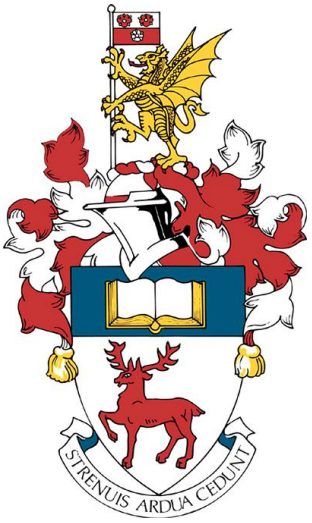
UNIVERSITY OF SOUTHAMPTON




# Probing the evolution of the most massive galaxies.

by

Stewart William Buchan

A thesis submitted in partial fulfillment for the
degree of Doctor of Philosophy

in the
Physical sciences and engineering
School of Physics and Astronomy

February 2018

UNIVERSITY OF SOUTHAMPTON

# *Abstract*

Physical sciences and engineering

School of Physics and Astronomy

Doctor of Philosophy

**Probing the evolution of the most massive galaxies**

by Stewart William Buchan

This thesis sets more stringent constraints on how the most massive ($\log M_{\rm star} \geq 11.5[M_{\odot}]$) local, central galaxies have assembled their stellar mass, especially the relative roles of in-situ growth versus later accretion via mergers. To achieve this, I developed a series of semi-empirical/phenomenological models which, by construction, have few theoretical assumptions or free parameters.

I firstly show that by comparing basic abundance matching predictions to observations of massive galaxies, the high-mass slope of the $M_{\rm star}$-$M_{\rm halo}$ relation appears to be substantially steeper than the commonly used relations in the literature, and the scatter is small ($\sim 0.15$ *dex*). With the aid of the semi-empirical model I developed, which is based around this steeper abundance matching relation, I find that massive galaxies can grow by a factor two in stellar mass between $z = 1$ and $z = 0$ and a factor of four in size over the same redshift range, providing that mergers are efficient.

Next, I populate a catalogue of massive dark matter haloes with galaxies through the above abundance matching relations, and compare the average stellar mass to total baryonic mass of their progenitors at their putative formation epoch, $z_{\rm form}= 2 - 4$. I find that the former is in fact equal to, if not greater than the latter. This would imply that if galaxies form in a strict monolithic collapse, the efficiency of converting baryons into stars needs to be extremely high, if not 100%. I also argue that this means they would need to be born extended, which may provide an observational test to discern between early versus late formation scenarios.

I then outlined a promising framework to set constraints on the evolution of the slope and scatter of the high-mass end of the $M_{\rm star}$-$M_{\rm halo}$ relation in a way designed to bypass observational systematics. I utilised the halo mass distribution of massive central and satellite galaxies at $z = 0.5$. The former is used to constrain the relation at $z = 0.5$ and the latter at their redshift of infall: $0.5 < z \lesssim 1.2$. By comparing the two relations, I find signs that individual massive galaxies have some stellar mass growth over this epoch.

Finally, I show the results of two complementary projects involving numerical simulations that are designed to enhance and check the results presented in the above works.

Using novel semi-empirical/phenomenological models, this thesis indicates that massive, central galaxies assemble a substantial amount of their stellar mass at later epochs rather than forming through extremely efficient star-bursts at high redshift.

# Contents

**Abstract** iii

**List of Figures** ix

**List of Tables** xvii

**Declaration of Authorship** xix

**Acknowledgements** xx

**1 Introduction** 1
- 1.1 Overview . . . 1
- 1.2 Modern cosmology . . . 2
  - 1.2.1 Cosmological distances, redshift and look back time . . . 2
  - 1.2.2 What is the universe made of? . . . 3
  - 1.2.3 The time line of the Universe . . . 4
- 1.3 Structure formation . . . 7
- 1.4 Observations of local galaxies . . . 11
  - 1.4.1 The Hubble galaxy classification . . . 11
  - 1.4.2 Measuring physical galaxy properties . . . 12
  - 1.4.3 Statistical properties of galaxies . . . 13
    - 1.4.3.1 Luminosity and stellar mass functions . . . 13
    - 1.4.3.2 The $M_{\rm star}$-$M_{\rm halo}$ relation . . . 14
    - 1.4.3.3 The Colour-mass distribution . . . 14
    - 1.4.3.4 The Size-mass relation . . . 17
- 1.5 Observations of high redshift galaxies . . . 18
  - 1.5.1 Detecting $z > 1$ galaxies . . . 19
  - 1.5.2 Redshift evolution . . . 20
    - 1.5.2.1 Morphology . . . 20
    - 1.5.2.2 Stellar mass . . . 20
    - 1.5.2.3 Effective radius . . . 22
- 1.6 Galaxy formation scenarios . . . 23
  - 1.6.1 Formation of spirals . . . 23
  - 1.6.2 Formation of giant ellipticals . . . 24

1.6.2.1 Formation of giant ellipticals: monolithic collapse . . . . 24
1.6.2.2 Formation of giant ellipticals: Mergers . . . . . 25
1.6.3 Formation of giant ellipticals: Observational challenges . . . . . 25
1.6.4 Formation of giant ellipticals: Modelling techniques . . . . . 26
1.6.4.1 Semi-analytic models . . . . . 26
1.6.4.2 Cosmological hydrodynamical simulations . . . . . 27
1.6.4.3 Semi-empirical models . . . . . 28
1.7 Thesis outline . . . . . 29

**2 Probing the evolution of massive galaxies using their abundances and validation through a semi-empirical model. 31**
2.1 Introduction . . . . . 31
2.2 Initial constraints from abundance matching. . . . . 32
2.2.1 Observations . . . . . 34
2.2.2 Constraints from abundance matching . . . . . 35
2.2.3 Constraints from galaxy clustering . . . . . 41
2.3 Semi-empirical model . . . . . 46
2.3.1 Merger trees . . . . . 46
2.3.2 Generating halo catalogues . . . . . 48
2.3.3 Adding galaxies . . . . . 48
2.3.4 Results . . . . . 53
2.3.4.1 Individual brightest cluster galaxy . . . . . 55
2.3.4.2 Stellar mass evolution . . . . . 55
2.3.4.3 Size evolution . . . . . 55
2.3.4.4 Number of mergers . . . . . 59
2.4 Summary . . . . . 60
2.5 Future work . . . . . 61

**3 Constraining the evolution of the most massive galaxies through their abundances and ages. 63**
3.1 Introduction . . . . . 63
3.2 Overview and Methodology . . . . . 64
3.2.1 Selecting descendant galaxies at z=0 . . . . . 65
3.2.2 Selecting their progenitors . . . . . 67
3.2.3 Ages of massive, early type galaxies: selecting a formation epoch. . 69
3.3 Results . . . . . 70
3.3.1 Constraints on the assembly scenario . . . . . 70
3.3.2 Are today's central, massive galaxies just outliers at the epoch of formation? . . . . . 72
3.4 Discussion . . . . . 73
3.4.1 Can massive galaxies grow solely through in-situ star formation? . 75
3.4.2 Can mergers drive mass evolution of massive galaxies? . . . . . 77
3.4.3 Can mergers drive size evolution of massive galaxies? . . . . . 79
3.4.4 Size growth from quasar mode feedback . . . . . 80
3.4.5 No size growth for very massive ETGs? . . . . . 81
3.5 Conclusions . . . . . 81

**4 Probing the stellar mass to halo mass relation at $0.5 \leq z \lesssim 1.2$ using the clustering of satellite galaxies. 83**
4.1 Introduction . . . 83
4.2 Overview and Methodology . . . 86
4.2.1 Dark matter halo catalogues . . . 87
4.2.2 Mock galaxy catalogues . . . 89
4.3 Probing the high-mass end slope $\delta$ and intrinsic scatter $\sigma$ with BOSS massive galaxies. . . . 92
4.3.1 Constraints at $z = 0.5$ . . . 94
4.3.2 The "frozen" model . . . 96
4.3.3 Models with tidal stellar stripping . . . 99
4.3.4 Stellar mass loss and star formation . . . 103
4.4 Discussion . . . 103
4.4.1 Implications for galaxy evolution since $z \sim 1$ . . . 103
4.4.2 Comparison with previous works . . . 107
4.5 Summary . . . 108

**5 Testing the mass and size growth of massive galaxies using numerical simulations. 111**
5.1 Introduction . . . 111
5.2 Quasi-adiabatic expansion of an isolated galaxy . . . 112
5.2.1 Numerical set-up . . . 113
5.2.2 Results . . . 115
5.2.2.1 Expansion due to mass-loss . . . 115
5.2.2.2 Testing the significance of this expansion . . . 116
5.2.3 Conclusions . . . 118
5.3 A cosmological zoom-in with RAMSES . . . 118
5.3.1 Numerical set-up . . . 118
5.3.2 Analysis and results . . . 119
5.3.2.1 Global properties . . . 120
5.3.2.2 Estimating stellar mass . . . 121
5.3.2.3 Estimating the half-mass radius . . . 126
5.3.3 Conclusions . . . 126

**6 Conclusion. 129**
6.1 Main conclusions . . . 129
6.1.1 Chapter 2 . . . 129
6.1.2 Chapter 3 . . . 130
6.1.3 Chapter 4 . . . 131
6.1.4 Chapter 5 . . . 132
6.1.5 Answering the key questions . . . 133
6.2 Future work . . . 133

**A Generating merger trees. 135**
A.1 Analytic trees . . . 135
A.2 Numerical trees . . . 137

A.3 Comparison . . . . . 138

**B Modelling the $z_{\rm inf}$ graph structure with Neo4j** **141**

**C Test the Guo et al. (2014) and Tinker et al. (2016) $dP/d\log M_{\rm halo}$ distributions** **145**

**Bibliography** **149**

# List of Figures


1.1 Velocity-Distance Relation among Extra-Galactic galaxies extracted from Hubble (1929) . . . . . 3
1.2 The relation between redshift and look back time for WMAP 9 cosmology (Bennett et al., 2013). . . . . 4
1.3 Contours at 68.3%, 95.4%, and 99.7% confidence level on $\Omega_\Lambda$ and $\Omega_M$ obtained from CMB, BAO, and the Union SN set (Kowalski et al., 2008), as well as their combination. Extracted from Kowalski et al. (2008). . . . 5
1.4 The constraints on the the baryon fraction $f_b = (\Omega_b/\Omega_M)$ and $\Omega_M h$ from VIPERS (Rota et al., 2017), WiggleZ (Parkinson et al., 2012) and Planck (Planck Collaboration et al., 2014). Extracted from Rota et al. (2017). . . 6
1.5 Schematic diagram of the history of the Universe, Credit NAOJ. . . . . 7
1.6 The formation of a dark matter halo in a high resolution N-body simulation of a $3 \times 10^{11}[h^{-1}M_\odot]$ halo (extracted from Baugh, 2006). The circle marks the present day virial radius. The colours reflect the density of dark matter, with redder colours indicating spatial higher density. . . . . 10
1.7 A schematic representation of the classifications of 'nebulae' (now called galaxies). Extracted from Hubble (1936). . . . . 11
1.8 The observed $z = 0$ luminosity (blue, top) and corresponding observed $z = 0$ stellar mass (blue, bottom) functions, separated into their morphological calcification (as labelled). Adapted from Bernardi et al. (2013). . . 15
1.9 The $u - r$ colourmass diagram of local galaxies. The left panel shows all galaxies. The right panels show galaxies separated into early-type (top) and late-type (bottom). Extracted from Schawinski et al. (2014). . . . . 16
1.10 The median and dispersion of the distribution of Seŕsic half-light radius, $R$, in the z band as a function of stellar mass extracted from (Shen et al., 2003). Triangles represent results for late-type galaxies, while the squares are for early-type galaxies. The error bars represent the scatter among 20 bootstrap samples. . . . . 18
1.11 Evolution of the star formation rate density based on rest-frame UV luminosity functions (extracted from Bouwens et al., 2010). The lower, blue shaded region shows the star formation rate density with no dust correction. Instead, the top, red shaded region shows the dust corrected rate. . . . . 19
1.12 Examples of CANDELS $H_{160}$-band images of galaxies with different morphologies at $1 < z < 3$. Extracted from Mortlock et al. (2013) . . . . . 20
1.13 The evolution of the relative abundances of different galaxy morphologies between $z \sim 0$ and $z \sim 3$; extracted from Huertas-Company et al. (2015). 21

1.14 The redshift evolution of the stellar mass density. This is found by integrating the stellar mass function from $10^8$ to $10^{13}$ $[M_\odot]$ at each redshift. Extracted from Marchesini et al. (2009) . . . . . 22
1.15 The sizestellar mass distribution of late-type (blue) and early-type (red) galaxies for six redshift bins, extracted from van der Wel et al. (2014). The solid lines in each panel are the best-fit $M_{\rm star}$-$M_{\rm halo}$ relation at that redshift. Instead the dashed line is the best-fit at $0 < z < 0.5$. . . . . 23
1.16 Nearby galaxies originally observed at $z \sim 0$ in B-band, simulated to how they would appear at $z = 2.5$ in rest-frame B-band. Extracted from Conselice (2014) . . . . . 26

2.1 A cartoon sketching how two contrasting redshift evolutions in the $M_{\rm star}$-$M_{\rm halo}$ relation predicts predict different evolutions in $M_{\rm star}$. Panel A shows the mean evolution of a local log $M_{\rm halo}$= $11[M_\odot]$ halo. Panels B and C show a gradually decreasing $M_{\rm star}$-$M_{\rm halo}$ relation and a static one respectively. The blue arrows on the relations show the evolution in $M_{\rm halo}$from panel A along with the corresponding evolutions in $M_{\rm star}$. Panels D and E explicitly show the corresponding evolutions in $M_{\rm star}$. . . 33
2.2 A cartoon of abundance matching. Panel 'A' shows an image from SDSS centred on a massive galaxy. All the local galaxies within this picture have their log $M_{\rm star}$ labelled. Instead, panel 'B' is a picture of a slice through a zoom-in cosmological simulation. A random sample of haloes have their log $M_{\rm star}$ labelled. Panel 'C' then shows a schematic of abundance matching, where samples of galaxies and haloes are selected from equal volumes, ranked by mass and mapped. . . . . 37
2.3 A more realistic schematic of abundance matching. The left panel show the Tinker et al. (2008) integrated halo mass function. The right panel shows the Bernardi et al. (2013) integrated stellar mass function. The y-axis is the same between the two panels. The horizontal lines show $10^{-2}$, $10^{-4}$, $10^{-6}$ and $10^{-8}$ galaxies/haloes $/Mpc^3h^{-3}$. The vertical lines then show the halo/stellar mass these number densities correspond to. . . . . 39
2.4 A cartoon to show the affect of adding scatter to the $M_{\rm star}$-$M_{\rm halo}$ relation. The top left panel shows the $M_{\rm star}$-$M_{\rm halo}$ relation of fictitious galaxies that follow a strict relation with zero scatter. The blue histogram in the bottom panels shows this fictitious mass function. The top right panel shows the relation when the stellar mass each galaxy in the top left panel is scattered by a random amount following a log-normal distribution. The standard deviation has been chosen such that 10% of the objects are scattered into a higher mass bin and 10% being scattered into a lower one (as shown by the arrows in the bottom left panel). The orange line in the bottom right panel shows the resulting mass function after this scattering. 41
2.5 Top panel: A plot showing the effect that changing the intrinsic scatter in the mean $M_{\rm star}$-$M_{\rm halo}$ relation at fixed halo mass has on the best fit $M_{\rm star}$-$M_{\rm halo}$ relation. Bottom panel: The input observed stellar mass function by (Bernardi et al., 2013) as well as the resulting best fits for each scatter. 42

2.6 The median stellar mass to halo mass relation for central galaxies at $< z >= 0.1, 0.4, 1.1$ and 0.7 going clockwise from top left. Each panel shows data points from observed massive galaxies at that redshift range. Specifically, the data points in the top left panel are local galaxies from Kravtsov et al. (2014) (orange) and Gonzalez et al. (2013) (blue). The open circles in the other three panels show galaxies from Huertas-Company et al. (2013) and the filled circles in each panel show their median. The black squares in the bottom right panel show a selection of targeted observations (see Section 2.2.1) at $0.8 < z < 1.4$ and the green triangles are galaxies in the Cl1604 supercluster from Ascaso et al. (2014). The lines in each panel correspond to the results of abundance matching from stellar mass functions found in the literature. For reference, each panel also shows reference slopes where $\delta = 1 - \gamma$ from Equation 2.2. . . . . 43

2.7 The predicted probability that a BCG with log $M_{\rm star}>$ 11.5$[M_{\odot}]$ will reside in a halo of mass $M$. The grey band is the results from Guo et al. (2014) for observed, central galaxies in BOSS and the coloured lines are the results of Shankar et al. (2014b) analysis for central galaxies using three different intrinsic scatters, as labelled. The left panel shows results at $z = 0.5$ and the right panel shows them at $z = 0.7$. . . . . 46

2.8 An example merger tree taken from Heitmann et al. (2014). The coloured discs represent dark matter haloes where the mass increases from pink to blue. As highlighted, forks show when two dark matter haloes merge. A major merger is when the two haloes are of compatible mass. Haloes can also grow via steady smooth accretion of dark matter from the cosmic web. 47

2.9 A comparison between the halo mass functions produced by two dark matter-only simulations and the analytic relation of Tinker et al. (2008) at $z = 0$. . . . . 49

2.10 A demonstration of our sampling strategy. The blue curve is the integrated number density of haloes above a given mass. The alternating coloured regions represent single haloes (when N in an integer) for a $10^4\ Mpc^3h^{-3}$ box. . . . . 49

2.11 A cartoon showing a portion of a single merger tree which is being simulated by the semi-empirical model. The coloured circles represent the underlying dark matter. The presence of a galaxy shows that that halo is populated. The central column shows the main progenitor branch and the others show smaller galaxies which merge onto it. . . . . 50

2.12 The evolution of a single, massive galaxy in the semi-empirical model. The top panel shows the stellar mass contained within the disc (orange), the bulge (blue) and total (green). Additionally, the cold gas mass is shown (red). The bottom panel shows the effective radius of the disc (orange), bulge (blue) and the mass weighted average (blue). There is a major merger at $z \sim 1.8$ which transforms the galaxy from disc dominated to a giant elliptical. . . . . 54

2.13 A detailed schematic of the semi-empirical model presented in this chapter (adapted from Grylls et al., in prep). Box 'A' contains the processes that load merger trees and assign dark matter haloes galaxies. Box 'B' is the section of the model which evolves the galaxies from their formation to $z = 0$. Box 'C' relates to storing the simulated galaxies and retrieving them as to populate new dark matter haloes. . . . . 56

2.14 The median and $1\sigma$ scatter of the stellar masses of the progenitors of galaxies which have a final stellar mass log$M_{\rm star}$> 11.5 at $z = 0$ evolved using the SEM. . . . . 57
2.15 A comparison between the observed size-mass relation for local galaxies in SDSS from (Bernardi et al., 2013, green line) and the results of the SEM assuming all satellites have $f_{orb} = 0$ (parabolic orbits, blue line) and an $f_{orb} = 0.5$ (intermediate orbit, orange). . . . . 57
2.16 The median and $1\sigma$ dispersion of the effective radii of the progenitors of galaxies with log$M_{\rm star}$> 11.5$[M_{\odot}]$ at $z = 0$. The error bars show the sizes of typical elliptical galaxies from SDSS (Bernardi et al., 2013, purple), DEEP2 (Trujillo et al., 2007, Grey) and CANDLES (green van der Wel et al., 2014) with the same stellar masses as the galaxies shown from the SEM. . . . . 58
2.17 Same as Figure 2.15 but showing the effect of including environmental quenching in the SEM. . . . . 59
2.18 The mean number of mergers between $z = 1$ and $z = 0$, above a given merger ratio, $\mu$ (defined at the time of the merger) a galaxy (left) or halo (right) has as a function of their $z = 0$ mass. . . . . 60

3.1 A comparison among estimates of the stellar mass-to-halo mass relations. In this chapter, I adopt the relations by Moster et al. (2013) as well a steeper version which matches the relations by Shankar et al. (2014b) and Kravtsov et al. (2014). . . . . 64
3.2 The redshift dependence of the $M_{\rm star}$-$M_{\rm halo}$ relation using the parameters of Model I from Table 3.1. . . . . 67
3.3 The top panels are visualizations of two dark matter merger trees and bottom panels show the mass evolutions of the main progenitor (solid) and the most massive progenitor at $z = 2$ (dashed line). The panels labelled A show the evolution of a halo where the main progenitor halo is not the most massive halo at high redshift. The panels labelled B instead shows a more idealised case where the main progenitor is the most massive at all the redshifts I consider. . . . . 68
3.4 The mass functions of the progenitor haloes that contain galaxies with log $(M_{\rm star}) > 11.5\ M_{\odot}$ at $z = 0$. The solid and dashed lines show the main progenitor and most massive progenitor mass functions at different redshifts, as labelled. . . . . 70
3.5 A plot showing the predicted mean stellar mass of galaxies with log $(M_{\rm star}) > 11.5\ M_{\odot}$ at $z = 0$ (blue triangle at $z = 0$) using the original Moster et al. (2013) $M_{\rm star}$-$M_{\rm halo}$ relation (bottom) and a modified version to match the latest relation from (Shankar et al., 2014b, top) (models I and II, respectively). The one sigma range of this $z = 0$ point has been extended to all redshifts (blue band) so the z=0 point can be compared to points at higher redshift. I show the mean and $1\sigma$ of the total baryonic mass associated with the the main progenitor haloes (red circles) and most massive progenitors (green circles). The decline in $M_{\rm star}$ after star formation has stopped in due to older stars dying (Behroozi et al., 2013b). For reference, I also show the stellar mass estimates of galaxies in the respective haloes using the given $M_{\rm star}$-$M_{\rm halo}$ relations (Model I or Model 2; blue triangles). 71

3.6 Top: the integrated number density of dark matter haloes in the Bolshoi simulation. The black line shows the distribution for all haloes in the simulation and blue shows those which are included in the stellar mass selection (log ($M_{\rm star}$) > 11.5 $M_{\odot}$). Middle: the differential mass function of dark matter haloes at z=3 which survive to z=0. The black line shows the total halo distribution and blue are those haloes which enter the stellar mass selection. Bottom: I apply a halo mass selection at z=3 (highlighted in the middle panel) and follow those haloes forward to z=0. The black shows the distribution of haloes when selected from the total distribution and blue when selected from those which are included in the stellar mass cut. . . . . 74

3.7 The mean star formation track of galaxies selected with log ($M_{\rm star}$) < 11.5 $M_{\odot}$ at $z = 0$ using the mass-dependant empirical star formation rates of Peeples & Somerville (2013, top panel) and Tomczak et al. (2016, middle panel). The y-axis shows the mean stellar mass of all galaxies in the sample. The band shows the systematic uncertainty in the $M_{\rm star}$-$M_{\rm halo}$ relation. The unfilled band shows the full evolutionary track from $z = 3$ to $z = 0$ and solid band shows the effects if the star formation is quenched at $z = 2$. The bottom panel instead shows a band bracketing the mean evolutionary path assuming a constant star formation of 500 and 1000 $M_{\odot}/yr$. For reference, I also show the total baryonic mass in the progenitor haloes as a red band between $z = 2$ and 3. . . . . 76

3.8 The evolution of the mean stellar mass of a galaxies with *log* ($M_{\rm star}$) > 11.5 at $z = 0$ evolved using the SEM in Shankar et al. (2014a). The shaded region shows the statistic dispersion in the galaxies' evolutionary histories. The error bars show the predictions from abundance matching for galaxies with log ($M_{\rm star}$) > 11.5 $M_{\odot}$ where model I and II are shown in yellow and red respectively. For reference, I also show the the mean evolution in the mass of the most massive galaxies in the illustris simulation (Torrey et al., 2015). . . . . 79

3.9 The median evolution of the effective radius of a galaxies with log ($M_{\rm star}$) > 11.5 at $z = 0$ evolved using the SEM in Shankar et al. (2014a). The shaded area represents the systematic uncertainty of the model when gas dispersion included. . . . . 80

4.1 A cartoon sketching how two contrasting redshift evolutions in the $M_{\rm star}$-$M_{\rm halo}$ relation predicts predict different evolutions in $M_{\rm star}$. Panel A shows the mean evolution of a local log $M_{\rm halo}$= 15[$M_{\odot}$] halo. Panels B and C show a gradually decreasing $M_{\rm star}$-$M_{\rm halo}$ relation and a static one respectively. The blue arrows on the relations show the evolution in $M_{\rm halo}$from panel A along with the corresponding evolutions in $M_{\rm star}$. Panels D and E explicitly show the corresponding evolutions in $M_{\rm star}$. . . 84

4.2 A plot showing the integrated halo mass function of Tinker et al. (2008) (orange), the integrated mass function of all subhaloes by George et al. (2011) (blue) and their sum (green). These are the principle mass functions used to build halo catalogues in this thesis. . . . . 88

4.3 A cartoon showing how I find the redshift of infall for a satellite galaxy. Here the orange circles represent a halo which starts as a distinct halo at $z > z_{\rm inf}$and falls into a new host halo, represented in blue, at $z = z_{inf}$ and remains a satellite until $z_{obs}$. . . . . 89

4.4 The distribution of redshift of infall, $z_{\rm inf}$, for all sub haloes in the Bolshoi simulation with $\log M_{\rm halo}/M_{\odot} > 12$ at $z = 0.5$. This distribution is randomly sampled to assign a $z_{\rm inf}$ to sub-haloes in the derived catalogue. 90

4.5 A comparison among estimates of the stellar $M_{\rm star}$-$M_{\rm halo}$ relations from the literature at $z = 0.1$. The black solid line shows the 'no evolution' relation which I define in section 4.2.2, in which I use the analytic relation by Moster et al. (2013) changing the high mass slope and the scatter to fit the Bernardi et al. (2013) stellar mass function. Additionally, I show three reference slopes with $\delta = 0.6$, 0.45 and 0.3. . . . . 92

4.6 Blue: the boss incompleteness at $z = 0.5$ from Leauthaud et al. (2016). Green: the expected communicative number of satellite galaxies that are in all the galaxy clusters in the $750^3[Mpc^3/h^3]$ box I use in this chapter (assuming the Moster et al. (2013) $M_{\rm star}$-$M_{\rm halo}$ relation) as a function of the satellite galaxy's mass. . . . . 93

4.7 The galaxy halo mass distributions $n_{\rm gal}(M_{\rm halo})$ of massive galaxies with stellar mass $\log M_{\rm star}/M_{\odot} > 11.5$. I show the effects of changing the high-mass end slope $\delta$ (top) and the intrinsic scatter $\sigma$ (bottom). . . . . 95

4.8 (top) The results of the MCMC for central galaxies at $z = 0.5$. The histograms show the probability distributions of the high mass slope, $\delta$ and intrinsic scatter, $\sigma$ at $z = 0.5$. The shaded areas shows the 1, 2 and 3 standard deviation confidence regions. The number quoted on top of each histogram is the maximum likelihood of that parameter with its associated error. Additionally, I show the value Moster et al. (2013) high mass slope (red dashed line) and the slope that would imply no evolution in stellar masses between $z = 0.5$ and $z = 0$ (solid blue line) in the top left histogram. (Bottom) The stellar mass function taken from the maximum likelihood parameters of the MCMC as compared to the $z = 0$ stellar mass function from Bernardi et al. (2013). . . . . 97

4.9 Same format as Figure 4.7, showing the $n_{\rm gal}(M_{\rm halo})$ parent halo mass distributions for *satellite* galaxies with stellar mass $\log M_{\rm star}/M_{\odot} > 11.5$. . 98

4.10 Similar format to Figure 4.8. Here I have added the constraints on the $n_{\rm gal}(M_{\rm halo})$ of massive satellite galaxies to probe higher redshifts $z > 0.5$ (with mean redshift of infall $z_{\rm inf}$= 0.7). I assume a "frozen" model, i.e., satellites maintain constant stellar mass after infall. I have also added for comparison the values for $\delta$ from Moster et al. (2013, red, long-dashed line) at $z > 0.5$, and the no-evolution model (solid, vertical lines) with two values of the scatter, as labelled. In the bottom panel we additionally show the implied stellar mass function derived from the maximum likelihood parameters of the $M_{\rm star}$-$M_{\rm halo}$ relation at $z = 0.5$ and at $z > 0.5$ (orange, dot-dashed and green, dotted lines, respectively). . . . . 100

4.11 Same as Figure 4.8 but now assuming that all satellites loose 30% of their stellar mass at $z_{\rm inf}$. . . . . 102

4.12 Top: the blue data points show the stellar masses of BCGs in the McDonald et al. (2016) sample. The horizontal line shows that $\log M_{\rm star}$= $12.1[M_{\odot}]$ is representative of the stellar masses of the sample. Bottom: A comparison between the star formation rates of the BCGs shown in the top panel (measured through the labelled technique) as well the empirical star formation rates by Tomczak et al. (2016) and Peeples & Somerville (2013) using a stellar mass of $\log M_{\rm star}$= $12.1[M_{\odot}]$ at all redshifts. . . . . 104

4.13 Same as Figure 4.11 but now also including star formation in satellite galaxies between their $z_{\rm inf}$and $z = 0.5$ at a rate given by Tomczak et al. (2016) and additional mass loss due to stellar evolution given by Behroozi et al. (2013b). . . . . 105
4.14 The predicted mean evolution of stellar mass in individual galaxies, anchored at z=0.5. I have selected haloes in the Bolshoi simulation at $z = 0$ that have log$M_{\rm star}\sim$ 11.7 using the Moster et al. (2013) $M_{\rm star}$-$M_{\rm halo}$ (blue line) and abundance matching to the Bernardi et al. (2013) stellar mass function (errorbars). I trace the progenitors of these haloes to higher redshift. In the blue line shows the median stellar mass of the progenitors when populated using the Moster et al. (2013) $M_{\rm star}$-$M_{\rm halo}$ relation. The error bars show the mean stellar mass when populated using the maximum likelihood $M_{\rm star}$-$M_{\rm halo}$relation assuming the frozen model. The horizontal line shows a constant stellar mass. The orange line shows the results of a full semi-empirical model from Shankar et al. (2015). . . . . 109

5.1 The fractional expansion of a galaxy after a portion of its mass has been removed. The blue, green and purple lines show when the stellar mass starts as 3, 2 and 1.5 times the final bulge mass (plus the disk mass). In other words, when 20%, 33% and 50% of the stellar mass is removed. The orange, red and brown lines show the same but for a thicker disc geometry (the disk height, $DH = h_z/R_{05m}$). Time starts when the particles are removed. . . . . 116
5.2 The 3D density profile of an over-massive galaxy with the initial conditions described in Section 5.2.1 immediately after its initialization (blue line) and after $0.5Gyr$ of time (orange line). The vertical lines show their respective half-mass radii. . . . . 117
5.3 An image of the projected 2D density of the dark matter halo I simulated at $z \sim 4.3$ within a cube with sides $6 \times R_{200c}$. Darker regions show higher densities. The orange circle shows $R_{200c}$. . . . . 120
5.4 The evolution of $M_{\rm halo}$ with redshift at the centre of each of Martizzi et al. (2014) zoom-in simulations (coloured lines). Additionally, I show $M_{\rm halo}$ of my simulated halo at $z \sim 4.3$ (blue star). . . . . 121
5.5 The purple line shows the 3D density profile of the central galaxy of the simulation run in this work at $z \sim 4.3$ (top) and one example fit from the Martizzi et al. (2014) simulations at $z = 0$ (bottom). The red line shows the best fit single Hernquist profile. Instead, the blue line shows the best fit double Hernquist profile, where the orange line shows the first 'galaxy' profile and the green shows the second 'ICL' profile. . . . . 123
5.6 The coloured lines show the evolution of $M_{\rm star}$ in each of the central galaxies of the zoomed-in dark matter halo in the Martizzi et al. (2014) simulations. The blue star shows the $M_{\rm star}$ of the central galaxy of the zoomed-in dark matter halo in my simulation at $z \sim 4.3$. The top panel shows $M_{\rm star}$ when only a single Hernquist profile is used. The middle panel shows $M_{\rm star}$ measured using the first 'galaxy' profile and the bottom panel shows the sum of the 'galaxy' and 'ICL' profiles when two Hernquist profiles are fitted. . . . . 124
5.7 The mean ratio of the baryonic (stars plus gas) mass to the dark matter mass within $R_{200c}$ in the zoomed-in dark halo in the Martizzi et al. (2014) simulations (coloured lines) and of this work (blue star). . . . . 125

5.8 The evolution in the stellar mass of the central galaxies in the zoomed-in halo in the Martizzi et al. (2014) simulations (dashed orange lines) and of this work (orange star). Each line represents a different simulation. Additionally, I show the evolution of the total baryonic (stellar and gas) mass within $R_{200c}$ (blue solid lines and blue star). . . . . 126
5.9 Same as Figure 5.6, but showing the evolution the half-mass radius, $R_{0.5M}$. 127

A.1 The median and one standard deviation mass accretion histories of main progenitor haloes with $z = 0$ masses: $\log M_{halo} = 12, 13, 14$ & 15. The bottom right panel is constructed using haloes in the Bolshoi simulation. The other panels are constructed using algorithms based on extended Press-Schector theory. In all cases except $\log M_{halo} = 15$ in the Bolshoi simulation (due to volume limitations), 100 haloes are simulated. . . . . 139
A.2 A comparison between the mass accretion histories of three extended Press-Schector bases algorithms to that of haloes in the Bolshoi simulations. . . . . 140
A.3 A comparison between the mean $1^{st}$ order subhalo mass functions predicted with three extended Press-Schector bases algorithms for producing merger trees to analytic relation of Giocoli et al. (2008) for two initial halo masses. . . . . 140

B.1 The mean $z_{\rm inf}$ for different host halo masses and sub-halo mass at infall. I fond that there is only a weak trend with host halo mass. . . . . 143

C.1 A comparison between the central (top) and satellite (bottom) $dP/d\log M_{\rm halo}$ distributions from Tinker et al. (2016, Blue) and Guo et al. (2014, red). . 146
C.2 The probability that a halo with a given mass contains a central galaxy with a mass $\log M_{\rm star} > 11.5[M_{\odot}]$ at $z = 0.5$ (blue) and $z = 0.7$ (orange) from Guo et al. (2014). . . . . 147
C.3 A comparison between the satellite $dP/d\log M_{\rm halo}$distributions from Tinker et al. (2016, Blue) and Guo et al. (2014, red) with the self consistency test described in this Appendix. . . . . 147

# List of Tables

2.1 The mapping between stellar mass had halo mass corresponding to the example shown in Figure 2.3. . . . . 38

3.1 The parameters used in equation 1. The top portion gives the $z = 0$ parameters of the $M_{\rm star}$-$M_{\rm halo}$ relation for both the original Moster et al. (2013) relation as well as the steeper version from Shankar et al. (2014b). The bottom portion gives the redshift dependence of the $M_{\rm star}$-$M_{\rm halo}$ relation. . . . . 66

B.1 Properties of the Neo4j database: the number of nodes and relationships broken down by label. . . . . 142

# Declaration of Authorship

I, Stewart William Buchan, declare that this thesis titled, 'Probing the evolution of the most massive galaxies' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: _____

Date: 13/02/2018

# *Acknowledgements*

First and foremost I offer my sincerest gratitude to my supervisor, Dr Francesco Shankar, who has supported me throughout my Ph.D study and research. Without his patience and relentless effort, this thesis would not have been completed. It was an honour to be your first solo PhD student and I feel as if we went on the journey together and both came out better.

I am grateful to my parents and brother for always giving me their full attention and unconditional support in all my pursuits in the past 26 years. I hope I have made you proud.

I thank my wonderful girlfriend, Amy, for her constant encouragement and project management skills which were vital in helping me to complete this thesis. I am eternally grateful to Adam for mentoring me in the ways of data science. He has unmeasurably improved my PhD and has been an inspiring role model. Furthermore, I'd like to thank the whole Pivigo team and everyone at the 2016 London S2DS program for showing me I have the skills to contribute outside of academia.

I thank Chris F and Andy C for coaching me through the hard times; my office mates Juan, James and Judith who created an amazing atmosphere to work in; Steve, Szymon and Henry for being awesome housemates; Boon and Chris R for satisfying my alcohol needs; Jamie for organising D&D nights; Mat and Aarron for pushing me to get fitter. You have all been an amazing people without who, I wouldn't have made it to the end.

I wish my Pip all the best in continuing my work with Francesco.

I thank the STFC for funding the research that went into this thesis.

