in parenthesis; (5) number of massive galaxies with reliable $z_{\rm phot}$ that host a major projected companion satisfying redshift proximity (Equation 1.8) as described in § 1.2.3.3; (6) the correction factor computed using Equation 1.4; (7) the number of massive galaxies hosting a major companion after statistically correcting $N_{\rm phy}$ for incompleteness owing to unreliable $z_{\rm phot}$ values from Equation 1.3; (8) correction factor to account for random chance pairing as described in § 1.3.3.3; (9) the fraction of massive galaxies hosting a major companion (*major companion* fraction); (10) random chance pairing corrected $f_{\rm mc}$, as described in § 1.3.3.3.

Name	Redshift	$N_{\rm m}$	$N_{\rm proj}$	$N_{\rm phy}$	C_1	$N_{\rm mc}$	C_3	$f_{\rm mc}$ (%)	$f_{\rm mc,c}$ (%)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	$0.5 \leq z \leq 1$	256	132 (16)	44	0.38	50		20 ± 5	
	$1 \leq z \leq 1.5$	304	130 (23)	34	0.32	41		14 ± 4	
UDS	$1.5 \leq z \leq 2$	290	130 (30)	25	0.25	33		11 ± 4	
	$2 \leq z \leq 2.5$	216	80 (31)	14	0.29	23		11 ± 4	
	$2.5 \leq z \leq 3$	157	82 (42)	3	0.07	6		4 ± 3	
	$0.5 \leq z \leq 1$	216	107 (21)	32	0.37	40		18 ± 5	
	$1 \leq z \leq 1.5$	252	116 (14)	34	0.33	39		15 ± 4	
GOODS-S	$1.5 \leq z \leq 2$	213	87 (22)	15	0.23	20		9 ± 4	
	$2 \leq z \leq 2.5$	138	56 (14)	3	0.07	4		3 ± 3	
	$2.5 \leq z \leq 3$	123	57 (15)	5	0.12	7		6 ± 4	
	$0.5 \leq z \leq 1$	333	195 (33)	56	0.35	67		20 ± 4	
	$1 \leq z \leq 1.5$	278	140 (27)	40	0.35	50		18 ± 4	

GOODS-N	$1.5 \leq z \leq 2$	209	99 (20)	15	0.19	19		9 ± 4	
	$2 \leq z \leq 2.5$	191	83 (17)	6	0.09	8		4 ± 3	
	$2.5 \leq z \leq 3$	122	61 (16)	5	0.11	7		6 ± 4	
	$0.5 \leq z \leq 1$	448	244 (38)	40	0.19	47		11 ± 3	
	$1 \leq z \leq 1.5$	270	128 (37)	1	0.01	1		1 ± 1	
COSMOS	$1.5 \leq z \leq 2$	350	157 (75)	15	0.18	29		8 ± 3	
	$2 \leq z \leq 2.5$	153	86 (54)	0	0.0	0		0 ± 0	
	$2.5 \leq z \leq 3$	86	36 (22)	6	0.43	15		18 ± 8	
	$0.5 \leq z \leq 1$	224	120 (19)	21	0.21	25		11 ± 4	
	$1 \leq z \leq 1.5$	304	137 (19)	36	0.31	42		14 ± 4	
EGS	$1.5 \leq z \leq 2$	331	171 (50)	39	0.32	55		17 ± 4	
	$2 \leq z \leq 2.5$	167	94 (18)	23	0.3	28		17 ± 6	
	$2.5 \leq z \leq 3$	67	35 (13)	2	0.09	3		5 ± 5	
	$0.5 \leq z \leq 1$	1477	798 (127)	193	0.29	230	0.15	16 ± 2	13 ± 2
	$1 \leq z \leq 1.5$	1408	651 (120)	145	0.27	173	0.17	12 ± 2	10 ± 2
All fields	$1.5 \leq z \leq 2$	1393	644 (197)	109	0.24	155	0.18	11 ± 2	9 ± 2
	$2 \leq z \leq 2.5$	865	399 (134)	46	0.17	63	0.25	7 ± 2	5 ± 2
	$2.5 \leq z \leq 3$	555	271 (108)	21	0.13	38	0.22	7 ± 2	5 ± 2

1.3.3 Redshift Evolution of Major Merging Frequency

We use the corrected counts of massive galaxies hosting a major companion to compute the companion fraction ($f_{\rm mc}$; see Equation 1.2) in the SDSS and CANDELS. We

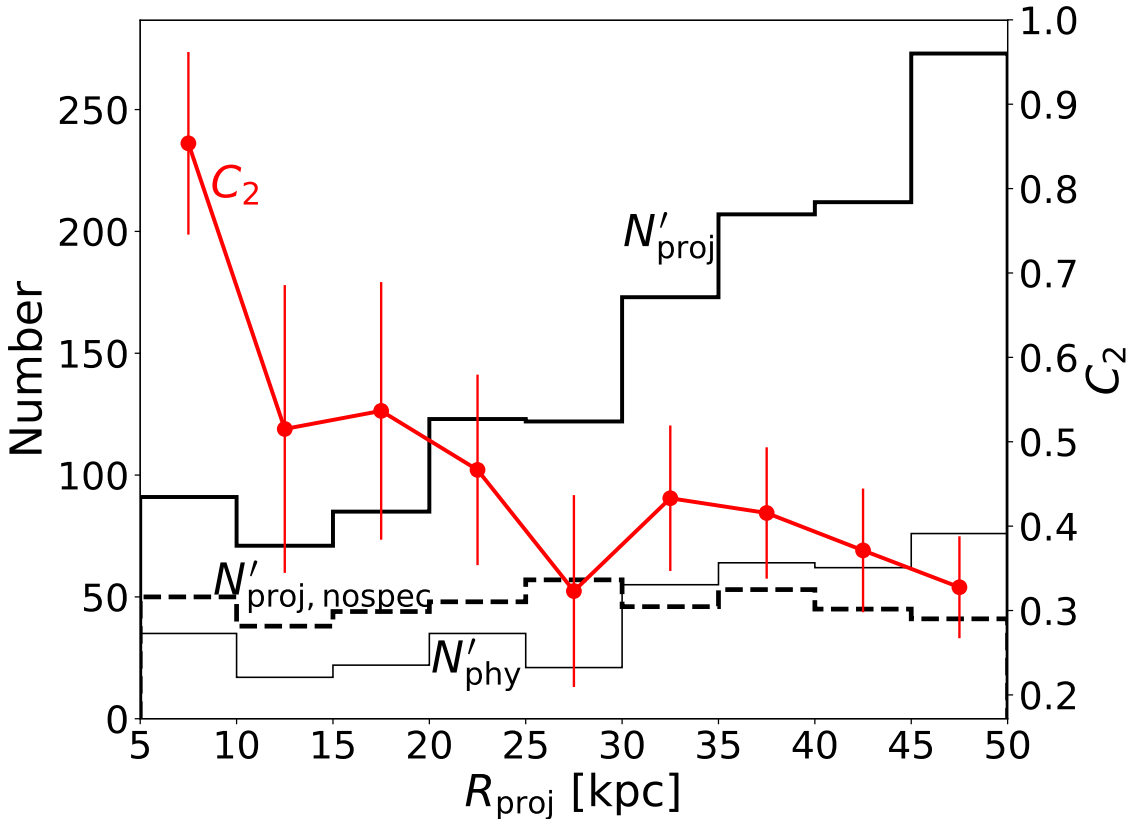

Figure 4: The number of massive galaxies hosting a major companion in projected pairs as a function of projected separation from the SDSS: total ($N'_{\rm proj}$; bold line), the subset of galaxies with no spectroscopic redshift information ($N'_{\rm proj,nospec}$; dashed line), and the subset of galaxies satisfying $\Delta v_{12} \leq 500\,{\rm km\,s^{-1}}$ ($N'_{\rm phy}$; thin line). The *prime* signifies that the quantities are derived using a larger SDSS sample spanning $0.01 \leq z \leq 0.05$ and $\sim 4000\,{\rm deg^2}$ (see § 1.3.2) for nine $R_{\rm proj}$ bins between $5 - 50\,{\rm kpc}$. The correction factor C_2 per $R_{\rm proj}$ bin (Equation 1.6) is given by the red circles connected by a red solid line. The error bars represent 95% binomial confidence of C_2.

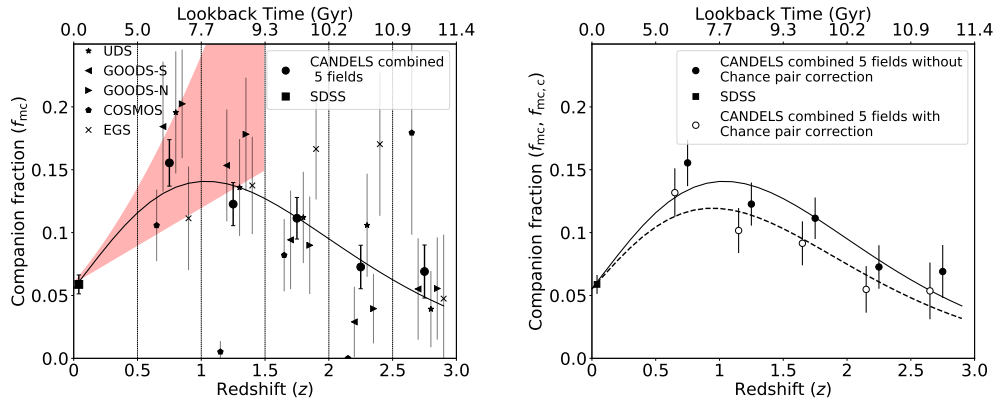

Figure 5: *Left-panel*: The redshift evolution of the *major companion* fraction $f_{\rm mc}$ shown for the five CANDELS fields UDS (star), GOODS-S (left triangle), GOODS-N (right triangle), COSMOS (pentagon), EGS (cross). The combined CANDELS fractions in five redshift bins (circles) and the SDSS low-redshift anchor (square) include 95% binomial confidence limit error bars. To place our finding in the context of common, close-pair-based evolutionary trends found in the literature, we plot the shaded region (red) encompassing a common range of power-law slopes $f_{\rm mc} = 0.06(1+z)^{1-2}$ at $0 < z < 1.5$. *Right-panel*: The random chance corrected fractions ($f_{\rm mc,c}$) for the five CANDELS Δz bins (open circles) are compared with the $f_{\rm mc}(z)$ from (a). For $f_{\rm mc,c}$, the binomial errors and scatter of C_3 (see S 1.3.3.3) are added in quadrature. Best-fit curves to the companion fraction ($f_{\rm mc}$) evolution data (see Equation 1.7 and § 1.3.3.2 for details) are shown in solid ($f_{\rm mc}$) and dashed ($f_{\rm mc,c}$) lines, respectively. In the case of SDSS, since the correction $C_3 \sim 0.01$, we only plot $f_{\rm mc}$ for simplicity. From this figure, we conclude that the major companion fraction increases strongly from $z \sim 0$ to $z \sim 1$, and decreases steeply towards $z \sim 3$ (see text for details).

Table 3: Photometric redshift uncertainty outlier limits that are used to determine reliable $z_{\rm phot}$ values for each CANDELS field. Columns: (1) name of the CANDELS field; (2,3) the 80% and 95% outlier clipped limits of the redshift normalized uncertainity $\sigma_{\rm z}/(1+z_{\rm best})$ distributions for galaxies in the mass-limited ($M_{\rm stellar} \geq 5\times10^{9}M_{\odot}$ sample for redshifts $0.5 \leq z \leq 3.0$ as shown in Figure 3.

Name (1)	80% limit (2)	95% limit (3)
UDS	0.033	0.051
GOODS-S	0.038	0.058
GOODS-N	0.04	0.061
COSMOS	0.024	0.037
EGS	0.041	0.062

compare field-to-field variations of $f_{\rm mc}$ in CANDELS, quantify the redshift evolution of the $f_{\rm mc}$ from $z = 3$ to $z = 0$, and measure the impact of random chance pairing on the observed major companion fraction evolutionary trends.

1.3.3.1 Field-to-Field Variations

In Figure 5(a), we plot $f_{\rm mc}(z)$ for the combined CANDELS fields and compare this with the individual fractions from each field at each redshift; these are also tabulated in Table 2. Despite noticeable variations between the fractions derived from each CANDELS field owing to small-number statistics, we find fair agreement between multiple fields at each redshift. We note that the combined CANDELS sample and three individual fields (UDS, GOODS-S, and GOODS-N) show consistent trends with the highest merger fractions at $z \sim 1$, which then steadily decrease with increasing redshift. The EGS and COSMOS companion fractions exhibit different behavior with redshift, the former peaks at $z \sim 2$ while the latter has no trend with redshift owing to a lack of galaxies hosting major companions in two different redshift bins. We compute the cosmic-variance

($\sigma_{\rm CV}$) on the combined CANDELS $f_{\rm mc}(z)$ values using the prescription by [115]. For $\log_{10}(M_{\rm stellar}/M_{\odot}) \gtrsim 10$ galaxies at $0.5 < z < 3.0$ populating the five CANDELS fields each with an area of $160\,{\rm arcmin}^2$ such that their cumulative area matches that of the total CANDELS coverage ($5 \times 160 = 800\,{\rm arcmin}^2$), we find that the number counts of galaxies hosting major companions have $\sigma_{\rm CV}$ ranging from 11% ($z = 0.75$) to 18% ($z = 2.75$)[7]. While most of the individual CANDELS-field fractions are consistent with the $\sigma_{\rm CV}$ within their large uncertainties (owing to small sample size), few $f_{\rm mc}$ values (e.g., at $z > 1.5$ for the COSMOS and EGS fields; see Figure 5a) are significantly above the possible cosmic-variance limits.

1.3.3.2 Analytical Fit to the Major Companion Fraction Evolution

To characterize the redshift evolution of the companion fraction during $0 < z < 3$, we anchor the combined CANDELS $f_{\rm mc}(z)$ measurements to the SDSS-derived data point at $z \sim 0$. As shown in Figure 5(a), the low-redshift fraction is $\sim 3\times$ lower than the maximum $f_{\rm mc} \sim 0.16$ value at $0.5 < z < 1$, which then decreases to $f_{\rm mc} \sim 0.07$ at $2.5 < z < 3$. This suggests a turnover in the incidence of merging sometime around $z \sim 1$, in agreement with some previous studies [116, 117]. Previous close-pair-based studies at $z \sim 0$ find fractions $f_{\rm mc} \sim 2\% \pm 0.5\%$ [60, 111, 118], but they used criteria that are different from our fiducial selection. Similarly, many empirical, close-pair-based studies in the literature broadly agree that $f_{\rm mc}$ rises at $0 < z < 1.5$ but with a range of

[7]We take into account that $\sigma_{\rm CV}$ is smaller for multiple, widely separated fields when compared to the $\sigma_{\rm CV}$ of a single contiguous field. For additional details, see [115]

evolutionary forms $(1+z)^{1\sim2}$ owing to varying companion selection criteria (for discussion, see [18]). After normalizing for these variations, we note that our SDSS-based $f_{\rm mc}$ is in good agreement with previous close-pair-based estimates, and our rising trend (see shaded region, Figure 5a) between $0 < z < 1.5$ is in broad agreement with other empirical trends. In § 1.5, we will present detailed comparisons to other studies by re-computing $f_{\rm mc}$ based on different companion selection choices that match closely with others.

All studies that measured redshift evolution of merger frequency at $0 < z < 1.5$, irrespective of the methodology, have used the power-law analytical form $f(z) \propto (1+z)^{\rm m}$ to represent the best-fit of $f(z)$. This functional form cannot be used to represent the observed rising and then decreasing $f_{\rm mc}(z)$ for redshift ranges $0 < z < 3$. Therefore, following [117] and initially motivated by [119], we use a modified power-law exponential function given by

$$f_{\rm mc}(z) = \alpha(1+z)^{\rm m} \exp^{\beta(1+{\rm z})} \ . \tag{1.7}$$

As demonstrated in Figure 5(a), this analytic function provides a good fit to the observed evolution. The best-fit curve to the fractions derived from the SDSS and CANDELS measurements from our fiducial companion selection criteria has parameters $\alpha = 0.5 \pm 0.2$, $m = 4.5 \pm 0.8$, and $\beta = -2.2 \pm 0.4$. We note that we will apply this fitting function for different companion selection choices throughout our comparative analysis described in § 1.4.

1.3.3.3 Correction for Random Chance Pairing

Finally, we note that a subset of massive galaxies hosting a major companion ($N_{\rm mc}$) can satisfy the companion selection criteria by random chance. To account for this contamination, we apply a statistical correction and recompute the counts for the combined CANDELS sample per redshift bin as $N_{\rm mc,c} = N_{\rm mc}(1 - C_3)$ in each redshift bin. To compute C_3, we generate 100 simulated Monte-Carlo (MC) randomized datasets[8]. We define C_3 in each redshift bin as the ratio of massive galaxy number counts hosting major companions which satisfy our projected separation and photometric redshift proximity criteria in the MC datasets (by random chance) to the measured $N_{\rm mc}$ (1.3.1 and § 1.3.2). For example, in redshift bin $1 < z < 1.5$, we find that 17% of $N_{\rm mc} = 173$ galaxies statistically satisfy the companion selection criteria by random chance. We tabulate C_3 values at each redshift for CANDELS in Table 2. We repeat this process for the SDSS ($0.03 < z < 0.05$) and find a very small correction of $\sim 1\%$ ($C_3 \sim 0.01$). This demonstrates the very low probability for two SDSS galaxies to satisfy both the close projected separation and stringent spectroscopic redshift proximity ($\Delta v_{12} \leq 500$ km s^{-1}) criteria.

In Figure 5(b), we compare the random chance corrected fractions $f_{\rm mc,c} = N_{\rm mc,c}/N_{\rm m}$ at each redshift bin from CANDELS, to the uncorrected $f_{\rm mc}$ values copied from the left panel. Owing to the insignificant 1% correction at $z \sim 0$, we anchor both the corrected and uncorrected fits to the same SDSS data point. We find that $f_{\rm mc,c}(z)$ follows the same evolutionary trend as $f_{\rm mc}(z)$, in which the best-fit curve rises to a maximum fraction at $z \sim 1$ and then steadily decreases to $z = 3$. At all redshifts, $f_{\rm mc,c}$ is within the statistical

[8]We generate these datasets by randomizing the positions of each galaxy in the $\log_{10}(M_{\rm stellar}/M_{\odot}) \geq 9.7$ mass-limited sample and repeating the selection process in § 1.2.3

errors of $f_{\rm mc}$. The qualitatively similar redshift evolutionary trends of $f_{\rm mc}(z)$ and $f_{\rm mc,c}(z)$ is due to the nearly redshift independent amount of statistical correction for random pairing ($|\Delta f|/f_{\rm mc} \sim 20\%$) at $1 < z < 3$. This is because of the nearly invariant angular scale in this redshift range, which results in similar random chance pairing probabilities. We note that some previous close-pair-based studies have applied this random chance correction [5, 6], while others have not [63].

1.4 Impact of Close-Companion Selection Criteria on Empirical Major Companion Fractions

So far, we have discussed the derivation and redshift evolution of the major companion fraction $f_{\rm mc}$ based on our *fiducial* selection criteria described in § 1.2.3. As illustrated in Table 5, previous studies have employed a variety of criteria to select companions. In this section, we study the impact of different companion selections on $f_{\rm mc}(z)$ derived from our sample. We systematically vary each criterion (projected separation, redshift proximity, and stellar mass ratio versus flux ratio) individually, while holding the other criteria fixed to their fiducial values. Then, we compare each recomputed $f_{\rm mc}(z)$ to the fiducial result from Figure 5. In § 1.3.3.3, we show that applying a statistical correction for random chance pairing will produce companion fractions that are $\sim 20\%$ lower at each redshift interval between $0.5 \leq z \leq 3$. Therefore, we focus the following comparative analysis on uncorrected fractions and note that our qualitative conclusions are robust to whether or not we apply this correction.

1.4.1 Projected Separation

To quantify the impact that changing *only* the criterion for companion projected separation will have on $f_{\rm mc}(z)$, we compare fractions based on the fiducial $R_{\rm proj} = 5 - 50$ kpc selection to those derived from two common criteria: $R_{\rm proj} = 14 - 43$ kpc [16,62] and $R_{\rm proj} = 5 - 30$ kpc [60, 63]. For each case, we hold all other companion selection criteria fixed such that we are strictly comparing fractions of major (stellar mass-selected) companions in close redshift proximity that are found within projected annular areas of $2/3$ and $1/3$ the fiducial selection window (2475π kpc^2), respectively.

In Figure 6, we plot the fiducial and non-fiducial $f_{\rm mc}(z)$ data and their best-fit curves. The key result from this figure is the observed redshift evolution of major companion fractions shown in Figure 5 are qualitatively robust to changes in the projected separation criterion. For each case, we find the fractions increase from low redshift to $z \sim 1$, then turn over and decrease fairly steadily to $z = 3$. Quantitatively, the non-fiducial fractions are well fit by the same $f_{\rm mc}(z)$ functional form as the fiducial results (see Table 4). Owing to the size of the best-fit confidence intervals, the redshifts of the peak fractions are statistically equivalent. Larger samples of massive galaxy-galaxy pairs at redshifts $0.5 < z < 1.5$ are needed to place stringent constraints on the peak or turnover redshift.

Besides the overall $f_{\rm mc}$ trends with redshift, we find that both smaller projected separation criteria select smaller $f_{\rm mc}$ values than the fiducial selection does, as we expect. Despite being a factor of two smaller in projected area, the conservative $R_{\rm proj} = 5 - 30$ kpc criterion coincidentally produces fractions that are statistically matched to those

from the larger $R_{\rm proj} = 14-43\,{\rm kpc}$ selection criterion at most redshift intervals. Only the $0.5 < z < 1$ bin has unequal companion fractions between the two non-fiducial criteria. The coincidental finding of similar fractions using different projected separation criteria is consistent with an increased probability of physical companions at smaller projected separations [7, 8].

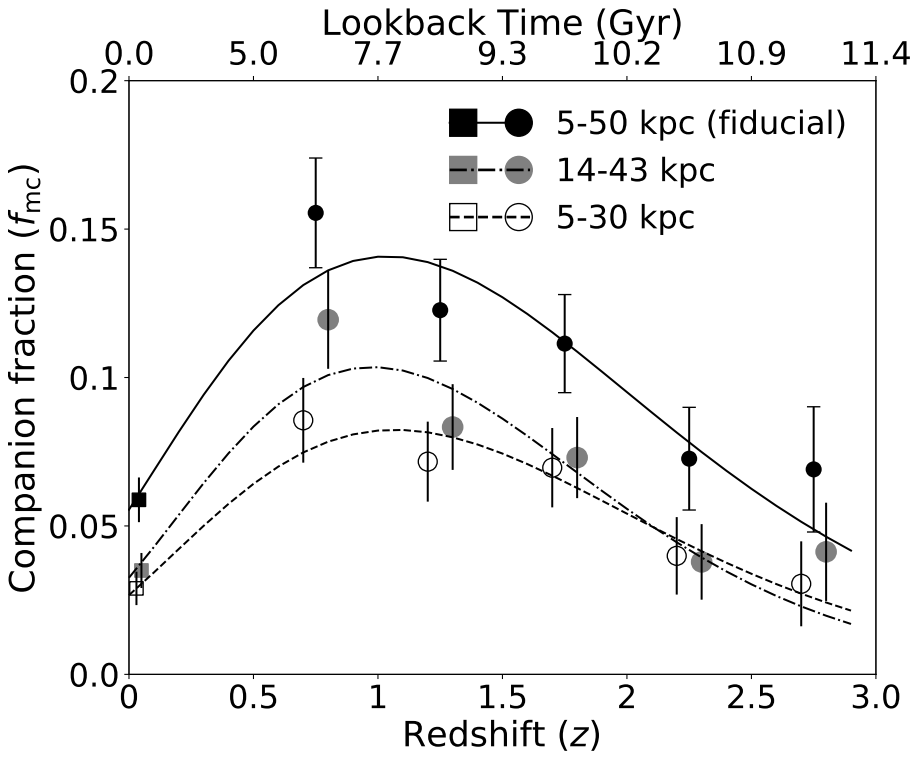

Figure 6: Comparison of the redshift evolution of the major companion fractions based on three projected separation criteria: $R_{\rm proj} = 5-50\,{\rm kpc}$ (fiducial; black symbols, solid line), $R_{\rm proj} = 14-43\,{\rm kpc}$ (grey symbols, dashed line), and $R_{\rm proj} = 5-25\,{\rm kpc}$ (open symbols, dot-dashed line). Best-fit curves for each $f_{\rm mc}(z)$ spanning $0 < z < 3$ are a modified power-law exponential (Equation 1.7; see § 1.3.3.2 for details). The low-redshift fractions are from the SDSS (squares) and the $z > 0.5$ fractions are from CANDELS (circles).The error bar on each $f_{\rm mc}$ data point represents 95% binomial confidence limit. The non-fiducial $f_{\rm mc}$ data points are offset by a small amount within each redshift bin for clarity.

Table 4: Modified power-law exponential function parameter values for best-fit models to major companion fraction evolution $f_{\rm mc}(z)$ based on different selection criteria; our fiducial criteria are given in bold. Columns: (1) major companion definition – stellar mass ratio $1 \leq M_1/M_2 \leq 4$ (MR), or H-band flux ratio $1 \leq F_1/F_2 \leq 4$ (FR); (2) projected separation $R_{\rm proj}$ criterion; (3) redshift proximity method applied to CANDELS samples – Equation 1.7 adopted from B09, Equation 1.8 modified from [63], or 'Hybrid' version of B09 (see text for details); and (4–6) best-fit parameter values (α, m and β) and their 1σ confidence limits for the model function given in Equation 1.7.

Major companion definition	$R_{\rm proj}$	Redshift Proximity	α	m	β
(1)	(2)	(3)	(4)	(5)	(6)
MR	$\mathbf{5-50}$ **kpc**	**B09**	$\mathbf{0.5 \pm 0.2}$	$\mathbf{4.6 \pm 0.9}$	$\mathbf{-2.3 \pm 0.5}$
MR	$14-43$ kpc	B09	0.7 ± 0.5	6.2 ± 1.4	-3.1 ± 0.8
MR	$5-30$ kpc	B09	0.4 ± 0.1	5.4 ± 0.8	-2.6 ± 0.4
MR	$5-50$ kpc	modM12	0.5 ± 0.2	4.6 ± 0.9	-2.2 ± 0.5
MR	$5-50$ kpc	hybB09	0.4 ± 0.1	3.8 ± 0.7	-1.9 ± 0.4
FR	$5-50$ kpc	B09	0.1 ± 0.1	2.1 ± 1.2	-0.5 ± 0.6

1.4.2 Redshift Proximity

Following a similar methodology as in § 1.4.1, we quantify the impact that changing *only* the criterion for redshift proximity will have on $f_{\rm mc}(z)$. As described in § 1.2.3.2, we apply selections based on photometric redshifts and their uncertainties only to CANDELS data, and we anchor our evolutionary findings to stringent velocity-separation-based companion fractions at $z \sim 0$ from the SDSS. As such, we first compare CANDELS fractions ($0.5 < z < 3$) based on our fiducial B09 selection given in Equation 1.1 to fractions based on two other related methods. Then we demonstrate that the low-redshift fraction from the SDSS changes very little ($\Delta f / f_{\rm fid} < 5\%$) if we select major companions using simulated photometric redshift errors that are similar in quality to the CANDELS σ_z values, rather than use our fiducial selection criterion ($\Delta v_{12} \leq 500\,{\rm km\,s^{-1}}$).

In Figure 7, we plot $f_{\rm mc}(z)$ data and best-fit curves for our fiducial selection and two additional methods. In each non-fiducial case, we hold the projected separation and mass ratio selection criteria fixed to our fiducial choices. The first method is a simple overlapping of the host and companion galaxy 1σ photometric redshift errors

$$|\Delta z_{12}| \leq \sigma_{z,1} + \sigma_{z,2}\,. \tag{1.8}$$

This criterion is modified from [63] who used redshift overlap at the 3σ level. As with fiducial selection, we apply this criterion only to those galaxies with reliable photometric redshifts (see § 1.2.3.2). Second, we incorporate available spectroscopic redshifts for the host and/or companion galaxies from CANDELS into a 'hybrid' of our fiducial B09 redshift proximity selection as follows:

a) if both galaxies have $z_{\rm spec}$ data, then the pair must satisfy $\Delta v_{12} \leq 500\,{\rm km\,s^{-1}}$;

b) or if only one galaxy has $z_{\rm spec}$ data, then $z_{\rm spec}$ must be within $z_{\rm phot} \pm \sigma_{\rm z}$ limit of the other galaxy;

c) otherwise use B09 criterion.

As with changing the project separation for selecting close companions, the key takeaway from Figure 7 is the redshift evolution trends for $f_{\rm mc}(z)$ are robust to slight modifications in the redshift proximity criterion. Both non-fiducial companion fractions peak in the lowest-redshift CANDELS bin, the decline at $z > 1$ and their redshift evolution anchored to the SDSS $z \sim 0$ fraction are well fit by the same functional form with statistically equivalent best-fit parameters (see Table 4). Besides the similar evolutionary trends, we notice that the modified Man et al. criterion results in $f_{\rm mc}$ values that are systematically $\sim 11 - 24\%$ higher than the fiducial fractions between $0.5 < z < 3$. This is expected since simple photometric redshift error overlap (Equation 1.1) is less stringent than a quadrature overlap; e.g., two galaxies with $\sigma_z = 0.06$ must have photometric redshifts within $\Delta z_{12} = 0.12$ compared to 0.085 (fiducial). On the other hand, the hybrid B09 method yields $9 - 17\%$ smaller major companion fractions than fiducial at $0.5 < z < 1.5$ owing to the availability of $z_{\rm spec}$ data for a fair fraction of CANDELS galaxies and the stringent Δv_{12} criterion for this subset, but nearly identical fractions at $z > 1.5$ for which the spectroscopic-redshift coverage becomes quite sparse.

We close our redshift proximity analysis by quantifying changes in the $f_{\rm mc}(0)$ measurement from the SDSS data if we employ either the B09 or modified Man et al. photometric redshift error overlap criteria. We opt to generate simulated $z_{\rm phot}$ and $\sigma_{\rm z}$

for the SDSS sample such that these values mimic the quality of the CANDELS photometric data. First, we construct the combined $\sigma_{\rm z}/(1+z_{\rm phot})$ distribution from all five CANDELS fields (see Fig. 3), apply a 3σ outlier clipping, and fit a normalized probability density function (PDF) to this distribution such that the area under the curve sums up to unity. Note that this PDF is the probability for each $\sigma_{\rm z}/(1+z_{\rm phot})$ value and is different from $P(z)$ discussed in § 1.2. For each galaxy, we set its SDSS redshift to $z_{\rm phot}$ and we assign it a $\sigma_{\rm z}$ value drawn randomly from the PDF distribution. Then, we repeat our close companion selections and find $f_{\rm mc}(0) = 0.057$ (B09) and 0.061 (modified Man et al.). The relative change in each fraction is roughly 3% compared to our fiducial selection of close velocity separation.

One can argue that actual $z_{\rm phot}$ data from the SDSS $ugriz$ photometry will produce larger redshift errors and higher companion fractions. Yet, the motivation of this exercise is to quantify the difference between tight spectroscopic velocity separation versus photometric redshift proximity selection matched to the $z > 0.5$ data, not to make our $z \sim 0$ companion fraction more uncertain. We note that if we repeat this analysis by resampling each SDSS galaxy redshift from a PDF defined by its simulated $\sigma_{\rm z}$, we find *smaller* major companion fractions since this effectively makes the redshift proximity worse (larger) and we demonstrated that companions have high probability of small velocity separations for the SDSS. Either way, all the SDSS $f_{\rm mc}$ anchors are significantly lower than the companion fractions at $z \sim 1$, which means the turnover trend in $f_{\rm mc}(z)$ is a robust result.

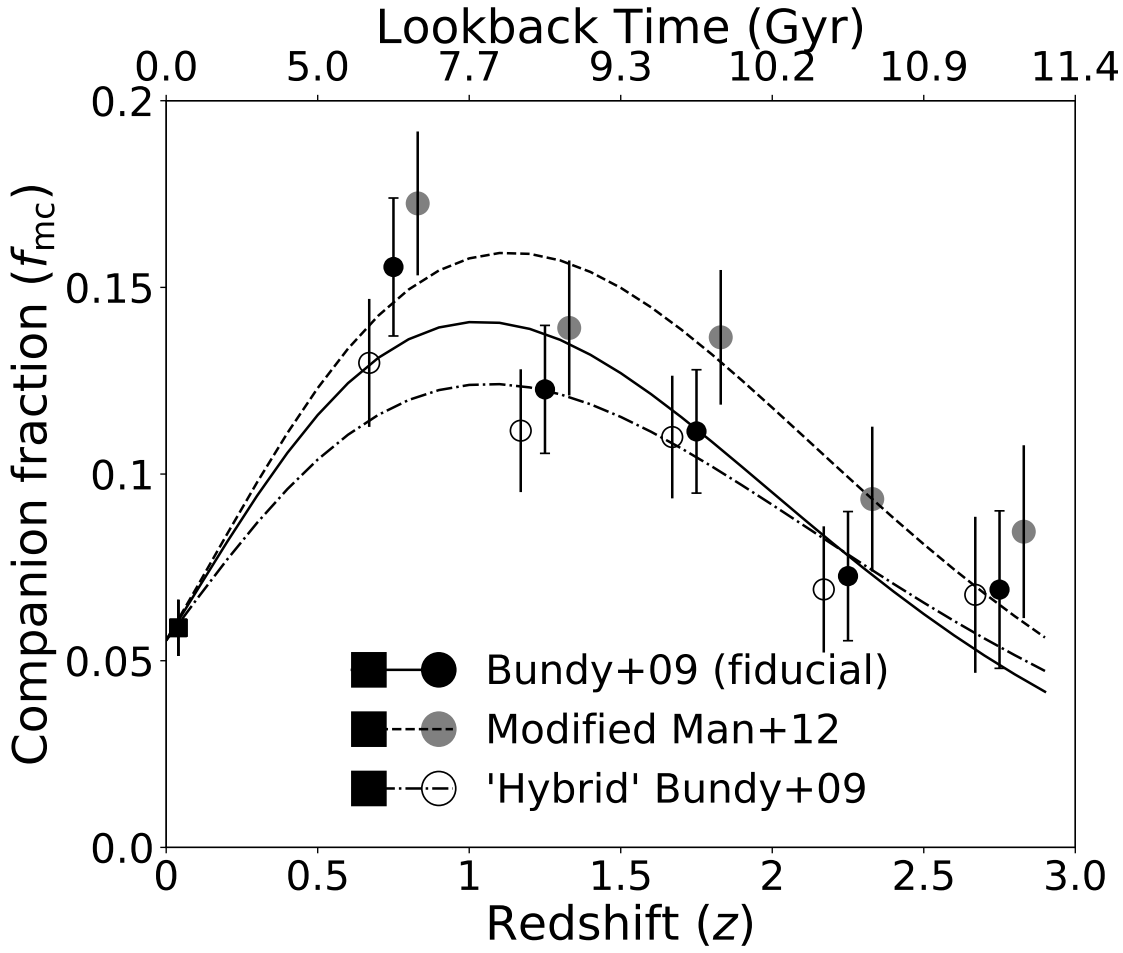

Figure 7: Comparison of the redshift evolution of the major companion fractions $f_{\rm mc}(z)$ based on three redshift proximity criteria applied to $z > 0.5$ galaxies from CANDELS: our fiducial (Equation 1.1; black circles, solid line) adopted from B09; Equation 1.8 (grey circles, dashed line) modified from [63]; and a hybrid of the B09 criterion plus close velocity separation for the subset of CANDELS galaxies with spectroscopic redshifts (open circles, dot-dashed line). All three best-fit curves (Equation 1.7) to the $z > 0.5$ data are anchored to the same $f_{\rm mc}(0)$ point (square symbol) based on SDSS companions with $\Delta v_{12} \leq 500$ km s^{-1} velocity separation. The error bars are defined as in Figure 6.

1.4.3 Stellar Mass vs Flux Ratios

Our fiducial choice for selecting major companions is the stellar mass ratio criterion: $1 \leq M_1/M_2 \leq 4$. Some previous studies, in which galaxies lack stellar mass estimates, have used observed-band flux ratios of $1 \leq F_1/F_2 \leq 4$ as a proxy for selecting major companions [7, 14]. [120] and B09 speculated that flux-ratio selection could preferentially lead to inflated companion counts especially at $z \gtrsim 2$ owing to increasing star formation activity. Two recent studies found conflicting trends in $f_{\rm mc}(z)$ estimates between $z \gtrsim 1$ and $z = 3$ [16, 63]. Here we quantify the impact that changing the major companion selection from 4:1 stellar mass ratio (MR) to 4:1 flux ratio (FR) has on companion fractions for massive galaxies in CANDELS and the SDSS while holding all other selection criteria fixed to fiducial values.

In Figure 8, we compare the fiducial MR-based $f_{\rm mc}(z)$ to fractions based on H-band (CANDELS) and r-band (SDSS) 4:1 flux ratios. We find the FR-based major companion fractions follow starkly different evolutionary trends compared to MR-based fractions at $z \gtrsim 1$. At $0 < z < 1$, the FR and MR fractions both increase with redshift and agree within their 95% confidence limits. However, FR produces increasingly larger fractions from $z \sim 1$ to $z = 3$ that diverge away from the fiducial $f_{\rm mc}(z)$ trends and grow $1.5 - 4.5$ times greater than MR-based fractions at these redshifts. We attempt to fit the FR-based $f_{\rm mc}(z)$ with the same function form (Eq. 1.7) that we employ for MR fraction redshift evolution, but the best-fit parameter values (see Table 4) for the steadily increasing fractions between $0 < z < 3$ are statistically consistent with a simple power law since the FR fractions do not peak nor turn over at $z > 1$. Therefore, we refit the FR fraction

evolution with a power law and find $f_{\rm mc}(z) = 0.07(1+z)^1$ for $0 < z < 3$.

To better understand the striking differences between the FR and MR $f_{\rm mc}(z)$ trends, we analyze stellar mass ratios and flux ratios of CANDELS and SDSS close-pair systems that satisfy our fiducial $5 - 50$ kpc separation and redshift proximity selections (see § 1.2.3.2) in Figure 9. At $z < 1$, we find good agreement between MR and FR values for a majority of major and *minor* ($M_1/M_2 > 4$) close-pair systems. The agreement suggests that r-band (H-band) flux ratios are a good approximation for stellar-mass ratios at $z \sim 0$ ($0.5 < z < 1$). In detail, 80% (72%) of SDSS (CANDELS) pairs are considered major pairs according to *both* FR and MR criteria. We find 17% (25%) of FR-based major pairs have a $> 4:1$ (minor) stellar mass ratio, and only 3% of MR-selected major pairs have minor flux ratios at $z < 1$.

In the bottom four panels of Figure 9, we notice that the FR-based major companion selection suffers a steadily increasing contamination by minor companions according to CANDELS stellar mass ratios. This contamination rises from 40% at $1 < z < 1.5$ to over three-quarters in the highest redshift interval. In contrast, our fiducial selection of major companions maintains a very small ($\sim 5\%$) constant contamination of $F_1/F_2 > 4$ pairs between $1 < z < 3$. This analysis clearly demonstrates that FR selection results in inflated major companion counts at $z \gtrsim 1$, confirming the result that significantly different mass-to-light ratio properties of the companion galaxies at $z > 1$ may be the main cause of the contamination by minor companions [16].

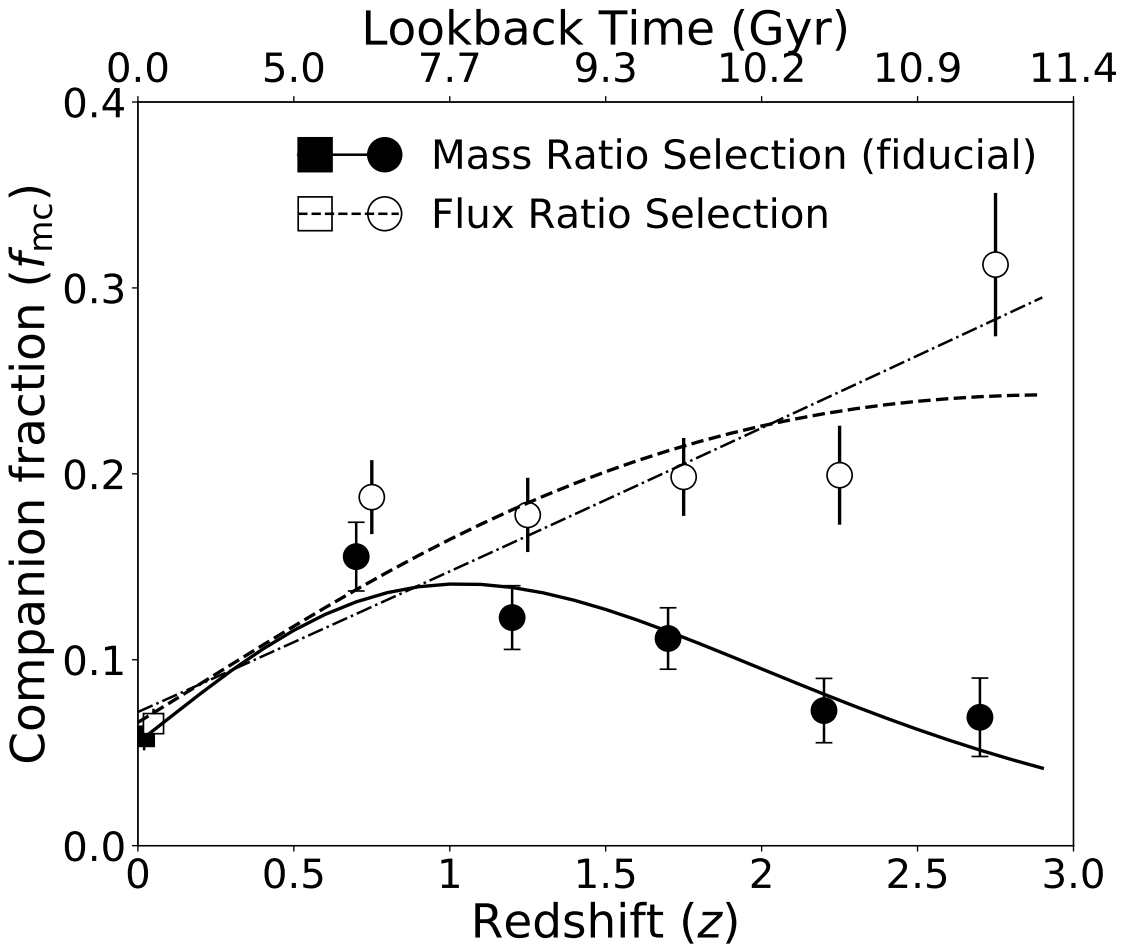

Figure 8: Comparison of the redshift evolution of major companion fractions $f_{\rm mc}(z)$ based on our fiducial selection (stellar mass ratio $1 \leq M_1/M_2 \leq 4$; black symbols, solid line) and based on H-band flux ratios ($1 \leq F_1/F_2 \leq 4$; open points; slightly offset in z direction for clarity). Best-fit curves to the flux-ratio fraction evolution are shown for modified power law (Eq. 1.7, dashed line) and a simple power law (dot-dashed line; $f_{\rm mc}(z) = 0.07(1+z)^1$). The data symbols distinguishing CANDELS and SDSS fractions, and the error bars are the same as in Fig. 6. The flux ratio selection results in larger fractions at $z > 1$ and an increasing power-law redshift dependence in contrast to the $f_{\rm mc}(z)$ derived from stellar-mass ratios.

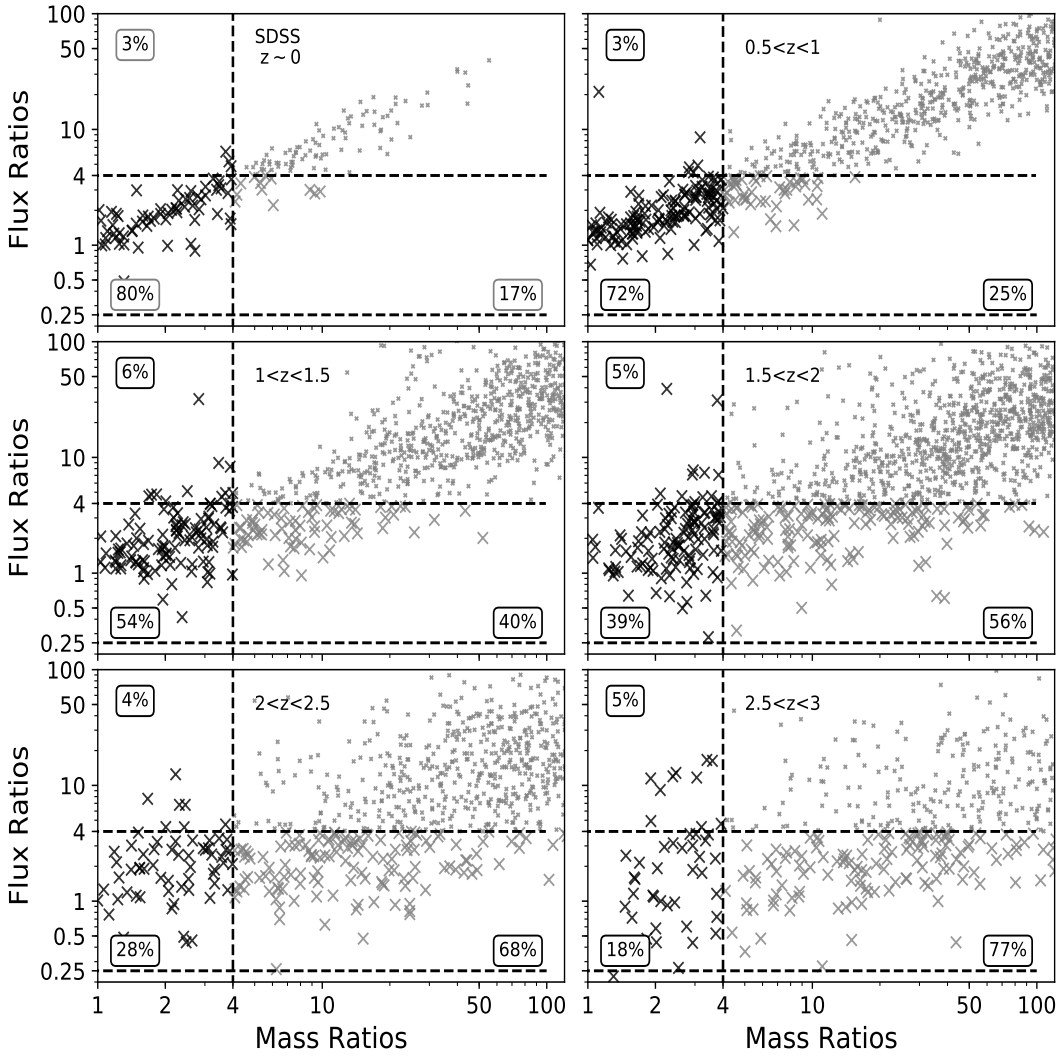

Figure 9: Stellar-mass-ratio vs. H-band flux-ratio for the close-pair systems satisfying our fiducial projected separation ($5 - 50$ kpc) and redshift proximity choices for SDSS ($\Delta v_{12} \leq 500\,\mathrm{km\,s^{-1}}$; $z \sim 0$; top left) and CANDELS (B09; $0.5 \leq z \leq 3$; rest of the panels). We show the pairs that satisfy our fiducial mass-ratio criterion ($1 \leq M_1/M_2 \leq 4$) in large-black points , those having a $M_1/M_2 > 4$ but satisfy the flux-ratio criterion ($1 \leq |F_1/F_2| \leq 4$) in large-grey points, the remaining points in small-grey markers. To visualize the contamination from minor pairs ($M_1/M_2 > 4$), introduced by the flux-ratio criterion in each redshift panel, we show the percentage of pairs that satisfy either mass-ratio or flux-ratio selection that fall in each of the three quadrants. We notice that the contamination increases steadily from $z \sim 0$ to $z = 3$, causing inflated $f_{\rm mc}$ values observed in Figure 8 (see § 1.4.3 for discussion).

Table 5: Compilation of companion selection criteria employed by previous empirical close-pair-based studies. Columns: (1) names of the studies – presented in the order of their mention in Figure 10; (2) the choice of projected separation $R_{\rm proj}$ annulus; (3) the criterion to choose major companions, where $\mu_{12} = M_1/M_2$ is the stellar mass ratio (MR), flux-ratio (FR) is given by F_1/F_2, and the luminosity ratio (LR) is given by L_1/L_2; (4) the choice of redshift proximity criterion to select companions in plausible close pairs, where Δz_{12} is the photometric redshift difference $z_1 - z_2$, Δv_{12} is the spectroscopic velocity difference, LOS is the statistical correction for line-of-sight projections, and $CDF(z_1, z_2)$ is the cumulative probability of the galaxies involved in a close pair; (5) the applied selection criterion to select the primary galaxy sample, which is either a stellar mass-limited or flux-limited selection; (6) the redshift range in which the study presented their findings. We divide the table into two parts, where studies based on mass ratios are above the solid-dashed line and vice-versa. [121] employed a redshift proximity where $\Delta z_{12}/1+z_1 < 0.1$ for $z < 1$ and $\Delta z_{12}/1 + z_1 > 0.2$ when $z_1 > 1$. The dagger represents that the studies derived a 'pair' fraction rather than the companion fraction ($f_{\rm mc}$), see § 1.5.1 for details.

Study	$R_{\rm proj}$ selection [kpc]	Major Companion	Redshift Proximity	Primary galaxy selection	Redshift (z)
(1)	(2)	(3)	(4)	(5)	(6)
[15]	$\in [5, 30]$	MR$\in$ $[1, 4]$	2σ $z_{\rm phot}$ error overlap	$M_{\rm stellar} \geq 10^{10}\, M_\odot$	$\in (0.5, 2)$
[6]†	$\in [5, 30]$	MR$\in$ $[1, 4]$	$\Delta z_{12}^2 \leq \sigma_{\rm z,1}^2 + \sigma_{\rm z,2}^2$	$M_{\rm stellar} \geq 10^{10}\, M_\odot$	$\in [0.4, 1.4]$
[17]†	$\in [5, 30]$	MR$\in$ $[1, 4]$	$CDF(z_1, z_2)$	$M_{\rm stellar} \geq 10^{10}\, M_\odot$	$\in (0, 2)$
[67]	$\in [14, 43]$	MR$\in$ $[1, 4]$	$\lvert\Delta z_{12}\rvert/1 + z_1 < 0.2$	$M_{\rm stellar} \geq 10^{10.5}\, M_\odot$	$\in [0.4, 2]$

[121]	$\in [14, 43]$	MR$\in$ $[1, 4]$	see caption	$M_{\rm stellar} \geq 10^{10.7}\, M_{\odot}$	$\in [0.4, 2]$
[16]	$\in [14, 43]$	MR$\in$ $[1, 4]$	same as Newman+12	$M_{\rm stellar} \geq 10^{10.8}\, M_{\odot}$	$\in [0, 3]$
[5]	$\in [5, 30]$	LR$\in [1, 4]$	$\Delta v_{12} \leq 500\,{\rm km\,s^{-1}}$	$M_{\rm V} < -19.8 - 1.0z$	$\in [0.1, 1.2]$
[6]†	$\in [5, 30]$	FR$\in [1, 4]$	LOS	$M_{\rm stellar} \geq 10^{10}\, M_{\odot}$	$\in [0.4, 1.4]$
[63]	$\in [5, 30]$	FR$\in [1, 4]$	3σ $z_{\rm phot}$ error overlap	$M_{\rm stellar} \geq 10^{11}\, M_{\odot}$	$\in [0, 3]$
[122]	$\in [5, 30]$	FR$\in [1, 4]$	LOS	$M_{\rm stellar} \geq 10^{11}\, M_{\odot}$	$\in [1.7, 3]$
[62]	$\in [14, 43]$	LR$\in [1, 4]$	$\Delta v_{12} \leq 500\,{\rm km\,s^{-1}}$	$-21 < M_{\rm B} + 1.3z < -19$	$\in [0, 1.2]$
[123]†	$\in [14, 43]$	LR$\in [1, 4]$	$\Delta v_{12} \leq 500\,{\rm km\,s^{-1}}$	$M_{\rm V} < -20 - 1.1z$	$z = 0.5$, $z = 0.8$
[16]	$\in [14, 43]$	FR$\in [1, 4]$	same as Newman+12	$M_{\rm stellar} \geq 10^{10.8}\, M_{\odot}$	$\in [0, 3]$

1.5 Discussion

Our extensive analysis of the full CANDELS sample and a well-defined selection of SDSS galaxies in the preceding sections provides a new baseline for the evolving frequency of massive galaxies with close major companions (and of major mergers by extrapolation) spanning epochs from the start of cosmic high noon to the present-day. In this section, we present a comprehensive comparison of our measurements to those from previous studies by taking advantage of our analysis of varying the companion selection criteria from § 1.4. We derive the empirical major merger rates based on a *constant* observability timescale and find significant variation in the rates depending on our stellar-mass

ratio or H-band flux-ratio choice. We discuss plausible reasons for these variations by re-computing merger rates based on theoretically-motivated, redshift-dependent timescale prescriptions. Finally, we describe the need for detailed calibrations of the complex and presumably redshift-dependent conversion factors required to translate from fractions to rates, which is key to improve major merger history constraints.

