

**On the key processes that drive galaxy evolution:  
the role of galaxy mergers, accretion, local  
environment and feedback in shaping the  
present-day Universe**

*Author:*

Garreth MARTIN

*Supervised by:*

Dr. Sugata KAVIRAJ

Prof. Martin HARDCASTLE

Centre for Astrophysics Research  
School of Physics, Astronomy and Mathematics  
University of Hertfordshire

*Submitted to the University of Hertfordshire in partial fulfilment of the requirements of the  
degree of Doctor of Philosophy.*

May 2019

# *Abstract*

The study of galaxy evolution is a fundamental discipline in modern astrophysics, dealing with how and why galaxies of all types evolve over time. The diversity of present-day galaxies is a reflection of the processes through which these populations were assembled and offers insights into how these processes influence and regulate their mass assembly over the lifetime of the Universe. The currently favoured hierarchical paradigm of structure formation hypothesises that much of a galaxy's evolution must be driven by mergers. It is therefore important to understand the role of the merger process in shaping the galaxy populations in today's Universe. Together with data from large observational surveys, statistical studies of galaxy evolution rely on comparison to simulations, which can be used to make realistic survey-scale predictions. Together these two approaches can offer powerful insights into the processes that drive galaxy evolution over cosmic time.

I have used the Horizon-AGN simulation to study the effect of galaxy mergers on the stellar populations and central super-massive black holes of galaxies over cosmic time. I have shown that, while mergers can enhance star formation and black-hole growth significantly in the low redshift Universe, these enhancements are small at high redshift when the cosmic SFH peaks. This is because galaxies are already gas-rich at early epochs and mergers are not able to increase gas densities in the central regions of the galaxy. As a result, mergers are directly responsible for creating only around 30 per cent of the stellar mass and black-hole mass found in today's galaxies and that mergers never dominate the budget (e.g.  $\sim 35$  and  $\sim 20$  per cent of star formation at  $z \sim 3$  and  $z \sim 1$  respectively are a result of mergers).

Notwithstanding their relatively minor role in driving stellar and BH mass growth, mergers are important drivers of morphological change, with major and minor mergers accounting for essentially all (95 per cent) of the morphological change experienced by massive present-day spheroids over their lifetime. However, at a given stellar mass, the average merger histories of discs and spheroids do not differ strongly enough to explain the survival of discs to the present day. Instead, their survival is largely due to a preponderance of prograde and gas rich mergers. Prograde mergers trigger milder morphological transformation than retrograde mergers – the average change due to retrograde mergers is around twice that due to their prograde counterparts at  $z \sim 0$  and remnant morphology also depends strongly on the gas fraction of a merger, with gas-rich mergers routinely re-growing discs. My results also emphasise the important role of minor mergers, which dominate the stellar mass and black-hole growth budget after  $z = 1$  and are a potentially important reservoir of cold gas which plays a role in the rejuvenation and survival of discs.

I have also investigated the biases that this morphological evolution produces in observational studies of galaxy populations. In particular, I have shown that 'progenitor bias' i.e. the bias produced by using only early-type galaxies to define the progenitor population of today's early-types, is a significant problem at all but the lowest redshifts and an important consideration for large, deep observational surveys (JWST,

LSST etc.). For example while early-types attain their final morphology at relatively early epochs – by  $z \sim 1$ , around 60 per cent of today’s early-types have had their last significant merger, progenitor bias is severe at all but the lowest redshifts. At  $z \sim 0.6$ , less than 50 per cent of the stellar mass in today’s early-types is actually in progenitors with early-type morphology, while, at the peak epoch of cosmic of star-formation ( $z \sim 2$ ), studying only early-types misses almost all (80 per cent) of the stellar mass that eventually ends up in local early-type systems.

I have explored the significance and formation mechanisms of low-surface-brightness galaxies (LSBGs). For  $M_\star > 10^8 M_\odot$ , LSBGs contribute 50 per cent of the local number density and exist in significant numbers across all environments. Their progenitors have stronger, burstier star formation at high redshift which causes stronger supernova feedback. This feedback flattens the gas-density profiles (but does not remove the gas reservoirs). This, in turn, gives rise to flatter stellar profiles, which are more susceptible to environmental processes and galaxy interactions, which produce today’s LSBG populations by driving the steady removal of cold gas and gradually increasing galaxy effective radii over time. The ability of these populations to elucidate key questions in the field of galaxy evolution and significantly alter our current paradigm is becoming increasingly clear, especially with the advent of new deep surveys.

Finally, I have implemented a new *unsupervised* machine learning technique (UML) on images from the Hyper-Suprime-Cam Subaru-Strategic-Program Ultra-Deep survey. The algorithm autonomously reduces galaxy populations down to a small number of ‘morphological clusters’, populated by galaxies with similar morphologies, which are then benchmarked using visual inspection. The morphological classifications reproduce known trends in key galaxy properties as a function of morphological type (e.g. stellar mass functions and colours). This study demonstrates the power of UML in performing accurate morphological analysis, which will become indispensable in the forthcoming era of deep-wide surveys.

# Declaration

I declare that no part of this work is being submitted concurrently for another award of the University or any other awarding body or institution. This thesis contains a substantial body of work that has not previously been submitted successfully for an award of the University or any other awarding body or institution.

The following parts of this submission have been published previously and/or undertaken as part of a previous degree or research programme:

1. Chapter 2: this has been published as Martin, G.; Kaviraj, S.; Devriendt, J. E. G.; Dubois, Y.; Pichon, C.; Laigle, C., 2018, *Monthly Notices of the Royal Astronomical Society*, **474**, 3140.
2. Chapter 3: this has been published as Martin, G.; Kaviraj, S.; Devriendt, J. E. G.; Dubois, Y.; Laigle, C.; Pichon, C., 2017, *Monthly Notices of the Royal Astronomical Society: Letters*, **472L**, 50.
3. Chapter 4: this has been published as Martin, G.; Kaviraj, S.; Volonteri, M.; Simmons, B. D.; Devriendt, J. E. G.; Lintott, C. J.; Smethurst, R. J.; Dubois, Y.; Pichon, C., 2018, *Monthly Notices of the Royal Astronomical Society*, **476**, 2801.
4. Chapter 5: this has been published as Martin, G.; Kaviraj, S.; Devriendt, J. E. G.; Dubois, Y.; Pichon, C., 2018, *Monthly Notices of the Royal Astronomical Society*, **480**, 2266.
5. Chapter 6: this has been published as Martin, G.; Kaviraj, S.; Laigle, C.; Devriendt, J. E. G.; Jackson, R. A.; Peirani, S.; Dubois, Y.; Pichon, C.; Slyz, A., 2019, *Monthly Notices of the Royal Astronomical Society*, **485**, 796.
6. Chapter 7: this has been submitted as Martin, G.; Kaviraj, S.; Hocking, A.; Read, S. C.; Geach, J. E., 2019 to the *Monthly Notices of the Royal Astronomical Society*.

This work makes use of the Horizon-AGN simulation produced by Yohan Dubois (PI), Christophe Pichon, Julien Devriendt (co-Is) and the Horizon collaboration ([horizon-simulation.org/people.html](http://horizon-simulation.org/people.html)). Yohan Dubois produced simulation catalogues used for some of the galaxy properties/measurements used throughout this thesis, Marta Volonteri was responsible for bulge-to-total ratio measurements used in Chapter 4, Clotilde Laigle was responsible for producing the mock galaxy images and Sébastien Peirani was responsible for producing catalogues matching DM haloes used in Chapter 6. Chapter 7 is based on code originally written by Alexander Hocking.

# *Acknowledgements*

*To everyone who made my time here special, thank you for a wonderful three and a half years.*

I would like to extend my thanks to all the people whose support has been invaluable to me and has made it a pleasure to study here at Hertfordshire. Firstly, I am indebted to my supervisor, Sugata Kaviraj who went above and beyond anything that I could have expected from and acted as an incredible teacher and mentor. Thank you for your endless time, encouragement, support and enthusiasm and your invaluable feedback and career advice. I am also grateful to all the academics and students at Hertfordshire, Paris, Oxford and elsewhere for productive discussions (and not-so-productive arguments), for your valuable advice and for your friendship and camaraderie. In particular, my thanks go to my long-time mentors and collaborators Christophe Pichon, Julien Devriendt and Clotilde Laigle and Yohan Dubois who have insightful and supportive in equal measure. Thank you also to Jim Geach, Brooke Simmons, Becky Smethurst, Chris Lintott, Alex Hocking and Marta Volonteri and Robert Lupton for your time and expertise.

To all of the postgraduate students who I have shared the last three and a half years with: Joanna Ramasawmy, Ryan Cheale, Vijay Mahatma, Ryan Jackson, Emma Lofthouse, Jerly Galeano, Elena Gonzalez, Calum Morris, Ben Law, Maya Horton, Mubela Mutale, Boris Nedelchev, Thomas Spriggs, Tracey Garratt, Piyamas Choochalerm, Will Cooper, Max Lisogorsky, Amy Harris, Jon Westcott, Shaun Read, Alex Jones, Jaime Vargas, Matt Doherty, Aiyuan Yang, George Reavley, Chris Haynes, Nancy Hine, Kushatha Ntwaetsile, thank you for your friendship and for many great times together. In particular I would like to thank Shaun Read for his many useful insights and for putting up with me for the last three years. A special thank you also goes to my academic siblings, Ryan Jackson and Emma Lofthouse. Thank you also to the many wonderful people that I have been fortunate to have shared lively and supportive office environment with over the years: Ryan Jackson, Tracey Garratt, Billy English, Patrick Yates, Tabassum Tanvir, Boris Nedelchev, John Tolday, Maddie Burchard and Emmanuel Manful, thank you for making my time here entertaining and productive.

And thank you to the many academics at Hertfordshire and elsewhere: Leigh Smith, Jan Forbrich, Sebastien Viaenne, Dan Smith, Adrienne Slyz, Martin Krause, Martin Hardcastle, David Valls-Gabaud, Wendy Williams, Carolyn Devereux, Zhen Guo, Sarah Jaffa, Elias Brinks, Marc Sarzi, Donna Rogers-Lee, Vivienne Wild, Kristen Coppin and Chiaki Kobayashi, for your support and for many interesting discussions and useful advice.

Finally, this work would not have been possible without the love and support of my family, especially my mother, who has always supported me and encouraged me to follow my ambitions.

# Contents

<b>Abstract</b>	<b>i</b>
<b>Acknowledgements</b>	<b>iv</b>
<b>Contents</b>	<b>v</b>
<b>1 Introduction</b>	<b>1</b>
1.1 The evolution of galaxies and their properties over cosmic time . . . . .	1
1.2 The processes that drive structural evolution and mass assembly over cosmic time . . . . .	4
1.2.1 Accretion of gas . . . . .	5
1.2.2 In-situ star formation . . . . .	6
1.2.3 Merging . . . . .	7
1.2.4 Regulating processes . . . . .	10
1.2.4.1 Feedback . . . . .	10
Reionisation . . . . .	10
Stellar feedback . . . . .	11
AGN feedback . . . . .	11
1.2.4.2 Environment . . . . .	12
1.3 Cosmological scale numerical simulations of galaxy formation and evolution . . . . .	12
1.3.1 Horizon-AGN . . . . .	14
1.3.1.1 Cosmology . . . . .	14
Cosmological parameters . . . . .	14
Initial conditions and evolution . . . . .	14
1.3.1.2 Hydrodynamics code and refinement scheme . . . . .	14
1.3.1.3 Gas physics . . . . .	15
Gas heating and cooling . . . . .	15
Pressure floor . . . . .	15
Metallicity . . . . .	16
1.3.1.4 Star formation . . . . .	16
Seeding star particles . . . . .	16
Feedback by SNIa SNII, winds . . . . .	16
1.3.1.5 Black Holes . . . . .	17
Seeding . . . . .	17
Accretion . . . . .	17
Merging . . . . .	18
Feedback by AGN . . . . .	18
1.3.1.6 Compatibility with current observations . . . . .	18

1.4	Overview . . . . .	19
<b>2</b>	<b>The Progenitors of present-day early-type galaxies</b>	<b>22</b>
2.1	Introduction . . . . .	22
2.2	The simulation . . . . .	26
2.2.1	Horizon-AGN . . . . .	26
2.2.2	Identifying galaxies and building merger trees . . . . .	27
2.2.2.1	Identifying the galaxy sample . . . . .	28
2.2.2.2	Producing merger histories . . . . .	28
2.2.3	Prediction of observables . . . . .	28
2.2.3.1	Morphology . . . . .	28
2.2.3.2	Local environment . . . . .	29
2.2.3.3	Star-formation rate . . . . .	31
2.3	Redshift evolution of the progenitors of local early-type galaxies . . . . .	31
2.4	Identifying late-type galaxies that are progenitors of present day early-types . . . . .	34
2.5	Summary . . . . .	39
<b>3</b>	<b>The limited role of galaxy mergers in driving stellar mass growth over cosmic time</b>	<b>41</b>
3.1	Introduction . . . . .	41
3.2	The Horizon-AGN simulation . . . . .	44
3.3	Star formation enhancement due to merging . . . . .	44
3.3.1	Defining and identifying mergers . . . . .	45
3.3.2	Star formation triggered by mergers . . . . .	45
3.4	The merger contribution to the cosmic star formation history . . . . .	48
3.5	Summary . . . . .	49
<b>4</b>	<b>Normal black holes in bulge-less galaxies: the largely quiescent, merger-free growth of black holes over cosmic time</b>	<b>50</b>
4.1	Introduction . . . . .	50
4.2	The Horizon-AGN simulation . . . . .	55
4.2.1	Treatment of baryons . . . . .	55
4.2.2	Identification of galaxies and mergers . . . . .	56
4.2.3	The growth of black holes and black-hole feedback on ambient gas . . . . .	56
4.2.4	Galaxy morphology: measurement of B/T ratios . . . . .	58
4.3	Mergers and production of bulges . . . . .	59
4.4	BH growth over cosmic time . . . . .	63
4.4.1	Is there a correlation between BH growth and merger history? . . . . .	63
4.4.2	Contribution of mergers to the cosmic BH accretion budget . . . . .	65
4.5	Summary . . . . .	70
<b>5</b>	<b>The role of mergers in driving morphological transformation over cosmic time</b>	<b>73</b>
5.1	Introduction . . . . .	73
5.2	The Simulation . . . . .	77
5.2.1	Baryons . . . . .	77
5.2.2	Identification of galaxies and mergers . . . . .	78
5.2.3	Galaxy morphology . . . . .	79
5.2.4	Treatment of black holes and black-hole feedback . . . . .	79
5.3	The effect of individual mergers . . . . .	81

5.3.1	Morphological change induced by major and minor mergers as a function of stellar mass and redshift . . . . .	83
5.3.2	The effect of gas fraction . . . . .	86
5.3.3	The effect of orbital configuration . . . . .	87
5.3.4	Average $\Delta$ morph in the progenitors of today's discs and spheroids . . . . .	89
5.4	The cumulative effect of mergers over cosmic time . . . . .	92
5.4.1	Galaxy morphology over cosmic time . . . . .	92
5.4.2	Average merger histories of spheroids and discs . . . . .	95
5.4.3	Cumulative impact of mergers on galaxy morphology over cosmic time . . . . .	97
5.4.4	The fraction of global morphological transformation triggered by major and minor mergers . . . . .	101
5.5	Conclusions . . . . .	103
<b>6</b>	<b>The formation and evolution of low-surface-brightness galaxies</b>	<b>106</b>
6.1	Introduction . . . . .	106
6.2	The Horizon-AGN simulation . . . . .	109
6.2.1	Baryons . . . . .	110
6.2.2	Identifying galaxies and merger trees . . . . .	110
6.2.3	Surface-brightness maps and selection of LSBGs . . . . .	111
6.3	The low-surface-brightness Universe at the present-day . . . . .	113
6.3.1	Contribution of LSBGs to the local number, stellar mass and luminosity densities	115
6.3.2	Properties of LSB galaxies at the present day . . . . .	117
6.4	Redshift evolution of LSBG progenitors . . . . .	121
6.4.1	Gas fractions and effective radii . . . . .	122
6.4.2	Density profiles . . . . .	124
6.5	How do low-surface-brightness galaxies form? . . . . .	128
6.5.1	Supernova feedback - a trigger for LSBG formation . . . . .	130
6.5.2	Perturbations due to the ambient tidal field - a key driver of LSBG evolution . . .	135
6.5.3	Ram pressure stripping - an additional mechanism of gas removal in cluster LSBGs	138
6.5.3.1	Ram pressure . . . . .	139
6.5.3.2	Bulk flow of gas . . . . .	140
6.6	Summary . . . . .	142
<b>7</b>	<b>Galaxy morphological classification in deep-wide surveys via unsupervised machine learning</b>	<b>146</b>
7.1	Introduction . . . . .	146
7.2	The Hyper Suprime Cam Subaru Strategic Program (HSC-SSP) . . . . .	148
7.2.1	Survey description . . . . .	148
7.2.2	Data . . . . .	149
7.3	An unsupervised machine-learning algorithm for morphological classification . . . . .	150
7.3.1	Feature selection . . . . .	150
7.3.2	Feature extraction . . . . .	151
7.3.2.1	Growing neural gas . . . . .	151
7.3.2.2	Hierarchical clustering . . . . .	152
7.3.3	Constructing feature vectors . . . . .	153
7.3.4	Producing morphological clusters . . . . .	153
7.3.5	Cross-matching the UML objects with the HSC-SSP . . . . .	155

7.4	Morphological catalogue . . . . .	155
7.4.1	Benchmarking of morphological clusters via visual classification . . . . .	155
7.4.2	Star-galaxy separation . . . . .	157
7.4.3	Completeness of the UML-classified galaxy sample . . . . .	159
7.4.4	Released data products . . . . .	162
7.5	Galaxy properties as a function of morphological type . . . . .	162
7.5.1	Sample selection and methodology . . . . .	164
7.5.1.1	Redshift binning . . . . .	164
7.5.1.2	$1/V_{max}$ weighting and simulation of uncertainties . . . . .	164
7.5.2	Stellar mass distributions as a function of morphological type . . . . .	166
7.5.3	Star-formation rates and rest-frame colours as a function of morphological type . . . . .	166
7.6	Summary . . . . .	170
<b>8</b>	<b>Conclusions</b>	<b>174</b>
8.1	How do mergers drive the evolution of galaxy properties over cosmic time? . . . . .	174
8.2	Progenitor bias . . . . .	175
8.3	The formation of low-surface-brightness (LSB) galaxies . . . . .	176
8.4	A novel unsupervised machine-learning technique for galaxy morphological classification in new and forthcoming deep-wide surveys . . . . .	177
<b>9</b>	<b>Future work</b>	<b>178</b>
9.1	Comparison and limitations of current work . . . . .	178
9.2	High resolution cosmological simulations . . . . .	179
9.3	The drivers of galaxy diversity in the low-surface-brightness Universe . . . . .	182
9.4	Visual morphologies at the intersections of the cosmic web . . . . .	185
9.5	The UVCANDELS survey . . . . .	187
9.5.1	UV, optical, IR and composite morphologies with UVCANDELS . . . . .	187
9.5.2	Elucidating the origins of the UV upturn in elliptical galaxies with UVCANDELS . . . . .	188
<b>A</b>		<b>189</b>
A.1	Tables of progenitor probabilities . . . . .	189
<b>B</b>		<b>191</b>
B.1	Robustness of results with respect to choice of merger timescale and mass ratio threshold . . . . .	191
<b>C</b>		<b>195</b>
C.1	Extrapolating the mass and surface-brightness functions . . . . .	195
C.2	Effect of the spatial resolution of the simulation . . . . .	195
C.3	Relevance of this study to observed LSBG populations . . . . .	200
<b>D</b>		<b>202</b>
D.1	Cross-matching the UML objects to HSC-SSP DR1 centroids . . . . .	202
D.2	Released data products - lists of morphological clusters and individual galaxy properties . . . . .	202
	<b>Bibliography</b>	<b>203</b>

# Chapter 1

## Introduction

### 1.1 The evolution of galaxies and their properties over cosmic time

Even before Hubble (1926b) was able to demonstrate conclusively that the galaxy in which we live is just one among many similar objects, the resolved nature of ‘spiral nebulae’ (e.g. Schaeberle, 1904; Franks, 1905) allowed for their classification based on visible structure. Early classification schemes were able to take advantage of photographic plates, which allowed the internal structure of nearby objects to be resolved over long exposure times (e.g. Wolf, 1908; Lundmark, 1926). These early schemes did not distinguish, however, between galactic planetary nebulae and galaxies (see Fig 1.1), whose distant origin had yet to be determined.

The Hubble classification scheme was introduced by Hubble (1926a) and was the first overall scheme for classifying galaxies alone. The final Tuning Fork form of the Hubble classification was established in (Hubble, 1936) (see Fig 1.2) and was later revised by de Vaucouleurs (1959) to include internal structure such as bars and rings. Under the Hubble scheme galaxies are classified into four major morphological types. Elliptical and lenticular galaxies are referred to as early-type galaxies; spiral and irregular galaxies are referred to as late-type. Although Hubble did not intend to imply an evolutionary sequence with this nomenclature (in fact, Hubble (1927) warned against such an interpretation), it is now generally accepted that the galaxy population must have undergone significant morphological evolution between the late elliptical dominated Universe and the early spiral dominated Universe (Buitrago et al., 2013; Conselice, 2014). It is apparent from current observations that there is a significant average evolution of galaxy morphology and mass with redshift. Galaxies observed in the early Universe are generally small and peculiar, with high rates of star formation (Conselice et al., 2007; Buitrago et al., 2013; Madau & Dickinson, 2014). They also tend to be more disc-like. Galaxies in the nearby Universe are large, massive and quiescent by comparison and tend to have elliptical morphologies (Buitrago et al., 2013; Madau & Dickinson, 2014).

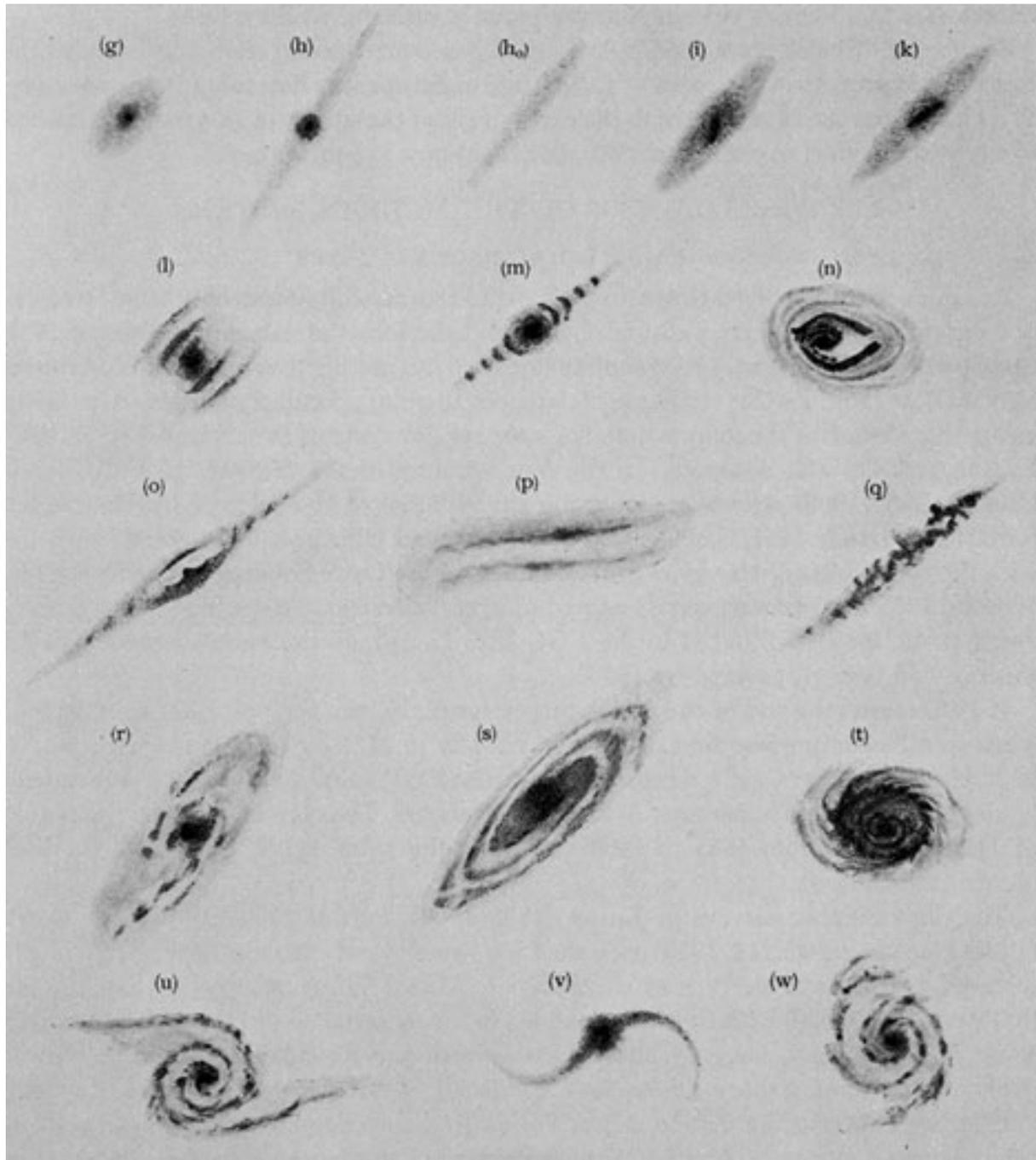


FIGURE 1.1: An early classification scheme by Wolf (1908) includes both planetary nebulas and galaxies.

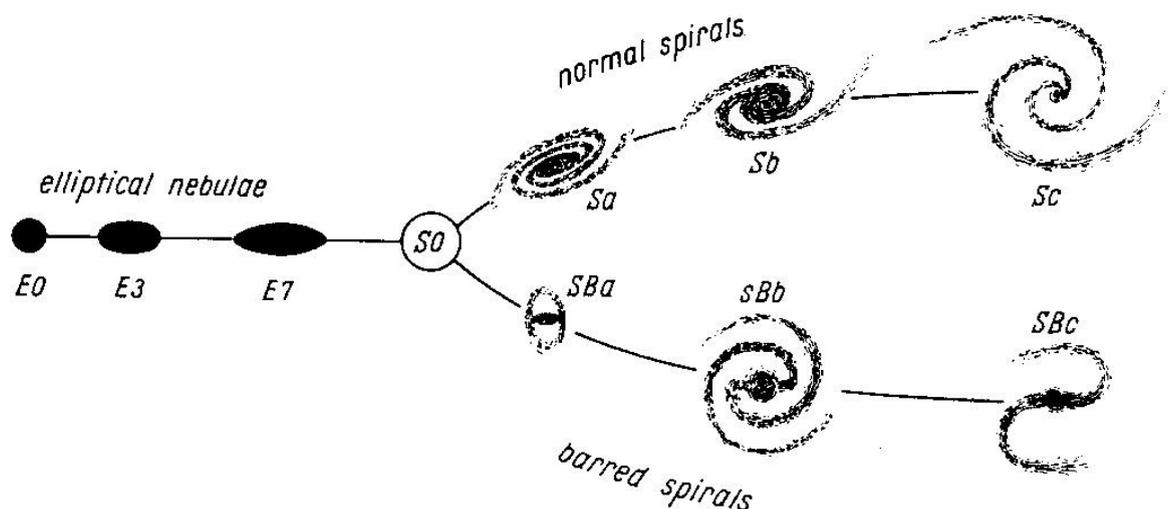


FIGURE 1.2: Classification scheme by Hubble (1936) which classifies only extragalactic nebulae.

Schemes based on the visual classification of galaxies are effective tools for astronomers because a galaxy's physical properties and evolutionary history are encoded within their morphology. The morphological type of galaxies at the present-day is therefore strongly aligned with their physical properties. There exists a robust correlation between morphology and fundamental physical properties such as stellar masses, star-formation rates, colours, merger histories and local environment (e.g. Dressler, 1980; Dressler et al., 1997; Strateva et al., 2001; Hogg et al., 2002; Bundy et al., 2005; Conselice, 2006; Skibba et al., 2009; Bluck et al., 2014; Smethurst et al., 2015; Whitaker et al., 2015), which all correlate strongly with galaxy morphology. For example, galaxies today can largely be classified as having elliptical or spiral morphologies, with each possessing distinctly different physical characteristics. Elliptical galaxies tend to be massive and red with old stellar populations (Strateva et al., 2001), whereas spirals tend to be less massive and blue with ongoing star formation. The clear segregation of galaxy properties by morphology is a clue as to the underlying physical processes that shape their evolution. A central question is, therefore, how the channels by which galaxies assemble their stellar mass, and their interplay with various regulating processes, are expressed in the diversity of galaxy sizes, gas fractions, morphologies and star formation histories at the present-day.

In recent decades, the convergence of wide-area surveys like the SDSS (Abazajian et al., 2009) and large-scale numerical simulations (e.g. Croton et al., 2006; Dubois et al., 2014; Vogelsberger et al., 2014) have had a transformational impact on our understanding of the physical drivers of galaxy formation. However, comparisons using large sample sizes have been largely restricted to low and intermediate redshift and relatively high surface-brightness objects. While studies examining morphology and structure in the local Universe are extensive, at high redshift, the decline in achievable spatial resolution renders studies of anything beyond basic morphological and structural properties of galaxies extremely challenging for current survey instruments in a majority of cases. As such, robust, statistical comparison of nearby and distant galaxies lies beyond the limits of current observational tools. Our observational picture of the evolution of their morphology and properties remains limited.

However, in the coming years, new deep-wide survey instruments and space observatories (e.g. HSC-SSP, LSST, JWST, EUCLID) will allow us to begin probing the distant and low-luminosity Universe, a new regime that has the potential to significantly alter our familiar understanding of galaxy evolution, particularly as observational and theoretical studies (e.g. Dalcanton et al., 1997; Impey & Bothun, 1997; Martin et al., 2019) have predicted a rich population of low surface-brightness (LSB) objects below the detection limits of contemporary wide surveys.

As the capability of new instruments to probe deeper and to higher redshift expands, simulations and numerical models continue to be a powerful method for probing galaxy evolution at high redshift. New, high-resolution simulations, e.g. New Horizon (Dubois et al. in prep) and Illustris TNG (Pillepich et al., 2018), will help revolutionise our understanding of the formation mechanisms of galaxies and allow us to make survey-scale predictions which offer powerful insights into the processes that drive galaxy evolution over cosmic times. Additionally, simulations that are already able to reproduce the observed low and intermediate redshift Universe to a high level of accuracy are also able to give us insights into the early Universe (Dubois et al., 2014; Kaviraj et al., 2016) and the evolutionary paths that lead to its present day composition, by providing a robust, time-resolved view of the evolution of galaxy properties over cosmic time.

## **1.2 The processes that drive structural evolution and mass assembly over cosmic time**

The first galaxies were likely cool gas discs assembled within dark matter (DM) haloes. These DM haloes were produced from the enhancement of over-densities in the matter field, created by quantum fluctuations present at the time of the Big Bang (Ma & Bertschinger, 1995). Following initial condensation into the dark matter haloes, baryonic matter (mainly pristine H and He) will tend to spin up and flatten into a disc like distribution as it cools and contracts. The initial source of angular momentum is postulated variously to be a consequence of bottom-up assembly (White & Rees, 1978), direct advection in the form of cold flows (Kereš et al., 2005) or tidal torques between gravitationally interacting dark matter haloes (Eggen et al., 1962). Regardless of the mechanism, once gas is sufficiently cool, stars are able to form as gas collects into dense clumps within the disc (Conselice, 2014; Somerville & Davé, 2015; Tonini et al., 2016). Galaxies can continue to assemble mass through two main channels: mergers with other galaxies and accretion from the intergalactic medium.

These channels are responsible for a number of major processes which shape the formation and evolution of galaxies:

### 1.2.1 Accretion of gas

In undisturbed environments, gas is accreted from the environment and onto the dark matter halo, where it is shock heated at the halo virial radius to the halo virial temperature. Because gas is shocked at the edges of the halo rather than close to the centre of the galaxy, the cooling rate is slow and gas collapses gradually (Rees & Ostriker, 1977; White & Rees, 1978). The result is a halo of hot gas that is trapped gravitationally within the dark matter halo (Croton et al., 2016; Tonini et al., 2016). ‘Hot’ mode accretion occurs as this hot halo gas slowly cools, resulting in gradual isotropic accretion. However, in some cases, a shock may instead form in the denser medium, closer to the center of the halo where the energy of the hot gas can be rapidly radiated away. This smooth or ‘cold’ mode of accretion allows gas to travel quickly into the centre of the galaxy, allowing the inside-out formation of a stellar disc (Tonini et al., 2016).

Theoretical evidence highlights the importance of such cold accretion flows bringing gas directly from the cosmic web (e.g. Kereš et al., 2005; Pichon et al., 2011; Kimm et al., 2011) and into the central regions of galaxies. These cold flows are thought to be important for the acquisition of angular momentum and in driving the formation and evolution of galaxies (Pichon et al., 2011). This is supported by observations of gas with significantly higher angular momentum than their DM haloes. For example, signatures of cold accretion are observed at low and intermediate redshifts in high spin, extended or warped HI and XUV discs (e.g. Bothun et al., 1987; Oosterloo et al., 2007; Holwerda et al., 2012; Rahmani et al., 2018) and in the early Universe in extended proto-galactic rings (e.g. Prescott et al., 2015; Martin et al., 2016). Evidence of past cold accretion may also persist in the morphology and statistical properties of galaxy populations, for example in galaxy spin alignment with filaments (e.g. Jones et al., 2010; Codis et al., 2012; Tempel et al., 2013; Rong et al., 2016) or in the colours and scale lengths of discs (Noguchi, 2018). Some observational evidence of the importance of accretion comes from the fact that observed star formation rates, when compared with the reservoir of fuel supplied by mergers alone, is observed to be deficient and requires additional gas accretion to make up the difference. For example, at high redshift Conselice et al. (2013) find that, in addition to mergers, a galaxy with  $M_* > 10^{11} M_\odot$  must accrete gas at a rate of around  $100 M_\odot \text{yr}^{-1}$  in order to make up the difference. The source of this gas can be explained by the accretion of gas from the intergalactic medium. This is roughly the same amount as is contributed by mergers, so the cold accretion of gas must also play a major role in the assembly of stellar mass over cosmic time.

Cold accretion flows are thought to be dominant at high redshift and for low mass haloes due to a combination of a lack of hot halo atmospheres and the high density of filamentary gas, which result in rapid cooling rates and prevent shocks from propagating (e.g. Kereš et al., 2005; Agertz et al., 2011; Benson & Bower, 2011). As haloes heat up and cosmic density drops towards later times, hot haloes are able efficiently shock accreting gas flows and prevent material from reaching the centre of the galaxy. This makes the hot mode of accretion increasingly dominant at later times. However, although cold

flows become less important, they may still provide some gas supply to lower mass haloes (e.g. Benson & Bower, 2011).

Finally, gas accretion can also have implications for galaxy structure, for example in the case of chaotic accretion. Chaotic gas accretion can occur as a result of high in-fall rates onto the galaxy due either to accretion from the intergalactic medium or accretion of matter during minor mergers. The high rate at which gas flows onto the galaxy results in a gravitational instability in the disc, the angular momentum of which gas must be dissipated in order for the disc to regain hydrodynamical equilibrium (Somerville & Davé, 2015; Tonini et al., 2016). Excess gas loses angular momentum and falls towards the centre of the disc. Gas clouds merge in the centre of the galaxy, triggering high levels of star formation and producing a bulge in the centre of the disc.

### 1.2.2 In-situ star formation

The accretion of gas is not, on its own, sufficient to increase the stellar mass of a galaxy as the reservoir of gas must first be converted into stars. While stellar mass assembly in galaxies is the result of several physical processes (chiefly in-situ star-formation from accreted gas or acquisition of ex-situ mass from mergers), ultimately, all stellar mass in the Universe must ultimately have formed in-situ within galaxy haloes.

In part because of the large dynamic range involved, the physics of star formation are difficult to model and are still relatively poorly understood. However the well established star formation laws, such as the Kennicutt-Schmidt law (Kennicutt, 1998), describe an empirical relationship between the integrated star formation rate density and gas surface density (or other gas properties) that appear to be relevant across a wide range of spatial scales (e.g. Roychowdhury et al., 2017; Viaene et al., 2018). In addition, most star-forming galaxies exist along a tight correlation in the stellar-mass–star-formation rate plane (the ‘main-sequence’) (e.g. Elbaz et al., 2007; Daddi et al., 2007), with only a small number galaxies lying above. The tightness of this relation and the relative dearth of outliers above it, may imply that the bulk of the star formation in the Universe proceeds at a steady pace. The star formation rate density is also observed correlate with the gas fractions of galaxies across cosmic time (e.g. Geach et al., 2011), peaking around  $z = 2$  (Madau & Dickinson, 2014).

Star formation is limited by the build up of dark matter haloes at high redshift and slowed at low redshift as feedback processes and decreasing average gas densities cause star formation rates to fall. However, individual gas-rich galaxies are highly star forming at high redshift, with-formation rates greater than  $100 M_{\odot} \text{ yr}^{-1}$  routinely observed (Daddi et al., 2005; Gruppioni et al., 2013). There is, therefore, good reason to believe that the gradual growth of stellar mass from in-situ processes has a significant role in the assembly of galaxies.

### 1.2.3 Merging

However, in addition to the in-situ conversion of gas into stars, mass can also be assembled by combining the existing stellar mass of two or more galaxies in the form of violent mergers. The favoured hierarchical paradigm of galaxy formation (White & Rees, 1978) implies that much of a galaxy's mass is assembled in this way, making them potentially important drivers of mass assembly and morphological evolution (as we will explore later in Chapters 3 and 6) (e.g. Bundy et al., 2007). The pattern of mergers is, nevertheless, encoded in the large-scale cosmic structure of the Universe. For example, observed anisotropies in the distribution of satellites in groups (e.g. Paz et al., 2011; Tempel et al., 2015) are the result of satellites being funnelled towards hosts along the axis the filaments (Welker et al., 2017; Shao et al., 2018). This may be an important effect since, as we explore later in this thesis (Chapter 6, merger outcomes may depend strongly on their geometry.

Usually, mergers are split into major and minor flavours, defined by the mass ratio of the two merging galaxies. Major mergers are mergers between galaxies of a similar mass and usually defined as having a mass ratio greater than 1:4. Minor mergers are mergers where one galaxy is significantly larger than the other and are usually defined by a mass ratio greater than 1:10 but less than 1:4 (Somerville & Davé, 2015). The merger rate, as well as the relative merger rate between major and minor mergers evolves as a function of cosmic time and environment, as the merger rate decreases with the expansion of the Universe.

Observationally, mergers are identified by disturbed morphology, or through observations of pairs of interacting (see Fig 1.3) or approaching galaxies (Conselice, 2014). None of these methods guarantees that a merger will or has taken place, so merger fractions calculated from this must be considered statistically (Buta, 2011; Conselice, 2014).

Through major and minor mergers, galaxies are able to increase their stellar mass through the existing stellar mass of the merging satellite galaxy. They can also increase their stellar mass through starbursts triggered by an enhancement of gas density in the host or by an increase in the amount of gas available to form stars from the reservoir of fuel provided by the merging satellite galaxy (Tonini et al., 2016). When gas clouds merge or their density is enhanced by local shocks, they may exceed the Jean's mass and collapse to form stars (Tonini et al., 2016).

For a given redshift, most galaxies with observed high star formation are mergers, although this does not necessarily imply that mergers always induce star formation (Mihos & Hernquist, 1996). In fact, while mergers do tend to enhance star formation of the host, there is a requirement that mergers are wet (gas rich) for any significant star formation to occur. That is, one of the merging galaxies must contain a significant amount of cold gas that can be used as a fuel source to produce new stars (Mihos & Hernquist, 1996; Tonini et al., 2016).



FIGURE 1.3: A pair of peculiar interacting galaxies NGC3808A and NGC3808B known collectively as Arp 87. Credit: NASA, ESA, and the Hubble Heritage Team

Typically wet mergers are observed to increase star formation rate of the host galaxy by a factor of 3 – 4. Very significant increases in star formation rate (greater than 10 times) are rarely observed either because they are intrinsically rare or because the duration of such starbursts is very short (Conselice, 2014).

In any case, the stellar mass of the host galaxy will always increase through the existing stellar mass of the satellite, even if there is a negligible net increase in the total stellar mass of the host and satellite.

As well as increasing their mass, mergers also induce changes in the morphology and kinematics of the galaxy. Mergers have the effect of increasing a galaxy’s dispersal velocity by randomising stellar orbits. This may reduce or entirely destroy the disc component and enhance the bulge component – inducing a change towards an elliptical morphology (Somerville & Davé, 2015). In addition to these more permanent changes, mergers can also produce temporarily peculiar morphologies, which remain until the merger remnant becomes dynamically relaxed (Conselice, 2014).

Major mergers are thought to be a significant driver of changes in galaxy morphology, with single major mergers between spirals typically being assumed to produce a remnant with elliptical morphology. Owing to high gravitational drag (Chandrasekhar, 1943), the duration of major mergers is relatively short (Bournaud, 2010; Somerville & Davé, 2015) (on order the rotation period of the outer galactic disc) and are therefore extremely violent. Lasting only a few orbits before the two galaxies coalesce, major mergers are able to induce changes in the gravitational potential of the galaxies that are more rapid than their dynamical timescale, thus efficiently redistributing the orbits and energies of individual stars in the process of ‘violent relaxation’ (Bournaud, 2010).

This occurs because individual stars experience a time-dependent potential as the two galaxies merge. Unlike masses orbiting in a static spherical potential (Saslaw, 2000) orbital energy is not conserved for

individual stars, leading to a non-Maxwellian distribution (Shlosman, 2012; Bournaud, 2010). Stars that lose energy from violent relaxation sink towards the centre of the galaxy while stars that gain energy move to further orbits or escape the system entirely. This has the effect of increasing concentration at the centre of the merger remnant while also extending the outer envelope of the galaxy (Hernquist & Mihos, 1995).

There is currently some debate as to whether gas-rich major mergers with subsequent rapid quenching of star formation by feedback from active galactic nuclei (AGN) is the primary mechanism by which spiral galaxies undergo the morphological transformation to elliptical – the so-called modern merger hypothesis (Hopkins et al., 2008). Evidence of this process is found in the observed centrally concentrated Sérsic profiles of many elliptical galaxies, suggesting violent relaxation has occurred (Bournaud, 2010; Conselice, 2014). However, it is not clear that the modern merger hypothesis is the main mechanism by which morphological transformation of ellipticals proceeds (Brennan et al., 2015) or even describes the formation of the majority of observed major merger remnants (Haines et al., 2015).

Because the number of galaxies increases towards the faint end of the luminosity function, mergers between galaxies of differing mass are more common and become increasingly common towards low redshift as the shape of the luminosity function becomes flatter (Man et al., 2012). Minor mergers are thought to have a less pronounced effect on the eventual structure of the galaxy, leaving the host galaxy relatively unaffected but still allowing the host to assemble stellar mass through the triggering of a starburst (in the case of wet mergers) and through the existing mass of the satellite (Tonini et al., 2016).

Although minor mergers are able to produce a bulge component, they do so via a different mechanism to major mergers. The rate of orbital decay in a minor merger system is much slower than for a major merger – around an order of magnitude longer. This makes minor mergers a comparatively gentle affair – resulting in more of a gradual accretion of the satellite than a conventional major merger (Toomre, 1977). In this scenario, the host strips material from its satellite, which can form into shells around the host galaxy (Naab, 2013; Tonini et al., 2016). Assuming the satellite is dense enough that it is not entirely ripped apart before the merger completes, it will impact with the host, scattering stellar orbits and thus thickening and warping the disc (Qu et al., 2011). In the case of disc-dominated spiral galaxies, minor mergers can also drive instabilities that grow the bulge through migration of stars to the galactic centre (Croton et al., 2016). Bulge dominated elliptical galaxies, on the other hand, may simply accumulate mass in shells around the central bulge (Tonini et al., 2016; Wellons et al., 2016).

Minor mergers are also a possible mechanism that is able to explain why massive spiral galaxies are able to survive to late redshifts despite the fact that they must assemble mass through mergers. Since they are able to increase the mass of the host while not significantly affecting its morphology, a host spiral galaxy undergoing a number of minor mergers may be able to assemble significant mass while retaining its spiral morphology (Schweizer, 2000; Somerville & Davé, 2015). Another possible explanation is that a disc is able to reform around the bulge through smooth accretion of gas subsequent to its destruction by a major merger or another mechanism (Kannappan et al., 2009; Somerville & Davé, 2015).

It should be noted that, as discussed in section 1.2.1, as well as through mergers, a bulge component may also build up through secular processes like accretion that can drive gravitational instabilities in the disc as discussed in section 1.2.1.

Therefore, dependent on the environment that they find themselves in, galaxies are able to grow in mass by one or a combination of evolutionary channels. The path that they take will determine the eventual morphology, kinematics and stellar population of the galaxy (Tonini et al., 2016; Wellons et al., 2016). The diversity of galaxy properties at the present day is, therefore, a reflection of the complex interplay of the underlying processes by which these populations were assembled. Processes including feedback, environmental mechanisms (e.g. cluster processes, Gnedin, 2003), large scale structure and gas dynamics (e.g. White & Rees, 1978; Pichon et al., 2011) and the evolution of the merger rate all influence or regulate the stellar mass assembly of galaxy populations, while also acting to shape their properties. Major feedback processes responsible for regulating their growth of galaxies include:

## 1.2.4 Regulating processes

The combination of evolutionary paths that a galaxy travels defines the dynamical and star-formation processes that take place. Assessing how these processes contribute to the evolution of the galaxy morphology, stellar mass and stellar populations helps us to explain the distribution of galaxy morphologies, masses and stellar populations seen at the present day.

### 1.2.4.1 Feedback

The formation and assembly of galaxies is regulated by a number of feedback processes, which may act to reduce the efficiency of star formation, quench it altogether or otherwise prevent galaxies from forming at all.

**Reionisation** Observations (e.g. Fan et al., 2006) indicate that the Universe was largely reionised early on in its history ( $z \sim 6$ ). Although their relative contributions are uncertain, the source of this ionising radiation is thought to be from a combination QSOs and young stars. The UV background radiation field produced by these sources heated the gas in the early Universe to around  $10^4$  K. The effect of this heating was to disrupt the formation of the lowest mass haloes (those that would otherwise grow to contain stellar masses of  $M_* < 10^7 M_\odot$ ) by ejecting their baryons, providing these baryons have not had a chance collapse earlier and begun to form stars (Silk, 2011). The source and timing of the ionising UV background as well as the evolution of its flux density and the local conditions of galaxies (e.g. Sawala et al., 2016) is subject to much uncertainty. Understanding the source of the ionising UV background is therefore important as it will determine the masses of the haloes that are disrupted – thus regulating the abundance of galaxies at the low-mass end of the galaxy stellar mass function (Efstathiou, 1992).

**Stellar feedback** At higher stellar masses ( $10^8 M_{\odot} < M_{\star} < 10^{10} M_{\odot}$ ), stellar feedback becomes important for expelling baryons in the early Universe (Silk, 2011). However supernovae and stellar winds remain important in all low mass haloes across cosmic time, where they continue to regulate their assembly. Low mass galaxies reduce their star formation efficiency through self-regulation, that is, the supernovae and stellar winds, that are a direct consequence of the formation of stars, act to regulate the levels of star-formation (Hayward & Hopkins, 2017). Stellar feedback operates by injecting energy into star forming gas, thus removing fuel for star formation by producing fountains of outflowing material. This material can then cool and infall, reducing star-formation efficiency (Dong et al., 2003; Agertz et al., 2011) and acting to prolong the timescale over which these galaxies form their stars.

Stellar feedback mechanisms include both type II and type Ia supernovae (SNII and SNIa) as well as stellar winds. SNII and stellar winds are associated with massive, young stars and, together, they account for the vast majority of the energy injection budget in the early history of a star forming region (Fierlinger et al., 2016) (a few tens of Myrs). On the other hand, stellar winds and SNIa operate over longer timescales, becoming the dominant source of energy injection on longer timescales (over Gyrs).

**AGN feedback** Finally, the energy released from central supermassive black holes (SMBH) as they accrete material is thought to be the dominant mechanism of feedback in the most massive galaxies ( $M_{\star} > 10^{11} M_{\odot}$ ) (e.g. Beckmann et al., 2017), whose potential wells are deep enough to bind gas against the effects of stellar feedback. Typically, AGN feedback is split into two major modes (e.g. Fabian, 2012; Best & Heckman, 2012). The quasar mode, which occurs during phases of rapid accretion onto the black hole (at close to the Eddington rate), typically in the most massive haloes and at higher redshifts (e.g. Hlavacek-Larrondo et al., 2013), and the jet mode, which typically occurs at lower accretion rates and outputs lower energies (Ishibashi et al., 2014).

During the quasar mode a portion of the orbital energy of the matter matter accreting onto the black hole is emitted in the form of radiation, which heats the medium surrounding the black hole. Quasar mode feedback is typically effective on small scales (pc or kpc), acting directly on the host galaxy to heat gas and quench star-formation (e.g. Veilleux et al., 2017). During the jet mode, most of the energy of the AGN is released in the form of kinetic energy, as the AGN drives powerful radio-emitting jets which heat the surrounding circumgalactic medium (e.g. Blandford & Königl, 1979; Rees, 1984).

The jet mode is thought to be important as a maintenance mode, acting to keep the surrounding medium hot and thus prevent the cooling and accretion of hot halo gas. Perhaps even more importantly, the influence of jet mode feedback is thought to extend to the large scale environment. For example, by counteracting the X-ray cooling of the hot gas, jet mode feedback can act to prevent cooling flows from entering clusters (e.g. McNamara & Nulsen, 2007; Hardcastle et al., 2019).

### 1.2.4.2 Environment

The environment of a galaxy is shown to have a strong effect on its structure. As Dressler (1980, 1984) first showed, morphology is observed to be strongly correlated with environment in the local Universe and at intermediate redshift (Dressler et al., 1999), with denser environments tending to have a greater fraction of elliptical galaxies. Additionally the tidal forces in these environments act to thicken discs, quench star formation and disrupt or destroy diffuse or low-mass galaxies.

However, the influence of environment is not so clear-cut at high redshifts (Papovich et al., 2012): there is also an observed correlation between morphology and stellar mass – with the most massive galaxies tending to be elliptical. In addition, the structure redshift relation sees the number of peculiar galaxies increase with redshift (Conselice, 2014). It is therefore not clear what is the main cause of galaxy morphological transformation: local environment, mass or gradual, secular evolution over time. This is the essence of the so-called ‘nature vs. nurture’ argument.

Comparatively little is understood about the effect of structure on cosmological scales. There is nevertheless a body of theoretical (e.g. Lee & Pen, 2000; Dubois et al., 2014; Aragon-Calvo & Yang, 2014; Musso et al., 2018) and observational (e.g. Tempel et al., 2015; Poudel et al., 2017; Kuutma et al., 2017) evidence that demonstrates the important role of large-scale structure on the evolution of galaxy properties, which extends beyond the effects of local density enhancements. Therefore, to obtain a complete picture of galaxy evolution, we must also understand how the evolution of galaxies and the large-scale structure of the Universe are interlinked.

## 1.3 Cosmological scale numerical simulations of galaxy formation and evolution

Recent observations above  $z = 2$  are now able to provide samples of sufficient size and quality that it is possible for matched samples at high and low redshift to be compared morphologically e.g. via the CANDELS survey (Lee et al., 2013). However, probing the evolution of many of the physical properties of galaxies at  $z = 2$  and morphology at higher redshifts remains difficult (Dubois et al., 2014; Somerville & Davé, 2015). The question of how galaxies evolve with cosmic time and how their environments drive changes to their physical properties has yet to be answered and, with current instrumentation, cannot be satisfactorily answered by observation alone (Moscardini & Dolag, 2011; Conselice, 2014).

Large-scale simulations consistent with the observed Universe can provide insight into the physical mechanisms involved in structure formation and galaxy formation and evolution, as well as make predictions that are testable by observation. They are a useful tool for validating cosmological models, as well as uncovering possible biases in real observations (Bertschinger, 2008; Moscardini & Dolag, 2011).

In the standard  $\Lambda$ CDM cosmology, primordial fluctuations in density were created when quantum fluctuations were expanded to macroscopic scales by a rapid inflationary phase (Ma & Bertschinger, 1995; Moscardini & Dolag, 2011). This produced small-scale inhomogeneities in the matter density field that enhanced over time; upon reaching a critical density, these regions of enhanced density became self-gravitating dark matter haloes, uncoupled from the expansion of the Universe (Ma & Bertschinger, 1995). Cosmological scale simulations attempt to evolve model universes from these small-scale inhomogeneities in order to probe structure formation and the evolution of galaxies across cosmic time.

This requires three ingredients. A cosmological model with an associated matter and radiation content, a model of the initial fluctuations in the density field and code which evolves a simulated universe from its initial conditions (Bertschinger, 2008).

Models of galaxy formation within a cosmological context present a significant challenge for simulators due to the breadth of spatial scales that must be simulated. A truly *ab initio* simulation would need to model cosmic filamentary structure on Mpc scales simultaneously with the sub-parsec scales associated with the accretion discs around black holes and feedback from supernovae. A realistic simulation of galaxy formation requires spatial resolutions better than 1 kpc and mass resolution better than  $10^6 M_{\odot}$  but within a volume of roughly 100 Mpc (the scale over which we conjecture the Universe is homogeneous and isotropic (Ellis, 1975; Ryden, 2003)) and containing over  $10^{17} M_{\odot}$  in total mass (Bertschinger, 2001). The other major obstacle is the large number of physical processes that must be modelled or whose effects must be approximated by numerical recipes (Somerville & Davé, 2015). As a result, many large-scale simulations today rely on uncertain and sometimes arbitrary sub-grid recipes for small scale and physical processes, even if they are poorly understood.

Advances in computer power and improved numerical techniques over the past few decades have led to remarkable advances in the scope, resolution and accuracy of cosmological scale numerical simulations. These developments have made it feasible to run increasingly larger scale  $N$ -body and hydrodynamical simulations at high enough resolution to model galaxy evolution (Somerville & Davé, 2015).

Numerical hydrodynamical simulations have an advantage over semi-analytical models because they allow the properties of dark matter, stars and gas to be obtained self-consistently. This approach allows for detailed predictions of galaxy spatial and kinematic properties across cosmic time in addition to more global properties. Processes at scales that cannot be simulated from first principles must instead be parameterised and matched to observations or else sub-grid recipes must be altered to match observations (Somerville & Davé, 2015). The evolution of galaxies within simulations can be traced through merger trees.  $N$ -body merger trees are produced by identifying DM haloes at each time step and attempting to match each halo's progenitors.

### 1.3.1 Horizon-AGN

Throughout this thesis, we make extensive use of the Horizon-AGN simulation (Dubois et al., 2014). Horizon-AGN is a large-scale hydrodynamical simulation adopting a standard  $\Lambda$ CDM cosmology with a box length of  $100 h^{-1} \text{coMpc}$ . In the following section, we detail the properties and recipes used by the simulation.

#### 1.3.1.1 Cosmology

**Cosmological parameters** Cosmological parameters are taken from WMAP7 results (Komatsu et al., 2011), which remain compatible to within 10% of more recent results such as Planck Collaboration et al. (2014). This corresponds to a flat  $\Lambda$ CDM cosmology with matter, dark energy and baryon density parameters of  $\Omega_m=0.272$ ,  $\Omega_\Lambda=0.728$  and  $\Omega_b = 0.045$  respectively and a Hubble constant of  $H_0 = 70.4 \text{ km s}^{-1} \text{Mpc}^{-1}$ . The amplitude and scalar spectral index of the matter power spectrum are  $\sigma_8=0.81$  and  $n_s = 0.968$  respectively (Dubois et al., 2014).

**Initial conditions and evolution** Initial conditions for the simulation are produced by the COSMICS package (Bertschinger, 2008) using the cosmological parameters listed above. Cosmological parameters go into selecting the shape of the power spectrum and producing an initial matter density and velocity field. Density fluctuations are evolved according to linearized equations of general relativity until the non-linear evolution of the density field must be treated numerically (Bertschinger, 2008).

#### 1.3.1.2 Hydrodynamics code and refinement scheme

The simulation employs the RAMSES tree-based adaptive mesh refinement (AMR) Eulerian hydrodynamics code (Teyssier, 2002). Refinement of an initially uniform  $1024^3$  cell grid is triggered according to a quasi Lagrangian criterion whenever the total mass in a cell exceeds 8 times the dark matter mass resolution of  $8 \times 10^7 \odot$ . Refinement of cells is allowed to continue down to a resolution of  $\Delta x \approx 1 \text{ kpc}$  in proper units. An additional maximum level of refinement is allowed at every doubling of the scale factor,  $a(t)$ , so that the minimum cell size is approximately constant (at 1 kpc) over cosmic time (Dubois et al., 2014; Kaviraj et al., 2016). Fig 1.4 demonstrates the variable resolution of the adaptive mesh where cells in areas of higher density have higher levels of refinement.

Gas properties are evolved by solving equations of hydrodynamics and thermodynamics over the AMR grid. AMR cells are also used to define the force softening and large scale force calculations in order to evolve the collisionless, gravitationally interacting dark matter and star particles (using a grid based  $N$ -body solver (Teyssier, 2002)).

The AMR approach, compared with other schemes like smoothed particle hydrodynamics (SPH), offers higher order consistency as well as greater ability to capture shocks and instabilities. AMR is also a more flexible refinement scheme. For example, Horizon-AGN is refined based on the DM mass within a cell rather than only the baryonic gas mass in a cell, which is advantageous if one wants to resolve gas flows within the relatively sparse halo of a galaxy. Additionally, because resolution is retained in regions of low density, AMR remains accurate in voids.

However, this approach also leads to some inaccuracies, such as the artificial torquing of discs into alignment with the grid (Hahn et al., 2010), lack of angular momentum conservation leading to degradation of orbits as well as inefficiencies and some inaccuracies relative to moving mesh or SPH based codes, which can more efficiently concentrate a greater number of resolution elements into dense regions since they have no limit to the coarseness of the resolution in regions that are empty.

### 1.3.1.3 Gas physics

Gas has an ideal monoatomic equation of state with an adiabatic index of  $\gamma = 5/3$ .

**Gas heating and cooling** Beginning at  $z = 10$ , heating from a uniform UV background begins. The evolution of the UV flux density follows (Haardt & Madau, 1996) based on the energy output from QSOs and hot, massive stars. Gas is also allowed to cool via H and He collisions with additional metal line cooling according to (Sutherland & Dopita, 1993). The minimum temperature that gas is allowed to cool down to is  $10^4$  K.

This relatively simple implementation of gas cooling and heating may introduce some sources of uncertainty. For example, as we mention above, since the UV background acts to suppress the formation of galaxies with low circular velocities and effectively prevents gas from condensing onto low mass haloes, the timing, uniformity, evolution and source (e.g. Trebitsch et al., 2018) of this background is important for regulating the abundances of low-mass dwarf galaxies. In practice this has little effect on the galaxy mass function in the mass ranges we probe in this thesis ( $M_* > 10^{8.5} M_\odot$ ), since the UV background is typically only effective at disrupting galaxies of much lower mass. Additionally, the temperature floor of  $10^4$  K may act to slow the cooling rate of accreting gas, can effectively reduce the rate of star formation further.

**Pressure floor** Above a critical density of  $\rho > \rho_0$ , where  $\rho_0$  is the density threshold for star formation described below, the gas pressure is artificially enhanced. This is in place in order to avoid excessive gas fragmentation and mimic the effect of stellar heating on the mean temperature of the interstellar medium, which is a process not otherwise captured by in the simulation. In order to implement the pressure floor, a polytropic equation of state ( $T = T_0(\rho/\rho_0)^{\kappa-1}$ ) is assumed, with a polytropic index  $\kappa = 4/3$ . While

necessary ensure the Jean's length is resolved so that gas is stabilised against collapse at small scales, the introduction of a pressure floor can produce spurious pressure support for low-mass (poorly resolved) objects, which may lead to overly thick discs.

**Metallicity** Gas metal species include C, O, N, Mg, Si and Fe and are tracked as passive variables. For cooling and heating calculations, the ratio of the abundances of metal species is not taken into account and are assumed to be solar. As described below, the metallicity of the gas is modified as a result of supernovae and stellar wind ejecta. The limited resolution (1 kpc) of the simulation means that metals are smeared out over relatively large volumes, equivalent to the assumption that metals become well mixed over very short timescales.

### 1.3.1.4 Star formation

**Seeding star particles** The process of star formation is modelled by the Schmidt law – the star formation rate density,  $\dot{\rho}_*$ , is given by

$$\dot{\rho}_* = \varepsilon_* \rho / t_{\text{ff}} \quad (1.1)$$

where  $\varepsilon_* = 0.02$  is the star formation efficiency,  $\rho$  is the gas density and  $t_{\text{ff}}$  is the free fall time of the gas. Star formation is only allowed to occur in gas cells where the number density of Hydrogen exceeds  $n_0 > 0.1 \text{ cm}^{-3}$ , equivalent to a critical density of around  $\rho_0 \lesssim 2 \times 10^{-25} \text{ g cm}^{-3}$ . Collisionless star particles are formed following a stochastic (Poissonian) approach, where star particles are generated in each star-forming gas cell with a probability proportional to the star formation rate density,  $\rho_*$ , at each time step (Rasera & Teyssier, 2006). Star particles are generated with a constant initial mass of  $m_* = \rho_0 \Delta x^3 \sim 2 \times 10^6 M_\odot$  and with the same initial chemical abundances as their parent gas cell. The corresponding mass, momentum and internal energy of the newly formed star particles are removed from the parent gas cell.

The Schmidt law formulation of the star formation efficiency does not take into account a number of effects that can act to reduce the efficiency. Other recipes, such as those that take into account turbulence (e.g. Federrath & Klessen, 2012), can reduce the effective star formation rate.

**Feedback by SNIa SNII, winds** The simulation employs continuous stellar feedback that includes momentum, mechanical energy and metals from stellar winds and both Type II and Type Ia supernovae (SNe). Momentum and mass as well as the chemical elements synthesised in stars are transferred to surrounding cells by supernova blast waves, sweeping up mass as they propagate. In order to realistically mimic the propagation of supernova bubbles, star particles may only inject energy, mass and momentum into the surrounding medium if the blast wave from all of the combined star particles in each cell is able to propagate further than  $2\Delta x$ . If the energy released in a given cell is not large enough to push the blast wave this far, the energy, momentum, and metals that would have been released are allowed to

accumulate until a large enough blast wave can be launched. This approach aims to prevent bubbles from expanding overly rapidly while also realistically propagating energy and metals.

The kinetic energy and metal enrichment contributed by Type II SNe by each star particle is calculated according to its mass, age and metallicity and is implemented using STARBURST99 (Leitherer et al., 1999, 2010), via the Padova model (Girardi et al., 2000) with thermally pulsating asymptotic giant branch stars (Vassiliadis & Wood, 1993). The ‘Evolution’ model of Leitherer et al. (1992) is used to calculate the kinetic energy of stellar winds. Matteucci & Greggio (1986) is used to determine the implementation of Type Ia SNe, assuming a binary fraction of 5% (Matteucci & Recchi, 2001), the SNe Ia rate is calculated according to a model by Greggio & Renzini (1983) and chemical yields are taken from the W7 model of Nomoto et al. (2007). Stellar feedback is assumed to be a heat source after 50 Myrs, because after this timescale the bulk of the energy is liberated via Type Ia SNe that have time delays of several hundred Myrs to a few Gyrs (e.g. Maoz et al., 2012). These systems are not susceptible to large radiative losses, since stars will disrupt or migrate away from their birth clouds after a few tens of Myrs (see e.g. Blitz & Shu, 1980; Hartmann et al., 2001).

### 1.3.1.5 Black Holes

**Seeding** Black hole ‘sink’ particles are created with an initial seed mass of  $10^5 M_{\odot}$  wherever both the gas and stellar density in a cell exceeds the star formation critical density,  $\rho_0$ , described above and when the stellar velocity dispersion exceeds  $100 \text{ km s}^{-1}$ . This criterion is used in order to mimic the conditions of the dense stellar clusters in which direct-collapse black holes are thought to form (e.g. Begelman et al., 2006).

It should be noted that the origin of SMBH seeds is not well understood and that the objects that originate these seeds and, therefore the seed mass is uncertain. Although many simulations assume a seed mass of  $10^5 M_{\odot}$ , based on a direct collapse scenario, other simulations (e.g. Taylor & Kobayashi, 2014; Wang et al., 2019) have been more successful at reproducing the observed evolution of certain scaling relations using lower mass seeds.

**Accretion** Black holes accrete matter at the Bondi-Hoyle-Lyttleton accretion rate (Edgar, 2004),  $\dot{M}_{BH}$ , given by,

$$\dot{M}_{BH} = 4\pi\alpha G^2 M_{BH}^2 \rho / (c_s^2 + u^2)^{3/2} \quad (1.2)$$

where  $c_s$  is the average sound speed and  $u$  is the average gas velocity.  $\alpha$  is the accretion efficiency or ‘boost’ factor.  $\alpha$  enhances accretion rate whenever gas density exceeds the critical value. The boost factor corrects for accretion of cold high-density phase ISM, since accretion is underestimated when the gas properties are not resolved (Booth & Schaye, 2009b). The accretion rate is not allowed to exceed the Eddington accretion rate with an assumed efficiency of  $\epsilon_r = 0.1$ . Above this, radiation pressure from

the black hole prevents any further accretion assuming uniform spherical accretion. This is likely a poor assumption as, in reality, accretion onto black holes is likely to be clumpy (e.g. Tremblay et al., 2016).

Additionally, rather than being anchored to the centre of their dark matter haloes as in some other simulations (e.g Taylor & Kobayashi, 2014; Schaye et al., 2015; Sijacki et al., 2015), BHs are allowed to move freely, with a drag force applied in order to mitigate unrealistic motions and spurious oscillations arising from the effect of a finite resolution. Without this drag force, the black hole may spend most of its time in regions of low density away from the galaxy disc, effectively preventing it from accreting at all.

**Merging** BHs are allowed to coalesce if they form a tight enough binary. Two black holes must be within four AMR cells of one another and have a relative velocity that is smaller than the escape velocity of the binary. The resulting mass of the merged binary is simply the sum of the masses of the two BHs.

**Feedback by AGN** AGN feedback operates with a radiative efficiency of  $\epsilon_r = 0.1$ . Feedback is modelled as a combination of two different modes, dependent on the ratio of the Bondi-Hoyle-Lyttleton accretion rate to the maximum Eddington accretion rate. During radio mode feedback, when  $\dot{M}_{BH} / \dot{M}_{Edd} < 0.01$ , the AGN deposits its energy mechanically in the form of bipolar outflows at a rate proportional to the square of the accretion rate and an efficiency of  $1 \times \epsilon_r$  and with jet velocities of  $10^4 \text{ km s}^{-1}$ . The outflows are modelled as a cylinder with a cross-sectional radius of  $\Delta x$  and height of  $2\Delta x$  following Omma et al. (2004). Quasar mode feedback, which operates when  $\dot{M}_{BH} / \dot{M}_{Edd} > 0.01$ , injects thermal energy into the surrounding nearby gas with a lower efficiency of  $0.15 \times \epsilon_r$  (Dubois et al., 2014). The efficiency of the quasar mode is chosen to reproduce the local observed  $M_{BH} - M_\star$  and  $M_{BH} - \sigma_\star$  relations and the local cosmic black-hole mass density (Dubois et al., 2012).

By imparting energy into the surrounding gas, AGN feedback acts to reduce the accretion rate of the black hole by altering its sound speed and density. The effect of radiation coupling to the ISM via radiation pressure on dust is not modelled as such radiative processes are costly to implement. The inclusion of radiation pressure driven outflows may act to further heat the surrounding medium and reduce accretion and star formation.

### 1.3.1.6 Compatibility with current observations

Kaviraj et al. (2016) and (Volonteri et al., 2016) compare the predictions of Horizon-AGN to observations from the local Universe over cosmic time. They show that, while Horizon-AGN is a naïve simulation with respect to the local Universe<sup>1</sup>, it reproduces galaxy properties related to mass growth from the late to early Universe with to good agreement as well as other key properties like the black

<sup>1</sup>i.e. all sub-grid models (except for feedback efficiency) use standard, fiducial values for their parameters, rather than being tuned to reproduce the observed Universe.

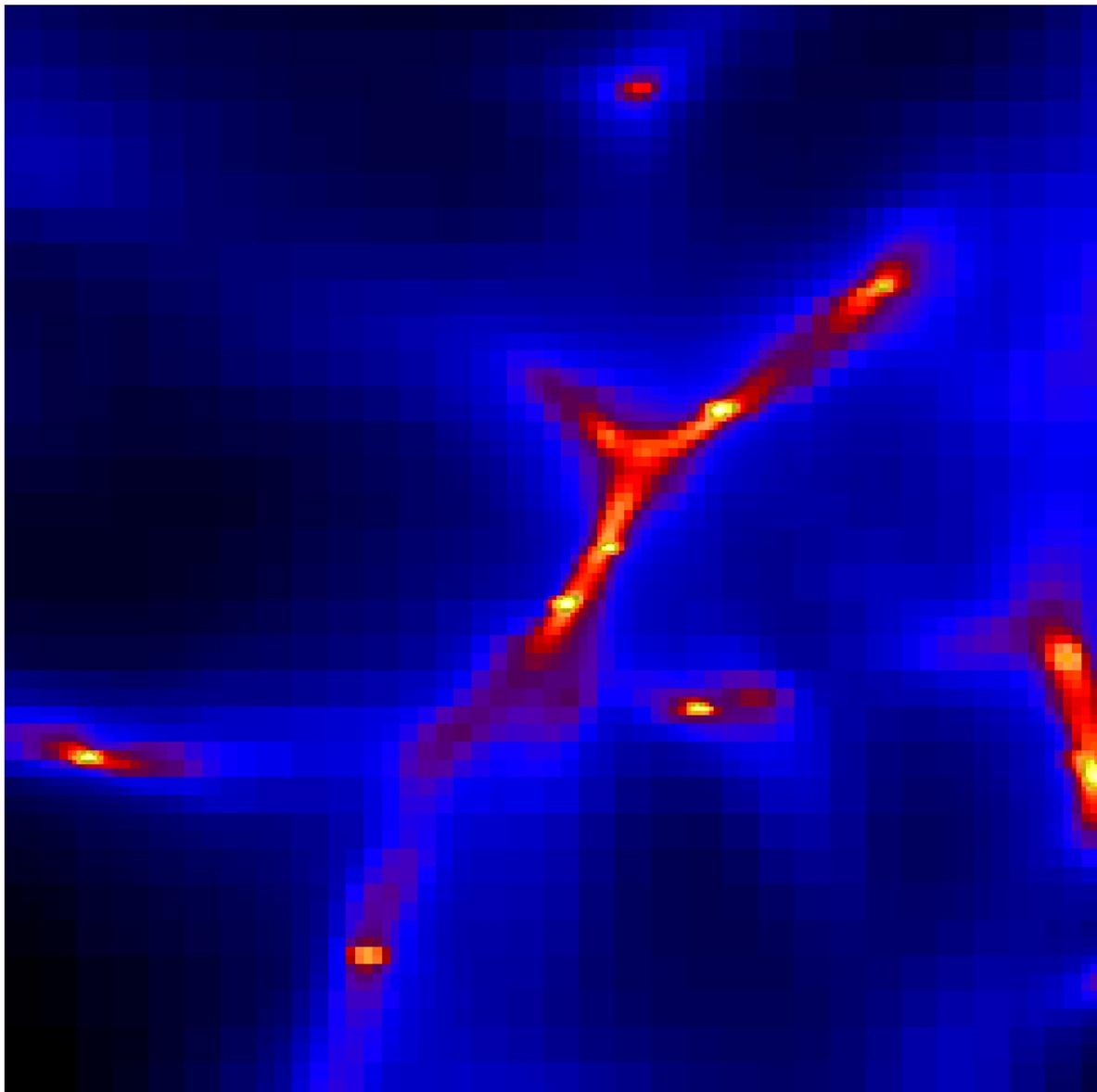


FIGURE 1.4: Projection of a slice through a small region of the Horizon-AGN simulation showing the gas density.

hole mass–stellar mass relation, galaxy colours, the evolution of the star-formation rate density and the star-formation rate main sequence. Thus, Horizon-AGN provides a means of exploring the processes that drive the evolution and mass assembly of galaxies over cosmic time.

## 1.4 Overview

The aim of this work is to use cosmological simulations to help explain how the diversity of galaxy properties at the present day arises from the underlying processes by which they are assembled. In particular, we have studied the role of mergers in driving some of the bulk quantities of the Universe and

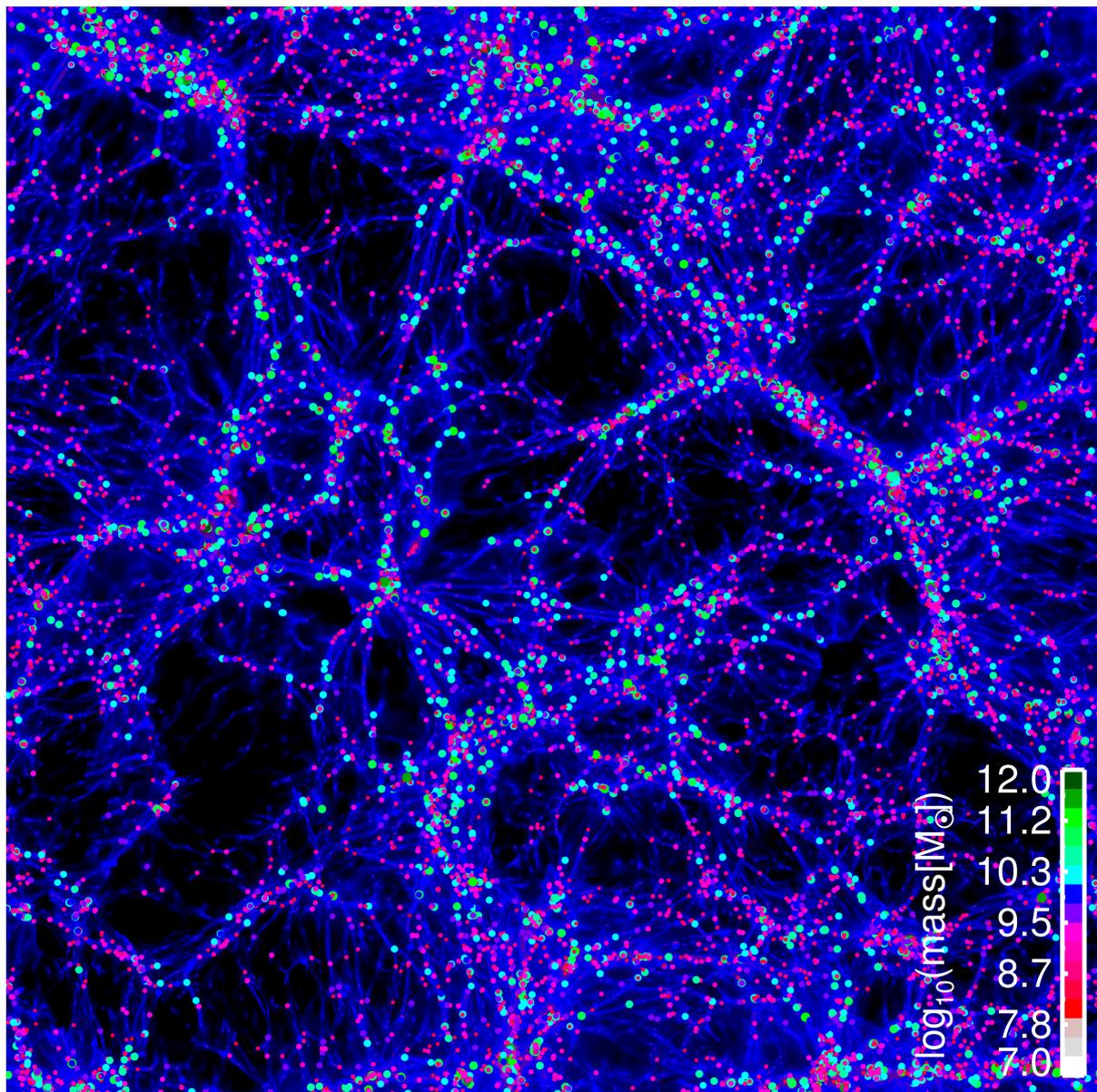


FIGURE 1.5: A slice through the Horizon-AGN volume at  $z = 0$  showing gas density. The locations of galaxies are overplotted and colour and size coded according to their mass.

developed theoretical tools for studying and quantifying the build up of galaxies and their morphology over cosmic time.

In Chapter 2, we present a framework for quantifying and alleviating the so called ‘progenitor bias’, which occurs if one uses only early-type galaxies to study the progenitor population of today’s early-type galaxies. We use the Horizon-AGN simulation to quantify the evolution of the progenitors of today’s early-type galaxies and provide a route for identifying late-type galaxies that are progenitors of present-day early-types in observational surveys by producing probabilistic prescriptions for finding progenitors based on their predicted stellar masses, star-formation rates, stellar kinematics and local environment.

In Chapter 3, we study the importance of galaxy mergers in driving stellar mass growth by quantifying

the contribution of galaxy merging to the cosmic stellar mass buildup of the universe over cosmic time. By calculating the mean sSFR in merging and non-merging systems, we define a merger-driven star formation enhancement and, using this quantity, calculate the fraction of star formation that is a *direct* result of minor or major mergers over cosmic time.

In Chapter 4, we study the importance of a quantity related to the star-formation budget, the BH mass accretion budget. We use Horizon-AGN to study the growth of BHs in the context of galaxy-BH co-evolution and specifically for the case of ‘bulge-less’ galaxies, which are thought to be merger-free. We investigate whether it is reasonable to assume that bulge-less galaxies have really undergone no significant episodes of merging, as well as the role of major and minor mergers in driving the BH mass accretion budget over cosmic time.

In Chapter 5, we attempt to elucidate the key processes that drive morphological transformation of discs into ellipticals at the present day and examine in detail the differences in the merger histories of discs and spheroids, including the geometry and gas content of the mergers that they undergo. We produce a detailed cumulative picture of the role that mergers play in the production of spheroids and the survival of discs and quantify the contribution of mergers to these processes.

In Chapter 6, we explore the significance and formation mechanisms of low-surface-brightness (LSB) galaxies (LSBGs), whose ability to elucidate key questions in galaxy evolution and potentially overturn many aspects of our current paradigm is becoming increasingly clear, especially with the advent of new deep survey instruments. We study formation and evolution of LSBGs and predict their frequency and properties at the present day. We explore the evolution of LSB and high surface-brightness (HSB) galaxies as a function of their local environment and attempt to determine the most likely scenario for their formation, including SN and AGN feedback, tidal forces, mergers and ram-pressure stripping.

In Chapter 7, we demonstrate an unsupervised machine learning (UML) technique applied to Ultra-deep Hyper Suprime-Cam data. The technique works by grouping pixels that have similar properties (and objects built from those pixels, like galaxies). Galaxies with similar morphologies are efficiently grouped together, so that a large galaxy sample can be reduced to small morphological groups which can then analysed via visual inspection. Our study demonstrates the power of UML techniques in performing accurate and efficient morphological analysis, which will become indispensable tool in the era of deep-wide surveys.

Finally, in Chapters 8 and 9 we present our conclusions and lay out plans for future work.

## Chapter 2

# The Progenitors of present-day early-type galaxies

### 2.1 Introduction

In the standard  $\Lambda$ CDM paradigm, galaxy formation proceeds hierarchically. Dark matter halos, which arise as a result of primordial fluctuations in the initial matter density field (Starobinsky, 1982; Guth & Pi, 1982; Hawking, 1982), merge to form progressively more massive haloes over cosmic time (e.g. Blumenthal et al., 1984; Kauffmann et al., 1993; Somerville & Primack, 1999). Cold gas condenses into these halos where it forms rotationally-supported discs (Franx & van Dokkum, 1996). The rate of star-formation is determined by the local density of this cold gas (Kennicutt, 1998), with feedback from supernovae (Scannapieco et al., 2008) and active galactic nuclei (AGN) (e.g. Silk & Rees, 1998; Kaviraj et al., 2016) regulating the process of stellar mass growth. A consequence of this paradigm is that the stellar mass of an individual galaxy is assembled through a combination of in-situ star-formation i.e. by gas turning into stars within a galaxy’s own halo, and ex-situ star-formation i.e. stars formed in another halo which have become members of the halo in question as a result of merging (e.g. Kauffmann et al., 1993). As ‘end-points’ of this hierarchical assembly process, local ‘early-type’ (i.e. elliptical and S0) galaxies are a particularly significant class of objects (e.g. Kaviraj et al., 2007). These galaxies dominate the stellar mass density in today’s Universe (e.g. Bernardi et al., 2003) and thus encode, in their stellar populations, the signatures of galaxy mass assembly over cosmic time (e.g. Worthey, 1994; van Dokkum & Franx, 2001). Studying these galaxies offers unique insights into the build up of the observable Universe (e.g. Barrientos & Lilly, 2003; Longhetti et al., 2005; McDermid et al., 2015) and significant effort in the literature has, therefore, been dedicated to understanding these systems.

Galaxies in the early universe form with disc-like (late-type) morphologies and, through interactions and secular processes, acquire more spheroidal (early-type) morphologies over time (e.g. Franx & van Dokkum, 1996; van Dokkum & Franx, 2001; Buitrago et al., 2014; Conselice et al., 2014; Kaviraj,

2014b). Consequently, at progressively higher redshift, the progenitors of today's early-types become increasingly dominated by late-type systems. Understanding the formation and evolution of today's early-types therefore requires us to consider their entire progenitor population, especially their late-type progenitors at earlier epochs. If, as is often assumed, early-type galaxies do not revert back to late-types (e.g. Hau et al., 2008; Fang et al., 2012)<sup>1</sup>, techniques that identify *late-type* galaxies that are progenitors of today's early-types in observational surveys are essential. This becomes particularly important at high redshift, e.g.  $z > 2$ , the redshift regime at which the early-type population is rapidly assembled (e.g. Conselice et al., 2014), and which can be routinely accessed by forthcoming facilities such as JWST (Gardner et al., 2006), EUCLID (Laureijs et al., 2011), etc.

Past observational studies (e.g. Gladders et al., 1998; Stanford et al., 1998) have attempted to trace the assembly of present-day early-types by focussing only on the population of early-type galaxies at high-redshift. Since the stellar populations of present-day early-types are largely in place in the early Universe (e.g. Trager et al., 2000; Thomas et al., 2005; Kaviraj et al., 2011; Kaviraj, 2014b), this is a reasonable approximation at low and intermediate redshift ( $z \sim 1$ ), assuming that galaxies, once they achieve early-type morphology, cannot revert back to being late-type systems. However, as noted above, at earlier epochs, an increasing proportion of the mass in present-day early-types is contained in late-type progenitors. Thus, considering only early-type galaxies introduces a bias (the so-called 'progenitor bias') in any study of their evolution, which becomes progressively more severe with increasing redshift.

In a similar vein, other observational work (e.g. Bell et al., 2004) has often used the optical red sequence (Faber et al., 2007; Barro et al., 2013) as a proxy for the population of progenitors of early-type galaxies. However, since the stellar populations in late-type galaxies tend to be younger (and therefore bluer), such a colour cut misses the majority of the late-type galaxies that are progenitors of local early-types. In addition, since a wide variety of star-formation histories are observed in early-type galaxies themselves, particularly at high redshift (Kaviraj et al., 2013a; Fitzpatrick & Graves, 2015; Lofthouse et al., 2017a), a large fraction of blue galaxies that already have early-type morphology will also be missed (Shankar et al., 2015), if such a colour cut is employed. While the red sequence traces the progenitors of early-type galaxies well at the highest end of the luminosity function (Kaviraj et al., 2009), it becomes less reliable at low masses (where galaxies are bluer) and fails to identify early and late-type progenitors (of all stellar masses) that lie blueward of the red sequence. In a general sense, therefore, a simple colour cut is not a reliable approximation for the progenitor population of today's early-type galaxies. And, in a similar vein to using an early-type selection to identify the progenitor population, the red-sequence approximation becomes progressively less effective with increasing redshift.

Progenitor bias is difficult to overcome observationally, since individual galaxies cannot be observed as they evolve and, therefore, in any given survey, it is difficult to directly identify objects that will end up in early-type galaxies at  $z \sim 0$ . Nevertheless, observational methods have been applied to reduce

<sup>1</sup>Dubois et al. (2016) have shown that transformation from early-type to late-type morphology is prevented by AGN feedback.

or mitigate this issue. For example, a widely-used technique, proposed by van Dokkum et al. (2010) assumes that cumulative co-moving number density is conserved as the galaxy mass function evolves with time (i.e. as a galaxy evolves, the number of galaxies more massive than it remains constant), so that a given galaxy maintains the same rank. However, while this method is able to account for the mass evolution of the galaxy population, the assumptions made may be too simplistic. For example, it ignores the fact that a galaxy's rank may change over time, resulting in the evolution of its co-moving number density. This may occur if, for instance, galaxies are removed from the population as a result of mergers (Ownsworth et al., 2014), or as a result of the spread in specific star-formation rate (sSFR) and its dependence on mass (Leja et al., 2013; Shankar et al., 2015).

Improvements on the method of van Dokkum et al. (2010) which include prescriptions for the number density evolution of galaxies have been proposed. For example, Behroozi et al. (2013) uses abundance matching in order to match observed galaxies to corresponding dark matter halos in  $\Lambda$ CDM simulations, allowing the median and dispersion of the cumulative co-moving number density tracks to be quantified for the progenitors of galaxies of a given mass. Work by Torrey et al. (2015, 2017) introduces an analytic framework, which accounts for the effects of the merger rate ('coagulation') and stochasticity of galaxy growth rates ('scatter rate'), and describes the median (Torrey et al., 2015) and intrinsic scatter (Torrey et al., 2017) of the evolution of the galaxy population in co-moving density space. Wellons & Torrey (2017) present a probabilistic method based on this framework, which they show is able to more effectively predict progenitor properties than the methods of van Dokkum et al. (2010) and Behroozi et al. (2013).

While this group of methods is widely applicable to any given galaxy property (e.g. Torrey et al., 2015; Clauwens et al., 2016), they may lose predictive power in cases where there is no expectation that the rank of the property of interest will be conserved (e.g. morphology). In such cases, leveraging the constraining power of additional galaxy properties becomes essential. Additionally, all of these methods retain a weakness of the van Dokkum et al. (2010) method, in that they still assume that early-type progenitors follow the same distribution of properties as the general population at a given redshift.

An appealing alternative is to employ a simulation that reproduces the properties of galaxies over cosmic time. Since the identities of the progenitors of local early-type galaxies are precisely known in the model, they can be used to calculate the probability that a galaxy is a progenitor, as a function of its observable properties (e.g. redshift, stellar mass, star-formation rate (SFR), local environment). Kaviraj et al. (2009) have previously addressed the problem of progenitor bias using the GalICS semi-analytical model (Hatton et al., 2003). However, while the semi-analytical approach has successfully reproduced the phenomenology of many aspects of the galaxy formation process (e.g. Somerville & Primack, 1999; Cole et al., 2000; Benson et al., 2003; Bower et al., 2006; Croton et al., 2006), the recent advent of hydrodynamical simulations in cosmological volumes provides a more accurate route to addressing the problem. Unlike their semi-analytical counterparts, hydrodynamical models resolve the gas and baryonic content of galaxies, typically on kpc scales, (e.g. Devriendt et al., 2010; Dubois et al., 2014; Vogelsberger

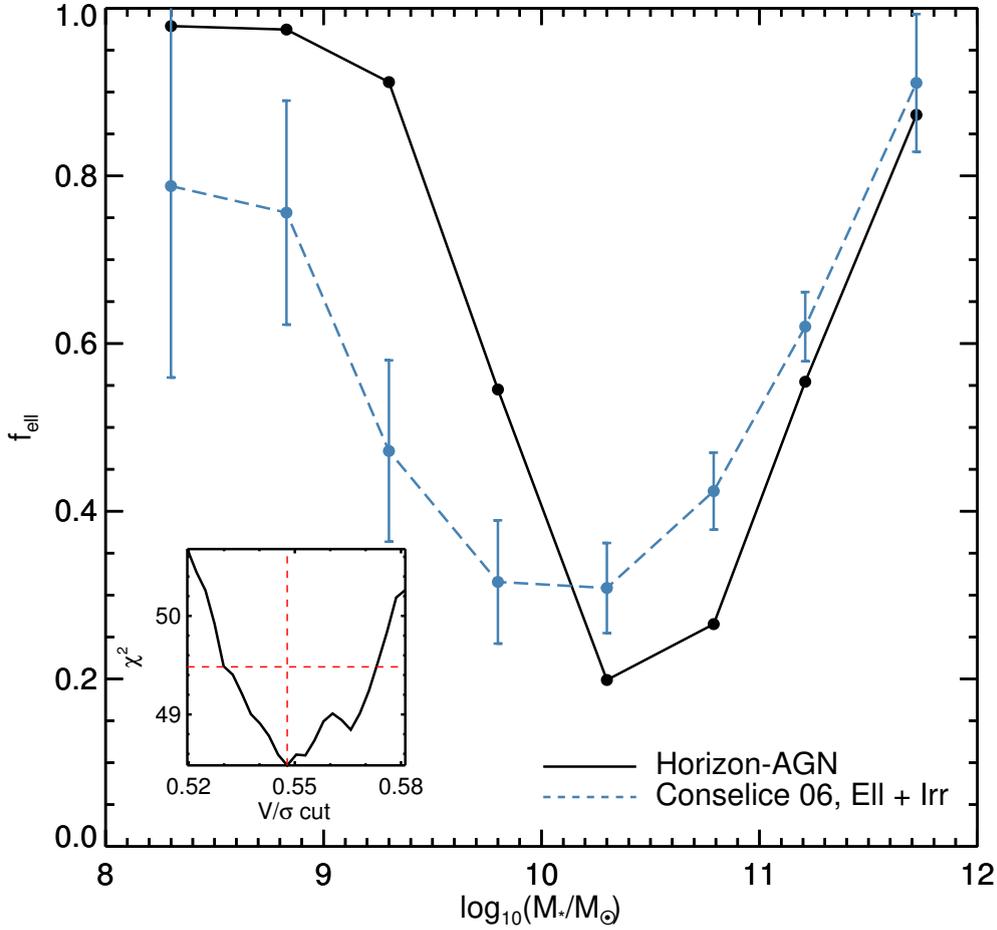


FIGURE 2.1: The predicted early-type fraction in Horizon-AGN, for the  $V/\sigma$  value ( $0.55 \pm_{0.02}^{0.03}$ ) that best reproduces the observed early-type fractions in the local Universe (Conselice, 2006). The inset shows the value of  $\chi^2$  between the predicted and observed early-type fractions for different  $V/\sigma$  thresholds (the red dashed lines show the minimum  $\chi^2$  and  $\Delta\chi^2 = 1$ ).

et al., 2014; Schaye et al., 2015; Khandai et al., 2015; Taylor & Kobayashi, 2016; Kaviraj et al., 2016). This enables them to more accurately model a greater range of physical processes, without the need for semi-analytical recipes (Schaye et al., 2015), although some processes, such as AGN feedback, that cannot be resolved, must still be described using sub-grid models (e.g. Katz, 1992; Booth & Schaye, 2009a; Kimm et al., 2015). Such hydrodynamical simulations typically rely on a smaller number of free parameters and offer a more realistic treatment of the physical processes involved in the formation and evolution of galaxies, yielding better agreement with the observed Universe without the need for tuning.

In this study, we use the Horizon-AGN<sup>2</sup> cosmological simulation (Dubois et al., 2014; Kaviraj et al., 2016) to (1) quantify the evolution of the progenitor population of today’s early-type galaxies and (2)

<sup>2</sup><http://www.horizon-simulation.org>

provide a route for identifying late-type galaxies that are progenitors of present-day early-types in observational surveys, by estimating the probability of a given late-type to be the progenitor of a local early-type system, as a function of measurable observables like redshift, stellar mass, star-formation rate and local density.

The structure of this chapter is as follows. In Section 2.2, we describe the main characteristics of the simulation, describe our simulated galaxy sample and define how observables are measured. In Section 2.3, we probe the redshift evolution of the progenitors of present-day early-type galaxies. In Section 2.4, we present probabilistic prescriptions to identify late-type galaxies that are progenitors of local early-types as a function of redshift, stellar mass, environment and star-formation rate. We summarize our findings in Section 2.5.

## 2.2 The simulation

We begin by briefly describing the Horizon-AGN simulation, the extraction of galaxies and merger trees and the prediction of observable quantities.

### 2.2.1 Horizon-AGN

Horizon-AGN is a cosmological hydrodynamical simulation (Dubois et al., 2014) that employs the adaptive mesh refinement Eulerian hydrodynamics code, RAMSES (Teyssier, 2002). The size of the simulation box is  $100 h^{-1} \text{coMpc}$ , which contains  $1024^3$  dark matter particles and uses initial conditions from a *WMAP7*  $\Lambda$ CDM cosmology (Komatsu et al., 2011). The simulation has a dark matter mass resolution of  $8 \times 10^7 M_{\odot}$ , a stellar-mass resolution of  $2 \times 10^6 M_{\odot}$  and a spatial resolution of  $\sim 1 \text{ kpc}$ . We direct readers to Kaviraj et al. (2016) for details of the recipes (e.g. star-formation and stellar and AGN feedback) employed to model the baryonic evolution of galaxies. Briefly, star-formation follows a standard Schmidt-Kennicutt law (Kennicutt, 1998), with the model implementing continuous stellar feedback, that includes momentum, mechanical energy and metals from stellar winds, Type II SNe and Type Ia SNe. Black-hole (BH) feedback on ambient gas operates via two separate channels, depending on the gas accretion rate. For Eddington ratios  $> 0.01$  (high accretion rates), 1.5 per cent of the accretion energy is injected as thermal energy (a quasar-like feedback mode), whilst for Eddington ratios  $< 0.01$  (low accretion rates), bipolar jets are employed with a 10 per cent efficiency. The parameters are chosen to produce agreement with the local cosmic black-hole mass density, and the  $M_{\text{BH}} - M_{\star}$  and  $M_{\text{BH}} - \sigma_{\star}$  relations (Dubois et al., 2012). Apart from choosing the BH-feedback parameters to match the  $M_{\text{BH}} - M_{\star}$  and  $M_{\text{BH}} - \sigma_{\star}$  relations at  $z = 0$ , Horizon-AGN is not otherwise tuned to reproduce the bulk properties of galaxies at  $z \sim 0$ . As described in Kaviraj et al. (2016), the simulation reproduces key quantities

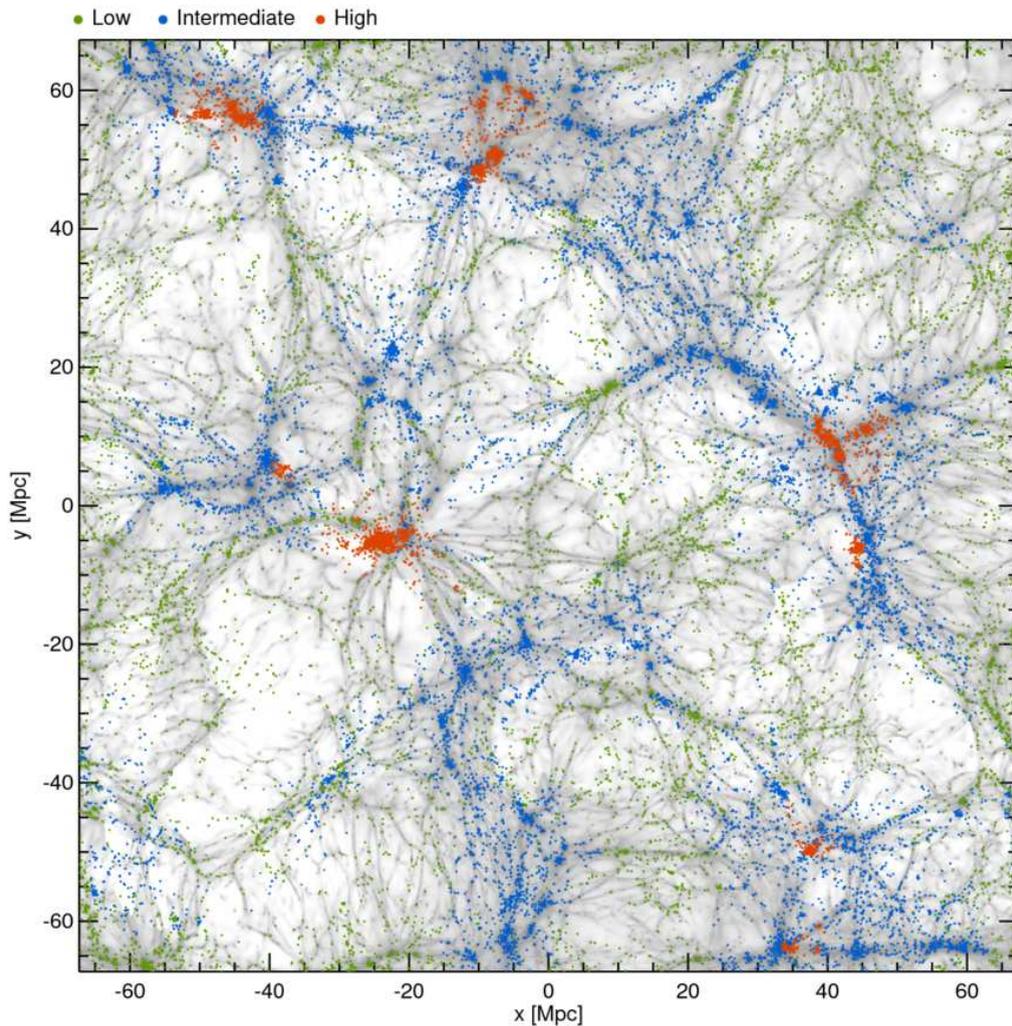


FIGURE 2.2: Greyscale map showing the projected gas density in a  $\sim 20$  Mpc slice through the simulation volume at the final time output ( $z = 0.06$ ). Green, blue and red points correspond to the positions of galaxies ( $M_* > 10^{9.5} M_\odot$ ) in low, intermediate and high density environments respectively (see text in Section 2.2.3.2 for more details). Axes are in units of proper Mpc.

that trace the aggregate stellar mass growth of galaxies: stellar mass and luminosity functions, rest-frame UV-optical-near infrared colours, the star-formation main sequence and the cosmic star-formation history.