# Chapter 1

# Introduction

## 1.1 Overview

The vast ocean of space is full of starry islands called galaxies. These distant objects are dynamically bound structures comprised of gas, stars and 'dark matter'. Understanding the exact nature of galaxies has puzzled astronomers since the late 18th century when they started to identify large numbers of extended objects which they called 'nebulae'. Modern astronomers, using large telescopes with CCD cameras and multi-object spectrographs have been able to observe millions of such objects, now known as galaxies. However, there are still difficulties in trying to understand how to connect galaxies at different distances (and therefore look-back times) to probe how an individual galaxy evolves. It is impossible for us to actually observe a galaxy changing in real time because the relevant timescales for galaxies evolution are vast, ranging from $\sim 10^4 - 10^{10}$ years.

Despite significant efforts both on the observational and theoretical fronts, a standard model for the evolution of all galaxy types has not yet been achieved. A clear theoretical framework is particularly lacking for the most massive galaxies, which suffer from a number of observational selection effects (as discussed in Section 1.6.3). The goal of this thesis is to place constraints on the evolution of the most massive galaxies via the aid of phenomenological and semi-empirical models. These are a branch of theoretical models which are developed from the ground up using empirical relationships and numerical dark matter merger trees. Since these models directly utilise observed relationships rather than idealised physics, by construction, they have few input assumptions and free parameters.

## 1.2 Modern cosmology

When trying to understand how galaxies form and evolve, it is important to first have knowledge of how the Universe has evolved. To this end, I first describe the basic principles of modern cosmology.

### 1.2.1 Cosmological distances, redshift and look back time

The first important concept is that galaxies are extremely distant[1]. This was debated in the early 20th century (Shapley & Curtis, 1921) since measuring distances to astronomical objects is very challenging. One of the original and still widely used methods for measuring distances to galaxies is to use the Cepheid variable stars in that galaxy as standard candles since they have a known period-luminosity relation (Leavitt, 1908; Leavitt & Pickering, 1912).

One important observation made around the same time was that atomic emission lines in a galaxy's optical spectra are shifted, usually to longer wavelengths, relative to those lines observed in the laboratory (eg. Slipher, 1917). This change is due to the doppler effect because most galaxies are moving away from us. This effect is known as redshift, $z$:

$$1 + z \equiv \frac{\lambda_{ob}}{\lambda_{lab}} = \sqrt{\frac{1 + \frac{v}{c}}{1 - \frac{v}{c}}}\,, \tag{1.1}$$

where $\lambda_{ob}$ is the observed wavelength of an emission line, $\lambda_{lab}$ is the wavelength of the same line in a laboratory, $v$ is the line-of-sight velocity of the galaxy and $c$ is the speed of light.

Hubble (1929) compared the distances and redshifts of galaxies (see Figure 1.1). It is immediately obvious that there is a strong correlation whereby the further away a galaxy is, the faster it is moving away from us. The gradient of this relation is known as Hubble's constant, $H = 100 \times h[km\ s^{-1}\ Mpc^{-1}]$. This correlation strongly implies that the universe is expanding. One simple thought experiment is to reverse time; if the Universe is expanding, reversing time means it's contracting. If the contraction is constant throughout all time, the Universe will start as a singularity. This thought experiment lead to the notion of "The Big Bang".

[1] Astronomers use the parsec (pc) as the unit of distance. It is defined as the distance which an object subtends 1 arcsecond of parallax.

Furthermore, matter should slow down the expansion, due to gravity. The possible evolutionary scenarios are as follows. If the density is high enough, expansion will stop and the Universe will collapse again. If the density is too low, the Universe will expand forever. This lead to the idea of a critical matter density, $\rho_{crit}$, (or energy density because of mass-energy equivalence) which is perfectly in between these two cases. This concept was upset by the discovery of the accelerating universe (Riess et al., 1998) using observations of supernova, which bought about the concept of 'dark energy', $\Lambda$, which adds a constant repulsive pressure.

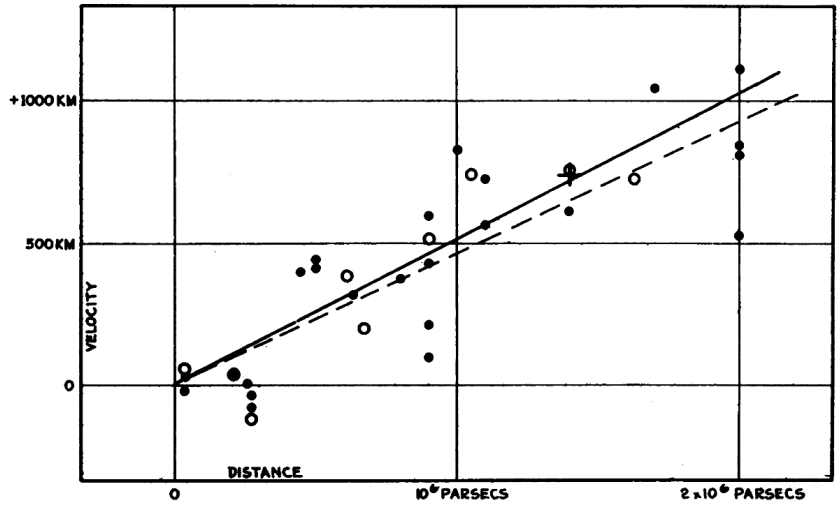



FIGURE 1.1: Velocity-Distance Relation among Extra-Galactic galaxies extracted from Hubble (1929)

One last note is that due to the finite speed of light, galaxies at high redshift are observed to be more closer in time to the big bang than more local ones. This means we can probe different evolutionary epochs by exploring galaxies at different redshifts (see Figure 1.2).

### 1.2.2 What is the universe made of?

With the advent of precision cosmology, the last decade has allowed cosmologists to accurately measure the energy density of the Universe. This is usually quoted as a dimensionless density parameter, $\Omega = \rho/\rho_{crit}$. Studies which utilise baryon acoustic oscillations (Alam et al., 2016), measurements of the cosmic microwave background (Planck Collaboration et al., 2014), and distances to supernova (Sullivan et al., 2011) have measured the current energy-density of the Universe and how it is divided between matter, $\Omega_m$ and dark energy, $\Omega_\Lambda$. Figure 1.3 shows recent measurements of $\Omega_m$ and $\Omega_\Lambda$ to be 0.3 and 0.7 respectively from a combined supernova, CMB and BAO dataset (taken from Kowalski et al., 2008). Indeed, the most precise constraints on the cosmological parameters come from Planck: $\Omega_m = 0.308 \pm 0.012$ and $\Omega_\Lambda = 0.692 \pm 0.012$.

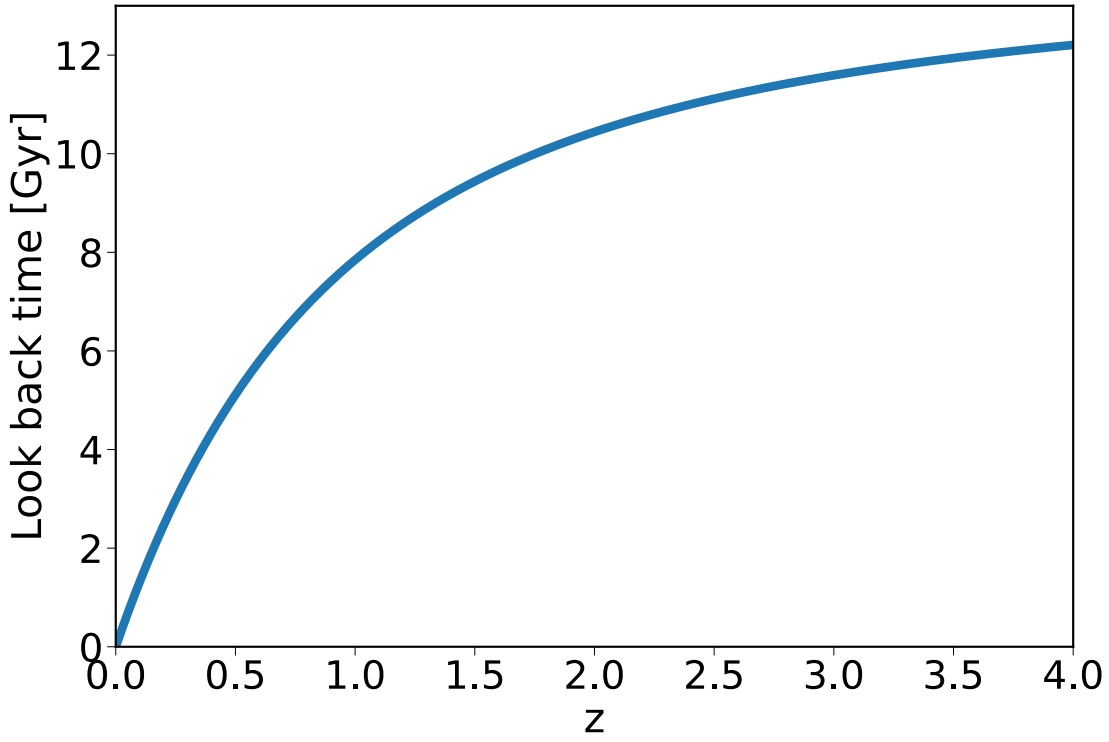



FIGURE 1.2: The relation between redshift and look back time for WMAP 9 cosmology (Bennett et al., 2013).

However, when one measures the density of baryons (in stars and gas), one finds it to be much lower than $\Omega_m = 0.3$ (Fukugita et al., 1998). Figure 1.4 shows a measurement of $\Omega_m$ against the fraction of mass in baryons, $f_b = \Omega_b/\Omega_m$ (extracted from Rota et al., 2017). It is clear from this figure that $f_b \lesssim 20\%$, implying that the majority of matter is not visible. Indeed, studies of motion of stars in the Milky Way (Kapteyn, 1922), of satellite galaxies in clusters (Zwicky, 1933) and the rotation curves of galaxies (Babcock, 1939) showed that there is a large amount of 'dark matter' around galaxies.

The most widely adopted theory in cosmology is that the Universe today is comprised primarily of dark energy and cold dark matter, known as ΛCDM (Lemaître, 1933; Riess et al., 1998). The exact nature of dark matter and dark energy are not critical for this thesis. To this end, dark matter is treated as a massive, weakly interacting particle that interacts with baryons via gravity but not via any other force.

### 1.2.3 The time line of the Universe

In this section, I briefly summarise the key phases in the evolution of the early Universe. For reference, a schematic is shown in Figure 1.5.

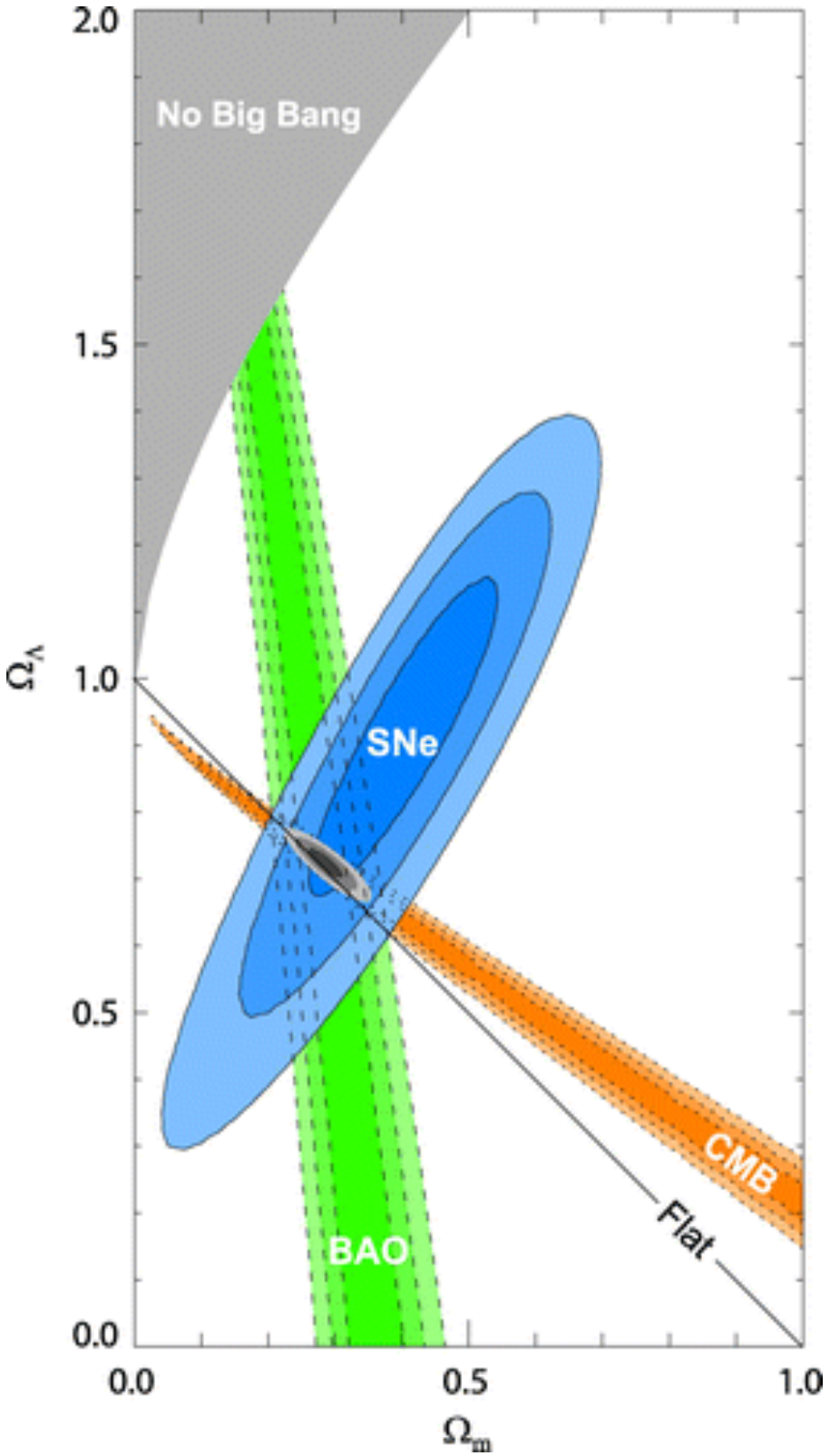



FIGURE 1.3: Contours at 68.3%, 95.4%, and 99.7% confidence level on $\Omega_\Lambda$ and $\Omega_M$ obtained from CMB, BAO, and the Union SN set (Kowalski et al., 2008), as well as their combination. Extracted from Kowalski et al. (2008).

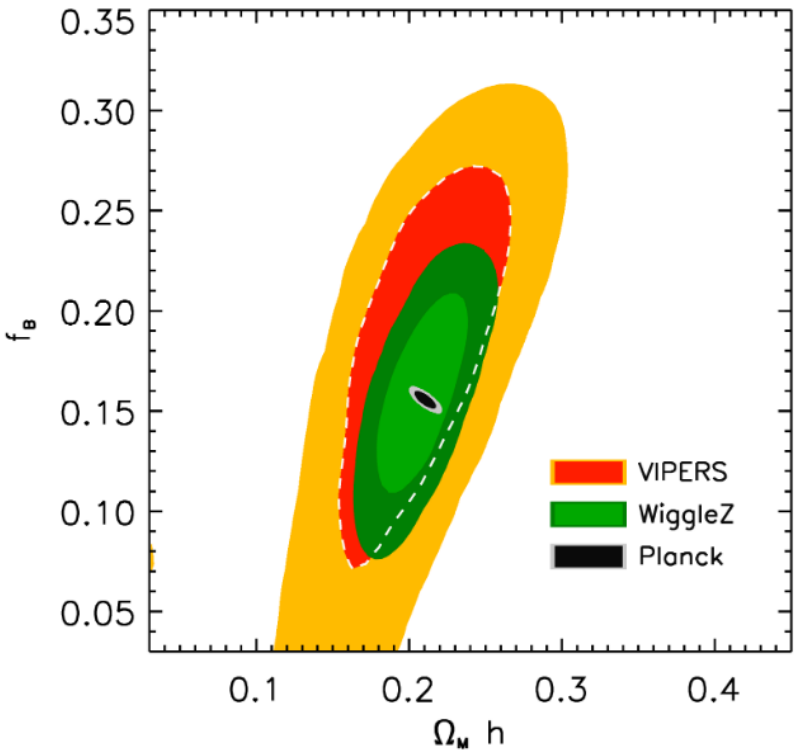



FIGURE 1.4: The constraints on the the baryon fraction $f_b = (\Omega_b/\Omega_M)$ and $\Omega_M h$ from VIPERS (Rota et al., 2017), WiggleZ (Parkinson et al., 2012) and Planck (Planck Collaboration et al., 2014). Extracted from Rota et al. (2017).

- Big bang: The "start" of the Universe. Immediately after the big bang, the universe is comprised of a plasma of elementary quantum particles.
- Inflation: In the first $\sim 10^{-34}s$ after the Big Bang, the Universe underwent a rapid expansion phase. Quantum fluctuations were frozen in place during inflation, giving rise to random density perturbation which are the initial seeds for structure formation.
- Structure formation begins: $\sim 70,000$ years after the Big Bang, the first dark matter structures begin to collapse.
- Surface of last scattering: $\sim 370,000$ years after the Big Bang, the Universe cooled significantly enough for atoms to form. At this time, photons are no longer coupled to the hot plasma and so the Universe went from being opaque to transparent. We observe this today as the cosmic microwave background.
- First stars: The fist stars form around 200-300 million years after the Big Bang. These stars can be extremely massive ($> 100M_\odot$) and thus are UV bright. They are then capable of re-ionising the primordial gas, dropping its pressure and thus promoting galaxy formation.

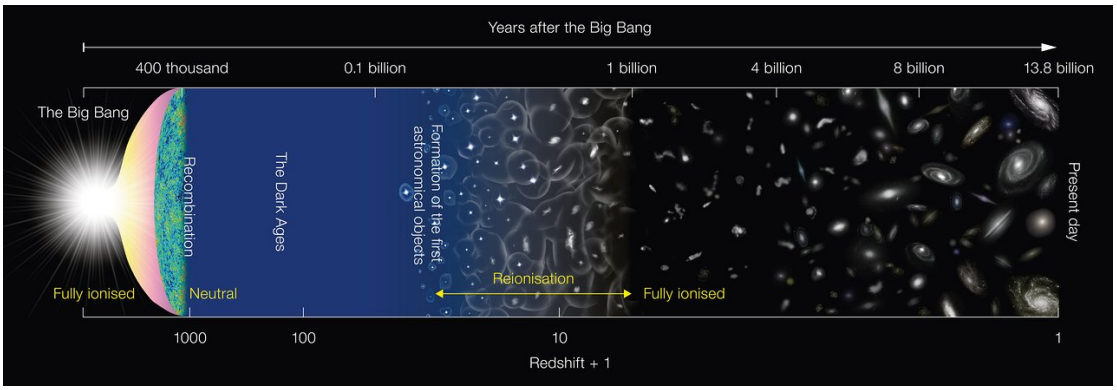



FIGURE 1.5: Schematic diagram of the history of the Universe, Credit NAOJ.

## 1.3 Structure formation

Having briefly established the main concepts in the current standard cosmological framework which describe the evolution of the Universe, I now discuss how dark matter structures form and grow. These dark matter structures, known as "haloes", form the potential wells that gas will gradually condense into to form galaxies. The latter would imply that there is an intimate connection between dark matter haloes and galaxies.

Over cosmic time, initial density perturbations (possibly arising from quantum fluctuations that expand during inflation) will collapse into bound structures, when they become dense enough. Small, high density regions collapse under gravity against the expansion of the Universe before larger, more massive regions. Bound structures can grow both via mergers and smooth mass accretion to form more massive haloes. This paradigm, known as the hierarchical growth of dark matter, is described in more detail below.

Here, I introduce two additional cosmological parameters which describe the distribution of dark matter in the universe. Firstly, the Universe can be described as a density field. The distribution of matter within this density field is normally described by its spatial Fourier transform. This indicates the power, $P$ at a given scale (the spatial wave number, $k$) and is characterised as a power law of the type:

$$P(k) = P(k_0)\left(\frac{k}{k_0}\right)^{n_s+(1/2)dn_s/d\ln k \ln k/k_0}, \tag{1.2}$$

where $n_s$ is the spectral slope at $k_0 = 0.05Mpc^{-1}$ and takes the value $n_s \sim 0.95$ (Kosowsky & Turner, 1995; Planck Collaboration et al., 2014). $P$ is a measure of the amount of variance in the density field that is accounted for by a wave with wave number $k$. The normalisation of this power law is more often than not quoted as $\sigma_8$, the linear,

rms fluctuation amplitude of the power spectrum within spheres of radius $8h^{-1}Mpc$, evaluated at $z = 0$. Planck gives one of the best estimate of $\sigma_8 = 0.829 \pm 0.012$.

The first attempt to quantitatively calculate the number of collapsed haloes at a given cosmic epoch was made by Press & Schechter (1974). They considered the Universe to be a comprised of a random Gaussian density density field, which can be expressed in terms of an overdensity field:

$$\delta(\vec{x}.t) = \frac{\rho(\vec{x}, t) - \bar{\rho}}{\bar{\rho}} \,. \tag{1.3}$$

According to the spherical collapse model, regions with $\delta(\vec{x}, t) > \delta_c \simeq 1.69$ will be a collapsed dark matter halo at time $t$. In linear theory; $\delta(\vec{x}, t) = D(t)\delta_0(\vec{x})$, where $D(t)$ depends on the cosmological parameters and is normalised to unity at the present day. Press & Schechter (1974) begins by smoothing the overdensity field with a spherical top-hat filter with characteristic radius $R$ to give the new smoothed field; $\delta_s(\vec{x}; R)$. By varying $R$, structures at different masses can be considered (with $M = 4\pi R^3\rho/3$). The ansatz of Press-Schechter formalism is that the probability that $\delta_s > \delta_c(t)$ is the same as the fraction of mass elements that, at time $t$, are contained in haloes with mass greater then $M$. Under the above assumptions, the probability that $\delta_s > \delta_c(t)$ is given by:

$$\begin{aligned} \mathscr{P}[> \delta_c(t)] =& \frac{1}{\sqrt{2\pi}\sigma(M)} \int_{\delta_c(t)}^{\inf} \exp\left[-\frac{\delta_s^2}{2\sigma^2(M)}\right] d\delta_s \,, \\ =& \frac{1}{2} \operatorname{erfc}\left[\frac{\delta_c(t)}{\sqrt{2}\sigma(M)}\right] \,, \\ \sigma_M^2 =& \left\langle \delta_s^2(\vec{x}; R) \right\rangle \,. \end{aligned} \tag{1.4}$$

Then following the Press-Schechter ansatz, equation 1.4 is equal to the mass fraction of collapsed objects with mass $> M$. At this stage, Press & Schechter (1974) also introduced the famous factor two 'fudge factor', such that $F(> M) = 2 \times \mathscr{P}[> \delta_c(t)]$. The number of objects between $M$ and $M + dM$ (the halo mass function) is then given by:

$$
\begin{aligned}
n(M,t) =& \frac{\bar{\rho}}{M}\frac{\partial F(>M)}{\partial M}dm = 2\frac{\bar{\rho}}{M}\frac{\partial \mathscr{P}[>\delta_c(t)]}{\partial \sigma}\left|\frac{d\sigma}{dm}\right| dm\,, \\
=& \sqrt{\frac{2}{\pi}}\frac{\bar{\rho}}{M^2}\frac{\delta_c}{\sigma_M}\exp\left[-\frac{\delta_M^2}{2\sigma_M^2}\right]\left|\frac{d\ln\sigma_M}{d\ln M}\right| dM\,.
\end{aligned}
\tag{1.5}
$$

During the 1990's, much effort was put into developing the extended-Press-Schechter formalism(Peacock & Heavens, 1990; Bond et al., 1991; Bower, 1991; Lacey & Cole, 1993; Jedamzik, 1995; Yano et al., 1996; Nagashima, 2001). In part, this was to explain the factor two but also to broaden the input assumptions.

In parallel to development of the extended-Press-Schechter formalism, substantial effort went into the study of structure formation using N-body simulations (Efstathiou et al., 1988; Lacey & Cole, 1994; Gross et al., 1998; Governato et al., 1999; Somerville et al., 2000). These initial works find that, despite the obvious approximations found in extended-Press-Schechter formalism, it does a remarkably good job of reproducing the evolution of the halo population.

With the increase in computational power, cosmological dark matter simulations could be run with greater spatial and mass resolution, revealing large numbers of sub-structures in each distinct halo (Klypin et al., 1999; Colín et al., 2000). Figure 1.6 shows snapshots from a high resolution, N-body simulation, extracted from Baugh (2006). Sub-structures originate as distinct haloes which 'fall into' a more massive halo. Each infalling halo is observed to orbit round the more massive host halo, slowly losing mass due to tidal effects, and eventually completely merging into it. The merging histories of haloes can be represented as a tree (Lacey & Cole, 1993), where the trunk is the 'main progenitor' (the most massive halo during any merging event) and the branches are mergers between haloes. The base of the tree is at $z = 0$ with the branches at higher redshifts. In Appendix A, I provide a detailed descriptions and comparisons of both extended-Press-Schechter based, analytic algorithms for generating merger trees and the results of N-body simulations. In this theses, I will utilise merger trees constructed using both methods.

An important assumption I make throughout this thesis is that every distinct halo has exactly one *central* galaxy and, by extension, each sub-halo will host a *satellite* galaxy. This is a safe assumption for the galaxy masses I probe in this thesis which are not affected by e.g., the missing satellite problem (Bullock, 2010, and refs. therein).

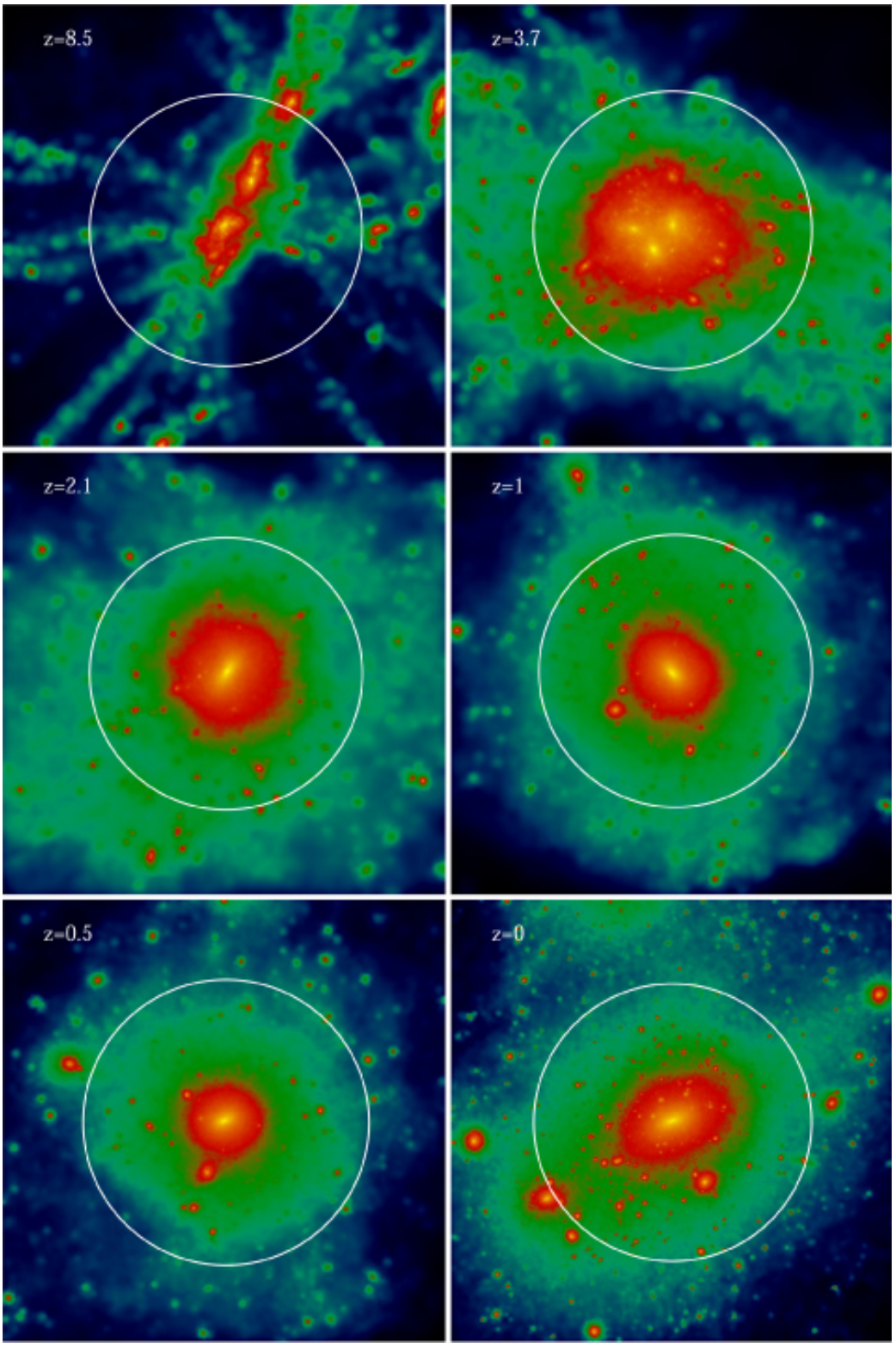



FIGURE 1.6: The formation of a dark matter halo in a high resolution N-body simulation of a $3 \times 10^{11}[h^{-1}M_{\odot}]$ halo (extracted from Baugh, 2006). The circle marks the present day virial radius. The colours reflect the density of dark matter, with redder colours indicating spatial higher density.

## 1.4 Observations of local galaxies

In Sections 1.2 and 1.3 I covered the theoretical framework for dark matter evolution. In this section, I discuss the current observational understanding of galaxies in the local Universe ($z \lesssim 0.1$). In the next section I discuss observations of galaxies at higher redshifts, closer to their formation epochs.

### 1.4.1 The Hubble galaxy classification

One of the most common methods of classifying galaxies is through their visual morphology. Figure 1.7, shows the widely used classification scheme, first laid out by Hubble (1936), known as the Hubble tuning fork. It shows the division between spiral galaxies, characterised as being disc-like and having spiral arms and dust lanes (right in the diagram) and elliptical galaxies, which are quasi-spherical and featureless (left). In between these two categories are lenticular galaxies, which are disc-like but usually lack spiral features or dust lanes. Ellipticals and lenticulars are then collectively known as "early-type" galaxies and spirals as "late-type". Anything that doesn't fit into these categories are known as 'irregular' galaxies.

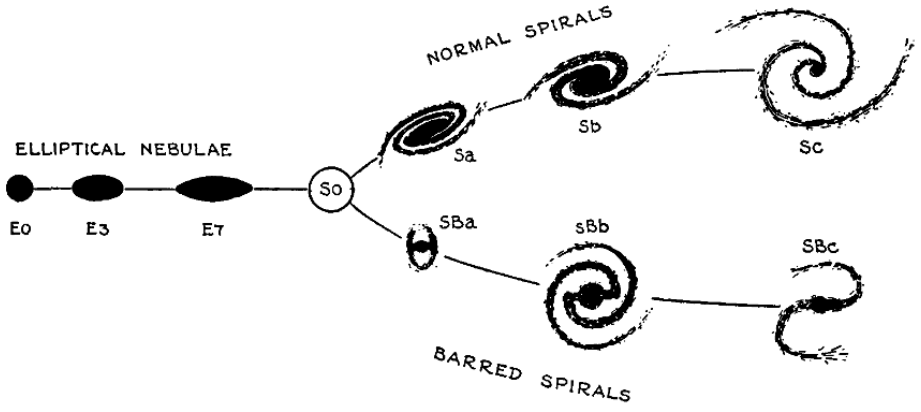



FIGURE 1.7: A schematic representation of the classifications of 'nebulae' (now called galaxies). Extracted from Hubble (1936).

The evolution of the most massive elliptical galaxies is the topic of this thesis. They are universally classified as having smooth elliptical morphologies, usually containing little gas and dust and are red in colour, meaning they have old stellar populations. They are $\sim 3 - 100$ times more luminous than the Milky Way and have characteristic sizes of tens of kiloparsecs. The stars in giant ellipticals show very little ordered motion (i.e. rotation) thus implying they are pressure supported.

### 1.4.2 Measuring physical galaxy properties

In the following section I describe how some key, relevant physical properties of galaxies are measured.

- Galaxy light profiles and effictive radii: A typical approach to characterise the structure of a galaxy is to measure its surface brightness as a function of radius, i.e., the integrated light through an annulus of radius, $r$. The surface brightness profile, $I(R)$, derived in this way can well fit by a Sérsic (1963) profile:

$$I(r) = I_0 \times \exp\left(-b(n) \times \left[\left(\frac{r}{r_e}\right)^{1/n} - 1\right]\right) , \tag{1.6}$$

  where $I_0$ is the normalisation, $n$ is the "Sersic index" and $b(n)$ is calibrated such that half of the galaxy's light is contained within the radius $r_e$. Disc galaxies (spirals) have $n = 1$ and ellipticals tend to have $n \gtrsim 4$, and up to $8 - 10$ for the most massive, central galaxies.

- Velocity dispersion and rotation: An additional probe of a galaxy's structure is the mean orbital velocity of its stars. In rotationally supported galaxies (spirals and lenticulars), a long slit spectragraph or, in recent years, an integral field unit can be used to probe rotation curves. In pressure supported galaxies (ellipticals), the mean light-of-sight velocity of stars can be extracted from the doppler broadening of atomic emission and absorption lines in the galaxy's spectra, such as $H_\alpha$ and $H_\beta$ .

- Stellar mass: The main method to measure the stellar content of a galaxy is to fit its spectrum to model stellar population templates (known as SED fitting; e.g. Maraston, 2005; Cid Fernandes et al., 2005; Ocvirk et al., 2006; Conroy et al., 2009). Groups have created such mock stellar populations by inputting a stellar initial mass function (IMF, e.g. Salpeter (1955); Kroupa (2001); Chabrier (2003)), into standard stellar evolution recipes. A synthetic spectral templates are then created by summing the spectra of the evolved stars assuming different metallicities and ages for the populations. These templates can then be fit to a galaxies spectrum to give its stellar mass.

- Gas fraction: The method used to observe gas in galaxies depends heavily on the gas' temperature. The most important gas phases are; firstly, the molecular gas which is used in star formation, and is observed using tracers such as CO (Israel, 1997). Secondly, a large proportion of the gas in early-type galaxies is locked up in atomic hydrogen. This is probed using radio observations of the 21cm line, cased

by the flip of the electron's spin. Lastly, hot X-ray gas can be used to measure dark matter halo masses and is detected through deep X-ray observations.

- Star formation rate: There are several methods for tracking the formation of new stars, all of which rely on tracing the number of massive, short lived stars (e.g. Kennicutt, 1983; Condon, 1992; Cowie et al., 1997). The most common methods use the UV continuum to directly probe the number of young, hot stars. Additionally, a fraction of the UV luminosity is absorbed by the interstellar medium and dust. The former can be traced by recombination lines (such as $H_\alpha$, $H_\beta$, $P_\alpha$, $P_\beta$, $Br_\alpha$, $Br_\gamma$, etc.) and the latter by the far-infrared emission. Thus both these can be used as proxies for the intrinsic UV luminosity and thus star formation rate. The star formation rate in local galaxies can range from virtually zero in early-type galaxies to $\lesssim 20$ $[M_\odot/yr]$ in late-type galaxies (although, it can be an order of magnitude higher in galaxies undergoing a "starburst").

- Dark matter mass: There are a number of ways to measure the host dark matter halo mass of a galaxy. Many studies have used the weak-lensing (Tyson et al., 1984; Hoekstra et al., 2005; Mandelbaum et al., 2006) and strong-lensing (Auger et al., 2010; Sonnenfeld et al., 2013; Dye et al., 2014; Jauzac et al., 2015) of background galaxies to probe the total line-of-sight mass. Other techniques include modelling the rotation curves of spiral galaxies (Sofue & Rubin, 2001), measuring the properties of hot X-ray gas (Vikhlinin et al., 2006) and the dynamics of globular clusters (Romanowsky et al., 2009).

### 1.4.3 Statistical properties of galaxies

In the previous section I described how to measure some relevant properties of individual galaxies. In this section I focus on the statistical properties of galaxies that will be used throughout this thesis.

#### 1.4.3.1 Luminosity and stellar mass functions

Possibly the most fundamental statistical property of galaxies is the number density of galaxies as a function of their intrinsic luminosity. The blue line in the top panel of Figure 1.8 displays one of the latest renditions of the local ($z \lesssim 0.1$) luminosity function, $\Phi(L)$, from Bernardi et al. (2013), defined as the number of galaxies per unit volume in the luminosity range $L \pm dL/2$. In general, the luminosity function is well described by a power law with an exponential cut-off at the high luminosity end. This is otherwise known as a Schechter function (Schechter, 1976) and has the form:

$$\Phi(L)dL = \Phi_\star \left(\frac{L}{L_\star}\right)^\alpha \exp\left[-\left(\frac{L}{L_\star}\right)\right] dL\,. \tag{1.7}$$

The other coloured lines in the top panel of Figure 1.8 show how the luminosity function is split between the different morphological types (as labelled). It is clear that elliptical galaxies dominate the high-luminosity end whereas spirals dominate the low-luminosity end.