1.5.1 Comparison to previous empirical close-pair-based studies

We compare our CANDELS+SDSS major companion fractions from to the results from previous stellar-mass and flux selected, empirical close-pair-based studies tabulated in Table 5. We note that some studies quote fraction of pairs ($f_{\rm pair}$), which is the fraction of close-pair systems among a desired massive galaxy sample of interest. This is different from the companion fraction $f_{\rm mc}$. To correctly compare to previous studies which used a $f_{\rm pair}$ definition, we derive a simple conversion between the two definitions by separately computing the ratio of $f_{\rm mc}/f_{\rm pair} = 1.5$, which is constant for the CANDELS and SDSS samples at $0 < z < 1.5$. Overall, our $f_{\rm mc}(z)$ results are in good agreement with previous studies when we properly convert the published pair fractions into companion fractions. We separate our comparisons into stellar mass-ratio and flux-ratio based studies in Figure 10. For each set of comparisons, we further split them into matching $5-30$ kpc and $14-43$ kpc projected separation selections. We discuss each comparison with respect to any differences in the details of the redshift proximity selection.

1.5.1.1 Stellar Mass-Ratio Selected Studies

We use the $5-30$ kpc and $14-43$ kpc fractions from § 1.4.1 (Figure 6) and plot them in Figure 10 alongside previous studies that employed a similar projected separation criterion and 4:1 stellar mass-ratio selection ($1 \leq M_1/M_2 \leq 4$). We start our discussion by comparing to the studies that employed a $R_{\rm proj} = 5-30$ kpc criterion. [15] studied major close pairs among galaxies more massive than $10^{10} M_{\odot}$ over the redshift range $0.5 < z < 2.5$, and found an un-evolving fraction ($\sim 10\%$). We agree with these estimates within their quoted uncertainties; however, we acknowledge that the consistency at $2 < z < 2.5$ is because of their large error bar ($> 50\%$). We also note that [15] used a $2\sigma_z$ overlap as their redshift proximity, which is less restrictive than our modified [63] redshift proximity criteria in Equation 1.8. Based on our analysis in § 1.4.2, this implies that the [15] fractions may be over-estimated. Next, we compare to B09 who studied $5-30$ kpc pairs among $\log_{10}(M_{\rm stellar}/M_{\odot}) \geq 10$ galaxies between $0.4 < z < 1.4$ using a redshift proximity that we have adopted as our fiducial choice. It is important to note that B09 elected a 4:1 K_s-band flux ratio as their fiducial major companion selection criterion. However, in their analysis, they found that 80% of their flux-ration-selected companions also met $< 4:1$ stellar-mass ratio major companion selection. To appropriately compare to B09 pair fractions, we include a 0.8 multiplicative factor to covert them to mass-ratio-based $f_{\rm mc}$. We find that their estimates are on average smaller than our fractions, marginally agreeing at $0.7 < z < 1.4$, and disagreeing at $0.4 < z < 0.7$.

We now compare to the most recent study by [17], who analyzed close pairs ($R_{\rm proj} = 5-30$ kpc) among a sample of $\log_{10}(M_{\rm stellar}/M_{\odot}) \geq 10$ galaxies at $0 < z < 2$

from the UDS, VIDEO, GAMA, and COSMOS datasets. In Figure 10, we show all their pair fractions, except for an upper-limit of 17.5% from COSMOS at $1.5 < z < 2$ and convert them to $f_{\rm mc}$. We find excellent agreement between GAMA-survey-based, spectroscopically derived fraction at $0 < z < 0.2$ and our SDSS-based fraction at $z \sim 0$. For the same GAMA sample, they also computed the fractions using photometric redshifts and found an average of $\sim 5\%$ at $0 < z < 0.2$. We separately computed the SDSS $f_{\rm mc}$ using B09 redshift proximity and matching their $R_{\rm proj} = 5 - 30$ kpc selection, and find that our fraction $f_{\rm mc}(0) \sim 2.8\%$ is ~ 1.8 times smaller and marginally disagrees within their quoted uncertainties. At $0.2 < z < 1$, we find that Mundy17 fractions are in good agreement with our $f_{\rm mc}(z)$, except for their COSMOS-survey-based value at $0.2 < z < 0.5$, which is marginally smaller than our $f_{\rm mc}$. At $1 < z < 1.5$, their fraction ($\sim 15\%$) is in disagreement with our findings, while their upper-limits are consistent with our results. Overall, we demonstrate good agreement with [17].

Now, we discuss our comparison to the studies that employed a $R_{\rm proj} = 14 - 43$ kpc separation selection. We start with [67], who studied major (<4:1 stellar-mass ratio) close pairs among $\log_{10}(M_{\rm stellar}/M_{\odot}) \gtrsim 10.5$ galaxies at $0.4 < z < 2$, and found a diminishing redshift evolution of the fractions. We find disagreement with the [67] fractions at $0.4 < z < 1.2$, where our $f_{\rm mc}$ are higher by 25%, however, we show good agreement within their diminished fractions at $1.2 < z < 2$. We note that our sample is almost two times larger than their galaxy sample, and a likely reason for this discrepancy is that their uncertainties may have been under-estimated. Next, we compare to [121], who quantified major companion fraction among $\log_{10}(M_{\rm stellar}/M_{\odot}) \gtrsim 10.7$ galaxies,

and found that $f_{\rm mc}$ is flat (at $\sim 10\%$) at $0.4 < z < 2$. We find good agreement with their findings within the quoted uncertainties and in that redshift range.

Finally, we compare to a recent study by [16] who study a massive galaxy sample of $\log_{10}(M_{\rm stellar}/M_{\odot}) \gtrsim 10.8$ galaxies to constrain major companion fractions at $0 < z < 3$ using two separate datasets, namely 3D-HST and UltraVISTA. Firstly, we find that our $f_{\rm mc}(z)$ values are in excellent agreement with the [16] 3D-HST estimates over their redshift range, which is not surprising as 3D-HST covers 75% of the CANDELS fields. We find that the [16] UltraVISTA fractions follow an increasing trend that is qualitatively similar to our $f_{\rm mc}(z)$ at $0 < z < 1$, however, we notice that the former fractions are quantitatively smaller than the latter. At $z > 1$, we find that the UltraVISTA fractions are in good agreement with our results, except at $z \sim 2.25$, where our $f_{\rm mc}$ is two times smaller. Overall, we note that our companion fractions are in good agreement with [16] estimates. Moreover, [16] applied a parabolic fitting function to their UltraVISTA major companion fractions and report that their redshift evolution may be peaking around $z \sim 1 - 1.5$, in contrast to our fractions that peak around $0.5 < z < 1$ and diminish up to $z = 3$. This may be a consequence of more-massive galaxies experiencing growth earlier in cosmic-time than the less-massive population, which is often referred to as the galaxy-downsizing phenomenon. In summary, we conclude that our stellar-mass ratio based major companion fractions $f_{\rm mc}(z)$ are in good agreement with previous empirical close-pair-based estimates, once they are closely-matched in selection criteria.

1.5.1.2 Flux Ratio Selected Studies

We separately derive the major companion fractions for $R_{\rm proj} = 5-30$ and $14-43$ kpc selections using H-band (r-band) flux ratio selection for CANDLES (SDSS) and plot them in Figure 10 along side some published studies that used a similar projected separation and $< 4:1$ flux (luminosity) ratio selection (see Table 5). First, we discuss the comparison to those studies that employ $R_{\rm proj} = 5 - 30$ kpc. We start with [5], who studied close pairs using a V-band luminosity-limited galaxy sample at $0.2 \leq z \leq 1.2$, and found that the companion fraction evolves as $f_{\rm mc} \propto (1+z)^{3.1}$. We find that [5] estimates are ~ 2 times smaller than our results at $0.5 < z < 1$, however, they start to converge at $z < 0.5$ and $z > 1$, and agree with our $f_{\rm mc} \sim 10\%$ at $z = 1.2$. We note that [5] uses a smaller separation annulus of $R_{\rm proj} = 5 - 20$ kpc, and in addition they also apply a random chance pairing correction. Based on our analysis from § **??**, we note [5] fractions may be systematically smaller, and most plausibly are causing the observed discrepancy at $0.5 < z < 1.0$.

We now compare to B09, and remind our reader that they used a 4:1 flux-ratio as their fiducial major companion criterion. We find good agreement with the [6] estimates at $0.7 < z < 1.4$, but find marginal disagreement at $0.4 < z < 0.7$. Next, we compare to [63], who analyzed close pairs among $\log_{10}(M_{\rm stellar}/M_{\odot}) \geq 11$ galaxies using 4:1 H-band flux-ratio criterion at $0 < z < 3$, and found fairly high fractions up to $z = 3$. We find good agreement with their fractions over the full redshift range, however, we acknowledge that a proper comparison with our highest-redshift bin results is not possible owing to their large uncertainty.

Finally, we compare to [14,122], who analyzed the incidence of close companions which satisfy a $4:1$ H-band flux ratio selection among a GOODS NICMOS survey sample of $\log_{10}(M_{\rm stellar}/M_{\odot} > 11)$ galaxies at $1.7 < z < 3$. [14] reports that the fraction of galaxies hosting a nearby major companion evolves steeply with redshift as $f \propto (1+z)^3$ (dashed line in Figure 10b), reaching up to $f = 0.19 \pm 0.07$ ($1.7 < z < 2.3$) and $f = 0.40 \pm 0.1$ ($2.3 < z < 3$), with a total fraction of $f = 0.29 \pm 0.06$. While we find consistent agreement at $1.7 < z < 2.3$, we note that their fractions are a factor of two higher at $2.5 < z < 3$. Similarly, we also find that their total fraction, which spans over $1.7 < z < 3$, is also larger by a factor of two when compared to our total $f_{\rm mc}$ at $2 < z < 3$. Although we qualitatively agree with the strongly increasing redshift evolution of [14], we note that our $f_{\rm mc}(z)$ trend is shallower with a simple power-law slope of $m \sim 1$ (see Figure 8). We argue that the steeply increasing fraction evolution by [14] may be a consequence of the increased contamination from minor companions by their flux-ratio selection (see § 1.4.3). Also, significant field-to-field variance due to their small sample size and the use of a relatively less restrictive line-of-sight contamination correction may be contributing towards the observed (factor of two) disagreement to our fractions.

We now focus our comparisons to studies that employ a $R_{\rm proj} = 14 - 43\,$kpc selection (see Table 5). We start our comparison with [62], who studied close pairs among luminosity-limited sample of galaxies at $0.12 < z < 1.1$. We find that our fractions are only marginally consistent with their results at $z < 0.5$ and disagree with them at $0.5 < z < 1$ as our fractions are ~ 2 times larger than their values. We note that [62] used a stringent spectroscopic redshift proximity criterion of $\Delta v_{12} \leq 500\,{\rm km\,s^{-1}}$, which

may be the reason for their smaller fractions and the observed discrepancy. Nevertheless, we notice that [62] fractions start to converge with our estimates at $z > 1$, which may be due to the rising minor-companion contamination by flux-ratio selection leading to larger fractions. We now compare to [124], who study a spectroscopically confirmed sample of close pairs among B-band luminosity-limited galaxy sample to quantify pair fractions at two redshifts $z = 0.5$ and $z = 0.8$. We find that our $f_{\rm mc}$ estimate at $z = 0.5$ is marginally higher than their fraction, but is in good agreement at $z = 0.8$. Also, we qualitatively agree with their increasing redshift evolutionary trend of the fractions at $0.5 < z < 0.8$.

Finally, we compare to [16], who also derived 4:1 H-band flux-ratio based companion fractions for $\log_{10}(M_{\rm stellar}/M_{\odot}) \gtrsim 10.8$ galaxies spanning $0.1 < z < 3$ among the 3D-HST and UltraVISTA datasets. We find that our $f_{\rm mc}(z)$ values trace very closely with the [16] 3D-HST estimates and are in excellent agreement within the uncertainties. We also find that [16] UltraVISTA estimates qualitatively agree with the increasing redshift evolution of our flux-ratio selected $f_{\rm mc}(z)$. However, we notice that their estimates at $z < 0.5$ are marginally smaller than our $f_{\rm mc}$ extrapolation between $z \sim 0$ and $0.5 < z < 1$. In summary, we conclude that our H-band flux-ratio based $f_{\rm mc}(z)$ are in good agreement with previous empirical studies once we match closely the choices of companion selection criteria.

1.5.2 Major Merger Rates

As outlined in the Introduction, empirical merger rates provide a fundamentally crucial measure of the importance of major merging in the evolution of massive galaxies.

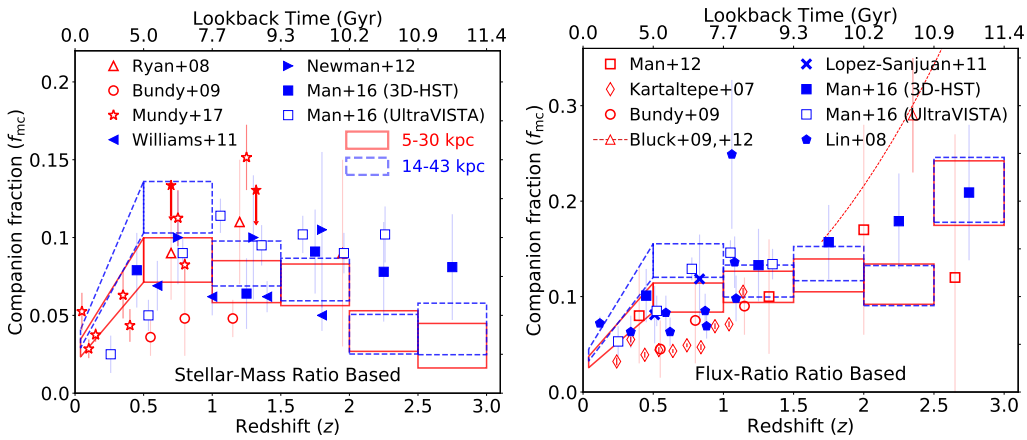

Figure 10: Comparison of major companion fractions from CANDELS+SDSS to those from previous studies that employed 4:1 stellar-mass ratio (a) and 4:1 flux-ratio (b) selections. In both panels, we outline the $f_{\rm mc}$ measurements data points as rectangles where their height represents the 95% binomial confidence limits per redshift bin (width of the rectangles). The data-points of previous empirical studies are given in the panel keys (see Table 5). We compare fractions based on different projected separation criteria as follows: $5-30$ kpc (solid-red line; open-red markers), and $14-43$ kpc (blue-dashed line, filled-blue markers). We find our major companion fraction estimates are in good agreement with previous empirical constraints when the companion selection criteria are closely matched. We off-set multiple fields from [17] by a small amount, and show the upper-limits in filled markers with bold arrow for clarity.

Our new major companion fractions from CANDELS+SDSS allow us to place clearly-defined new constraints on the evolution of major merger rates between $z = 3$ and $z = 0$. Here, we elect to focus on a straightforward comparison of merger rates derived from our major companion fractions with the largest systematic differences: stellar mass-ratio versus flux-ratio selection. Then we compare our merger rates to several recent empirical studies and cosmologically-motivated predictions that employ similar companion selection criteria as we do.

The major merger rate based on close-companion (pair) statistics is defined as the

number of major galaxy-galaxy merger events per unit time per selected galaxy (see [16]):

$$R_{\rm merg,pair} = \frac{C_{\rm merg,pair} \times f_{\rm mc}}{T_{\rm obs,pair}} \, [{\rm Gyr}^{-1}] \, , \tag{1.9}$$

where $T_{\rm obs,pair}$ is the average observability timescale during which a galaxy-galaxy pair satisfies a given selection criterion, and the multiplicative factor $C_{\rm merg,pair}$ attempts to account for the fact that not all such companions will merge within the $T_{\rm obs,pair}$ interval [18]. It is clear that the $C_{\rm merg,pair}$ and $T_{\rm obs,pair}$ factors are *crucial* assumptions involved in converting observed companion fractions into merger rates. As such, these factors are a major source of systematic uncertainty in merger rate calculations. A few analyses of theoretical simulations have provided fairly broad guidelines for these key assumptions. [13] find that different merger timescale assumptions can induce up to a factor of 2 uncertainty in the merger rates. Others have attempted to constrain the fraction-to-rate conversion factors and quote limits of $C_{\rm merg,pair} = 0.4$ to 1 [125, 126]. Despite the need for improved constraints on these key merger rate factors, attempts to calibrate $C_{\rm merg,pair}$ and $T_{\rm obs,pair}$ in detail with regard to changing companion selection criteria and as a function of fundamental galaxy properties such as redshift and stellar mass, are lacking.

In practice, previous empirical studies have used merger rate factors that span a wide range. For example, some studies have adopted $C_{\rm merg,pair} = 1$ for simplicity [16], while others have employed an intermediate value of 0.6 [18] based on limits from simulations. Similarly, some empirical studies adopt their $T_{\rm obs,pair}$ values based on analytical fitting function provided by [125], while others adopt values provided by [127] that range from 0.3 to 2 Gyr depending on the close-companion selection criteria. For the major merger rate calculations below, we make a simple assumption that these factors

are constant over all redshifts and masses we probe. We adopt $C_{\rm merg,pair} = 0.6$ and $T_{\rm obs,pair} = 0.65\,$Gyr, which is suitable for our fiducial $R_{\rm proj} = 5 - 50\,$kpc criterion [127]. Exploring the detailed impact of these assumptions is beyond the scope of this paper. However, in § 1.5.3.2, we test two theoretically motivated redshift-dependent timescale prescriptions to re-derive the merger rates and discuss their implications.

In Figure 11, we show the major merger rate evolution during $0 < z < 3$, which we derive from CANDELS+SDSS major companion fractions for stellar mass-ratio (MR) and flux-ratio (FR) selections with matched fiducial projected separation and B09 redshift proximity criteria. Holding aside the factor-of-two systematic uncertainty contribution from $T_{\rm obs,pair}$ to merger rate calculations, we find both the MR and FR-based merger rates evolutionary trends mimic their respective $f_{\rm mc}(z)$ trends discussed in § **??**, which is expected given our simplistic fraction-to-merger-rate conversion factors. We notice that both the MR and FR selections yield consistent major merger rates at $z < 1$, where they increase strongly from ~ 0.05 ($z \sim 0$) to $\sim 0.15\,$Gyr^{-1} ($0.5 < z < 1$). At $z > 1$, the different selections have divergent merger rates. The MR-based merger rates decline with redshift to $z = 3$, indicating a turnover at $z \sim 1$. In contrast, the inflated companion fractions owing to contamination by minor companions, the flux-ratio selection yields $1.5-4.5$ times larger rates between $1 \lesssim 3$ with an increasing power-law trend $R_{\rm merg,pair} \propto (1+z)^1$.

Next, we compare our empirical MR-based merger rates to previous empirical estimates (Figure 11). For $z \leq 1.5$, we compare to [18] who compile close-pair fractions from several observational stellar-mass and luminosity-selected studies (magenta line)

and found strongly increasing major merger rate evolution with redshift $R \propto (1+z)^{1.7-2.1}$, which qualitatively agrees with both our MR and FR rate evolutions over this redshift range. The [18] merger rate evolutions have systematically smaller normalization than ours because their compilation includes close-pair-based fractions that may be systematically smaller than our $f_{\rm mc}$ due to the use of different selection criteria. For example [5,62] use a stringent proximity selections (see § 1.5.1.2, and Table 5) and [60] uses a stringent major companion selection criterion [9]. For merger rates at $0.5 < z < 3$, we compare our fiducial MR-based merger rate to a recent study of close pairs by [16]. We find that our individual $R_{\rm merg,pair}$ data points agree well with the man16 results (blue) within the measurement uncertainties at these redshifts. Yet, our standalone rising-diminishing merger rate evolution disagrees with the [16] flat trend. If we were to extend their rates to $z \sim 0$, we would disagree with them; however, we note that [16] do not attempt to anchor to a low-redshift (for their 3D-HST sample), as such, claim a flat trend with redshift in agreement with [17]. Recently, [17] in conjunction with Duncan et al., in prep find that the [27] semi-analytic model mock light-cones accurately reproduce a flat pair-fraction trend that is consistent with the recent findings by [16, 128], and the results of this work. Employing an un-evolving close-pair observability timescale assumption, [17] reports a flat major merger rate evolution. Our fiducial MR-selected merger rates shown in Figure 11 agree with the [17] result, which is not surprising owing to the assumption of constant observability timescale.

Here, we discuss our comparison to theoretical major merger predictions. Since

[9] [60] selects major companions using a 2:1 luminosity ratio as opposed to our 4:1 criterion from a luminosity-limited galaxy sample at $z \sim 0$.

these are based on stellar-mass ratios, we refrain from comparing them to our FR-based rates. [13] provides a comprehensive analysis of major galaxy-galaxy merging using ΛCDM motivated simulation and Semi-Analytic Modeling to derive major merger rate predictions, and quantified systematic error contribution from various theoretical model-dependent assumptions. We find that the [13] major merger rate evolution of massive ($\log_{10}(M_{\rm stellar}/M_{\odot}) \geq 10.3$) galaxies ($R = 0.04(1+z)^{1.35}$; green line in Figure 11), within a factor of two uncertainty, agrees qualitatively and quantitatively with our MR-based rates up to $z < 1.5$. At $z > 1.5$, our fiducial MR-selected, diminishing merger rate trend disagrees with the [13] predictions. We also compare our merger rates to a recent theoretical prediction by [19] from the Illustris numerical hydrodynamic simulation [129]. We note that their $\log_{10}(M_{\rm stellar}/M_{\odot}) \geq 10.3$ galaxy merger rates (black line in Figure 11) follow an even stronger increasing redshift dependence ($R \propto (1+z)^{2.4-2.8}$) than the [13] predictions, which also agrees qualitatively with our $f_{\rm mc}(z)$ trend at $z < 1.5$. However, at $z < 1.5$, we note that the [130] predictions are smaller than our empirical estimates. Similar to the conclusion from the [13] comparison, our fiducial MR-based merger rates disagree with predictions from [130] at $z > 1.5$. We present further discussion on the plausible reason for the observed discrepancy of flat vs. rising merger rate evolution in § 1.5.3.2.

Being the primary driver for galaxy-galaxy merging, halo-halo merger rates serve as a ceiling for both empirical measurements and theoretical galaxy merging predictions, and their qualitative trend is expected to mimic the galaxy merger rates closely. We

find that our empirical MR-selected merger rate trend broadly agrees with the analytical merger rates of $M_{\rm halo} \sim 10^{12}\,M_{\odot}$ dark-matter halos ($R \propto (1+z)^{2.5}$; brown line in Figure 11) predicted by [24–26] (also see [131]) at $z \leq 1$ but starts to deviate significantly at $z > 1$. Since the translation from a halo-halo merger rate to a galaxy-galaxy merger rate depends on the redshift-dependent mapping of galaxies onto their respective dark-matter halos (halo-occupation statistics), the galaxy merger rates are expected to follow a shallower redshift evolution than the ones set by the halo merger rates (for discussion, see [13]). Although different assumptions of halo-occupation models may contribute up to a factor of two uncertainty in the theoretically predicted galaxy merger rates (e.g., [132]), they all hint towards a rising incidence with increasing redshift. As such, a deviation of more than factor of five (e.g., see redshift bin $2 < z < 3$ in Figure 11) between the constant timescale-based fiducial MR-selected empirical rates (diminishing) and the analytical halo merger rate evolution (rising; $R \propto (1+z)^{2.5}$) is puzzling. Based on this analysis, we conclude that a straight-forward MR-based estimation of major merger rates using a constant fraction-to-merger-rate observability timescale ($T_{\rm obs,pair} = 0.65\,$Gyr) are in disagreement with the theoretical predictions at $z > 1.5$. We discuss the implications of these conclusions in § 1.5.3.

1.5.3 Implications of the Disagreement between Simplistic Empirical Merger Rates and Model Predictions

So far, we have demonstrated that major merger rates based on a rather straightforward MR-selection start to disagree with theoretical predictions of major galaxy-galaxy

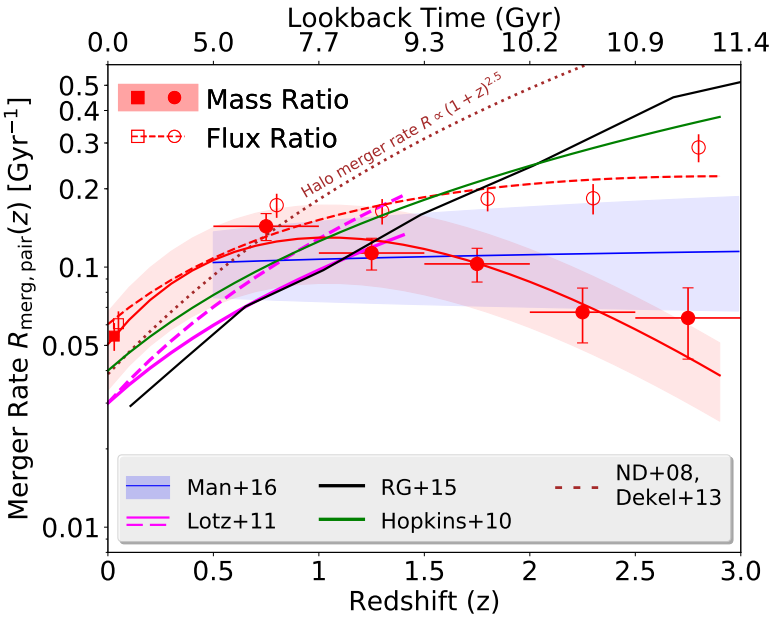

Figure 11: Comparison of CANDELS+SDSS galaxy-galaxy major merger rates $R_{\rm merg,pair}(z)$ (number of mergers per galaxy per Gyr) for massive ($M_{\rm stellar} \geq 2\times10^{10} M_{\odot}$) galaxies at $0 < z < 3$, to rates from previous empirical studies and theoretical model predictions. We show the $R_{\rm merg,pair}(z)$ computed using major companion fractions ($f_{\rm mc}$) based on fiducial projected separation ($5-50$ kpc) and redshift proximity (CANDELS: Equation 1.1; SDSS: $\Delta v_{12} \leq 500\,{\rm km\,s^{-1}}$) split into stellar mass-ratio (filled points; solid line) and flux ratio (open points; dashed line) for CANDELS (circles) and SDSS (square). We employ simplistic assumptions for fraction-to-rate conversion factors $T_{\rm obs,pair} = 0.65$ Gyr and $C_{\rm merg,pair} = 0.6$, and show the variation of $R_{\rm merg,pair}(z)$ for $C_{\rm merg,pair} = 0.4$ to 0.8 in red-shading. The error-bars on the data points indicate their 95% binomial confidence limits, solely based on the observed number counts. The solid (dashed) magenta line is the empirical merger rate evolution from [18] for stellar-mass (luminosity) limited selections, the empirical merger rate from [16] is shown in solid blue line along with its blue shaded uncertainty. We compare to theoretical galaxy-galaxy major merger rate predictions of $M_{\rm stellar} \geq 2\times10^{10} M_{\odot}$ galaxies by [13] (solid green lines), [130] (solid black), and to the analytical halo major merger rate prediction of $M_{\rm halo} \sim 10^{12}\,M_{\odot}$ dark-matter halos by [24–26] ($R \propto (1+z)^{2.5}$; dashed brown line). We strongly advise against trusting flux-ratio-based galaxy merger rates despite their agreement with the theoretical models, owing to the notable contamination from non-major mergers (see § 1.4.3). Our fiducial mass-ratio based empirical major merger rates show broad agreement up to $z \sim 1.5$, but demonstrate a strong tension at $z > 2$ when compared to previous empirical and theoretical constraints.

merging at $z \gtrsim 1.5$. This indicates there may be one or many redshift-dependent effects plaguing the key variables ($C_{\rm merg,pair}$, $f_{\rm mc}$, and $T_{\rm obs,pair}$) in Equation 1.9, causing the observed merger rates to disagree with simulations. Here, we first investigate for plausible systematic under-estimation of $f_{\rm mc}$ values at $z \gtrsim 1.5$ as a reason for the observed discrepancy by testing if we are missing close companions by not choosing the full $z_{\rm phot}$ PDF when applying the B09 redshift proximity. Then, we discuss another possibility that the simplistic, constant value assumptions of $T_{\rm obs,pair}$ may be incorrect, and test two theoretically-motivated, redshift-dependent $T_{\rm obs,pair}$ prescriptions to re-derive merger rates. We also elaborate on the remnant formation times, which is a key difference between the merger rates measured in theoretical simulations and empirically calculated close-pair-based rates, and comment on how it may be affecting the data-theory comparison.

1.5.3.1 Are We Missing Companions by Not Accounting for the Full $z_{\rm phot}$ PDFs?

One plausible reason for the $z > 2$ data-theory discrepancy may be that the adopted B09 redshift proximity criterion (Eq. 1.1) is missing plausible physical close-pair systems because we are using only the 1σ photometric uncertainties and not the full $z_{\rm phot}$ PDF. To test if this is the case we carry out a simple test, where we start by simplifying Equation 1.1 as $\xi = \Delta z_{12}/\sqrt{\sigma_{\rm z,1}^2 + \sigma_{\rm z,2}^2}$; therefore, our fiducial criterion equates to $\xi \leq 1$. The distribution of ξ values for physical close-pair systems is Gaussian, assuming $\sigma_{\rm z}$ values are Gaussian. Therefore, using $\xi \leq 1$ is equivalent to selecting 68% of the

full ξ distribution. Ideally, if one assumes the contribution from non-physical pairs (hereafter called as the background) per ξ bin is negligible, as much as 32% of the close pairs (thereby companions) can be missed when using $\xi \leq 1$. For this exercise, we allow for a non-zero background to properly estimate the missing pairs at five redshift bins between $0.5 < z < 3.0$.

We begin with the cumulative distribution function (CDF) of ξ for all the close-pair systems selected using our fiducial $R_{\rm proj} = 5 - 50\,{\rm kpc}$ projected separation criterion. To avoid errors from small number statistics, we do not limit close-pair systems with a stellar-mass-ratio selection. However, we require that both the host and companion galaxies to be brighter than $H = 25\,{\rm mag}$ to exclude large ξ values from affecting the statistics. We now model CDF with a Gaussian integral function in conjunction with a constant background contribution, following the fitting formula: $N(\xi \leq c) = A/\sqrt{2\pi}\int_0^c \exp^{-x^2/2}\,dx + Bc$. Here, $N(\xi \leq c)$ is the number of close-pair systems that satisfy a cutoff significance (c) as $\xi \leq c$, and the variables A and B are the contributions from plausible physical close pairs and the background, respectively. For demonstrative purposes, we present the fits to the CDF and their best-fit parameter values of A and B in Figure 14 for two redshift bins between $1.5 < z < 3$. At each redshift bin, we define the ratio $(A/2)/N(\xi \leq 1)$ as the multiplicative correction factor to our fiducial $f_{\rm mc}$ that corrects for the missing close-pair systems due to Method I. When the background is negligible ($B \sim 0$), the correction factor is ~ 1.44. However, if B is significant, then the correction factor may become less than unity. In Figure 15, we plot the corrected $f_{\rm mc}$ alongside the fiducial fractions. We note that fractional change in the major companion

fraction after applying the correction factor is less than 5% at $0.5 < z < 2$. However, at $z > 2$ we find the relative contribution from the background starts to dominate the Gaussian contribution and therefore causes the fiducial $f_{\rm mc}$ to decrease by more than 10%. Despite these changes, we note that the overall corrected $f_{\rm mc}(z)$ values do not significantly deviate from the fiducial $f_{\rm mc}(z)$ values. Therefore, we conclude that we are not missing close pairs (thereby companions), by using 68% photometric-redshift errors for our fiducial redshift proximity criteria.

1.5.3.2 Evolving Observability Timescale and Remnant Formation Times

It is evident by now that the close-pair observability timescale ($T_{\rm obs,pair}$) is a key parameter that dictates the measurement of merger rates. Yet, a wide range of values can be found in the literature (see § 1.5.2), often assumed to be constant over a wide range of redshifts. Therefore, with the growing evidence [16, 17] for flat or possibly diminishing close-pair-based fractions at $1 < z < 3$, it is expected that a constant timescale assumption-based merger rates to disagree with the continually rising merger-rate trends from the theoretical predictions. A possible implication of this disagreement is that the timescale may be evolving with redshift, such that an ongoing merger can be observed as a close pair for a shorter period at earlier cosmic times. This may not be surprising as simulation-based merging-timescale calibrations [125, 126], motivated by dynamical friction time for a satellite DM-halo to merge with the parent, are complex functions that depend on key merger properties such as the mass ratio, galaxy mass, $R_{\rm proj}$, and

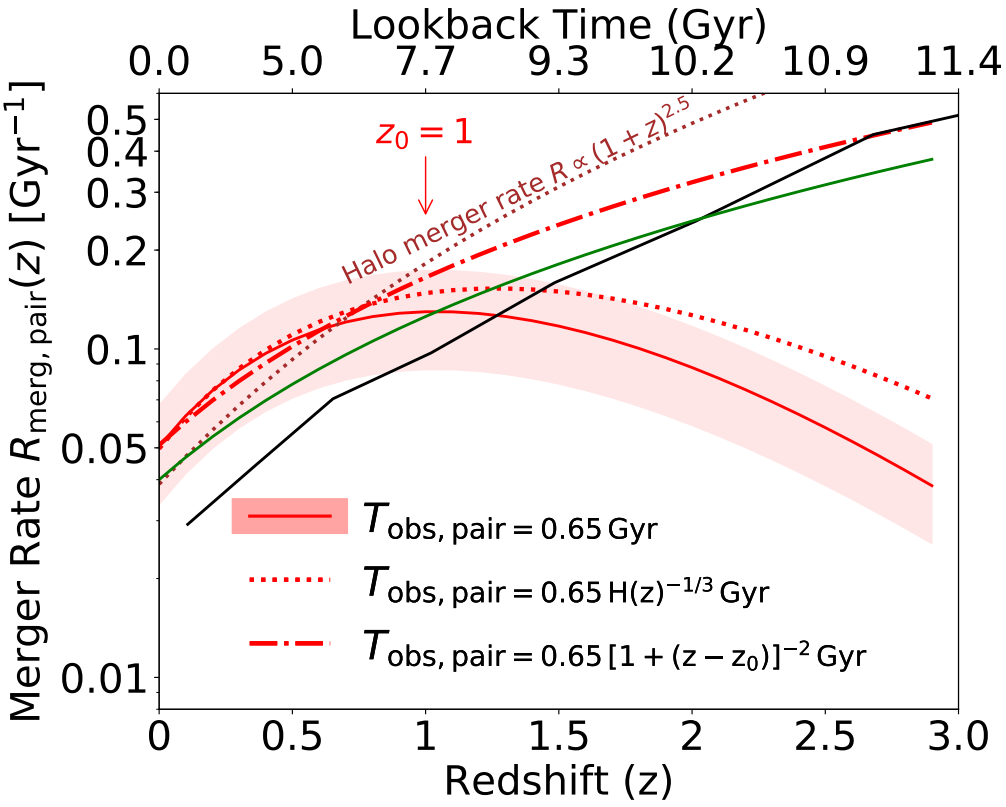

Figure 12: Comparing the redshift evolution of major merger rate of $M_{\mathrm{stellar}} \geq 2 \times 10^{10} M_{\odot}$ galaxies based on our fiducial close-pair timescale assumption $T_{\mathrm{obs,pair}} = 0.65\,\mathrm{Gyr}$ copied from Figure 11 (solid red line, shading) to rates from different timescale choices. We show the rates based on [126] scaling relation $T_{\mathrm{obs,pair}} \propto H(z)^{-1/3}$ in red dotted line, and [128] relation $T_{\mathrm{obs,pair}} \propto (1+z)^{-2}$ in red dot-dashed line (starting at $z_0 = 1$). We also plot the theoretical merger rate predictions shown in Figure 11.

$H(z)^{-1/3}$ ($H(z) = H_0 E(z)$; H_0 is the present-day Hubble constant; $E(z)$ is the evolution of H_0 as a function of redshift). Moreover, the dynamical timescale based on simple cosmological arguments is also redshift dependent, and can be approximated as $T \propto (1+z)^{-1.5}$ [128]. However, as per the discussion in [128], any processes that can act faster than the dynamical timescale and may impact the variables of close companion selection criteria, which in turn can change the $T_{\rm obs,pair}$. In this context, the galaxies are undergoing a rapid transformation concerning their stellar-mass assembly, owing to rapid ongoing star-formation [133] at $z > 1$. [128] found that this rapid mass assembly can shorten the close-pair observability timescale towards earlier redshifts, which evolves as $T \propto (1+z)^{-2}$ at $z > 1$. Moreover, they measured the close-pair fractions within the Illustris simulation [129] and inferred that the timescale should evolve as $T \propto (1+z)^{-2}$ to match the intrinsic merger rates trend from [130].

In Figure 12, we re-compute the merger rates for the fiducial mass-ratio based companion fractions by adopting the two evolving timescale assumptions $T_{\rm obs,pair} \propto H(z)^{-1/3}$ and $T_{\rm obs,pair} \propto (1+z)^{-2}$ from [126, 128], respectively. Since [128] asserts that the close-pair observability is mainly dynamical time dominated at $z < 1$, which results in an un-evolving timescale value, we choose to evolve the $T_{\rm obs,pair}$ only at $z > 1$ when re-computing the merger rates. We find that the major merger rates based on [126] evolving timescale still does not sufficiently agree with the theoretical predictions. On the other hand, when we use simulation-tuned snyder17 prescription of $T_{\rm obs,pair} \propto (1+z)^{-2}$, we find that the merger rates agree closely to model predictions of [13, 130], and closely mimic the [25, 26] halo-halo merger rates. The fact that our evolving timescale based

empirical merger rates agree with the theoretical predictions from [130] implies a very good agreement between our $f_{\rm mc}(z)$ and [128] fractions. Additionally, we argue that the disagreement between [17, 27] flat merger rate evolution and the [13, 130] theoretical predictions (see § 5.2) may be due to the un-evolving close-pair observability timescale assumption. These results imply that an evolving timescale assumption may be more appropriate when measuring merger rates at $z > 1$, and it may be necessary to explain the discrepancy between a constant timescale based rates and theoretical model predictions.

Additionally, we also stress the difference in the measured quantities between the simulations and empirical close-pair-based studies. In theoretical simulations, a merger is counted towards the intrinsic merger rate when the merging process is concluded, when the remnant of the progenitor galaxies is formed, and therefore it is an instantaneous measure. On the other hand, the close-pair method probes future merging systems, within some time-frame after observing them. [128] found that the remnant formation time also evolves as $(1 + z)^{-2}$, which is defined as the time prior to the formation of the remnant, when the merger is observed as a close pair. We follow Section 4.1 in [128] and compute instantaneous remnant formation rates at $z > 1$, which is conceptually similar to the instantaneous merger rates quoted by the simulations. We find that remnant formation rate follows an increasing redshift trend, which is in qualitative agreement with finding by [128]. We acknowledge that a comprehensive analysis of the empirical and theoretical measurement methods is necessary to make quantitative data-theory comparisons confidently. However, it is beyond the scope of the current study to fully address them. We reserve further discussions on this analysis to an accompanying work (Chapter 3).

1.5.3.3 Checking Consistency of CANDELS team Redshifts to Single Participant Estimates

In § 1.2.3.2, we have discussed adopting the σ_z values based on single participant $P(z)$ as the photometric-redshift errors of $z_{\rm phot}$ to be able to apply the redshift proximity criterion successfully. To choose an appropriate participant for this analysis, we test the consistency of redshift estimates from six participants to the CANDELS team $z_{\rm phot}$ values for the sample of $\log_{10}(M_{\rm stellar}/M_{\odot} \geq 9.7)$ galaxies. We find the redshifts based on S. Wuyts $P(z)$s agree best with the CANDELS team $z_{\rm phot}$ and available $z_{\rm spec}$ values, where the median (of the CANDELS five fields) outlier fraction[10] of 1.5% and 3%, respectively. This agreement is key to our analysis, and thus we choose the S. Wuyts σ_z values centered on their redshifts to estimate the σ_z for the CANDELS team $z_{\rm phot}$ values. For simplicity, in Figure 13, we only show the S. Wuyts redshift estimates vs. team $z_{\rm phot}$.

1.5.4 Future Work

So far, we have discussed plausible evolutionary observability timescale prescriptions that may be necessary when deriving merger rates at $z > 1$. While this may be true, $C_{\rm merg,pair}$ is another key parameter in Equation 1.9 which may also depend on the companion selection criteria. In fact, recent studies (e.g., [109, 134]) find that $C_{\rm merg,pair}$ varies as a function of $R_{\rm proj}$, redshift, and local over-density (often represented by δ). Although, a redshift-dependent timescale prescription sufficiently brings empirical and theoretical merger rates to agreement, it is possible that $T_{\rm obs,pair}$ may be a simultaneous function of

[10]It is defined as the fraction of $\log_{10}(M_{\rm stellar}/M_{\odot} \geq 9.7)$ galaxies that are outliers. We consider a galaxy to be an outlier if $(z_{\rm participant} - z_{\rm phot})/(1 + z_{\rm phot}) > 0.1$

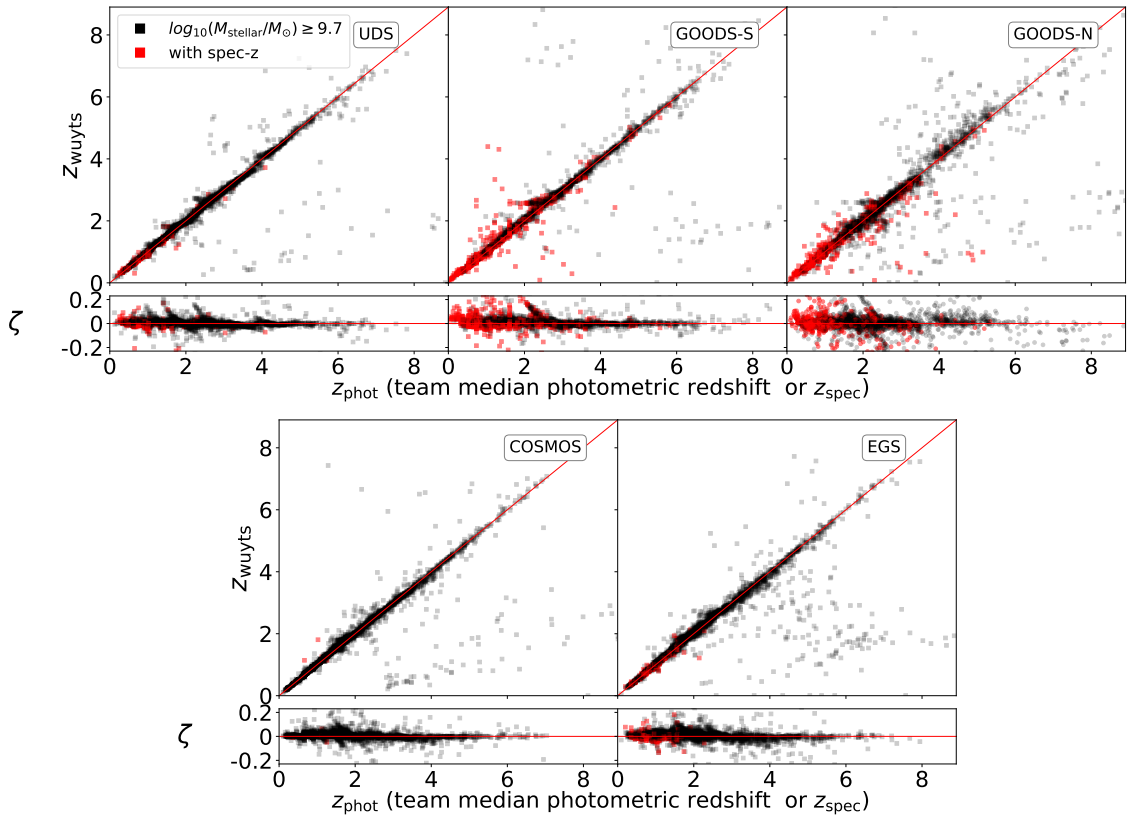

Figure 13: Comparison between the single participant (S. Wuyts) $P(z)$-based redshifts ($z_{\rm wuyts}$) to the CANDELS team $z_{\rm phot}$ values for $\log_{10} M_{\rm stellar}/M_{\odot} \geq 9.7$ galaxies (black squares) and those with spectroscopic redshifts (red squares) for five CANDELS fields. In each panel, the top portion of the figure compares both redshift estimates with one-to-one correspondence (red) line , and the bottom portion visualizes the redshift normalized scatter defined as $\zeta = (z_{\rm wuyts} - z_{\rm phot})/(1 + z_{\rm phot})$ centered on zero. We find that the S. Wuyts redshift estimates best match to the CANDELS team values with a median outlier fraction of 1.5% and 3% for $z_{\rm phot}$ and $z_{\rm spec}$ samples, respectively (see text in Appendix B for details).

several key close-pair variables. As such, these conversion factors may be better represented as $C'_{\rm merg,pair} = C_{\rm merg,pair}(z, R_{\rm proj}, \delta), T'_{\rm obs,pair} = T_{\rm obs,pair}(z, R_{\rm proj}, M_1/M_2, M_{\rm stellar}, \delta)$, where $C'_{\rm merg,pair}$ and $T'_{\rm obs,pair}$ are the evolving prescriptions of $C_{\rm pair}$ and $T_{\rm obs,pair}$, respectively. To fully quantify the interplay among the companion selection criteria and the fraction-to-merger-rate conversions, it is important to derive detailed calibrations for $C'_{\rm merg,pair}$ and $T'_{\rm obs,pair}$ using large-scale cosmological simulations.

Furthermore, it has been speculated that stellar-mass ratio may be a poor representer of ongoing "significant" mergers, owing to the dominant cold-gas contribution to the total baryonic mass (gas+stellar mass) of the galaxy at $z \sim 2-4$ [16, 18]. If this is true, a galaxy merger with $M_1/M_2 > 4$ may in fact be significant if the total baryonic (stellar+cold gas) mass ratio of the galaxies is considered ($M_{1,\rm bar}/M_{2,\rm bar} \leq 4$). This may cause for many of the mergers that can be significant contributors to several aspects of galaxy evolution to be missed because of the selection bias. Empirical confirmation of this speculated selection bias deserves a dedicated analysis of its own.

Semi-Analytic Models (SAMs) are showing a great promise as the testing grounds for many galaxy evolution related questions (e.g., [59, 135]). Three studies, [136–138] have independently developed SAMs based on Bolshoi-Plank N-body simulation [139] and produced mock datasets that mimic observational data in the five CANDELS fields. Additionally, multiple realizations of each CANDELS mock field are also made available. We will take advantage of the intrinsic merger history information and apply close pair analysis on these mock datasets to derive the necessary calibrations. With the help of multiple realizations per field, we will be able to quantify the impact of sample variance

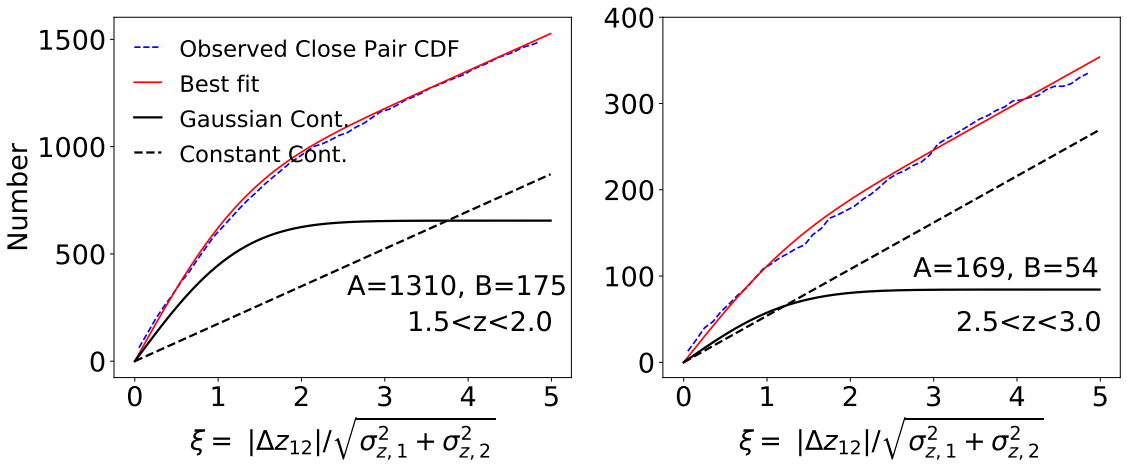

Figure 14: The cumulative distribution functions (CDFs) of $\xi = \Delta z_{12}/\sqrt{\sigma^2_{z,1} + \sigma^2_{z,2}}$ (from Eq. 1.1) for projected pairs satisfying our fiducial $5\,\mathrm{kpc} \leq R_{\mathrm{proj}} \leq 50\,\mathrm{kpc}$ selection (blue dashed, narrow line) as described in § 1.5.3.1 for two redshift bins: $1.5 < z < 2$ (left) and $2.5 < z < 3$ (right). In each redshift-bin panel, following the equation $N(\xi \leq c) = A/\sqrt{2\pi}\int_0^c \exp^{-x^2/2}\,dx + Bc$, the best-fit curve to the CDF is shown in narrow solid-red line. The break-down of the best-fit curve into Gaussian and background contributions are shown in bold solid-black and bold-dashed lines, respectively, and we print the best-fit values (A and B) in each panel.

towards close-pair statistics and merger rates. Also, by introducing realistic, CANDELS-like random and systematic errors onto simulated z_{phot} and M_{stellar} quantities, we will carry out a comprehensive investigation of their impact on the measured merger rate evolution. These calibrations and analyses will help fill the gaps in our understanding of the merger rate measurements and will provide most up-to-date major merger rate constraints with future state-of-the-art telescopes. We will discuss this in our Chapter 3.