## 2.2.2 Identifying galaxies and building merger trees

In order to track their progenitors, we build merger histories for each early-type galaxy in the final snapshot of the simulation ( $z = 0.06$ ). In the sections below, we describe the process of galaxy identification,

followed by the process of building merger trees.

### 2.2.2.1 Identifying the galaxy sample

For each snapshot, we produce a catalogue of galaxies, using the ADAPTAHOP structure finder (Aubert et al., 2004; Tweed et al., 2009), operating directly on the star particles. Galactic structures are selected using a local threshold of 178 times the average matter density, where the local density of individual particles is calculated using the 20 nearest neighbours. Only galactic structures with more than 100 star particles are considered, corresponding to a minimum galaxy stellar mass of  $M_{\star} \sim 2 \times 10^8 M_{\odot}$ , which is a consequence of a minimum star particle mass of  $m_{\star} \sim 2 \times 10^6 M_{\odot}$ . We identify an average of  $\sim 150,000$  galaxies above the 100 particle threshold in each snapshot. We restrict our study to galaxies with stellar mass  $M_{\star} > 10^{9.5} M_{\odot}$ , for reasons outlined in Section 2.2.2.2.

### 2.2.2.2 Producing merger histories

Using the catalogue of galaxies identified by ADAPTAHOP, we extract merger histories for each early-type galaxy at the final snapshot. We produce merger trees using 91 snapshots in the range  $z \in [0.06, 7.04]$ , with an average time-step of  $\sim 130$  Myr. Merger trees are produced for each early-type, by identifying their progenitors at each snapshot, using the method described in Tweed et al. (2009). Since our threshold for identifying structures is 100 star particles, only mergers where the satellite galaxy has  $M_{\star} \gtrsim 2 \times 10^8 M_{\odot}$  are considered, regardless of mass ratio. Given that our sample excludes galaxies less massive than  $M_{\star} \sim 10^{9.5} M_{\odot}$ , and the minimum galaxy mass identified is  $M_{\star} \sim 2 \times 10^8 M_{\odot}$ , our sample is complete for mergers down to a mass ratio of at least 1:15.

## 2.2.3 Prediction of observables

We produce observables that can be used in conjunction with contemporary and future observational datasets. These are stellar mass (derived using the total mass of the star particles in a galaxy), star-formation rate, local number density and stellar kinematics (which we use as a proxy for morphology). The following sections describe how we derive each of these measures.

### 2.2.3.1 Morphology

The morphology of each model galaxy in our analysis is inferred using its stellar kinematics. Morphology is defined using  $v/\sigma$ , which is the ratio of the mean rotational velocity ( $V$ ) to the mean velocity dispersion ( $\sigma$ ), both measured using the entire star particle distribution. Higher values of this ratio correspond to more late-type (disc-like) morphologies.  $v/\sigma$  is calculated by first rotating the coordinate

system so that the  $z$ -axis is oriented along the angular momentum vector of the galaxy. Rotational velocity is defined as the mean tangential velocity component in cylindrical co-ordinates,  $V_\theta$ , and velocity dispersion is computed using the standard deviations of the radial, tangential and vertical star particle velocities,  $\sigma_r$ ,  $\sigma_\theta$  and  $\sigma_z$ , summed in quadrature.  $V/\sigma$  is then given by

$$V/\sigma = \frac{\sqrt{3}\bar{V}_\theta}{\sqrt{\sigma_r^2 + \sigma_\theta^2 + \sigma_z^2}}. \quad (2.1)$$

To separate galaxies morphologically into early and late types using  $V/\sigma$ , we consider a range of values for  $V/\sigma$  and compare how the resulting predicted early-type fractions compare to their observed counterparts at low redshift. The value of  $V/\sigma$  which produces the best agreement with the observational data is then selected as the threshold value that separates early-types and late-types in the model. Fig 2.1 shows the predicted early-type fractions in Horizon-AGN for this best-fitting  $V/\sigma$  value,  $0.55 \pm_{0.02}^{0.03}$ , compared with early-type fractions derived from observations (Conselice, 2006). The largest discrepancy between the observations and the simulation occurs at the low mass end, and is likely a result of insufficient mass resolution (Dubois et al., 2016). Nevertheless, over the mass range considered in this study ( $M_\star > 10^{9.5} M_\odot$ ), the early-type fractions predicted by Horizon-AGN, for our  $V/\sigma$  threshold of 0.55, are in reasonable agreement with the observations. We note that the progenitor fractions presented in Section 2.4 are resistant even to relatively large changes in our  $V/\sigma$  threshold. Varying the  $V/\sigma$  threshold by as much as 50 per cent introduces only a  $\pm 0.05$  variation in our calculated progenitor fractions.

We note that the minimum refinement of the AMR grid is increased at  $z = 4, 1.5$  and  $0.25$ , in order to keep the minimum physical cell size approximately constant (Dubois et al., 2014; Peirani et al., 2016). While it is possible that this refinement may result in the production of sudden instabilities in previously stable discs that increase galaxy velocity dispersions, the smooth nature of the  $V/\sigma$  evolution of the galaxy population (Dubois et al., 2014), indicates that this is not a significant effect.

### 2.2.3.2 Local environment

We compute the local environment of each galaxy using an estimate of the 3-D local number density. For each galaxy, we estimate the surrounding number density of galaxies above our mass cut ( $10^{9.5} M_\odot$ ). This is achieved using the adaptive kernel density estimation method of Breiman et al. (1977) which utilises a finite-support Epanechnikov kernel (instead of the more typically-used infinite-support Gaussian kernel) and using the density estimator itself as a pilot to steer the local kernel size (Wilkinson & Meijer, 1995; Ferdosi et al., 2011). Galaxies at our final snapshot ( $z = 0.06$ ) are classified into ‘high’, ‘intermediate’ and ‘low’ density environments, by comparing the density percentile they occupy to the corresponding percentiles that observed galaxies inhabiting cluster, group and field environments occupy in the low-redshift universe.

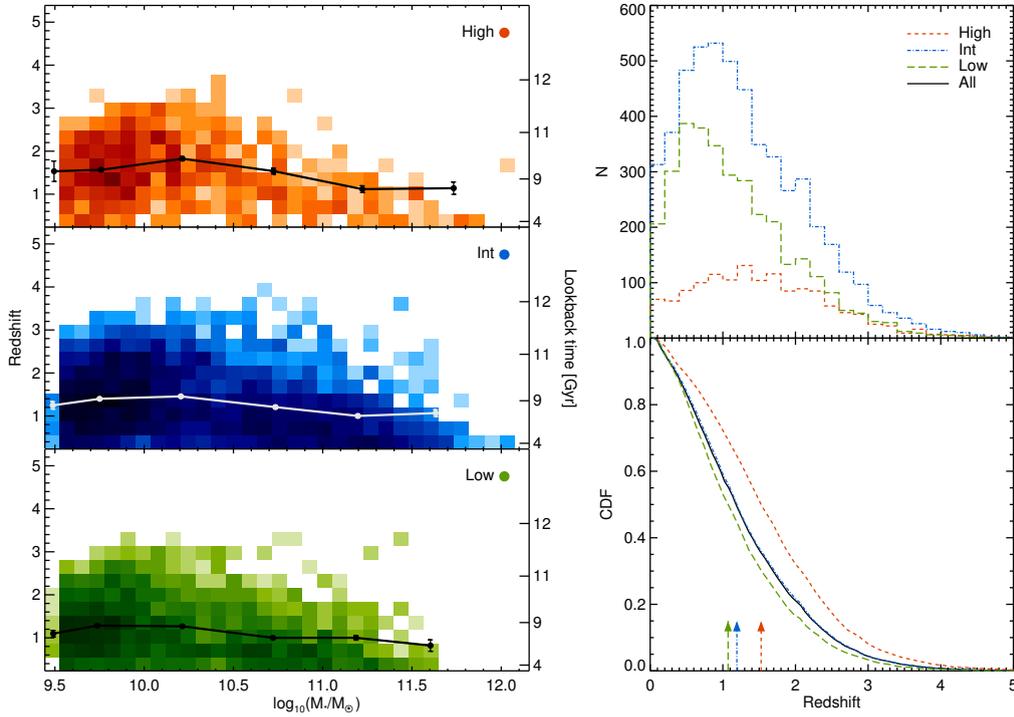


FIGURE 2.3: Left: Density plots showing the redshift at which local early-type galaxies in different environments had their last significant merger (i.e. a merger with a mass ratio greater than 1:10). Green, blue and red colours correspond to low, intermediate and high density environments. Over-plotted are lines showing the mean redshift in each bin and its error. Right: Histograms (top) and associated cumulative distribution functions (bottom) of last merger redshifts. The arrows in the bottom panel indicate the redshift where the cumulative distribution functions reach values of 0.5 (i.e the median value) for each environment.

Observations indicate that around 10 per cent of galaxies occupy rich clusters, while around 50 per cent of galaxies occupy groups and poor clusters (e.g Bahcall, 1996; Dekel & Ostriker, 1999; Tempel et al., 2012), with the remaining 40 per cent of galaxies occupying the field. We use these values as a guide to make density cuts, in order to separate our simulated galaxies into our three density bins (high, intermediate and low), that are roughly analogous to cluster, group and field environments. To allocate our simulated galaxies into these density bins, we first rank the objects by density. Then, the 10 per cent of galaxies with the highest local density (i.e. the 90<sup>th</sup> – 100<sup>th</sup> percentile range) are classified as occupying high density environments, galaxies in the 40<sup>th</sup> – 90<sup>th</sup> range are classified as occupying intermediate density and the remaining galaxies are classified as occupying low-density environments. Fig 2.2 shows the environment classifications of galaxies with  $M_{\star} > 10^{9.5} M_{\odot}$  superimposed over a map of the gas density. Not unexpectedly, galaxies classified as being in high density environments lie at the nodes of the cosmic web, while intermediate density galaxies lie largely along filaments, with low-density galaxies typically being found in voids. Regions of high gas density also host a high number density of galaxies which, as we show later in Section 2.3, leads to more rapid morphological transformation. More massive galaxies

(those above our detection threshold of  $2 \times 10^8 M_{\odot}$ ) first appear in the simulation around nodes, owing to a greater abundance of gas from which to accrete, and also undergo a higher incidence of interactions or mergers. These galaxies are thus likely to experience earlier and accelerated evolution compared to galaxies in less dense environments.

In all of our analysis below, we always cast the local environment in terms of the density percentile of individual galaxies. This is because observers (and theorists) inevitably use different metrics for measuring local density. However, while the absolute values of density depends on the actual metric being used, the density percentile that a galaxy occupies is likely to be roughly independent of the actual estimation method (as we demonstrate later in Section 2.4).

### 2.2.3.3 Star-formation rate

Star-formation rates are calculated by computing the change in stellar mass,  $m_{\star}$ , of the galaxy in question between two snapshots, dividing by the time difference between those snapshots, and subtracting the mass of stars formed *ex-situ* that has merged with the galaxy between the two snapshots,  $m_{\star,merged}$ :

$$\text{SFR} = \frac{m_{\star,t=t_i} - m_{\star,t=t_{i-1}} - m_{\star,merged}}{\Delta t}, \quad (2.2)$$

where  $t_i$  is the time at the current snapshot and  $t_{i-1}$  is the time at the previous snapshot.  $\Delta t$  is equal to  $t_i - t_{i-1}$  and, as discussed in Section 2.2.2.2, the average time-step used in the simulation is  $\sim 130$  Myr. We note that the SFR is not sensitive to the exact value of  $\Delta t$ . For example, if we double or halve the value  $\Delta t$ , in either case, our calculated SFRs change by less than 30 per cent.

## 2.3 Redshift evolution of the progenitors of local early-type galaxies

We begin our analysis by performing a broad exploration of the redshift evolution of the progenitors of local early-type galaxies in the simulation. The left-hand panel of Fig 2.3 presents density plots showing the last-merger redshift of each local early-type galaxy. We define this as a galaxy's last significant merger, i.e. a merger which has a mass ratio greater than 1:10. The right-hand panels of Fig 2.3 show histograms (top-right) and their associated cumulative density functions (bottom-right) as a function of redshift for each environment. The arrows in the bottom-right panel indicate the redshift where the cumulative distribution functions reach values of 0.5 (i.e. the median value) for each environment.

Local early-type galaxies that inhabit denser environments tend to have higher *dynamical ages* i.e. their last significant mergers take place at earlier epochs. While 50 per cent of galaxies in the high density bin have their last significant merger by  $z = 1.5$ , this is only the case for galaxies in the lowest density bin at  $z = 1.1$  (as indicated by the coloured arrows in the cumulative density function plot (bottom-right

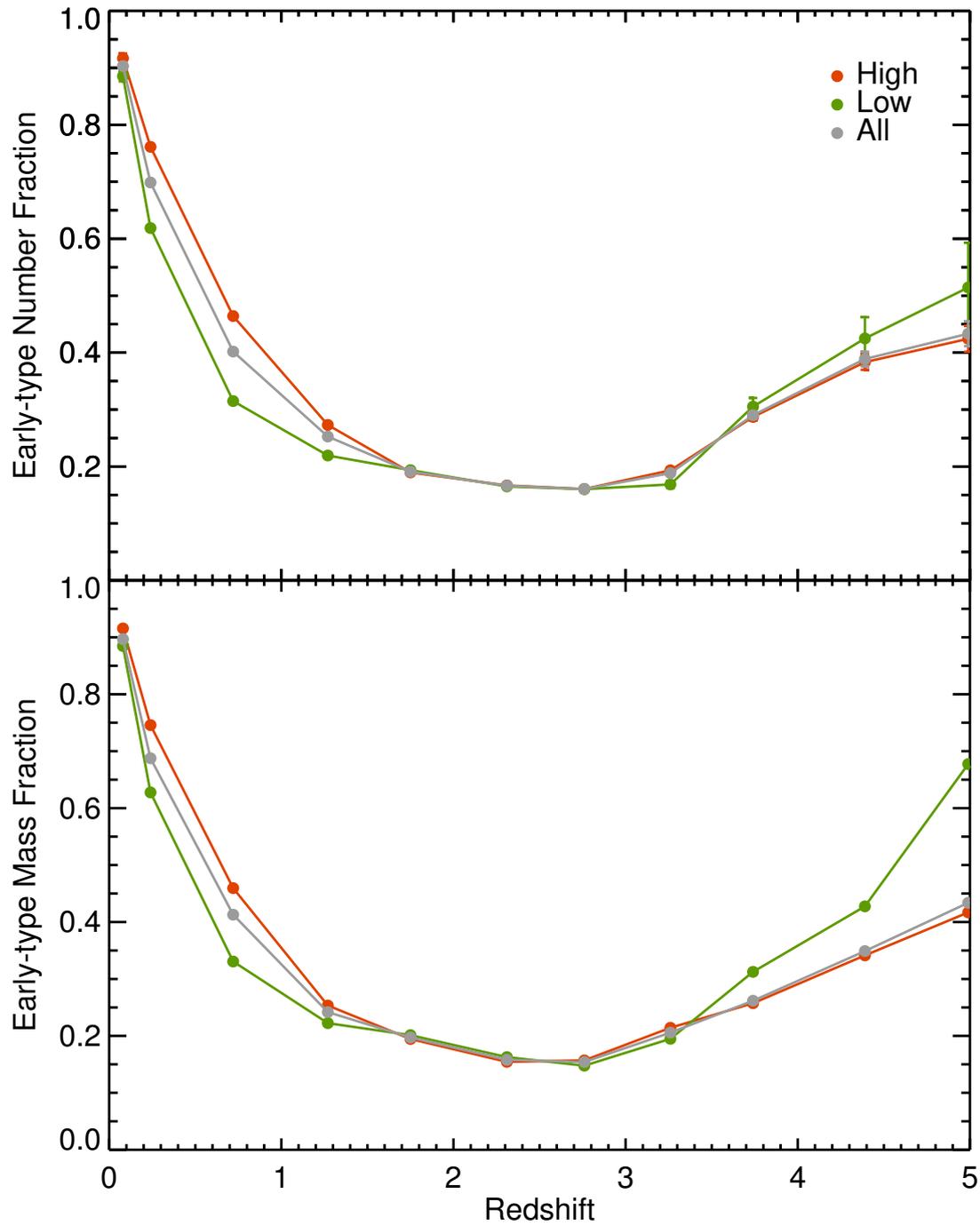


FIGURE 2.4: Top: The number fraction of progenitors that already have early-type morphology, split by the local environment of the early-type galaxy that the progenitor ends up in at the present day. Bottom: The mass fraction contained in progenitors that have early-type morphology, split by the environment of the early-type galaxy that the progenitor ends up in at the present day. Poisson error bars are shown (note that most lie within the data points).

panel in Fig 2.3)). Local early-types in higher-density environments also tend to have higher final masses indicating that the bulk of their evolution takes place earlier and is more rapid, although within a specific environment (high, intermediate or low), the most massive galaxies finish assembling at later epochs (see also De Lucia & Blaizot, 2007; Dubois et al., 2016).

In Fig 2.4, we quantify the extent of progenitor bias at various redshifts, as a function of galaxy mass and local environment. We show both the fraction of progenitor galaxies that have already acquired early-type morphology (top panel) and the mass fraction in the progenitor population that is contained in progenitors with early-type morphology (bottom panel). We find that, across all environments, only  $\sim 50$  per cent of the progenitors have acquired early-type morphology by  $z \sim 0.6$ . This is also true of the mass fraction contained in these early-type progenitors i.e. at  $z \sim 0.6$  only half of the stellar mass that eventually ends up in early-type galaxies today is contained in progenitors that have early type morphology. In other words, looking *only* at early-type systems to trace the evolution of today's early-types would miss half of the progenitor population at  $z \sim 0.6$ .

Since morphological transformation is more rapid in regions of higher density, progenitor bias is less severe. Thus, 50 per cent of the progenitors of local early-types in high-density environments (i.e. today's clusters) have already acquired early-type morphology by  $z = 0.7$  (compared to  $z = 0.6$  across all environments). Note, however, that the bias remains reasonably high regardless of environment. Very similar trends are seen when quantifying progenitor bias as a function of stellar mass (not shown in Fig 2.4), with more massive galaxies ( $M_* > 10^{11.5} M_\odot$ ) following the same trend as the 'high' density environment in Fig 2.4. This is simply because the most massive galaxies occur overwhelmingly in dense environments.

Finally, we note that the early-type fraction appears to decrease (somewhat counter-intuitively) from  $z = 5$  and begins to increase again around  $z = 3$ . This is not an artefact of our  $2 \times 10^8 M_\odot$  detection threshold, which might cause the most rapidly evolving (and more massive) galaxies to be detected first, potentially biasing our result. If we limit our study to narrow mass bins or follow only the evolution of galaxies that are detected by the structure finder at  $z \sim 5$ , we observe the same non-monotonic evolution. This is partially the result of generally more clumpy star formation and more disturbed morphologies at high redshift (e.g. Ceverino et al., 2010), and is consistent with observational work (e.g. Kassin et al., 2012, 2014), which has shown that star-forming galaxies steadily settle into flat, rotationally supported discs, through the process of 'disc settling' at these epochs. This kinematic settling is driven by the fact that many processes which are able to 'puff up' (i.e. increase the dispersional motion) of gas in the disc without disturbing stellar orbits significantly, become less frequent and/or less intense with time - examples of this include (minor) mergers and gas accretion episodes (e.g. Covington et al., 2010; Bournaud et al., 2011; Lofthouse et al., 2017b; Martin et al., 2017), strong stellar feedback as a result of the high star-formation rates at high redshift (e.g. Silk & Norman, 2009) and the high gas fractions at early epochs that lead to increased disc instability. In essence, the more gentle evolution that galaxies undergo at later times is thought to allow the gas in the disc to settle into a more ordered state (Kassin

et al., 2012). Star formation then proceeds primarily in a planar disc, gradually reducing the mean  $v/\sigma$  as more stars form. Indeed, for Horizon-AGN galaxies that still host a significant disc at  $z = 0.06$ , we find that old stars (those formed before  $z = 3$ ) are more likely to be found in orbits outside of the plane of the disc, symptomatic of the fact that, at these early epochs, gas fractions and merger rates were typically higher on average. Disc settling has also been observed in other simulations (e.g. Kassin et al., 2014; Ceverino et al., 2017). In the case of Horizon-AGN, this effect may be compounded slightly by changes to the maximum refinement of the gas grid, which, as mentioned in Section 2.2.3.1 is increased towards lower redshifts in order to keep the minimum physical cell size approximately constant. This can have the effect of artificially thickening discs of some galaxies, although we still observe non-monotonic evolution regardless of refinement level.

We conclude this section by exploring the morphologies of the progenitors of early-type galaxies that are involved in mergers. We focus only on binary mergers because, although non-binary mergers do occur, they are rare (around two orders of magnitude less frequent than binary mergers). In Fig 2.5, we show the fraction of mergers at a given redshift that involve two late-type galaxies ('late-late'), one late-type and one early-type galaxy ('mixed') and two early-type galaxies ('early-early'). Mergers between two late-type galaxies dominate in the early Universe i.e. around the epoch of peak cosmic star-formation and beyond ( $z > 1.5$ ). The fraction of mergers involving two early-type galaxies climbs rapidly in the low redshift Universe ( $z < 0.5$ ). However, at all redshifts, the majority of mergers involve at least one late-type galaxy.

## 2.4 Identifying late-type galaxies that are progenitors of present day early-types

We proceed by constructing probabilistic prescriptions for identifying late-type progenitors of local early-type galaxies in observational surveys, as a function of quantities that are measurable in today's datasets: redshift, stellar mass, local environment and star-formation rate. As noted in the introduction, the overall aim is to provide a means for correcting progenitor bias in observational studies, by allowing for the inclusion of late-type progenitors of today's early-type systems. We do this by calculating the fraction of late-type galaxies that are progenitors of local early-types, as a function of the measurable quantities mentioned above. This fraction can then be thought of as a probability that a galaxy with the given properties is the progenitor of an early-type galaxy at present day. Observers who wish to include late-type progenitors of early-type galaxies can then use these probabilities to 'weight' objects in observational surveys, thus enabling them to reduce progenitor bias by including, in a probabilistic sense, the late-type members of the progenitor population. These probabilistic prescriptions are likely to be particularly useful in the new era of deep-wide surveys (e.g. DES (Abbott et al., 2005), EUCLID (Laureijs et al., 2011), LSST (Abell et al., 2009), JWST (Gardner et al., 2006) etc.) which will routinely offer large datasets that probe the early Universe, where progenitor bias becomes most severe,

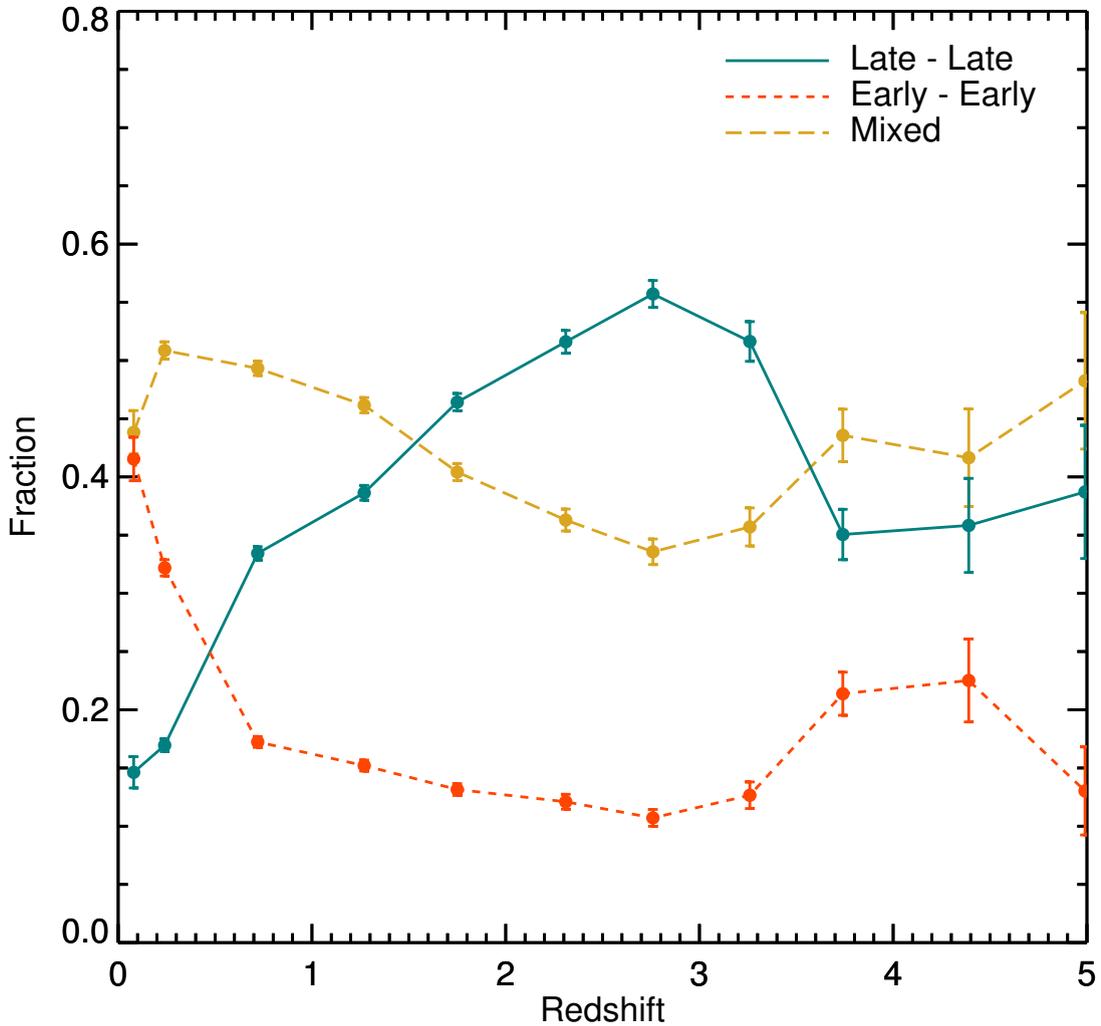


FIGURE 2.5: The morphological composition of binary mergers within the progenitor population, that have mass ratios greater than 1:10, as a function of redshift. ‘Late-late’ indicates mergers between two late-type galaxies, ‘mixed’ indicates mergers between one early-type and one late-type galaxy and ‘early-early’ indicates a merger between two early-type galaxies. Poisson errors are shown.

and simplifying assumptions, such as using only early-type galaxies or the red sequence to trace the progenitor population, break down. In what follows, we first explore 1-D progenitor probabilities as a function of stellar mass (split by local environment) and then 2-D probabilities as a function of mass and environment and mass and star-formation rate.

Since we are interested in probing progenitor probabilities as a function of local environment, we calculate, at each redshift of interest, the 3-D local number density using the method described in Section 2.2.3.2. As in Section 2.3, we consider galaxies in the 90<sup>th</sup> – 100<sup>th</sup> percentile range to be inhabiting high

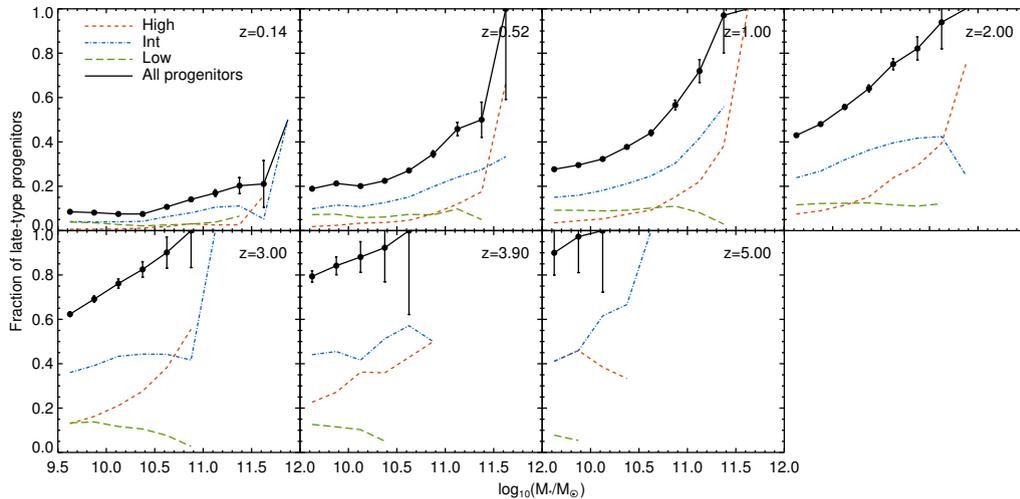


FIGURE 2.6: The evolution with redshift of the fraction of late-type galaxies that are progenitors of early-type galaxies at  $z \sim 0$ . Each panel shows, in black, the fraction of late-type galaxies that are progenitors of local early-types for a given redshift, as a function of the stellar mass of the progenitor. The sample of late-type galaxies is split further into high (red), intermediate (blue) and low (green) density environments. Error bars show Poisson errors. The sample becomes smaller (leading to a corresponding increase in the size of the Poisson error bars) towards higher redshifts, because there are fewer galaxies with stellar masses above  $M_* > 10^{9.5}$  in the simulation. For clarity, errors bars and points are not shown where the errors are larger than 0.5.

density environments, those in the 40<sup>th</sup> – 90<sup>th</sup> percentiles to be inhabiting intermediate-density environments and those in the 0<sup>th</sup> – 40<sup>th</sup> percentile range to be inhabiting low density environments. As noted before, the density percentile in which a galaxy lies (which is driven by its rank in density) is likely to be reasonably resistant to the exact method used for the density estimation.

To check this, we compare two different density estimation methods. These are the adaptive kernel density estimator used in Section 2.2.3.2 and the  $k^{\text{th}}$  nearest neighbours density estimator, that is commonly used in many observational studies (e.g. Dressler, 1980; Baldry et al., 2006; Ferdosi et al., 2011; Shattow et al., 2013). Note that for consistency with the adaptive kernel method, we choose a definition for the  $k^{\text{th}}$  nearest neighbour algorithm whereby each galaxy is considered to be its own neighbour. Specifically, we choose,  $k = 6$ , which is almost equivalent<sup>3</sup> to the commonly used case where  $k = 5$  and each galaxy is not considered to be its own neighbour (e.g. Baldry et al., 2006). We test the two methods on a 7 Mpc (proper) slice through the simulation snapshot at  $z = 0.5$ , which corresponds to a difference in velocity due to the Hubble flow of  $\Delta V = 500 \text{ km s}^{-1}$ , and implies a requisite precision in redshift of  $\Delta z = 0.002$ . Such precision will be achievable at intermediate and high redshift using spectroscopic and grism redshifts from future instruments, such as MOONS (Cirasuolo et al., 2011), PFS (Takada et al., 2014), 4MOST (de Jong et al., 2012) and JWST (Gardner et al., 2006). We explore estimates of both

<sup>3</sup>i.e. both methods calculate the density within the same radius, but the number of objects differs by 1, meaning the  $k = 5$  (not-own-neighbour) number density estimate simply differs by a factor of 5/6 from the  $k = 6$  (own-neighbour) estimate.

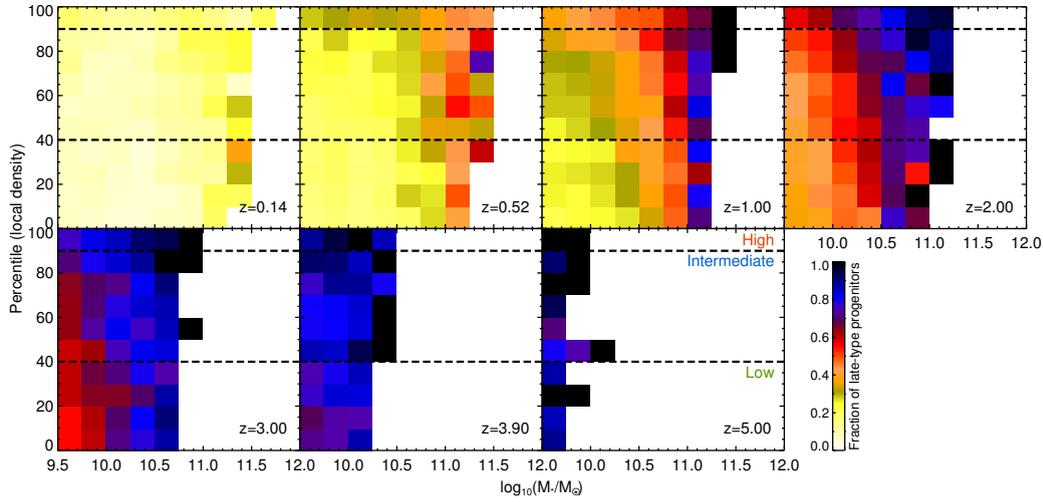


FIGURE 2.7: Density plots showing the redshift evolution of the fraction of late-type galaxies that are progenitors of early-types at the present day, as a function of the stellar mass and density percentile of the late-type galaxies in question. The late-type progenitor fraction is represented by the colour bar. We do not plot bins containing 3 or fewer galaxies.

the 2-D surface density and the 3-D density. We find that the rank of each galaxy indeed remains approximately constant, regardless of either the exact estimator used, or whether we consider the 2-D or 3-D densities. Typically, the rank of a galaxy does not change by more than 10 per cent and, therefore, changing the density estimator leaves our conclusions unchanged.

Fig 2.6 shows the fraction of late-type galaxies at a given redshift that are the progenitors of a local early-type, as a function of stellar mass and split by local environment. We show the 2-D progenitor probability as a function of both stellar mass and local density in Fig 2.7, with the colour bar indicating the progenitor probabilities. At all redshifts, there is a positive trend of progenitor probability with stellar mass i.e. more massive late-type galaxies are more likely to be progenitors of local early-type remnants.

At high redshifts, almost all massive galaxies, regardless of their local environment, are progenitors of present-day early-types. While the progenitor probabilities increase with redshift, for the most massive galaxies the progenitor probability remains close to  $\sim 1$  until  $z \sim 0.5$ . The principal reason for an increase in the progenitor fraction with redshift is simply the fact that late-type galaxies have more time to merge with other galaxies and undergo morphological transformation before the present day. The rate of morphological transformations is regulated by the merger rate per galaxy, which rises with redshift (e.g. Welker et al., 2015; Rodriguez-Gomez et al., 2015; Kaviraj et al., 2015b) and thus controls the rate of change in the progenitor fraction as a function of redshift. Lower-mass ( $M_{\star} < 10^{10.5} M_{\odot}$ ) galaxies can also exhibit high progenitor probabilities at high redshift, but *only* if they occupy regions of high density (e.g. the 80-100<sup>th</sup> density percentile, see Fig 2.7). Note that the progenitor probabilities decline for all galaxies towards low redshifts, because these systems will not have had time to undergo enough merging to achieve early-type morphology.

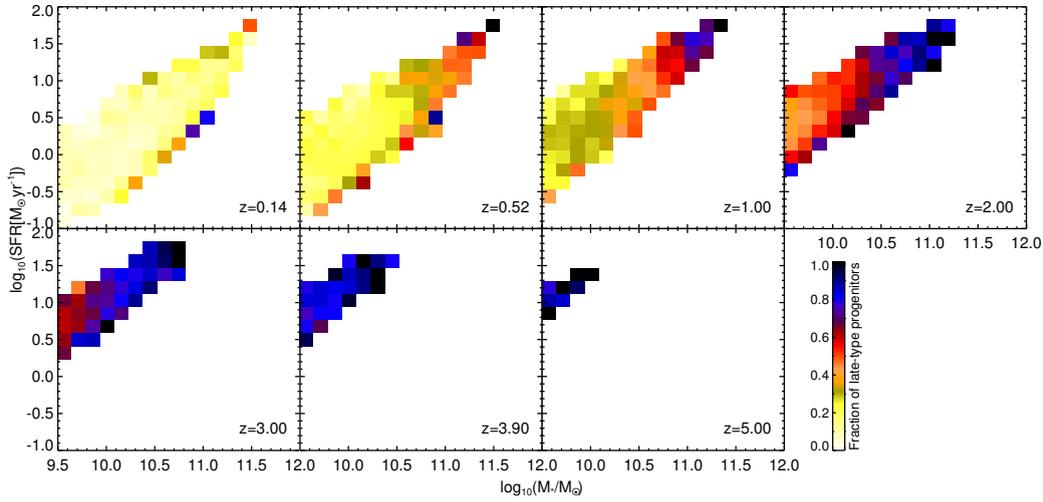


FIGURE 2.8: Left: Star-formation main sequence (i.e. star-formation rate plotted against stellar mass) for late-type galaxies, as a function of redshift, colour-coded by the fraction of late-type galaxies that are progenitors of early-types at the present day. We do not plot bins containing 3 or fewer galaxies.

We proceed, in Fig 2.8, by exploring the progenitor probabilities as a function of stellar mass and star-formation rate (the parameter space that is commonly referred to as the ‘star-formation main sequence’). The colour bar indicates the progenitor probabilities. Mirroring the trends found earlier, massive late-type galaxies are more likely to be progenitors of early-types. At high redshift, these massive late-type progenitors can be some of the most highly star-forming systems in the Universe. However, this is simply a consequence of star-formation activity being, on average, more vigorous in the early Universe. It is worth noting, however, that at all redshifts, progenitor fractions increase *at fixed stellar mass* as the star-formation rate decreases. In other words, at fixed stellar mass, late-type galaxies with lower specific star-formation rates are more likely to be progenitors. For example, at  $z \sim 2$  (top-right hand panel of Fig 2.8), a late-type galaxy with a stellar mass of  $10^{10.7} M_{\odot}$ , which resides at the upper end of the star-formation main sequence, has a progenitor probability of around 70 per cent. A galaxy with a similar mass which sits at the bottom of the star-formation main sequence has a progenitor probability of close to 100 per cent.

We release the progenitor probabilities calculated here, as a function of different observables, as described in the Appendix A.1. For a series of redshifts, we provide tables of progenitor probabilities as a function of the three principal observables studied here: stellar mass, star-formation rate and local environment. Since not all observational datasets may offer access to all three quantities, we also provide separate tables for projections of this 3-D parameter space i.e. progenitor probabilities as a function of stellar mass only, stellar mass and local environment and stellar mass and star-formation rate. As noted above, the properties of individual galaxies in current and future observational surveys which provide these observables can be compared to these tables in order to estimate the probability that they are progenitors of an early-type galaxy in the local Universe. The structure of the files containing these tables, and scripts to read them, are provided in the Appendix A.1.

## 2.5 Summary

As end-points of the hierarchical mass assembly process, early-type galaxies host more than half of the stellar mass density in the local Universe, their stellar populations encoding the assembly history of galaxies over cosmic time. Studying these galaxies in the local Universe and probing their progenitors at earlier epochs offers a unique perspective on the evolution of the observable Universe. However, since morphological transformations progressively convert late-type (disc-like) galaxies into early-type systems, the progenitors of today’s early-type galaxies become increasingly dominated by late-types at high redshift. Understanding the evolution of early-types over cosmic time therefore requires a reliable method for identifying these late-type progenitors of local early-types. Here, we have used the Horizon-AGN cosmological hydrodynamical simulation, which produces good agreement with the observed properties of galaxies in the redshift range  $0 < z < 5$ , to study how the progenitors of local early-type galaxies evolve over cosmic time.

We have studied the merger histories of local early-types and the morphologies of galaxies that are involved in these mergers and traced how the morphological mix of galaxies in the progenitor population changes over cosmic time. We have then used the simulation to study the fraction of late-type galaxies that are progenitors of present-day early-types, as a function of redshift, stellar mass, local environment and star-formation rate: observables that can be routinely measured in current and future datasets. As noted earlier, these fractions can be treated as probabilities that can then be used to include late-type progenitors of local early-types in observational surveys by ‘weighting’ these late-type systems by these probabilities. The benefit of this approach is to alleviate progenitor bias i.e. the bias that occurs if one considers only early-type galaxies (or proxies like the red sequence) to study the progenitor population of today’s early-types. Our key conclusions are as follows:

- The merger history of early-type galaxies indicates that these systems finish assembling their stellar mass at relatively early epochs. By  $z \sim 1$ , around 60 per cent of today’s massive early-types, averaged over all environments, have had their last significant merger (i.e. a merger with mass ratio greater than 1:10). For early-type galaxies that inhabit high density environments at the present day (e.g. clusters) this value is 70 per cent, while it is  $\sim 50$  per cent in early-types that inhabit low-density environments (e.g. the field). On average, morphological transformation is  $\sim 50$  per cent faster in high-density environments compared to low-density regions.
- Progenitor bias is significant at all but the lowest redshifts. Until  $z \sim 0.6$  less than half of the progenitors of today’s early-types actually have early-type morphology. Similarly, less than half of the stellar mass that ends up in an early-type today is actually hosted by a progenitor that has early-type morphology at this redshift. Around the epoch of peak cosmic star-formation, which is also the epoch at which morphological transformation occurs most rapidly, studying only early-type galaxies misses almost all (at least 80 per cent) of the stellar mass that eventually ends up in

early-types at the present day.

- The morphological mix of progenitor galaxies that are involved in mergers evolves over time. At all redshifts, the majority of mergers have at least one late-type progenitor. Mergers between two late-type galaxies dominate at early times i.e. around the epoch of peak cosmic star-formation and beyond ( $z > 1.5$ ) and the fraction of mergers involving two early-type galaxies climbs rapidly at low redshift ( $z < 0.5$ ).
- At all redshifts, late-type galaxies with larger stellar masses are more likely to be progenitors of local early-type remnants. At high redshifts, almost all massive ( $M_{\star} < 10^{11} M_{\odot}$ ) late-type galaxies, regardless of their local environment, are progenitors of present-day early-type galaxies. While the progenitor probabilities increase with redshift, for these massive galaxies, the progenitor probability remains close to  $\sim 1$  until  $z \sim 0.5$ . Lower-mass ( $M_{\star} < 10^{10.5} M_{\odot}$ ) galaxies also exhibit high progenitor probabilities at high redshift, as long as they occupy regions of high density (e.g. the 80-100<sup>th</sup> percentiles in density).
- At high-redshift, massive late-type galaxies that are progenitors of present-day early-types can be some of the most highly star-forming systems in the Universe, simply because star-formation activity is, on average, more vigorous in the early Universe. However, at fixed stellar mass, progenitor fractions increase as the star-formation rate decreases i.e. late-type galaxies with lower specific star-formation rates are more likely to be progenitors of early-type galaxies.

In the impending era of large observational surveys (e.g. LSST, EUCLID, JWST), this chapter provides a framework for studies of how the stellar mass hosted by the local early-type galaxy population is built up over cosmic time.

## Chapter 3

# The limited role of galaxy mergers in driving stellar mass growth over cosmic time

### 3.1 Introduction

Understanding the processes that drive stellar mass growth over cosmic time is a key topic in observational cosmology. Since the cosmic star formation rate (SFR) density peaked at  $z \sim 2$  and dropped by more than an order of magnitude towards the present day (e.g. Madau & Dickinson, 2014; González et al., 2014), almost half of the stellar mass hosted by today's galaxies formed at  $z \gtrsim 1.3$  (Madau & Dickinson, 2014), making this epoch particularly important in the evolution of the observable Universe.

Galaxy mergers are often considered to be important drivers of stellar mass growth (e.g. van Dokkum et al., 2010; Kaviraj et al., 2011; López-Sanjuan et al., 2012; Ferreras et al., 2014). For example, mergers can produce orders-of-magnitude enhancements in SFRs in the nearby Universe (e.g. Duc et al., 1997; Elbaz & Cesarsky, 2003), implying that a significant fraction of the stellar mass formed in these episodes is a direct consequence of the merger event. Since the merger rate increases towards high redshift, it is reasonable to consider whether a significant fraction of the stellar mass in today's galaxies may, therefore, have been created in enhanced star-formation episodes associated with galaxy mergers (e.g. Somerville et al., 2001; Conselice et al., 2008). In other words, if galaxy mergers are frequent and routinely enhance SFRs when they take place, then much of the stellar mass at the present day could be directly attributable to the merging process.

However, while mergers are clearly capable of triggering bursts of star formation (e.g. Mihos & Hernquist, 1996; Di Matteo et al., 2008), and strongly star-forming systems are often coincident with ongoing interactions (e.g. Sanders et al., 1988; Bell et al., 2006), the empirical picture remains unclear, especially

at high redshift. Recent observational studies of galaxies around the epoch of peak cosmic star formation (e.g. Rodighiero et al., 2011; Stott et al., 2013; Lofthouse et al., 2017a) indicate that ‘major’ mergers (i.e. mergers with mass ratios  $> 1 : 4$ ) are unlikely to be responsible for the bulk of the stellar mass growth at these epochs, as the SFR enhancements in major mergers – compared to the non-merging population – are relatively low (e.g. Lofthouse et al., 2017a). This implies that there must be other processes that fuel these high SFRs and drive the production of stellar mass at these epochs.

Given that the frequency of ‘minor’ mergers (mass ratios  $< 1 : 4$ ) is several times that of major mergers (e.g. Lotz et al., 2011; Kaviraj et al., 2015b), and that mergers of moderate mass ratios are also capable of producing large SFR enhancements (e.g. Cox et al., 2008), minor merging could potentially make an important contribution to the star formation budget (Kaviraj, 2014a,b). Alternatively, the high SFRs may simply be the result of high molecular-gas fractions, fuelled by intense cosmological gas accretion (e.g. Tacconi et al., 2010; Geach et al., 2011; Béthermin et al., 2015).

While quantifying the role of mergers in driving cosmic stellar mass growth is an important exercise, an empirical determination of this issue brings with it several difficulties. Selecting mergers based on morphological disturbances is not a simple task, since disturbed morphologies can also result from internal processes, especially in the early Universe (e.g. Bournaud et al., 2008; Agertz et al., 2009; Förster Schreiber et al., 2011; Cibinel et al., 2015; Hoyos et al., 2016). Furthermore, since the surface brightness of merger-induced tidal features declines with the mass ratio of the merger, minor mergers are less likely to produce observable asymmetries, especially at high redshift, even in today’s deep surveys (Kaviraj et al., 2013a). Finally, given the depth and areal coverage of current and past facilities, samples of mergers are often small, and both the galaxy populations studied and star formation indicators employed can be heterogeneous, making it difficult to compare results across a large range in redshift.

With these issues in mind, an appealing alternative is to employ a simulation that reproduces the observed properties of galaxies over cosmic time (e.g. Lemastra et al., 2013; Vogelsberger et al., 2014; Schaye et al., 2015; Khandai et al., 2015; Taylor & Kobayashi, 2016; Kaviraj et al., 2017). A major advantage of this approach is that, since the identities of the progenitors of each galaxy in the simulation are precisely known, it is straightforward to separate merging galaxies from their non-merging counterparts. This then allows us to integrate over the star formation history of each merger (including any subsequent post-starburst decrease in the SFR), study the properties of the induced star formation and make quantitative statements about the overall role of merging in creating the stellar mass in today’s Universe.

In this Letter, we use the hydrodynamical cosmological simulation, Horizon-AGN<sup>1</sup> (Dubois et al., 2014; Kaviraj et al., 2017), to quantify the contribution of mergers to the star formation budget since  $z = 6$ . In Section 3.2, we describe the simulation and the prediction of observable quantities in the model. In Section 3.3, we describe our identification of mergers and calculate the merger contribution to the star formation budget as a function of redshift. In Section 3.4, we quantify the contribution of both major and minor mergers to the cosmic star formation history. We summarize our findings in Section 3.5.

---

<sup>1</sup><http://www.horizon-simulation.org>

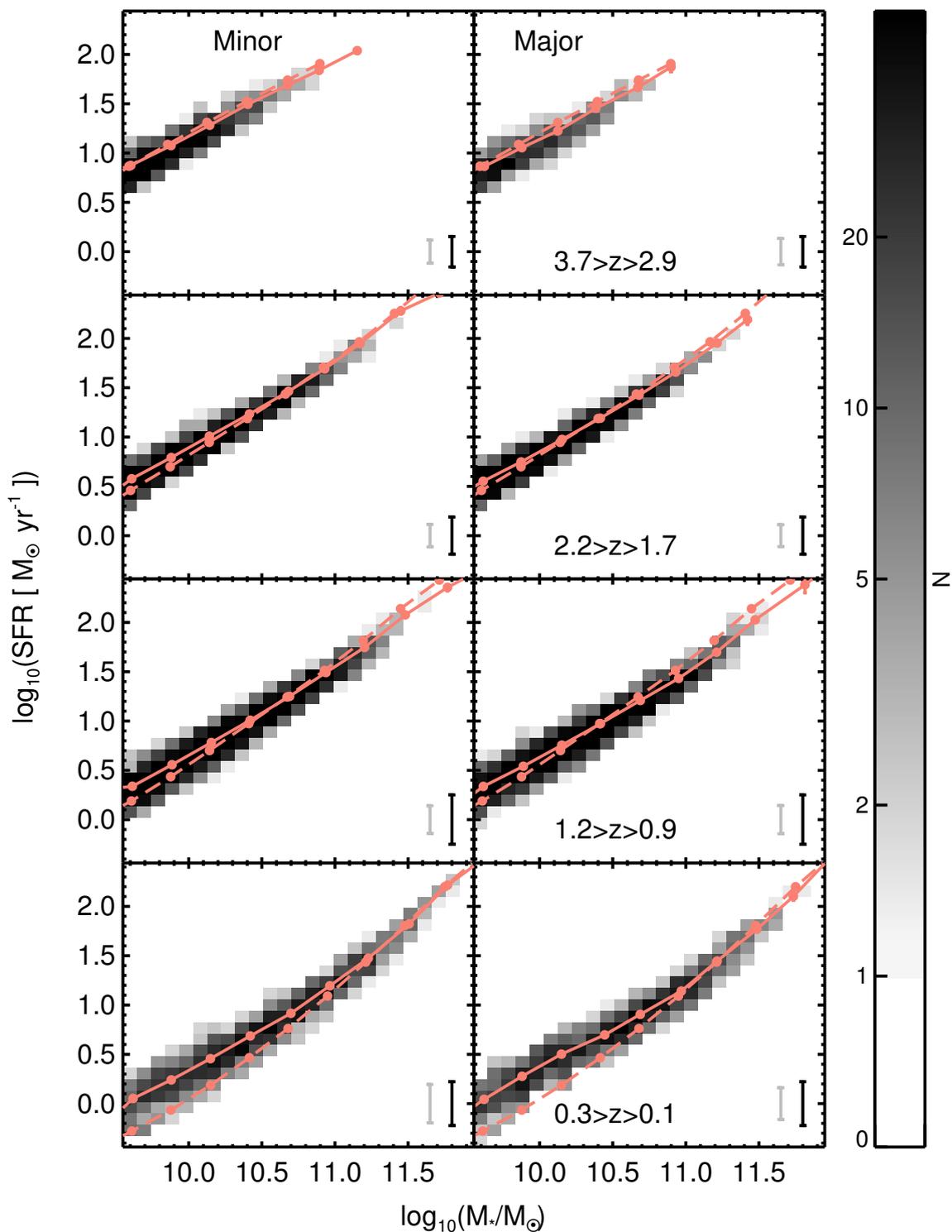


FIGURE 3.1: SFR as a function of stellar mass for the merging and non-merging populations in various redshift ranges. Greyscale density maps represent the minor (left) and major (right) merger populations. Solid pink lines shows the mean SFR of the merging population, while the dashed pink lines show the mean SFR of the non-merging population in each redshift bin. The offset between the solid and dashed lines in every panel therefore indicates the average enhancement due to a merger. The grey and black error bars indicate the typical standard deviations for the merging and non merging populations respectively.

### 3.2 The Horizon-AGN simulation

We begin with a brief description of the Horizon-AGN simulation and the prediction of observable quantities in the model. Horizon-AGN is a cosmological hydrodynamical simulation (Dubois et al., 2014) that employs RAMSES (Teyssier, 2002), an adaptive mesh refinement Eulerian hydrodynamics code. It simulates a volume of  $(100 h^{-1} \text{coMpc})^3$  containing  $1024^3$  DM particles ( $M_{\text{DM}} = 8 \times 10^7 M_{\odot}$ ) and uses initial conditions from a *WMAP7*  $\Lambda$ CDM cosmology (Komatsu et al., 2011). The initial gas mass resolution is  $10^7 M_{\odot}$ , with a maximum grid refinement of  $\Delta x = 1$  kpc. Horizon-AGN includes sub-grid prescriptions for star formation and stellar/AGN feedback. Star formation proceeds with a standard 2 per cent efficiency per free-fall time (Kennicutt, 1998), once the Hydrogen gas density reaches  $n_0 = 0.1 \text{ H cm}^{-3}$ . Continuous stellar feedback is employed which includes momentum, mechanical energy and metals from Type II SNe, stellar winds, and Type Ia SNe (Kaviraj et al., 2017), with the Type Ia SNe implemented following Matteucci & Greggio (1986), assuming a binary fraction of 5 per cent. Black-hole feedback on ambient gas operates via a combination of two channels and depends on the ratio of the gas accretion rate to the Eddington luminosity,  $\chi = \dot{M}_{\text{BH}}/\dot{M}_{\text{Edd}}$ . For Eddington ratios greater than 0.01 (high accretion rates) a ‘quasar’ mode is active with 1.5 per cent of the accretion energy being injected isotropically into the gas as thermal energy. For Eddington ratios less than 0.01 (low accretion rates) a ‘radio’ mode is active, where cylindrical bipolar outflows are employed with a jet velocity of  $10^4 \text{ km s}^{-1}$ . The efficiency of the radio mode is higher, at 10 per cent of the accretion energy. The quasar mode efficiency is chosen to reproduce observed relations between  $M_{\text{BH}} - M_{\star}$  and  $M_{\text{BH}} - \sigma_{\star}$  relations as well as the local cosmic black-hole mass density (Dubois et al., 2012).

Horizon-AGN reproduces key observables that trace the aggregate cosmic stellar mass growth of galaxies: stellar mass and luminosity functions, rest-frame UV-optical-near infrared colours, the star formation main sequence and the cosmic star formation history (Kaviraj et al., 2017). It also reproduces galaxy merger histories (Kaviraj et al., 2015b) and the demographics of black holes (BHs): the BH luminosity and mass functions, the BH mass density versus redshift, and correlations between BH and galaxy mass (Volonteri et al., 2016).

We use the ADAPTAHOP structure finder (Aubert et al., 2004; Tweed et al., 2009) to identify galaxies in the final snapshot of the simulation ( $z = 0.06$ ), and build merger histories for each galaxy. We produce a catalogue of galaxies with  $M > 10^{9.5} M_{\odot}$  from  $z = 0.06$  to  $z = 6$  and calculate the stellar mass formed in each galaxy between timesteps. Since the minimum galaxy mass identified by the structure finder is  $M_{\star} \approx 2 \times 10^8 M_{\odot}$ , our sample is complete for mergers down to a mass ratio of at least 1:15.

### 3.3 Star formation enhancement due to merging

We begin our analysis by identifying mergers in the simulation and measuring their mass ratios (Section 3.3.1). We then compare the SFRs of merging galaxies with those of the non-merging galaxy population,

so as to estimate (and ‘subtract’) the star formation that would have taken place anyway in the absence of merging (Section 3.3.2). This then enables us to calculate the stellar mass growth that is directly attributable to the merger process.

### 3.3.1 Defining and identifying mergers

To measure the SFR of the merging system, we calculate the total stellar mass formed in a 2 Gyr window, centred around the time that the two galaxies coalesce (i.e. when both galaxies are identified as being part of the same structure). We note that the size of the window is chosen to encompass the star formation history of the system around the merger, and that the exact choice of timescale (e.g. increasing it to 3 Gyrs or even reducing it to 1 Gyr) does not alter our conclusions.

It is also worth noting that how the mass ratio is defined can influence the minor and major merger rate and therefore the results of such an analysis (e.g. Rodriguez-Gomez et al., 2015). For this study, we use the mass ratio calculated when the satellite is at its maximum mass prior to coalescence – i.e. before material begins to be transferred between the merging companions – because this measures the ‘true’ mass ratio of the system, before the merger process begins to alter the properties of the merging progenitors. Only mergers with mass ratios greater than 1 : 10 are considered since, in agreement with previous studies (e.g. Cox et al., 2008), we find that smaller mass ratio mergers have a negligible effect on the star formation rate.

### 3.3.2 Star formation triggered by mergers

Fig 3.1 compares the star-formation main sequence of the merging and non-merging galaxy populations in four redshift bins. The mergers are further split into minor mergers (mass ratios  $< 1 : 4$ ; left-hand column) and major mergers (mass ratios  $\gtrsim 1 : 4$ ; right-hand column). The solid pink lines show the mean SFRs of the merging populations, while the dashed pink lines shows the mean SFRs of the non-merging populations in each panel. The difference between the solid and dashed lines therefore indicates the (average) enhancement of star formation in merging galaxies at a given epoch. We note that, in common with other theoretical work (e.g. Davé, 2008; Lemastra et al., 2013), the normalisation of the star-formation main sequence in Horizon-AGN is underestimated compared to its observational counterparts by  $\sim 0.2$  dex (Kaviraj et al., 2017). However, since the baryonic recipes used are not altered during merging, any star-formation enhancement in merging galaxies will be proportionally reduced to the same extent, leaving our conclusions unchanged.

It is interesting to note that the SFR enhancement due to minor mergers does not differ significantly from that of major mergers, consistent with the findings of recent observational studies (see e.g. Willett et al., 2015; Carpineti et al., 2015). This is likely driven by the fact that the gas inflows which underpin the SFR enhancements in mergers (Di Matteo et al., 2007) can be of similar magnitude in minor mergers

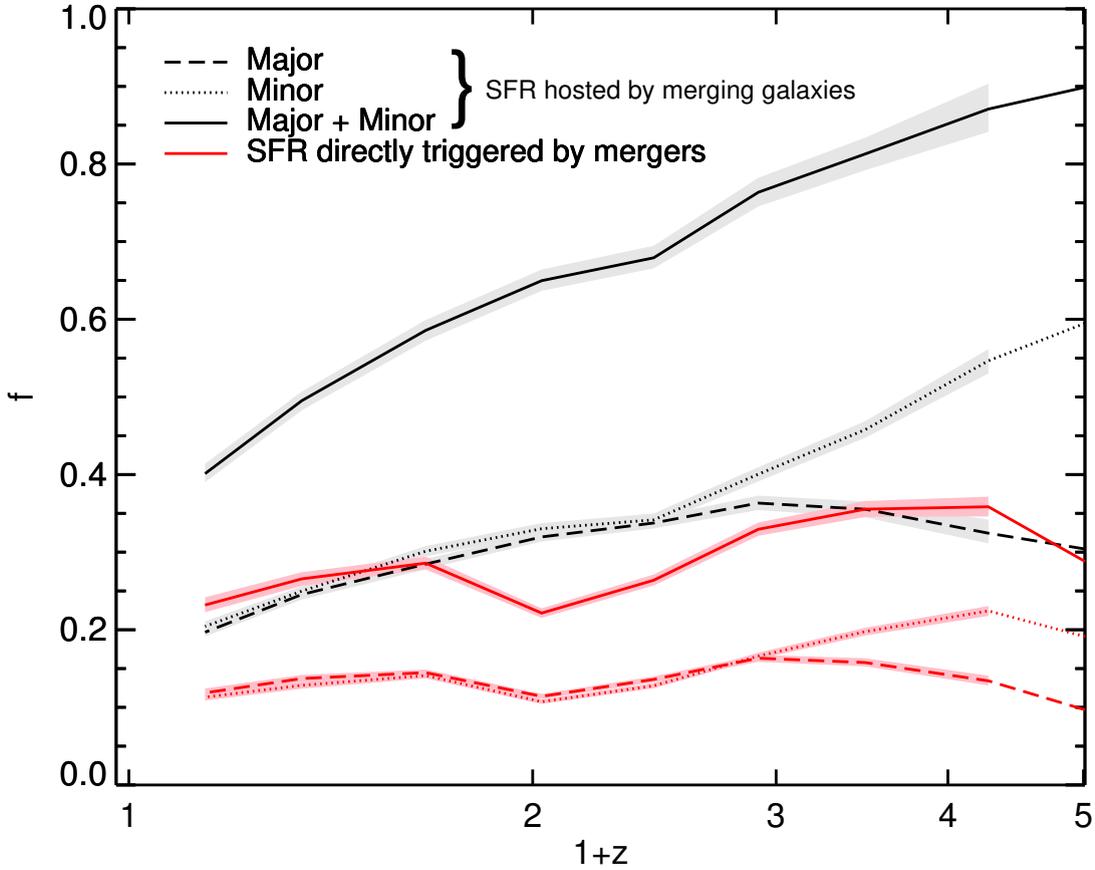


FIGURE 3.2: The fraction of star formation in merging systems (black) and the fraction of the star formation budget that is directly triggered by merging (red) as a function of redshift. Dotted lines indicate the contribution of minor mergers; dashed lines indicate the contribution of major mergers; solid lines indicate the combined contribution of major and minor mergers. Filled lighter colour regions indicate the  $1\sigma$  errors obtained from bootstrap re-sampling ( $n = 1000$ ).

as they are in their major counterparts (e.g. Hernquist & Mihos, 1995). Furthermore, merger-driven SFR enhancement is most efficient in the local universe, because the ‘ambient’ level of star formation due to secular processes is much lower, which allows violent events like mergers to produce significant enhancements in the SFR (e.g. Mihos & Hernquist, 1996). We define the merger-driven enhancement of star formation,  $\xi$ , as the ratio of the mean specific star formation rate (sSFR) in the non-merging population to that in the merging population. We measure  $\xi$  in bins of both redshift and stellar mass (since the sSFR has a dependence on this parameter (e.g. Whitaker et al., 2012)):

$$\xi(m_*, z) = \frac{\langle \text{sSFR}_m(m_*, z) \rangle}{\langle \text{sSFR}_{non}(m_*, z) \rangle}. \quad (3.1)$$

The enhancement can be used to estimate the fraction of star formation that would have occurred in the merger progenitors anyway, had they not been in the process of merging. For example, if  $\xi$  is a factor of

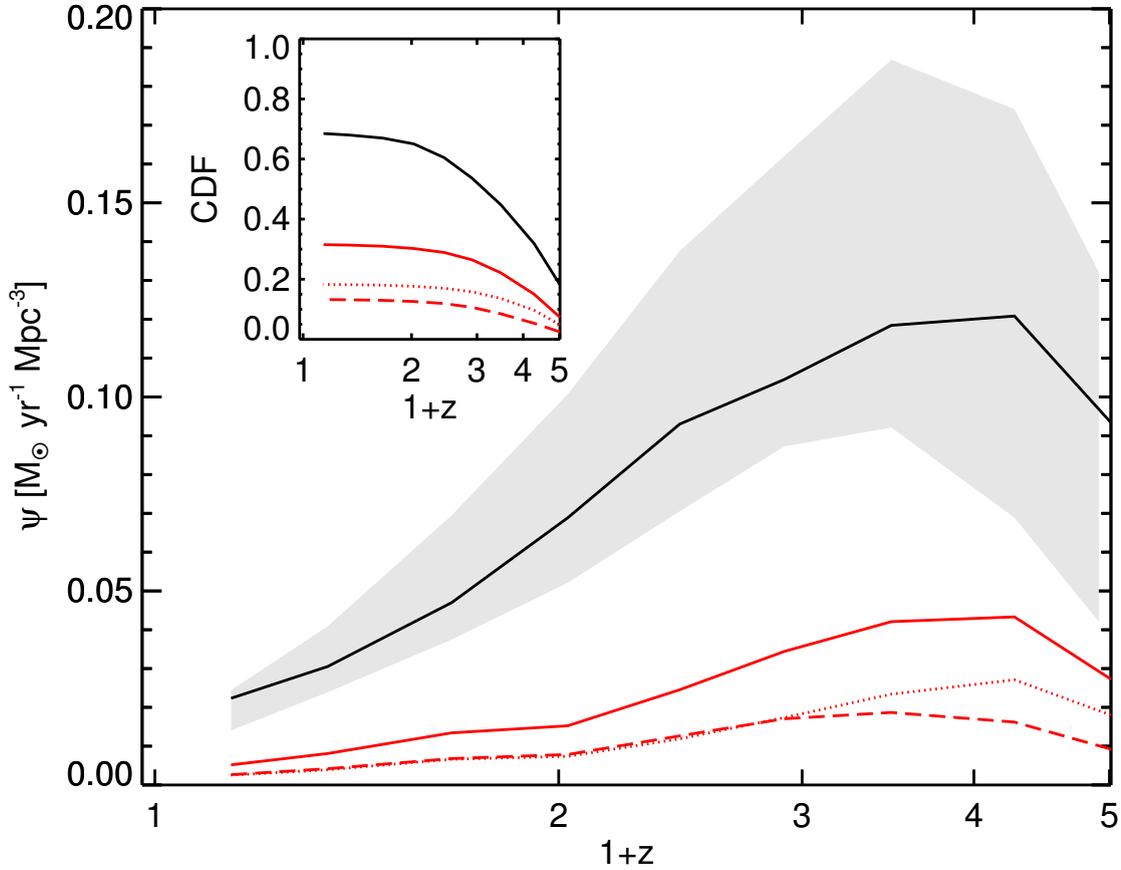


FIGURE 3.3: The cosmic star formation rate density from the Horizon-AGN simulation (black) and the contribution due to major and minor mergers (red). The inset shows a cumulative version of this plot i.e. the cumulative fraction of stellar mass formed due to mergers (red) and due to other processes (black). The grey filled area indicates the  $3\sigma$  confidence region from observations (Hopkins & Beacom, 2006).

2 then, on average, around half the star formation in the merging system in question is likely driven by other processes (see e.g. Kaviraj et al., 2013b; Lofthouse et al., 2017a). By subtracting the star formation that would have occurred anyway had the merger not taken place, we can then measure the fraction of star formation that is *directly due* to mergers ( $f$ ) as follows:

$$f = \frac{m_{new,m}(m_*, z) [1 - 1/\xi(m_*, z)]}{m_{new,total}(m_*, z)}, \quad (3.2)$$

where  $m_{new,m}(m_*, z)$  is the total stellar mass formed in mergers in a given stellar mass and redshift bin and  $m_{new,total}(m_*, z)$  is the total stellar mass formed in the simulation in the stellar mass and redshift bin in question.

As Fig 3.2 shows, the fraction of star formation *in* merging galaxies increases towards high redshift, reflecting the increasing merger rate. However, the fraction of star formation that is *directly due* to

mergers (shown by the red lines in Fig 3.2) does not increase to the same extent, which is a consequence of a decreasing merger-driven SFR enhancement towards high redshift, as shown in Fig 3.1. The fraction of star formation triggered by merging peaks around  $z \sim 3$  ( $\sim 35$  per cent), and then decreases to  $\sim 20$  per cent by  $z \sim 1$ . We find that, on average, 65 per cent of the enhanced star formation due to a merger takes place prior to coalescence, with the star formation rate in the post-merger remnant returning to that of the non-merging population in less than 1 Gyr for galaxies at  $z > 1$ . It is worth noting that our results are consistent with recent observational and theoretical work that has probed the contribution of major mergers to the cosmic SFR density in selected redshift ranges. For example, Lamastra et al. (2013) and Robaina et al. (2009) indicate that the major-merger contribution to cosmic star formation at low/intermediate redshift ( $0.4 < z < 2$ ) is around 10 per cent, with only modest SFR enhancements at these epochs (Robaina et al., 2009; Fensch et al., 2017), as indicated by Fig 3.1.

### 3.4 The merger contribution to the cosmic star formation history

We proceed by studying the merger contribution to the overall build-up of stellar mass over cosmic time, by multiplying the fraction of star formation directly due to mergers from Section 3.3 (red lines in Fig 3.2) by the cosmic star formation rate density ( $\psi$ ). We present, in Fig 3.3, the cosmic star formation rate density in Horizon-AGN (black solid line). Since our sample of simulated galaxies is limited to masses above  $10^{9.5} M_{\odot}$ , and the merger-driven enhancement of star formation increases for galaxies with lower stellar mass (Fig 3.1), it is important to ask if galaxies less massive than our mass threshold could contribute significantly to the star formation budget. To explore this, we multiply the star formation rate vs stellar mass trend at  $z \sim 0$  (Elbaz et al., 2007) and  $z \sim 2$  (Daddi et al., 2007) with the galaxy stellar mass functions at the same redshifts from Baldry et al. (2008) and Tomczak et al. (2014a), in order to produce star formation rate densities per dex in stellar mass down to  $10^7 M_{\odot}$ . We find that only  $\sim 22$  per cent and  $\sim 16$  per cent of stellar mass at  $z \sim 0$  and  $z \sim 2$  respectively is formed in galaxies less massive than  $10^{9.5} M_{\odot}$ . It appears reasonable, therefore, to assume that considering the full stellar mass range would not significantly alter our conclusions.

Fig 3.3 indicates that the proportion of the cosmic star formation budget that is directly attributable to merging is small at all redshifts. Following the trends in Fig 3.2, it peaks around  $z \sim 3$  ( $\sim 0.04 M_{\odot} \text{ yr}^{-1} \text{ Mpc}^{-3}$ ) and then steadily declines towards the present day. The inset shows a cumulative version of this plot, indicating that only 25 per cent of the star formation budget since  $z \sim 6$  is attributable to mergers ( $\sim 10$  per cent from major mergers and  $\sim 15$  per cent from minor mergers). Recall that the contribution by very low mass ratio ( $< 1 : 10$ ) mergers is expected to be negligible, so that this result should hold generally for all mergers over cosmic time. While a detailed study of the role of secular processes is beyond the scope of this Letter, our results indicate that an overwhelming majority ( $\sim 75$  per cent) of the cosmic star formation budget is unrelated to merging and a result of secular evolution, driven simply by cosmological accretion of molecular gas, in line with the suggestions of recent observational work (e.g. Tacconi

et al., 2010; Béthermin et al., 2015) and previous theoretical work which has suggested that non-merging systems dominate the SFR density at all redshifts (e.g. Hopkins et al., 2010).

### 3.5 Summary

We have used the Horizon-AGN cosmological hydrodynamical simulation, to quantify the contribution of galaxy mergers to stellar mass growth over cosmic time. Our key results are as follows:

- *Mergers enhance star formation most efficiently at low redshift.* Mergers are most effective at increasing the star formation rate of the host galaxy at  $z < 1$ , when the ‘ambient’ level of star formation due to secular processes is low.
- *Both major and minor mergers enhance star formation, on average, by similar amounts at any given redshift.* e.g. minor mergers enhance SFRs, on average, by a factor of 1.69 at  $z \sim 2$ , while the corresponding value for major mergers is 1.75. At  $z \sim 3.3$ , minor mergers enhance SFRs, on average, by a factor of 1.68 while major mergers enhance SFRs by a factor of 1.69.
- *Merger-driven enhancement of star formation decreases with increasing redshift.* While the merger rate increases with redshift, the SFR enhancement due to mergers decreases with redshift. This means that, while the fraction of star formation hosted *in* merging systems increases with look-back time (due to the increasing merger rate), the fraction of star formation *directly due* to mergers increases at a much slower rate.
- *Episodes of enhanced star formation typically occur prior to coalescence.* On average, 65 per cent of the enhanced star formation in a merger episode occurs prior to coalescence. Star formation in the post-merger remnant returns to levels found in the non-merging population on short timescales of around 1 Gyr.
- *Only 25 per cent of the stellar mass growth since  $z \sim 6$  is directly attributable to galaxy mergers.* Major and minor mergers together account for just 25 per cent of the stellar mass formed since  $z = 6$ . Only  $\sim 10$  per cent of today’s stellar mass is directly due to major mergers, while  $\sim 15$  per cent is due to minor mergers. While individual minor mergers are less efficient enhancers of star formation, the minor merger rate outstrips the major merger rate at all redshifts, leading to a greater minor merger contribution over cosmic time. Thus, smooth accretion, not merging, is the dominant driver of stellar mass growth over the lifetime of the Universe.

## Chapter 4

# Normal black holes in bulge-less galaxies: the largely quiescent, merger-free growth of black holes over cosmic time

### 4.1 Introduction

The co-evolution of galaxies and their black holes (BHs) is a central theme of our galaxy formation paradigm. In the nearby Universe, several correlations are observed between BH mass and the properties of the host galaxy, such as its velocity dispersion (Magorrian et al., 1998; Ferrarese & Merritt, 2000), the mass of its bulge (e.g. Marconi & Hunt, 2003; Häring & Rix, 2004) and its total stellar mass (e.g. Cisternas et al., 2011a; Marleau et al., 2013), which suggest that the evolution of galaxies and their central BHs may be linked.

However, the processes that underpin these correlations have remained a matter of debate. For example, the correlation between BH and bulge mass is often considered to be a product of galaxy mergers (e.g. Sanders et al., 1988; Croton et al., 2006; Hopkins et al., 2006). Simulations show that mergers (in particular ‘major’ mergers, i.e. those with near-equal mass ratios) are efficient at building bulges (e.g. Toomre & Toomre, 1972; Barnes, 1992), although some bulges may form via other processes, such as disk instabilities (e.g. Dekel et al., 2009; Kaviraj et al., 2013a) and, in cases where gas fractions are particularly high, disks may reform from residual gas even after a major merger (see e.g. Springel & Hernquist, 2005; Kannappan et al., 2009; Hopkins et al., 2009a). Combined with the fact that active galactic nuclei (AGN), and thus growing BHs, are often observed in systems undergoing major mergers (e.g. Urrutia et al., 2008; Bessiere et al., 2014; Chiaberge et al., 2015; Glikman et al., 2015; Trakhtenbrot et al., 2017), it is reasonable to suggest that this process could create the observed  $M_{\text{BH}}-M_{\text{Bulge}}$  correlation, by simultaneously building the BH and the galaxy bulge (e.g. Sanders et al., 1988; Hopkins et al., 2006; Peng, 2007).

While much of the past literature on BH–galaxy correlations has focussed on early-type (i.e. bulge dominated) galaxies, recent work has started to probe how these correlations may behave in the general galaxy population. Many studies now indicate that a broad correlation exists across the general population of galaxies, if the relationship between BH mass and the *total* stellar mass (e.g. Grier et al., 2011; Cisternas et al., 2011a,b; Marleau et al., 2013; Reines & Volonteri, 2015) or the relationship between BH mass and halo mass (e.g. Booth & Schaye, 2010; McAlpine et al., 2017) of the host galaxy is considered. The  $M_{\text{BH}}-M_{\text{Bulge}}$  correlation is then likely to be just a subset of this general trend, since early-type galaxies are bulge-dominated, and therefore their total stellar mass is largely the same as their bulge mass.

The origin of a  $M_{\text{BH}}-M_{\star}$  correlation, that exists irrespective of morphological type, is difficult to explain via major mergers alone, since the bulk of the material contained in galaxy disks is likely to have been formed via secular processes (e.g. Martig et al., 2012; Conselice et al., 2013). However, the recent observational literature suggests that building up such a correlation, via processes other than major mergers, is plausible. AGN, particularly those with moderate accretion rates typical of normal galaxies (Hasinger et al., 2005), are often found in systems that are not associated with major mergers (e.g. Grogin et al., 2005; Gabor et al., 2009; Pipino et al., 2009; Kaviraj et al., 2012b; Kocevski et al., 2012; Shabala et al., 2012; Kaviraj, 2014a; Kaviraj et al., 2015a). Recent studies have shown that minor mergers can enhance star-formation and nuclear-accretion rates (e.g. Kaviraj, 2014a,b; Comerford et al., 2015; Capelo et al., 2015; Smethurst et al., 2015; Steinborn et al., 2016; Martin et al., 2017) and could, therefore, produce BH growth while leaving the disk intact. Certain secular processes which are connected to, or responsible for, fuelling star formation – e.g. bar driven inflows of gas (Regan & Teuben, 2004; Lin et al., 2013), disc instabilities (Bournaud et al., 2011) or cosmological cold flows (Feng et al., 2014) – may also be capable feeding the BH by driving gas towards the central regions of galaxies.

There is evidence that spiral galaxies with low central velocity dispersions, and therefore low bulge masses, tend to have over-massive BHs (Sarzi et al., 2002; Beifiori et al., 2009) when considering the  $M_{\text{BH}}-M_{\text{Bulge}}$  correlation, which suggests that the processes that build the BH and the bulge may be different (e.g. Grupe & Mathur, 2004; Mathur & Grupe, 2004, 2005a,b). It is also worth noting that a general dearth of major mergers in the AGN population is found around the epoch of peak cosmic star formation (e.g. Simmons et al., 2011; Schawinski et al., 2012; Kocevski et al., 2012), when the bulk of the stellar and BH mass in today’s galaxies was assembled (e.g. Hopkins et al., 2006).

Furthermore, recent work (Martin et al., 2017) has shown that a majority ( $\sim 90$  percent) of the stellar mass in today’s Universe is likely to be unrelated to major merging. If BH and stellar mass growth move in lockstep with each other, then it is reasonable to suggest that the BH accretion rate budget may also be decoupled from the major-merger process, which would then lead to the  $M_{\text{BH}}-M_{\star}$  relation observed at low redshift.

Some caveats to the arguments presented above are worth considering. Given that BHs comprise, on average, only  $\sim 0.2$  per cent of their host galaxy’s stellar mass (e.g. Häring & Rix, 2004), it is still possible that they form preferentially in major mergers, since this process is responsible for  $\sim 10$  per

cent of cosmic stellar growth (Martin et al., 2017). In addition, since major mergers with high gas fractions may result in reformed disks, it may be possible for ‘normal’ BHs to form in systems that do not have early-type morphology, but which have had gas-rich major mergers in their formation history. Additionally, recent work (e.g. Sani et al., 2011; Mathur et al., 2012; Kormendy & Ho, 2013) have found that galaxies with pseudo-bulges – which are often interpreted to be the result of minor mergers and secular processes (e.g. Kormendy & Kennicutt, 2004) – lie below the  $M_{\text{BH}}-M_{\text{Bulge}}$  relation, suggesting that another process, such as major merging, may still be an important channel for BH growth (although it is worth noting that BH masses in pseudo-bulged galaxies do not differ greatly from those in galaxies that exhibit classical bulges).

A compelling counter to the hypothesis of (major) merger-driven black hole growth is the presence of massive black holes in disc-dominated galaxies. Studies of the BHs in such systems, (e.g. Filippenko & Ho, 2003; Ghosh et al., 2008; Araya Salvo et al., 2012) show that supermassive BHs cannot be associated exclusively with bulges. A particularly stringent test of whether major mergers preferentially build BHs is to compare the BH masses in bulge-less galaxies (i.e. those that are unlikely to have had many major mergers) to those in the general galaxy population. While such galaxies are rare in the nearby Universe, Simmons et al. (2017) have recently performed such a test on disk-dominated and bulge-less systems drawn from the SDSS. Their study shows that disk-dominated and bulge-less galaxies lie offset above the main locus in the  $M_{\text{BH}}-M_{\text{Bulge}}$  correlation. However, these galaxies fall on the main locus of the  $M_{\text{BH}}-M_{\star}$  relation, like the rest of the galaxy population. In other words, bulge-less galaxies appear to have normal BHs, yet are systems that are unlikely to have had many major mergers in their evolutionary histories. In these systems, at least, major mergers appear unlikely to have been the dominant drivers of BH growth.

While recent observational work hints at the possibility that BH growth does not require major mergers (e.g. Kaviraj, 2014b), it remains challenging to address this issue via empirical work alone, at least using current surveys. Current observational datasets are heterogeneous and relatively small and the measurement of precise BH masses remains a difficult exercise. Furthermore, while major mergers are generally expected to build bulges, there is a possibility that disk rebuilding after very gas-rich major mergers (especially at high redshift) may preserve some disk structure - differentiating such systems from ‘normal’ disk galaxies that have evolved in the absence of major mergers is difficult observationally.

Given the observational challenges described above, an alternative approach is to appeal to a theoretical model that reproduces both the stellar mass growth and BH demographics of the galaxy population over cosmic time. In this study, we use Horizon-AGN, a cosmological hydrodynamical simulation, to probe the BH–galaxy correlations that naturally arise in the standard model. We specifically probe the evolution of bulge-less galaxies in the simulation, explore how galaxies with varying contributions of major and minor mergers in their evolutionary histories differ in their positions on these correlations, and quantify how much of the BH accretion budget is directly attributable to merging over cosmic time.

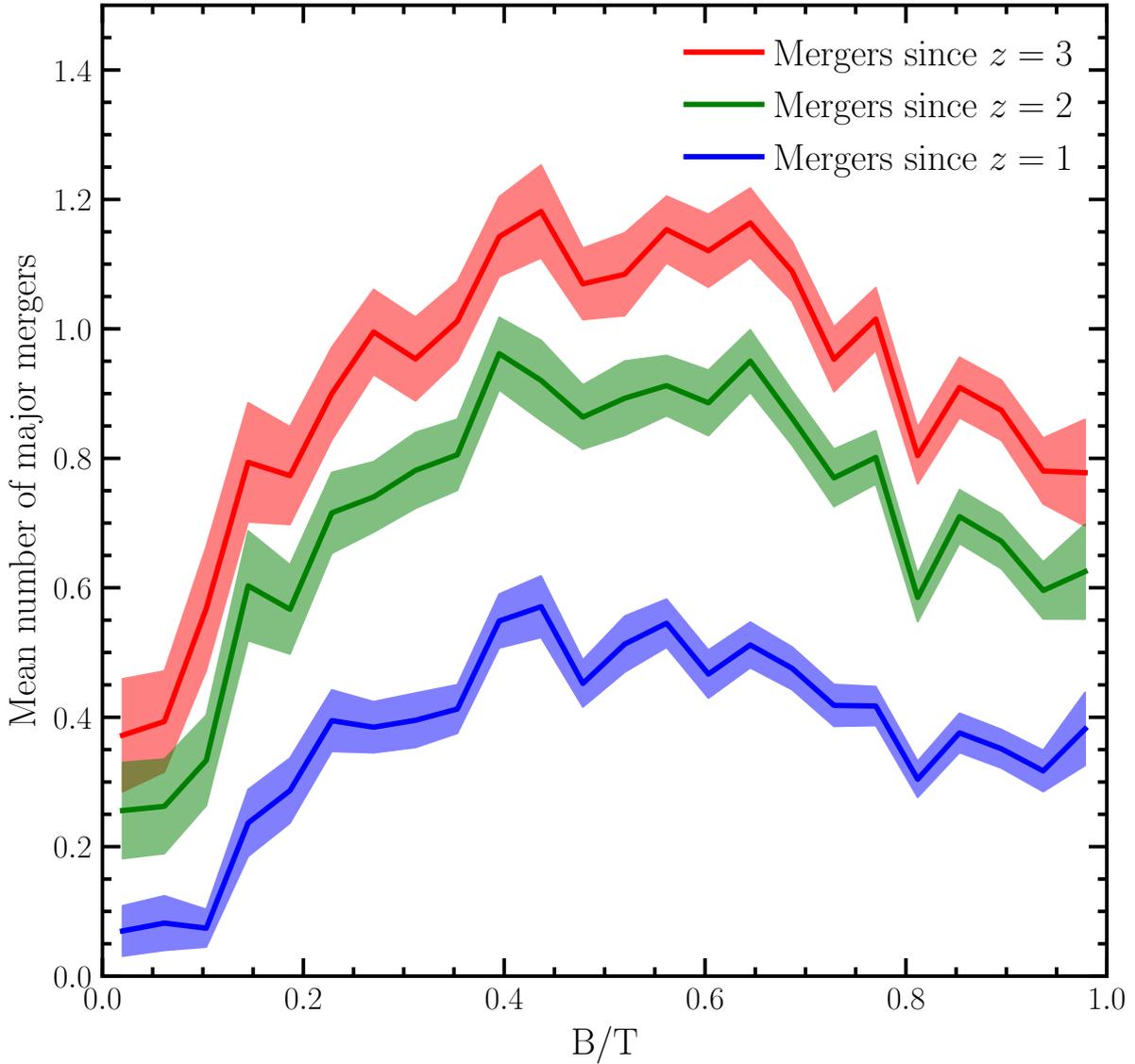


FIGURE 4.1: The mean number of major mergers that local massive galaxies ( $M_* > 10^{10.5}$  at  $z = 0$ ) have undergone after a given redshift, as a function of their bulge to total stellar mass (B/T) ratios. The colour corresponding to a given redshift is indicated by the legend. Filled polygons indicate the standard error on the mean from Poisson errors. Major mergers are defined as mergers with mass ratios greater than 1 : 4. Note that the downward trend in the mean number of major mergers at high B/T values is driven by the fact that these galaxies are typically massive early-type galaxies. Since these systems tend to be some of the highest-mass systems at a given epoch, there are, by definition, not many systems of similar mass. Thus, these galaxies are less likely to experience major mergers. The number of minor mergers (not shown) is typically a factor of 2.5 times greater (e.g. Kaviraj et al., 2015a).

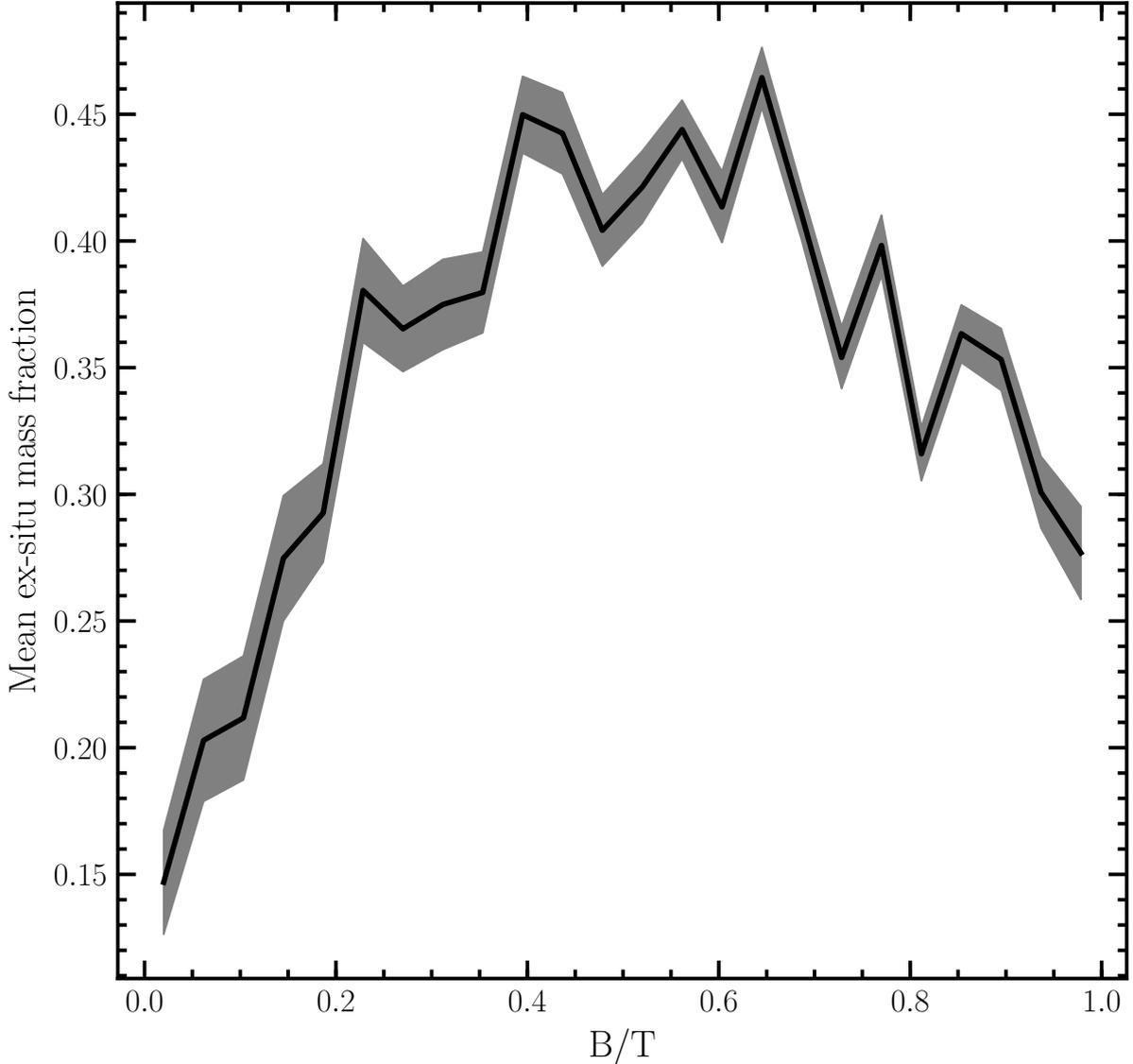


FIGURE 4.2: The mean fraction of ex-situ mass in today’s massive galaxies ( $M_* > 10^{10.5}$  at  $z = 0$ ), as a function of their B/T ratio. Galaxies with low B/T ratios are likely to have low ex-situ mass fractions, indicating that the majority of their stellar mass formed via secular processes.

This chapter is organised as follows. In Section 4.2, we describe the simulation employed by this study. Although the model is described in detail in Dubois et al. (2014) and Kaviraj et al. (2017), we outline the treatment of baryons and BHs, as these are relevant to the observable quantities that are being studied in this chapter. In Section 4.3, we briefly outline the role of mergers (and major mergers in particular) in the production of bulges. In Section 4.4, we explore the correlations between BH, bulge and total stellar mass that are produced by the simulation, the potential role of mergers in the formation of these correlations and quantify the fraction of the BH accretion budget that is directly attributable to mergers over cosmic time. We summarize our findings in Section 4.5.

## 4.2 The Horizon-AGN simulation

Horizon-AGN is a hydrodynamical simulation (Dubois et al., 2014) in a cosmological volume that employs the adaptive mesh refinement (AMR) code, RAMSES (Teyssier, 2002). The simulation box is  $100 h^{-1}$  coMpc on a side, with  $1024^3$  dark matter particles, and uses initial conditions from a *WMAP7*  $\Lambda$ CDM cosmology (Komatsu et al., 2011). It has a dark matter mass resolution of  $8 \times 10^7 M_{\odot}$ , a stellar-mass resolution of  $2 \times 10^6 M_{\odot}$  and a spatial resolution of  $\sim 1$  kpc. A quasi Lagrangian criterion is used to refine the initially uniform  $1024^3$  grid, when 8 times the initial total matter resolution is reached in a cell, down to a minimum cell size of 1 kpc in proper units.

In the following sections, we describe some aspects of the simulation that are central to this study: the treatment of baryons, the identification of galaxies and mergers, the growth of BHs and BH feedback on ambient gas. As described in detail in Kaviraj et al. (2017), Horizon-AGN reproduces key observables that trace the aggregate cosmic stellar mass growth of galaxies since  $z \sim 6$ : stellar mass and luminosity functions, rest-frame UV-optical-near infrared colours, the star formation main sequence, the cosmic star formation history and galaxy merger histories (Kaviraj et al., 2015b). It also reproduces the demographics of black holes (BHs) over cosmic time: the BH luminosity and mass functions, the BH mass density versus redshift, and correlations between BH and galaxy mass (Volonteri et al., 2016).

### 4.2.1 Treatment of baryons

Following Sutherland & Dopita (1993), gas cools via H, He and metals, down to  $10^4$  K. A UV background is switched on at  $z = 10$ , following Haardt & Madau (1996). Star formation takes place via a standard 2 per cent efficiency (Kennicutt, 1998), when the hydrogen density reaches a critical threshold of  $n_0 = 0.1 \text{ H cm}^{-3}$ . A subgrid model for stellar feedback is implemented, that includes all processes that may impart thermal and kinetic feedback on ambient gas.

Horizon-AGN implements continuous stellar feedback that incorporates momentum, mechanical energy and metals from stellar winds and Type II/Type Ia SNe. When considering stellar winds and Type II SNe, the *STARBURST99* model (Leitherer et al., 1999, 2010) is employed to generate look-up tables as a function of metallicity and age. The model employs the Padova tracks (Girardi et al., 2000), with thermally pulsating asymptotic branch stars (see e.g. Vassiliadis & Wood, 1993). The kinetic energy of the stellar winds is calculated using the ‘Evolution’ model of Leitherer et al. (1992).

The implementation of Type Ia SNe follows Matteucci & Greggio (1986) and assumes a binary fraction of 5% (Matteucci & Recchi, 2001). The chemical yields are taken from the W7 model of Nomoto et al. (2007). Stellar feedback is modelled as a heat source after 50 Myr (mainly to reduce computational cost). This is reasonable, given that, after 50 Myr, the bulk of the energy from stellar feedback is liberated via Type Ia SNe that have long time delays i.e. several hundred Myrs to a few Gyrs (e.g. Maoz et al., 2012).

These systems are not prone to strong radiative losses, as stars disrupt their dense birth clouds, or move away from them, after a few tens of Myrs (e.g. Blitz & Shu, 1980; Hartmann et al., 2001).

### 4.2.2 Identification of galaxies and mergers

We identify galaxies using the ADAPTAHOP structure finder (Aubert et al., 2004; Tweed et al., 2009), which is applied to the distribution of star particles. Structures are identified using a local threshold of 178 times the average matter density. The local density of individual star particles is measured using the 20 nearest neighbours, with structures that have more than 50 particles being considered as galaxies. This corresponds to a minimum identifiable stellar mass of  $10^{8.5} M_{\odot}$  and yields a catalogue of  $\sim 100,000$  galaxies with  $M_{\star} > 10^9 M_{\odot}$  at  $z = 0.06$ . We then produce merger trees for each galaxy, tracking their progenitors to  $z = 6$ . The average length of timesteps is  $\sim 130$  Myr.

We use the merger trees to identify the major (mass ratios  $> 1 : 4$ ) and minor (mass ratios between  $1 : 4$  and  $1 : 10$ ) mergers that each galaxy has undergone. It is worth noting that how the mass ratio is defined influences the identification of major and minor mergers (e.g. Rodriguez-Gomez et al., 2015). Here, we adopt the mass ratio when the mass of the less massive galaxy is at its maximum prior to coalescence – i.e. before material is transferred between the merging companions. This effectively measures the ‘true’ mass ratio of the system, before the merger process alters the properties of galaxies involved in the merger.

We note that BH growth is not prescriptively linked to mergers in the simulation. The growth in BH mass is simply a result of accretion from ambient gas, but will naturally respond to changes in the geometry and dynamics of the gas that is induced by a merger e.g. if major mergers efficiently funnel gas into the centre of a remnant then BH growth could be accelerated. However, the model is not set up to preferentially build BHs during mergers.

The minimum galaxy mass of  $10^{8.5} M_{\odot}$  imposes a limit on the minimum merger mass ratio that is detectable for a galaxy of a given mass. For example, of galaxies that have stellar masses around  $10^{9.5} M_{\odot}$  at  $z = 0.06$  (the final snapshot of the simulation), 96, 72 and 37 per cent are massive enough to detect a merger with a mass ratio of  $1 : 4$ , at  $z = 1$ ,  $z = 2$  and  $z = 3$  respectively. For mergers with mass ratios of  $1 : 10$ , the corresponding values are 84, 47 and 20 per cent at the same redshifts. For galaxies with stellar masses above  $10^{11} M_{\odot}$ , the merger history is at least 85 per cent complete for mass ratios greater than  $1 : 10$ , up to  $z = 3$ .

### 4.2.3 The growth of black holes and black-hole feedback on ambient gas

BH or ‘sink’ particles are seeded in the simulation wherever the gas density in a cell exceeds a critical threshold of  $\rho > \rho_0$  and the stellar velocity dispersion exceeds  $100 \text{ km s}^{-1}$ , where  $\rho_0 = 1.67 \times$

$10^{-25} \text{ g cm}^{-3}$  and corresponds to  $0.1 \text{ H cm}^{-3}$ , the minimum threshold for star formation. To prevent the formation of multiple BHs within the same galaxy, BHs cannot form while there is another BH within 50 kpc. BHs have an initial mass of  $10^5 M_{\odot}$ , which is chosen to match BH masses predicted a direct collapse scenario (e.g. Begelman et al., 2006). However, BH masses quickly become self regulated, so that the exact choice of seed mass is not important (Dubois et al., 2012).

BH seeding continues until  $z = 1.5$ , after which no new BHs are allowed to form. This is purely to prevent an unmanageable number of BHs from being formed, and has a negligible effect on the growth of massive BHs. Almost all late forming BHs do so in low mass galaxies, and by  $z = 0$ , the BH occupation fractions of massive galaxies are in agreement with observational estimates (e.g. Trump et al., 2015).

Following their formation, each BH is able to grow through gas accretion, or through coalescence with another black hole (Dubois et al., 2014, 2016). Accretion is modelled using the Bondi-Hoyle-Lyttleton rate:

$$\dot{M}_{BH} = \frac{4\pi\alpha G^2 M_{BH}^2 \bar{\rho}}{(\bar{c}_s^2 + \bar{u}^2)^{3/2}}, \quad (4.1)$$

where  $M_{BH}$  is the mass of the BH,  $\bar{\rho}$  is the mass-weighted average gas density,  $\bar{c}_s$  is the mass-weighted average sound speed,  $\bar{u}$  is the mass-weighted average gas velocity relative to the BH and  $\alpha$  is a dimensionless boost factor which accounts for the inability of the simulation to capture the cold high-density inter-stellar medium and corrects for accretion that is missed due to unresolved gas properties (Booth & Schaye, 2009a). The effective accretion rate of the BH is capped at the Eddington accretion rate:

$$\dot{M}_{Edd} = \frac{4\pi G M_{BH} m_p}{\epsilon_r \sigma_T c}, \quad (4.2)$$

where  $m_p$  is the mass of a proton,  $\epsilon_r$  is the radiative efficiency, assumed to be  $\epsilon_r = 0.1$  for Shakura & Sunyaev (1973) accretion onto a Schwarzschild BH,  $\sigma_T$  is the Thompson cross-section and  $c$  is the speed of light. BHs are allowed to coalesce if they form a tight enough binary. Two black holes must be within four AMR cells of one another and have a relative velocity that is smaller than the escape velocity of the binary. The resulting mass of the merged binary is simply the sum of the masses of the two BHs.