The bottom panel of Figure 1.8 shows the corresponding local stellar mass function, which will be utilised extensively throughout this thesis. As evident from this panel, the stellar mass function has, as expected, a very similar morphological trend to the luminosity function.

#### 1.4.3.2 The $M_{\rm star}$-$M_{\rm halo}$ relation

As conveyed in Section 1.3, galaxies are born and evolve within their host dark matter haloes. This implies that there should be an intimate connection between the evolutionary pathways of (central) galaxies and their dark matter hosts, especially in terms of the stellar and halo mass growth. Indeed, direct measurements of massive galaxy stellar masses and their halo mass have revealed a strong correlation between the two properties. More statistical approaches, such as abundance matching whereby the integrated stellar mass function and halo mass functions are equated (e.g., Vale & Ostriker, 2004; Shankar et al., 2006; Moster et al., 2010; Leauthaud et al., 2012; Behroozi et al., 2013b; Moster et al., 2013; Gu et al., 2016; van Uitert et al., 2016, and see Section 2.2.2) have shown that this relationship is best fit by a double power law, and there is remarkably small scatter about the relation ($0.15 - 0.20\ dex$). The relationship between a galaxy's stellar mass and its host halo mass will be utilised heavily throughout this thesis. More specific details will be given in Section 2.2.2.

#### 1.4.3.3 The Colour-mass distribution

Another key observed quantity is a galaxy's colour. In this context, colour is defined as the ratio of the galaxy's flux in two optical bands (or the difference in magnitudes in two bands) which is usually redshift and dust corrected. Figure 1.9 shows the distribution of colours and stellar masses for galaxies in the Sloan Digital Sky Survey (Eisenstein et al., 2011; SDSS Collaboration et al., 2016, SDSS). From this figure, it is clear that there is a bimodality in galaxy colours, where the red galaxies belong to the "red sequence" and blue galaxies to the "blue cloud", separated by the "green valley" (Strateva et al., 2001; Hogg et al., 2002; Blanton et al., 2003; Baldry et al., 2006). The blue band of a

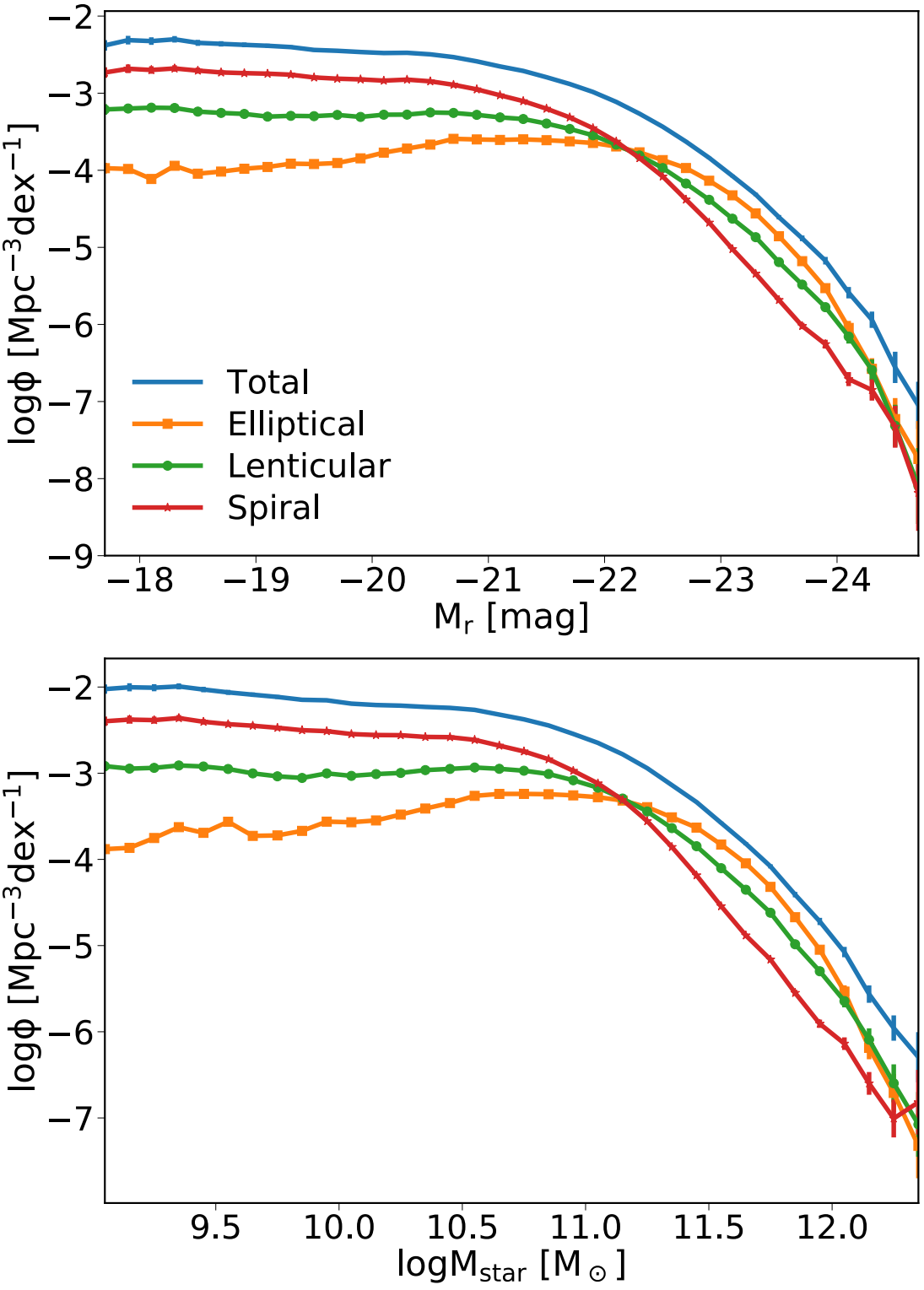



FIGURE 1.8: The observed $z = 0$ luminosity (blue, top) and corresponding observed $z = 0$ stellar mass (blue, bottom) functions, separated into their morphological calcification (as labelled). Adapted from Bernardi et al. (2013).

galaxy's spectrum is dominated by the young, hot stars which are tracers of active star formation, meaning galaxies in the blue cloud are actively star forming whereas those

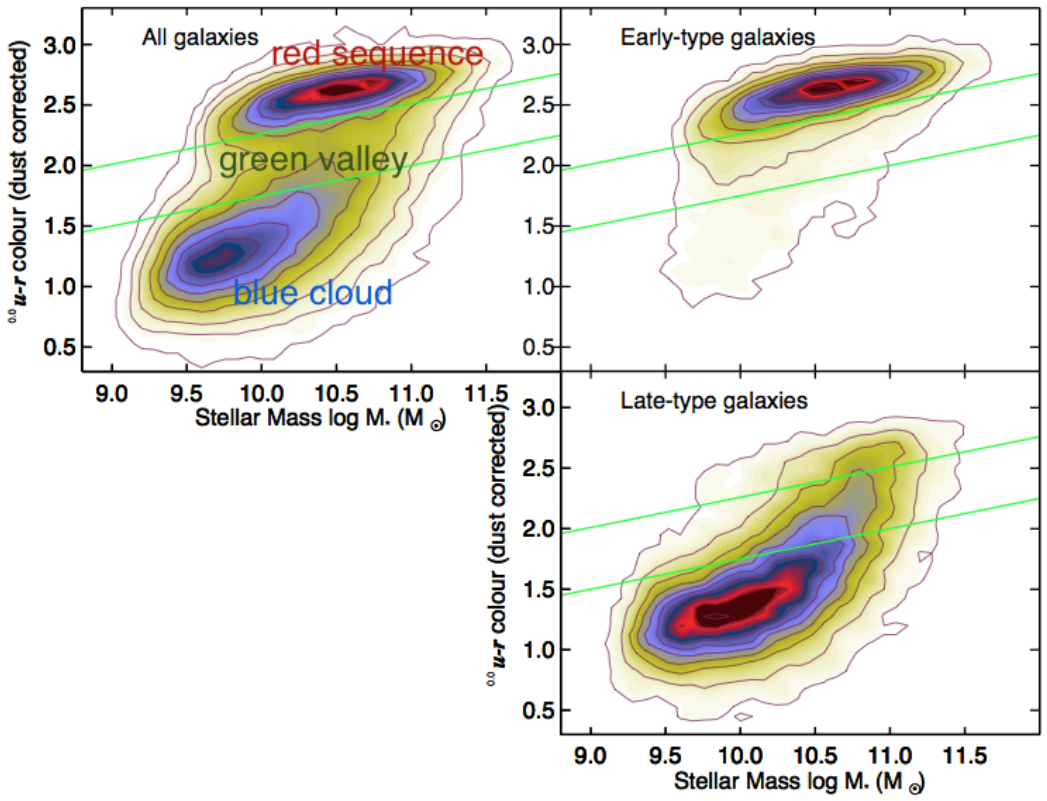



FIGURE 1.9: The $u-r$ colourmass diagram of local galaxies. The left panel shows all galaxies. The right panels show galaxies separated into early-type (top) and late-type (bottom). Extracted from Schawinski et al. (2014).

in the red sequence are not. The lack of galaxies in the green valley suggests that this transition should be fast, a process refereed to as "quenching".

The observation that more massive galaxies tend to be red gave rise to the idea of "mass quenching' whereby a galaxy is quenched when it reaches a mass threshold. This could be due to the energy and/or momentum input from the central supermassive black hole (Granato et al., 2004; Shankar et al., 2006; Hopkins et al., 2008). It could also be because the galaxy's host dark matter halo, when more massive than given threshold, $\log(M_{\rm halo}) \sim 12$, prevents new cold gas from reaching the central galaxy, a mechanism known as "halo quenching" (Dekel & Birnboim, 2006; Dekel et al., 2009). Physically, above this critical halo mass, cold gas is shock heated at the virial radius to the virial temperature ($\sim 10^6 k$), which dramatically decreases the star formation rate of the central galaxy.

Additionally, It was noted by, e.g. Hubble & Humason (1931) that the mix of morphologies and colours of galaxies in a given region of space depends on their density (Kauffmann et al., 2004; Baldry et al., 2006; Weinmann et al., 2006). The densest environments with the most massive dark matter haloes (known as galaxy clusters) host

a larger fraction of quenched, early-type satellite galaxies. This is labelled as "environmental quenching" (Peng et al., 2010; Ilbert et al., 2013; Pontzen et al., 2017). The physical mechanisms behind environment quenching are still debated but could be ram pressure stripping, strangulation, harassment to name a few (for a full review, see Boselli & Gavazzi, 2006).

#### 1.4.3.4 The Size-mass relation

The last key statistical property of galaxies I discuss is the relationship between a galaxy's stellar mass and its radius. A galaxies size is quantified by its half-light radius (the circle containing half the galaxy's projected light) or as the half-mass radius (the sphere containing half the galaxy's total stellar mass). Figure 1.10 shows the median half-light-stellar mass relation for pressure supported (ellipticals; squares) and rotationally supported (spirals and lenticulars; triangles) galaxies. It is clear from this figure that the size of a galaxy is correlated with its stellar mass. Additionally, the relationship is different for different galaxy morphologies, with ellipticals having a much steeper size-mass relation.

More generally, elliptical galaxies are observed to follow a tight (typically only a $15-20\%$ scatter) "fundamental plane" in the size-stellar mass-velocity dispersion phase space, which is of the form:

$$\log R_e = a \log \sigma_0 + b \log \langle I \rangle_e + constant\,, \tag{1.8}$$

where $R_e$ is the effective radius, $\sigma_0$ is the central velocity dispersion and $\langle I \rangle_e$ is the mean surface brightness within $R_e$. The origin of the fundamental plane can be traced back to the virial theorem:

$$\frac{GM}{\langle R \rangle} = \left\langle v^2 \right\rangle\,, \tag{1.9}$$

where $M$ is the mass of the system and $\langle R \rangle$ is the average radius. The left hand side of equation 1.9 defines the potential energy per unit mass, and the right hand side, $\left\langle v^2 \right\rangle$, is the mean squared velocity, i.e. twice the kinetic energy per unit mass. Equation 1.9 would lead to the constants in Equation 1.8 having values of $a = 2$ and $b = -1$. However, typically the best fit values range from $a \sim 1.2$ for blue bands to $a \sim 1.5$ for IR bands, and $b \sim -0.8$ (Jorgensen et al., 1996; Pahre et al., 1998; Colless et al., 2001; Bernardi et al., 2003). This departure from the virial theorem is known as the "tilt" of the fundamental plane. The tilt is usually associated to non-homology, dark matter contributions, and/or variations in $M_\star/L_\star$ ratios.

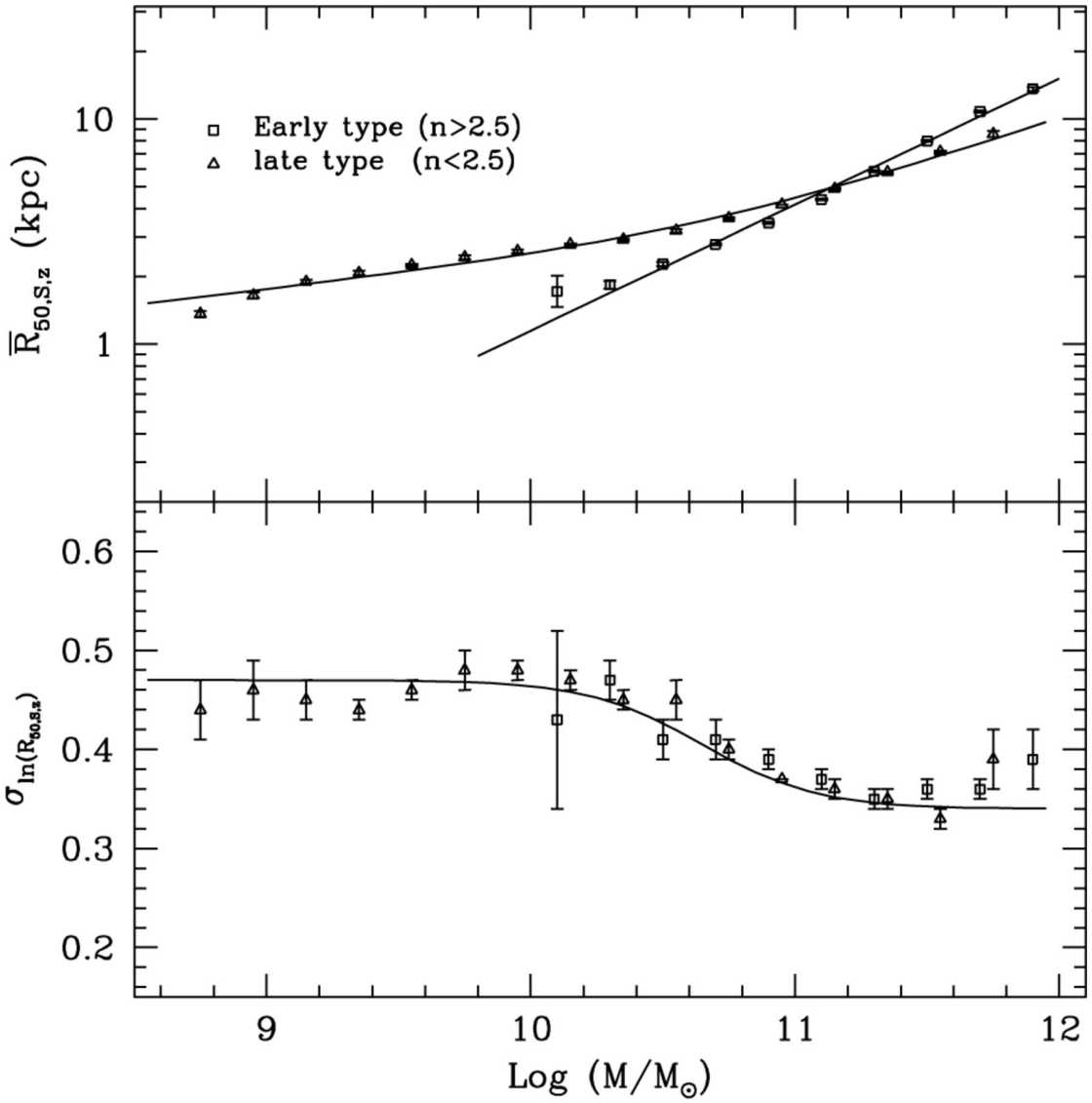



FIGURE 1.10: The median and dispersion of the distribution of Seŕsic half-light radius, $R$, in the z band as a function of stellar mass extracted from (Shen et al., 2003). Triangles represent results for late-type galaxies, while the squares are for early-type galaxies. The error bars represent the scatter among 20 bootstrap samples.

## 1.5 Observations of high redshift galaxies

The results of fitting synthetic stellar populations to the spectra of massive early-type galaxies show that their stars are, on average $\gtrsim 8Gyr$ old, corresponding to formation epochs $z_{\rm form}\gtrsim 2$ (Thomas et al., 2005). Indeed, the epoch $2 \lesssim z \lesssim 4$ corresponds to the peak of the star formation density (see Figure 1.11) when the Universe was most active, therefore, catching the formation of massive galaxies "in the act" requires observations at these early times (Conselice, 2014).

Deep optical and infra-red surveys (e.g. CANDELS; Grogin et al. 2011; Koekemoer et al. 2011 and UltraVista McCracken et al. 2012) have revealed that galaxies at this time are

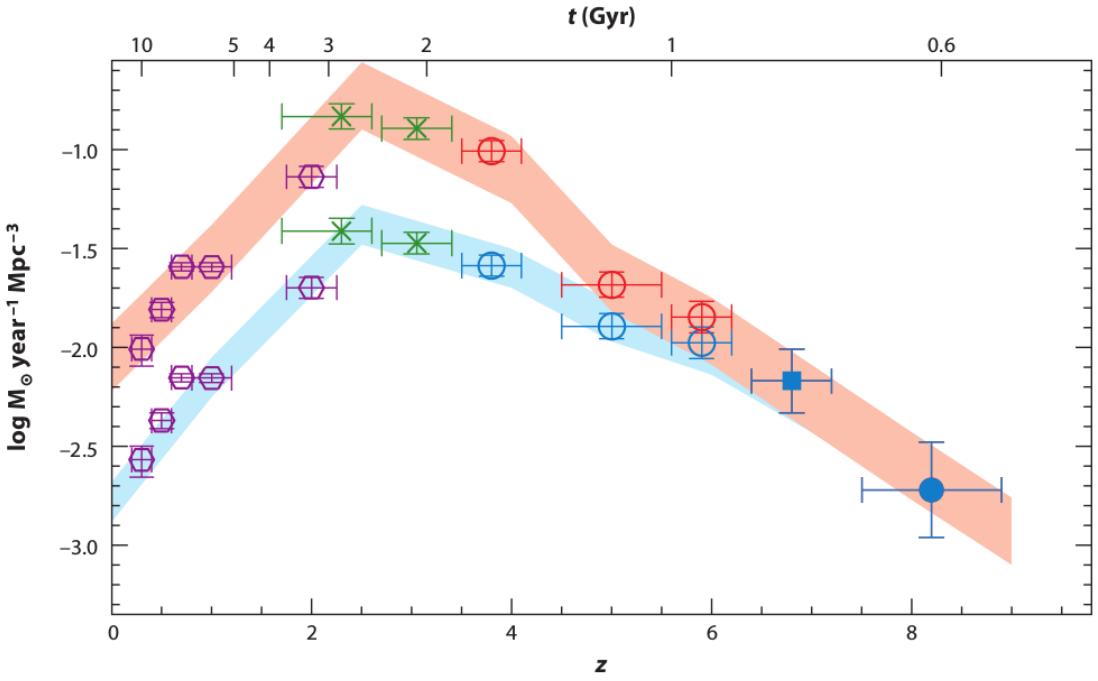



FIGURE 1.11: Evolution of the star formation rate density based on rest-frame UV luminosity functions (extracted from Bouwens et al., 2010). The lower, blue shaded region shows the star formation rate density with no dust correction. Instead, the top, red shaded region shows the dust corrected rate.

distinctly different to local ones. Galaxies at these early epochs do not fit into the standard Hubble classification scheme used for local galaxies of discs and ellipticals. Instead, the appearance of high redshift galaxies ranges from extremely compact spheroids to clumpy, diffuse discs (Huertas-Company et al., 2015; Kartaltepe et al., 2015).

### 1.5.1 Detecting $z > 1$ galaxies

With the exception of a few deep spectroscopic surveys (e.g. VIRMOS; Le Fèvre et al., 2005), a handful of novel methods are utilised to detect high redshift galaxies. Although effective at isolating high-redshift galaxies, all of these selection methods suffer from incompleteness which is very difficult to estimate. For example, carefully tuned colour cuts can isolate, e.g., the Lyman-$\alpha$ break (these are known as Lyman-break galaxies; Steidel et al. 1996, 2003), the Balmer break (Franx et al., 2003; van Dokkum et al., 2003), or combinations thereof (Daddi et al., 2004).

The increase in star-formation leads to massive galaxies having a high rest frame UV luminosity. However, they are generally heavily dust enshrouded and so their UV light is absorbed and re-emitted in the IR. Consequently, massive, high redshift, star forming galaxies are observed to be bright in the sub-mm band and so can be efficiently detected by, e.g., SCUBA or MIPS (known as sub-mm galaxies; SMG; Smail et al. 1997).

### 1.5.2 Redshift evolution

In this section, I compare the physical properties of galaxies, observed at different epochs.

#### 1.5.2.1 Morphology

Galaxies at high redshift appear very different to local galaxies. Figure 1.12 shows the types of morphologies observed at $1 < z < 3$, extracted from Mortlock et al. (2013). Figure 1.13 shows the fraction of each galaxy morphology at $0 < z < 3$ at fixed number density, extracted from Huertas-Company et al. (2015). In general, morphologies at high redshift are more disturbed, asymmetrical and clumpy than local galaxies (Lotz et al., 2006; Ravindranath et al., 2006). Additionally, the fraction of massive discs is much higher (Huertas-Company et al., 2015; Mortlock et al., 2013). Also, star-forming galaxies are observed to have strong outflows (Pettini et al., 2001; Shapley et al., 2003; Steidel et al., 2010).

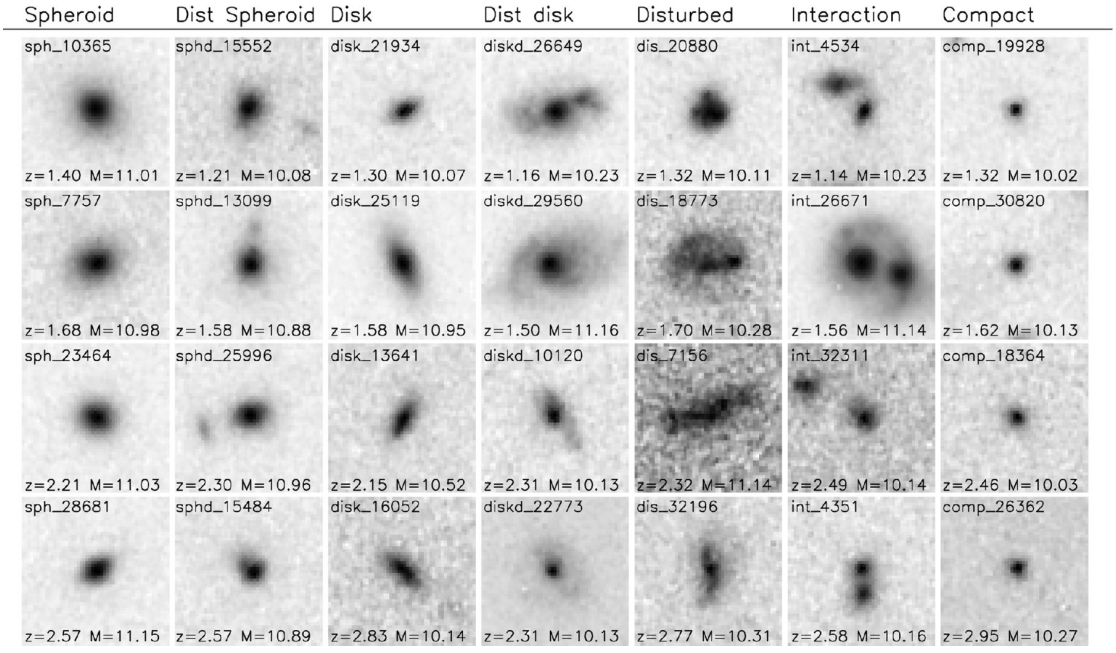



FIGURE 1.12: Examples of CANDELS $H_{160}$-band images of galaxies with different morphologies at $1 < z < 3$. Extracted from Mortlock et al. (2013)

#### 1.5.2.2 Stellar mass

The stellar mass function has been measured up to $z \gtrsim 5$ (e.g. Fontana et al., 2004, 2006; Drory et al., 2005; Pozzetti et al., 2007; Elsner et al., 2008). Figure 1.14 shows the integrated stellar mass in galaxies above $\log M_{\rm star} > 8$ $[M_{\odot}]$ per unit volume, taken from (Marchesini et al., 2009, and references therein). This figure shows that only 50% of the

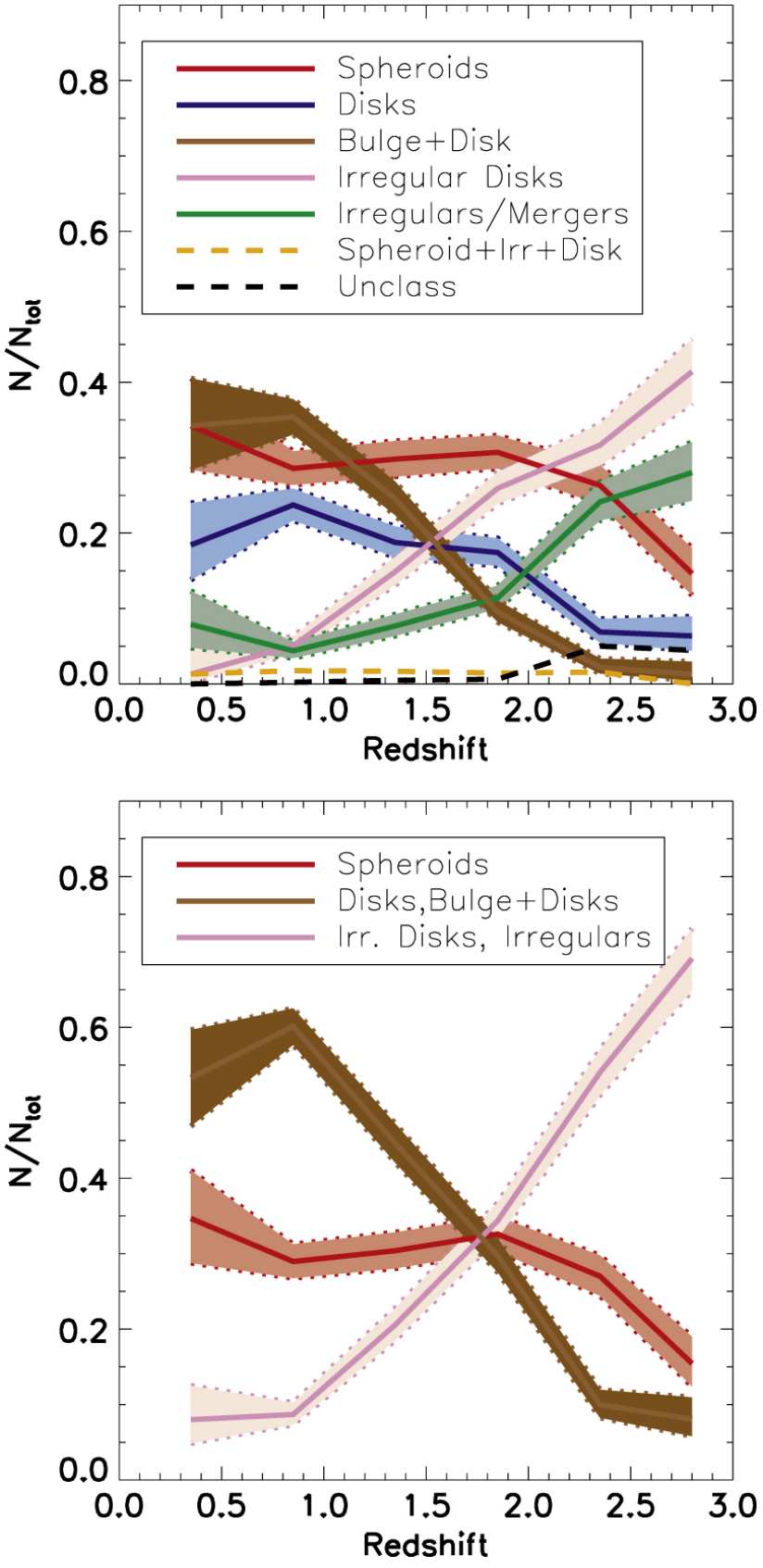



FIGURE 1.13: The evolution of the relative abundances of different galaxy morphologies between $z \sim 0$ and $z \sim 3$; extracted from Huertas-Company et al. (2015).

stellar mass in the Universe was formed $z \gtrsim 1$ and only 10% was formed $z \gtrsim 2.5$. While there is clear evidence that there is substantial growth in the stellar mass density of the Universe, it is still debated how the new stars are distributed between existing and newly forming galaxies as well as between the different morphologies and stellar mass regimes.

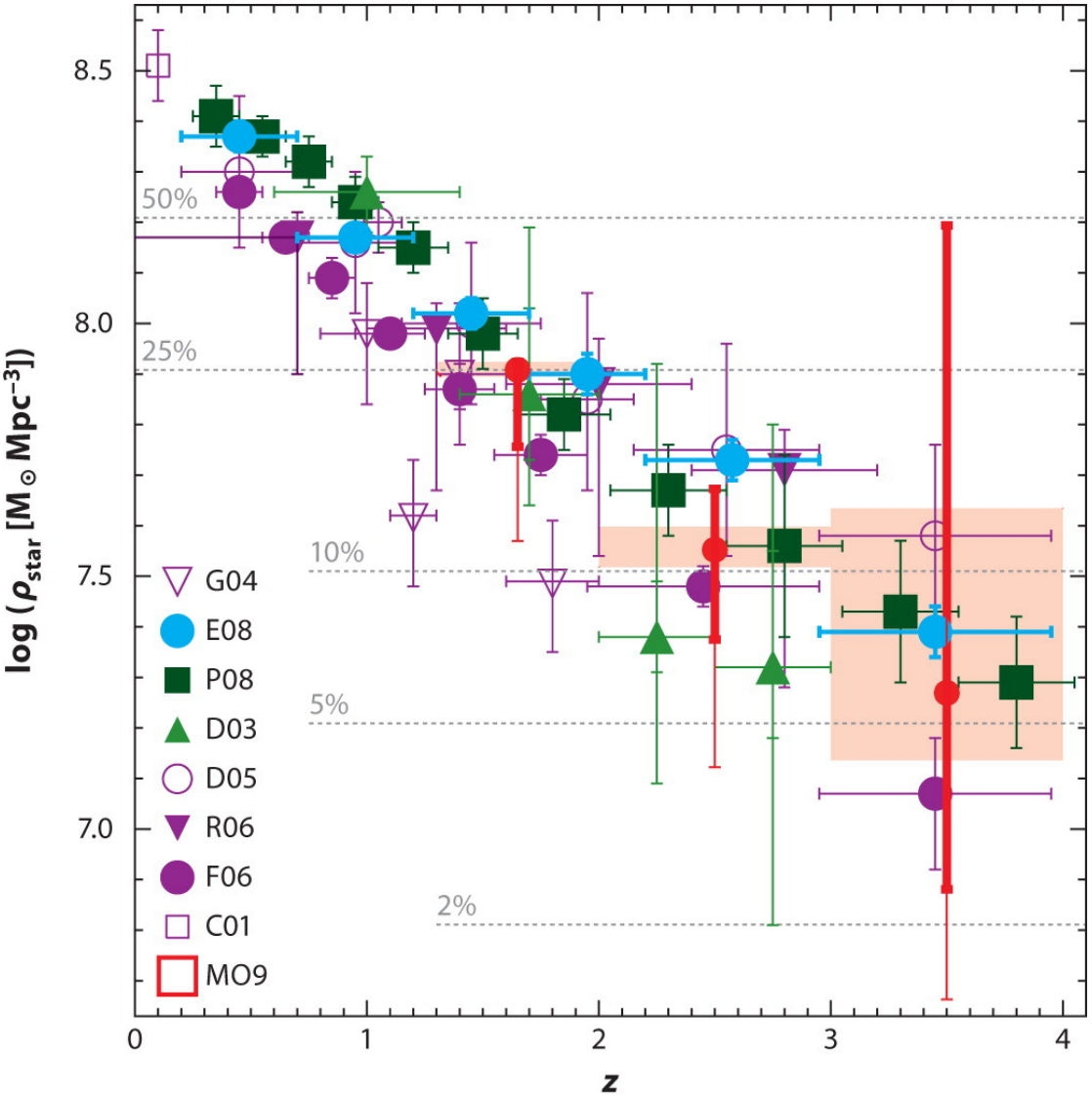



FIGURE 1.14: The redshift evolution of the stellar mass density. This is found by integrating the stellar mass function from $10^8$ to $10^{13}$ $[M_\odot]$ at each redshift. Extracted from Marchesini et al. (2009)

#### 1.5.2.3 Effective radius

As noted by several groups (e.g. Daddi et al., 2005; Toft et al., 2007; Trujillo et al., 2007; Buitrago et al., 2008; Cimatti et al., 2008; van Dokkum et al., 2008; Damjanov et al., 2009), $z \sim 2$ quiescent galaxies have effective radii that are $\sim 3-5$ times smaller than

those of local early-type galaxies of the same stellar mass. Figure 1.15 shows the effective radii of disc and spiral galaxies and their stellar mass at $0.25 \leq z \leq 2.75$; extracted from van der Wel et al. (2014). Of important note is that the slope and $1\sigma$ scatter of this size-mass relation remains constant with redshift, and the mean relation only changes in normalisation. Furthermore, the change in normalisation may be different for spirals and ellipticals. At fixed stellar mass, the effective radius of spirals varies as $R_e \propto (1+z)^{-0.75}$, whereas the redshift dependants is more strong in ellipticals; $R_e \propto (1+z)^{-1.5}$ (van der Wel et al., 2014).

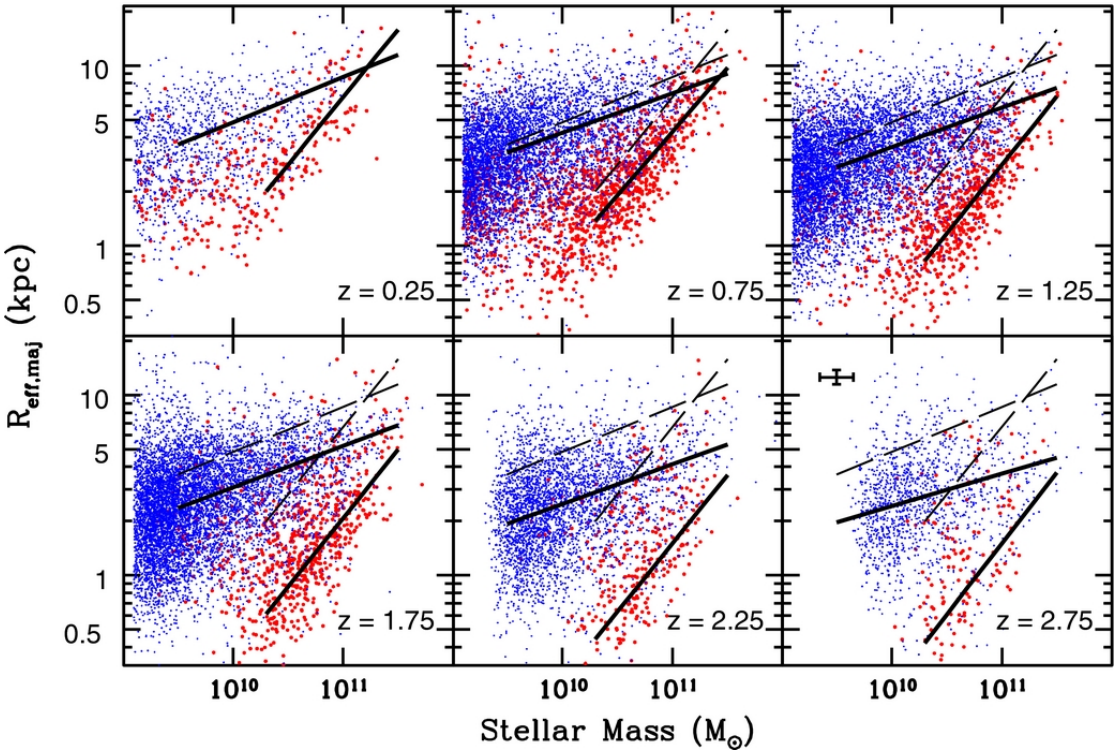



FIGURE 1.15: The sizestellar mass distribution of late-type (blue) and early-type (red) galaxies for six redshift bins, extracted from van der Wel et al. (2014). The solid lines in each panel are the best-fit $M_{\rm star}$-$M_{\rm halo}$ relation at that redshift. Instead the dashed line is the best-fit at $0 < z < 0.5$.