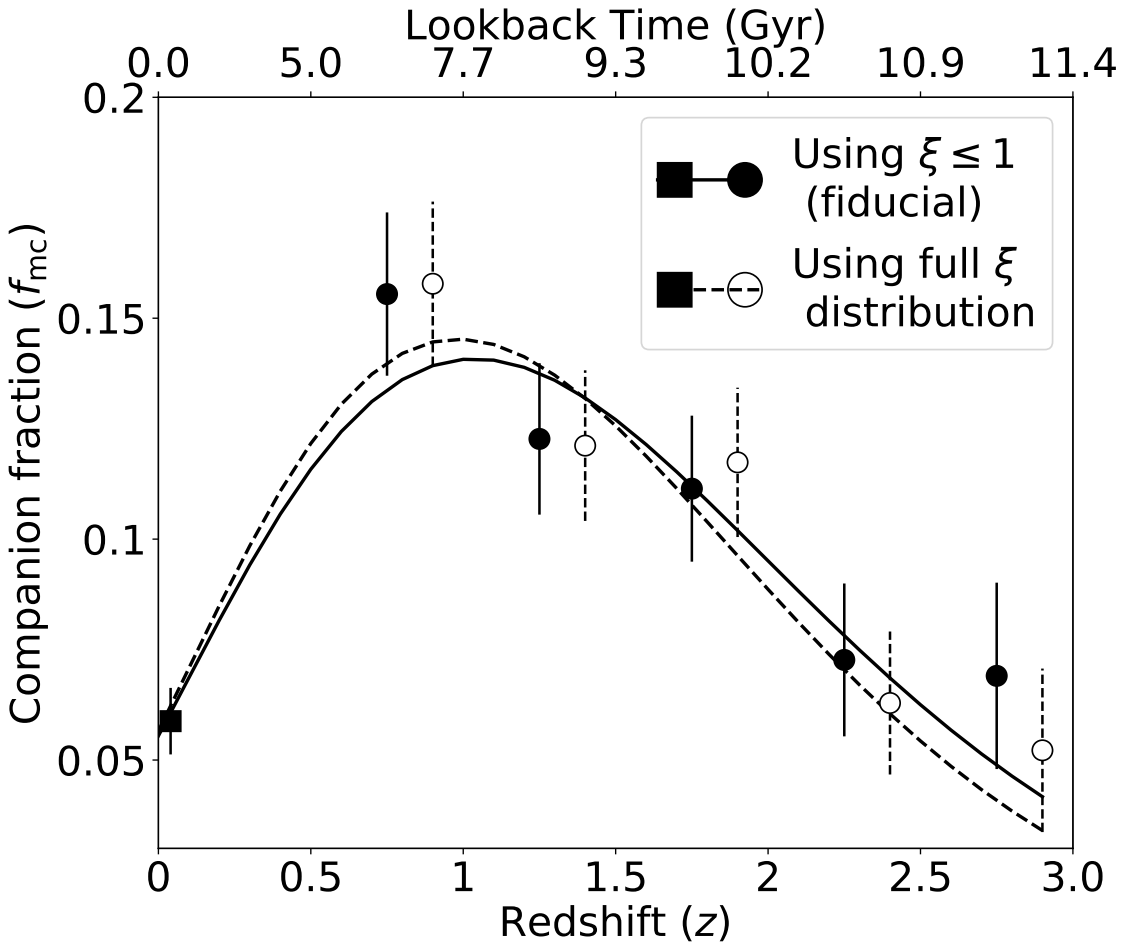

Figure 15: Comparing the redshift evolution of the fiducial major companion fraction $f_{\rm mc}(z)$ (filled circles; copied from Figure 5) to the fractions after correcting for full $\xi = |\Delta z_{12}|/\sqrt{\sigma_{z,1}^2 + \sigma_{z,2}^2}$ distribution usage (open circles; see § 1.5.3.1 for details) . For this exercise, we anchor both CANDELS $f_{\rm mc}(z)$ to the same fiducial selection based SDSS data-point at $z \sim 0$. The error-bars correspond to 95% binomial confidence limits on the $f_{\rm mc}$.

1.6 Conclusions

In this work we analyze a large sample of nearly 9800 *massive* ($M_{\rm stellar} \geq 2 \times 10^{10} M_{\odot}$) galaxies spanning $0 < z < 3$ using the five *Hubble Space Telescope* CANDELS fields (totaling $\sim 0.22\,{\rm deg}^2$) and volume-matched region of the SDSS surveys to quantify the redshift evolution of major companion fraction $f_{\rm mc}(z)$. We adopt a fiducial selection criteria of projected separation ($5 \leq R_{\rm proj} \leq 50\,{\rm kpc}$), redshift proximity $\Delta z_{12} \leq \sqrt{\sigma_{z,1}^2 + \sigma_{z,2}^2}$ (CANDELS) and $\Delta v_{12} \leq 500\,{\rm km\,s^{-1}}$ (SDSS), and stellar mass-ratio major companion selection criterion $1 \leq M_1/M_2 \leq 4$ (MR).

Our key result is the MR-based major companion fraction increases from 6% ($z \sim 0$) to $\sim 16\%$ ($0.5 < z < 1$) and decreases to $\sim 7\%$ ($2.5 < z < 3$), indicating a turnover at $z \sim 1$ with power-law exponential (Equation 1.7) best-fit values $\alpha = 0.5 \pm 0.2$, $m = 4.6 \pm 0.9$, and $\beta = -2.3 \pm 0.5$. We perform comprehensive tests demonstrating that these evolutionary trends are robust to changes in companion selection criteria except for H-band flux ratio (FR) selection, where FR-based $f_{\rm mc}(z)$ agree with MR-based fractions up to $z \sim 1$, but disagree at $z > 1$ as they increase steadily with redshift as $f_{\rm mc}(z) \propto (1+z)^1$ up to $z = 3$, confirming previous speculations [6, 120]. This disagreement is due to increasing contamination of FR selection by minor companions (MR> 4) from 40% at $1 < z < 1.5$ to over three-quarters at $2.5 < z < 3$, confirming the result by [16] that significantly different mass-to-light ratio properties of the companion galaxies at $z > 1$ may be the main cause for the contamination by minor companions.

We compute major merger rates for MR and FR-based fractions by using a constant fraction-to-merger-rate conversion $C_{\rm merg,pair} = 0.6$ and a non-evolving close-pair

observability timescale $T_{\rm obs,pair} = 0.65\,$Gyr [127]. We find that the MR-based rates follow an increasing trend with redshift up to $z \sim 1.5$, after which they decline up to $z \sim 3$. On the other hand, the FR-based rates follow a rising trend with redshift over our full redshift range $0 < z < 3$. We also re-compute the MR-based rates using two additional theoretically-motivated, redshift dependent timescales: $T_{\rm obs,pair} \propto H(z)^{-1/3}$ [126] and $T_{\rm obs,pair} \propto (1+z)^{-2}$ [128].

If recent cosmologically-motivated merger simulations are representative of nature, our results indicate the strong need for improved understanding of how MR and FR estimates trace the galaxy mass (gas+stars) and halo mass ratios of merging systems as a function of redshift. This analysis underscores the strong need for detailed calibrations of these complex, presumably evolving prescriptions of $C_{\rm merg,pair}$ and $T_{\rm obs,pair}$ to constrain the empirical major merger rates confidently. In addition, $C_{\rm merg,pair}$ and $T_{\rm obs,pair}$ may be evolving not only as a function of redshift, but also stellar-mass, companion selection criteria ($R_{\rm proj}$, M_1/M_2), and environment (local over-density), which we pursue through a comprehensive analysis of close-pair statistics using mock datasets from the Semi-Analytic Models in Chapter 3.

Finally, we recompute $f_{\rm mc}(z)$ for different companion selections and we find good agreement with previous empirical estimates when we match their companion selection criteria. This confirms the report by [18] that part of the reason for the observed study-to-study differences is due to close-pair selection mismatch. Additionally, our comprehensive analysis of the impact of different selection criteria and statistical corrections on $f_{\rm mc}(z)$ measurements yields the following results:

- The redshift evolution of major companion fractions is qualitatively robust to changes in projected separation criterion, with smaller $R_{\rm proj}$ resulting in lower fractions as expected. However, the $5-30$ kpc selection ($1/2$ the area of $5-50$ kpc selection) yields similar results to $14-43$ kpc criterion owing to the increased probability of physical companions at smaller projected separations.

- Moreover, the redshift evolution of major companion fractions is qualitatively robust to changes in redshift proximity criterion. While the modified man2012 (§ 1.4.2) proximity-based fractions are systematically $\sim 11-24\%$ higher than the fiducial values, the hybrid B09-based proximity (§ 1.4.2) yields $9-17\%$ smaller major companion fractions at $0.5 < z < 1.5$ and nearly identical values at $z > 1.5$ when compared to fiducial B09 (Equation 1.1) based fractions, owing to the sparsity of spectroscopic-redshift coverage in CANDELS at these redshifts.

- We recompute the SDSS $f_{\rm mc}(0)$ evolutionary anchor using the modified man2012 and B09 redshift proximity criteria applied to simulated $z_{\rm phot}$ errors similar to those for $z > 0.5$ CANDELS galaxies. We find the recomputed $z \sim 0$ companion fractions are statistically equivalent ($\Delta f/f_{\rm fid} \sim 3\%$) to the fiducial $f_{\rm mc}(0)$ values, which demonstrates a high probability of small velocity separations ($\Delta v_{12} \leq 500\,{\rm km\,s^{-1}}$) for companions in $5-50$ kpc projected pairs in the SDSS.

- When major companion fractions are corrected for random chance pairing, we find negligible corrections (1%) for SDSS owing to our stringent $\Delta v_{12} \leq 500\,{\rm km\,s^{-1}}$ proximity selection, and approximately equal corrections of $\sim 20\%$ at $0.5 < z < 3$

for CANDELS. Importantly, our key results regarding the evolutionary trends of major companion fractions are qualitatively the same with or without this statistical correction.

- We also demonstrate that the B09 redshift proximity method does not exclude close-pair systems owing to our use of a simple 68% confidence limit to represent the full $z_{\rm phot}$ PDF.

- Finally, we demonstrate that MR-selected merger rates using un-evolving prescriptions of $C_{\rm merg,pair}$ and $T_{\rm obs,pair}$ agree with the theoretical predictions at $z < 1.5$ but strongly disagree at $z \gtrsim 1.5$. If we re-compute our MR-based rates using an evolving, dynamical-time motivate timescale by jiang14, we find the merger rates are still lower than and disagree with the theoretical merger rate predictions at $z > 1.5$. However, if we use a redshift-dependent timescale $T_{\rm obs,pair}(z) \propto (1+z)^{-2}$ motivated by [128], our MR-based rates agree with the theory at redshifts $0 < z < 3$.

CHAPTER 2

STUDYING THE PHYSICAL PROPERTIES OF TIDAL FEATURES I. EXTRACTING MORPHOLOGICAL SUBSTRUCTURE IN CANDELS OBSERVATIONS AND VELA SIMULATIONS

Published in *Mantha et al., 2019, MNRAS, 486, 2643*

DOI: 10.1093/mnras/stz872

Abstract

The role of major mergers in galaxy evolution remains a key open question. Existing empirical merger identification methods use non-parametric and subjective visual classifications which can pose systematic challenges to constraining merger histories. As a first step towards overcoming these challenges, we develop and share publicly a new Python-based software tool that identifies and extracts the flux-wise and area-wise significant contiguous regions from the model-subtracted "residual" images produced by popular parametric light-profile fitting tools (e.g., GALFIT). Using Hubble Space Telescope (*HST*) H-band single-Sérsic residual images of 17 CANDELS galaxies, we demonstrate the tool's ability to measure the surface brightness and improve the qualitative identification of a variety of common residual features (disk structures, spiral substructures, plausible tidal features, and strong gravitational arcs). We test our method on synthetic *HST* observations of a $z \sim 1.5$ major merger from the VELA hydrodynamic simulations. We extract *H*-band residual features corresponding to the birth, growth, and fading of tidal

features during different stages and viewing orientations at CANDELS depths and resolution. We find that the extracted features at shallow depths have noisy visual appearance and are susceptible to viewing angle effects. For a VELA $z \sim 3$ major merger, we find that James Webb Space Telescope NIRCam observations can probe high-redshift tidal features with considerable advantage over existing *HST* capabilities. Further quantitative analysis of plausible tidal features extracted with our new software hold promise for the robust identification of hallmark merger signatures and corresponding improvements to merger rate constraints.

2.1 Introduction

Merging of two similar-mass (stellar-mass ratio $\leq 4:1$) galaxies is often referred to in the literature as major merging. Identifying and quantifying the rate at which galaxies experience such mergers (merger rates) over cosmic history is a key step to empirically quantify the role of major mergers in galaxy evolution. Conceptually, two broad observational approaches are employed to quantify major merging, namely: (1) *close-pair methods* and (2) *morphological methods*. These approaches yield redshift-evolutionary trends that broadly agree with theoretical merger rate predictions that major mergers become more frequent with increasing redshift out to $z < 1.5$. However, both methodologies suffer from unique, yet analogous systematic biases from selection effects and uncertain observability timescale assumptions, making it difficult to interpret the merger rates at $z > 1.5$. [127] investigated the effects of such systematics in binary merger simulations using non-parametric identifiers of plausible merging (e.g., Gini-M_{20} and CAS metrics).

In this study, we attempt to help overcome the subjective identification of 'hallmark' merger signatures (e.g., tidal arms, tails, bridges, and extended fans) and to robustly quantify their observability timescales – the largest sources of uncertainty towards merger rate estimation in morphology-based methods – by developing a new public `Python` analysis tool to *extract and quantify* the morphological substructure of galaxies. When applied to both empirical and simulated data, this tool will facilitate future efforts to calibrate the observables associated with galaxy merging signatures and lead to improved merger rate constraints.

Substantial efforts have attempted to constrain major merger rates as a function of cosmic time using empirical "merger fractions" based on close-pair statistics, typically identifying galaxy-galaxy pairs in close physical proximity [17, 140] or by measuring clustering statistics [7, 8]. This method leverages on the long-standing idea (supported by numerical expectations) that galaxies in close proximity become gravitationally bound and merge into a more-massive system [111, 125, 141–143]. Many previous close-pair-based works broadly agree that merger rates grow more frequent at earlier cosmic times during the redshift range $0 \lesssim z \lesssim 1.5$, albeit with a wide range of redshift dependencies $\propto (1+z)^{0.5-3}$ [3–5,61,62,109,144–146]. Such variance is largely caused by different study-to-study close-pair selection assumptions [18]. In [140], we systematically quantified the effect of different close pair selection choices on the derived merger fractions.

Recent efforts extend the close pair method to higher redshifts and find that merger rates may be un-evolving or diminishing with increasing redshift between $1.5 \leq z \leq 3$ [15–17, 140]. These measurements disagree with theoretical predictions of steadily

increasing merger rates from $z = 0$ to $z = 3$ [13, 130]. Leveraging recent simulation-based work by [128], observational studies argue that the data-theory discrepancy at $z > 1.5$ may be due to the assumption of simplistic (non-evolving) close-pair observability timescale employed in merger rate calculations [140, 147]. What is needed is a thorough analysis of the observability timescale evolution for both close-pair and morphological methods. In a related study (Chapter 3), we are constraining the close-pair observability timescale as a function of different pair selection variables. The residual feature extraction tool that we introduce in this paper will aid new efforts to constrain the morphological feature observability timescales.

Besides merger rates derived from close pairs, rates based on morphology have also been measured motivated by simulations showing that merging galaxies exhibit morphologically disturbed appearance [69, 70, 148]. These merger constraints can be broadly categorized into rates based on visual identification of disturbed morphologies [42, 149, 150] and quantitative metrics of large-scale galaxy asymmetries or morphology such as CAS (first introduced by [151, 152] and used by [78, 116, 153], and Gini-M_{20} introduced by [154] and applied extensively by [9, 10]. Early morphology-based studies find a broad range of merger rate evolutions $R \propto (1+z)^{2-5}$ at $z \lesssim 1.5$ [18], albeit with significant study-to-study scatter where some studies find no redshift dependence of merger rates [9, 79], and sometimes finding up to $25\% - 50\%$ of their samples to be merging [116]. These studies suffer from low-number statistics due to small-volume pencil-beam surveys, redshift-dependent systematic biases induced by cosmological surface brightness dimming and strong morphological *k*-corrections when using rest-frame

UV images, which can impact the purity and completeness of merger selection. Although the advent of large Hubble Space Telescope (*HST*) surveys like CANDELS [20, 21] alleviate some issues, the observability timescales for morphology-based methods are highly uncertain and may be varying with redshift.

A possible solution to overcome the systematic, purity, and timescale related challenges faced by the aforementioned morphology-based approaches is to employ more focused morphological signatures as merger identifiers. For example, [155] used the presence of multiple-nuclei separated by $\lesssim 8\,\mathrm{kpc}$ to select merging galaxies during a narrow window straddling the early and post-merger stages ([few hundred Myr timescale; [18]), while aiming to minimize the contamination from spurious non-merging interlopers. Tidal features are a specific sub-class of morphological substructures that are predicted to be prevalent among merging galaxies [148, 156, 157]. A fruitful technique for improving the detection of faint and transient tidal features is through the analysis of residual images [1] either by visual inspection [42, 43] or quantitative parameters [73, 159]. In fact, galaxy light-profile fitting is a popular step employed in many large surveys to provide key insights into the structural (size, shape) evolution of galaxies [121, 160, 161]. In this context, our novel tool is designed to analyze the by-product residual images from galaxy image fitting routines and extract useful information about additional (often complex) substructures.

Detailed objective distinction of *hallmark* tidal features from other non-merging signatures (e.g., lopsided disks, asymmetric or one-armed spiral structures) is yet to be

[1] Produced by subtracting the best-fit light profile of the host galaxy using tools such as `GALFIT` [158].

achieved. Existing empirical merger-identification methods only provide a *plausible* indication of tidal signatures and do not attempt to quantitatively capture key information such as structure (length, width, and pitch), morphology (e.g., tails and fans), color, and stellar mass of the features. Furthermore, constraining the role of observational effects (e.g., depth, cosmological surface-brightness dimming) and observability timescales $T_{\rm obs}$ for *hallmark* tidal features is an open question, and is a key hurdle in achieving robust tidal-feature-based merger rates.

As a first step towards quantitatively identifying and extracting tidal signatures of galaxy merging, in this study, we introduce a new technique and its corresponding public `Python` software pipeline[2] to extract and quantify *residual substructures* and we demonstrate its application with a select sample of 16 CANDELS galaxies that exhibit a range of merging and non-merging morphological substructures. We also test our technique's ability to identify plausible interaction signatures by applying it to mock Hubble Space Telescope (*HST*) observations of a galaxy merger from the VELA zoom-in hydrodynamic simulations Ceverino14,zolotov15. We structure this paper as follows: In § 2.2, we introduce our CANDELS test galaxy sample and their GALFIT-derived data products, followed by a brief description of a galaxy merger from the VELA hydro-dynamic simulation and the generation of its mock *HST* observations. In § 2.3, we describe the step-by-step methodology of our residual substructure analysis pipeline, and we apply this method to the select CANDELS galaxies in § 2.4. In § 2.5, we apply our methodology to mock observations of a merger simulation and investigate the role of image depth

[2]The data products discussed in this chapter along with the software are publicly available at `https://github.com/AgentM-GEG/residual_feature_extraction`

and viewing angle on the feature extraction process. In § 2.6, we present possible applications and limitations of our method. We present our conclusions in § 2.7. Throughout this work, we adopt the [162] cosmological framework ($H_0 = 67.7\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$), and use the AB magnitude system [85].

2.2 Data and Test Samples

To develop and demonstrate our novel method for extracting and quantifying residual substructures, we use *HST*/WFC3 H-band (F160W) images and single-Sérsic `GALFIT` residual images from [160] for a select sample of galaxies from the CANDELS survey [20,21] that exhibit different residual substructures including plausible tidal signatures. To explore the application of our method for repeatable identification and extraction of bonafide tidal features, we also analyze mock images derived from a merging galaxy from the VELA zoom-in simulations [164].

2.2.1 CANDELS Data

We make use of the H-band Sérsic-fitting data products from [160] (hereafter vdW12) for galaxies bright than $H = 24.5\,\mathrm{mag}$ in the CANDELS survey. Briefly, vdW12 used GALFIT [158] to generate an image-cube containing the *original* image used for fitting, the best-fit *model* image with the Sérsic model information stored in its header, and the *residual* (original−model) image. Following standard procedures outlined in [165], vdW12 used `SExtractor` [166] to extract sources from the CANDELS imaging [21], estimated the background sky, and determined GALFIT parameter input values, and produced a rectangular image cutout of each galaxy based on its SExtractor parameters: x, y

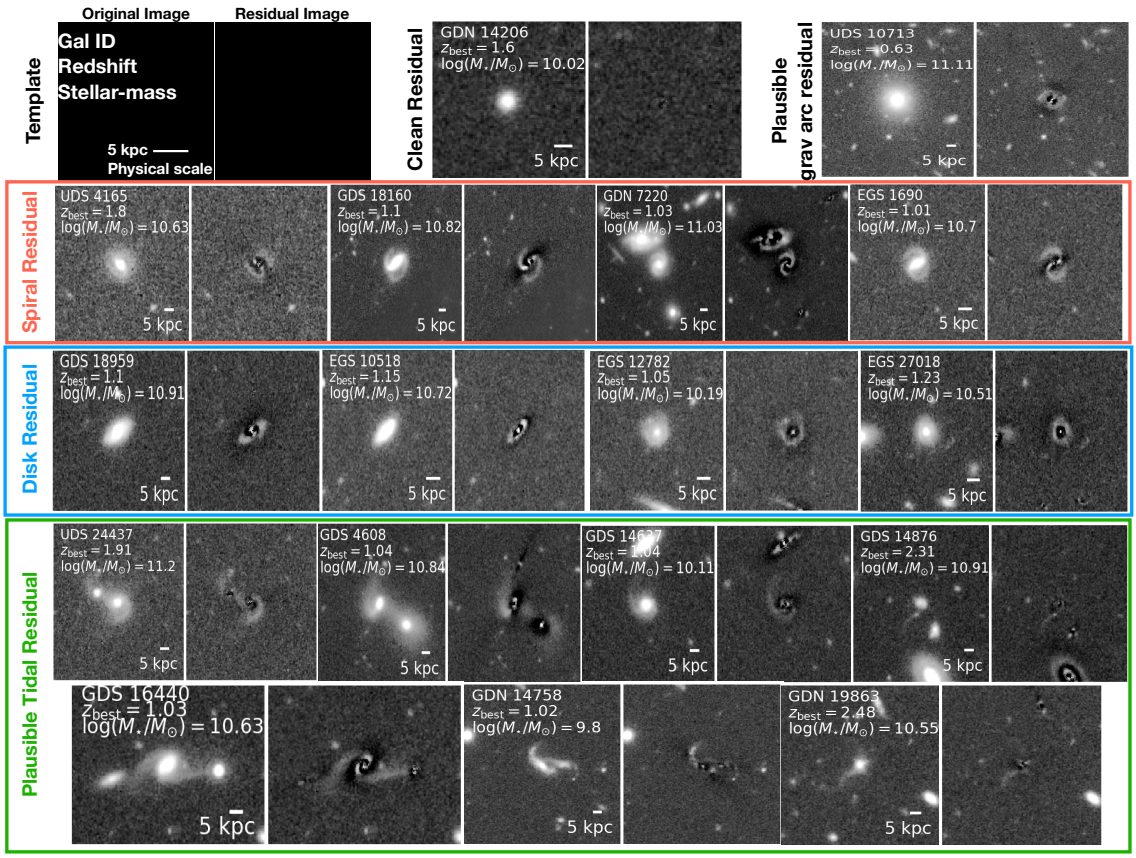

Figure 16: Visualization of our demonstrative galaxy sample (§ 2.2.1, Table 6) hosting different residual features – Clean (consistent with the background; *top-row, middle panel*), plausible gravitational arc residual from [163] (*top row, right panel*), spiral arms (*second row*), left-over disk (*third row*), and plausible interaction signatures (*forth row*). We also visually illustrate each panel with a template, where for each galaxy, we show the H-band (F160W) image (left) and its corresponding single-Sérsic residual (model-subtracted) image from vdW12 (right). We also report the official CANDELS identification number, the $5\,\mathrm{kpc}$ physical scale at the galaxy's best-available redshift (z_{best}), and stellar mass.

centroid, long side length of five (5) times the Kron radius, and orientated by the position angle θ. In practice, the postage stamp images typically have sides $20 - 30$ times the galaxy's effective size ($r_{\rm e}$). To ensure optimal galaxy profile fitting with GALFIT, vdW12 masked neighboring objects 4 mag fainter than the central galaxy and simultaneously fit all remaining objects in the image cutout.

The methodology we describe in § 2.3 starts with the vdW12 image cubes for a select sample of 16 galaxies from an ongoing effort (HST-AR 15040; PI: McIntosh) to visually characterize the H-band residual substructures hosted by CANDELS galaxies with $9.8 \lesssim \log_{10}(M_{\rm stellar}/M_{\odot}) \lesssim 11$ and $0.5 \lesssim z \lesssim 2.5$. As shown in Table 6, these galaxies sample different kinds of residual features: disk (4 galaxies), spiral arms (4 galaxies), plausible tidal signatures (7 galaxies), and gravitational strong-lensing arc (1 galaxy). We also note that our galaxies span two observational depth regimes – CANDELS/Wide (*HST* 2-orbit; 6 galaxies) and CANDELS/Deep (*HST* 10-orbit; 10 galaxies) depths as described in [20, 21]. For each galaxy in our sample, we extract residual substructures using the images with best-available depth, unless otherwise specified. In Figure 16, we provide example postage-stamps of our test set galaxies and their respective residual features to be extracted and analyzed in this paper. For reference, we also show an example 'clean' (i.e., visually no left-over) residual image.

Table 6: Test sample of galaxies hosting different residual substructures which we use to demonstrate our feature extraction technique. Columns: (1) the CANDELS field identifier – UDS, GOODS-South (GDS), GOODS-North (GDN), EGS, followed by the galaxy's official CANDELS identification number; (2–3) celestial coordinates in degrees; (4) the best-available redshift for the galaxy (see for details [93]); (5) the stellar mass [94,95]; (6) the type of residual substructure hosted by the galaxy; (7) the observational depth based on the official CANDELS exposure maps as described in [21]. † denotes that galaxy is involved in a close-pair system. ‡ indicates that we use the galaxy to demonstrate a deep vs. shallow feature-extraction comparison.

ID	R.A.	Decl.	$z_{\rm best}$	Stellar Mass	Residual Feature	Obs. Depth
(1)	(2)	(3)	(4)	(5)	(6)	(7)
UDS 4165	34.580786	−5.25419	0.985	10.63	spiral	shallow
UDS 24437†	34.339863	−5.149623	1.823	11.2	plausible tidal	shallow
GDS 4608†	53.081307	−27.871586	1.067	10.84	plausible tidal	deep‡
GDS 14637	53.04408	−27.785048	1.045	10.11	plausible tidal	deep‡
GDS 14876	53.11879	−27.782818	2.309	10.91	plausible tidal	deep‡
GDS 16440†	53.050929	−27.772409	1.033	10.63	plausible tidal	deep
GDS 18160	53.163589	−27.758952	1.095	10.82	spiral	deep
GDS 18959	53.202349	−27.751279	1.113	10.91	disk	deep

GDN 7220	189.172373	62.191548	1.019	11.03	spiral	deep
GDN 14758[†]	189.358539	62.241522	1.021	9.8	plausible tidal	deep
GDN 19863	189.154657	62.274604	2.482	10.55	plausible tidal	deep
EGS 1690	215.092284	52.922091	1.081	10.7	spiral	shallow
UDS 10713	34.275997	−5.22157	0.627	11.11	lensing arc	shallow
EGS 10518	214.877124	52.819527	1.192	10.72	disk	shallow
EGS 12782	215.20786	53.064729	1.103	10.19	disk	deep
EGS 27018	215.006664	52.996171	1.244	10.51	disk	shallow

2.2.2 VELA Datasets

To demonstrate our new methodology on the extraction of residual substructure signatures hosted by a galaxy-galaxy major merger, we apply it to mock images from VELA#01[3] of the VELA zoom-in hydro-dynamic simulations [164, 167]. Briefly, in this simulation baryonic physics[4] are solved in conjunction with gravitational N-body interactions using the Adaptive Refinement Tree (ART) code [168–170] with spatial resolution

[3]The VELA simulation-suite comprises of 35 intermediate-mass galaxies ($\log_{10}(M_{\rm stellar}/M_{\odot}) \sim 10$ at $z \sim 2$) and they are referred using an ID starting from 01 to 35.

[4]The VELA simulations include many physical prescriptions related to galaxy evolution – gas cooling, star formation, ISM metal enrichment, photo-ionization heating from the Ultra-Violet background radiation, stellar-mass loss, stellar feedback, tracing of smooth cosmological accretion of other galaxies and gas.

of $17.5\,\mathrm{pc}$ to $35\,\mathrm{pc}$. The VELA simulation-suite has been extensively used in the context of galaxy morphology such as elongated galaxies [82,171], giant clumps [172–174], compaction [167,175,176], galaxy spin and size Jiang18, H-alpha morphology [177,178], and structural parameters [179]. Starting at a cosmological scale factor $a = 0.125$ ($z \sim 7$), the simulation information is recorded at equal increments of $\Delta a = 0.01$, which approximately corresponds to a time frame $\Delta t = 100\,\mathrm{Myr}$ at $z = 2$. For the merging system VELA 01, we use the information during 15 time intervals spanning $0.37 \leq a \leq 0.47$ ($1.7 \lesssim z \leq 1.1$), bracketing roughly 0.6 (0.8) Gyr before (after) the coalescence stage (occurring at $a = 0.41$; $z = 1.44$). We note that this system experiences a gas-rich, major merger (stellar mass ratio $\sim 1.33 : 1$, gas mass ratio $\sim 1 : 1$) resulting in post-merger stellar and gas masses of $\log_{10}(M_{\mathrm{stellar}}/M_{\odot}) = 10.05$ and $\log_{10}(M_{\mathrm{gas}}/M_{\odot}) = 9.55$, respectively.

At each time step, we use mock *HST* WFC3/F160W observations of the merging system generated using the methodology described in [179, 180]. Briefly, they use the raw-simulation information at each time-step and assign stellar population synthesis model-informed (e.g., [96,97]) spectral energy distributions (SEDs) to the stellar particles using `SUNRISE` [181,182] with dust scattering and absorption assuming a dust-to-metals ratio of 0.4 [183] and a Milky Way-like dust grain size distribution ($R_{\mathrm{v}} = 3.1$; [184,185]). At each time-step, they integrate this 3-dimensional data onto a 2D camera plane to produce an idealized image at 19 (camera) orientation angles (for details see Table 1 in [176]), which are then convolved with the point spread function (PSF) generated by `Tiny Tim` (WFC3/F160W; [186]) to reach desired spatial resolution. These mock idealized images

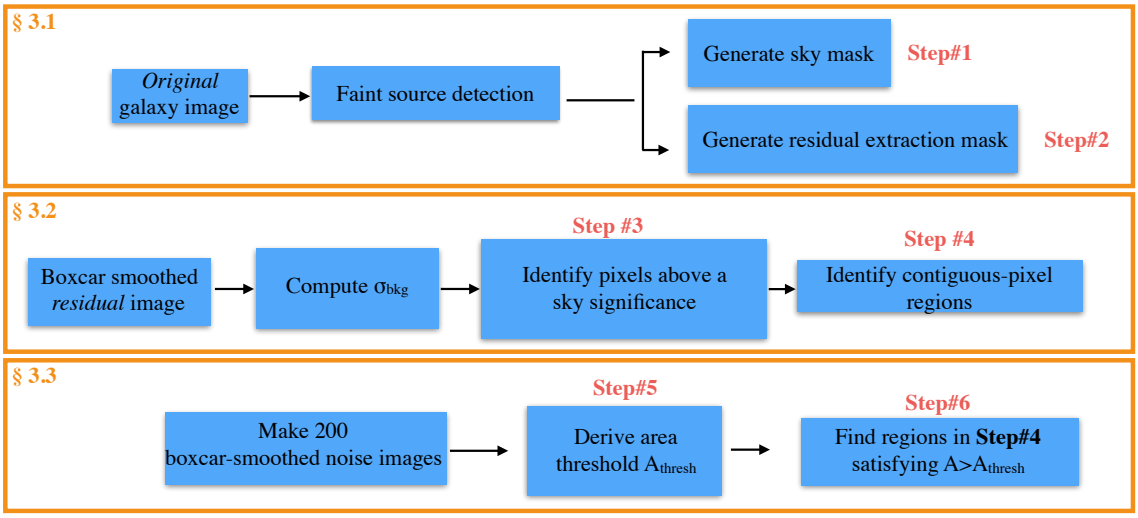

Figure 17: Flow diagram illustrating our detailed step-by-step residual substructure extraction process, which is divided into three main parts (orange boxes), described in their respective sub-sections (§ 2.3). We refer to the key steps (red text) when presenting our method in § 2.3.

are available on MAST[5] [180].

To match the empirical CANDELS F160W 5σ limiting surface-brightness sensitivities at both Wide ($25.25\,\mathrm{mag\,arcsec^{-2}}$) and Deep ($26.25\,\mathrm{mag\,arcsec^{-2}}$) depths, we add Gaussian random noise to the F160W PSF convolved images of VELA 01. This process of generating mock *HST* CANDELS images from simulated data products is often referred to as *CANDELization* Huertas-Company18 and we call the resultant depth-matched images as *CANDELized* images. We apply `GALFIT` to the VELA 01 *CANDELized* images to generate single-Sérsic residual images consistent with the vdW12 empirical data at each time step, viewing angle, and depth realization for a total of 380 unique residual images.

[5] `https://archive.stsci.edu/prepds/vela/`

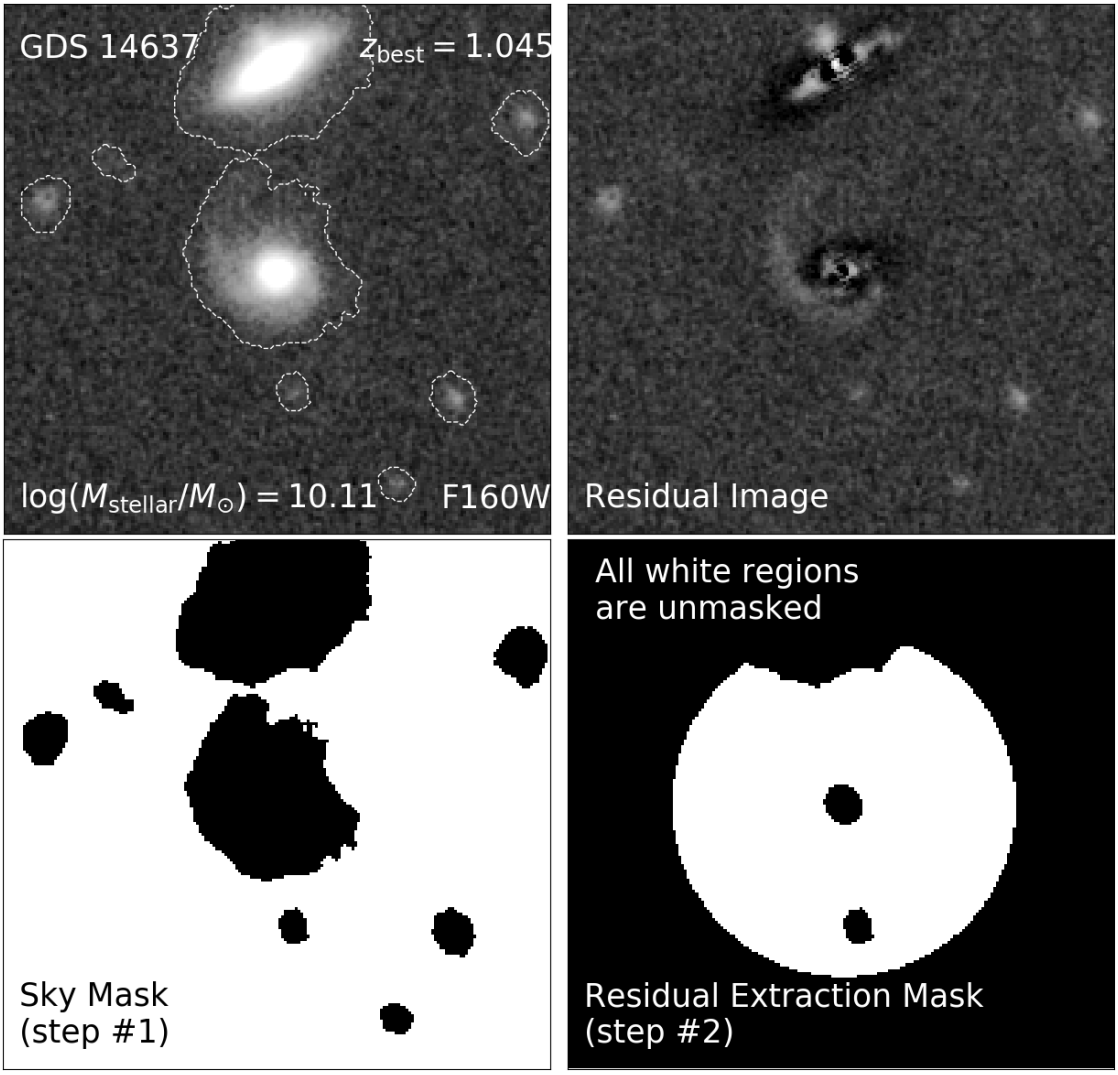

Figure 18: Visualization of the dual masking scheme described in § 2.3.1 for an example galaxy (GDS 14637; at the center of image; see Table 6). *Top left:* WFC3/F160W ($1.6\,\mu$m) image with faint source detections shown with white outlines. *Top right:* single-Sérsic based residual image from vdW12. *Bottom left:* mask generated by step#1 for measuring background sky in unmasked regions (white; pixel values$=1$); masked regions are shown in black (pixel value$=0$). *Bottom right:* mask generated by step#2 for extracting residual substructure.

2.3 Methodology

We demonstrate our residual substructure extraction method by analyzing the GALFIT-based residual images of a selective sample of galaxies hosting different left-over substructures (§ 2.2.1). We organize the functionality of our method into three main objectives as follows. Starting with a *H*-band original image, we generate two masks to facilitate estimation of local-sky background ($\sigma_{\rm bkg}$) and residual feature extraction. Next, we estimate $\sigma_{\rm bkg}$ in the smoothed single-Sérsic-based residual image and search (in proximity to the galaxy-of-interest) for contiguous-pixel regions that satisfy a sky-significance. Finally, we identify the flux-wise significant regions whose areas are larger than a pixel-contiguity area threshold ($A_{\rm thresh}$) informed using noise-only Monte-Carlo (MC) simulations to quantify their *H*-band surface brightness. We present a detailed flow-chart of our residual extraction process in Figure 17 and label the key steps to be discussed in this section.

We acknowledge that our methodology is closely dependent on the outputs generated by different software routines such as GALFIT and Source Extractor. Although these software are extensively used in the astronomy community, their specific functionality may change in the future updated versions. In this context, we emphasize that our modular algorithmic approach described in § 2.3 and illustrated in the flow chart (Figure 17) can be treated as a road map for future applications using the updated or new versions of the software analogous to GALFIT or Source Extractor.

Table 7: Source extraction configuration parameters to inform our masking routine. Columns: (1) name of the parameter; (2) value employed in this study.

Parameter (1)	Value Used (2)
DETECT_THRESH	0.7
MIN_AREA	7
DETECT_NTHRESH	32
DEBLEND_MINCONT	0.0001
CONV_FILTER	2D Tophat (radius = 5)

2.3.1 Masking

We devise a dual masking scheme to facilitate the extraction and isolation of residual substructures at radii outside the each galaxy center. This requires robust estimation of the local-sky background, detection of faint sources not fit by vdW12, and a definition of galaxy center. The magnitude criterion imposed by vdW12 to mask nearby faint objects (§ 2.2.1) is a reasonable choice in the context of quantifying galaxy structural properties. However, such faint un-fit sources can have comparable surface brightness to our desired faint residual substructure (e.g., tidal features). In addition, central galactic regions may house complex substructures (especially in young post-mergers) that may not be well represented by a Sérsic profile and may lead to significant central left-over light [e.g.,][]Hopkins09 that can obscure the detection of faint tidal features of interest.

2.3.1.1 (Step#1) Generate sky mask

We start by performing faint source detection within the postage stamp of the galaxy (original image) using a python-based implementation (SEP; [187]) of the standard source-extraction tool SExtractor [166]. We find that the source extraction parameter

values provided in Table 7 yield optimal identification of small faint sources in the images, without deblending faint extended features (e.g., plausible tidal signatures) into separate objects[6]. Compared to the SExtractor configurations used to generate the CANDELS source catalogs (e.g., see [86, 87]), we employ a smaller convolution filter size (five vs. nine) for better detection of smaller and fainter sources.

In the top panels of Figure 18, we demonstrate the thorough detection of all faint sources seen in the vdW12 residual image of an example galaxy (GDS 14637; center of image). Hereafter, we use this galaxy to demonstrate all of our analysis steps outlined in the flow-chart (Figure 17). In the bottom left panel, we provide the resultant background sky mask. Masked regions are shown in black (pixel value$= 0$). The unmasked regions (white) represent the sky contribution in the vicinity of the galaxy, which use during our feature extraction in § 2.3.2.

2.3.1.2 (Step#2) Generate residual extraction mask

We also generate a mask for extraction of residual substructures within a physically-motivated proximity of $R_{\rm proj} < 30\,{\rm kpc}$ from the centroid of the galaxy-of-interest. This corresponds to nearly five (ten) times the effective size of late (early) type galaxies at $z \sim 1$ [161]. We find that this extraction radius encompasses all faint residual substructure for the galaxies in our sample. As shown in Figure 18 (bottom-right panel), we mask contributions from the central core regions of each galaxy residual so that we extract only outer features. We define the inner masking by an ellipse centered on the galaxy, with

[6]A small value of DETECT_THRESH ensures that the resultant segmentation maps envelop the outermost light of the sources An alternative to this step is to use a higher detection threshold value and grow the resultant segmentation regions to fully encompass the extended light.

semi-major and semi-minor axes equal to 1.5 times those of the SExtractor detection ellipse from step#1, and orientation along the SExtractor position angle. Finally, we also mask any faint companions from step#1 that fall within the $R_{\rm proj} < 30\,{\rm kpc}$ extraction region. In summary, we use this mask during the extraction of residual substructures (§ 2.3.2).

2.3.2 Final Extraction of Residual Features

We start the feature extraction process by analyzing a Boxcar-smoothed version of the residual image (*top-left panel*; Figure 19) to facilitate the robust identification of faint features. We utilize the masks described in § 2.3.1 to identify all unmasked pixels within the residual extraction mask region that are significantly above the background sky. Then, we identify contiguous groupings of such pixels to define unique residual features.

2.3.2.1 (Step#3) Identify flux-wise significant pixels

The first step in extracting residual features is to quantify the background sky. We compute the background $\sigma_{\rm bkg}$ as the standard deviation of the best-fit Gaussian profile representing the normalized distribution of pixel values in the unmasked regions of the smoothed residual image after applying the sky mask (§ 2.3.1). In Figure 19 (*top-right panel*), we show the sky pixel-value distribution (blue histogram) with its corresponding Gaussian curve. We denote the $2\sigma_{\rm bkg}$ with a vertical dashed line.

Next, we apply the residual extraction mask and identify all unmasked pixels in the smoothed-residual image (red histogram in Figure 19, *top-right panel*). In Figure 19 (*bottom-left panel*), we plot all such pixels satisfying a fiducial flux-wise significance cut

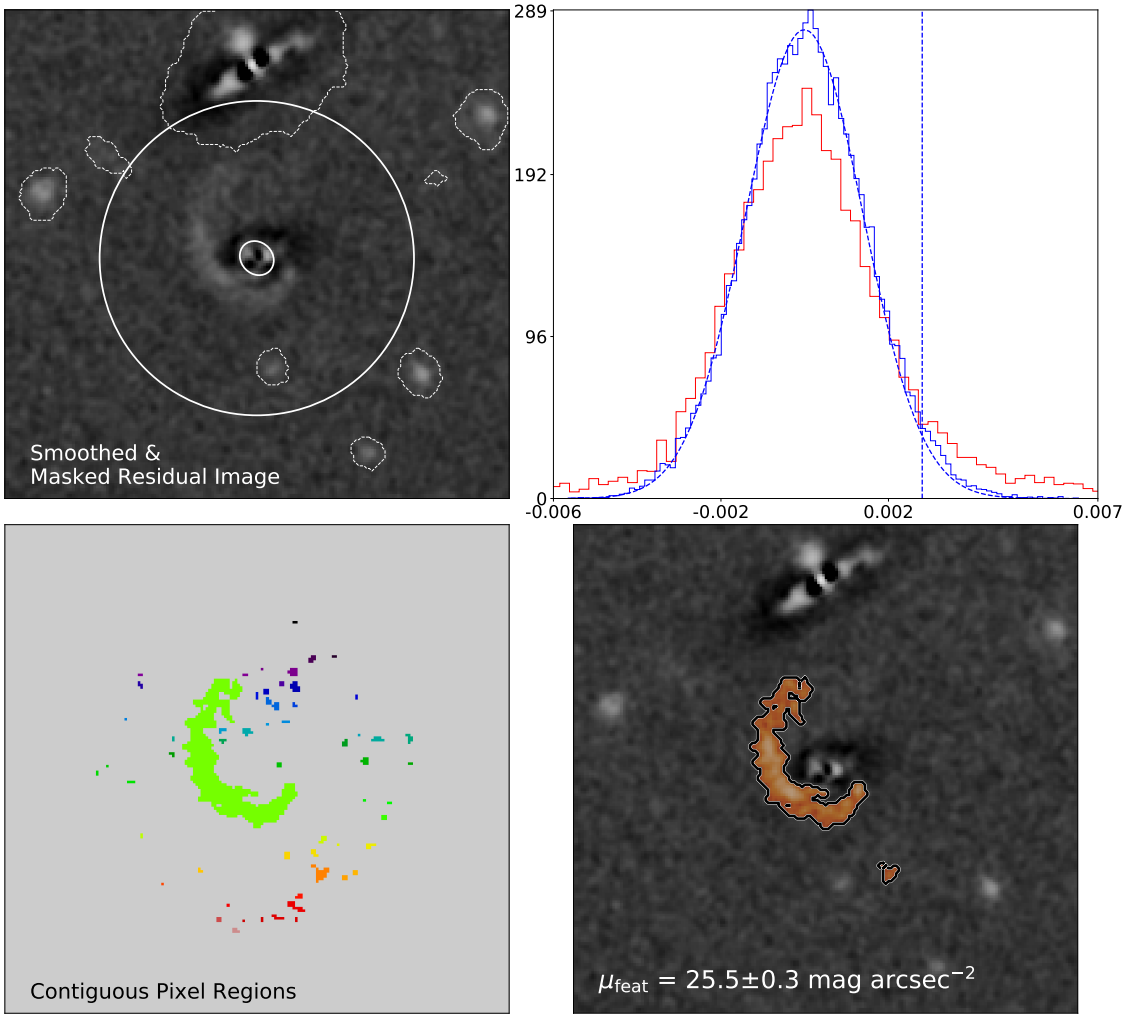

Figure 19: Visualization of the residual feature extraction steps described in § **??** for an example galaxy (shown in Figure 18). *Top-left:* 2D Boxcar kernel-smoothed ($3\,{\rm pix} \times 3\,{\rm pix}$; side length $\sim 0.18''$) residual image with the inner and outer residual feature mask region boundaries (step#1) given in solid white, and faint-source masking (step#2) shown as dashed white outlines. *Top-right:* the normalized pixel-value distribution of the sky background (blue histogram) in units of electrons/sec, its corresponding best-fit Gaussian curve (dashed-blue curve), and $2\sigma_{\rm bkg}$ value (vertical blue line) are compared to the distribution of pixel values from the residual extraction region (red histogram). *Bottom-left:* the pixels residing in the residual extraction region satisfying a flux-wise significance cut $f_{\rm pix} > 2\sigma_{\rm bkg}$ (step#3) are color coded according to their unique contiguity (described in detail in § 2.3.2.2; step#4). *Bottom-right* the final residual features (red shading) after rejecting smaller area regions as described in § 2.3.3 (steps#5 and #6); the feature surface brightness and its 68% confidence level are given.

$f_{\rm pix} > 2\sigma_{\rm bkg}$, where $f_{\rm pix}$ is flux value of the pixel. Visually, we observe one large flux-wise significant residual feature and many small areas of $> 2\sigma_{\rm bkg}$ pixels. We note that a less stringent significance cut of $f_{\rm pix} > 1\sigma_{\rm bkg}$ increases both the number of small pixel groupings and the typical area per pixel-group, albeit with decrease in the overall signal-to-noise ratio. On the other hand, a more strict $f_{\rm pix} > 3\sigma_{\rm bkg}$ cut identifies fewer pixels (with high signal-to-noise), but with significant structural discontinuities (an extended feature is broken into multiple smaller regions). As such, we elect to use $f_{\rm pix} > 2\sigma_{\rm bkg}$ as our fiducial criterion for flux-wise significant pixels in our analysis.

2.3.2.2 (Step#4) Identify contiguous-pixel regions

To isolate and separate larger, contiguous residual substructures from smaller groupings of flux-wise significant pixels, we use a python-based 2D pixel-connectivity algorithm `skimage: label` [188,189]. This tool assigns unique identification number to each contiguous pixel region using a binary image in which all the flux-wise significant pixels are set to unity and the remainder are set to zero. In Figure 19 (*bottom-left panel*), we show the contiguous flux-wise significant pixels (color coded regions).

Next, we compute the area and enclosed flux of these contiguous regions using a companion module `skimage: regionprops`[7] [190, 191]. Re-affirming our visual interpretation, we find a contiguous substructure with large area (shown in green) in the immediate vicinity of our example galaxy and many regions with smaller areas (Figure 19; *bottom-left panel*). We note that correlated-noise fluctuations can cause several pixels to

[7]The module computes cumulative flux per unique contiguous region in the residual image at their respective pixel coordinates.

be flux-wise significant and contiguous by random chance. We quantify this effect and reject such spurious regions in the following section.

2.3.3 Final Feature Selection

We isolate the final, flux-wise and area-wise significant residual features following a MonteCarlo approach. We estimate an area threshold ($A_{\rm thresh}$) based on a noise-only pixel contiguity expectation and select those flux-wise significant contiguous features with areas larger than $A_{\rm thresh}$.

2.3.3.1 (Step#5) Quantify area threshold $A_{\rm thresh}$

As a first step to quantify an area threshold above which the noise contribution to pixel contiguity is minimum, we generate 200 Boxcar smoothed images with Gaussian noise matching the $\sigma_{\rm bkg}$ value from step#3. In each random-noise image, we identify contiguous regions satisfying our fiducial flux-wise significance and quantify their area (as described in § 2.3.2.1). In Figure 20, we show an example smoothed random noise image (*left panel*) and its corresponding flux-wise significant contiguous regions (*right panel*; color coding). We notice that the noise-generated contiguous areas visually mimic the smaller area regions identified in step#4 (Figure 19; *bottom-left panel*). Combining the result of 200 MonteCarlo iterations for our example galaxy, we find an extraction area threshold $A_{\rm thresh} = 20$, where $> 99\%$ of the contiguous pixel regions in smoothed noise-only images have areas smaller than 20 pixels. We repeat this MonteCarlo process for all galaxies in our analysis that span different CANDELS fields and find a range of $A_{\rm thresh}$ values between 17 to 23 pixels, owing to varying field-to-field noise properties.

2.3.3.2 (Step#6) Identify and Quantify Final Residual Features

To select final significant features, we identify the contiguous regions with areas larger than the threshold set by the noise properties ($A > A_{\rm thresh}$). We then quantify the surface brightness of the final feature regions ($\mu_{\rm feat}$) as:

$$\mu_{\rm feat} = -2.5\log_{10}(F) + zp + 2.5\log_{10}(A')\ [{\rm mag\,arcsec^{-2}}]\ , \tag{2.1}$$

where F is the cumulative flux (in Jansky) of all the regions satisfying our criteria, $zp = 8.9$ is the zero-point magnitude, and A' is the cumulative area of all these significant contiguous regions in units of arcsec^2. We also report the photometric error on the quantified surface brightness by adding the noise from $\sigma_{\rm bkg}$ and Poisson noise from the feature in quadrature. In Figure 19 (*bottom-right panel*; shading), we show the final residual features (an extended tidal-tail structure) and quantify its H-band surface brightness to be $\mu_{\rm feat} = 25.5 \pm 0.3\,{\rm mag\,arcsec^{-2}}$.

2.4 Application to CANDELS Galaxies

In this section, we demonstrate the application of our residual substructure extraction method on example CANDELS galaxies hosting a range of features – disk, spiral, and plausible interaction signatures (see Table 6). We also illustrate the role of image depth by comparing results at *HST* 2-orbit and 10-orbit depths for three galaxies (GDS 4608, GDS 14637, and GDS 14876).