BH feedback on ambient gas operates via a combination of two channels and depends on the ratio of the gas accretion rate to the Eddington luminosity,  $\chi = \dot{M}_{BH}/\dot{M}_{Edd}$ . For Eddington ratios  $\chi > 0.01$  (which represent high accretion rates) a ‘quasar’ mode is implemented, with 1.5 per cent of the accretion energy being injected isotropically into the gas as thermal energy. For Eddington ratios  $\chi < 0.01$  (which represent low accretion rates) a ‘radio’ mode is active, where cylindrical bipolar outflows are implemented with a jet velocity of  $10^4 \text{ km s}^{-1}$ . The quasar mode efficiency is chosen to reproduce the local  $M_{BH}-M_{\star}$  and  $M_{BH}-\sigma_{\star}$  relations, as well as the local cosmic black-hole mass density (Dubois et al.,

2012). Horizon-AGN is not otherwise tuned to reproduce the bulk observable properties of galaxies in the nearby Universe.

The effect of AGN feedback in Horizon-AGN is to regulate BH growth and star formation by preventing the accumulation of cold gas (Dubois et al., 2012, 2016). Rapid cosmological accretion in the early Universe leads to enhanced quasar mode activity and is the dominant mode of feedback for high redshift, gas-rich galaxies. As galaxies grow, they expel or consume their supply of cold gas leading to reduced BH accretion rates. As a result, the radio mode becomes increasingly important towards lower redshifts, eventually becoming the dominant mode of feedback in the low redshift Universe (Krongold et al., 2007; Best & Heckman, 2012; Dubois et al., 2012; Volonteri et al., 2016; Peirani et al., 2017).

Rather than being anchored to the centre of their dark matter haloes as in some other simulations (e.g Taylor & Kobayashi, 2014; Schaye et al., 2015; Sijacki et al., 2015), BHs are allowed to move freely, with a drag force applied in order to mitigate unrealistic motions and spurious oscillations arising from the effect of a finite particle resolution. BHs must, therefore, be matched with a host galaxy, since they are not explicitly assigned to a host galaxy by the simulation. We assign a BH to a host galaxy only if it lies within twice the effective radius of a galaxy structure and within 10 per cent of the virial radius of its dark matter halo. By this definition, a majority of massive galaxies ( $M_{\star} > 10^{10}M_{\odot}$ ) at  $z \sim 0$  are host to a BH (Volonteri et al., 2016). In practice, almost all single luminous ( $L_{bol} > 10^{43}$ ) BHs are found at the centres of their host galaxies. Binary BHs account for a significant fraction of the off-centre BH population, with single off-centre BHs accounting for less than 1 per cent of the total population of luminous BHs (Volonteri et al., 2016).

#### 4.2.4 Galaxy morphology: measurement of B/T ratios

We employ bulge-to-total (B/T) ratios calculated by Volonteri et al. (2016). Sérsic fits to the stellar mass profiles of our simulated galaxies are performed, which include a disc component with index  $n = 1$ , plus a second ‘bulge’ component with index  $n = [1, 2, 3, 4]$ , with the best-fitting component used for our analysis. The mass associated with each component is measured, from which the B/T ratio is calculated.

Observational studies of bulge-less galaxies can differ slightly in their definition of a ‘bulge-less’ system (e.g. Kormendy et al., 2010; Jiang et al., 2011; Simmons et al., 2012, 2013; Secret et al., 2012; Marleau et al., 2013). Here, we follow Marleau, Clancy & Bianconi (2013) and define objects as bulge-less if they have  $B/T < 0.1$ . In the final snapshot of the simulation ( $z = 0.06$ ), 2.8 per cent of galaxies are classified as bulge-less. At  $z = 0.5$  and  $z = 2.5$ , the corresponding values are 2.5 per cent and 5.5 per cent respectively. Note that the second component of the fit in around half of these bulge-less galaxies has an index of  $n = 2$  or below, (with a similar fraction for galaxies that are not classified as bulge-less), possibly indicating a pseudo-bulge, in broad agreement with Simard et al. (2011) for sufficiently resolved galaxies.

### 4.3 Mergers and production of bulges

The role of galaxy mergers in driving morphological transformation, as a function of the properties of the merging progenitors (e.g. stellar mass, gas fraction, orbital configuration, local environment and redshift) will be addressed in detail in a forthcoming paper (Martin et al. in prep). Here, we outline some aspects of merger-driven bulge formation that are relevant to this study.

We begin by exploring the hypothesis that mergers are primarily responsible for the production of bulges and, as is often assumed in observational studies (e.g. Kormendy & Kennicutt, 2004; Satyapal et al., 2009, 2014; Schawinski et al., 2011; Bizzocchi et al., 2014), that galaxies that do not contain significant bulge components (e.g. those with B/T ratios less than 0.1) must not have undergone significant merger activity. This assumption is typically motivated by idealised simulations of isolated mergers (e.g. Di Matteo et al., 2005; Hopkins et al., 2006) which do not, therefore, place the merging system in a cosmological context or realistically sample the parameter space. While they share similar physics to their cosmological counterparts, e.g. in terms of prescriptions for BH growth (typically Eddington-limited Bondi accretion), star formation and implementation of other baryonic processes, idealised simulations do not allow for statistical studies of galaxy evolution, nor do they model a galaxy’s wider environment. They cannot, for example, account for cosmological accretion from filaments or cooling of hot halo gas, which may contribute to continued stellar mass growth and rebuilding of disks subsequent to the merger.

Disks could also regrow simply from residual gas from the merger progenitors, in cases where the initial gas fractions are extremely high. Such processes could act to increase the total mass of the galaxy, without necessarily growing the bulge, and therefore work to reduce the B/T ratio of the galaxy. The assumption that bulge-less galaxies have undergone no major mergers could, therefore, depend somewhat on the epoch at which a merger takes place, and the accretion and star formation history of the galaxy (e.g. Sparre & Springel, 2017). For example, galaxies in the high-redshift Universe exhibit high gas fractions (e.g. Tacconi et al., 2010; Geach et al., 2011), which may enable (gas-rich) merger remnants to regrow disks, either via cosmological accretion and/or gas left over after the merger (e.g. Springel & Hernquist, 2005; Athanassoula et al., 2016; Font et al., 2017).

In Fig 4.1, we study the effect of major mergers (mass ratios  $> 1 : 4$ ) on the B/T fraction of galaxies in the local Universe. For the purposes of the analysis in this section, we limit ourselves to galaxies with stellar masses greater than  $10^{10.5}M_{\odot}$ , because their merger histories are relatively complete. Our sample is  $> 80$  per cent complete for mergers of mass ratios  $1 : 4$  at  $z = 3$  (i.e. more than 80 per cent of  $10^{10.5}M_{\odot}$  galaxies were massive enough at  $z = 3$  that a  $1 : 4$  merger would be detectable) and  $> 95$  per cent complete at  $z = 1$  and  $z = 2$ .

Fig 4.1 shows the mean number of major mergers as a function of the final B/T ratio of the local galaxy. Blue, green and red lines indicate the mean number of mergers since  $z = 1$ ,  $z = 2$  and  $z = 3$  respectively. The probability that bulge-less galaxies ( $B/T < 0.1$ ) have undergone any major mergers between  $z = 1$  and the present day is essentially zero. However, while the assumption that bulge-less galaxies have

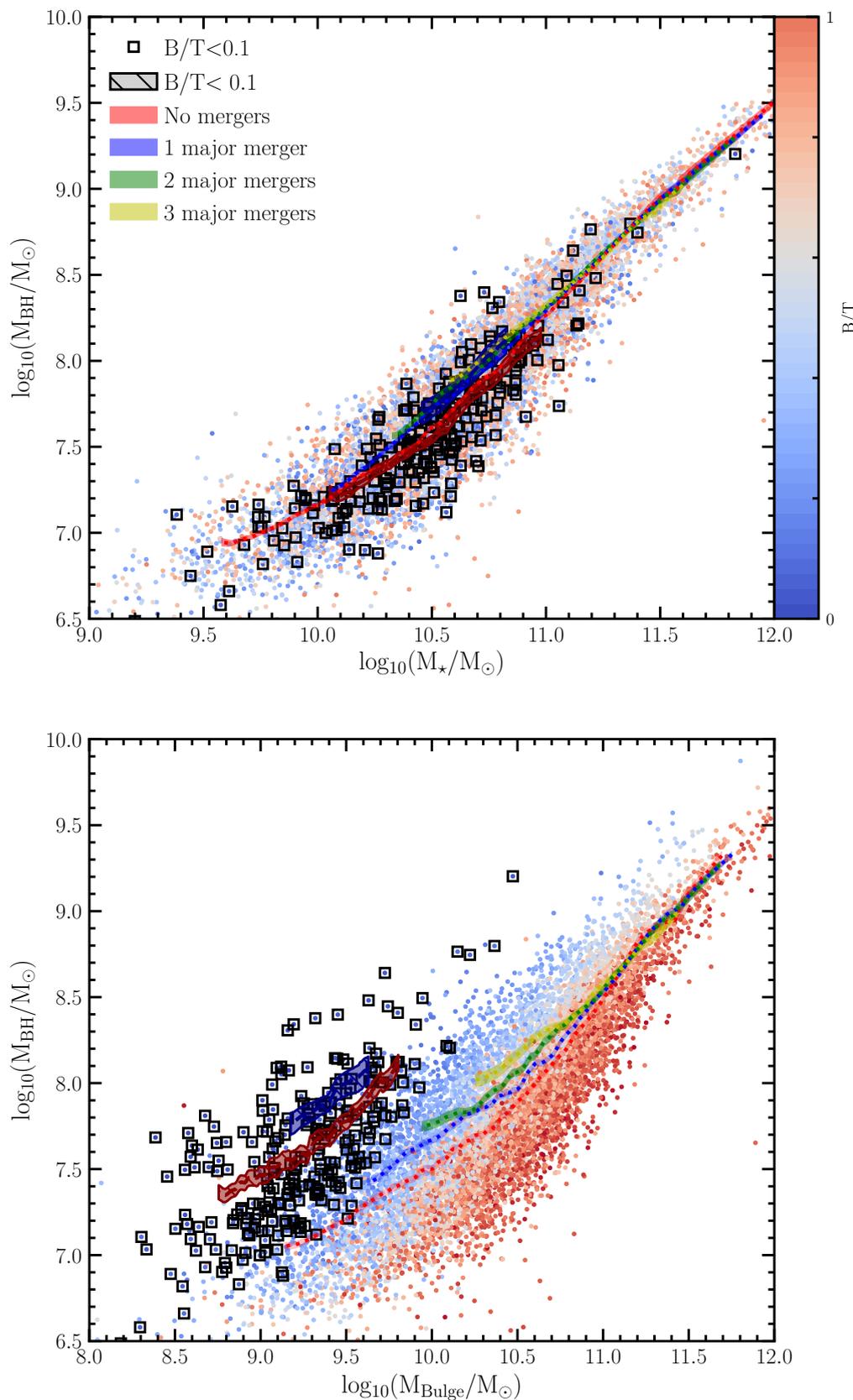


FIGURE 4.3: **Top panel:**  $M_{\text{BH}}$  vs  $M_{\star}$  relation from Horizon-AGN for local massive galaxies. Galaxies are indicated by points and bulge-less galaxies are indicated by squares. Dotted coloured lines show a running mean for galaxies that have undergone 0, 1, 2 and 3 major mergers since  $z = 3$  (see legend for colour coding), where the region around the line indicates the standard error on the mean. Darker coloured dashed lines with hatched regions indicate the bulge-less population only. **Bottom panel:** Same as the top panel but now with bulge mass on the  $x$ -axis instead of total stellar mass.

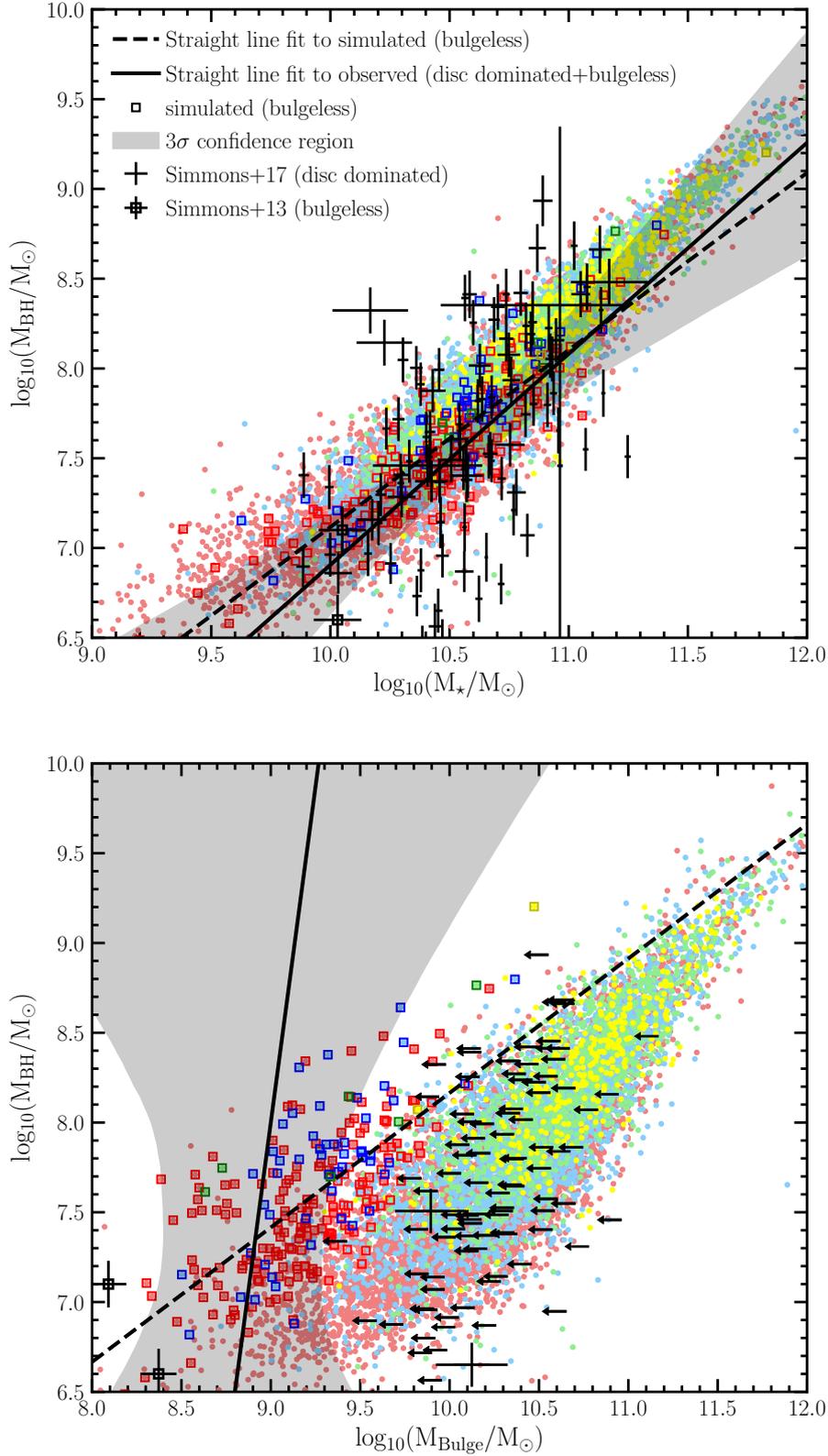


FIGURE 4.4: **Top panel:**  $M_{\text{BH}}$  vs  $M_{\star}$  relation for local, massive galaxies in Horizon-AGN, with points colour coded by the number of mergers as in Fig 4.3. Square symbols indicate bulge-less ( $B/T < 0.1$ ) galaxies. Black symbols indicate observational data from Simmons et al. (2013, 2017). Dashed and solid lines show linear fits to the simulated and observed data points respectively, with the grey filled region indicating the  $3\sigma$  confidence region from the fit to the observed points (Simmons et al., 2017). **Bottom panel:** Same as the top panel with bulge mass on the  $x$ -axis instead of total stellar mass. The arrows represent upper limits on the bulge mass. The fit properly incorporates the bulge-mass upper limits as censored data, which results in a large confidence region due to the large uncertainty on the bulge mass of these galaxies (Simmons et al., 2017).

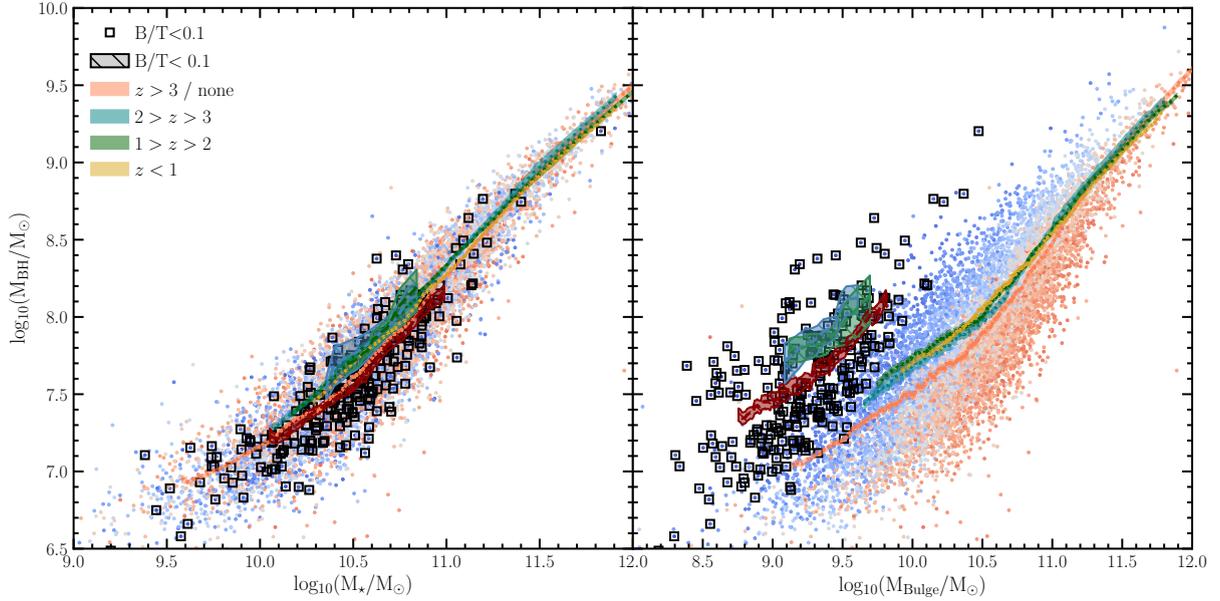


FIGURE 4.5:  $M_{\text{BH}}$  vs  $M_*$  (left) and  $M_{\text{BH}}$  vs  $M_{\text{Bulge}}$  (right) relations for local massive galaxies, with the dotted coloured lines showing a running mean for galaxies that underwent their last major merger before  $z = 3$ , between  $z = 3$  and  $z = 2$ , between  $z = 2$  and  $z = 1$  and after  $z = 1$  (see legend for colour coding). The widths of the lines indicate the standard errors on the mean. Darker coloured dashed lines with hatched regions indicate the same for bulge-less galaxies.

undergone no major mergers after  $z = 1$  appears to be a good one, it is necessary to relax this somewhat towards higher redshifts. For example, bulge-less galaxies have, on average,  $\sim 0.25$  and  $\sim 0.35$  major mergers since  $z = 2$  and  $z = 3$  respectively. In other words, around one in four and one in three of these galaxies have undergone a major merger since  $z = 2$  and  $z = 3$  respectively. Note that a qualitatively similar picture emerges when considering minor mergers (mass ratios between  $1 : 4$  and  $1 : 10$ ). A small fraction ( $\sim 20$  per cent) of galaxies with low B/T values have had a minor merger since  $z = 1$ , indicating that some galaxies have survived recent low mass ratio mergers, without producing a significant bulge component.

The major-merger histories described above suggest that disk rebuilding plays some role in the merger history of even those galaxies that do not exhibit a strong bulge component at the present day (although such events are relatively rare). The effect of disk rebuilding (e.g. from cosmological accretion of cold gas or continued stellar mass growth from residual gas after a merger) is largely to wash out some of the morphological (i.e. disk to bulge) transformation produced by high-redshift mergers. We note that, for high values of B/T ( $> 0.7$ ), the number of major mergers decreases. This is because the most massive galaxies (e.g.  $M_* > 10^{11} M_\odot$ ) typically dominate the high B/T population (e.g. Dubois et al., 2016) and there are, by definition, not many systems of similar mass. Thus, these galaxies are less likely to experience major mergers.

It is also instructive to directly consider the fraction of stellar mass in a galaxy that did not form in-situ. Using a raw number of mergers could be misleading, because the impact of a merger on the final morphology of a galaxy at  $z = 0$  depends, to some extent, on the final mass of the galaxy, in addition

to the mass of the galaxy at the time of the merger, since, as discussed above, subsequent secular stellar mass growth, in effect, dilutes the merger’s contribution to the bulge mass. We define the ex-situ mass fraction as  $M_{\text{exsitu}}/M_*$ , where  $M_{\text{exsitu}}$  is the total stellar mass accreted from other galaxies, calculated from each galaxy’s merger tree. Fig 4.2 shows the mean ex-situ mass fraction as a function of the B/T ratio. We find that, similar to Fig 4.1, bulge-less galaxies host very low ex-situ mass fractions – less than 15 per cent, on average. While a fraction of bulge-less galaxies have experienced major mergers at high redshift, continued stellar mass growth significantly diminishes the contribution of these events to the final mass of the galaxy at the present-day.

Overall, the assumption that bulge-less galaxies have not undergone significant major-merger activity at recent ( $z < 2$ ) epochs is robust. Progressively lower B/T ratios show rapidly diminishing probabilities for merger activity, with bulge-less galaxies indeed showing comparatively little merger activity over cosmic time.

## 4.4 BH growth over cosmic time

### 4.4.1 Is there a correlation between BH growth and merger history?

We proceed by studying the  $M_{\text{BH}}-M_{\text{Bulge}}$  and  $M_{\text{BH}}-M_*$  correlations in the local Universe. If mergers are primarily responsible for feeding BHs, we would expect the population of bulge-less galaxies to fall on the same  $M_{\text{BH}}-M_{\text{Bulge}}$  and  $M_{\text{BH}}-M_*$  relations as the rest of the galaxy population. This is because, in the absence of mergers, the bulge-less population would have both small bulges and small BHs. Conversely, if BH feeding was preferentially produced by secular processes and accretion onto the host galaxy and *not* by galaxy mergers, then we would expect the bulge-less systems to lie on  $M_{\text{BH}}-M_*$  populated by the general galaxy population, but to be offset from the main  $M_{\text{BH}}-M_{\text{Bulge}}$  locus. This is because, while secular processes steadily build their BH and stellar mass over cosmic time, their bulges will be under-massive due to the lack of major mergers. Here, we use our full sample of galaxies down to  $10^9 M_\odot$ , so there is some incompleteness in terms of detecting high redshift mergers towards the low mass end. However, the majority of our galaxy sample have stellar masses of  $10^{10} M_\odot$  and greater, and are therefore almost entirely complete in their merger histories.

The top panel of Fig 4.3 shows the BH mass as a function of the total stellar mass of the galaxy ( $M_*$ ). Solid coloured lines show a running mean for galaxies that have undergone 0, 1, 2 and 3 major mergers since  $z = 3$  (see legend for colour coding), where the width of the line indicates the standard error on the mean. Hatched regions indicate the same for the bulge-less population only. The general galaxy population is shown using the coloured dots, with the colours indicating the B/T of the galaxy in question.

The top panel of Fig 4.3 indicates that the number of major mergers that a galaxy has undergone does not significantly alter its position on the main locus of the  $M_{\text{BH}}-M_{\star}$  correlation (offsets are visible for low-mass galaxies but these are small,  $< 0.1$  dex per major merger). Additionally, the hatched lines, which indicate the bulge-less population, are completely consistent with the main locus of the correlation. This is evidence that mergers are not the principal driver of BH feeding, since, if that were the case, galaxies with a larger number of mergers would exhibit relatively over-massive BHs and be offset from the main locus. The bottom panel of Fig 4.3 shows the corresponding plot for the  $M_{\text{BH}}-M_{\text{Bulge}}$  correlation. In this plot, the population of bulge-less galaxies lies offset above the locus traced by the general population, driven by the fact that these galaxies have under-massive bulges (due to a smaller number of mergers). As the colour coding of the points shows, in general, galaxies with lower bulge masses tend to have higher BH masses and bulge-less galaxies simply represent the tail of this trend. The fact that this trend is not present in the  $M_{\text{BH}}-M_{\star}$  is strong evidence that processes that grow the bulge are not also responsible for BH growth.

Fig 4.4 shows the same galaxies as in Fig 4.3, now with each point colour-coded by the number of mergers the galaxy has undergone, using the same colour scheme as the lines in Fig 4.3. The bulge-less population is indicated by squares. Over-plotted are the sample of bulge-less galaxies from Simmons et al. (2013) and the sample of disc dominated galaxies from Simmons et al. (2017), with linear best fits to the observed (solid line) and simulated (dashed line) galaxies. The fits to the observed sample of disc-dominated and bulge-less galaxies are a linear regression performed by Simmons et al. (2017), incorporating errors and limits in both dimensions. The fit to the points in the bottom panel properly incorporates the bulge mass upper limits as censored data, which results in a large confidence region, due to the large uncertainty on the bulge mass of these galaxies (see Simmons et al., 2017, for more details).

As the top panel in this figure shows, the linear fit to the simulated bulge-less galaxies matches the slope and normalisation of the fit to the observed data, within the parameter space defined by Simmons et al. (2017). In the bottom panel, both simulated and observed bulge-less or disk-dominated galaxies lie above the  $M_{\text{Bulge}}-M_{\text{BH}}$  relation. Although the slope of the fit to the simulated datapoints does not exactly match that of the observed data, both describe the same qualitative picture. Note that the bulk of the observed bulge masses in the bottom panel are limits. The simulated galaxies are consistent with those limits, and 3 out of 4 of the observed bulge-less galaxies which have precise measurements lie along the locus defined by the simulated bulge-less galaxies. Additionally, the majority of the simulated data points lie within the parameter space defined by the observed points.

Fig 4.5 again shows these two correlations for local simulated galaxies, but this time indicates how the positions of galaxies may vary, given the redshift at which their *last* major merger took place. In a similar vein to Fig 4.3, we find that the position of local galaxies remains largely unchanged in either correlation, irrespective of when they had their last major-merger event. Indeed, galaxies that have had major mergers around the epoch of peak cosmic star formation ( $2 < z < 3$ ) do not deviate from the main

locus of the correlation, indicating that the gas richness of these major mergers have little impact on the overall growth of their BHs.

Our analysis so far has focussed on galaxies in the local Universe and has shown that mergers are unimportant in terms of the *cumulative* evolution of BHs over cosmic time. It is also instructive to study whether merger activity might have a transient impact on the  $M_{\text{BH}}-M_{\text{Bulge}}$  and  $M_{\text{BH}}-M_{\star}$  correlations at high redshift. We complete our analysis by studying the redshift evolution of these correlations, and exploring whether the impact of major mergers may be higher in the high-redshift Universe. In Figures 4.6 and 4.7, we show the redshift evolution of these correlations in the simulation, with mean locii indicated for galaxies that have had 0, 1, 2 and 3 major mergers before the redshift in question ( $z = 0, 0.5$  and  $2.5$ ; which correspond to look-back times of 0, 5 and 11 Gyrs respectively) shown using the coloured lines. The colour coding is the same as that used in Fig 4.3.

This figure shows that the number of major mergers a galaxy experiences does not alter its position on the *evolving* correlations as a function of redshift. Indeed, if major mergers were the principal driver of BH growth, then galaxies would be expected to show large offsets from the mean locus (which would induce a large scatter), before enough merging has taken place to put them on the relation at the present day. However, Fig 4.6 indicates a persistently tight correlation, as these relations build up steadily over cosmic time, the opposite to what would be expected if BH growth were episodic and driven by largely stochastic events like major mergers. *Thus, major-merger activity of any kind is unlikely to be driving significant BH growth at any epoch.*

Our analysis suggests that whatever processes dominate the *overall* stellar-mass growth of the galaxy population over cosmic time, also drive the growth of their constituent BHs. Furthermore, BH mass does not correlate as well with the part of the galaxy, i.e. the bulge, that is preferentially built in mergers. Together, this indicates that BH growth tends to occur largely by secular means, without recourse to mergers.

#### 4.4.2 Contribution of mergers to the cosmic BH accretion budget

So far, we have shown that galaxies that have undergone mergers do not lie on a different  $M_{\text{BH}}$  vs  $M_{\star}$  relation to their non-merging counterparts. While this is evidence that BH growth does not preferentially take place in mergers, it is useful to precisely quantify the fraction of the cosmic BH accretion budget which is directly attributable to the merger process over cosmic time. Recent work, that has studied the proportion of the star formation budget that is directly driven by major and minor mergers (Martin et al., 2017), has shown that only 25 per cent of the stellar mass in today's Universe is directly triggered by merging, with major and minor mergers accounting for 10 and 15 per cent of this value respectively. Here, we perform a corresponding study of BH growth and quantify the proportion of the BH accretion budget that is attributable to major and minor mergers.

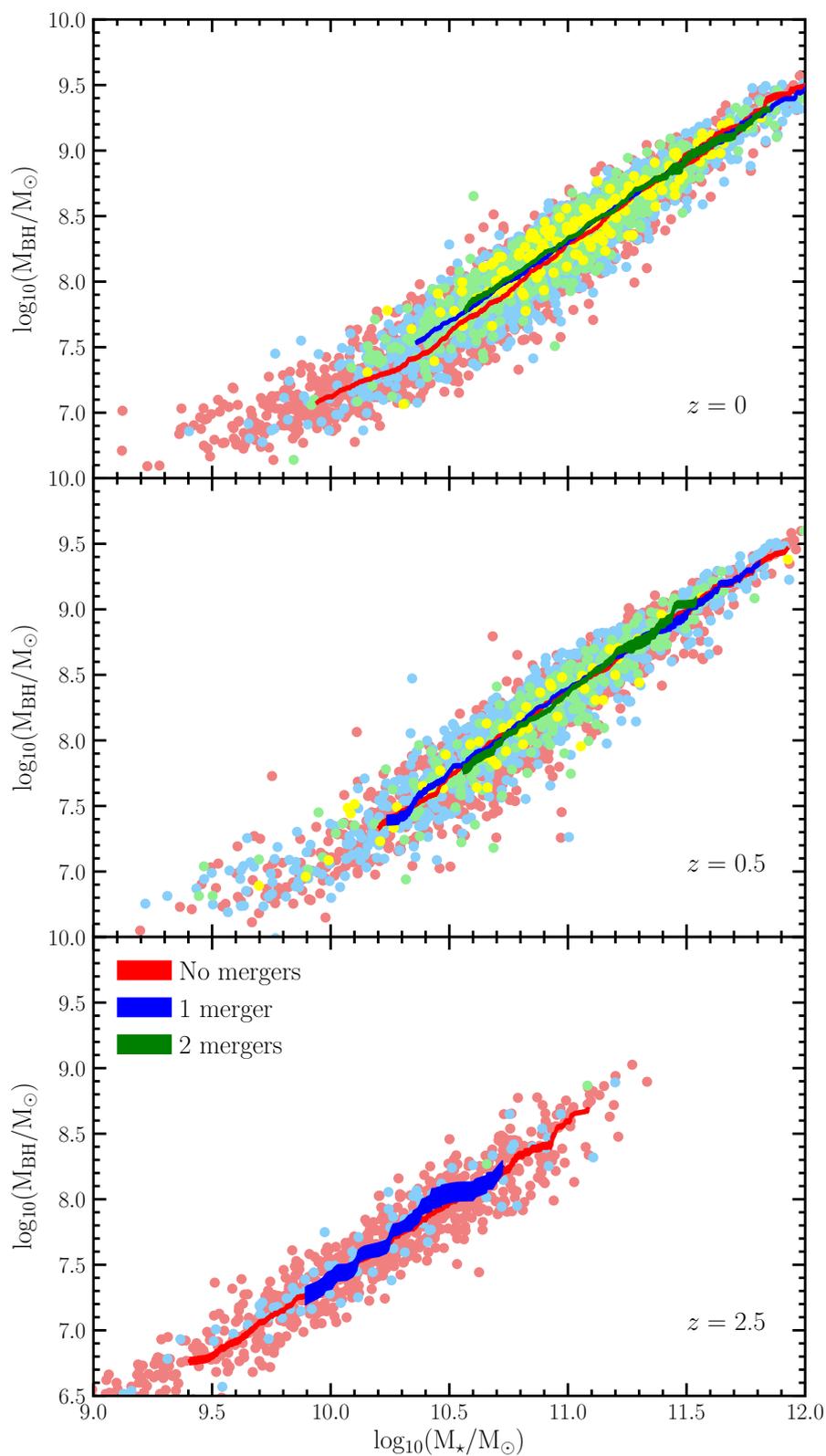


FIGURE 4.6: Evolution of the  $M_{\text{BH}}-M_*$  relation in Horizon-AGN for local massive galaxies. Solid coloured lines show a running mean for galaxies that have undergone 0 (red), 1 (blue) and 2 (green) major mergers before the redshift indicated in each panel, where the width of the line indicates the standard error on the mean.

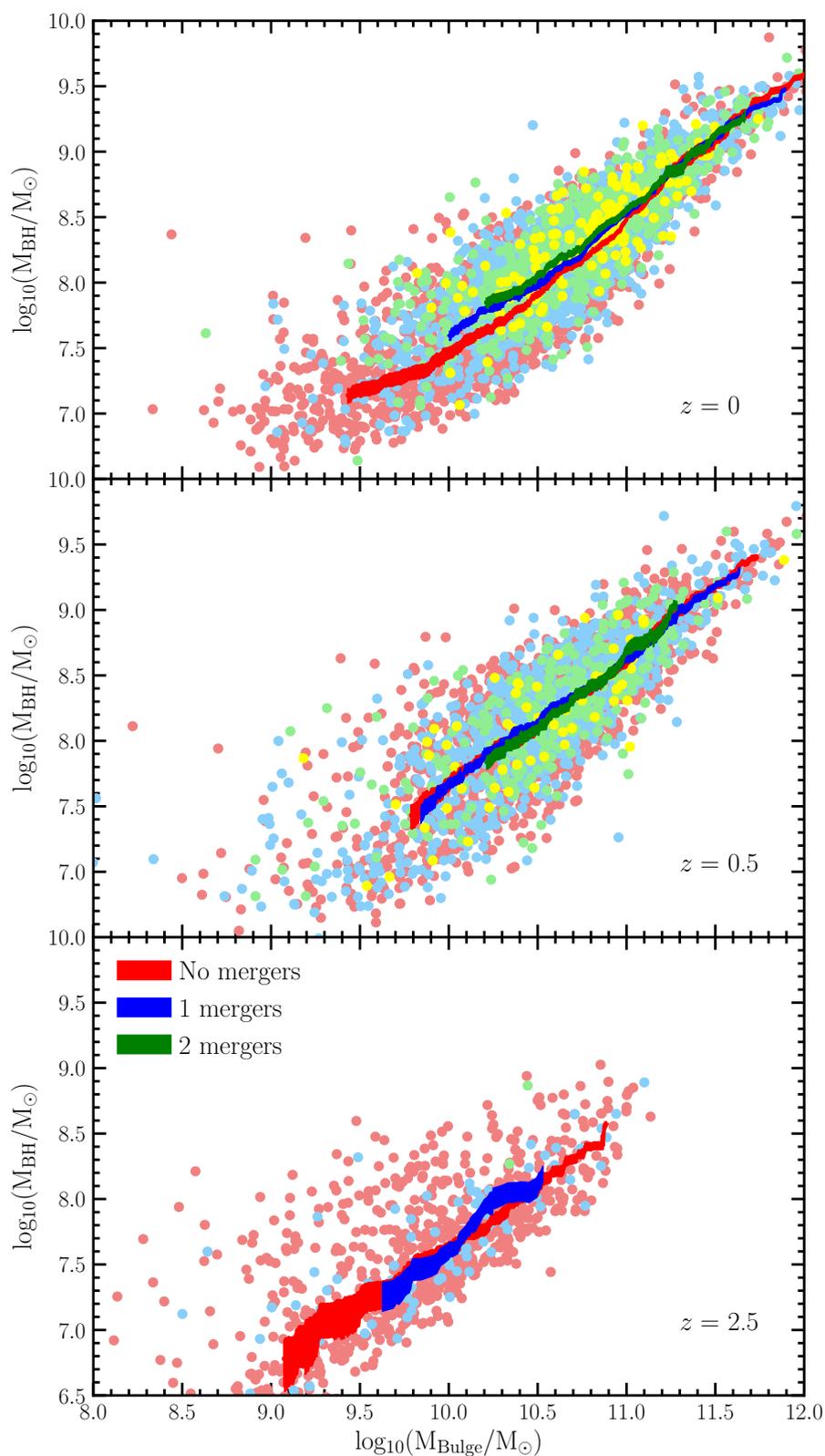


FIGURE 4.7: Evolution of the  $M_{\text{BH}}-M_{\text{Bulge}}$  relation in Horizon-AGN. As in Fig 4.6, solid coloured lines show a running mean for galaxies that have undergone 0 (red), 1 (blue) and 2 (green) major mergers before the redshift indicated in each panel, where the width of the line indicates the standard error on the mean.

We perform our analysis by tracking the mass evolution of each of the BHs hosted by one of our galaxies at  $z = 0$ . In a similar vein to Martin et al. (2017) who studied merger-driven star formation activity, we first define a merger-driven enhancement of the BH accretion rate,  $\xi$ , as the ratio of the mean specific BH accretion rate in the merging galaxies to that in their non-merging counterparts:

$$\xi(M_{\text{BH}}, z) = \frac{\langle \dot{M}_{\text{BH}}/M_{\text{BH}}(M_{\text{BH}}, z) \rangle_m}{\langle \dot{M}_{\text{BH}}/M_{\text{BH}}(M_{\text{BH}}, z) \rangle_{\text{non}}}, \quad (4.3)$$

where  $\dot{M}_{\text{BH}}$  is the BH accretion rate.  $\langle \dot{M}_{\text{BH}}/M_{\text{BH}}(M_{\text{BH}}, z) \rangle_m$  is the mean specific accretion rate for the merging population at a given redshift,  $z$ , and  $\langle \dot{M}_{\text{BH}}/M_{\text{BH}}(M_{\text{BH}}, z) \rangle_{\text{non}}$  is the same for galaxies that are not merging. Galaxies are defined as merging if they have had undergone a merger (major or minor) within the last Gyr or will undergo a merger in the next Gyr. Our results are robust to changes in this timescale: doubling or halving this number changes the contribution of mergers to the cosmic star formation budget by less than 5 per cent.

We use this enhancement to estimate the fraction of BH accretion that would have occurred in the merger progenitors anyway, had they not been in the process of merging. For example, if  $\xi$  is a factor of 2 then, on average, around half the BH accretion in the merging system in question is likely driven by other processes (see e.g. Kaviraj et al., 2013b; Lofthouse et al., 2017a; Martin et al., 2017, for a similar discussion of star formation activity). By subtracting the BH accretion that would have occurred anyway, had the merger not taken place, we can then measure the fraction of BH accretion that is *directly triggered* by mergers ( $f$ ) as follows:

$$f = \frac{m_{\text{new},m}(M_{\text{BH}}, z) [1 - 1/\xi(M_{\text{BH}}, z)]}{m_{\text{new},\text{total}}(M_{\text{BH}}, z)}, \quad (4.4)$$

where  $m_{\text{new},m}(M_{\text{BH}}, z)$  is the total mass accreted onto BHs in merging systems in a given BH mass and redshift bin and  $m_{\text{new},\text{total}}(M_{\text{BH}}, z)$  is the total mass accreted onto BHs in the stellar mass and redshift bin in question<sup>1</sup>. To ensure that our sample is complete down to a merger mass ratio of 1:10, we restrict ourselves to galaxies with stellar masses  $M_{\star} > 10^{9.5} M_{\odot}$  at all redshifts.

Fig 4.8 shows the cosmic BH accretion rate density in Horizon-AGN as a function of redshift. The BH accretion rate density decreases with redshift. The contribution due to mergers increases towards  $z = 1$ , peaking at around  $z = 1.1$ , and decreasing towards the present day. At all times, major mergers outweigh the contribution of minor mergers to the BH accretion rate density, even though minor mergers account for the majority of galaxy interactions (e.g. Lotz et al., 2011; Kaviraj et al., 2015b).

Since black hole accretion in Horizon-AGN is modelled using the Bondi-Hoyle-Lyttleton rate (Equation 4.1), the increase in gas density around the BH corresponds directly to an increase in the accretion rate.

<sup>1</sup>Equation 4 above is the BH-accretion equivalent of Equation 2 in Martin et al. (2017).

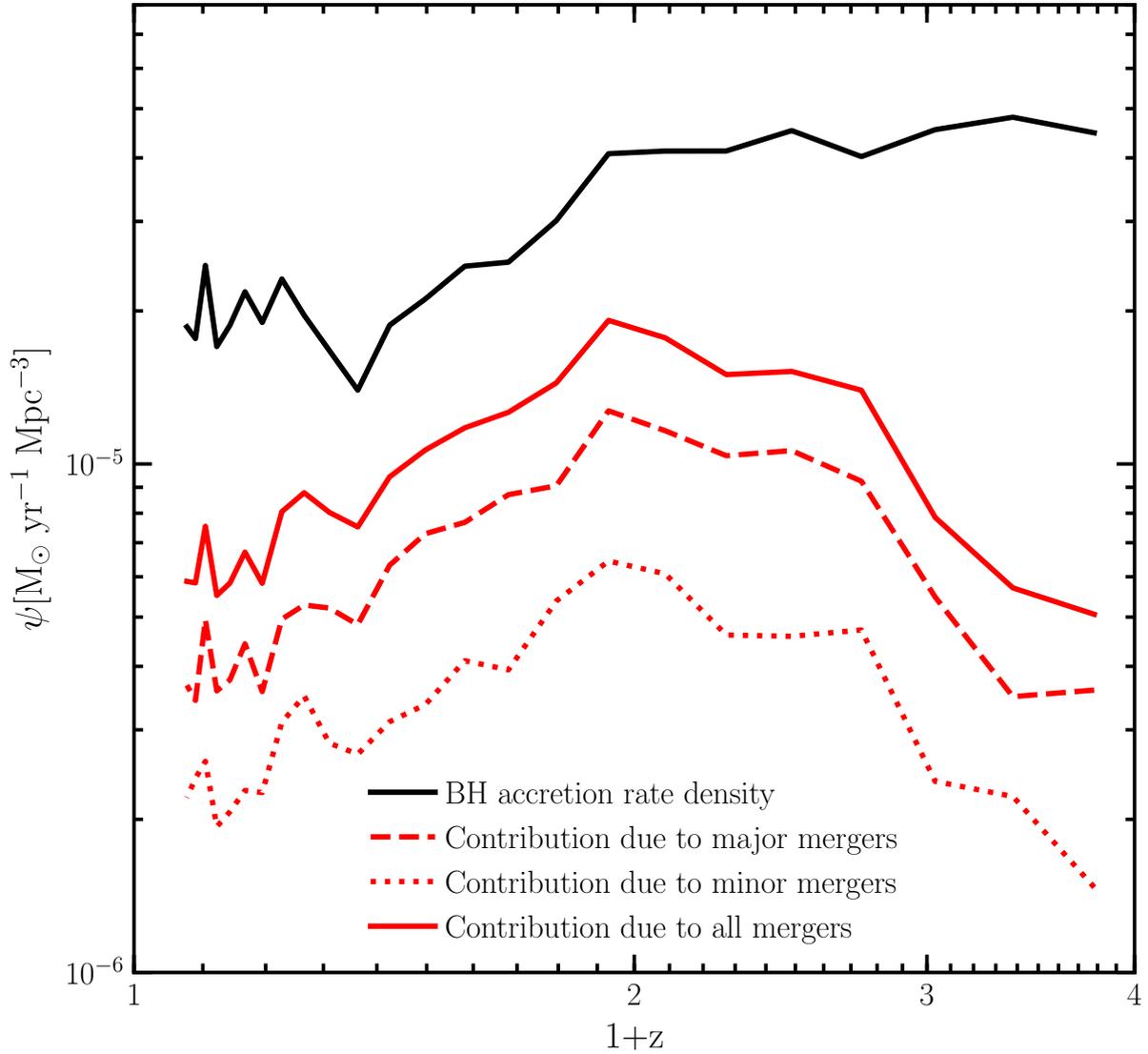


FIGURE 4.8: The BH accretion rate density for galaxies with  $M_* > 10^{9.5} M_\odot$  as a function of redshift from Horizon-AGN (black). The red lines indicate the portion of the BH accretion rate density that is a direct result of major (dashed line), minor (dotted line) and major + minor (all) mergers (solid line). The small jump in accretion at low redshift corresponds to the introduction of an additional grid refinement level at  $z = 0.26$ .

Mergers are least significant at high redshift, where galaxies already host high densities of gas (e.g. Geach et al., 2011), which enables efficient BH growth through the secular accretion of low angular momentum gas over short timescales (Dubois et al., 2012).

The small jump in accretion rate density observed at low redshift is due to the implementation of an additional AMR grid refinement at  $z = 0.26$ . This increases the local density in gas cells, thus increasing the accretion rate onto the black hole. Mass accretion after  $z = 0.26$  only accounts for 12 per cent of total mass accreted by black holes since  $z = 3$ , so the effect of grid refinement does not alter our qualitative conclusions.

Fig 4.9 shows the cumulative fraction of BH mass (in galaxies more massive than  $10^{9.5}M_{\odot}$ ) that is triggered by major and minor mergers as a function of redshift. At the present day, only  $\sim 35$  per cent of the BH mass in massive galaxies is directly attributable to the merger process – of this  $\sim 22$  per cent is driven by major mergers while the rest ( $\sim 13$  per cent) is driven by minor mergers. Mergers are, therefore, minority contributors to the BH accretion budget over cosmic time. It is worth noting that these values are not a strong function of galaxy mass. The fraction of BH mass that is created as a direct result of mergers increases from  $\sim 25$  per cent in galaxies with stellar masses around  $10^{10}M_{\odot}$  to  $\sim 40$  per cent in galaxies with stellar masses of  $10^{11.5}M_{\odot}$  or greater. However, across the range of stellar masses considered in this study, the majority of the BH mass is created via secular processes, not mergers.

Finally, we note that, while only  $\sim 25$  per cent of black hole growth globally is the direct result of major mergers, a small fraction of galaxies do grow most of their BH mass during major mergers. 28 per cent of galaxies that have undergone at least one major merger since  $z = 3$  have more than half of their total black hole mass built up as a direct result of major mergers during this time; this number is reduced to just 12 per cent when all galaxies are taken into account. The fact that the BH growth of a small fraction of galaxies is dominated by merging is likely responsible for the small increase in scatter towards higher redshift indicated by Fig 4.7.

## 4.5 Summary

A consistent picture is now emerging of the role that galaxy mergers play in driving stellar mass and BH growth across cosmic time, and particularly in the early Universe. Both theoretical and observational work now indicates that major mergers (and mergers in general) do not enhance star-formation activity around the epoch of peak cosmic star formation (e.g. Lofthouse et al., 2017a; Fensch et al., 2017). In other words, the bulk of the star formation that takes place at these epochs is driven secularly via cosmological accretion and not triggered by merging. And since the bulk of the stellar mass in today’s galaxies forms around this epoch, the majority of today’s stellar mass ( $\sim 75$  per cent, see Martin et al., 2017) is also unrelated to merging.

This particular study has used Horizon-AGN, a cosmological hydrodynamical simulation, to extend this analysis to BH growth. Our results indicate that a similar picture to that for star formation activity likely holds for accretion on to BHs. The majority ( $\sim 65$  per cent) of the cumulative BH growth in today’s massive galaxies takes place via secular processes, with the remaining  $\sim 35$  per cent attributable to either major or minor mergers. Our key findings can be summarised as follows:

- *Almost all bulge-less galaxies have undergone no major mergers since  $z = 1$ .* However,  $\sim 25$  per cent of such systems have had a major merger since  $z = 3$  (although, on average, more than 85 per cent of their stellar mass at  $z = 0$  is formed in-situ), indicating that disk rebuilding in gas-rich

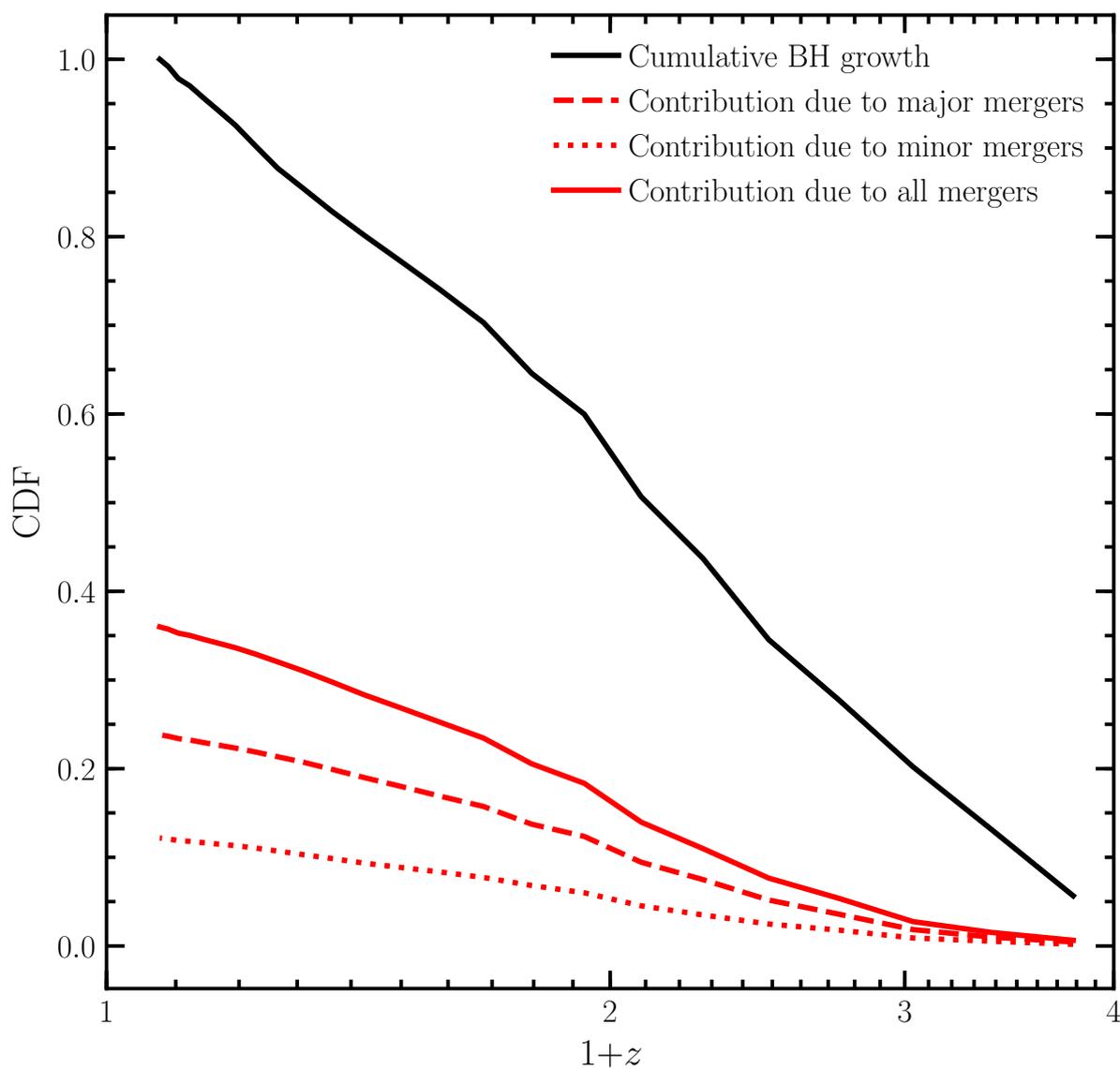


FIGURE 4.9: The cumulative fraction of BH mass in today's massive galaxies that has already been assembled as a function of redshift (black line). The contribution from major mergers, minor mergers and major + minor (all) mergers are shown using the dashed, dotted and solid red lines respectively. Only  $\sim 35$  per cent of the BH mass in massive galaxies at the present day is directly attributable to merger activity.  $\sim 22$  per cent is driven by major mergers and  $\sim 13$  per cent is driven by minor mergers. The bulk ( $\sim 65$  per cent) of the BH mass build-up over cosmic time is unrelated to merging and is therefore be a result of secular processes.

mergers may play a role in building these systems. However, the assumption made in many observational studies, that bulge-less galaxies have undergone no major mergers over most of cosmic time, is typically robust.

- *Bulge-less galaxies lie on the same  $M_{\text{BH}}-M_{\star}$  relation as the general galaxy population.* The number of major (mass ratios greater than 1 : 4) or minor mergers (mass ratios between 1 : 4 and 1 : 10) that a galaxy has undergone does not alter a galaxy's position on the  $M_{\text{BH}}-M_{\star}$  relation, indicating that mergers are not a significant mechanism for feeding the BH.
- *Bulge-less galaxies lie offset from the  $M_{\text{BH}}-M_{\text{Bulge}}$  relation observed in the general population.* This relation is not as tight as the  $M_{\text{BH}}-M_{\star}$  relation, with the number of mergers having a larger effect on the position of a galaxy on the  $M_{\text{BH}}-M_{\text{Bulge}}$  relation than on the  $M_{\text{BH}}-M_{\star}$  relation. The offset of the bulge-less galaxies is driven by the fact that these galaxies have normal black holes but under-massive bulges (due to a smaller number of mergers).
- *Mergers are directly responsible for a minority of BH growth over cosmic time.* Only  $\sim 35$  per cent of the BH mass in galaxies more massive than  $10^{9.5}M_{\odot}$  in today's Universe is directly attributable to mergers.  $\sim 22$  per cent is driven by major mergers and  $\sim 13$  per cent is driven by minor mergers. Secular processes, therefore, account for the creation of the majority ( $\sim 65$  per cent) of BH mass over the lifetime of the Universe.

## Chapter 5

# The role of mergers in driving morphological transformation over cosmic time

### 5.1 Introduction

As predicted by hierarchical structure formation scenarios (e.g. Fall & Efstathiou, 1980; van den Bosch et al., 2002; Agertz et al., 2011), high-redshift observations of massive galaxies indicate that the early Universe was dominated by systems possessing disc-like morphologies (e.g. Buitrago et al., 2014; Shibuya et al., 2015). In contrast, the morphological mix of today's Universe is dominated by massive galaxies with spheroidal morphologies (e.g. Bernardi et al., 2003; Conselice et al., 2014), with a majority of objects at low redshift hosting significant bulge components (e.g. Lintott et al., 2011). This disparity is evidence for significant structural evolution in the galaxy population over cosmic time, as a result of which discy, rotationally-supported galaxies are steadily transformed into spheroidal, dispersion-supported systems (e.g. Butcher & Oemler, 1984; Dressler et al., 1997; Postman et al., 2005; Smith et al., 2005; Conselice et al., 2008; Buitrago et al., 2014). Understanding the processes that drive this morphological transformation is, therefore, central to our comprehension of how galaxies have evolved over the lifetime of the Universe.

While empirical morphological classification schemes (e.g. Hubble, 1936; Cappellari et al., 2011) are largely defined using only visual or kinematic criteria, the morphological type of galaxies at the present-day is strongly aligned with their physical properties. Stellar masses, star-formation rates, colours, merger histories and local environment (e.g. Dressler, 1980; Dressler et al., 1997; Strateva et al., 2001; Hogg et al., 2002; Bundy et al., 2005; Conselice, 2006; Skibba et al., 2009; Bluck et al., 2014; Smethurst et al., 2015; Whitaker et al., 2015) all correlate strongly with galaxy morphology. This points towards a

picture of galaxy morphology that does not depend on a single mechanism for morphological change, at least across a broad range of masses and environments.

However, disentangling the role of different mechanisms in triggering morphological change remains difficult. Many processes are likely to be involved in the transformation of discs to spheroids, and the relative contribution of these processes is not well understood. For example, the theoretical literature has long highlighted the role of mergers in the creation of spheroidal systems, as the gravitational torques can remove stars from ordered rotational orbits in discy progenitors to chaotic orbits that form dispersion-supported spheroidal remnants. Major mergers i.e. those that involve progenitors with roughly equal mass, are considered to be particularly efficient at producing spheroidal systems (e.g. Toomre, 1977; Negroponte & White, 1983; Di Matteo et al., 2007; Hopkins et al., 2009a; Ferreras et al., 2009; Conselice et al., 2009; Taranu et al., 2013; Naab et al., 2014; Deeley et al., 2017), although minor mergers (i.e. those with unequal progenitor mass ratios) are also likely to play a role in the transformation of morphologies, by either producing chaotic stellar orbits as major mergers do, or by triggering disc instabilities (e.g. Dekel et al., 2009; Fiacconi et al., 2015; Zolotov et al., 2015; Welker et al., 2017). Processes other than galaxy mergers may also play a role in inducing morphological transformation. For example, in very dense environments, fly-bys (harassment) may act to make systems more spheroidal, and processes like ram-pressure stripping may act to suppress gas accretion (Moore et al., 1998; Abadi et al., 1999; Choi & Yi, 2017) which would otherwise spin galaxies up.

It is worth noting that, while the global morphological trend in the Universe is for discs to transform into spheroids, the reverse transformation is also possible (in individual events) through the accretion of gas, as this gas settles into rotational orbits and creates stars that add to the rotational component of the system. Indeed, in very gas-rich major mergers the residual gas may reform a disc, so that the remnant may be discy rather than spheroidal (e.g. Springel & Hernquist, 2005; Hau et al., 2008; Kannappan et al., 2009; Font et al., 2011; Aumer et al., 2013; Rodrigues et al., 2017; Sparre & Springel, 2017). At high redshift ( $z > 2$ ), cosmological accretion likely plays a dominant role in building up and reforming discs, especially in galaxies fainter than  $L^*$  (e.g. Brooks et al., 2009). During these epochs the dominant fuel for star formation and source of angular momentum acquisition in discs are filamentary inflows of cold gas, rather than accretion of shock-heated gas or hierarchical merging (e.g. Murali et al., 2002; Kereš et al., 2005; Brooks et al., 2009; Kimm et al., 2011; Pichon et al., 2011; Stewart et al., 2013; Martin et al., 2016; Welker et al., 2017). Coherent cold flows appear capable of reforming discs up to a critical mass of  $10^{10.5} M_{\odot}$ , after which the coherence of the flow is lost and the galaxy morphology is frozen in (Welker et al., 2017).

The orbital parameters of mergers and the spins of accreted satellites may also be an important factor in the morphological evolution of galaxies (e.g. Taylor et al., 2018). The alignment or misalignment of both the orbit and the spin of the satellite, relative to the spin of a massive accreting galaxy, may be an important factor in determining the evolution of their angular momentum at later times. The orbits of satellites have been shown to align progressively with the major axis (e.g. Yang et al., 2006) and spin

(e.g. Ibata et al., 2013; Welker et al., 2014, 2015) of the more massive galaxy during infall. A preference for prograde mergers may be important for the survival of discs, because in cases where the satellite's orbit is in the same direction as the spin of the more massive merging companion, a merger remnant where disc morphology is preserved may be more probable (e.g. Hopkins et al., 2009b).

Observational studies generally support the predictions of theoretical work. For example, broad morphological change from discs to spheroids has been observed in many studies, across a range in redshift (e.g. Butcher & Oemler, 1984; Dressler et al., 1997; Conselice et al., 2014; Huertas-Company et al., 2015b). Many spheroids show signatures of violent and sudden morphological change in their stellar populations (e.g. Blake et al., 2004; Bundy et al., 2005; Goto, 2005; Kaviraj et al., 2008, 2009, 2011; Kaviraj, 2014a; Wild et al., 2016), internal dynamics (e.g. Tacconi et al., 2008; Perret et al., 2012; Cappellari, 2016; Rodrigues et al., 2017) and structure (e.g. McIntosh et al., 2008; Conselice & Arnold, 2009; Kaviraj et al., 2012a,b; Huertas-Company et al., 2015b, 2016), indicating a major merger in their recent history. However, recent work has also demonstrated that many spheroids (especially at  $z \sim 2$ ) appear to be forming without recourse to major mergers, indirectly supporting the potentially important role of minor mergers in driving morphological transformation (e.g. Bundy et al., 2007; Pracy et al., 2009; Kaviraj et al., 2013a; Haines et al., 2015; Lofthouse et al., 2017a).

Nevertheless, while today's surveys are able to provide datasets of sufficient quality that it is possible for galaxy populations across a large range in redshift to be compared morphologically, an empirical determination of the role that mergers and other processes may play in the morphological evolution of galaxies remains difficult. For example, given the limited depth and/or survey area of past surveys, samples of mergers are typically small (e.g. Darg et al., 2010a,b). And since the surface brightness of tidal features induced by mergers decreases with the mass ratio of the merger (e.g. Peirani et al., 2010), most surveys are too shallow to detect the signatures of low mass ratio mergers (see e.g. Kaviraj et al., 2013a; Kaviraj, 2014b). Furthermore, disturbed morphologies may result naturally from internal processes, especially in the early Universe (e.g. Bournaud et al., 2008; Agertz et al., 2009; Förster Schreiber et al., 2011; Cibinel et al., 2015; Hoyos et al., 2016), making it difficult to accurately separate merger remnants from the non-interacting population. Thus, even as we enter an era of deep-wide observational surveys (e.g. DES (Dark Energy Survey Collaboration et al., 2016), EUCLID (Laureijs et al., 2011), LSST (Tyson, 2002; Robertson et al., 2017) and JWST (Gardner et al., 2006)), a purely empirical study of the processes that contribute to the morphological evolution of galaxies remains a challenge.

While theoretical studies offer a better avenue for exploring morphological transformation, many theoretical explorations of this issue have focussed on isolated and idealised simulations of galaxy mergers (e.g. Barnes, 1988; Hernquist, 1992; Bois et al., 2011). However, such simulations lack a realistic context, and so exclude the effects of environment and gas accretion from the cosmic web. Additionally, since the parameter space explored by these studies is small and is not informed by a cosmological

model, it is not possible to make statistical statements about the importance of mergers and other processes to morphological transformation globally. While ‘zoom-in’ studies from cosmological simulations (e.g. Sales et al., 2012; Wuyts et al., 2014; Fiacconi et al., 2015; Sparre & Springel, 2016) do offer a way of placing merging systems into a realistic environment, without requiring significant increases in computing power, both approaches are generally limited by small sample sizes and restricted parameter spaces. Cosmological volumes are essential for a statistical study of morphological transformation.

In the recent literature, semi-analytical models (e.g. Kauffmann et al., 1993; Somerville et al., 2001; Menci et al., 2002; Hatton et al., 2003; Lu et al., 2011) have played an important role in exploring galaxy evolution, using large, statistically-significant samples. While these models have been able to reproduce broad trends in galaxy formation, including the evolution of morphology, stellar mass and gas content (e.g. Somerville & Primack, 1999; Cole et al., 2000; Benson et al., 2003; Bower et al., 2006; Croton et al., 2006; Khochfar et al., 2011; Lamastra et al., 2013; Tonini et al., 2016), they are essentially phenomenological and lack realistic baryonic physics, relying instead on simple numerical recipes for sub-galaxy-scale processes, including morphological transformation. However, recent advances in computing power mean that it has now become possible to simulate the resolved baryonic physics (e.g. gas content and stellar populations) of individual galaxies within cosmological volumes. Modern cosmological, hydrodynamical simulations (e.g. Dubois et al., 2014; Vogelsberger et al., 2014; Khandai et al., 2015; Schaye et al., 2015; Taylor & Kobayashi, 2016; Dubois et al., 2016; Kaviraj et al., 2017) are typically capable of resolving baryonic physics on kpc scales, allowing for the detailed study of small-scale processes within large populations of galaxies. Such simulations offer an unprecedented route to understanding the relative role of different mechanisms in driving the evolution of the morphological mix of the Universe (e.g. Welker et al., 2017; Rodriguez-Gomez et al., 2017; Clauwens et al., 2017).

In this chapter, we use Horizon-AGN (Dubois et al., 2014; Kaviraj et al., 2017), a cosmological hydrodynamical simulation, to investigate key open questions in our understanding of the evolution of the morphological mix of the Universe: what is the magnitude of morphological change imparted by major and minor mergers as a function of redshift and stellar mass? what is the impact of gas fraction on these morphological changes? are the properties of the remnants dependent on the orbital configurations (e.g. prograde vs retrograde) of mergers? what fraction of the total morphological change over cosmic time is attributable to major and minor mergers and other processes?

The structure of this chapter is as follows. In Section 5.2 we present an overview of Horizon-AGN, outlining the treatment of baryonic physics and black holes, the identification of galaxies and mergers, and the definition of morphology used in this study. In Section 5.3 we explore the effect that individual mergers have in driving changes in morphology as a function of redshift, merger mass ratio, gas fraction and orbital configuration. In Section 5.4 we study the average merger histories of discs and spheroids, quantify the cumulative effect of major and minor mergers over cosmic time and outline the role of environment in producing morphological change in regions of high density (e.g. clusters). We summarise our findings in Section 5.5.

## 5.2 The Simulation

Horizon-AGN is a cosmological-volume hydrodynamical simulation (Dubois et al., 2014), based on RAMSES (Teyssier, 2002), an adaptive mesh refinement (AMR) Eulerian hydrodynamics code. It simulates a  $100 h^{-1} \text{coMpc}$  length box, using initial conditions from a *WMAP7*  $\Lambda\text{CDM}$  cosmology (Komatsu et al., 2011). The simulation contains  $1024^3$  dark matter particles, with a mass resolution of  $8 \times 10^7 M_{\odot}$ . An initially uniform  $1024^3$  cell gas grid is refined, according to a quasi Lagrangian criterion (when 8 times the initial total matter resolution is reached in a cell) and the refinement can continue until a minimum cell size of 1 kpc in proper units is reached.

As shown in Kaviraj et al. (2017) and Kaviraj et al. (2015b), Horizon-AGN produces good agreement to key observables that trace the aggregate evolution of galaxies across cosmic time e.g. stellar mass and luminosity functions, rest-frame UV-optical-near infrared colours, the star formation main sequence, galaxy merger histories and the cosmic star formation history. It also reproduces the demographics of BHs, including BH luminosity and mass functions, the BH mass density as a function of redshift, and correlations between BH and galaxy mass in the local Universe (Volonteri et al., 2016). Finally, Horizon-AGN reproduces the morphological mix of the local Universe, with predicted galaxy morphologies in good agreement with observed morphological fractions for intermediate and high mass galaxies (Dubois et al., 2016; Martin et al., 2018a).

In the following sections, we briefly describe aspects of the simulation that are particularly relevant to this study: the treatment of baryonic matter (gas and stars), the identification of galaxies and mergers, the measurement of galaxy morphology and the treatment of BHs and BH feedback on ambient gas.

### 5.2.1 Baryons

Gas cooling proceeds via H, He and metals (Sutherland & Dopita, 1993) down to a temperature of  $10^4$  K and a uniform UV background is switched on at  $z = 10$ , following Haardt & Madau (1996). Star formation is implemented via a standard 2 per cent efficiency (e.g. Kennicutt, 1998), when the density of hydrogen gas reaches  $0.1 \text{ H cm}^{-3}$ . The stellar-mass resolution in the simulation is  $\sim 2 \times 10^6 M_{\odot}$ .

Continuous stellar feedback is employed, including momentum, mechanical energy and metals from stellar winds and Type II and Type Ia supernovae (SNe). Energetic feedback from stellar winds and Type II SNe is applied via STARBURST99 (Leitherer et al., 1999, 2010), implemented using the Padova model (Girardi et al., 2000) with thermally pulsating asymptotic branch stars (Vassiliadis & Wood, 1993). The kinetic energy of stellar winds is calculated via the ‘Evolution’ model of Leitherer et al. (1992). The implementation of Type Ia SNe follows Matteucci & Greggio (1986) and assumes a binary fraction of 5% (Matteucci & Recchi, 2001), with chemical yields taken from the W7 model of Nomoto et al. (2007). Stellar feedback is modelled as a heat source after 50 Myrs. This is because after 50 Myrs the bulk of the energy is liberated via Type Ia SNe that have time delays of several hundred Myrs to a few Gyrs (e.g.

Maoz et al., 2012). These systems do not suffer large radiative losses, as stars disrupt or move away from their dense birth clouds after around a few tens of Myrs (see e.g. Blitz & Shu, 1980; Hartmann et al., 2001).

## 5.2.2 Identification of galaxies and mergers

The ADAPTAHOP structure finder (Aubert et al., 2004; Tweed et al., 2009), applied to the distribution of star particles, is used to identify galaxies. The selection of structures requires that the local density exceeds 178 times the average matter density. The local density is calculated using the 20 nearest particles. A minimum number of 50 particles is required to identify a structure, which imposes a minimum galaxy stellar mass of  $2 \times 10^8 M_{\odot}$ . We produce merger trees for each individual galaxy and track their progenitors to  $z = 3$ . The average timestep in the merger histories is  $\sim 130$  Myr. A major merger is defined as a merger where the mass ratio of the merging progenitors is greater than or equal to 1 : 4. A minor merger is defined as a merger where the mass ratio of the merging progenitors is between 1 : 4 and 1 : 10.

The choice of a threshold mass ratio for minor mergers of 1 : 10 is not arbitrary, but driven by previous work which indicates that this is typically a threshold below which the impact of mergers generally becomes negligible. For example, below this threshold, star formation and black-hole accretion rates are not detectably enhanced in mergers (Martin et al., 2017; Martin et al., 2018b). In the Appendix, we demonstrate this point by quantifying the effect of varying the minimum mass ratio down to 1 : 40 for our results in Sections 3 and 4, and showing that, in mergers with mass ratios below 1 : 10, there is negligible morphological change, compared to galaxies that are not merging.

The requirement of 50 particles for the definition of a galaxy, imposes a limit on the minimum merger mass ratio that is detectable for a galaxy of a given mass at a given redshift. Since galaxies contain less stellar mass at higher redshift, we detect a smaller proportion of mergers at earlier times. Fig 5.1 presents detectability limits for mergers of various mass ratios in the merger histories of galaxies, as a function of the stellar mass of galaxies at  $z = 0$ . For each galaxy, we calculate the mass of its main progenitor at a redshift of interest. This then determines the mass ratio limit of detectable mergers for the galaxy in question.

In Fig 5.1, we show the fraction of galaxies of a given stellar-mass at  $z = 0$  that have progenitors that are massive enough for mergers of various mass ratios to be detectable at different redshifts. For example, for galaxies that have a mass of  $10^{9.5} M_{\odot}$  at the present day, 96, 72 and 37 per cent of their progenitors are massive enough for a merger with a 1 : 4 mass ratio to be detectable at  $z = 1$ ,  $z = 2$  and  $z = 3$  respectively. For mergers with mass ratios of 1 : 10, the corresponding values are 84, 47 and 20 per cent at the same redshifts. For galaxies with stellar masses above  $10^{11} M_{\odot}$ , the merger history is at least 85 per cent complete for mass ratios greater than 1 : 10, up to  $z = 3$ . We note that, while mergers will not be detectable for a large proportion of very low-mass galaxies, the merger rate is expected to fall with

decreasing stellar mass (e.g. Stewart et al., 2008; Rodriguez-Gomez et al., 2015), so that the importance of mergers is lower in the regime where the sample is most incomplete.

### 5.2.3 Galaxy morphology

Following Martin et al. (2018a), we estimate morphology using a galaxy's stellar kinematics. We use  $V/\sigma$ , the ratio of the mean rotational velocity ( $V$ ) and the mean velocity dispersion ( $\sigma$ ), measured using the entire star particle distribution of the galaxy. Higher values of  $V/\sigma$  correspond to systems that are more rotationally-supported i.e. those that have more late-type (disc-like) morphologies.  $V/\sigma$  is calculated by rotating the coordinate system so that the  $z$ -axis is oriented along the stellar angular momentum vector.  $V$  is then defined as the mean tangential velocity component in cylindrical co-ordinates,  $V_\theta$ . The velocity dispersion ( $\sigma$ ) is computed using the standard deviations of the radial, tangential and vertical star particle velocities,  $\sigma_r$ ,  $\sigma_\theta$  and  $\sigma_z$ , summed in quadrature.  $V/\sigma$  is defined as:

$$V/\sigma = \frac{\sqrt{3}\bar{V}_\theta}{\sqrt{\sigma_r^2 + \sigma_\theta^2 + \sigma_z^2}}. \quad (5.1)$$

As in Martin et al. (2018a), we use 'spheroid' and 'disc' to refer to galaxies that are dominated by their dispersional and rotational velocities respectively. Following Martin et al. (2018a), we choose a  $V/\sigma$  threshold value of 0.55, which best reproduces the observed spheroid and disc fractions of the Universe at low redshift (Conselice, 2006). In other words, galaxies with  $V/\sigma$  values above 0.55 are considered to be discs, while those with values below this threshold are spheroids.

### 5.2.4 Treatment of black holes and black-hole feedback

BH are seeded as 'sink' particles with an initial mass of  $10^5 M_\odot$  until  $z = 1.5$ , wherever the local gas density exceeds  $\rho > \rho_0$  and the stellar velocity dispersion exceeds  $100 \text{ km s}^{-1}$ , where  $\rho_0 = 1.67 \times 10^{-25} \text{ g cm}^{-3}$  and corresponds to  $0.1 \text{ H cm}^{-3}$  (the minimum density threshold required for star formation). To prevent multiple BHs from forming within the same galaxy, BHs cannot form if there is another BH within 50 kpc. Each BH grows through gas accretion, or coalescence with another black hole (Dubois et al., 2014, 2016). An Eddington-limited Bondi-Hoyle-Lyttleton rate is used to model BH accretion:

$$\dot{M}_{\text{BH}} = \frac{4\pi\alpha G^2 M_{\text{BH}}^2 \bar{\rho}}{(\bar{c}_s^2 + \bar{u}^2)^{3/2}}, \quad (5.2)$$

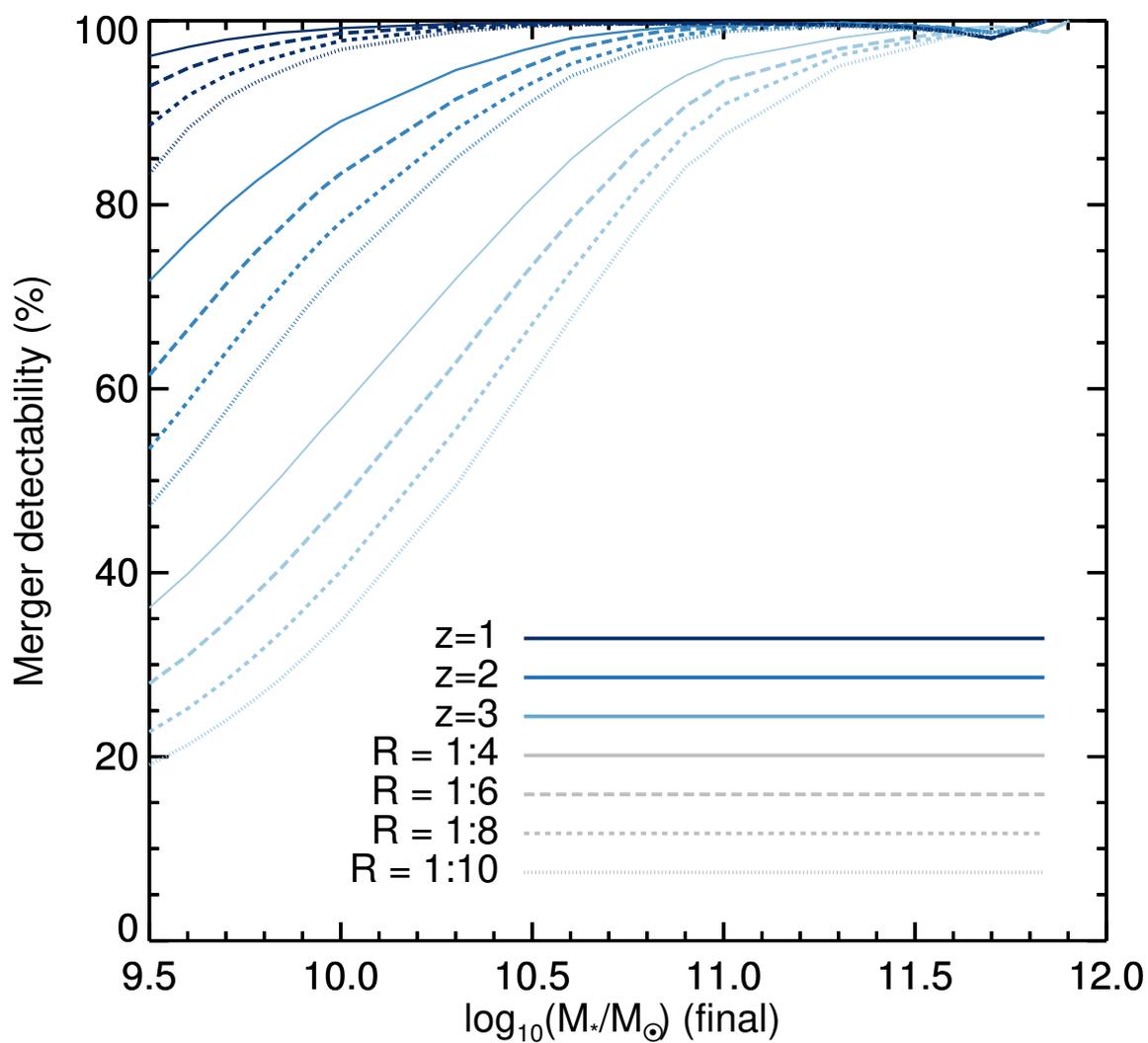


FIGURE 5.1: The proportion of galaxies for which mergers of different mass ratios are detectable at various redshifts (see legend), as a function of the stellar mass of the galaxy at  $z = 0$ .  $R$  is the stellar mass ratio of the merger. For example, for galaxies that have a mass of  $10^{9.5} M_\odot$  at the present day, 96, 72 and 37 per cent of these systems have progenitors that are massive enough for a merger with a 1 : 4 mass ratio to be detectable at  $z = 1$ ,  $z = 2$  and  $z = 3$  respectively. For mergers with mass ratios of 1 : 10, the corresponding values are 84, 47 and 20 per cent at the same redshifts.

where  $M_{\text{BH}}$  is the mass of the BH,  $\bar{\rho}$  is the mass-weighted average gas density,  $\bar{c}_s$  is the mass-weighted average sound speed,  $\bar{u}$  is the mass-weighted average gas velocity relative to the BH and  $\alpha$  is a dimensionless boost factor accounting for the inability of the simulation to capture the cold (high-density) inter-stellar medium (e.g. Booth & Schaye, 2009a).

BHs impart feedback on gas via two different channels that depend on the ratio of the gas accretion rate and the Eddington luminosity,  $\chi = \dot{M}_{\text{BH}}/\dot{M}_{\text{Edd}}$ . For Eddington ratios  $\chi > 0.01$  (that represent high accretion rates), BH feedback operates via a ‘quasar’ mode, with 1.5 per cent of the accretion energy injected as thermal energy into the gas isotropically. For Eddington ratios  $\chi < 0.01$  (which represent low accretion rates), BH feedback is modelled as a ‘radio’ mode, where bipolar outflows are implemented with jet velocities of  $10^4 \text{ km s}^{-1}$ . The efficiency of the quasar mode is chosen to reproduce the local observed  $M_{\text{BH}} - M_*$  and  $M_{\text{BH}} - \sigma_*$  relations and the local cosmic black-hole mass density (Dubois et al., 2012).

BH feedback principally quenches external gas accretion (e.g. Dubois et al., 2014, 2016), but does not couple to star particles. Gas accretion typically acts to increase  $V$ , since new stars forming from the gas inherit its angular momentum and add to the rotational component of the galaxy. Thus, the effect of BH feedback is essentially to prevent the value of  $V$  from increasing and, therefore, to also lock in the value of  $V/\sigma$ , since it doesn’t alter  $\sigma$ , the stellar velocity dispersion. BH feedback therefore plays an important, if indirect role, in the morphological evolution of galaxies. It is necessary for reproducing the observed morphological diversity of the present-day Universe (e.g. Dubois et al., 2016), as it can *maintain* the morphology of a system, in between events like mergers that alter it (e.g. Dubois et al., 2016; Pontzen et al., 2017). It is worth noting, however, that BH feedback cannot, by itself, create spheroidal systems. Indeed in Horizon-AGN, we do not find significant differences in the aggregate BH accretion rates in spheroid and disc progenitors, again indicating that BH feedback alone is not capable of producing spheroidal morphologies.

### 5.3 The effect of individual mergers

We begin by investigating how key properties of the progenitors affect the morphology of the merger remnant. The large volume of the Horizon-AGN simulation allows us to explore a realistic  $\Lambda$  CDM parameter space of mergers, across a broad range of properties, including redshift, stellar mass, merger mass ratio, gas fraction and orbital configuration. In this section, we first describe our method for calculating the change in morphology due to a merger, study the effect of major and minor mergers on the morphology of galaxies as a function of stellar mass and redshift (Section 5.3.1), explore the effect of gas fraction on the morphology of merger remnants (Section 5.3.2) and probe how orbital configurations influence the properties of merger remnants (Section 5.3.3).

In order to quantify the morphological change in a merging system, we measure the change in  $V/\sigma$  of the main (i.e. the more massive) progenitor in a 2 Gyr window, centred around the time that the two galaxies coalesce (i.e. when both galaxies are identified as being part of the same structure). As we elaborate below, the size of the window is chosen to ensure that we measure the morphology of the main companion before it is affected by gravitational torques in the merger, and to allow time for the merger remnant to relax, at least in its inner regions (low-surface-brightness features in galaxy outskirts, such as shells and loops, can last for many dynamical timescales, e.g. Mihos & Murdin (2000); Kaviraj (2014b), but make up a negligible proportion of the galaxy's stellar mass). Note that, since we consider the collection of individual merger events in this section, incompleteness does not affect our analysis, on the assumption that the statistical properties of mergers that are not visible are similar to those that are observed. Incompleteness is a larger issue when studying the cumulative impact of mergers over cosmic time, and we return to this point in the next section.

For each galaxy merger, we measure the morphological change,  $\Delta\text{morph}$ , defined as the fractional change in the  $V/\sigma$  of the main progenitor over the course of the merger. We assume a timescale of 2 Gyrs, measuring the change in  $V/\sigma$  between  $t = -1$  Gyr and  $t = +1$  Gyr relative to coalescence:

$$\Delta\text{morph} = \frac{V/\sigma_{t=1 \text{ Gyr}} - V/\sigma_{t=-1 \text{ Gyr}}}{V/\sigma_{t=-1 \text{ Gyr}}}, \quad (5.3)$$

where  $\Delta t = t_{1 \text{ Gyr}} - t_{-1 \text{ Gyr}}$ , and is approximately equal to 2 Gyrs, with the exact value depending on the coarseness of the merger-tree timesteps. The choice of a 2 Gyr timescale is driven by the fact that, for the merger mass ratios we will consider in this study ( $>1:10$ ), the merger process is typically complete over this timescale (e.g. Jiang et al., 2008; Kaviraj et al., 2011). We note that an important issue when selecting a timescale is to use a value that encompasses the merger event completely. In particular, choosing timescales that are too short will lead to spurious results, because merger remnants may not have relaxed at the point at which they are observed. We explore our choice of timescale in more detail in the Appendix and show that choosing a slightly longer 3 Gyr or even a 1 Gyr timescale does not alter the conclusions of this chapter, although, as we discuss below, 1 Gyr may be too short for mergers closer to the lower end ( $\sim 1 : 10$ ) of our mass ratio range.

Finally, galaxies are considered to be not merging if they have not undergone either a major or minor merger within the last Gyr, and will not undergo such a merger in the next Gyr. All galaxy properties, such as  $M_*$ ,  $m_{\text{gas}}$  etc., are calculated at the initial ( $t = -1$  Gyr) snapshot. We also calculate  $\Delta\text{morph}$  for galaxies that are not undergoing mergers in the same way as for the merging galaxies, again using a timescale of 2 Gyr.

### 5.3.1 Morphological change induced by major and minor mergers as a function of stellar mass and redshift

In Fig 5.2, we investigate  $\Delta\text{morph}$  as a function of the stellar mass and redshift of the main (more massive) progenitor. Galaxies undergoing major and minor mergers are indicated using the solid and dashed lines respectively, while galaxies that are not undergoing any mergers are indicated using the dotted lines. Recall that a major merger is defined as a merger where the mass ratio of the merging progenitors is greater than or equal to 1 : 4, while a minor merger is defined as one where the mass ratio of the merging progenitors is between 1 : 4 and 1 : 10. We further separate the main progenitors into spheroids (left-hand column) and discs (right-hand column). Positive values of  $\Delta\text{morph}$  indicate that the merger remnant has spun up (i.e. become more rotationally-supported or discy), while negative values of  $\Delta\text{morph}$  indicate that the remnant has spun down (i.e. become more dispersion-supported or spheroidal).

Mergers in which the main progenitor is a disc galaxy almost exclusively spin the system down, and result in remnants with lower  $V/\sigma$ , i.e. systems that are more spheroidal. Spinning up as a result of such mergers is rare and happens only in  $\sim 5$  per cent of cases for main progenitor masses of  $10^{10.5} < M_*/M_\odot < 10^{11}$ . The values of  $\Delta\text{morph}$  in Fig 5.2 indicate that major mergers where the main progenitor is a disc produce larger morphological changes than minor mergers. In the nearby Universe, individual major and minor events with main progenitor masses of  $10^{10.5} < M_*/M_\odot < 10^{11}$  reduce  $V/\sigma$  by around  $\sim 28$  per cent and  $\sim 13$  per cent respectively over the course of the merger. The corresponding values at  $z \sim 1$  and  $z \sim 2$  are 44/27 per cent, 57/39 per cent respectively. Typically, the magnitude of the morphological change induced by major mergers is around a factor of 2 greater than that in minor mergers.

While the effect of individual mergers is largely insensitive to the stellar mass of the main progenitor, it is dependent on the dispersional component of the stellar velocity distribution in a galaxy. As galaxies grow larger bulge components towards the present day, more of a merger's ability to induce morphological change is removed, as larger proportions of stellar mass have already been removed from circular orbits and re-arranged into random orbits. This leads to the gradual decrease in  $\Delta\text{morph}$  in discs towards lower redshift.

The impact of mergers on spheroids is qualitatively different. Unlike discs, which are efficiently destroyed by major and minor mergers, there is little preference for spin down over spin up in spheroids. As indicated by the  $1\sigma$  dispersion regions, mergers where the main progenitor is a spheroid produce a greater range of outcomes, and can produce both positive and negative values for  $\Delta\text{morph}$ . In high redshift ( $z > 1$ ) mergers that involve massive spheroidal galaxies ( $M_*/M_\odot > 10^{11}$ ), mergers tend, on average, to spin remnants down. The magnitude of this morphological change is similar to what is observed in mergers where the main progenitor is a disc, although the scatter in the range of  $\Delta\text{morph}$

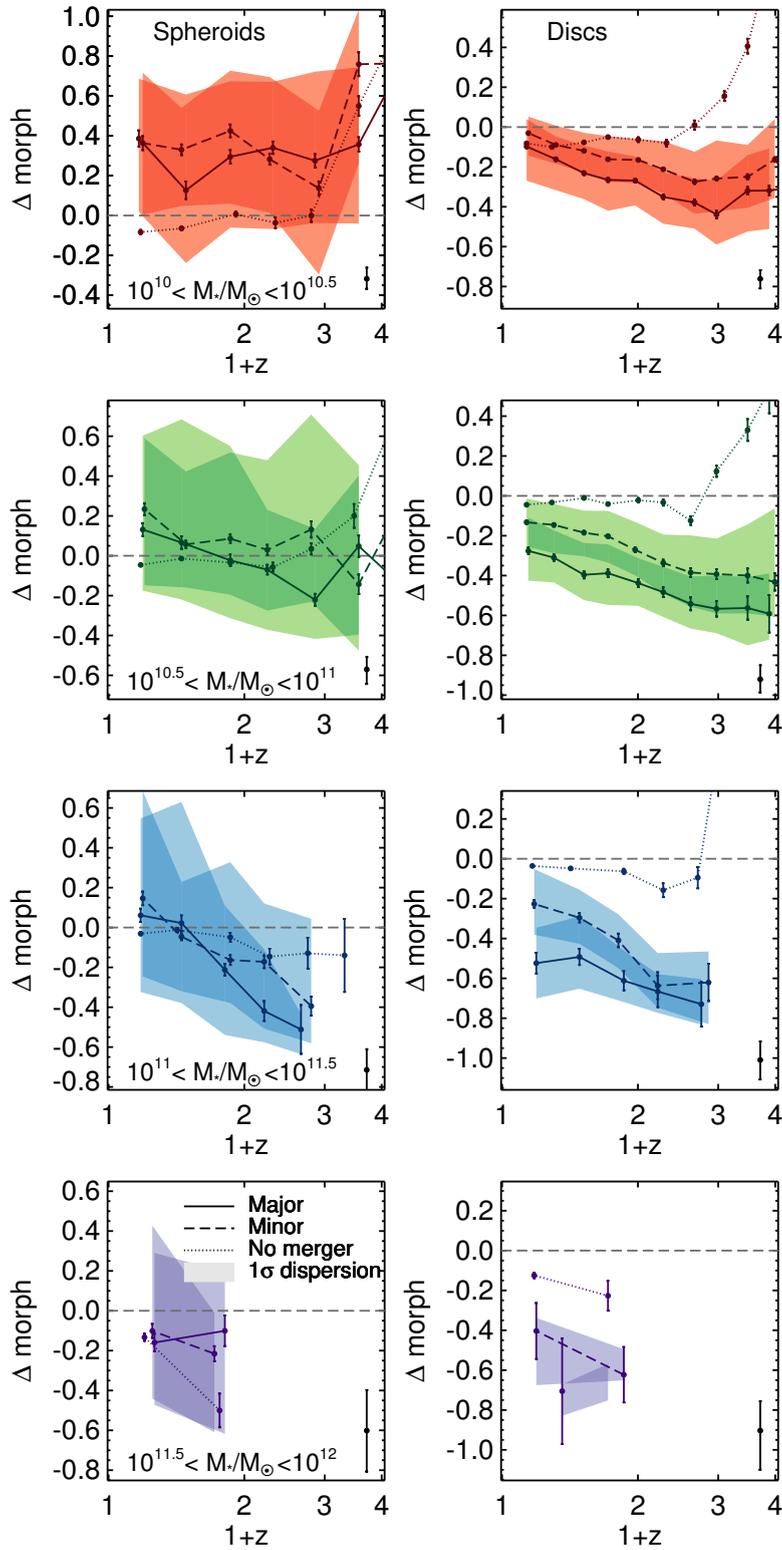


FIGURE 5.2: Median  $\Delta\text{morph}$  as a function of redshift and stellar mass. Positive values of  $\Delta\text{morph}$  indicate that the merger remnant has spun up (i.e. become more rotationally-supported), while negative values of  $\Delta\text{morph}$  indicate that the remnant has spun down (i.e. become more dispersion-supported). The left-hand column indicates mergers where the main (i.e. more massive) progenitor is a spheroid, while the right-hand column shows mergers where the main progenitor is a disc. Error bars indicate the standard error on the median and filled regions indicate  $\pm 1\sigma$  dispersions. For non-merging galaxies, typical dispersions are indicated by a black error bar. The solid and dashed dark lines indicate the median  $\Delta\text{morph}$  for major and minor mergers respectively, while the dotted line indicates  $\Delta\text{morph}$  for non-merging galaxies. Galaxies with  $M_* > 10^{11.5} M_\odot$  only begin to appear in the simulation after  $z \sim 2$ , so there are no datapoints for higher redshifts.

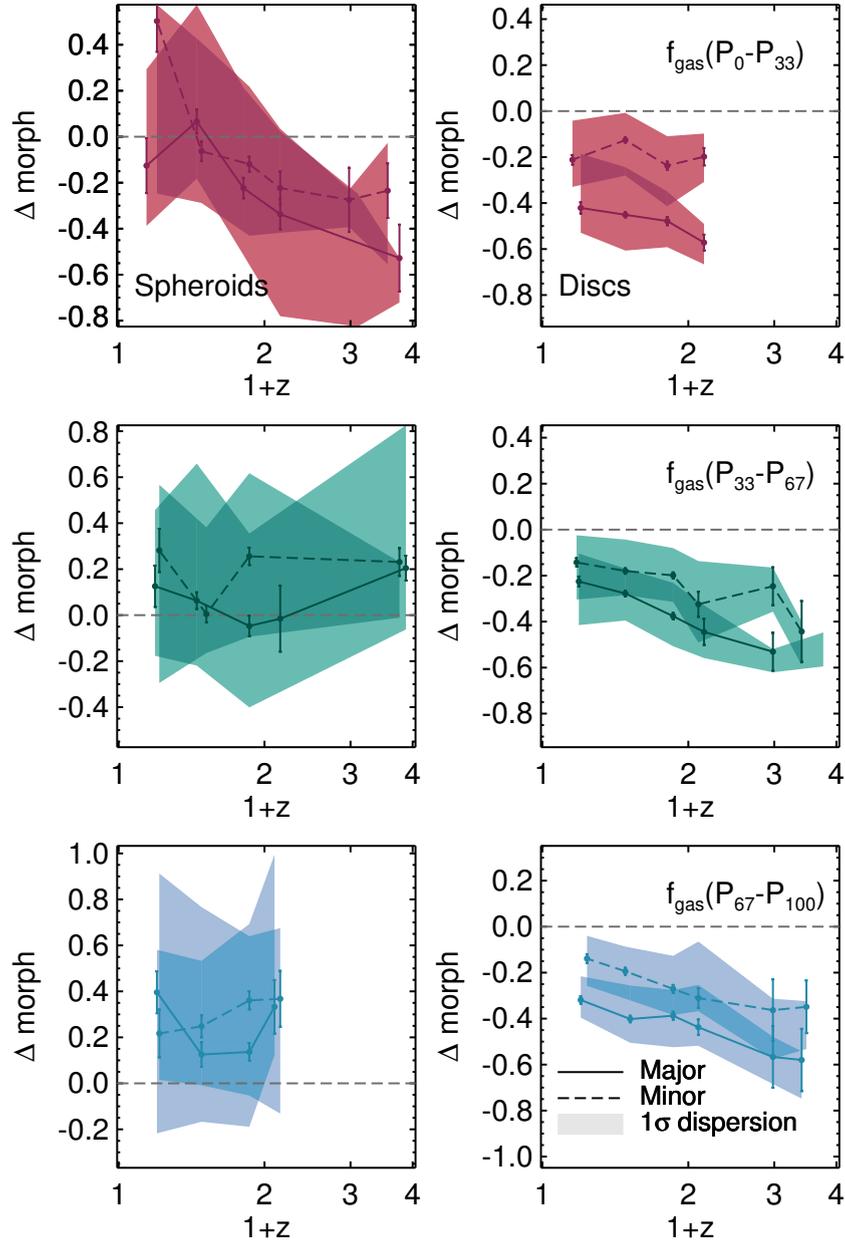


FIGURE 5.3: Median  $\Delta \text{morph}$  as a function of redshift and split by the gas fraction percentile they inhabit for galaxies in the mass range  $10^{10.5} < M_*/M_\odot < 10^{11}$ . The 33rd and 67th percentiles are used to create the percentile ranges, and are typically 0.85 times and 1.15 times the average values indicated by the green lines in the top panel of Fig 5.4. The left-hand column indicates mergers where the main (i.e. more massive) progenitor is a spheroid, while the right-hand column shows mergers where the main progenitor is a disc. Error bars indicate the standard error on the median, while filled regions indicate the  $\pm 1\sigma$  dispersions. The solid and dashed dark lines indicate the median  $\Delta \text{morph}$  for major and minor mergers respectively. Note that the behaviour is similar regardless of the mass of the main progenitor.

values is higher, so that it becomes more likely that some mergers spin remnants up (this happens in approximately 30 per cent of major mergers and 40 per cent of minor mergers).

In low redshift ( $z < 1$ ) mergers that involve such massive spheroidal galaxies, the average value of  $\Delta\text{morph}$  is close to zero i.e. mergers do not, on average, produce strong morphological changes, although, as the  $1\sigma$  dispersion regions indicate, both spinning up and spinning down is possible from such events. The impact of mergers that involve less massive spheroidal main progenitors is somewhat different. While for intermediate stellar masses ( $10^{10.5} < M_*/M_\odot < 10^{11}$ ) the average values of  $\Delta\text{morph}$  remain close to zero across all redshifts, at the low-mass end ( $10^{10} < M_*/M_\odot < 10^{10.5}$ ) remnants of both major and minor mergers tend to spin remnants up.

For non-merging galaxies, similar trends are observed in mergers where the main progenitors are spheroids or discs, largely regardless of stellar mass.  $\Delta\text{morph}$  is large and positive at high redshift ( $z > 2$ ), indicating intense cosmological gas accretion that imparts angular momentum to the galaxy and spins it up (e.g. Brooks et al., 2009; Pichon et al., 2011; Stewart et al., 2013). At lower redshifts,  $\Delta\text{morph}$  in non-merging galaxies is close to zero (indicating no morphological change), but is typically slightly negative ( $\Delta\text{morph} \gtrsim -0.08$ ) within errors, possibly indicating some morphological impact from very low mass ratio ( $< 1 : 10$ ) interactions.

### 5.3.2 The effect of gas fraction

In Fig 5.3, we study the value of  $\Delta\text{morph}$  as a function of the redshift and gas fraction of the merging system. We again separate our analysis into main progenitors that are spheroids (left-hand column) and discs (right-hand column). The gas fraction ( $f_{\text{gas}}$ ) is defined as the combined cold gas fraction of the two merging companions:

$$f_{\text{gas}} = \frac{m_{\text{gas},\text{main}} + m_{\text{gas},\text{sat}}}{m_{\text{gas},\text{main}} + m_{*,\text{main}} + m_{\text{gas},\text{sat}} + m_{*,\text{sat}}}, \quad (5.4)$$

where  $m_{\text{gas},\text{main}}$  is the mass of cold gas within  $2 R_{\text{eff}}$  of the main (more massive) companion 1 Gyr prior to coalescence, and  $m_{\text{gas},\text{sat}}$  is the corresponding value for the lower mass companion (i.e. the satellite). Similarly,  $m_{*,\text{main}}$  is the stellar mass within  $2 R_{\text{eff}}$  of the more massive companion and  $m_{*,\text{sat}}$  is the corresponding value for the smaller companion. For this analysis, we restrict our sample to a narrow range in stellar mass,  $10^{10.5} < M_*/M_\odot < 10^{11}$ , since the mean gas fraction evolves with galaxy stellar mass. However, the behaviour we find here is similar regardless of the stellar mass range considered. We split our sample by gas fraction percentile after splitting the sample by redshift and the morphology of the main progenitor. The top and bottom panels represent the more extreme objects in terms of gas fraction for a given redshift. The 33rd and 67th percentiles are used to create the percentile ranges, and are typically 0.85 times and 1.15 times the average values indicated by the green lines in the top panel of Fig 5.4.