## 1.6 Galaxy formation scenarios

In the previous section, I described some key observational probes of galaxy evolution. I now discuss the key question of how they form and evolve.

### 1.6.1 Formation of spirals

The long standing theory to explain the formation of spiral galaxies is that gas is accreted from the cosmic web into the centres of host dark matter haloes. The specific angular

momentum of the gas is conserved thus forming a disc (Fall & Efstathiou, 1980; Mo et al., 1998).

The formation of gas clumps in high redshift disc galaxies can be explained in terms of the Toomre (1964) $Q$ parameter;

$$Q \equiv \frac{c_s \Omega}{\pi G \Sigma}\,, \tag{1.10}$$

Where $c_s$ is the sound speed of the gas, $\Omega$ is the angular velocity and $\Sigma$ is the local surface density. If $Q < 1$, then local gravitational collapse can overcome the sheering disruption of the disc. Clumps are less prominent in the local universe because discs are more settled (Dekel et al., 2009), possibly because of a decrease in the cosmic gas accretion rate (Genel et al., 2012).

### 1.6.2 Formation of giant ellipticals

Unlike Disc galaxies, there is no singular hypothesis on how giant ellipticals form. Instead, there are broadly two, almost polar opposite ideas which I describe in the following sections.

#### 1.6.2.1 Formation of giant ellipticals: monolithic collapse

The monolithic collapse model claims that massive galaxies formed and assembled most of their final stellar mass in strong bursts of star formation at high redshifts (Partridge & Peebles, 1967; Larson, 1975). These starbursts can have star formation rates as high as several thousands of solar masses per year (Chapman et al., 2005). After the starburst has quenched, possibly induced by an efficient AGN feedback, the galaxy is assumed to evolve almost passively until the present day (Granato et al., 2004, 2006; Carollo et al., 2013; Zolotov et al., 2015). Massive ellipticals have remarkably homogeneous and have uniform, old stellar populations. The morphology and size of the galaxy depends critically on the redshift at which all the stars formed (Granato et al., 2004; Cirasuolo et al., 2005). In addition, they are observed to be enhanced in alpha-elements relative to their iron content which is evidence for short bursts of intense star formation (Thomas et al., 2005; Pipino et al., 2009; Conroy et al., 2014; Citro et al., 2016).

In principle, one might expect the collapse of this gas to form a disc (like lower mass galaxies). However, early studies showed that it is possible for the gas to transfer angular momentum to the halo, naturally creating pressure supported spheroids (Katz & Gunn, 1991; Katz, 1992).

In this framework, there would be little room for any structural evolution along cosmic time which is inconstant with observations (see Section 1.5.2.3). Fan et al. (2008) proposed a method for expanding galaxies at $z < z_{\rm form}$ in which strong winds/jets from AGN can possibly expel large quantities of baryons from the galaxy. As a reaction to drop in the central gravitational potential well, the galaxy will expand quasi-adiabatically (Ragone-Figueroa & Granato, 2011). Fan et al. (2010) showed that in cosmological context, including all galaxies formed at different epochs, quasi-adiabatic expansion may substantially contribute to the observed size increase in massive elliptical galaxies.

#### 1.6.2.2 Formation of giant ellipticals: Mergers

Hierarchical merger models instead predict that massive galaxies have assembled most of their final stellar mass via a sequence of mergers following their host dark matter haloes (e.g., Toomre, 1977; Naab et al., 2009; Shankar & Bernardi, 2009; van Dokkum et al., 2010; Guo et al., 2011; Shankar et al., 2013; Montes et al., 2014). There is no doubt that massive galaxies must have merged at some point as tidal tails and concentric shells are observed around massive, local galaxies (Duc et al., 2015). However, the two pressing questions are: 1) Can mergers between real galaxies actually create ellipticals like those observed in the local Universe? 2) Is the merger rate and mixture of progenitor properties as a function of environment, integrated over cosmic time, able to reproduce the observed properties of local ellipticals. The first issue has been addressed using N-body simulations of isolated merging galaxies (Gerhard, 1981; Farouki & Shapiro, 1982; Negroponte & White, 1983; Barnes, 1988). More recent simulations have shown that dissipationless (dry) mergers between equal mass galaxies can result in massive ellipticals with properties similar to observed ones (Khochfar & Burkert, 2005; Cox et al., 2006; Naab et al., 2006), whereas more unequal mergers between disc galaxies tend to preserve the disc-like morphology (Naab & Burkert, 2003). The second issue will be addressed over the course of this thesis.

### 1.6.3 Formation of giant ellipticals: Observational challenges

From the observational perspective, it is currently difficult to distinguish between these two modes of massive galaxy formation. As several groups have noted, the perceived evolution in physical properties could simply be due to the large systematics in mass estimates (Marchesini et al., 2009; Kravtsov et al., 2014; Shankar et al., 2014b; Bernardi et al., 2016b, 2017), progenitor bias (van Dokkum & Franx, 1996; Saglia et al., 2010; Newman et al., 2012; Carollo et al., 2013) and environmental effects (Poggianti et al., 2006; Shankar et al., 2013; Delaye et al., 2014; Shankar et al., 2014a; Stringer et al., 2015).

These all particularly effect the extreme high-mass end of the stellar mass function (e.g., Marchesini et al., 2014; Shankar et al., 2014a; Leauthaud et al., 2016; Bernardi et al., 2016b). The left panel in Figure 1.16 shows Hubble space telescope (HST) images of local galaxies. The right panel shows mock HST observations of those same galaxies if they were at $z = 2.5$. It is clear from this figure that it is often difficult to discern the structure of galaxies at high redshifts. In the future, telescopes such as the James Webb space telescope might be able to set direct constrains on massive galaxy evolution.

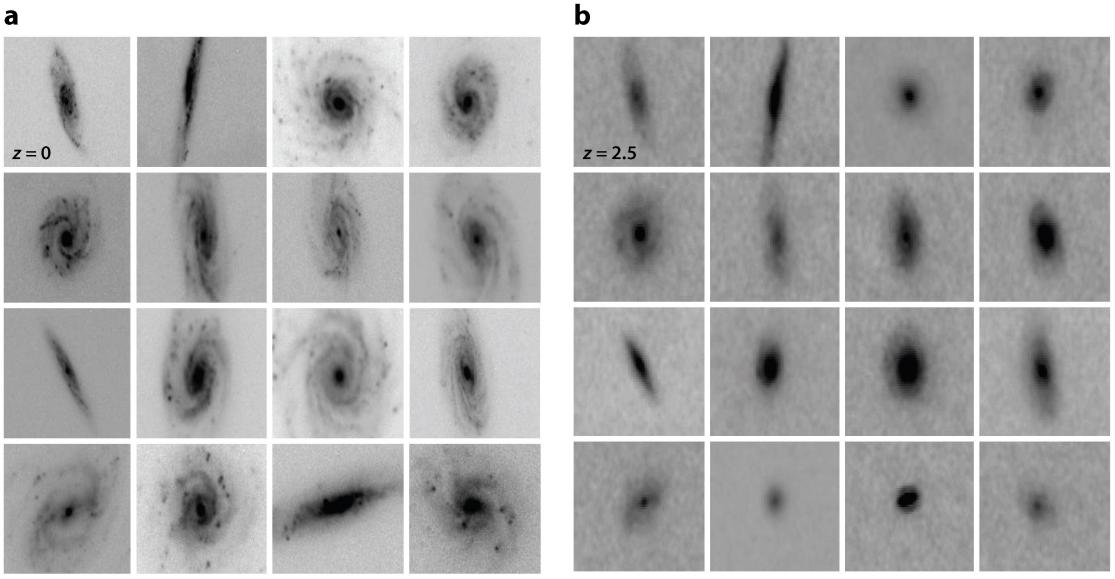



FIGURE 1.16: Nearby galaxies originally observed at $z \sim 0$ in B-band, simulated to how they would appear at $z = 2.5$ in rest-frame B-band. Extracted from Conselice (2014)

### 1.6.4 Formation of giant ellipticals: Modelling techniques

Instead, a theoretical understanding of galaxy evolution can be obtained through detailed physical models. In this section, I discuss the results of semi-analytic, hydrodynamical and semi-empirical models, paying particular attention to their input into the early versus late formation scenarios.

#### 1.6.4.1 Semi-analytic models

Semi-analytic models (SAM hereafter) take the approach of treating various physical processes associated with galaxy formation using approximate, analytic prescriptions (Cole et al., 2000; Baugh, 2006; Guo et al., 2011; Benson, 2012; Lacey et al., 2015). SAMs start with dark matter merger trees, constructed using either N-body simulations or analytic extended-Press-Schechter algorithms (Appendix A gives a detailed comparison of

merger tree algorithms). Gas is allowed to flow and relax into the halo. Heating/cooling prescriptions are added to trace the hot/cold gas components. Star formation follows the Kennicutt-Schmidt law (Kennicutt, 1998) with a free parameter for normalisation. SAMs use similar analytic recipes for black hole growth (Marulli et al., 2008), AGN and supernova feedback (used for reheating cold gas and driving outflows Menci et al., 2006; Guo et al., 2011), mergers (Cole et al., 2000), starbursts (Somerville et al., 2008) and quenching (Cattaneo et al., 2006). Each of these recipes has a series of free parameters which are tuned so the model matches observations of local galaxies.

SAMS are an effective way of probing the diverse physical processes that are believed to drive galaxy formation and evolution. However, a range of SAMs can provide degenerate solutions while still reproducing key statistical properties such as the galaxy stellar mass function (see review discussion in Mo et al., 2010). This is because they have a large numbers of free parameters and a wide array of input assumptions which can lead to a range of interpretations of their outputs. Broadly speaking, SAMs can predict both that mergers can dominate the evolution (De Lucia et al., 2011; Gonzalez et al., 2011; Guo et al., 2011; Shankar et al., 2013; Wilman et al., 2013) and that in-situ processes are more important (Lapi et al., 2011; Ragone-Figueroa & Granato, 2011; Chiosi et al., 2012; Merlin et al., 2012; Posti et al., 2014) depending on their exact parametrisations.

#### 1.6.4.2 Cosmological hydrodynamical simulations

With the rise in computing power, full cosmological hydrodynamical simulations are becoming an increasingly popular way of studying galaxy evolution (e.g. Dubois et al., 2013; Hopkins et al., 2014; Schaye et al., 2015). They aim to directly simulate both the dark matter and baryon components in a cosmological box. High redshift random density perturbations are sampled in both gas and dark matter (either using grids or particles) to give the initial conditions of the simulation. These initial conditions are then allowed to evolve by solving gravitational and hydrodynamical equations. Because of the finite resolution of hydrodynamical simulations, 'sub-grid' physics is included to take into account the intricate physics that is important on scales that cannot be resolved. In general, these are similar to the prescriptions used in SAMs. Hydro simulations have the advantage that they make fewer assumptions about the interactions between the dark matter and gas components. However, they are generally more complex to analyse and more computationally expensive to run.

Modern SAMs and hydrodynamical simulations are capable of fitting a handful of key observations such as the local stellar mass function (Schaye et al., 2015). Both agree on the need for strong AGN and supernova feedback (Shankar et al., 2006). For the most

part, models agree that there is at least some mass and structural evolution in massive galaxies, however, they disagree on the exact cause. For instance, some groups suggest that all the evolution is driven by mergers (Gonzalez et al., 2011) whereas other take the line that mergers are unimportant (Lapi et al., 2011). Disc instabilities (Bournaud et al., 2011) have been put forward to explain the formation of bulges. Hydrodynamical zoom simulations (e.g., Hirschmann et al., 2012) have converged on the idea that there are two phases to massive galaxy evolution where in-situ star formation dominates the early assembly and mergers become more important at lower redshifts (Naab et al., 2009; Oser et al., 2010). Hydrodynamical simulations in a full cosmological box continue to support this two-stage evolutionary patten at least for the most massive galaxies (Hirschmann et al., 2012; Torrey et al., 2015; Welker et al., 2015).

#### 1.6.4.3 Semi-empirical models

In recent years, a third complementary approach has been put forward to more securely probe and constrain the possible evolutionary pathways of massive galaxies. Semi-empirical models aim to explore the fundamentals of galaxy formation by utilising empirical relationships (such as the $M_{\rm star}$-$M_{\rm halo}$, $M_{\rm star}$-$m_{gas}$, $M_{\rm star}$-$r_e$, $M_{\rm star}$-sSFR relations) thus minimising the number of assumptions and free parameters. A successful model has been the 'bath tub' model which treats galaxies as a gas reservoir (e.g. Lilly et al., 2013; Dekel & Mandelker, 2014). van Dokkum et al. (2010), Marchesini et al. (2014) and Huertas-Company et al. (2015) have adopted number conservation techniques to track the putative main progenitors of massive galaxies. Other techniques are based on continuity equation models for the stellar population (e.g., Peng et al., 2010; Aversa et al., 2015). Also, Lidman et al. (2012) and Shankar et al. (2015) followed the main progenitor track of the host haloes to identify potential proto-galaxies as progenitors. All of these semi-empirical approaches broadly agree in assessing the primary role of in-situ growth for galaxies below $M_{\rm star} \lesssim 10^{11}\ M_{\odot}$. However, models become generally more discordant when predicting the evolution of the most massive galaxies. One of the main reasons for such discrepancies can be traced back to the growing significance of the systematics associated with observations such as surface brightness variations, estimates of the proper background, cosmic variance, stellar mass estimates, the number of mergers and the initial mass function (van Dokkum et al., 2010; Marchesini et al., 2009; Behroozi et al., 2013b; Maraston et al., 2013; Bernardi et al., 2014; Shankar et al., 2014a; Aversa et al., 2015; Leauthaud et al., 2016; Bernardi et al., 2016a). In particular, Bernardi et al. (2016b) have recently shown that even when some of the more serious systematic uncertainties are removed by creating a self-consistent observational dataset

of massive galaxies at a range of redshifts, a clear interpretation of how the most massive galaxies evolve is still elusive.

## 1.7 Thesis outline

This thesis aims to set more stringent constraints on the still debated formation and evolution scenarios for the most massive galaxies. To achieve this, I have developed a series of semi-empirical models to answer the following questions:

1. On average, how much do massive galaxies grow in stellar mass since their formation epoch?
2. Can mergers alone account for such growth in stellar mass and in size?
3. Can a basic monolithic collapse scenario account for the observations?

This thesis contains the research I conducted to explore the above questions, and is structured as follows:

- In Chapter 2 I set constraints on the evolution of the most massive galaxies at $z \sim 1$ to $z = 0$ using the relationship between their stellar mass and halo mass, constructed using abundance matching (work presented in Shankar et al. 2014b). This is enriched by the results of a cosmological semi-empirical model I developed based around the abundance matching results (work presented in Shankar et al. 2015).
- In Chapter 3 I present more strong, model independent constraints on the evolution of the most massive galaxies from $z \geq 2$ to $z = 0$ using their abundances and ages (work presented in Buchan & Shankar 2016).
- In Chapter 4 I present a novel, systematic-free method for evaluating the evolution of the most massive galaxies between $z \lesssim 1$ and $z = 0.5$ utilising the clustering properties of central and satellite galaxies (work presented in Buchan et al. 2017).
- In Chapter 5 I compare the above results to numerical simulations. Firstly, I revisit the adiabatic-expansion model presented by Fan et al. (2010) using idealised numerical simulations to evaluate its role in growing galaxies in size. Secondly, I utilise a sub-sample of the cosmological zoom-in simulations run by Martizzi et al. (2014), as well as my own, to study the growth of a massive galaxy.
- In Chapter 6 I summarise this thesis and conclude.

# Chapter 2

# Probing the evolution of massive galaxies using their abundances and validation through a semi-empirical model.

## 2.1 Introduction

The size and mass evolution of massive, spheroidal galaxies has become one of the hottest topics in modern cosmology. Local early-type galaxies are observed to obey a tight half-light radius-stellar mass relation with an intrinsic scatter of less than 0.3 dex (Bernardi et al., 2011b,a; Nair et al., 2011). This observational feature still represents a major challenge for hierarchical models of galaxy formation that evolve proto-galaxies via a sequence of stochastic mergers, often resulting in the incorrect scaling relation and/or much greater scatter (Bower et al., 2006; Nipoti et al., 2009; Shankar et al., 2010a, 2013). While there are promising ideas on how massive galaxies assembled to give tight scaling relations, a comprehensive evolutionary model is still missing.

From the observational side, a significant fraction of studies are conducted by comparing galaxies at different redshifts at fixed stellar mass. This procedure inevitably includes contributions from both pre-existing galaxies and ones which have evolved to enter the selection criteria at later times. As emphasised by several groups (Hopkins et al., 2009; Carollo et al., 2013), the impact of these newcomers, usually termed "progenitor bias" (van Dokkum & Franx, 1996; Saglia et al., 2010), may account for most of the observed structural evolution. Indeed, there is strong evidence for a significant increase in number

density of massive galaxies at all redshifts (Buitrago et al., 2013; Carollo et al., 2013; Huertas-Company et al., 2013).

The goal of this chapter is to provide initial constraints on the mass and structural evolution of very massive galaxies in ways that are, by construction, free from progenitor bias. Further more, they are largely independent of the complexities of modern cosmological models instead relying (mostly) on direct observations and abundance matching (see Section 2.2.2).

To further validate the above constraints, I also utilise a semi-empirical models of galaxy evolution (eg. Hopkins et al., 2009) which I developed following the methodology of Shankar et al. (2014a). The latter, by design, are an extremely effective way of making use of the known properties of galaxies together with minimal theoretical inputs to make testable predictions on a set of observables, and set unique, independent constraints on the major processes that drive galaxy evolution. These models are fast enough to allow a large volume of galaxies to be quickly simulated, and to test a wide range of input assumptions to with strong statistical significance.

## 2.2 Initial constraints from abundance matching.

I start by discussing work presented in Shankar et al. (2014b) which is an initial step in understanding the evolution of massive galaxies and further motivates the work in this thesis. Shankar et al. (2014b) compares the direct measurements of the halo mass and stellar mass of very massive galaxies to the results of abundance matching at various redshifts. Specifically, Shankar et al. (2014b) utilise a sample of massive galaxies and haloes to test the null hypothesis that the stellar mass of massive galaxies remains constant between z=0 and z=1. Abundance matching (AM hereafter) is a semi-empirical technique for statistically mapping stellar masses onto dark matter halo masses. The null hypothesis that all galaxies do not change in stellar mass will imply that the number densities of massive galaxies also remains constant.

In principle, by tracking the redshift-dependent $M_{\rm star}$-$M_{\rm halo}$ relation, it is possible to extract valuable constraints on the average stellar mass growth of (especially) the most massive galaxies, the focus of this work. In Figure 2.1 I provide a sketch to explain this point. In the top panels of Figure 2.1 I assume an extreme scenario in which the number density of massive galaxies remains constant at all cosmic times. The halo mass function strongly evolves with cosmic time (due to the hierarchical growth of dark matter haloes), thus the mean halo mass at fixed number density substantially increases with decreasing redshift. As a consequence, massive galaxies at high redshifts get mapped

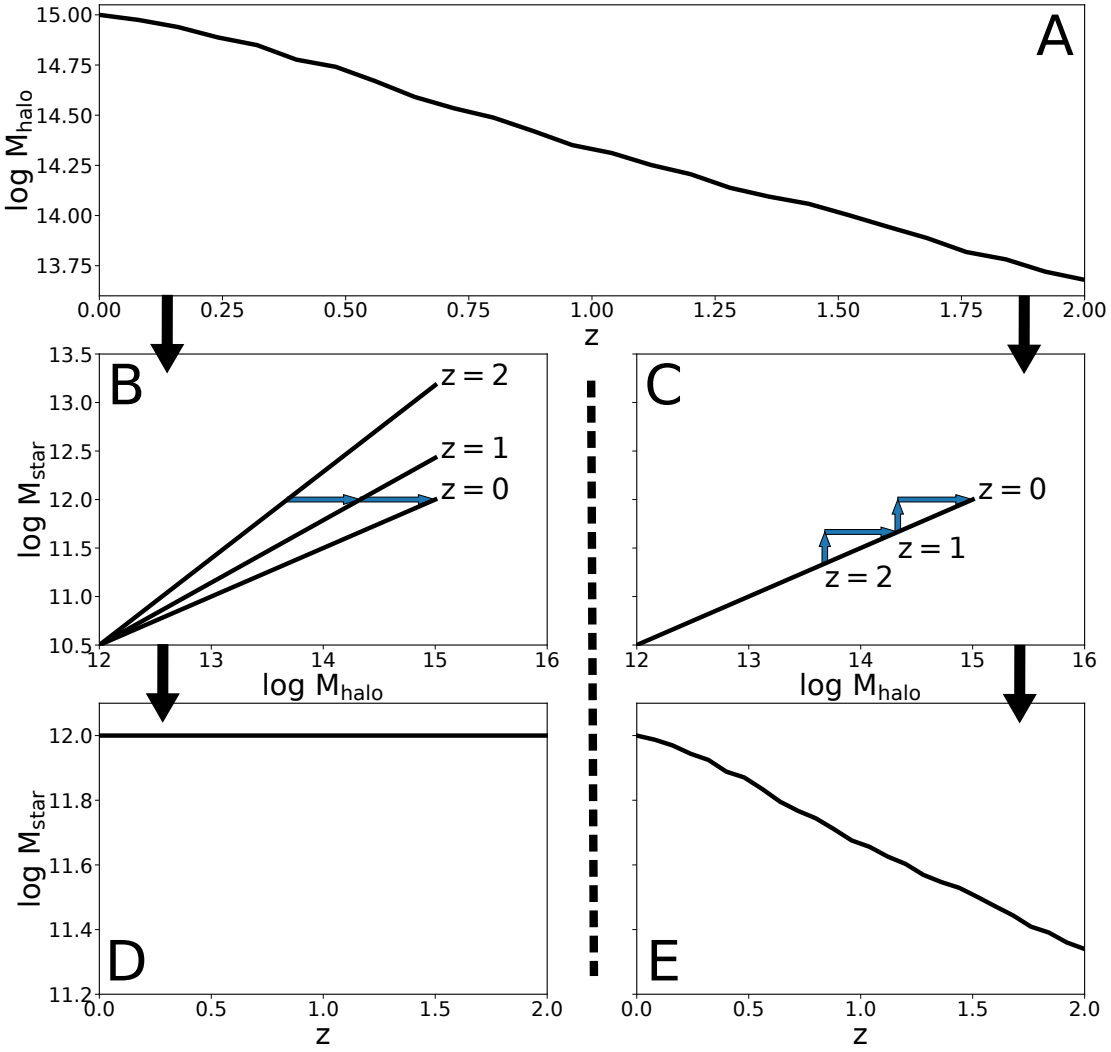



FIGURE 2.1: A cartoon sketching how two contrasting redshift evolutions in the $M_{\rm star}$-$M_{\rm halo}$ relation predicts predict different evolutions in $M_{\rm star}$. Panel A shows the mean evolution of a local log $M_{\rm halo}$= 11[$M_{\odot}$] halo. Panels B and C show a gradually decreasing $M_{\rm star}$-$M_{\rm halo}$ relation and a static one respectively. The blue arrows on the relations show the evolution in $M_{\rm halo}$from panel A along with the corresponding evolutions in $M_{\rm star}$. Panels D and E explicitly show the corresponding evolutions in $M_{\rm star}$.

onto increasingly lower mass haloes since the mean mass of haloes is decreasing, at fixed number density, but the mean stellar mass is remaining constant. This naturally predicts that the high mass slope of the $M_{\rm star}$-$M_{\rm halo}$ relation should increase with higher redshift (top, left panel).

Conversely, in the bottom panels of of Figure 2.1 I allow for substantial mass growth in massive galaxies. In this scenario, the number density of massive galaxies is decreasing with increasing redshift. This would imply that the high-mass end slope of the $M_{\rm star}$-$M_{\rm halo}$ relation should steepen less than the former scenario or even get shallower. In fact if, on average, massive galaxies and their host dark matter haloes are growing in such a way that their respective change in number density at fixed mass is the same,

the high-mass end of the $M_{\rm star}$-$M_{\rm halo}$ relation will remain roughly constant in slope, at fixed intrinsic scatter.

The idea is therefore to compare the $M_{\rm star}$-$M_{\rm halo}$ relationship constructed using a static $z = 0$ stellar mass function and a redshift dependent halo mass function to observed galaxies at a range of redshifts. Comparing simple, semi-empirical models such as this to direct observations has the advantage that the conclusions are not clouded by the complexities and parametrisations of more sophisticated semi-analytic/numerical models. In what follows, I (briefly) present the observations adopted in Shankar et al. (2014b), and then move onto technique of AM. I will then discuss the constraints when combining the direct measurements with AM.

### 2.2.1 Observations

Firstly, Shankar et al. (2014b) constructed an observational dataset of massive, central galaxies at different redshifts. They gather measurements of the host halo masses of massive galaxy clusters and the stellar mass of their respective central galaxies. Here, I list the datasets used:

- Clusters at $z < 0.3$, selected from the Sloan Digital Sky Survey (Eisenstein et al., 2011; SDSS Collaboration et al., 2016, SDSS hereafter) by Kravtsov et al. (2014) and Gonzalez et al. (2013). These two studies use X-ray measurements of the hot inter-cluster gas to constrain the total mass of the cluster and they obtain estimates of the stellar mass of the central galaxies using SED fitting from multi-wavelength spectral energy distribution (SED hereafter) observations.

- Galaxies in the The Cosmic Evolution Survey (Scoville et al., 2007, COSMOS hereafter) fields at $0.2 < z < 1$ from Huertas-Company et al. (2013) and refs. therein. They measure halo masses through a combination of weak lensing x-ray observations. Galaxy stellar masses through multi-wavelength SED fitting.

- Brightest cluster galaxies (BCGs hereafter) from the Spitzer Adaptation of the Red-Sequence Cluster Survey (Muzzin et al., 2009, SpARCS hereafter) spanning $0.03 < z < 1.63$ from Lidman et al. (2012) and (van der Burg et al., 2013). They use optical/IR measurements to constrain stellar masses and x-ray observations to constrain the total cluster mass.

- Targeted observations of a clusters at $z \sim 1.2 - 1.4$ by Strazzullo et al. (2010); Raichoor et al. (2011); Strazzullo et al. (2010). They measure halo masses through X-ray observations and galaxy properties through IR/optical imaging.

- BCGs in the Cl1604 supercluster (and other structures) from the ORELSE survey (Ascaso et al., 2014). The halo virial masses are measured through the dynamics of the member galaxies an the stellar mass of the central galaxies are measured through SED fitting.

Galaxies and haloes from these catalogues are grouped together into four redshift bins: $z < 3$, $0.3 < z < 0.6$, $0.6 < z < 0.8$ and $0.8 < z < 1.4$ so that they can compared to results of AM at those redshifts.

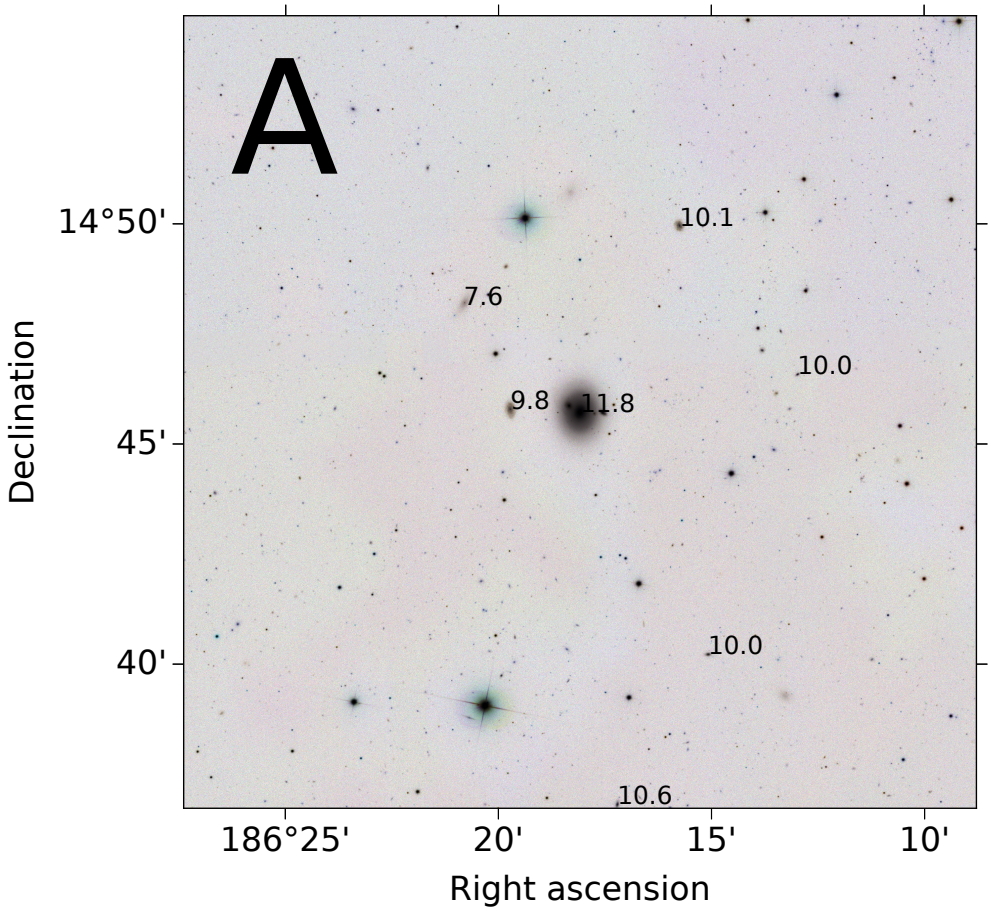
### 2.2.2 Constraints from abundance matching

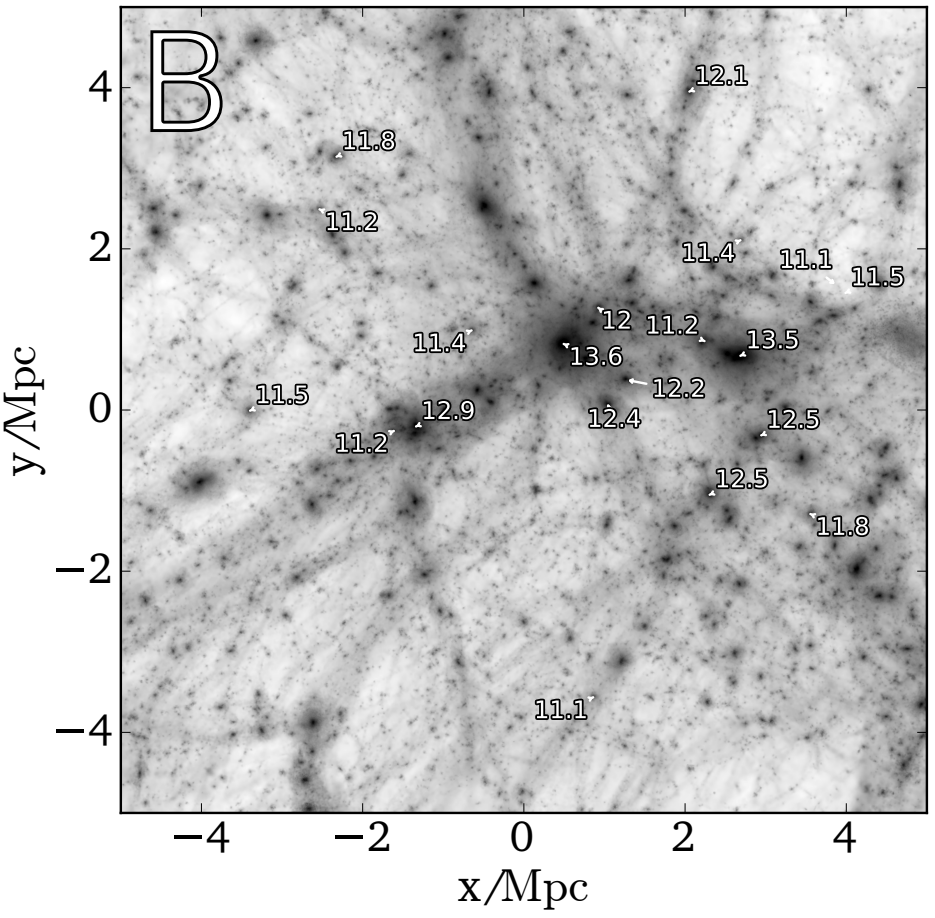
As a second step, Shankar et al. (2014b) compute the mean $M_{\rm star}$-$M_{\rm halo}$ relation from abundance matching, to compare with the data. Figure 2.2 shows a simplified cartoon of the abundance matching (AM) procedure. Panel 'A' shows an image from SDSS centred on a massive galaxy. All the local galaxies in this image that are in the NASA-Sloan Atlas have their stellar masses labelled. Panel 'B' shows a zoom-in simulation at $z = 0$ where a random selection of collapsed dark matter haloes have their halo masses labelled. Panel 'C' (right) then shows the basic AM procedure where galaxies and haloes contained in equal co-moving volumes are ranked by their mass and mapped from one to the other. In other words, the most massive galaxies in a given volume is mapped onto the most massive halo in the same volume and the $100^{th}$ most massive galaxy is mapped onto the $100^{th}$ most massive halo, etc.

Note that most stellar mass functions are the combined number densities of central and satellite galaxies at a given mass. This means that the AM procedure also requites a list of the masses of sub-haloes that contain satellite galaxies.

A major issue with this procedure is that constructing complete, unbiased catalogues of galaxies and haloes at equal volumes is challenging. Figure 2.3 shows a more practical method for AM. The left panel shows the sum of the integrated halo mass function (ie. the number of distinct haloes above a given mass per unit volume) from Tinker et al. (2008) and the integrated unevolved subhalo mass function (same but for the unstripped masses of subhaloes that host satellites) from Giocoli et al. (2008). The right panel shows the integrated stellar mass function (centrals plus satellites) from Bernardi et al. (2013), in the same units. The horizontal lines track constant number densities. The vertical lines track instead the halo/stellar mass these integrated number densities correspond to. Table 2.1 reports the implied mapping between $M_{\rm star}$ and $M_{\rm halo}$ over a wide range of number densities, as shown in Figure 2.3.