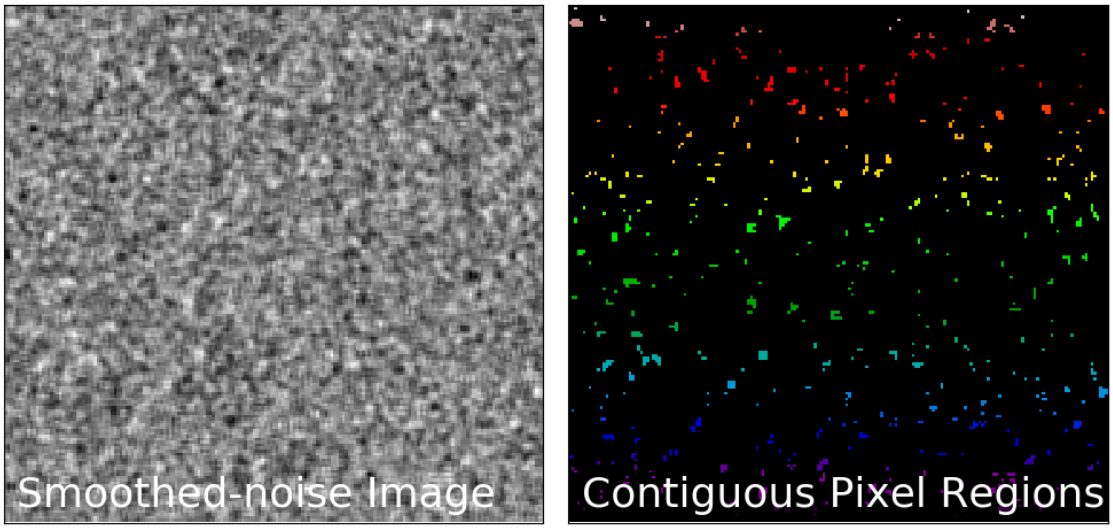

Figure 20: Visualization of the pixel contiguity by random chance (step#5). *Left panel:* the Boxcar smoothed random noise image for one randomly chosen MonteCarlo iteration generated as described in § 2.3.3. *Right panel:* the contiguous pixel regions satisfying our fiducial flux-wise significance cut $f_{\rm pix} > 2\sigma_{\rm bkg}$ (color coding).

2.4.1 Extracting disk and spiral features

In Figure 21 (*top-two rows*), we show the results of applying our feature extraction method on galaxies visually identified to be hosting residual disk structure. In the cases of EGS 10518 and GDS 18959, we find similar looking, inclined disk-like residual structures, whereas the residuals of EGS 27018 and EGS 12782 exhibit circular features (likely owing to their face-on orientation). We also note hints of clump-like substructures embedded within the disk feature of EGS 12782. We find that these features have surface brightness measures ranging between $23.4 - 24.5\,{\rm mag\,arcsec^{-2}}$. Re-affirming our visual interpretation, we find that our method extracts features that are consistent with a residual disk, likely owing to an underlying bulge+disk morphology being fit with a single-Sérsic model.

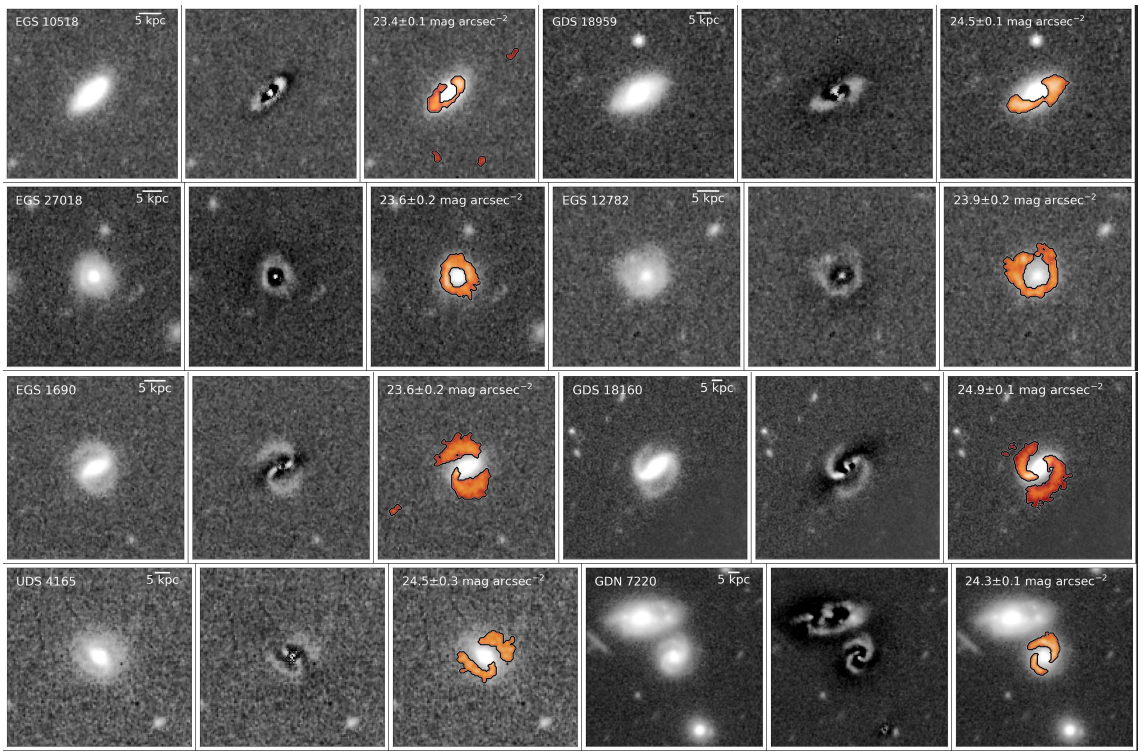

Figure 21: Example residual feature extraction and quantification for select CANDELS galaxies exhibiting disk structure (*top-two rows*) and spiral substructure (*bottom-two rows*). For each galaxy, we show the *original* WFC3/F160W ($1.6\,\mu\mathrm{m}$) image (left) and include the official CANDELS ID and a $5\,\mathrm{kpc}$ physical scale; the single-Sérsic residual image from vdW12 (center); and the extracted residual features (solid-black outlines with red shading) overlaid on its respective *original* image (right), and we include the feature surface brightness measurement.

In Figure 21 (*bottom-two rows*), we show the results of our feature extraction on galaxies hosting residual spiral substructures (EGS 1690, GDS 18160, UDS 4165, and GDN 7220). We extract two-arm spiral patterns for all the four galaxies and find their surface brightness ranges between $23.6 - 24.9\,\mathrm{mag\,arcsec^{-2}}$. We note that the spiral substructures exhibited by GDS 18160 and GDN 7220 are visually apparent (despite having fainter surface brightness) than EGS 1690 and UDS 4165, owing to the difference in their observational depths (deep vs. shallow; see Table 6). We also note that all except GDN 7220 appear to be hosting a central bar. Extracting such spiral substructures with our method will help future analyses quantify their physical parameters such as pitch angle, bar structure.

2.4.2 Extracting plausible tidal features

In Figure 22, we show the results of our feature extraction method on galaxies exhibiting plausible tidal interaction signatures. These galaxies broadly fall into one of the following four categories – i) close-pair systems in which both galaxies exhibit plausible signatures of tidal interaction (GDN 14758, UDS 24437, and GDS 4608); ii) close-pair systems in which only one galaxy exhibits a plausible interaction feature (GDS 16440 and GDS 14637); one example galaxy experiencing a plausible minor interaction (GDS 14786); and one plausible post-merger system (GDN 19863). The extracted residual substructures in these examples highlight our visual interpretations.

First, we apply our feature extraction method to each galaxy in the three example close-pair systems with plausible dual interaction signatures. We note that all three

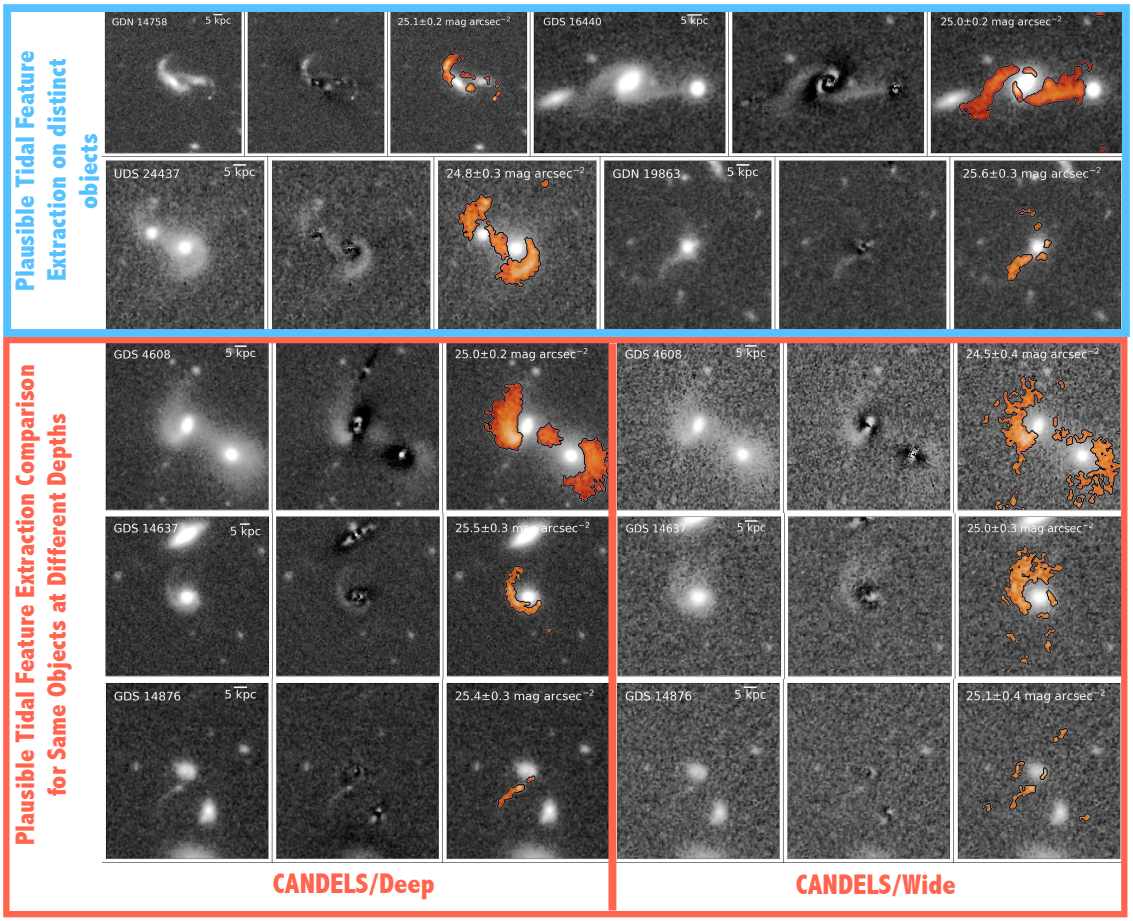

Figure 22: Example residual structure extraction and quantification for select CANDELS galaxies hosting plausible tidal signatures. For each galaxy, we show the *original* image, single-Sérsic residual image, and the extracted residual features using the same format as in Figure 21. For three example galaxies (bottom three rows), we compare the feature extraction process at CANDELS Deep (*HST* 10-orbit; left panels) and Wide (*HST* 2-orbit; right panels) depth configuration.

pairs have $|\Delta z|/(1+z) \leq 0.025$ and residual substructure consistent with theoretical predictions for tidal features produced by major mergers [192]. We find a striking structural similarity between the features exhibited by UDS 24437 – UDS 23110 pair ($z \sim 1.8$) and the GDS 4608 – GDS 4529 pair ($z \sim 1$), such that each close pair hosts two tidal arms and an overlapping region between the host and companion galaxy. These broad and extended features are similar to those found in observations and simulations of major mergers between spheroid-dominated galaxies [42]. The residual substructure midway between each galaxy core in these two pairs may be a stellar bridge or may be an artifact of the simultaneous Sérsic fitting of two galaxies in very close proximity. In contrast, each galaxy in pair GDN 14758–GDN 14516 appears to have two tidal arms which are thin and clumpy in some places. The characteristics of the latter tidal arms are likely the result of interacting disk-dominated galaxies.

Next, we demonstrate the feature extraction on two pairs in which only one galaxy exhibits plausible interaction signatures. GDS 14637 has a single arm-like residual that is similar in appearance to the broad and extended tidal arms discussed above. In contrast, GDS 16440 shows two arm-like features, where one extends outward away from the host and is aligned with a nearby (compact) galaxy (at right) with a spectroscopic redshift difference of only $|\Delta cz| \sim 100\,\mathrm{km\,s^{-1}}$ from GDS 16440. This picture is consistent with theoretical merger simulations (e.g., Figure 6 from [192]) of a disk-dominated galaxy experiencing a major merger with a nearby compact companion. We note that GDS 16440 has a clear inner spiral pattern that extends to a radius of $R \sim 5$–$10\,\mathrm{kpc}$. Despite its point-like appearance suggesting a plausibly interaction-fueld AGN, the companion has

no known X-ray detection to corroborate this suggestion. Finally, we acknowledge the possibility that the extracted arm-lke features of GDS 16440 may be extensions of the inner spiral structure, yet, their diffuse appearance and physical extent out to $R > 25\,\mathrm{kpc}$ is much more consistent with tidal interaction signatures. Admittedly, quantitatively distinguishing spiral versus interaction features is critical for robust identification of galaxy merging. While beyond the scope of this work, we believe that further quantification of extracted residual features will provide improved methods for distinguishing distinct physical processes.

Finally, we demonstrate our residual feature extraction technique on two additional galaxies with plausible signs of tidal activity. First, GDS 14786 shows a minor companion galaxy with a narrow $\sim 10\,\mathrm{kpc}$ tail-like feature, consistent with numerical predictions of features caused by minor mergers [192, 193]. Second, GDN 19863 appears to have two tidal-tail residuals, each extending $\sim 10\,\mathrm{kpc}$ from the galaxy center. One is much fainter than the other. Multiple faint tidal tails are found briefly after coalescence in simulations of major mergers.

2.4.3 Role of Depth during Feature Extraction

To illustrate the role of imaging depth on the extraction and quantification of *different* plausible tidal features, we compare the results of residual substructures extracted from deep (CANDELS/Deep) and shallow (CANDELS/Wide) images for three examples in Figure 22. We perform this comparison on the following examples from § 2.4.2: GDS 4608 (close-pair system where both galaxies show plausible tidal arms), GDS 14637

(close-pair system where on galaxy shows plausible interaction feature), and GDS 14876 (galaxy experiencing a plausible minor interaction).

We find that the morphologies of the extracted residuals at a shallow 2-orbit depth are qualitatively similar to their 10-orbit depth counterparts in all the three cases, albeit with noticeably noisy visual-appearance. We note that the features extracted from the shallow depth images tend to be confined to the brighter surface brightness (by $0.3 - 0.5\,\mathrm{mag\,arcsec^{-2}}$) regions of the same features extracted from the deep images. The shallow residuals are also lower in signal-to-noise when compared to features extracted at a deeper depth. This is expected as an increase in noise per pixel reduces the structural contiguity, owing to the reduction in number of pixels meeting flux-wise significance cut, thereby, resulting in a less significant yet brighter surface brightness. Although we re-calibrate the feature extraction employing a larger area threshold value for shallow depth images, we notice additional small faint sources in the nearby projected vicinity that are contaminating the extracted features. This is because such faint sources are not significantly detected during the source extraction and masking steps, but are extracted from the residuals by our method. This deep-shallow comparison highlights the role of observational depth in identifying faint, high-redshift plausible tidal signatures.

2.5 Extracting Tidal Features in a major merger from cosmological simulation

In this section, we demonstrate the application of our residual substructure extraction method on synthetic *HST* images of a major merger from the VELA simulation [164].

We extract the residual features and quantify their areas during 10 time steps of the merging process spanning $1.7 \lesssim z \leq 1.1$ ($\Delta t \sim 1.4\,\mathrm{Gyr}$), each at 19 viewing orientation angles, and two observational depth configurations matching empirical CANDELS observations (Wide and Deep). We note that the first pericentric passage and the coalescence stages occur at $z = 1.63$ and $z = 1.44$, respectively. To visually comprehend the quantitative trends, we also discuss the qualitative appearance of example extracted features at one fixed viewing orientation.

2.5.1 Quantitative Analysis of Extracted Features

In Figure 23, we show the median extracted feature areas as a function of time before and after the coalescence stage ($t = 0\,\mathrm{Gyr}$) for CANDELS Wide and Deep depths at all 19 viewing angles. We also quantify the detection probability as the percentage of viewing sightlines during which we detect features with their areas larger than the A_{thresh} and μ_{feat} brighter than the limiting surface brightness (Wide – $25.25\,\mathrm{mag\,arcsec^{-2}}$; Deep – $26.25\,\mathrm{mag\,arcsec^{-2}}$). We find the deep and shallow extracted residuals follow qualitatively similar trends in terms of their changing *prominence* (i.e., feature area) during different pre- and post-merger stages. Owing to coarse time resolution ($\sim 150\,\mathrm{Myr}$) between each time step, we are only able to discuss broad trends and differences between the deep and shallow results.

At deeper image depth, we detect residual substructure at nearly all viewing angles and time steps. Prior to first pericentric passage ($t = -0.46\,\mathrm{Gyr}$), the feature area is minimal ($A = 3 \pm 1\,A_{\mathrm{thresh}}$). At first pericentric passage, we find a plausible increase in

the feature prominence, albeit with a large spread between different viewing orientations. This increase clearly peaks at the next time step 150 Myr later, hinting at the formation of tidal interaction-triggered features. If true, these features abruptly diminish in significance (by a factor of two) at $t = -0.15\,\mathrm{Gyr}$, after which extracted feature areas steadily increase through the coalescence stage until they peak at 300 Myr later with a value 3–7 times greater than the initial minimum. This temporal behavior strongly implies the presences of tidal merger signatures. During the three time steps that follow ($0.47 \leq t \leq 0.79$), we note that the feature prominence follows a stochastic rising and falling trend with a range of area values between those of coalescence and the last maximum, or even slightly more at some viewing angles. As we discuss later, these stochastic fluctuations during different time steps correspond to noticeable changes in the visual appearance of the extracted residual features.

In the case of shallow depth images, we typically find a similar temporal time evolution trend, albeit with some key distinctions. First, we note that the trends at shallow depths are shifted to systematically smaller areas when compared to the same features observed in deeper images. This is because of the decreased signal-to-noise per pixel causing less number of contiguous pixels to satisfy our feature extraction criteria. Next, contrary to what we observe in deeper images, we find no increase in the prominence of the extracted feature at the $t = -0.31\,\mathrm{Gyr}$ time step following first pericentric passage. This is consistent with the idea that tidal interaction features are faint and their detection is highly sensitive to image depth. Indeed, we find the detection probability of any residual substructure in the shallow images is less than 50% until after first passage, and

only reaches unity at the coalescence stage when the first noticeable increase in feature area occurs. During $t > 0\,\mathrm{Gyr}$, we note that post-merger time steps with detection probability $< 100\%$ correspond to smaller feature areas. Likewise, maximum feature areas correspond to high detection probability. These subtle trends highlight the role played by viewing lines-of-sight during the detection of certain merger features, especially in shallow depth images. Nevertheless, during the time steps with high detection probabilities (irrespective of the depth), we note that the viewing angles can induce a scatter of $20 - 40\%$ in the measured feature prominences.

2.5.2 Qualitative Interpretation of Extracted Features

To illustrate the temporal quantitative trends shown in Figure 23, we discuss the qualitative appearances of the features extracted from one example viewing orientation. We show the extracted features at shallow and deep depths alongside their idealized images during 10 time steps in Figure 24. We discuss the extracted features from different image depths with respect to the distinct structures evident in their idealized (no noise added) images, and we compare the morphologies of the features extracted from the VELA mock data to those from CANDELS images of plausible merging systems shown in Figure 22.

During the first two time steps ($t \leq -0.46\,\mathrm{Gyr}$), we extract small residual structures in both deep and shallow images that correspond to faint sources in proximity to the merging galaxies within their idealized images. At $t = -0.31\,\mathrm{Gyr}$ ($\sim 150\,\mathrm{Myr}$ after the first pericentric passage), we extract features at deeper depths that correspond to the

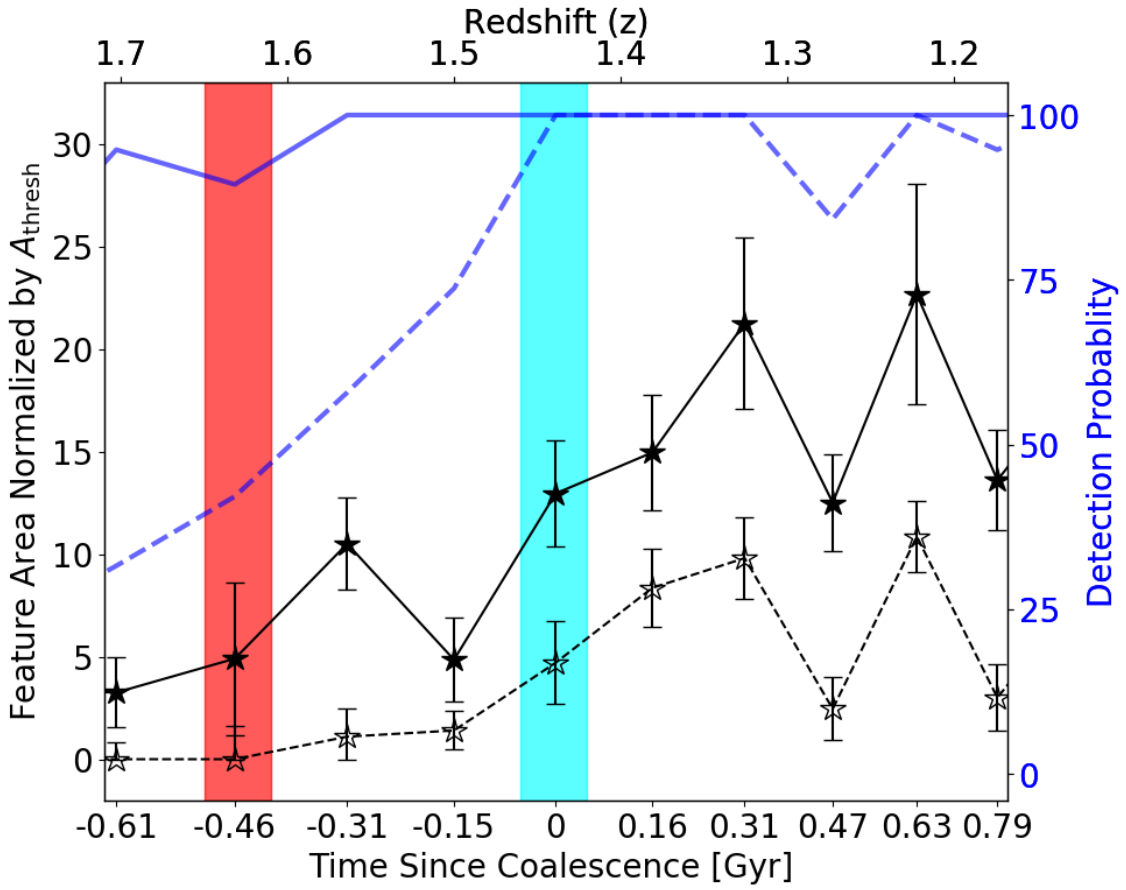

Figure 23: Area of the extracted residual features with $A \geq A_{\rm thresh}$ and $\mu_{\rm feat}$ brighter than a depth-dependent limiting surface brightness, as a function of time since the coalescence phase from the synthetic mock observations of a galaxy merger simulation. We normalize the feature area with $A_{\rm thresh}$ value employed for each CANDELS depth – Wide (filled points) and Deep (open points). At a fixed time step and depth, we show the median feature normalized area with a star; its corresponding standard deviation across different viewing orientations is shown with a error-bar. In blue, we give the detection probability which is defined as the percentage of camera angles during which the extracted residual features satisfy our area and surface brightness criteria at shallow depth (dashed line) and deep depth (solid line). We emphasize the first pericentric passage and coalescence stages with red and cyan shading, respectively.

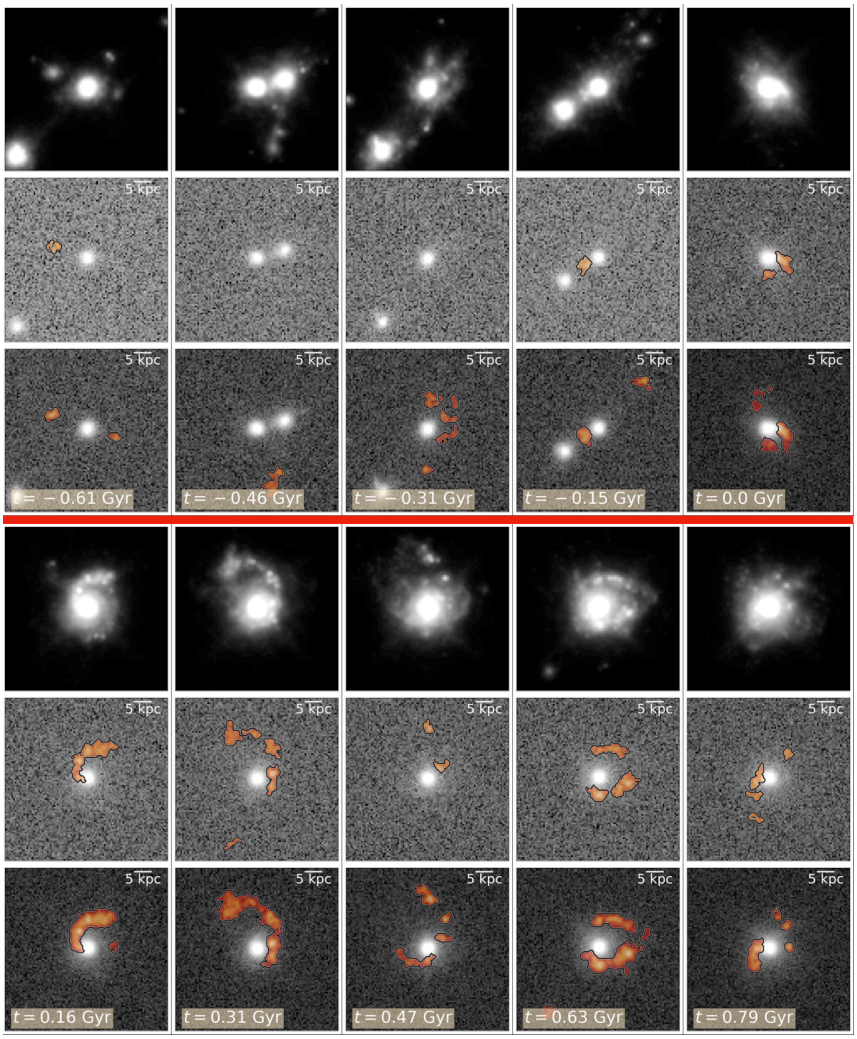

Figure 24: Extracted residual features during the 10 time steps shown in Figure 23 from a galaxy merger simulation (VELA; see text for details). At each time step (left to right), we present the idealized (noise-less) mock *HST* WFC3/F160W images for one viewing angle (cam00; *top*), the CANDELS/Wide depth images (*middle*), and the CANDELS/Deep depth images (*bottom*). The extracted features are shown with red shading for the shallow and deep data. For each trio of images, we indicate the $5\,\mathrm{kpc}$ physical scale and the time elapsed since (or prior to) the coalescence stage ($z = 1.44$; $t = 0\,\mathrm{Gyr}$).

interaction debris surrounding the merging galaxies in the idealized images. However, we do not detect these features in shallow depth images, which aligns with the noticeable difference in feature prominence at $t = -0.31\,\mathrm{Gyr}$ in Figure 23. Just prior to coalescence ($t = -0.15\,\mathrm{Gyr}$), we extract a small feature midway between the two merging galaxies at both depths, which corresponds to a plausible bridge-like structure in the idealized image. We find similar excess-light features in the extracted residuals for UDS 24437 and GDS 4608 in Figure 22. Yet, interestingly, the VELA merger does not exhibit tidal arms during its close-pair stage, which is in contrast to all three example plausible interacting systems (GDS 4608, UDS 24437, and GDN 14758).

Starting at coalescence ($t = 0\,\mathrm{Gyr}$), we extract features at both depths that highlight the birth of a post-merger tidal-arm that grows in area over the next $\sim 300\,\mathrm{Myr}$. This feature is similar in appearance to the single arm-like structure we extract from one of the two example CANDELS pairs in which only one galaxy has a plausible interaction signature (GDS 14637; Fig. 22). As such, GDS 14637 is also consistent with a post-merger in close proximity to another galaxy. At $t = 0.47\,\mathrm{Gyr}$, we extract two narrow tail features with clumpy visual appearance in the deep image, which resemble the plausible post-merger signature we extract for GDN 19863. This picture is consistent with the theoretical expectations that major gas-rich mergers produce diffuse tidal features with clumpy knots Fensch17. We note that we only marginally detect a fraction of this tidal feature in the shallow depth image, which aligns with the abrupt decrease in feature prominence noted in Figure 23. Finally, during the last two time steps in both shallow and deep images we extract a large and extended plausible tidal loop or two-tailed feature (at $t = 0.63\,\mathrm{Gyr}$)

that becomes less prominent at $t = 0.79\,\mathrm{Gyr}$. The changes in these features during the last two time steps illustrate the stochastic fluctuations in the extracted feature areas we find for all viewing angles in Figure 23. This exercise illustrates the interplay between interaction/merger feature extraction and observational depth.

2.6 Discussion

So far, we have presented the development of a residual substructure extraction tool and discussed its application to example galaxies from the CANDELS survey and mock *CANDELized* observations of a galaxy merger simulation. In this section, we discuss several additional possible applications of our residual extraction method.

2.6.1 Possible Applications

We present some possible applications of our residual feature extraction method towards improving popular structural-fitting routines, studying multi-wavelength properties of different single-component Sérsic residual features (e.g., disk structures from host bulge+disk systems) and substructures (e.g., spiral arms, plausible interaction signatures), and extracting strong gravitational arcs.

2.6.1.1 Improving structural fitting

An important step in many galaxy structural-fitting studies (e.g., [161, 194]) is to choose the number of required parametric components during `GALFIT` modeling (or other similar tools), starting with a simple single-Sérsic model and advancing to more complex models (e.g., bulge+disk) including even the inclusion of Fourier modes (e.g.,

to model spiral features). For upcoming large-scale surveys, developing machinery to make such decisions is key for automated structural modeling for very large samples. One way to approach this is to use machine learning to inform the number of parametric components required to best represent a galaxy morphology [195]. Alternatively, we propose that our method can be used to inform additional residual components and repeat `GALFIT` with updated number of components.

2.6.1.2 Multi-wavelength application

Empirical studies often perform multi-wavelength decomposition of galaxy structure into their bulge and disk components to separately inform their physical parameters (e.g., stellar mass, color; [196, 197]), which can provide key insights into the physics of galaxy evolution. We propose that extending our feature extraction method to multiple wavelengths can inform the physical properties for disk, central bulge, spiral arms, and interaction signatures. Furthermore, recent studies are developing resolution-element-based spectral energy distribution fitting in conjunction with spectroscopic measurements to produce maps of physical parameters (stellar-mass, star formation rates; [198]. Using our method in conjunction with these semi-resolved maps can enable future investigations to study structurally-resolved properties (e.g., colors, star-formation, chemical composition, kinematics) of different galactic substructures, especially for tidal features (e.g., see [199]).

2.6.1.3 Extracting Gravitational Arcs

Yet another possible application of our method is the extraction of strong gravitational lensing arcs. Strong gravitational lensing provides an unique opportunity to observe the otherwise inaccessible high-redshift universe. Several previous studies have used strong lensing to study cosmology, the physical properties of lensing systems, and the properties of distant background lensed sources (e.g., [200–202]). Identifying and extracting the gravitational lensing signatures is a key step in such works, which often use `GALFIT` to subtract the host galaxy light to unearth such lensing features (e.g., [203, 204]).

In Figure 25, we show the result of our feature extraction method on an example CANDELS strong-lensing candidate (UDS 10713), which is identified using a machine learning based classification scheme by [163]. We use the galaxy's single-Sérsic residual image by vdW12 and extract significant contiguous substructure above a $3\sigma_{\rm bkg}$ threshold (instead of our fiducial $2\sigma_{\rm bkg}$). We find arc-like features with surface brightness $\mu_{\rm feat} = 24.3 \pm 0.2\,{\rm mag\,arcsec^{-2}}$ in close vicinity to the galaxy, which qualitatively aligns with the blue-arc in its corresponding false-color RGB image (at roughly 10 o'clock). However, we note that the single-Sérsic residual of UDS 10713 shows excess galactic light and additional sources in its projected proximity, which are contributing to the surface brightness of our extracted features. Isolating such contaminants is beyond the scope of this current work and we recommend that future studies should employ more complex Sérsic models for better galactic light subtraction and use multi-wavelength information

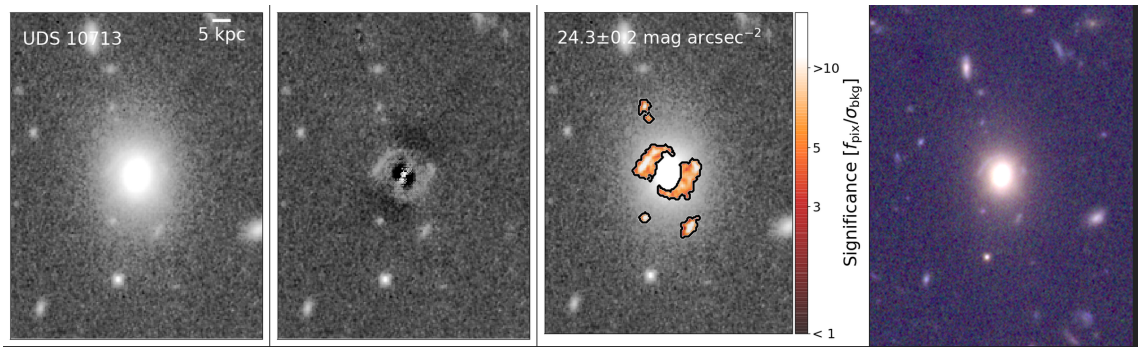

Figure 25: Example extraction of a candidate strong lensing arc hosted by the CANDELS galaxy UDS 10713 (center of the image), identified by [163]. *First panel:* the WFC3/F160W (1.6μm) *original* image with the official CANDELS ID and a 5 kpc physical scale. *Second panel:* the single-Sérsic residual image by vdW12. *Third panel:* the extracted residual features (shading) overlaid on the *original* image; the signal-to-noise is indicated with the color-bar. *Fourth panel:* false-color F160W–F125W–F606W RGB image of UDS 10713; strong lensing arc appears at the top-left of UDS 10713.

of the features for better automated lensing feature detection and extraction. Furthermore, future works can extend our residual extraction method to galaxy cluster environments where strong lensing can be more prevalent (usually with high magnifications) as a means to derive cluster mass measurements.

2.6.2 Future Prospects with the James Webb Space Telescope

In preparation to probe high-redshift galaxy merging using the James Webb Space Telescope (*JWST*), we analyze synthetic *JWST* observations of a different merger in the VELA cosmological simulation suite (VELA 21), which experiences a major merger (post-merger stellar mass $\log_{10}(M_{\rm stellar}/M_{\odot}) = 10.5$; stellar-mass ratio $\sim 3:1$) at $z \sim 3.2$ (coalescence). In Figure 26, we compare and contrast the feature extraction process on its mock *JWST* and *HST* observations.

Motivated by our simulated observations of a major merger in § 2.5, we focus this exercise on one specific merger stage $z = 3$ ($\sim 120\,\mathrm{Myr}$ after coalescence), where we expect the galaxy to exhibit tidal features. We generate idealized mock *HST* WFC3/F160W and *JWST* NIRCam/F277W images using the process described in § 2.2.2. We add Gaussian noise to them such that they match the CANDELS/Deep depth and the proposed Cosmic Evolution Early Release Science Survey (CEERS; PI: S. Finkelstein, *JWST* PID: 1345) depth [8], respectively. Finally, we generate the single-Sérsic residuals by applying `GALFIT` to the *HST* and *JWST* noise-added images following vdW12, and apply our feature extraction technique on these residuals (§ 2.3).

While we detect no significant residual features at CANDELS/Deep *HST* depth, we extract extended tidal-fan features in the NIRCam/F277W *JWST* CEERS depth image, which mimic the structures exhibited by our example CANDELS close-pair system (GDS 4608; see Figure 22). Our demonstration asserts the capabilities of *JWST* to probe faint tidal features out to high-redshifts with considerably less exposure time and at a longer rest-frame wavelengths than the current extremes accessible to *HST*. This will enable future studies to probe high-redshift merging systems that are key towards understanding the role of galaxy merging in galaxy evolution.

2.7 Conclusions

As a *first step* towards quantitatively identifying 'hallmark' signatures of galaxy merging (e.g., tidal arms, tails, bridges, and extended fans), in this paper, we introduce a

[8] The proposed depth of the CEERS survey reaches greater than a 10σ point-source and 5σ extended-source sensitivity of $28\,\mathrm{mag}$ with a ~ 2800 second exposure in NIRCam/F277W.

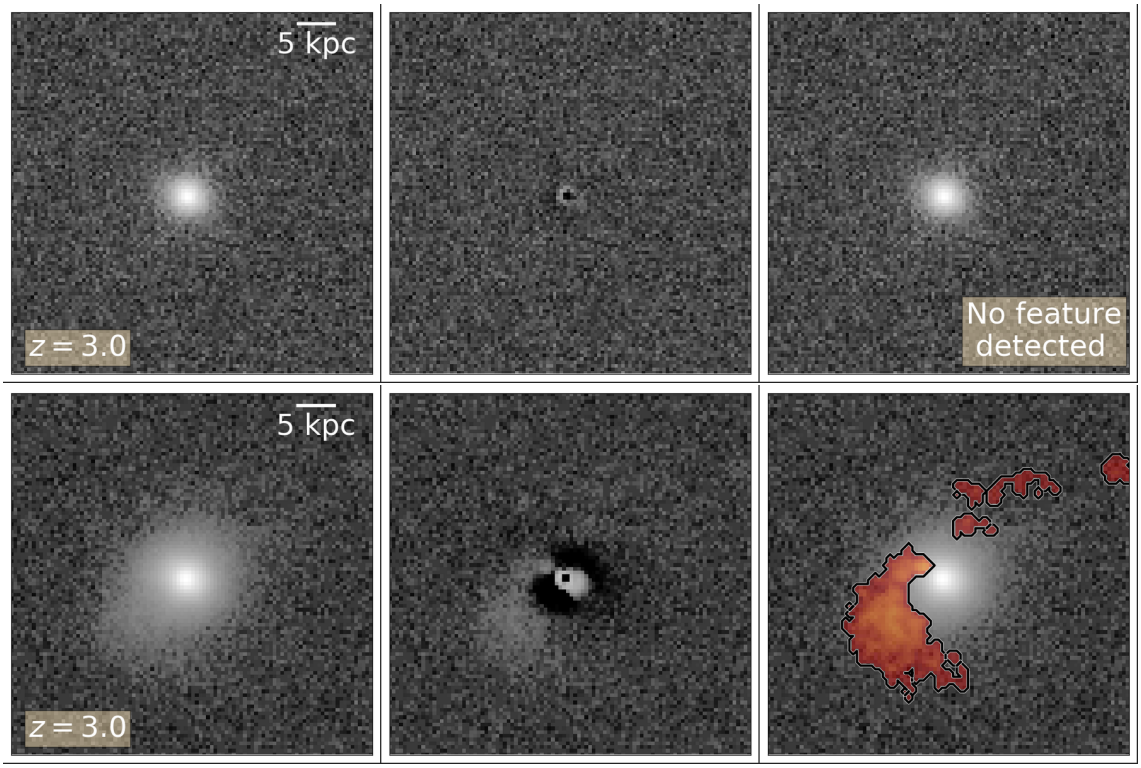

Figure 26: A simulated massive post-merger at $z = 3$ (VELA 21; see text for details) with mock CANDELS/Deep *HST* imaging (*top row*), and mock *JWST*-ERS CEERS imaging (*bottom row*). In each row, we present the synthetic noise-added images matching their respective empirical depths (*left*), the corresponding single-Sérsic GALFIT residual images (*middle*), and the extracted feature (if detected) shown with red shading (*right*). This figure demonstrates that *JWST* will enable us to probe tidal features out to high redshifts in a considerably smaller amount of exposure time and at longer rest-frame wavelengths when compared to *HST*.

new analysis tool to extract and quantify residual substructures hosted by galaxies. Our method analyzes residual images produced by any standard light-profile fitting tools and extracts contiguous features that are significant above the sky background and a noise-only expectation of pixel-contiguity. When applied to both empirical and simulated data, this tool will facilitate future efforts to calibrate the observables associated with galaxy merging signatures and provide improved merger rate constraints.

We apply our residual feature extraction method to single-Sérsic residual images of 16 select galaxies from the CANDELS survey (see Table 6 and [160]) and extract different structures hosted by these galaxies – disk structures (4 galaxies), spiral substructures (4 galaxies), and plausible interaction/merger-triggered tidal features (8 galaxies). The main conclusions of our empirical exercise are:

• We extract residual disk features and two-armed spiral substructures for all example galaxies visually identified as hosting disk and spiral features, respectively.

• We extract plausible structurally-similar, dual tidal interaction signatures hosted by *three* example close-pair systems, *two* close-pair systems where only one galaxy exhibits plausible interaction signatures, *one* system experiencing a plausible minor merger event, and *one* galaxy hosting plausible post-merger signature.

• To illustrate the role of imaging depth during feature extraction, we compare the residual substructures extracted from deep (*HST* 10-orbit) and shallow (*HST* 2-orbit) images of three example galaxies hosting plausible tidal features. We find that the features extracted at shallow depths are qualitatively similar but noisier in appearance compared to their deeper counterparts owing to the limited sensitivity of shallow CANDELS data to

only the higher-surface brightness substructure of the features.

We also apply our tool to synthetic *HST* observations of a galaxy merger from the VELA hydro-dynamic simulations [164]. We extract the residual features and quantify their areas during 10 time steps of the merging process spanning $1.7 \lesssim z \leq 1.1$ ($\Delta t \sim 1.4\,\mathrm{Gyr}$), each at 19 viewing orientation angles, and two observational depth configurations matching empirical CANDELS observations (Wide and Deep). We also illustrate the qualitative appearance of these residual features at one fixed orientation. The main conclusions of our theoretical exercise are:

• We find that the features typically rise in *prominence* (feature area) $\sim 150\,\mathrm{Myr}$ after the first pericentric passage ($z = 1.63$; time since coalescence $t = -0.46\,\mathrm{Gyr}$), which visually correspond to close-pair interaction signatures. These features abruptly diminish by a factor of two in significance during $t = -0.15\,\mathrm{Gyr}$ before coalescence.

• Starting with the coalescence stage ($z = 1.44$; $t = 0\,\mathrm{Gyr}$), we extract post-merger features that grow in prominence until peaking at $3 - 7$ times the initial minimum value after $300\,\mathrm{Myr}$. These stages visually represent the birth and growing prevalence of tidal arm features. At later time steps $0.47 \leq t \leq 0.79\,\mathrm{Gyr}$, the feature prevalence follows a stochastic rising and falling trend, where the extracted tidal features appear structurally fragmented and clumpy appearance.

• Imaging depth and viewing orientation play a key role in detecting certain merger features. At deep image depths, we extract all the features at nearly all viewing angles ($\sim 100\%$ detection probability). However, the same features in shallow depth images are detected at $\sim 50 - 75\%$ of the viewing sightliness during $-0.46 \leq t \leq -0.15\,\mathrm{Gyr}$

prior to coalescence. They only reach 100% detection probability at $t > 0\,\mathrm{Gyr}$, when the first noticeable increase in feature prominences occur. Additionally, the temporal trends in shallow depth images are shifted to systematically smaller areas than their deeper depth counterparts. At a fixed depth, the measured feature prominences between different viewing orientations can vary by $20 - 40\%$.

While the application of feature extraction method provides important insights about the onset and evolution of tidal features hosted by an example galaxy major merger from the VELA hydro-simulations, we caution the extrapolation of our quantitative (e.g., observability timescales) and qualitative (e.g., tidal feature shapes) interpretations to a general sample of mergers. This is because the strength and appearance of tidal debris may depend on the intrinsic properties of merging galaxies such as: stellar mass-ratio, redshift (driven by the evolution gas-fractions), merger orbital configurations (e.g., prograde vs. retrograde approach). Moreover, it is also important to note that the tidal feature observability carried out in a cosmological context may differ from a similar exercise on idealized binary-merger simulations.

Lastly, we discuss two additional possible future applications of our general purpose tool. First, we illustrate the extraction of strong gravitational arcs in a candidate strong lensing system identified by [163]. Second, we demonstrate the future prospect of using our method to probe high-redshift tidal features by applying it to synthetic *JWST* observations of a $z \sim 3$ simulated major merger. We find that *JWST* can probe faint, high-redshift tidal features with a smaller exposure time investment and longer rest-frame

wavelengths ($2.7\,\mu\mathrm{m}$) than *HST*/WFC3 *H*-band 10-orbit (i.e., CANDELS/Deep) observations.

CHAPTER 3

MAJOR CLOSE-PAIR FRACTION CALIBRATIONS USING MOCK REALIZATIONS FROM SEMI-ANALYTICAL MODELS

Unpublished

Abstract

Observational selection effects may be one of the primary reasons for the study-to-study close-pair frequency measurements during the key epoch of galaxy growth $1 < z < 3$, yet their impact is poorly constrained. In this study, we analyze the redshift evolution of the close-pair frequency and quantify the impact of different observational effects (photometric redshifts and their uncertainties, stellar-mass errors, sample completeness, and survey area sizes) using mock light cone data from the leading SantaCruz Semi-Analytical Model (SC SAM). Our main findings are that (i) Only $\sim 30-35\%$ of the pairs selected by common photmetric redshift proximity criteria have small relative velocities $\Delta v \leq 500\,\mathrm{km\,s^{-1}}$; (ii) The observed major pair counts require a relative reduction of $5-25\%$ to account for the impact of stellar-mass errors and to accurately represent the true major counts; (iii) a relative increment of $10-35\%$ at $z \sim 2-3$ is required to correct the observed photometric redshift based pair fractions for sample completeness at a typical CANDELS survey depth setting; (iv) A typical pair fraction variation of $\sim 0.01-0.015$ is expected among CANDELS-UDS sized fields ($\sim 160\,\mathrm{arcmin^2}$). By applying these calibration corrections for observational effects, we compute latest CANDELS close-pair

fractions among ~ 9500 massive galaxies ($\log_{10}(M_{\star,\mathrm{can}}/M_{\odot}) \geq 10$) during $0.5 < z < 3$. We find that the calibrated CANDELS pair fraction evolves mildly from $\sim 6\%$ to $\sim 9\%$ during $0.5 < z < 1.5$ and remains considerably flat at $\sim 9\%$ over $1.5 < z < 3$.

3.1 Introduction

Major galaxy merging (mass ratio < 4) is a fundamental aspect of the hierarchical structure-growth scenario of the Universe. As such, it is theoretically predicted to contribute to the empirically well-documented stellar-mass growth of high-mass galaxies (e.g., $M_{\mathrm{stellar}} > 10^{11} M_{\mathrm{sun}}$), enhance global galactic star-formation, and facilitate central super-massive black hole growth and triggering. Observationally quantifying the major merger frequency and its evolution is a key methodological step in empirically verifying the "merging – galaxy evolution" connection. Several studies over the past two decades have quantified the redshift evolution of major merger incidence at $z \lesssim 1.5$ using close-proximity galaxy pairs [3–6, 144], and more recent works have extended the technique's use out to $z \sim 3 - 6$ [16, 17, 140, 147, 205]. These studies broadly agree that major close-pair fractions (f_{pair}) increases moderately to strongly during $0 < z \lesssim 1.5$, albeit with some study-to-study variations owing to differences in pair selection choices (see [140] for discussion). However, the empirical results at $1 < z < 3$ vary considerably between moderately rising [205] to flattening or diminishing diminishing trends [15, 16, 140], despite choosing closely-matching pair-selection criteria. These variations may be due to important observational biases that dominate during the $1 < z < 3$ regime such as the photometric redshift and stellar-mass measurement uncertainties, sample incompleteness,

intrinsic number count variance from field-to-field. The impact of such observational effects on the measurement of $f_{\rm pair}$ evolution is yet to be systematically explored and doing so is a crucial step towards robustly quantifying the close-pair statistics during the key epoch of galaxy evolution. To address some of these key open questions, In this study, we carryout a detailed systematic calibration of the impact of such observational effects by analyzing the redshift evolution of $f_{\rm pair}$ within the mock galaxy light cone data based on the SantaCruz Semi-Analytical Model [136, 206, 207] with realistic CANDELS survey [20, 21] like observational effects. This comprehensive analysis and its methodological framework will enable future studies to derive robust pair statistics out to $z \sim 3$ and beyond.

Major galaxy-galaxy mergers are theoretically expected as a natural consequence of the gravitationally driven dark-matter halo merging in a hierarchical structure growth paradigm of the universe. Theoretical galaxy evolution simulations advocate for the importance of major mergers in driving key aspects of galaxy growth and transformation such as bulid-up of massive elliptical galaxies [28, 31, 32], enhancement of star formation [33–35], and growth and triggering of central super-massive black hole activity (SMBH) [13, 37–39]. Many empirical studies have attempted to validate these theoretical expectations and have indeed found that major mergers can explain the stellar-mass and number density buildup of massive galaxies at $z < 1$ [7, 43, 44], enhancement of SF [46, 150] and SMBH activity [47–49]. However, some studies have also found contradicting results of weak or no merger-SF-AGN connection [50, 51, 54, 56], suggesting that alternate physical processes may be at simultaneous play (e.g., Violent Disk Instabilities

(VDI); [57, 58]), indicating that the role of major mergers in galaxy evolution remains a key open question. Robust empirical identification of major merging systems and quantifying their frequency (i.e., $f_{\rm pair}$) is an important step in such studies and biases with this step can significantly hinder the ability to correctly interpret the merging – galaxy evolution connection. As such, accurately quantifying the major merger frequency is crucial to address the role played by mergers in galaxy evolution.

Motivated by theoretical simulation predictions that similar-mass galaxies in close physical proximity interact gravitationally and coalesce into a more massive system, many studies have used close pairs as a reliable probe for identifying ongoing or future major mergering systems. These studies commonly select host (1) galaxies that have at least one companion (2) meeting a set of selection criteria: a) proximity in 2D projected distance ($R_{\rm proj}$); b) proximity in redshift ($\Delta z_{12} = z_1 - z_2$) or relative velocity space ($\Delta v_{12} = v_1 - v_2$); c) stellar-mass ratio (M_1/M_2). A wide range of choices have been employed in the literature for these criteria. For $R_{\rm proj}$, an annular region is often chosen with $R_{\rm proj} \in [R_{\rm min}, R_{\rm max}]$, where the minimum ($R_{\rm min}$) and the maximum ($R_{\rm max}$) projected separation ranges between $\sim 0 - 14\,{\rm kpc}$ [14, 16] and $\sim 30 - 140\,{\rm kpc}$ [60, 61], respectively. To ensure that the host and companion lie in close (plausible) physical proximity, it is a standard practice in the literature to either choose stringent spectroscopic velocity difference (commonly, $\Delta v_{12} \leq 500\,{\rm km\,s^{-1}}$, e.g., [5, 62]) when data is available or use photometric redshift error (σ_z) weighted differences ($\Delta z_{12}/\sigma_z$; e.g., [6, 16, 128, 140]). Finally, the choices for the stellar-mass ratio based major pair selection ranges between

$M_1/M_2 \leq 2$ [64] and $M_1/M_2 \leq 5-6$ [65, 147], where the most common choice being $M_1/M_2 \leq 4$ (see [140]). Such variation in pair selection choices yielded a range of $f_{\rm pair}(z)$ evolutionary trends $\propto (1+z)^{0.5-3}$ at $0 < z < 1.5$. Indeed, through a systematic analysis of close-pair selection effects, [140] suggested that the variations between different studies can be explained by their use of systematically stringent/lenient selection criteria. However, despite using closely matched pair selection criteria, recent measurements of $f_{\rm pair}(z)$ evolution at $z > 1.5$ vary from moderately increasing to flat, or diminishing trends [16, 17, 67, 140, 147, 205], highlighting the strong need to understand the study-specific observational selection effects at play.