Mergers where the main progenitor is a disc galaxy typically produce remnants with negative  $\Delta\text{morph}$  i.e. typically spin systems down, largely regardless of the gas fraction. However, the behaviour in spheroids is different. While mergers with low gas fractions are most likely to spin remnants down (apart from in the nearby Universe, where spin up or down appear equally likely), mergers that involve high gas fractions produce significant spin up, particularly in the high-redshift Universe, in line with the results of recent observational and theoretical studies (e.g. Robertson et al., 2006; Athanassoula et al., 2016; Rodriguez-Gomez et al., 2017; Font et al., 2017).

Fig 5.4, which shows the evolution of the average galaxy gas fraction over cosmic time, helps explain the mass dependence of  $\Delta\text{morph}$  for discs and spheroids seen in Fig 5.2. As this figure indicates, average gas fractions evolve as a function of mass and redshift, with lower mass galaxies exhibiting higher values, consistent with the results of recent observational work (e.g. Geach et al., 2011; Tacconi et al., 2013). Lower mass galaxies tend to be more gas-rich (at any epoch). Since high gas fractions tend to produce remnants with more positive  $\Delta\text{morph}$  values (Fig 5.3), i.e. milder morphological transformation, mergers involving lower mass galaxies are more gas rich (e.g. Kaviraj et al., 2009; Struve et al., 2010), which explains the trend in  $\Delta\text{morph}$  becoming more positive for both discs and spheroids at lower stellar masses.

### 5.3.3 The effect of orbital configuration

We proceed by investigating how  $\Delta\text{morph}$  varies as a function of the orbital configuration of the merging system, at the point where the satellite enters the virial radius of the main (more massive) progenitor. We consider both the angular momentum of the satellite galaxy,  $L_{\text{sat}}$ , measured using its star particle distribution, and the angular momentum of the orbit of the satellite relative to the main progenitor,  $L_{\text{orb}} = M_{\text{sat}}(r \times v)$ , where  $r$  and  $v$  are the position and velocity of the satellite relative to that of the main progenitor. The meaning of the angular momentum vectors is illustrated in Fig 5.5. We use these quantities to define the total angular momentum that is imparted by the merging satellite to the main progenitor, relative to its spin:

$$L_{\text{external}} = |L_{\text{sat}}| \cos(\theta_{L_{\text{main}}, L_{\text{sat}}}) + |L_{\text{orb}}| \cos(\theta_{L_{\text{main}}, L_{\text{orb}}}) \quad (5.5)$$

where  $\theta_{L_{\text{main}}, L_{\text{sat}}}$  is the angle between the angular momentum vector of the main progenitor and the satellite, so that an angle below  $\pi/2$  denotes two co-rotating galaxies (i.e. a ‘prograde’ merger), while angles above  $\pi/2$  denote counter-rotating galaxies (i.e. a ‘retrograde’ merger).  $\theta_{L_{\text{main}}, L_{\text{orb}}}$  is the angle between the angular momentum vector of the main progenitor and the angular momentum vector of the satellite’s orbit. Thus, an angle that is close to zero denotes a merger where the orbit of the merging satellite is co-planar with the disc of the main progenitor and in the same direction as its rotation, while a value of  $\pi$  denotes a merger that impacts the disc of the main progenitor in the opposite direction to its rotation,

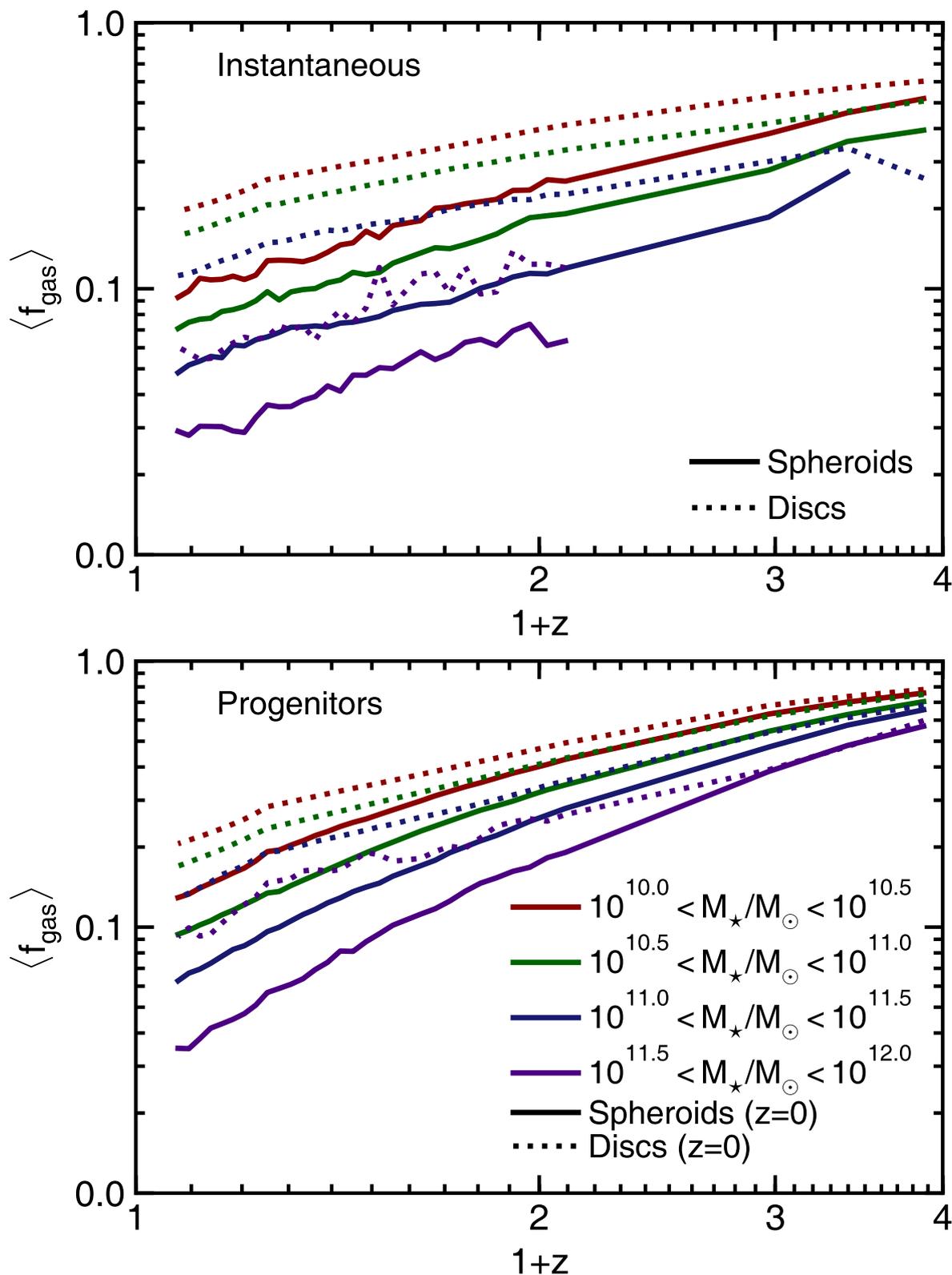


FIGURE 5.4: The redshift evolution of the mean cold gas fraction ( $f_{gas}$ ) in spheroidal and disc galaxies in bins of stellar mass. **Top:** Mean value of  $f_{gas}$  in spheroids (solid lines) and discs (dotted lines) as a function of redshift, shown in bins of stellar mass. **Bottom:** Mean value of  $f_{gas}$  in the main (i.e. the most massive) progenitors of today's spheroids (solid lines) and discs (dotted lines), shown as a function of the stellar mass of the final spheroid or disc at  $z = 0$ .

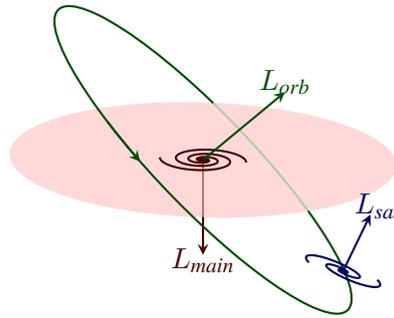


FIGURE 5.5: Sketch illustrating the vectors used to describe the orbital configuration of merging systems. The main progenitor refers to the more massive merging companion. The angles between vectors are defined in the standard way as the angles between two vectors in their common plane.

most efficiently removing angular momentum from the system. Note that our definition of prograde or retrograde includes both the angular momentum of the satellite's orbit as well as its spin relative to the spin of the main progenitor.

Fig 5.6 shows the average  $\Delta\text{morph}$  in prograde and retrograde mergers where the main progenitor is a disc galaxy, as a function of redshift. Major mergers efficiently destroy discs in almost all cases, but are most effective when the angular momentum of the spin of the satellite and its orbit is counter to the rotation of the main progenitor's disc (i.e.  $L_{\text{external}} < 0$ ). The difference is fairly modest, especially at high redshift, where major mergers produce a mean  $\Delta\text{morph}$  of 0.52 and 0.50 for retrograde and prograde mergers respectively. The difference is more significant at low redshifts (mean  $\Delta\text{morph}$  of 0.35 and 0.20 for retrograde and prograde mergers) when average gas fractions are low (see Fig 5.4) and mergers are less effective at rebuilding discs. Low-redshift minor mergers are almost equally likely to produce positive or negative changes to  $\Delta\text{morph}$  when the merger is prograde. In the case of retrograde minor mergers, spin down of discs remains significant in almost all cases.

Fig 5.7 shows the fraction of prograde events in mergers that involve progenitors of today's spheroids (solid line) and today's discs (dotted line). The prograde fractions are shown as a function of the stellar mass of the spheroid or disc at  $z = 0$ . The progenitors of today's discs have undergone more prograde mergers than the progenitors of spheroids in all but the highest stellar mass bin, where the prograde fractions are relatively similar within errors (note that there are very few discs with such high stellar masses, which leads to a large error bar in the prograde fraction). The tendency of disc progenitors to have had more prograde mergers over cosmic time is therefore likely to be a contributing factor to the continued survival of these discs.

### 5.3.4 Average $\Delta\text{morph}$ in the progenitors of today's discs and spheroids

While the previous sections have explored the impact of mass ratios, gas fractions and orbital configurations, we complete our analysis of the effect of individual mergers by taking an average, aggregate

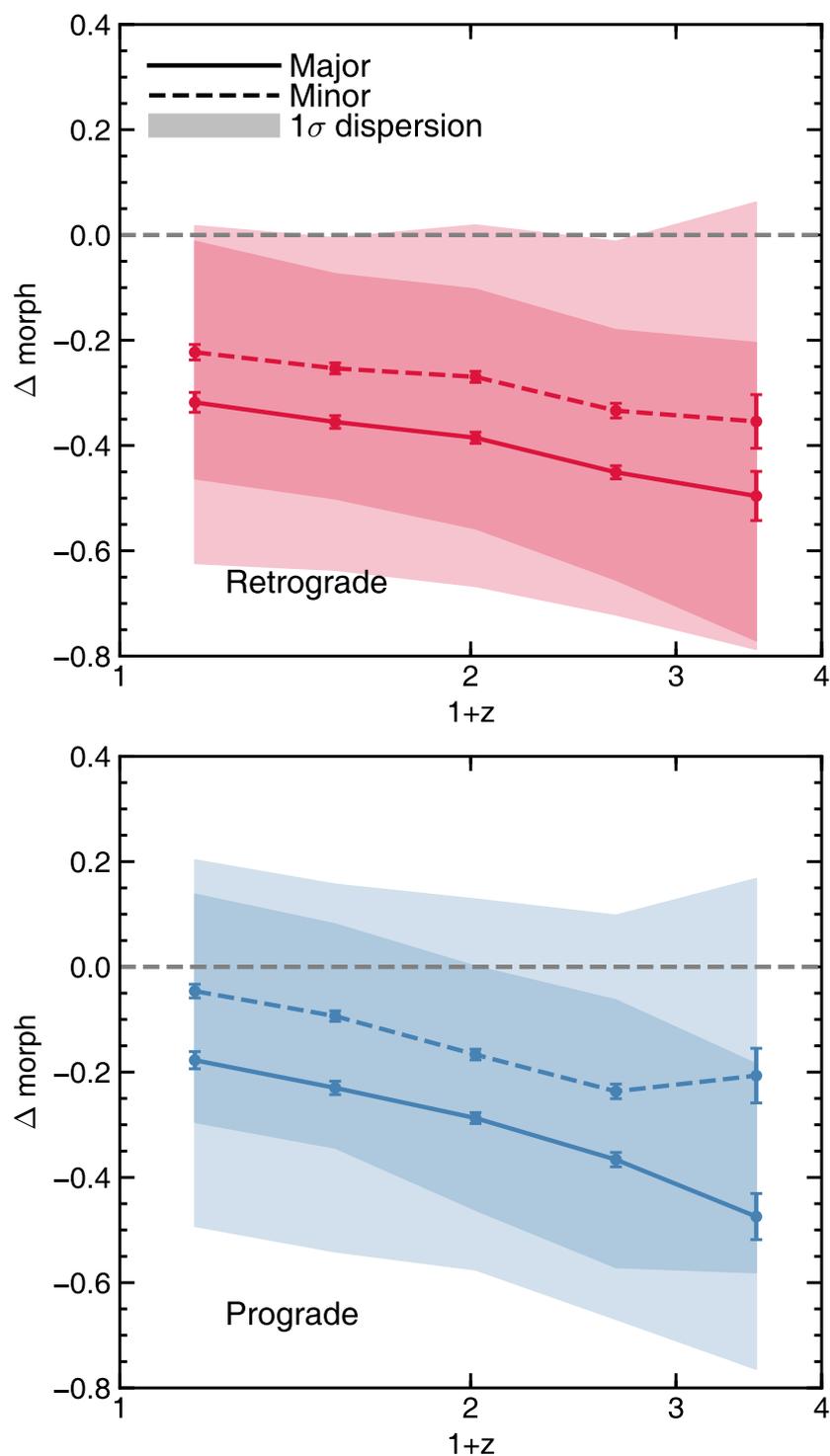


FIGURE 5.6: Median  $\Delta\text{morph}$  for discs as a function of redshift for prograde ( $L_{\text{external}} > 0$ ) and retrograde ( $L_{\text{external}} < 0$ ) mergers. The left-hand column indicates retrograde mergers, where the angular momentum is imparted counter to the spin of the main (more massive) merging progenitor and the right-hand column indicates prograde mergers where the angular momentum is imparted in the same direction as the spin of the main progenitor. Error bars indicate the standard error on the median and solid and dashed lines indicate the median  $\Delta\text{morph}$  for major mergers and minor mergers respectively.

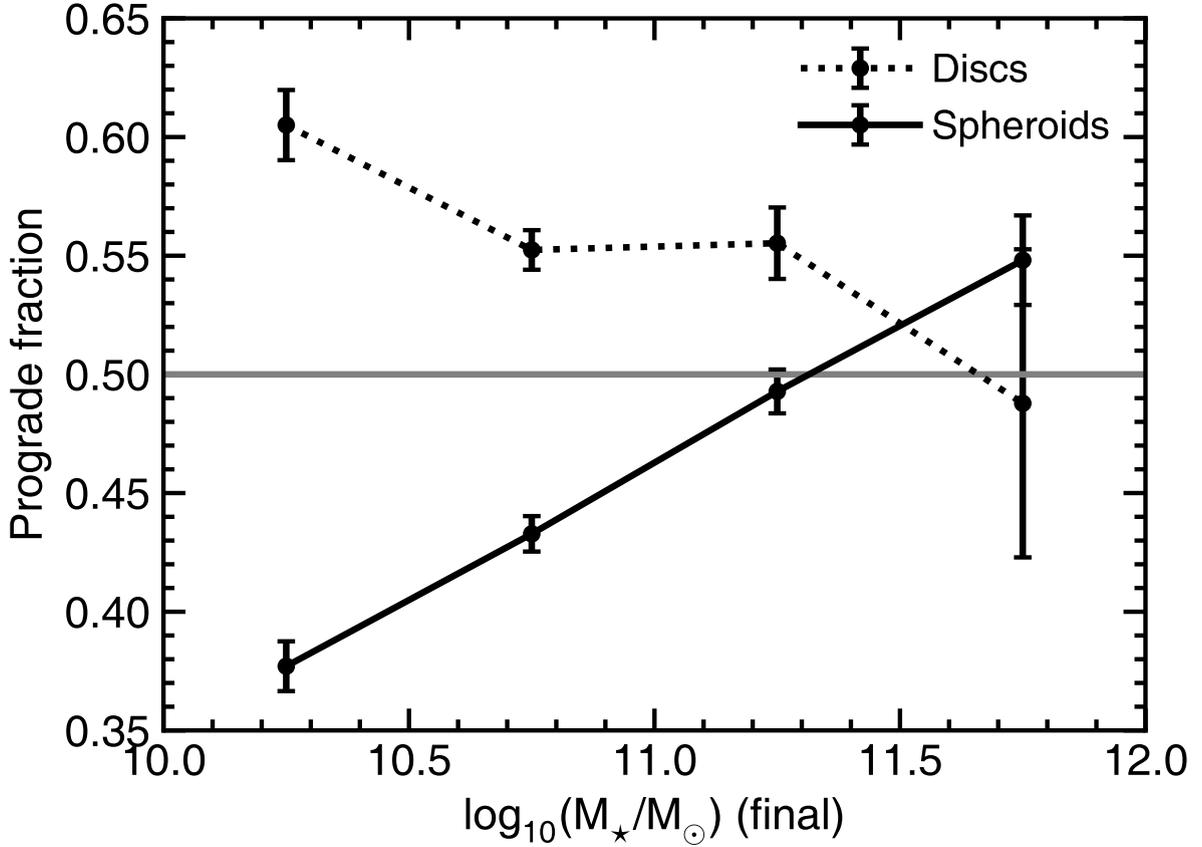


FIGURE 5.7: The fraction of mergers since  $z = 3$  that are prograde in the merger histories of galaxies that are spheroids (solid line) and discs (dotted line) at  $z = 0$ , as a function of the stellar mass of the galaxy.

TABLE 5.1: Mean (median) properties of spheroids (white rows) and discs (shaded rows) at  $z = 0$ .  $\pm$  indicates the  $1\sigma$  dispersion. (i) Number of galaxies in mass bin, (ii) average  $V/\sigma$ , (iii) average fractional change in  $V/\sigma$  between  $z = 3$  and  $z = 0$ , (iv) average local density percentile (see text and Martin et al. (2018a)), (v) the average fraction of time that the galaxies have spent with the morphology that they have at  $z = 0$ .

$\log_{10}(M_*/M_\odot)$	$N$ (i)	$V/\sigma$ (ii)	$\Delta V/\sigma$ (iii)	Environment (iv)	time with morphology (v)
10.5–11.0	2064	0.333(0.359) $^{+0.078}_{-0.107}$	-0.544(-0.629) $^{+0.187}_{-0.032}$	49.16(48.23) $^{+18.17}_{-16.99}$	0.498(0.493) $^{+0.137}_{-0.156}$
	6056	0.898(0.896) $^{+0.111}_{-0.099}$	0.110(-0.051) $^{+0.313}_{-0.013}$	49.84(49.83) $^{+17.68}_{-17.94}$	0.952(0.987) $^{+0.019}_{-0.048}$
11.0–11.5	1189	0.268(0.254) $^{+0.103}_{-0.088}$	-0.627(-0.707) $^{+0.183}_{-0.038}$	47.51(47.16) $^{+16.99}_{-15.45}$	0.579(0.573) $^{+0.123}_{-0.155}$
	987	0.856(0.812) $^{+0.135}_{-0.073}$	0.149(-0.110) $^{+0.408}_{-0.060}$	53.18(55.29) $^{+18.43}_{-20.69}$	0.897(0.973) $^{+0.032}_{-0.103}$
11.5–12.0	299	0.193(0.156) $^{+0.091}_{-0.050}$	-0.566(-0.809) $^{+0.305}_{-0.115}$	49.68(49.43) $^{+17.01}_{-17.36}$	0.693(0.720) $^{+0.133}_{-0.173}$
	21	0.697(0.675) $^{+0.067}_{-0.006}$	0.094(-0.023) $^{+0.277}_{-0.043}$	55.07(58.29) $^{+23.87}_{-20.87}$	0.650(0.662) $^{+0.086}_{-0.221}$

view on the morphological transformation that takes place in mergers undergone by the *progenitors* of present-day discs and spheroids.

Fig 5.8 shows the mean  $\Delta\text{morph}$  in events that constitute the merger history of massive ( $M_*/M_\odot > 10^{10.5}$ ) systems at the present day. In other words, while Fig 5.2 considered the morphology of the main progenitors at the time of the merger itself, Fig 5.8 considers the set of events that make up the merger history of galaxies that are spheroids and discs at the present day. Typically, we find that events in the merger history of today's spheroids spin remnants down. For most of cosmic time, major mergers produce  $\Delta\text{morph}$  values of 0.5, while minor mergers produce values of 0.3. On the other hand, events in the merger history of today's discs produce much smaller morphological changes, with  $\Delta\text{morph}$  values of less than 0.2 and 0.1 for major and minor mergers respectively. The discrepancy between the average outcomes of mergers between the progenitors of spheroids and discs shows that the morphological evolution of galaxies must be shaped to some extent by the properties of the merging galaxies themselves, particularly the cold gas fraction and the direction of the angular momentum of the spin and orbit of the merging satellite with respect to the spin of the more massive progenitor.

## 5.4 The cumulative effect of mergers over cosmic time

In this section, we investigate the cumulative effect that mergers have on the morphological evolution of galaxies over cosmic time. We discuss the cumulative evolution of  $V/\sigma$  as a function of mass and redshift (Section 5.4.1), present the average merger histories of spheroids and discs (Section 5.4.2), explore the cumulative effect of mergers on  $V/\sigma$  in both spheroids and discs (Section 5.4.3) and quantify the relative role of major and minor mergers (and other potential processes) in driving the overall evolution of galaxy morphology over cosmic time (Section 5.4.4).

### 5.4.1 Galaxy morphology over cosmic time

Fig 5.9 shows a projection through the Horizon-AGN simulation volume, with the  $V/\sigma$  and stellar mass of galaxies shown using the colour and size of the points respectively. At  $z = 3$ , the galaxy population is relatively homogeneous in a morphological sense - at this epoch, the mean  $V/\sigma$  of the progenitors of today's disc and spheroidal galaxies are 0.95 and 0.9 respectively and very few galaxies have gained spheroidal morphologies. This indicates that significant morphological transformation is not yet underway at this redshift. On average,  $V/\sigma$  decreases towards the present day, with the most massive galaxies ( $M_* > 10^{11}M_\odot$ ) dominated by spheroidal morphologies in the local Universe. On the other hand, intermediate mass galaxies ( $10^{10.5} < M_*/M_\odot < 10^{11}$ ) undergo relatively little morphological change ( $V/\sigma$  is reduced by less than 5 per cent between  $z = 3$  and today on average) and the population remains dominated by discs at  $z = 0$ .

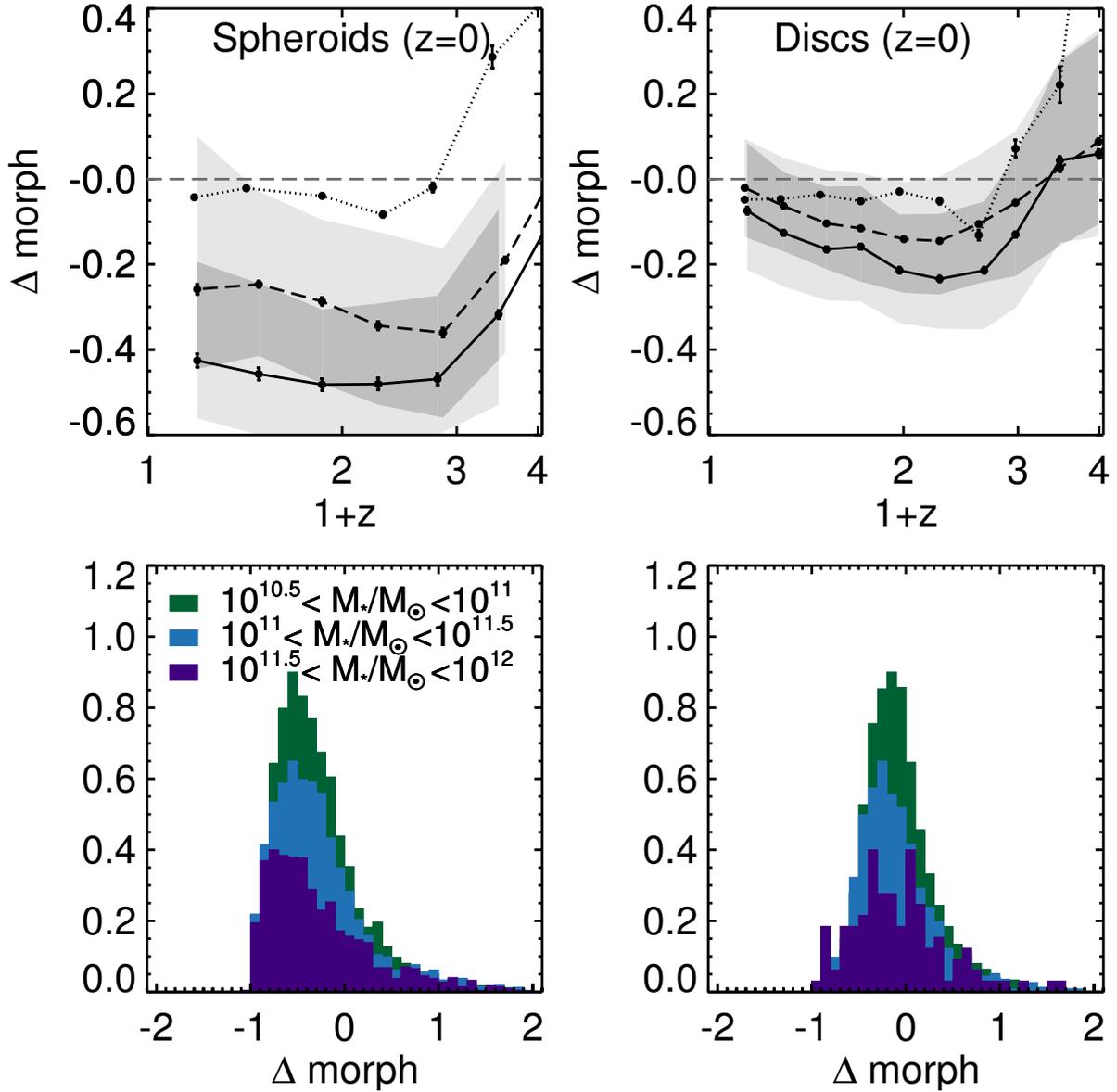


FIGURE 5.8: **Top:** Median  $\Delta \text{morph}$ , as a function of redshift, in mergers that involve *progenitors* of galaxies that have stellar masses of  $M_*/M_\odot > 10^{10.5}$  at  $z = 0$ . The left-hand column shows the redshift evolution of  $\Delta \text{morph}$  for mergers that involve progenitors of spheroids at  $z = 0$ , while the right-hand column is the corresponding plot for galaxies that are discs at  $z = 0$ . The filled regions indicate the  $\pm 1\sigma$  dispersions and the dark solid and dashed lines indicate the median  $\Delta \text{morph}$  for major and minor mergers respectively. The dotted lines indicate the median  $\Delta \text{morph}$  when these progenitors were not merging. **Bottom:** Histograms showing the corresponding distributions of  $\Delta \text{morph}$  values for major + minor mergers since  $z = 3$ . Colours indicate the final stellar masses of galaxies at  $z = 0$  (see legend).

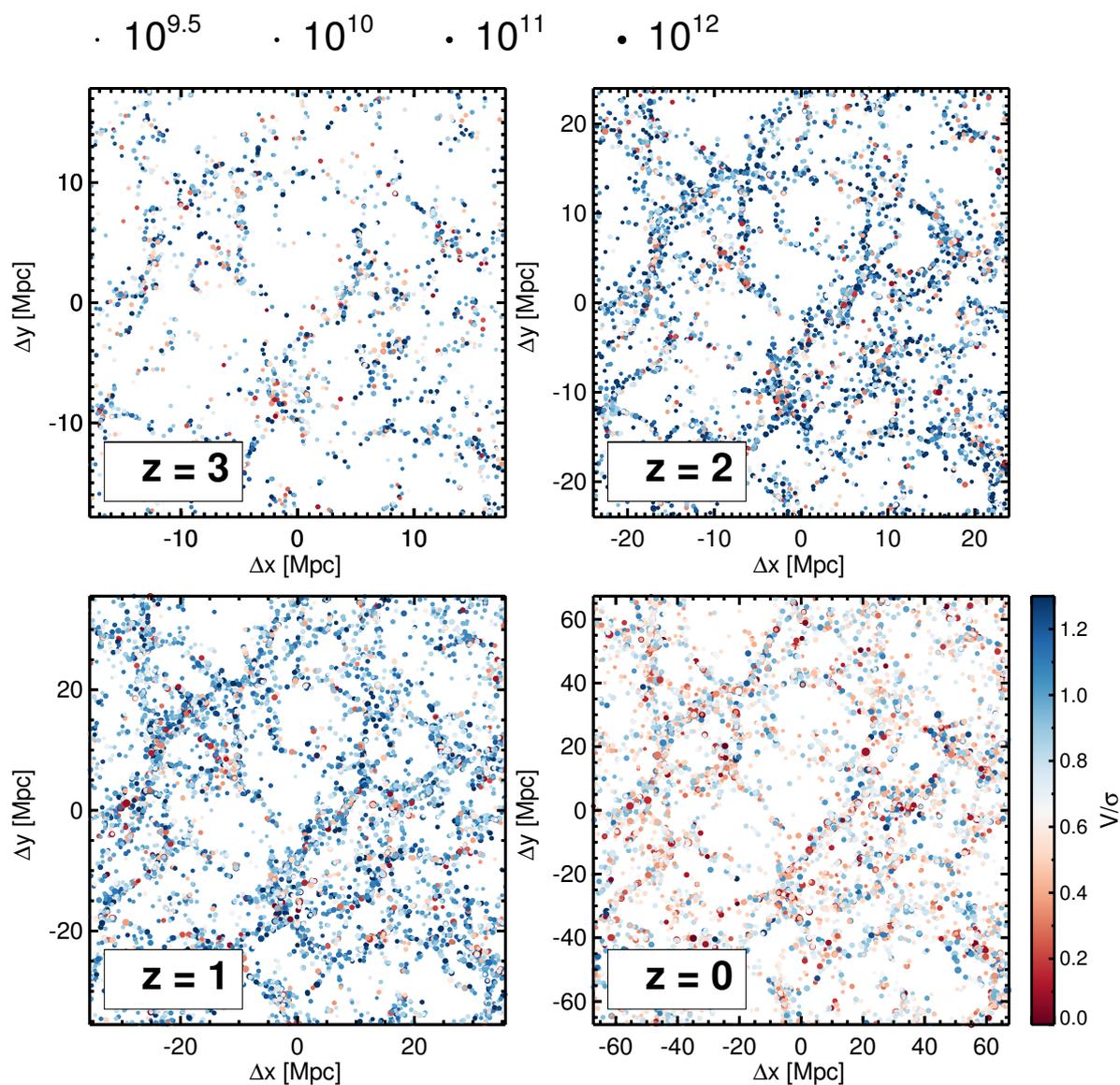


FIGURE 5.9: A projection through a 20 coMpc slice of the Horizon-AGN simulation volume showing the position (in proper Mpc) and  $v/\sigma$  of galaxies with stellar masses greater than  $10^{9.5}M_{\odot}$ . Each panel shows the simulation at a different redshift.  $v/\sigma$  is represented by the colour of each point (see colour bar). The stellar mass of each galaxy is represented by the size of the symbol, as indicated in the legend above the plots.

TABLE 5.2: Mean (median) merger histories of spheroids (white rows) and discs (shaded rows) for galaxies at  $z = 0$ .  $\pm$  indicates the  $1\sigma$  dispersion. For columns (iii) and (iv) we report the mean and standard deviation only. (i) The average redshift at which the largest mass ratio merger occurred, (ii) the average mass ratio of the largest mass ratio merger, (iii) the average number of major mergers undergone, (iv) the average number of minor mergers undergone, (v) the average ex-situ mass fraction at  $z = 0$ .

$\log_{10}(M_*/M_\odot)$	$z$ of largest (i)	largest mass ratio (ii)	# major (iii)	# minor (iv)	ex situ mass fraction (v)
10.5–11.0	1.224(1.092) $^{+0.509}_{-0.341}$	2.864(2.485) $^{+0.996}_{-0.238}$	1.178 $\pm$ 0.921	1.118 $\pm$ 1.063	0.412(0.401) $^{+0.089}_{-0.070}$
	1.169(1.027) $^{+0.614}_{-0.396}$	3.249(2.719) $^{+1.413}_{-0.444}$	0.726 $\pm$ 0.803	0.816 $\pm$ 0.872	0.201(0.176) $^{+0.091}_{-0.050}$
11.0–11.5	1.203(1.027) $^{+0.488}_{-0.322}$	2.815(2.367) $^{+0.978}_{-0.284}$	1.355 $\pm$ 1.032	1.504 $\pm$ 1.185	0.542(0.533) $^{+0.078}_{-0.061}$
	1.121(0.968) $^{+0.601}_{-0.364}$	3.244(2.679) $^{+1.291}_{-0.419}$	0.946 $\pm$ 0.981	1.150 $\pm$ 1.052	0.336(0.321) $^{+0.093}_{-0.059}$
11.5–12.0	1.235(1.092) $^{+0.520}_{-0.369}$	3.076(2.523) $^{+1.141}_{-0.282}$	1.217 $\pm$ 1.114	1.548 $\pm$ 1.246	0.609(0.597) $^{+0.081}_{-0.067}$
	0.901(0.632) $^{+0.573}_{-0.538}$	2.210(2.099) $^{+0.538}_{-0.128}$	1.238 $\pm$ 0.921	1.714 $\pm$ 1.201	0.536(0.555) $^{+0.070}_{-0.119}$

Table 5.1 summarises the average properties of galaxies at  $z = 0$  as well as aspects of their morphological evolution. Massive spheroids have undergone significant morphological transformation between  $z = 3$  and today (the value of  $V/\sigma$  today is at least 50 per cent of the value they had at  $z = 3$ , for  $M_* > 10^{10.5}M_\odot$ ). Lower mass spheroids tend to have attained spheroidal morphology later in their lifetime. Discs on the other hand undergo almost no morphological change over this period. Furthermore, as might be expected, discs are unlikely to have spheroidal morphologies at any point in their lifetime, although, interestingly, the main progenitors of extremely massive discs, that have  $10^{11.5} < M_*/M_\odot < 10^{12}$  today, spend around a third of their time as spheroids. We return to this point in the next section.

We note that galaxy morphology does not appear to be strongly correlated with galaxy environment (column (iv) of Table 1). Following Martin et al. (2018a), we estimate environment by first ranking each galaxy by their local number density, calculated using an adaptive kernel density estimation method (Breiman et al., 1977). Galaxies are then sorted into density percentiles, so that galaxies in e.g. the 0-10<sup>th</sup> percentile range represent those in the least dense environments and those in the 90-100<sup>th</sup> percentile range represent the most dense (we refer readers to Martin et al., 2018a, for more details of this procedure). The average environments of spheroids and discs are reasonably similar, when controlled for stellar mass. As we show later in Section 4.3, environment is only important for the morphological transformation of intermediate mass galaxies in the most extreme environments i.e. clusters (where a minority of the overall galaxy population is found).

## 5.4.2 Average merger histories of spheroids and discs

Table 5.2 shows the average merger histories of today’s spheroids and discs in different mass ranges. On average, today’s spheroids have undergone a greater number of major and minor mergers than discs. For example, present day spheroids in the stellar mass range  $10^{10.5} < M_*/M_\odot < 10^{11}$  have experienced an average of 1.18 (2.30) mergers with mass ratios  $> 1 : 4$  (mass ratios  $> 1 : 10$ ), since  $z = 3$ . This corresponds to 20 (34) per cent of time spent in merging episodes on average. Present-day discs in the same mass range spend less time merging, undergoing 0.73 (1.54) such mergers since  $z = 3$ , corresponding to 11 (22) per cent of their lifetime since  $z = 3$ . Similarly, present day spheroids in the stellar

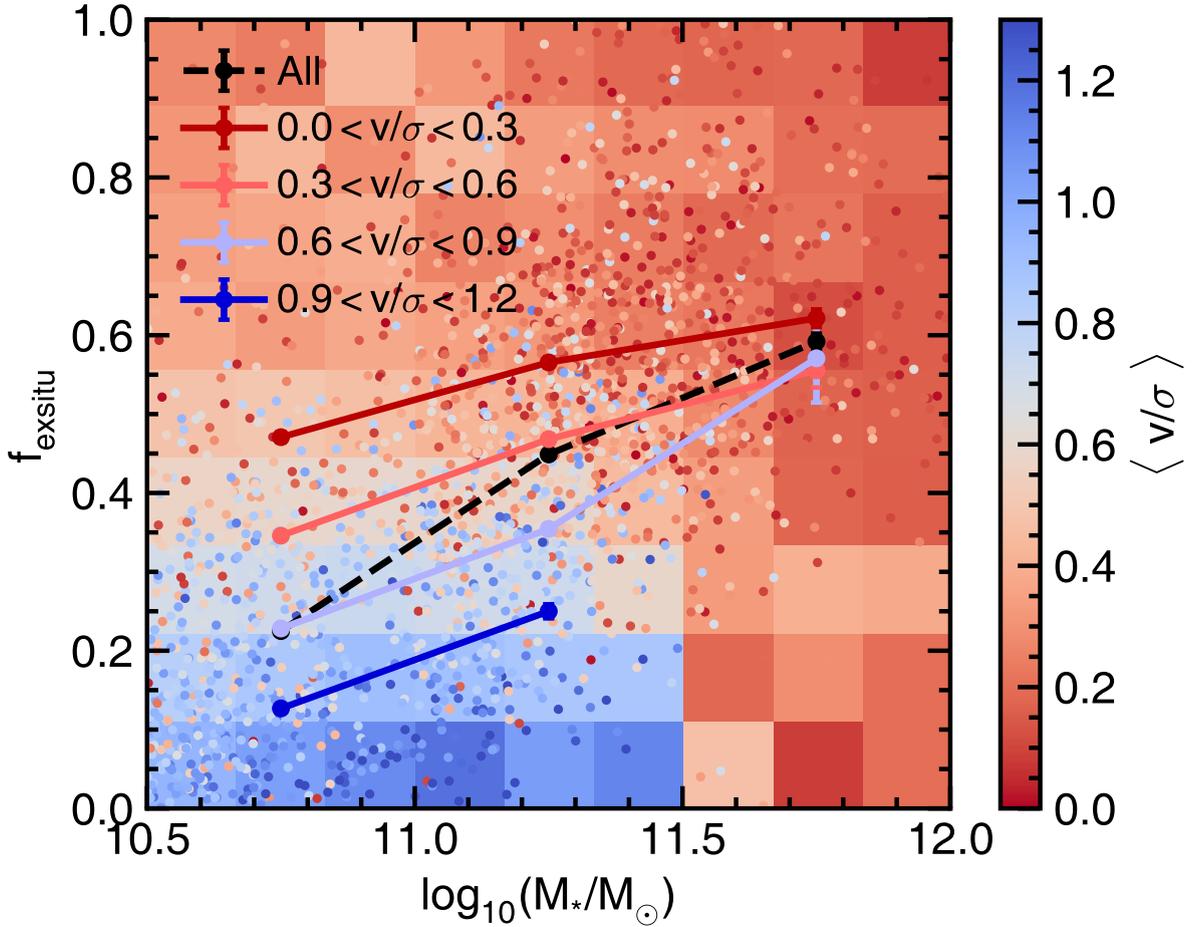


FIGURE 5.10: 2-D histogram indicating the mean  $v/\sigma$  as a function of stellar mass and ex-situ mass fraction ( $f_{\text{exsitu}}$ ). Overlaid is a scatter plot of 2000 randomly selected galaxies, and lines indicating the mean ex-situ mass fraction in four  $v/\sigma$  bins (see legend) and the whole population (black).  $v/\sigma$  values for the histogram and points are indicated by the colour bar.

mass range  $10^{11} < M_*/M_\odot < 10^{11.5}$  undergo 1.35 (2.86) mergers, whereas discs undergo 0.95 (2.10) mergers, corresponding to 27 (47) per cent and 17 (34) per cent of their lifetime for spheroids and discs respectively. Spheroids also tend to undergo mergers of larger mass ratios – on average, the largest mass ratio merger undergone by a  $10^{10.5} < M_*/M_\odot < 10^{11}$  spheroid is 1:2.9, whereas for discs it is 1:3.2. For  $10^{11} < M_*/M_\odot < 10^{11.5}$  these values are 1:2.8 and 1:3.2 for spheroids and discs respectively.

Fig 5.10 is a 2-D histogram showing the mean value of  $v/\sigma$  (indicated by the colour-bar) as a function of the stellar mass of galaxies at the present day and their ex-situ mass fractions ( $f_{\text{exsitu}}$ ), i.e. the fraction of stellar mass directly accreted from other objects via mergers. Coloured points in Fig 5.10 show the  $v/\sigma$  of a randomly selected sample of galaxies. Not unexpectedly, discs dominate the region of parameter space where  $f_{\text{exsitu}}$  is low ( $< 0.3$ ), except for  $M_* > 10^{11.5}$ , because galaxies are typically unable to reach the highest stellar masses through secular growth alone. At low  $f_{\text{exsitu}}$ , stellar mass is not a strong predictor of morphology and, as the blue region does not evolve appreciably towards higher stellar mass. This

indicates that stellar mass does not correlate significantly with morphology beyond the trend between stellar mass and ex-situ mass fraction.

Spheroids are the dominant morphological type at high  $f_{\text{exsitu}}$ , reflecting the important role that mergers play in morphological transformation. More massive galaxies are, therefore, more spheroidal on average, because mergers are the primary means by which they grow their stellar mass. However, some interesting sub-populations become apparent in this figure - at high  $f_{\text{exsitu}}$ , a significant population of discs (high  $V/\sigma$ ) remains and there is a population of massive, low  $f_{\text{exsitu}}$  slowly-rotating spheroids. For example, 13 per cent of galaxies with exsitu mass fractions greater than 0.8 still have disc-like morphologies, and 6 per cent of galaxies with exsitu mass fractions less than 0.2 have spheroidal morphologies. Although these populations are in the minority, this reflects the diversity of formation channels for discs and spheroids. We will study the formation of these sub-populations in detail in two forthcoming papers. These will show that extremely massive discs (which all have high ex-situ mass fractions) are the result of very recent disc rejuvenation from gas-rich mergers (Jackson et al. in prep), while slowly-rotating spheroids with low ex-situ mass fractions are typically the result of single minor-merger events where the orbits of the satellites at coalescence are close to being co-planar with the disc of the more massive merger progenitor (Jackson et al. in prep).

### 5.4.3 Cumulative impact of mergers on galaxy morphology over cosmic time

In Fig 5.11, we study the evolution, over cosmic time, of the mean  $V/\sigma$  of galaxies, and the contribution of mergers to the overall morphological transformation in galaxies that are spheroids (left-hand column) and discs (right-hand column) at the present day. As noted in Section 2.2 above, the accuracy of such a cumulative analysis is sensitive to the completeness of the merger history. Hence, we restrict this analysis to galaxies in the regime where completeness is high ( $M_\star > 10^{10.5} M_\odot$ ).

In each panel, we show the average change in  $V/\sigma$  of the population in question (dark solid line), the average change contributed just by major mergers (light solid line) and the average change contributed by both major and minor mergers (dashed line). For example, if the mean evolution of  $V/\sigma$  for a galaxy population, shown by the dark solid line, overlaps perfectly with that due to mergers (i.e. the dashed line), then mergers are responsible for all the morphological transformation in these galaxies. Similarly, if the mean  $V/\sigma$  at the present day is higher than what would be expected due to mergers (i.e. the dark solid line is above the dashed line at  $z \sim 0$ ), then there must be other processes (e.g. accretion) that are responsible for spinning the galaxies up over their lifetimes.

For spheroids that have stellar masses greater than  $10^{10.5} M_\odot$  at  $z = 0$ , major and minor mergers (i.e. mergers with mass ratios greater than 1 : 10) together explain essentially all the morphological evolution. However, it is important to note that the morphological evolution is not induced by major mergers alone since, if that had been the case, the dark solid lines would have overlapped with the lighter solid lines (which correspond to just major mergers). Indeed, the overall change in  $V/\sigma$  of such massive spheroids

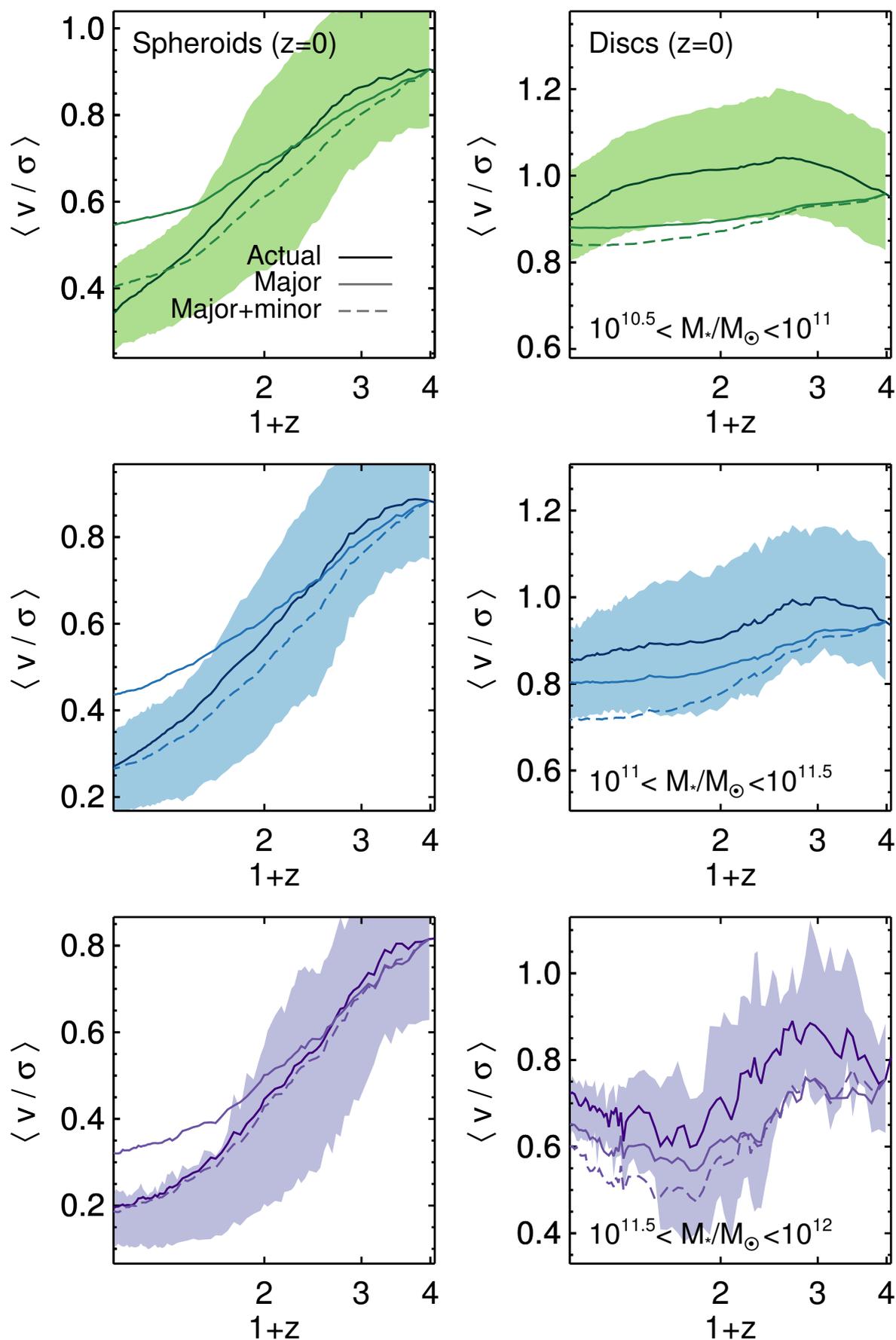


FIGURE 5.11: Evolution of the mean  $V/\sigma$ , as a function of redshift and stellar mass, of the main (i.e. more massive) progenitors of spheroids (left-hand column) and discs (right-hand column). The dark solid line indicates the actual  $V/\sigma$  evolution, the light solid line indicates the  $V/\sigma$  evolution due to major mergers alone and the dashed line shows the  $V/\sigma$  evolution due to major + minor mergers.

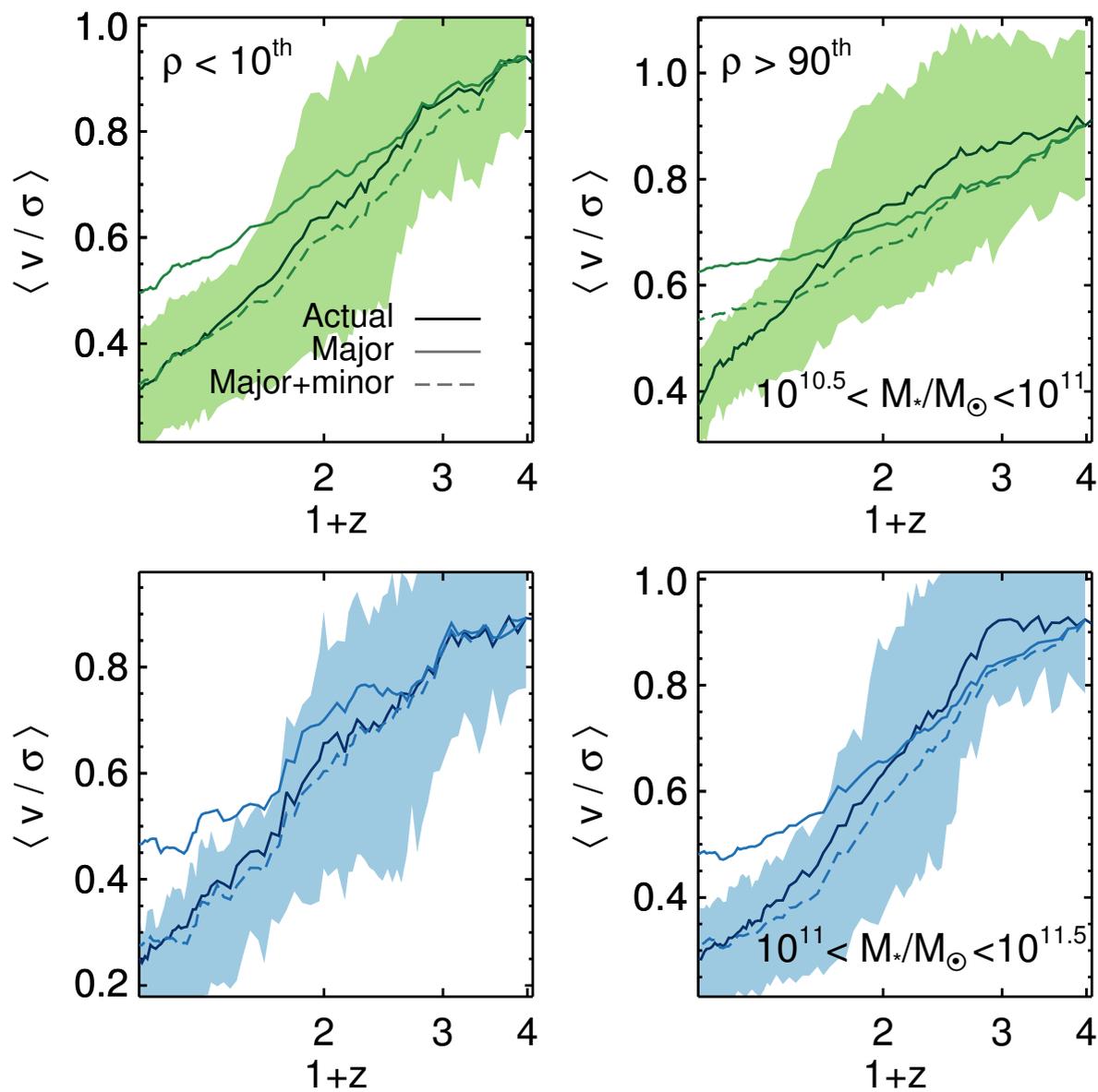


FIGURE 5.12: Evolution of the mean  $v/\sigma$ , as a function of redshift and stellar mass, of the main (i.e. more massive) progenitors of spheroids that fall in the lowest 10<sup>th</sup> percentile in local density (left-hand column) and those that fall in the highest 10<sup>th</sup> percentile in local density (right-hand column). The highest 10<sup>th</sup> percentile in local density typically correspond to cluster environments (see Martin et al., 2018a).

is larger than what can be produced by major mergers alone, indicating that *minor mergers have a significant role to play in the transformation of morphology (disc to spheroid) over cosmic time*. Indeed, after  $z \sim 1$  (where the light solid and dashed lines start diverging), minor mergers are responsible for the majority of the morphological transformation in spheroids. It is worth noting here that, since all the morphological transformation can be accounted for by major and minor mergers, other processes, such as fly-bys, or the formation of low  $V/\sigma$  stars from low-angular momentum gas (e.g. via feeding from counter-rotating filaments where the net angular momentum is close to zero, e.g. Danovich et al. (2015)) are unlikely to be significant drivers of morphological transformation over cosmic time.

The morphological evolution of discs is qualitatively different at all stellar masses, and is not dominated by mergers. On average, the  $V/\sigma$  values of disc galaxies today are higher than they would be if their evolution was being driven by mergers. This indicates that discs differ from spheroids, in the sense that they are consistently spun up by gas accretion at all epochs. This accretion counteracts the decrease in  $V/\sigma$  due to merging, especially at high redshift when the Universe is more gas rich.

It is important to note that, while mergers explain the majority of the morphological evolution that leads to the formation of spheroids, the morphological evolution of the disc population *cannot* be entirely explained by a *lack* of mergers. As described above in Table 5.2, while spheroids do tend to have more mergers than discs at a given stellar mass, the average merger histories of the two morphological classes are not too dissimilar. This is not surprising, since the merger history of a galaxy is expected to be a strong function of its stellar mass (e.g. Stewart et al., 2008; Rodriguez-Gomez et al., 2015). At a given stellar mass, therefore, spheroids do not undergo many more mergers than discs, nor do they undergo mergers of appreciably higher mass ratios (Table 5.2). However, the mergers that discs do undergo clearly do not produce the same morphological change as that seen in mergers that involve spheroids. For example, mergers with mass ratios  $> 1 : 10$  produce a mean fractional change  $((\Delta V/\sigma)/(V/\sigma)_{z=3})$  of -0.72 for spheroids of stellar mass  $10^{11} < M_*/M_\odot < 10^{11.5}$  between  $z = 3$  and  $z = 0$ , yet only produce a fractional change of -0.12 in discs, where  $\Delta V/\sigma$  is the  $V/\sigma$  at  $z = 0$  subtracted from that at  $z = 3$ .

The explanation lies largely in the actual properties of the galaxies and mergers themselves. As described in Section 5.4.1, discs have higher gas fractions, for a given stellar mass, than spheroids. Since gas rich mergers have a higher likelihood of producing  $\Delta \text{morph} > 0$  (Fig 5.3), any mergers they undergo tend not to significantly decrease  $V/\sigma$ , as new stars formed from the residual gas act to counteract some of the morphological transformation, by adding to the rotational component of the system. Disc rejuvenation, especially at early epochs, can also be assisted by cosmological accretion, which enables galaxies to reacquire cold gas and reform discs, at least until a critical mass of  $\sim 10^{10.5} M_\odot$  (e.g. Welker et al., 2017). Finally, as we have shown in Section 5.3.3, the spin of the satellite and the orbital configuration of the merger, relative to the spin of the more massive progenitor, plays a role in determining the properties of the remnant. On average, present-day discs have undergone more prograde mergers, whereas their spheroidal counterparts of comparable stellar mass have undergone more retrograde mergers. This also plays a part in preserving the morphological properties of discs.

It is worth noting the fraction of time galaxies are considered to be in a merging phase, given our chosen timescale (2 Gyr). Since more massive galaxies undergo a greater number of mergers, they spend more of their time merging on average. Galaxies in the stellar mass range  $10^{11.5} < M_*/M_\odot < 10^{12}$  undergo major and major+minor mergers for 32 and 55 per cent of their lifetimes respectively. For galaxies with stellar masses in the range  $10^{10.5} < M_*/M_\odot < 10^{11}$ , the corresponding values are 13 and 25 per cent. Importantly, our choice of a 2 Gyr timescale does not result in a scenario where most of galaxy's evolution takes place during merging episodes, thereby underestimating the role of secular evolution.

We complete this section by exploring, in Fig 5.12, the role of environment in producing morphological transformation, by considering the  $V/\sigma$  evolution of spheroid progenitors in the densest and least dense environments. While the morphological transformation of the most isolated galaxies (those in the bottom 10<sup>th</sup> percentile of local number density) is entirely accounted for by major and minor mergers, other processes must be invoked in order to explain the evolution of morphology in the densest environments (those in the top 10<sup>th</sup> percentile, which correspond to clusters, see Martin et al. (2018a)). Mergers account for around two-thirds of the morphological change in intermediate-mass galaxies ( $10^{10.5} < M_*/M_\odot < 10^{11}$ ) in the densest environments, while the remaining one-third must be due to other processes e.g. harassment. Our result is consistent with the results of Choi & Yi (2017), who find that mechanisms other than mergers contribute to the morphological transformation of such intermediate-mass galaxies in cluster environments. We note that, unlike the simulations of Choi & Yi (2017), rich cluster environments are rare in the cosmological Horizon-AGN volume (there is only one rich Coma-like cluster in the entire volume). Thus, the Horizon-AGN volume is skewed towards low-density environments, where environmental processes are, by definition, less important.

#### 5.4.4 The fraction of global morphological transformation triggered by major and minor mergers

We complete our study by quantifying the fraction of overall morphological change in massive galaxies, since  $z = 3$ , that is attributable to mergers. We use the following definition for the fraction of the morphological transformation that is driven by mergers:

$$f = \Delta(V/\sigma)_m / \Delta(V/\sigma), \quad (5.6)$$

where  $\Delta(V/\sigma)_m$  is the total change in  $V/\sigma$  during merging episodes since  $z = 3$  and  $\Delta(V/\sigma)$  is the total change in  $V/\sigma$  since  $z = 3$ . By this definition,  $f$  can be positive or negative (in cases where the  $V/\sigma$  of galaxies increases between  $z = 3$  and today). It may also exceed 1 in cases where the  $\Delta(V/\sigma)_m$  is larger than  $\Delta(V/\sigma)$ . In Fig 5.13 we present 2-D histograms that show the mean value of  $f$ , as a function of stellar mass and  $V/\sigma$ , for major mergers (mass ratios  $> 1 : 4$ ; top panel) and major + minor mergers (mass ratios  $> 1 : 10$ ; bottom panel). Values in brackets indicate the mean fraction of ex-situ stellar mass

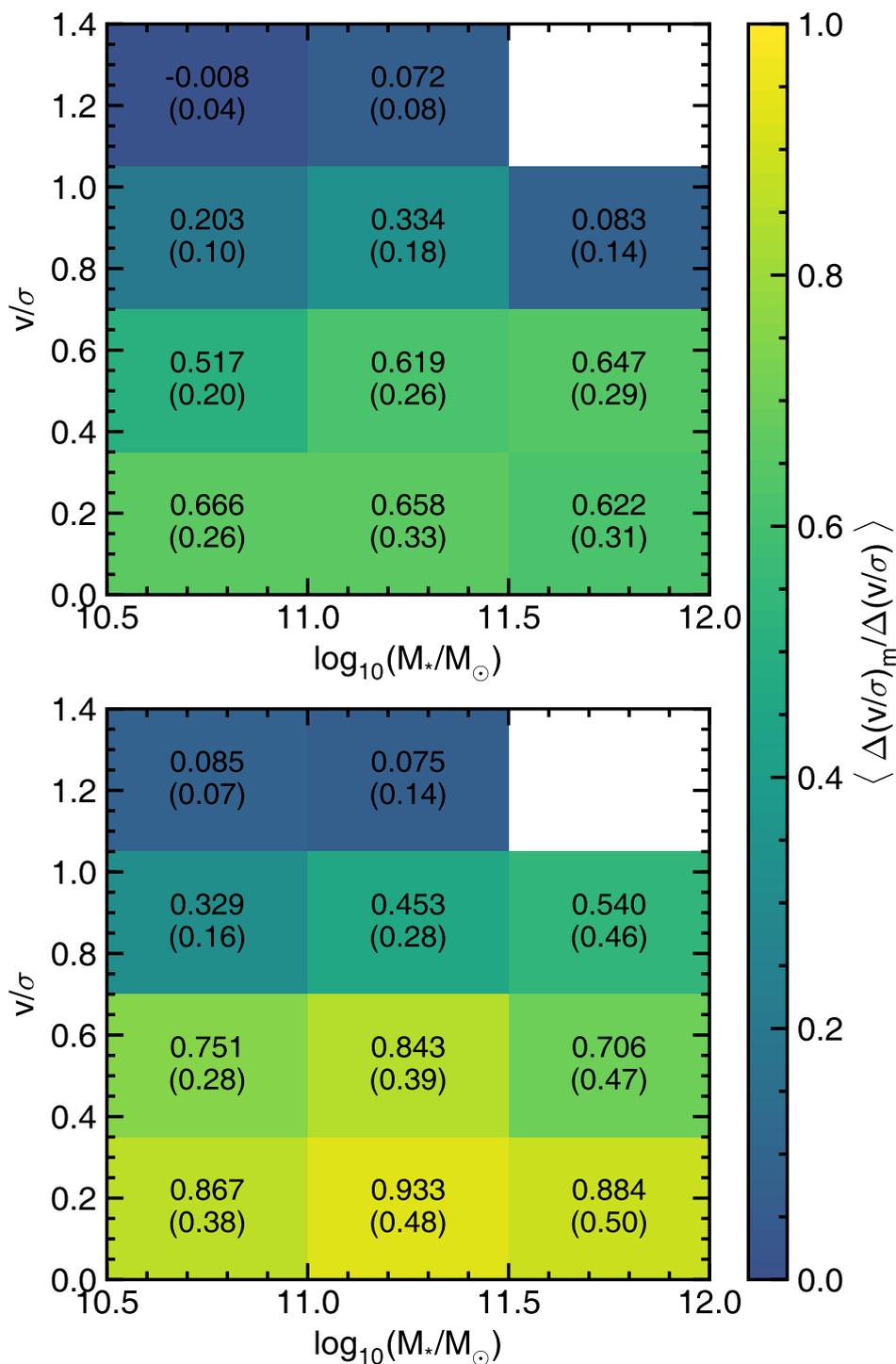


FIGURE 5.13: 2-D histograms showing the mean fraction of morphological change that occurs during merging episodes, as a function of the stellar mass and  $v/\sigma$  of the remnant at  $z = 0$ . This mean fraction is indicated by the colour-bar and also shown in the centre of each bin. The bracketed value indicates the ex-situ stellar mass fraction of the remnant. **Top:** The mean fraction of morphological change and ex-situ mass fractions produced by major mergers. **Bottom:** The mean fraction of morphological change and ex-situ mass fractions produced by major + minor mergers.

that is accreted in mergers with these mass ratios in each bin. The most spheroidal galaxies are those with the highest values of  $f$ . Of the morphological transformation produced by mergers, minor mergers are responsible for a significant minority (around 30 per cent) over cosmic time. And, as noted in the previous section, the role of minor mergers becomes dominant at later epochs ( $z < 1$ ), where they drive almost all the morphological transformation in the spheroid population.

Mergers are the dominant means of morphological transformation in high mass spheroids ( $M_* > 10^{10.5} M_\odot$ ), where they are responsible for at least 75 per cent of the change. The vast majority (90 per cent) of the morphological change in the most spheroidal galaxies ( $V/\sigma < 0.3$ ) is the result of mergers. In discs with significant dispersion-dominated components, a large proportion of morphological change is a result of mergers ( $\sim 40$  per cent), particularly minor mergers. Our results are consistent with those of Choi et al. (2018), who find that the spin change since  $z = 1$  in 94 per cent of massive central early-type galaxies is dominated by mergers.

While discs typically have non-negligible ex-situ mass fractions, a larger fraction of this mass is derived from minor mergers (averaged over all masses, around 70 per cent of ex-situ mass in spheroids is derived from major mergers, compared to 50 – 60 per cent for discs). Minor and major mergers do not play a significant role either in the morphological transformation or mass assembly of the most rotationally-supported discs ( $V/\sigma > 1$ ). Taken together, major and minor mergers account for only  $\sim 10$  per cent of the stellar mass of such discs and are responsible for less than 10 per cent of their morphological transformation. What little morphological transformation takes place in these galaxies is instead dominated by spin up due to accretion of gas.

## 5.5 Conclusions

We have used Horizon-AGN, a cosmological hydrodynamical simulation, to study the processes that drive morphological transformation across cosmic time. In particular, we have (1) studied the average merger histories of discs and spheroids over cosmic time, (2) quantified the magnitude of the morphological change (i.e. disc to spheroid) that is imparted by major and minor mergers as a function of redshift and stellar mass, (3) explored the effect of gas fraction on these morphological changes, (4) studied the effect of orbital configuration in determining the properties of merger remnants and (5) quantified the overall contribution of major/minor mergers and other processes to the creation of spheroidal galaxies. Our key conclusions are as follows:

- *The morphological evolution of spheroids with stellar masses greater than  $10^{10.5} M_\odot$  at  $z = 0$  can be largely explained by relatively high-mass-ratio ( $> 1 : 10$ ) mergers.* Essentially all of the morphological evolution in galaxies that are spheroids at  $z = 0$  took place in mergers with mass ratios greater than 1 : 10. However, major mergers (mass ratios  $> 1 : 4$ ) *alone* are not sufficient. Around a third of the overall morphological change in massive spheroids is driven by minor mergers (mass

ratios between 1 : 4 and 1 : 10). Furthermore, minor mergers become the dominant channel for morphological change at late epochs, driving the bulk of the morphological transformation in spheroid progenitors at  $z < 1$ . Finally, across the *general* galaxy population (i.e. across all masses and environments), other processes, such as fly-bys, harassment or the formation of low  $V/\sigma$  stars from direct accretion of low-angular momentum gas, are relatively insignificant in transforming galaxy morphology, since all the morphological change in massive spheroids is accounted for by mergers with mass ratios greater than 1 : 10.

- *In clusters, environmental processes like harassment do play a role in morphological transformation for intermediate mass galaxies ( $10^{10.5} < M_*/M_\odot < 10^{11}$ ).* Around a third of the morphological transformation of  $10^{10.5} < M_*/M_\odot < 10^{11}$  spheroids in the *densest* environments is a result of processes other than mergers. However, since the vast majority of galaxies do not reside in clusters, almost all morphological transformation in the general population of massive spheroids can be explained by mergers alone.
- *The outcome of a merger is strongly influenced by the gas fraction of the merging pair.* Mergers involving spheroids with higher gas fractions are more likely to produce remnants with increased  $V/\sigma$ . For example, at  $z \sim 0$ , gas poor major mergers ( $f_{gas} < 0.2$ ) typically produce fractional morphological changes ( $\Delta$  morph) of -0.1 in spheroids, whereas gas rich major mergers ( $f_{gas} > 0.3$ ) typically produce significantly more positive changes of 0.38 over the course of each merger for galaxies in the mass range  $10^{10.5} < M_*/M_\odot < 10^{11}$ .
- *The orbital configuration of a merger has a measurable impact on the properties of the remnant.* Mergers that are prograde (i.e. where the spin of the satellite and the orbital angular momentum are aligned with the spin of the more massive galaxy) typically produce smaller decreases in the spin of discs (i.e. milder morphological transformation) than their retrograde counterparts. While both types of events produce similar morphological changes at  $z > 2$ , the average change due to retrograde mergers is around twice as large as that due to their prograde counterparts at  $z \sim 0$ .
- *Spin up due to cosmological accretion is an important effect, especially at early epochs.* In the early universe ( $z > 2$ ), stellar mass forms in discs more rapidly than it can be removed by mergers. At later times, gas-rich minor mergers and existing reservoirs of gas in the halo become increasingly important for spinning up galaxies as the effectiveness of cosmological accretion declines.
- *On average, spheroids have undergone more mergers (and mergers with higher mass ratios) since  $z = 3$  compared to discs of equivalent stellar mass (Table 5.2).* The fraction of stellar mass formed

ex-situ (i.e. accreted directly via mergers) is around 1.5–2 times higher in massive spheroids. In addition, the average fractional morphological transformation ( $\Delta\text{morph}$ ) induced per merger is around a factor of 2 larger in disc progenitors compared to in spheroid progenitors.

- *However, the survival of discs to  $z = 0$  cannot be explained solely by a lack of mergers.* Although the progenitors of discs undergo fewer mergers than equivalent mass spheroids, the disparity between their merger histories is not large enough to account for their relative lack of morphological evolution. The discrepancy instead stems from the properties of the mergers themselves. Mergers involving the progenitors of today’s discs tend to be more gas-rich, which promotes disc rejuvenation subsequent to a merger. In addition, discs typically undergo a greater fraction of prograde mergers (compared to spheroids of a similar stellar mass) which induce milder morphological transformation and thus assist in disc survival.

Finally, while we have quantified the creation of spheroids in a broad sense in this chapter, it is worth recalling some interesting sub-populations identified in this study, which will be the subject of forthcoming work. First is the existence of a small population of very massive discs ( $M_*/M_\odot > 10^{11.5}$ ), which have rich merger histories like their spheroidal counterparts (see Table 5.2) and, as a result, high ex-situ mass fractions. Given that mergers are significant drivers of morphological transformation, a key question is: why do such extremely massive discs exist at all? As we will show in a forthcoming paper (Jackson et al. in prep), the existence of these disks is due to very recent disk rejuvenation via gas-rich mergers. Equally interesting is the population of massive, slowly-rotating (spheroidal) galaxies with low ex-situ mass fractions, which indicate that these systems have not undergone many mergers, yet are not rotationally-supported as one might expect of a galaxy whose assembly history is largely merger-free. As we will demonstrate in another paper (Jackson et al. in prep), these systems are created by a minor-merger, where the orbit of the satellite at coalescence is almost co-planar with the disc of the more massive merger progenitor (Jackson et al. in prep).

## Chapter 6

# The formation and evolution of low-surface-brightness galaxies

### 6.1 Introduction

Our understanding of galaxy evolution is intimately linked to the part of the galaxy population that is visible at the surface-brightness limits of past and current wide-area surveys. Not only do these thresholds determine the extent of our empirical knowledge, but the calibration of our theoretical models (and therefore our understanding of the physics of galaxy evolution) is strongly influenced by these limits. In recent decades, a convergence of wide-area surveys like the SDSS (Abazajian et al., 2009) and large-scale numerical simulations (e.g. Croton et al., 2006; Dubois et al., 2014; Vogelsberger et al., 2014) has had a transformational impact on our understanding of galaxy evolution. While these surveys have mapped the statistical properties of galaxies, comparison to cosmological simulations – first via semi-analytical models (e.g. Somerville & Primack, 1999; Cole et al., 2000; Benson et al., 2003; Bower et al., 2006; Croton et al., 2006) and more recently via their hydrodynamical counterparts (e.g. Dubois et al., 2014; Vogelsberger et al., 2014; Schaye et al., 2015; Kaviraj et al., 2017) – has enabled us to understand the physical drivers of galaxy formation over much of cosmic time.

The SDSS, which has provided much of the discovery space at low and intermediate redshift, starts becoming incomplete at an  $r$ -band effective surface-brightness,  $\langle\mu\rangle_e^1$ , of  $\sim 23$  mag arcsec $^{-2}$  (e.g. Driver et al., 2005; Blanton et al., 2005; Zhong et al., 2008; Bakos & Trujillo, 2012). This is primarily due to the lack of depth of the survey but also due, in part, to the standard SDSS pipeline not being optimised for structures that are close to the sky background. Indeed, while bespoke sky subtraction on SDSS images is able to mitigate some of these issues and reveal low-surface-brightness galaxies (LSBGs), these objects do not form the bulk of the population that are visible in such surveys (e.g. Kniazev et al., 2004; Williams et al., 2016). Thus, while it is clear that a (largely) hidden Universe exists just below the

---

<sup>1</sup>The effective surface-brightness,  $\langle\mu\rangle_e$ , is defined as the mean surface-brightness within an effective radius.

surface-brightness limits of current large-area surveys, the detailed nature of galaxies in this LSB domain remains largely unexplored, both observationally and in our theoretical models of galaxy evolution. Indeed, the existence of large numbers of faint, undiscovered galaxies has deep implications for our understanding of galaxy evolution. Since our current view of how galaxies evolve is largely predicated on high-surface-brightness galaxies (HSBGs;  $\langle\mu\rangle_e < 23$  mag arcsec<sup>-2</sup>), this almost certainly leads to potentially significant biases in our understanding of the evolution of the baryonic Universe. Mapping the LSB domain empirically, and exploring the mechanisms by which galaxies in this regime form and evolve, is central to a complete understanding of galaxy evolution.

The existence of a population of faint, diffuse, (typically) low-mass galaxies has been known since the mid-1980s (e.g. Sandage & Binggeli, 1984). However, in the decades following their discovery, very few additional examples were identified (e.g. Impey et al., 1988; Bothun et al., 1991; Turner et al., 1993; Dalcanton et al., 1997), largely due to the surface-brightness limits of contemporary observations. Only very recently, thanks to advances in the sensitivity and field of view of modern instruments (e.g. Miyazaki et al., 2002; Kuijken et al., 2002; Miyazaki et al., 2012; Diehl & Dark Energy Survey Collaboration, 2012; Abraham & van Dokkum, 2014; Torrealba et al., 2018) and the introduction of new observational and data-analysis techniques (e.g. Akhlaghi & Ichikawa, 2015; Prole et al., 2018), has the identification of significant samples of LSBGs become possible (e.g. van Dokkum et al., 2015; Koda et al., 2015; Muñoz et al., 2015; van der Burg et al., 2016; Janssens et al., 2017; Venhola et al., 2017; Greco et al., 2018b).

While modern instruments are enabling the study of systems at significantly fainter surface-brightnesses than was previously possible, deep-wide surveys and spectroscopic follow-up of areas large enough to contain significant populations of LSBGs outside dense, cluster environments remain prohibitively expensive. As a result, the LSB domain remains poorly explored in groups (e.g. Smith Castelli et al., 2016; Merritt et al., 2016; Román & Trujillo, 2017a,b) and the field (e.g. Martínez-Delgado et al., 2016; Papastergis et al., 2017; Leisman et al., 2017). This is particularly true for the extremely faint, diffuse end of the LSB population, often referred to, in the contemporary literature, as ‘ultra-diffuse’ galaxies (UDGs; van Dokkum et al. (2015)).

Recent work suggests that, while LSBGs may be ubiquitous in clusters (e.g. Koda et al., 2015), they occur across all environments (Román & Trujillo, 2017a; Merritt et al., 2016; Papastergis et al., 2017). However, the contribution of the LSB population to the number, mass and luminosity density of the Universe remains unclear. A number of studies (e.g. Davies et al., 1990; Dalcanton et al., 1997; O’Neil & Bothun, 2000; Minchin et al., 2004; Haberzettl et al., 2007) have argued that LSBGs represent a significant fraction of objects at the faint end of the luminosity function and dominate the number density of galaxies at the present day. They may also account for a significant fraction of the dynamical mass budget ( $\sim 15$  per cent) (e.g. Driver, 1999; O’Neil & Bothun, 2000; Minchin et al., 2004) and the neutral hydrogen density (Minchin et al., 2004) in today’s Universe, although they are thought to contribute a

minority (a few per cent) of the local luminosity and stellar mass density (Bernstein et al., 1995; Driver, 1999; Hayward et al., 2005).

While new observations are opening up the LSB domain, the formation mechanisms of LSBGs and their relationship to the HSBG population, on which our understanding of galaxy evolution is predicated, remains poorly understood. Compared to the HSBG population, LSBGs, and UDGs in particular, appear to be relatively quenched, dispersion-dominated systems which largely occupy the red sequence (van Dokkum et al., 2015, 2016; Ferré-Mateu et al., 2018; Ruiz-Lara et al., 2018). In lower-density environments, however, they are typically bluer (i.e. unquenched) possibly reflecting a wide range of formation scenarios across different environments (e.g. Román & Trujillo, 2017b; Zaritsky et al., 2019). LSBGs are typically extremely extended systems for their stellar mass, with low ( $n \lesssim 1$ ) Sérsic indices (Koda et al., 2015). While there does not appear to be a single evolutionary path that is able to explain the formation of these objects, a number of mechanisms capable of producing such extended, relatively quenched systems have been proposed.

For example, van Dokkum et al. (2015) have proposed that UDGs may be failed Milky Way-like ( $L^*$ ) galaxies, which were quenched at high redshift as a result of gas stripping. However, observational evidence using globular cluster abundances (Beasley & Trujillo, 2016; Peng & Lim, 2016; Amorisco et al., 2018), velocity dispersions (e.g Toloba et al., 2018), weak lensing measurements (e.g Sifón et al., 2018), stellar populations (e.g Ferré-Mateu et al., 2018; Ruiz-Lara et al., 2018), and the spatial distributions and abundances of the galaxies themselves (e.g Román & Trujillo, 2017a), largely supports the idea that the vast majority of LSBGs are low-mass (i.e. dwarf) galaxies that are hosted by correspondingly low mass dark-matter haloes, except perhaps in a small number of extreme cases (e.g van Dokkum et al., 2016; Beasley et al., 2016).

UDGs, for example, have been suggested to form as the result of various channels, including anomalously high spin (e.g Amorisco & Loeb, 2016; Amorisco et al., 2016; Rong et al., 2017; Leisman et al., 2017), gas outflows due to supernova (SN) feedback (e.g Di Cintio et al., 2017; Chan et al., 2018) and strong tidal fields or mergers (e.g. Carleton et al. 2018; Conselice 2018; Abraham et al. 2018; Baushev 2018, but see Mowla et al. 2017). Thus, while the exact mechanisms responsible for producing UDGs are still debated, there is broad consensus that the progenitors of the majority of UDGs are galaxies in low mass haloes, rather than ‘failed’ high mass haloes where galaxies were prevented from forming in the first place.

In this chapter, we use Horizon-AGN, a cosmological hydrodynamical simulation (Dubois et al., 2014; Kaviraj et al., 2017), to perform a comprehensive study of galaxies in the LSB domain. The use of a cosmological simulation is essential for this exercise, since it enables us to study baryonic processes that are likely to drive LSBG formation (e.g. SN feedback, ram-pressure stripping and tidal perturbations) within fully resolved cosmological structure. We explore the predicted properties of a complete sample of LSBGs in today’s Universe across all environments, investigate the evolution of their progenitors over

cosmic time and study the role of key processes (e.g. SN feedback, tidal perturbations and ram-pressure stripping) in creating these systems.

This chapter is structured as follows. In Section 6.2, we present an overview of the Horizon-AGN simulation, including the treatment of baryonic physics, the definition of galaxies and their merger trees, and the identification of LSBGs. In Section 6.3, we compare the present-day properties of LSBGs to a sample of their HSB counterparts that have the same distribution of stellar masses. In Section 6.4, we explore the evolution of key properties in which LSBGs and HSBGs diverge the most (gas fractions, effective radii and density profiles) and which are, therefore, central to the formation of LSB systems. In Section 6.5, we quantify the processes (SN feedback, ram pressure stripping and tidal perturbations) that are responsible for creating LSBGs over cosmic time. We summarise our results in Section 6.6.

## 6.2 The Horizon-AGN simulation

In this study we employ Horizon-AGN, a cosmological-volume hydrodynamical simulation (Dubois et al., 2014), that is based on RAMSES (Teyssier, 2002), an adaptive mesh refinement (AMR) Eulerian hydrodynamics code. Horizon-AGN simulates a box with a length of  $100 h^{-1}$  coMpc. Initial conditions are taken from a *WMAP7*  $\Lambda$ CDM cosmology (Komatsu et al., 2011), using  $1024^3$  dark matter (DM) particles, with a mass resolution of  $8 \times 10^7 M_{\odot}$ . An initially uniform  $1024^3$  cell grid is refined, according to a quasi Lagrangian criterion (when 8 times the initial total matter resolution is reached in a cell), with the refinement continuing until a minimum cell size of 1 kpc in proper units is achieved. Additional refinement is allowed at each doubling of the scale factor, in order to keep the resolution constant in physical units. Note that, in addition to the hydrodynamics, the AMR cells also define the force softening for the dark matter and baryons. We direct readers to Appendix C.2 for a discussion of the effect of the resolution of Horizon-AGN on the sizes of galaxies.

Horizon-AGN produces good agreement with key observables that trace the cumulative evolution of galaxies across at least 95% of cosmic time: stellar mass/luminosity functions, the star formation main sequence, rest-frame UV-optical-near infrared colours and the merger and star formation histories of galaxies (Kaviraj et al., 2015b, 2017). The simulation also reproduces black-hole (BH) demographics, such as the luminosity and mass functions of BHs, the evolution of BH mass density over cosmic time and correlations between BH and galaxy mass from  $z = 3$  to  $z = 0$  (Volonteri et al., 2016; Martin et al., 2018b). Finally, Horizon-AGN produces good agreement with the morphological mix of the local Universe, with the predicted galaxy morphologies reproducing the observed fractions of early and late-type galaxies that have intermediate and high stellar masses (Dubois et al., 2016; Martin et al., 2018a).

In the following sections, we describe aspects of the simulation that are particularly relevant to this study: the treatment of baryonic matter (gas and stars), the identification of galaxies, construction of their merger trees and the selection of LSBGs.

### 6.2.1 Baryons

Gas cooling is assumed to take place via H, He and metals (Sutherland & Dopita, 1993), down to a temperature of  $10^4$  K. A uniform UV background is switched on at  $z = 10$ , following Haardt & Madau (1996). Star formation proceeds via a standard 2 per cent efficiency (e.g. Kennicutt, 1998), when the hydrogen gas density reaches  $0.1 \text{ H cm}^{-3}$ . The stellar-mass resolution in Horizon-AGN is  $4 \times 10^6 M_{\odot}$ .

The simulation employs continuous stellar feedback that includes momentum, mechanical energy and metals from stellar winds and both Type II and Type Ia supernovae (SNe). Feedback from stellar winds and Type II SNe is implemented using STARBURST99 (Leitherer et al., 1999, 2010), via the Padova model (Girardi et al., 2000) with thermally pulsating asymptotic giant branch stars (Vassiliadis & Wood, 1993). The ‘Evolution’ model of Leitherer et al. (1992) is used to calculate the kinetic energy of stellar winds. Matteucci & Greggio (1986) is used to determine the implementation of Type Ia SNe, assuming a binary fraction of 5% (Matteucci & Recchi, 2001), with chemical yields taken from the W7 model of Nomoto et al. (2007). Stellar feedback is assumed to be a heat source after 50 Myrs, because after this timescale the bulk of the energy is liberated via Type Ia SNe that have time delays of several hundred Myrs to a few Gyrs (e.g. Maoz et al., 2012). These systems are not susceptible to large radiative losses, since stars will disrupt or migrate away from their birth clouds after a few tens of Myrs (see e.g. Blitz & Shu, 1980; Hartmann et al., 2001).

We note that using an AMR refinement scheme based on total matter density allows us to resolve the gas content of galaxies out to larger radii, since the resolution in the outskirts of the galaxy is principally set by the DM mass, where it dominates rather than the gas mass, which is generally small (as would be the case in smoothed particle hydrodynamics schemes, for example). This is important for the study of diffuse galaxies, particularly those with small gas fractions.

### 6.2.2 Identifying galaxies and merger trees

To identify galaxies we use the ADAPTAHOP structure finder (Aubert et al., 2004; Tweed et al., 2009), applied to the distribution of star particles. Structures are identified if the local density exceeds 178 times the average matter density, with the local density being calculated using the 20 nearest particles. A minimum number of 50 particles is required to identify a structure. This imposes a minimum galaxy stellar mass of  $2 \times 10^8 M_{\odot}$ . We then produce merger trees for each galaxy in the final snapshot ( $z \sim 0.06$ ), with an average timestep of  $\sim 130$  Myr, which enables us to track the main progenitors (and thus the assembly histories) of individual galaxies.

We note that, due to the minimum mass limit described above ( $2 \times 10^8 M_{\odot}$ ), the LSBGs we study in this chapter have masses in excess of this threshold. These systems are, therefore, typically at the higher mass end of the LSBG populations that have been studied in recent observational work.

### 6.2.3 Surface-brightness maps and selection of LSBGs

We use the Bruzual & Charlot (2003, BC03 hereafter) stellar population synthesis models, with a Chabrier (2003) initial mass function, to calculate the intrinsic spectral energy distribution (SED) for each star particle within a galaxy, given its metallicity. We assume that each star particle represents a simple stellar population, where all stars are formed at the same redshift and have the same metallicity. The SEDs are then multiplied by the initial mass of each particle to obtain their intrinsic flux.

We use the SUNSET code to measure dust attenuation, as described in Kaviraj et al. (2017). Briefly, we first extract the density and metallicity of the gas cells in the galaxy and convert the gas mass within each cell to a dust mass, assuming a dust-to-metal ratio of 0.4 (e.g. Draine et al., 2007). The column density of dust is used to compute the line-of-sight optical depth for each star particle, and dust-attenuated SEDs are then calculated assuming a dust screen in front of each star particle. As shown in Kaviraj et al. (2017), for optical filters, this produces comparable results to a full radiative transfer approach. The attenuated SEDs are then convolved with the SDSS  $r$  band filter response curve and binned to a spatial resolution of 1 kpc.

Following the convention in the observational literature, we identify LSBGs using their effective surface-brightness,  $\langle\mu\rangle_e$ , defined as the average surface-brightness within the effective radius ( $R_{\text{eff}}$ ). We calculate  $R_{\text{eff}}$  by performing photometry using isophotal ellipses as apertures, with  $R_{\text{eff}}$  defined as the semi-major axis of an isophote containing half of the total galaxy flux. The effective surface-brightness is then calculated using the total flux contained within this ellipse divided by the area of the aperture. We note that the  $r$  band surface-brightness is largely insensitive to the specific dust attenuation recipe, especially for LSBGs, which are largely dust poor.

It is worth noting that the labelling of galaxies as ‘LSB’ systems is strongly determined by the surface-brightness limits of surveys that were available, when the term was coined (e.g. Disney, 1976). Galaxies we define as LSBGs in this study are those that are largely invisible at the depth of *current* wide-area surveys, like the SDSS. Indeed, if contemporary large surveys were deeper (e.g. like the forthcoming LSST survey, which will be 5 magnitudes deeper than the SDSS) then our definition of an LSB galaxy would be very different. Surveys like the SDSS start becoming incomplete around  $\langle\mu\rangle_e < 23$  mag arcsec<sup>-2</sup> (e.g. Kniazev et al., 2004; Bakos & Trujillo, 2012; Williams et al., 2016) in the  $r$  band. The nominal completeness of the survey is  $\sim 70$  per cent at  $\sim 23$  mag arcsec<sup>-2</sup> (e.g. Zhong et al., 2008; Driver et al., 2005), falling rapidly to  $\sim 10$  per cent for galaxies that are fainter than  $\sim 24$  mag arcsec<sup>-2</sup> (e.g. Kniazev et al., 2004). In our analysis below, we split our galaxies into three categories, defined using effective surface-brightness:

1. ‘*High-surface-brightness galaxies*’ (HSBGs): These are defined as galaxies with  $\langle\mu\rangle_e < 23$  mag arcsec<sup>-2</sup> in the  $r$  band. They represent the overwhelming majority of galaxies that are detectable in past

surveys like the SDSS, and which underpin our current understanding of galaxy evolution.

2. ‘Classical low-surface-brightness galaxies’ (Cl. LSBGs): These are defined as galaxies with  $\langle \mu \rangle_e > 23 \text{ mag arcsec}^{-2}$  in the  $r$  band. They represent the brighter end of the LSBG population and are the ‘classical’ LSB galaxy populations that have been studied in the past literature, particularly that which preceded the SDSS.

3. ‘Ultra-diffuse galaxies’ (UDGs): These are defined as galaxies with  $\langle \mu \rangle_e > 24.5 \text{ mag arcsec}^{-2}$  in the  $r$  band (e.g. Laporte et al., 2018). They represent the fainter end of the LSB galaxy population.

We note that there is no standard definition in the literature of what constitutes a UDG, owing to the often specialised nature of the instruments and techniques involved in their detection. However, most definitions are roughly equivalent. For example, van Dokkum et al. (2015) and Román & Trujillo (2017b) both use a  $g$  band central surface-brightness ( $\mu_0$ ) of  $24 \text{ mag arcsec}^{-2}$ , Koda et al. (2015) use an  $R$  band effective surface-brightness of  $24 \text{ mag arcsec}^{-2}$  and van der Burg et al. (2016) use an  $r$  band effective surface-brightness of  $24 \text{ mag arcsec}^{-2}$ . Often, UDGs are also selected using an effective radius threshold of  $R_{\text{eff}} \gtrsim 1.5$  in order to differentiate them from more compact, lower mass objects with equivalent surface-brightnesses (e.g. van Dokkum et al., 2015; Koda et al., 2015; van der Burg et al., 2016; Román & Trujillo, 2017b). While this is an important consideration over the mass ranges that these observational studies examine ( $M_\star < 10^8 M_\odot$ ), the range of masses that we consider in Section 6.3.2 onwards ( $10^{8.5} - 10^{10} M_\odot$ ) precludes such objects.

Note that, in the following sections, we use ‘low surface-brightness galaxy’ (LSBG) to refer to any galaxy in Horizon-AGN with  $\langle \mu \rangle_e > 23 \text{ mag arcsec}^{-2}$  (i.e. any galaxy that falls in either the Cl. LSBG or UDG categories). As we describe below, the threshold  $\langle \mu \rangle_e \sim 24.5 \text{ mag arcsec}^{-2}$  between our two LSBG categories (Cl. LSBGs and UDGs), appears to demarcate two galaxy populations that are reasonably distinct, both in terms of the redshift evolution of their properties and their formation mechanisms. The Cl. LSBGs are much closer to the HSBGs in terms of their formation histories, with the real distinctions emerging between HSBGs and UDGs. The differences between the evolution of HSBGs and UDGs is therefore the principal focus of this study.

Figure 6.1 shows an example of a galaxy from our three populations, with the dashed ellipses indicating the apertures used to calculate the effective surface-brightness. Figure 6.2 shows the effective radii and stellar masses of a random selection of Horizon-AGN galaxies that fall into each of the three categories described above. For comparison, we show observed galaxy populations in the nearby Universe. We

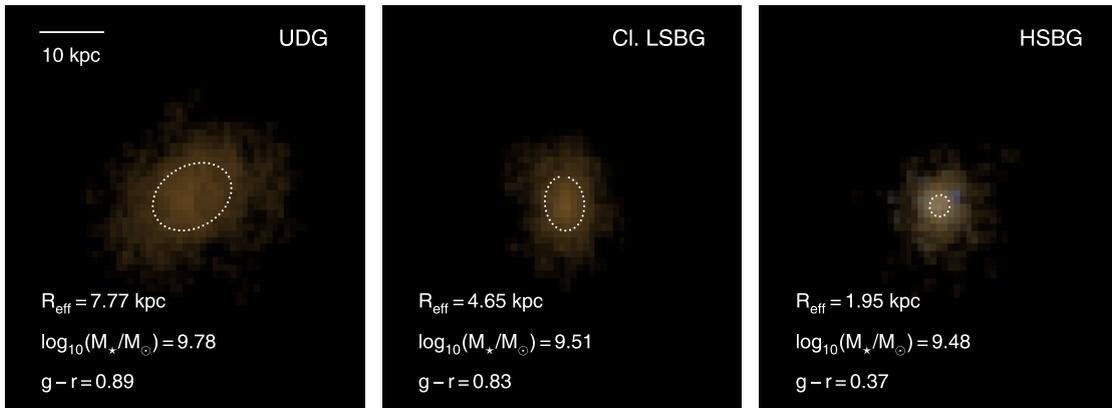


FIGURE 6.1: Example  $g r i$  band false colour images of low-mass Horizon-AGN galaxies. The left, middle and right hand panels show a typical example of galaxies identified as UDGs, Cl. LSBGs and HSBGs respectively. The dotted white ellipses are isophotes which contain half of the galaxy’s  $r$  band flux. A common spatial scale (indicated in the top-left corner of the left-hand panel) is used for all three images.

note that, even for relatively low stellar masses ( $M_* \sim 10^{8.5} M_\odot$ ), the LSBGs in Horizon-AGN are well-resolved enough to recover accurate effective radii. However, depending on the implementation of sub-grid physics (e.g. prescriptions for feedback), effects other than resolution can produce some systematic offset in galaxy sizes (see Appendix C.2 for a full discussion).

Our simulated HSBGs fall along the same locus as observed HSBGs and dwarf ellipticals from Cappellari et al. (2011) and Dabringhausen & Kroupa (2013). Although the mass resolution of Horizon-AGN ( $2 \times 10^8 M_\odot$ ) does not allow us to probe the stellar mass regime where the majority of UDGs have been discovered observationally, many observed UDGs from e.g. van Dokkum et al. (2015), Mihos et al. (2015) and Yagi et al. (2016) that are massive enough do occupy the same region in parameter space as their model counterparts. Furthermore, as we describe in Appendix C.3, while past observational studies are dominated by low-mass LSBGs, this is largely due to the small volumes probed in these works. These small volumes do not preclude the existence of massive LSBGs in new and forthcoming deep-wide surveys. Indeed, some massive LSBGs, such as Malin 1 and UGC 1382, are already known (see Fig 6.2 below), although the small observational volumes probed so far mean that such objects are rare in current (and past) datasets.

### 6.3 The low-surface-brightness Universe at the present-day

We begin by studying the contributions of LSBGs to the number, mass and luminosity densities at low redshift (Section 6.3.1). We then compare key properties of LSBGs (effective radii, local environments, dark matter fractions, stellar ages and star-formation histories) to their HSB counterparts at  $z \sim 0$  (Section 6.3.2).

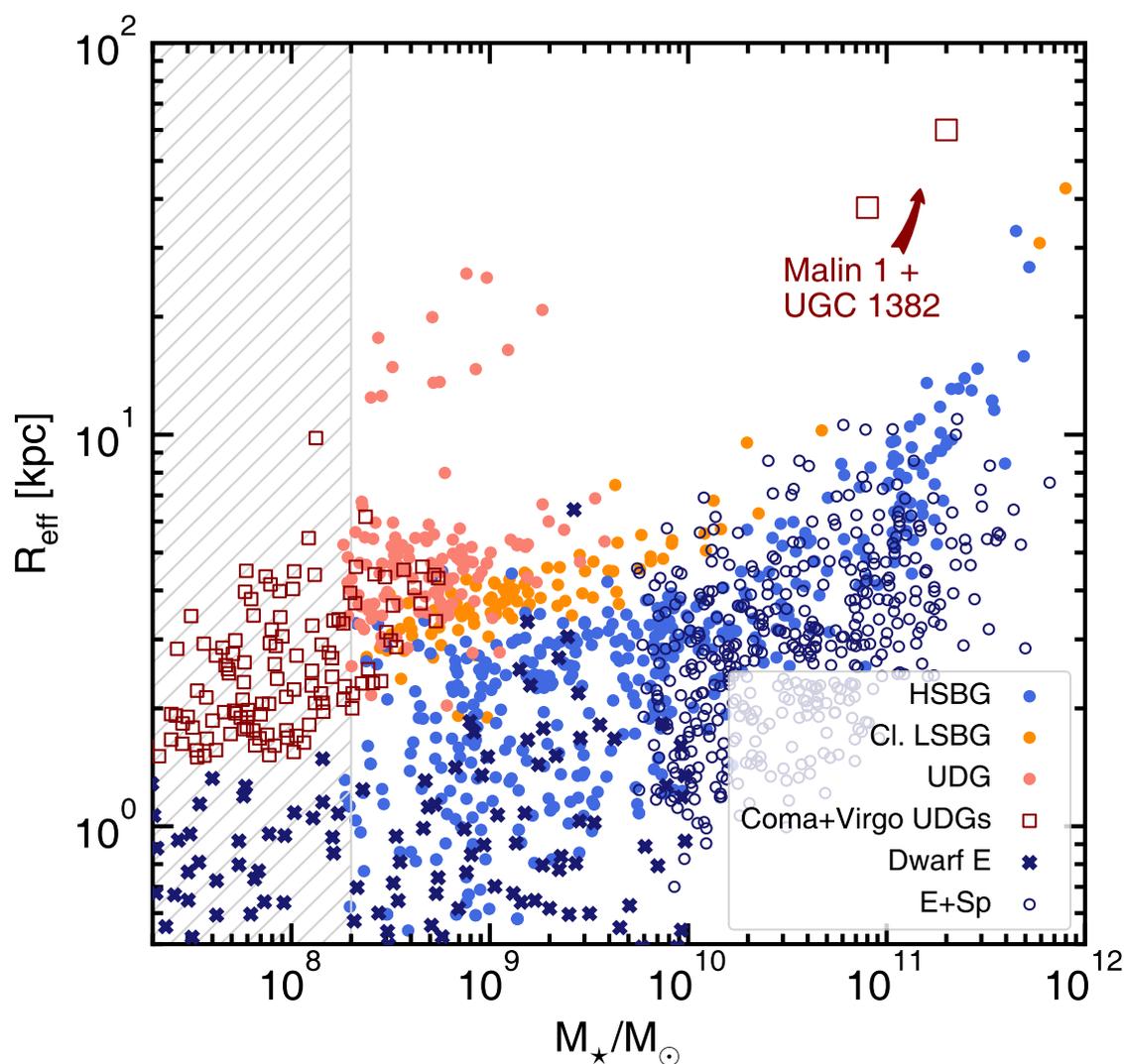


FIGURE 6.2: Effective radius ( $R_{\text{eff}}$ ) vs. stellar mass ( $M_*$ ) for a random selection of galaxies from Horizon-AGN, compared to observed galaxies in the local Universe. Blue, orange and red filled circles show simulated galaxies identified as HSBGs, Cl. LSBGs and UDGs respectively. Open red squares show UDGs from the Coma and Virgo clusters (van Dokkum et al., 2015; Mihos et al., 2015; Yagi et al., 2016; Gu et al., 2018). Dark blue crosses indicate dwarf ellipticals, and open dark blue circles indicate high mass ellipticals and spirals, from Dabringhausen & Kroupa (2013) and Cappellari et al. (2011). Large open red squares show the giant LSBGs Malin 1 and UGC 1382 (Bothun et al., 1987; Hagen et al., 2016). The grey hatched region falls below the mass resolution limit of the simulation.