This procedure can be summed up with the equation:


A
14°50'
45'
40'
Declination
186°25'
20'
15'
10'
Right ascension
10.1
7.6
10.0
9.8
11.8
10.0
10.6


B
12.1
11.8
11.2
11.4
11.1
11.5
12
11.2
13.5
11.4
13.6
12.2
11.5
12.9
12.4
12.5
11.2
12.5
11.8
11.1
4
2
0
−2
−4
y/Mpc
−4
−2
0
2
4
x/Mpc

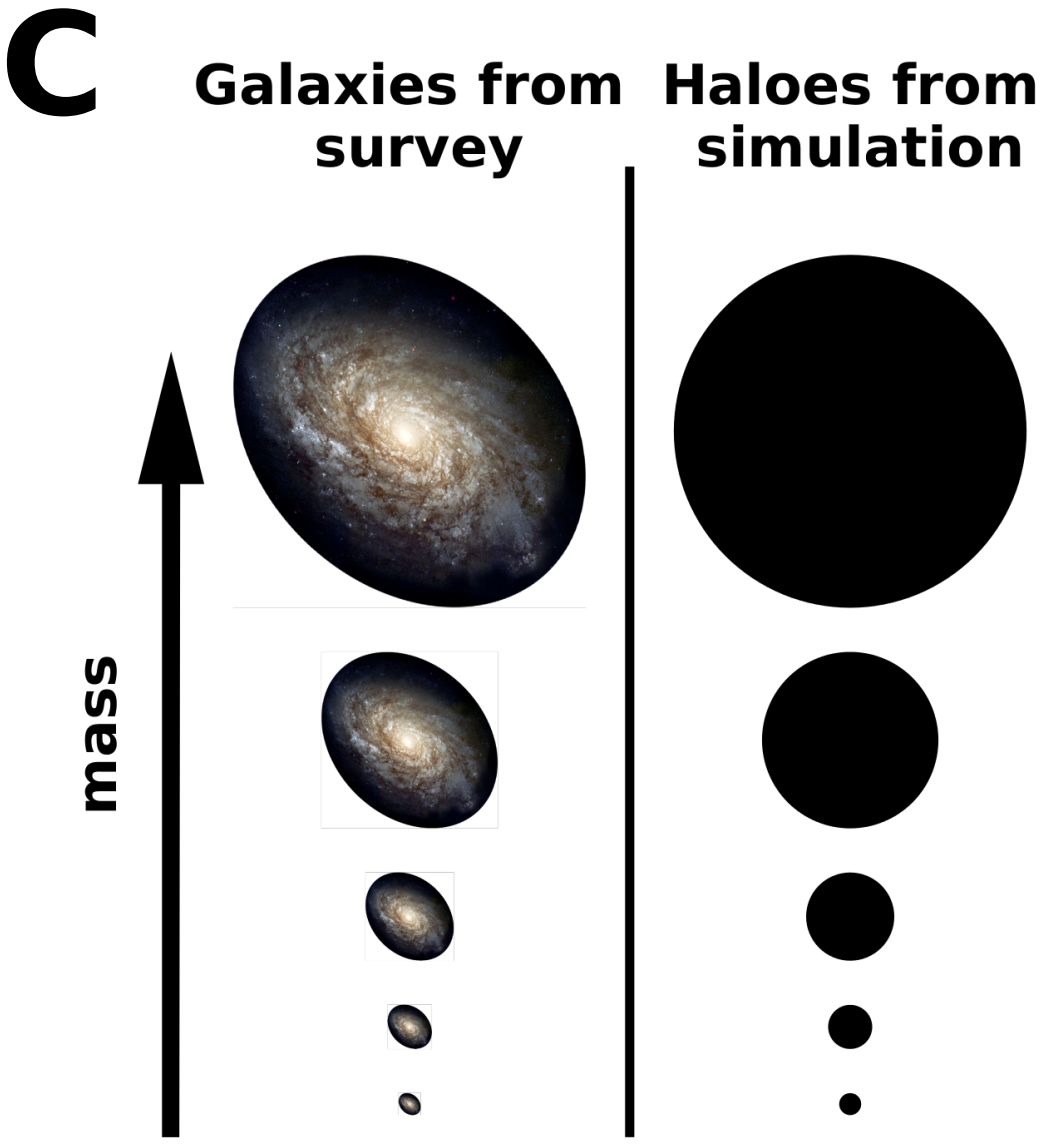



FIGURE 2.2: A cartoon of abundance matching. Panel 'A' shows an image from SDSS centred on a massive galaxy. All the local galaxies within this picture have their log $M_{\rm star}$ labelled. Instead, panel 'B' is a picture of a slice through a zoom-in cosmological simulation. A random sample of haloes have their log $M_{\rm star}$ labelled. Panel 'C' then shows a schematic of abundance matching, where samples of galaxies and haloes are selected from equal volumes, ranked by mass and mapped.

TABLE 2.1: The mapping between stellar mass had halo mass corresponding to the example shown in Figure 2.3.

| Number $(>M)/Mpc^3h^{-3}$ | $10^{-2}$ | $10^{-4}$ | $10^{-6}$ | $10^{-8}$ |
|---|---|---|---|---|
| $M_{\rm halo}[M_\odot\ h^{-1}]$ | $4\times10^{11}$ | $2.2\times10^{13}$ | $2.9\times10^{14}$ | $1.1\times10^{15}$ |
| $M_{\rm star}[M_\odot\ h^{-1}]$ | $1.8\times10^{10}$ | $2.9\times10^{11}$ | $1.0\times10^{12}$ | $2.5\times10^{12}$ |

$$\Phi(>M_{star},z)=\Phi(>M_{200c},z)\,, \tag{2.1}$$

Where $\Phi(>M_{star},z)$ is the integrated number density of galaxies above $M_{star}$ at $z$ and $\Phi(>M_{200c},z)$ is the integrated number density host haloes above $M_{200c}$ at $z$. Equation 2.1 yields a monotonically increasing function which can conveniently be fit by a double power-law of the type:

$$\frac{M_{\rm star}}{M_{\rm halo}}=2N\left[\left(\frac{M_{\rm halo}}{M_1}\right)^{-\beta}+\left(\frac{M_{\rm halo}}{M_1}\right)^{\gamma}\right]^{-1}\,, \tag{2.2}$$

where $M_1$ is the overall normalisation, $N$ controls the location of the break, $\beta$ and $\gamma$ are the low and high mass power-law indices (or the slopes of the relation in log-space).

An important caveat the the AM procedure is the intrinsic scatter in the $M_{\rm star}$-$M_{\rm halo}$ relation. This scatter has the effect of mixing up the the one-to-one correlation described above. To outline the effect the scatter will have on the AM results presented above, I will first consider a fixed halo mass function and a fixed median mapping between $M_{\rm star}$ and $M_{\rm halo}$. In this case, the scatter will manifest itself as a scatter in stellar masses of galaxies at fixed halo mass. Figure 2.4 is a cartoon showing the effect that a scatter in stellar mass at fixed halo mass will have on the stellar mass function. The top left panel shows a scatter plot of the stellar masses of fictitious galaxies and their associated halo masses, assuming that there is a monotonically increasing relation between them with no scatter. The resulting stellar mass function is shown in blue in the bottom panels. The top right panel shows the result if the stellar mass of each galaxy is scattered by a random amount following a log-normal distribution. It can be seen that the mean relation is the same. I have chosen the standard deviation of this scatter such that 10% of the galaxies enter a higher mass bin in the stellar mass function and 10% enter a lower one (as shown by the green and orange arrows in the bottom left panel). The comparison between the initial stellar mass function (blue) and the resulting one after the scatter (orange) can be seen in the bottom right panel. It can be seen that the scatter has no effect in the left bins where the mass function is flat. However at the

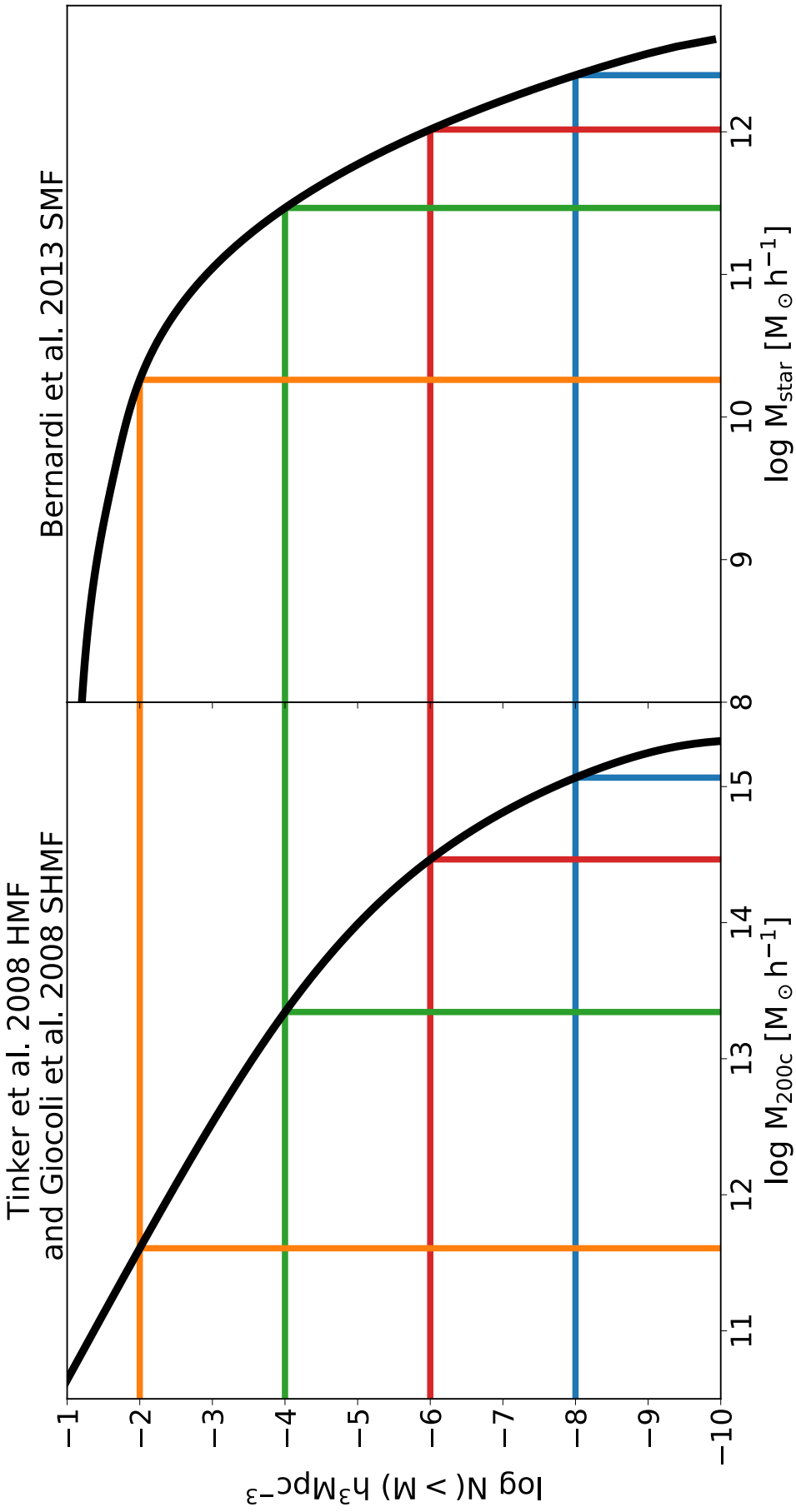



FIGURE 2.3: A more realistic schematic of abundance matching. The left panel show the Tinker et al. (2008) integrated halo mass function. The right panel shows the Bernardi et al. (2013) integrated stellar mass function. The y-axis is the same between the two panels. The horizontal lines show $10^{-2}$, $10^{-4}$, $10^{-6}$ and $10^{-8}$ galaxies/haloes $/Mpc^3h^{-3}$. The vertical lines then show the halo/stellar mass these number densities correspond to.

higher mass bins, the mass function is flattened out since more galaxies are scattered up than are scattered down. This effect is known as Eddington bias.

In practice, there is an intrinsic, non-zero but **unknown** scatter in the observed relation between the masses of galaxies and their host halo. Figure 2.5 shows a real case where the parameters in equation 2.2 are fit using the AB technique with the Tinker et al. (2008) halo mass function, Giocoli et al. (2008) subhalo mass function and the Bernardi et al. (2013) stellar mass function assuming the intrinsic scatter is 0, 0.15 and $0.3dex$ (blue, orange and green respectively). As it can be seen in the bottom panel, all three scatters can provide good fits to the stellar mass function. However, as indicated in the top penal, increasing the scatter leads to a shallower high-mass slope in the mean $M_{\rm star}$-$M_{\rm halo}$ relation. However, the other parameters in the $M_{\rm star}$-$M_{\rm halo}$ relation are roughly unchanged. Thus, an efficient (but not unique) way to include scatter in AM is to first fit Equation 2.2 first with zero scatter then "tilt" the best-fit $\delta = 1 - \gamma$ until the input stellar mass function (are other statistical properties) are reproduced. I will (mostly) follow this method throughout this thesis.

In Shankar et al. (2014b), AM is done using the Tinker et al. (2008) halo mass function which is well calibrated ($< 5\%$ systematic uncertainty) over a large range of halo masses ($10^{11}h^{-1}M_{\odot} < M < 10^{15}h^{-1}M_{\odot}$). Additionally, a variety of stellar mass functions are adopted from the literature to verify that the key results are not induced by the systematics from stellar mass/volume estimates. Figure 2.6 shows four panels of galaxy stellar mass versus its host halo mass for each of the redshift bins. The observed massive galaxies described in Section 2.2.1 are plotted in their respective panels. Additionally, the mappings from $M_{\rm star}$ to $M_{\rm halo}$ derived from AM are shown.

It is clear from Figure 2.6 that all $M_{\rm star}$-$M_{\rm halo}$ relations derived by Shankar et al. (2014b) (Yang et al., 2011; Bernardi et al., 2013; Maraston et al., 2013; Muzzin et al., 2013) are substantially steeper than those widely adopted in the literature (Behroozi et al., 2013b; Moster et al., 2013) despite all using the AB technique (although using different halo mass and stellar mass functions). Additionally, it is also evident from Figure 2.6 that in the limit where the stellar mass function does not change, AB produces a too-steep slope with respect to the data at $z \sim 1$. This conclusion holds on the assumption that the scatter is small and does not change with redshift and that the data does not suffer from high redshift systematics such as surface brightness dimming. I will show that the former is holds via clustering measurements, which is a technique independent of the AM procedure presented above. The latter cannot be excluded the latter using the methodology presented so far, but I will present a new technique in Chapter 4 to probe the evolution of the $M_{\rm star}$-$M_{\rm halo}$ relation which is free from redshift-dependant observational systematics.

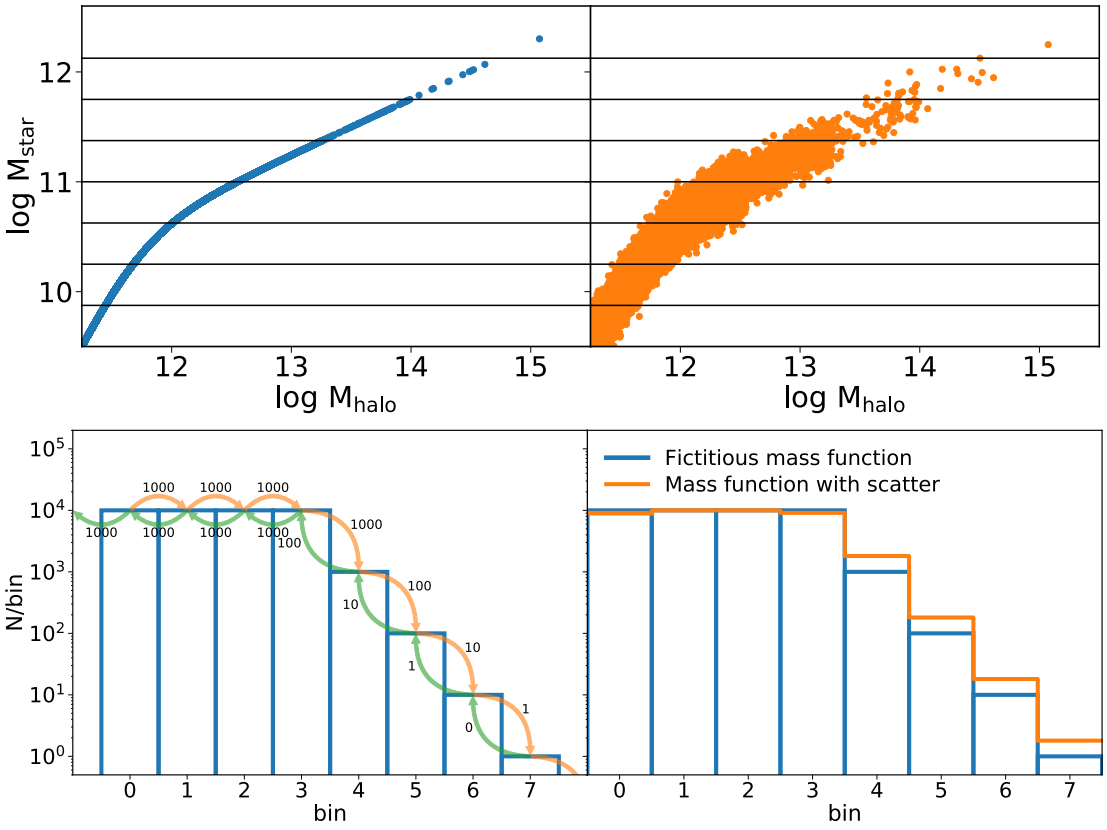



FIGURE 2.4: A cartoon to show the affect of adding scatter to the $M_{\rm star}$-$M_{\rm halo}$ relation. The top left panel shows the $M_{\rm star}$-$M_{\rm halo}$ relation of fictitious galaxies that follow a strict relation with zero scatter. The blue histogram in the bottom panels shows this fictitious mass function. The top right panel shows the relation when the stellar mass each galaxy in the top left panel is scattered by a random amount following a log-normal distribution. The standard deviation has been chosen such that 10% of the objects are scattered into a higher mass bin and 10% being scattered into a lower one (as shown by the arrows in the bottom left panel). The orange line in the bottom right panel shows the resulting mass function after this scattering.

The top right and bottom left panels show galaxies at $0.3 < z < 0.6$ and $z < 0.6 < 0.8$ respectively from Huertas-Company et al. (2013) The open circles show individual galaxies whereas the solid points shows their median.

### 2.2.3 Constraints from galaxy clustering

To constrain the scatter in the $M_{\rm star}$-$M_{\rm halo}$ relation, Shankar et al. (2014b) use the probability distribution $dP/d\log M_{\rm halo}$, independently extracted from the modelled halo occupation distribution (HOD hereafter) of massive galaxies. As input, these models take the two point correlation function (2PCF hereafter), which is the excess probability that two galaxies (from within a sample of galaxies) are found at a distance $r$ apart versus a random distribution. They then give in output the average number of central

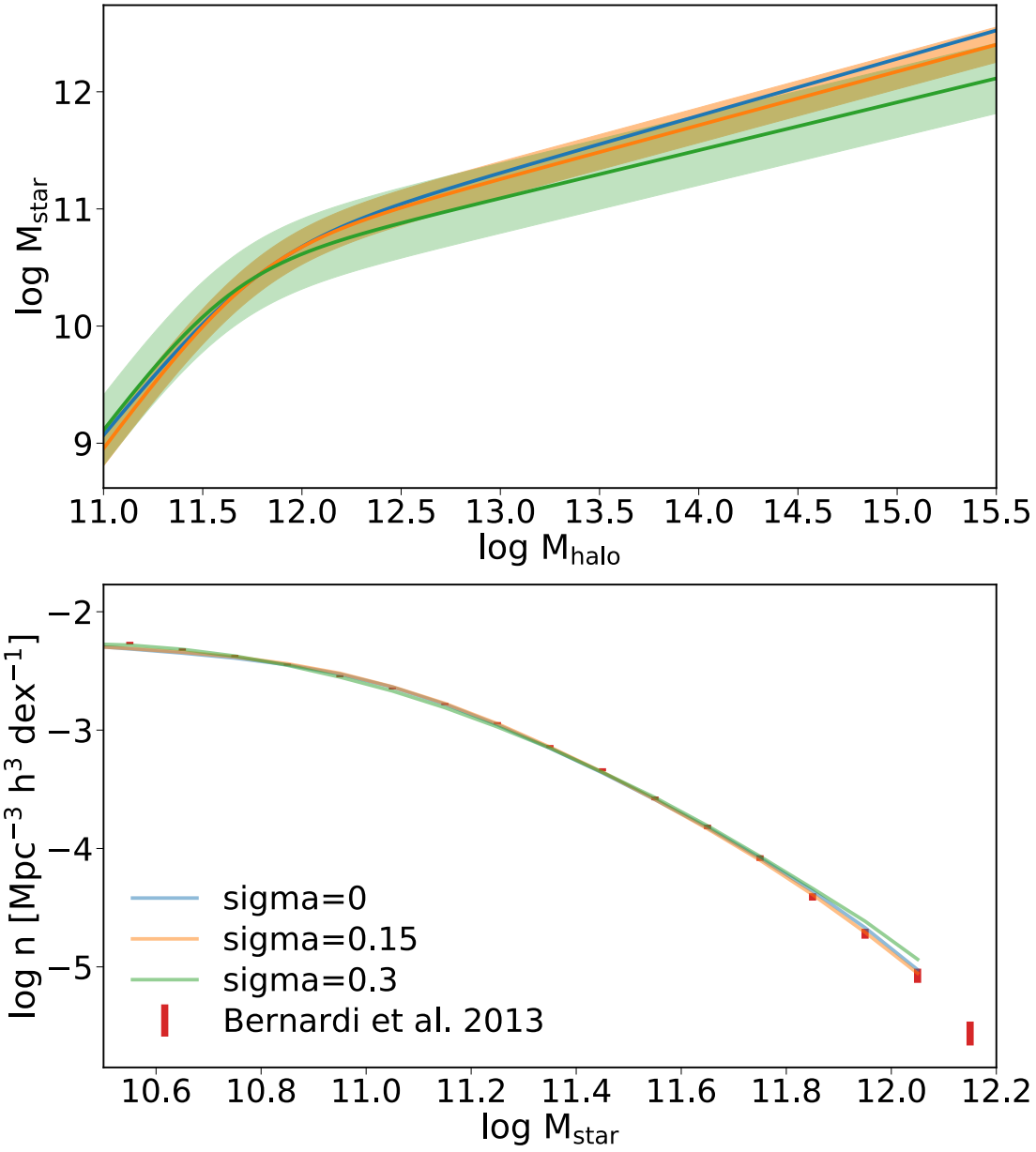



FIGURE 2.5: Top panel: A plot showing the effect that changing the intrinsic scatter in the mean $M_{\rm star}$-$M_{\rm halo}$ relation at fixed halo mass has on the best fit $M_{\rm star}$-$M_{\rm halo}$ relation. Bottom panel: The input observed stellar mass function by (Bernardi et al., 2013) as well as the resulting best fits for each scatter.

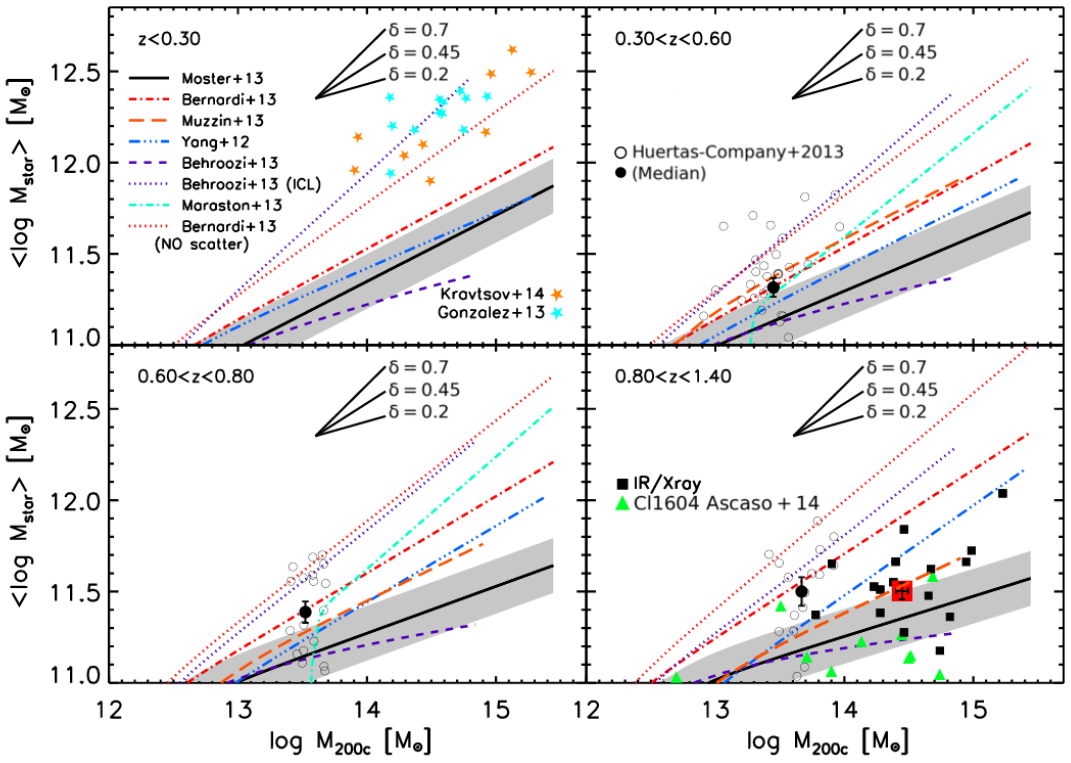



FIGURE 2.6: The median stellar mass to halo mass relation for central galaxies at $< z >= 0.1, 0.4, 1.1$ and 0.7 going clockwise from top left. Each panel shows data points from observed massive galaxies at that redshift range. Specifically, the data points in the top left panel are local galaxies from Kravtsov et al. (2014) (orange) and Gonzalez et al. (2013) (blue). The open circles in the other three panels show galaxies from Huertas-Company et al. (2013) and the filled circles in each panel show their median. The black squares in the bottom right panel show a selection of targeted observations (see Section 2.2.1) at $0.8 < z < 1.4$ and the green triangles are galaxies in the Cl1604 supercluster from Ascaso et al. (2014). The lines in each panel correspond to the results of abundance matching from stellar mass functions found in the literature. For reference, each panel also shows reference slopes where $\delta = 1-\gamma$ from Equation 2.2.

and satellite galaxies (from within the galaxy sample) at fixed halo mass $\langle N(M_{\rm halo})\rangle$, which can be renormalised to give the probability distribution, $dP/d\log M_{\rm halo}$.

In practice, the 2PCF is found by measuring the positions of all galaxies within a sample. The galaxy-galaxy ($gg$) 2PCF is then computed using, eg, the Landy & Szalay (1993) estimator:

$$\xi_{gg}(r) = \frac{n_R(n_R-1)}{n_D(n_D-1)}\frac{DD}{RR} - 2\frac{n_R-1}{n_D}\frac{DR}{RR} + 1 \tag{2.3}$$

where $DD$ is the histogram of the euclidean distance between every data point in sample. $RR$ and $DR$ are the same but now are histograms of the distances between all the points in a random and the distances between a random sample and the original data respectively. $n_D$ and $n_R$ are the number of data points in the data and random samples

respectively. The difficulty is that 2PCF is computed in 3D but measuring the distances to galaxies is very difficult. This can either be measured by directly measuring the distances to galaxies using, eg. Cepheid variable stars or by converting the galaxies redshift to a luminosity distance assuming that the galaxy is in the hubble flow. Conversely, the line-of-sight 2PCF is often used which is simply the 3D 2PCF projected onto a 2D plane. This is calculated as:

$$w_p(r_p) = 2\int_0^{r_{\pi,max}} \xi(r_p, r_\pi) dr_\pi \tag{2.4}$$

In what follows, it does not matter if $\xi$ or $w_p$ is chosen. The 2PCF is formed of two components, the 'one halo' which represents the clustering of galaxies within one halo and the 'two halo' term which is the clustering between galaxies in different haloes:

$$\xi_{gg}(r) = \xi_{gg}^{1h}(r) + \xi_{gg}^{2h}(r) + 1 \tag{2.5}$$

At large radii (beyond the virial radii of the most massive clusters) the two halo term dominates and at small radii the single halo term dominates. At extremely large radii the clustering tends to zero (in accordance with the cosmological principle of homogeneity).

In turn, the one halo term is given by

$$1 + \xi_{gg}^{1h}(r) = \frac{1}{2}\bar{n}_g^{-2} \int n(m)\langle N(N-1)\rangle_M \lambda(r|M) dm \tag{2.6}$$

and the two halo term is given by

$$\begin{aligned} \xi_{gg}^{2h} = \xi_{mm}^{lim}(r)\bar{n}_g^{-2} &\int n(M_1) b_h(M_1) \langle N\rangle_{M_1} dM_1 \\ &\times \int n(M_2) b_h(M_2) \langle N\rangle_{M_2} \lambda(r|M_1, M_2) dM_2 \end{aligned} \tag{2.7}$$

where $\bar{n}_g$ is the mean number density of galaxies, $n(M)$ is the halo mass function, $\lambda(r|M)$ relates to the radial mass profile of a halo with mass $M$ and $\lambda(r|M_1, M_2)$ relates to the convolution of the mass radial profiles of haloes with $M_1$ and $M_2$. $b_h$ is the halo bias which is a measurement of how clustered dark matter haloes are relative to the underlying dark matter distribution. The latter is found from high resolution, large scale n-body simulations.

The key to HOD modelling is that most of these parameters in equations 2.6 and 2.7 are well constrained and the only unknowns are the mean number of galaxies per halo of a given mass, $\langle N \rangle_M$ and the mean number of satellites with a given mass $M_2$ in a halo with mass $M_1$, $\langle N(N-1) \rangle_M$. For a sample of galaxies with a minimum mass, Kravtsov et al. (2004) and Zheng et al. (2005) show that the best analytical description for $\langle N \rangle_M$ and $\langle N(N-1) \rangle_M$ are of the type:

$$\langle N_{cen}(M) \rangle = \frac{1}{2} \left[ 1 + erf \left( \frac{\log M - \log M_{min}}{\sigma_{\log M}} \right) \right] \tag{2.8}$$

$$\langle N_{sat}(M) \rangle = \langle N_{cen}(M) \rangle \left( \frac{M - M_0}{M_1'} \right)^{\alpha} \tag{2.9}$$

these can be convolved with the halo mass function and renormalised to give the probability that a galaxy within the sample is in a halo of mass $M$, ie. $dP(M)/dM$.

Shankar et al. (2014b) uses the host halo mass probability distributions, $dP/d\log M_{\rm halo}$ from Guo et al. (2014) as an independent constraint on the scatter in the $M_{\rm star}$-$M_{\rm halo}$ relation. Figure 2.7 shows the predicted $dP/d\log M_{\rm halo}$ distributions of galaxies with $\log M_{\rm star} > 11.5 [M_{\odot}]$ at $0.4 < z < 0.6$ (left panel) and $0.6 < z < 0.8$ (right panel). The grey bands show the results of the HOD analysis by Guo et al. (2014) who use a sample of galaxies from BOSS with that mass cut at those redshifts. The three lines in each panel show instead the results of the analysis by Shankar et al. (2014b) who use AM to fit the Muzzin et al. (2013) stellar mass function, which matches the number densities of Guo et al. (2014). Different values of the intrinsic scatter are used, $\sigma = 0.15, 0.2$ & $0.25$ dex. As evident from Figure 2.7, an intrinsic scatter of $\sigma \lesssim 0.15$ dex best matches the clustering measurements.

Overall, the results of Shankar et al. (2014b) can be summarised as follows:

- To fit the clustering data of central galaxies from Guo et al. (2014), a scatter of $\sigma \sim 0.15$ *dex* is required (at least at $z \lesssim 1$).
- The latest determinations of $z \lesssim 1$ stellar mass functions point to the high-mass slope of the $M_{\rm star}$-$M_{\rm halo}$ relation being substantially steeper than previous estimates in the literature (Behroozi et al., 2013b; Moster et al., 2013) in the redshift range $0 \leq z \lesssim 1$.
- The null hypothesis that the stellar mass function does not evolve is not supported by direct observations of the stellar and halo halo masses of massive, central galaxies in the redshift range $0 \leq z \lesssim 1$, when compared to the expectations from AM.

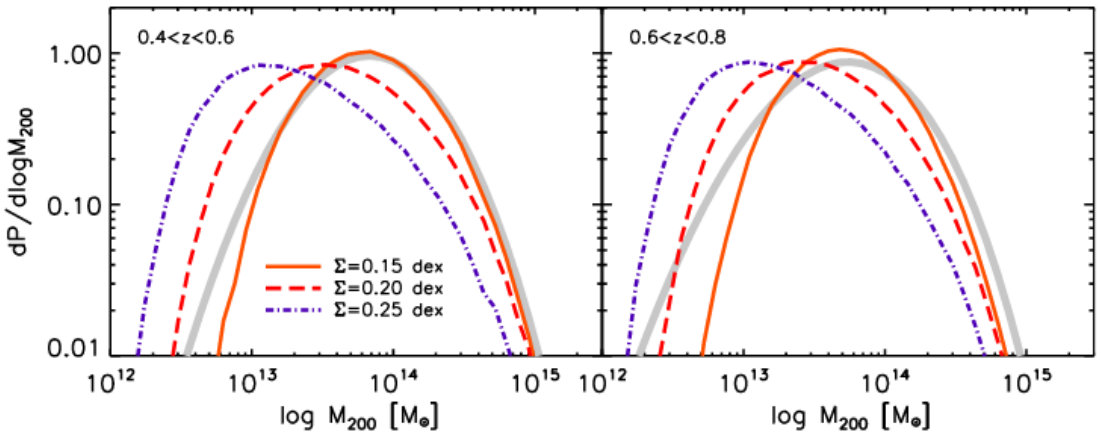



FIGURE 2.7: The predicted probability that a BCG with log $M_{\rm star}> 11.5[M_{\odot}]$ will reside in a halo of mass $M$. The grey band is the results from Guo et al. (2014) for observed, central galaxies in BOSS and the coloured lines are the results of Shankar et al. (2014b) analysis for central galaxies using three different intrinsic scatters, as labelled. The left panel shows results at $z = 0.5$ and the right panel shows them at $z = 0.7$.

The validity of the above results partly relies on the reliability of the measurements of high redshift stellar masses and volumes. While a full discussion of this will be presented in Chapter 4, the work presented in the rest of this chapter will show that these preliminary results on the evolution of the most massive galaxies are consistent with the predictions of a state-of-the-art cosmological evolution model.

## 2.3 Semi-empirical model

In the rest of this chapter, I present a state-of-the-art model for growing galaxies through cosmic time. The semi-empirical model (SEM hereafter) is designed as a light-weight, flexible, fast, efficient model based on few input parameters and assumptions; In contrast with the SEM, semi-analytic models populate haloes with gas and allow them to evolve via complex parametrisations which attempts to capture the core physics. Instead, SEM populates haloes with galaxies using empirical AM relationships before evolving them through empirically inspired recipes. In the following section, I describe the recipes used to evolve galaxies in the SEM after which I present a handful of key results.

### 2.3.1 Merger trees

Dark matter halo merger trees have become the backbone of most (if not all) models of galaxy formation. This is because dark matter haloes define the potential well in which high redshift gas condenses into to form galaxies. The depth of this well controls the flow

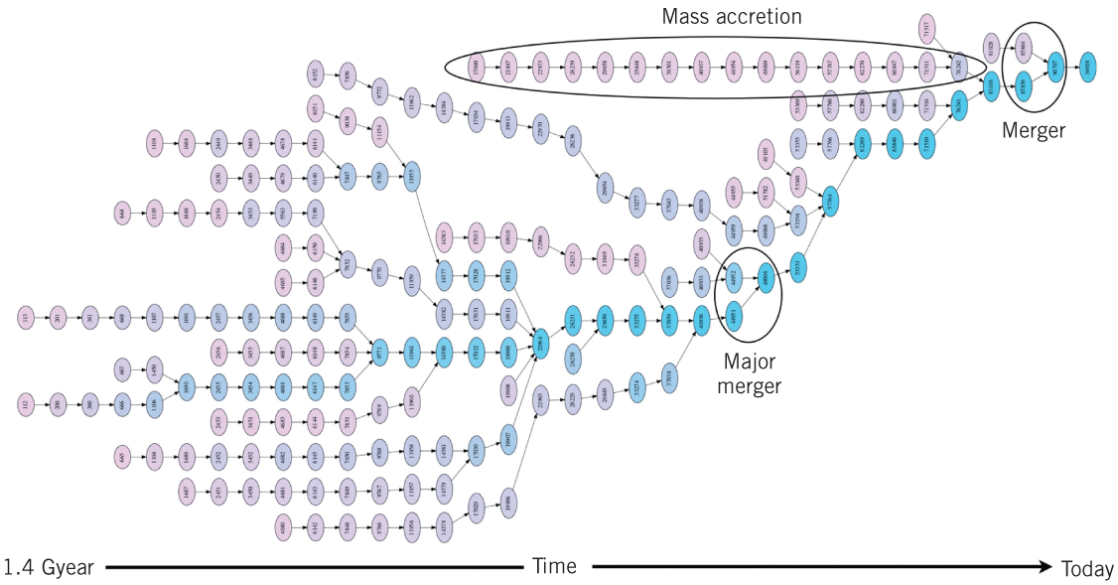



FIGURE 2.8: An example merger tree taken from Heitmann et al. (2014). The coloured discs represent dark matter haloes where the mass increases from pink to blue. As highlighted, forks show when two dark matter haloes merge. A major merger is when the two haloes are of compatible mass. Haloes can also grow via steady smooth accretion of dark matter from the cosmic web.

rate of gas into its centre and the rate at which this gas cools to form stars. Additionally, dark matter haloes grow hierarchically with smaller haloes merging together into larger ones. The mergers between haloes give rise to mergers between galaxies and thus the merger history of the halo itself possibly becomes a driver of galaxy evolution.

Mergers between haloes can be represented as trees; In this analogy, the trunk of the tree is the "so-called" main progenitor branch, which represents the direct evolution of the most massive halo. Off the trunk are the "branches" which represent the evolution of the lower mass haloes merging onto the main progenitor.