Photometric redshift ($z_{\rm phot}$) based close-pair proximity selection is a common technique used by high redshift pair studies, where the $z_{\rm phot}$ uncertainties play a crucial role. For e.g., [6] impose $\Delta z_{12}^2 \leq \sigma_{z,1}^2 + \sigma_{z,2}^2$ and [128] uses $\Delta z_{12} \leq \sigma_{z,1}(1+z_1)$, where $\sigma_{z,1}$ and $\sigma_{z,2}$ are the 1σ host and companion redshift uncertainties, respectively. These $z_{\rm phot}$ based criteria translate to a redshift-dependent relative velocity separation selection, where a $\Delta z \lesssim 0.02(1+z)$ corresponds to $\Delta v \lesssim 18,000\,{\rm km\,s^{-1}}$ (see [128]). The accuracy of photometric redshift estimations vary amongst different studies and is dependent on several survey-related and methodological aspects (e.g., number of wavelength bands, template-fitting assumptions, etc). For e.g., [17] reports $\sigma_z \sim 0.01-0.05$ for different survey fields in their study (also see [205]), and [140] reports $\sigma_z \sim 0.03$ for the CANDELS fields. Such study-to-study differences in the $z_{\rm phot}$ uncertainties could lead to systematic variations in the $f_{\rm pair}$ measurements, and this motivates the strong need to quantify and calibrate their impact.

In simultaneity to the $z_{\rm phot}$ uncertainties, the galaxy's stellar mass plays a crucial role in the selection of the primary, mass-limited host-galaxy sample, and in selecting major, similar-mass companions using a stellar-mass ratio M_1/M_2 criterion. Systematic and random uncertainties of stellar mass can impact the selection of primary host sample and major pair selection by scattering the galaxies into/out of the selection function boundaries, thereby impacting the subsequent pair frequency measurements. Therefore, study-to-study variations in stellar-mass uncertainty properties could be contributing to the differences in close pair frequency measures, and this highlights the strong need to calibrate their impact.

The choice of the stellar-mass limited host galaxy sample selection employed by several close-pair studies closely ties with the sample's stellar-mass/detection completeness as a function of redshift. It is a common practice to select a mass limit for the host sample such that their corresponding potential companion galaxy sample is highly complete in the redshift range of study. For e.g., [140, 205] choose $\log_{10} M_{\rm stellar}/M_{\odot} \geq 10.3$ for their host galaxy sample, where their corresponding $M_1/M_2 \leq 4$ companions ($\log_{10} M_{\rm stellar}/M_{\odot} \geq 9.7$) are $\gtrsim 90\%$ complete out to $z \sim 3$. While, majority of studies assume that the high host and companion sample completeness will have minimal impact on the resultant close-pair frequency measures and do not apply an explicit correction to them, some recent studies (e.g., [205]) have applied a stellar-mass function based statistical correction for missing pairs due incompleteness. Quantifying the direct impact of sample incompleteness on the measured close-pair statistics and deriving systematic correction calibrations are crucial to accurate pair frequency measures at high redshift

($z \sim 2-3$, and beyond), where such completeness effects might be especially dominant in current common extra-galactic survey settings.

Different pair studies in the literature have used ground-based and space-based observations of different survey fields of interest, which have a wide range of sky area and redshift coverage. Intrinsic variations in the galaxy number counts can occur due to cosmic sample variance, which can become a dominant source of uncertainty especially for smaller fields and at high redshifts [115]. Such variations in the number counts of host and companion samples can directly impact the close-pair incidence statistics, and this may be one of the contributing effects towards study-to-study pair frequency variation at $z > 1.5$. This motivates the need for deriving detailed calibrations of the impact of intrinsic sample variance when deriving close-pair statistics.

Semi-analytical models (SAMs) of galaxy evolution offer a powerful way towards understanding several key aspects of the galaxy formation and development physics. Conceptually, the SAM framework involves assuming analytical prescriptions for baryonic physics (e.g., star formation, stellar and black-hole feedback, stellar evolution, etc) and forward models them using the merger histories of dark matter haloes (often referred to as merger trees) derived from large N-body simulations. Decades of research has yielded several novel SAMs following different assumptions for baryonic physics [27, 136–138, 206, 208–210], and more recent works have updated their frameworks with new modeling assumptions as informed by recent advancements in hydro-dynamic simulations and empirical observations [211–213]. The Santa-Cruz Semi Analytical Model

(SC SAM; e.g., [136,206,207]) is a leading galaxy evolution framework used by many observational studies [59,135] to physically interpret various aspects of galaxies and provide forecasts for future studies with upcoming surveys [213, 214]. Recent studies have leveraged the power of the SAMs frameworks to build mock light cone catalogs of galaxies and constructed observational celestial footprints that mimic popular surveys [215, 216]. Such SAM light cone data enables a direct path towards measuring the close-pair statistics and a systematic quantification of the impact of different observational selection effects in a realistic observational setting.

In this work, we analyze the redshift evolution of the close-pair frequency using the mock light cone data of SC SAMs over $0.5 < z < 3.0$ while assuming commonly-used pair selection criteria . By systematically incorporating CANDELS-like observational effects such as photometric redshifts and their uncertainties, stellar-mass errors, sample completeness, and survey area sizes, we assess their impact on the pair frequency measurements and derive detailed correction calibrations. We also compute latest empirical CANDELS pair frequency evolution during $0.5 < z < 3$ by appropriately correcting for the observational effects.

This work is structured as follows: In Section 2, we provide a description of the SC SAM raw light cone catalogs and discuss our methodology of applying the CANDELS-like observational effects (photometric redshifts and their errors, stellar-mass errors, sample completeness curves). We also briefly discuss the CANDELS survey photometric and high-level data products (redshifts and stellar masses). In Section 3, we discuss different close-pair selection criteria explored in this study and derive the redshift evolution of the

close-pair frequency with and without the different incorporated observational effects. We also systematically quantify the impact of each observational effect and provide statistical correction calibrations. In Section 4, we derive the latest empirical CANDELS close-pair frequency evolution measurement after applying the appropriate correction calibrations derived in Section 3. Finally, we discuss our main conclusions in Section 5.

3.2 Raw Santa-Cruz SAM Mock Data and Derived Parameters

To better understand and quantify the role of observational systematics when measuring close-pair fractions, we analyze galaxy-galaxy close pair fractions in mock light cone catalogs generated by the SantaCruz Semi-Analytic Model (SC SAM; [136]). In this section, we introduce with the raw SC SAM light cone mock data and discuss the derived parameters with realistic CANDELS survey like observational effects such as photometric redshifts and their errors, stellar masses uncertainties, and detection completeness limits.

3.2.1 Santa-Cruz SAM Raw Light cone Catalogs

We use the SC SAM based mock light cone catalogs generated to match the celestial footprints of the CANDELS survey by [136]. Briefly, the SC SAM model maps the baryonic physics onto root dark-matter halo information drawn from the Bolshoi-Planck N-body simulations [217]. For details regarding different analytical prescriptions used in the SC SAM, we refer the reader to [137, 218]. Following the framework by [125], this SC SAM information is further processed to create eight light cone realizations sampling different sight lines of the N-body simulation box. These light cones are then used to

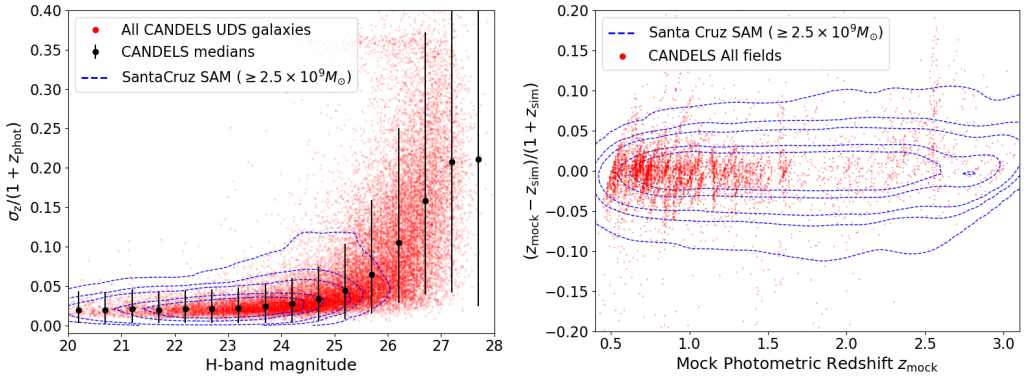

Figure 27: *Left-panel*: *H*-band magnitude vs. photometric redshift uncertainty $\sigma_{\rm z,can}/(1+z_{\rm phot,can})$ for all the galaxies (with good data; PhotFlag$=0$) from the CANDELS UDS field (red points), with their medians and 1σ error bars in bins of *H*-band magnitude ($\Delta H_{\rm 160w} = 0.5$mag). We also show the analogous quantity $\sigma_{\rm z,mock}/(1+z_{\rm mock})$ and for $M_{\rm stellar} \geq 2.5\times10^9 M_\odot$ galaxies from the one mock SantaCruz SAM lightcone (blue contours) assigned based on the CANDELS galaxies information (see Sec 3.2.1). *Right-panel*: We show the variation of mock photometric redshift estimates ($z_{\rm mock}$) from the true simulated redshifts ($z_{\rm sim}$) for the SAMs galaxies: $(z_{\rm mock}-z_{\rm sim})/(1+z_{\rm sim})$ as a function of $z_{\rm mock}$ (blue contours). For reference, we show analogous information $(z_{\rm phot,can}-z_{\rm spec})/(1+z_{\rm spec})$ for all galaxies in the CANDELS fields with available spectroscopic redshift information (red points). In both the panels, the contours levels correspond to counts of 50, 68, 90, 95, and 99 percentile levels of the 2D histogram number counts in the respective parameter space.

generate emulated galaxy mock catalogs spanning the celestial footprint setting of the five CANDELS fields (UDS, GOODS-S, GOODS-N, COSMOS, and EGS), and sky coordinates (RA and DEC) are assigned to each galaxy in the mock catalog. Alongside this information, each galaxy in the mock catalog holds its true simulated redshift ($z_{\rm sim}$), stellar-mass ($M_{\rm stellar,sim}$), and apparent magnitude in *HST* WFC3/F160W (*H*-band; $H_{\rm sim}$), which we use throughout this analysis.

We note that the raw SC SAM mock catalogs span $\sim 5-8$ times larger sky area than the native CANDELS field footprint sizes. In conjunction with eight different realizations, this amounts to 64 unique CANDELS UDS field sized regions, which enables us with ample statistical power when measuring the close pair statistics and directly assessing their field-to-field variations. Therefore, for our analysis, we choose CANDELS UDS field mock catalogs, and note that choosing mock data from a different field (e.g., CANDELS GOODS-S) does not impact the results discussed in this work.

3.2.2 Derived Parameters from Mock Data

To quantify the effect of different observational effects on the measured redshift evolution of close-pair fractions, we mimic the photometric redshift errors, stellar-mass systematic and random measurement errors, and detection incompleteness onto the mock lightcone catalogs.

3.2.2.1 Photometric Redshifts and their uncertainties

Photometric redshifts uncertainties (σ_z) are commonly used in the close-pair studies to isolate plausible physical pairs by imposing a $z_{\rm phot}$-based proximity selection [6, 16,

140, 205]. While the effects of different redshift proximity selection choices was explored in [140], a comprehensive look into the role of σ_z on measured pair fractions is yet to be explored. To quantify the effect of $z_{\rm phot}$ errors in selecting primary galaxy samples for close-pair analysis and quantifying the pair fractions, we incorporate the statistical behavior of CANDELS photometric redshift properties such that their $\sigma_z/(1+z_{\rm phot})$ distributions are mimicked by the SC SAM mock galaxies. To achieve this, we use the latest CANDELS photometric redshift data ($z_{\rm phot,can}$) produced by [114]. Briefly, the CANDELS photometric redshifts were computed by different teams using different codes and template-fitting prescriptions, which were then complied by [93] to derive a "CANDELS-team" redshift as the median of all individual team's $z_{\rm phot,can}$ values. Building on initial efforts by [93], [114] used a hierarchical Bayesian approach to combine individual participant's $z_{\rm phot,can}$ probability density functions (PDFs) for each galaxy and derived more accurate point estimates and self-consistent $z_{\rm phot}$ uncertainties ($\sigma_{\rm z,can}$).

In Figure 27 (left-panel), we show the $\sigma_{\rm z,can}/(1+z_{\rm phot,can})$ as a function of the *HST* CANDELS *H*-band magnitude ($H_{\rm 160w,can}$) for all the galaxies (with good data; $\rm PhotFlag = 0$) in the CANDELS UDS field. For each galaxy in the SC SAM mock lightcone data, we use its *H*-band magnitude ($H_{\rm 160w,sam}$) to assign a $\sigma_{\rm z}/(1+z_{\rm phot})$ value by randomly sampling the CANDELS $\sigma_{\rm z,can}/(1+z_{\rm phot,can})$ distribution in bins of $\Delta H_{\rm 160w} = 0.5\,{\rm mag}$. We then apply the sampled SAM redshift uncertainty ($\sigma_{\rm z,sam}$) to the true SC SAM redshift ($z_{\rm sam}$) of each galaxy and then generate its corresponding SAM photometric redshift $z_{\rm phot,sam}$ value by re-sampling a Gaussian distribution centered on the true redshift $z_{\rm sam}$ and with a standard deviation of $\sigma_{\rm z,sam}$. In Figure 27 (left-panel), we also show the

$\sigma_{\rm z,sam}/(1+z_{\rm phot,sam})$ vs $H_{\rm 160w,sam}$ for $\log_{10}(M_{\star,\rm sam}/M_{\odot}) \gtrsim 9.4$ galaxies[1] and illustrate how they statistically mimic the CANDELS distribution in this parameter space. In Figure 27 (right panel), we show the redshift normalized difference between the assigned $z_{\rm phot,sam}$ and $z_{\rm sam}$ as a function of $z_{\rm phot,sam}$ for SAM galaxies in the *left panel* in conjunction with analogous $(z_{\rm phot,can} - z_{\rm spec})$ quantity for all CANDELS galaxies with good quality $z_{\rm spec}$ information. From this diagnostic, we illustrate that our $z_{\rm phot,sam}$ and $\sigma_{\rm z,sam}$ assignment also self-consistently mimics the empirical CANDELS $(z_{\rm phot} - z_{\rm spec})$ parameter space.

3.2.2.2 Stellar-mass errors

Stellar-mass measurements of galaxies play a critical role in selecting massive galaxy samples of interest for close-pair analysis. Moreover, the stellar-mass ratio M_1/M_2 is also an important variable used to select major pairs ($M_1/M_2 \leq 4$). To understand the role played by stellar-mass errors on the pair fraction measurements, for each galaxy in our SAM mock catalog, we mimic an empirically constrained, redshift dependent stellar-mass systematic ($\mu_{\log M}$) and a Gaussian scatter ($\sigma_{\log M}$) derived by [219] as:

$$\mu_{\log M} = -0.044\frac{z}{1+z} + 0.041\ [\rm dex], \tag{3.1}$$

and

$$\sigma_{\log M} = min[0.07 + 0.07z, 0.3][\rm dex], \tag{3.2}$$

where we select either $\sigma_{\log M} = 0.07 + 0.07z$, or $\sigma_{\log M} = 0.3$, whichever is the minimum value at a given redshift.

[1]We choose this sub-sample as they are potential $< 4:1$ companions to $\log_{10}(M_{\star,\rm sam}/M_{\odot}) \sim 10$ galaxies.

For each galaxy in our mock lightcone, we fold in the redshift errors simultaneously by assigning a $\mu_{\log M}$ and $\sigma_{\log M}$ based on its $z_{\rm phot,sam}$. Assuming a log-normal Gaussian centered at $\log_{10} M_{\rm stellar,sim} + \mu_{\log M}$ and a standard deviation of $\sigma_{\log M}$, we randomly assign a mock stellar-mass $\log_{10}(M_{\rm stellar,mock})$ for each galaxy. In Figure 28 (left-panel), we show the $\log_{10}(M_{\rm stellar,sim}) - \log_{10}(M_{\rm stellar,mock})$ behaviour of a subset of galaxies ($0.5 < z < 1.0$ and $|z_{\rm mock} - z_{\rm sim}| < 0.1$) as a function of $\log_{10}(M_{\rm stellar,sim}/M_{\odot})$ in the SC mocks. We note that the median offset and standard deviations are consistent with prescription given by equations 3.1 and 3.2. During our close-pair analysis (§ 3.4), we quantify the impact of using $\log_{10} M_{\rm stellar,mock}$ on the selection of primary galaxy samples and major pair selection.

3.2.2.3 Detection Completeness

To understand the role of missing pairs due to source detection completeness limits while measuring the pair fraction, we incorporate empirically-motivated, magnitude-dependent detection completeness limits to our mock lightcones, which conceptually translate into stellar-mass dependent completeness limits. We use the H-band magnitude based completeness fractions of the CANDELS fields by [205] at two survey configurations: Wide1 and Wide2. Briefly, Wide1 configuration corresponds to the shallowest mode of CANDELS observations with a 80% magnitude completeness limit ($H_{80} = 25.3\,{\rm mag}$), whereas the Wide2 scenario is at a moderate depth with $H_{80} = 25.8\,{\rm mag}$. Note that $\sim 57\%$ of the CANDELS fields have been observed at a Wide2 setting, however, with considerable amount of area ($\sim 20\%$) in Wide1 configuration (see [205]).

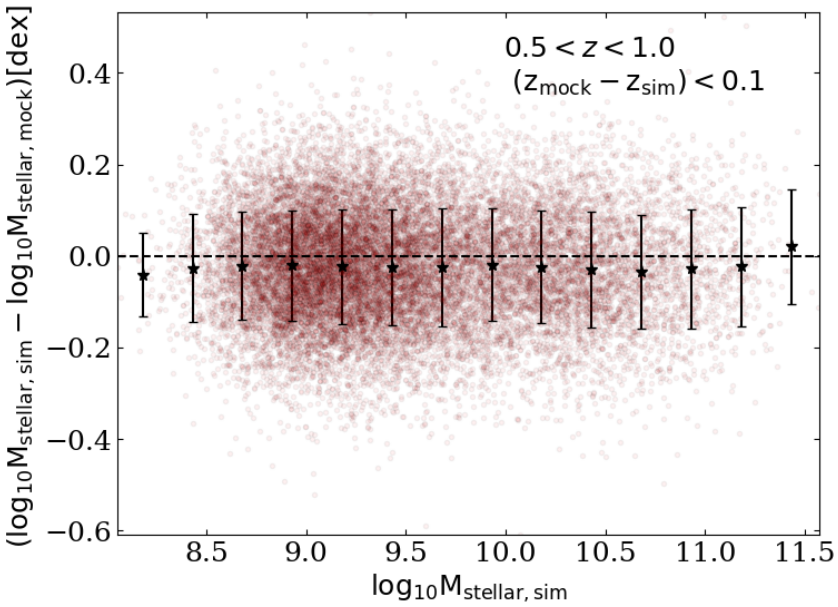

Figure 28: The difference between the true and realistically mimicked observed stellar masses ($\log_{10} M_{\rm stellar,sim} - \log_{10} M_{\rm stellar,mock}$) in bins of $\log_{10} M_{\rm stellar,sim}$ at $0.5 < z < 1.0$ for galaxies that satisfy $|z_{\rm mock} - z_{\rm sim}| < 0.1$ in the SantaCruz SAM mock lightcones (red background points). We show the running medians and standard deviation of the difference in 0.25 dex bins of $\log_{10} M_{\rm stellar,sim}$ (black points with errorbars).

To assign completeness probabilities to galaxies across a wide range of magnitudes in our mock catalogs, we analytically approximate their completeness fractions using a modified (reverse) sigmoid function of the form:

$$c = \frac{1}{1 + p \times \exp\left[q(H_{\rm mag} - r)\right]} \ . \tag{3.3}$$

We choose this parametric functional form because it mathematically follows the general shape of the common completeness functions in the literature (e.g., see [16, 140, 205]). The parameters p, q conceptually control the decaying shape of the completeness curve, whereas the r parameter controls the shift along the magnitude axis. An alternative to our approach is to use the empirical completeness curves with spline interpolation between

the magnitude bins and directly adopting them onto the simulated data. However, we expect our adopted choice of the modified sigmoid function will have a minimum impact on our analysis results.

In Figure 29 (left-panel), we show the CANDELS H-band completeness limits for the Wide1 and Wide2 configurations from [205]. We perform least-square fit to this data with equation 3.3 and find best-fit (p, q, r) parameters $(2.5, 2.0, 27.0)$ and $(2.5, 2.0, 26.5)$ for Wide1 and Wide2 configurations, respectively. In Figure 29, we show the best-fit completeness curves alongside the CANDELS survey completeness estimates [205]. For both the Wide1 and Wide2 configuration based best-fit completeness curves, we also quote the 80% completeness magnitude H_{80} by substituting their corresponding best-fit p, q, r values in equation 3.3 and equating it to a completeness value $= 0.8$.

We assign detection probabilities to each galaxy in our SC-SAM mock catalogs at two CANDELS-survey depth configurations (Wide1 and Wide2) by substituting the galaxy's simulated H-band magnitude $H_{\rm sim}$ and their corresponding best-fit (p, q, r) values in equation 3.3. In Figure 29 (right panel), we show the fraction of galaxies that are missed due to the detection completeness fraction within three stellar-mass bins $\log_{10}(M_{\rm stellar}/M_{\odot}) \in [9.4, 10]$ for CANDELS Wide1 and Wide2 depth configurations. Our notable finding based on this figure is that up to $\sim 10\%$ ($\sim 5\%$) of the $\log_{10}(M_{\rm stellar}/M_{\odot}) \in [9.4, 9.7]$ galaxies, which are potential major companions to $\log_{10}(M_{\rm stellar}/M_{\odot}) \sim 10$, can be missed at the $1.5 < z < 3$ in the CANDELS Wide1 (Wide2) configuration. We also find that nearly all $\log_{10}(M_{\rm stellar}/M_{\odot}) > 9.4$ galaxies are retained at $z < 1$ irrespective of the depth configuration, however, at $z > 1$, the completeness limits start to show an effect

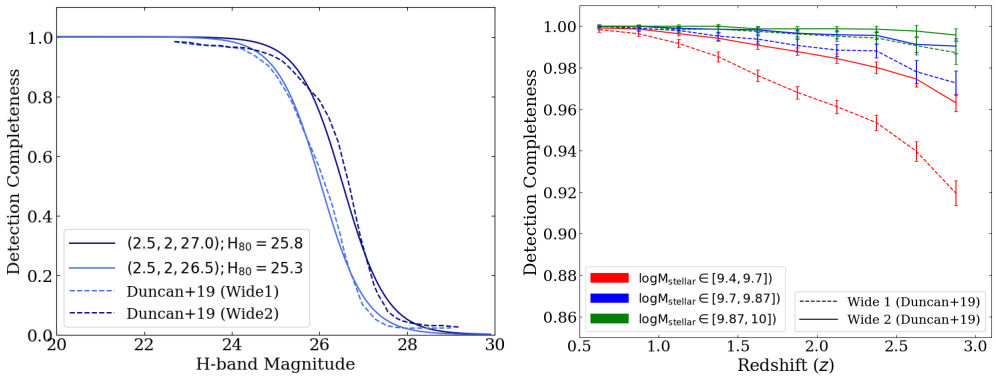

Figure 29: *Left-panel:* Extended source detection completeness as a function of H-band magnitude by [205] for two survey depth configurations: Wide1 (sky blue dashed line) and Wide2 (dark blue dashed line). We also show the best-fit reverse sigmoid function (equation 3.3) to their respective empirical completeness in solid lines, and report the best-fit equation 3.3 parameters (p, q, r) and magnitude limit at which the detection completeness is 80% (H_{80}) in the legend. *Right-panel:* Fraction of galaxies in three stellar-mass bins $\log_{10}(M_{\rm stellar}/M_{\odot}) \in [9.4, 10]$ (red, blue, and green colors) that are missed due to the incorporated source detection completeness (completeness fraction) as a function of redshift for CANDELS Wide1 (solid lines) and Wide2 (dashed lines) configurations. The error bars correspond to the standard deviation of the completeness fraction from 100 bootstrap resamplings.

on these trends, where $\leq 5\%$ of the $\log_{10}(M_{\rm stellar}/M_{\odot}) \in [9.7, 10]$ galaxies are missed at $2.5 < z < 3$ in the Wide1 configuration. Furthermore, we also note that the effect of incorporated completeness limits is less severe for a deeper, Wide2 configuration, which is expected because it has better detection completeness at fainter magnitude limits. In § 3.3.2.3, we investigate how the missing of potential major companions as a function of different survey depth settings affects the pair fraction measurements.

3.3 Impact of Observational Effects on Close-pair Measurements

To quantify the impact of observational effects on the pair fraction measurements,

first we discuss five commonly employed pair selection criteria explored in our analysis, and the computation of $f_{\rm pair}(z)$. Next, we quantify the impact of the observational effects (photometric redshifts and their uncertainties, stellar-mass systematic and random measurement errors, source detection completeness, and intrinsic field-to-field variations) by assessing the detection of close pairs and measuring the $f_{\rm pair}$ in SC-SAM mocks with and without such observational effects.

3.3.1 Close-pair selection functions & $f_{\rm pair}$ derivation

To assess the impact of different observational effects for different pair selection criteria, we analyze close pairs within the SC-SAM mock light cone data for a selective set of five commonly employed pair selection functions shown in Table 8.

Table 8: Close-pair selection criteria explored in our analysis. All the following selections use a common stellar-mass ratio based major-pair selection criterion $M_1/M_2 \leq 4$. For each selection, we indicate the information used for redshift and stellar mass.

Selection	Stellar-mass Limit $\log_{10}(M_{\rm stellar}/M_{\odot})$	Proj. Distance $R_{\rm proj} \in$ [kpc]	Physical Proximity Δv or Δz	Information used
A	≥ 10	$[5, 50]$	$\Delta v \leq 500$ km/s	$z_{\rm sim}$ & $M_{\rm stellar,sim}$
B	≥ 10	$[5, 30]$	$\Delta v_{12} \leq 500$ km/s	$z_{\rm sim}$ & $M_{\rm stellar,sim}$
C	≥ 10	$[5, 50]$	$\Delta z_{12} \leq \sqrt{\sigma_{z,1}^2 + \sigma_{z,2}^2}$	$z_{\rm mock}$ & $M_{\rm stellar,mock}$

D	≥ 10	$[5, 50]$	$\Delta z_{12}/(1+z) \leq 0.05$	$z_{\rm mock}$ & $M_{\rm stellar,mock}$
E	≥ 10.7	$[5, 50]$	$\Delta v_{12} \leq 500$ km/s	$z_{\rm sim}$ & $M_{\rm stellar,sim}$

Briefly, our selection criteria spans two stellar-mass limited choices $\log_{10}(M_{\rm stellar}/M_{\odot}) \geq 10$ and $\log_{10}(M_{\rm stellar}/M_{\odot}) \geq 10.7$, two projected distance selections $R_{\rm proj} \in [5-30]\,{\rm kpc}$ and $\in [5-50]\,{\rm kpc}$, one stellar-mass ratio based major-pair selection $M_1/M_2 \leq 4$. Additionally, we also explore three physical proximity choices – spectroscopic velocity difference $\Delta v_{12} \leq 500\,{\rm km\,s^{-1}}$ [5], photometric redshift proximity: $\Delta z_{12} = |z_1 - z_2|/(1+z) \leq 0.05$ [128] (hearafter S17 criteria) and $\Delta z_{12} \leq \sqrt{\sigma_{z,1}^2 + \sigma_{z,2}^2}$ [6] (hearafter B09 criteria).

Following the standard practice in the literature, we quantify the number of massive galaxies that have a major companion in close physical proximity ($N_{\rm pair}$) in each redshift bin for a chosen close-pair selection criteria, and compute the close-pair fraction as:

$$f_{\rm pair}(z) = \frac{N_{\rm pair}(z)}{N_{\rm m}(z)}\,, \tag{3.4}$$

where $N_{\rm m}$ is the number of massive galaxies satisfying one of the aforementioned stellar-mass limit. We also compute the 68.3% binomial confidence limits for the $f_{\rm pair}$ using the number counts of $N_{\rm pair}$ and $N_{\rm m}$.

3.3.2 Impact of Observational Effects

We analyze close pairs among the SC-SAM mock catalogs with realistically mimicked observational effects to quantify the impact of photometric redshift uncertainties, stellar-mass errors, source detection completeness, and intrinsic field-to-field pair fraction variations.

3.3.2.1 Impact of Photometric Redshift Errors in Selecting Physical Proximity Pairs

The choice of the redshift proximty selection can impact the quantification of close-pair counts, especially during the epochs where redshift uncertainties are high ($z \gtrsim 1.5$). Conceptually, photometric redshift proximity criteria selects pairs that have small redshift separations permitted within the redshift uncertainties of the paired host and companion galaxies. The typical photometric redshift errors translate to several $\sim 1000\,\mathrm{km\,s^{-1}}$ in relative velocity differences, especially at high redshifts ($z \gtrsim 1.5-2$), thereby leading to a systematically higher number of selected close pairs than the stringent $\Delta v_{12} \leq 500\,\mathrm{km\,s^{-1}}$ selection. In Figure 30 (*left panel*), we show the velocity difference Δv_{12} vs absolute redshift difference $|\Delta z_{12}|$ for close pairs at $0.5 < z < 1$. In this figure, we visually demonstrate how several pairs with a typical redshift difference $\Delta z \leq 0.05$ at $0.5 < z < 1.0$ have high relative velocities ($\Delta v_{12} > 500\,\mathrm{km\,s^{-1}}$).

To understand the impact of photometric redshift proximity criterion in identifying $\Delta v_{12} < 500\,\mathrm{km\,s^{-1}}$ pairs, we quantify a redshift proximity calibration weight factor (w_{zprox}):

$$w_{\mathrm{zprox}} = \frac{N_{\mathrm{pair}}\,[(\mathrm{B09\ or\ S17})\ \mathbf{and}\ \Delta v_{12} \leq 500\,\mathrm{km\,s^{-1}}]}{N_{\mathrm{pair}}\,[\mathrm{B09}\ \mathbf{or}\ \mathrm{S17}]}\ . \tag{3.5}$$

Conceptually, in the the above equation, $w_{\rm zprox}$ translates to the fraction of pairs selected using either B09 or S17 photometric redshift selection that also satisfy $\Delta v_{12} \leq 500\,{\rm km\,s^{-1}}$.

In Figure 30 (*right panel*) we show $w_{\rm zprox}$ as a function of redshift and find that only $\sim 35\%$ of the B09 or S17 selected pairs at $z \sim 1$ have small velocities ($< 500\,{\rm km\,s^{-1}}$) and the $w_{\rm zprox}$ falls to 30% at $z \sim 3$. We also note that the $w_{\rm zprox}$ values for B09 proximity selection are higher than the S17 criterion at $z \lesssim 2$, however, they converge to the same values at $z \gtrsim 2$. This behavior is a consequence of both B09 and S17 redshift proximity selections probing the same Δv_{12} pairs at $z \gtrsim 2$. Our finding highlights that the choice of redshift proximity selection, especially based on photometric redshifts and their uncertainties can bear significant systematic offset in terms of selecting gravitationally-bound pair counts by up to ~ 3 times. In § 3.5, we demonstrate the application of this $w_{\rm zprox}(z)$ correction to empirical pair fraction data from CANDELS to infer an expected empirical $\Delta v_{12} \leq 500\,{\rm km\,s^{-1}}$ pair fraction measurement.

3.3.2.2 Quantitative Accuracy of Stellar-Mass Uncertainties in Choosing True Major Pairs

In addition to the photometric redshifts and their uncertainties, incorporating realistic observational systematic and random uncertainties for stellar-masses can result in some major pairs to not satisfy the major pair criterion ($M_1/M_2 \leq 4$), and some to falsely satisfy it. To quantify the extent of this effect, we measure the accuracy of mock stellar-mass ratio $[M_1/M_2]_{\rm mock} \leq 4$ criterion in selecting pairs that intrinsically satisfy the simulated stellar-mass ratio criterion $[M_1/M_2]_{\rm sim} \leq 4$. We quantify the major pair correction

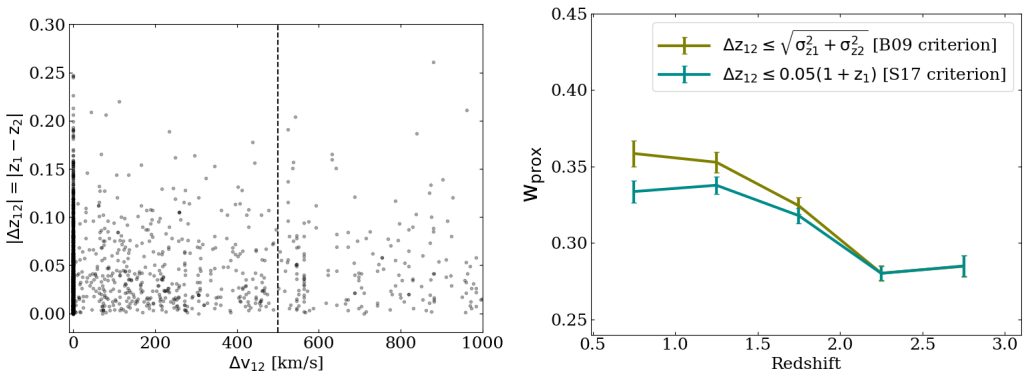

Figure 30: *Left-panel:* Velocity difference Δv_{12} vs. the absolute mock photometric redshift difference Δz_{12} for major close pairs during $0.5 < z < 1$ from the SC-SAM mocks. We also indicate $\Delta v_{12} = 500\,\mathrm{km\,s^{-1}}$ using the vertical line. *Right panel:* The fraction of pairs selected based on photometric redshift proximity selections: i) $\Delta z_{12} \leq \sqrt{\sigma_{z1}^2 + \sigma_{z2}^2}$ (olive-green line; B09 criteria) and ii) $\Delta z_{12} \leq 0.05(1 + z_1)$ (cyan line; S17 crietria) that satisfy spectroscopic velocity difference $\Delta v_{12} \leq 500\,\mathrm{km\,s^{-1}}$ as a function of redshift between $0.5 < z < 3.0$.

weight $w_{\rm maj}$ as the multiplicative factor to the mock major pair counts to account for false positives and false negatives as:

$$w_{\rm maj} = \frac{TP - FP + FN}{TP + FP}, \tag{3.6}$$

where the TP, FP, and FN are true positives, false positives, and false negatives, respectively. They are defined as follows:

$$\begin{aligned} TP &= N_{\rm pair}([M_1/M_2]_{\rm mock} \leq 4\,|\,[M_1/M_2]_{\rm sim} \leq 4) \\ FP &= N_{\rm pair}([M_1/M_2]_{\rm mock} \leq 4|[M_1/M_2]_{\rm sim} > 4). \\ FN &= N_{\rm pair}([M_1/M_2]_{\rm mock} > 4|[M_1/M_2]_{\rm sim} \leq 4), \end{aligned} \tag{3.7}$$

where $N_{\rm pair}({\rm condition1}|{\rm condition2})$ is the number of pairs satisfying condition 1, given that they already meet condition 2. Conceptually, TP pairs satisfy both the mock and

simulated stellar-mass ratio criteria ($M_1/M_2 \leq 4$). FP pairs have their mock mass ratios $[M_1/M_2]_{\rm mock} < 4$, but have their simulation $[M_1/M_2]_{\rm sim} > 4$. Similarly, FN pairs are those which have their simulation $[M_1/M_2]_{\rm sim} < 4$, but have a mock mass ratio $[M_1/M_2]_{\rm mock} > 4$.

In Figure 31 (left-panel), we show a demonstrative visualization of the pairs in the simulation mass ratio (using $M_{\rm stellar,sim}$) vs. mock stellar-mass ratio (using $M_{\rm stellar,mock}$) parameter space, satisfying $R_{\rm proj} \in [5, 50]\,{\rm kpc}$ and $\Delta v_{12} \leq 500\,{\rm km\,s^{-1}}$ (using $z_{\rm sim}$ in the $0.5 < z < 1.0$ redshift bin. In this example, we find that $\sim 15\%$ of the simulated major pairs ($[M_1/M_2]_{\rm sim} \leq 4$) are scatter to $[M_1/M_2]_{\rm mock} > 4$ mock mass ratios. We also note that $\sim 17\%$ of the mock major pairs $[M_1/M_2]_{\rm mock} \leq 4$ are contaminant minor pairs ($[M_1/M_2]_{\rm sim} > 4$). Using equations 3.6 & 3.7, we find that this corresponds to a $w_{\rm maj} = 0.8$, indicating that the mock pair counts at $0.5 < z < 1$ need to be reduced by 20% to accurately represent the simulation pair counts.

In Figure 31 (right-panel), we show the $w_{\rm maj}$ as a function of redshift for the three redshift proximity selections (Δv_{12}, B09, and S17). Our generalized interpretation of this suggests that $w_{\rm maj}$ follows a flat trend with redshift and ranges between $\sim 0.75 - 0.95$ depending on the redshift bin. Additionally, we also note that the $w_{\rm maj}$ values based on Δv_{12} criteria are systematically lower at each redshift bin than the photometric redshift proximity selections (B09 and S17), likely as a consequence of propagating redshift errors into $M_{\rm stellar,mock}$ assignment. Furthermore, the $w_{\rm maj}$ values when using the B09 and S17 are consistent with each other at all redshifts, indicating that the $w_{\rm maj}$ measurement is less sensitive to the nature of photometric redshift proximity selection used.

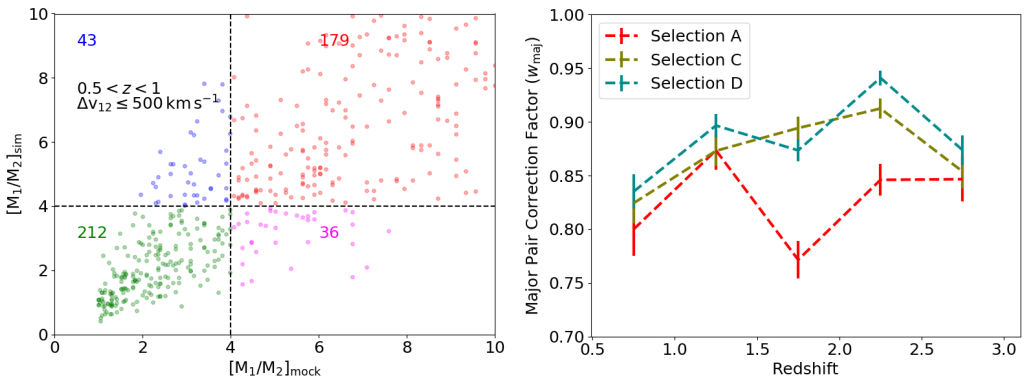

Figure 31: *Left-panel:* The mock stellar-mass ratio $[M_1/M_2]_{\rm mock}$ vs. simulation mass ratio $[M_1/M_2]_{\rm sim}$ for a sample of pairs satisfying $\Delta v_{12} \leq 500\,{\rm km\,s^{-1}}$ and $R_{\rm proj} \in [5 - 50\,{\rm kpc}]$ (Selection A, see Table 8) at $0.5 < z < 1.0$. We divide the axes into four quadrants (colored points) based on $[M_1/M_2]_{\rm mock} \leq 4$ and $[M_1/M_2]_{\rm sim} \leq 4$ lines and indicate the number of pairs per quadrant in text. *Right-panel:* The redshift evolution of the major pair correction factor $w_{\rm maj}$ (see equations 3.6 & 3.7) for pairs satisfying $\Delta v_{12} \leq 500\,{\rm km\,s^{-1}}$ (red lines), $\Delta z_{12} \leq \sqrt{\sigma_{z1}^2 + \sigma_{z2}^2}$ (B09 criteria; olive-green lines), and $\Delta z_{12} \leq 0.05(1+z_1)$ (S17 criteria, cyan lines). The errorbars of $w_{\rm maj}$ correspond to its 1σ binomial confidence limits.

Within the assumed $M_{\rm stellar}$ error budget incorporated into our mock light cones, our findings imply that the observed major pair counts require a reduction of $\sim 5 - 25\%$ to accurately represent the true major pair counts and account for false contaminant minor pairs and missed major pairs due to the stellar-mass errors. It is that this important effect be taken into account when empirically quantifying the pair fractions and in § 3.5 we show an application of the $w_{\rm maj}$ correction factors to CANDELS pair fraction data.

3.3.2.3 Impact of Detection Completeness on the $f_{\rm pair}(z)$

To quantify the impact of sample completeness on the pair fraction measurements, we incorporate the magnitude-dependent detection probabilities for each galaxy in the SC mock light cone catalogs (see § 3.2.2.3). We propagate this information to quantify the

pair fraction *with completeness* ($f^{'}_{\rm pair}$), such that both host and companion galaxies in a pair are deemed "detected" as per a weighted coin-toss based rejection sampling, where detection and non-detection ($1-$detection) probabilities are assigned using the completeness equation 3.3. We then quantify completeness correction weight factor ($w_{\rm comp}$) as the ratio of pair fraction measurements with and without the incorporated completeness effects ($f^{'}_{\rm pair}/f_{\rm pair}$).

We quantify the $w_{\rm comp}$ as a function of redshift for three selections A, C, and D (see Table 8), and assess the $w_{\rm comp}$ dependence on the nature of the imposed completeness function by quantifying $w_{\rm comp}$ as a function of H_{80} values spanning $H_{80} \in [24, 26]\,{\rm mag}$. We achieve this by changing the r parameter between $25 \lesssim r \lesssim 27$ in equation 3.3 and fixing the p, q variables to their best-fit values. Conceptually, in this step we preserve the overall shape of the completeness function by holding the p, q values constant and shifting the overall completeness function to higher/lower H_{80} values by changing the r values. We quantify w_{comp} as a function of H_{80} as follows:

$$w_{\rm comp}(z, H_{80}) = \frac{f^{'}_{\rm pair}}{f_{\rm pair}}(z, H_{80}), \tag{3.8}$$

where the $f^{'}_{\rm pair}$ and $f_{\rm pair}$ are the SC-SAM mock data based pair fraction measurements with and without incorporating the detection completeness effects, respectively.

In Figure 32 (top-left panel), we show the $w_{\rm comp}$ as a function of H_{80} for pair selection A (see Table 8) at five redshift slices between $0.5 < z < 3$ (see figure legend), highlighting the sole effect of source-detection completeness on the measured pair fractions. For pair selection A, we find that the $w_{\rm comp} \sim 1$ for both Wide1 ($H_{80} = 25.3\,{\rm mag}$)

and Wide2 $H_{80} = 25.8\,\mathrm{mag}$) CANDELS-depth configurations at $z \lesssim 1.5$. This suggests that nearly all expected close pairs are detected at the Wide1 and Wide2 depths, and the detection completeness effects have a minimal impact on the f_{pair} measures at these epochs. However, at $z \gtrsim 2$, the incorporated incompleteness starts to show its notable effect where the some lower-mass companions are no longer detected. As a result, for the Wide1 configuration, we find that the $w_{\mathrm{comp}} \sim 0.9$ and $w_{\mathrm{comp}} \sim 0.8$ at $z \sim 2$ and $z \sim 3$, respectively. We also note that the impact of detection completeness effects are less severe for the deeper Wide2 depth configuration with $w_{\mathrm{comp}} \sim 0.9$ only at $z \sim 2.5 - 3$.

To assess the impact of sample completeness alongside other observational effects such as photometric redshifts and stellar-mass uncertainties, we analyze the w_{comp} as a function of H_{80} for pair selections C and D, and show it in Figure 32. We find that at a fixed redshift slice (e.g., at $z \sim 3$ and $H_0 = 25.3\,\mathrm{mag}$), the w_{comp} for selection C case has a smaller value (~ 0.65) than for the selection D (~ 0.8). This effect although exists at $H_0 = 25.8\,\mathrm{mag}$, it is less severe with $w_{\mathrm{comp,obs}} \sim 0.8$ vs. ~ 0.9 for selections C and D, respectively.

Our results based on Figure 32 conceptually suggest that regardless of the pair selection choices, the w_{comp} depends strongly on the H-band detection completeness limit for a fixed stellar-mass selected sample, and the mock pair fractions need to be corrected by a relative f_{pair} amount of $10 - 20\%$ depending on the redshift. They also illustrate the simultaneous and compound interplay between the redshift proximity and detection incompleteness limits on the measurement of pair fractions. We encourage future close-pair studies to incorporate the w_{comp} trends shown in Figure 32 for their specific pair

selection functions and survey detection limits to correct for the effect of incompleteness on the measurement of pair fractions.

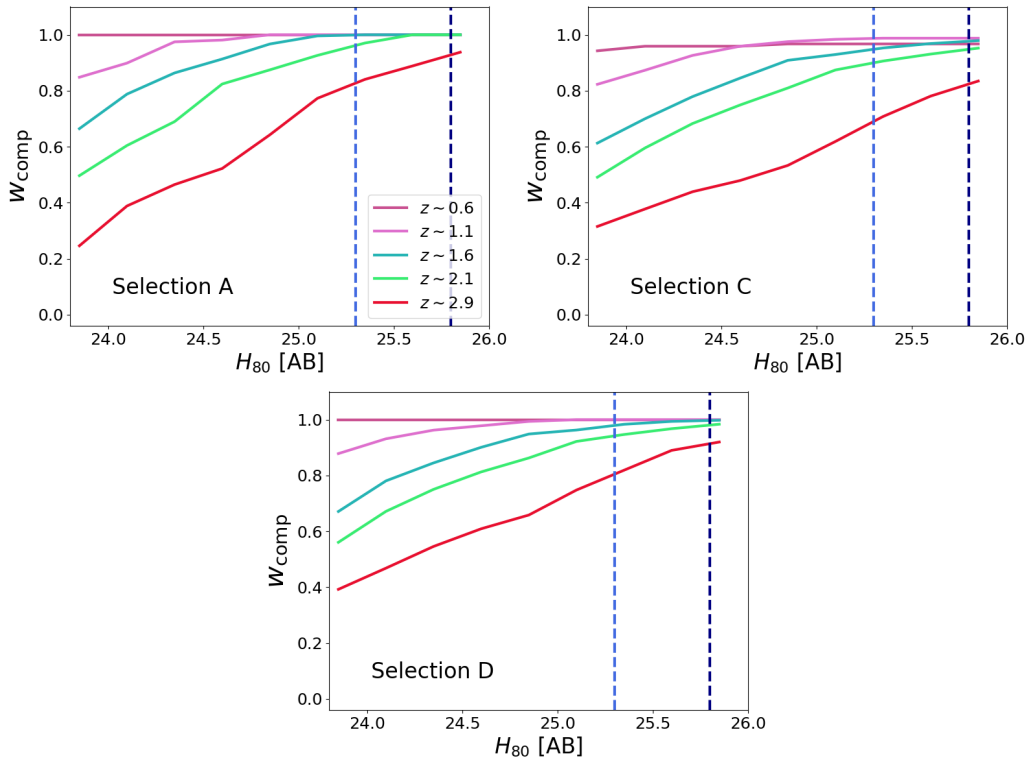

Figure 32: The pair fraction completeness correction $w_{\rm comp}$ as a function of the 80% H-band detection completeness limit for three pair selections A, C, and D (see Table 8), which adopt the following redshift proximity selections: i) $\Delta v_{12} \leq 500\,{\rm km\,s^{-1}}$ (top-left panel), ii) $\Delta z_{12} \leq \sqrt{\sigma_{z1}^2 + \sigma_{z2}^2}$ (top-right panel), and iii) $\Delta z_{12} \leq 0.05(1+z_1)$ (bottom panel), respectively. In each panel, we show the $w_{\rm comp}$ trends at five redshift slices indicated by different colored lines (see legend in first panel).

3.3.2.4 Impact of Field-to-Field Pair Fraction Variations

The pair fraction measurements can vary from field to field owing to the intrinsic variations in the host and companion galaxy number counts from sample variance, and this may explain the $f_{\rm pair}$ differences between different previous empirical studies, especially

among those that probe smaller survey volumes. To quantify this effect and provide calibration corrections for future empirical studies, we measure the $f_{\rm pair}$ at $0.5 < z < 3.0$ for five pair selections (A,B,C,D, and E in Table 8) using 64 unique mock fields (sized $\sim 160\,{\rm arcmin}^2$ each) sampled from the multiple realisations of $\sim 8\times$ larger SC-SAM's CANDELS-UDS mock light cone footprints (see § 2.2).

To quantify the field-to-field variation of the pair fraction measurement, we generate 1000 bootstrap resamplings of the $f_{\rm pair}$ values per redshift bin for each of our pair selections. We then compute their corresponding standard deviations $\sigma(f_{\rm pair})$, and measure the field-to-field variance of pair fractions as the median $\sigma(f_{\rm pair})$ value across 1000 bootstraped values.

To investigate the dependence of $\sigma(f_{\rm pair})$ on the choice of projected distance and stellar-mass limit, in Figure 33 (left panel), we show the median $\sigma(f_{\rm pair})$ measurements along with their errorbars for the pair selections A, B, and E. For selection A, we find that $\sigma(f_{\rm pair})$ increases from $\sim 0.7\%$ to $\sim 1.5\%$ between $0.5 < z < 3.0$. For selection B, which has a smaller projected distance annulus ($5 - 30\,{\rm kpc}$) when compared to selection A, we find that $\sigma(f_{\rm pair})$ values are systematically lower by $\sim 2\times$, albeit with similar redshift dependence. On the other hand, for a selection E, which adopts a more massive galaxy selection ($\log(M_{\rm stellar}/M_{\odot}) \geq 10.7$) than selection A, $\sigma(f_{\rm pair})$ remains flat over $0.5 < z < 3$ at $\sim 0.5\%$ and is $\sim 2 - 3$ times smaller than the selection A.

Observational effects such as the photometric redshift and stellar-mass uncertainties can aggravate the galaxy number count variations by scattering of galaxies in and out of the redshift, stellar-mass, stellar mass-ratio criteria, thereby subsequently impacting

the $f_{\rm pair}$ measurements. In Figure 33 (right panel), we assess the $\sigma(f_{\rm pair})$ for selections A, C, and D, which span the $\Delta v_{12} \leq 500\,{\rm km\,s^{-1}}$, B09, and S17 redshift proximity selection choices, respectively (see Table 8). We find that $\sigma(f_{\rm pair})$ for both selection C and D are higher than selection A by $\sim 1.5-2$ times, reaffirming our aforementioned motivation statement. We also find that a selection D that adopts a liberal redshift proximity selection S17 has a systematically higher $\sigma(f_{\rm pair})$ value than selection C that uses the B09 criterion. Our investigation provides important calibrations for pair fraction field-to-field variations and offers key insights into their dependence on key pair selection variable choices.