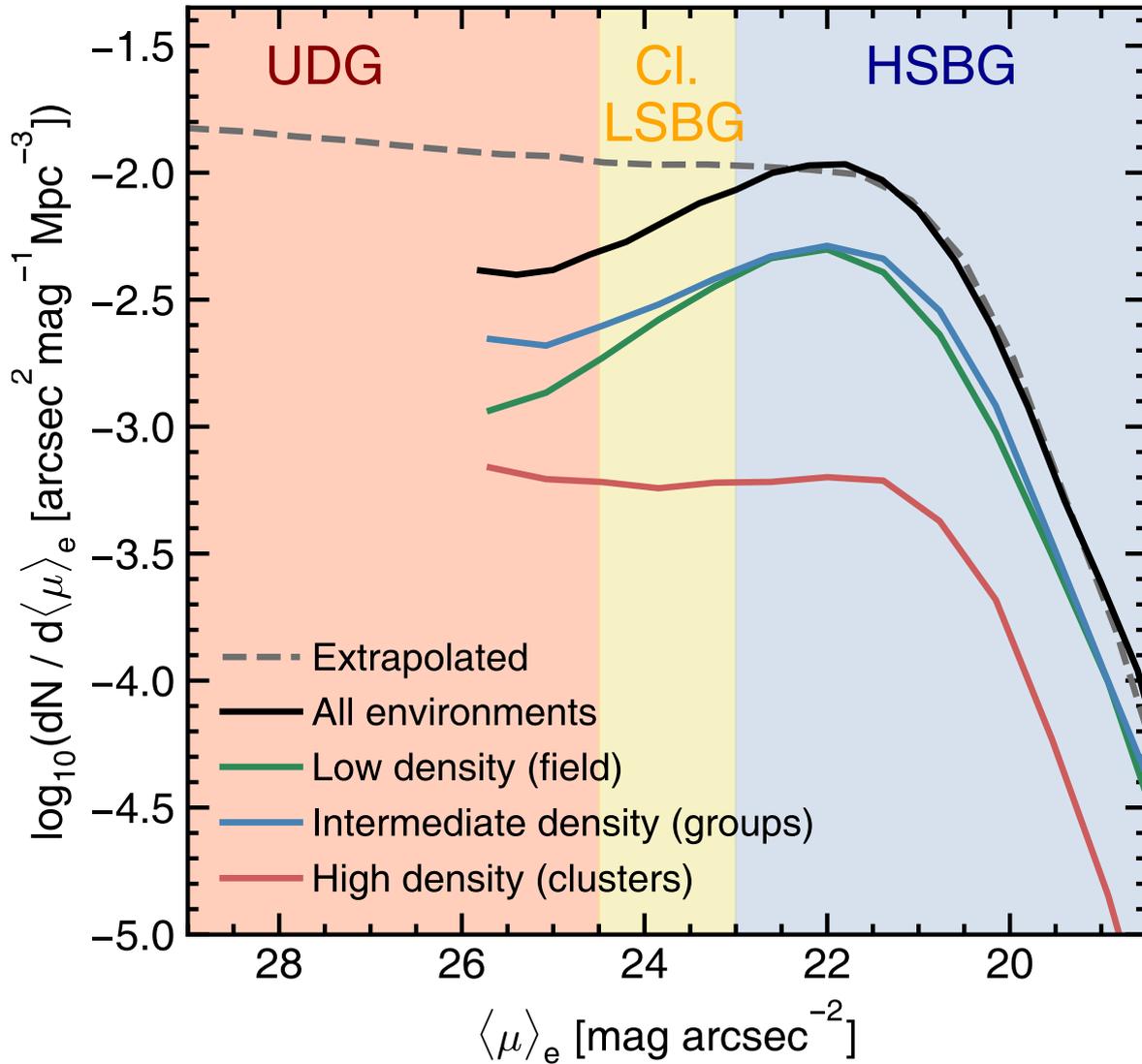


FIGURE 6.3: The surface-brightness function, showing the number density of galaxies as a function of their  $r$  band effective surface-brightness at  $z = 0$ . We show separate curves for low (green), intermediate (blue) and high (red) density environments and all environments (black). Low, intermediate and high density environments roughly correspond to the field, groups and clusters respectively. The dashed line shows the surface-brightness function that is produced by extrapolating the stellar mass function down to  $10^7 M_{\odot}$ , as described in Appendix C.1.

### 6.3.1 Contribution of LSBGs to the local number, stellar mass and luminosity densities

Figure 6.3 shows the surface-brightness function in Horizon-AGN i.e. the number density of galaxies as a function of  $\langle\mu\rangle_e$  in the  $r$  band (solid line). The coloured lines indicate galaxies in different environments. Following Martin et al. (2018a), environment is defined according to the 3-D local number density of objects around each galaxy. Local density is calculated using an adaptive kernel density estimation method<sup>2</sup> as described in Breiman et al. (1977); Ferdosi et al. (2011); Martin et al. (2018a). The

<sup>2</sup>The sharpness of the kernel used for multivariate density estimation is responsive to the local density of the region, such that the error between the density estimate and the true density is minimised.

TABLE 6.1: The frequency (col 2, 3) and number (col 4, 5) of LSBGs of different surface-brightnesses (in  $r$  band mag arcsec $^{-2}$ ) as a function of environment in the present-day universe in the Horizon-AGN simulation, for stellar masses greater than  $2 \times 10^8 M_{\odot}$ . The numbers in brackets indicate the corresponding fractions produced by extrapolating the stellar mass function down to  $10^7 M_{\odot}$ . The ‘low’ (local number density in the 0th – 40th density percentile), ‘intermediate’ (40th – 90th density percentile) and ‘high’ (90th – 100th density percentile) density bins correspond roughly to field, group and cluster environments respectively (see Martin et al., 2018a).

	$f(24.5 > \langle \mu \rangle_e > 23)$	$f(\langle \mu \rangle_e > 24.5)$	$N(24.5 > \langle \mu \rangle_e > 23)$	$N(\langle \mu \rangle_e > 24.5)$
Low density (Field)	0.23 (0.09)	0.18 (0.77)	10760	5634
Intermediate density (Groups)	0.21 (0.09)	0.27 (0.74)	12691	12119
High density (Clusters)	0.19 (0.07)	0.46 (0.83)	2310	4572

TABLE 6.2: The fraction of the local stellar mass, luminosity and number density budget contributed by galaxies of different  $r$  band surface-brightnesses (in units of mag arcsec $^{-2}$ ) in the Horizon-AGN simulation, for stellar masses greater than  $2 \times 10^8 M_{\odot}$ . The numbers in brackets indicate the corresponding fractions produced by extrapolating the stellar mass function down to  $10^7 M_{\odot}$ .

	$\langle \mu \rangle_e < 23$	$24.5 > \langle \mu \rangle_e > 23$	$\langle \mu \rangle_e > 24.5$
$f_{M_{\star}}$	0.924 (0.902)	0.059 (0.067)	0.014 (0.030)
$f_L$	0.939 (0.892)	0.049 (0.071)	0.012 (0.037)
$f_N$	0.534 (0.145)	0.214 (0.093)	0.252 (0.762)

density estimate takes into account all galaxies above  $2 \times 10^8 M_{\odot}$ .

Galaxies are then split into three bins in local density: ‘low density’ corresponds to galaxies in the 0th – 40th density percentiles, ‘intermediate density’ correspond to the 40th – 90th percentiles and ‘high density’ corresponds to galaxies in the 90th – 100th percentiles. The low, intermediate and high density bins roughly correspond to the field, groups and clusters (see Martin et al. (2018a) for more details). Typically, galaxies in the intermediate and high density bins are found in halos with masses of  $10^{12.5} < M_{halo} < 10^{13.5} M_{\odot}$  and  $M_{halo} > 13.5 M_{\odot}$  respectively. In the low density bin, most galaxies ( $\sim 70$  per cent) are isolated (i.e. they are not a sub-halo of a larger halo). Of the galaxies in the low-density bin that are satellites, typical halo masses are  $\sim 10^{12} M_{\odot}$ . We note that there is no perfect correspondence between number density and halo mass - for example, at fixed density, UDGs are typically hosted by haloes that are  $\sim 0.5$  dex more massive than HSBGs.

Since we do not consider objects with stellar masses below  $2 \times 10^8 M_{\odot}$ , the predicted surface-brightness function starts becoming incomplete as we approach this limit. In order to account for this when estimating the LSBG contribution to the local number, mass and luminosity densities, we extrapolate the galaxy stellar-mass function down to  $10^7 M_{\odot}$  (as described in Appendix C.1). The dashed black line indicates the corresponding extrapolated surface-brightness function, using a combination of surface-brightnesses drawn from the extrapolated fits (between  $10^7 M_{\odot}$  and  $10^9 M_{\odot}$ ) and the raw simulation data ( $10^8 M_{\odot}$  to  $10^{12} M_{\odot}$ ) - see Appendix C.1 for more details.

Table 6.1 summarises the absolute numbers and number fractions of HSBGs and LSBGs in the present-day Universe, as a function of local environment. The numbers in brackets indicate the corresponding

values using the extrapolated mass function. For galaxies with stellar masses above the resolution limit of the simulation ( $2 \times 10^8 M_\odot$ ), LSBGs account for a significant fraction (over half) of the galaxy population in clusters and a significant minority (40-50 per cent) of objects in low-density environments (groups and the field).

However, for stellar masses down to  $10^7 M_\odot$ , LSBGs are expected to overwhelmingly dominate the number density of the Universe, accounting for more than 70 per cent of galaxies, irrespective of the local environment being considered. It is worth noting that the *absolute* numbers of LSBGs across different environments (see col 4, 5 in Table 6.1) are similar. For example, the absolute numbers of UDGs in the Horizon-AGN volume that inhabit the field and those that inhabit clusters are predicted to be almost the same (col 5 in Table 6.1). This is because, although the LSBG fraction is higher in clusters, the total number of galaxies that inhabit low-density environments (e.g. the field) is much larger.

Table 6.2 summarises the contribution of HSBGs and LSBGs to the mass, luminosity and number density budgets of the local Universe. For galaxies with stellar masses greater than  $2 \times 10^8 M_\odot$ , LSBGs contribute around 47 per cent of the total number density and make a small but non-negligible contribution to the stellar mass (7.5 per cent) and luminosity (6 per cent) budgets. These numbers change to 85 (number density), 10 (mass density) and 11 (luminosity density) per cent respectively, when we extrapolate down to a stellar mass of  $10^7 M_\odot$ . Although they account for the majority of the number density budget (76 per cent with extrapolation to  $10^7 M_\odot$ ) at low redshift, the extreme end of the LSBG population, i.e. UDGs ( $\langle \mu \rangle_e > 24.5$ ), account for only a small fraction of the mass or luminosity budget (less than 4 per cent in both cases).

We note that the extrapolated quantities above are used only to estimate the overall contribution of LSBGs to the number, stellar mass and luminosity density down to a stellar mass of  $10^7 M_\odot$ . For the rest of the analysis that follows, we use galaxies that are actually resolved in the simulation and for which the minimum stellar mass is  $2 \times 10^8 M_\odot$ .

### 6.3.2 Properties of LSB galaxies at the present day

In this section we compare the properties of LSBGs to their HSB counterparts at the present day. Figure 6.4 shows the spatial distribution of a random selection of UDGs, Cl. LSBGs and HSBGs within the cosmic web. The contours indicate the surface density of galaxies calculated using all objects in the simulation. Although they appear to exist preferentially in regions of high number density, many UDGs occur in regions of much lower density. On the other hand HSBGs appear to be essentially uniformly distributed.

In Figure 6.5, we show contour plots of the distribution of galaxies as a function of  $r$  band effective surface-brightness,  $\langle \mu \rangle_e$ , and stellar mass at  $z = 0$ , split by local environment. The histogram for all galaxies across all environments is bimodal. However, the bimodality varies strongly with environment.

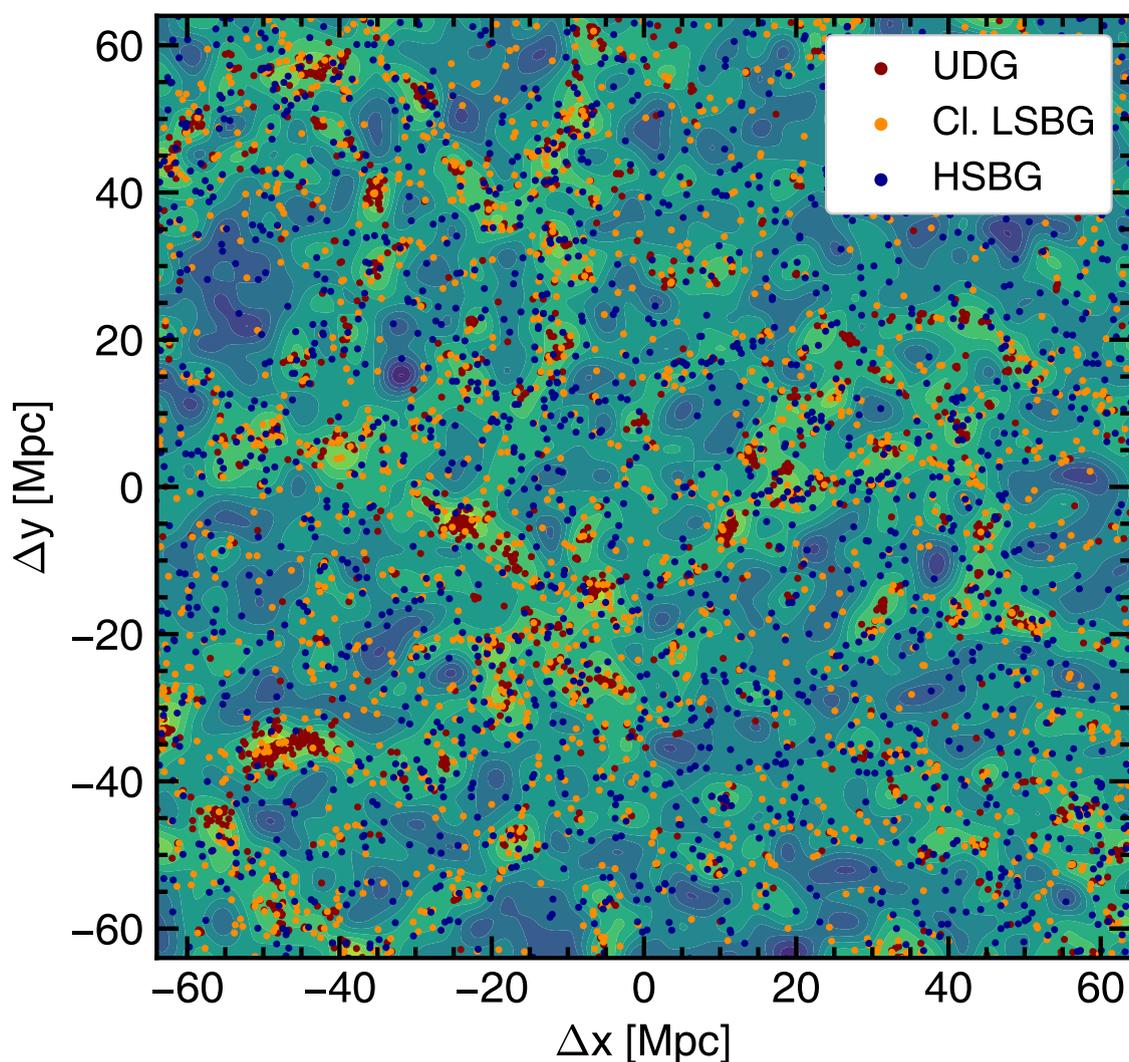


FIGURE 6.4: The spatial distribution, within the cosmic web, of the UDG, Cl. LSBG and HSBG populations. Red, orange and blue coloured points show the positions of individual UDGs, Cl. LSBGs and HSBGs. Contours indicate the surface density, calculated using all objects in the simulation, with lighter colours indicating higher densities.

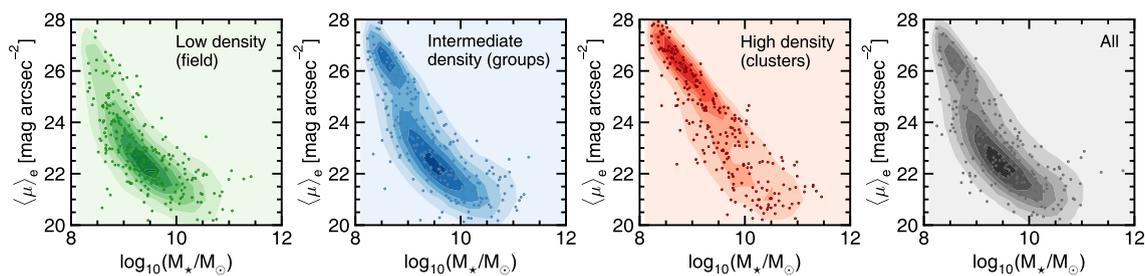


FIGURE 6.5: Contour plots showing the number density of galaxies as a function of effective surface-brightness ( $\langle \mu \rangle_e$ ) and stellar mass ( $M_*$ ) at  $z = 0.06$ , split by local environment. Low, intermediate and high density environments correspond roughly to field, group and cluster environments respectively. A random sample of galaxies is plotted using points in each panel. The right-hand panel shows the same for all galaxies in the simulation.

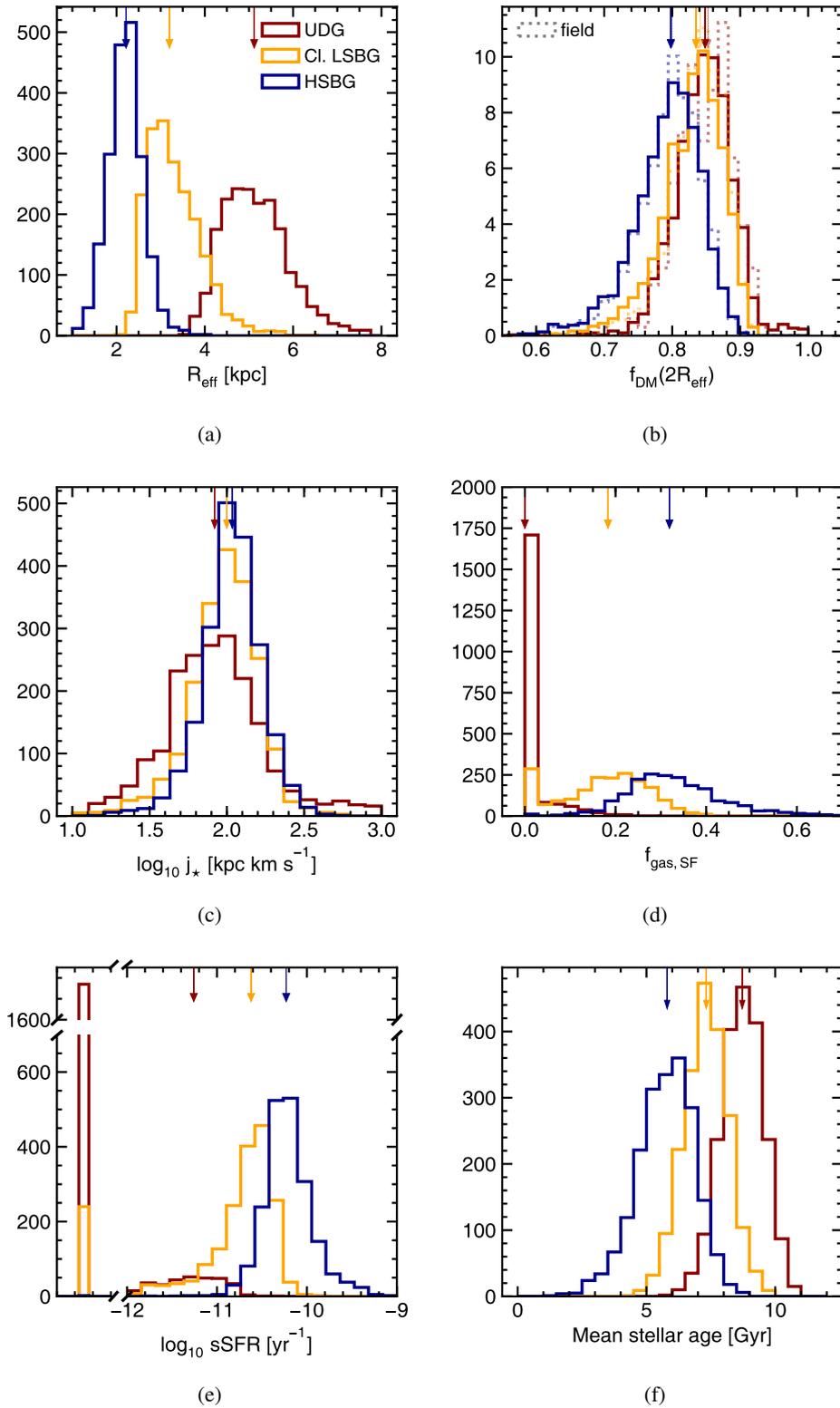


FIGURE 6.6: Properties of present-day UDGs and Cl. LSBGs compared to those of HSBGs at  $z \sim 0$  (red, orange and blue histograms respectively). Coloured arrows indicate the median values of each population. Fainter dashed arrows indicate the median values for field populations only. Panels are as follows: **(a)** effective radius measured using the stellar distribution of each galaxy **(b)** the dark-matter fraction ( $M_{\text{DM}}/(M_{\text{DM}} + M_*)$ ) measured within the central  $2 R_{\text{eff}}$ , for all galaxies (solid line) and galaxies in the field (dotted lines) - note that the histograms are normalised in order to easily compare the two populations **(c)** stellar specific angular momentum **(d)** star-forming ( $\rho_{\text{gas, cell}} > 0.1 \text{H cm}^{-3}$ ) gas fraction measured within  $2 R_{\text{eff}}$  ( $M_{\text{gas,SF}}/(M_* + M_{\text{gas,SF}})$ ) **(e)** specific star formation rate - the bar to the left indicates galaxies with sSFRs of 0 **(f)** mass-weighted mean stellar age ( $(\sum_i \text{age}_i m_{*,i}) / \sum_i m_{*,i}$ ).

At a given stellar mass, the frequency of LSBGs is higher in denser environments. While in the field most galaxies inhabit the HSB peak, the LSB peak progressively dominates as we move to higher density environments. Indeed, for low-mass galaxies, in clusters, the LSB peak overwhelmingly dominates the population (this is partly the reason why much of the UDG literature has been focussed on clusters to date).

Since the frequency of LSBGs is a strong function of stellar mass (see e.g. Figure 6.5), we first construct mass-matched samples of 2000 HSBGs, LSBGs and UDGs with stellar masses between  $10^9 M_\odot$  and  $10^{10} M_\odot$ , each of which have the same distribution in stellar mass. Due to the shape of the UDG mass function (see Appendix C.3), the stellar mass distribution of our sample peaks close to  $10^9 M_\odot$  and declines such that  $\sim 95$  per cent of galaxies are less massive than  $10^{9.5} M_\odot$ . We then use these mass-matched samples to explore key properties of LSB systems – effective radii, dark-matter fractions, specific angular momenta, gas densities, specific star formation rates and mean stellar ages. Note that the analysis presented in all subsequent sections, which explore how LSBG progenitors evolve with time, is also based on these mass-matched samples.

Figure 6.6 shows histograms of these properties. LSBGs have larger effective radii (panel (a)), with the mean effective radii of UDGs around 2.5 times larger than HSBGs. The dark-matter (DM) fractions in LSBGs and HSBGs (panel (b)) are similar, with the median value for LSBGs predicted to be slightly ( $\sim 5$  per cent) higher than in HSBGs. The overwhelming majority of LSBGs are, therefore, not devoid of DM, nor do they have anomalously large DM fractions for their stellar mass. Contamination due to galaxies being embedded in more massive DM haloes does not appear to have a significant impact on the ratios shown - when we restrict our sample to field galaxies only (dotted histograms), there is no difference in the median DM to stellar mass ratio. This suggests that high-DM-fraction UDGs (i.e. failed  $L^*$  galaxies) (e.g. van Dokkum et al., 2016; Beasley et al., 2016) are extremely uncommon, at least in the stellar mass range we study here ( $10^9 M_\odot < M_\star < 10^{10} M_\odot$ ).

It is worth noting here that, while recent observations have suggested that at least some UDGs may have very low dark matter fractions (e.g. van Dokkum et al. 2018, but see Laporte et al. 2018; Trujillo et al. 2018), a small fraction of low mass DM-free galaxies can form naturally within the LCDM paradigm as tidal dwarf galaxies in galaxy mergers (e.g. Barnes, 1992; Okazaki & Taniguchi, 2000; Bournaud & Duc, 2006; Kaviraj et al., 2012a). However, mergers typically produce tidal dwarfs with very low stellar masses (Kaviraj et al., 2012a), and the mass range that we consider ( $M_\star > 10^9 M_\odot$ ) precludes significant numbers of these objects in our sample. It may not be surprising, therefore, that we do not find any evidence of UDGs with anomalously low DM fractions in Horizon-AGN, even if this were a significant channel for their production.

The distribution of the stellar specific angular momenta (panel (c)) of LSBGs and HSBGs is similar, indicating that the formation of LSBGs, and UDGs in particular, is not primarily due to them being the high spin tail of the angular momentum distribution (e.g. Yozin & Bekki, 2015; Amorisco & Loeb, 2016; Amorisco et al., 2016; Rong et al., 2017). The LSBGs in this study typically have spins that are

not significantly different from, or indeed, are slightly below, those seen in HSBGs (see also Di Cintio et al., 2017; Chan et al., 2018).

Finally, we consider quantities that trace the star formation properties of galaxies. Panel (d) shows the ‘star-forming’ gas fraction, defined as the ratio of the gas mass that is dense enough to form stars ( $\rho_{\text{gas cell}} > 0.1 \text{H cm}^{-3}$ ) to the stellar mass ( $M_{\text{gas,SF}}/(M_{\text{gas,SF}} + M_{\star})$ ), measured within the central  $2 R_{\text{eff}}$ <sup>3</sup>. Gas fractions in LSBGs are lower than those in their HSB counterparts. For example, the gas fractions of UDGs mostly lie around zero, with 4 out of 5 UDGs being completely devoid of star forming gas in their central  $2 R_{\text{eff}}$ . HSBGs, on the other hand, still retain fairly significant fractions of star-forming gas ( $f_{\text{gas,SF}} \sim 0.3$ ). UDGs that do contain some star-forming gas at the present day have median values that are around one sixth of this value. The lower gas fractions are reflected in lower specific star formation rates (sSFRs; panel (e)) and higher mass-weighted mean stellar ages ( $\sum_i \text{age}_{i,m_{\star,i}}/\sum_i m_{\star,i}$ ; panel (f)) in LSB systems. For example, the sSFRs in UDGs are an order of magnitude lower than in HSBGs, when galaxies with zero sSFR (again, around 4 out of 5 UDGs) are neglected. The median age of UDG stellar populations is 9 Gyrs, 50 per cent older than their HSB counterparts. The large age differences between LSBGs and HSBGs indicates that the LSB nature of these systems must be partly driven by gas exhaustion at *early* epochs and consequently a more quiescent recent star history.

We note here that the production of UDGs may be too efficient in clusters leading to quenched HSB galaxies being relatively unrepresented. Additionally, since the quenched fraction (especially at low redshift) is somewhat inconsistent with observations, and produces an offset in the star formation main sequence between the observed and theoretical populations in low-mass galaxies (e.g. see Kaviraj et al., 2017), this may lead to relatively diffuse HSB or LSB galaxies becoming UDGs due to fading stellar populations.

## 6.4 Redshift evolution of LSBG progenitors

We proceed by comparing the redshift evolution of LSBG progenitors to the progenitors of their HSB counterparts. We focus, in particular, on the evolution of the effective radii and gas fractions which, as we showed in Section 6.3, are the quantities in which HSBGs and LSBGs diverge the most at the present day. We note that, since we restrict our study to resolved progenitors, There is some incompleteness in the sample at higher redshifts. This is due to the limit of 50 particles that we impose on the structure finder (see Section 6.2.2), which renders their merger trees incomplete after galaxies fall below this level. The merger trees of the LSBG and UDG samples are largely complete after  $z = 2$  (80 and 90 per cent of main progenitors at  $z = 2$  are accounted for respectively) owing to their rapid assembly histories (see Section 6.5.1 below). For the HSBG sample, around 60 per cent of main progenitors are accounted for at  $z = 2$  (rising to 100 per cent by  $z = 1$ ), which may lead to the exclusion of more slowly evolving HSBGs before  $z = 1$ .

<sup>3</sup>We note that calculating the gas fraction within a fixed radius does not alter our conclusions.

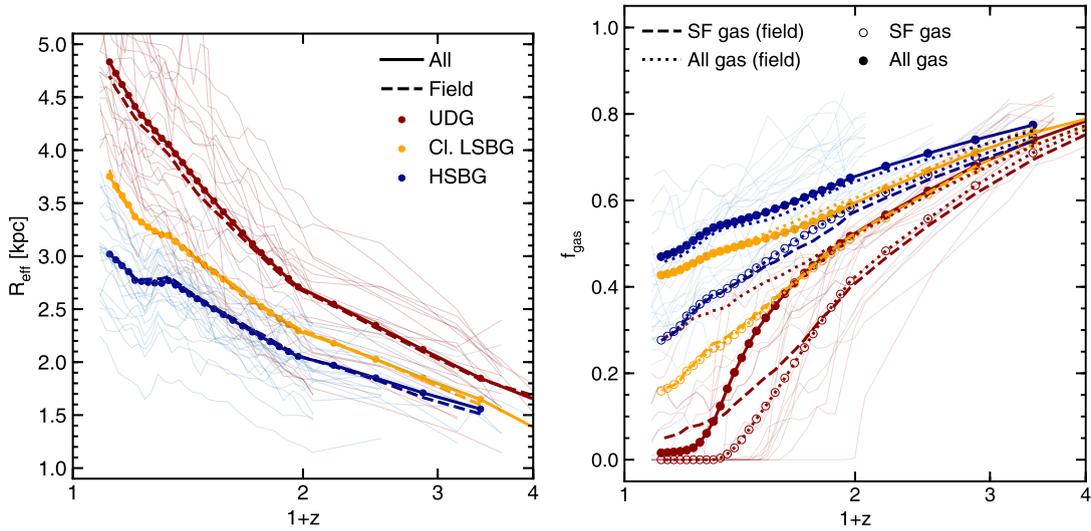


FIGURE 6.7: **Top:** The redshift evolution of effective radii. Solid blue, orange and red coloured points show the redshift evolution of the median effective radii of the HSBG, Cl. LSBG and UDG populations respectively. The dashed lines shows the evolution of  $R_{\text{eff}}$  for galaxies in the field only. Note that rather than attempt to emulate observational methods to calculate  $R_{\text{eff}}$  at all redshifts, we instead use the average projected half mass radius in the  $xy$ ,  $xz$  and  $yz$  planes here **Bottom:** The redshift evolution of the median gas fraction, defined as  $M_{\text{gas}}/(M_{\text{gas}} + M_{\star})$ , for total gas (solid coloured points) and star-forming gas (open coloured points) for the HSBG, LSBG and UDG populations. Dashed and dotted lines without points show the evolution of  $f_{\text{gas}}$  for total gas and star-forming gas respectively for field galaxies only. Pale red and blue lines show tracks for the effective radii and star-forming gas fractions of a random sample of individual UDGs and HSBGs.

### 6.4.1 Gas fractions and effective radii

The top panel of Figure 6.7 describes how the effective radii of the main progenitors of LSBGs and HSBGs evolve as a function of redshift. LSBGs, and UDGs in particular, are consistently larger, on average, than their HSB counterparts. Furthermore, after  $z \sim 1$ , the rate of increase in the effective radii of UDGs is higher compared to that in HSBGs. Figure 6.7 shows that the evolution of the effective radii of all galaxy populations is not abrupt but relatively steady and smooth with time, both galaxy by galaxy (pale lines) and as a population. *It is unlikely, therefore, that the large radii of LSBGs today are the result of single, violent events at early epochs.*

The dashed lines indicate the evolution of galaxy populations in field environments only. As the dashed red line indicates, the evolution of the effective radii of field UDGs proceeds almost identically to the general UDG population, despite the frequency of UDGs being higher in very dense (cluster) environments. This implies that the process(es) that produce the large sizes seen in today's UDGs are the same regardless of environment (although they may occur less frequently in the field). In particular, the principal mechanism for UDG production is not cluster-specific i.e. galaxies do not have to inhabit cluster environments to be the progenitors of UDGs at the present day.

The bottom panel of Figure 6.7 describes how the gas fractions of the main progenitors of LSBGs and HSBGs evolve as a function of redshift. While the gas fractions are similar for progenitors of all galaxies at high redshift, they begin to diverge rapidly at  $z \sim 2$ . The *total* gas fractions in HSBGs and Cl. LSBGs evolve similarly to each other and both HSBGs and Cl. LSBGs retain relatively high total gas fractions at  $z = 0$ . In these populations the reduction in the average gas fraction is primarily due to gas being converted into stars, rather than as a result of gas being expelled from the galaxy. As we will also show in Section 6.5, most of this gas in HSBGs that is turned into stars is not replenished, at least after  $z = 1$ , so that the decreasing gas fractions are due the gas masses steadily decreasing rather than the stellar masses simply increasing in these galaxies. There is a more pronounced divergence in terms of the fraction of *star-forming* gas. By  $z = 0$ , Cl. LSBGs have significantly lower fractions of star-forming gas compared to their HSB counterparts.

While Cl. LSBGs and HSBGs retain relatively significant reservoirs of gas as they evolve, the same is not true of UDGs. By  $z = 0.5$ , the majority of UDGs have lost almost all of their star-forming gas, essentially terminating star formation, and by  $z = 0.25$ , the majority of UDGs have been almost completely stripped of all of their gas. In around half of the cases, the gas fractions of the main progenitors of UDGs do not evolve linearly with time. Instead they undergo a phase of rapid gas loss lasting a few Gyrs around  $z \sim 0.5$ , which significantly reduces their gas content towards the present day.

The evolution of UDGs in field environments (dotted red lines) is slightly different from that of the global UDG population. There is no phase of rapid gas stripping and both the total and star-forming gas fractions in field UDGs evolve with a similar pattern to their HSB counterparts, albeit much more rapidly. Ultimately, the rate of gas heating is intense enough that the *star-forming* gas fraction is still reduced to similar levels to the wider UDG population ( $< 5$  per cent by  $z = 0$ ) by the present day. Note that the loss of star-forming gas is not due to gas being physically removed (i.e. gas stripping), since field UDGs retain fairly high *total* gas fractions ( $\sim 30$  per cent on average, as shown by the dotted red line).

The complete removal of gas is, therefore, not a necessary criterion for the production of UDGs. Gas heating alone produces the low star-forming gas fractions in these objects (regardless of local environment), without requiring that the gas be removed from the galaxy entirely. Whether UDGs have had their gas entirely removed or have just undergone heating makes little difference to their stellar populations at  $z = 0$ . The median stellar ages of UDGs that have been completely stripped of gas, and those in field environments that have only undergone heating, are 8.7 Gyrs and 8.5 Gyrs respectively. In Section 6.5, we explore the processes that lead to the removal or heating of gas in the LSBG population.

Note that although some galaxies ( $\sim 30$  per cent) in the low-density ‘field’ environment are actually satellites of another galaxy, the average properties of field UDGs (or LSBGs and HSBGs) do not change significantly if we select genuinely isolated galaxies only (i.e. those that are not satellites). Isolated UDGs have typical effective radii that are only slightly larger than field UDGs generally (5.15 kpc) and have slightly higher gas fractions (0.11).

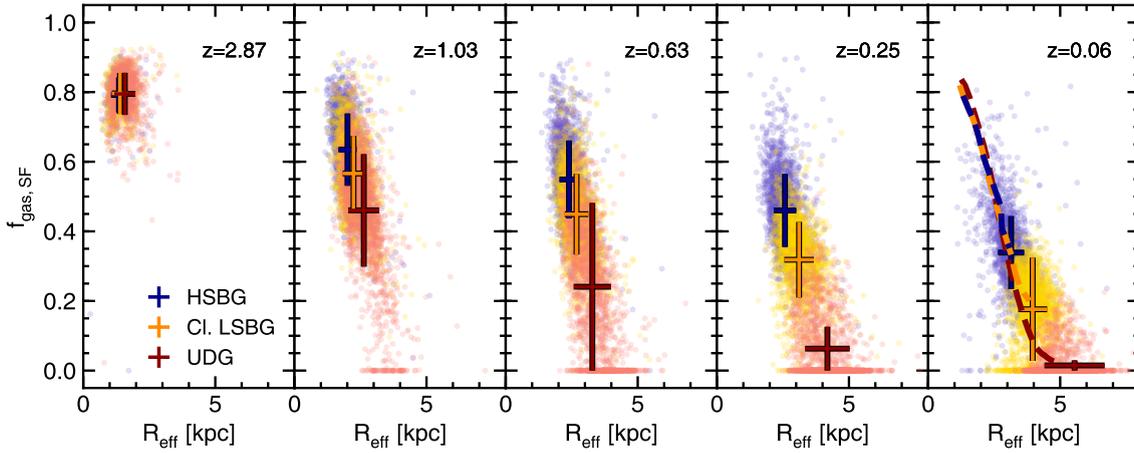


FIGURE 6.8: Redshift evolution of the star-forming gas fraction, defined as  $M_{\text{gas,SF}}/(M_{\text{gas,SF}} + M_{\star})$ , and effective radii of the progenitors of the HSBG (blue), Cl. LSBG (orange) and UDG (red) populations. The time between each redshift snapshot is  $\sim 2.5$  Gyr. Coloured points indicate the position of individual galaxies in the  $f_{\text{gas}}-R_{\text{eff}}$  plane. The error bars in each panel show the median values and  $1\sigma$  dispersions for the distributions of the HSBG, Cl. LSBG and UDG populations at each redshift. The dashed lines in the right-hand panel indicate the average locii followed by the main progenitors of HSBGs, Cl. LSBGs and UDGs in the  $f_{\text{gas}}-R_{\text{eff}}$  plane over cosmic time.

In Figure 6.8, we show the redshift evolution of LSBGs and HSBGs in the star-forming gas fraction vs effective radius plane. As shown in the left-hand panel, the main progenitors of the different populations are very similar at high redshift ( $z \sim 3$ ). Although they differ somewhat in terms of their other properties (e.g. stellar mass and environment), the progenitors of today’s LSBGs and HSBGs share essentially identical effective radii and gas fractions in the early Universe. *This indicates that LSBGs emerged from a common population of progenitors as HSBGs.* The three populations only begin to diverge significantly around  $z \sim 2$  (see Figure 6.7) and then separate rapidly at intermediate redshifts ( $z < 1$ ). UDGs, in particular, diverge quickly from their HSB counterparts, both in terms of rapidly increasing their effective radii and losing significant fractions of their gas reservoirs at these redshifts.

We note that LSBGs appear to be part of a smooth distribution of properties across the general galaxy population. The dashed blue, orange and red lines in the right-hand panel show the average evolutionary tracks followed by HSBGs, Cl. LSBGs and UDGs respectively over cosmic time. LSBGs do not take a different route through the  $f_{\text{gas}}-R_{\text{eff}}$  plane. Instead, they follow very similar locii, although their evolution (particularly for the UDG population) is more rapid. Together with the fact that their high-redshift progenitors share very similar properties with the progenitors of HSBGs, this suggests that *LSBGs are not a special class of object in terms of the populations from which they originate.*

## 6.4.2 Density profiles

Our mass-matched population of LSBGs exhibit somewhat larger effective radii compared to their HSB counterparts, even at high redshift. This can either be a result of processes that directly influence the

distribution of the stellar component of the galaxy, or a result of processes that influence the distribution of the gas from which these stars form. Establishing which of these is the case is important for understanding what triggers the formation of LSBGs at early epochs.

In this section, we consider how the slope of the median gas and stellar density profiles of the different galaxy populations evolve over time. The slope of the stellar density profile determines the measured effective radius of the galaxy, with shallower slopes typically resulting in larger effective radii at a given stellar mass. Shallower density slopes (and therefore shallower gravitational potentials) also reduce the energy required to displace material in the system. In the case of the gas content, the shape of the potential defines the distribution of stellar mass that forms from this gas. Galaxies with shallower slopes are more vulnerable to the effects of encounters with other galaxies or interactions between the galaxy and the intergalactic medium (tidal heating, harassment, gas stripping etc.), which may be important factors in their subsequent evolution.

We calculate the mass-weighted log-log slope of each galaxy's gas and stellar outer density profile between  $0.5R_{\text{eff}}$  and  $3R_{\text{eff}}$ . We calculate the density profile using radial bins of 30 particles. The log-log density slope is parametrised by  $\gamma'$  (Dutton & Treu, 2014):

$$\gamma' = \frac{1}{M(3R_{\text{eff}}) - M(0.5R_{\text{eff}})} \int_{0.5R_{\text{eff}}}^{3R_{\text{eff}}} \gamma(r) 4\pi r^2 \rho(r) dr, \quad (6.1)$$

where  $\gamma = -d \log(\rho) / d \log(r)$  is the local log-log slope of the density profile,  $M(R)$  is the mass enclosed within a radius  $R$ , and  $\rho(r)$  is the local density at radius  $r$ . Lower values of  $\gamma$  indicate shallower density slopes. The density slopes that we recover are consistent with previous studies using the Horizon-AGN simulation (Peirani et al., 2017).

The main panel of Figure 6.9 shows the redshift evolution of the median stellar density slopes for HSBGs, Cl. LSBGs and UDGs. The inset shows the evolution of the median gas density slopes for the same populations between  $z = 3$  and  $z = 1$ . This is an epoch at which galaxies are forming significant fractions of their present-day stellar mass. This is particularly true of UDGs which, as we show in Section 6.5.1 below, form the bulk ( $\sim 75$  per cent) of their stellar mass by  $z = 1$ . At these early epochs, therefore, the gas distribution is actively driving the creation of the stellar distribution. In calculating the median gas density slope, we exclude any galaxies with star-forming gas fractions ( $M_{\text{gas,SF}} / (M_{\text{gas,SF}} + M_{\star})$ ) smaller than 0.05, so as to remove galaxies where the gas is no longer influencing the stellar distribution (since the star-forming gas mass is negligible and star-formation has effectively ceased).

At high redshift, the median value of  $\gamma'(\text{gas})$  is lower (i.e. the gas density slopes are shallower) in UDGs compared to both Cl. LSBGs and HSBGs ( $\sim 1.56$  at  $z = 2$  compared with  $\sim 1.8$  for Cl. LSBGs and HSBGs). Between  $z = 3$  and  $z = 1$ , the gas density slopes in the UDGs remain at a level significantly below the Cl. LSBG and HSBG populations, while their stellar density slopes decline faster than those of the Cl. LSBG and HSBG populations. Thus, at the epochs where UDGs are actively forming the bulk

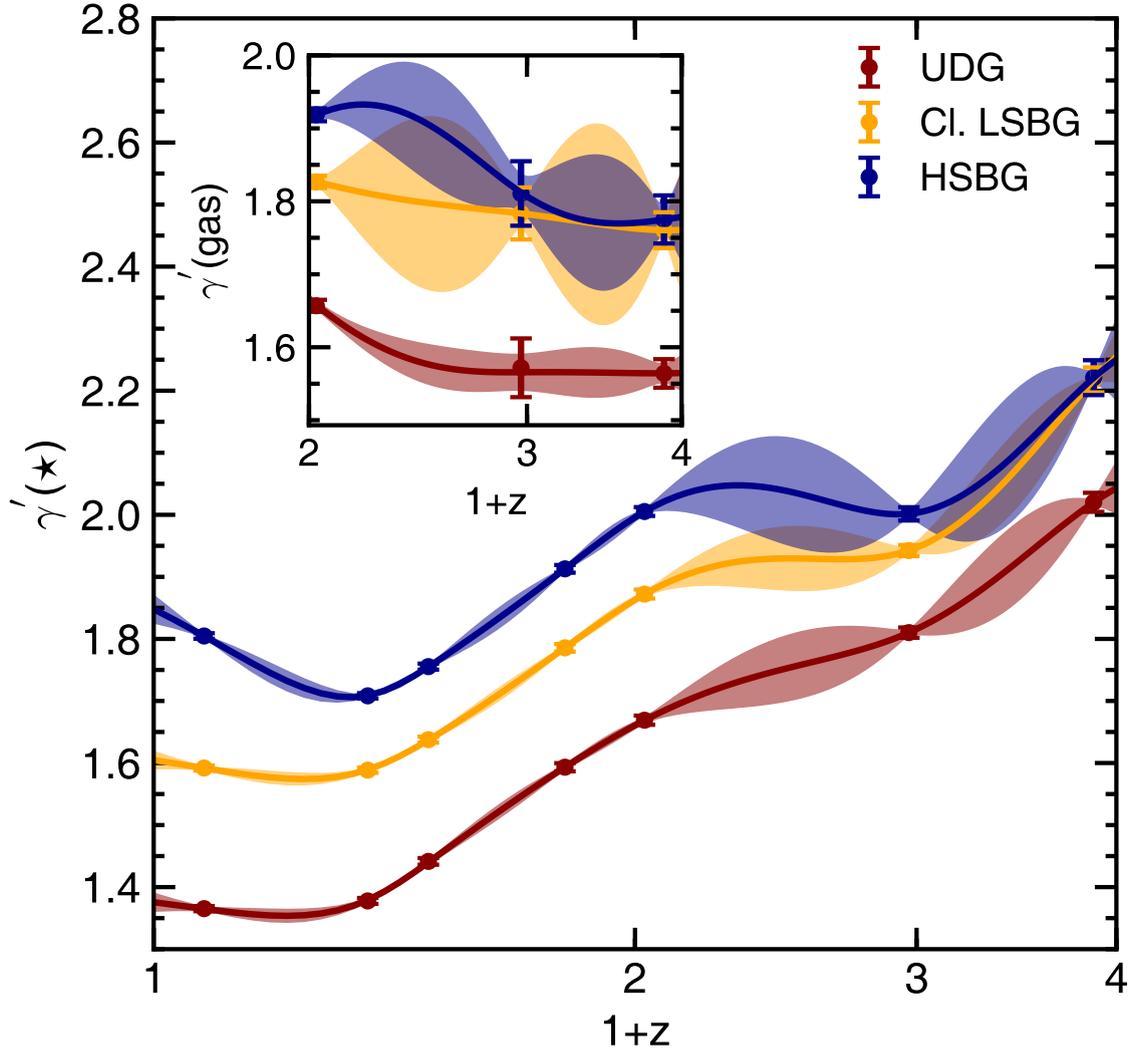


FIGURE 6.9: The evolution of the median log-log stellar density slope,  $\gamma'(\star)$ , calculated within  $0.5 < R_{\text{eff}} < 3$  for UDGs, Cl. LSBGs and HSBGs as a function of redshift. Error bars indicate the error on the median value of  $\gamma'$  at each redshift and solid filled regions show the  $1\sigma$  confidence interval for a Gaussian process regression to these points. **Inset:** the corresponding plots for the log-log star-forming gas density slope,  $\gamma'(\text{gas})$ . Note that, in the case of the gas density slope, galaxies with very low star-forming gas fractions, i.e. those less than 0.05, are excluded when we calculate the median values of  $\gamma'(\text{gas})$ .

of their stellar mass, their gas density profiles are significantly flatter than that of the HSBGs (and also the Cl. LSBGs).

After  $z = 1$ , the stellar density slopes decline rapidly, even though most LSBGs have assembled the majority of their stellar mass by this time. By  $z = 0.06$  the median value of  $\gamma'(\star)$  for UDGs has fallen by  $\sim 0.32$ , from 1.67 at  $z = 1$  to 1.35. The median value of  $\gamma'(\star)$  for HSBGs (most of which have not yet assembled the majority of their stellar mass at  $z = 1$ ), falls from 2.0 to 1.8 between  $z = 1$  and  $z = 0.06$ .

Figure 6.10 shows the distributions of the gas and stellar density slopes at two epochs:  $z = 0.06$  and  $z = 1.03$  (where the divergence in the effective radii, gas fractions and stellar density slopes between the

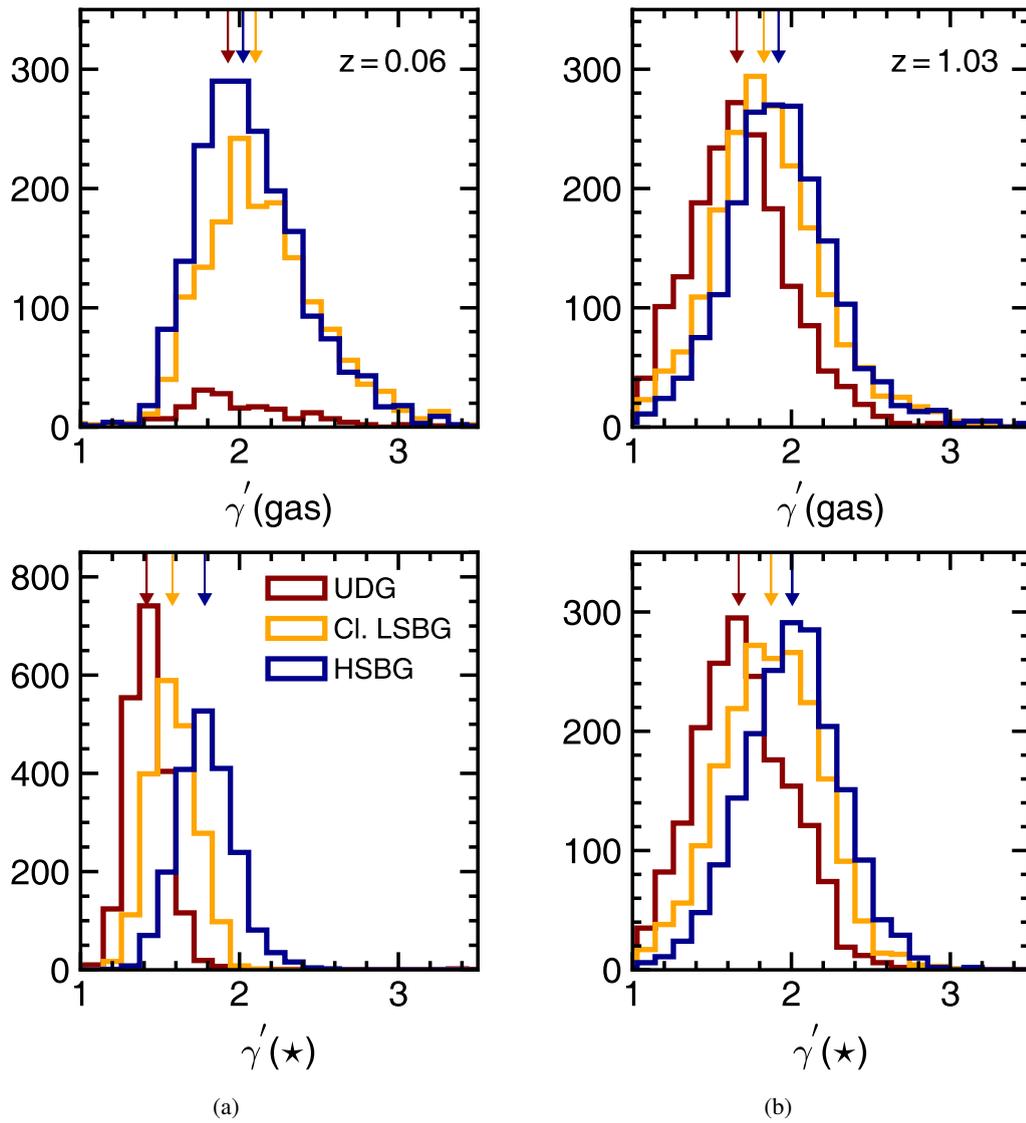


FIGURE 6.10: **(a)** The distribution of the gas (top) and stellar (bottom) log-log density slopes,  $\gamma'$ , calculated within  $0.5 < R_{\text{eff}} < 3$  for UDGs, Cl. LSBGs and HSBGs at  $z = 0.06$ . Coloured arrows indicate the median value for each histogram. **(b)** The same for the distribution of the log-log density slopes at  $z = 1.03$ . As in Figure 6.9, we exclude galaxies with star-forming gas fractions smaller than 0.05 when plotting the gas density slopes.

LSB and HSB populations accelerates). At  $z = 1.03$ , the distribution of  $\gamma'(\text{gas})$  strongly resembles that of  $\gamma'(\star)$  for all three populations. For example, for the UDG gas and stellar density slope distributions, a two-sample Kolmogorov–Smirnov test (Smirnov, 1939) yields a  $D$ -statistic of 0.033 and a  $p$ -value of 0.28, indicating a strong likelihood that the two samples are drawn from the same distribution. This is a natural consequence of the fact that, at early epochs ( $z > 1$ ), the gas distribution is the principal factor driving the development of the stellar profile, especially in UDGs, which form the bulk of their stellar mass at these redshifts. The stellar density slope is, therefore, gradually driven towards the gas density slope over this epoch.

After  $z \sim 1$  the gas and stellar slopes progressively diverge, with the divergence being fastest in UDG

progenitors. By  $z = 0.06$ , the stellar density slopes in UDGs have decoupled completely from the gas density slopes, with the average stellar density slope becoming much shallower than the average gas density slope. Thus, the trigger for the initial divergence of HSBGs and UDGs at high redshift is likely to be processes that act on the gas profiles in UDGs to make them shallower, rather than those that directly affect the stellar components of these galaxies.

In the next section we explore some of the processes that lead to the divergence in the evolution in effective radius, gas fraction and density slopes of LSBGs compared to their HSB counterparts.

## 6.5 How do low-surface-brightness galaxies form?

The analysis presented above shows that the formation mechanisms that produce LSBGs act to both increase the effective radii of their progenitors and drive the steady loss of star-forming gas (either by ejection from the galaxy or by heating). This produces diffuse systems with low SFRs and older stellar populations which, together, result in systems that exhibit low surface-brightnesses. In this section, we study the mechanisms which drive these changes over cosmic time: SN feedback, perturbations due to the ambient tidal field and ram-pressure stripping.

We begin our analysis by taking an aggregate view of the role of key processes that could drive LSBG formation. Figure 6.11 shows distributions of the change in star-forming gas and stellar mass (in units of  $10^9 M_{\odot} \text{Gyr}^{-1}$ ), for approximately evenly spaced simulation outputs ( $\sim 250$  Myrs), in the redshift range  $1 < z < 3$  (left) and at  $z < 1$  (when the HSB and LSB populations diverge most rapidly; right). The top, middle and bottom panels show distributions for the progenitors of HSBGs, Cl. LSBGs and UDGs respectively.

Different regions in this plot indicate different processes that act to produce each of these galaxy populations. For example, star formation will increase the stellar mass while decreasing the gas mass, as it fuels the star formation. Galaxies undergoing star formation will, therefore, populate the upper-left quadrant of this plot. Mergers increase both stellar and gas mass (upper right quadrant), with dry mergers towards the left-hand side of this quadrant. The signature of gas removal (e.g. ram-pressure stripping and/or gas heating) is a decrease in gas mass which is not accompanied by a corresponding change in stellar mass (i.e. the negative half of the  $x$ -axis), while gas accretion causes points to accumulate close to the positive half of the  $x$ -axis. Tidal stripping (which is driven by tidal heating) results in stripping of both stellar and gas mass (lower left quadrant), typically from the outskirts of a galaxy. Tidal heating will also cause the entire distribution of stars to expand, although this is not possible to show in this plot. Finally, the lower right quadrant is typically forbidden, because galaxies tend not to increase their gas mass while simultaneously losing stars.

The top panel in Figure 6.11 indicates that HSBG evolution at all epochs is largely driven by gas accretion, star formation and mergers, with little impact from processes such as ram-pressure or tidal

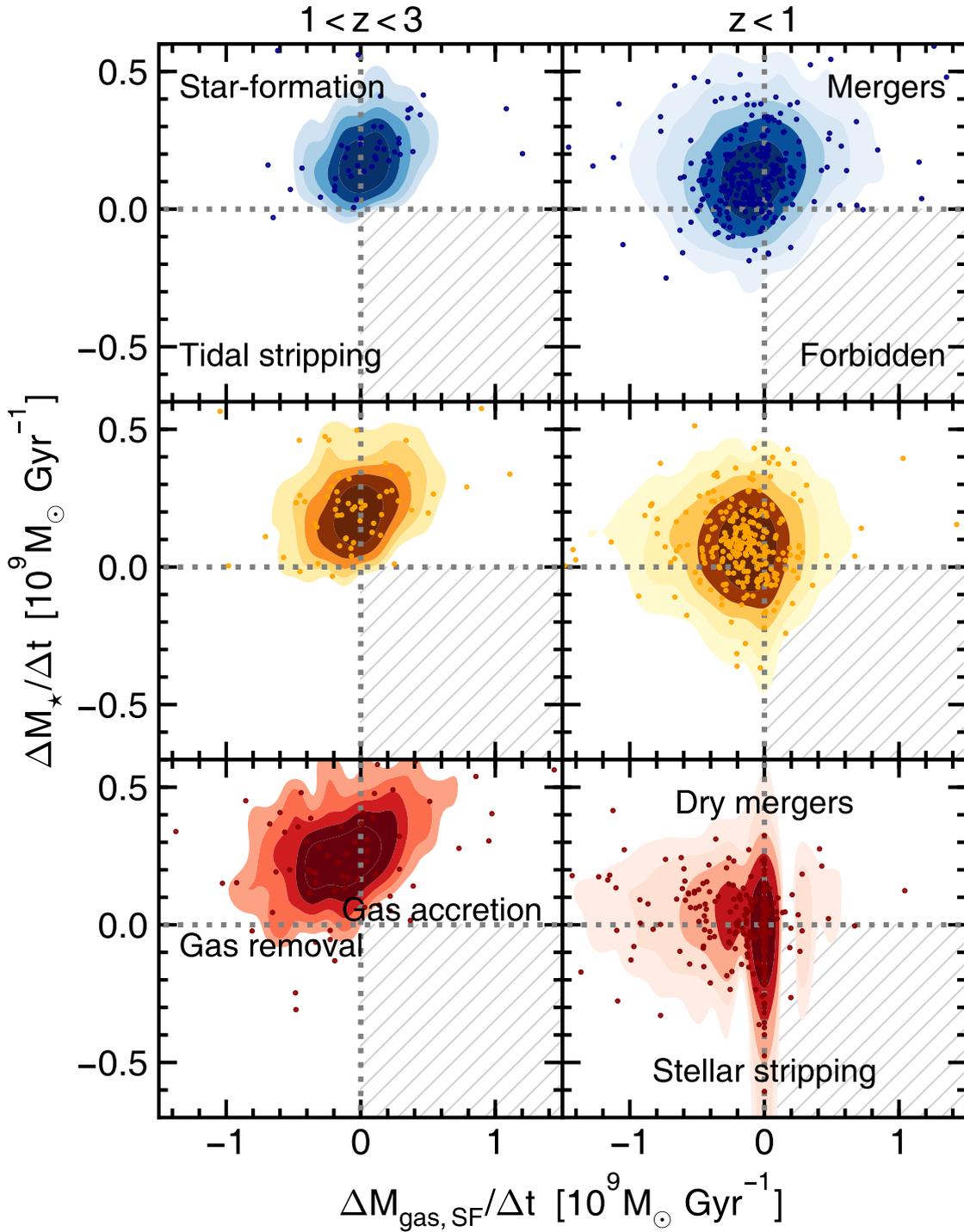


FIGURE 6.11: Contour plots showing the distribution of the rate of change in star-forming gas mass vs the rate of change in stellar mass, in units of  $10^9 M_{\odot} \text{Gyr}^{-1}$ . A random selection of the data is plotted using the coloured points in each panel. Each point represents the change between two consecutive timesteps. The left panels show changes in the redshift range  $1 < z < 3$  and the right panels show the same for  $z < 1$ . The top, middle and bottom rows show distributions for HSBGs, Cl. LSBGs and UDGs respectively. Labels in each quadrant of the top panels indicate the main processes operating in that section of the parameter space. Labels in the bottom panels indicate the processes that cause points to fall close to the horizontal and vertical axes.

stripping/heating. Star formation at high redshift is smooth and the gas mass lost to star formation is typically replenished by accretion. At lower redshifts, star formation remains at similar levels, but accretion is typically no longer fast enough to offset the gas that is transformed into stars. The plots show that the degree of gas stripping and heating experienced by HSBGs must be small as, in the vast majority of cases, any decrease in gas mass is accompanied by an increase in stellar mass of a similar magnitude.

However, as we transition to populations that have lower surface-brightnesses at the present day, the relative role of these processes changes. Cl. LSBGs and UDGs both show similar evolution to HSBGs at  $z > 1$ , the epoch at which the bulk of their stellar mass forms (see Section 6.5.1 below). However at  $z < 1$ , Cl. LSBGs and, in particular, UDGs, have both experienced large decreases in gas mass that are not the result of star formation. In the case of UDGs, ram pressure stripping and tidal stripping/heating of both stars and gas are clearly important processes in their evolutionary history (particularly at lower redshifts), as shown by the much higher fraction of such systems that inhabit the lower left quadrant compared to HSBGs. In the following sections, we study the mechanisms that drive these processes and explore their relative role in creating LSB systems over cosmic time.

### 6.5.1 Supernova feedback - a trigger for LSBG formation

Theoretical studies by Di Cintio et al. (2017) and Chan et al. (2018) show that, at least at low stellar masses, SN feedback may be capable of producing UDGs by fuelling outflows which create flattened total density profiles (e.g. Navarro et al., 1996; Governato et al., 2010; Pontzen & Governato, 2012; Teyssier et al., 2013; Errani et al., 2015; Oñorbe et al., 2015; Carleton et al., 2018; Sanders et al., 2018). These outflows may be effective, not only at removing gas from the galaxy, but also at producing shallower gas density profiles (Brook et al., 2011, 2012; Pontzen & Governato, 2014; Di Cintio et al., 2014a,b; Dutton et al., 2016; Di Cintio et al., 2017) and through the dynamical heating of stars, increasing their effective radii (Chan et al., 2015; El-Badry et al., 2016) (although there may be some tension with observations e.g. Patel et al., 2018). It may also be the case, as we show later, that, rather than directly influencing the size and gas content of galaxies, SN feedback instead allows other processes to work more efficiently (e.g. tidal heating and ram-pressure stripping).

It has been shown using the Horizon suite of simulations that the inclusion of baryons and their associated feedback processes results in shallower stellar density slopes, compared to an otherwise identical DM-only simulation (Peirani et al., 2017). As we have already shown in Section 6.3, the DM masses of UDGs and Cl. LSBGs are not dissimilar to that in their HSB counterparts. It is therefore not the case that UDGs are massive haloes that have been quenched before their reservoir of star-forming gas has been used up. Instead, it is worth considering whether differences in the actual stellar assembly history of LSBGs, especially at early epochs where the bulk of the stellar mass was formed, may be a contributing factor to creating their LSB nature over cosmic time.

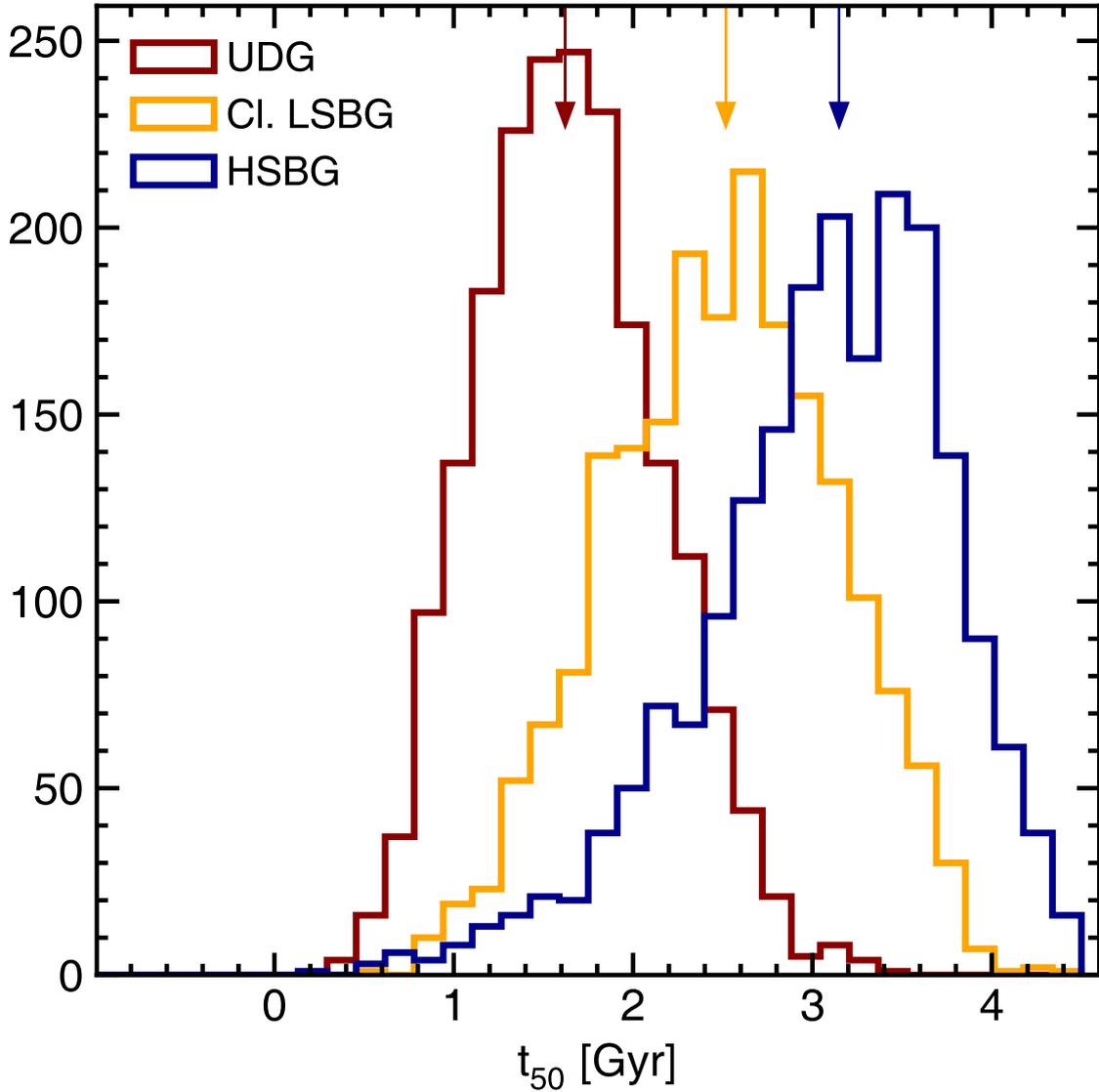


FIGURE 6.12: Distribution of  $t_{50}$ , (i.e. the minimum time required to form 50 per cent of a galaxy’s present-day stellar mass) for UDGs, Cl. LSBGs and HSBGs. Coloured arrows indicate the median value for each histogram.

Since the amount of SN feedback energy deposited in the potential well will be sensitive to the star formation history, we consider the burstiness of star formation in our different galaxy populations. To quantify the burstiness we define  $t_{50}$ , which measures the minimum amount of time required to form 50 per cent of the stellar mass in a galaxy<sup>4</sup>. Figure 6.12 shows the distribution of  $t_{50}$  values for the HSBG, Cl. LSBG and UDG populations. The median value of  $t_{50}$  for HSBGs is typically  $\sim 3$  Gyrs, while for UDGs it is  $\sim 1.5$  Gyr (with the value for Cl. LSBGs falling in between these values). This indicates that the formation of UDGs is much more rapid than HSBGs of similar stellar masses. UDGs typically

<sup>4</sup>In order to calculate  $t_{50}$ , we first produce a histogram of the distribution of star-particle ages in each galaxy at  $z = 0.06$ , using bin widths of 100 Myrs. The bins in each histogram are then re-sorted, in order of decreasing frequency, and the cumulative distribution function (CDF) is calculated.  $t_{50}$  is the time at which this CDF reaches 50 per cent.

assemble earlier and, on average, they have already formed 75 per cent of their stellar mass by  $z = 1$  (i.e. as a result of halo assembly bias, Sheth & Tormen, 2004).

On the other hand, HSBGs have formed only 30 per cent of their stellar mass by this time. The median SFR for HSBGs falls only modestly between  $z = 3$  and the present day. As a result, energy released by supernovae (SNe) and stellar winds is distributed over most of the lifetime of the galaxy, whereas feedback energy is almost entirely concentrated before  $z = 1$  in the case of UDGs. This, in turn, means that the maximum instantaneous energy imparted into the gas is much larger in UDGs than in their HSB counterparts.

We proceed by quantifying the impact that SN and stellar feedback may have on the galaxy populations due to their disparate formation histories. We define the total mechanical and thermal energy released by stellar processes between two timesteps,  $t_0$  and  $t_1$ , by summing the energy released by each star particle within this interval:

$$E_{\text{SN}} = \sum_i m_{*,i} (E(z_1, Z)_i - E(z_0, Z)_i) \quad (6.2)$$

where  $m_{*,i}$  is the mass of a star particle and  $E(z, Z)_i$  is the cumulative mechanical and thermal energy released by that star particle as a result of Type Ia SNe, Type II SNe and stellar winds per unit stellar mass, for a metallicity  $Z$ , and between the time of its formation and a redshift of  $z$ .

Figure 6.13 shows the median mechanical and thermal energy released by stars over the last 100 Myr, as a function of redshift. Since our samples are matched in stellar mass, the total cumulative feedback energy for each sample reaches the same value at  $z = 0.06$  but the pattern of energy injection differs between the populations. Since they form the majority of their stellar mass early on (75 per cent before  $z = 1$ ), the progenitors of UDGs release energy over a shorter period of time.

As a result, UDGs experience high levels of SN feedback at early times. Between  $z = 3$  and  $z = 1$ , UDGs have already released 75 per cent of their integrated stellar feedback energy, compared with 50 per cent for Cl. LSBGs and only 30 per cent for HSBGs. The SN energy released in HSBGs remains roughly constant as a function of redshift, decreasing by only 0.25 dex between the peak at  $z \sim 1$  and  $z = 0$ . In comparison, the SN energy released in UDGs peaks between  $z = 2$  and  $z = 1$  and then declines rapidly towards  $z = 0$  to a value 1.5 dex lower.

We note that the same patterns are *not* observed for AGN feedback. As Figure 6.14 shows, the evolution of AGN feedback energy in UDGs, Cl. LSBGs and HSBGs proceeds similarly at high redshift ( $z > 1$ ), falling rapidly for low redshift UDGs as hot gas in cluster environments quenches the Bondi accretion rates of their BHs. Additionally, BH growth in low-mass haloes is regulated by SN feedback (e.g. Volonteri et al., 2015; Habouzit et al., 2017; Bower et al., 2017), so that SN feedback is the principal feedback process in the UDG population.

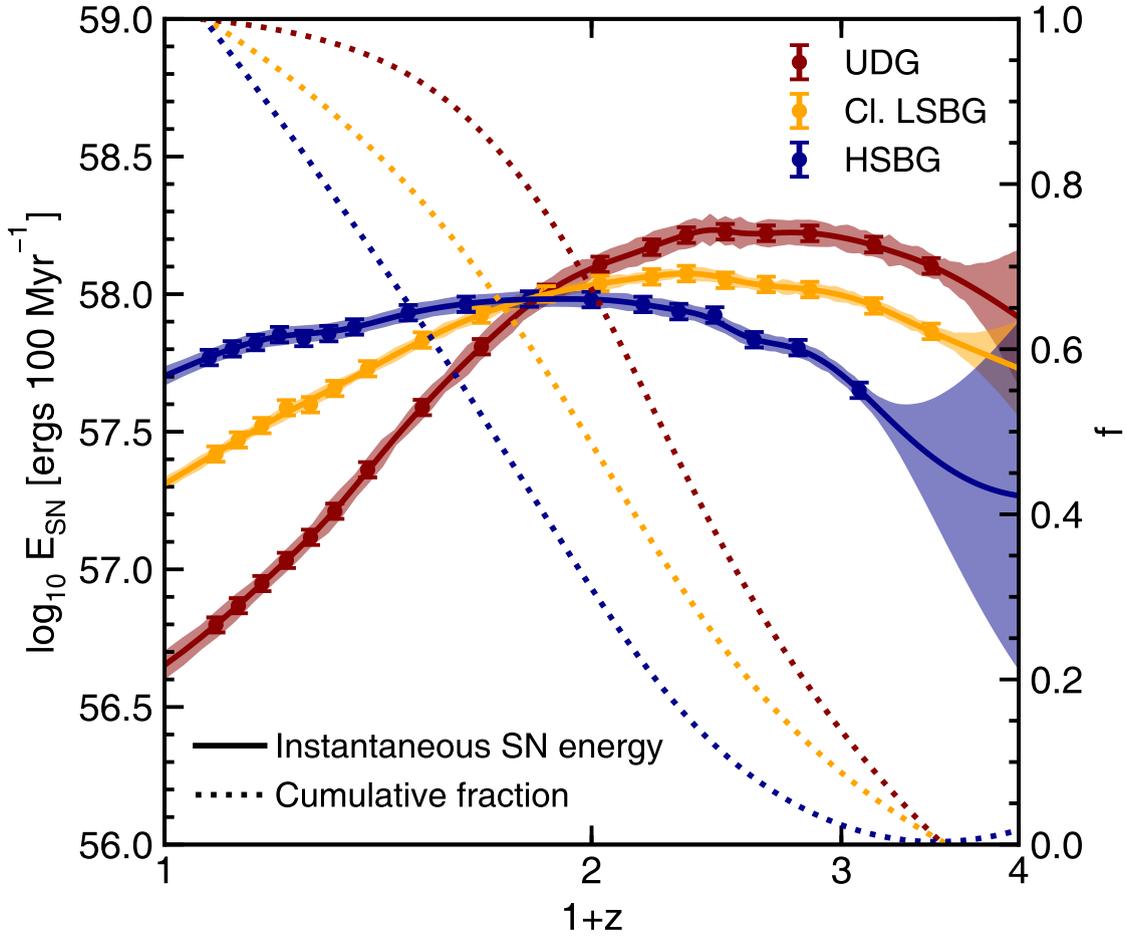


FIGURE 6.13: Median mechanical and thermal energy injected as a result of Type Ia and Type II SNe and stellar winds. Solid lines indicate the SN energy released per 100 Myr as a function of redshift. Dotted lines show the average cumulative energy as a fraction of the value at  $z = 0.06$ . The fractions are indicated by the values on the right-hand axis.

As we have shown in Section 6.4.2 (Fig 6.9), the gas density slopes of UDG progenitors are significantly and consistently shallower than their HSB counterparts in the early Universe (between  $z = 1$  and  $z = 3$ ). This coincides with the period where instantaneous SN feedback energy is at its peak in the UDG progenitor population. It is worth noting that the profiles of HSBG progenitors behave very differently at these early epochs. Both their gas and stellar density slopes tend to *increase* with time at these early epochs, as baryons accumulate in the centres of their gravitational potential wells. SN feedback therefore has a much greater impact on LSBG progenitors than it does on their HSB counterparts.

We note that, while large amounts of energy are released into the gas in UDG progenitors at these early epochs, the fraction of star-forming gas (Figure 6.7, bottom panel) in UDG progenitors remains significant ( $f_{\text{gas,SF}} > 0.4$ ) at these times. This indicates that, while the slope of the gas density profile is made shallower due to this SN feedback, the feedback is not so strong that the gas is completely removed and star-formation quenched.

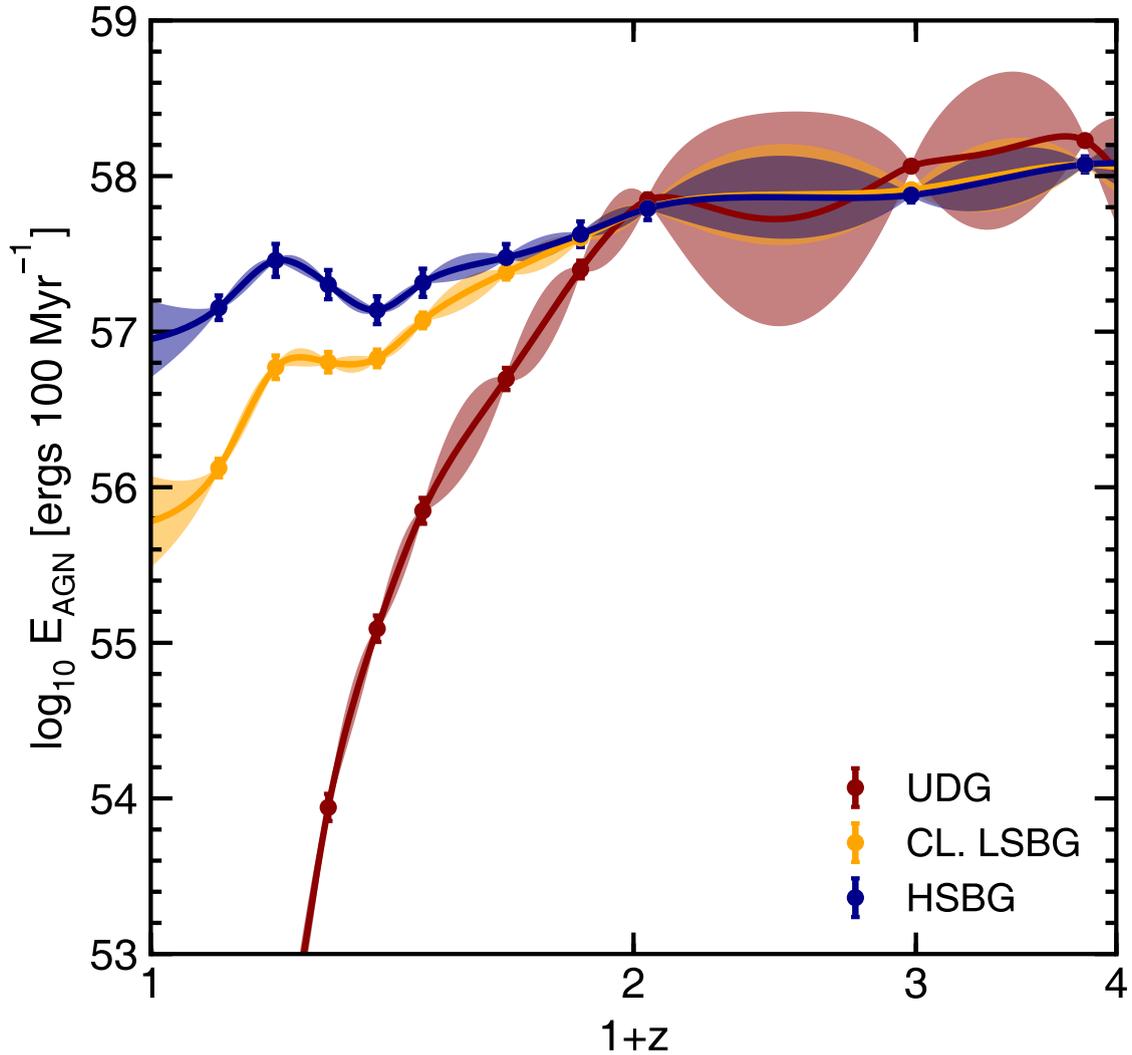


FIGURE 6.14: Median mechanical and thermal energy released as a result of AGN feedback. Solid lines indicate the AGN energy released per 100 Myr as a function of redshift.

As was noted in Section 6.4.2, stars forming from this gas progressively flatten the stellar density slopes, leading to the decrease in  $\gamma'(\star)$  shown in Fig 6.9. SN feedback, therefore, appears to be the mechanism that drives the creation of shallower gas and stellar density slopes in UDG progenitors at high redshift, which leaves these systems more vulnerable to tidal processes (e.g. tidal heating and, additionally, ram-pressure stripping in dense environments) over cosmic time. It is worth noting here that the specific angular momenta of LSBG and HSBG progenitors are very similar at  $z \sim 3$ , indicating that the flatter density profiles of LSBG progenitors is not due to them initially forming with higher values of spin.

Although UDGs clearly increase in size (Figure 6.7, top panel) and gain flatter density slopes (Figure 6.9, bottom panel) compared to HSBGs and other LSBGs at  $z > 1$ , the difference is fairly modest compared to the much greater divergence in effective radii and gas fractions seen *after*  $z = 1$  (Figure 6.7, top panel). Thus, SN feedback appears to be the initial *trigger* for the divergence of UDGs from the rest of the galaxy population, rather than the principal cause of their large sizes at  $z = 0$ . A combination

of a shallower potential and a broader distribution of stars is likely to contribute to the steep rise in the effective radii of UDG progenitors, in contrast to their HSB counterparts seen after  $z = 1$ .

Much of this evolution must be due to external processes that act to increase the effective radii steadily over cosmic time. Since they would be expected to operate more efficiently on systems where galaxies have shallower gravitational potentials (and where the material, at least in the outer regions, is more weakly bound), environmental processes such as perturbations from the ambient tidal field and ram-pressure stripping are likely to amplify the initial divergence produced by SN feedback (Pontzen & Governato, 2012; Errani et al., 2015; Carleton et al., 2018; Sanders et al., 2018). We explore the effect of these processes in the next two sections.

It is worth noting here that processes other than SN feedback could assist in the initial creation of shallower density slopes in UDG progenitors. For example, an accretion history that is rich in low-mass-ratio (i.e. minor) mergers may also act to broaden the stellar distribution (e.g. Naab et al., 2009; Bezanson et al., 2009; Hopkins et al., 2010; Bédorf & Portegies Zwart, 2013). However, while there is some evidence that LSBG progenitors do exhibit some level of enhancement in their merger histories in the *early* Universe ( $\sim$  twice the number of major mergers undergone by HSBGs between  $z = 3$  and  $z = 1$ ), it is difficult to draw concrete conclusions, as the merger histories of low mass-ratio mergers are typically highly incomplete in the simulation at high redshift (Martin et al., 2018c)<sup>5</sup>.

In order to quantify the relative (and probably additive) roles of feedback (e.g. Dashyan et al., 2018) and minor mergers (e.g. Di Cintio et al., 2019) in triggering the initial shallower gas density profiles, a higher resolution simulation is required. In a forthcoming paper (Jackson et al. in prep) we will use New-Horizon (Dubois et al. in prep), a 4000 Mpc<sup>3</sup> zoom-in of a region of Horizon-AGN, which has 64 times better spatial resolution to probe this ‘trigger epoch’ in more detail.

## 6.5.2 Perturbations due to the ambient tidal field - a key driver of LSBG evolution

Recall first from the arguments above that the processes that produce LSBGs operate steadily over cosmic time (since the effective radii and gas fractions change gradually with redshift) and are not specific to cluster environments (since UDGs are found in all environments). Mergers and tidal interactions with nearby objects offer an attractive mechanism for LSBG formation because they act to dynamically heat galaxies and destroy cold, ordered structures (Moore et al., 1996, 1998; Gnedin, 2003; Johansson et al., 2009). These processes are therefore likely contributors to both the observed increase in the effective radii and the decrease in the star-forming gas fractions seen in the LSBG population, regardless of local environment.

---

<sup>5</sup>Due to the stellar mass resolution of the simulation, only objects that are more massive than  $2 \times 10^8 M_{\odot}$  are detectable. As a result, only 50 per cent and 20 per cent of the ( $z = 0$ ) progenitors of  $10^{9.5} M_{\odot}$  galaxies are massive enough for a 1:10 mass ratio merger to be detectable at  $z = 2$  and  $z = 3$  respectively (Martin et al. 2018c, Figure 1)

It is worth noting first that, compared to HSBGs and Cl. LSBGs, UDGs in our sample are considerably more ‘spheroidal’ (i.e. a larger fraction of their stars are on random orbits compared to ordered, rotational ones). While the median value of the ratio of rotational to dispersal velocities of the stellar component,  $(V/\sigma)_*$ , is 0.4 for Cl. LSBGs, it is only 0.15 for UDGs. In comparison, late-type i.e. disc-dominated galaxies typically exhibit  $(V/\sigma)_* > 0.55$  (Martin et al., 2018a). Since mergers and interactions are efficient drivers of (disc-to-spheroid) morphological transformation (Martin et al., 2018c), this is evidence that the UDGs have indeed undergone a larger number of interactions (but not necessarily actual mergers) that have shaped their structural evolution.

Recent observational work lends support to the idea that the formation of LSBGs is connected to the tidal effects of nearby galaxies. Some studies have pointed to the idea that UDG progenitors may be more massive star-forming dwarfs that are destroyed as a result of interactions within a cluster environment (e.g. Conselice, 2018). Alternatively, they may be less massive dwarfs that have undergone considerable expansion (e.g. Carleton et al., 2018) due to tidal interactions. It has also been suggested that at least some UDGs may be tidal dwarfs (e.g. van Dokkum et al., 2018; Ogiya, 2018; Greco et al., 2018a), formed when material is stripped from larger galaxies. However, since mergers typically produce tidal dwarfs with low stellar masses (less than 1 per cent of the mass of the merging progenitors, see e.g. Barnes (1992); Okazaki & Taniguchi (2000); Kaviraj et al. (2012a)), the mass range that we consider in this study ( $M_{\odot} > 10^8$ ) precludes significant numbers of these objects in our sample.

In the context of mergers (i.e. interactions which result in the actual coalescence of the interacting progenitors) it is worth noting that both LSBGs and HSBGs undergo very few actual mergers at low redshift, where the effective radii and star-forming gas fractions change significantly. Indeed, only a few percent of galaxies have undergone mergers of mass ratios larger than 1:4 since  $z = 1$ ; see e.g. Darg et al. 2010a; Martin et al. 2018b; Martin et al. 2018c). While UDGs do undergo more mergers than HSBGs at high redshift (as was noted earlier in Section 6.5.1), they experience a relative dearth of mergers (a factor of 2.5 fewer major mergers) than their HSB counterparts between  $z = 1$  and the present day, when much of the increase in radii and decrease in gas content takes place. Galaxy mergers, therefore, are unlikely to be the principal driver of LSBG evolution over cosmic time.

However, tidal interactions (or fly bys) between galaxies can produce similar effects to that due to actual mergers (e.g. Martin et al., 2018c; Choi et al., 2018). To explore the effect of tidal interactions on LSBGs and HSBGs, we employ a perturbation index (PI) which quantifies the environmental tidal field due to objects in the vicinity of the galaxy in question. We define the PI (e.g. Byrd & Valtonen, 1990; Choi et al., 2018) between  $z = 3$  and the redshift in question, by calculating the *cumulative* contribution of all galaxies within 3 Mpc:

$$PI = \int_{z=3}^z \sum_i \left( \frac{M_i}{M_{gal}} \right) \left( \frac{R_{eff}}{D_i} \right)^3 dt / \text{Gyr} \quad (6.3)$$

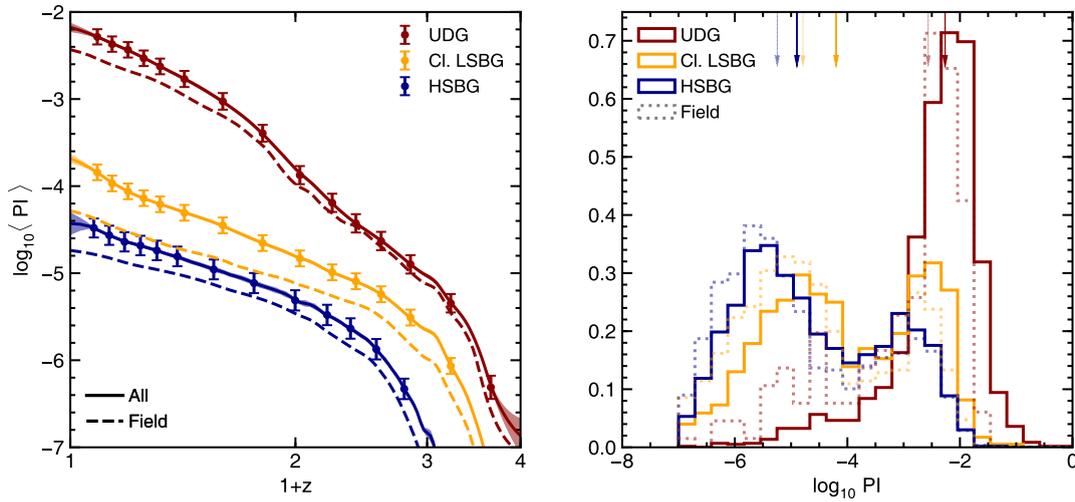


FIGURE 6.15: **Left:** Median perturbation index (PI), as defined by Equation 6.3, between  $z = 3$  and the redshift in question. Error bars indicate the errors on the median value of the PI at each redshift and solid filled regions show the  $1\sigma$  confidence intervals for a Gaussian process regression to these points. Dashed lines indicate the same for galaxies in the field only. **Right:** Distribution of the perturbation index between  $z = 3$  and  $z = 0$ . Dotted lines show the distribution for field galaxies only. Coloured arrows indicate the median value for each histogram and fainter arrows indicate the median values for field galaxies. Note that the histograms are normalised so that the field and general populations can be easily compared.

where  $M_{gal}$  is the stellar mass of the galaxy in question and  $M_i$  is the stellar mass of the  $i$ th perturbing galaxy.  $R_{eff}$  is the effective radius as defined in Section 6.2.3,  $D_i$  is the distance from the  $i$ th perturbing galaxy and  $dt$  is in units of Gyrs. By this definition, galaxies that are more massive and/or approach more closely will contribute more to the PI, with each galaxy's contribution dropping off steeply with distance. For example, a perturbation index  $PI = 10^{-1}$  is equivalent to a single 1:10 mass ratio merger or an equal mass galaxy moving within 2 effective radii. We note that our definition of PI is a cumulative one, so that we integrate the perturbations felt by individual galaxies between  $z = 3$  and the redshift in question ( $z$ ). The PI is calculated at evenly spaced timesteps of  $\sim 130$  Myr and we do not attempt to integrate galaxy orbits, as the relatively coarse time resolution makes this unreliable.

In the top panel of Figure 6.15 we plot the median value of the PI in each of our populations, as a function of redshift. At all redshifts galaxies that have lower surface-brightnesses exhibit consistently higher PI values. The discrepancy between the median PI values in the LSBG and HSBG populations becomes more pronounced with time. Compared with HSBGs, UDGs in all environments undergo more frequent or violent perturbations, exhibiting PI values more than 2 dex higher towards low redshift (with Cl. LSBGs reaching values around 1 dex higher). Not unexpectedly, for all populations, galaxies that inhabit the field exhibit lower PI values.

In the bottom panel, we show the PI over the entire redshift range of the top panel ( $0 < z < 3$ ), i.e. Equation 6.3 evaluated at the present day, for each of the galaxy populations. In other words, this is the cumulative impact of the tidal field experienced by the galaxy over around 90 per cent of cosmic time.

The PI values for UDGs are significantly larger, with the median of the UDG distribution being around 2 orders of magnitude greater than that for the HSBGs.

We note that, if the definition of the perturbation index is changed so that it is independent of  $R_{\text{eff}}$  (by fixing  $R_{\text{eff}}$  to 1 kpc), the average perturbation index for UDGs remains significantly larger than for equivalent HSB galaxies. With such a change in definition, the median for UDGs remains 40 times higher than for HSBGs (compared to 160 times higher when radius is considered), indicating that the PI is a genuine result of stronger perturbations, rather than simply an effect of galaxy size.

It is important to note that the perturbations felt by UDGs are not a strong function of environment. As the dashed red line in the top panel and the dotted histograms in the bottom panel indicate, the majority of UDGs in field environments have still undergone very large perturbations compared with their HSB counterparts. Indeed the PI values of field UDGs are not dissimilar to that of the general UDG population (which is dominated by UDGs in groups and clusters). Finally, it is worth noting that if we only consider galaxies in low-density field environments which are not satellites, i.e. those that are truly isolated, the cumulative PI of such UDGs remains more than 10 times higher than that of field HSBGs. Together with the fact that field UDGs have similar effective radii and star-forming gas fractions at the present day to UDGs in clusters (Fig 6.7), this indicates that *tidal interactions are likely to be the primary mechanism that drives LSBG evolution and causes these systems to both expand and lose their reservoir of star-forming gas over cosmic time.*

### 6.5.3 Ram pressure stripping - an additional mechanism of gas removal in cluster LSBGs

While tidal perturbations are capable of acting on galaxies regardless of their environment, ram-pressure provides an additional process that can shape the evolution of galaxies in denser environments, particularly in clusters. The ram pressure exerted on the gas in a galaxy as it travels through a hot intra-cluster medium (ICM) or intra-group medium (IGM) can remove gas from the galaxy and quench star formation (Gunn & Gott, 1972). This represents an appealing mechanism for explaining the transformation of galaxies from gas-rich, star-forming objects to quiescent systems that might resemble LSBGs at the present day. Indeed, the interaction between the ICM/IGM and the inter-stellar media of galaxies that are traversing hot, dense environments has often been used to explain the deficiency of gas and the redder colours of galaxies in clusters (e.g. Chamaroux et al., 1980; Lee et al., 2003; Sabatini et al., 2005; Boselli et al., 2008; Gavazzi et al., 2013; Habas et al., 2018). This is a particularly effective mechanism in low-mass galaxies ( $M_{\star} < 10^{10}M_{\odot}$ ), as gravitational potentials are typically shallow enough to allow the efficient removal of gas (e.g. Vollmer et al., 2001). In this section, we explore whether ram pressure stripping may play a role in the gas exhaustion that creates our sample of LSBGs.

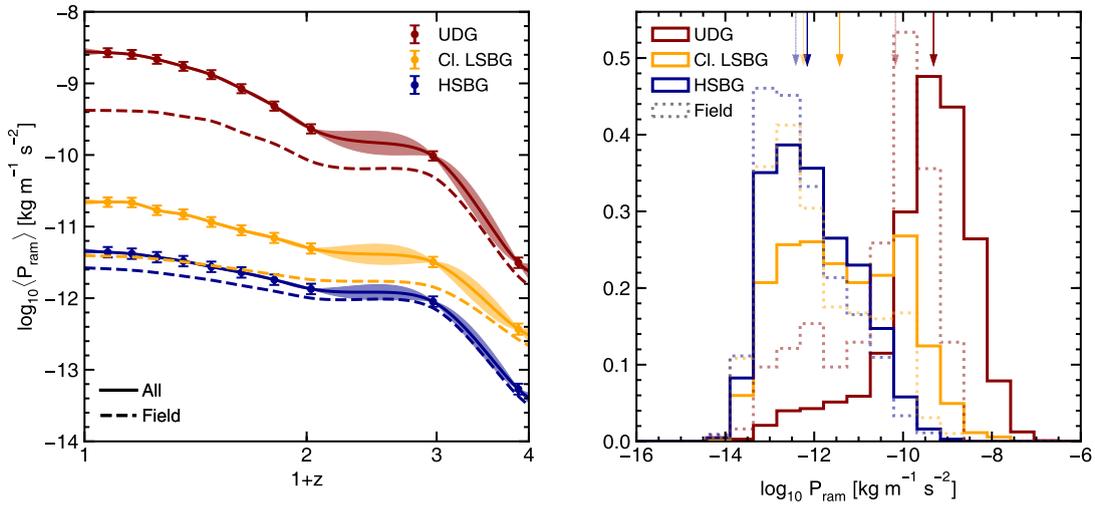


FIGURE 6.16: **Top:** The cumulative ram pressure felt by galaxies between  $z = 3$  and the redshift in question, as defined by Equation 6.4. Error bars indicate the errors on the median value of  $P_{ram}$  at each redshift and solid filled regions show the  $1\sigma$  confidence intervals for a Gaussian process regression to these points. Dashed lines show the cumulative ram pressure felt by field galaxies only. **Bottom:** Distribution of the cumulative ram pressure felt by galaxies between  $z = 3$  and  $z = 0$ . Coloured arrows indicate the median value for each histogram and fainter arrows indicate the median values for galaxies in the field. Dotted lines indicate the total integrated ram pressure for field galaxies only. Note that the histograms are normalised so that the field and general populations can be easily compared.

### 6.5.3.1 Ram pressure

The cumulative ram pressure,  $P_{ram}$ , felt between  $z = 3$  and  $z$  by a galaxy moving through the local medium is given by

$$P_{ram} \sim \int_{z=3}^z \rho_{IGM} v_{gal}^2 dt / \text{Gyr} \quad (6.4)$$

where  $v_{gal}$  is the velocity of the galaxy relative to the bulk velocity of the surrounding medium and  $\rho_{IGM}$  is the mean gas density of the surrounding medium within 10 times the maximum extent of the stellar distribution of the galaxy.

The top panel of Figure 6.16 shows the median cumulative value of  $P_{ram}$  for the HSBG, Cl. LSBG and UDG populations as a function of redshift. The average ram-pressure continues to increase towards the present day for UDGs, Cl. LSBGs and HSBGs. However, the average ram-pressure felt by HSBGs and Cl. LSBGs is relatively small at all redshifts (around 2–3 orders of magnitude smaller than that of the UDG population). Ram-pressure stripping begins to have a significantly stronger impact on UDG progenitors around  $z = 1$ . This is consistent with the typical infall epoch of galaxies into clusters (e.g. Tormen, 1998; Muldrew et al., 2015; Mistani et al., 2016; Muldrew et al., 2018).

The cumulative ram pressure experienced by the progenitors of UDGs in the field (dashed red line) is significantly lower (by an order of magnitude) than the general population of UDG progenitors. Although the level of ram pressure in these field UDGs is high compared to that in Cl. LSBGs and HSBGs,

it is low enough that significant gas stripping does not occur (as indicated by the relatively high total gas fractions retained by field UDGs at  $z = 0$ , shown in the bottom panel of Fig 6.7). The bottom panel of Figure 6.16 shows the cumulative ram pressure experienced by HSBGs, Cl. LSBGs and UDGs between  $z = 0$  and  $z = 3$ . Again, the cumulative ram pressure felt by UDGs is, on average, several orders of magnitude higher than that felt by either the Cl. LSBGs or HSBGs.