Figure 2.8 shows a merger tree where the circles represent dark matter haloes, and cosmic time increases from left to right. The colour of each circle indicates the mass of the halo. Importantly, it highlights the key features that define a merger tree. Along the top of the tree is a halo which has no mergers but sill increases in mass via smooth accretion of dark matter from the cosmic web. In the middle of the tree, a major merger is highlighted which is a merger of two haloes with equal masses.

Producing dark matter merger trees can be generally achieved in two ways. The first is to use the Extended Press-Schechter formalism to randomly fragment a $z = 0$ seed halo into its progenitors at higher redshift. The second is to utilise large volume N-body cosmological simulations. Because dark matter plays a dominant role in the evolution of its central galaxy, it is very important to construct statistically correct merger trees before building any galaxy evolution models on top of them. In Appendix A, I discuss

merger trees from numerical simulations and analytic algorithms and their respective advantages and disadvantages for assessing massive galaxy evolution. In summary, I find that the Parkinson et al. (2008) algorithm is the best choice for the SEM.

### 2.3.2 Generating halo catalogues

Having selected to use Parkinson et al. (2008) merger trees, the first step is to construct a catalogue of merger trees to use as the backbone for my galaxy evolution models. Firstly, a sufficiently large catalogue of massive haloes needs to be constructed at $z = 0$. Each halo in the catalogue will form the base of a distinct merger tree, created via the Parkinson et al. (2008) algorithm. For a given volume, I integrate the halo mass function, finding the masses where the number of haloes is an integer, i.e., I perform the operation:

$$n = V \int_{M_{\rm halo}}^{\infty} \Phi(m) dm \quad n \in \mathbb{Z}\,. \tag{2.10}$$

Similarly to a random selection, the approach in equation 2.10 efficiently produces large catalogues of dark matter haloes. Figure 2.10 demonstrates this procedure for a volume, $V = 10^4\ Mpc^3 h^{-3}$. Each coloured band has an area equal to one and therefore is a distinct halo.

The initial catalogue of haloes is extracted from the Tinker et al. (2005) halo mass function. Figure 2.9 shows the Tinker et al. (2005) halo mass function is a very good fit to two large, dark matter only simulations, Bolshoi (Klypin et al., 2011) and MultiDark (Riebe et al., 2011). It is clear that these simulations have limited statistics at high halo masses which is the focus of this thesis. This is one of the main reasons why the SEM was constructed on top of analytic merger trees. However, I will often refer back to numerical simulations when appropriate.

### 2.3.3 Adding galaxies

In this section, I describe the details of the semi-empirical model.

Initially, for a given volume at $z = 0$, a catalogue of dark matter haloes is constructed using the method described in section 2.3.2. For each halo in the catalogue, I construct a merger tree using the Parkinson et al. (2008) algorithm to a maximum redshift, $z_{max} = 3$ (since the empirical AM relationships are defined best below this redshift Huertas-Company et al. 2015) and with a mass resolution of $M_{min} = 10^{11}[M_{\odot}h^{-1}]$ (as this best

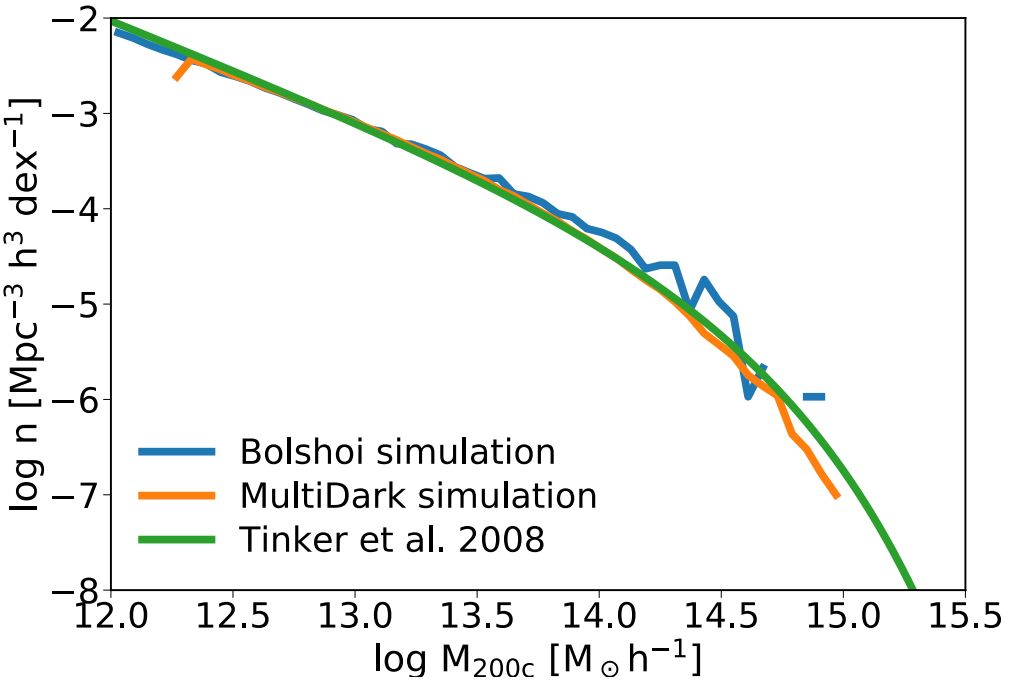



FIGURE 2.9: A comparison between the halo mass functions produced by two dark matter-only simulations and the analytic relation of Tinker et al. (2008) at $z = 0$.

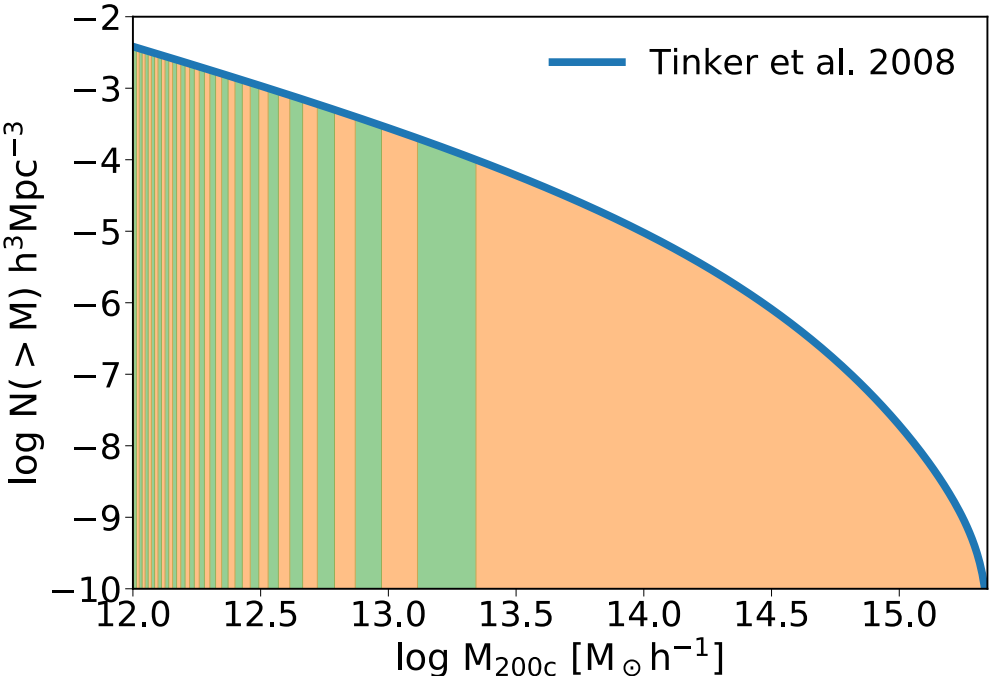



FIGURE 2.10: A demonstration of our sampling strategy. The blue curve is the integrated number density of haloes above a given mass. The alternating coloured regions represent single haloes (when N in an integer) for a $10^4$ $Mpc^3h^{-3}$ box.

balances good mass resolution while allowing for large numbers of massive galaxies to be simulated). The SEM then starts with the lowest mass haloes in the $z = 0$ halo catalogue and sequentially works to the more massive ones.

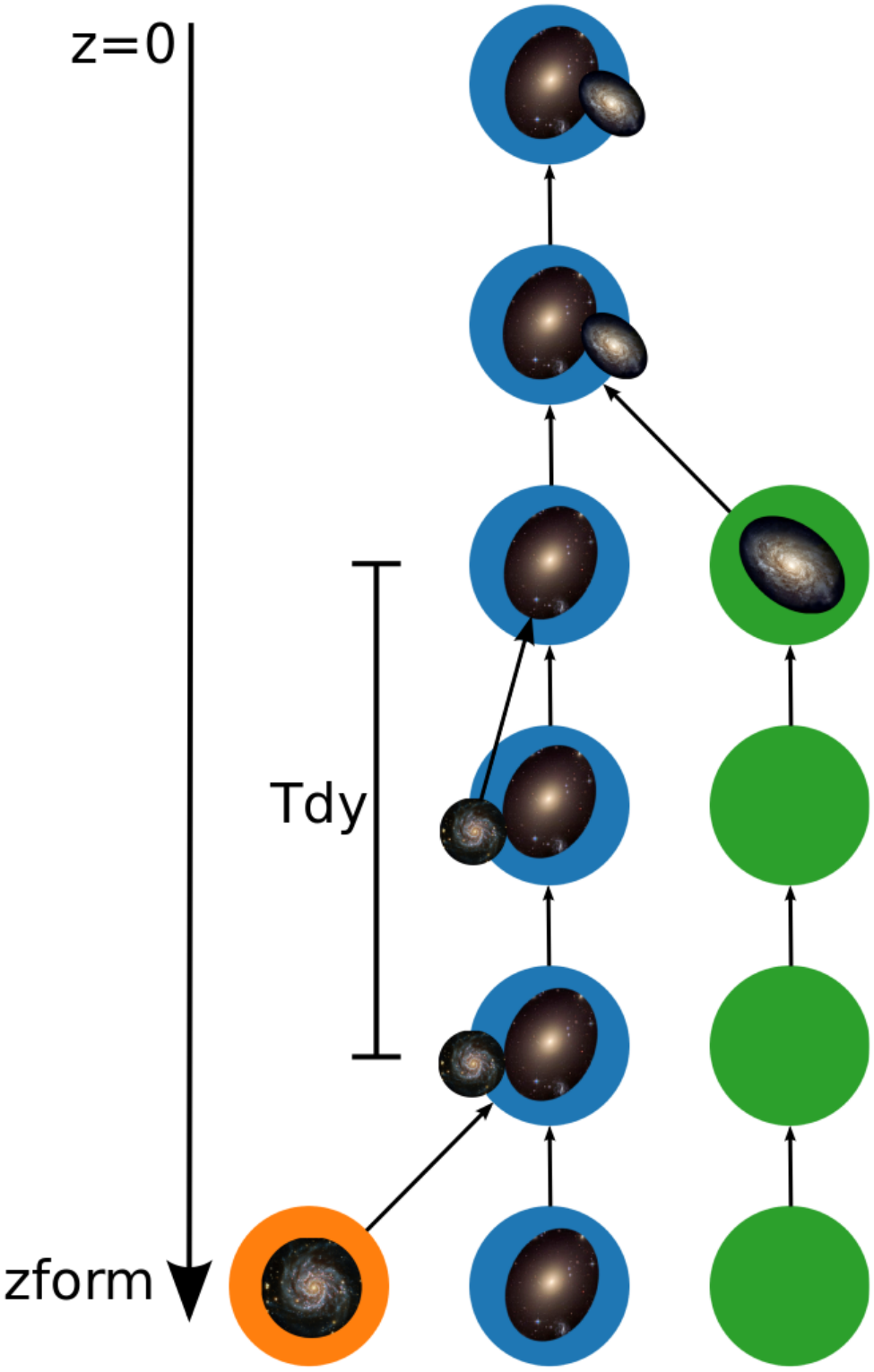



FIGURE 2.11: A cartoon showing a portion of a single merger tree which is being simulated by the semi-empirical model. The coloured circles represent the underlying dark matter. The presence of a galaxy shows that that halo is populated. The central column shows the main progenitor branch and the others show smaller galaxies which merge onto it.

Figure 2.11 shows a cartoon of the inner workings of the SEM for a single $z = 0$ central galaxy. The coloured circles show the underlying dark matter merger tree from some arbitrary $z_{\rm form}$ to $z = 0$. The central column represents the main progenitor branch of the merger tree and the galaxy at its centre is the the main focus of the SEM. The other branches are mergers between the main progenitor and other, (usually) less massive haloes (see Section 3.2.2 a detailed discussion).

Firstly, the main progenitor is populated with a galaxy at the chosen $z_{\rm form}$. This galaxy is assumed to be a typical, star forming, disc-like (see section 1.5.2.1) main sequence galaxy (Huertas-Company et al., 2015, showed this to be the case for the galaxies the SEM is focused on at $z \gtrsim 3$.) and thus, haloes are populated using empirical relationships for $M_{\rm star}$-$M_{\rm halo}$ (eg. Moster et al., 2013), $M_{\rm star}$-$M_{gas}$ (eg. Stewart et al., 2009), $M_{\rm star}$-specific star formation rate (sSFR; e.g. Peeples & Somerville, 2013) and $M_{\rm star}$-effective (half-light) radius ($r_e$; e.g. Shen et al., 2003) with $r_e(z)/r_e(0)$ from (Hopkins et al., 2009).

Next, the haloes that merge directly with the main progenitor get populated with galaxies at their redshift of infall, $z_{\rm inf}$. This is initially done assuming that the satellite galaxies are also disc-like and follow the main sequence. However, when the simulation has built up a statistically significant population of galaxies, the model searches through through the simulated main progenitors to find one with a similar halo mass and redshift and copies the properties of that galaxy into the infalling halo. This makes the simulation self-consistent as the model merges galaxies that are representative of the galaxy population at lower redshift.

Once the main progenitor and $1^{st}$ order progenitors in the merger tree have been populated with galaxies, the model works on evolving the central galaxy. At each timestep, the model will evolve the galaxy through any in-situ processes that have been included[1]. Additionally, as detailed below, any satellite galaxy that merges with the central in that timestep gets its properties 'merged' with the central.

When dark matter haloes merge according to the merger tree, the model waits an additional dynamical friction timescale, before allowing the galaxies to merge. The dynamical friction timescale is given by:

$$t_{df} = t_{dyn} \times T(M_{halo}/M_s, orbit)\,, \tag{2.11}$$

[1]The semi-empirical model presented in this chapter does not include processes such as disc instabilities (Bower et al., 2006) or disc regrowth after major mergers (Hammer et al., 2009). However, semi-analytic models have shown these processes only play a significant role in intermediate mass galaxies (De Lucia et al., 2011; Wilman et al., 2013) which is below the stellar mass threshold we consider in this work.

where $t_{dyn}$ is the dynamical time given by $t_{dyn} = 0.1H(z)^{-1}$ and $T(M_{halo}/M_s, orbit)$ is a general function of the host-to-satellite mass ratio at infall and the orbit of the satellite. $T(M_{halo}/M_s, orbit)$ is normally explored using high resolution simulations of idealised merging galaxies (e.g. Lotz et al., 2008). There are several analytical treatments of $t_{df}$ in the literature (see De Lucia et al., 2010, for review). The SEM presented in this chapter adopts the parametrisation presented by McCavana et al. (2012). The dynamical friction timescale ranges from $\sim 0.1Gys$ to $\sim 10Gys$ (but typically few $Gys$) depending on the mass ratio of the central and satellite.

After the delay of $t_{df}$, the merger gets classified into one of three categories based on the ratio of the central and satellites stellar mass, $\mu = M_{star,sat}/M_{star,cen}$.

- If $\mu < 0.01$, the merger is classified as a "micro" merger. All the stellar mass and gas mass of the satellite get added to the central galaxy's disc.
- If $0.01 \leq \mu < 0.3$, the merger is classified as a "minor" merger. All the stellar mass of the satellite gets added to the bulge of the central and the gas from the satellite gets added to the disc of the central. Additionally, there is a starburst during which a fraction of the total gas gets converted to stars. This is due to the satellite causing turbulence in the gas in the central, inducing star formation (Sanders et al., 1988; Bell et al., 2006). Specifically, I use the prescription from Somerville et al. (2001):
$$e_{burst} = 0.56 \left(\frac{M_2}{M_1}\right)^{0.7}, \tag{2.12}$$
where Somerville et al. (2001) chose these parameters to reproduce the detailed numerical simulations by Mihos & Hernquist (1994, 1996).
- If $\mu \geq 0.3$, the merger is classified as a "major" merger. In this case, the gas from both the central and satellite is converted into stars in a maximal starburst. The stellar mass from this new starburst, the total stellar mass from the satellite and the stellar mass from the disc of the central are added to the bulge of the central such that only the bulge remains. This simulates a complete morphological transformation to an elliptical galaxy.

In all cases, the new size of the central galaxy is calculated by conserving the binding and orbital energies of the central and satellite galaxies.

$$\frac{(M_1 + M_2)^2}{R_{new}} = \frac{M_1^2}{R_1} + \frac{M_2^2}{R_2} + \frac{f_{orb}}{c}\frac{M_1 M_2}{R_1 + R_1}, \tag{2.13}$$

Where $M_i$ and $R_i$ are, respectively, the total baryonic masses and half light radii of the merging galaxies. $c$ is the form factor and takes the value $c \sim 0.5$. In principle, $c$

can be calculated analytically for a given density profile or can be found experimentally through numerical simulations (Cole et al., 2000). $f_{orb}$ parametrises the average mutual orbit between the merging galaxies where 0 are parabolic orbits and 1 are circular.

A final important procedure in the model is to re-initialise the central galaxy at every timestep. In practice, this means that if the stellar mass is too low relative to AM relations, stellar mass gets added to the disc which accounts for extra, hidden star formation or any other unspecified physics that drives galaxy evolution. Additionally, if the gas fraction is too low then extra gas is added to replicate cold flows. This step is only done when necessary to preserve the modelled scatter in the $M_{\rm star}$-$M_{\rm halo}$ relation and the $M_{\rm star}-r_e$ relation which are two key outputs of the model used to compare to data. There is no clear way to re-assign sizes and thus the process of updating stellar masses slightly biases the $M_{\rm star}-r_e$ relation. Because of this, it is important to reinitialise stellar mass as little as possible and therefore, the star formation and merger routines are critical. It is currently not possible to extract the breakdown of how much a galaxy's stellar mass is grown by mergers, star formation and reinitialisation (because of the volume of data this would add to the model's output) but it is an important feature being explored by Grylls et al., in prep.

However, the re-initialisation routine can be vetoed if the galaxy is 'quenched'. I implement two forms of quenching in the SEM. The first is when the central galaxy has a major merger. The second is if the halo grows larger than $10^{12}[M_{\odot}]$ simulating the shock heating of cold flows (Dekel et al., 2009).

Figure 2.13 shows a more detailed schematic of the model. The box labelled 'A' shows the section which controls loading merger trees and populating them with galaxies. Instead, the box labelled 'B' is the section that evolves galaxies from $z_{\rm form}$ to $z = 0$. Finally, the box labelled 'C' controls the storage and querying of central galaxies simulated earlier in the model.

### 2.3.4 Results

In this section I present key results from the SEM I constructed. I first present the SEM in action on a single merger tree. I then present the median stellar mass and effective (half-light; $R_e$) radius evolution of massive galaxies simulated within a large cosmological box as well as the number of mergers they have at fixed halo/stellar mass. These results were used to validate the Shankar et al. (2014a) SEM in Shankar et al. (2015). Additionally, I show the effect of changing the orbit between the central and satellite galaxy has on the $M_{\rm star}$-$R_e$ relation, validating a key result of Shankar et al.

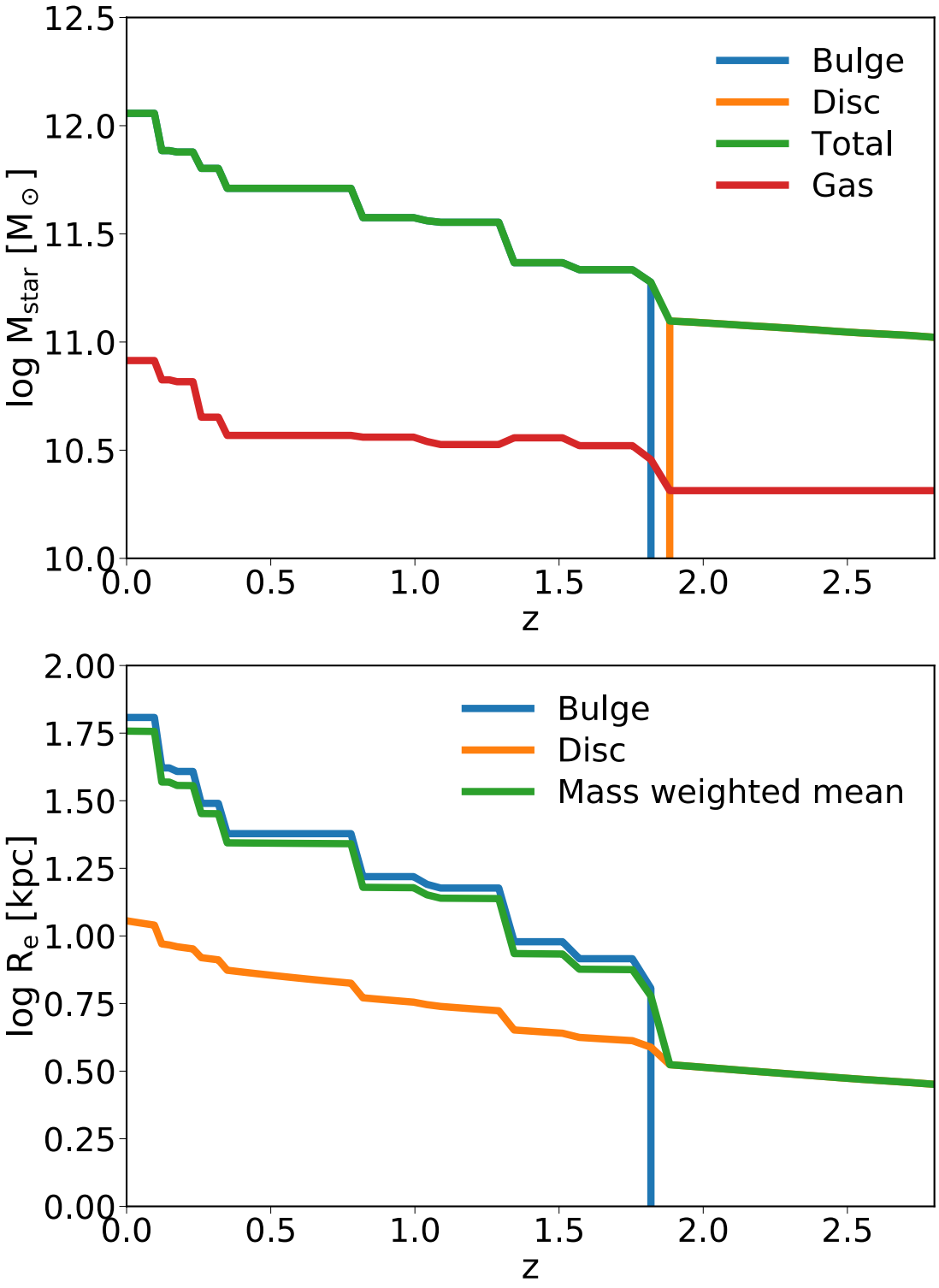



FIGURE 2.12: The evolution of a single, massive galaxy in the semi-empirical model. The top panel shows the stellar mass contained within the disc (orange), the bulge (blue) and total (green). Additionally, the cold gas mass is shown (red). The bottom panel shows the effective radius of the disc (orange), bulge (blue) and the mass weighted average (blue). There is a major merger at $z \sim 1.8$ which transforms the galaxy from disc dominated to a giant elliptical.

(2014a). Additionally, I will present an extra, new result assessing the effect 'halo quenching' has on the $M_{\rm star}$-$R_e$ relation.

#### 2.3.4.1 Individual brightest cluster galaxy

Figure 2.12 shows the evolution of the stellar mass (left panel) and the effective radius (right panel) of one massive galaxy out of $\sim 30,000$ galaxies in a $256^3$ $[Mpc^3]$ cosmological box. Galaxies are comprised of a disc, a bulge and a gas component. This specific galaxy starts as a disc and grows initially in size and radius via gas accretion and in-situ star formation. At $z \sim 1.8$, the galaxy has a major merger which causes the disc to be fully disrupted and become a pure elliptical galaxy. Below this redshift, the stellar mass and bulge size grow dramatically via a string of dry (gas-poor) mergers.

#### 2.3.4.2 Stellar mass evolution

Figure 2.14 shows the median stellar mass growth of all central massive galaxies with $\log M_{\rm star} > 11.5 [M_\odot]$ in a $256^3$ $[Mpc^3]$ cosmological box. The model predicts that very massive galaxies have a mass evolution of $\sim 0.3$dex between $z = 1$ and $z = 0$ which is consistent with the original findings by Shankar et al. (2014b, Figure 2.6 here), more advanced abundance matching (Shankar et al., 2015), other semi-analytic models (eg. Guo et al., 2011) and numerical simulations (eg. Torrey et al., 2015).

#### 2.3.4.3 Size evolution

The SEM also predicts a degree of size evolution of massive galaxies. As discussed in section 2.1, one key observable is the tight correlation between the stellar mass and size of galaxies. I have explored the effect changing the orbital energy parameter ($f_{orb}$ in equation 2.13) has on the size-mass relation. Figure 2.15 shows the observed size-mass relation of local galaxies from SDSS (Bernardi et al., 2013) as a blue line. Additionally, it shows the median size of central, $z = 0$ galaxies from two outputs of the SEM where $f_{orb} = 0$ (efficient parabolic orbits) and $f_{orb} = 0.5$ (less efficient, more circularised orbits) are adopted. Clearly, the model favours the more efficient merging orbits confirming the findings of Shankar et al. (2014a).

Having established that the SEM can faithfully replicate the local size-mass relation, Figure 2.16 shows the median and $1\sigma$ dispersion of the radii of the progenitors of galaxies with $\log M_{\rm star} > 11.5$ at $z = 0$. The model predicts that with efficient mergers, massive galaxies may grow by up to a factor of four between $z = 1$ and $z = 0$.

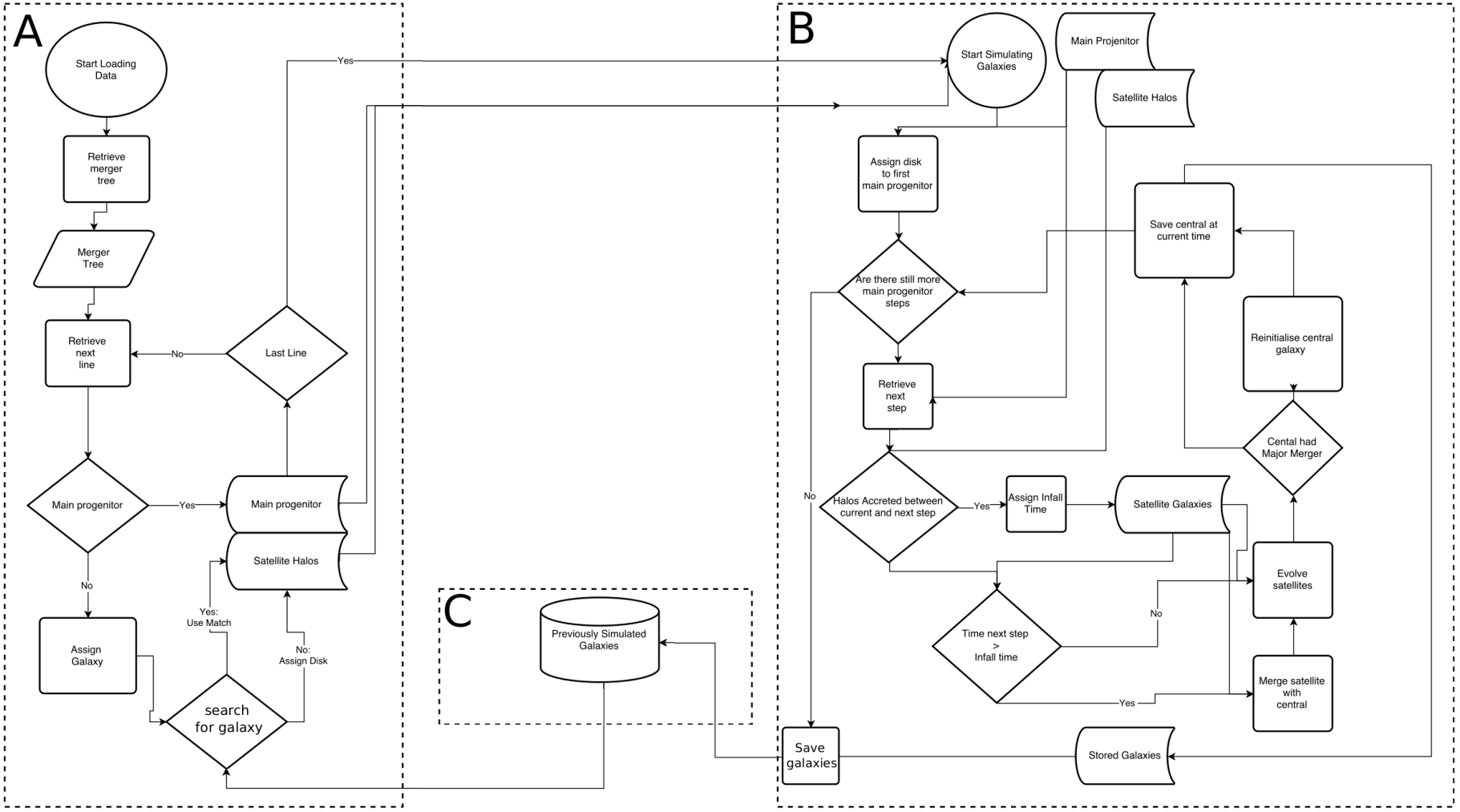



FIGURE 2.13: A detailed schematic of the semi-empirical model presented in this chapter (adapted from Grylls et al., in prep). Box 'A' contains the processes that load merger trees and assign dark matter haloes galaxies. Box 'B' is the section of the model which evolves the galaxies from their formation to $z = 0$. Box 'C' relates to storing the simulated galaxies and retrieving them as to populate new dark matter haloes.

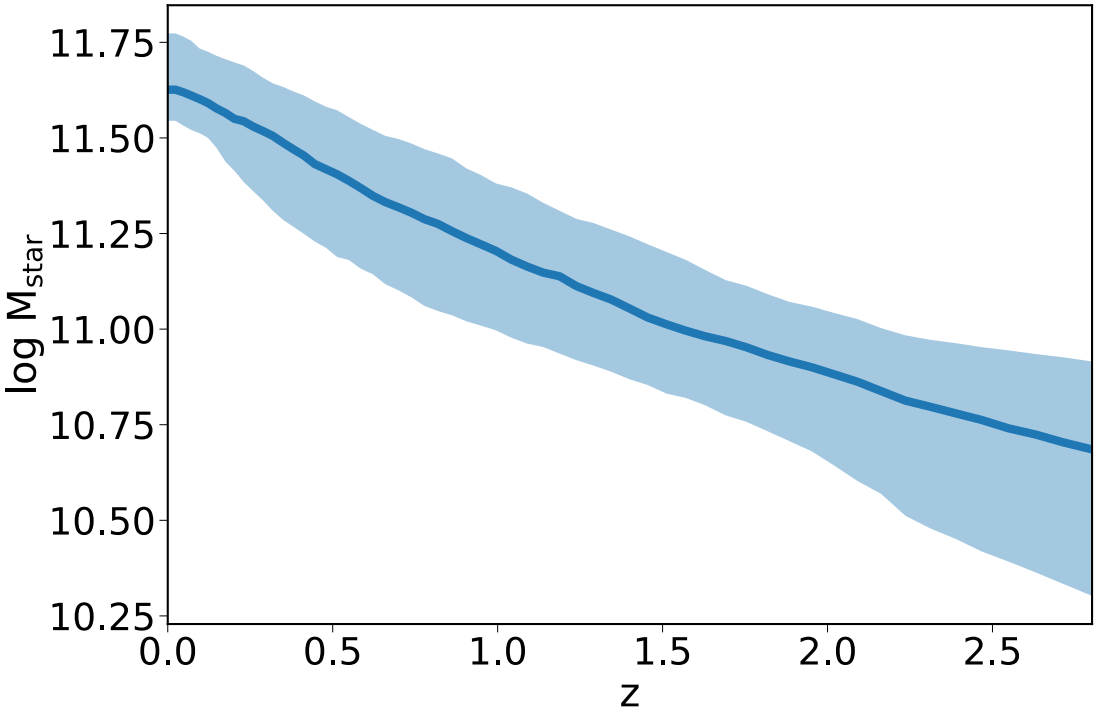



FIGURE 2.14: The median and $1\sigma$ scatter of the stellar masses of the progenitors of galaxies which have a final stellar mass $\log M_{\rm star} > 11.5$ at $z = 0$ evolved using the SEM.

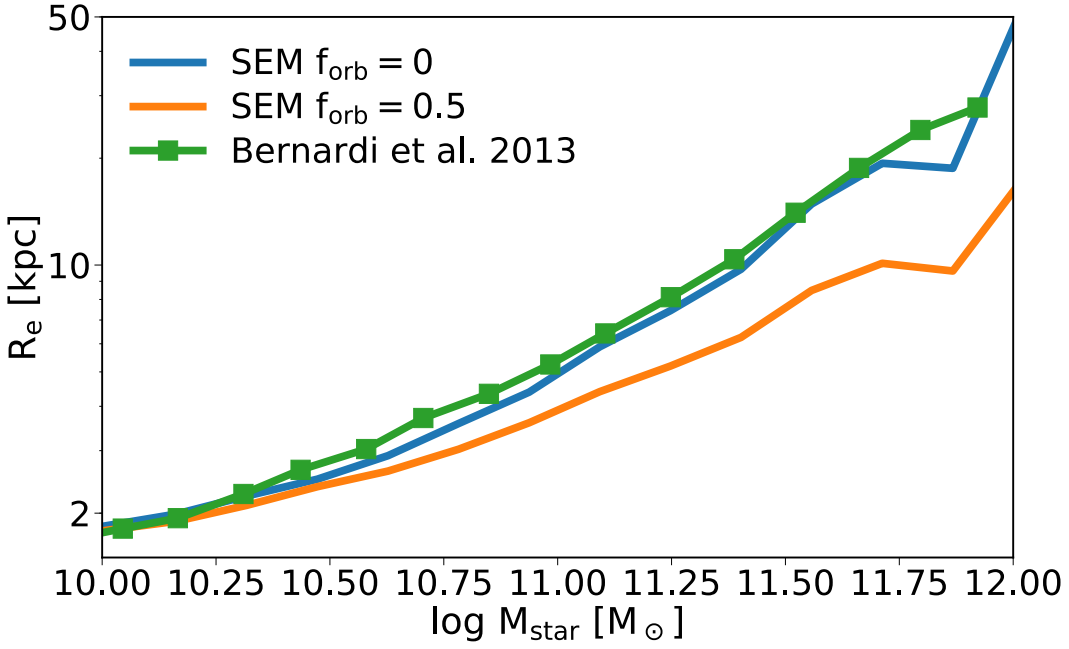



FIGURE 2.15: A comparison between the observed size-mass relation for local galaxies in SDSS from (Bernardi et al., 2013, green line) and the results of the SEM assuming all satellites have $f_{orb} = 0$ (parabolic orbits, blue line) and an $f_{orb} = 0.5$ (intermediate orbit, orange).

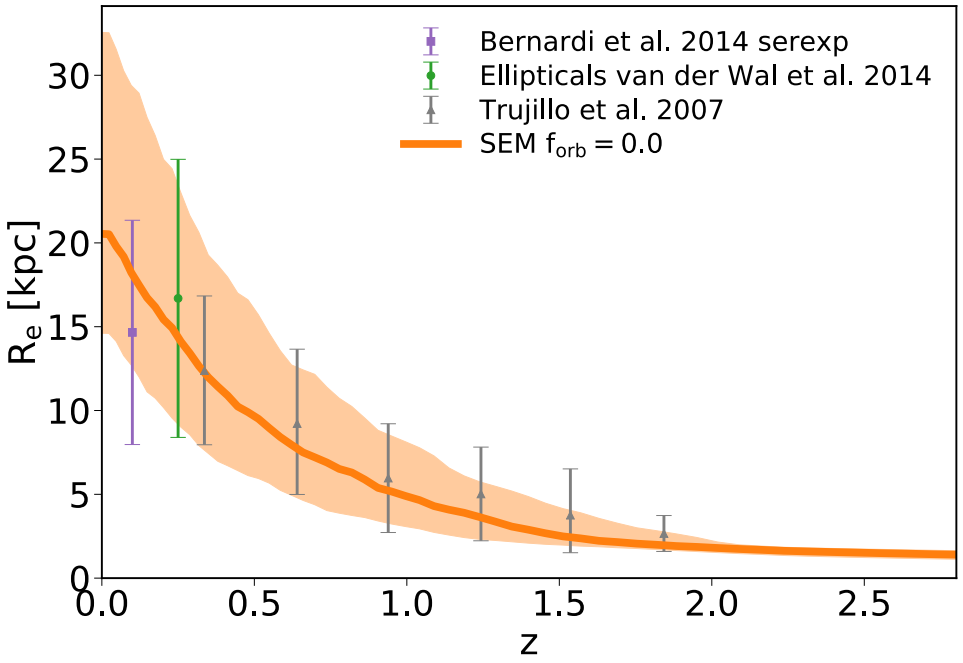



FIGURE 2.16: The median and $1\sigma$ dispersion of the effective radii of the progenitors of galaxies with $\log M_{\rm star} > 11.5[M_\odot]$ at $z = 0$. The error bars show the sizes of typical elliptical galaxies from SDSS (Bernardi et al., 2013, purple), DEEP2 (Trujillo et al., 2007, Grey) and CANDLES (green van der Wel et al., 2014) with the same stellar masses as the galaxies shown from the SEM.