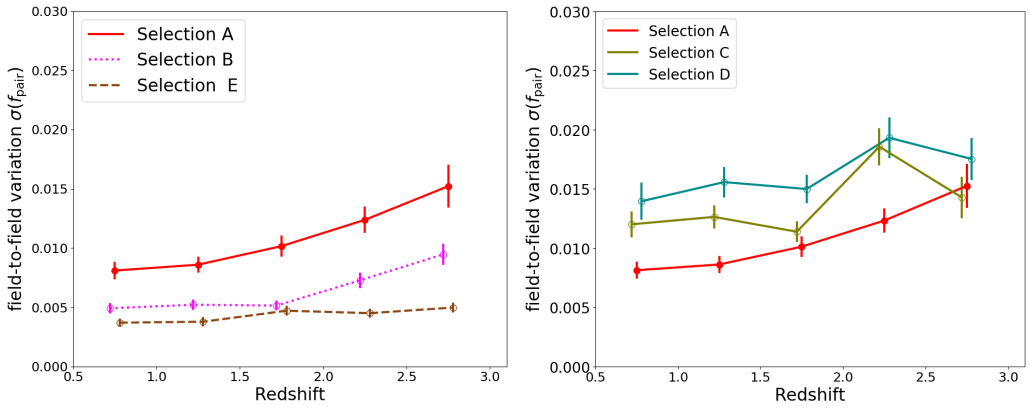

Figure 33: *Left-panel:* Field-to-field "true" pair fraction variation $\sigma_{f_{\rm pair,sam}}$ as a function of redshift ($0.5 < z < 3.0$) for $\log_{10}(M_{\star,\rm sam}/M_{\odot}) > 10$, $R_{\rm proj} = 5-50\,{\rm kpc}$ (filled markers; red solid lines) and $5-30\,{\rm kpc}$ (open markers, magenta dotted lines), and for $\log_{10}(M_{\star,\rm true}/M_{\odot}) > 10.7$, $5-50\,{\rm kpc}$ (open markers, brown dashed lines). *Right-panel:* Field-to-field variance of the "observed" close-pair fraction $\sigma(f_{\rm pair,sam,obs})$ as a function of redshift for $\Delta z_{12} \leq 0.05(1+z_1)$ (cyan line; S17 criteria) and $\Delta z_{12} \leq \sqrt{\sigma_{z1}^2+\sigma_{z2}^2}$ (olive line; B09 criteria). For comparison, we also show the corresponding $\sigma(f_{\rm pair,true})$ from the left panel (red line). In both panels, we show the median and standard deviation of the $\sigma(f_{\rm pair,*})$ from 1000 bootstrap resamplings of $f_{\rm pair}$ values in CANDELS-UDS sized fields ($\sim 160\,{\rm arcmin^2}$). For details, see § 3.3.2.4.

3.4 Close-Pair Fractions in SC-SAM Mocks

To assess the redshift evolution of the pair fraction $f_{\rm pair}(z)$, with and without various realistic observational selection effects, we analyze close pairs within the SC-SAM using a selective set of five pair selection criteria (see Table 8), spanning common choices adopted in the literature for stellar-mass, projected distance, redshift proximity.

3.4.1 $f_{\rm pair}(z)$ for Pair Selections A, B, and E

To assess the pair fraction evolution in the SC-SAM mocks without any observational effects, we quantify the $f_{\rm pair}(z)$ for selections A, B, and E, using the simulation information for redshift, and stellar-mass $z_{\rm sim}$ and $\log_{10} M_{\rm stellar,sim}$, respectively. The selections A and B only differ in their choice of projected distance annulus ($5 - 50\,{\rm kpc}$ vs. $5 - 30\,{\rm kpc}$, respectively), whereas the selections A and E differ in their choice of stellar-mass limit to select massive galaxy samples ($\log_{10}(M_{\rm stellar}/M_{\odot}) \geq 10$ vs. $\log_{10}(M_{\rm stellar}/M_{\odot}) \geq 10.7$, respectively).

To assess the $f_{\rm pair}(z)$ for two stellar-mass limit selected samples of massive galaxies, in Figure 34 (left panel), we show the $f_{\rm pair}(z)$ for selections A and E, and find that the pair fraction for $\log_{10}(M_{\rm stellar}/M_{\odot}) \geq 10$ galaxies rises from $\sim 3\%$ to $\sim 10\%$ during $0.5 \leq z \leq 3$. On the other hand, $f_{\rm pair}$ for more massive galaxies with $\log_{10}(M_{\rm stellar}/M_{\odot}) \geq 10.7$ evolves weakly with redshift from 0.5% to 1%, which is $\sim 5-10$ times smaller than $f_{\rm pair}$ values for selection A. This systematic $f_{\rm pair}(z)$ difference between selections A and E is a direct consequence of the steep shape of the stellar-mass function of SC-SAM galaxies at high-mass end, where at the number density of $\log_{10}(M_{\rm stellar}/M_{\odot}) \geq 10.7$

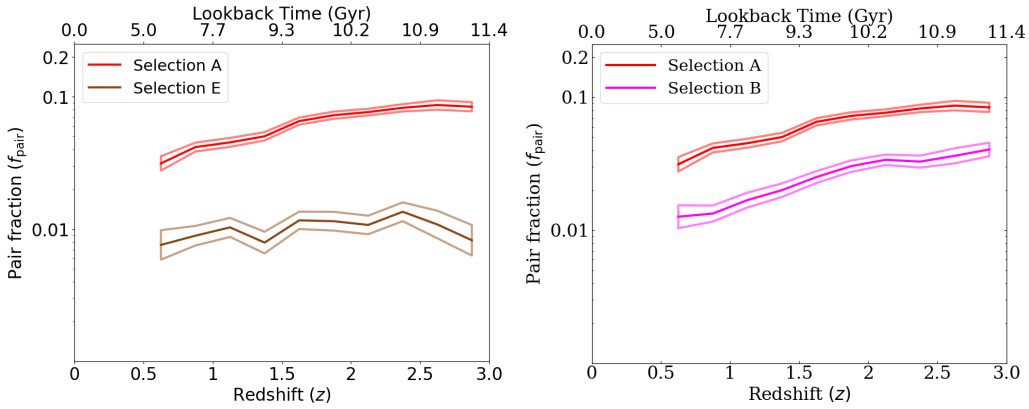

Figure 34: *Left-panel:* The $f_{\rm pair}(z)$ for two selections A and B using different mass-limited samples $\log_{10}(M_{\rm stellar}/M_{\odot}) > 10$ (red solid line) and $\log_{10}(M_{\rm stellar}/M_{\odot}) > 10.7$ (brown solid line). *Right-panel:* The $f_{\rm pair,sam}(z)$ evolution for pair selections A and B, which use two projected distance selections: $5-50\,{\rm kpc}$ (solid red lines) and $5-30\,{\rm kpc}$ (solid magenta lines), respectively. The transparent outer boundary to each opaque line corresponds to the 1σ binomial confidence level of $f_{\rm pair}$.

galaxies is smaller than the $\log_{10}(M_{\rm stellar}/M_{\odot}) \geq 10$ counterparts.

To assess the impact of using different projected distance annulus choice on the $f_{\rm pair}(z)$ trends, in Figure 34 (right panel), we show the $f_{\rm pair}(z)$ for selections A and B, which use $R_{\rm proj} = 5-30\,{\rm kpc}$ and $R_{\rm proj} = 5-50\,{\rm kpc}$, respectively. We find that the $f_{\rm pair}(z)$ evolution for both these selections evolves similarly, albeit with a systematic offset where $f_{\rm pair}$ for selection B is smaller than the selection A values. Such a difference is expected as as a larger selection annulus yields in a larger $N_{\rm pair}$ at each redshift bin, yielding higher $f_{\rm pair}$ values. This result is consistent with a similar analysis discussed in Chapter 1 [140].

3.4.2 $f_{\rm pair}(z)$ for Pair Selections C and D

To quantify the impact of observational effects such as photometric redshift, stellar-mass uncertainties, and detection completeness on the pair fraction measurements, we assess the $f_{\rm pair}(z)$ in the SC-SAM mocks for pair selections C and D (see Table 8), using the mock derived quantities for redshift and stellar masses $z_{\rm mock}$ and $\log_{10} M_{\rm stellar,mock}$, respectively. Briefly, these two pair selections differ in their choice of the redshift proximity (B09 vs. S17). For comparison purposes, we assess the $f_{\rm pair}(z)$ trends for selections C and D in conjunction with the selection A $f_{\rm pair}(z)$ measurements.

In Figure 35, we show the $f_{\rm pair}(z)$ evolution for selections C and D, alongside the pair fraction evolution for selection A (from Figure 34). We find that the $f_{\rm pair}(z)$ for both selections C and D rises from $\sim 6-7\%$ at $z \sim 0.5$ to $\sim 15\%$ at $z \sim 3$. This evolutionary trend is qualitatively comparable to the $f_{\rm pair}$ evolution for selection A, however, the $f_{\rm pair}(z)$ values for selections C and D are ~ 3 times larger than the selection A ones. As discussed in § 3.3.2.1, this is a consequence of the photometric redshift proximity criteria selecting high velocity pairs ($> 500\,{\rm km\,s^{-1}}$). We also notice that regardless of the selection choice, all the three trends (selections A, C, and D) stay flat at $z > 1.5$ relative to the rising pair fractions at $z < 1.5$. This is especially interesting as recent empirical (e.g., [205]) and theoretical (e.g., [128]) close pair studies find similar flattening of pair fractions at $z \gtrsim 1.5$. We also note that the selection D $f_{\rm pair}(z)$ stays systematically higher than the selection C values at $z \lesssim 2.5$, suggesting that the selection C (using B09 redshift proximity) is more restrictive than selection D in that redshift range. However, we find that the selection C and selection D $f_{\rm pair}$ values become nearly equal

at $z \gtrsim 2$, owing to the B09 and S17 criteria probing similar redshift difference (Δz_{12}) pairs when the typical photometric redshift uncertainty of the galaxies at $z > 2$ makes the B09 criterion behave quantitatively equivalent to the S17 selection (also see discussion in § 3.3.2.1). Our results based on Figure 35 illustrate that using photometric redshift proximity selections can inflate the derived pair fractions, where a significant portion of the identified pairs can be high relative velocity contaminant pairs.

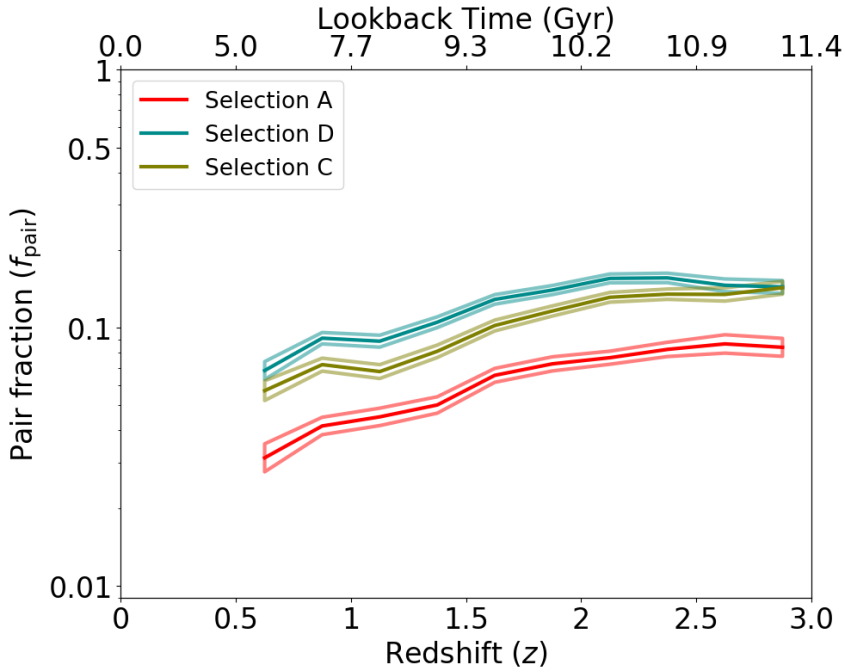

Figure 35: Redshift evolution of the close-pair fraction $f_{\rm pair}(z)$ during $0.5 < z < 3.0$. We show the $f_{\rm pair}$ evolution for pair selection C (olvie-green line) and D (cyan line), which use B09 and S17 redshift proximity criteria, respectively. For comparison purposes, we also show the $f_{\rm pair})(z)$ trend of pair selection A (red line,same as in Figure 34). The transparent lines encompassing the medians are 68.3% binomial confidence levels of pair fraction values.

3.5 Application of Pair-Fraction Caibrations to the CANDELS Pair Fractions

Thus far, we measured the redshift evolution of pair fractions in the SantaCruz SAM mock lightcones with realistically mimicked observational effects, and derived comprehensive calibrations that quantify the effect of photometric redshift and stellar mass errors, detection completeness, and field-to-field variance of the measured pair fractions. As a demonstrative application exercise for future close-pair studies to use calibrations derived in our work, we derive latest major close pair fractions $f_{\rm pair}$ among a massive sample in the five CANDELS fields, and apply the corrections to derive calibrated CANDELS pair fraction evolution measurements.

3.5.1 CANDELS Data

For our demonstrative application of observational selection effect calibrations, we derive latest CANDELS close-pair fraction among a sample of of 9540 massive galaxies $\log_{10}(M_{\rm stellar}/M_{\odot}) > 10$ spanning $0.5 < z < 3$ from the five CANDELS legacy fields: UDS, GOODS-N, GOODS-S, COSMOS, EGS.

We use the H-band (*HST*/WFC3 F160W) based source catalogs for each CANDELS field by [220] (UDS), [87] (GOODS-S), [88] (GOODS-N), [89] (COSMOS), [90] (EGS) and the latest photometric redshift information for them ($z_{\rm phot}$) data by [114]. In addition to the aforementioned details regarding the $z_{\rm phot}$ compilation (see § 3.2.2.1), [114] also compiled public spectroscopic redshift ($z_{\rm spec}$) data and defined a "best" available redshift estimate ($z_{\rm best}$) for each CANDELS galaxy using $z_{\rm spec}$ wherever available. Additionally, we use the CANDELS "team" stellar-mass estimates compiled by [94, 95]

which were generated by median combining individual mass estimates (of ten participants) based on different prescriptions for SED-fitting assumptions (see [95]). For our empirical analysis, we search for close companions hosted by 9540 massive galaxies (with photometric good data; PhotFlag$= 0$) satisfying $\log_{10}(M_{\rm stellar}/M_{\odot}) \geq 10$ and spanning $0.5 \leq z_{\rm best} \leq 3.0$.

3.5.2 Raw $f_{\rm pair}(z)$ in CANDELS

Using the CANDELS data described in § 3.5.1, we analyze close pairs among massive galaxies ($\log_{10}(M_{\rm stellar}/M_{\odot}) > 10$) satisfying pair selection C in Table 8. We quantify the raw (i.e., uncorrected for observation effects) CANDELS pair fraction evolution as a function of redshift ($f_{\rm pair}(z)$) using the equation 3.4 and show it in Figure 36. We find that the $f_{\rm pair}$ increases from ~ 0.08 (at $z \lesssim 1.5$) to ~ 0.11 (at $z \gtrsim 1.5$). At $1.5 \leq z_{\rm best} \leq 3.0$, we find that the $f_{\rm pair}(z)$ evolution is consistent with being flat. Later in this section, we will correct the raw CANDELS $f_{\rm pair}(z)$ for observational effects using calibrations from Section 3.4 to derived calibrated empirical pair fraction evolution.

It is important to note that the trends shown in Chapter I are not to be directly compared with those shown in Figure 36, as in Chapter I, we use a different quantity – the fraction massive galaxies in major pairs ($f_{\rm mp}$) that is different (but analogous) from $f_{\rm pair,can}$. The main difference between $f_{\rm mp}$ and $f_{\rm pair}$ is that in the former case both primary and companion galaxies are counted towards the numerator in equation 3.4 while simultaneously imposing a stellar-mass limit for massive galaxy selection on both of them. In an ideal scenario where every primary galaxy hosts a massive and major companion,

$f_{\rm mp} \sim 2 f_{\rm pair,can}$. However in practice, the ratio of $f_{\rm mp}$ and $f_{\rm pair,can}$ is redshift dependent owing to imposition of a fixed mass cut on a redshift dependent stellar-mass function. For example, while the companion galaxies satisfy the stellar-mass cut at $z \lesssim 1.5$, an increasing number of galaxy pairs at higher redshifts ($z \sim 2-3$) are between a massive primary galaxy and a less-massive one (with $4:1$) that do not necessarily satisfy the massive galaxy stellar-mass cut; In such cases $f_{\rm mp} \sim f_{\rm pair,can}$. Moreover, at a fixed redshift bin, the $f_{\rm mp}$ definition tends to probe similar (or nearly-equal) mass companions for galaxies near the edge of mass cut.

3.5.3 Corrected $f_{\rm pair}(z)$ in CANDELS

We derive the calibrated CANDELS pair fraction evolution $f_{\rm pair,corr}(z)$ by applying SC-SAM based pair fraction calibrations (§ 3.3) to the raw CANDELS $f_{\rm pair}$ values from above as:

$$f_{\rm pair,corr}(z) = (\frac{w_{\rm maj} \times w_{\rm prox}}{w_{\rm comp}}) \times f_{\rm pair}(z). \tag{3.9}$$

Briefly, $w_{\rm maj}$ accounts for the effect of stellar-mass systematic and random measurement uncertainties in selecting true major pairs (i.e., $[M_1/M_2]_{\rm sam} \leq 4:1$; § 3.3.2.2), $w_{\rm zprox}$ statistically corrects photometric redshift proximity selection to spectroscopic velocity difference criterion (§ 3.3.2.1), and $w_{\rm comp}$ corrects for missing pairs due to detection incompleteness (§ 3.3.2.3). Additionally, we propagate the effect of field-to-field pair fraction variation (§ 3.3.2.4) by adding the 1σ binomial $f_{\rm pair}$ errors and $\sigma(f_{\rm pair})$ in quadrature.

We first derive the corrected, photometric redshift based CANDELS $f_{\rm pair,corr}$ using equation 3.9 and $w_{\rm maj}$ calibration from § 3.3.2.2, $w_{\rm comp}$ calibration for CANDELS Wide2

depth configuration (§ 3.3.2.3), and by setting $w_{\rm zprox} = 1$. We also add the binomial errors of raw CANDELS $f_{\rm pair}$ and the $\sigma(f_{\rm pair})$ derived in § 3.3.2.4 in quadrature. In Figure 36, we show the corrected CANDELS $f_{\rm pair,corr}(z)$ and find that the statistical calibrations systematically lower the raw CANDELS $f_{\rm pair}$ values, where the dominant contribution is from $w_{\rm maj}$. After correcting for observational effects, we find that the corrected pair fraction $f_{\rm pair,corr}$ evolves weakly at $0.5 \lesssim z \lesssim 1.5$, and is consistent with being flat at $\sim 10\%$ during $1.5 \leq z_{\rm best} \leq 3.0$.

Leveraging our $w_{\rm zprox}$ calibration (§ 3.3.2.1) for selection C, we infer the redshift evolution of spectroscopic pair fraction in CANDELS by multiplying the above corrected photometric redshift based $f_{\rm pair,corr}$ with $w_{\rm zprox}$ for selection C (from § 3.3.2.1). We find that the CANDELS inferred spectroscopic pair fraction remains flat at $\sim 2.5\%$ during the entire redshift range $0.5 \leq z_{\rm best} \leq 3.0$.

3.5.4 Implications of Corrected Pair Fraction Evolution Measurements

Thus far, we quantified the impact of common observational selection effects on the measurement of close-pair fraction measurements, and derived calibration corrections for them. We applied these corrections to empirical pair fractions from the CANDELS survey to derive the latest calibrated $f_{\rm pair}(z)$ measurement over $0.5 < z < 3.0$. A key observation of our work is that the overall calibrated pair fractions are systematically smaller than the uncorrected $f_{\rm pair}$. Here, we discuss the implications our results on the overarching role and contribution of galaxy mergers towards stellar-mass assembly of high-mass galaxies.

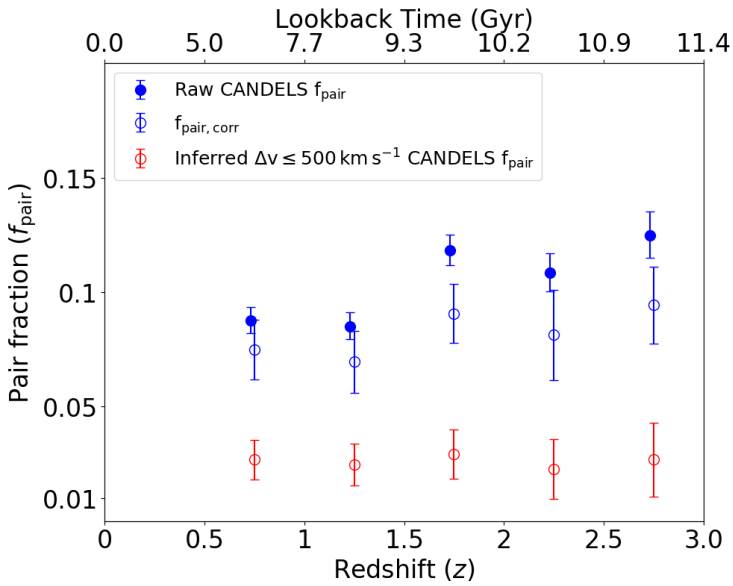

Figure 36: Redshift evolution of the major ($< 4:1$) close pair fraction $f_{\rm pair}(z)$ among ~ 9500 massive galaxies from the CANDELS survey (blue filled points with 68% binomial errors) using pair selection C from Table 8. We also show the corrected pair fraction evolution $f_{\rm pair,corr}(z)$ (open blue markers) after applying the major pair correction ($w_{\rm maj}$; from Figure 31), detection incompleteness ($w_{\rm comp}$; from Figure 32), and field-to-field pair fraction variation in CANDELS-sized fields ($\sigma(f_{\rm pair})$; from Figure 33). We also infer the spectroscopic ($\Delta v_{12} \leq 500\,{\rm km\,s^{-1}}$) pair fraction evolution (open red markers) after additionally correcting the *open blue points* with $w_{\rm zprox}$ from Figure 35. Note that the errors on the open points are quadrature sum of binomial confidence limits and $\sigma(f_{\rm pair})$.

Assuming the pair observability timescales doesn't depend on the observational effects, an overall systematic reduction in the $f_{\rm pair}(z)$ can lower the resultant major merger rate evolution. Conceptually, the consequence of major mergers on the stellar-mass growth can be thought of as mass deposition by the companion galaxy onto the host galaxy. As such, a systematically smaller merger rate would reduce the stellar-mass accretion rate caused major mergers. Recent theoretical works (e.g., [216]) have found that major mergers are a dominant source of external stellar-mass accretion and that they play

an increasingly important role in building up the present-day ($z \sim 0$) high-mass galaxies ($\log_{10}(M_{\rm stellar}/M_{\odot}) \geq 11$). However, if the observational selection effect corrected merger rates are to be used, then the major merger contribution to the build-up of such high-mass galaxies will be reduced. We reserve the detailed analysis on this topic for future work.

3.6 Conclusions

In this work, we analyze the redshift evolution of the close-pair fractions $f_{\rm pair}$ using the mock light cone data of SantaCruz SAMs during $0.5 < z < 3.0$, while assuming commonly-used pair selection criteria from the literature, and using the intrinsic simulation information of the galaxies (redshift – $z_{\rm sim}$; stellar-mass – $M_{\rm stellar,sim}$. Simultaneously, we also quantify the pair fraction evolution $f_{\rm pair}$ by systematically incorporating realistic CANDELS-like observational effects such as photometric redshifts and their uncertainties, stellar-mass errors, sample completeness, and survey area sizes. We quantify the impact of such observational effects and derive detailed calibration corrections for them. Finally, we also compute latest empirical pair fraction evolution among $\log_{10}(M_{\rm stellarr}/M_{\odot}) \geq 10$ galaxies during $0.5 < z < 3$ using the CANDELS survey, by appropriately applying the calibrations derived in this study to the raw CANDELS pair fraction measurements. The key analysis items and their results are as follows:

- The pair fraction evolution $f_{\rm pair}(z)$ for massive galaxies with $\log_{10}(M_{\star \rm sam}/M_{\odot}) > 10$ in the SC SAM mocks satisfying a relative velocity selection $\Delta v_{12} \leq 500\,{\rm km\,s^{-1}}$ evolves moderately with redshift from $\sim 3\%$ to $\sim 10\%$, albeit with considerably

flattening behavior at $z \gtrsim 2$. On the other hand, the $f_{\rm pair}(z)$ evolution for a heavier stellar-mass limited sample of $\log_{10}(M_{\star \rm sam}/M_{\odot}) > 10.7$ galaxies is systematically smaller with $5-10$ times lower pair fraction values than the less-massive selection.

- The $f_{\rm pair}$ measurements for a wider annular selection ($R_{\rm proj} \in [5, 50]\,{\rm kpc}$) are systematically higher than a smaller projected distance selection $R_{\rm proj} \in [5, 30]\,{\rm kpc}$, however, they follow a comparable redshift evolutionary trends.

- The pair fraction evolution $f_{\rm pair}(z)$ using two photometric redshift proximity based selections (B09 and S17, see § 3.3.1) and using incorporated photometric redshift errors evolves consistently from $\sim 6-7\%$ (at $z \sim 0.5$) to $\sim 15\%$ (at $z \sim 3$), albeit with a considerable flattened trend during $2 < z < 3$. Furthermore, these photometric redshift selected pairs have large velocity separations ($> 500\,{\rm km\,s^{-1}}$).

- To assess the impact of photometric redshift selection choices on the pair fraction measurements, we quantify the proximity correction factor ($w_{\rm prox}$) as the fraction of photometric redshift proximity selected pairs (for B09 and S17 selection) that satisfy a small (true) velocity separation $\Delta v_{12} \leq 500\,{\rm km\,s^{-1}}$. We find that $w_{\rm prox}$ for both B09 and S17 selections falls from ~ 0.35 (at $z \sim 0.5$) to ~ 0.3 ($z \sim 3$), indicating that only $\sim 30-35\%$ of the photometric proximity pairs have small velocity differences indicative of gravitationally-bound pairs.

- To calibrate for the effect of stellar-mass systematic and random uncertainties on the selection of 'true" mass ratio pairs ($[M_1/M_2]_{\rm sim} \leq 4$), we quantify the major pair correction factor ($w_{\rm maj}$). We find that $w_{\rm maj}$ ranges between $\sim 0.75-0.95$

depending on the redshift bin ($0.5 < z < 3$) and redshift proximity selection cases (Δv_{12} vs. B09 or S17). Our findings indicate that the observed major pair counts require a relative reduction of $5-25\%$ to accurately represent the true major counts.

- We quantify the impact of sample completeness on the $f_{\rm pair}(z)$ measurements by measuring the completeness correction factor ($w_{\rm comp}$) as the ratio of pair fractions with and without imposed completeness function ($f'_{\rm pair}/f_{\rm pair}$). We assess the $w_{\rm comp}$ as a function of redshift and survey completeness limits (H_{80}; see § 3.2.2.3) for redshift proximity selections (Δv_{12}, B09, or S17). We find that a relative pair fraction correction is $10-20\%$ at $z \sim 2-3$ for Δv_{12} based pairs, whereas it is up to $\sim 10-35\%$ for the B09 and S17 selection cases at a typical CANDELS depth configuration (see § 3.3.2.3). Our exercise highlights the compound dependence of redshift proximity selection choices on the $w_{\rm comp}$, and its strong dependence on the survey depth limits (see Figure 32).

- Leveraging the 64 unique CANDELS-UDS sized fields from the larger SC-SAM footprint, we assess the field-to-field pair fraction variation ($\sigma(f_{\rm pair})$; see § 3.3.2.4) for different redshift proximity, stellar-mass, and projected distance choices. We find that the variation of the $\sigma(f_{\rm pair})$ among $\log_{10}(M_{\star,\rm sam}/M_{\odot}) \geq 10$ galaxies and $R_{\rm proj} \in [5, 50]\,{\rm kpc}$ rises moderately with redshift from 0.01 ($z \sim 0.5$) to 0.015 ($z \sim 3$). We also find that $\sigma(f_{\rm pair})$ follows a similar rising trend with redshift for a smaller annular separation of $R_{\rm proj} \in [5, 30]\,{\rm kpc}$, but with $\sim 1.5\times$ smaller values at each redshift bin (see Figure 33). On the other hand, we find that the $\sigma(f_{\rm pair})$ values for a heavier mass selection of $\log_{10}(M_{\star,\rm sam}/M_{\odot}) \geq 10.7$ galaxies stay constant

around ~ 0.005 over $0.5 < z < 3$.

- The field-to-field variation of the $\sigma(f_{\rm pair})$ when using the B09 and S17 redshift proximity selections also follow a moderately rising redshift trend, ranging between $0.01 - 0.02$, and are systematically higher ($\sim 1.5 - 2\times$) than the $\sigma(f_{\rm pair})$ values at a fixed stellar-mass and projected selection case. Our analysis results highlight that the intrinsic pair fraction variation from field to field is also a complex function of the pair selection choices.

- Finally, we use the CANDELS data to identify major close pairs ($M_1/M_2 \leq 4$) among ~ 9500 massive galaxies with $\log_{10}(M_{\star,\rm can}/M_{\odot} \geq 10)$ during $0.5 < z < 3$. We quantify the raw CANDELS close-pair fractions using the B09 photometric redshift proximity selection and then compute the latest, calibrated CANDELS close-pair fractions ($f_{\rm pair,corr}$) after applying the major pair ($w_{\rm maj}$), sample completeness ($w_{\rm comp}$), and field-to-field variation ($\sigma(f_{\rm pair})$) corrections derived using the SC SAM mocks. We find that $f_{\rm pair,corr}$ evolves mildly from $\sim 6\%$ to $\sim 9\%$ during $0.5 < z < 1.5$ and remains considerably flat at $\sim 9\%$ over $1.5 < z < 3$. By additionally applying the proximity correction ($w_{\rm prox}$), we statistically infer the spectroscopic pair fraction evolution (satisfying $\Delta v \leq 500$ km s^{-1}) and find that it remains flat at $\sim 2.5 - 3\%$ over the redshift range $0.5 < z < 3$.

Our comprehensive close-pair analysis in the Santa-Cruz Semi-Analytical Model based light cones provides key insights into the role of observational effects when measuring close-pair statistics, and offers statistical calibration corrections to account for them.

These systematically-derived calibrations will help future close-pair studies to derive robust pair fraction measurements, and will help better understand the role of mergers in various aspects of galaxy evolution. Moreover, the methodological framework of this work will enable future close pair studies to calibrate observational effects at $z > 3$ in preparation for upcoming telescope surveys (e.g., with *James Webb Space Telescope*).

CHAPTER 4

CHARACTERIZATION OF RESIDUAL MORPHOLOGICAL SUBSTRUCTURE USING SUPERVISED AND UNSUPERVISED DEEP LEARNING

Unpublished, Interdisciplinary (CSEE) Research Focused Work

Abstract

As a Deep Learning (DL) based exploratory exercise towards demonstrating the inter-disciplinary focus of this dissertation, we develop supervised CNN and unsupervised variational CAE (CvAE) networks to characterize different kinds of residual substructures hosted by a large sample galaxies from the CANDELS survey. We use the single-Sérsic profile fitting based residual images of 9973 bright and massive galaxies ($H < 24.5\,\mathrm{mag}$ and $M_{\mathrm{stellar}} \geq 10^{9.5}\,M_{\odot}$) spanning $1 < z < 3$ from the CANDELS survey, and their visual-based classification labels (five classes) indicating the nature of residual substructures hosted within them. Using our unique data pre-processing approach, we prepare our residual image data such that the input images to our DL networks only comprise the "galaxy of interest", and augment the data set such that our training sample spans uniformly across different residual characteristics and relative viewing angles. We train a 5-layer deep supervised CNN network and a 3-layered unsupervised CvAE framework on our data and extract the learned deep latent space features for each case. We assess the latent space using Principle Component Analysis (PCA) along with qualitative (i.e.,

visual based) labels and independently quantified metrics of residual strength (significant pixel flux – SPF, Bumpiness – B, and Residual Flux Fraction – RFF). We also employ an unsupervised Gaussian Mixture Modeling (GMM) based clustering and Support Vector Classification (SVC) to identify groupings (and their decision boundaries) in PCA space that correspond to similar residual substructure. We find that our supervised CNN latent features in PCA space correlate with the SPF values and distinguish between qualitatively strong and weak residual substructures. While our unsupervised CvAE latent space also correlates with visual and quantitative residual characteristics, it lacks clear discriminatory power (compared to supervised CNN) when characterizing different residual substructures.

4.1 Introduction

Major galaxy merging (mass ratio < 4) is a fundamental aspect of the hierarchical structure-growth scenario of the Universe. As such, it is theoretically predicted to contribute to the empirically well-documented stellar-mass growth of high-mass galaxies (e.g., $M_{\rm stellar} > 10^{11} M_{\odot}$), enhancing global galactic star-formation, and facilitating central super-massive black hole growth and triggering. Observationally identifying major mergers is a key methodological step in empirically verifying the "merging – galaxy evolution" connection. Commonly, *close-pair* and *morphological* methods are used to identify major mergers, among which quantitative metrics (e.g., CAS, $G - M_{20}$ parameters) and qualitative visual classification of galaxy images or light-profile subtracted residual images (e.g., from `GALFIT`) are frequently-employed approaches to assess the presence

of disturbed morphological substructure. While these methods yield broadly-consistent, rising merger frequency evolution measurements out to $z \sim 2$, each approach selects from a distribution of merger stages (e.g., pre-coalescence vs. late-stage), and suffers from systematic selection biases caused by cosmological surface-brightness dimming, different observability timescales, and subjective interpretation of low surface-brightness signatures. Advancements in machine learning has enabled recent efforts to overcome some of these biases through supervised frameworks and unsupervised deep learning (DL) networks to better identify mergers out to $z \sim 3$ and accelerate visual classifications. In this study, we develop two new public unsupervised and supervised DL frameworks and demonstrate their applicability on `GALFIT`-based residual images of $\sim 10,000$ massive galaxies spanning $1 < z < 3$ to perform automated residual substructure characterization and accelerate visual classification of especially interpretive classes (e.g., plausible hallmark tidal features) that impose subjective assessment of physical processes. The machine learning tools developed in this work are timely in the era of big-data astrophysics and applicability to large-scale surveys by upcoming telescope facilities.

Galaxy-galaxy mergers are theoretically predicted to happen as a natural consequence of the hierarchical, gravitation-driven merging of dark-matter halos. Numerical simulations predict galaxy merging as an important pathway for several key aspects of galaxy evolution such as stellar-mass growth and buildup of massive galaxies [28,31,32], enhancement of star-formation [33–35], growth and triggering of central super-massive

black hole [13, 37–39]. Indeed, many observational studies on these fronts have found evidence for a strong "merger-galaxy evolution" connection supporting the theoretical predictions [7, 43, 44, 47, 48]. However, some studies have also found contradictory results that contest the hypothesized strong role of mergers in galaxy evolution [50, 51, 54, 56], hinting at alternate physical processes that may be at simultaneous play (e.g., Violent Disk Instabilities and Cold-flow accretion; [57, 58, 221])). This highlights that the role of major mergers in galaxy evolution is a key open question, and robustly identifying major merging systems and quantifying their incidence is a key methodological step.

Motivated by the long-standing numerical expectation based idea that galaxies in close physical proximity will interact gravitationally and merge into a more-massive system, several studies over the past two decades have used *close-proximity pairs* as probes for ongoing or future merging systems and quantified their incidence over a wide redshift range $0 \lesssim z \lesssim 6$. These broadly agree that close pair frequency rises between the redshifts $0 \lesssim z \lesssim 1.5$, albeit with a wide range of redshift dependencies $\sim (1+z)^{0.5-3}$ [3,4,16,17,61,109,140,144,146,205], owing to the differences in employed pair selection choices. In Chapter I (also [140]), we quantified the effects of different close-pair selection choices on the derived pair fractions. However, despite using closely matched pair selection criteria, recent measurements of $f_{\rm pair}(z)$ evolution at $z > 1.5$ vary from moderately increasing to flat, or diminishing trends [16, 17, 67, 140, 147, 205]. In Chapter III, we quantified the impact of important study-to-study variant observational effects such as photometric redshifts and their errors, stellar-mass uncertainties, sample incompleteness, and intrinsic field-to-field pair fraction variations, and found that they can contribute to

the observed differences during $z \gtrsim 1.5$.

Simultaneously, several studies have also used *morphological methods* to identify merging systems, motivated by the theoretical merger simulation predictions that merging galaxies exhibit morphological disturbances induced by gravitational tidal forces during the interaction. Most morphology-based studies often adopt subjective visual identification of disturbed morphology [42, 149, 150] or quantitative metrics of large-scale galaxy morphology such as Gini-M_{20} [9, 10, 154], CAS [75, 76, 78, 151, 152] parameters. These morphology based studies find a wide range of (often strongly evolving) merger frequency evolutionary trends $\propto (1+z)^{2-5}$ at $z \lesssim 1.5$ [18], albeit with some finding no redshift dependence [9, 79], and some noting $\sim 25-50\%$ merger fractions at $z \sim 2-3$ [116]. Such variations are due to morphological-study specific systematic effects such as small number statistics owing to narrow pencil-beam surveys, redshift dependent cosmological surface brightness dimming, morphological *k*-corrections when using bluer rest-frame wavelengths, that can impact the completeness and purity of the selected merger samples. Some recent studies have used latest *Hubble Space Telescope* (*HST*) based data (e.g., CANDELS survey [20, 21]) in conjunction with more focused indicators of mergers (e.g., multiple nuclei in close proximity, faint and transient tidal features; [155, 222]) as a means to evade from some of such systematics.

Tidal features are theoretically predicted to be prevalent among merging galaxies [148, 156, 157, 192], and previous studies have identified galaxies hosting merging

activity by detecting tidal features using qualitative visual inspections [42, 43] and quantitative residual metrics [159, 223] of the residual images produced by parametric light-profile fitting software (e.g., GALFIT; [158]). However, these studies often rely on the subjective visual interpretations of human classifiers that can be time intensive and non-repeatable (especially for large-scale surveys), or use quantitative metrics that only indicate the plausibility of tidal signatures and do not quantitatively capture their key properties (strength, shape, color, etc). In an attempt to overcome some of these pitfalls, in Chapter II ([222]), we developed a more direct quantitative way to extract and quantify the strengths of residual substructure, including tidal features hosted by CANDELS galaxies. However, the sample used to demonstrate our feature extraction tool is still based on select galaxies identified through visual inspection. This motivates the strong need to develop automated methods for characterizing residual substructures and providing an accelerated path towards performing a more focused visual/quantitative assessment of galaxies hosting highly interpretive signatures (e.g., tidal features). One methodological avenue that has seen recent booming advancements in the field of image-based automated characterization is *Deep Learning*.

Deep learning (DL) conceptually allows the user to automatically extract low-level pixel features (often referred to as latent space) by iteratively adjusting a series of tuneable parameters within the "deep" layers such that the DL network attempts to achieve an overall global (user-defined) goal. This automated extraction of latent space can be broadly categorized into these approaches – user-guided (i.e., supervised) learning, unguided (unsupervised) learning, and sparsely-guided (semi-supervised) learning. The

extracted latent space can be used for further regression/classification tasks or for dimensionality reduction purposes. Many recent studies have adopted DL frameworks for various astronomical applications such as galaxy morphology classifications [176, 224–226], gravitational lens finding [163, 227, 228], galaxy merger identifications [229, 230]. Several of these morphological or merger related DL-based studies often train Convolutional Nueral Networks (CNNs; for supervised learning; [231, 232]) or Convolutional Autoencoders (CAE; for unsupervised learning; [233]) on the single or multi-wavelength band images of galaxies from the telescopic imaging. However, such an approach may not be ideal in the case of galaxy merger identifications where their signatures are faint and are sub-dominant when compared to the global galactic light profile. As such, devising DL networks to train on the residual images (after subtracting the galactic light profile) may be much more fruitful, yet this approach is relatively unexplored. Only recently, [234] developed Generative Adversarial Networks to model multi-band galaxy images and used the resultant model-subtracted images to identify "anomalous" galaxies. Moreover, it is a common practice among large astronomy survey science communities to run light-profile fitting software (e.g., GALFIT; [158]) and generate residual images for a large galaxy samples. Therefore, analyzing residual images using DL frameworks remains a worthwhile exploratory path, which can provide provide insightful information and automated assessment of the morphological substructure hosted by large galaxy samples.

In this DL-based exploratory study, we develop supervised CNN and unsupervised variational CAE (CvAE) networks to characterize different kinds of residual substructures

hosted by a large sample of $\sim 10,000$ massive and bright galaxies from the *HST* CANDELS survey spanning $1 < z < 3$. Our analysis focuses primarily on these networks' deep latent space extraction and we use Principle Component Analysis (PCA) to assess the extracted latent space of our galaxies using qualitative (i.e., visual based labels) and quantitative metrics. We also explore unsupervised clustering in dimensionality-reduced PCA space and assess the supervised and unsupervised DL networks' ability to automatically distinguish different residual substructures at a latent space level.

We structure this chapter as follows – In § 4.2, we describe our main CANDELS massive and bright galaxy sample selection, its corresponding imaging data products, and visual characteristic catalogs of residual images used for training and assessment of our DL models. In § 4.3, we describe our training and testing sample construction, data augmentation steps, imaging data preparatory steps involving creation of scaled "object-only" residual images that serve as inputs to our DL frameworks, and computation of independent residual quantitative metrics used for latent space assessments. In § 4.4, we introduce our supervised and unsupervised architectures and discuss their training and optimization strategy. In § 4.5, we analyze the deep latent space from our supervised and unsupervised frameworks using PCA, followed by unsupervised clustering in PCA space, and assess of our networks' abilities to characterize different residual substructure features. Finally, we present our concluding statements in § 4.6.

4.2 CANDELS Data

To implement our supervised and unsupervised DL frameworks for an automated

residual substructure characterization, we use *HST*/WFC3 *H*-band (F160W) images and *GALFIT*-based single-Sérsic model-subtracted residual images by [160] for a sample of $10,046$ massive galaxies ($M_{\rm stellar} \geq 10^{9.5} M_{\odot}$) spanning the key epoch of galaxy evolution $1 < z < 3$. For training our supervised model and independently assessing the learning of our unsupervised network, we also use residual characteristic flags based on visual inspection from an ongoing effort (*HST*-AR 15040; PI: McIntosh).

4.2.1 Source Catalogs and Sample Selection

To select our main galaxy sample used in this work, we use the source catalogs and high-level science data products from the CANDELS survey [20, 21], which spans a total sky area of $\sim 800\,{\rm arcmin}^2$ over five *HST* legacy fields – UDS, GOODS-S, GOODS-N, COSMOS, and EGS. We use the standard source extraction [166] based photometric source catalogs generated using the *HST H*-band (2-orbit) imaging by [86–90] for the five fields, respectively. Each identified source in these catalogs are provided with a quality flag (PhotFlag) following a robust automated routine by [86] to identify contaminated sources due to their proximity to nearby stars or image edges. We use a PhotFlag$= 0$ to exclude such contaminant sources ($\sim 3-5\%$ of the overall identified sources). To ensure that our sources have robust structural parameter fits and reliable visual characterization of residual substructure (as discussed later in this section), we enforce a *H*-band magnitude cut and select sources that are brighter than $H = 24.5\,{\rm mag}$.

For these *H*-band bright sources, we then use the latest best-available redshift ($z_{\rm best}$) information by [114], who used bayesian methods to combine the photometric

redshift probability distributions from several participants in a previous CANDELS-team effort [93], and provided latest photometric redshifts and their uncertainties while giving precedence to the relatively more precise *HST* Grism or spectroscopic redshifts wherever available. Additionally, we also use the CANDELS-team stellar-mass estimates ($M_{\rm stellar}$) generated by [94, 95], who performed SED-fitting of multi-band photometric data spanning optical to near-IR wavelengths over a range of template-fitting codes and physical parameter assumptions (see [95]). Using the redshift ($z_{\rm best}$) and stellar-mass ($M_{\rm stellar}$) information, we select our main galaxy sample of $10,046$ massive galaxies with $M_{\rm stellar} \geq 10^{9.5}\,M_{\odot}$ and $H < 24.5\,{\rm mag}$, and spanning a redshift range $1 \leq z_{\rm best} \leq 3$.

4.2.2 Imaging Products and GALFIT Residuals

We use the *H*-band structural fitting data by [160], who used a popular light-profile fitting software GALFIT [158] to model the galaxy's structural parameters identified in the CANDELS source catalogs. Briefly, [160] employed the standard procedures outlined in [165] and used the CANDELS *H*-band image mosaics [21] in conjunction with a comprehensive SExtractor [166] routine to identify sources, make postage stamp cutouts, mask unwanted sources[1], compute initial structural parameter guesses, compute a robust sky-background, and fit each source with a single Sérsic profile. For each CANDELS galaxy, [160] generated an image-cube using GALFIT [158], which contains the *original* image used for fitting, the *model* image with the best-fit structural parameter information

[1] [160] masked neighbouring sources $4\,{\rm mag}$ fainter than the primary object, and all remaining sources were simultaneously fit.

stored in its header, and the *residual* (original−model) image. In Figure 37, we show example *H*-band original images of our galaxies along with their corresponding residuals by [160]. In a later section (§ 4.3), we further process each galaxy's residual images to prepare them for our DL-based analysis.

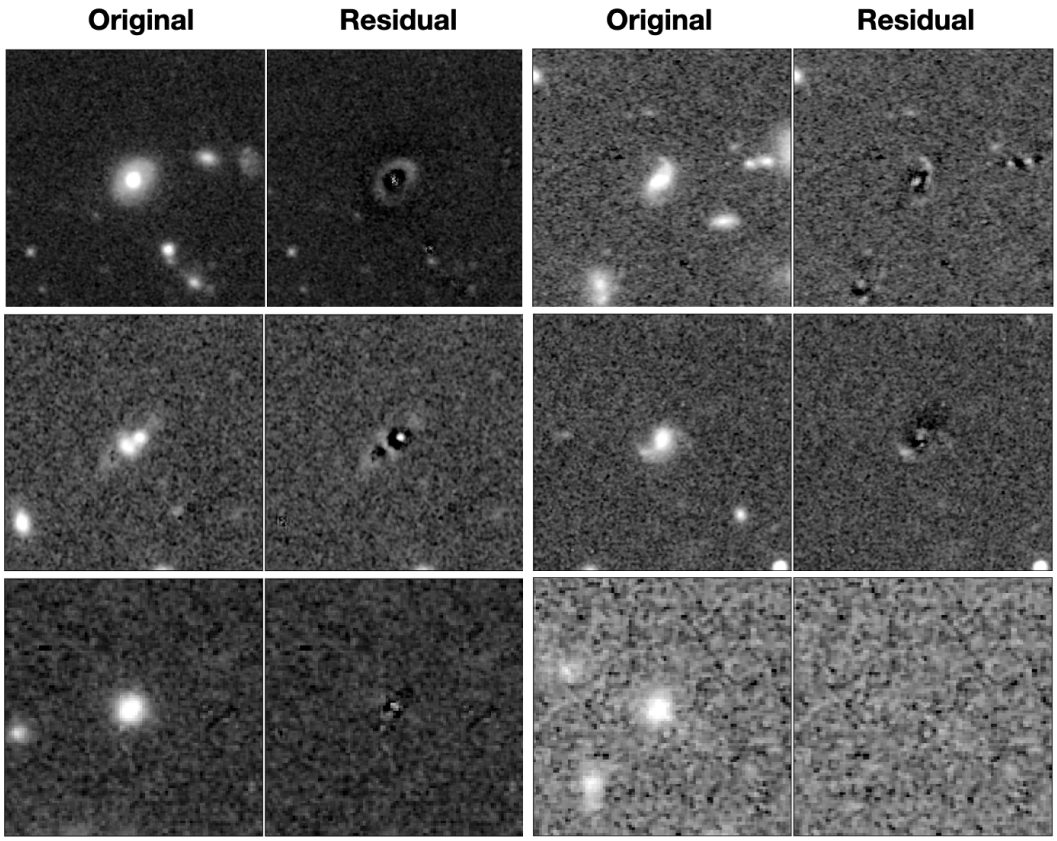

Figure 37: Visualization of the *H*-band original images (left in each column pair) and their corresponding residual images (right panel in each column pair) for example CANDELS galaxies from our main sample, generated by performing single-Sérsic light-profile fitting using `GALFIT` [158] by [160].

4.2.3 Visual Inspection of Residual Substructure

To guide the training of our supervised DL network with user-defined labels that describe the residual substructure and qualitatively assess the outcomes of our unsupervised approach, we construct a catalog of visually-identified residual characteristics hosted for our main galaxy sample. Briefly, we visually inspected all the residual images, following a process in which at least three human classifiers simultaneously inspected each galaxy's original and residual images to characterize the visual nature of the hosted residual substructure. This information is validated by an expert and is further processed using a $2/3^{\rm rd}$ majority voting to generate a catalog of residual characteristics for our galaxy sample. We broadly bin them into five classes that span a range of residual feature characteristics and qualitative strengths, and an additional quality check flag indicating plausible fit-quality issues. In Table 9, we show the definitions used while visually inspecting our sample residual images and we report the percentage breakdown of different residual characteristic groupings. Briefly, the "Clean" residuals have no significant left-over light and are consistent with a background noise appearance. The "General" residuals have strong left-over light, usually corresponding to a disk or spiral shape, or clumpy substructure associated with the overall disk or spiral morphology of the galaxy. Peculiar residuals also have strong under-fit residuals, however, they are usually asymmetric and align with the overall disturbed morphological structure of the galaxy. The "Core" residuals have a bright point-like under-fit structure towards the central region of the galaxy. Finally, "Asymmetric" residuals have concentrated or diffuse non-symmetric signatures with or without the presence of central excess.

Table 9: Breakdown of the five residual characteristic classes based on the visual inspection of our $10,046$ massive galaxy sample from the CANDELS survey. Columns: (1) Name of the residual characteristic; (2) Adopted definition of the residual characteristic for the visual inspection; (3) Number of galaxies in the sample with the corresponding residual characteristic; (4) Percentage contribution of the residual characteristic out of $10,046$ galaxies.

Residual Characteristic (1)	Definition (2)	Sample Size (3)	Percentage (4)
Clean	The residual has no apparent features or structures pertaining to the galaxy and is consistent with the background noise.	3815	38.0%
General	The residual hosts strong left-over light with either a symmetric disk or spiral-like structure, or clumpy underfit substructure associated to underlying plausible disk or spiral galaxy morphology.	608	6.0%
Core	The residual contains one or more bright point-like grouping of pixels near to the central region (by eye) of the galaxy, and no excess away from the galaxy's central region.	1958	19.5%

Asymmetric	The residual contains concentrated or diffuse, asymmetric excess light outside the central region of the galaxy, with or without the presence of central excess light.	3043	30.3%
Peculiar	The residual contains strong asymmetric excess light with or without multiple central point excess that aligns with the overall disturbed morphology of the galaxy.	512	5.1%
Fit-Quality Issue	The galaxy's residual region of interest is very close to the image edges or artefacts, or is dominated/contaminated by a nearby star's light.	110	1.1%

We note that a dominant portion of our main sample are qualitatively Clean and Asymmetric residuals ($\sim 38\%$ and $\sim 30\%$, respectively), and $\sim 19\%$ of our sample hosts cored residuals (see Table 9). A sub-dominant, but notable fraction of our sample ($\sim 11\%$) are within the General and Peculiar classes, hosting qualitatively strong residual characteristics. In Figure 38, we show example *original* and *residual* images of our galaxies per residual characteristic class, where different residual substructures as per the class-wise definitions are visually evident. We note that the visual inspection based residual characteristics information used in this work is preliminary and pending full validation by an expert. While a future fully-validated sample could change the results discussed here, our current analysis and methodological framework are meant to serve as an informative

exercise to aid the full validation of our residual characterization exercise.

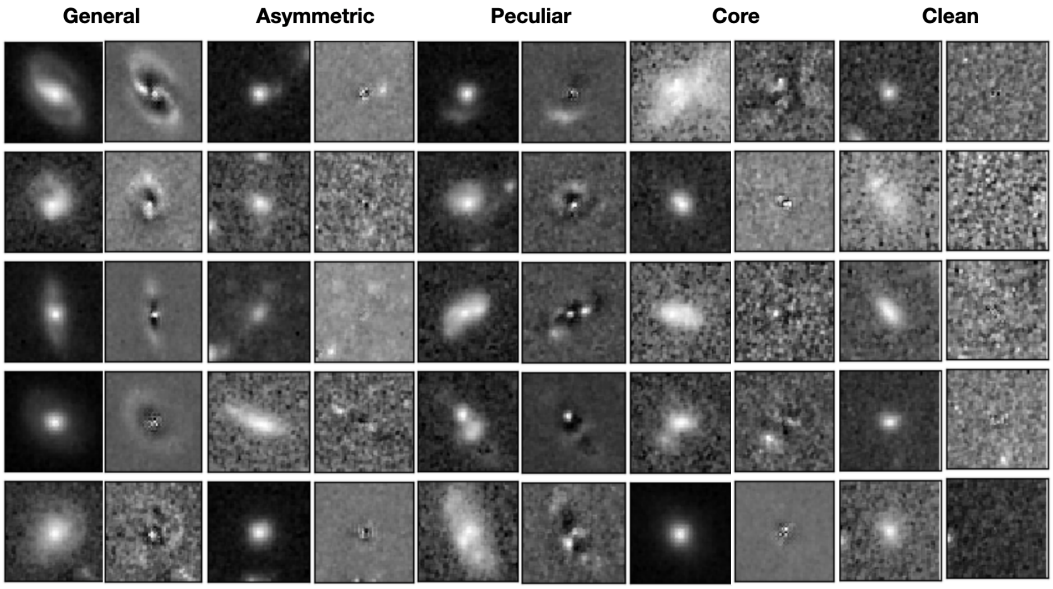

Figure 38: An example view of our galaxies and their hosted residual substructure. In each column pairs, we show *HST* WFC3/F160W *H*-band original images (left) and their corresponding GALFIT-based residual images from [160] (right) for five example galaxies per class.