It is worth noting here that the ram pressure experienced by UDGs in the field is higher than that experienced by Cl. LSBGs and HSBGs. This is a consequence of the fact that a larger fraction ( $\sim 65$  per cent) of local UDGs are satellites (i.e. their haloes are identified as sub-structures of a more massive halo) while a majority of low-mass field HSBGs at  $z = 0$  are not (only  $\sim 25$  per cent of these galaxies are satellites). UDGs are therefore typically found in regions of slightly higher gas density and experience ram pressure due to the host halo they are embedded in (e.g. Simpson et al., 2018). When genuinely isolated UDGs are selected (i.e. those that are not satellites), the ram pressure felt falls significantly so that the median cumulative ram pressure felt by completely isolated UDGs, LSBGs and HSBGs agrees to within 0.2 dex.

### 6.5.3.2 Bulk flow of gas

Studying the bulk flow of gas within galaxies also allows us to quantify the degree to which ram-pressure stripping is experienced by our different galaxy populations. We explore the density weighted average angle,  $\theta$ , between the relative velocity between the gas and stars ( $\mathbf{v}_{rel}$ ) and the bulk motion of the stellar component in the observed frame ( $\mathbf{v}_*$ ):

$$\cos(\theta) = \frac{1}{\sum \rho_i} \sum_i \frac{\mathbf{v}_{rel,i} \cdot \langle \mathbf{v}_* \rangle}{|\mathbf{v}_{rel,i}| \cdot |\langle \mathbf{v}_* \rangle|} \rho_i \quad (6.5)$$

where  $\mathbf{v}_{rel} = \mathbf{v}_{gas} - \langle \mathbf{v}_* \rangle$  is the velocity of each gas cell relative to the average velocity of the galaxy's stellar component. In the case where the bulk motion of the gas is in the opposite direction to the stars,  $\theta$  will be close to  $\pi$  radians (and  $\cos(\theta)$  will be close to -1). When the gas and stellar components are moving together at roughly the same velocity, the angle between a given component of  $\mathbf{v}_{rel}$  and  $\langle \mathbf{v}_* \rangle$  is essentially randomly distributed and therefore  $\theta$  will be close to  $\pi/2$  (i.e.  $\cos(\theta) = 0$ ). If the gas is either moving ahead of the stellar component of the galaxy, or being accreted in a wake behind the galaxy (e.g. Sakelliou, 2000), then  $\theta$  will be close to 0. When ram-pressure stripping occurs, we therefore expect  $\theta$  to be close to  $\pi$  radians and  $\cos(\theta)$  to be close to -1. Note that gas loss as a result of mechanisms other than ram pressure stripping does not produce the same signature. For example, in the case of gas loss driven by harassment or feedback processes, gas moves out of the galaxy either in a random direction or approximately isotropically, so the average value of  $\cos(\theta)$  will be close to 0.

Figure 6.17 shows the *minimum* value of  $\cos(\theta)$  that galaxies exhibit over cosmic time. Thus, minimum  $\cos(\theta)$  values close to 0 would indicate that the ram pressure has not operated on the galaxy at any

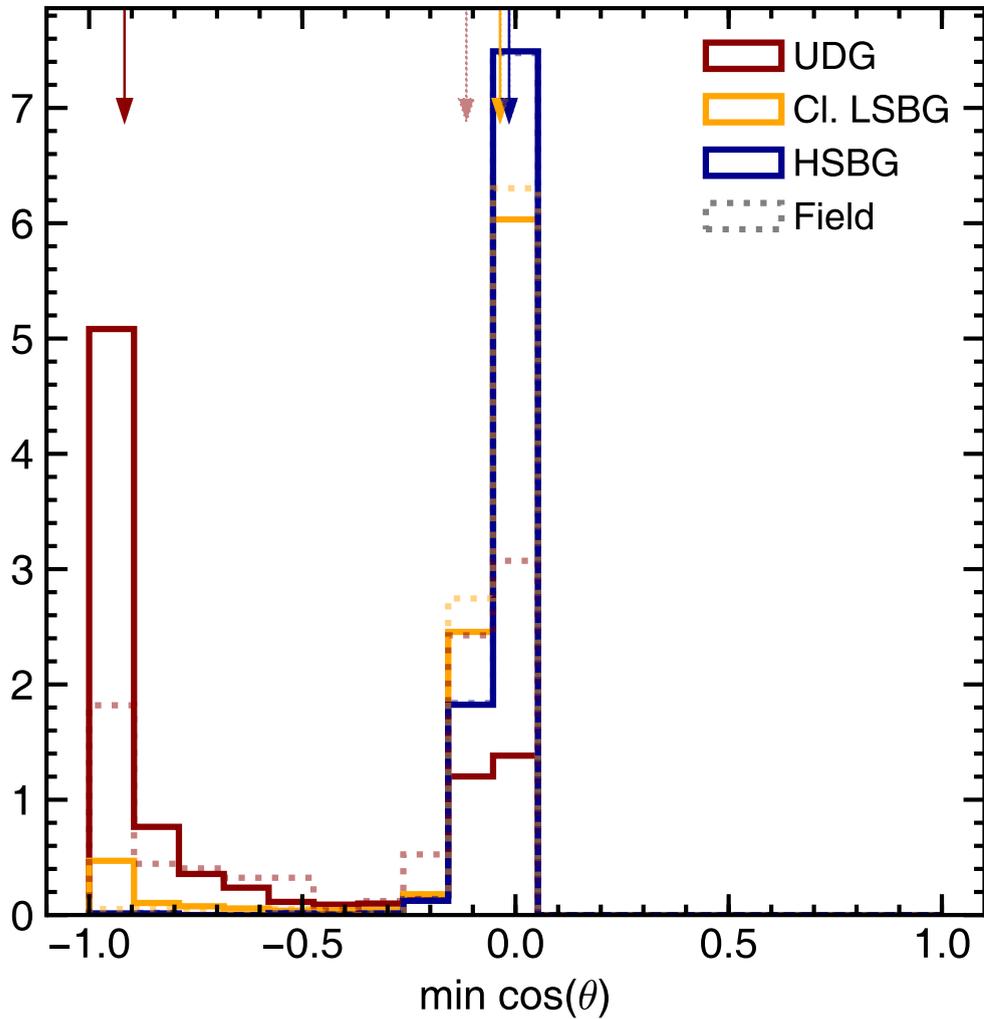


FIGURE 6.17: The minimum value of  $\cos \theta$  between  $z = 3$  and  $z = 0$ , where  $\theta$  is the angle between the average direction of the bulk motion of gas relative to that of the stellar component in galaxies. Dotted lines indicate the largest value of  $\cos \theta$  for field galaxies only. Coloured arrows indicate the median value for each histogram and fainter dotted arrows indicate the median values for field galaxies only. Note that the histograms are normalised so that the field and general populations can be easily compared.

point over cosmic time. On the contrary, if we consider galaxies to have undergone some ram-pressure stripping when the minimum value of  $\cos(\theta)$  is less than  $-0.75$ , then a large majority (65 per cent) of UDGs have undergone ram pressure stripping at some point in their history. The same is not true of Cl. LSBG or HSBG progenitors (or to a large extent, field UDGs). By the same definition, almost none of the HSBGs in our sample (0.3 per cent) have ever undergone significant ram-pressure stripping and a small minority of Cl. LSBGs have (6 per cent). In the field, only a modest fraction (25 per cent) of field UDGs have been ram-pressure stripped.

Taken together, Figure 6.16 and Figure 6.17 indicate that ram-pressure stripping make a significant contribution to the quenching of UDG progenitors in dense environments. However, UDGs in the field

are not as significantly stripped (as shown by both panels of Figure 6.16) but still have very low *star-forming* gas fractions at  $z = 0$  (panel d of Figure 6.6). This indicates that ram-pressure stripping is not a necessary ingredient for the low star formation rates seen in today's UDGs. The high total gas fractions and low *star-forming* gas fractions of field UDGs indicates that, for this subset of UDGs, their gas has been heated by other processes rather than been entirely removed from the galaxy. Thus, in cases where ram-pressure stripping is absent, other processes still act to quench UDGs by heating their gas. While ram-pressure stripping is an important mechanism for removing gas from UDGs in dense environments, UDGs (in all environments) lose their star-forming gas through tidal perturbations, even in the absence of this process. Ram-pressure stripping is, therefore, an *additional* process, to tidal perturbations, that assists in the removal of gas in LSBGs, particularly in clusters, but is not necessary for quenching their star formation. Interaction with the tidal field remains the principal driver of LSBG evolution in all environments.

## 6.6 Summary

In the forthcoming era of deep-wide observational surveys, the low-surface-brightness Universe represents an important new frontier in the study of galaxy evolution. While largely uncharted, due to the lack of depth of past wide-area datasets like the SDSS, low-surface-brightness galaxies (LSBGs) are essential to a complete understanding of galaxy evolution. Recent work using small deep surveys has hinted at the significant contribution that LSBGs may make to the galaxy number density of the local Universe and highlighted the need to understand the evolution of these objects across all local environments. Given the current dearth of data on LSBGs, theoretical insights, using cosmological simulations, into their demographics, the redshift evolution of their properties and the principal mechanisms that drive their formation is highly desirable.

Here, we have used the Horizon-AGN hydrodynamical cosmological simulation to perform a comprehensive study of the formation and evolution of LSBGs. We have (1) studied the demographics and properties of local LSBGs and compared them to that of their high-surface-brightness (HSB) counterparts, (2) explored the evolution of the properties of LSBG progenitors with redshift and (3) quantified the role of key processes, in particular SN feedback, tidal perturbations and ram-pressure stripping, that lead to the formation of LSB systems. Our main conclusions are as follows:

- *LSBGs are significant contributors to the number density of galaxies in the local Universe.* For  $M_{\star} > 10^8 M_{\odot}$ , LSBGs contribute 47 per cent of the local number density ( $\sim 85$  per cent for  $M_{\star} > 10^7 M_{\odot}$ ). They are, however, minority contributors to the local stellar mass and luminosity densities. For  $M_{\star} > 10^8 M_{\odot}$  ( $M_{\star} > 10^7 M_{\odot}$ ), the LSBGs contribute 7 (11) per cent and 6 (10) per cent to the stellar mass and luminosity densities respectively.

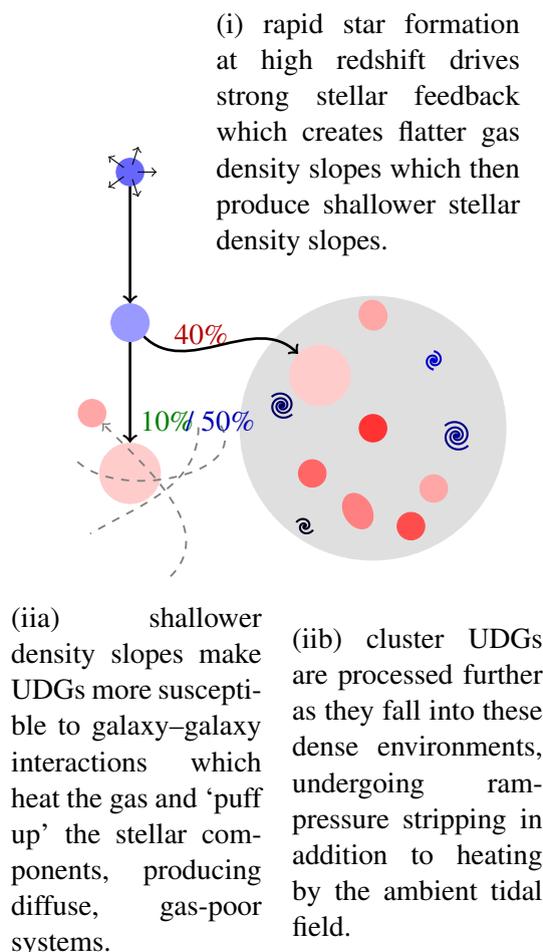


FIGURE 6.18: A summary of the formation mechanisms of our sample of UDGs ( $M_{\star} > 2 \times 10^8 M_{\odot}$ , Section 6.3.2). 40 per cent of these galaxies are found in high-density (cluster) environments at  $z = 0$ , while 10 and 50 per cent are found low density (field) and intermediate density (group) environments, as indicated by the text next to each arrow.

- *Local LSBGs have similar dark matter fractions and angular momenta as their HSB counterparts but exhibit larger effective radii ( $2.5\times$  for UDGs), older stellar populations ( $1.6\times$  for UDGs), lower gas fractions (no star-forming gas remaining in most UDGs) and shallower density profiles.*
- *LSBGs evolve from the same progenitor population as HSBGs at high redshift. HSBGs and LSBGs originate from populations with almost identical gas fractions and effective radii at  $z = 3$  and evolve along the same locii in the  $f_{gas} - R_{eff}$  plane. However, the evolution of LSBGs (and UDGs in particular) is much more rapid, especially at  $z < 1$ .*
- *UDGs experience more rapid star formation between  $z = 3$  and  $z = 1$ , which triggers their creation and ultimate divergence from the HSB population. More rapid star formation in UDG progenitors produces more concentrated SN feedback which, in turn, leads to shallower gas density profiles at high redshift ( $z > 1$ ) without quenching star formation. The star formation fuelled by this gas*

then produces systems which have shallow stellar density slopes (and larger effective radii). These systems are more susceptible to processes like tidal heating of both stars and gas by the ambient tidal field, and ram-pressure stripping of gas in denser environments.

- *External processes (tidal perturbations and ram-pressure stripping) that drive most of the evolution of LSBGs are principally effective at low and intermediate redshifts.* At  $z < 1$ , the total and star-forming gas fractions and effective radii of LSBGs, and UDGs in particular, change drastically after fairly gradual evolution between  $z = 3$  and  $z = 1$ .
- *Tidal heating (regardless of local environment) is able to produce the large sizes and low star-forming gas fractions of today's UDGs.* Flattened density profiles, produced via stronger SN feedback, are amplified by the ambient tidal field, further broadening the stellar distributions. UDGs, regardless of environment, undergo tidal perturbations of similar magnitude, with field UDGs exhibiting similar effective radii to their group/cluster counterparts at the present day. In a similar vein, tidal heating is also able to prevent gas from forming stars in UDG progenitors, regardless of their local environment. Even in field environments, where field UDGs remain star-forming down to low redshift, the tidal field is able to continually heat the gas in a large number of these systems, effectively quenching their star formation by  $z = 0.25$ .
- *In clusters, ram-pressure stripping is a significant additional mechanism that removes gas from in-falling UDG progenitors, starting around  $z = 1$ .* Although ram-pressure stripping is very effective at stripping gas in dense environments, it acts as a secondary mechanism to tidal heating outside of these environments, for creating the low fractions of *star-forming* gas found in UDGs at the present day. Our analysis shows that tidal heating would likely produce the low gas fractions found in cluster UDGs, even in the absence of ram-pressure stripping.

Figure 6.18 shows a summary of the evolutionary channels for LSBG formation described above. Our results offer insights into the formation of galaxies in the LSB regime which, given their dominance of the galaxy number densities, are essential pieces of the puzzle of galaxy evolution. Furthermore, as we have demonstrated in the analysis above, LSBGs are much more sensitive tracers of key processes that shape galaxy evolution (e.g. SN feedback, tidal perturbations and ram-pressure stripping) than their HSB counterparts. Without an understanding of the formation and evolution of LSBGs, therefore, our comprehension of galaxy evolution remains incomplete.

The new era of deep-wide surveys like the Hyper Suprime Cam Subaru Strategic Program (HSC-SSP), and forthcoming datasets from instruments like LSST, Euclid and WFIRST will revolutionize the study of LSBGs, by yielding statistical samples of these systems, for the first time, across all environments.

These datasets will enable us to perform the first statistical census of LSBG properties and their evolution with redshift, producing stringent tests of current theoretical predictions, such as those presented in this study. Together, this will create a platform for constructing a new generation of cosmological simulations, which offer a better understanding of processes (e.g. SN feedback, ram pressure stripping and tidal perturbations) to which the LSBG population is particularly sensitive, and a better reproduction of galaxies in the as-yet-unexplored LSB regime. This convergence of deep-wide surveys and cosmological hydrodynamical simulations is likely to have a transformational impact on our understanding of galaxy evolution in the coming years.

## Chapter 7

# Galaxy morphological classification in deep-wide surveys via unsupervised machine learning

### 7.1 Introduction

The measurement of galaxy morphology is a fundamental topic in observational cosmology. Morphology is a strong function of the dynamical state of a galaxy, encodes the physical processes that dominate its evolutionary history (e.g. Martin et al., 2018b) and is strongly aligned with physical properties like stellar mass (e.g. Bundy et al., 2005), star-formation rate (e.g. Bluck et al., 2014; Smethurst et al., 2015), colour (e.g. Strateva et al., 2001; Skibba et al., 2009) and local environment (e.g. Dressler, 1980; Dressler et al., 1997; Postman et al., 2005). For example, bulge-dominated galaxies typically have assembly histories that are richer in mergers (e.g. Conselice, 2006), with the strength of the bulge correlating with the number of mergers (e.g. Hatton et al., 2003). In comparison, the presence of a disc at the present day is a signature of a more quiescent formation history, with the buildup of stellar mass likely to be driven mainly by gas accretion and secular processes (Codis et al., 2012; Kaviraj, 2014a; Martin et al., 2018c). In a similar vein, at a given stellar mass, lower surface brightnesses or redder colours may indicate a larger role of tidal processes, like galaxy interactions or ram-pressure stripping, in the evolution of the galaxy in question (e.g. Dressler, 1980; Moore et al., 1999; Weisz et al., 2011; Martin et al., 2019). Finally, morphological details, such as extended tidal features, are signposts of recent mergers and/or strong interactions (e.g. Kaviraj, 2014a,b), with the surface-brightness of these tidal features typically scaling with the mass ratios of the mergers in question (e.g. Peirani et al., 2010; Kaviraj, 2010).

In addition to its key role in the study of galaxy evolution, morphological information is useful for a vast array of astrophysical science. For example, it is used as a prior in photometric-redshift pipelines (e.g. Soo et al., 2018; Menou, 2018), forms key contextual data in the classification of transient lightcurves

(e.g. Djorgovski et al., 2012; Wollaeger et al., 2018) and is important for identifying the processes that trigger the onset of AGN activity in galaxies (e.g. Schawinski et al., 2014). The measurement of accurate galaxy morphologies, particularly in large surveys which underpin our statistical endeavour is, therefore, a critical exercise.

Over the past few decades a rich literature has emerged on methods for measuring galaxy morphology, especially in large observational surveys. These methods range from parametric techniques, which attempt to describe galaxy light profiles using small sets of parameters (e.g. Sérsic, 1963; Simard et al., 2002; Odewahn et al., 2002; Lackner & Gunn, 2012), to non-parametric methods that reduce these light distributions to single values such as in the ‘CAS’ parameters (e.g. Abraham et al., 1994; Conselice, 2003; Menanteau et al., 2006), the Gini- $M_{20}$  coefficients (e.g. Lotz et al., 2004; Scarlata et al., 2007; Peth et al., 2016) or other non-parametric statistics such as the MID system (e.g. Freeman et al., 2013). Recent work has increasingly harnessed the power of machine-learning to perform morphological analysis. Although the use of machine-learning in astronomy can be traced back at least as far as Lahav et al. (1995), the recent literature has seen an explosion in the use of such techniques applied to a wide variety of problems in astrophysics (e.g. Huertas-Company et al., 2015a; Ostrovski et al., 2017; Schawinski et al., 2017; Hocking et al., 2018; Goulding et al., 2018; D’Isanto & Polsterer, 2018; Siudek et al., 2018; An et al., 2018; Ay et al., 2019).

While automated classification techniques, such as the ones described above, are particularly well-suited to efficiently processing large survey datasets, they are typically benchmarked against visual inspection (e.g. Kaviraj, 2010; Lintott et al., 2011; Simmons et al., 2017), which produces arguably the most powerful and accurate measures of galaxy morphology. While time-consuming to perform, the development of the Galaxy Zoo (GZ) platform has, in recent years, revolutionized the collection of visual classifications for large surveys. Using more than a million citizen-science volunteers, GZ has classified several contemporary surveys like the SDSS and the HST legacy surveys (e.g. Lintott et al., 2011; Willett et al., 2017). Automated methods, especially those that exploit machine-learning, have been routinely benchmarked against visual classifications from databases like GZ, and are now commonly deployed on large-scale survey data (e.g. Huertas-Company et al., 2015a; Dieleman et al., 2015; Beck et al., 2018; Walsley et al., 2019; Ma et al., 2019). In some cases, these automated methods may produce higher-quality classifications than those of expert or citizen-science classifiers alone (Cabrera-Vives et al., 2018).

Notwithstanding the variety of techniques on offer, forthcoming ‘Big Data’ surveys, e.g. LSST (Abell et al., 2009), present unprecedented challenges for performing morphological classification (Robertson et al., 2017). The sheer volume of data makes such surveys intractable for visual inspection, even via massively-distributed platforms like GZ. New techniques, which either combine visual and automated classification (e.g. Beck et al., 2018; Dickinson et al., 2019) or perhaps remove the need for visual classification altogether (e.g. Siudek et al., 2018; Hocking et al., 2018; Hendel et al., 2018; D’Isanto et al., 2018), will be crucial in dealing with the unprecedented data volumes expected from these new surveys. The short cadence of rapidly-changing datasets like LSST represents an additional hurdle and could

make supervised machine-learning techniques challenging to deploy, as it may become impractical to repeatedly produce large, reliable training sets on short timescales, as the survey becomes progressively deeper.

*Unsupervised* machine-learning (UML) algorithms are ideally suited to the morphological analysis of surveys like LSST. Unsupervised techniques do not require visually-classified training sets and can, in principle, autonomously compress an arbitrarily large galaxy population into a small number of ‘morphological clusters’ comprised of galaxies with similar properties (e.g. Hocking et al., 2018). These groups can then be benchmarked against visual classification which, if the number of groups is relatively small, becomes tractable even for individual researchers (and can be tackled easily using distributed systems like GZ).

In this chapter, we employ such a UML algorithm, which works by grouping pixels with similar properties and objects constructed from those pixels, like galaxies. Originally developed using HST data from the CANDELS survey (Hocking et al., 2017; Hocking et al., 2018), we apply the algorithm to the Ultradeep layer of the Hyper Suprime-Cam Subaru-Strategic Program (HSC-SSP) Data Release 1 (DR1). We release a catalog of morphological classifications which can be used in conjunction with the HSC-SSP DR1 catalog, explore the robustness of these classifications and discuss the applicability of the algorithm to surveys from forthcoming instruments like LSST (whose deep-wide-fast dataset will reach the same depth as the HSC-SSP Ultradeep survey after  $\sim 10$  years).

This chapter is structured as follows. In Section 7.2, we describe the UML algorithm that underpins this study. In Section 7.3, we outline the properties of the HSC-SSP and ancillary data used in this study. In Section 7.4, we describe the benchmarking of the UML outputs using visual classification, the completeness of the resultant morphological catalogue and the contents of the released data products. In Section 7.5, we explore the robustness of the classifications, by comparing the properties of galaxies in different morphological groups compare to known trends in these properties as a function of morphology, at  $z < 1$ . We summarise our results in Section 7.6.

## 7.2 The Hyper Suprime Cam Subaru Strategic Program (HSC-SSP)

### 7.2.1 Survey description

The HSC-SSP (Aihara et al., 2018a) is a multi-layered imaging survey in *grizy* (and 4 narrow-band filters), using the Hyper Suprime-Cam (HSC, Miyazaki et al., 2012) on the 8.2m Subaru Telescope. HSC has a 1.5 degree diameter field of view and a 0.168 arcsec pixel scale, with a median *i*-band seeing of  $\sim 0.6$  arcsec. The survey, which began in 2014, is being carried out using 300 nights over 5-6 years. The fields are chosen to be low in Galactic dust extinction and to have overlap with several well-known multi-wavelength data-sets including SDSS/BOSS (Eisenstein et al., 2011), X-ray surveys from XMM

(Jansen et al., 2001) and eROSITA (Merloni et al., 2012) and near-/mid-infrared imaging surveys e.g. VIKING/VIDEO (Jarvis et al., 2013) and UKIDSS (Lawrence et al., 2007).

The final HSC-SSP dataset (expected in 2021) will provide three layers: a ‘Wide’ layer covering an area of  $1400 \text{ deg}^2$  with a target  $i$ -band depth of  $26.2 \text{ mag arcsec}^{-2}$ , a ‘Deep’ layer covering an area of  $27 \text{ deg}^2$  with a target  $i$ -band depth of  $27.1 \text{ mag arcsec}^{-2}$  and an ‘Ultradeep’ layer covering an area of  $3.5 \text{ deg}^2$  with a target  $i$ -band depth of  $27.7 \text{ mag arcsec}^{-2}$  (Aihara et al., 2018a). The layers are nested, so that the Ultradeep layers are included in the Deep fields and the Deep regions are included in the Wide fields.

Here we use the HSC-SSP DR1<sup>1</sup>, which has released  $108 \text{ deg}^2$ ,  $26 \text{ deg}^2$ , and  $3.5 \text{ deg}^2$  in the Wide, Deep and Ultradeep layers, with current depths of  $i \sim 26.4$ ,  $\sim 26.5$ , and  $\sim 27$ . mag, respectively ( $5\sigma$  for point sources) (Aihara et al., 2018b). The survey is split into a number of  $1.5 \text{ deg}$  wide square ‘tracts’, each covering approximately a single HSC pointing. Each tract is further separated into  $9 \times 9$  patches, consisting of approximately  $4200 \times 4200$  pixels. Here we use stacked, sky-subtracted images, with WCS co-ordinate corrections applied and calibrated magnitude zero-points.

## 7.2.2 Data

For object centroids and observed colours and photometry, we use `cModel` magnitudes (Stoughton et al., 2002), which are released as part of the HSC-SSP DR1 forced catalogue. These are computed using the HSC-SSP reduction pipeline, using the  $i$ -band as the primary reference wavelength. We additionally calculate surface brightnesses using the Kron radius, by dividing the flux within this radius by the area of the aperture.

In order to infer physical properties and photometric redshifts for the galaxies in our sample, we use results from the MIZUKI (Tanaka, 2015) template-fitting code, that have been released as part of the HSC-SSP DR1 (Tanaka et al., 2018). Redshifts are derived purely from HSC  $g$ ,  $r$ ,  $i$ ,  $z$  and  $y$  band `cModel` magnitudes, for all primary objects detected in at least three bands. The MIZUKI code uses spectral energy distribution (SED) fitting to templates generated from the Bruzual & Charlot (2003) stellar population synthesis models, in order to self-consistently estimate redshifts and physical properties of individual galaxies. Redshift-dependent Bayesian priors are applied to physical parameters like stellar mass and star-formation rate (SFR). We use values from the public HSC-SSP DR1 `photoz_mizuki` catalogue for photometric redshifts, SFRs, stellar masses and rest-frame magnitudes and colours. For full details of the HSC-SSP DR1 we direct readers to (Aihara et al., 2018b).

<sup>1</sup><https://hsc.mtk.nao.ac.jp/ssp/data-release/>

TABLE 7.1: Parameters used for the feature extraction step.

Parameter	Description	Value
$r$	Side length of a square sub-image patch in pixels	16
$n$	Number of bins in the radial power spectrum	8

### 7.3 An unsupervised machine-learning algorithm for morphological classification

The UML algorithm that underpins this study is described in detail in Hocking et al. (2017) and Hocking et al. (2018). In the following sections, we describe the main components of the algorithm.

#### 7.3.1 Feature selection

The ultimate aim of the method is to automatically identify different groups of galaxies using HSC pixel data. Although the source data may be used directly, it is more useful to transform the data in a way that removes any irrelevant information. As our aim is to morphologically classify galaxies, we therefore transform the data so that irrelevant information like galaxy orientation is removed.

To do this we first extract  $r \times r$  pixel sub-image patches around each detected pixel in each HSC tract, where  $r$  is the patch size.

In order to reduce the time that the algorithm takes to run, and to avoid including pixels that contain no useful information, we only extract pixels that are  $1\sigma$  above the noise level, determined by a simple sigma-clipping (e.g. as implemented by Bertin & Arnouts, 1996; Astropy Collaboration et al., 2013). See Appendix D.1 for a discussion of potential improvements.

Following the initial detection and extraction step, we produce a rotationally invariant representation of each patch, by evaluating the radially averaged power spectrum, with  $n$  bins for each of the five bands ( $g, r, i, z, y$ ). It is, therefore, important that the patch size is large enough to sample the spatial scales over which the data varies (e.g. that it is larger than the PSF). Each  $n$  element power spectrum is concatenated into a  $5 \times n$  element feature vector,  $\mathbf{p}$ , which effectively encodes pixel intensity, colour and spatial frequency information for each sub-image patch in a rotationally invariant manner. Each feature vector is then combined into a patch data matrix,  $\mathbf{P}$ , which contains the feature vectors for every patch. Table 7.1 presents the values of  $r$  and  $n$  used for this feature extraction step.

TABLE 7.2: Parameters used for the growing neural gas (GNG)s, hierarchical clustering (HC)s and morphological classification steps.

Parameter	Description	Value
$N$	Maximum number of nodes in the graph	200,000
$\lambda$	Samples processed before new node added	100
$a_{max}$	Maximum age before an edge is removed	50
$\epsilon_b$	Size of the adjustment in step (i)	0.2
$\epsilon_n$	Size of adjustment for neighbours in step (i)	0.006
$\alpha$	Error reduction to node with the largest error	0.5
$\beta$	Error reduction to all nodes	0.995
$N_g$	Target number of HC groups	1500
$k$	Number of groups produced by $k$ -means	160

### 7.3.2 Feature extraction

The next step is to use clustering methods to first learn an accurate topological map (model) of the patch data matrix,  $\mathbf{P}$ , and then sub-divide the nodes within this map into coarser groups of feature vectors, thus producing a library of distinct ‘patch types’.

#### 7.3.2.1 Growing neural gas

We use a growing neural gas (GNG, Fritzke, 1995) algorithm to learn the optimal representation of the data, based on the patch data matrix,  $\mathbf{P}$ . The GNG algorithm produces a graph representation of the data, by iteratively growing a graph of nodes with *topological* neighbouring nodes in the graph connected by edges. The result is a topology-preserving map with an induced Delaunay triangulation (Okabe et al., 2009). Edges that are no longer part of the induced Delaunay triangulation must, however, be removed. This is achieved by removing edges that have reached a given *age*,  $a_{max}$ , without being connected to another node. The GNG algorithm is applied to  $\mathbf{P}$  using the following steps:

1. First, two nodes are initialized with positions using two randomly selected feature vectors from the patch data matrix,  $\mathbf{P}$ . Each node is, therefore, located within a  $5 \times n$  dimensional feature space with the same dimensionality as the number of elements of  $\mathbf{p}$ . A new random feature vector,  $\mathbf{p}'$ , is then drawn from  $\mathbf{P}$  and the following steps applied:
  - The two nearest nodes to the feature vector, whose positions in the feature space we designate  $\mathbf{s}_0$  and  $\mathbf{s}_1$ , are identified such that the Euclidean distance from  $\mathbf{p}'$ , is minimised.  $\mathbf{s}_0$  is the closest node to  $\mathbf{p}'$  and  $\mathbf{s}_1$  is the second closest.
  - If an edge connecting the two nodes,  $\mathbf{s}_0$  and  $\mathbf{s}_1$ , does not already exist it is created. The two connected nodes are called *topological* neighbours. Whenever two nodes are connected by an edge, the edge is also assigned an age,  $a$ , which is initially set to 0, and the age of all other edges connected to  $\mathbf{s}_0$  are incremented by 1.

- The closest node to  $\mathbf{p}'$ ,  $\mathbf{s}_0$ , is assigned an error equal to the square of their separation:

$$\sigma(\mathbf{s}_0) = \|\mathbf{p}' - \mathbf{s}_0\|^2. \quad (7.1)$$

- $\mathbf{s}_0$  and its direct topological neighbours (i.e. those directly connected by edges) are all moved towards  $\mathbf{p}'$  by a fraction ( $\varepsilon_b$  and  $\varepsilon_n$  respectively) of their separation from  $\mathbf{p}'$ , thus causing adaptation of the map towards the input data:

$$\begin{aligned} \Delta \mathbf{s}_0 &= \varepsilon_b(\mathbf{p}' - \mathbf{s}_0) \\ \Delta \mathbf{s}_n &= \varepsilon_n(\mathbf{p}' - \mathbf{s}_n). \end{aligned} \quad (7.2)$$

- All edges with ages larger than the maximum age (where  $a > a_{max}$ ) are removed. Any nodes that no longer have topological neighbours are also removed.

2. This procedure is repeated until  $\lambda$  feature vectors have been processed, after which:

- A new node,  $\mathbf{s}_r$ , is inserted at the mid-point between the node with the highest error,  $\mathbf{s}_q$ , and its highest error topological neighbour,  $\mathbf{s}_f$ .
- The edges connecting the two nodes are removed and new edges are created connecting  $\mathbf{s}_q$  and  $\mathbf{s}_f$  to  $\mathbf{s}_r$ .
- The error of  $\mathbf{s}_q$  and  $\mathbf{s}_f$  is decreased by multiplying their errors with the parameter,  $\alpha$ , and the error of  $\mathbf{s}_r$  is initialised with the same error as  $\mathbf{s}_q$ .
- The error of every node is decreased by multiplying their errors with the parameter  $\beta$ .

3. This is continued until the stopping criterion is met (i.e.  $N$  nodes has been reached).

The accumulation of errors in step (ii) ensures that the algorithm places new nodes in areas of the parameter space where the mapping from the model to the data is poor. Once the stopping criterion is met, we take a matrix containing the final positions of all the nodes within the feature space,  $\mathbf{N}$ , as the output. Table 7.2 presents the values of the parameters used for the GNG step. We note that the exact value of these parameters is not important for the outcome, but does affect the time it takes for the graph to converge. Any sensible choice of parameters will always result in adaptation towards the input data, but a poor choice of parameters may result in inefficient performance, requiring a large number of iterations to finish.

### 7.3.2.2 Hierarchical clustering

Agglomerative ('bottom-up') hierarchical clustering (HC; Johnson, 1967) of the GNG output is used to produce a hierarchical representation of the nodes in the topological map. At each iteration, the HC algorithm initially tries to cluster the most similar nodes into pairs, with similarity measured, in this

case, by the Pearson correlation. The Pearson correlation between the positions of nodes  $\mathbf{a}$  and  $\mathbf{b}$  in the feature space is given by their co-variance divided by the product of their standard deviations:

$$\rho(\mathbf{a}, \mathbf{b}) = \frac{\text{cov}(\mathbf{a}, \mathbf{b})}{\sigma_{\mathbf{a}}\sigma_{\mathbf{b}}}, \quad (7.3)$$

where  $\text{cov}(\mathbf{a}, \mathbf{b})$  is the co-variance between the two node position vectors and  $\sigma_{\mathbf{a}}$  and  $\sigma_{\mathbf{b}}$  are the standard deviations of the two position vectors.

At each subsequent iteration the algorithm merges clusters into pairs of similar clusters and so on, until only a single cluster remains. A particular advantage of this method is that it enables us to select the desired level of detail that we use to segment the GNG graph. The clusters can have disparate sizes and separations and therefore the method makes no assumptions about the structure of the data.

### 7.3.3 Constructing feature vectors

After a library of patch types has been produced by the GNG algorithm and then reduced via HC, it is possible to construct *object* feature vectors. Individual patches must be assembled into objects, either using existing detection maps or, as we use in this case, connected component labelling algorithms (e.g. Galler & Fisher, 1964). They can then be described using a histogram of patch types i.e. an object feature vector. The feature vector describes the frequency of different patch types that the object consists of, thereby encoding an easily manipulated description of that object.

The number of groups that patch types are clustered into, and therefore the length of the feature vector, can be changed according to the complexity of the data that is being classified. In this case it has a value  $N_g = 1500$ , equal to the number of clusters produced by the HC algorithm. The feature vector of an object should, therefore, encode the basic visual characteristics of that object, making it possible to identify visually similar objects. Fig 7.1 illustrates the process of extracting patches from multi-band survey data (Section 7.3.1), assembling a library of patch types (Section 7.3.2) and, finally, constructing feature vectors for each object (Section 7.3.3).

### 7.3.4 Producing morphological clusters

In order to finally classify galaxies into morphological clusters, we first define the similarity between feature vectors. Again, we use the Pearson correlation (Equation 7.3), in order to define the distance between feature vectors in this new feature space, although other distance measures (e.g. Euclidean distance or cosine distance) may be used and may accentuate different features. Additionally, we apply ‘term frequency-inverse document frequency’ (TF\*IDF, Rajaraman & Ullman, 2011) weightings when calculating the distance, in order to increase the importance of patch types with the greatest discriminatory power, and reduce the importance of patch types that are relatively common between all objects.

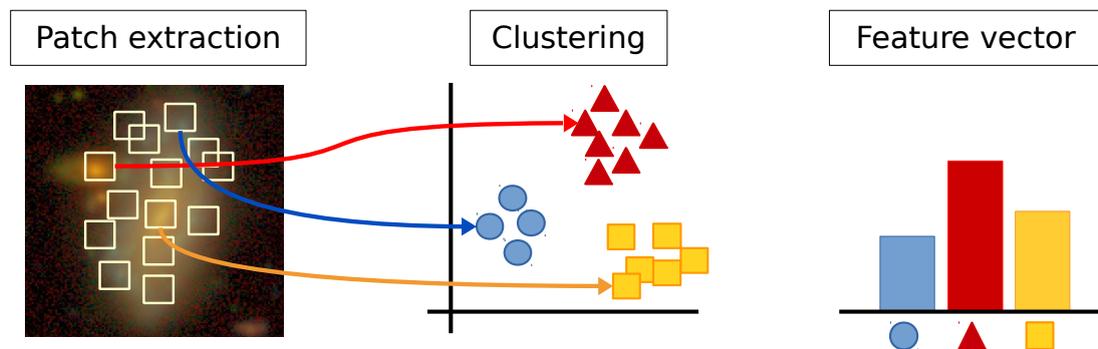


FIGURE 7.1: Cartoon view of the classification process. Patches are extracted around detected pixels in survey images and clustering methods are used to group these patches into a library of patch types. Galaxy feature vectors can then be constructed by creating a histogram for each object which describes the frequency of each patch ‘type’.

Once we have produced a feature vector that encodes the visual characteristics of each object, and defined a distance measure for these feature vectors, it is finally possible to group these objects by their visual similarity. This can be done either by direct comparison, or a similarity search, of individual feature vectors e.g. searching for other objects that are most similar, or closest in the feature space, to the feature vector of a given object, or by applying a clustering algorithm to the feature vectors in order to group them.

In order to ensure cleaner classifications, we exclude any objects that are comprised of fewer than 15 pixels. Using  $k$ -means clustering (e.g. MacQueen et al., 1967), we separate our object feature vectors into  $k$  morphological clusters (in this chapter we choose  $k$  to be 160). We calculate silhouette scores (Rousseeuw, 1987) for the objects in each morphological cluster in order to evaluate the overall quality of the clustering as well as the correspondence of individual objects to the average properties of the group they are assigned to. Silhouette scores range from -1 to 1: a high silhouette score indicates that the object is well matched to its own cluster and distinct from neighbouring clusters and a suitable value of  $k$  will produce more positive average silhouette scores. With  $k = 160$ , we obtain a mean silhouette score of 0.26.

Using the parameters described above, the algorithm takes around 40 ms per pixel in order to perform feature extraction, generate a model from training data and perform the classification. Feature extraction and classification using an existing model applied to unseen data takes only around 1-2 ms per pixel, on a single thread of execution on a contemporary desktop computer with an Intel CPU. The feature extraction and classification steps can be easily split up and executed concurrently (Herlihy & Shavit, 2011). This property makes the algorithm efficient, even on very large volumes of data (e.g. surveys like LSST or SDSS). Even without parallelisation of the extraction and classification steps, the algorithm performs well on large datasets. For example, the entire  $3.5 \text{ deg}^2$  of the HSC UltraDeep dataset used in this chapter was processed in under 40 CPU hours, including feature selection, extraction and classification. Scaling up to much larger data volumes will also be possible. For example, under the conservative assumption

that 1 per cent of the approximately  $10^{12}$  pixels that make up the SDSS are detected, the entirety of SDSS could be processed in under 3000 CPU hours (assuming a modeling/feature extraction step has already been performed). Assuming the same set of assumptions for LSST (although LSST images will have more detected pixels than SDSS due to higher depth), the smaller pixel size and larger area of LSST would require around 16,000 CPU hours.

### 7.3.5 Cross-matching the UML objects with the HSC-SSP

We cross-match the galaxy centroids from the HSC-SSP DR1 Ultradeep catalogue with the object centroids from the UML algorithm, excluding objects that do not have a match within  $0.8''$ , which is approximately the PSF of the worst HSC *i*-band seeing (note that the median *i*-band seeing is  $0.6''$ ). Of the 89,257 objects produced by the UML classification, 53,003 have more than 15 pixels, which we consider to be large enough for reliable classification, as a sufficient range of spatial scales can be captured. Of these, 41,062 (77 per cent) have centroids that match an object in the HSC catalogue within  $0.8''$ . Mismatch between centroids arises because, at present, we use a simple connected component labelling algorithm to identify objects, rather than use the individual segmentation maps used by the HSC-SSP pipeline. The mismatch becomes increasingly worse for very large objects (see Fig D.1) and is, therefore, principally a problem in the very local Universe. However, in our analysis below, we study more distant objects ( $z \gtrsim 0.3$ ), with much smaller projected sizes. The fraction of matched objects is, therefore, much larger as, on average 90 per cent of objects smaller than 100 pixels are successfully cross-matched compared with only 15 per cent of objects larger than 1000 pixels.

## 7.4 Morphological catalogue

### 7.4.1 Benchmarking of morphological clusters via visual classification

The UML algorithm effectively compresses the galaxy population into a small number of morphological clusters. Crucially, the number is small enough to make visual classification of these clusters tractable for individual researchers. To generate a usable morphological catalogue, we benchmark the outputs of the UML algorithm via visual classification of each of the  $k = 160$  morphological clusters. These classifications are based on a subset of *g-r-i* false-colour images of the 10 highest silhouette-score objects in each cluster, plus a sample of 10 objects selected at random, in order to assess the morphological purity of the cluster. We do not classify individual galaxies but perform visual classification on the cluster as a whole.

We classify each cluster into one of three broad Hubble (Hubble, 1936) morphological types: elliptical galaxies, S0/Sa galaxies or spiral galaxies. We also store finer morphological information, e.g. the type of spiral morphology (Sb, Sc, Sd) and noteworthy colour or structural features (e.g. when spirals appear

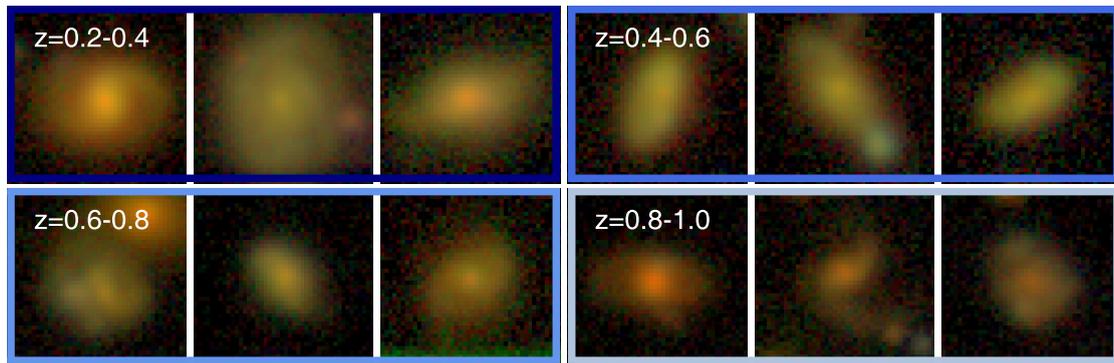
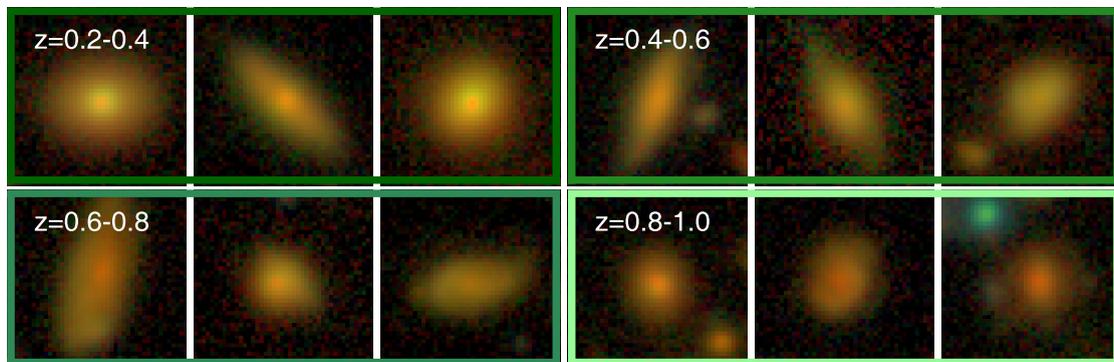
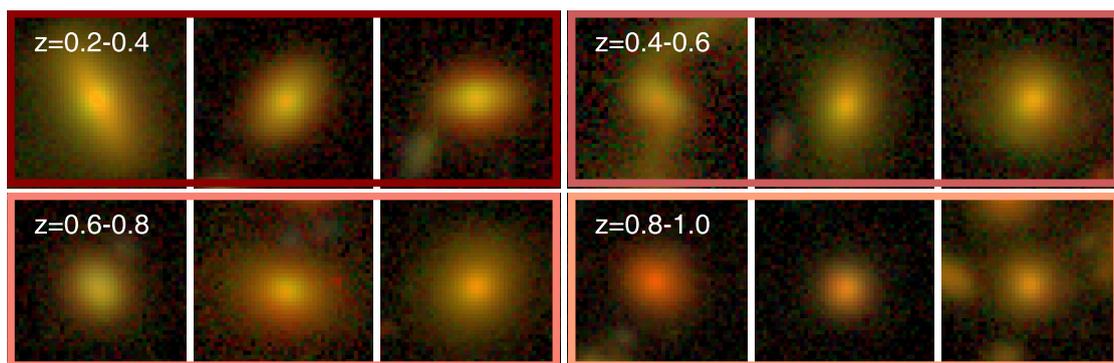
(a) *Spiral galaxies.*(b) *S0/Sa galaxies.*(c) *Elliptical galaxies.*

FIGURE 7.2:  $g-r-i$  false colour images showing a random selection of galaxies from each morphological group. The samples are further split into bins of redshift, indicated by the label in the top right of each coloured box. Panel (a) shows objects classified as spirals, panel (b) shows objects classified as S0/Sa and panel (c) shows objects classified as ellipticals.

unusually red or show clumpy structure, or when elliptical galaxies appear unusually blue). Except in Section 7.4.2, we only consider objects which are extended (based on the difference between the PSF magnitude and the `cModelMag` magnitude; Eisenstein et al., 2001). We indicate the total number of objects in each morphological cluster that are not extended in Table D.1. Fig 7.2 shows a random selection of objects that are classified as having spiral, S0/Sa and elliptical morphologies, split into four redshift ranges. Note that, although a sample of individual objects in each cluster are visually classified in order to determine a morphological type for that cluster, the majority of objects in each cluster are unseen.

Fig 7.3 shows some individual morphological clusters identified by the algorithm. For example, cluster 10 contains galaxies identified as Sc/Sd Hubble types (Fig 7.3(a)), cluster 14 is comprised of systems that appear to be high-redshift mergers (Fig 7.3(b)), cluster 122 contains galaxies which show blue ring-like features indicative of the recent accretion of gas-rich satellites (Fig 7.3(c)) and cluster 127 is composed of clumpy discs (Fig 7.3(d)). As described in Section 7.4.4 below, the visual classifications of each morphological cluster, and the average properties of objects in these clusters, are presented in Appendix D.2 and Table D.1.

## 7.4.2 Star-galaxy separation

Fig 7.4 presents a colour-colour diagram, showing the  $g - r$  and  $r - i$  colours for spirals (blue), S0/Sa (green) galaxies, ellipticals (red) and stars (orange). The stellar locus is clearly delineated, occupying a distinct region of colour-colour space compared to spirals, ellipticals and S0/Sa galaxies. The objects that are morphologically identified as stars by the UML algorithm occupy the same region as objects that are identified as being not extended by the HSC pipeline (see Section 7.4.1). The region of colour-colour space containing objects that are not extended is indicated by a black contour, which contains 95 per cent of all such objects in our sample.

It is worth noting that optical colours alone may not encode sufficient information to effectively separate stars from galaxies (Fadely et al., 2012). However, the UML algorithm employed here is able to distinguish between stars and resolved galaxies, even within the region where they share the same colours, because resolved galaxies and unresolved stars do not share the same (spatial) power spectra or distribution of patch types, and therefore, do not fall into the same morphological clusters.

We note that the relatively simple method used in Eisenstein et al. (2001) (see also Section 7.4.1) to determine extendedness is not always a good proxy for stellarity. Although star-galaxy separation has traditionally used purely morphometric information to classify stars and galaxies in optical survey data (e.g. Kron, 1980; Eisenstein et al., 2001; Henrion et al., 2011), new ground-based deep-wide surveys, which contain many more unresolved galaxies than stars at faint apparent magnitudes (Fadely et al., 2012; Soumagnac et al., 2015), represent an emerging challenge. Further work is therefore needed in order to determine whether the algorithm can effectively distinguish faint, unresolved galaxies from stars

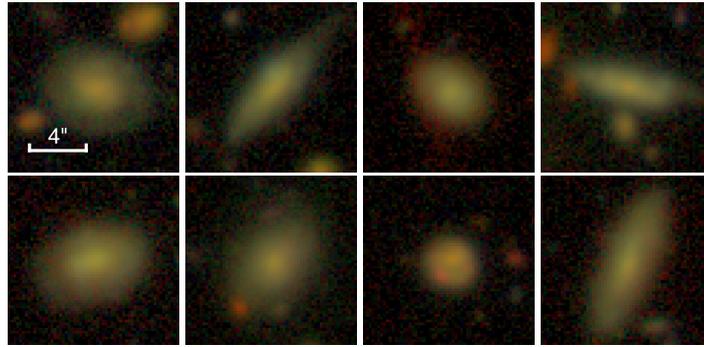
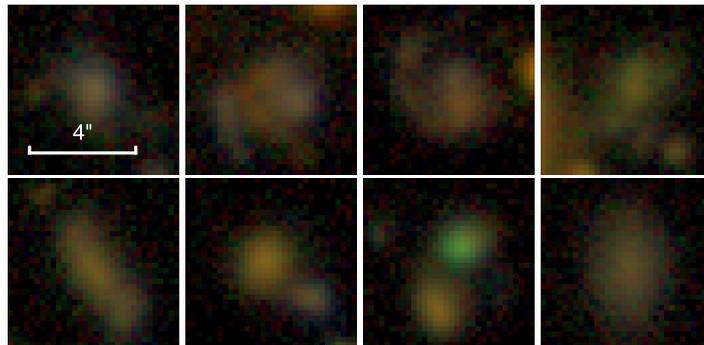
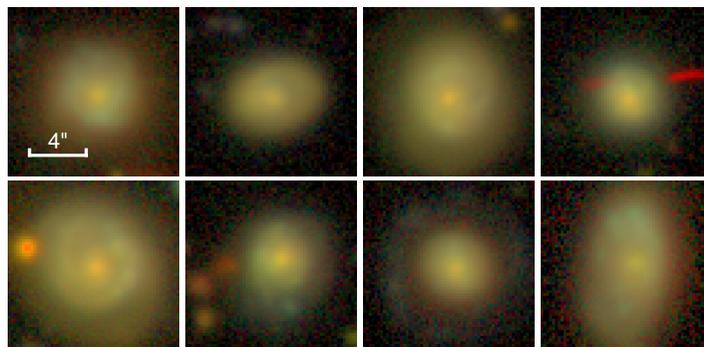
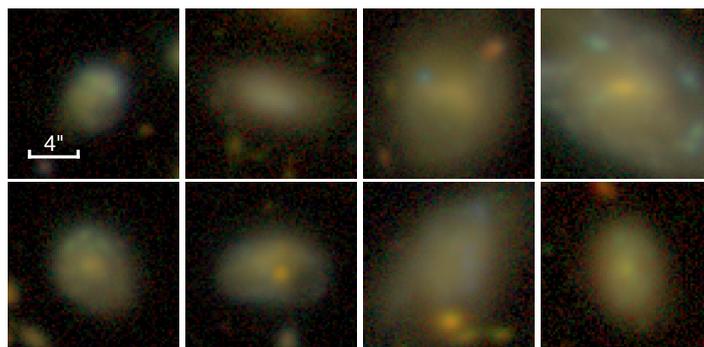
(a) Cluster 10: *Sc/Sd galaxies.*(b) Cluster 14: *High-z mergers.*(c) Cluster 122: *Disks with blue rings, possibly indicative of the recent accretion of blue satellites.*(d) Cluster 127: *Clumpy discs.*

FIGURE 7.3: Examples of interesting morphological clusters produced by the UML algorithm: (a) Sc/Sd galaxies (b) merging systems at high-redshift (c) disks that have blue ring-like structures that might be the result of the recent accretion of gas-rich satellites (d) Clumpy disks in the nearby Universe. Spatial scales are indicated by the white bar in the top-left panel of each cluster.

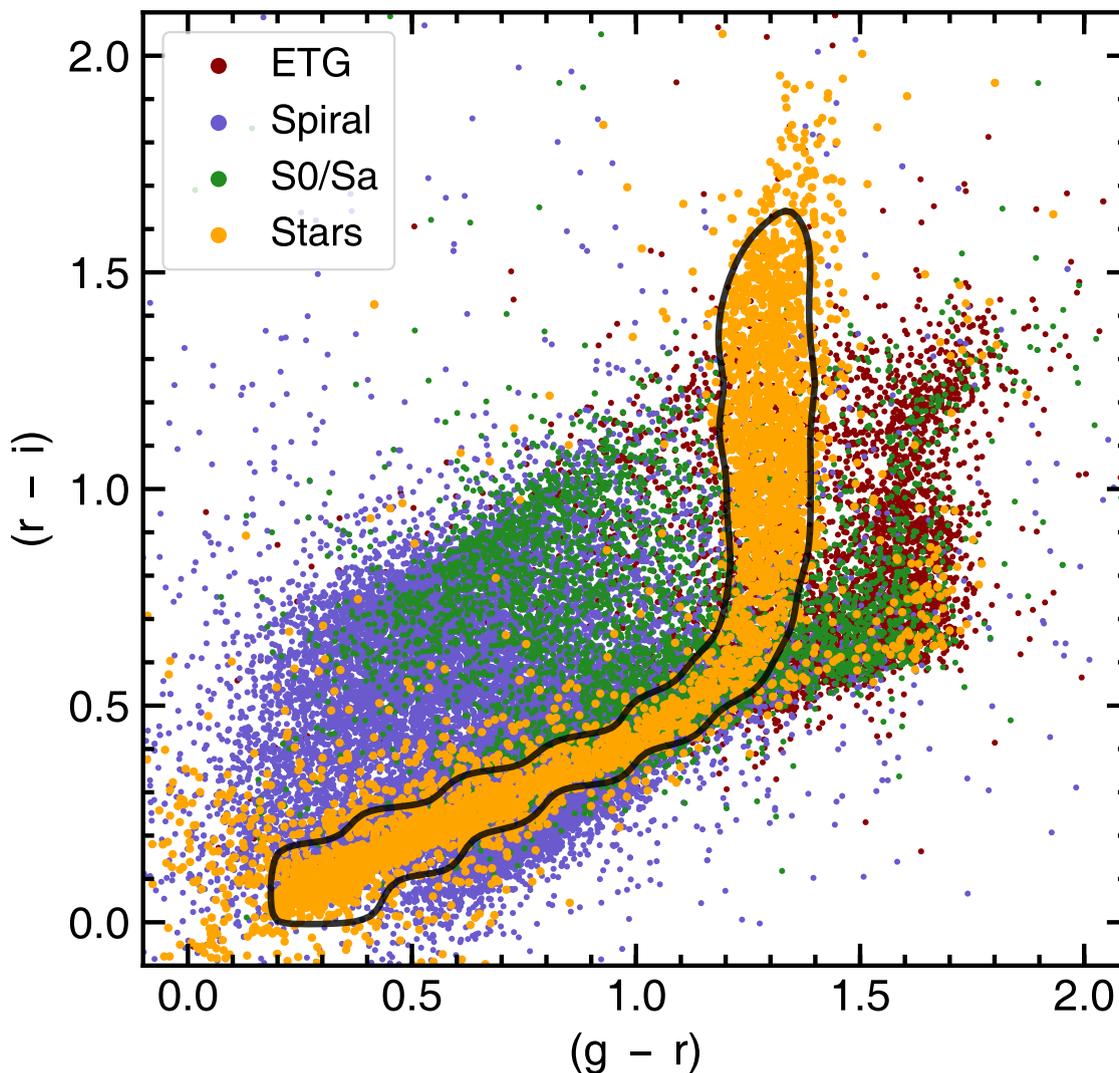


FIGURE 7.4: The positions of ellipticals (red), spirals (blue), S0/Sa galaxies (green) and stars (orange) in the rest-frame  $g-r$  vs  $r-i$  plane. The black contour contains 95 per cent of objects classified as non-extended in the HSC-SSP catalogue.

in very deep images. We note, however, that this does not affect the analysis in this study, since we are focused on bright objects.

### 7.4.3 Completeness of the UML-classified galaxy sample

In Fig 7.5, we compare the distribution of  $i$ -band magnitudes from the HSC-SSP DR1 Ultradeep survey (grey) and the distribution for objects that are large enough to be classified by the UML algorithm and then successfully matched to the Ultradeep catalogue (dark blue). The black line indicates the completeness, i.e. the fraction of all galaxies in each magnitude bin that can be classified by the UML algorithm and then matched to the HSC-SSP. The completeness values are indicated by the right-hand  $y$ -axis.

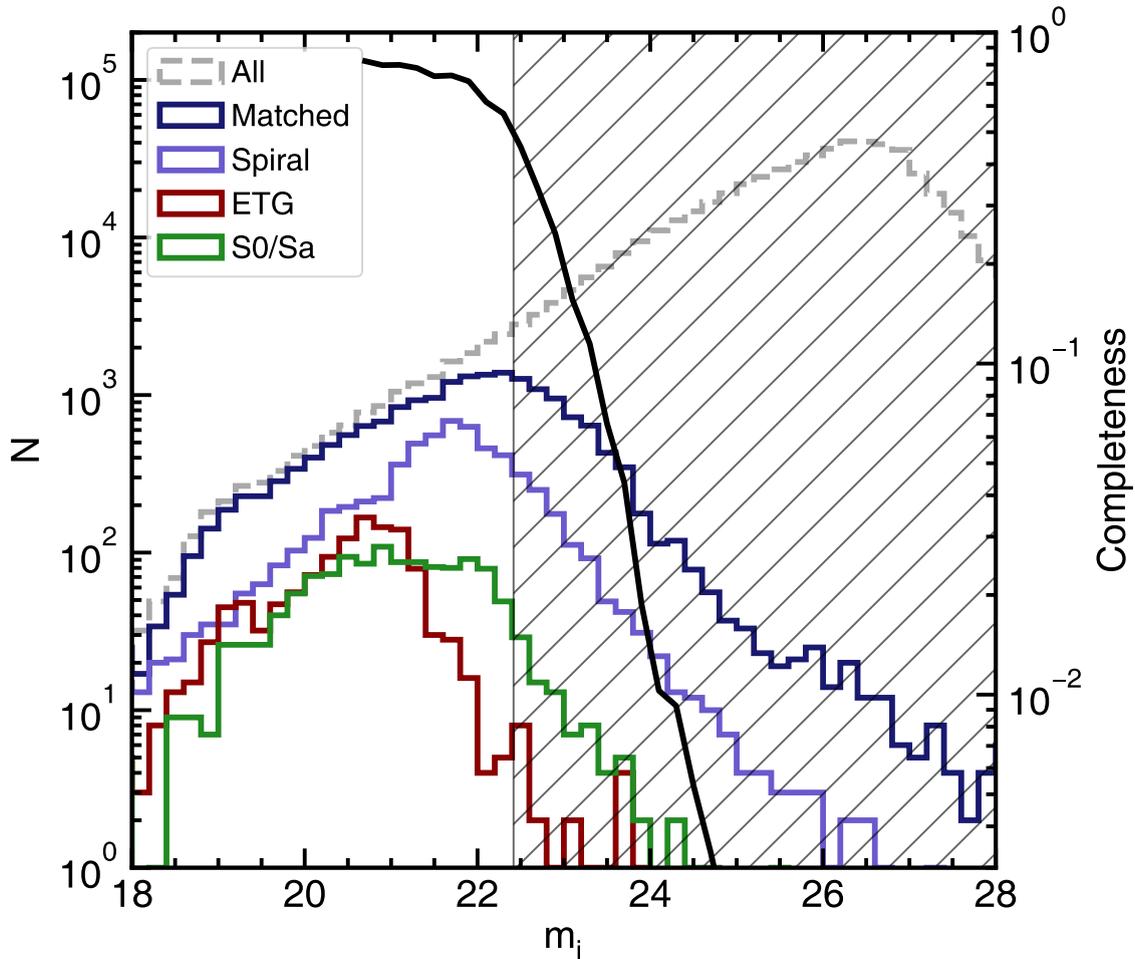
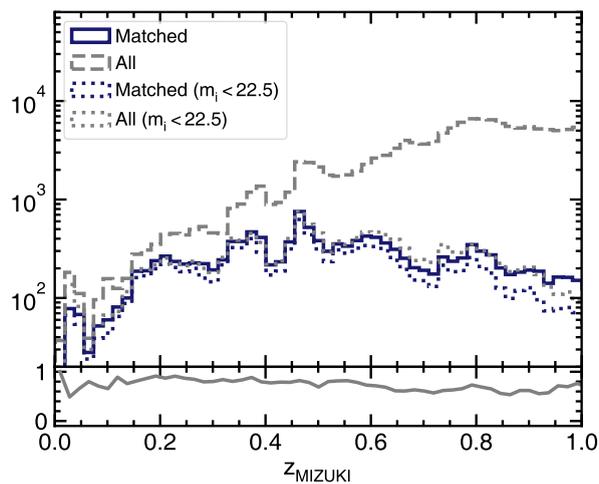


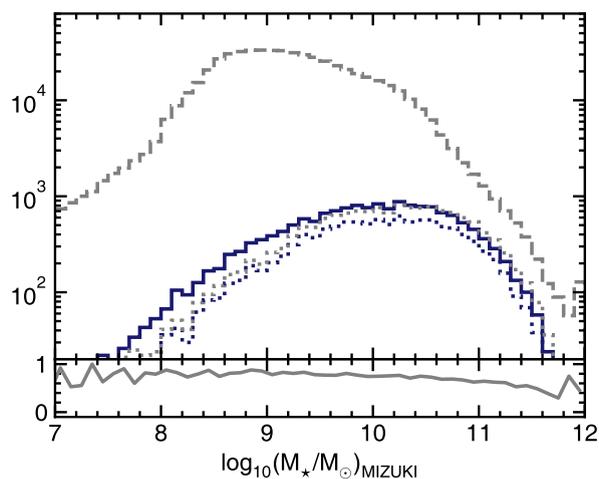
FIGURE 7.5: The  $i$ -band apparent-magnitude distribution in the HSC-SSP Ultradeep catalogue (dashed grey histogram) and the distribution of the subset of these sources that has been matched and classified by the UML algorithm (solid dark-blue histogram). Light-blue, red and green histograms indicate the distribution of objects that are identified as spirals, ellipticals or S0/Sa galaxies respectively. The black solid line indicates the completeness as a function of apparent magnitude (with values indicated on the  $y$ -axis) and the grey hatching indicates the region where the completeness falls below 50 per cent.

While the completeness of the full sample only begins to decline significantly around  $m_i > 27$  mag, a magnitude cut of  $m_i < 22.5$  mag ensures that a majority (i.e. more than 50 per cent) of objects in the Ultradeep survey have large enough sizes for robust morphological classification using the UML algorithm. We note that this cut appears to vary as a function of galaxy morphology, as demonstrated by the brighter limiting magnitudes for some morphologies, particularly for ellipticals. This is likely the result of different average projected sizes, with ellipticals being typically more compact than spiral galaxies that have similar magnitudes, particularly at low luminosities (Lange et al., 2015). In the subsequent figures in this section, we consider galaxies brighter than this  $m_i = 22.5$  mag threshold. We also show that the size criterion imposed by the UML algorithm does not produce biased galaxy populations as a result of the classification and matching procedure.

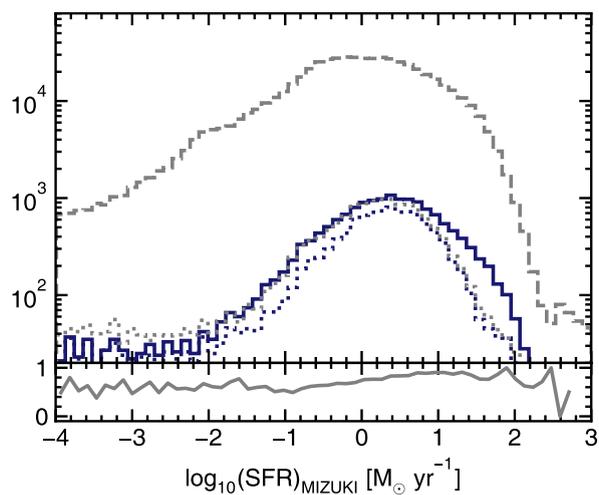
Fig 7.6(a) shows the distribution of photometric redshifts derived using the MIZUKI code, for the full Ultradeep catalogue (grey) and the matched UML catalogue (dark blue). Dotted grey and dark blue



(a)



(b)



(c)

FIGURE 7.6: Distributions of galaxy properties ((a) MIZUKI photometric redshifts, (b): stellar masses, (c): SFRs) in the full HSC-SSP Ultradeep catalogue (grey) and for the subset of galaxies that has been matched and classified by the algorithm (blue). Dotted blue and grey lines show the distributions for galaxies with  $m_i < 22.5$  mag. **Bottom:** The fraction of matched and classified objects with  $m_i < 22.5$  mag as a function of photometric redshift.

lines show the same for all galaxies with  $i$ -band absolute magnitudes brighter than 22.5 mag for the full and matched samples respectively. The lower panel shows the fraction of galaxies with  $m_i < 22.5$  mag that are matched as a function of photometric redshift. As might be expected, the matched sample, the magnitude limited matched sample and the full magnitude limited sample all share similar distributions, but the matched sample falls off more quickly compared to the full sample, as their projected sizes increase with redshift, making more objects unclassifiable.

Figs 7.6(b) and 7.6(c) show the corresponding analyses for stellar masses and SFRs respectively. Again, the histograms show the distributions of stellar masses and SFRs for the full HSC-SSP Ultradeep catalogue (grey), the distribution for matched objects only (dark blue) and the full and matched distributions for  $m_i < 22.5$  mag (shown using grey and blue dotted lines). The lower panels again show the fraction of galaxies with  $m_i < 22.5$  mag that are matched as a function of photometric redshift. While the redshift distributions of objects is influenced by the size cut, for objects brighter than the magnitude cut, the full and matched samples have very similar distributions of physical properties. This indicates that the size cut does not introduce any bias in such galaxy properties (e.g. high star-formation rates), so that a comparison of average properties as a function of redshift is possible.

#### 7.4.4 Released data products

The released data products are contained in two tables. The first table comprises a list of morphological clusters with their associated visual classifications and median values of key galaxy properties within the cluster (surface-brightness, stellar mass, specific SFR, rest-frame  $(g - r)$  colour and absolute  $r$ -band magnitude). This table is presented in its entirety in Appendix D.2. The second table is a list of individual HSC-SSP galaxies with their associated morphological cluster number and useful ancillary information, including their coordinates HSC DR1 ID, extendedness, size in pixels and silhouette score. Note that some morphological clusters can have some contamination from stars. Users should discard objects which are classified as not extended and which are, therefore, likely to be stars. The first ten rows of this table is presented in Table D.2 of Appendix D.2.

Both tables are available in electronic from the MNRAS website, and also at the following URL: [https://github.com/garrethmartin/HSC\\_UML](https://github.com/garrethmartin/HSC_UML). At the same website, we provide the feature vectors for each galaxy as well as code for performing searches on these feature vectors to find similar objects (i.e. a ‘similarity search’, Hocking et al., 2018). We plan to release morphological catalogues for future versions of HSC datasets and those from LSST on the same website).

## 7.5 Galaxy properties as a function of morphological type

In this section, we explore the robustness of the morphological classifications produced by our algorithm. We study the distributions of key galaxy properties (e.g. stellar masses, star-formation rates,

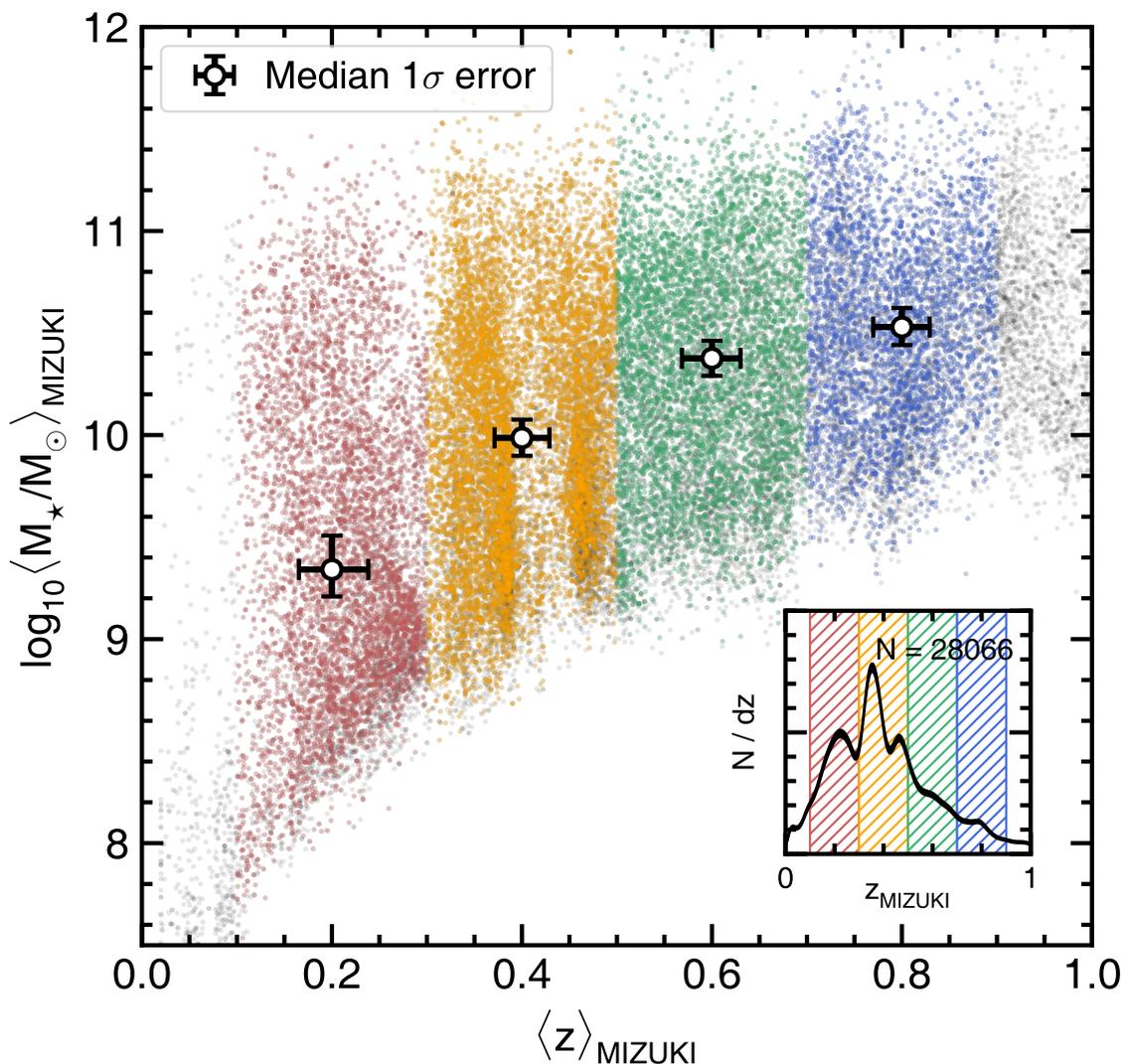


FIGURE 7.7: **Main panel:** The redshift and stellar mass distributions of our sample for  $m_i < 22.5$ . The  $x$ - and  $y$ -axes indicate the median redshift and the median stellar mass respectively, derived from the MIZUKI fits to the optical SED of each object. Error bars indicate corresponding  $1\sigma$  errors on the medians in each of the three redshift bins. **Inset:** the *unweighted* redshift distribution of our sample. The thickness of the line indicates the  $2\sigma$  confidence interval, calculated using 10,000 draws from the redshift probability distribution of each galaxy, assuming a two-sided Gaussian error around the median value. Red, orange and green hatched regions indicate the three redshift bins used.

rest-frame colours) as a function of morphological type, as a test of the veracity of our UML classifications. We demonstrate that the distributions of such galaxy properties in well-known morphological groups follow expected trends from studies using traditional visual morphological classification methods (e.g Menanteau et al., 2006; Kelvin et al., 2014; Khim et al., 2015; Willett et al., 2017).

## 7.5.1 Sample selection and methodology

### 7.5.1.1 Redshift binning

We first bin our galaxies into three redshift ranges:  $0.3 < z < 0.5$ ,  $0.5 < z < 0.7$  and  $0.7 < z < 0.9$ . A minimum redshift of  $z = 0.3$  is chosen to ensure that the  $3.5 \text{ deg}^2$  Ultradeep footprint encompasses a cosmological co-moving area greater than  $85 \text{ Mpc} \times 85 \text{ Mpc}$  in the lowest redshift bin, so that our galaxy populations are large enough to be statistically representative, and unlikely to be significantly biased by large-scale structure. For completeness, we do additionally consider a lower redshift bin ( $0.1 < z < 0.3$ ) when considering stellar mass functions in Section 7.5.2, but this bin is likely to be strongly affected by cosmic variance.

Fig 7.7 presents a scatter plot showing the distribution of median stellar masses of individual galaxies as a function of their median redshifts. Points are colour coded by their redshift bin. Open circles with error bars indicate the central redshift and median stellar mass of each redshift bin and the  $1\sigma$  error of these quantities in each bin. The inset shows the redshift distribution of galaxies. The thickness of the line indicates the  $2\sigma$  confidence interval, derived using 10,000 draws from the redshift probability distribution, which assumes a two-sided Normal error around the median redshift, with a standard deviation equal to the upper and lower redshift errors.

### 7.5.1.2 $1/V_{max}$ weighting and simulation of uncertainties

In order to correct for Malmquist bias (Malmquist, 1922), we weight galaxy counts using  $1/V_{max}$ , the inverse of the maximum volume in which it would be possible to detect an object of a given luminosity (e.g. Schmidt, 1968; Weigel et al., 2016). We do this by first making 10,000 random draws from the redshift probability distribution for each object. We assume that the probability density function (PDF) follows a two-sided Normal distribution, with a central value equal to the median MIZUKI redshift,  $\langle z \rangle$ , but with different standard deviations ( $\sigma_{upper}$  and  $\sigma_{lower}$ ) on either side of the central value. Each redshift,  $z_{draw}$ , is therefore drawn from the following distribution:

$$z_{draw} \sim \begin{cases} \mathcal{N}(z | \langle z \rangle, \sigma_{lower}^2) & \text{if } z \leq \langle z \rangle \\ \mathcal{N}(z | \langle z \rangle, \sigma_{upper}^2) \times (\sigma_{lower}/\sigma_{upper}) & \text{if } z > \langle z \rangle \end{cases} \quad (7.4)$$

where  $\mathcal{N}(z | \mu, \sigma^2)$  is a Normal distribution with a central value equal to the median MIZUKI redshift,  $\langle z \rangle$ , and a variance of  $\sigma^2$ .  $\sigma_{upper}$  is the 84th percentile of the redshift PDF, and  $\sigma_{lower}$  is the 16th percentile of the redshift PDF. The factor of  $\sigma_{lower}/\sigma_{upper}$  ensures that the distribution remains continuous.

In the MIZUKI fitting, the dominant source of uncertainty in the inferred stellar mass and absolute magnitude is the luminosity distance, rather than the model template weights or dust attenuation. The equivalent absolute magnitude,  $M_{i,draw}$ , and stellar mass,  $M_{\star,draw}$ , at a given redshift can therefore be well approximated by only varying their value by square of the ratio of the luminosity distance,  $D_L(z)$ , at the redshift of the draw and the median redshift. We can therefore calculate the new stellar mass for each drawn redshift as follows:

$$M_{\star,draw} \approx \langle M_{\star} \rangle \left[ D_L(z_{draw}) / D_L(\langle z \rangle) \right]^2 \quad (7.5)$$

and similarly for the absolute magnitude:

$$M_{i,draw} \approx \langle M_i \rangle \left[ D_L(z_{draw}) / D_L(\langle z \rangle) \right]^2. \quad (7.6)$$

We then find the maximum redshift,  $z_{max}$  at which an object with absolute magnitude  $M_{i,draw}$  will fall below the detection limit (the redshift where where the distance modulus,  $\mu$ , is equal to  $m_{lim,i} - M_{i,draw}$ ) and thus obtain  $V_{max}$ , which is proportional to the co-moving volume out to  $z_{max}$ .

$$V_{max} \propto D_c(z_{max})^3, \quad (7.7)$$

where  $D_c(z)$  is the comoving distance at  $z$ .

Note that the minimum size (15 pixels) that we impose influences the limiting apparent magnitude. Since the average size of objects at a given magnitude varies between morphological types, we use different values of  $m_{lim,i}$  when calculating  $z_{max}$ , corresponding to the limiting magnitude found for the morphological type in question (e.g. as in Fig 7.5). Since we are primarily interested in the relative distribution of galaxy properties between morphological types, rather than the exact normalisation of the number density, we do not take into account the area of the survey when calculating  $V_{max}$ .

Following the method of weighting and simulating uncertainties described above, we take draws from the redshift distribution for individual objects in each morphological sample of galaxies 10,000 times. For each of the 10,000 draws, we calculate new stellar masses and  $i$  band absolute magnitudes and thus the value of  $V_{max}$  for each galaxy. After binning our sample into four redshift bins, based on the draws from the redshift distribution (with central redshifts of 0.2 0.4, 0.6 and 0.8), we use  $1/V_{max}$  weighted univariate Gaussian kernel density estimation (e.g. Klein & Moeschberger, 2006) with a kernel bandwidth of 0.1 dex to produce a galaxy stellar mass function for each redshift bin. We use the median value and 1 and 2  $\sigma$  dispersions (defined by the central 68 and 95 per cent of values around the median) to characterise the galaxy stellar mass function and its uncertainty for each morphological type.

### 7.5.2 Stellar mass distributions as a function of morphological type

Fig 7.8 shows the evolution of the galaxy stellar mass function (left-hand column) and the evolution of the morphological fractions (right-hand column) as a function of redshift. We include a redshift bin in the range  $0.1 < z < 0.3$  for completeness, however we avoid drawing any conclusions at these epochs, as the volume of this subsample is not large enough to be statistically representative since it is likely to be strongly affected by cosmic variance.

As shown in previous work (e.g. Conselice et al., 2008; Ilbert et al., 2010; Conselice et al., 2014), there is a general trend for elliptical and S0/Sa fractions at a given stellar mass to increase towards lower redshifts. These systems increasingly dominate the number density at high stellar masses towards the present day (e.g. Wilman & Erwin, 2012; Kelvin et al., 2014), as spiral galaxies are quenched to form S0/Sa systems and/or undergo morphological transformation to form ellipticals. In the highest redshift bin ( $0.7 < z < 0.9$ ), ellipticals almost entirely dominate at masses greater than  $10^{11}M_{\odot}$ , whereas S0/Sa galaxies become more important in the same mass range towards lower redshifts.

While S0/Sa galaxies and ellipticals share similar mass functions, at least at lower redshifts, the dominance of ellipticals at high stellar mass in the early Universe indicates that a distinct, more gradual, evolutionary channel may be responsible for producing S0/Sa populations. In particular, ellipticals likely form at epochs that predate those where the mechanisms that produce S0/Sa populations (e.g. Dressler et al., 1997; Cerulo et al., 2017) are most efficient. This is likely to particularly for the most massive ellipticals, which must have formed rapidly at high or intermediate redshift (e.g. Jaffé et al., 2011; Tomczak et al., 2014b; Huertas-Company et al., 2015b).

The high-mass end of the elliptical mass function does not evolve significantly over redshift and is already in place in the highest redshift bin. The S0/Sa mass function appears instead to be built up from lower-mass systems, indicating a different evolutionary channel from their elliptical counterparts. At all redshifts, S0/Sa type galaxies typically dominate at intermediate masses, between spirals and ellipticals (e.g. Vulcani et al., 2011; Kelvin et al., 2014), with the peak of the S0/Sa fraction moving towards lower stellar masses at lower redshift.

### 7.5.3 Star-formation rates and rest-frame colours as a function of morphological type

Figs 7.9 and 7.10 show the star formation main sequence and the  $M_i$  vs. rest-frame  $g - i$  colour-magnitude diagram, for three redshift bins (with central redshifts of 0.4, 0.6 and 0.8). Contours show the density of objects weighted by  $1/V_{max}$ . Galaxies classified as spirals inhabit a well defined main sequence, while ellipticals dominate a cloud below this sequence. S0/Sa galaxies lie somewhere between these two populations. Many S0/Sa galaxies are not quenched and remain on the main locus of the star formation main sequence, with a small number lying further below. Similarly, the colour-magnitude diagram shows a clear bi-modality, with galaxies classified as ellipticals occupying the ‘red sequence’

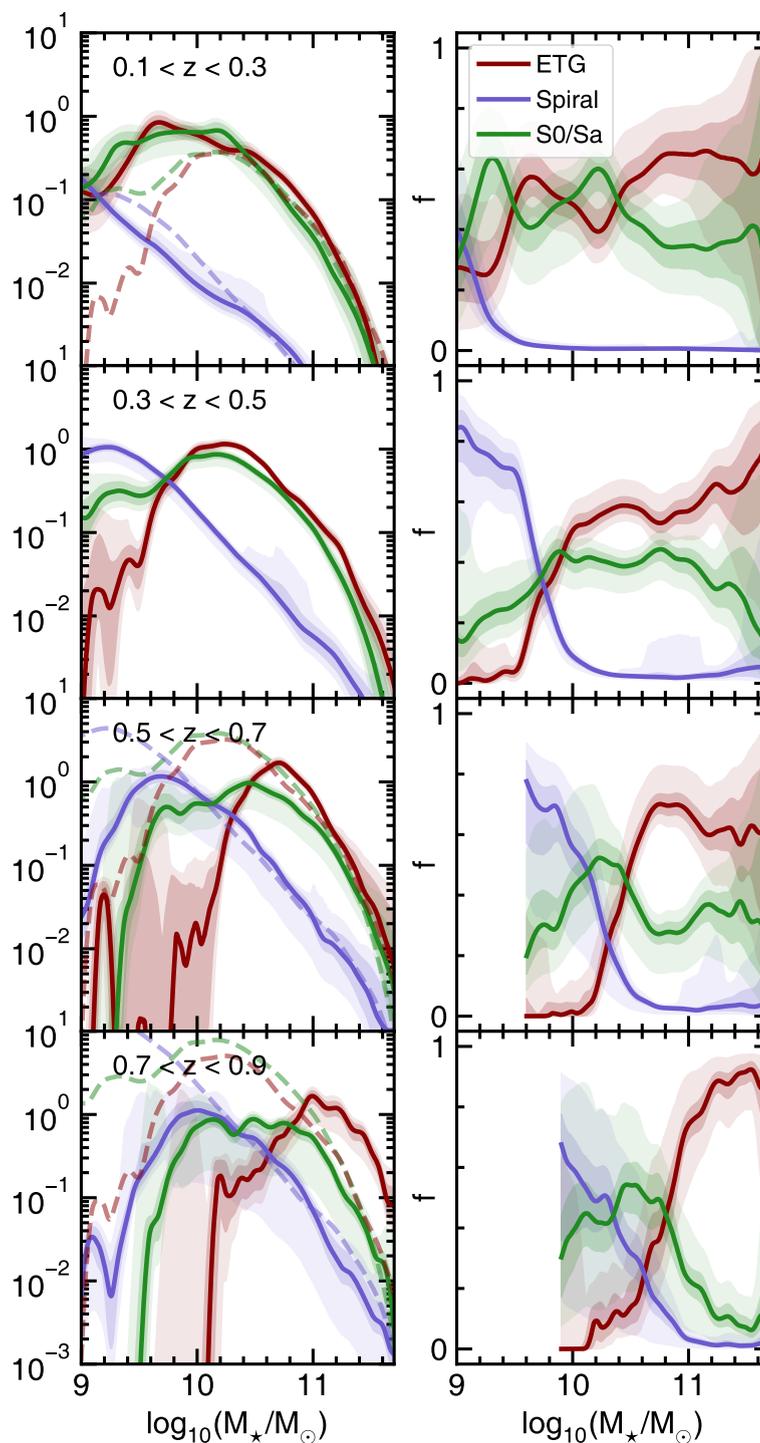


FIGURE 7.8: **Left:** Galaxy stellar mass functions for spirals (blue), S0/Sa galaxies (green) and ellipticals (red) in four redshift bins with arbitrary normalisation. Light and dark coloured regions indicating the  $1\sigma$  and  $2\sigma$  confidence intervals respectively, based on 10,000 draws from the redshift distribution of each galaxy. To enable comparison between the stellar mass functions at various redshifts, the pale dashed lines show the stellar mass function for the  $0.3 < z < 0.5$  bin, normalised to the mass function in each redshift bin. **Right:** The evolution of the spiral, S0/Sa and elliptical fractions between  $z = 0.1$  and  $z = 0.9$ . Blue, green and red lines show the fraction of galaxies that are spirals, S0/Sa and ellipticals, calculated from the galaxy stellar mass functions on the left. Light and dark coloured regions indicate the  $1\sigma$  and  $2\sigma$  confidence intervals respectively, based on 10,000 draws from the redshift distribution of each galaxy. Fractions are only plotted up to the point that the stellar mass function remains complete.

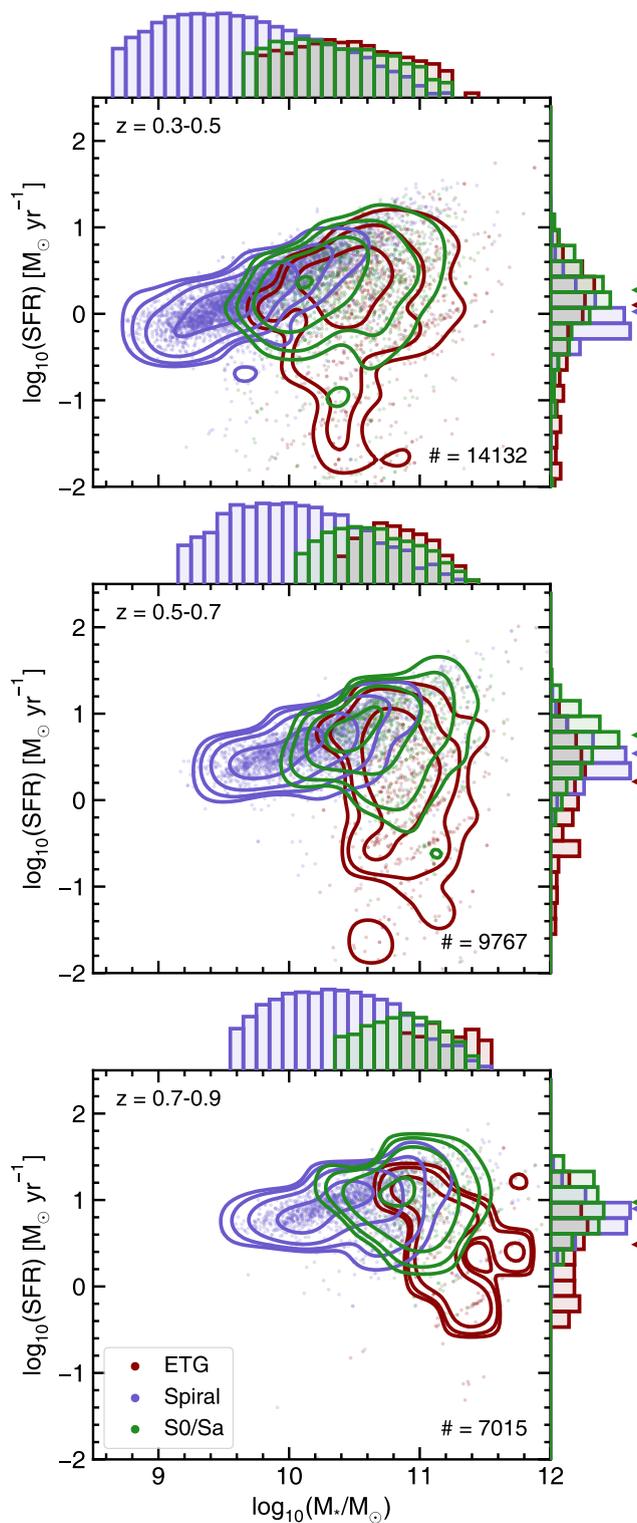


FIGURE 7.9: Scatter plots with contours overlaid, showing the distribution of galaxies as a function of SFR and stellar mass, for galaxies classified as elliptical (red), S0/Sa (green) and spiral (blue). The dots show individual galaxies, while contours show the  $1/V_{\text{max}}$  weighted density, with  $\log_{10}$  distributed levels. Each panel shows a different redshift range (using the MIZUKI derived photometric redshifts) which is indicated in the top-left corner. Histograms at the top and right-hand side of each panel show the distribution of stellar mass with  $\log$  scaling and SFRs with linear scaling respectively, for each morphological type. Coloured triangles indicate the  $1/V_{\text{max}}$  weighted median SFR for ellipticals, S0/Sa galaxies and spirals. The number in the bottom right corner of each panel indicates the total number of objects in each redshift bin.

and galaxies classified as spirals occupying the ‘blue cloud’ (e.g. Baum, 1959; Visvanathan, 1981). S0/Sa galaxies inhabit both parts of the diagram, but largely occupy the space in between the two distributions defined by the spiral and elliptical populations.

The histograms above each panel in Fig 7.9 show the distribution of stellar masses for each morphological type. In agreement with other studies (e.g. Kelvin et al., 2014), we find that the stellar mass function of S0/Sa galaxies is much closer to that of ellipticals than spirals. Spirals are much less massive, on average, than ellipticals and S0/Sa galaxies, while S0/Sa galaxies have marginally lower stellar masses than ellipticals.

The histograms on the right-hand side of each panel in Fig 7.9 show the distributions of SFRs. Coloured arrow heads indicate the  $1/V_{max}$  weighted median values (e.g. Edgeworth, 1888) of the SFRs in each population. While S0/Sa galaxies typically have SFRs that are comparable to spirals and higher than those found in ellipticals, they are typically more massive and therefore inhabit an intermediate range of values of specific SFRs. They remain redder and less star-forming than the majority of spirals, although the majority retain fairly high levels of star formation compared to ellipticals (e.g. Thronson et al., 1989; Pogge & Eskridge, 1993).

The histograms above each panel in Fig 7.10 show the distributions of absolute  $i$ -band magnitudes for each morphological type, while the histograms to the right of each panel show distributions of rest-frame  $g - i$  colours. Again, coloured arrow heads indicate the  $1/V_{max}$  weighted median values for each population. Galaxies classified as ellipticals and spirals inhabit opposite ends of a bi-modal distribution in  $g - i$  colour, with galaxies classified as S0/Sa typically lying between the two populations (e.g. Wilman & Erwin, 2012; López Fernández et al., 2018).

Given that different morphologies show some separation in integrated properties (e.g. stellar mass and SFR), it may be tempting, particularly when faced with the data volumes expected from future surveys, to use these properties as proxies for morphology. However, as previous studies (e.g. Fadely et al., 2012; Vika et al., 2015) have shown, spatial information is essential for the robust morphological classification of both stars and galaxies (see also Section 7.4.2). We use our morphological classifications to explore this point in more detail.

Fig 7.11 shows the positions of a random selection of objects classified as ellipticals, spirals and S0/Sa galaxies within the colour-colour (Fig 7.11(a)), colour-magnitude (Fig 7.11(b)), magnitude-magnitude (Fig 7.11(c)) and stellar mass-SFR (Fig 7.11(d)) planes. Regions of contiguous colour in each plot indicate parts of the parameter space which are dominated by objects of a given morphological group i.e. the parameter space is colour-coded by the modal group in each hexagonal bin. It is clear that a significant fraction of objects of different morphology can fall into the same regions of parameter space, regardless of the exact plane being considered. Thus, a large degree of overlap exists in the integrated properties of S0/Sa galaxies, spirals and ellipticals, not only in colour-colour, colour-magnitude and magnitude-magnitude space, but also in physical properties like stellar mass and SFR. Such integrated

properties *alone* are therefore not sufficient to separate objects morphologically. The spatial information contained in the power spectrum of each patch type, as well as the spatial distribution of patch types across each object, are essential ingredients of accurate morphological classification.

## 7.6 Summary

Morphology is a fundamental quantity that encodes the principal mechanisms that drive the evolution of individual galaxies. Essential for the full spectrum of galaxy-evolution studies, morphology is an important parameter for an array of topics in astrophysics, e.g. as a prior in photometric redshift pipelines and as contextual data in transient lightcurve classifications. A rich literature exists on morphological-classification techniques, with methods ranging from automated classification (e.g. via parametric and non-parametric reductions of galaxy images and machine-learning techniques) to direct visual classification by human classifiers, which is typically used to benchmark automated techniques.

Notwithstanding the array of techniques on offer, the forthcoming era of ‘Big Data’ deep-wide surveys poses unique challenges for measuring galaxy morphologies. The sheer volume of data expected from surveys like LSST makes visual classification intractable for such datasets (even via massively-distributed systems like Galaxy Zoo) and makes some degree of automation essential for this exercise. The short cadence of surveys like LSST presents an additional challenge, because repeatedly producing training sets, that are required for supervised machine-learning techniques, on short timescales may be impractical.

Unsupervised machine-learning (UML) offers an attractive solution to these issues and an ideal route for the morphological classification of galaxies in next-generation surveys. An effective UML algorithm can autonomously compress an arbitrarily large galaxy population into a small set of morphological clusters whose members have similar morphology. If the number of clusters is small enough (e.g. in the hundreds), then this makes it tractable to benchmark them using visual classification by individual researchers. The resultant classifications can thus combine both the speed of automation and the accuracy of visual classification.

Here, we have employed such a UML algorithm, which automatically identifies distinct groups of galaxy types from survey pixel data, to separate galaxies in the HSC-SSP DR1 Ultradeep layer into 160 morphological clusters. This technique extracts sub-image patches from multi-band HSC data, each of which are transformed into a rotationally-invariant representation of a small region of the survey data, efficiently encoding colour, intensity and spatial frequency information. Utilising growing neural gas and hierarchical clustering algorithms, it then groups patches into a library of patch types, based on their similarity, and assembles feature vectors for each object, which describe the frequency of each patch type. A  $k$ -means algorithm is then used to separate objects into morphological clusters, based on the similarity of their feature vectors.

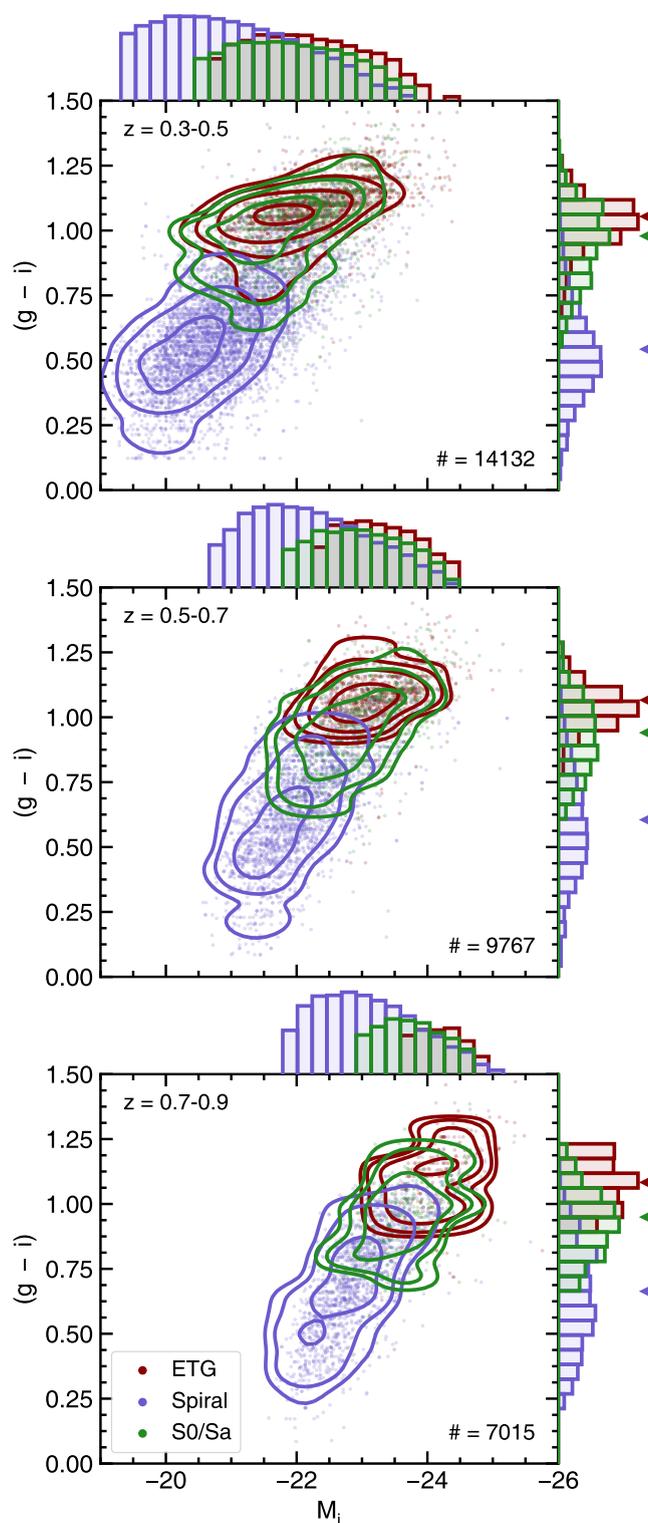


FIGURE 7.10: Contour plots showing the distribution of galaxies as a function of  $g-i$  colour and rest-frame  $i$ -band absolute magnitude, for galaxies that have been classified as elliptical (red), S0/Sa (green) and spiral (blue). Dots show individual galaxies, while contours show the  $1/V_{max}$  weighted density with  $\log_{10}$  distributed levels. Each panel shows a different redshift range (using the MIZUKI derived photometric redshifts) indicated in the top left corner. Histograms at the top and right hand side of each panel show the distribution of rest-frame  $i$ -band magnitudes with log scaling and  $g-i$  colour with linear scaling respectively, for each morphological type. Coloured triangles indicate the  $1/V_{max}$  weighted median  $g-i$  colour for ellipticals, S0/Sa galaxies and spirals. The number in the bottom right corner of each panel indicates the total number of objects in each redshift bin.