Additionally, I test the effect that halo quenching has on the size-mass relation. Halo quenching is implemented by 'shutting down' the galaxy (vetoing the galaxy's reinitialisation). There are several physical mechanisms that could result in the central galaxy quenching. Here, I test the hypothesis that halo quenching is the main driver thus only quench galaxies if they are above the halo mass threshold given by Cattaneo et al. (2006):

$$M_{crit} = M_{shock} \times max\{1,\ 10^{1.3(z-z_c)}\}[M_\odot]\,, \tag{2.14}$$

where $M_{shock} = 10^{12}$ and $z_c$ is a free parameter in the range $\sim 1-3$. Figure 2.17 shows the observed size-mass relation of local galaxies in SDSS from Bernardi et al. (2013) (green line). Additionally, it shows the median size-mass relation of central $z = 0$ galaxies from two runs of the SEM including halo quenching with $z_c = 3$ (orange line) and without (blue line). No significant change to this relation is evident between the two relations indicating late stellar growth in massive central galaxies is dominated by (dry) mergers.

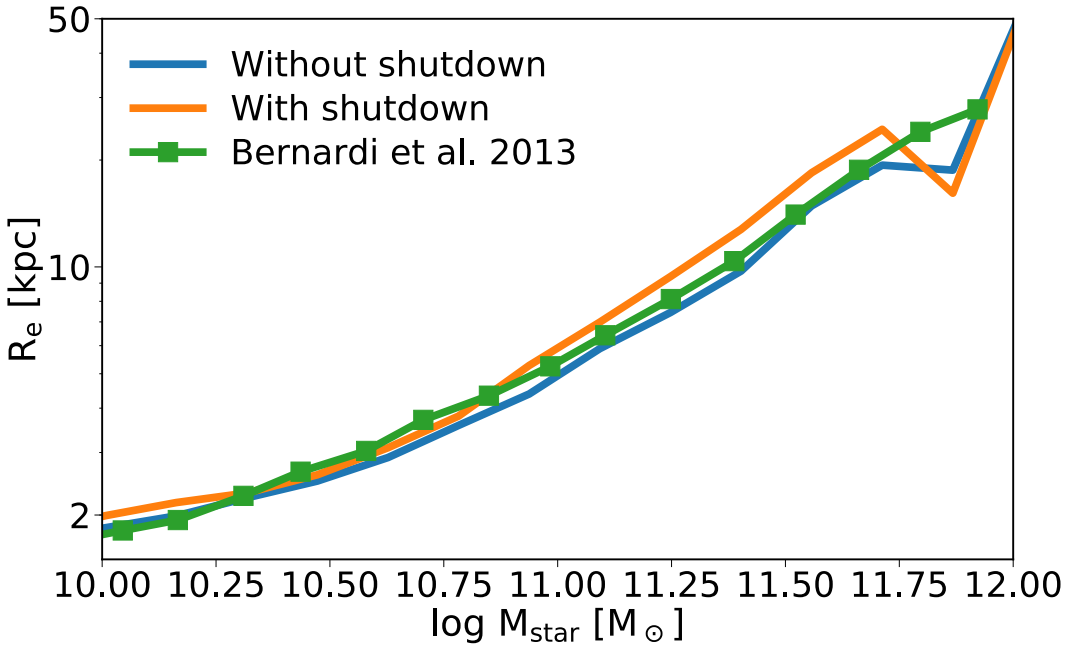



FIGURE 2.17: Same as Figure 2.15 but showing the effect of including environmental quenching in the SEM.

#### 2.3.4.4 Number of mergers

One last interesting output of the SEM is the number of mergers each massive, central galaxy has undergone. Figure 2.18 shows a comparison between the numbers of mergers galaxies (left panel) and haloes (right panel) have as a function of their mass between $z = 1$ and $z = 0$. Specifically, the colours in each plot represent the mean number of mergers above a given merger ratio, $\mu \equiv M_{sat}/M_{cen}$ at the redshift of the merger. I found that massive galaxies with $\log M_{\rm star} > 11.5[M_\odot]$ have, on average, $\sim 1$ major merger and $\sim 8$ minor mergers. This number drops as the $z = 0$ stellar mass and halo mass decreases.

Observationally, the number of mergers that massive galaxies have is still debated. For instance, Liu et al. (2015) inferred that merger rate for BCGs is $(0.55 \pm 0.27)Gyr^{-1}$ at $z \sim 0.43$ from deep, high resolution imaging. From this merger rate, they find BCGs should assemble $\sim (35 \pm 15)\%$ of their present stellar mass below $z = 0.6$ via dry, minor mergers which is fully consistent with the results presented in this chapter. Conversely, a detailed study of galaxy pairs by Mundy et al. (2017) find the major merger rate ($\mu > 0.3$) is $0.04^{+0.025}_{-0.008}Gyr^{-1}$ and clustering measurements (eg. Wake et al., 2008; White et al., 2008, and references therein) infer a merger rate that is considerably lower still ($\sim 0.024Gyr^{-1}$). This observational discrepancy is, in part, due to the assumptions in converting from galaxy pairs to a merger rate. Namely, the time it takes for the

galaxies to merge ($T_{obs}$) and the fraction of galaxies which are spatially close that are real merging events ($C_{merge}$; Mundy et al., 2017).

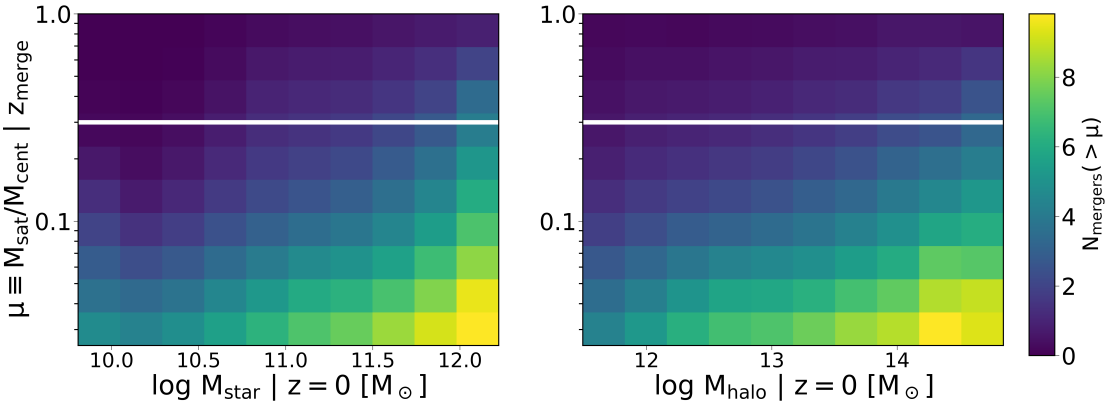



FIGURE 2.18: The mean number of mergers between $z = 1$ and $z = 0$, above a given merger ratio, $\mu$ (defined at the time of the merger) a galaxy (left) or halo (right) has as a function of their $z = 0$ mass.

## 2.4 Summary

In this chapter, I have shown that:

- Comparing basic abundance matching predictions to observations of massive galaxies showed, firstly, that the high mass slope of the $M_{\rm star}$-$M_{\rm halo}$ relation should be substantially steeper than the commonly used relations in the literature and the scatter should be small. Secondly, this analysis showed that there are signs of substantial growth in stellar mass in massive, central galaxies. However, this latter conclusion rests on the accuracy of measurements of $M_{\rm star}$ and estimates of co-moving volumes at $z \sim 1$.

- A state-of-the-art, semi-empirical model has been developed from scratch, following the general methodology laid out in Shankar et al. (2014a). The model takes dark matter merger trees as input, and gives the stellar mass, gas fraction and effective radius as functions of redshift as output.

- By inputting merger trees which are drawn from a cosmologically significant sample, I showed that the model predicts that massive galaxies can grow by a factor two in stellar mass between $z = 1$ and $z = 0$, consistent with other studies.

- Additionally, massive galaxies can grow in size by (up to) a factor of four over the same redshift range, providing that mergers are "efficient" (satellites have parabolic orbits).

- Finally, massive galaxies have on average $\sim 1$ major mergers between $z = 1$ and $z = 0$ and $\sim 8$ minor mergers.

## 2.5 Future work

The semi-empirical model I designed and constructed has been left as a legacy project to the galaxy formation group at the University of Southampton. Already, a graduate student has started to modify and extend the model to include velocity dispersion to study the $M - \sigma$ relation and has plans to additionally include black holes.

# Chapter 3

# Constraining the evolution of the most massive galaxies through their abundances and ages.

## 3.1 Introduction

In the previous chapter, I discussed an initial probe of massive galaxy evolution using abundance matching as well as a cosmological semi-empirical model. I now present the work published in Buchan & Shankar (2016). The aim of this chapter is to set more stringent and secure constraint on the evolution of the most massive, central galaxies in the local universe for $M_{\rm star}$>$3 \times 10^{11}$ $M_{\odot}$ for which data are still incomplete and/or uncertain, especially at high redshifts. In this chapter, I use a series of observationally-driven models that, by design, rely on very few assumptions and thus provide us with constraints less clouded by more complex modelling.

This chapter is structured as follows: In Section 3.2 I give an overview of the methodology and describe the sample selection. In Section 3.3 I discuss the constraints I set on the assembly scenario of massive galaxies. In Section 3.4 I investigate the relative importance of in-situ processes and mergers in driving the evolution of massive ETGs in a late assembly scenario using both observationally informed models as well as a full cosmological, semi-empirical model.

In this chapter I adopt a flat ΛCDM cosmological with $\Omega_M = 0.3$, $h = 0.7$, $\Omega_B = 0.045$, $\sigma_8 = 0.8$, $d_c^0 = 1.69$, and assume a Chabrier initial mass function (IMF: Chabrier, 2003). Throughout this chapter, I define the halo mass as $M_{\rm halo}$= $M_{200c}$, 200 times the critical density at redshift $z$.

## 3.2 Overview and Methodology

In a dark-matter dominated universe, large scale structures are formed from the collapse of primordial density fluctuations (White & Frenk, 1991). Over cosmic time, cold primordial gas condenses within these density perturbations forming the baryonic portion of galaxies (see Mo et al., 2010, for a detailed review). However, the processes of turning primordial gas into the galaxies we observe today are still debated. In this chapter, I circumvent the complexities of baryon physics by tracing the evolution of the host dark matter haloes, which is more transparent and secure as it relies only on gravitational physics. I then map galaxies to haloes in a statistical sense using semi-empirical relationships. I selected dark matter haloes from the dark matter-only Bolshoi simulation (Klypin et al., 2011), which provides the full dark matter merging history.

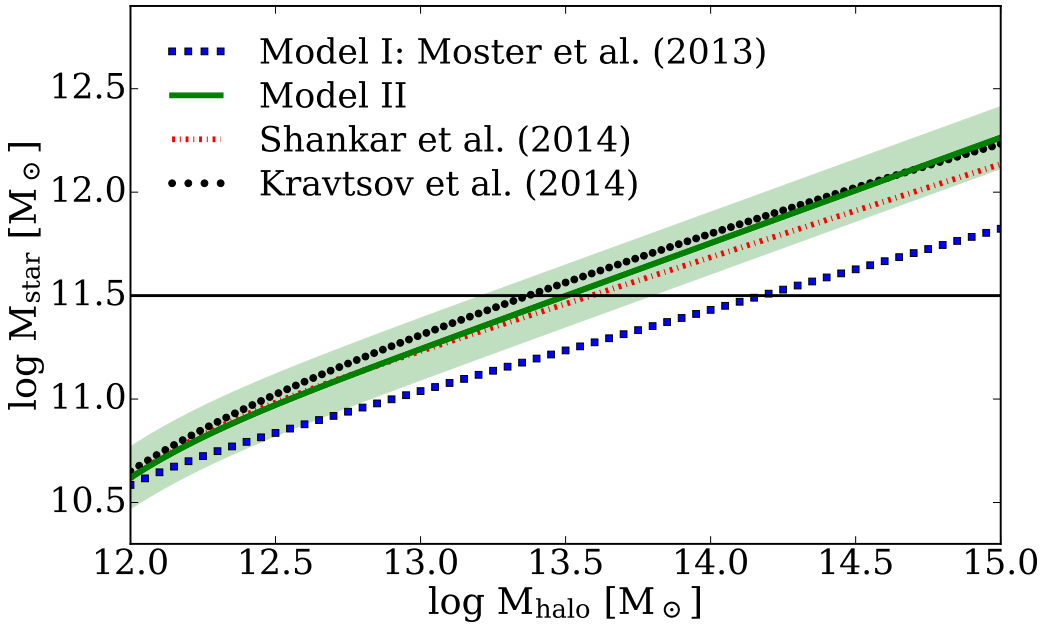



FIGURE 3.1: A comparison among estimates of the stellar mass-to-halo mass relations. In this chapter, I adopt the relations by Moster et al. (2013) as well a steeper version which matches the relations by Shankar et al. (2014b) and Kravtsov et al. (2014).

More specifically, to set more stringent constraints on the evolutionary patterns of local massive galaxies, I use an observationally-driven model that works as follows:

1. I extract all central haloes from the Bolshoi simulation and assign them a stellar mass, $M_{\rm star}$, using the $M_{\rm star}$-$M_{\rm halo}$ relation found in Section 2.2.2 (Equation 3.6 here with parameters presented in table 3.1).

2. I select those haloes where the central galaxy has log $(M_{\rm star}) > 11.5\ M_\odot$.

3. For each halo, I track its progenitors backwards in time until the putative formation epoch, $z_{\rm form}$= 2, 3 & 4.

4. I estimate their total baryonic mass and stellar mass at $z_{\rm form}$ from the global baryon fraction ($M_{baryonic} = 0.9 \times M_{\rm halo} \times f_b$; Crain et al. 2007) and from abundance matching relations, respectively.

5. I finally compare the estimated baryonic mass to their descendent galaxy's stellar mass at $z = 0$.

### 3.2.1 Selecting descendant galaxies at z=0

Techniques such as abundance matching allow us to connect galaxies to their dark matter haloes in a statistical sense. Abundance matching works by matching the cumulative number densities of dark matter haloes to the observed number densities of galaxies. This is a powerful technique in predicting the mean stellar content of dark matter haloes, especially for massive, central galaxies with $M_{\rm star} > 2 \times 10^{11} M_{\odot}$, where the scatter in stellar mass reduces to $\leq 0.15 dex$, and the dispersion in assembly histories due to, e.g., environment, age spread, specific star formation history, becomes less important (Shankar et al., 2014a; Gu et al., 2016; Clauwens et al., 2016). In this chapter, I use the parametrizations of the stellar mass to halo mass relation of Moster et al. (2013):

$$\frac{M_{\rm star}}{M_{\rm halo}} = 2N \left[ \left( \frac{M_{\rm halo}}{M_1} \right)^{-\beta} + \left( \frac{M_{\rm halo}}{M_1} \right)^{\gamma} \right]^{-1}, \tag{3.1}$$

$$\tag{3.2}$$

$$\log M_1(z) = M_{10} + M_{11}(1-a) = M_{10} + M_{11} \frac{z}{z+1}, \tag{3.3}$$

$$N(z) = N_{10} + N_{11}(1-a) = N_{10} + N_{11} \frac{z}{z+1}, \tag{3.4}$$

$$\beta(z) = \beta_{10} + \beta_{11}(1-a) = \beta_{10} + \beta_{11} \frac{z}{z+1}, \tag{3.5}$$

$$\gamma(z) = \gamma_{10} + \gamma_{11}(1-a) = \gamma_{10} + \gamma_{11} \frac{z}{z+1}. \tag{3.6}$$

where $M_{10}$, $M_{11}$, $N_{10}$, $N_{11}$, $\beta_{10}$, $\beta_{11}$, $\gamma_{10}$ and $\gamma_{11}$ are constants.

One of the main sources of systematic uncertainties in the high mass end of the $M_{\rm star}$-$M_{\rm halo}$ relation comes about from the exact shape of the stellar mass function. It has recently been shown that the high mass end of the stellar mass function has significantly higher number densities than earlier measurements (Bernardi et al., 2013; D'Souza et al.,

TABLE 3.1: The parameters used in equation 1. The top portion gives the $z = 0$ parameters of the $M_{\rm star}$-$M_{\rm halo}$ relation for both the original Moster et al. (2013) relation as well as the steeper version from Shankar et al. (2014b). The bottom portion gives the redshift dependence of the $M_{\rm star}$-$M_{\rm halo}$ relation.

| | $M_{10}$ | $N_{10}$ | $\beta_{10}$ | $\gamma_{10}$ |
|---|---|---|---|---|
| Model I | 11.59 | 0.0351 | 1.376 | 0.608 |
| Model II | 11.70 | 0.0380 | 1.25 | 0.490 |

| | $M_{11}$ | $N_{11}$ | $\beta_{11}$ | $\gamma_{11}$ |
|---|---|---|---|---|
| | 1.195 | -0.0247 | -0.826 | 0.329 |

2015; Bernardi et al., 2016a; Thanjavur et al., 2016). Bernardi et al. (2016a) in particular, have shown that possible systematics in photometry are now of the order $\sim 0.1dex$. Recent results by Kravtsov et al. (2014) and Shankar et al. (2014b), based on the new stellar mass function of Bernardi et al. (2013), coupled with direct measurements of the stellar masses and host halo masses of individual brightest group and cluster galaxies, conclude that the mean stellar mass of massive central galaxies is systematically a factor of ∼3-4 higher at fixed halo mass than previously estimated by, e.g., Behroozi et al. (2013b) and Moster et al. (2013). In what follows, to bracket the possible residual systematics in the $M_{\rm star}$-$M_{\rm halo}$ relation, I will adopt Equation 3.6 with *both* the original parameters found by Moster et al. (2013, hereafter model I) as well as with updated parameters to match the results of Kravtsov et al. (2014) and Shankar et al. (2014b) in the stellar mass range of interest in this thesis (hereafter model II; see table 1 for the new parameters).

Figure 3.1 shows a comparison between models I, II and the latest relations by Shankar et al. (2014b) and Kravtsov et al. (2014). I choose to specifically consider galaxies with $M_{\rm star} > 3 \times 10^{11}\ M_{\odot}$ as this is the threshold in stellar mass where there is most disagreement in the assembly history among different galaxy evolutionary models (e.g., Bernardi et al., 2016b). This stellar mass cut of $\log(M_{\rm star}) > 11.5\ M_{\odot}$ is shown as a horizontal line in Figure 3.1.

When applying the $M_{\rm star}$-$M_{\rm halo}$ relation to high redshift progenitor haloes, I keep the original redshift dependence of Moster et al. (2013) for both models I and II. The redshift dependants of the $M_{\rm star}$-$M_{\rm halo}$ relation using Model I is presented in Figure 3.2. Maintaining the Moster et al. (2013) formalism has the advantage to directly extend the abundance matching to $z > 0.5 - 1$ a redshift regime beyond the one probed by Kravtsov et al. (2014) and Shankar et al. (2014b). It is important to note that at higher redshift, I am probing haloes with $M_{\rm halo} \sim 10^{12}\ M_{\odot}$, which sit around the knee of the

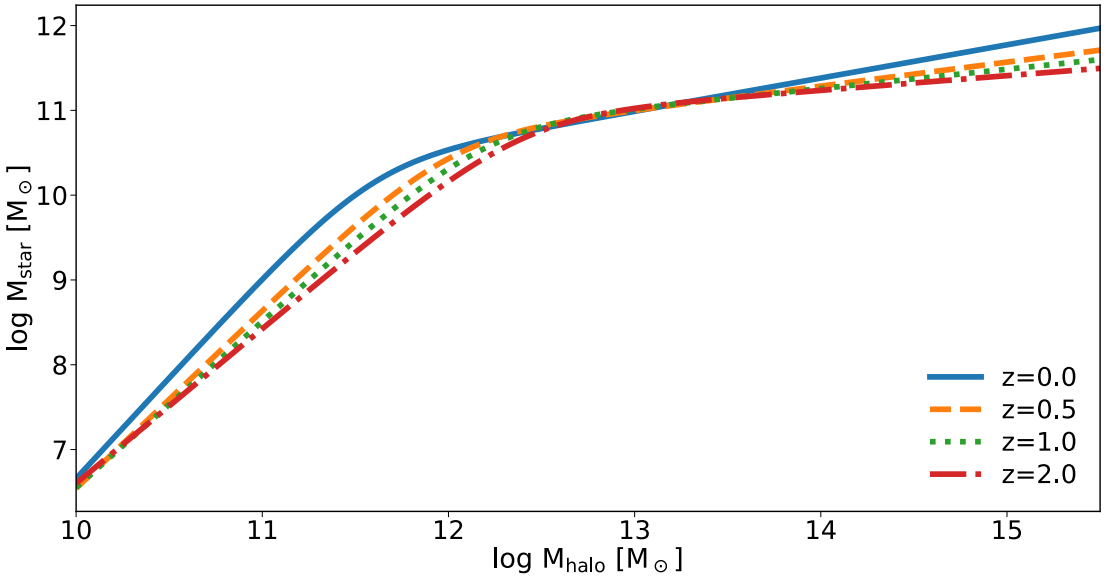



FIGURE 3.2: The redshift dependence of the $M_{\rm star}$-$M_{\rm halo}$ relation using the parameters of Model I from Table 3.1.

stellar mass function and are thus significantly less prone to the above mentioned systematic uncertainties (for example, in stellar masses) characterizing the high-mass end of the $M_{\rm star}$-$M_{\rm halo}$ relation. I anyway stress that the main conclusions do not depend on the exact redshift dependencies in Moster et al. (2013). I also note that Behroozi et al. (2013b) are consistent with Moster et al. (2013) within the uncertainties.

### 3.2.2 Selecting their progenitors

The progenitor of a massive galaxy is usually considered to follow the so-called "main progenitor" halo. This is defined as the most massive progenitor of a parent halo (Jiang & van den Bosch, 2014). The main progenitor branch is therefore a chain of haloes constructed by finding the most massive progenitor of the previous main progenitor, starting at $z = 0$ and working backwards in time. However, this definition does not necessarily imply that the main progenitor is the most massive progenitor at all times. This is because of the statistic nature of dark matter halo growth; sometimes the initial mass of the main progenitor can be lower than the initial mass a halo that it mergers with if the main progenitor has gained more mass over time (this will be discussed in more detail in the following sections). In fact, studies such as those by Lapi et al. (2013), which are based on the excursion-set formalism (Bond et al., 1991), have pointed out that possibly the most massive progenitor haloes at $z_{\rm form}$ are more relevant than the main progenitor for the evolution of today's central, massive galaxies. This is an important distinction to make as a more massive halo has more potential to form a more massive galaxy in a single burst.

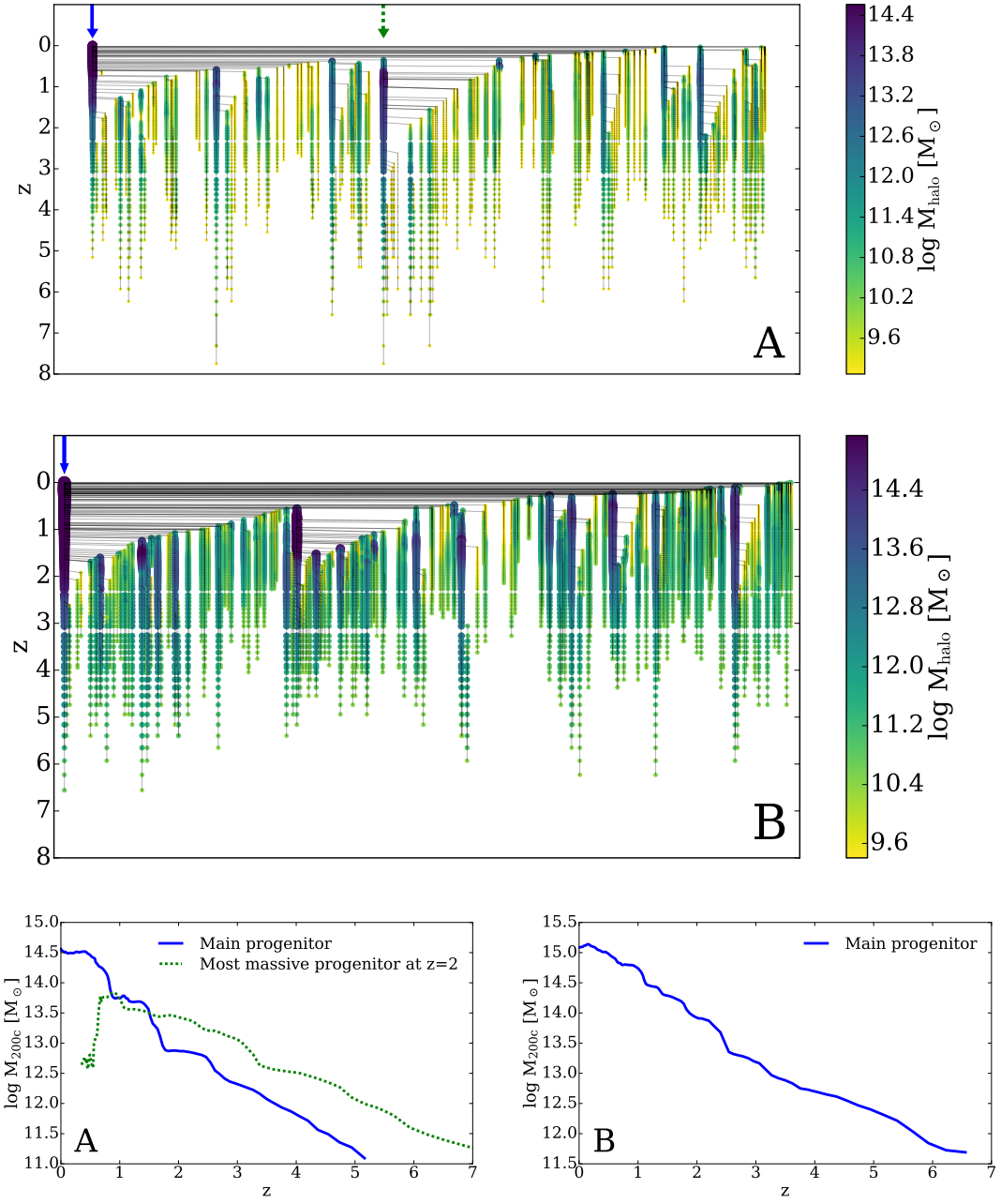



FIGURE 3.3: The top panels are visualizations of two dark matter merger trees and bottom panels show the mass evolutions of the main progenitor (solid) and the most massive progenitor at $z = 2$ (dashed line). The panels labelled A show the evolution of a halo where the main progenitor halo is not the most massive halo at high redshift. The panels labelled B instead shows a more idealised case where the main progenitor is the most massive at all the redshifts I consider.

To visualise the difference between the main progenitor and most massive progenitor branches, Figure 3.3 shows the merger trees of two representative dark matter haloes in the Bolshoi simulation. The panels labelled "A" show a merger tree whereby the main progenitor is *not* the most massive progenitor at all redshifts. The panels labelled "B" instead show a more idealised case where the main progenitor is the most massive

progenitor at all epochs. The top two panels show a visualisation of the merger trees with redshift along the $y$-axis and the branches of the tree separated out along the $x$-axis. The main progenitor branch is positioned at the far left-end of the plots and is indicated with a blue, solid arrow. I also indicate the branch which contains the most massive progenitor at $z = 2$ with a green, dashed arrow. The horizontal lines show merging events between the branches. The size of the circles is proportional to the mass of the progenitors at that redshift, as encoded in the colour legend. The bottom panels then show the mass evolution of the main progenitor and of the most massive $z = 2$ progenitor.

The relevant question one needs to answer is how frequent the main progenitor remains indeed the most massive progenitor at all times, as in the panels labelled "A" of Figure 3.3. To this purpose, I carefully analyse the merger trees of each halo in the sample[1] of galaxies with log $(M_{\rm star}) > 11.5\ M_{\odot}$. In Figure 3.4, I show the mass functions of the main progenitors (solid) and most massive progenitors (dashed) at the labelled redshifts. From this figure, it is clear that the choice in definition of progenitors has little impact on the mean evolution in halo mass and at most is only relevant for the low mass wings of the distributions at $z > 3$. This is because only $\lesssim 25\%$ of the haloes follow the "most massive" progenitor track. I checked that this conclusion still holds even if the analysis is restricted to only the most massive haloes in the sample with log $(M_{\rm halo}) > 14\ M_{\odot}$ where the effect could be most prominent (Lapi et al., 2013). In the following, I use the main progenitor as the reference, though I also show results using the most massive progenitor, where relevant.

### 3.2.3 Ages of massive, early type galaxies: selecting a formation epoch.

From stellar population synthesis modelling it is possible to estimate the mass-weighted age of the stars within a galaxy. The general method to constrain the age of a galaxy is to fit the galaxy's spectra with either a single or a combination of synthetic stellar populations with varying star formation rates, for a given initial mass function (Vincoletto et al., 2012; McDermid et al., 2015; Mendel et al., 2015; Citro et al., 2016). For the galaxies of interest in this thesis ($M_{\rm star} > 3 \times 10^{11}\ M_{\odot}$), the majority of the stars form at or above $z_{\rm form} = 2$, with less massive galaxies having, on average, younger stellar populations (Thomas et al., 2005). This is the so-called 'top-down' mass assembly scenario for massive ETGs.

I explore the consequences of a formation redshift, $z_{\rm form} = 2-4$. For reference, McDermid et al. (2015) find that $\geq 50\%(90\%)$ of the stellar mass is older than $z = 3(2)$ for the stellar

[1] I verified that most massive progenitor distributions are very similar in the MultiDark Planck 2 simulation (Klypin et al., 2016) which has a larger volume and different cosmology than Bolshoi.

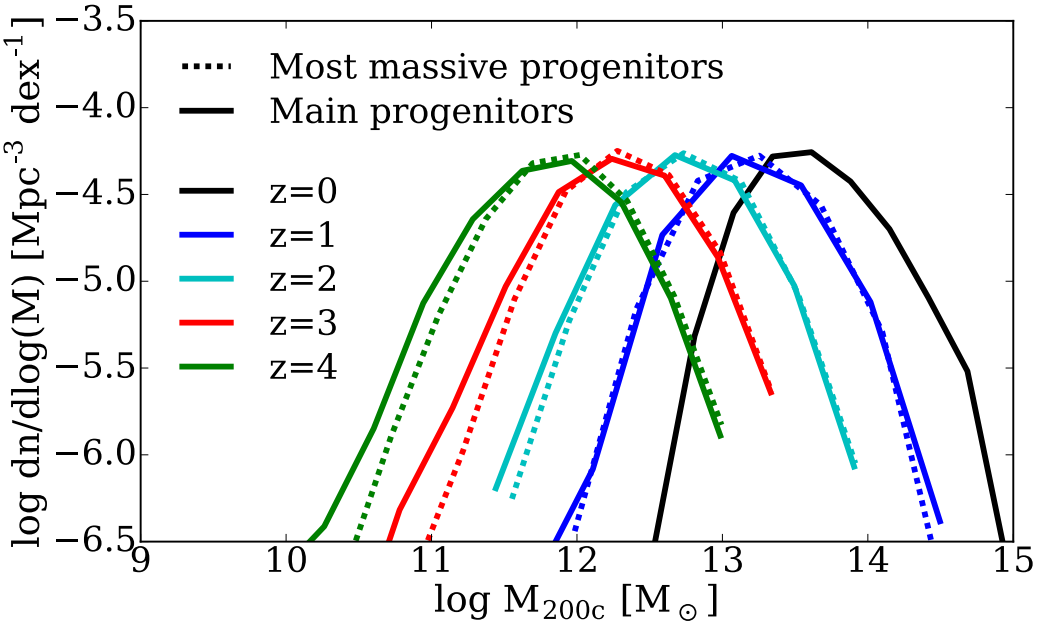



FIGURE 3.4: The mass functions of the progenitor haloes that contain galaxies with log $(M_{\rm star}) > 11.5\ M_\odot$ at $z = 0$. The solid and dashed lines show the main progenitor and most massive progenitor mass functions at different redshifts, as labelled.

mass I consider in this chapter[2]. At these formation epochs, I compare the amount of baryons in the progenitor haloes to the stellar content of the descendants.

## 3.3 Results

### 3.3.1 Constraints on the assembly scenario

Having constructed the $z = 0$ dark matter halo catalogue and traced their progenitors to $z_{\rm form}$, I am now in a position to compare the baryonic content between progenitor haloes at $z_{\rm form}$ and descendent haloes at $z = 0$. The total baryonic mass is computed using the cluster baryon fraction extracted from numerical simulations. Crain et al. (2007), in particular, find that the baryon fraction inside the virial radius of dark matter haloes is 90% of the cosmic mean fraction, independent of halo mass and redshift

$$M_{baryon} = 0.9 \times M_{halo} \times f_b \tag{3.7}$$

[2]Note that strictly speaking the stellar masses in McDermid et al. (2015) are dynamical masses from jeans modelling. Cappellari et al. (2013) discuss that these masses are closer to those measured assuming a Salpeter IMF. This would imply that the galaxies I have selected for this chapter have higher stellar mass at fixed halo mass and hence I could be, if anything, underestimating their age.