4.3 Data Preparation

Thus far, we have discussed the selection of our main galaxy sample and their imaging and cataloged data products. To prepare this data for our DL analysis, we carryout the following key pre-processing steps – i) generation of residual image cutouts which only contain information corresponding to the "galaxy-of-interest" (GOI); ii) construction of training and testing sub-samples and generating data augmentation applied training sample; iii) computation of independent residual metrics to aid the assessment of the learning of our DL frameworks.

4.3.1 Creating Object-Only Image for Training and Testing Samples

Each galaxy's image may contain other objects in it's field of view, which are not the main focus of the residual characteristics for the GOI (by design at the center of the image, see [160]). Furthermore, such interlopers may confuse the low-level, feature learning of the DL model. Therefore, we process the residual images such that only the pixel data corresponding to the GOI is preserved and the rest is matched with the sky background.

In Figure 39, we visually demonstrate the steps involved in the generation of object-only images for one example galaxy. We first identify sources using `SEP` [187], which is a python adaptation of the standard source extraction software SExtractor [166]. We adopt the source-extraction settings used in [222], where we use a 2D Tophat convolution kernel (radius$= 5\,\mathrm{pix}$), a minimum source significance of 0.75, and a minimum area of $7\,\mathrm{pix}$. These adopted settings help identify small and faint sources, and enable their segmentation regions to encompass the light reaching far out from the galaxy's center. This enables robust quantification of sky background and masking of close projected-sky proximity interlopers (see [222] for more details). We then use the segmentation region map generated during the source-extraction step (top-center panel in Figure 39) to generate a binary GOI-mask (top-right in Figure 39), where the pixels corresponding to the GOI are set to one, and rest all other interlopers including the sky background are set to zero. Simultaneously, we also generate a generate a binary sky-mask (middle center panel in Figure 39) by making all the source segmentation regions to zero, and the background to one. We then generate a source-masked, background sky-only image

by multiplying the sky-mask with the residual image (center-right panel). We randomly sample 5×5 pixel regions from this background sky image to construct a contiguous background-sky image (bottom-left panel) matching the residual image size, where the native small-scale pixel-to-pixel correlation information is preserved. We then generate the GOI residual image (bottom-center panel) by multiplying the GOI-mask with the residual image data. Finally, we generate an "object-only" residual image (bottom-right panel) by adding the GOI residual image and the GOI-masked background-sky image (background-sky $\times$ GOI-mask^{c}), where GOI-mask^{c} is the complement (inverted) image of the GOI-mask.

Following the standard practice used in DL-based literature (e.g., [224, 227, 230]), we use the cutouts of the object-only images as inputs to our DL frameworks. As will be discussed in § 4.4, our DL-based frameworks are inspired by the shallow network architecture of [235], and therefore we choose $50\,\mathrm{pix} \times 50\,\mathrm{pix}$ as our cutout dimensions centered at the `GALFIT`-based centroid values provided by [160], which is slightly smaller than [235] cutout size choice ($64\,\mathrm{pix} \times 64\,\mathrm{pix}$). Our cutout dimensions correspond to $3\,\mathrm{arcsec}$ on the side ($\sim 26\,\mathrm{kpc}$ at $z \sim 2$) and is at least $\sim 5\times$ the size of $\sim 90\%$ of the galaxies in our sample. We visually randomly inspected our $50\,\mathrm{pix} \times 50\,\mathrm{pix}$ cutout images and found that a majority of them encompass the bulk of the galaxy's residual substructure within them. However, it is worth noting that some galaxy's residuals exceed outside our cutout size, which may impact the learning of the DL frameworks. To test for this, we repeated our analysis using a 64×64 pix input image size and found no significant differences in the results and conclusions discussed in this chapter. Therefore, we resort

to using our initial $50\,\mathrm{pix} \times 50\,\mathrm{pix}$ cutout size choice as it is relatively faster to train our model and offers slightly lower dimensionality exploration space for our network.

One possible alternative to our approach (i.e., using object-only residual image) is to use just the GOI-only residual image (bottom-center panel of Figure 39) as inputs to our networks. However, this may impose unforeseen priors on the model learning, where the network focuses more on the shapes of GOI segmentation regions. To avoid such biases, we adopt a object-only based approach that is free from the GOI segmentation shape information.

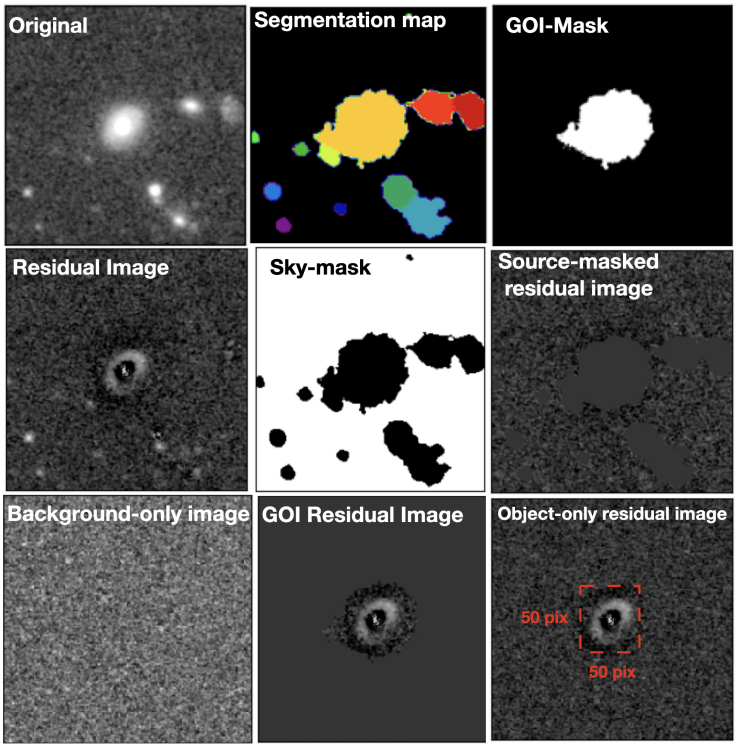

Figure 39: Visual illustration of our preprocessing step to prepare our residual images for the DL framework. We show the original image, source extraction based segmentation map highlighting the detected sources, residual image, region of interest corresponding the central galaxy in the image, and the GOI region of the residual image with a $50\,\mathrm{pix} \times 50\,\mathrm{pix}$ postage stamp size for reference.

Finally, it is a standard practice in the DL-based studies (e.g., [235]) to normalize the input data so that the learning of the deep latent-space vectors are well behaved (i.e., bound and normalized). To conserve the underfit-overfit (i.e., dark-bright structures) appearance of input residual images, we choose to normalize our data to span between $[-1, 1]$ using the linear transformation scheme – `MaxAbs`, where the maximum absolute value of the data sample is set 1. We tested various available non-linear data transformation schemes[2] (e.g., `PowerTransformer`, `QuintileSclaer`, etc) and found that the choice of data scaling has no impact on the results discussed in this work.

4.3.2 Training-Testing Samples and Data Augmentation

We sub-divide our main sample information of $10,046$ galaxies and their corresponding object-only images into training and testing sub-samples using a commonly-employed splitting ratio of $80\% - 20\%$, yielding 7948 training and 1987 testing galaxies. In Figure 40, we show the class-wise distribution of our total parent sample, subdivided into training and testing samples. We also quote the fractional contribution of each class towards the data sets. From this figure, we illustrate the class-wise proportionate sampling of the training and testing data from the main parent sample. However, a class-wise imbalance in the parent sample, where certain classes (e.g., Clean, Asymmetric) dominate in numbers, propagates into the training and testing sample construction. Such an imbalance in the training sample will lead to class-specific biasing of the DL model

[2]see `https://scikit-learn.org/stable/auto_examples/preprocessing/plot_all_scaling.html#sphx-glr-auto-examples-preprocessing-plot-all-scaling-py` for more details

learning. Furthermore, despite our galaxy sample hosting a diversity of the residual substructures (see Figure 38), inclination-induced effects (e.g., similar substructure viewed at different inclinations) can induce a dataset-specific bias and a lack of stochasticity at the deep latent-space level, which may limit their generalized applicability to other data sets. Data augmentation using sample-specific invariant properties (e.g., rotation, brightness/intensity) is a popular choice in the literature to alleviate from such biases (e.g., see [176, 225]). Conceptually, data augmentation provides a robust low-level learning by inducing a data-level stochasticity, by generating additional training samples perturbed randomly by changing one or more of their invariant properties.

Motivated by a similar methodology employed in [235], we adopt the data augmentation step on our training (object-only) image sample such that each image is augmented with a random horizontal flip and random rotation by $45\,\mathrm{deg}$. We generate our "augmented" training image sample such that each residual characteristic class contains 5000 images, which are randomly sampled from a large pool of augmented realizations of our parent sample object-only images. In total, our augmented training sample amounts to 25000 images. In Figure 41, we show an example visualization of this data augmentation step for example residuals from our training sample. In Figure 40, we also show the uniform class-wise distribution of augmented training sample alongside the lopsided, non-augmented raw training set. We use the augmented training images to train both of our supervised and unsupervised DL networks.

It is worth noting that although our augmentation step enables data-level stochasticity, it does not alleviate the underlying human subjectivity of residual characterization,

which is especially relevant to our supervised learning case. Recent advancements in the DL literature has proposed various ways to account for human classifier biases and errors. For e.g., Bayesian convolutional neural networks (e.g., [236]) have the capability to incorporate the biases during learning phase and propagate it through to the output posterior class-probability distributions. It is beyond the scope of this current work to investigate the role of human subjectivity on our analysis results, and we leave this for future work.

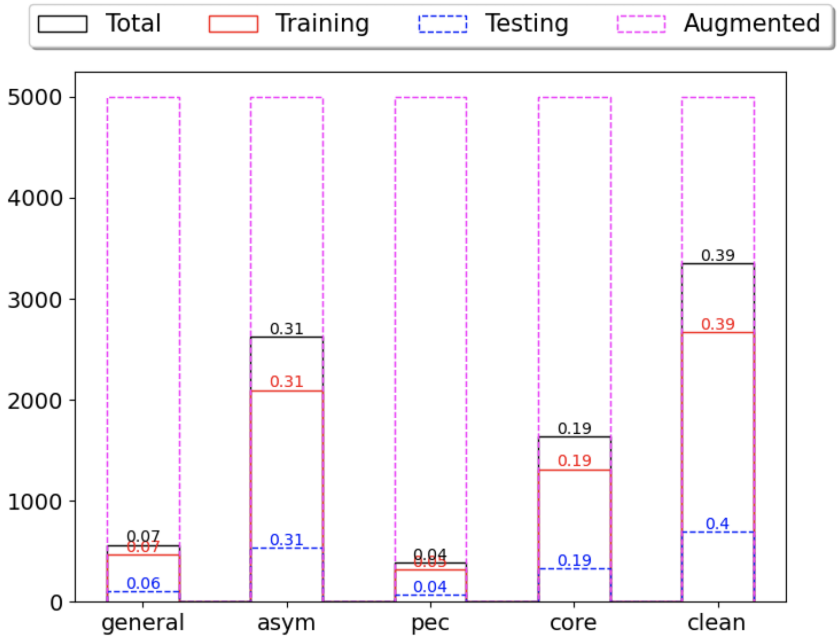

Figure 40: The histograms of our five residual characteristic classes among our parent sample of $\sim 10,000$ galaxies (black), training sub-sample (red), testing sub-sample (blue), and augmented training data sets (magenta). For each class, we also show its fractional contribution to the data set with respective colored text.

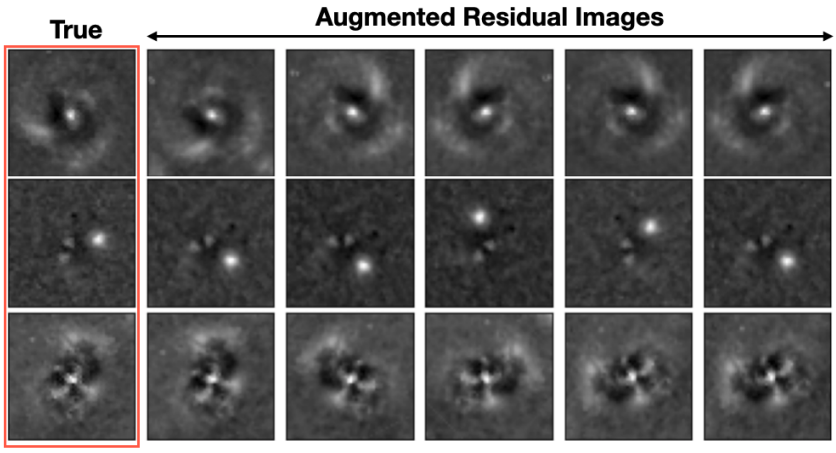

Figure 41: Example visualization of our data augmentation step (see § 4.3.2). For three example residual images from our sample, we show five random horizontal flip and $45\,\mathrm{deg}$ rotation steps.

4.3.3 Residual Quantitative Metrics

In addition to the visual-based residual characterization information, we also derive additional residual quantification metrics: *Significant Pixel Flux* (*SPF*), commonly-used quantities in the literature – *Bumpiness* (*B*; [237]), and *Residual Flux Fraction* (*RFF*; [159]) to assess the extracted latent space of our supervised and unsupervised approaches and independently interpret their learning (see § 4.5). We describe the derivation of these metrics for our parent galaxy sample as the following.

The cumulative flux of significant pixels above the sky background serves as a straight-forward quantitative indicator for the strength of left-over residual light. As such, for each galaxy in our sample, we quantify the *SPF* among the pixels corresponding to the GOI that satisfy $> 3\sigma$ significant above the background as:

$$SPF = \sum_{x,y \in GOI} \mid f_{\mathrm{pix}>3\sigma}(x,y) \mid, \tag{4.1}$$

where $f_{x,y}$ is the flux of the pixel corresponding to pixel coordinate (x, y), and the GOI region is the product of the binary GOI mask (top-right panel in Figure 39) and the residual data. In Figure 42, we show an example visual illustration of the *SPF* calculation. For each galaxy, we identify sources in the original image using a python-based implementation `SEP` [187] of a standard source-extraction software `SExtractor` [166]. We generate a binary sky mask such that all the source segmentation regions are set to zero and the background sky regions are set to unity (top-right panel in Figure 42). We then generate a binary *GOI* mask as described in § 4.3.2 and produce a GOI residual region by multiplying the *GOI* mask with the residual data (bottom-left panel in Figure 42). We extract the distribution of background sky pixels from the background sky image (sky mask $\times$ residual image), and fit it with a Gaussian to derive its upper and lower 3σ bounds. We select the pixels in GOI residual region that have $f(x, y) > 3\sigma$ and $f(x, y) < 3\sigma$, and derive its *SPF* value as the cumulative absolute flux sum of these pixels. For the example shown in Figure 42, the total positive flux of $> 3\sigma$ pixels is $7.56\,\mathrm{e/s}$ (native, exposure time normalized image units) and the total negative flux is $-3.73\,\mathrm{e/s}$, which corresponds to a $SPF = 11.28\,\mathrm{e/s}$.

We also quantify the *Bumpiness* ([237]) parameter, which conceptually quantifies the strength of second-order light-profile moments and deviations of galaxy's structure from a smooth parametric model. It is defined as the ratio of the residual rms to the best-fit sérsic mean model:

$$B = \frac{\sqrt{(1/N)\ \sum_{(x,y)\in GOI}\ [R_s^2(x,y) - \sigma_{\mathrm{sky}}^2]}}{(1/N)\ \sum_{(x,y)\in GOI}\ [S(x,y)]}, \tag{4.2}$$

where R_s is the smoothed residual image after convolution with a 2D BoxCar filter of size

$3 \times 3\,\mathrm{pix}$, $\sigma^2_{\rm sky}$ is the variance of background sky computed using the best-fitted Gaussian to the background-sky distribution in the above SPF calculation steps, S is the best-fit sérsic model generated by [160], and N is the number of pixels corresponding to the GOI segmentation region (i.e., white regions in GOI-mask panel of Figure 39).

Finally, we also compute the *Residual Flux Fraction* (*RFF*; [159]), which conceptually quantifies the additional residual light that cannot be accounted by background noise fluctuations. It is defined as:

$$RFF = \frac{\sum_{x,y \in GOI} R(x,y) - 0.8 \times \sum_{x,y \in GOI} \sigma_{\rm sky}}{\sum_{x,y \in GOI} S(x,y)}, \tag{4.3}$$

where R is the residual image, $0.8 \times \sum \sigma_{\rm sky}$ is the expectation of residual light from Gaussian noise fluctuations, and S is the best-fit light profile model. In § 4.5), we assess the deep latent space features of our trained DL models in conjunction with SPF, B, and RFF metrics to interpret their learning.

4.4 Proposed Deep Learning Models

To better automate the characterization of residual galactic substructures and improve the identification of plausible merging signature hosting galaxies, we develop two frameworks – i) a supervised Convolutional Neural Network (CNN) model; ii) an unsupervised Convolutional variational Autoencoder (CvAE) model – to learn the latent space features describing the variety of residual substructures hosted by our sample of $\sim 10,000$ massive galaxies from the *HST* CANDELS survey. In this section, we describe the architecture of our CNN and CvAE frameworks and discuss their training and optimization strategy.

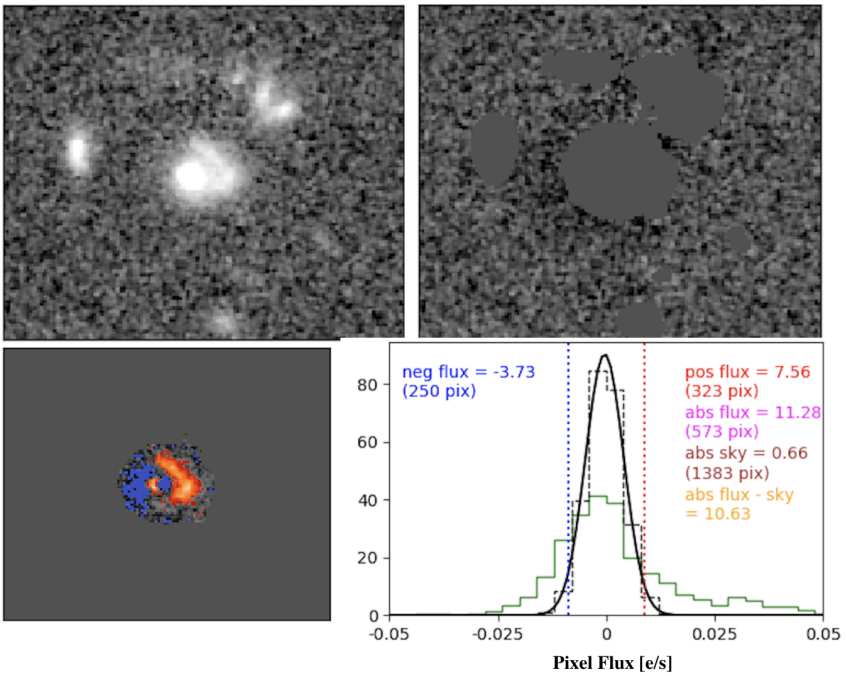

Figure 42: Visual illustration of our Significant Pixel Flux (*SPF*) calculation using an example galaxy image (top-left) from our sample. We show the source-masked image (top-right) and the Region of Interest (*ROI*; bottom-left) highlighting the Galaxy of Interest (*GOI*). We show the histograms of pixel fluxes (bottom-right) in the ROI (green), the sky background distribution (dashed black) fit with a Gaussian (solid black) and their lower and upper 3σ bounds with blue and red vertical lines, respectively. We highlight those pixels that are above and below the 3σ bounds in bottom-left panel with their respective colors. For this example, *SPF*$= 11.28$.

4.4.1 Supervised Framework

The architecture of our supervised framework is inspired from the CNN-based classifier developed by [224]. In Figure 43, we show the visual overview of our supervised CNN-based framework. Briefly, our network takes the "object-only" residual image stamps as single channel inputs (i.e., one object-only image per sample), followed by three sets of Convolutional and MaxPooling layers for feature extraction, and two Dense Fully-Connected layers. Given that our input data is scaled and spans [-1,1], we choose a "*Tanh*" activation function for all the convolutional layers, whose resultant activations also span [-1,1]. For our deep latent space feature layer, we choose to extract 512 vectors, motivated by a separate dimensionality reduction exercise, where we apply a direct PCA decomposition of the object-only training images and find that 512 eigen vectors hold $\gtrsim 90\%$ of the explained sample variance (i.e., $> 90\%$ of information is captured by reduced, 512 dimensions). We choose *Tanh* activation for this Dense layer and a softmax activation for the final classification layer (of size 5) such that the output is probabilistic prediction per target class (here visual residual class labels) that sums up to unity. Following the standard practices to avoid over-fitting scenarios, we introduce a 50% dropout layer after each Convolutional+MaxPooling block, which conceptually omits (randomly) half of the layer features from propagating further. We experimented with different dropout percentage values and found that using lower values (e.g., $\sim 5 - 10\%$) resulted in worsened model performance with reduced testing set accuracy due to overfitting. Furthermore, we also introduce a Gaussian Noise layer with a standard deviation $\sigma = 0.02$ at the entrance of the network to facilitate learning of latent space that is less susceptible to image noise

patterns, and ensures that the network never sees the exact same image during model training. Using a higher value for the entrance random noise layer (e.g., $\sigma \sim 0.1$) induces too much stochasticity into the network and leads to an oscillating model convergence (i.e., cyclic learning-unlearning patterns).

To optimize our supervised CNN model during the training, we use *categorical cross-entropy* loss function with *Adam* optimizer [238] with starting learning rate of 10^{-4}. We train our network with the augmented training sample of 25000 images for a maximum of 100 epochs in batch sizes for weight updation of 32 images, and decay the learning rate by $2\times$ for every $1/3^{\text{rd}}$ of the total steps to aid the model convergence. Our model achieved an accuracy of $\sim 95\%$ after 75 training epochs and it yielded a testing accuracy of $\sim 75\%$. We note that the model achieves an accuracy of $\sim 91\%$ on the Clean class, but the overall (lower) testing accuracy is due to considerable cross-class confusion between the Peculiar, General, and the Asymmetric classes. We acknowledge that this stage needs further improvement and will one of the main items of future work. In the later section, we assess the latent space features learned by our CNN model.

4.4.2 Unsupervised Framework

There are several approaches to learning the deep latent space in an unsupervised fashion, amongst which Autoencoders, specifically Convolutional Autoencoders (CAE; [233]) are a popular option. Briefly, they comprise of an Encoder-Decoder framework, where the objective is to learn the dimensionality-reduced latent space vectors via

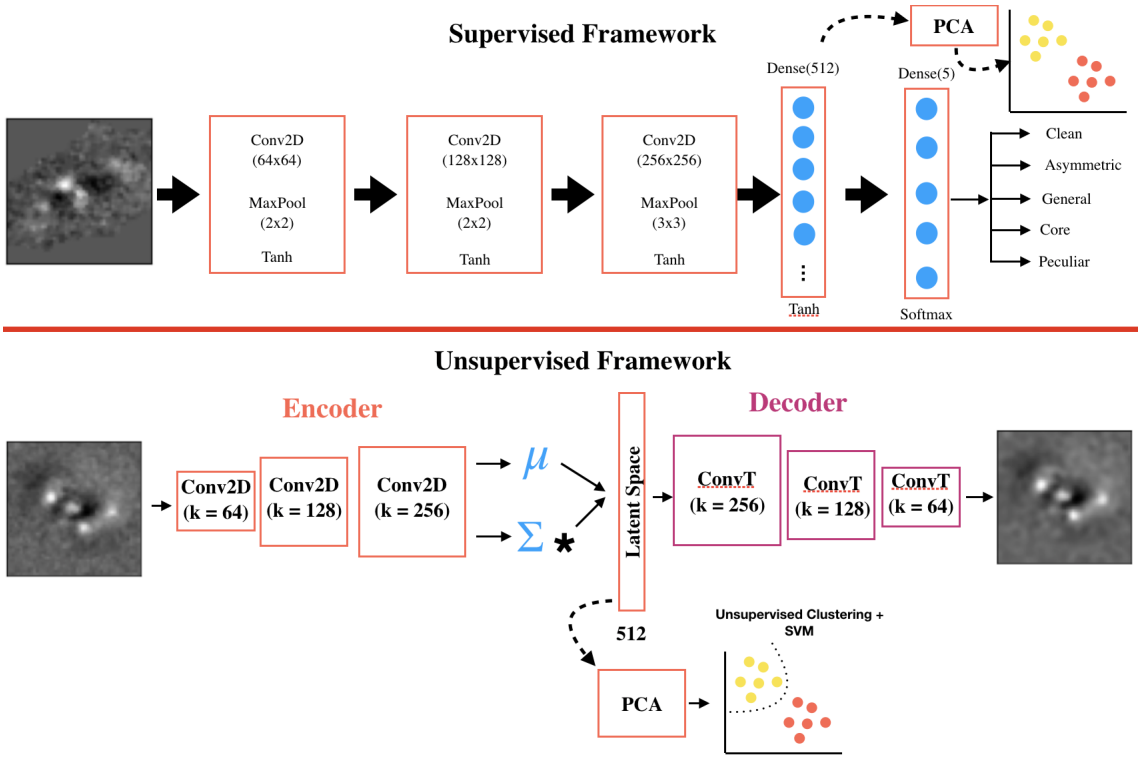

Figure 43: Overview of our supervised CNN (top) and unsupervised Convolutional Variational Autoencoder (bottom) frameworks used to analyze the galactic residual substructures (see § 4.4.1 & § 4.4.2). For both networks, we also visually illustrate our analysis of the latent space features using Principle Component Analysis and unsupervised clustering (see § 4.5).

iterative adjusting of layer weights such that the output reconstruction (by a Decoder network) matches closely to the input image. Motivated by this ideology, we implement an unsupervised Convolutional Autoencoder based framework to learn the deep latent space that characterizes different galactic residual substructures. Specifically, we adopt a Convolutional Variational Autoencoder (CvAE), which is an improvement over the simple CAE, where the network learns a distribution of latent space vectors described by a unit Gaussian distribution instead of simple latent space point vector, which makes the learned

latent space less sensitive to unknown nuisance in the input images. In practice, we noticed that the output reconstruction from the CvAE is qualitatively (i.e., by visually by eye) closer to the input images than CAE counterparts (which are often more "blurry" in appearance).

We model our CvAE framework based on the previously mentioned supervised CNN architecture (see § 4.4.1). In Figure 43 (bottom panel), we show our CvAE architecture, where our Encoder network involves three 2D Convolutional layers, whose end features are flattened and passed to a fully-connected, latent space feature layer of size 512 (following our aforementioned PCA-based suggestions). We then connect this latent feature layer to two more fully-connected layers (each of size 512) to describe the mean (μ_l) and variance (σ_l^2) of the latent space features (l), which are then passed to a custom unit Gaussian Sampling layer. The output of this layer (and the encoder network) is a latent space feature vector of size 512, where each feature is sampled from the Gaussian(μ_l,σ_l^2).

Our decoder network starts with an input layer of size 512 (i.e., matching the Encoder output size), followed by three 2D transpose convolutional layers that mirror the sizes of our encoder networks' convolutional layers. We use *ReLu* activation function for all the layers in the Encoder-Decoder framework, except for the output layer of the decoder network where we use the *Tanh* activation. Finally, similar to our CNN framework, we also introduce a Gaussian Noise layer ($\sigma = 0.02$) at the start of the encoder framework to enable latent-space features that are less susceptible to image noise patters. It is worth noting that we do not use feature dropout in between the convolutional layers (unlike in our CNN framework), as in doing so, the output reconstructions qualitatively deviated

considerably from their input counterparts, where dropping features resulted in blurrier reconstructions.

To optimize our network during the training, we jointly minimize the input image *reconstruction loss* and the *Kullback-Leibler* (*KL*) divergence loss of the latent space feature distribution. Conceptually, our optimization process corresponds to requiring the reconstructed outputs to be closer to the input images while simultaneously ensuring that their underlying latent space features (describing the input-output images) are also closely matched. Our *total loss* function, which is the summation of the image reconstruction loss and the *KL*-divergence loss is as follows:

$$\mathcal{L}_{\rm total} = \mathcal{L}_{\rm recon} + \mathcal{L}_{\rm KL}, \tag{4.4}$$

where we choose $\mathcal{L}_{\rm recon}$ as the mean-squared error (*MSE*) between the input (I) and the reconstructed image ($\hat{I}$) as:

$$\mathcal{L}_{\rm recon} = \frac{1}{N}\sum_{i,j}(I - \hat{I})^2. \tag{4.5}$$

The *KL* divergence loss of the latent space vectors (l) with mean (μ_l) and variance (σ_l^2) is given by:

$$-0.5 \times [1 + \log_{10}(\sigma_l^2) - \mu_l^2 - \exp\{\log_{10}(\sigma_l^2)\}] \, . \tag{4.6}$$

To optimize our CvAE network during the training, we adopt the same strategy as in our supervised approach (see § 4.4.1) with the *Adam* optimizer and an initial learning rate of 10^{-4} with a decay factor of two every $1/3^{\rm rd}$ of the steps over a maximum of 100 training epochs. In Figure 44, we show the input-reconstructed image pairs of example galaxy

residuals from the testing data. We notice that the overall visual structure of reconstructed residuals is preserved, albeit with a globally "smoothed" appearance.

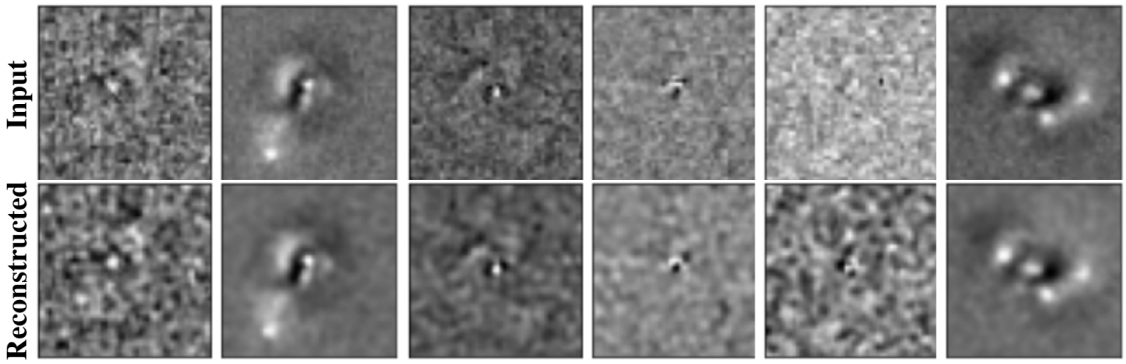

Figure 44: Example visualization of our unsupervised Convolutional Variational Autoencoder network's input (top) and reconstruction (bottom) images.

4.5 Discussion: Assessment of Supervised and Unsupervised Latent Space using PCA and Unsupervised Clustering

Thus far, we have developed two DL frameworks, a supervised CNN and an unsupervised CvAE network, and trained them on our augmented training image sample of residual images to learn the latent space representations of different residual substructure characteristics spanning in our data set. In this section, we assess the supervised and unsupervised latent space features with Principle Component Analysis (PCA; [239,240]) along with supplemental, independently derived quantitative metrics (*SPF*, *B*, and *RFF*). We also explore unsupervised clustering algorithms and Support-Vector Machines [241, 242] to identify classification boundaries in the PCA space that can distinguish different residual structures.

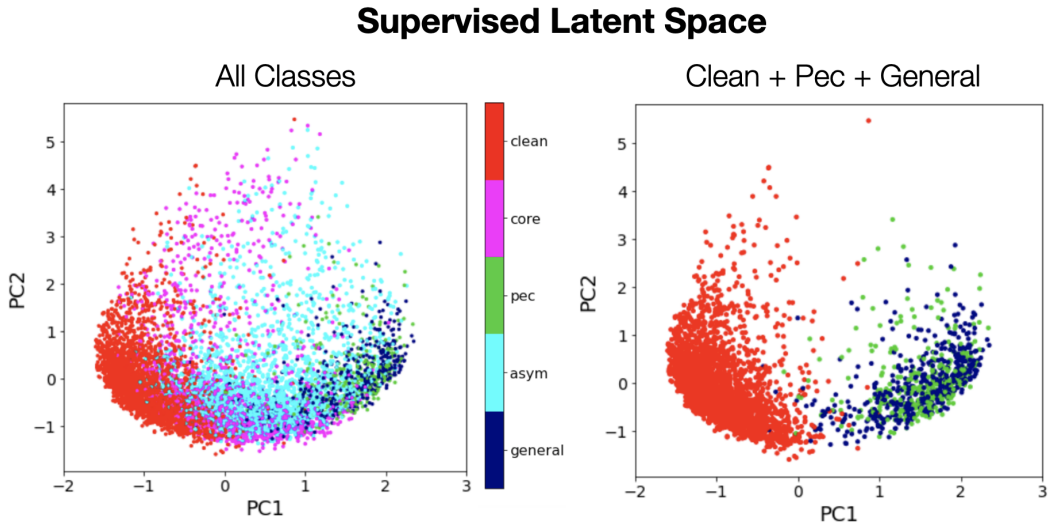

Figure 45: Visualization of principle component axes of the supervised latent space color coded with five residual class labels (left panel). In the right panel, we show the PC1 vs. PC2 only for the Clean, Peculiar, and General Classes.

4.5.1 PCA Analysis of Supervised CNN Latent Space

Using our trained CNN network, we extract the latent space features representing the *unaugmented* training image set by running them through the network (with frozen, trained weights) and extracting information from the fully-connected layer penultimate to the classification output (see Figure 43). We then run PCA on these latent features using the python-based implementation `scikit-learn` [243] and extract three top-most information bearing eigen components capturing $\gtrsim 90\%$ of the explained variance. In Figure 45 (left panel), we show the training sample projected in the first two principle component axes (PC1 vs. PC2) and color code their corresponding visual residual class labels. We notice that the Clean vs. Peculiar and General classes appear distinctly separate (by eye) in the PCA space, where the bulk of Clean class objects are clustered

alongside the "left-arm" and the Peculiar and General classes are along the "right-arm" in the apparent "U" shape of the PC1 vs. PC2 distribution. This distinction can be clearly seen in the right panel of Figure 45, where we show the PC1 vs. PC2 distribution for peculiar and general classes (i.e., galaxies hosting strong residuals) as these objects are of particular scientific interest, in the context of merging induced tidal features. On the other hand, we notice that the core and asymmetric classes lie in the "saddle" region of the PC1 vs. PC2 distribution ($PC1 \in \sim [0, 1]$), and are considerably overlapping with the peculiar and general classes (at $PC1 \gtrsim 1$). Our results based on Figure 45 conceptually illustrate that the CNN learns a higher-order measure of residual strength (especially along PC1) and uses it as a primary leverage to characterize residual substructures into strong, intermediate, and clean groupings. This is intuitive as the CNN's learning mimics that of a human's interpretation of the residual strengths when characterizing them.

To test our above interpretation that the PC1 vs. PC2 may be correlated with a measure of quantitative residual strength, we assess the learned latent space with independently quantified measures of residual strength – SPF, B, and RFF. In Figure 46, we show the PC1 vs. PC2 distribution for the Clean, Peculiar, and General residual classes along with their quantitative metrics SPF, B, and RFF. We find that all the three quantities correlate distinctly with the with the PC1 and weakly with the PC2 axis. More notably, we find a strong bimodal correlation of PC1 and PC2 with the SPF, where the objects with $PC \gtrsim 0.5$ have higher SPF with $\sim 10 - 100\times$ than the objects at $PC1 \lesssim 0$ ($\log_{10}(SPF) \sim -0.5 \,\mathrm{to}\, -1.5$). We also note a relatively weaker, but noticeable correlation with the Bumpiness and Residual Flux Fractions, where the $PC \gtrsim 0.5$ have higher

B and RFF values than the objects spanning $PC \lesssim 0$. Our observations based on Figure 46 conceptually suggest that the latent space learned by our CNN network physically corresponds to a combination of different residual strength metrics.

It is worth noting that the training sample distribution along the third principle component (PC3) is nearly identical to the PC2 distribution. As a result, our analysis of the above trends in PC1 vs. PC3 space are very similar to the PC1 vs. PC2 based conclusions. Therefore, we just focus on the PC1 vs. PC2 space to make our main points.

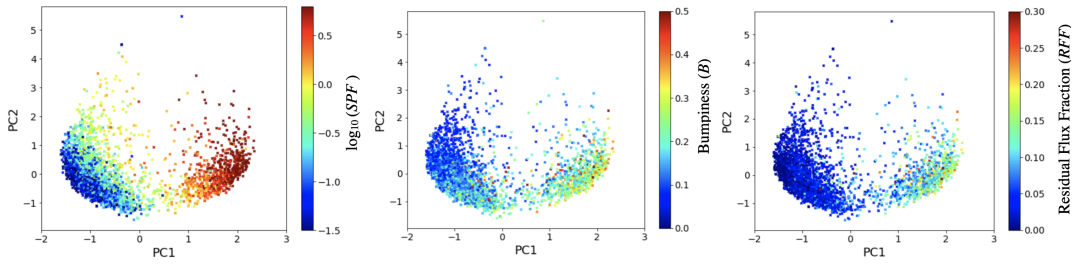

Figure 46: The supervised CNN based latent space visualized in the PCA eigen axes for a subset of classes (Clean, Peculiar, and General; shown in Figure 45) and color coded by different residual quantitative metrics – $\log_{10}(SPF)$ (left panel), Bumpiness (B; middle panel), and Residual Flux Fraction (RFF; right panel).

4.5.2 PCA Analysis of Unsupervised CvAE Latent Space

Using our encoder frontend of the trained unsupervised CvAE network, we extract the latent space features corresponding to our unaugmented training image sample. Analogous to our supervised case, we then perform PCA analysis on these latent vectors and analyze trends in the first two principle components PC1 and PC2. In Figure 47 (left panel), we show the PC1 vs. PC2 distributions for all five visual residual classes, and only

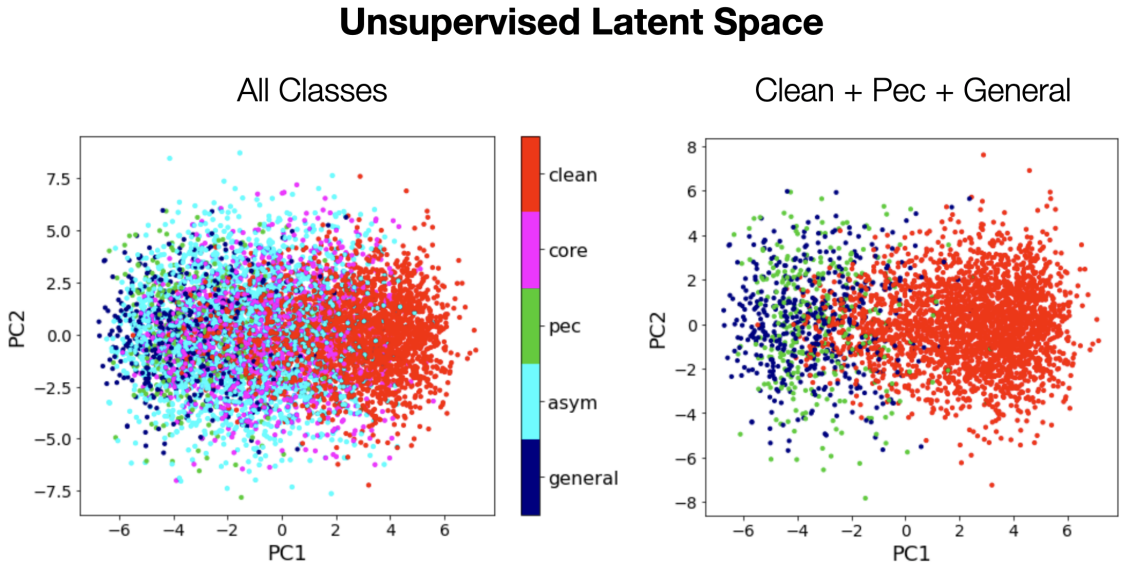

Figure 47: Visualization of the unsupervised CvAE latent space in the PCA eigen axes (PC1 vs. PC2), where each data point is color coded by its visual residual class label (left panel). In the right panel, we only show the Clean, Peculiar, and General classes.

focus on the Clean, Peculiar, and General classes in its right panel. Immediately, we notice that the overall structure of the PC1 vs. PC2 distribution is less distinctive and more contiguous than our supervised case, where the Clean, Core, and Asymmetric classes follow a continuum with no distinguishing correlation across PC1 and PC2 axes. These results conceptually suggest that although our unsupervised CvAE framework learns a residual strength equivalent measure, it lacks the additional discriminatory information from the visual-based residual characterizations. Nevertheless, it is worth noting that the distribution of General and Peculiar classes along the PC1 axis fall towards the lower tail-end ($PC \lesssim -2$) of the gaussian-like distribution of the Clean class objects.

Similar to our supervised latent space exercise discussed in § 4.5.1, we also assess the unsupervised latent features in PC1 vs. PC2 with the residual quantitative metrics

(SPF, B, and RFF). In Figure 48, we show the PC1 vs. PC2 for our Clean, Peculiar, and General Classes and color code each data point based on their SPF, B, and RFF values. We find a bimodality in the SPF values along the PC1 axis, where $PC1 \gtrsim 2$ ($PC1 \lesssim -2$) have smaller (larger) SPF values, where there is continuum of objects with intermediate SPF value during $-2 \lesssim PC1 \lesssim 2$. Similarly, we notice a weak correlation between the B and RFF quantities and $PC1$, where smaller $PC1$ values correspond to higher B and RFF measures. It is worth mentioning that there is no distinct correlation between the three metrics and the $PC2$ axis. Our exercise illustrates that the unsupervised CvAE latent space is only somewhat informative (relative to our supervised case) in terms of distinguishing different residual characteristics, where only one informative eigen axis ($PC1$) correlates with (independent) physical measures of residual strength.

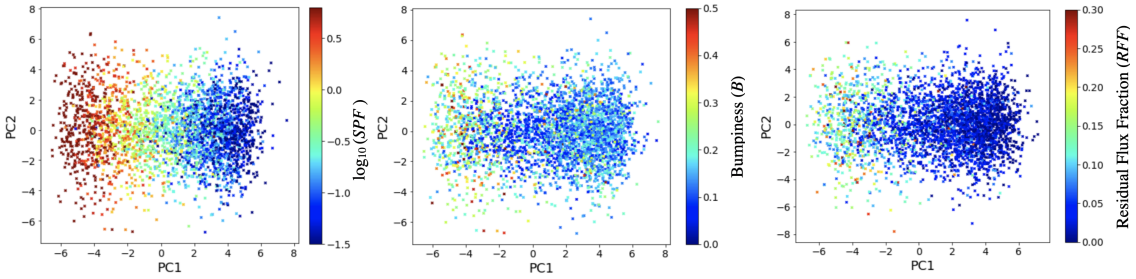

Figure 48: Visualization of the unsupervised CvAE latent space in PCA eigen axes along with residual quantitative metrics SPF, B, and RFF following a similar layout as in Figure 46.

4.5.3 Unsupervised Clustering of the CNN and CvAE Latent Features in PCA Space

To assess the discriminatory power of our supervised and unsupervised latent space features to naturally distinguish between different residual characteristics, we investigate unsupervised clustering in their corresponding PCA space along with Support

Vector Classifiers (SVC; [244]). There are various popular unsupervised clustering techniques to identify data-driven natural clusters in N-dimensional parameter space such as *k*-means clustering [245, 246], Spectral Clustering [247], Density based spatial clustering (DBscan; [248]), and Gaussian Mixture Modeling (GMM, [249]). For our exercise, we choose the posterior probability based cluster assignment approach – GMM, informed by a preliminary k-means clustering assessment of the data distribution in PC1 vs. PC2 space.

First, we repeatedly run k-means clustering on the principle components of the supervised CNN latent features with different number of permissible clusters ($k \in [1, 15]$) and compute the "inertia" as the sum of squared distances of the data points to their nearest cluster centers. In Figure 49 (left panel), we show the k-means inertia for different number of clusters, which falls sharply up to an optimal value of $k = 6$ clusters and shifts to a more gradual fall off. This test to find data-driven optimal number of clusters is dubbed the "Elbow" method. Using this information, we run the unsupervised GMM clustering on our CNN latent features in PC1 vs. PC2 space and identify 6 clusters.

In Figure 49 (right panel), we show the PC1 vs. PC2 space with data points color coded based on their GMM identified cluster labels ($n_{\rm gmm,cnn}$). We notice that the clusters $n_{\rm gmm,cnn} = 2, 4$ have been naturally identified, which quantitatively correspond to residuals with high and low SPF values, respectively. We also find that the clusters $n_{\rm gmm,cnn} = 0, 1, 5$ span the intermediate region between the $n_{\rm gmm,cnn} = 2, 4$ clusters,

which also quantitatively correlate to intermediate SPF values. Finally, our method identified the $n_{\rm gmm,cnn} = 3$ cluster data as "outliers" to the remaining clusters. To visually assess the GMM-based clusters, in Figure 50, we show the residual images of the randomly selected GMM-based cluster members. Interestingly, we notice that $n_{\rm gmm,cnn} = 2$ indeed hosts galaxies with strong and often interesting residual characteristics, and the residuals of the $n_{\rm gmm,cnn} = 4$ objects are nearly "clean" in appearance. On the other hand, we notice that the residuals of objects spanning the clusters $n_{\rm gmm,cnn} = 0, 1, 5$ host weaker (compared to $n_{\rm gmm,cnn} = 2$), diffuse asymmetric signatures. Finally, we find that a majority of the $n_{\rm gmm,cnn} = 3$ residuals host distinct (yet visually-similar) central residual patterns, with no obvious signatures outside the central region. Our exercise suggests that the supervised CNN model learns latent space which can be segregated into qualitatively and quantitatively similar residual characteristics. To automatically identify quantitative boundaries amongst the identified GMM clusters, we use the SVC with a radial-basis function kernel ($\gamma = 0.7$ and $C = 1$). In Figure 49, we show the SVM-based classification boundaries of the GMM clusters. Our GMM-based cluster identities and SVM-based classification boundaries can be further used to refine/validate our visual residual class characteristics used for training purposes.

Following a similar approach to our supervised case, we run an initial k-means inertia test for our unsupervised CvAE latent features in PCA space, and find that it allows an optimal $k = 2$ clusters (see Figure 51. Using this information, we run GMM-based clustering to identify 2 clusters in the CvAE-based PCA space (see right panel in

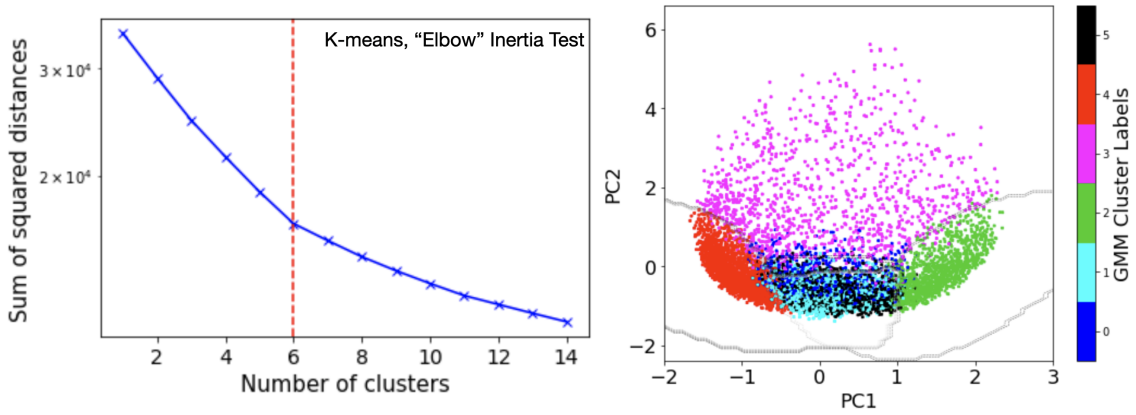

Figure 49: *Left panel*: The sum of squared distances between the data points using k-means clustering technique as a function of number of permissible clusters, indicating a transition from a sharp to gradual fall off at an optimal number of 6 clusters. *Right panel*: We show unsupervised Gaussian Mixture Modeling based clustering of the PCA embedding of the CNN latent features, where the GMM identified clusters are labelled with unique color coding (see color bar).

Figure 51). After running an SVC classifier with a "Linear" kernel (owing to the simplistic nature of the PC1 vs PC2 distribution), we find that these two GMM-based clusters ($n_{\rm gmm,vae}$) are split roughly at $PC1 \sim 0$, where $n_{\rm gmm,vae} = 1$ ($= 0$) cluster quantitatively corresponds to objects with smaller (larger) SPF values, respectively (see Figure 48). In Figure 52, we visually show example residual images randomly sampled per GMM-identified cluster. We find that the $n_{\rm gmm,vae} = 0$ residuals host a mixture of visually strong and intermediate strength residuals, and the $n_{\rm gmm,vae} = 1$ residuals are mostly "clean" with some hosting weak diffuse signatures. This exercise reaffirms our previous discussion that the unsupervised latent features are less informative and lack clear discriminatory power in terms of characterizing residual substructure.

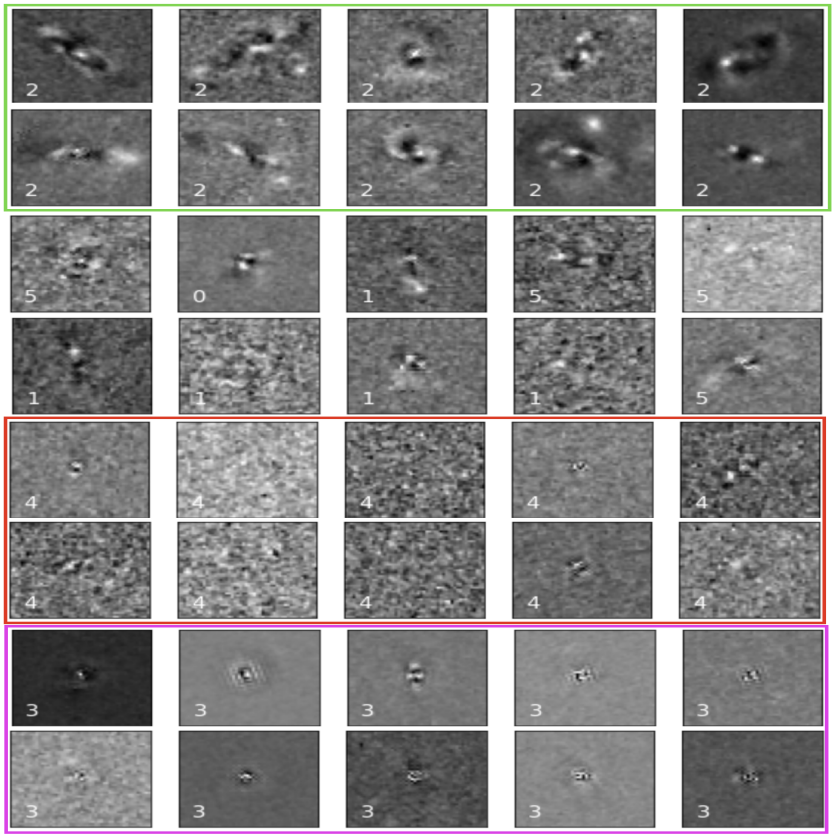

Figure 50: Visualization of the residual images sampled randomly from the GMM-based clusters identified in the supervised CNN-based PCA space (in Figure 49). We outline the row pairs with borders corresponding to the color coding used in Figure 49, except for the second and third rows, which are random samplings from multiple clusters (GMM labels $= 0, 1, 5$). On each residual image, we indicate the GMM cluster label.