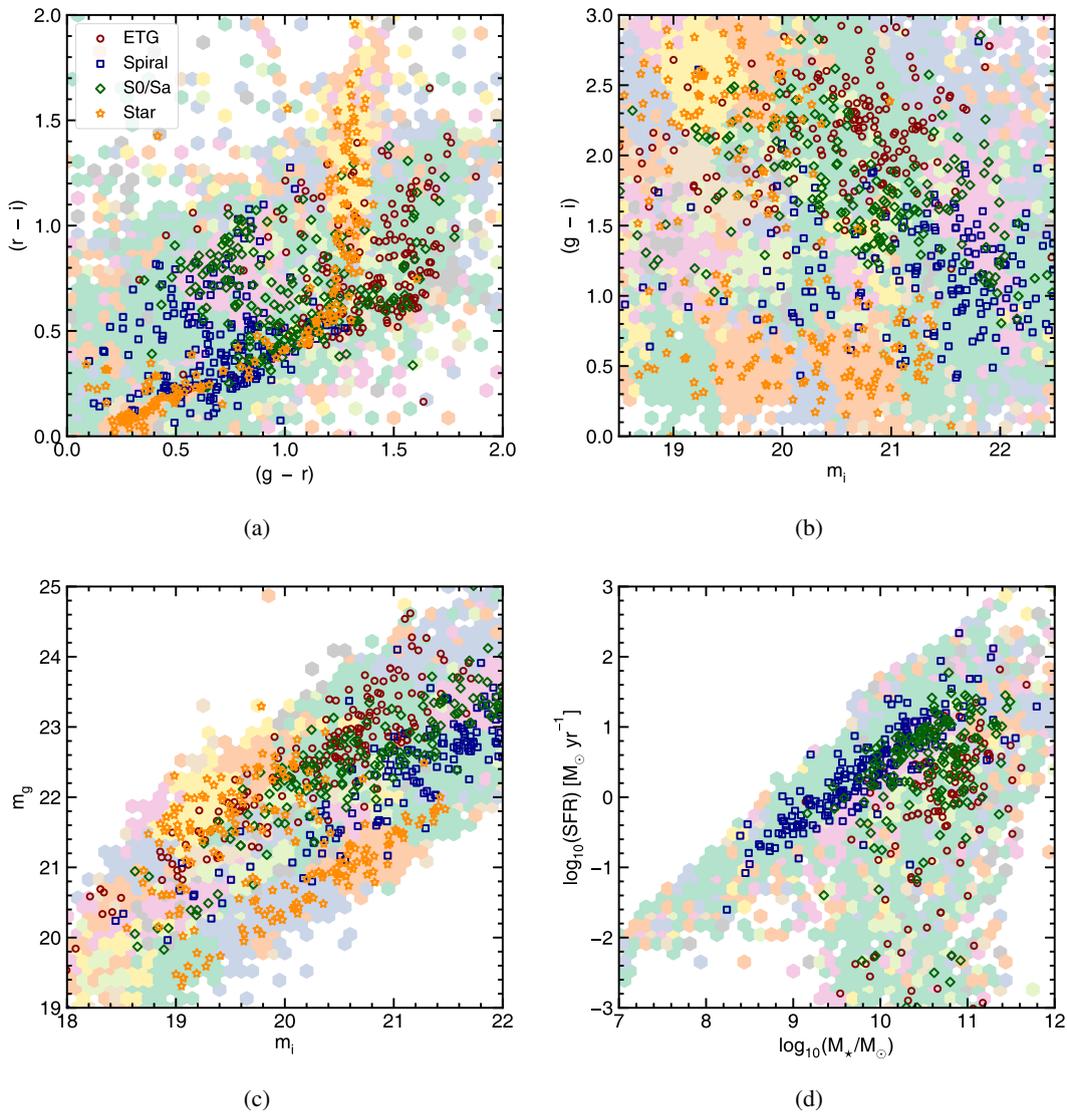


FIGURE 7.11: Morphological clusters as a function of various parameters. Contiguous hexagonal bins with the same colour indicate regions in the parameter space which share the same dominant group. We show the most frequent morphological clusters in colour-colour **(a)**, colour-magnitude **(b)**, magnitude-magnitude **(c)** and stellar mass vs. star-formation rate **(d)** space. Open red circles, blue squares, green diamonds and orange stars show the positions of a random sample of 200 ellipticals, spirals, S0/Sa galaxies and stars within each parameter space.

We have visually inspected a representative sample of objects in each morphological cluster to classify them into three broad morphological types: elliptical galaxies, S0/Sa galaxies and spiral galaxies. We also provide finer morphological information e.g. the type of spiral morphology (Sb, Sc, Sd) and noteworthy colour or structural features (e.g. when spirals appear unusually red or show clumpy structure, or when elliptical galaxies appear unusually blue). To test the robustness of the classifications, we have shown that galaxies in different morphological classes reproduce known trends in key galaxy properties as a function of morphological types at  $z < 1$ , e.g. stellar mass functions, rest-frame magnitudes and colours and the position of galaxies on the star formation main sequence.

Our study demonstrates the potential of UML in the morphological analysis of forthcoming deep-wide surveys. The combination of initial UML-driven automation, followed by benchmarking via visual classification, is likely to become an optimal tool for the morphological analysis of surveys like LSST. While this study has focused on bright galaxies at  $z < 1$ , it is worth noting that a significant fraction of objects, especially at low masses, inhabit the low-surface-brightness (LSB) Universe (e.g. Martin et al., 2019). In forthcoming papers, we will optimize the algorithm for the morphological classification of LSB galaxies and the detection of LSB structures, such as faint merger-induced tidal features, which will be routinely detectable in future surveys like LSST. Furthermore, while our morphological classifications are limited to  $z < 1$ , due to the ground-based nature of the HSC images, implementation of this UML algorithm on forthcoming higher-resolution data, e.g. from EUCLID, will enable virtually all-sky morphological classification of galaxies out to high redshift.

## Chapter 8

# Conclusions

In this thesis, we have investigated the diversity of galaxy properties at the present day and how this arises from the underlying processes by which they are assembled. In particular, we have studied the effect of galaxy mergers on the stellar populations, central super-massive black holes and the morphologies of galaxies over cosmic time and developed tools for quantifying the biases that this evolution produces in observational studies of galaxy populations. We have also explored the significance and formation mechanisms of LSB galaxies, whose ability to elucidate key questions in the field of galaxy evolution and significantly alter our current paradigm is becoming increasingly clear, especially with the advent of new deep-wide surveys.

### **8.1 How do mergers drive the evolution of galaxy properties over cosmic time?**

Mergers are often considered important drivers of stellar mass growth (e.g. van Dokkum et al., 2010; Kaviraj et al., 2011; Ferreras et al., 2014), the co-evolution of black hole (BH) mass and bulge mass (e.g. Sanders et al., 1988; Croton et al., 2006; Hopkins et al., 2006) and galaxy morphology. While today's surveys are able to provide datasets of sufficient quality that it is possible to compare galaxy populations across a large range in redshift, given the limited depth and/or area of past surveys, samples of mergers are typically small (e.g. Darg et al., 2010a,b) and an empirical determination of the role that mergers play remains difficult, particularly as most surveys are too shallow to detect the signatures of low mass-ratio mergers (see e.g. Peirani et al., 2010; Kaviraj et al., 2013a). Exploiting the Horizon-AGN simulation, we have performed detailed studies of the role that mergers play in driving stellar mass growth, BH growth and morphological evolution.

Chapter 3 probed the contribution of mergers to the cosmic star formation (SF) budget, showing that, while mergers can enhance star formation significantly in the low redshift Universe, this enhancement

is small at high redshift when the cosmic SFH peaks. As a result, mergers are directly responsible for under 30% of the stellar mass today, with minor mergers being responsible for more than 60% of merger-driven mass growth. Interaction driven starbursts are often considered important for the build up of stellar mass, based on the high incidence of starbursts around the cosmic peak of the star formation activity. However, this work offers a counter to this idea, showing that mergers are not important drivers of the *integrated* stellar mass growth of the Universe and less than 30 per cent of cumulative stellar mass growth of the Universe is attributable to mergers.

Similarly, Chapter 4 investigated whether mergers drive the observed correlation of galaxy (bulge) and BH mass. We showed that galaxy merger histories have little effect on their position on and evolution across the  $M_*-M_{BH}$  relation and ‘bulge-less’ (bulge-to-total mass ratio  $< 0.1$ ) galaxies, which are virtually merger free, have BH masses considerably larger than would be expected based on the  $M_{bulge}-M_{BH}$  relation. We showed that most (65%) of the BH mass in the general population of today’s massive galaxies is, in fact, the result of secular processes, counter to the hypothesis that the bulge and central BH co-evolve as the result of merger-driven growth. Together, these studies show that the majority (more than two-thirds) of the stellar and BH mass in today’s massive galaxies was formed through direct accretion of gas around the epoch of peak cosmic star formation.

Notwithstanding their relatively minor role in driving stellar and BH mass growth, we found that mergers are important drivers of morphological change. In Chapter 5 we analysed the role of mergers in creating today’s spheroids and discs. We showed that essentially all morphological transformation that spheroids undergo is due to mergers with mass ratios greater than 1:10. However, major mergers *alone* are not sufficient to produce the observed morphological mix at the present day. Minor mergers contribute around a third of the overall cosmic disc-to-spheroid morphological transformation and become the dominant drivers of morphological change after  $z = 1$ . We also showed that the basic merger histories of galaxies do not have a strong influence on their present-day morphologies. Instead, the gas content and geometry of mergers are particularly important predictors of remnant morphology, with gas-rich or prograde mergers spinning galaxies up in many cases. The survival of discs to the present day is therefore driven by a preponderance of prograde and/or gas-rich mergers in their assembly histories.

## 8.2 Progenitor bias

Methods for overcoming progenitor bias (i.e. the bias that occurs if one uses only early-type galaxies (ETGs) to study the progenitor population of today’s early-types) are essential for studies of ETG mass assembly, particularly at high redshift where the majority of galaxies are discs. As deep observational surveys become more commonplace, and galaxy samples include lower stellar masses at higher redshifts, properly quantifying progenitor bias becomes ever more crucial for accurately quantifying the true evolution of ETGs.

Chapter 2 explored the merger histories of local ETGs in order to quantify the effects of progenitor bias as a function of redshift, environment and galaxy properties (e.g. stellar mass). We showed that progenitor bias is a significant problem at all but the lowest redshifts and, importantly for large, deep observational surveys (JWST, LSST etc.),  $> 80\%$  of the stellar mass that will end up in early-types at the present day is found in late-type galaxies at the cosmic peak of star-formation. We have released software that enables users to quantify progenitor bias in new and forthcoming large surveys.

### 8.3 The formation of low-surface-brightness (LSB) galaxies

New deep surveys have illuminated a low luminosity Universe filled with new types of objects including compact ellipticals (cEs), ultra-compact dwarfs (UCDs) and ultra-diffuse galaxies (UDGs). The demographics and formation mechanisms of UDGs are particularly uncertain, especially as a function of environment. While UDGs appear to be ubiquitous in clusters (e.g. Koda et al., 2015), they are found across all environments (Román & Trujillo, 2017a; Merritt et al., 2016; Papastergis et al., 2017). However deep surveys and spectroscopic follow-up of areas large enough to contain significant populations of UDGs outside dense cluster environments are often prohibitively expensive. The formation mechanisms of these systems and their relationship to the high-surface-brightness galaxy (HSBG) population, on which our understanding of galaxy evolution is predicated, remains (very) poorly understood. Theoretical studies into the formation of these new types of objects will, therefore, be key to a complete understanding of galaxy evolution, and will be central to observational work using new deep-wide surveys.

In Chapter 6, we used Horizon-AGN to explore the formation and evolution of low-surface-brightness galaxies (LSBGs), particularly ultra-diffuse galaxies (UDGs; e.g. van Dokkum et al., 2015). While our understanding of galaxy evolution is dominated by HSBGs (since they are visible in current surveys like the SDSS), a majority of galaxies inhabit the LSB regime (e.g. Dalcanton et al., 1997), meaning it is vital that we understand how these systems fit into the galaxy evolution paradigm, in order to understand galaxy evolution as a whole. We showed that for  $M_* > 10^8 M_\odot$ , LSBGs contribute 50 per cent of the local number density and, as Fig 6.3 shows, exist in significant numbers across all environments. They cannot therefore be ignored when formulating our wider paradigm of galaxy evolution. We showed that their progenitors have stronger, burstier star formation at high redshift which causes stronger supernova feedback that, in turn, produces flatter gas-density profiles. This then gives rise to flatter stellar profiles which are more susceptible to environmental processes and galaxy interactions which produce today's LSB population by driving the steady removal of cold gas and gradually increasing their effective radii over time.

## 8.4 A novel unsupervised machine-learning technique for galaxy morphological classification in new and forthcoming deep-wide surveys

Galaxy morphology is a fundamental quantity, that is essential for galaxy-evolution studies as well as a plethora of science in observational cosmology (e.g. as a prior for photometric-redshift measurements and as contextual data for transient lightcurve classifications). However, the unprecedented data volumes, coupled, in some cases, with the short cadences of forthcoming ‘Big-Data’ surveys (e.g. LSST) present novel challenges, even for established supervised machine learning techniques. The large volumes of data produced by these surveys make them intractable for visual inspection, even for citizen science platforms like Galaxy Zoo. And the short (three day) cadence of LSST in particular will make it difficult to employ techniques like supervised machine-learning which rely on carefully selected, unbiased training sets, since it may be impractical to repeatedly produce such training sets on short timescales. *Unsupervised* machine learning (UML), which does not require training sets, may be ideally suited to the morphological analysis of such datasets.

In Chapter 7 we demonstrated the implementation of a UML algorithm – which works by collecting pixels with similar properties (and objects constructed from those pixels, like galaxies) – on images from the Hyper-Suprime-Cam Subaru-Strategic-Program Ultra-Deep survey, which is a precursor survey to LSST. We use the algorithm to autonomously reduce the galaxy population down to a small number ( $\sim 160$ ) of ‘morphological clusters’, populated by galaxies with similar visual morphologies in the optical HSC images. By benchmarking each of these morphological clusters using visual inspection, we can produce a catalogue of meaningful galaxy morphologies with minimal human effort. The morphological classifications exhibit a high level of purity, and reproduce known trends in key galaxy properties as a function of morphological type at  $z < 1$  (e.g. stellar mass functions, rest-frame magnitude and colours and the position of galaxies on the star-formation main sequence).

Our study has demonstrated the potential of UML in performing accurate morphological analysis of forthcoming deep-wide surveys. The combination of initial UML-driven automation, followed by benchmarking via visual classification, is likely to become indispensable in the forthcoming era of deep-wide surveys and is an optimal tool for the morphological analysis of surveys like LSST. In the future, implementation of the UML algorithm on forthcoming higher-resolution data, like EUCLID, WFIRST and JWST, will enable virtually all-sky morphological classification of galaxies out to high redshift.

## Chapter 9

# Future work

### 9.1 Comparison and limitations of current work

The Horizon-AGN simulation has been successful at reproducing a wide range of the observed present-day properties of galaxies, generally in good agreement with other contemporary simulations (e.g. Vogelsberger et al., 2014; Schaye et al., 2015). Despite some significant differences in methodology (including the hydrodynamics schemes used, the implementation of sub-grid recipes, feedback processes and the treatment of shocks, shear flows and instabilities), all of these simulations have been relatively successful at reproducing the basic properties of galaxies over cosmic time.

This agreement largely extends to the specific results presented in this thesis. For example studies using similar large cosmological simulations to Horizon-AGN (e.g. McAlpine et al., 2018; Steinborn et al., 2018) also find a limited link between galaxy mergers and the growth of the central SMBH. Higher resolution simulations like the ROMULUS simulation (70 pc minimum AMR cell size with cooling via metals below  $10^4$  K (Tremmel et al., 2017)), which are better able to resolve angular momentum loss close to the black hole, also find no evidence of a link between galaxy mergers and black hole accretion rates (Ricarte et al., 2019). A good consensus between simulations and observations is also beginning to emerge on the importance of mergers for star formation and stellar mass growth, with only limited contributions expected (Lofthouse et al., 2017a; Laganá & Ulmer, 2018; Wilson et al., 2019; Lisenfeld et al., 2019) as a result of major mergers or environment.

However a number of more subtle disagreements between simulations and observations remain. These include, for example, whether ‘intrinsic alignments’ (the spin alignment of galaxies with each other and their environment) is captured (Velliscig et al., 2015; Chisari et al., 2016) and disagreements in the abundances of disc galaxies and passive satellites as a function of stellar mass (e.g. Bahé et al., 2017). More importantly, a number of uncertainties and shortcomings also remain in the methodology and sub-grid models of all of these simulations. These include a lack of gas cooling below  $10^4$  K, which may limit the

accretion rate of gas, an inability to resolve disc scale heights or the small scale motions and properties of the CGM, which inhibit our ability to understand disc formation and the advection of angular momentum respectively, and uncertain SN and AGN feedback recipes also introduce uncertainties in the mass and the velocity launched or swept up by these processes (e.g. Hu et al., 2016). These uncertainties, as well as somewhat low time and spatial resolutions, can result in tensions between the predictions of simulations and observations (e.g. Kaviraj et al., 2017; Pillepich et al., 2019).

One particular case where it is unclear whether observations fully line up with theoretical predictions is the formation mechanisms of LSBGs. Although other studies have come to similar conclusions regarding the role of supernovae and environment (e.g. Di Cintio et al., 2017, 2019; Jiang et al., 2019), for higher mass field UDGs/LSBGs, these models rely on the ability of SN to drive rapid, non-adiabatic changes in order to produce permanent flattening of galaxy matter distribution. Additionally, in terms of correspondence with observations UDGs in the field are generally observed to have blue colours (Greco et al., 2018b), while those observed in Horizon-AGN are typically quenched. There is therefore still work to do in robustly confronting the abundances and properties of simulated galaxies (including surface brightness distribution and density profiles) with new deep observational surveys and instruments like LSST, which will allow us to constrain and improve our models of the physics of galaxy evolution (e.g. reionisation, SN feedback, models of DM). Other new facilities including JWST, Euclid, ALMA, E-ELT and the SKA will soon furnish observers with large amounts of deep, high resolution and multi-wavelength data out to high redshifts. More complex theoretical modelling is therefore required in order to produce testable predictions for this incoming data and thereby enable a richer observational picture and theoretical understanding of galaxy evolution.

## 9.2 High resolution cosmological simulations

New high-resolution cosmological simulations, like the New Horizon (NH) simulation (Dubois et al. in prep.) and TNG50 (Pillepich et al., 2019; Nelson et al., 2019) aim to meet the challenge of this new data by way of higher spatial resolutions and improved sub-grid recipes, albeit over somewhat smaller volumes than previous simulations. This allows them to more accurately model the processes relevant to galaxy evolution (and with greater detail), while still retaining volumes of tens of Mpc. The large contiguous volume of these simulations offers advantages over many equivalent resolution single-halo zoom simulations (e.g. NIHAO, FIRE; Di Cintio et al., 2017; Chan et al., 2018). In particular, it becomes possible to study both realistic, unbiased statistical populations of galaxies and details of the environment and filamentary structure on all scales and within their proper cosmological context. Fig 9.1 demonstrates this high dynamic range in the case of the NH simulation, which is able to model the clumpy discs of high redshift galaxies and dwarfs while also resolving the structure and kinematics of filaments and the CGM.

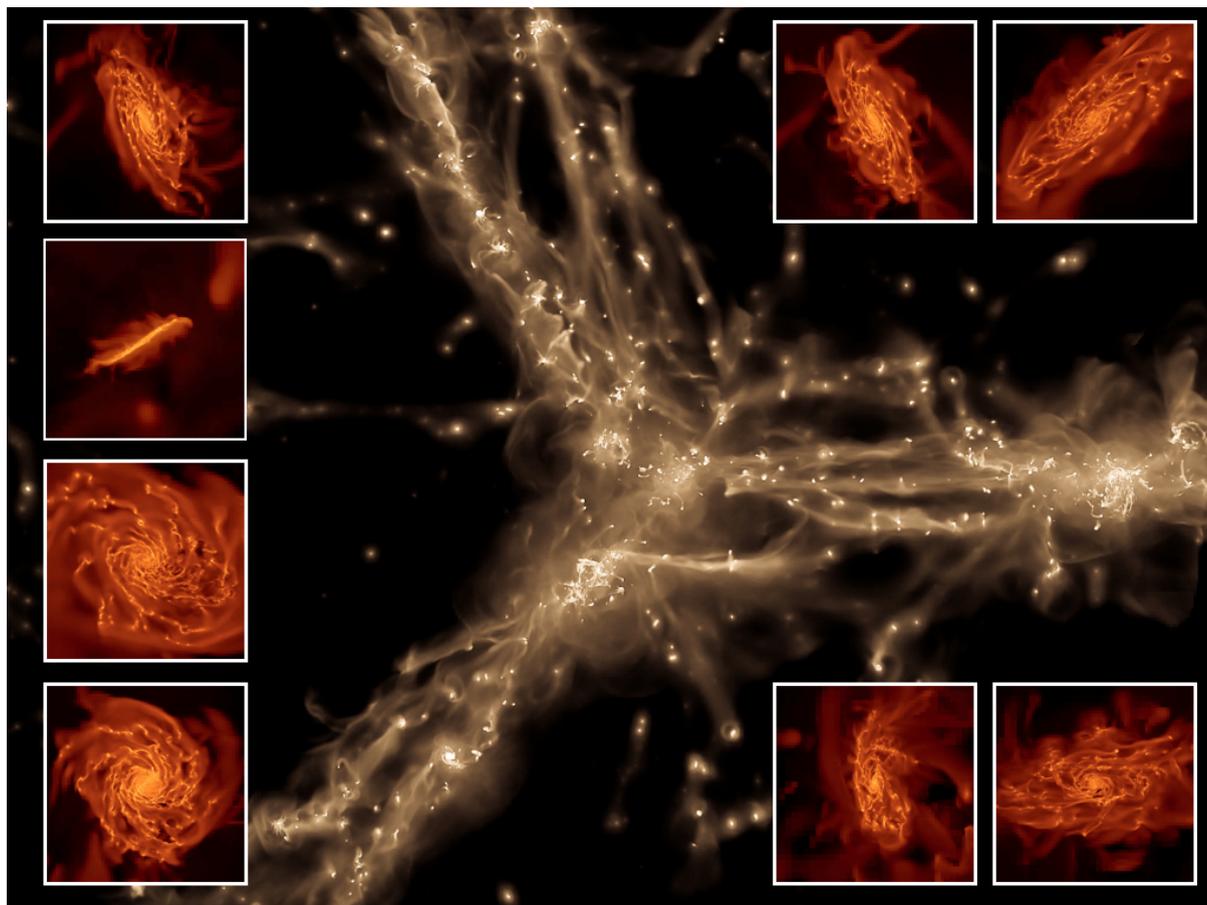


FIGURE 9.1: **Main plot:** gas density map showing a slice through a portion of the NH volume at the confluence of three filaments. The scale is a few Mpc across. **Surrounding panels:** zoom-in showing individual galaxies. Credit: NH collaboration.

NH is a high-resolution re-simulation of a large ( $4 \text{ Gpc}^3$ ) region of the Horizon-AGN simulation, with a spatial resolution of 35 pc. It will allow us to perform statistical studies of galaxy evolution within the context of fully resolved cosmological large scale structure, while also resolving the small scale baryonic processes that influence the evolution of low mass haloes, this allows us to probe large populations of dwarfs, as well as LSB galaxies and the effect of supernova-driven outflows, minor mergers and cored dark matter haloes on their formation (Bédorf & Portegies Zwart, 2013; Di Cintio et al., 2017; Carleton et al., 2018). For comparison, Fig 9.2 shows the parameter space covered by the NH simulation compared to the Horizon-AGN simulation. With the resolution of New Horizon, it is now possible to resolve the region of parameter space which, for example, the majority of LSB galaxies and dwarfs occupy, as well as begin to resolve the regime of ultra compact dwarfs (which may be related to nucleated LSB galaxies; Janssens et al., 2017).

The implementation of improved physical recipes is essential for gaining an better insight into the processes that drive the evolution of galaxies. NH includes new physics including metal-dependent gas cooling down to 0.1 K and new models of gas turbulence driven star formation (Devriendt et al., in

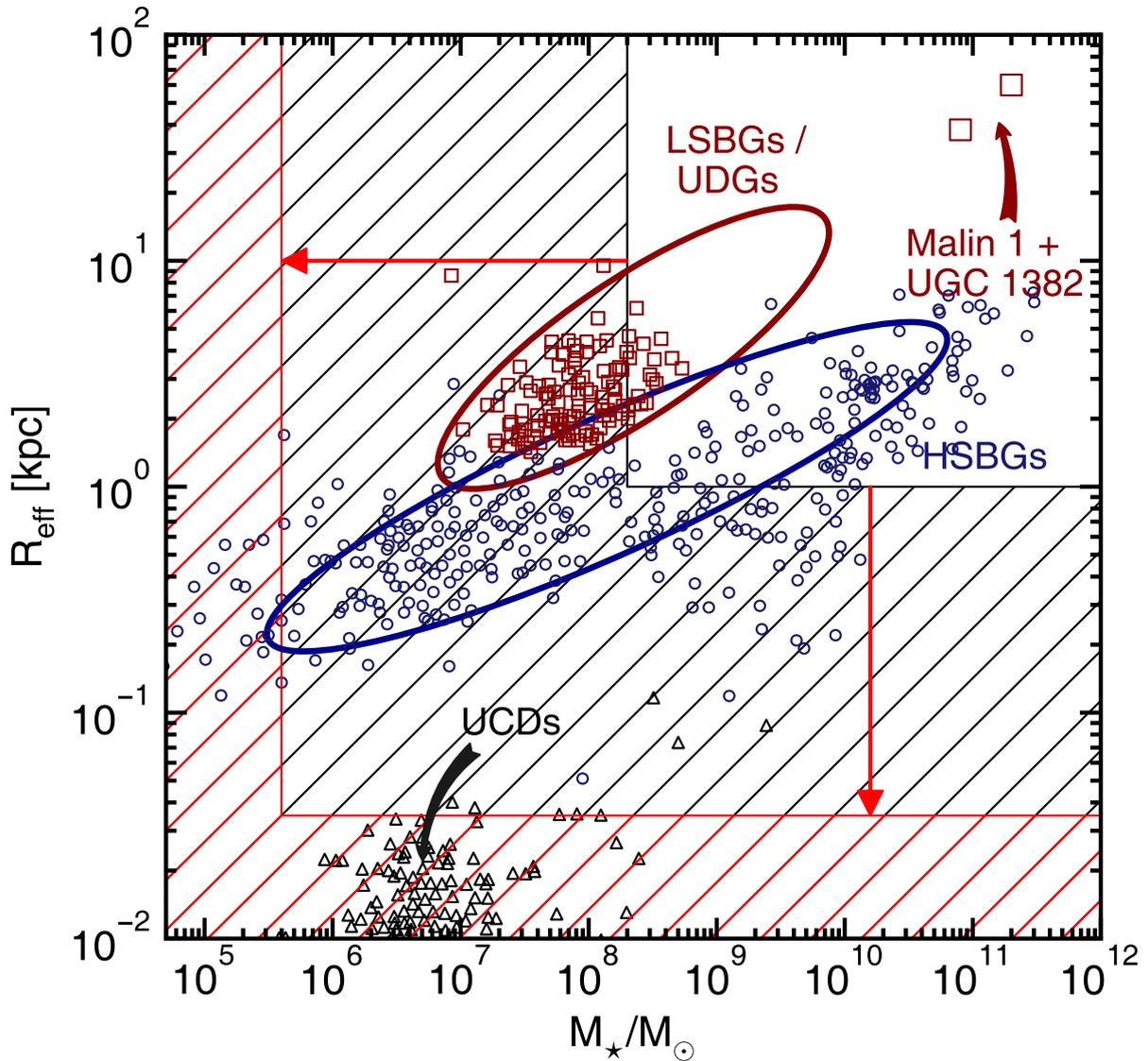


FIGURE 9.2: The parameter space covered by the Horizon-AGN and New Horizon simulations. The grey hatched region shows the parameter space that is beyond the mass and spatial resolution limits of the Horizon-AGN simulation. The red hatched region shows the same for the New Horizon simulation. The positions of observed LSB and HSB populations in this parameter space are plotted as open points.

prep), motivated by high-resolution magneto-hydrodynamic simulations of interstellar medium turbulence (e.g. Padoan & Nordlund, 2011) where the star-formation efficiency depends on the local turbulent mach number and virial parameter. Following Kimm et al. (2015), the model for type II SN explosions now includes the direct modelling of energy-conserving and momentum-conserving phase, injecting the correct amount of radial momentum expected at each stage, while also correcting for the non-uniform density distribution of the ISM below the resolution scale. Importantly, this model successfully reproduces the stellar-to-halo mass relation in low-mass galaxies in the high redshift Universe. More generally, such simulations will allow us to resolve the scale-height for a statistical sample of Milky-Way-like galaxies, making detailed morphological analysis possible while also resolving details of the CGM and enabling us to quantify the dynamics of the cosmic web and its implications for galaxy morphology. We

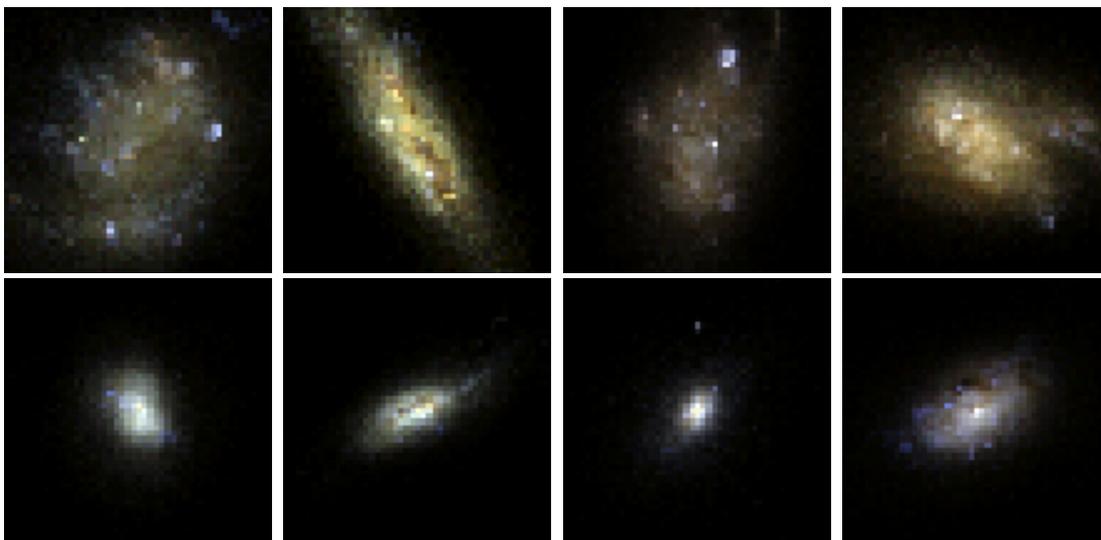


FIGURE 9.3: Mock observations processed with SKIRT radiative transfer code using JWST NIRCам F070W, F115W and F150W filters and pixel size. The top row shows candidate LSB progenitor galaxies and the bottom shows HSB progenitors at  $z \sim 1$ .

will also be able to, for example, determine what processes are responsible for the disruption of cold flows and probe the processes that lead to the production of cores in galaxy density profiles, allowing us to probe the relative contribution of these processes as a function of cosmic environment as they pertain to the production on low-surface-brightness galaxies.

The high spatial resolution of these new simulations will also enable predictions and mock observations of the high redshift Universe in regimes that will be probed by upcoming instruments. Fig 9.3 shows examples of mock observations of dwarf galaxies ( $\sim 10^8 M_{\odot}$  at  $z = 1$ ), indicating the kind of level of detail that will be accessible by these instruments for these classes of object. The images shown here are processed using the SKIRT radiative transfer code and convolved to match the JWST NIRCам instrument.

### 9.3 The drivers of galaxy diversity in the low-surface-brightness Universe

This new generation of high-resolution simulations will allow the theoretical study of the statistics and formation mechanisms of different LSB galaxy populations within a fully cosmological context. Coupled with new deep and high-resolution survey instruments and machine learning techniques it will become possible to probe the formation, evolution and properties of LSB populations at the present day and across cosmic time, confronting theoretical predictions with robust comparison to observations.

Although an active field of research for the past few decades (e.g. Bothun et al., 1987; Dalcanton et al., 1997; Impey & Bothun, 1997; O’Neil & Bothun, 2000), significant progress in the characterisation and detection of a significant numbers and varieties of LSBG populations has only recently become possible. In particular, improved data reduction techniques and bespoke instruments have allowed the discovery of extremely low-surface-brightness ( $\mu_r > 24.5$  mag arcsec<sup>2</sup>) and ultra diffuse galaxies (UDGs; e.g. van Dokkum et al., 2015) in large numbers. This has produced renewed interest in the field, particularly as UDGs may be important laboratories for answering a number of key open questions in astronomy, including the nature of dark matter, the physics of stellar and black hole feedback and the role of environment in galaxy evolution (e.g. Hagen et al., 2016; Di Cintio et al., 2017; Pahwa & Saha, 2018; Martin et al., 2019). A complete census of LSBG populations across all environments (e.g. cluster, group, field) has remained elusive, however, as the search for UDGs has largely been limited to deep surveys of dense cluster environments (e.g. van Dokkum et al., 2015; Koda et al., 2015), meaning our current understanding of LSBGs in isolation is limited. As such, an understanding of how LSBG demographics vary as a function of environment remains incomplete.

Although LSB galaxies may dominate in terms of number, they have so far eluded attempts to paint a comprehensive theoretical picture for how they form and evolve, especially in the context of our current understanding of galaxy evolution which is informed by the HSB Universe. Previous studies have offered a wide range of potential drivers of LSBG evolution including interactions, stellar feedback and cluster processes (e.g. via the strong ambient tidal field) (e.g. Amorisco & Loeb, 2016; Di Cintio et al., 2017; Carleton et al., 2018). My previous work (Chapter 6, Martin et al., 2019) suggests that SN feedback may flatten the gas density profiles of some galaxies at high redshift ( $z > 2$ ). This can then produce flatter stellar density profiles, which are more susceptible to environmental processes, which then remove their gas and increase their effective radii over time, producing today’s LSBGs. In particular, it is not clear if LSB galaxies are ‘special’ objects, distinct from HSB galaxies, or if they exist as part of a common continuum of properties along with HSB galaxies.

New simulations with higher resolution and more realistic SN recipes will enable us to better study LSBG formation mechanisms while also offer a statistical explanation for LSBG evolution and produce set of theoretical predictions that can be rigorously tested with the next generation of deep-wide surveys like the currently available HSC survey and ultimately LSST and JWST. These simulations now possess sufficient spatial resolution to, in combination with instruments like JWST, explore the epoch where the progenitors of LSBGs begin to diverge from the high surface-brightness population and test formation mechanisms, including the potentially important role of minor mergers and AGN feedback (which may be more efficient in low-mass haloes than has traditionally been assumed e.g. Dashyan et al., 2018; Kaviraj et al., 2019).

Combining the state-of-the-art New Horizon simulation, novel unsupervised machine learning techniques and deep-wide survey data will allow the comprehensive study of the formation, evolution and present-day properties of LSB populations, answering a number of key questions:

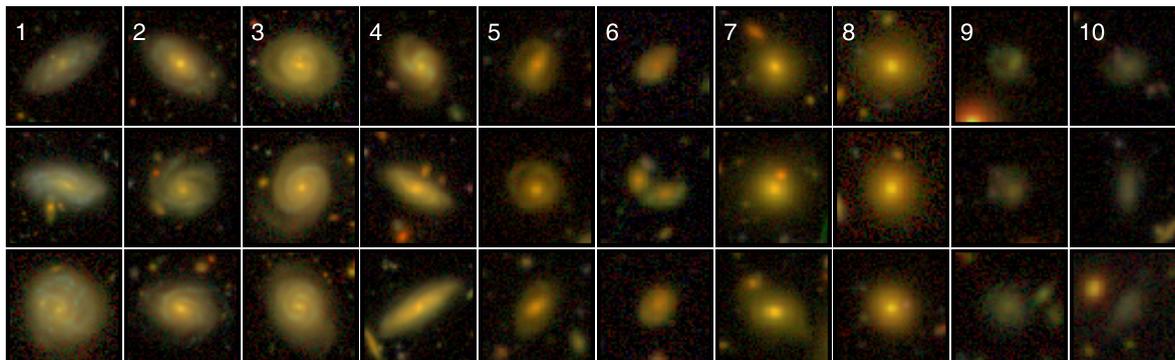


FIGURE 9.4: The first eight columns show late-type to early-type HSB galaxies, each column shows a group of visually similar galaxies identified by our machine learning algorithm. The final two columns show groups of LSB galaxies.

- What gives rise to the bursty star formation histories and subsequent flattened density profiles around the time that LSB and HSB galaxy populations begin to diverge at high redshift?
- What are the properties and formation mechanisms of observed LSB galaxies today? In particular, how do they vary as a function of morphological type?
- How well do our models reproduce the LSB galaxy population, and, in particular, where do they fail?

By performing high-resolution theoretical studies of the statistics and formation mechanisms of different LSB galaxy populations within a fully cosmological context, and producing a complete census of LSB galaxy morphologies from new deep survey multi-band survey data using novel unsupervised machine learning techniques, it will be possible to probe the formation, evolution and properties of LSB populations at the present day and across cosmic time, confronting theoretical predictions with observations. In particular, unsupervised machine learning represents a powerful method for automatically identifying and classifying the large numbers of faint objects that will be present in new deep surveys. Fig 9.4 shows an example of the algorithm used to finely classify galaxies from ultra-deep ( $\mu_r \sim 28$ ) HSC data (as in Chapter 7). Classifications go from late-type to early-type HSB morphologies with classifications of LSB galaxies in the final two columns. The resulting classifications are unbiased, sensitive to features that human classifiers have difficulty detecting and can be connected to known morphologies through classification of a broad group of object rather than individual galaxies.

Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP) (full release by early 2021), and later the LSST commissioning survey (available from early 2021) will make it possible to confront theoretical predictions with reality. The HSC survey will cover 1400 square degrees down to a depth of  $\mu_r = 26$  mag arcsec<sup>-2</sup> and a further 27 and 3.5 square degrees down to  $\mu_r = 27$  and 28 mag arcsec<sup>-2</sup> respectively. LSST commissioning data will cover 1500 and 200 square degrees down to  $\mu_r = 26.5$  and 27.8 mag arcsec<sup>-2</sup> respectively. Together, this data already represents a sample of several million galaxies at depths up to 6 mags deeper than the SDSS. Having established a morphological sequence of

LSB galaxies, we will be able to study the evolution of LSB physical properties as a function of local environment and morphological type. Within this framework, it will be possible study the size, stellar mass and SFR evolution of these galaxies down to  $z = 0.5$ . Comparison in this redshift range with predictions from simulations will allow us to confront our theoretical picture with the observational reality. Importantly, confronting the abundances and properties of simulated LSB galaxies with deep observational surveys will allow us to constrain and improve our models of the physics that LSB and low-mass galaxies are highly sensitive to (e.g. SN feedback).

## 9.4 Visual morphologies at the intersections of the cosmic web

As mentioned above, the ability of new simulations to resolve the detailed motions of gas within filaments as well as within galaxy haloes make it possible to study the dynamics of cosmic filaments and their implications for the eventual morphology of galaxies. While the existence of a strong correlation between galaxy morphology and the environment in which they form is well known (e.g. Dressler, 1980), studies of this kind will help elucidate the role of cold filamentary flows and large scale structure, as well as make predictions for how evidence of these processes may persist in the properties of galaxies today (e.g. Jones et al., 2010; Codis et al., 2012; Tempel et al., 2013; Rong et al., 2016; Noguchi, 2018). Observations, as well as theoretical evidence (e.g. Dubois et al., 2014; Aragon-Calvo & Yang, 2014; Musso et al., 2018; Tempel et al., 2015; Kuutma et al., 2017) demonstrates that the large-scale structure of the Universe influences the acquisition of angular momentum and mass over all of cosmic time. But the way in which specific configurations and structures within the network of voids, filaments, walls and nodes, that make up the complex multi-scale structure of the cosmic web, shape the properties of galaxies is poorly studied.

NH, among others, allows the statistical study of galaxy properties and morphologies within the cosmic web over the lifetime of the Universe. Using astrophysical structure identification techniques like DisPerSE (Sousbie, 2011; Sousbie et al., 2011) it is possible to trace the cosmic web by extracting a ‘skeleton’ which describes the large scale structure of matter: identifying voids, walls, filaments, clusters and their configuration within the metric of the cosmic web. Fig 9.5 shows an example of DisPerSE applied to the Horizon-AGN simulation, the greater resolution afforded by NH will allow much more detailed determinations of the structure of the cosmic web at smaller scales. By combining this technique with robust morphological classifications based on the ML method presented in Chapter 7.

It will be possible to identify simulated galaxies within specific configurations and structures of the cosmic web and classify the visual morphologies of these populations. This will allow us to test, for example, whether galaxies that share common large-scale environments also share common morphological features or properties, how morphologies vary with specific filamentary arrangements particularly around nodes of the cosmic web where galaxies are fed by multiple filaments and to identify the kind of

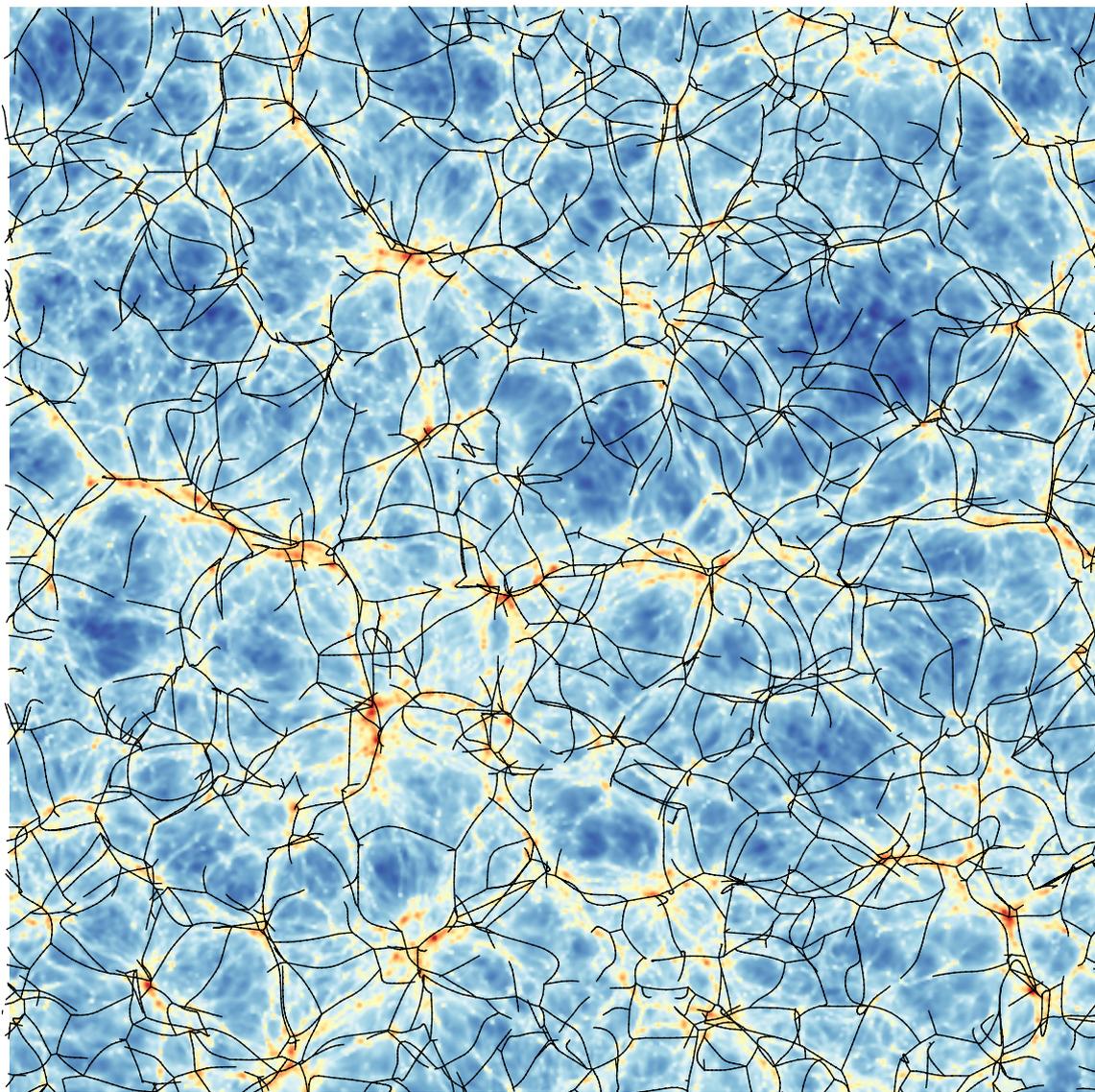


FIGURE 9.5: Figure from Dubois et al. (2014) showing the cosmic skeleton of Horizon-AGN.

high redshift signatures of different accretion scenarios. This will help to paint a precise theoretical picture of the role that the cosmic web plays in shaping the morphology of present day galaxy populations, help explain how specific structures in the cosmic web may be encoded in the morphology and structure of galaxies and establish whether such signatures that will be observable with the next generation of instruments.

Investigating these kinds of statistical signals will be possible with instruments like LSST, JWST and EUCLID. The next generation of space based observatories (e.g. JWST) will allow us to test the signatures of different angular momentum acquisition scenarios based on detailed predictions of galaxy structure and morphology. Additionally, by utilising deep imaging capabilities of ground based surveys like LSST it will be possible to produce a detailed picture of the cosmic web with similar accuracy to the COSMOS survey ( $\sim 75$  Mpc slices based on photometric redshifts; Laigle et al. 2018). Additionally,

shallower space-based observations from EUCLID will be able to make use of spectroscopic redshifts to map the cosmic web with even greater accuracy and also have the advantage of highly resolved observations (0.1 arcsec angular resolution), enabling the determination of highly detailed galaxy morphologies, including accurate measurement of Sérsic indices, which can provide an observational test for theoretical predictions.

## 9.5 The UVCANDELS survey

### 9.5.1 UV, optical, IR and composite morphologies with UVCANDELS

The shapes, sizes and morphologies of galaxies depend significantly on the wavelength in which they are observed (e.g. Kuchinski et al., 2000; Petty et al., 2014; Vika et al., 2015). Since different wavelengths trace different processes and populations, the visual characteristics of objects can vary strongly between bands, making traditional Hubble type classifications impossible to reconcile (e.g Kuchinski et al., 2000). New space-based, multi-wavelength surveys like UVCANDELS (Teplitz, 2018), in combination with the previous CANDELS survey (Grogin et al., 2011), offer high spatial resolution photometry out to high redshift spanning observed-frame wavelengths from the mid-IR to the mid-UV. This wide spectral range, in combination with high spatial resolution, will allow robust determinations of galaxy properties and morphologies not possible with ground based optical instruments. Unsupervised machine learning techniques offer us a route to quantify and summarise the visual properties of these objects across a wide range of wavelengths.

Morphologies from composite UV-optical-IR photometry, like those that will be available from UVCANDELS and CANDELS can offer a greater insight into the visual properties of galaxies. By grouping together galaxies with similar morphologies we can reduce a large galaxy sample to a small number of morphological groups, which can then be analysed via by visual inspection. For example, one can separate galaxies in UVCANDELS into different UV morphological classes, producing a UV defined morphological sequence or do the same using a combination of UV, IR and optical bands, producing robust and nuanced morphological classifications that combine information from across the electromagnetic spectrum. Correlating these morphologies with AGN signatures will allow us to study the effect of AGN on star formation in massive galaxies. For example, negative and positive AGN feedback are expected to produce different signatures in UV morphologies, allowing for a detailed picture of the impact of AGN on star formation and structure in the intermediate redshift Universe ( $z < 1$ ). Another possible application of this technique is to use narrow band survey data like J-PAS (Benitez et al., 2014), which will allow the determination of the morphology of galaxies in specific spectral lines, thereby tracing the resolved physical processes encoded by these spectral features.

### 9.5.2 Elucidating the origins of the UV upturn in elliptical galaxies with UVCANDELS

Although they are typically assumed to be ‘red and dead’, a large proportion of elliptical galaxies may exhibit signatures of UV excess in their optical-UV colours. The origin of this ‘UV upturn’ is not well understood, but is thought to be the result of hot helium-burning stars. In the model of Yi et al. (1999), the UV upturn is the result of long-lived, metal-rich, UV-bright horizontal branch stars. However the models of Han et al. (2003) imply it may instead be the result of post-main-sequence stars which were formed more recently. Another possibility is that this flux comes as a result of young stars. These stars would likely have formed in the last few Gyrs (e.g. Kaviraj et al., 2007) and their formation may be fuelled by recent mergers (Kaviraj et al., 2013c; George & Zingade, 2015; George, 2017). Finally the UV excess may result from AGN activity.

UVCANDELS would make it possible to discriminate between these scenarios through resolved UV imaging and produce resolved 2-D maps of the UV-optical colours of elliptical galaxies, which will enable the study of processes which lead to excess UV flux. With high-angular-resolution imaging, the existence of AGN or star formation can be easily determined from the structure of UV flux. For example, AGN feedback produces a ring of suppressed star formation and bluer FUV–NUV colours (e.g. George et al., 2018). Recent star formation would instead manifest itself in distributed, clumpy structures across the object. On the other hand, galaxies without ongoing star formation should appear smooth in the UV. These three scenarios will also exhibit vastly different UV morphologies that can be effectively separated using unsupervised machine learning. Finally, in the case of past star formation, the Yi et al. (1999) and Han et al. (2003) models can be discriminated between by tracing the evolution of the rest-frame UV-optical colours out to intermediate redshift ( $z < 0.4$ ). Compared with the Yi et al. (1999) model, the Han et al. (2003) model predicts a milder evolution in average rest-frame UV-optical colours and dispersions, which UVCANDELS has the requisite sensitivity to precisely characterise.

# Appendix A

## A.1 Tables of progenitor probabilities

The joint progenitor probability distribution is tabulated as a function of redshift, stellar mass, local environment and star-formation rate and stored in an  $a \times b \times c \times d$  binary file. The python routine `tabulate_progenitor_probability.py` outputs progenitor probabilities from the joint probability distribution and can be run as follows:

```
python tabulate_progenitor_probability.py -z <redshift> -m <mass> -p <percentile>
-s <SFR>,
```

where mass is in units of  $\log_{10}(M_*/M_\odot)$  and SFR is in units of  $M_\odot \text{ yr}^{-1}$ . At least 1 of the 4 keywords must be supplied, if fewer than 4 keywords are supplied the routine outputs the progenitor probability marginalised over the missing dimension(s).

e.g. calling `python tabulate_progenitor_probability.py -z 2.0 -m 10.0 -s 3` returns the joint progenitor probability for  $z = 2$ ,  $M_* = 10^{10} M_\odot$  and  $SFR = 3 M_\odot \text{ yr}^{-1}$  with environment marginalised out.

We also provide progenitor probabilities in ASCII format. For each redshift  $z \in [0,5]$  progenitor probabilities are tabulated in 4 files:

- `m_rho_SFR.txt` contains 4 columns listing stellar mass [ $\log_{10}(M_*/M_\odot)$ ], percentile, star-formation rate [ $M_\odot \text{ yr}^{-1}$ ] and the joint progenitor probability.
- `m_rho.txt` contains 3 columns listing stellar mass, percentile and the progenitor probability marginalised over SFR.

- `m_SFR.txt` contains 3 columns listing stellar mass, SFR and the progenitor probability marginalised over percentile.
- `m.txt` contains 2 columns listing stellar mass and the progenitor probability marginalised over percentile and SFR.

The routine and files above are available from <http://www.star.herts.ac.uk/~gmartin/bias>.

# Appendix B

## B.1 Robustness of results with respect to choice of merger timescale and mass ratio threshold

In this section, we consider the robustness of our results to both changes in the assumed merger timescale and the range of mass ratios considered in this study. Fig B.1 shows the average fractional morphological change ( $\Delta$  morph) for the progenitors of today's spheroids and discs as a function of merger mass ratio. This is shown both for our adopted merger timescale of  $\tau = 2$  Gyrs, as well as for a shorter 1 Gyr window ( $t = \pm 0.5$  Gyrs centred around coalescence) and a longer 3 Gyr window ( $t = \pm 1.5$  Gyrs). Black lines indicate the median morphological change for all mergers since  $z = 0.5$  once the average fractional morphological change in the non-merging population over the same timescale has been subtracted. We restrict the plot to  $z < 0.5$  in order to exclude, as much as possible, the effect of cosmological accretion and because of the dependence of  $\Delta$  morph on redshift. However, the curves presented in Fig B.1 reach an asymptote at similar mass ratios regardless of the redshift range chosen.

We note first that an important consideration when selecting a timescale is to use a value that encompasses the merger event completely. In particular, choosing a timescale that is too short will lead to spurious results, because the merger remnant will not have relaxed at the point at which it is observed. For mergers that we are interested in studying here (mass ratios around 1:10 or greater), merger timescales are  $\sim 2$  Gyr (e.g. Jiang et al., 2008; Kaviraj et al., 2011). Fig B.1 shows excellent convergence in the  $\Delta$  morph values for timescales of 2 and 3 Gyrs. It also shows that for these timescales mergers with very low mass ratios (i.e. less than  $\sim 1 : 10$ ) do not induce large changes in galaxy morphology (we reinforce this point in our discussion of Fig B.3 below). For such low mass ratios, the morphological change during mergers for either spheroid and disc progenitors is not appreciably larger than the morphological change observed in the non-merging population over the same time period.

We also show the values of  $\Delta$  morph for a 1 Gyr timescale, which, as noted above, may be too short for the mass ratios of interest in this study. For the most part, the 1 Gyr timescale gives the same result as the 2 Gyr and 3 Gyr timescales, although they diverge slightly at large mass ratios. Visual inspection of the stellar mass distribution of a subset of galaxies undergoing mergers indicates that, 500 Myrs after

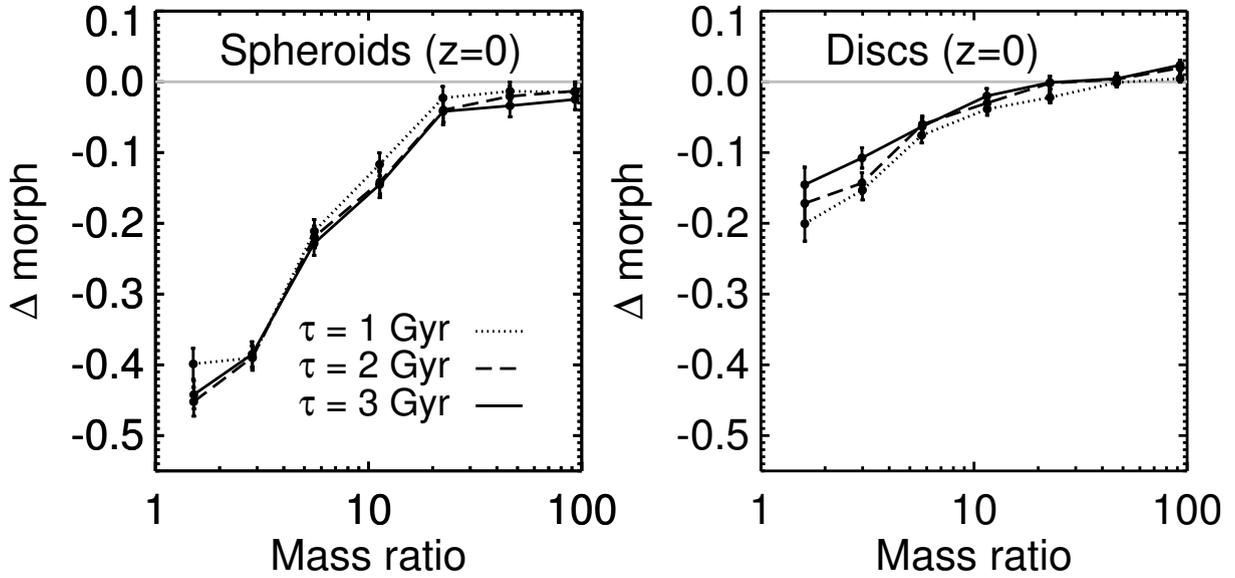


FIGURE B.1: The median fractional change in  $v/\sigma$  ( $\Delta$  morph) over the course of a merger, as a function of merger mass ratio, for mergers at  $z < 0.5$  where the more massive galaxy has a stellar mass of at least  $10^{10.5} M_{\odot}$  at the time of the merger. Note that, in order to control for the morphological change that is due to processes other than mergers, we have calculated the average fractional change in  $v/\sigma$  for non-merging systems over the same time period and subtracted this value from  $\Delta$  morph. Dotted, dashed and solid black lines indicate the median value of  $\Delta$  morph when merger timescales of 1 Gyr, 2 Gyrs or 3 Gyrs are used respectively. The left-hand panel shows the result for mergers involving the progenitors of today's spheroids, while the right-hand panel shows the same for the progenitors of today's discs.

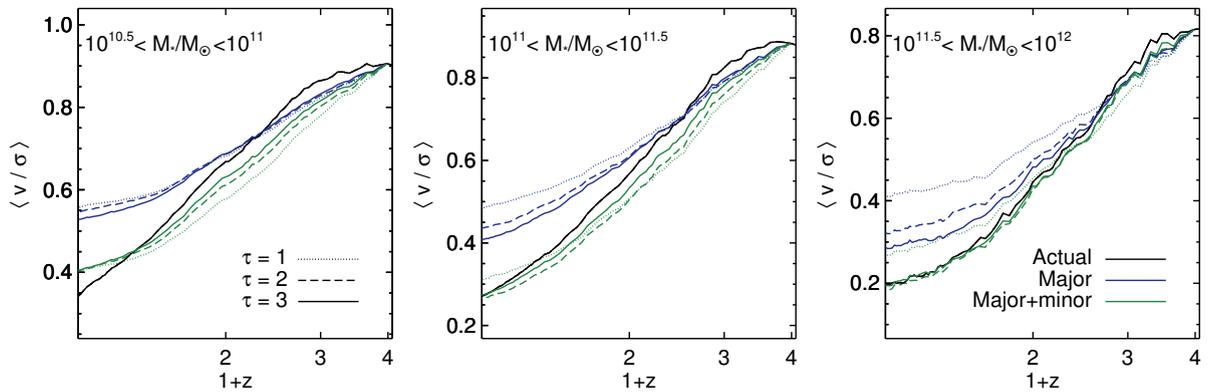


FIGURE B.2: Left-hand panel of Fig 5.11, now plotted for merger timescales of 1 Gyr, 2 Gyrs or 3 Gyrs, indicated by dotted, dashed and solid lines respectively. Blue lines indicate morphological transformation due to major mergers alone, while green lines indicate morphological change that is due to major+minor mergers.

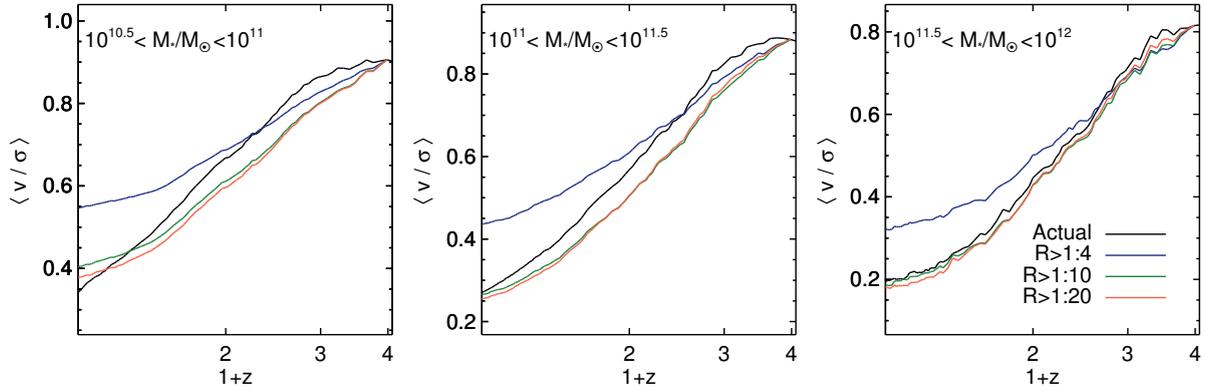


FIGURE B.3: Left-hand panel of Fig 5.11 plotted for additional mass ratio cuts.  $R$  represents the stellar mass ratio of mergers. Blue lines indicate morphological transformation due to major mergers alone ( $R > 1 : 4$ ) and green lines indicate morphological change that is due to major+minor mergers ( $R > 1 : 10$ ). Red lines indicate the morphological change due all mergers down to a mass ratio of 1:20.

coalescence, a large number of these systems have obvious asymmetries (i.e. the merger is not yet complete and the remnant has not relaxed), which will produce spurious values of  $v/\sigma$ . Thus, while timescales need to be short enough that significant amounts of morphological change are not missed or erroneously counted as merger-induced, it is important that the timescale is long enough that it probes the full duration of the merger.

Fig B.2 shows a version of the left-hand panels of Fig 5.11 (which describe the cumulative change in  $v/\sigma$  in the progenitors of today's spheroids) using different merger timescales. We show both the timescale we have adopted for this study (2 Gyr), as well as timescales of 3 Gyr and 1 Gyr (although note from the arguments above that a 1 Gyr timescale is inappropriately short for the mass ratios probed in this study). We find that reducing the merger timescale to 1 Gyr or increasing it to 3 Gyrs introduces only minimal change. We find good convergence between the lines showing the change in average  $v/\sigma$  due to mergers for the 2 Gyr and 3 Gyr timescales. The absolute difference between these lines is less than 0.05 for both major and major+minor mergers at all redshifts and in every mass bin). While there is good convergence, the difference between the 1 Gyr and 2 Gyr lines is slightly larger for the highest mass bin. The lines showing the change in average  $v/\sigma$  due to mergers, when calculated using 1 Gyr and 2 Gyr timescales, differ by 0.1 by  $z = 0$ , for  $M_*/M_\odot > 10^{11}$ ). Nevertheless, the choice of a 3 Gyr or even a 1 Gyr timescale does not qualitatively affect our results.

For large stellar masses ( $10^{11} < M_*/M_\odot < 10^{12}$ ), where the  $v/\sigma$  evolution attributed to mergers is most sensitive to the timescale chosen, the fraction of morphological transformation that we attribute to mergers (Equation 5.6, Fig 5.13) remains essentially unchanged. In the case of spheroids, there is a typical absolute increase of only 0.01 in the fraction of morphological change that we attribute to mergers, when a 3 Gyr timescale is used compared to a 2 Gyr timescale. When decreasing the timescale to 1 Gyr, there is a typical absolute decrease of 0.08 in terms of the fraction of morphological change attributed to mergers when compared with a 2 Gyr timescale. The values are similar in the case of discs in the same mass range. There is a typical absolute increase of 0.03 for a 3 Gyr timescale and typical absolute

decrease of 0.06 for a 1 Gyr timescale. The exact choice of timescale (1 - 3 Gyrs) is, therefore, relatively unimportant.

Finally, in Fig B.3 we show a version of the left-hand panels of Fig 5.11, now with an additional mass-ratio cut of 1:20. In line with the results of Fig B.1, we find that considering mass ratios less than 1:10 does not change the results of this study. In other words, mass ratios less than 1:10 produce negligible amounts of morphological transformation.

# Appendix C

## C.1 Extrapolating the mass and surface-brightness functions

Since the Horizon-AGN mass function becomes incomplete as we approach the mass resolution limit of the simulation ( $M_* \sim 10^8 M_\odot$ ), it is necessary to extrapolate the stellar mass and surface-brightness functions in order to obtain estimates of the contribution of LSBGs to the number, stellar mass and luminosity densities down to  $M_* \sim 10^7 M_\odot$ .

To do this, we first fit a Schechter function (Schechter, 1976) to the raw Horizon-AGN mass function yielding a slope of  $1.23 \pm_{0.05}^{0.03}$  (Figure C.1). We then use a Gaussian mixture model to fit the distribution of galaxies in the stellar mass – surface-brightness plane, allowing us to estimate the shape and variance of the data. Beyond the resolution limit ( $2 \times 10^8 M_\odot$ ), we linearly extrapolate the variance down to  $10^7 M_\odot$  (Figure C.2).

In order to obtain the extrapolated surface-brightness function, we split the mass function into narrow mass bins and draw  $N$  times from a Gaussian distribution with a variance and mean defined by our fit to the stellar mass – surface-brightness distribution, where  $N$  is the number of objects in that mass bin. Where the mass function is complete (above  $10^9 M_\odot$ ), we fill in the surface-brightness function using the raw data. The resultant surface-brightness is shown in Figure 6.3.

## C.2 Effect of the spatial resolution of the simulation

In this section, we discuss the effect of the spatial resolution of the simulation (1 kpc) on the sizes (and therefore surface-brightnesses) of low-mass galaxies. Although the locus of the  $M_*$ – $R_{\text{eff}}$  distribution only barely reaches 1 kpc at our mass limit ( $2 \times 10^8 M_\odot$ ; e.g. see Figure 6.2), it is still possible that the resolution may produce some spurious dynamical support. This problem would likely be compounded if the maximum resolution is not satisfied in the cells at the centres of our galaxies.

We first check the refinement level (i.e. the accuracy used by the gravity and hydrodynamics solvers; see Teyssier (2002)) of the AMR grid within the central  $R_{\text{eff}}$ . As noted in Section 6.2, the AMR grid

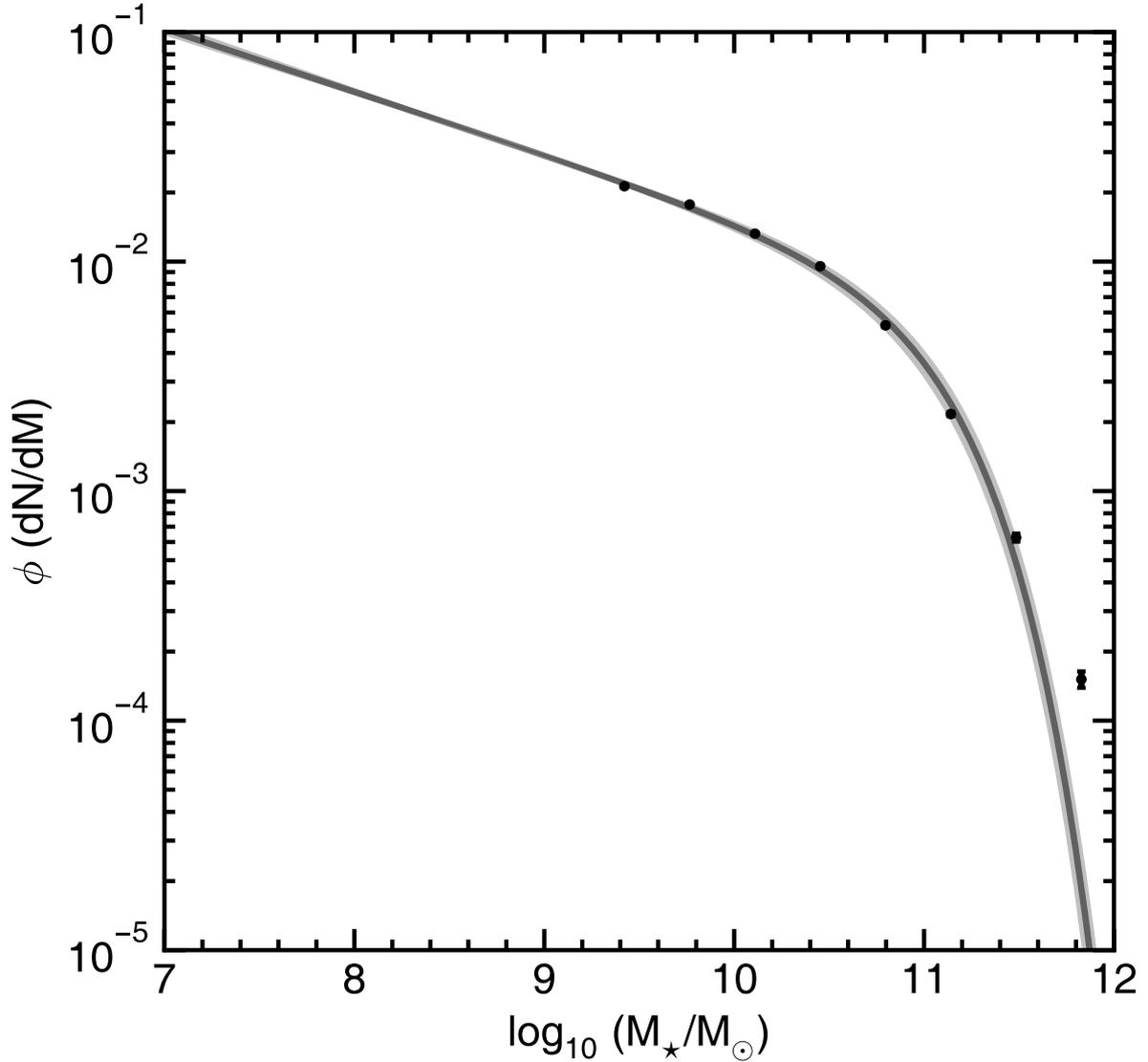


FIGURE C.1: Schechter function fit to the Horizon-AGN mass function. Points with error bars indicate binned data with Poisson errors. The grey region indicates the 99 per cent confidence interval for the fit to the data.

is refined according to a semi-Lagrangian criterion, where the refinement of a cell is approximately proportional to the total mass within the cell. Table C.1 shows the refinement of the AMR cells within 1 and  $2 R_{\text{eff}}$  of each galaxy used for our sample of UDGs and HSBGs in Section 6.4. On average almost 100 per cent of the AMR cells within  $1 R_{\text{eff}}$  of each galaxy are refined to the maximum level (level 17; 1 kpc) for both UDGs and HSBGs and within  $2 R_{\text{eff}}$ . The value falls to 21 per cent for UDGs, owing to their larger effective radii (i.e. they extend much further from the centre of the total mass distribution). In both cases, all cells are refined to at least the second highest level (level 16; 2 kpc).

We also check how the effective radii of equivalent galaxies in a higher resolution,  $4000 \text{ Mpc}^3$  zoom-in of a region of Horizon-AGN (New Horizon; Dubois et al. in preparation) differ from those in the Horizon-AGN simulation. New Horizon has a spatial resolution 35 pc ( $\times 30$  Horizon-AGN) but uses

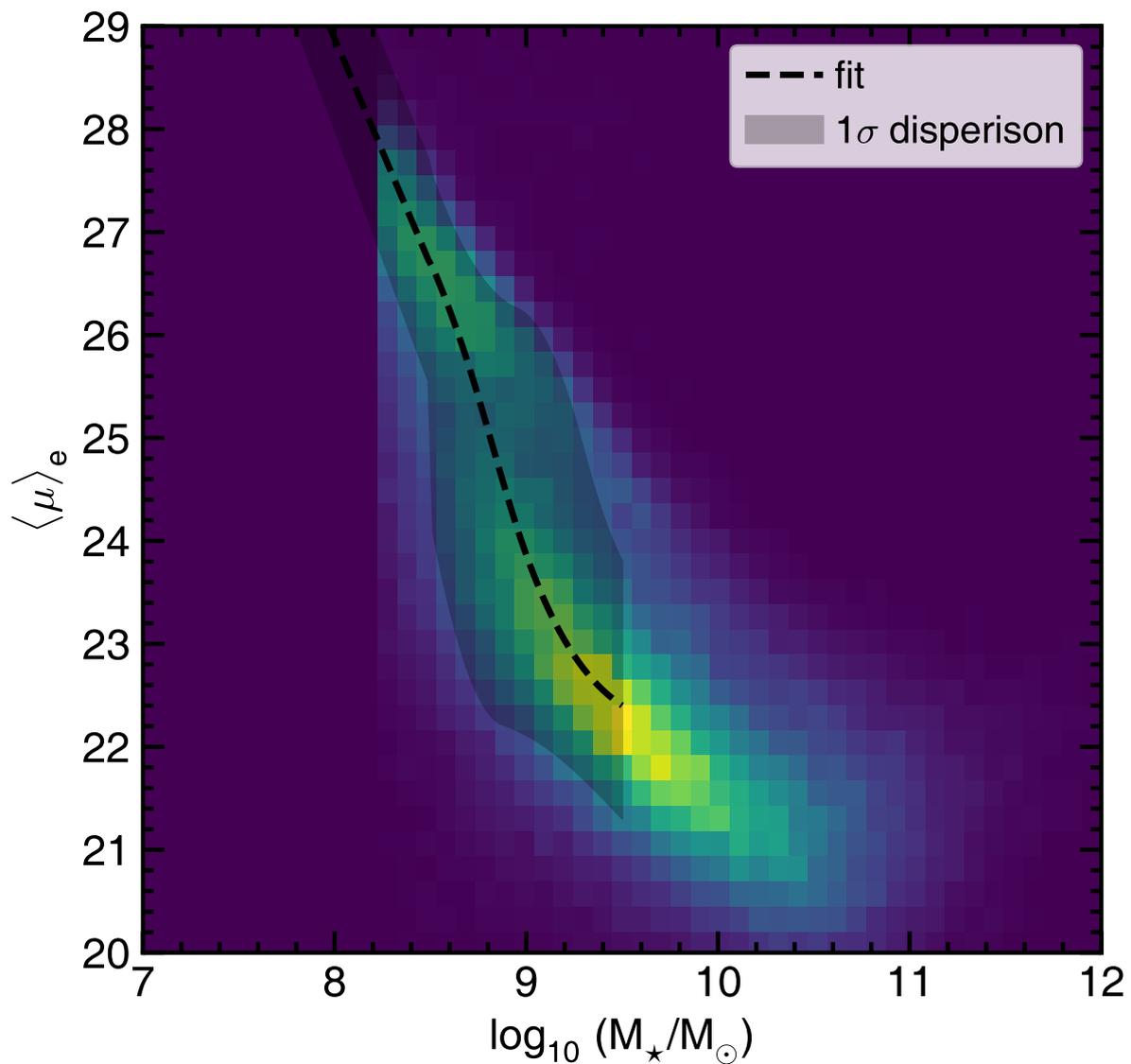


FIGURE C.2: Density plot showing the distribution of galaxies in the stellar mass – surface-brightness plane. The dashed black line indicates the fit to the data using a mixture of Gaussians and the filled grey region indicates the  $1\sigma$  dispersion.

UDGs	level 17 (1 kpc)	$\geq$ level 16 ( $\geq 2$ kpc)
$R < R_{\text{eff}}$	98%	100%
$R_{\text{eff}} < R < 2 R_{\text{eff}}$	21%	100%
HSBs		
$R < R_{\text{eff}}$	100%	100%
$R_{\text{eff}} < R < 2 R_{\text{eff}}$	88%	100%

TABLE C.1: The percentage of AMR cells within  $1 R_{\text{eff}}$  or  $2 R_{\text{eff}}$  that are refined to level 17 or at least level 16. The table is split between the UDG sample and the HSB sample.

the same underlying code (RAMSES Teyssier, 2002) and implements similar sub-grid prescriptions. The comparison is made at  $z = 0.7$ , the lowest redshift to which the New Horizon simulation has been run.

In order to produce a matching catalogue of galaxies, at the initial snapshot, the particle IDs of multiple (64) DM particles in the high resolution simulation were mapped onto each DM particle in Horizon-AGN. This allows us to match galaxy haloes between simulations and thereby attempt to find each galaxy’s ‘twin’ in the New Horizon simulation. We limit ourselves to haloes that share at least 75 per cent of the same DM particles, have at least 75 per cent of the mass of their matching halo and which host galaxies with stellar masses that are no more than a factor of two different from their twin. This yields a sample of 50 galaxies with masses between  $2 \times 10^8 M_{\odot}$  and  $10^{10} M_{\odot}$ .

Figure C.3 shows the effective radii and stellar masses of galaxies that we were able to robustly match between the two simulations, with each pair of twin galaxies joined by a dashed grey line. While the much higher resolution of the New Horizon simulation produces differences in the accretion histories and star formation of haloes compared with their twin haloes in the Horizon-AGN simulation, the lower resolution of the Horizon-AGN simulation does not produce a significant systematic offset in galaxy sizes. On average, galaxies in the Horizon-AGN simulation have only marginally larger sizes. The mean of the distribution of size offsets in Figure C.3 (denoted by a red arrow) is  $0.1 \pm 0.04$ , so that galaxies in Horizon-AGN are only 10 per cent larger, on average, than their twin in New Horizon. Note that the higher-resolution twin is often the larger of the two (26 per cent of higher-resolution galaxies are slightly larger).

The black lines in Figure C.3 show the trend in  $R_{\text{eff}}$  vs  $M_{\star}$  for the whole sample of galaxies within the same volume as New Horizon, regardless of whether they are reliably matched. Again, the average sizes of galaxies in Horizon-AGN are only around 10 per cent larger than equivalent mass galaxies in New Horizon.

We note however, that there is some degree of systematic offset between the average sizes of the simulated and observed galaxies (e.g. see blue filled points vs blue open points in Fig 6.2). This is especially pronounced at the high stellar mass end ( $\sim 10^{11.5} M_{\odot}$ ), where, compared to Cappellari et al. (2011, open circles), simulated galaxy sizes are around 1.5 – 2 times larger than observed galaxies of equivalent mass. Towards lower masses ( $\sim 10^{10} M_{\odot}$ ), the typical sizes of simulated galaxies are only 1.15 times larger than those of observed galaxies. This may be an indication that the AGN feedback prescription produces artificially large galaxies at high stellar mass (where AGN feedback is most efficient), but is not such an important effect at low masses, where AGN feedback becomes relatively unimportant compared to stellar feedback. For example, Peirani et al. (2018, see their Figure 1) show that the AGN feedback implementation used in Horizon-AGN produces galaxies that are larger than observed galaxies compared to the same simulation with no AGN feedback at masses of  $M_{\star} \approx 10^{11.5} M_{\odot}$ .

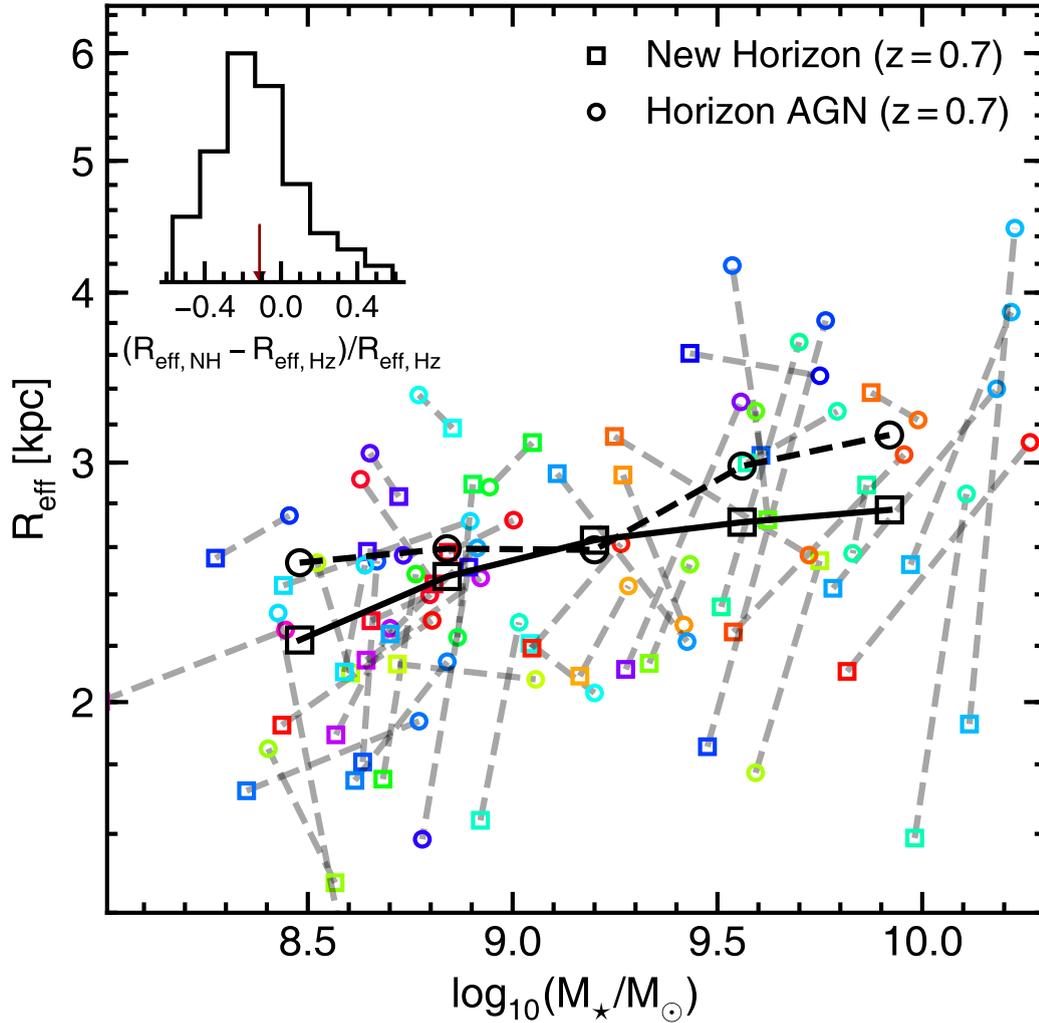


FIGURE C.3: Comparison of the effective radius and stellar mass of galaxies with haloes matched between the Horizon-AGN simulation and higher-resolution (35 pc) New Horizon simulation. Square and circle markers of the same colour linked by a dashed line indicate the stellar mass and effective radius of a matched galaxy in the New Horizon simulation (open square) and the Horizon-AGN simulation (open circle). The black solid and dashed lines show the mean trend in  $R_{\text{eff}}$  vs  $M_{\star}$  for New Horizon and the matching volume in Horizon-AGN respectively. Errors are not shown in the interest of legibility, but the typical error on the mean is  $\pm 0.1$  kpc in each bin. The inset plot shows the distribution of the fractional difference in  $R_{\text{eff}}$  between the two simulations and the red arrow shows the mean fractional difference.

### C.3 Relevance of this study to observed LSBG populations

In this section, we discuss the relevance and applicability of this study to observed UDG populations. Figure C.4 compares the mass distribution of cluster UDGs in the Horizon-AGN simulation (with a correction for incompleteness applied as detailed in Appendix C.1) to that of observed UDGs in the Coma and Virgo clusters (van Dokkum et al., 2015; Mihos et al., 2015; Yagi et al., 2016; Gu et al., 2018). Within the mass range where both samples overlap, there is good agreement between the mass distribution of Horizon-AGN UDGs (red histogram) and the observed cluster UDGs (blue histogram). Assuming a log-normal distribution for the observed sample, we also find that Horizon-AGN agrees within a factor of a few with the extrapolated value for the observed sample for  $M_* > 10^9 M_\odot$ .

Although high mass UDGs ( $M_* > 10^9 M_\odot$ ) are largely missing from observations, the very limited volumes explored observationally to date do not preclude the existence of galaxies with significantly larger stellar masses that satisfy the same low-surface-brightness criteria as their less massive counterparts. Indeed, examples of such massive LSBGs are already known e.g. Malin 1 and UGC 1382 (see large open red squares in Fig 6.2). Furthermore, the dashed black line in Figure C.4 indicates the galaxy stellar mass function from Baldry et al. (2008). A decline in the UDG fraction towards higher stellar masses should be expected and is likely driven by a combination of a mass dependence in the efficiency of the physical processes (e.g. SN feedback) that drive the formation of LSBGs (e.g. Brook & Di Cintio, 2015; van Dokkum et al., 2016; Toloba et al., 2018) and the steep decline in the galaxy stellar mass function towards higher stellar masses (which is exacerbated by the small observational volumes probed so far).

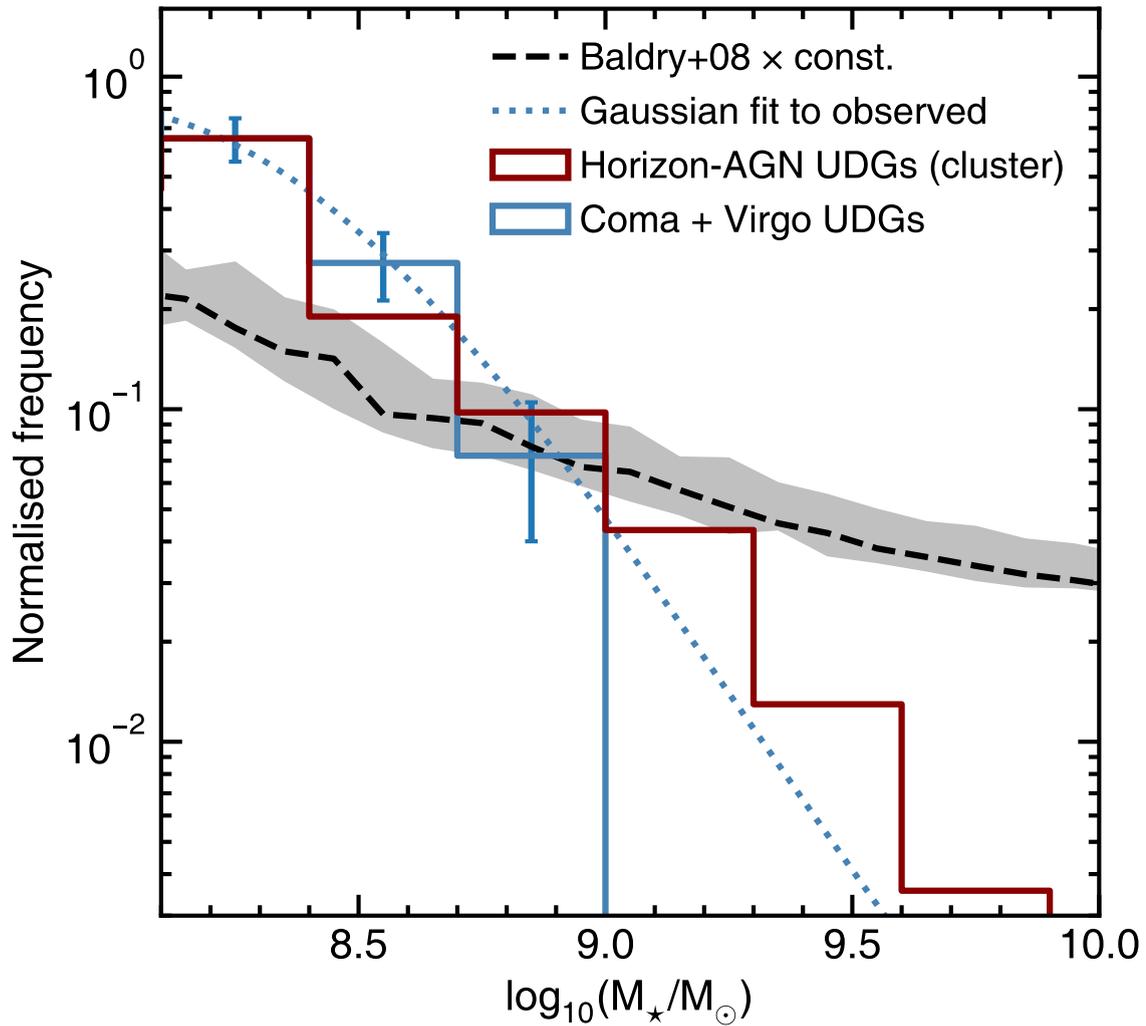


FIGURE C.4: The blue histogram shows the normalised stellar mass distribution for UDGs in the Coma and Virgo clusters (red open squares in Figure 6.2) and the blue dotted line shows a log-normal fit to the full distribution of masses. The red histogram shows the stellar mass distribution for Horizon-AGN cluster UDGs after an extrapolation of the stellar mass function has been performed as detailed in Appendix C.1. The Horizon-AGN UDG and observed cluster UDG histograms are both normalised by dividing by the number of counts in the three bins where the datasets overlay (between  $10^8 M_{\odot}$  and  $10^9 M_{\odot}$ ). The dashed black line indicates the non-parametric galaxy stellar mass function from Baldry et al. (2008) multiplied by a constant for clarity and the shaded region indicates the error.

# Appendix D

## D.1 Cross-matching the UML objects to HSC-SSP DR1 centroids

Fig D.1 shows the cumulative frequency of objects detected by the UML algorithm within a single tract of HSC-SSP data, as a function of their size. The size of each object is measured by the number of pixels they consist of, which is determined by the number of connected components above the sigma-clipping level in each object. The grey dashed histogram shows the cumulative number of objects larger than 10 pixels detected by the UML algorithm. The blue solid histogram indicates the cumulative number of these objects whose centroids can be matched to objects in the HSC-SSP DR1 catalogue within  $0.8''$ . Although the number of objects successfully matched is close to the total number of objects detected by the UML algorithm for small sizes ( $n_{\text{pixels}} \lesssim 50$ ), objects with larger sizes are significantly less likely to be matched.

The mismatch between centroids becomes a significant problem for large objects, with almost no objects with sizes larger than 200 pixels being matched. We note, however, that this mismatch does not present a significant problem for our analysis, as we consider only intermediate redshift ( $z > 0.3$ ) objects, which typically have small sizes. It would, however, be advantageous to select detected pixels from the object footprints taken directly from a catalogue that we hope to match to (i.e. in this case, from the stacked `calexp` images from HSC-SSP DR1) – as discussed at the beginning of Section 7.5. This is likely to yield more reliable cross-matching, especially for large, nearby objects and might be necessary for a perfect one-to-one matching.

## D.2 Released data products - lists of morphological clusters and individual galaxy properties

In this Appendix, we present the tables that form the data release from this chapter. Table D.1 describes the morphological clusters with their associated visual classifications and median values of key galaxy properties within the cluster (surface-brightness, stellar mass, specific SFR, rest-frame ( $g - r$ ) colour

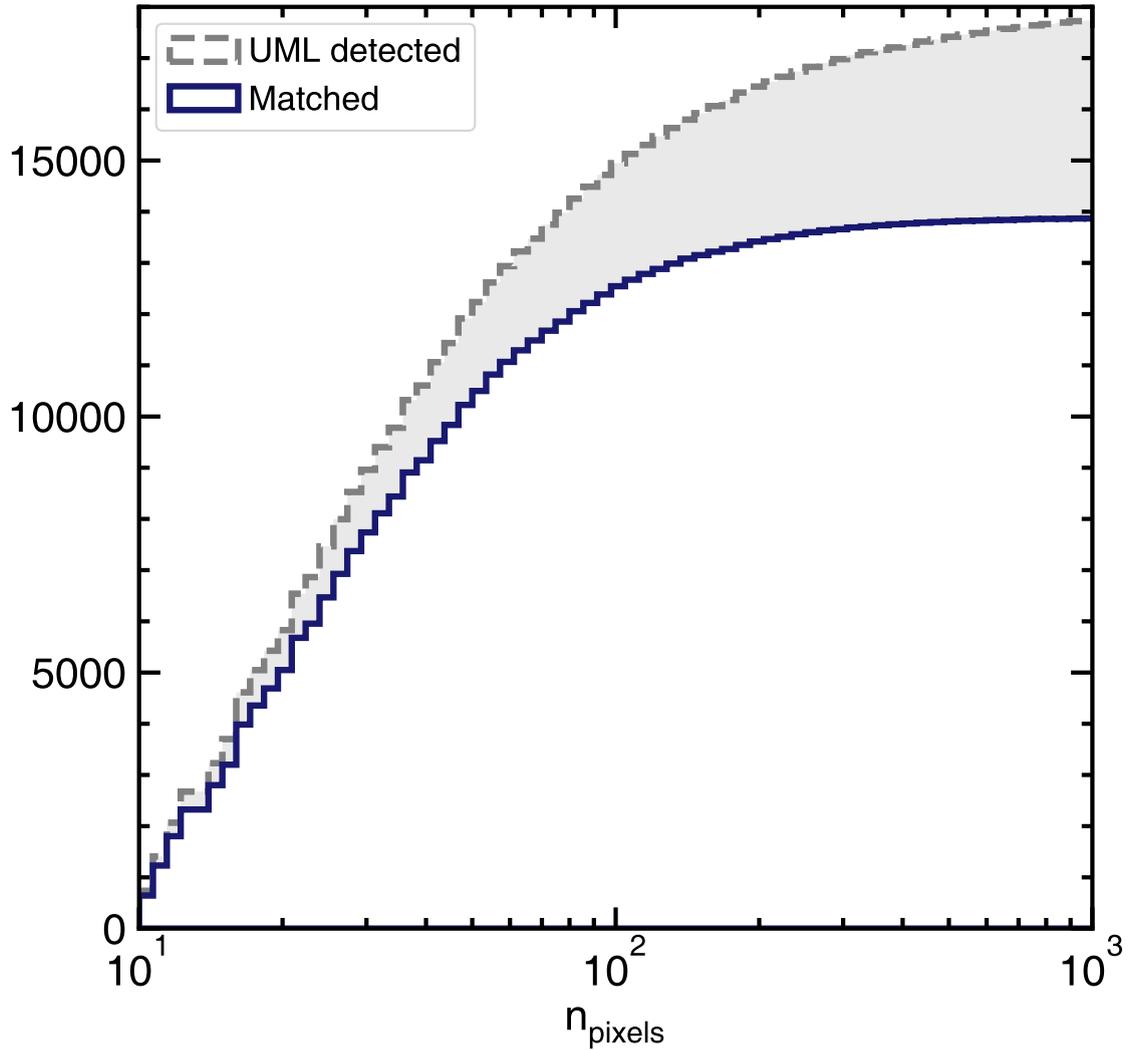


FIGURE D.1: The grey dotted histogram indicates the cumulative number of objects within a single tract, with sizes larger than 10 pixels, that are detected by the UML algorithm as a function of their size in pixels. The blue solid histogram indicates the same for objects whose centroid is matched to within  $0.8''$  of an object in the HSC-SSP DR1 catalogue. The shaded region shows the difference between the two histograms.

and absolute  $r$ -band magnitude). Table D.2 (only the first ten rows are shown here) provides a list of individual HSC-SSP galaxies with their associated morphological cluster IDs and ancillary information. As noted above, users should discard objects which are classified as not extended, as they are likely to be stars.

TABLE D.1: Average quantities (and their  $1\sigma$  dispersions) for objects in individual morphological clusters (the cluster ID is indicated by the first column, #). Columns are as follows: (a) the number of *matched* objects in the morphological cluster, (b) the number of objects identified as extended by the HSC-SSP pipeline, (c) median redshift, (d) median surface brightness in mag arcsec $^{-2}$ , (e) median stellar mass, (f) median sSFR in  $M_{\odot} \text{ yr}^{-1}$ , (g) median rest-frame  $g - r$  colour, (h) median  $g$ -band absolute magnitude. This final column details the dominant classified morphology (Hubble type or ‘St’ for star or ‘Sp’ for spirals where specific Hubble type could not be decided) of each morphological cluster including any other notable features.