4.6 Conclusions

In this chapter, we carryout a Deep Learning (DL) based exploratory study of the residual substructures hosted by a large sample of $\sim 10,000$ massive and bright galaxies from the *Hubble Space Telescope* (*HST*) CANDELS survey spanning a redshift range

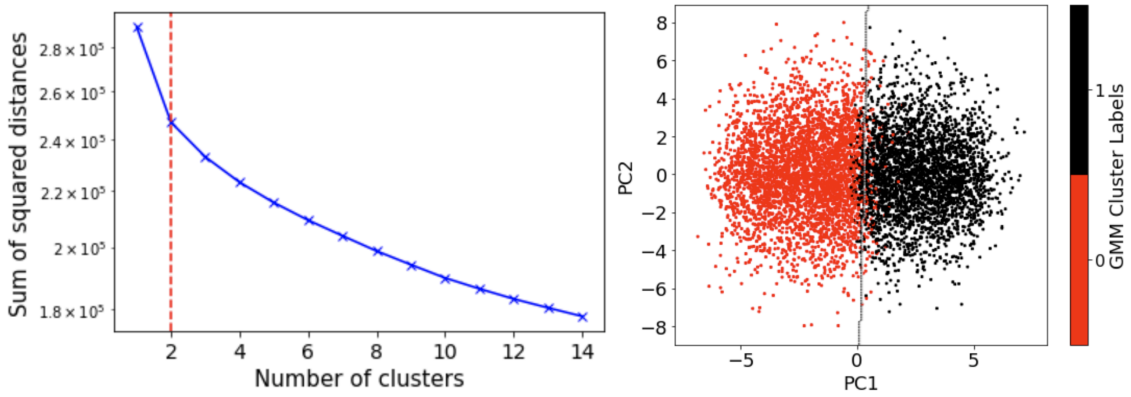

Figure 51: Same as Figure 49, but for our unsupervised CvAE latent features in PCA space.

$1 < z < 3$. We develop supervised Convolutional Neural Network (CNN) and unsupervised Convolutional Variational Autoencoder (CvAE) frameworks to primarily extract the deep latent space features representing the different residual substructure characteristics. We analyze the latent space using Principle Component Analysis in conjunction with independently quantified metrics of residual strength. We assess the DL-based latent features' ability to distinguish different residual characteristics by using Gaussian Mixture Modeling (GMM) based unsupervised clustering and Support Vector Classifiers (SVC). The key methodological steps and their corresponding results are as follows:

- We select a sample of 9973 massive galaxies ($M_{\rm stellar} \geq 10^{9.5}\,M_{\odot}$) spanning $1 < z < 3$ from the CANDELS survey that are also *HST H*-band bright ($H < 24.5\,{\rm mag}$) and have good photometric data (PhotFlag$= 0$). For this sample, we procure the single-Sérsic profile-fitting based residual images generated by [160], carryout a visual-based characterization of different residual features hosted by them, and compile them into five residual classes – Clean, Core, Peculiar, Asymmetric, and

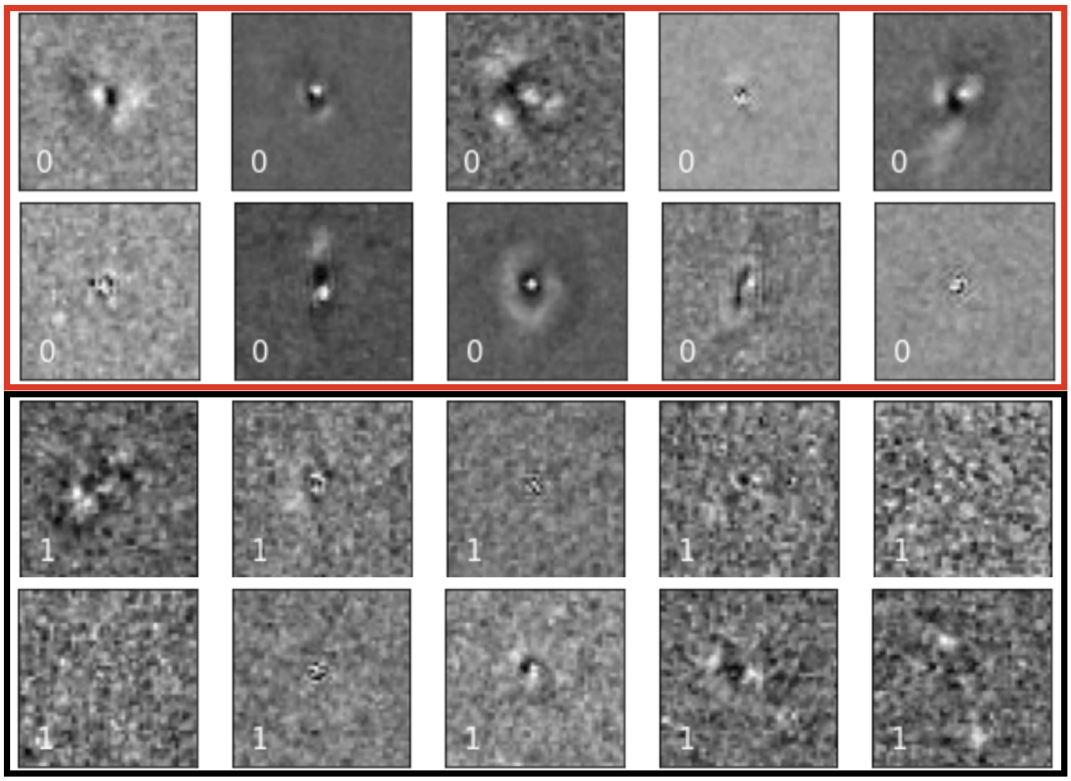

Figure 52: Following the similar layout as in Figure 50, we show the residual images sampled randomly from the GMM-based clusters identified in the unsupervised CvAE based PCA space.

General.

- We prepare our sample's residual images to serve as inputs to our DL networks by pre-processing them such that the resultant image set comprises of postage-stamp cutouts of the raw residual images which have been modified to contain only the "galaxy of interest". We split our main sample into training and testing subsets, and augment our training data to span equi-proportionately across the five residual characteristic classes. Finally, we derive three residual metrics that quantitatively

describe their strengths: Significant Pixel Flux (SPF), Bumpiness (B), and Residual Flux Fraction (RFF), which we use in our analysis to assess the latent space features of our DL networks.

- We implement a 5-layer deep CNN network and train it on the augmented training image sample and visual-based residual class labels as targets. We extract the latent space features from the pre-classification fully-connected layer and perform PCA analysis on it. We assess the latent features in the first two principle component axes and find that the Clean vs. Peculiar and General classes are distinctly separable, while the remaining Core and Asymmetric classes lie intermediate to them with considerable (by eye) overlap.

- By correlating the supervised CNN-based latent features in PCA space with the residual quantitative metrics (SPF, B, and RFF), we find that the $PC1$ eigen axis correlates strongly with SPF and relatively weakly with B and RFF, where objects with $PC1 \gtrsim 1$ have higher values in all three metrics, and vice-versa.

- Informed by a k-means clustering inertia test on the supervised CNN-based latent features in PCA space, we identify 6 clusters ($n_{\rm gmm,cnn}$) using GMM and find separating decision boundaries using SVC. We find that $n_{\rm gmm,cnn} = 2$ group exclusively contains objects hosting strong symmetric and asymmetric residuals, whereas the $n_{\rm gmm,cnn} = 3, 4$ clusters host centrally dominant and "Clean" residuals, respectively. On the other hand, the clusters $n_{\rm gmm,cnn} = 0, 1, 5$ comprise weaker diffuse

residual signatures. We find that our CNN-based latent space offers promise towards naturally distinguishing different residual characteristics.

- We also implement an unsupervised CvAE with 3-layered Encoder-Decoder framework and train it on our augmented training image data. We extract the learned latent space features from the encoder network and perform PCA on it. We assess this latent space in the prominent PCA eigen axes PC1 vs. PC2 and find that its overall visual-class wise distribution follows a continuum, where the Peculiar and General classes appear as outliers to the Clean distribution, and the Asymmetric and Core classes substantially overlap with the remaining remaining.

- We correlate the unsupervised latent features in PCA space with the quantitative residual metrics (SPF, B, and RFF) and find that only the $PC1$ eigen axis correlates strongly with the SPF values and weakly with B and RFF values.

- Using a k-means inertia test, we find that the CvAE-based latent features are best described by utmost 2 clusters. We apply GMM-based clustering to identify 2 clusters in the PCA axes and define a decision boundary using a SVC classifier. We find that the two identified clusters ($n_{\mathrm{gmm,vae}}$) are split at $PC1 \sim 0$, where the $n_{\mathrm{gmm,vae}} = 0$ $(= 1)$ cluster hosts a mixture of strong and intermediate strength residuals (clean and weak residuals), respectively. We find that our unsupervised CvAE based latent features although offer some insights, lack clear discriminatory power when characterizing residual substructure.

Our methodological framework and the informative results from our supervised and unsupervised DL models offers a promising path towards implementing automated methods for residual feature assessment. These are especially essential in the context of future large-scale telescope surveys.

CHAPTER 5

SUMMARY AND FUTURE WORK

As a key step towards better identifying galaxy mergers and quantifying their merger rate history, in this dissertation work, we carryout a comprehensive analysis of close-proximity galaxy pairs and plausible merging induced tidal features (and in general galactic substructures) using forefront observational data from the *Hubble Space Telescope* (*HST*) and realistic mock observations from leading theoretical simulations. Here, we briefly summarize the overarching outcomes of our analysis and discuss important open questions and plans for future work.

5.1 Close-pair method based work

5.1.1 Summary

Two decades of close-pair based merger rate measurements vary by up to a factor of five, especially during the early epoch of rapid galaxy development ($\sim 7-11$ Billion years ago), highlighting that the merger contribution to galaxy growth remains poorly constrained. In Chapter 1, we constrain the major merger rate evolution of similar-mass (mass ratio<4), massive ($\log M_{\rm stellar}/M_{\odot} \geq 10.3$) galaxies over $\sim 11\,{\rm Gyr}$ in cosmic history by analyzing the incidence of major close pairs among a large sample of ~ 9800 massive galaxies from the *HST*-CANDELS survey. A key result of this analysis is that the redshift evolution of the pair fraction $f_{\rm pair}(z)$ at $z \gtrsim 1.5-3$ is likely flat or even

diminishing (Figure 10). Furthermore, we also find that the major merger rate evolution measurement based on our empirical $f_{\rm pair}(z)$ values and a commonly-used, non-evolving assumption for the close-pair observability timescale $T_{\rm obs,pair}$ disagree up to a factor of five with theoretical merger rate expectations at $2 < z < 3$ (Figure 11). We conclude that systematic errors on $T_{\rm obs,pair}$ are a dominant source of uncertainty (up to 50%, also see [18]) while measuring the merger rates, and a physically-motivated, evolving $T_{\rm obs,pair}$ may be required to explain the merger rate evolution discrepancy (Figure 12).

Recent pair fraction measurements made by different close-pair studies vary substantially during $1.5 \lesssim z \lesssim 3$ despite choosing similar pair selection choices, suggesting that study-dependent observational effects may be at play. In Chapter 3, we quantify the impact of common observational selection effects such as photometric redshift uncertainty ($\sigma_{\rm z}$), stellar-mass errors ($\sigma_{\log M_{\rm stellar}}$), sample completeness ($c$), and survey field area ($A_{\rm survey}$). We derive calibrations for them by analyzing close pairs among theoretical mock light cones generated using the leading SantaCruz Semi-Analytical Model (SC-SAM) with realistic *HST*-CANDELS survey like observational setting. A key result of our work is that SC-SAM calibrated empirical $f_{\rm pair}(z)$ measurement based on the *HST*-CANDELS survery is considerably flat during $1.5 < z < 3$ (Figure 36). Our results indicate that the previous $f_{\rm pair}(z)$ measurements need to be systematically lowered to correct for the substantial impact of observational selection effects at $z \gtrsim 2$.

5.1.2 Future Work

Despite active research on the topic of pair observability timescales, it remains

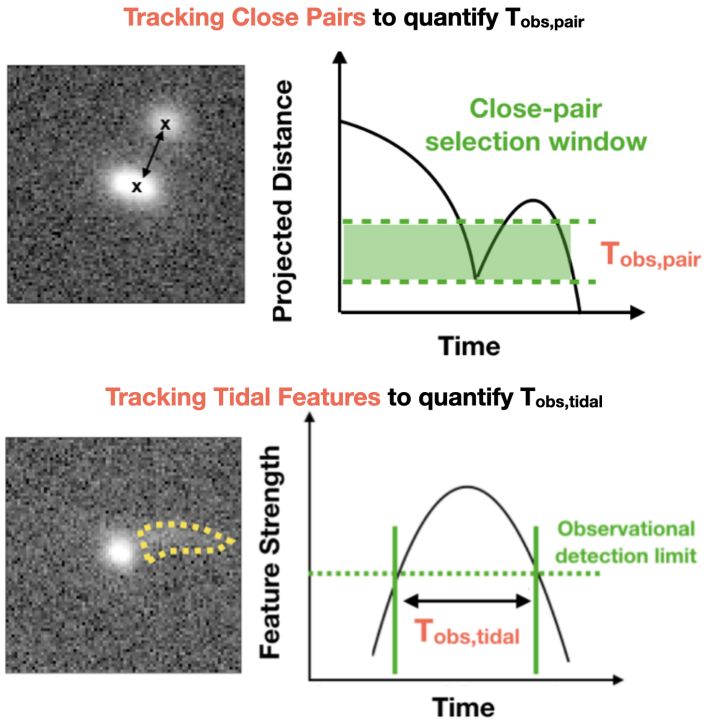

Figure 53: A conceptual illustration of our future steps to constrain the close-pair and tidal feature observability timescale by tracking mergers from the Horizon-AGN cosmological simulation [250].

a key open question in the field of galaxy merger rate quantification. As such, one of the key motivations moving forward is to robustly quantify the $T_{\rm obs,pair}$ values as a function of key physical parameters that govern the pair criteria – redshift (z), stellar-mass ($M_{\rm stellar}$), stellar-mass ratio (M_1/M_2), pair projected distance ($R_{\rm proj}$), and physical proximity (Δz or Δv). A direct and intuitive way of quantifying the $T_{\rm obs,pair}$ is to "follow" mergers during their close-pair stages using merger simulations that span a wide range of selection criteria (e.g., stellar-mass, mass ratio, etc). For example, [127] carried out a series of idealized binary merger simulations and quantified a $T_{\rm obs,pair}$ for different mass

ratio, projected distance selections, and assessed the impact of gas fractions on the measured timescales. Although this work has been the basis for many empirical merger rate studies using close pairs, their timescale values were constant (i.e., didn't vary as function of merger parameters) and they were computed in a controlled scenario of merging and not in a cosmological context. Another less-direct approach to infer the $T_{\rm obs,pair}$ is to use both the close-pair fractions and the intrinsic merger rates from the SAMs (e.g., [215]) or cosmological simulations (e.g., based on Illustris simulation [129] by [128]). However robustly estimating the true galaxy merger rate in the simulations can be quite challenging, especially in the case of semi-analytic assumptions of different physical mechanisms (including merging; see Chapter 3).

Robustly quantifying $T_{\rm obs,pair}$ requires a systematic analysis of a large sample of mergers from cosmological simulations. For our future work, we aim to quantify a thorough statistical prescription of the $T_{\rm obs,pair}$ evolution as a function of merger properties using 10^6 *simulated* mock observations of 10^4 major mergers from the Horizon-AGN [250] cosmological simulation, where each simulated merger is viewed at high-resolution timesteps ($\Delta t \sim 20\,{\rm Myr}$). In Figure 53 (left panel), we visually illustrate a key timescale measurement methodological step, where we will track each simulated merger in our sample to measure the cumulative time during which it satisfies a close-pair selection window (i.e., $T_{\rm obs,pair}$). After computing the $T_{\rm obs,pair}$ values for all the mergers in our sample, we will analyze evolutionary trends of $T_{\rm obs,pair}$ as a function of key merger properties (z, $M_{\rm stellar}$, M_1/M_2, and Δz) and provide best-fitting model prescriptions that describe the $T_{\rm obs,pair}$ evolution for common pair-selection choices used in the literature.

The outcomes of this work will yield improved merger rate measurements and will enable future theories to better incorporate the major merger contribution into galaxy evolution framework and provide improved modeling of the massive galaxy growth.

In the case of our SAM-based close-pair calibration analysis of observational selection effects discussed in Chapter 3, we focused on one specific model of galaxy evolution (i.e., the SC SAM) and adopted one observational survey setting (*HST* CANDELS). However, the derived pair fraction calibrations can vary among different SAM models, driven by their intrinsic variations in their resultant stellar-mass functions (number density of galaxies as a function stellar mass) at fixed redshifts. Furthermore, the pair fraction corrections derived in Chapter 3 are also a simultaneous function of the study-dependent variables $\sigma_{\rm z}$, $\sigma_{\log M_{\rm stellar}}$, c, and $A_{\rm survey}$. Therefore, a key motivation for future work is to assess the variability of different close-pair corrections by repeating our Chapter 3 analysis on mock light cones from *different* SAMs (e.g., Carnegie SAM [137], SAGE SAM [138]), and derive calibration corrections over for a grid of $\sigma_{\rm z}$, $\sigma_{\log M_{\rm stellar}}$, c, and $A_{\rm survey}$ values that are common among different existing observational surveys. The outcomes of this work will help future empirical studies to derive an improved and calibrated measurement of close-pair fraction evolution, which will yield accurate merger rate history constraints.

5.2 Tidal feature based work

5.2.1 Summary

Isolating plausible tidal signatures of merging can help improve identification of

interacting, merging, and post-merger systems. In Chapter 2, we developed a new Python-based software tool that identifies flux-wise and area-wise significant contiguous regions and extracts the different galactic substructures hosted within the model-subtracted "residual" images produced by popular parametric light-profile fitting tools (e.g., GALFIT). We demonstrated the applicability of our software by measuring the surface brightness of a variety of common residual features (disk structures, spiral substructures, plausible tidal features, and strong gravitational arcs) hosted by 17 galaxies using their *HST*-CANDELS survey H-band single-Sérsic residual images (Figures 21 & 22). We also showcase our tool's ability to track quantitative properties of tidal features using synthetic *HST* observations of a $z \sim 1.5$ major merger from the VELA hydrodynamic simulations (Figure 24). Our new software provides a promising approach to overcoming the subjectivity of identifying mergers via the presence of tidal features and a more direct way to extract and quantify their strength.

To overcome the subjective identification of merging related features and provide an automated and repeatable way of identifying them, in Chapter 4, we explore supervised Convolutional Neural Networks (CNNs) and unsupervised auto-encoders (CvAE) frameworks to characterize different kinds of residual substructures hosted by a sample of $\sim 10,000$ bright and massive galaxies from the *HST* CANDELS survey (Figure 43). We assess the latent space learned by our CNN and CvAE frameworks using Principal Component Analysis (PCA) along with qualitative (i.e., visual based) labels of residual substructure and independently quantified metrics of residual strength (significant pixel flux – SPF, Bumpiness – B, and Residual Flux Fraction – RFF). We find that our supervised

CNN latent features in PCA space can distinguish between qualitatively strong and weak residual substructures (Figure 45), whereas the unsupervised CvAE latent space lacks clear discriminatory power (compared to supervised CNN) when characterizing different residual substructures (Figure 47). Our DL-based exploratory exercise demonstrating the inter-disciplinary focus of this dissertation work serves as a key initial step for our future investigations to address the key open questions pertaining to tidal feature based merger identification.

5.2.2 Future Work

The tidal observability timescale $T_{\rm obs,tidal}$ during which mergers host detectable tidal signatures is an important parameter when quantifying morphology-based merger rates. Yet, the tidal observability timescales remain poorly constrained and there is a strong need to systematically quantify them as a function different merger properties (e.g., z, $M_{\rm stellar}$, M_1/M_2). Furthermore, the relationship between the observed properties of tidal features (e.g., strength, shapes) and the physical properties of merging galaxies causing them remains poorly understood. With this as the primary motivation for our future work, we aim to quantify the strength and timescales of tidal features hosted by simulated merging galaxies by analyzing them with feature extraction software developed in Chapter 2.

A conceptually straight-forward way to quantify $T_{\rm obs,tidal}$ is to apply quantitative (or qualitative) indicators of tidal features on a suite of merger simulations and identify the timescales over which the features are detected by the indicators. For example, [127]

also analyzed the $G - M_{20}$ metrics among a (small) suite of idealized binary merger simulations and quantified a timescale over which they satisfied a selection window in $G - M_{20}$ space. However, [251] found that non-parametric quantities such as $G - M_{20}$ and CAS do not effectively capture merging signatures (no better than a random guess), further complicating the issue. For our future work, we will track the strength of tidal features hosted by our 10^4 simulated mergers from the Horizon-AGN [250] cosmological simulation using our residual feature extraction software (Chapter 2). In Figure 53 (right panel), we visually illustrate our key tidal feature timescale measurement step. We will quantify the $T_{\rm obs,tidal}$ for each simulated merger as the cumulative time during which the extracted tidal feature surface brightness (flux per unit sky area) is brighter than a detection threshold set by observational survey image depth. Finally, we will also analyze the evolutionary dependence of $T_{\rm obs,tidal}$ and the feature surface brightness on the merger properties (z, $M_{\rm stellar}$, M_1/M_2) and provide their best-fit model prescriptions. The outcomes of this work will produce comprehensive tidal timescale prescriptions and will enable future studies to derive accurate tidal feature based merger rate measurements.

To understand the relationship between the observed properties of tidal features and merger properties, we will explore the application of supervised and unsupervised DL-based methods on the extracted tidal feature shapes and strengths of 10^4 simulated merger sample from the Horizon-AGN simulation such that the trained DL networks will predict merger progenitor properties based on observed tidal features. In Figure 54, we conceptually illustrate our supervised CNN and unsupervised CvAE based frameworks, which are inspired from our exploratory exercise discussed in Chapter 4. We will apply

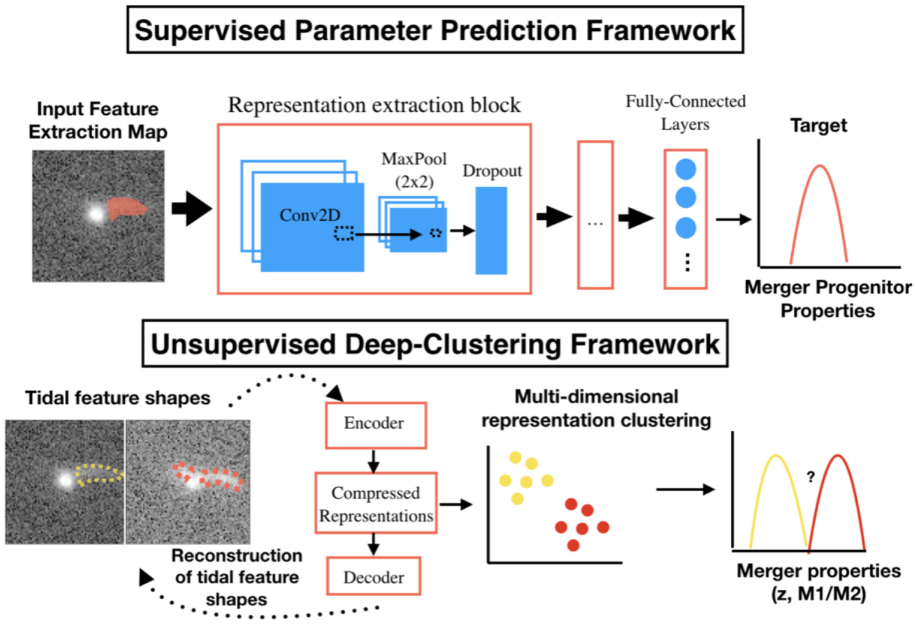

Figure 54: An overview of our future deep-learning based supervised Convolutional Neural Network (top) and unsupervised Convolutional Auto-Encoder (bottom) frameworks to map the observed properties of tidal features to the intrinsic properties of merging progenitor systems (z, $M_{\rm stellar}$, and M_1/M_2).

the feature extraction software (from Chapter 2) to our Horizon-AGN simulated merger sample to obtain their 2D tidal feature maps, and will train the CNN and CvAE frameworks to predict the properties of their merging systems. The outcomes of this work will improve our understanding of the relationship between tidal feature shape-similarity and the merging progenitor properties for the first time. This will help future studies to better quantify the major merger contribution to stellar-mass growth of high-mass galaxies.

5.3 Concluding Statements

Our efforts using observational and simulation based data in this dissertation work help improve existing merger identification methodologies (close pairs and tidal features)

and serve as important stepping stones towards improving constraints of the cosmic history of the rate of major mergers between comparable high-mass galaxies. The methodological framework and pipeline tools developed while meeting several of these research objectives are timely for upcoming large-scale surveys and latest cosmological simulations to further aid automated galaxy merger identification. The outcomes of our current work serve as foundational pillars for our future goals, which are the next key steps towards improving our understanding of the role of mergers in the overarching context of galaxy development.

REFERENCE LIST

[1] M. B. Taylor, TOPCAT STIL: Starlink Table/VOTable Processing Software, in *Astronomical Data Analysis Software and Systems XIV*, edited by P. Shopbell, M. Britton, and R. Ebert, volume 347 of *Astronomical Society of the Pacific Conference Series*, page 29, 2005.

[2] Astropy Collaboration et al., **558**, A33 (2013).

[3] D. R. Patton, C. J. Pritchet, H. K. C. Yee, E. Ellingson, and R. G. Carlberg, ApJ **475**, 29 (1997).

[4] L. Lin et al., ApJ **617**, L9 (2004).

[5] J. S. Kartaltepe et al., ApJ Supplement Series **172**, 320 (2007), arXiv: 0705.2266.

[6] K. Bundy et al., ApJ **697**, 1369 (2009), arXiv: 0902.1188.

[7] E. F. Bell et al., ApJ **652**, 270 (2006).

[8] A. R. Robaina et al., ApJ **719**, 844 (2010), arXiv: 1002.4193.

[9] J. M. Lotz et al., ApJ **672**, 177 (2008).

[10] C. J. Conselice, C. Yang, and A. F. L. Bluck, Monthly Notices of the Royal Astronomical Society **394**, 1956 (2009), arXiv: 0812.3237.

[11] S. Gottlöber, A. Klypin, and A. V. Kravtsov, ApJ **546**, 223 (2001).

[12] R. G. Bower et al., MNRAS **370**, 645 (2006).

[13] P. F. Hopkins et al., ApJ **715**, 202 (2010), arXiv: 0906.5357.

[14] A. F. L. Bluck et al., Monthly Notices of the Royal Astronomical Society: Letters **394**, L51 (2009), arXiv: 0812.0926.

[15] R. E. Ryan, Jr., S. H. Cohen, R. A. Windhorst, and J. Silk, ApJ **678**, 751 (2008).

[16] A. W. S. Man, A. W. Zirm, and S. Toft, ApJ **830**, 89 (2016).

[17] C. J. Mundy et al., ArXiv e-prints (2017).

[18] J. M. Lotz et al., ApJ **742**, 103 (2011), arXiv: 1108.2508.

[19] V. Rodriguez-Gomez et al., MNRAS **449**, 49 (2015).

[20] N. A. Grogin et al., ApJ Supplement Series **197**, 35 (2011), arXiv: 1105.3753.

[21] A. M. Koekemoer et al., ApJ Supplement Series **197**, 36 (2011).

[22] O. Fakhouri and C.-P. Ma, MNRAS **386**, 577 (2008).

[23] S. Genel, R. Genzel, N. Bouché, T. Naab, and A. Sternberg, ApJ **701**, 2002 (2009).

[24] O. Fakhouri, C.-P. Ma, and M. Boylan-Kolchin, MNRAS **406**, 2267 (2010).

[25] E. Neistein and A. Dekel, MNRAS **388**, 1792 (2008).

[26] A. Dekel et al., MNRAS **435**, 999 (2013).

[27] B. M. B. Henriques et al., MNRAS **451**, 2663 (2015).

[28] V. Springel, MNRAS **312**, 859 (2000).

[29] S. Khochfar and A. Burkert, ApJ **597**, L117 (2003).

[30] S. Khochfar and A. Burkert, MNRAS **359**, 1379 (2005).

[31] T. Naab, S. Khochfar, and A. Burkert, ApJ **636**, L81 (2006).

[32] T. J. Cox, P. Jonsson, R. S. Somerville, J. R. Primack, and A. Dekel, MNRAS **384**, 386 (2008).

[33] D. B. Sanders et al., ApJ **325**, 74 (1988).

[34] P. Di Matteo, F. Combes, A.-L. Melchior, and B. Semelin, **468**, 61 (2007).

[35] P. Di Matteo et al., **492**, 31 (2008).

[36] M. Martig and F. Bournaud, MNRAS **385**, L38 (2008).

[37] P. F. Hopkins et al., ApJ **652**, 864 (2006).

[38] J. D. Younger et al., MNRAS **396**, L66 (2009).

[39] D. Narayanan et al., MNRAS **407**, 1701 (2010).

[40] T. Di Matteo, V. Springel, and L. Hernquist, **433**, 604 (2005).

[41] P. F. Hopkins, T. J. Cox, D. Kereš, and L. Hernquist, ApJ Supplement Series **175**, 390 (2008).

[42] E. F. Bell et al., ApJ **640**, 241 (2006), arXiv: astro-ph/0506425.

[43] D. H. McIntosh et al., MNRAS **388**, 1537 (2008).

[44] A. van der Wel, H.-W. Rix, B. P. Holden, E. F. Bell, and A. R. Robaina, ApJ **706**, L120 (2009).

[45] S. Jogee et al., ApJ **697**, 1971 (2009), arXiv: 0903.3700.

[46] D. R. Patton, S. L. Ellison, L. Simard, A. W. McConnachie, and J. T. Mendel, MNRAS **412**, 591 (2011).

[47] E. Treister, K. Schawinski, C. M. Urry, and B. D. Simmons, ApJ **758**, L39 (2012).

[48] M. E. Weston et al., MNRAS **464**, 3882 (2017).

[49] T. Hewlett et al., MNRAS **470**, 755 (2017).

[50] A. R. Robaina et al., ApJ **704**, 324 (2009).

[51] A. M. Swinbank et al., MNRAS **405**, 234 (2010).

[52] T. A. Targett et al., MNRAS **412**, 295 (2011).

[53] N. A. Grogin et al., ApJ **627**, L97 (2005).

[54] D. D. Kocevski et al., ApJ **744**, 148 (2012), arXiv: 1109.2588.

[55] C. Villforth et al., MNRAS **439**, 3342 (2014).

[56] C. Villforth et al., MNRAS **466**, 812 (2017).

[57] F. Bournaud et al., ApJ **741**, L33 (2011).

[58] A. Dekel and A. Burkert, MNRAS **438**, 1870 (2014).

[59] R. Brennan et al., MNRAS **451**, 2933 (2015).

[60] D. R. Patton and J. E. Atfield, ApJ **685**, 235 (2008).

[61] L. de Ravel et al., **498**, 379 (2009).

[62] L. Lin et al., ApJ **681**, 232 (2008), arXiv: 0802.3004.

[63] A. W. S. Man, S. Toft, A. W. Zirm, S. Wuyts, and A. van der Wel, ApJ **744**, 85 (2012).

[64] R. De Propris et al., ApJ **666**, 212 (2007), arXiv: 0705.2528.

[65] E. K. Lofthouse, S. Kaviraj, C. J. Conselice, A. Mortlock, and W. Hartley, MNRAS **465**, 2895 (2017).

[66] C. R. Bridge et al., ApJ **659**, 931 (2007), arXiv: astro-ph/0701040.

[67] R. J. Williams, R. F. Quadri, and M. Franx, ApJ **738**, L25 (2011), arXiv: 1106.6054.

[68] J. E. Barnes and L. Hernquist, ApJ **471**, 115 (1996).

[69] F. Bournaud and P.-A. Duc, **456**, 481 (2006).

[70] S. Peirani et al., MNRAS **405**, 2327 (2010).

[71] D. W. Darg et al., MNRAS **401**, 1043 (2010).

[72] J. S. Kartaltepe et al., ApJ Supplement Series **221**, 11 (2015).

[73] T. Tal, P. G. van Dokkum, J. Nelan, and R. Bezanson, **138**, 1417 (2009).

[74] J. M. Lotz, J. Primack, and P. Madau, **128**, 163 (2004).

[75] C. J. Conselice, ApJ Supplement Series **147**, 1 (2003).

[76] C. López-Sanjuan et al., Astronomy and Astrophysics **501**, 505 (2009), arXiv: 0905.2765.

[77] Z. Z. Wen and X. Z. Zheng, ApJ **832**, 90 (2016).

[78] C. J. Conselice, S. Rajgor, and R. Myers, Monthly Notices of the Royal Astronomical Society **386**, 909 (2008), arXiv: 0711.2333.

[79] P. Cassata et al., MNRAS **357**, 903 (2005).

[80] A. Dekel, R. Sari, and D. Ceverino, ApJ **703**, 785 (2009).

[81] M. Cacciato, A. Dekel, and S. Genel, MNRAS **421**, 818 (2012).

[82] D. Ceverino, J. Primack, and A. Dekel, MNRAS **453**, 408 (2015).

[83] D. G. York et al., **120**, 1579 (2000).

[84] J. K. Adelman-McCarthy et al., ApJ Supplement Series **162**, 38 (2006).

[85] J. B. Oke and J. E. Gunn, ApJ **266**, 713 (1983).

[86] A. Galametz et al., ApJ Supplement Series **206**, 10 (2013), arXiv: 1305.1823.

[87] Y. Guo et al., ApJ Supplement Series **207**, 24 (2013), arXiv: 1308.4405.

[88] G. Barro et al., ApJ Supplement Series **243**, 22 (2019).

[89] H. Nayyeri et al., ApJ Supplement Series **228**, 7 (2017).

[90] M. Stefanon et al., ApJ Supplement Series **229**, 32 (2017).

[91] V. G. Laidler et al., **119**, 1325 (2007).

[92] M. L. N. Ashby et al., ApJ Supplement Series **218**, 33 (2015).

[93] T. Dahlen et al., ApJ **775**, 93 (2013), arXiv: 1308.5353.

[94] P. Santini et al., ApJ **801**, 97 (2015), arXiv: 1412.5180.

[95] B. Mobasher et al., ApJ **808**, 101 (2015).

[96] G. Chabrier, **115**, 763 (2003).

[97] G. Bruzual and S. Charlot, MNRAS **344**, 1000 (2003).

[98] L. Pozzetti et al., **523**, A13 (2010).

[99] X. Yang et al., ApJ **671**, 153 (2007).

[100] E. F. Bell, D. H. McIntosh, N. Katz, and M. D. Weinberg, ApJ Supplement Series **149**, 289 (2003), arXiv: astro-ph/0302543.

[101] M. R. Blanton et al., The Astronomical Journal **129**, 2562 (2005), arXiv: astro-ph/0410166.

[102] M. Fioc and B. Rocca-Volmerange, **326**, 950 (1997).

[103] R. S. de Jong and E. F. Bell, Astrophysics and Space Science Proceedings **3**, 107 (2007).

[104] J. Moustakas et al., ApJ **767**, 50 (2013).

[105] K. N. Abazajian et al., ApJ Supplement Series **182**, 543 (2009).

[106] M. A. Strauss et al., **124**, 1810 (2002).

[107] M. R. Blanton et al., **125**, 2276 (2003).

[108] M. Cebrián and I. Trujillo, MNRAS **444**, 682 (2014).

[109] L. de Ravel et al., arXiv:1104.5470 [astro-ph] (2011), arXiv: 1104.5470.

[110] C. J. Conselice, ApJ **638**, 686 (2006).

[111] D. R. Patton et al., ApJ **536**, 153 (2000).

[112] A. W. S. Man, A. Zirm, and S. Toft, arXiv:1112.3764 [astro-ph] (2011), arXiv: 1112.3764.

[113] G. B. Brammer, P. G. van Dokkum, and P. Coppi, ApJ **686**, 1503 (2008), arXiv: 0807.1533.

[114] D. Kodra, *The Galaxy Morphology-Density Relation at High Redshift with Candels*, PhD thesis, University of Pittsburgh, 2019.

[115] B. P. Moster, R. S. Somerville, J. A. Newman, and H.-W. Rix, ApJ **731**, 113 (2011).

[116] C. J. Conselice, M. A. Bershady, M. Dickinson, and C. Papovich, **126**, 1183 (2003).

[117] C. J. Conselice, S. Rajgor, and R. Myers, MNRAS **386**, 909 (2008).

[118] D. L. Domingue, C. K. Xu, T. H. Jarrett, and Y. Cheng, ApJ **695**, 1559 (2009).

[119] R. G. Carlberg, ApJ **359**, L1 (1990).

[120] K. Bundy, M. Fukugita, R. S. Ellis, T. Kodama, and C. J. Conselice, ApJ **601**, L123 (2004).

[121] A. B. Newman, R. S. Ellis, K. Bundy, and T. Treu, ApJ **746**, 162 (2012), arXiv: 1110.1637.

[122] A. F. L. Bluck et al., ApJ **747**, 34 (2012).

[123] C. López-Sanjuan et al., **548**, A7 (2012).

[124] C. López-Sanjuan et al., **530**, A20 (2011).

[125] M. G. Kitzbichler and S. D. M. White, MNRAS **391**, 1489 (2008).

[126] C. Y. Jiang, Y. P. Jing, and J. Han, ApJ **790**, 7 (2014).

[127] J. M. Lotz, P. Jonsson, T. J. Cox, and J. R. Primack, Monthly Notices of the Royal Astronomical Society **404**, 575 (2010), arXiv: 0912.1590.

[128] G. F. Snyder et al., MNRAS **468**, 207 (2017).

[129] M. Vogelsberger et al., MNRAS **444**, 1518 (2014).

[130] V. Rodriguez-Gomez et al., Monthly Notices of the Royal Astronomical Society **449**, 49 (2015), arXiv: 1502.01339.

[131] C. Maulbetsch et al., ApJ **654**, 53 (2007).

[132] P. F. Hopkins et al., ApJ **724**, 915 (2010).

[133] P. Madau and M. Dickinson, **52**, 415 (2014).

[134] L. Lin et al., ApJ **718**, 1158 (2010).

[135] V. Pandya et al., ArXiv e-prints (2016).

[136] R. S. Somerville, P. F. Hopkins, T. J. Cox, B. E. Robertson, and L. Hernquist, MNRAS **391**, 481 (2008).

[137] Y. Lu et al., ApJ **795**, 123 (2014).

[138] D. J. Croton et al., ApJ Supplement Series **222**, 22 (2016).

[139] A. Klypin, G. Yepes, S. Gottlöber, F. Prada, and S. Heß, MNRAS **457**, 4340 (2016).

[140] K. B. Mantha et al., MNRAS **475**, 1549 (2018).

[141] A. Toomre, Mergers and Some Consequences, in *Evolution of Galaxies and Stellar Populations*, edited by B. M. Tinsley and R. B. G. Larson, D. Campbell, page 401, 1977.

[142] J. E. Barnes, ApJ **331**, 699 (1988).

[143] R. G. Carlberg, C. J. Pritchet, and L. Infante, ApJ **435**, 540 (1994).

[144] S. E. Zepf and D. C. Koo, ApJ **337**, 34 (1989).

[145] B. C. Hsieh, H. K. C. Yee, H. Lin, M. D. Gladders, and D. G. Gilbank, ApJ **683**, 33 (2008).

[146] A. S. G. Robotham et al., MNRAS **444**, 3986 (2014).

[147] E. Ventou et al., **608**, A9 (2017).

[148] J. E. Barnes and L. Hernquist, ApJ **471**, 115 (1996).

[149] C. Wolf et al., ApJ **630**, 771 (2005).

[150] S. Jogee, Galaxy Interactions, Star Formation History, and Bulgeless Galaxies, in *The Galaxy Disk in Cosmological Context*, edited by J. Andersen, Nordströara, B. m, and J. Bland-Hawthorn, volume 254 of *IAU Symposium*, pages 67–72, 2009.

[151] R. G. Abraham et al., The Astrophysical Journal Supplement Series **107**, 1 (1996).

[152] R. G. Abraham et al., MNRAS **279**, L47 (1996).

[153] C. López-Sanjuan et al., **501**, 505 (2009).

[154] R. G. Abraham, S. van den Bergh, and P. Nair, ApJ **588**, 218 (2003).

[155] C. N. Lackner et al., **148**, 137 (2014).

[156] A. Toomre and J. Toomre, ApJ **178**, 623 (1972).

[157] T. M. Eneev, N. N. Kozlov, and R. A. Sunyaev, **22**, 41 (1973).

[158] C. Y. Peng, L. C. Ho, C. D. Impey, and H.-W. Rix, **124**, 266 (2002).

[159] C. Hoyos et al., MNRAS **419**, 2703 (2012).

[160] A. van der Wel et al., ApJ Supplement Series **203**, 24 (2012).

[161] A. van der Wel et al., ApJ **788**, 28 (2014), arXiv: 1404.2844.

[162] Planck Collaboration et al., **594**, A13 (2016).

[163] A. Hocking, J. E. Geach, Y. Sun, and N. Davey, MNRAS **473**, 1108 (2018).

[164] D. Ceverino et al., MNRAS **442**, 1545 (2014).

[165] M. Barden, B. Häußler, C. Y. Peng, D. H. McIntosh, and Y. Guo, MNRAS **422**, 449 (2012).

[166] E. Bertin and S. Arnouts, **117**, 393 (1996).

[167] A. Zolotov et al., MNRAS **450**, 2327 (2015).

[168] A. V. Kravtsov, A. A. Klypin, and A. M. Khokhlov, ApJ Supplement Series **111**, 73 (1997).

[169] A. V. Kravtsov, ApJ **590**, L1 (2003).

[170] D. Ceverino and A. Klypin, ApJ **695**, 292 (2009).

[171] M. Tomassetti et al., MNRAS **458**, 4477 (2016).

[172] C. E. Moody et al., MNRAS **444**, 1389 (2014).

[173] S. Inoue et al., MNRAS **456**, 2052 (2016).

[174] N. Mandelker et al., MNRAS **464**, 635 (2017).

[175] S. Tacchella et al., MNRAS **458**, 242 (2016).

[176] M. Huertas-Company et al., ArXiv e-prints (2018).

[177] D. Ceverino et al., MNRAS **457**, 2605 (2016).

[178] D. Ceverino et al., MNRAS **460**, 2731 (2016).

[179] G. F. Snyder et al., MNRAS **454**, 1886 (2015).

[180] R. C. Simons et al., arXiv e-prints , arXiv:1902.06762 (2019).

[181] P. Jonsson, MNRAS **372**, 2 (2006).

[182] P. Jonsson, B. A. Groves, and T. J. Cox, MNRAS **403**, 17 (2010).

[183] E. Dwek, ApJ **501**, 643 (1998).

[184] J. C. Weingartner and B. T. Draine, ApJ **548**, 296 (2001).

[185] B. T. Draine and A. Li, ApJ **657**, 810 (2007).

[186] J. E. Krist, R. N. Hook, and F. Stoehr, 20 years of Hubble Space Telescope optical modeling using Tiny Tim, in *Optical Modeling and Performance Predictions V*, volume 8127 of , page 81270J, 2011.

[187] K. Barbary et al., kbarbary/sep: v1.0.0, 2016.

[188] C. Fiorio and J. Gustedt, Theoretical Computer Science **154**, 165 (1996).

[189] K. Wu, E. Otoo, and A. Shoshani, Optimizing connected component labeling algorithms, in *Medical Imaging 2005: Image Processing*, volume 5747, pages 1965–1977, International Society for Optics and Photonics, 2005.

[190] T. H. Reiss, *Recognizing planar objects using invariant image features*, Springer-Verlag New York, Inc., 1993.

[191] W. Burger, M. J. Burge, M. J. Burge, and M. J. Burge, *Principles of digital image processing*, Springer, 2009.

[192] P.-A. Duc and F. Renaud, Tides in Colliding Galaxies, in *Lecture Notes in Physics, Berlin Springer Verlag*, edited by J. Souchay, S. Mathis, and T. Tokieda, volume 861 of *Lecture Notes in Physics, Berlin Springer Verlag*, page 327, 2013.

[193] P. M. S. Namboodiri and R. K. Kochhar, Bulletin of the Astronomical Society of India **13**, 363 (1985).

[194] V. A. Bruce et al., MNRAS **427**, 1666 (2012).

[195] D. Tuccillo et al., MNRAS **475**, 894 (2018).

[196] V. A. Bruce et al., MNRAS **444**, 1001 (2014).

[197] P. Dimauro et al., MNRAS **478**, 5410 (2018).

[198] S. Wuyts et al., ApJ **779**, 135 (2013).

[199] E. Kado-Fong et al., ArXiv e-prints (2018).

[200] S. Refsdal, MNRAS **128**, 307 (1964).

[201] C. S. Kochanek, ApJ **445**, 559 (1995).

[202] T. Treu and L. V. E. Koopmans, ApJ **611**, 739 (2004).

[203] G. B. Brammer et al., ApJ Supplement Series **200**, 13 (2012).

[204] A. van der Wel et al., ApJ **777**, L17 (2013).

[205] K. Duncan et al., ApJ **876**, 110 (2019).

[206] R. S. Somerville and J. R. Primack, MNRAS **310**, 1087 (1999).

[207] R. S. Somerville, R. C. Gilmore, J. R. Primack, and A. Domínguez, MNRAS **423**, 1992 (2012).

[208] S. Cole, ApJ **367**, 45 (1991).

[209] D. J. Croton et al., MNRAS **365**, 11 (2006).

[210] A. J. Benson and R. Bower, MNRAS **405**, 1573 (2010).

[211] G. Popping, R. S. Somerville, and S. C. Trager, MNRAS **442**, 2398 (2014).

[212] G. Popping et al., MNRAS **454**, 2258 (2015).

[213] L. Y. A. Yung, R. S. Somerville, G. Popping, and S. L. Finkelstein, MNRAS **494**, 1002 (2020).

[214] L. Y. A. Yung et al., MNRAS **496**, 4574 (2020).

[215] J. A. O'Leary, B. P. Moster, T. Naab, and R. S. Somerville, MNRAS **501**, 3215 (2021).

[216] J. A. O'Leary, B. P. Moster, and E. Krämer, MNRAS **503**, 5646 (2021).

[217] A. Rodríguez-Puebla et al., MNRAS **462**, 893 (2016).

[218] R. S. Somerville and R. Davé, **53**, 51 (2015).

[219] P. Behroozi, R. H. Wechsler, A. P. Hearin, and C. Conroy, MNRAS **488**, 3143 (2019).

[220] A. Galametz et al., ApJ Supplement Series **206**, 10 (2013), arXiv: 1305.1823.

[221] D. Ceverino, A. Dekel, D. Tweed, and J. Primack, MNRAS **447**, 3291 (2015).

[222] K. B. Mantha et al., MNRAS **486**, 2643 (2019).

[223] T. Tal et al., ApJ **789**, 164 (2014).

[224] M. Huertas-Company et al., ApJ Supplement Series **221**, 8 (2015).

[225] S. Dieleman, K. W. Willett, and J. Dambre, MNRAS **450**, 1441 (2015).

[226] H. Domínguez Sánchez, M. Huertas-Company, M. Bernardi, D. Tuccillo, and J. L. Fischer, MNRAS **476**, 3661 (2018).

[227] R. B. Metcalf et al., **625**, A119 (2019).

[228] T.-Y. Cheng et al., MNRAS **494**, 3750 (2020).

[229] C. Tohill, L. Ferreira, C. J. Conselice, S. P. Bamford, and F. Ferrari, arXiv e-prints , arXiv:2012.09081 (2020).

[230] L. Ferreira et al., ApJ **895**, 115 (2020).

[231] K. Fukushima and S. Miyake, Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition, in *Competition and Cooperation in Neural Nets*, edited by S.-i. Amari and M. A. Arbib, pages 267–285, Berlin, Heidelberg, 1982, Springer Berlin Heidelberg.

[232] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, Proceedings of the IEEE **86**, 2278 (1998).

[233] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, Stacked convolutional auto-encoders for hierarchical feature extraction, in *Artificial Neural Networks and Machine Learning – ICANN 2011*, edited by T. Honkela, W. Duch, M. Girolami, and S. Kaski, pages 52–59, Berlin, Heidelberg, 2011, Springer Berlin Heidelberg.

[234] K. Storey-Fisher et al., arXiv e-prints , arXiv:2105.02434 (2021).

[235] M. Huertas-Company et al., MNRAS **489**, 1859 (2019).

[236] M. Walmsley et al., **491**, 1554 (2020).

[237] J. P. Blakeslee et al., ApJ **644**, 30 (2006).

[238] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, 2017.

[239] I. T. Jolliffe, *Principal Components in Regression Analysis*, pages 129–155, Springer New York, New York, NY, 1986.

[240] J. de la Calleja and O. Fuentes, MNRAS **349**, 87 (2004).

[241] M. A. Hearst, IEEE Intelligent Systems **13**, 18â28 (1998).

[242] M. Huertas-Company, J. A. L. Aguerri, M. Bernardi, S. Mei, and J. Sánchez Almeida, **525**, A157 (2011).

[243] F. Pedregosa et al., Journal of Machine Learning Research **12**, 2825 (2011).

[244] B. E. Boser, I. M. Guyon, and V. N. Vapnik, A training algorithm for optimal margin classifiers, in *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, pages 144–152.

[245] S. Lloyd, IEEE Trans. Inform. Theor.(1957/1982) **18** (1957).

[246] J. MacQueen et al., Some methods for classification and analysis of multivariate observations, in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297, Oakland, CA, USA, 1967.

[247] C. J. Alpert, A. B. Kahng, and S.-Z. Yao, Discrete Applied Mathematics **90**, 3 (1999).

[248] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, in *Proceedings of the Second*

International Conference on Knowledge Discovery and Data Mining, KDD'96, page 226â231, AAAI Press, 1996.

[249] G. J. McLachlan and K. E. Basford, *Mixture models: Inference and applications to clustering*, volume 38, M. Dekker New York, 1988.

[250] Y. Dubois, M. Volonteri, and J. Silk, MNRAS **440**, 1590 (2014).

[251] M. W. Abruzzo, D. Narayanan, R. Davé, and R. Thompson, arXiv e-prints , arXiv:1803.02374 (2018).

VITA

Kameswara Bharadwaj Mantha was born on April 28, 1992, in Hanamkonda, Andhra Pradesh, India. He completed his early education at Kennedy High School in 2002, and graduated at Bhashyam Public School, Vijayawada, in 2005. He completed his intermediate studies with Narayana IIT Academy, Vijayawada, in 2007. He joined Koneru Lakshmaiah University, Vijayawada, in 2009, and graduated with a Bachelors in Electronics and Communications Engineering in 2013. Mr. Mantha was motivated during his childhood and schooling days to study physics and especially astronomy-related topics. He self-taught all of undergraduate-level physics and successfully attempted the Physics GRE test in 2013.

In 2014, Mr. Mantha joined the Master's program in Physics at the University of Missouri Kansas City. He successfully passed the departmental qualifier at a Ph.D. level in 2015, and started pursing his inter-disciplinary Ph.D (iPhD). in Physics and Astronomy, and Computer Science with with Dr. Daniel McIntosh and the UMKC Galaxy Evolution Group. Mr. Mantha's research interests are to understand the role of galaxy mergers and in the grand context of galaxy evolution, and he is keen on applying forefront artificial intelligence concepts for his research. During his iPhD pursuits, Mr. Mantha published 2 first-author professional papers in leading astronomical journals, co-authored on 7 collaborative papers, presented at 24 major conferences and gave seminars at 7 invited sessions, received 14 merit scholarships from UMKC, won the 2018 Graduate Chambliss Medal from the American Astronomical Society. He is currently leading 2 first-author projects

(Chapters 3 and 4 in this dissertation), which he anticipates to submit upon completion of this degree. Mr. Mantha will be joining the Minnesota Institute for Astrophysics, University of Minnesota, Minneapolis, as post-doctoral researcher.

Mr. Mantha is an active member of the American Astronomical Society and the Astronomical Society of India.