#	$N^a$	$N_{ext}^b$	$z^c$	$\langle \mu_g \rangle_{Kron}^d$	$\log_{10}(M_*/M_{\odot})^e$	$\log_{10}(sSFR)^f$	$(g - r)^g$	$M_g^h$	Comment
0	1379	1122	$0.41 \pm 0.14$	$23.05 \pm 0.41$	$10.52 \pm 0.40$	$-13.11 \pm 2.00$	$0.67 \pm 0.09$	$-21.09 \pm 0.94$	E
1	110	110	$0.76 \pm 0.18$	$24.60 \pm 0.55$	$10.77 \pm 0.56$	$-10.41 \pm 1.03$	$0.64 \pm 0.15$	$-22.08 \pm 1.14$	S0/Sa
2	418	418	$0.72 \pm 0.24$	$24.08 \pm 0.37$	$10.02 \pm 0.46$	$-9.45 \pm 0.55$	$0.37 \pm 0.10$	$-21.36 \pm 1.13$	S0/Sa
3	267	267	$0.52 \pm 0.19$	$23.70 \pm 0.33$	$9.94 \pm 0.47$	$-9.53 \pm 0.64$	$0.39 \pm 0.11$	$-20.74 \pm 2.89$	Sp
4	107	88	$0.31 \pm 0.16$	$21.87 \pm 0.36$	$10.17 \pm 0.41$	$-9.62 \pm 1.28$	$0.45 \pm 0.12$	$-20.81 \pm 4.58$	E
5	169	168	$0.28 \pm 0.33$	$23.70 \pm 0.50$	$9.09 \pm 0.48$	$-9.37 \pm 0.43$	$0.25 \pm 0.07$	$-19.08 \pm 1.45$	Sb/Sc
6	519	507	$0.34 \pm 0.11$	$22.55 \pm 0.29$	$9.89 \pm 0.43$	$-9.53 \pm 0.47$	$0.39 \pm 0.10$	$-20.54 \pm 1.15$	Sa/Sd
7	171	148	$0.35 \pm 0.06$	$22.44 \pm 0.25$	$10.73 \pm 0.21$	$-13.65 \pm 1.77$	$0.70 \pm 0.05$	$-21.35 \pm 0.56$	S0
8	87	83	$1.06 \pm 0.57$	$25.80 \pm 1.40$	$10.17 \pm 0.89$	$-9.68 \pm 1.57$	$0.40 \pm 0.19$	$-21.61 \pm 1.98$	Sp, diffuse
9	443	440	$0.72 \pm 0.22$	$24.29 \pm 0.54$	$10.66 \pm 0.46$	$-9.85 \pm 0.46$	$0.54 \pm 0.10$	$-22.06 \pm 1.06$	Sb/Sc
10	253	253	$0.20 \pm 0.19$	$22.71 \pm 0.43$	$9.47 \pm 0.46$	$-9.52 \pm 0.38$	$0.33 \pm 0.10$	$-19.43 \pm 1.43$	Sc/Sd
11	1383	1349	$0.54 \pm 0.16$	$23.88 \pm 0.44$	$10.62 \pm 0.49$	$-11.52 \pm 1.95$	$0.67 \pm 0.09$	$-21.42 \pm 2.08$	E
12	0	—	— ± —	— ± —	— ± —	— ± —	— ± —	— ± —	St
13	755	371	$0.32 \pm 0.31$	$22.24 \pm 0.55$	$9.36 \pm 0.72$	$-9.35 \pm 2.30$	$0.27 \pm 0.13$	$-19.74 \pm 6.73$	St
14	308	299	$1.09 \pm 0.78$	$24.63 \pm 0.73$	$9.52 \pm 1.12$	$-9.26 \pm 0.49$	$0.23 \pm 0.12$	$-21.00 \pm 3.04$	Sp, disturbed
15	726	720	$0.29 \pm 0.47$	$23.92 \pm 0.53$	$8.93 \pm 0.83$	$-9.30 \pm 0.26$	$0.23 \pm 0.08$	$-18.71 \pm 2.39$	Sa/Sb/Sc, many edge-on
16	1921	1918	$0.76 \pm 0.40$	$24.35 \pm 0.56$	$9.79 \pm 0.72$	$-9.38 \pm 0.53$	$0.32 \pm 0.12$	$-21.07 \pm 1.79$	Sp
17	307	59	$0.04 \pm 0.68$	$19.10 \pm 0.82$	$9.22 \pm 0.89$	$0.00 \pm 4.05$	$0.36 \pm 0.14$	$-16.92 \pm 20.93$	St
18	210	209	$0.44 \pm 0.14$	$23.06 \pm 0.65$	$10.43 \pm 0.41$	$-9.69 \pm 0.88$	$0.53 \pm 0.09$	$-21.25 \pm 1.06$	S0/Sa
19	2513	2396	$0.46 \pm 0.24$	$23.15 \pm 0.42$	$9.68 \pm 0.60$	$-9.42 \pm 0.72$	$0.31 \pm 0.12$	$-20.38 \pm 2.17$	Sa/Sb/Sc
20	83	4	$0.26 \pm 0.13$	$20.47 \pm 1.33$	$9.05 \pm 0.96$	$-9.13 \pm 3.99$	$0.23 \pm 0.06$	$-19.88 \pm 3.31$	St
21	19	19	$0.09 \pm 0.64$	$22.73 \pm 1.20$	$8.72 \pm 1.12$	$-9.10 \pm 0.67$	$0.19 \pm 0.14$	$-18.43 \pm 3.71$	St
22	996	139	$0.23 \pm 0.32$	$21.55 \pm 0.73$	$9.45 \pm 1.10$	$-9.33 \pm 3.14$	$0.28 \pm 0.16$	$-19.60 \pm 10.42$	St
23	1	0	— ± —	— ± —	— ± —	— ± —	— ± —	— ± —	St
24	29	7	$0.00 \pm 0.27$	$21.19 \pm 0.56$	$10.29 \pm 1.66$	$0.00 \pm 4.46$	$1.21 \pm 0.40$	$21.31 \pm 19.45$	St
25	628	109	$0.41 \pm 0.25$	$22.71 \pm 0.81$	$10.83 \pm 0.54$	$-14.62 \pm 2.47$	$0.70 \pm 0.21$	$-21.56 \pm 4.52$	St
26	114	114	$0.14 \pm 0.08$	$20.83 \pm 0.54$	$10.45 \pm 0.52$	$-10.48 \pm 2.09$	$0.64 \pm 0.12$	$-20.43 \pm 3.92$	Sp
27	596	596	$0.42 \pm 0.23$	$23.63 \pm 0.42$	$9.84 \pm 0.53$	$-9.45 \pm 0.36$	$0.35 \pm 0.09$	$-20.60 \pm 1.37$	Sa/Sb
28	215	213	$0.65 \pm 0.26$	$23.93 \pm 0.32$	$10.06 \pm 0.47$	$-9.47 \pm 0.37$	$0.38 \pm 0.11$	$-21.19 \pm 1.11$	Sp, asymmetries
29	582	578	$0.91 \pm 0.66$	$25.15 \pm 0.92$	$9.63 \pm 0.89$	$-9.41 \pm 0.81$	$0.28 \pm 0.14$	$-20.79 \pm 2.26$	Sp, diffuse
30	276	275	$0.80 \pm 0.14$	$24.10 \pm 0.41$	$11.08 \pm 0.31$	$-10.11 \pm 0.75$	$0.61 \pm 0.10$	$-22.83 \pm 0.66$	S0/Sa
31	172	172	$0.29 \pm 0.07$	$22.14 \pm 0.52$	$10.96 \pm 0.25$	$-14.65 \pm 1.80$	$0.73 \pm 0.05$	$-21.53 \pm 0.68$	E
32	22	22	$0.50 \pm 0.26$	$24.59 \pm 0.62$	$9.33 \pm 0.66$	$-9.43 \pm 0.17$	$0.31 \pm 0.09$	$-19.78 \pm 1.67$	Sb/Sc, asymmetries
33	322	313	$0.49 \pm 0.13$	$23.32 \pm 0.47$	$10.50 \pm 0.47$	$-10.48 \pm 2.21$	$0.62 \pm 0.09$	$-21.32 \pm 1.09$	S0/Sa
34	112	110	$0.79 \pm 0.21$	$25.24 \pm 0.75$	$11.09 \pm 0.55$	$-11.21 \pm 1.71$	$0.70 \pm 0.14$	$-22.48 \pm 1.02$	S0/Sa, asymmetries
35	229	229	$0.38 \pm 0.41$	$23.53 \pm 0.52$	$9.17 \pm 0.89$	$-9.24 \pm 0.33$	$0.22 \pm 0.10$	$-19.48 \pm 2.52$	Sb/Sc
36	28	19	$0.08 \pm 0.59$	$18.69 \pm 0.24$	$8.96 \pm 0.66$	$0.00 \pm 5.06$	$0.36 \pm 0.12$	$-18.69 \pm 17.99$	St
37	22	16	$0.18 \pm 0.14$	$19.82 \pm 0.78$	$10.86 \pm 0.46$	$-10.35 \pm 4.83$	$0.71 \pm 0.24$	$-20.41 \pm 19.18$	St
38	97	94	$0.90 \pm 0.42$	$25.19 \pm 0.79$	$10.77 \pm 0.50$	$-10.35 \pm 0.97$	$0.59 \pm 0.13$	$-22.35 \pm 1.20$	Sp, red
39	546	45	$0.21 \pm 0.37$	$20.83 \pm 1.78$	$9.64 \pm 0.98$	$-9.55 \pm 4.01$	$0.37 \pm 0.19$	$-19.95 \pm 13.89$	St
40	178	152	$0.35 \pm 0.67$	$23.02 \pm 0.74$	$8.81 \pm 1.28$	$-9.08 \pm 0.96$	$0.15 \pm 0.13$	$-19.15 \pm 3.91$	E, blue
41	0	—	— ± —	— ± —	— ± —	— ± —	— ± —	— ± —	St
42	186	186	$0.36 \pm 0.10$	$22.83 \pm 0.21$	$10.74 \pm 0.37$	$-10.46 \pm 1.64$	$0.67 \pm 0.06$	$-21.35 \pm 0.95$	Sa/Sb
43	404	402	$0.56 \pm 0.15$	$23.72 \pm 0.31$	$10.81 \pm 0.44$	$-10.03 \pm 0.96$	$0.62 \pm 0.09$	$-21.97 \pm 1.04$	Sa
44	274	138	$0.00 \pm 0.40$	$20.79 \pm 1.16$	$11.77 \pm 0.82$	$0.00 \pm 5.45$	$1.21 \pm 0.36$	$20.71 \pm 21.99$	St
45	275	270	$0.89 \pm 0.49$	$25.02 \pm 0.67$	$10.11 \pm 0.75$	$-9.52 \pm 0.76$	$0.40 \pm 0.14$	$-21.33 \pm 1.76$	Sa/Sb/Sc
46	275	20	$0.12 \pm 0.55$	$20.76 \pm 0.88$	$8.88 \pm 1.14$	$-9.08 \pm 4.09$	$0.25 \pm 0.21$	$-18.11 \pm 12.01$	St
47	544	522	$0.74 \pm 0.35$	$24.81 \pm 0.59$	$10.91 \pm 0.52$	$-11.01 \pm 1.67$	$0.68 \pm 0.15$	$-22.23 \pm 2.39$	E
48	179	179	$0.20 \pm 0.13$	$22.83 \pm 0.29$	$8.90 \pm 0.54$	$-9.37 \pm 0.24$	$0.23 \pm 0.07$	$-18.47 \pm 1.60$	Sc/Sd
49	192	192	$0.84 \pm 0.30$	$24.64 \pm 0.78$	$10.64 \pm 0.46$	$-9.84 \pm 1.24$	$0.51 \pm 0.13$	$-22.13 \pm 1.00$	Sa/Sb
50	191	191	$0.18 \pm 0.18$	$21.80 \pm 0.30$	$9.27 \pm 0.53$	$-9.39 \pm 0.79$	$0.26 \pm 0.09$	$-19.33 \pm 1.76$	Sp
51	248	248	$0.36 \pm 0.11$	$22.63 \pm 0.27$	$10.49 \pm 0.34$	$-9.80 \pm 1.08$	$0.56 \pm 0.09$	$-21.11 \pm 1.08$	Sa
52	271	270	$0.42 \pm 0.29$	$23.85 \pm 0.39$	$9.41 \pm 0.70$	$-9.38 \pm 0.25$	$0.28 \pm 0.10$	$-19.89 \pm 1.83$	Sb/Sc
53	363	362	$0.28 \pm 0.22$	$22.95 \pm 0.44$	$9.29 \pm 0.74$	$-9.40 \pm 0.23$	$0.27 \pm 0.11$	$-19.51 \pm 1.88$	Sa/Sb
54	456	454	$0.47 \pm 0.14$	$23.49 \pm 0.34$	$10.83 \pm 0.45$	$-10.35 \pm 1.30$	$0.67 \pm 0.09$	$-21.73 \pm 1.09$	S0/Sa
55	1	1	$0.05 \pm 0.00$	$24.13 \pm 0.00$	$7.47 \pm 0.00$	$-9.42 \pm 0.00$	$0.25 \pm 0.00$	$-14.34 \pm 0.00$	St
56	4	4	$0.30 \pm 0.29$	$20.14 \pm 0.38$	$11.26 \pm 1.48$	$-10.06 \pm 4.43$	$0.55 \pm 0.22$	$-19.59 \pm 17.41$	St
57	260	256	$1.01 \pm 0.41$	$24.82 \pm 0.72$	$10.84 \pm 0.61$	$-10.00 \pm 0.92$	$0.56 \pm 0.14$	$-22.69 \pm 1.38$	S0/Sa
58	173	173	$0.49 \pm 0.38$	$24.12 \pm 0.48$	$9.50 \pm 0.79$	$-9.39 \pm 0.29$	$0.29 \pm 0.09$	$-20.28 \pm 2.15$	Sb/Sc

#	$N^a$	$N_{ext}^b$	$z^c$	$\langle \mu_g \rangle_{Kron}^d$	$\log_{10}(M_*/M_{\odot})^e$	$\log_{10}(sSFR)^f$	$(g-r)^g$	$M_g^h$	Comment
59	40	40	0.12 ± 0.24	23.09 ± 1.41	9.44 ± 0.85	-9.63 ± 1.25	0.39 ± 0.15	-19.02 ± 2.16	St
60	85	80	0.87 ± 0.67	25.32 ± 0.87	9.60 ± 0.91	-9.44 ± 0.51	0.30 ± 0.15	-20.60 ± 2.27	Sp, diffuse
61	2184	2176	0.66 ± 0.22	23.98 ± 0.39	10.17 ± 0.52	-9.62 ± 0.70	0.45 ± 0.12	-21.30 ± 1.56	E
62	111	109	0.25 ± 0.21	21.92 ± 1.00	10.96 ± 0.38	-14.76 ± 1.83	0.73 ± 0.09	-21.54 ± 0.83	St
63	3	3	1.49 ± 0.54	24.39 ± 1.00	11.59 ± 0.57	-11.31 ± 0.89	0.28 ± 0.06	-26.10 ± 2.11	Sb/Sc
64	196	196	0.65 ± 0.21	24.31 ± 0.44	9.89 ± 0.42	-9.41 ± 0.32	0.34 ± 0.09	-21.02 ± 1.02	St
65	91	43	0.00 ± 0.34	21.30 ± 0.20	11.98 ± 0.03	0.00 ± 4.55	1.37 ± 0.30	21.58 ± 18.31	Sa
66	162	160	0.21 ± 0.09	22.02 ± 0.44	10.68 ± 0.35	-13.13 ± 2.05	0.73 ± 0.08	-20.60 ± 0.94	St
67	1	1	0.02 ± 0.00	nan ± nan	7.45 ± 0.00	-11.45 ± 0.00	0.54 ± 0.00	-12.37 ± 0.00	S0/Sa
68	102	100	0.89 ± 0.36	25.18 ± 0.95	10.90 ± 0.53	-10.59 ± 1.43	0.64 ± 0.13	-22.49 ± 1.20	Sa/Sb/Sc
69	420	419	0.36 ± 0.15	22.62 ± 0.40	9.81 ± 0.42	-9.48 ± 0.42	0.35 ± 0.10	-20.43 ± 1.19	Sc/Sd
70	202	199	0.62 ± 0.51	23.55 ± 0.58	9.28 ± 1.07	-9.17 ± 0.52	0.21 ± 0.11	-20.74 ± 3.06	Sp, asymmetries
71	358	334	0.22 ± 0.11	21.72 ± 0.39	10.27 ± 0.42	-9.83 ± 1.21	0.53 ± 0.12	-20.39 ± 1.32	S0/Sa
72	57	56	0.12 ± 0.12	20.76 ± 1.97	10.57 ± 0.67	-13.64 ± 2.19	0.74 ± 0.16	-20.42 ± 1.23	E
73	0	-	- ± -	- ± -	- ± -	- ± -	- ± -	- ± -	St
74	507	507	0.70 ± 0.29	24.47 ± 0.42	9.78 ± 0.64	-9.39 ± 0.43	0.32 ± 0.11	-20.94 ± 1.53	Sb/Sc
75	128	125	0.82 ± 0.31	23.42 ± 0.69	10.02 ± 0.77	-9.23 ± 0.61	0.26 ± 0.11	-22.02 ± 2.22	Sa/Sb
76	71	71	0.30 ± 0.79	24.37 ± 0.70	8.72 ± 1.09	-9.22 ± 0.28	0.21 ± 0.11	-18.27 ± 3.09	Sc/Sd
77	0	-	- ± -	- ± -	- ± -	- ± -	- ± -	- ± -	St
78	194	194	0.16 ± 0.06	21.55 ± 0.33	9.79 ± 0.38	-9.64 ± 0.81	0.40 ± 0.12	-19.86 ± 0.79	Sa/Sb
79	314	314	0.25 ± 0.41	23.66 ± 0.39	8.75 ± 0.73	-9.29 ± 0.28	0.21 ± 0.07	-18.27 ± 2.21	Sc/Sd
80	147	147	1.03 ± 0.23	25.16 ± 0.72	10.64 ± 0.53	-10.00 ± 0.97	0.51 ± 0.13	-22.26 ± 1.13	Sp, red
81	419	417	0.52 ± 0.27	23.49 ± 0.37	10.38 ± 0.43	-9.67 ± 0.51	0.50 ± 0.09	-21.29 ± 1.11	S0/Sa
82	843	842	0.68 ± 0.20	24.26 ± 0.42	10.49 ± 0.43	-9.92 ± 0.57	0.54 ± 0.10	-21.67 ± 1.06	S0/Sa
83	15	13	0.70 ± 0.46	24.52 ± 1.40	11.20 ± 0.44	-11.81 ± 2.34	0.68 ± 0.18	-22.49 ± 0.63	St
84	229	229	0.38 ± 0.17	22.89 ± 0.29	10.09 ± 0.34	-9.55 ± 0.58	0.42 ± 0.10	-20.92 ± 0.94	Sa
85	0	-	- ± -	- ± -	- ± -	- ± -	- ± -	- ± -	St
86	276	267	0.22 ± 0.16	22.62 ± 0.46	9.13 ± 0.47	-9.39 ± 0.61	0.25 ± 0.08	-19.05 ± 2.87	S0/Sa, blue
87	35	33	0.31 ± 0.84	24.47 ± 1.49	9.41 ± 0.90	-9.39 ± 0.62	0.26 ± 0.16	-20.04 ± 2.82	E, blue
88	131	131	0.13 ± 0.06	21.09 ± 0.76	10.61 ± 0.44	-11.45 ± 2.01	0.71 ± 0.10	-20.36 ± 1.15	S0/Sa
89	2	2	1.12 ± 0.00	nan ± nan	11.33 ± 0.00	-11.31 ± 0.00	0.31 ± 0.00	-25.24 ± 0.00	St
90	430	429	0.34 ± 0.27	23.43 ± 0.45	9.56 ± 0.62	-9.44 ± 0.31	0.31 ± 0.10	-20.05 ± 1.62	Sc/Sd
91	436	434	0.33 ± 0.13	22.65 ± 0.39	10.14 ± 0.37	-9.62 ± 0.40	0.46 ± 0.09	-20.87 ± 1.07	Sb/Sc
92	3	3	0.55 ± 0.06	21.78 ± 0.56	11.41 ± 0.35	-14.62 ± 2.45	0.76 ± 0.16	-22.98 ± 0.19	St
93	93	91	1.19 ± 0.98	25.28 ± 1.10	9.15 ± 1.06	-9.26 ± 0.52	0.23 ± 0.14	-20.05 ± 2.94	Sb/Sc/Sd, diffuse
94	235	235	0.18 ± 0.08	22.17 ± 0.46	10.26 ± 0.38	-9.95 ± 1.70	0.56 ± 0.12	-20.16 ± 0.99	Sb/Sc
95	205	171	0.34 ± 0.12	22.43 ± 0.45	10.34 ± 0.36	-10.16 ± 1.69	0.62 ± 0.10	-20.92 ± 1.04	E
96	900	893	0.44 ± 0.11	23.27 ± 0.42	10.74 ± 0.39	-10.75 ± 1.85	0.68 ± 0.09	-21.46 ± 0.92	S0/Sa
97	152	152	0.29 ± 0.11	22.56 ± 0.29	10.74 ± 0.36	-10.29 ± 1.93	0.70 ± 0.10	-21.17 ± 1.10	Sa
98	284	284	0.46 ± 0.29	23.79 ± 0.46	9.56 ± 0.66	-9.37 ± 0.28	0.29 ± 0.10	-20.29 ± 1.73	E
99	602	593	0.54 ± 0.24	23.88 ± 0.49	10.72 ± 0.43	-10.71 ± 1.69	0.67 ± 0.11	-21.63 ± 0.98	Sp, clumpy
100	55	12	0.71 ± 0.43	24.46 ± 1.20	10.35 ± 1.04	-9.70 ± 1.71	0.42 ± 0.22	-21.86 ± 2.09	St
101	161	161	0.18 ± 0.06	21.28 ± 0.23	10.44 ± 0.33	-11.79 ± 1.96	0.68 ± 0.09	-20.42 ± 0.85	Sa
102	23	22	0.07 ± 0.39	20.38 ± 1.23	10.49 ± 0.56	-11.64 ± 2.23	0.74 ± 0.17	-19.67 ± 1.42	saturated
103	244	244	0.51 ± 0.31	23.88 ± 0.37	9.57 ± 0.53	-9.32 ± 0.25	0.29 ± 0.08	-20.44 ± 1.43	Sb/Sc/Sd
104	350	347	0.38 ± 0.19	23.15 ± 0.39	9.79 ± 0.65	-9.42 ± 0.76	0.33 ± 0.09	-20.50 ± 1.72	Sb/Sc
105	0	-	- ± -	- ± -	- ± -	- ± -	- ± -	- ± -	St
106	0	-	- ± -	- ± -	- ± -	- ± -	- ± -	- ± -	St
107	60	56	1.10 ± 0.66	25.94 ± 1.33	10.52 ± 0.62	-10.47 ± 1.85	0.55 ± 0.18	-22.01 ± 1.75	E
108	0	-	- ± -	- ± -	- ± -	- ± -	- ± -	- ± -	St
109	152	152	0.15 ± 0.07	21.93 ± 0.40	9.57 ± 0.59	-9.54 ± 0.85	0.33 ± 0.13	-19.56 ± 1.53	Sp, clumpy
110	73	71	0.43 ± 0.21	23.16 ± 0.65	10.28 ± 0.56	-9.59 ± 1.32	0.44 ± 0.14	-21.21 ± 1.54	Sb/Sc
111	4	3	1.95 ± 0.65	25.74 ± 0.84	9.92 ± 0.26	-9.00 ± 0.09	0.13 ± 0.05	-22.63 ± 0.62	S0/Sa
112	80	78	0.80 ± 0.16	24.19 ± 0.51	11.16 ± 0.57	-10.38 ± 0.81	0.64 ± 0.15	-22.82 ± 0.96	saturated
113	719	49	0.37 ± 0.37	21.52 ± 1.27	11.16 ± 1.14	-9.37 ± 5.19	0.76 ± 0.45	-20.56 ± 21.43	Sa/Sb, red
114	344	325	0.34 ± 0.19	22.42 ± 0.47	9.65 ± 0.58	-9.45 ± 0.78	0.33 ± 0.10	-20.15 ± 2.83	St
115	45	34	0.13 ± 0.86	19.02 ± 0.31	9.61 ± 0.59	0.00 ± 4.48	0.41 ± 0.20	-19.86 ± 20.14	saturated
116	312	312	0.33 ± 0.15	23.05 ± 0.41	9.97 ± 0.47	-9.58 ± 0.93	0.41 ± 0.12	-20.49 ± 1.19	St
117	171	171	0.58 ± 0.52	24.46 ± 0.60	9.22 ± 0.92	-9.27 ± 0.57	0.23 ± 0.12	-19.93 ± 2.41	Sb/Sc
118	262	261	0.48 ± 0.14	23.28 ± 0.25	10.47 ± 0.35	-9.66 ± 0.33	0.50 ± 0.07	-21.50 ± 0.94	Sb/Sc/Sd
119	442	441	0.48 ± 0.38	24.10 ± 0.36	9.31 ± 0.75	-9.30 ± 0.33	0.25 ± 0.09	-19.95 ± 2.08	S0/Sa, asymmetries, Sb/Sc
120	155	77	0.24 ± 0.10	21.55 ± 0.35	10.70 ± 0.23	-14.28 ± 4.42	0.71 ± 0.19	-20.78 ± 12.86	E
121	141	141	0.82 ± 0.42	24.49 ± 0.43	9.72 ± 0.80	-9.37 ± 0.27	0.28 ± 0.10	-21.08 ± 2.07	Sp, companions
122	123	123	0.11 ± 0.05	21.04 ± 0.44	9.78 ± 0.65	-9.61 ± 1.13	0.38 ± 0.14	-19.87 ± 1.50	Sa/Sb/Sd, blue ring

#	$N^a$	$N_{ext}^b$	$z^c$	$\langle\mu_g\rangle_{Kron}^d$	$\log_{10}(M_*/M_\odot)^e$	$\log_{10}(sSFR)^f$	$(g-r)^g$	$M_g^h$	Comment
123	0	—	— ± —	— ± —	— ± —	— ± —	— ± —	— ± —	St
124	12	7	0.00 ± 0.29	20.20 ± 0.22	11.62 ± 0.23	0.00 ± 4.77	1.20 ± 0.20	20.26 ± 21.78	St
125	1	1	0.28 ± 0.00	19.87 ± 0.00	11.34 ± 0.00	−10.72 ± 0.00	0.80 ± 0.00	−22.08 ± 0.00	St
126	209	208	0.31 ± 0.09	22.36 ± 0.24	10.50 ± 0.32	−10.59 ± 1.72	0.65 ± 0.08	−21.02 ± 0.93	E
127	158	154	0.13 ± 0.06	22.59 ± 0.41	9.05 ± 0.69	−9.38 ± 0.38	0.25 ± 0.11	−18.94 ± 1.76	Sc/Sd
128	146	146	0.11 ± 0.29	23.20 ± 0.89	8.57 ± 0.85	−9.32 ± 0.47	0.21 ± 0.11	−17.83 ± 2.48	Sd
129	301	275	0.47 ± 0.62	23.99 ± 1.80	10.02 ± 1.16	−9.48 ± 2.43	0.35 ± 0.24	−20.86 ± 7.33	Sp
130	193	193	0.35 ± 0.08	22.73 ± 0.34	10.99 ± 0.43	−12.47 ± 2.01	0.73 ± 0.07	−21.75 ± 1.04	E
131	1	1	1.53 ± 0.00	24.35 ± 0.00	11.38 ± 0.00	−10.04 ± 0.00	0.63 ± 0.00	−24.35 ± 0.00	St
132	183	174	0.26 ± 0.07	22.04 ± 0.30	10.62 ± 0.33	−14.23 ± 1.74	0.70 ± 0.06	−20.76 ± 0.84	E
133	1	0	— ± —	— ± —	— ± —	— ± —	— ± —	— ± —	saturated
134	271	238	0.18 ± 0.25	22.72 ± 0.42	8.94 ± 0.70	−9.36 ± 1.12	0.23 ± 0.10	−18.60 ± 4.32	Sp
135	82	81	1.24 ± 0.32	25.20 ± 1.19	10.62 ± 0.60	−9.47 ± 1.32	0.45 ± 0.16	−22.64 ± 1.26	Sp
136	1	1	1.07 ± 0.00	27.54 ± 0.00	9.53 ± 0.00	−9.61 ± 0.00	0.31 ± 0.00	−20.33 ± 0.00	saturated
137	1	1	0.02 ± 0.00	20.93 ± 0.00	7.45 ± 0.00	−11.45 ± 0.00	0.54 ± 0.00	−12.39 ± 0.00	St
138	931	930	0.40 ± 0.20	23.52 ± 0.33	9.67 ± 0.53	−9.44 ± 0.27	0.33 ± 0.09	−20.19 ± 1.36	Sa/Sb/Sc
139	141	141	0.63 ± 0.20	24.47 ± 0.47	10.85 ± 0.37	−11.21 ± 1.57	0.69 ± 0.11	−21.98 ± 0.80	E, asymmetries
140	148	141	0.42 ± 0.08	23.06 ± 0.35	10.92 ± 0.38	−14.58 ± 1.87	0.71 ± 0.08	−21.70 ± 0.72	E, companions
141	154	154	0.33 ± 0.29	23.96 ± 0.38	8.99 ± 0.77	−9.35 ± 0.47	0.25 ± 0.08	−18.85 ± 2.18	Sp, companions
142	2	2	0.21 ± 0.17	27.62 ± 0.00	8.11 ± 0.15	−10.46 ± 1.49	0.30 ± 0.23	−16.26 ± 1.71	saturated
143	117	116	0.91 ± 0.58	24.76 ± 0.76	10.05 ± 0.95	−9.35 ± 0.81	0.31 ± 0.13	−21.46 ± 2.38	Sb/Sd/Sc
144	129	129	0.95 ± 0.23	24.18 ± 0.46	10.63 ± 0.48	−9.37 ± 0.55	0.44 ± 0.12	−22.67 ± 1.01	Sp, clumpy, red
145	59	58	0.89 ± 0.34	25.28 ± 0.55	10.45 ± 0.69	−9.67 ± 1.21	0.49 ± 0.15	−21.62 ± 1.69	S0
146	258	257	0.74 ± 0.32	24.57 ± 0.60	9.86 ± 0.68	−9.39 ± 0.27	0.32 ± 0.11	−21.15 ± 1.59	Sa/Sb/Sc, companions
147	400	64	0.31 ± 0.26	22.00 ± 0.75	10.42 ± 0.41	−12.13 ± 1.84	0.65 ± 0.13	−20.87 ± 1.29	St
148	248	64	0.00 ± 0.83	20.04 ± 0.69	10.38 ± 0.33	0.00 ± 5.62	0.74 ± 0.27	19.82 ± 21.41	St
149	0	—	— ± —	— ± —	— ± —	— ± —	— ± —	— ± —	St
150	1	1	0.15 ± 0.00	20.58 ± 0.00	8.67 ± 0.00	−8.98 ± 0.00	0.09 ± 0.00	−18.63 ± 0.00	St
151	252	252	0.39 ± 0.09	22.88 ± 0.41	11.03 ± 0.26	−14.65 ± 1.81	0.73 ± 0.05	−21.90 ± 0.61	E
152	1307	1303	0.42 ± 0.20	23.36 ± 0.46	9.64 ± 0.50	−9.42 ± 0.28	0.31 ± 0.10	−20.20 ± 1.26	Sp, companions
153	88	86	1.11 ± 0.83	24.86 ± 1.06	9.78 ± 1.08	−9.34 ± 0.45	0.25 ± 0.11	−21.28 ± 3.02	Sd, asymmetries
154	39	18	0.70 ± 0.33	24.44 ± 0.57	11.42 ± 1.29	−11.63 ± 2.44	0.71 ± 0.19	−23.05 ± 2.73	St
155	244	244	0.22 ± 0.15	22.11 ± 0.29	9.64 ± 0.45	−9.51 ± 0.28	0.35 ± 0.10	−19.67 ± 1.23	Sa/Sb
156	116	107	0.60 ± 0.13	23.91 ± 0.38	11.17 ± 0.50	−14.91 ± 1.87	0.70 ± 0.09	−22.52 ± 1.08	E
157	141	140	0.88 ± 0.47	25.28 ± 0.79	9.76 ± 0.69	−9.44 ± 0.77	0.30 ± 0.13	−20.93 ± 1.77	Sb/Sc/Sd, asymmetries
158	0	—	— ± —	— ± —	— ± —	— ± —	— ± —	— ± —	St
159	353	353	0.82 ± 0.37	24.71 ± 0.59	9.97 ± 0.63	−9.46 ± 0.70	0.36 ± 0.12	−21.23 ± 1.53	Sp

TABLE D.2: Example of 10 entries from the catalogue showing the position and morphological cluster membership individual galaxies. Columns are as follows: (a) the RA of the centroid from the UML detection, (b) the declination of the centroid from the UML detection, (c) the RA of the centroid for the matched HSC object, (d) the declination of the centroid for the matched HSC object, (e) the ID of the matched HSC object, (f) whether the matched HSC object is extended or not, (g) the morphological cluster membership of the object, (h) the number of pixels that make up the UML detection, (i) the silhouette score for the object. Where there is no matching object in within  $0.8''$  in the HSC catalogue, the 3rd to 6th columns are left blank.

UML <sub>RA</sub> <sup>a</sup>	UML <sub>DEC</sub> <sup>b</sup>	HSC <sub>RA</sub> <sup>c</sup>	HSC <sub>DEC</sub> <sup>d</sup>	HSC ID <sup>e</sup>	extended <sup>f</sup>	cluster # <sup>g</sup>	$n_{pix}$ <sup>h</sup>	silhouette <sup>i</sup>
35.0484	−5.3316	—	—	—	—	33	116	0.1559
35.0484	−5.4808	35.0485	−5.4808	37489923418246937	False	132	112	−0.1047
35.0480	−5.4613	35.0479	−5.4612	37489923418227875	True	98	19	0.1081
35.0472	−5.3657	35.0471	−5.3657	37489923418254158	True	91	232	0.0831
35.0475	−5.3088	35.0475	−5.3086	37489923418258428	True	16	30	0.1947
35.0445	−5.3545	35.0444	−5.3544	37490060857210025	True	10	91	−0.0550
35.0438	−5.4894	35.0437	−5.4894	37490056562236660	True	75	50	0.3934
35.0436	−5.4011	—	—	—	—	152	53	0.0033
35.0433	−5.3099	35.0432	−5.3099	37490060857213384	True	79	67	0.1252
35.0429	−5.4777	35.0428	−5.4776	37490060857201353	True	58	26	0.3622

# Bibliography

- Abadi M. G., Moore B., Bower R. G., 1999, MNRAS, 308, 947
- Abazajian K. N., et al., 2009, The Astrophysical Journal Supplement Series, 182, 543
- Abbott T., et al., 2005, preprint, (arXiv:astro-ph/0510346)
- Abell P. A., et al., 2009, preprint, (arXiv:0912.0201)
- Abraham R. G., van Dokkum P. G., 2014, PASP, 126, 55
- Abraham R. G., Valdes F., Yee H. K. C., van den Bergh S., 1994, ApJ, 432, 75
- Abraham R., et al., 2018, Research Notes of the American Astronomical Society, 2, 16
- Agertz O., Teyssier R., Moore B., 2009, MNRAS, 397, L64
- Agertz O., Teyssier R., Moore B., 2011, MNRAS, 410, 1391
- Aihara H., et al., 2018a, Publications of the Astronomical Society of Japan, 70, S4
- Aihara H., et al., 2018b, Publications of the ASJ, 70, S8
- Akhlaghi M., Ichikawa T., 2015, The Astrophysical Journal Supplement Series, 220
- Amorisco N. C., Loeb A., 2016, MNRAS, 459, L51
- Amorisco N. C., Monachesi A., Agnello A., White S. D. M., 2016, preprint, (arXiv:1610.01595)
- Amorisco N. C., Monachesi A., Agnello A., White S. D. M., 2018, MNRAS, 475, 4235
- An F. X., et al., 2018, ApJ, 862, 101
- Aragon-Calvo M. A., Yang L. F., 2014, MNRAS, 440, L46
- Araya Salvo C., Mathur S., Ghosh H., Fiore F., Ferrarese L., 2012, ApJ, 757, 179
- Astropy Collaboration et al., 2013, A&A, 558, A33
- Athanassoula E., Rodionov S. A., Peschken N., Lambert J. C., 2016, ApJ, 821, 90

- Aubert D., Pichon C., Colombi S., 2004, MNRAS, 352, 376
- Aumer M., White S. D. M., Naab T., Scannapieco C., 2013, MNRAS, 434, 3142
- Ay F., İnce G., Kamaşak M. E., Ekşi K. Y., 2019, arXiv e-prints, p. arXiv:1904.04204
- Bahcall N. A., 1996, ArXiv Astrophysics e-prints,
- Bahé Y. M., et al., 2017, MNRAS, 470, 4186
- Bakos J., Trujillo I., 2012, preprint, (arXiv:1204.3082)
- Baldry I. K., Balogh M. L., Bower R. G., Glazebrook K., Nichol R. C., Bamford S. P., Budavari T., 2006, MNRAS, 373, 469
- Baldry I. K., Glazebrook K., Driver S. P., 2008, MNRAS, 388, 945
- Barnes J. E., 1988, ApJ, 331, 699
- Barnes J. E., 1992, ApJ, 393, 484
- Barrientos L. F., Lilly S. J., 2003, ApJ, 596, 129
- Barro G., et al., 2013, ApJ, 765, 104
- Baum W. A., 1959, Publications of the ASJ, 71, 106
- Baushev A. N., 2018, NewA, 60, 69
- Beasley M. A., Trujillo I., 2016, ApJ, 830
- Beasley M. A., Romanowsky A. J., Pota V., Navarro I. M., Martinez Delgado D., Neyer F., Deich A. L., 2016, ApJ, 819, L20
- Beck M. R., et al., 2018, MNRAS, 476, 5516
- Beckmann R. S., et al., 2017, preprint, (arXiv:1701.07838)
- Bédorf J., Portegies Zwart S., 2013, MNRAS, 431, 767
- Begelman M. C., Volonteri M., Rees M. J., 2006, MNRAS, 370, 289
- Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57
- Beifiori A., Sarzi M., Corsini E. M., Dalla Bontà E., Pizzella A., Coccato L., Bertola F., 2009, ApJ, 692, 856
- Bell E. F., et al., 2004, ApJ, 608, 752
- Bell E. F., Phleps S., Somerville R. S., Wolf C., Borch A., Meisenheimer K., 2006, ApJ, 652, 270

- Benitez N., et al., 2014, arXiv e-prints, p. arXiv:1403.5237
- Benson A. J., Bower R., 2011, MNRAS, 410, 2653
- Benson A. J., Bower R. G., Frenk C. S., Lacey C. G., Baugh C. M., Cole S., 2003, ApJ, 599, 38
- Bernardi M., et al., 2003, AJ, 125, 1817
- Bernstein G. M., Nichol R. C., Tyson J. A., Ulmer M. P., Wittman D., 1995, AJ, 110, 1507
- Bertin E., Arnouts S., 1996, Astronomy and Astrophysics Supplement Series, 117, 393
- Bertschinger E., 2001, ApJ, 137, 1
- Bertschinger E., 2008, COSMICS: Cosmological initial conditions and microwave anisotropy codes, Astrophysics Source Code Library (ascl:9910.004)
- Bessiere P. S., Tadhunter C. N., Ramos Almeida C., Villar Martín M., 2014, MNRAS, 438, 1839
- Best P. N., Heckman T. M., 2012, MNRAS, 421, 1569
- Béthermin M., et al., 2015, A&A, 573, A113
- Bezanson R., van Dokkum P. G., Tal T., Marchesini D., Kriek M., Franx M., Coppi P., 2009, ApJ, 697, 1290
- Bizzocchi L., et al., 2014, ApJ, 782, 22
- Blake C., et al., 2004, MNRAS, 355, 713
- Blandford R. D., Königl A., 1979, ApJ, 232, 34
- Blanton M. R., Lupton R. H., Schlegel D. J., Strauss M. A., Brinkmann J., Fukugita M., Loveday J., 2005, ApJ, 631, 208
- Blitz L., Shu F. H., 1980, ApJ, 238, 148
- Bluck A. F. L., Mendel J. T., Ellison S. L., Moreno J., Simard L., Patton D. R., Starkenburg E., 2014, MNRAS, 441, 599
- Blumenthal G. R., Faber S. M., Primack J. R., Rees M. J., 1984, Nature, 311, 517
- Bois M., et al., 2011, MNRAS, 416, 1654
- Booth C. M., Schaye J., 2009a, MNRAS, 398, 53
- Booth C. M., Schaye J., 2009b, in Heinz S., Wilcots E., eds, American Institute of Physics Conference Series Vol. 1201, American Institute of Physics Conference Series. pp 21–24, doi:10.1063/1.3293041
- Booth C. M., Schaye J., 2010, MNRAS, 405, L1

- Boselli A., Boissier S., Cortese L., Gavazzi G., 2008, *ApJ*, 674
- Bothun G. D., Impey C. D., Malin D. F., Mould J. R., 1987, *AJ*, 94, 23
- Bothun G. D., Impey C. D., Malin D. F., 1991, *ApJ*, 376, 404
- Bournaud F., 2010, in Smith B., Higdon J., Higdon S., Bastian N., eds, *Astronomical Society of the Pacific Conference Series Vol. 423, Galaxy Wars: Stellar Populations and Star Formation in Interacting Galaxies*. p. 177 (arXiv:0909.1812)
- Bournaud F., Duc P.-A., 2006, *A&A*, 456, 481
- Bournaud F., Duc P.-A., Emsellem E., 2008, *MNRAS*, 389, L8
- Bournaud F., Dekel A., Teyssier R., Cacciato M., Daddi E., Juneau S., Shankar F., 2011, *ApJ*, 741, L33
- Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, *MNRAS*, 370, 645
- Bower R. G., Schaye J., Frenk C. S., Theuns T., Schaller M., Crain R. A., McAlpine S., 2017, *MNRAS*, 465, 32
- Breiman L., Meisel W., Purcell E., 1977, *Technometrics*, 19, 135
- Brennan R., et al., 2015, *MNRAS*, 451, 2933
- Brook C. B., Di Cintio A., 2015, *MNRAS*, 450, 3920
- Brook C. B., et al., 2011, *MNRAS*, 415, 1051
- Brook C. B., Stinson G., Gibson B. K., Roškar R., Wadsley J., Quinn T., 2012, *MNRAS*, 419, 771
- Brooks A. M., Governato F., Quinn T., Brook C. B., Wadsley J., 2009, *ApJ*, 694, 396
- Bruzual G., Charlot S., 2003, *MNRAS*, 344, 1000
- Buitrago F., Trujillo I., Conselice C. J., Häußler B., 2013, *MNRAS*, 428, 1460
- Buitrago F., Conselice C. J., Epinat B., Bedregal A. G., Grützbauch R., Weiner B. J., 2014, *MNRAS*, 439, 1494
- Bundy K., Ellis R. S., Conselice C. J., 2005, *ApJ*, 625, 621
- Bundy K., Treu T., Ellis R. S., 2007, *ApJ*, 665, L5
- Buta R. J., 2011, preprint, (arXiv:1102.0550)
- Butcher H. R., Oemler Jr. A., 1984, *Nature*, 310, 31
- Byrd G., Valtonen M., 1990, *ApJ*, 350, 89

- Cabrera-Vives G., Miller C. J., Schneider J., 2018, *AJ*, 156, 284
- Capelo P. R., Volonteri M., Dotti M., Bellovary J. M., Mayer L., Governato F., 2015, *MNRAS*, 447, 2123
- Cappellari M., 2016, *Annual Review of Astronomy and Astrophysics*, 54, 597
- Cappellari M., et al., 2011, *MNRAS*, 416, 1680
- Carleton T., Errani R., Cooper M., Kaplinghat M., Peñarrubia J., 2018, preprint, p. arXiv:1805.06896 (arXiv:1805.06896)
- Carpinetti A., Kaviraj S., Hyde A. K., Clements D. L., Schawinski K., Darg D., Lintott C. J., 2015, *A&A*, 577, A119
- Cerulo P., et al., 2017, *MNRAS*, 472, 254
- Ceverino D., Dekel A., Bournaud F., 2010, *MNRAS*, 404, 2151
- Ceverino D., Primack J., Dekel A., Kassin S. A., 2017, *MNRAS*, 467, 2664
- Chabrier G., 2003, *Publications of the Astronomical Society of the Pacific*, 115, 763
- Chamaraux P., Balkowski C., Gerard E., 1980, *A&A*, 83, 38
- Chan T. K., Kereš D., Oñorbe J., Hopkins P. F., Muratov A. L., Faucher-Giguère C. A., Quataert E., 2015, *MNRAS*, 454, 2981
- Chan T. K., Kereš D., Wetzel A., Hopkins P. F., Faucher-Giguère C.-A., El-Badry K., Garrison-Kimmel S., Boylan-Kolchin M., 2018, *MNRAS*, 478, 906
- Chandrasekhar S., 1943, *ApJ*, 97, 255
- Chiaberge M., Gilli R., Lotz J. M., Norman C., 2015, *ApJ*, 806, 147
- Chisari N., et al., 2016, *MNRAS*, 461, 2702
- Choi H., Yi S. K., 2017, *ApJ*, 837, 68
- Choi H., Yi S. K., Dubois Y., Kimm T., Devriendt J. E. G., Pichon C., 2018, preprint, (arXiv:1802.05916)
- Cibinel A., et al., 2015, *ApJ*, 805, 181
- Cirasuolo M., Afonso J., Bender R., Bonifacio P., Evans C., Kaper L., Oliva E., Vanzi L., 2011, *The Messenger*, 145, 11
- Cisternas M., et al., 2011a, *ApJ*, 726, 57
- Cisternas M., et al., 2011b, *ApJ*, 741, L11

- Clauwens B., Franx M., Schaye J., 2016, MNRAS, 463, L1
- Clauwens B., Schaye J., Franx M., Bower R. G., 2017, preprint, (arXiv:1711.00030)
- Codis S., Pichon C., Devriendt J., Slyz A., Pogosyan D., Dubois Y., Sousbie T., 2012, MNRAS, 427, 3320
- Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168
- Comerford J. M., Pooley D., Barrows R. S., Greene J. E., Zakamska N. L., Madejski G. M., Cooper M. C., 2015, ApJ, 806, 219
- Conselice C. J., 2003, ApJ, 147, 1
- Conselice C. J., 2006, MNRAS, 373, 1389
- Conselice C. J., 2014, Annual Review of Astronomy and Astrophysics, 52, 291
- Conselice C. J., 2018, Research Notes of the American Astronomical Society, 2, 43
- Conselice C. J., Arnold J., 2009, MNRAS, 397, 208
- Conselice C. J., et al., 2007, MNRAS, 381, 962
- Conselice C. J., Rajgor S., Myers R., 2008, MNRAS, 386, 909
- Conselice C. J., Yang C., Bluck A. F. L., 2009, MNRAS, 394, 1956
- Conselice C. J., Mortlock A., Bluck A. F. L., Grützbauch R., Duncan K., 2013, MNRAS, 430, 1051
- Conselice C. J., Bluck A. F. L., Mortlock A., Palamara D., Benson A. J., 2014, MNRAS, 444, 1125
- Covington M. D., et al., 2010, ApJ, 710, 279
- Cox T. J., Jonsson P., Somerville R. S., Primack J. R., Dekel A., 2008, MNRAS, 384, 386
- Croton D. J., et al., 2006, MNRAS, 365, 11
- Croton D. J., et al., 2016, ApJ, 222, 22
- D’Isanto A., Polsterer K. L., 2018, A&A, 609, A111
- D’Isanto A., Cavauoti S., Gieseke F., Polsterer K. L., 2018, A&A, 616, A97
- Dabringhausen J., Kroupa P., 2013, MNRAS, 429, 1858
- Daddi E., et al., 2005, ApJ, 631, L13
- Daddi E., et al., 2007, ApJ, 670, 156
- Dalcanton J., Spergel D. N., Gunn J. E., Schmidt M., Schneider D. P., 1997, AJ, 114, 2178

- Danovich M., Dekel A., Hahn O., Ceverino D., Primack J., 2015, MNRAS, 449, 2087
- Darg D. W., et al., 2010a, MNRAS, 401, 1043
- Darg D. W., et al., 2010b, MNRAS, 401, 1552
- Dark Energy Survey Collaboration et al., 2016, MNRAS, 460, 1270
- Dashyan G., Silk J., Mamon G. A., Dubois Y., Hartwig T., 2018, MNRAS, 473, 5698
- Davé R., 2008, MNRAS, 385, 147
- Davies J. I., Phillipps S., Disney M. J., 1990, MNRAS, 244, 385
- De Lucia G., Blaizot J., 2007, MNRAS, 375, 2
- Deeley S., et al., 2017, MNRAS, 467, 3934
- Dekel A., Ostriker J. P., eds, 1999, Formation of Structure in the Universe
- Dekel A., Sari R., Ceverino D., 2009, ApJ, 703, 785
- Devriendt J., et al., 2010, MNRAS, 403, L84
- Di Cintio A., Brook C. B., Macciò A. V., Stinson G. S., Knebe A., Dutton A. A., Wadsley J., 2014a, MNRAS, 437, 415
- Di Cintio A., Brook C. B., Dutton A. A., Macciò A. V., Stinson G. S., Knebe A., 2014b, MNRAS, 441, 2986
- Di Cintio A., Brook C. B., Dutton A. A., Macciò A. V., Obreja A., Dekel A., 2017, MNRAS, 466, L1
- Di Cintio A., Brook C. B., Macciò A. V., Dutton A. A., Cardona-Barrero S., 2019, arXiv e-prints, p. arXiv:1901.08559
- Di Matteo T., Springel V., Hernquist L., 2005, Nature, 433, 604
- Di Matteo P., Combes F., Melchior A.-L., Semelin B., 2007, A&A, 468, 61
- Di Matteo P., Bournaud F., Martig M., Combes F., Melchior A.-L., Semelin B., 2008, A&A, 492, 31
- Dickinson H., Fortson L., Scarlata C., Beck M., Walmsley M., 2019, arXiv e-prints, p. arXiv:1903.07776
- Diehl T., Dark Energy Survey Collaboration 2012, Physics Procedia, 37, 1332
- Dieleman S., Willett K. W., Dambre J., 2015, MNRAS, 450, 1441
- Disney M. J., 1976, Nature, 263, 573
- Djorgovski S. G., Mahabal A. A., Donalek C., Graham M. J., Drake A. J., Moghaddam B., Turmon M., 2012, arXiv e-prints,

- Dong S., Lin D. N. C., Murray S. D., 2003, *ApJ*, 596, 930
- Draine B. T., et al., 2007, *ApJ*, 663, 866
- Dressler A., 1980, *ApJ*, 236, 351
- Dressler A., 1984, *Annual Review of Astronomy and Astrophysics*, 22, 185
- Dressler A., et al., 1997, *ApJ*, 490, 577
- Dressler A., Smail I., Poggianti B. M., Butcher H., Couch W. J., Ellis R. S., Oemler Jr. A., 1999, *ApJ*, 122, 51
- Driver S. P., 1999, *ApJ*, 526, L69
- Driver S. P., Liske J., Cross N. J. G., De Propriis R., Allen P. D., 2005, *MNRAS*, 360, 81
- Dubois Y., Pichon C., Haehnelt M., Kimm T., Slyz A., Devriendt J., Pogosyan D., 2012, *MNRAS*, 423, 3616
- Dubois Y., et al., 2014, *MNRAS*, 444, 1453
- Dubois Y., Peirani S., Pichon C., Devriendt J., Gavazzi R., Welker C., Volonteri M., 2016, *MNRAS*, 463, 3948
- Duc P.-A., Brinks E., Wink J. E., Mirabel I. F., 1997, *A&A*, 326, 537
- Dutton A. A., Treu T., 2014, *MNRAS*, 438, 3594
- Dutton A. A., et al., 2016, *MNRAS*, 461, 2658
- Edgar R., 2004, *New Astronomy Review*, 48, 843
- Edgeworth F. Y., 1888, *Journal of the Royal Statistical Society*, 51, 346
- Efstathiou G., 1992, *MNRAS*, 256, 43P
- Eggen O. J., Lynden-Bell D., Sandage A. R., 1962, *ApJ*, 136, 748
- Eisenstein D. J., et al., 2001, *AJ*, 122, 2267
- Eisenstein D. J., et al., 2011, *AJ*, 142, 72
- El-Badry K., Wetzel A., Geha M., Hopkins P. F., Kereš D., Chan T. K., Faucher-Giguère C.-A., 2016, *ApJ*, 820, 131
- Elbaz D., Cesarsky C. J., 2003, *Science*, 300, 270
- Elbaz D., et al., 2007, *A&A*, 468, 33
- Ellis G. F. R., 1975, *Quarterly Journal of the RAS*, 16, 245

- Errani R., Penarrubia J., Tormen G., 2015, MNRAS, 449, L46
- Faber S. M., et al., 2007, ApJ, 665, 265
- Fabian A. C., 2012, Annual Review of Astronomy and Astrophysics, 50, 455
- Fadely R., Hogg D. W., Willman B., 2012, ApJ, 760, 15
- Fall S. M., Efstathiou G., 1980, MNRAS, 193, 189
- Fan X., et al., 2006, AJ, 132, 117
- Fang J. J., Faber S. M., Salim S., Graves G. J., Rich R. M., 2012, ApJ, 761, 23
- Federrath C., Klessen R. S., 2012, ApJ, 761, 156
- Feng Y., Di Matteo T., Croft R., Khandai N., 2014, Monthly Notices of the Royal Astronomical Society, 440, 1865
- Fensch J., et al., 2017, MNRAS, 465, 1934
- Ferdosi B. J., Buddelmeijer H., Trager S. C., Wilkinson M. H. F., Roerdink J. B. T. M., 2011, A&A, 531, A114
- Ferrarese L., Merritt D., 2000, ApJ, 539, L9
- Ferré-Mateu A., et al., 2018, MNRAS, p. 1525
- Ferreras I., Lisker T., Pasquali A., Khochfar S., Kaviraj S., 2009, MNRAS, 396, 1573
- Ferreras I., et al., 2014, MNRAS, 444, 906
- Fiacconi D., Feldmann R., Mayer L., 2015, MNRAS, 446, 1957
- Fierlinger K. M., Burkert A., Ntormousi E., Fierlinger P., Schartmann M., Ballone A., Krause M. G. H., Diehl R., 2016, MNRAS, 456, 710
- Filippenko A. V., Ho L. C., 2003, ApJ, 588, L13
- Fitzpatrick P. J., Graves G. J., 2015, MNRAS, 447, 1383
- Font A. S., McCarthy I. G., Crain R. A., Theuns T., Schaye J., Wiersma R. P. C., Dalla Vecchia C., 2011, MNRAS, 416, 2802
- Font A. S., McCarthy I. G., Le Brun A. M. C., Crain R. A., Kelvin L. S., 2017, preprint, (arXiv:1710.00415)
- Förster Schreiber N. M., et al., 2011, ApJ, 739, 45
- Franks W. S., 1905, MNRAS, 65, 228

- Franx M., van Dokkum P. G., 1996, in Bender R., Davies R. L., eds, IAU Symposium Vol. 171, New Light on Galaxy Evolution. p. 233 (arXiv:astro-ph/9603029)
- Freeman P. E., Izbicki R., Lee A. B., Newman J. A., Conselice C. J., Koekemoer A. M., Lotz J. M., Mozena M., 2013, MNRAS, 434, 282
- Fritzke B., 1995, in Advances in neural information processing systems. pp 625–632
- Gabor J. M., et al., 2009, ApJ, 691, 705
- Galler B. A., Fisher M. J., 1964, Communications of the ACM, 7, 301
- Gardner J. P., et al., 2006, Space Science Reviews, 123, 485
- Gavazzi G., Fumagalli M., Fossati M., Galardo V., Grossetti F., Boselli A., Giovanelli R., Haynes M. P., 2013, A&A, 553
- Geach J. E., Smail I., Moran S. M., MacArthur L. A., Lagos C. d. P., Edge A. C., 2011, ApJ, 730, L19
- George K., 2017, A&A, 598, A45
- George K., Zingade K., 2015, A&A, 583, A103
- George K., et al., 2018, A&A, 613, L9
- Ghosh H., Mathur S., Fiore F., Ferrarese L., 2008, ApJ, 687, 216
- Girardi L., Bressan A., Bertelli G., Chiosi C., 2000, A&A Supp., 141, 371
- Gladders M. D., López-Cruz O., Yee H. K. C., Kodama T., 1998, ApJ, 501, 571
- Glikman E., Simmons B., Maily M., Schawinski K., Urry C. M., Lacy M., 2015, ApJ, 806, 218
- Gnedin O. Y., 2003, ApJ, 582, 141
- González V., Bouwens R., Illingworth G., Labbé I., Oesch P., Franx M., Magee D., 2014, ApJ, 781, 34
- Goto T., 2005, MNRAS, 357, 937
- Goulding A. D., et al., 2018, Publications of the Astronomical Society of Japan, 70, S37
- Governato F., et al., 2010, Nature, 463, 203
- Greco J. P., et al., 2018a, Publications of the Astronomical Society of Japan, 70, S19
- Greco J. P., et al., 2018b, ApJ, 857, 104
- Greggio L., Renzini A., 1983, A&A, 118, 217
- Grier C. J., Mathur S., Ghosh H., Ferrarese L., 2011, ApJ, 731, 60

- Grogin N. A., et al., 2005, *ApJ*, 627, L97
- Grogin N. A., et al., 2011, *ApJ*, 197, 35
- Grupe D., Mathur S., 2004, *ApJ*, 606, L41
- Gruppioni C., et al., 2013, *MNRAS*, 432, 23
- Gu M., et al., 2018, *ApJ*, 859, 37
- Gunn J. E., Gott J. Richard I., 1972, *ApJ*, 176, 1
- Guth A. H., Pi S.-Y., 1982, *Physical Review Letters*, 49, 1110
- Haardt F., Madau P., 1996, *ApJ*, 461, 20
- Habas R., Fadda D., Marleau F. R., Biviano A., Durret F., 2018, *MNRAS*, 475, 4544
- Haberzettl L., Bomans D. J., Dettmar R. J., 2007, *A&A*, 471, 787
- Habouzit M., Volonteri M., Dubois Y., 2017, *MNRAS*, 468, 3935
- Hagen L. M. Z., et al., 2016, *ApJ*, 826, 210
- Hahn O., Teyssier R., Carollo C. M., 2010, *MNRAS*, 405, 274
- Haines T., McIntosh D. H., Sánchez S. F., Tremonti C., Rudnick G., 2015, *MNRAS*, 451, 433
- Han Z., Podsiadlowski P., Maxted P. F. L., Marsh T. R., 2003, *MNRAS*, 341, 669
- Hardcastle M. J., et al., 2019, *A&A*, 622, A12
- Häring N., Rix H.-W., 2004, *ApJ*, 604, L89
- Hartmann L., Ballesteros-Paredes J., Bergin E. A., 2001, *ApJ*, 562, 852
- Hasinger G., Miyaji T., Schmidt M., 2005, *A&A*, 441, 417
- Hatton S., Devriendt J. E. G., Ninin S., Bouchet F. R., Guiderdoni B., Vibert D., 2003, *MNRAS*, 343, 75
- Hau G. K. T., Bower R. G., Kilborn V., Forbes D. A., Balogh M. L., Oosterloo T., 2008, *MNRAS*, 385, 1965
- Hawking S. W., 1982, *Physics Letters B*, 115, 295
- Hayward C. C., Hopkins P. F., 2017, *MNRAS*, 465, 1682
- Hayward C. C., Irwin J. A., Bregman J. N., 2005, *ApJ*, 635, 827
- Hendel D., Johnston K. V., Patra R. K., Sen B., 2018, arXiv e-prints, p. arXiv:1811.10613

- Henrion M., Mortlock D. J., Hand D. J., Gand y A., 2011, MNRAS, 412, 2286
- Herlihy M., Shavit N., 2011, The art of multiprocessor programming. Morgan Kaufmann
- Hernquist L., 1992, ApJ, 400, 460
- Hernquist L., Mihos J. C., 1995, ApJ, 448, 41
- Hlavacek-Larrondo J., Fabian A. C., Edge A. C., Ebeling H., Allen S. W., Sanders J. S., Taylor G. B., 2013, MNRAS, 431, 1638
- Hocking A., Sun Y., Geach J. E., Davey N., 2017, in 2017 International Joint Conference on Neural Networks (IJCNN). pp 4179–4186
- Hocking A., Geach J. E., Sun Y., Davey N., 2018, MNRAS, 473, 1108
- Hogg D. W., et al., 2002, AJ, 124, 646
- Holwerda B. W., Pirzkal N., Heiner J. S., 2012, MNRAS, 427, 3159
- Hopkins A. M., Beacom J. F., 2006, ApJ, 651, 142
- Hopkins P. F., Somerville R. S., Hernquist L., Cox T. J., Robertson B., Li Y., 2006, ApJ, 652, 864
- Hopkins P. F., Cox T. J., Kereš D., Hernquist L., 2008, ApJ, 175, 390
- Hopkins P. F., et al., 2009a, MNRAS, 397, 802
- Hopkins P. F., Cox T. J., Younger J. D., Hernquist L., 2009b, ApJ, 691, 1168
- Hopkins P. F., Younger J. D., Hayward C. C., Narayanan D., Hernquist L., 2010, MNRAS, 402, 1693
- Hoyos C., Aragón-Salamanca A., Gray M. E., Wolf C., Maltby D. T., Bell E. F., Böhm A., Jogee S., 2016, MNRAS, 455, 295
- Hu C.-Y., Naab T., Walch S., Glover S. C. O., Clark P. C., 2016, MNRAS, 458, 3528
- Hubble E. P., 1926a, ApJ, 64
- Hubble E., 1926b, Contributions from the Mount Wilson Observatory / Carnegie Institution of Washington, 310, 1
- Hubble E. P., 1927, The Observatory, 50, 276
- Hubble E. P., 1936, Realm of the Nebulae. Yale University Press
- Huertas-Company M., et al., 2015a, ApJ, 221, 8
- Huertas-Company M., et al., 2015b, ApJ, 809, 95
- Huertas-Company M., et al., 2016, MNRAS, 462, 4495

- Ibata R. A., et al., 2013, *Nature*, 493, 62
- Ilbert O., et al., 2010, *ApJ*, 709, 644
- Impey C., Bothun G., 1997, *Annual Review of Astronomy and Astrophysics*, 35, 267
- Impey C., Bothun G., Malin D., 1988, *ApJ*, 330, 634
- Ishibashi W., Auger M. W., Zhang D., Fabian A. C., 2014, *MNRAS*, 443, 1339
- Jaffé Y. L., Aragón-Salamanca A., De Lucia G., Jablonka P., Rudnick G., Saglia R., Zaritsky D., 2011, *MNRAS*, 410, 280
- Jansen F., et al., 2001, *A&A*, 365, L1
- Janssens S., Abraham R., Brodie J., Forbes D., Romanowsky A. J., van Dokkum P., 2017, *ApJ*, 839, L17
- Jarvis M. J., et al., 2013, *MNRAS*, 428, 1281
- Jiang C. Y., Jing Y. P., Faltenbacher A., Lin W. P., Li C., 2008, *ApJ*, 675, 1095
- Jiang Y.-F., Greene J. E., Ho L. C., Xiao T., Barth A. J., 2011, *ApJ*, 742, 68
- Jiang F., Dekel A., Freundlich J., Romanowsky A. J., Dutton A. A., Macciò A. V., Di Cintio A., 2019, *MNRAS*, p. 1490
- Johansson P. H., Naab T., Ostriker J. P., 2009, *ApJ*, 697, L38
- Johnson S. C., 1967, *Psychometrika*, 32, 241
- Jones B. J. T., van de Weygaert R., Aragón-Calvo M. A., 2010, *MNRAS*, 408, 897
- Kannappan S. J., Guie J. M., Baker A. J., 2009, *AJ*, 138, 579
- Kassin S. A., et al., 2012, *ApJ*, 758, 106
- Kassin S. A., Brooks A., Governato F., Weiner B. J., Gardner J. P., 2014, *ApJ*, 790, 89
- Katz N., 1992, *ApJ*, 391, 502
- Kauffmann G., White S. D. M., Guiderdoni B., 1993, *MNRAS*, 264, 201
- Kaviraj S., 2010, *MNRAS*, 406, 382
- Kaviraj S., 2014a, *MNRAS*, 437, L41
- Kaviraj S., 2014b, *MNRAS*, 440, 2944
- Kaviraj S., et al., 2007, *ApJ*, 173, 619
- Kaviraj S., et al., 2008, *MNRAS*, 388, 67

- Kaviraj S., Peirani S., Khochfar S., Silk J., Kay S., 2009, MNRAS, 394, 1713
- Kaviraj S., Tan K.-M., Ellis R. S., Silk J., 2011, MNRAS, 411, 2148
- Kaviraj S., Darg D., Lintott C., Schawinski K., Silk J., 2012a, MNRAS, 419, 70
- Kaviraj S., et al., 2012b, MNRAS, 423, 49
- Kaviraj S., et al., 2013a, MNRAS, 428, 925
- Kaviraj S., et al., 2013b, MNRAS, 429, L40
- Kaviraj S., et al., 2013c, MNRAS, 435, 1463
- Kaviraj S., Shabala S. S., Deller A. T., Middelberg E., 2015a, MNRAS, 452, 774
- Kaviraj S., Devriendt J., Dubois Y., Slyz A., Welker C., Pichon C., Peirani S., Le Borgne D., 2015b, MNRAS, 452, 2845
- Kaviraj S., et al., 2016, preprint, (arXiv:1605.09379)
- Kaviraj S., et al., 2017, MNRAS, 467, 4739
- Kaviraj S., Martin G., Silk J., 2019, arXiv e-prints, p. arXiv:1906.10697
- Kelvin L. S., et al., 2014, MNRAS, 444, 1647
- Kennicutt Jr. R. C., 1998, ApJ, 498, 541
- Kereš D., Katz N., Weinberg D. H., Davé R., 2005, MNRAS, 363, 2
- Khandai N., Di Matteo T., Croft R., Wilkins S., Feng Y., Tucker E., DeGraf C., Liu M.-S., 2015, MNRAS, 450, 1349
- Khim H.-g., Park J., Seo S.-W., Lee J., Smith R., Yi S. K., 2015, The Astrophysical Journal Supplement Series, 220, 3
- Khochfar S., et al., 2011, MNRAS, 417, 845
- Kimm T., Devriendt J., Slyz A., Pichon C., Kassin S. A., Dubois Y., 2011, preprint, (arXiv:1106.0538)
- Kimm T., Cen R., Devriendt J., Dubois Y., Slyz A., 2015, MNRAS, 451, 2900
- Klein J. P., Moeschberger M. L., 2006, Survival analysis: techniques for censored and truncated data. Springer Science & Business Media
- Kniazev A. Y., Grebel E. K., Pustilnik S. A., Pramskij A. G., Kniazeva T. F., Prada F., Harbeck D., 2004, AJ, 127, 704

- Kocevski D. D., et al., 2012, *ApJ*, 744, 148
- Koda J., Yagi M., Yamanoi H., Komiyama Y., 2015, *ApJ*, 807, L2
- Komatsu E., et al., 2011, *ApJ*, 192, 18
- Kormendy J., Ho L. C., 2013, *Annual Review of Astronomy and Astrophysics*, 51, 511
- Kormendy J., Kennicutt Jr. R. C., 2004, *Annual Review of Astronomy and Astrophysics*, 42, 603
- Kormendy J., Drory N., Bender R., Cornell M. E., 2010, *ApJ*, 723, 54
- Kron R. G., 1980, *The Astrophysical Journal Supplement Series*, 43, 305
- Krongold Y., Nicastro F., Elvis M., Brickhouse N., Binette L., Mathur S., Jiménez-Bailón E., 2007, *ApJ*, 659, 1022
- Kuchinski L. E., et al., 2000, *ApJ*, 131, 441
- Kuijken K., et al., 2002, *The Messenger*, 110, 15
- Kuutma T., Tamm A., Tempel E., 2017, *A&A*, 600, L6
- Lackner C. N., Gunn J. E., 2012, *MNRAS*, 421, 2277
- Laganá T. F., Ulmer M. P., 2018, *MNRAS*, 475, 523
- Lahav O., et al., 1995, *Science*, 267, 859
- Laigle C., et al., 2018, *MNRAS*, 474, 5437
- Lamastra A., Menci N., Fiore F., Santini P., 2013, *A&A*, 552, A44
- Lange R., et al., 2015, *MNRAS*, 447, 2603
- Laporte C. F. P., Agnello A., Navarro J. F., 2018, preprint, p. arXiv:1804.04139 (arXiv:1804.04139)
- Laureijs R., et al., 2011, preprint, (arXiv:1110.3193)
- Lawrence A., et al., 2007, *MNRAS*, 379, 1599
- Lee J., Pen U.-L., 2000, *ApJ*, 532, L5
- Lee H., McCall M. L., Richer M. G., 2003, *AJ*, 125, 2975
- Lee B., et al., 2013, *ApJ*, 774, 47
- Leisman L., et al., 2017, *ApJ*, 842, 133
- Leitherer C., Robert C., Drissen L., 1992, *ApJ*, 401, 596

- Leitherer C., et al., 1999, *ApJ*, 123, 3
- Leitherer C., Ortiz Otálvaro P. A., Bresolin F., Kudritzki R.-P., Lo Faro B., Pauldrach A. W. A., Pettini M., Rix S. A., 2010, *ApJ*, 189, 309
- Leja J., van Dokkum P., Franx M., 2013, *ApJ*, 766, 33
- Lin L.-H., Wang H.-H., Hsieh P.-Y., Taam R. E., Yang C.-C., Yen D. C. C., 2013, *ApJ*, 771, 8
- Lintott C., et al., 2011, *MNRAS*, 410, 166
- Lisenfeld U., Xu C. K., Gao Y., Domingue D. L., Yun M. S., Zuo P., 2019, arXiv e-prints, p. arXiv:1906.00682
- Lofthouse E. K., Kaviraj S., Conselice C. J., Mortlock A., Hartley W., 2017a, *MNRAS*, 465, 2895
- Lofthouse E. K., Houghton R. C. W., Kaviraj S., 2017b, *MNRAS*, 471, 2311
- Longhetti M., et al., 2005, *MNRAS*, 361, 897
- López Fernández R., et al., 2018, *A&A*, 615, A27
- López-Sanjuan C., et al., 2012, *A&A*, 548, A7
- Lotz J. M., Primack J., Madau P., 2004, *AJ*, 128, 163
- Lotz J. M., Jonsson P., Cox T. J., Croton D., Primack J. R., Somerville R. S., Stewart K., 2011, *ApJ*, 742, 103
- Lu Y., Mo H. J., Weinberg M. D., Katz N., 2011, *MNRAS*, 416, 1949
- Lundmark K., 1926, *Arkiv for Matematik, Astronomi och Fysik*, 19, 1
- Ma C.-P., Bertschinger E., 1995, *ApJ*, 455, 7
- Ma Z., et al., 2019, *The Astrophysical Journal Supplement Series*, 240, 34
- MacQueen J., et al., 1967, in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. pp 281–297
- Madau P., Dickinson M., 2014, *Annual Review of Astronomy and Astrophysics*, 52, 415
- Magorrian J., et al., 1998, *AJ*, 115, 2285
- Malmquist K. G., 1922, *Meddelanden fran Lunds Astronomiska Observatorium Serie I*, 100, 1
- Man A. W. S., Toft S., Zirm A. W., Wuyts S., van der Wel A., 2012, *ApJ*, 744, 85
- Maoz D., Mannucci F., Brandt T. D., 2012, *MNRAS*, 426, 3282
- Marconi A., Hunt L. K., 2003, *ApJ*, 589, L21

- Marleau F. R., Clancy D., Bianconi M., 2013, *MNRAS*, 435, 3085
- Martig M., Bournaud F., Croton D. J., Dekel A., Teyssier R., 2012, *ApJ*, 756, 26
- Martin D. C., Matuszewski M., Morrissey P., Neill J. D., Moore A., Steidel C. C., Trainor R., 2016, *ApJ*, 824, L5
- Martin G., Kaviraj S., Devriendt J. E. G., Dubois Y., Laigle C., Pichon C., 2017, *MNRAS*, 472, L50
- Martin G., Kaviraj S., Devriendt J. E. G., Dubois Y., Pichon C., Laigle C., 2018a, *MNRAS*, 474, 3140
- Martin G., et al., 2018b, *MNRAS*, 476, 2801
- Martin G., Kaviraj S., Devriendt J. E. G., Dubois Y., Pichon C., 2018c, *MNRAS*, 480, 2266
- Martin G., et al., 2019, *MNRAS*, 485, 796
- Martínez-Delgado D., et al., 2016, *AJ*, 151, 96
- Mathur S., Grupe D., 2004, in Storchi-Bergmann T., Ho L. C., Schmitt H. R., eds, *IAU Symposium Vol. 222, The Interplay Among Black Holes, Stars and ISM in Galactic Nuclei*. pp 501–504 (arXiv:astro-ph/0405095), doi:10.1017/S1743921304003035
- Mathur S., Grupe D., 2005a, *A&A*, 432, 463
- Mathur S., Grupe D., 2005b, *ApJ*, 633, 688
- Mathur S., Fields D., Peterson B. M., Grupe D., 2012, *ApJ*, 754, 146
- Matteucci F., Greggio L., 1986, *A&A*, 154, 279
- Matteucci F., Recchi S., 2001, *ApJ*, 558, 351
- McAlpine S., Bower R. G., Harrison C. M., Crain R. A., Schaller M., Schaye J., Theuns T., 2017, *MNRAS*, 468, 3395
- McAlpine S., Bower R. G., Rosario D. J., Crain R. A., Schaye J., Theuns T., 2018, *MNRAS*, 481, 3118
- McDermid R. M., et al., 2015, *MNRAS*, 448, 3484
- McIntosh D. H., Guo Y., Hertzberg J., Katz N., Mo H. J., van den Bosch F. C., Yang X., 2008, *MNRAS*, 388, 1537
- McNamara B. R., Nulsen P. E. J., 2007, *Annual Review of Astronomy and Astrophysics*, 45, 117
- Menanteau F., Ford H. C., Motta V., Benítez N., Martel A. R., Blakeslee J. P., Infante L., 2006, *AJ*, 131, 208
- Menci N., Cavaliere A., Fontana A., Giallongo E., Poli F., 2002, *ApJ*, 575, 18

- Menou K., 2018, arXiv e-prints, p. arXiv:1811.06374
- Merloni A., et al., 2012, arXiv e-prints, p. arXiv:1209.3114
- Merritt A., van Dokkum P., Danieli S., Abraham R., Zhang J., Karachentsev I. D., Makarova L. N., 2016, *ApJ*, 833, 168
- Mihos J. C., Hernquist L., 1996, *ApJ*, 464, 641
- Mihos C., Murdin P., 2000, *Encyclopedia of Astronomy and Astrophysics*
- Mihos J. C., et al., 2015, *ApJ*, 809, L21
- Minchin R. F., et al., 2004, *MNRAS*, 355, 1303
- Mistani P. A., et al., 2016, *MNRAS*, 455, 2323
- Miyazaki S., et al., 2002, *Publications of the Astronomical Society of Japan*, 54, 833
- Miyazaki S., et al., 2012, in *Ground-based and Airborne Instrumentation for Astronomy IV. Proceedings of the SPIE*, Volume 8446, article id. 84460Z, 9 pp. (2012).. , doi:10.1117/12.926844
- Moore B., Katz N., Lake G., Dressler A., Oemler A., 1996, *Nature*, 379, 613
- Moore B., Governato F., Quinn T., Stadel J., Lake G., 1998, *ApJ*, 499, L5
- Moore B., Lake G., Quinn T., Stadel J., 1999, *MNRAS*, 304, 465
- Moscardini L., Dolag K., 2011, in , *Dark Matter and Dark Energy*. Springer, pp 217–237
- Mowla L., van Dokkum P., Merritt A., Abraham R., Yagi M., Koda J., 2017, *ApJ*, 851, 27
- Muñoz R. P., et al., 2015, *ApJ*, 813
- Muldrew S. I., Hatch N. A., Cooke E. A., 2015, *MNRAS*, 452, 2528
- Muldrew S. I., Hatch N. A., Cooke E. A., 2018, *MNRAS*, 473, 2335
- Murali C., Katz N., Hernquist L., Weinberg D. H., Davé R., 2002, *ApJ*, 571, 1
- Musso M., Cadiou C., Pichon C., Codis S., Kraljic K., Dubois Y., 2018, *MNRAS*, 476, 4877
- Naab T., 2013, in Thomas D., Pasquali A., Ferreras I., eds, *IAU Symposium Vol. 295, The Intriguing Life of Massive Galaxies*. pp 340–349 (arXiv:1211.6892), doi:10.1017/S1743921313005334
- Naab T., Johansson P. H., Ostriker J. P., 2009, *ApJ*, 699, L178
- Naab T., et al., 2014, *MNRAS*, 444, 3357
- Navarro J. F., Eke V. R., Frenk C. S., 1996, *MNRAS*, 283, L72

- Negroponte J., White S. D. M., 1983, MNRAS, 205, 1009
- Nelson D., et al., 2019, arXiv e-prints,
- Noguchi M., 2018, ApJ, 853, 67
- Nomoto K., Saio H., Kato M., Hachisu I., 2007, ApJ, 663, 1269
- Oñorbe J., Boylan-Kolchin M., Bullock J. S., Hopkins P. F., Kereš D., Faucher-Giguère C.-A., Quataert E., Murray N., 2015, MNRAS, 454, 2092
- O’Neil K., Bothun G., 2000, ApJ, 529, 811
- Odewahn S. C., Cohen S. H., Windhorst R. A., Philip N. S., 2002, ApJ, 568, 539
- Ogiya G., 2018, preprint, p. arXiv:1804.06421 (arXiv:1804.06421)
- Okabe A., Boots B., Sugihara K., Chiu S. N., 2009, Spatial tessellations: concepts and applications of Voronoi diagrams. Vol. 501, John Wiley & Sons
- Okazaki T., Taniguchi Y., 2000, ApJ, 543, 149
- Omma H., Binney J., Bryan G., Slyz A., 2004, MNRAS, 348, 1105
- Oosterloo T. A., Morganti R., Sadler E. M., van der Hulst T., Serra P., 2007, A&A, 465, 787
- Ostrovski F., et al., 2017, MNRAS, 465, 4325
- Ownsworth J. R., Conselice C. J., Mortlock A., Hartley W. G., Almaini O., Duncan K., Mundy C. J., 2014, MNRAS, 445, 2198
- Padoan P., Nordlund Å., 2011, ApJ, 730, 40
- Pahwa I., Saha K., 2018, MNRAS, 478, 4657
- Papastergis E., Adams E. A. K., Romanowsky A. J., 2017, A&A, 601, L10
- Papovich C., et al., 2012, ApJ, 750, 93
- Patel S. G., Kelson D. D., Diao N., Tonnesen S., Abramson L. E., 2018, preprint, p. arXiv:1807.02118 (arXiv:1807.02118)
- Paz D. J., Sgró M. A., Merchán M., Padilla N., 2011, MNRAS, 414, 2029
- Peirani S., Crockett R. M., Geen S., Khochfar S., Kaviraj S., Silk J., 2010, MNRAS, 405, 2327
- Peirani S., et al., 2016, preprint, (arXiv:1611.09922)
- Peirani S., et al., 2017, MNRAS, 472, 2153
- Peirani S., et al., 2018, preprint, (arXiv:1801.09754)

- Peng C. Y., 2007, *ApJ*, 671, 1098
- Peng E. W., Lim S., 2016, *ApJ*, 822
- Perret V., et al., 2012, preprint, (arXiv:1212.1356)
- Peth M. A., et al., 2016, *MNRAS*, 458, 963
- Petty S. M., et al., 2014, *AJ*, 148, 111
- Pichon C., Pogosyan D., Kimm T., Slyz A., Devriendt J., Dubois Y., 2011, *MNRAS*, 418, 2493
- Pillepich A., et al., 2018, *MNRAS*, 473, 4077
- Pillepich A., et al., 2019, arXiv e-prints,
- Pipino A., Kaviraj S., Bildfell C., Babul A., Hoekstra H., Silk J., 2009, *MNRAS*, 395, 462
- Planck Collaboration et al., 2014, *A&A*, 571, A16
- Pogge R. W., Eskridge P. B., 1993, *AJ*, 106, 1405
- Pontzen A., Governato F., 2012, *MNRAS*, 421, 3464
- Pontzen A., Governato F., 2014, *Nature*, 506, 171
- Pontzen A., Tremmel M., Roth N., Peiris H. V., Saintonge A., Volonteri M., Quinn T., Governato F., 2017, *MNRAS*, 465, 547
- Postman M., et al., 2005, *ApJ*, 623, 721
- Poudel A., Heinämäki P., Tempel E., Einasto M., Lietzen H., Nurmi P., 2017, *A&A*, 597, A86
- Pracy M. B., Kuntschner H., Couch W. J., Blake C., Bekki K., Briggs F., 2009, *MNRAS*, 396, 1349
- Prescott M. K. M., Martin C. L., Dey A., 2015, *ApJ*, 799, 62
- Prole D. J., Davies J. I., Keenan O. C., Davies L. J. M., 2018, *MNRAS*, p. 970
- Qu Y., Di Matteo P., Lehnert M. D., van Driel W., 2011, *A&A*, 530, A10
- Rahmani H., et al., 2018, *MNRAS*, 474, 254
- Rajaraman A., Ullman J. D., 2011, Cited on, p. 139
- Rasera Y., Teyssier R., 2006, *A&A*, 445, 1
- Rees M. J., 1984, *Annual Review of Astronomy and Astrophysics*, 22, 471
- Rees M. J., Ostriker J. P., 1977, *MNRAS*, 179, 541

- Regan M. W., Teuben P. J., 2004, *ApJ*, 600, 595
- Reines A. E., Volonteri M., 2015, *ApJ*, 813, 82
- Ricarte A., Tremmel M., Natarajan P., Quinn T., 2019, arXiv e-prints, p. arXiv:1904.10116
- Robaina A. R., et al., 2009, *ApJ*, 704, 324
- Robertson B., Bullock J. S., Cox T. J., Di Matteo T., Hernquist L., Springel V., Yoshida N., 2006, *ApJ*, 645, 986
- Robertson B. E., et al., 2017, preprint, (arXiv:1708.01617)
- Rodighiero G., et al., 2011, *ApJ*, 739, L40
- Rodrigues M., Hammer F., Flores H., Puech M., Athanassoula E., 2017, *MNRAS*, 465, 1157
- Rodriguez-Gomez V., et al., 2015, *MNRAS*, 449, 49
- Rodriguez-Gomez V., et al., 2017, *MNRAS*, 467, 3083
- Román J., Trujillo I., 2017a, *MNRAS*, 468, 703
- Román J., Trujillo I., 2017b, *MNRAS*, 468, 4039
- Rong Y., Liu Y., Zhang S.-N., 2016, *MNRAS*, 455, 2267
- Rong Y., Guo Q., Gao L., Liao S., Xie L., Puzia T. H., Sun S., Pan J., 2017, *MNRAS*, 470, 4231
- Rousseuw P. J., 1987, *Journal of computational and applied mathematics*, 20, 53
- Roychowdhury S., Chengalur J. N., Shi Y., 2017, *A&A*, 608, A24
- Ruiz-Lara T., et al., 2018, *MNRAS*, 478, 2034
- Ryden B., 2003, *Introduction to cosmology*. Addison Wesley
- Sabatini S., Davies J., van Driel W., Baes M., Roberts S., Smith R., Linder S., O'Neil K., 2005, *MNRAS*, 357, 819
- Sakelliou I., 2000, *MNRAS*, 318, 1164
- Sales L. V., Navarro J. F., Theuns T., Schaye J., White S. D. M., Frenk C. S., Crain R. A., Dalla Vecchia C., 2012, *MNRAS*, 423, 1544
- Sandage A., Binggeli B., 1984, *AJ*, 89, 919
- Sanders D. B., Soifer B. T., Elias J. H., Madore B. F., Matthews K., Neugebauer G., Scoville N. Z., 1988, *ApJ*, 325, 74
- Sanders J. L., Evans N. W., Dehnen W., 2018, *MNRAS*, 478, 3879

- Sani E., Marconi A., Hunt L. K., Risaliti G., 2011, MNRAS, 413, 1479
- Sarzi M., et al., 2002, ApJ, 567, 237
- Saslaw W., 2000, Encyclopedia of Astronomy and Astrophysics, 1, 2085
- Satyapal S., Böker T., McAlpine W., Gliozzi M., Abel N. P., Heckman T., 2009, ApJ, 704, 439
- Satyapal S., Secrest N. J., McAlpine W., Ellison S. L., Fischer J., Rosenberg J. L., 2014, ApJ, 784, 113
- Sawala T., et al., 2016, MNRAS, 456, 85
- Scannapieco C., Tissera P. B., White S. D. M., Springel V., 2008, MNRAS, 389, 1137
- Scarlata C., et al., 2007, The Astrophysical Journal Supplement Series, 172, 406
- Schaeberle J. M., 1904, Nature, 69, 248
- Schawinski K., Treister E., Urry C. M., Cardamone C. N., Simmons B., Yi S. K., 2011, ApJ, 727, L31+
- Schawinski K., Simmons B. D., Urry C. M., Treister E., Glikman E., 2012, MNRAS, 425, L61
- Schawinski K., et al., 2014, MNRAS, 440, 889
- Schawinski K., Zhang C., Zhang H., Fowler L., Santhanam G. K., 2017, MNRAS, 467, L110
- Schaye J., et al., 2015, MNRAS, 446, 521
- Schechter P., 1976, ApJ, 203, 297
- Schmidt M., 1968, ApJ, 151, 393
- Schweizer F., 2000, in Astronomy, physics and chemistry of  $H^+$ . p. 2063  
(arXiv:astro-ph/0002263), doi:10.1098/rsta.2000.0630
- Secrest N. J., Satyapal S., Gliozzi M., Cheung C. C., Seth A. C., Böker T., 2012, ApJ, 753, 38
- Sérsic J. L., 1963, Boletín de la Asociación Argentina de Astronomía La Plata Argentina, 6, 41
- Shabala S. S., et al., 2012, MNRAS, 423, 59
- Shakura N. I., Sunyaev R. A., 1973, A&A, 24, 337
- Shankar F., et al., 2015, ApJ, 802, 73
- Shao S., Cautun M., Frenk C. S., Grand R. J. J., Gómez F. A., Marinacci F., Simpson C. M., 2018, MNRAS, 476, 1796
- Shattow G. M., Croton D. J., Skibba R. A., Muldrew S. I., Pearce F. R., Abbas U., 2013, MNRAS, 433, 3314

- Sheth R. K., Tormen G., 2004, *MNRAS*, 350, 1385
- Shibuya T., Ouchi M., Harikane Y., 2015, *ApJ*, 219, 15
- Shlosman I., 2012, arXiv preprint arXiv:1212.1463
- Sifón C., van der Burg R. F. J., Hoekstra H., Muzzin A., Herbonnet R., 2018, *MNRAS*, 473, 3747
- Sijacki D., Vogelsberger M., Genel S., Springel V., Torrey P., Snyder G. F., Nelson D., Hernquist L., 2015, *MNRAS*, 452, 575
- Silk J., 2011, in Carignan C., Combes F., Freeman K. C., eds, *IAU Symposium Vol. 277, Tracing the Ancestry of Galaxies*. pp 273–281 (arXiv:1102.0283), doi:10.1017/S1743921311022939
- Silk J., Norman C., 2009, *ApJ*, 700, 262
- Silk J., Rees M. J., 1998, *A&A*, 331, L1
- Simard L., et al., 2002, *The Astrophysical Journal Supplement Series*, 142, 1
- Simard L., Mendel J. T., Patton D. R., Ellison S. L., McConnachie A. W., 2011, *ApJ*, 196, 11
- Simmons B. D., Van Duyne J., Urry C. M., Treister E., Koekemoer A. M., Grogin N. A., GOODS Team 2011, *ApJ*, 734, 121
- Simmons B. D., Urry C. M., Schawinski K., Cardamone C., Glikman E., 2012, *ApJ*, 761, 75
- Simmons B. D., et al., 2013, *MNRAS*, 429, 2199
- Simmons B. D., Smethurst R. J., Lintott C., 2017, *MNRAS*, 470, 1559
- Simpson C. M., Grand R. J. J., Gómez F. A., Marinacci F., Pakmor R., Springel V., Campbell D. J. R., Frenk C. S., 2018, *MNRAS*, 478, 548
- Siudek M., et al., 2018, *A&A*, 617, A70
- Skibba R. A., et al., 2009, *MNRAS*, 399, 966
- Smethurst R. J., et al., 2015, *MNRAS*, 450, 435
- Smirnov N. V., 1939, *Bull. Math. Univ. Moscou*, 2, 3
- Smith Castelli A. V., Faifer F. R., Escudero C. G., 2016, *A&A*, 596, A23
- Smith G. P., Treu T., Ellis R. S., Moran S. M., Dressler A., 2005, *ApJ*, 620, 78
- Somerville R. S., Davé R., 2015, *Annual Review of Astronomy and Astrophysics*, 53, 51
- Somerville R. S., Primack J. R., 1999, *MNRAS*, 310, 1087
- Somerville R. S., Primack J. R., Faber S. M., 2001, *MNRAS*, 320, 504

- Soo J. Y. H., et al., 2018, MNRAS, 475, 3613
- Soumagnac M. T., et al., 2015, MNRAS, 450, 666
- Sousbie T., 2011, MNRAS, 414, 350
- Sousbie T., Pichon C., Kawahara H., 2011, MNRAS, 414, 384
- Sparre M., Springel V., 2016, MNRAS, 462, 2418
- Sparre M., Springel V., 2017, MNRAS, 470, 3946
- Springel V., Hernquist L., 2005, ApJ, 622, L9
- Stanford S. A., Eisenhardt P. R., Dickinson M., 1998, ApJ, 492, 461
- Starobinsky A. A., 1982, Physics Letters B, 117, 175
- Steinborn L. K., Dolag K., Comerford J. M., Hirschmann M., Remus R.-S., Teklu A. F., 2016, MNRAS, 458, 1013
- Steinborn L. K., Hirschmann M., Dolag K., Shankar F., Juneau S., Krumpel M., Remus R.-S., Teklu A. F., 2018, MNRAS, 481, 341
- Stewart K. R., Bullock J. S., Wechsler R. H., Maller A. H., Zentner A. R., 2008, ApJ, 683, 597
- Stewart K. R., Brooks A. M., Bullock J. S., Maller A. H., Diemand J., Wadsley J., Moustakas L. A., 2013, ApJ, 769, 74
- Stott J. P., Sobral D., Smail I., Bower R., Best P. N., Geach J. E., 2013, MNRAS, 430, 1158
- Stoughton C., et al., 2002, AJ, 123, 485
- Strateva I., et al., 2001, AJ, 122, 1861
- Struve C., Oosterloo T., Sancisi R., Morganti R., Emonts B. H. C., 2010, A&A, 523, A75
- Sutherland R. S., Dopita M. A., 1993, ApJ, 88, 253
- Tacconi L. J., et al., 2008, ApJ, 680, 246
- Tacconi L. J., et al., 2010, Nature, 463, 781
- Tacconi L. J., et al., 2013, ApJ, 768, 74
- Takada M., et al., 2014, Publications of the ASJ, 66, R1
- Tanaka M., 2015, ApJ, 801, 20
- Tanaka M., et al., 2018, Publications of the Astronomical Society of Japan, 70, S9

- Taranu D. S., Dubinski J., Yee H. K. C., 2013, *ApJ*, 778, 61
- Taylor P., Kobayashi C., 2014, *MNRAS*, 442, 2751
- Taylor P., Kobayashi C., 2016, *MNRAS*, 463, 2465
- Taylor P., Federrath C., Kobayashi C., 2018, *MNRAS*, 479, 141
- Tempel E., Tago E., Liivamägi L. J., 2012, *A&A*, 540, A106
- Tempel E., Stoica R. S., Saar E., 2013, *MNRAS*, 428, 1827
- Tempel E., Guo Q., Kipper R., Libeskind N. I., 2015, *MNRAS*, 450, 2727
- Teplitz H., 2018, Ultraviolet Imaging of the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey Fields (UVCANDELS), HST Proposal
- Teyssier R., 2002, *A&A*, 385, 337
- Teyssier R., Pontzen A., Dubois Y., Read J. I., 2013, *MNRAS*, 429, 3068
- Thomas D., Maraston C., Bender R., Mendes de Oliveira C., 2005, *ApJ*, 621, 673
- Thronson Jr. H. A., Bally J., Hacking P., 1989, *AJ*, 97, 363
- Toloba E., et al., 2018, *ApJ*, 856
- Tomczak A. R., et al., 2014a, *ApJ*, 783, 85
- Tomczak A. R., et al., 2014b, *ApJ*, 783, 85
- Tonini C., Mutch S. J., Croton D. J., Wyithe J. S. B., 2016, *MNRAS*, 459, 4109
- Toomre A., 1977, in Tinsley B. M., Larson D. Campbell R. B. G., eds, *Evolution of Galaxies and Stellar Populations*. p. 401
- Toomre A., Toomre J., 1972, *ApJ*, 178, 623
- Tormen G., 1998, *MNRAS*, 297, 648
- Torrealba G., et al., 2018, preprint, p. arXiv:1811.04082 (arXiv:1811.04082)
- Torrey P., et al., 2015, *MNRAS*, 454, 2770
- Torrey P., Wellons S., Ma C.-P., Hopkins P. F., Vogelsberger M., 2017, *MNRAS*, 467, 4872
- Trager S. C., Faber S. M., Worthey G., González J. J., 2000, *AJ*, 119, 1645
- Trakhtenbrot B., Lira P., Netzer H., Cicone C., Maiolino R., Shemmer O., 2017, *ApJ*, 836, 8
- Trebitsch M., Volonteri M., Dubois Y., Madau P., 2018, *MNRAS*, 478, 5607

- Tremblay G. R., et al., 2016, *Nature*, 534, 218
- Tremmel M., Karcher M., Governato F., Volonteri M., Quinn T. R., Pontzen A., Anderson L., Bellovary J., 2017, *MNRAS*, 470, 1121
- Trujillo I., et al., 2018, preprint, (arXiv:1806.10141)
- Trump J. R., et al., 2015, *ApJ*, 811, 26
- Turner J. A., Phillipps S., Davies J. I., Disney M. J., 1993, *MNRAS*, 261, 39
- Tweed D., Devriendt J., Blaizot J., Colombi S., Slyz A., 2009, *A&A*, 506, 647
- Tyson J. A., 2002, in Tyson J. A., Wolff S., eds, *Proceedings of the SPIE Vol. 4836, Survey and Other Telescope Technologies and Discoveries*. pp 10–20 (arXiv:astro-ph/0302102), doi:10.1117/12.456772
- Urrutia T., Lacy M., Becker R. H., 2008, *ApJ*, 674, 80
- Vassiliadis E., Wood P. R., 1993, *ApJ*, 413, 641
- Veilleux S., Bolatto A., Tombesi F., Meléndez M., Sturm E., González-Alfonso E., Fischer J., Rupke D. S. N., 2017, *ApJ*, 843, 18
- Velliscig M., et al., 2015, *MNRAS*, 454, 3328
- Venhola A., et al., 2017, *A&A*, 608
- Viaene S., Forbrich J., Fritz J., 2018, *MNRAS*, 475, 5550
- Vika M., Vulcani B., Bamford S. P., Häußler B., Rojas A. L., 2015, *A&A*, 577, A97
- Visvanathan N., 1981, *A&A*, 100, L20
- Vogelsberger M., et al., 2014, *MNRAS*, 444, 1518
- Vollmer B., Cayatte V., Balkowski C., Duschl W. J., 2001, *ApJ*, 561, 708
- Volonteri M., Capelo P. R., Netzer H., Bellovary J., Dotti M., Governato F., 2015, *MNRAS*, 449, 1470
- Volonteri M., Dubois Y., Pichon C., Devriendt J., 2016, *MNRAS*, 460, 2979
- Vulcani B., et al., 2011, *MNRAS*, 412, 246
- Walmsley M., Ferguson A. M. N., Mann R. G., Lintott C. J., 2019, *MNRAS*, 483, 2968
- Wang E. X., Taylor P., Federrath C., Kobayashi C., 2019, *MNRAS*, 483, 4640
- Weigel A. K., Schawinski K., Bruderer C., 2016, *MNRAS*, 459, 2150
- Weisz D. R., et al., 2011, *ApJ*, 739, 5

- Welker C., Devriendt J., Dubois Y., Pichon C., Peirani S., 2014, *MNRAS*, 445, L46
- Welker C., Dubois Y., Pichon C., Devriendt J., Chisari E. N., 2015, preprint, (arXiv:1512.00400)
- Welker C., Dubois Y., Devriendt J., Pichon C., Kaviraj S., Peirani S., 2017, *MNRAS*, 465, 1241
- Wellons S., Torrey P., 2017, *MNRAS*, 467, 3887
- Wellons S., et al., 2016, *MNRAS*, 456, 1030
- Whitaker K. E., van Dokkum P. G., Brammer G., Franx M., 2012, *ApJ*, 754, L29
- Whitaker K. E., et al., 2015, *ApJ*, 811, L12
- White S. D. M., Rees M. J., 1978, *MNRAS*, 183, 341
- Wild V., Almaini O., Dunlop J., Simpson C., Rowlands K., Bowler R., Maltby D., McLure R., 2016, *MNRAS*, 463, 832
- Wilkinson M., Meijer B., 1995, *Computer methods and programs in biomedicine*, 47, 35
- Willett K. W., et al., 2015, *MNRAS*, 449, 820
- Willett K. W., et al., 2017, *MNRAS*, 464, 4176
- Williams R. P., et al., 2016, *MNRAS*, 463, 2746
- Wilman D. J., Erwin P., 2012, *ApJ*, 746, 160
- Wilson T. J., et al., 2019, *ApJ*, 874, 18
- Wolf M., 1908, *Publikationen des Astrophysikalischen Instituts Koenigstuhl-Heidelberg*, 3, 109
- Wollaeger R. T., et al., 2018, *MNRAS*, 478, 3298
- Worthey G., 1994, *ApJ*, 95, 107
- Wuyts E., Rigby J. R., Gladders M. D., Sharon K., 2014, *ApJ*, 781, 61
- Yagi M., Koda J., Komiyama Y., Yamanoi H., 2016, *The Astrophysical Journal Supplement Series*, 225, 11
- Yang X., van den Bosch F. C., Mo H. J., Mao S., Kang X., Weinmann S. M., Guo Y., Jing Y. P., 2006, *MNRAS*, 369, 1293
- Yi S., Lee Y.-W., Woo J.-H., Park J.-H., Demarque P., Oemler Augustus J., 1999, *ApJ*, 513, 128
- Yozin C., Bekki K., 2015, *MNRAS*, 452, 937
- Zaritsky D., et al., 2019, *The Astrophysical Journal Supplement Series*, 240, 1

- Zhong G. H., Liang Y. C., Liu F. S., Hammer F., Hu J. Y., Chen X. Y., Deng L. C., Zhang B., 2008, MNRAS, 391, 986
- Zolotov A., et al., 2015, MNRAS, 450, 2327
- de Jong R. S., et al., 2012, in Ground-based and Airborne Instrumentation for Astronomy IV. p. 84460T (arXiv:1206.6885), doi:10.1117/12.926239
- de Vaucouleurs G., 1959, Handbuch der Physik, 53, 275
- van Dokkum P. G., Franx M., 2001, ApJ, 553, 90
- van Dokkum P. G., et al., 2010, ApJ, 709, 1018
- van Dokkum P. G., Abraham R., Merritt A., Zhang J., Geha M., Conroy C., 2015, ApJ, 798, L45
- van Dokkum P., et al., 2016, ApJ, 828
- van Dokkum P., et al., 2018, Nature, 555, 629
- van den Bosch F. C., Abel T., Croft R. A. C., Hernquist L., White S. D. M., 2002, ApJ, 576, 21
- van der Burg R. F. J., Muzzin A., Hoekstra H., 2016, A&A, 590