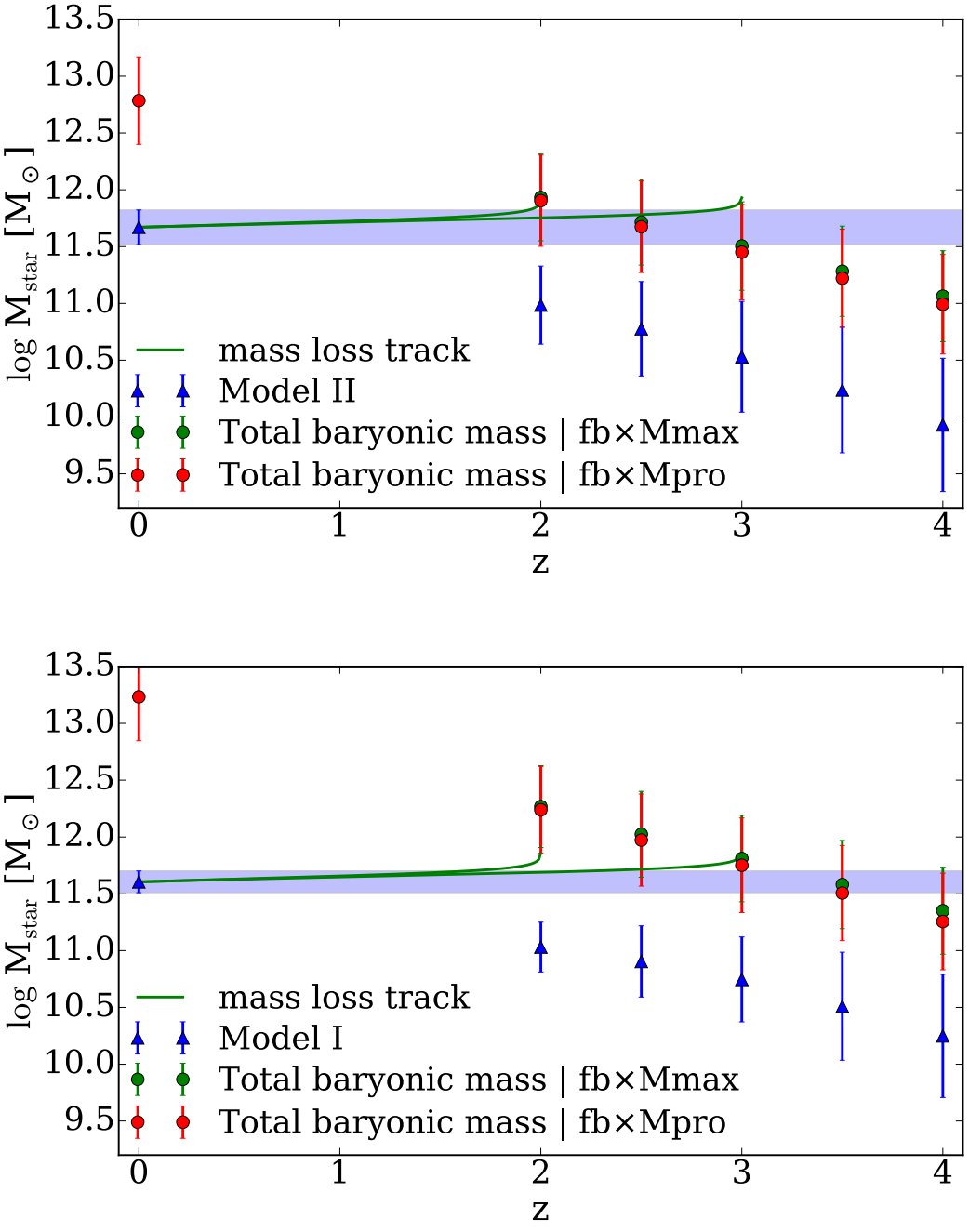



FIGURE 3.5: A plot showing the predicted mean stellar mass of galaxies with log ($M_{\rm star}$) > 11.5 $M_\odot$ at $z = 0$ (blue triangle at $z = 0$) using the original Moster et al. (2013) $M_{\rm star}$-$M_{\rm halo}$ relation (bottom) and a modified version to match the latest relation from (Shankar et al., 2014b, top) (models I and II, respectively). The one sigma range of this $z = 0$ point has been extended to all redshifts (blue band) so the z=0 point can be compared to points at higher redshift. I show the mean and $1\sigma$ of the total baryonic mass associated with the the main progenitor haloes (red circles) and most massive progenitors (green circles). The decline in $M_{\rm star}$ after star formation has stopped in due to older stars dying (Behroozi et al., 2013b). For reference, I also show the stellar mass estimates of galaxies in the respective haloes using the given $M_{\rm star}$-$M_{\rm halo}$ relations (Model I or Model 2; blue triangles).

where $f_b$ in equation 3.7 is the global mean ratio between the baryon and dark matter density[3]. I show the implied total baryon masses in Figure 3.5 for the main progenitor haloes (red circles) and most massive progenitor (green circles) at the putative formation epochs $z_{\rm form}$= 2 to 4 in steps of 0.5. The blue triangles show instead the mean stellar mass computed along the main progenitors adopting model I (the $M_{\rm star}$-$M_{\rm halo}$ relation using the original parameters found by (Moster et al., 2013, botton panel) and Model II (the updated parameters, top panel). Additionally, for reference, I show the mean mass evolutionary tracks expected in a passive evolutionary model, taking into account the ageing stellar population, following equation 14 by Behroozi et al. (2013b) who also use a Chabier IMF and Bruzual & Charlot (2003) stellar evolution tracks. The green solid lines show two evolutionary tracks that are needed to reproduce the $z = 0$ stellar mass distribution starting at $z_{\rm form}$= 2 and 3.

The first point to make regarding Figure 3.5 is that the local stellar mass of massive ETG's is comparable to, if not greater than, the total baryonic mass contained in their main progenitor haloes at $z_{\rm form}$. This is especially evident in the top panel, which utilise the steeper high mass slope in model II to map lower mass haloes at fixed stellar mass (Figure 3.1), while it only becomes evident at $z > 3$ in model I. Thus the updated abundance matching relations would imply that if a galaxy is formed via a strict monolithic collapse[4] at the epoch of formation, star formation should have been extremely efficient, if not 100%, to account for the (high) stellar mass content observed today in the descendant haloes. Moreover, this scenario would preclude any substantial loss of baryons due to stellar winds and/or quasar mode feedback during this rapid star formation phase. Note that this result is not changed and, in fact, possibly strengthened by assuming a Salpeter IMF, which some authors (e.g., Cappellari et al., 2013) have suggested to be more representative of massive galaxies. A Salpeter IMF would imply a higher stellar mass at $z = 0$ for the same set of host haloes of interest here ($M_{\rm halo} > 3 \times 10^{12}\ M_{\odot}$), worsening the tension with the available baryons at $z_{\rm form}$.

### 3.3.2 Are today's central, massive galaxies just outliers at the epoch of formation?

Figure 3.5 also shows that $z = 0$ stellar masses are a factor of at least five times greater than the typical stellar mass of the progenitor galaxies at $z_{\rm form}$ as predicted by abundance matching (blue triangles). There are two possible conclusions from this finding. If the

[3]This will be verified in Section 5.3.2.2.

[4]In the monolithic collapse scenario, I assume that the dark matter still grows hierarchically via mergers, as indicated from N-body simulations. However I assume that that mergers between galaxies are inefficient (because, e.g., the merging timescale is long or because mergers grow the intercluster light rather than the central galaxy due to high amounts of ram pressure stripping) and hence the central galaxy maintains its initial mass.

progenitors of massive galaxies at $z_{\rm form}$ are representative of galaxies at that halo mass, then clearly a later mass growth is needed to match the stellar mass of the descendants to their local counterparts. On the other hand, the progenitors of very massive galaxies might not be representative of the general population of galaxies at $z_{\rm form}$ at fixed halo mass. For instance, they could be extreme outliers with a stellar mass much greater than what is predicted by abundance matching relations. This could arise in strictly monolithic models where stars are formed in an extremely efficient and fast mode around $z_{\rm form}$.

To probe the latter possibility, I proceed as follows. In the top panel of Figure 3.6, I show as black solid and blue dashed lines respectively, the full $z = 0$ cumulative mass functions of all massive dark matter haloes with $M_{\rm halo} > 3 \times 10^{12}\ M_{\odot}$, and of those hosting the stellar mass-selected sample ($\log(M_{\rm star}) > 11.5\ M_{\odot}$). I then trace the main progenitors of the selected galaxies back to $z = 3$. In the middle panel of Figure 3.6, I show the differential mass functions of the main progenitors of all massive $z = 0$ haloes (solid black) and of those in the stellar mass selection (dashed blue). Here, it is evident that the mean mass of these progenitor haloes at $z = 3$ is $M_{\rm halo} \sim 2 \times 10^{12}\ M_{\odot}$, consistent with Figure 3.4, which I highlight as a blue band in the middle panel. I select those haloes at $z = 3$ which are between $12.25 < \log(M_{\rm halo}) < 12.45$ and follow them forward to $z = 0$. The black solid line in the bottom panel of Figure 3.6 is the cumulative number density at $z = 0$ of haloes which have a $z = 3$ progenitor mass $M_{\rm halo} \sim 2 \times 10^{12}\ M_{\odot}$. The blue dashed line instead are haloes which are in the same mass range at z=3 but also become massive enough to enter the $z = 0$ stellar mass selection.

By comparing the cumulative mass functions, I find that on average, only $\sim 30\%$ of the haloes with mean host mass $M_{\rm halo} \sim 2 \times 10^{12}\ M_{\odot}$ at $z = 3$ will host galaxies with $\log(M_{\rm star}) > 11.5\ M_{\odot}$ at $z = 0$. The selected brightest cluster galaxies are thus only a relatively minor fraction of the galaxies residing in $M_{\rm halo} \sim 2 \times 10^{12}\ M_{\odot}$ at $z = 3$. I thus conclude that the progenitors of massive galaxies could indeed be outliers with respect to the general population of galaxies at $z_{\rm form} = 3$ and with $M_{\rm halo} \sim 2 \times 10^{12}\ M_{\odot}$ (for instance, these galaxies might not abide by the $M_{\rm star}$-$M_{\rm halo}$ or $M_{\rm star}$-$sSFR$ relations at $z = 3$). However, they represent the *majority* of galaxies that will end up as centrals in haloes with $M_{\rm halo} > 5 \times 10^{13}\ M_{\odot}$, within the mass scale of massive groups and clusters.

## 3.4 Discussion

From the previous sections I conclude that very massive central galaxies are either extreme outliers with respect to their counterparts at $z_{\rm form}$ or, alternatively, have assembled most of their final stellar mass at $z < z_{\rm form}$(where $z_{\rm form} \sim 2 - 4$). Here, I probe

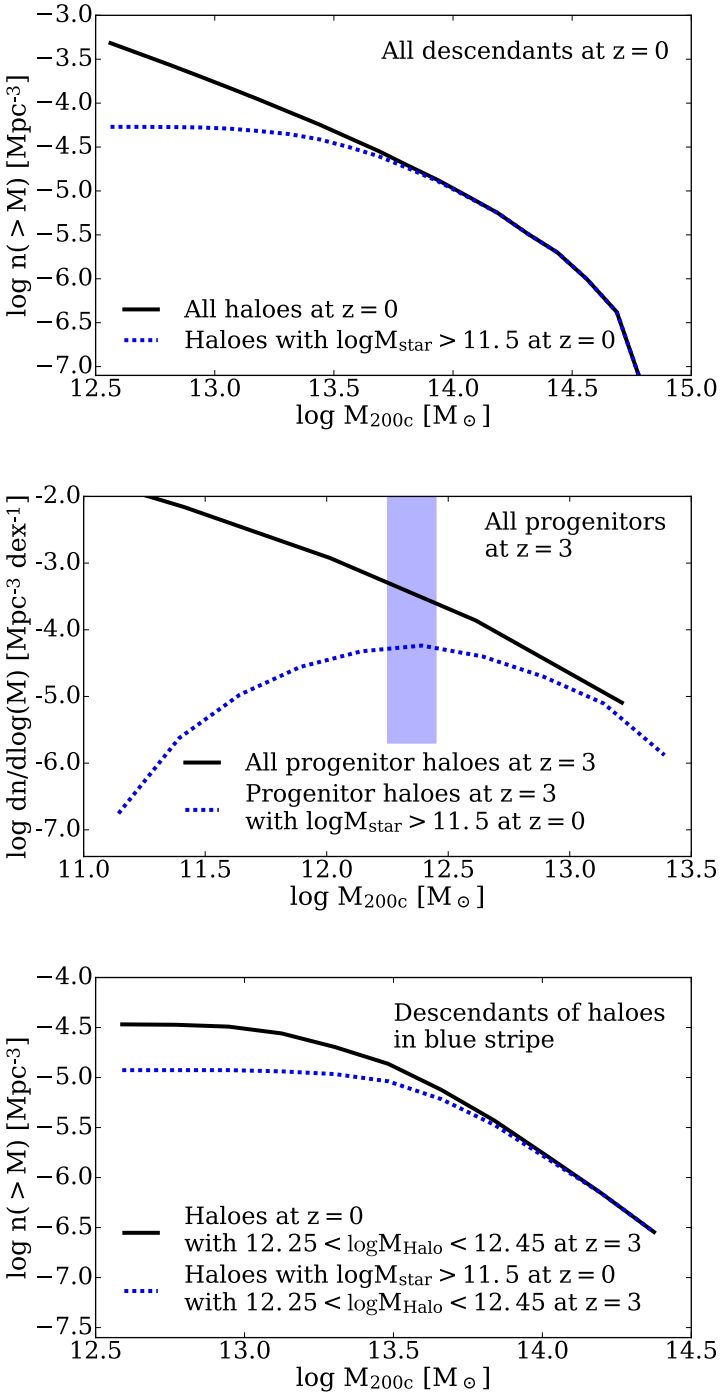



FIGURE 3.6: Top: the integrated number density of dark matter haloes in the Bolshoi simulation. The black line shows the distribution for all haloes in the simulation and blue shows those which are included in the stellar mass selection (log ($M_{\rm star}$) > 11.5 $M_\odot$). Middle: the differential mass function of dark matter haloes at z=3 which survive to z=0. The black line shows the total halo distribution and blue are those haloes which enter the stellar mass selection. Bottom: I apply a halo mass selection at z=3 (highlighted in the middle panel) and follow those haloes forward to z=0. The black shows the distribution of haloes when selected from the total distribution and blue when selected from those which are included in the stellar mass cut.

the relative roles of star formation and mergers in growing massive galaxies in a late assembly scenario. For the former, I utilise empirical estimates of the star formation rates of massive galaxies as functions of redshift and stellar mass to evaluate if it is sufficient to grow the galaxy up to the stellar mass we observe today. For the latter, I utilise the more sophisticated semi-empirical model described in Section 2.3 to evaluate the effectiveness of mergers in evolving the galaxy's stellar mass and size.

### 3.4.1 Can massive galaxies grow solely through in-situ star formation?

In this section, I utilise empirically-informed models to assess whether log $(M_{\rm star}) > 11.5\ M_{\odot}$ galaxies at $z = 0$ could have grown to their final stellar mass mostly through in-situ star formation, without the need for numerous mergers. I start by assuming that the progenitor galaxies are lying on the $M_{\rm star}$-$M_{\rm halo}$ relation and the specific star formation rate (sSFR)-$M_{\rm star}$ relation of typical main sequence galaxies at $z =$$z_{\rm form}$. I take each of the $z = 3$ progenitor haloes and assign to them a stellar mass using the $M_{\rm star}$-$M_{\rm halo}$ relation as described in Section 3.2.2. For each of the galaxies, I evolve the stellar mass via redshift and mass dependent star formation rates but also accounting for the stellar mass loss of the evolving stellar population assuming a Chabrier (2003) IMF. Specifically, I use the redshift-dependant, empirical star formation rates by Peeples & Somerville (2013) and Tomczak et al. (2016), and use the Behroozi et al. (2013b) prescription (their Equation 14) for the stellar mass loss. I assume that the galaxy can efficiently form stars, at least up to the maximum baryonic content, assuming that the cold gas reservoir can be replenished via, e.g., cold flows (Dekel et al., 2009).

Figure 3.7 shows the mean evolutionary track of the sub-sample of galaxies with log $(M_{\rm star}) > 11.5\ M_{\odot}$ at $z = 0$. The blue band represents the systematic uncertainty in the $M_{\rm star}$-$M_{\rm halo}$ relation (model I and II as described in Section 3.2.2). I find that assuming these galaxies remain on the star forming main sequence until $z = 0$, the Peeples & Somerville (2013) star formation rate (top panel) can fully account for the observed mass measured in the local universe (dotted lines). The Tomczak et al. (2016) star formation (middle panel) predicts a final stellar mass which is a factor of two to three lower. However, systems grown via a very prolonged star formation episode would be inconsistent with the observed ages (and colours) of very massive, central galaxies at $z = 0$. As discussed in Section 3.2.3, McDermid et al. (2015) claim that from their spectral fitting, 90% of stellar mass in galaxies with log $(M_{\rm star}) > 11.3 - 11.5\ M_{\odot}$ was formed at $z > 2$ with a slight dependence on environment. Also, as mentioned in section 3.1, very massive galaxies have enhanced alpha element abundances relative to iron. In a closed box model, values of $[\alpha/Fe] \sim 0.17$, which are consistent with those measured in BCGs (eg, Oliva-Altamirano et al., 2015), imply the length of star formation, $\Delta t$ is $< 0.5 - 1$ Gyrs,

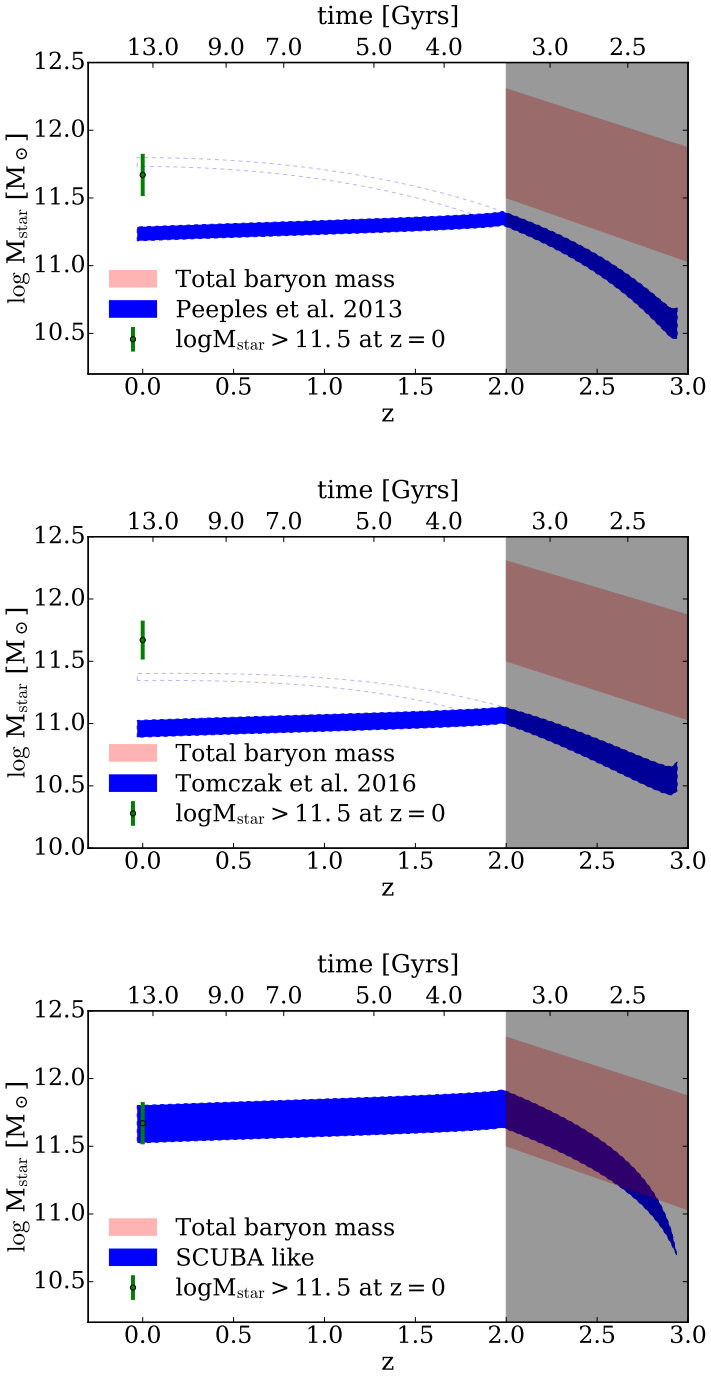



FIGURE 3.7: The mean star formation track of galaxies selected with log ($M_{\rm star}$) < 11.5 $M_\odot$ at $z = 0$ using the mass-dependant empirical star formation rates of Peeples & Somerville (2013, top panel) and Tomczak et al. (2016, middle panel). The y-axis shows the mean stellar mass of all galaxies in the sample. The band shows the systematic uncertainty in the $M_{\rm star}$-$M_{\rm halo}$ relation. The unfilled band shows the full evolutionary track from $z = 3$ to $z = 0$ and solid band shows the effects if the star formation is quenched at $z = 2$. The bottom panel instead shows a band bracketing the mean evolutionary path assuming a constant star formation of 500 and 1000 $M_\odot/yr$. For reference, I also show the total baryonic mass in the progenitor haloes as a red band between $z = 2$ and 3.

according to the approximation

$$[\alpha/Fe] \sim \frac{1}{5} - \frac{1}{6} \times \Delta t(\mathrm{Gyrs})\,, \tag{3.8}$$

from Thomas et al. (2005) and Citro et al. (2016), which is in broad agreement with other studies of chemical evolution of massive galaxies (e.g., Granato et al., 2004; Conroy et al., 2014). If star formation is longer than $\Delta t \sim 1Gyr$ then the alpha element enhancement will be rapidly washed out by type Ia supernovae. The filled blue bands in Figure 3.7 show the predicted stellar mass evolution in the hypothesis that massive galaxies form by $z \simeq 3$ and quench by $z \simeq 2$ satisfying the conditions that the burst of star formation is limited to $\Delta t \leqslant 1$ Gyrs and that most of the stars are formed by $z \simeq 2$. It can be seen that, in the assumption that no new star formation takes place at $z < 2$, the resultant stellar mass is at most log $(M_{\rm star}) \lesssim 11.2$ which is a factor of at least three less that what is measured at $z = 0$.

Alternatively, massive galaxies could be, as mentioned earlier, extreme outliers in both the $M_{\rm star}$-$M_{\rm halo}$ and/or sSFR-$M_{\rm star}$ relations. For example, if the progenitors were closer to SCUBA/ULIRGs with star formation rates up to 500-1000 $M_{\odot}/yr$ between $z = 3$ and $= 2$ (which is $\sim 1Gyr$), they could easily reach the stellar mass measured at z=0. This is illustrated in the bottom panel of Figure 3.7 which shows with a blue band the mean mass growth of galaxies assuming a constant star formation of 500 and $1000M_{\odot}/yr$. To maintain such a high star formation rates, progenitor galaxies would need to turn about all of their initial baryonic content in the host dark matter halo (red regions) into stars. Additionally, this scenario would not allow for the observed, significant mass loss by stellar and/or AGN winds seen in a number of ULIRGs (e.g., Smail et al., 2003; Swinbank et al., 2005).

### 3.4.2 Can mergers drive mass evolution of massive galaxies?

I now turn to explore the possibility that (mainly dry, minor) mergers are the main driver behind the mass and size evolution of massive, central galaxies. To this purpose, I utilise the more sophisticated, state of the art semi-empirical model (SEM). A full description of the SEM is given by Shankar et al. (2014a, 2015) and in Section 2.3 but I also provide a brief overview here.

The SEM is constructed on top of dark matter merger trees extracted from the Millennium simulation (Springel, 2005a) and analytic trees constructed using the Parkinson et al. (2008) algorithm. At the formation redshift, the main progenitors haloes are 'populated' with star forming, disc galaxies which, by construction, follow empirical relations for stellar mass (Moster et al., 2013), gas fraction (Stewart et al., 2009), disc radius (Shen

et al., 2003), and star formation rate (Peeples & Somerville, 2013). At each timestep, If a main progenitor galaxy's stellar mass is too low with respect to a fiducial abundance matching relation, it's stellar mass and gas fraction is re-initialised using these empirical relations thus bypassing the need to model the full, complex and still unclear aspects of galaxy formation, such as cooling and feedback. However, because there is no clear way to re-initialise the galaxy's size, adding substantial amounts of stellar mass will bias the scatter in the $M_{\rm star}$-$M_{\rm halo}$ relation and the mean $M_{\rm star}-R_e$ relation, two key outputs from the model which can be directly compared to observations. Therefore, it is important that the re-initialisation is kept to a minimum and thus processes such as star formation are still included.

When sub-haloes infall into the halo hosting the massive galaxy, I assign to them a satellite galaxy with stellar mass given by abundance matching relations at the redshift of infall. The structural properties of this new satellite are equal to a previously simulated, random central galaxy extracted from the model with equal stellar mass at the redshift of infall. The new satellite galaxy is allowed to orbit for a dynamical friction timescale, calculated at $z_{\rm inf}$using the analytic recipe by McCavana et al. (2012). Over this time, the satellite can grow in stellar mass and size according to its available gas and star formation rate at infall. If a merger between the central and satellite galaxy occurs, the stellar mass and gas mass of the satellite are added to the bulge and disc of the central galaxy, respectively. The new radius of the bulge is calculated by conserving the sum of the binding energies and the mutual orbital energy of the two merging galaxies (Cole et al., 2000).

Figure 3.8 shows the mean mass evolution of the most massive galaxies evolved using my SEM. The shaded region represents the $1\sigma$ dispersion in the mean stellar mass at any redshift. I find that the merger-driven SEM is capable of reproducing the median mass evolution of the most massive galaxies. This result confirms, and extends, what was found by Shankar et al. (2015) from the evolution of brightest cluster galaxies from $z \sim 1$ to 0. This result is also in agreement with previous works mainly based on high-resolution N-body simulations which showed that the history of the centre of clusters is highly affected by frequent mergers. For example, Gao et al. (2004) discussed that a typical massive brightest cluster galaxy should have undergone a significant number of merging events even at $z < 1$. More recently this has been further discussed and confirmed by Laporte & White (2015). As shown in Figure 3.8, the results from the semi-empirical model are also in agreement with the full, cosmological hydro-simulations of Torrey et al. (2015).

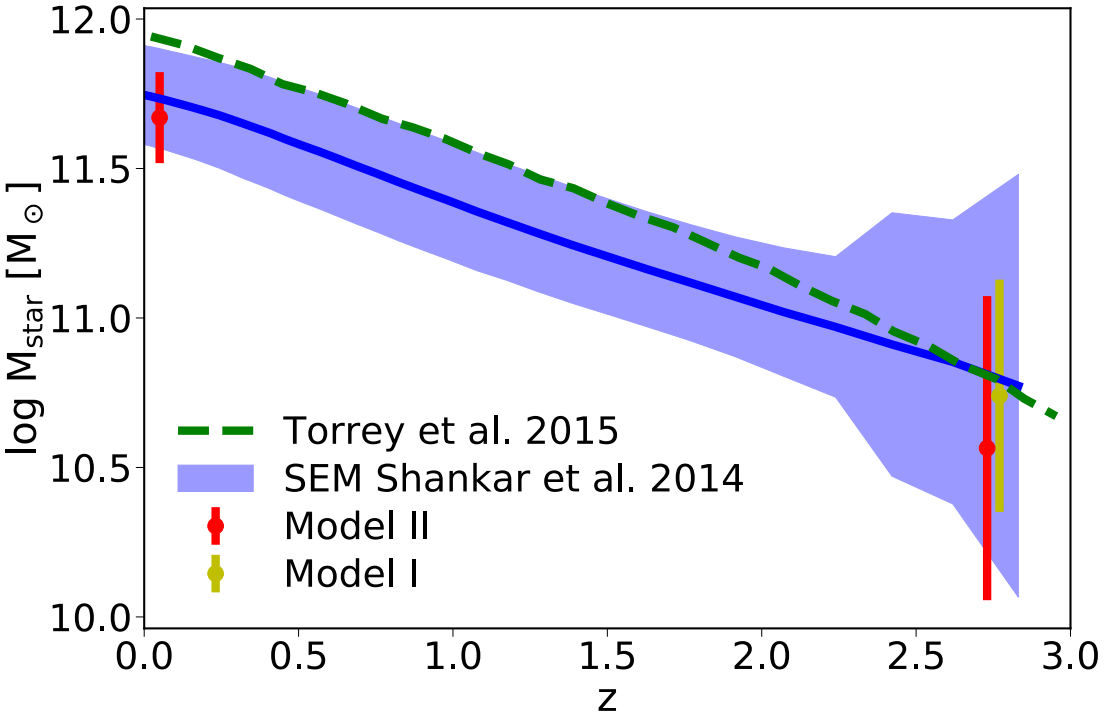



FIGURE 3.8: The evolution of the mean stellar mass of a galaxies with *log* $(M_{\rm star}) > 11.5$ at $z = 0$ evolved using the SEM in Shankar et al. (2014a). The shaded region shows the statistic dispersion in the galaxies' evolutionary histories. The error bars show the predictions from abundance matching for galaxies with log $(M_{\rm star}) > 11.5$ $M_\odot$ where model I and II are shown in yellow and red respectively. For reference, I also show the the mean evolution in the mass of the most massive galaxies in the illustris simulation (Torrey et al., 2015).

### 3.4.3 Can mergers drive size evolution of massive galaxies?

Additional hints come from the size evolution of the central galaxies. Figure 3.9 shows a comparison between the size evolution of massive ETGs evolved using my SEM to the observations of Bernardi et al. (2014) in the local universe. I also plot the observed size of the putative progenitor at $z = 0.25 - 2.75$ in steps of 0.5 by using the stellar masses predicted by the SEM and using the size-mass relation for disc galaxies found by van der Wel et al. (2014). I show the sizes these galaxies would have if they were all spirals and ellipticals since, as discussed previously, all galaxies start as disk-like galaxies at high redshift and transition to ellipticals. As described in Section 2.3.3, galaxies in the SEM are grown in size by conserving the binding energies and the mutual orbital energies of central and satellite galaxies when they merge. The shaded region in Figure 3.9 shows the systematic uncertainty in the predicted mean size evolution caused by allowing for some stellar stripping at the level suggested by observations (see Cattaneo et al., 2011; Shankar et al., 2014a, for full details). I find that, irrespective of the exact level of (stellar) stripping, the SEM is fully capable of reproducing the observed mean size

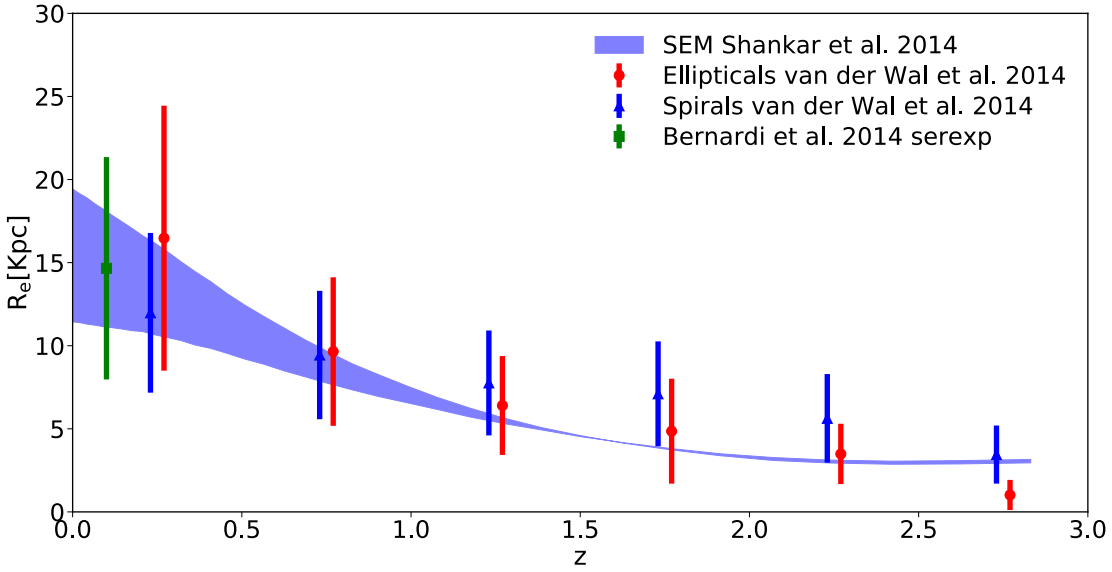



FIGURE 3.9: The median evolution of the effective radius of a galaxies with log $(M_{\rm star}) > 11.5$ at $z = 0$ evolved using the SEM in Shankar et al. (2014a). The shaded area represents the systematic uncertainty of the model when gas dispersion included.

evolution of the most massive galaxies, in line with the conclusions of Shankar et al. (2015) at $z < 1$.

### 3.4.4 Size growth from quasar mode feedback

The extreme assumption of a very efficient collapse in which the majority of initial baryons are converted into stars, as discussed in Section 3.4.1 and Section 1.6.2.1, would clearly not allow for any size growth from $z_{\rm form} > 2$ to $z = 0$. Even in the puffing up scenario proposed by Fan et al. (2008) and discussed in Section 3.1, a significant fraction of the baryons must be lost via stellar winds and/or quasar mode feedback at $z_{\rm form}$ for the galaxy to react quasi-adiabatically and expand. Analytic arguments by Fan et al. (2010), backed up by numerical simulations by Ragone-Figueroa & Granato (2011), have shown that the increase in effective radius is roughly proportional to the amount of mass lost. This would imply that to allow for a factor of at least three increase in mean size since $z \sim 3$, as observationally inferred by van der Wel et al. (2014), $\gtrsim 70\%$ of the total initial baryons must be expelled from the galaxy (Fan et al., 2010). A mass loss of $\gtrsim 70\%$ at $z > 2$ would place the residual baryonic mass in the progenitor halo significantly below the descendants' stellar mass at $z = 0$. Thus, even in an efficient quasar-feedback scenario, the progenitor-descendant evolutionary tracks would still require substantial late assembly of stellar mass via, e.g., mergers.

### 3.4.5 No size growth for very massive ETGs?

Interestingly, there is mounting observational evidence for very massive galaxies at the centre of high redshift clusters that sit already on the local size-mass relation of early-type galaxies (Strazzullo et al., 2013; Delaye et al., 2014; Newman et al., 2014). This might be in support of the hypothesis of very rapid and efficient bursts of star formation. However, even if massive galaxies are indeed true outliers with respect to the general population of central galaxies with $M_{\rm halo}\sim 10^{12} M_{\odot}$ at $z_{\rm form}\geq 2$, they would still require an initial star formation burst capable of converting $\approx 100\%$ of the initial baryons into stars. Thus, one clear observational prediction I can make is that the progenitors of massive galaxies should either be moderately massive and compact, or very massive and extended.

## 3.5 Conclusions

In this chapter, I have set tighter constrains on the assembly and evolution of massive, central galaxies. I utilise a catalogue of dark matter haloes created from the Bolshoi simulation. I populate these haloes with galaxies with a stellar mass given by the most recent rendition of the stellar mass-to-halo mass relation by Kravtsov et al. (2014) and Shankar et al. (2014b) at $z = 0$, and select haloes with log $(M_{\rm star}) > 11.5\ M_{\odot}$. I then trace host haloes back to their putative formation epoch, $z_{\rm form}= 2-4$, as inferred from the stellar ages of massive ETGs. At this epoch, I estimate the total mass in baryons within the host halo from the universal baryon fraction. I find that the stellar mass of the ETG in the local universe is comparable to, if not higher than, the total baryonic mass contained within the progenitor halo. From this comparison, I draw the following important conclusions.

1. In-situ formation: For massive galaxies with log $(M_{\rm star}) > 11.5\ M_{\odot}$ and log $(M_{\rm halo}) > 13.5\ M_{\odot}$ to fully assemble at the formation epoch, the efficiency of converting baryons into stars needs to be extremely high if not 100%. I also show that this formation scenario would lead to all ETGs being extreme outliers with respect to what is predicted by abundance matching at $z_{\rm form}$.

2. Size: Even when assuming an extremely efficient star formation at $z_{\rm form}$, the galaxy would not be allowed any size growth since the formation epoch. Even an in-situ expansion would in fact require a mass loss of $\geq 70\%$ of the initial baryon content to be sufficiently efficient. Thus, in a strictly monolithic scenario, progenitors of massive galaxies should already be extended systems at their formation

epoch. Measurements of the structure of massive galaxies in massive haloes with $\log\,(M_{\rm halo}) \sim 12.5\ M_{\odot}$ at $z \gtrsim 2$ will be critical to assess this possibility.

3. Late assembly: Star formation could contribute to the stellar mass growth of the progenitors of massive galaxies, but cannot explain their full evolution. I show through state-of-the-art, cosmological, semi-empirical models that massive galaxies could have indeed assembled most of their final mass via late mergers and be consistent with available data on their size evolution. It remains to be seen the impact of mergers on other (tight) galaxy scaling relations involving velocity dispersion (e.g., Bernardi et al., 2011a,b; Shankar et al., 2016, and references therein).

More secure and statistically relevant measurements of the stellar mass and structure of high redshift brightest cluster galaxies will be of key relevance to discern between merger scenarios and extremely efficient starbursts events.

# Chapter 4

# Probing the stellar mass to halo mass relation at $0.5 \leq z \lesssim 1.2$ using the clustering of satellite galaxies.

## 4.1 Introduction

In this Chapter, I extend the methodology presented at the start of Chapter 2 in which I set initial constraints on massive galaxy evolution using an empirically derived, redshift dependant $M_{\rm star}$-$M_{\rm halo}$ relation. Knowledge of the exact mapping between galaxy stellar mass and host dark matter halo mass as a function of time and, possibly, environment, has become one of the hottest topics in cosmology. This is because it potentially represents a valuable probe of the relevant, but still not fully understood, physical processes driving galaxy evolution, such as mergers, tidal stripping, or in-situ growth.

As discussed in Section 2.2, in principle, by tracking the redshift-dependent $M_{\rm star}$-$M_{\rm halo}$ relation, it is possible to extract valuable constraints on the average stellar mass growth of (especially) the most massive galaxies, the focus of this work. In this section, I reiterate the concept outlined in Section 2.2. In Figure 4.1 I provide a sketch to explain this point. In panel A of Figure 4.1, I show the mean evolution of a local log $M_{\rm halo}$= $15[M_{\odot}]$ halo from $z = 2$ to $z = 0$. In the left panels (B and D), I assume an extreme scenario in which the number density of massive galaxies remains constant at all cosmic times. As reflected in panel A, a halo mass strongly evolves with cosmic time (due to the hierarchical growth of dark matter haloes), thus the mean halo mass at fixed number density substantially increases with decreasing redshift. As a consequence, massive galaxies at high redshifts get mapped onto increasingly lower mass haloes since the mean mass of haloes is decreasing, at fixed number density, but the mean stellar mass

remains constant. This naturally predicts that the high mass slope of the $M_{\rm star}$-$M_{\rm halo}$ relation should increase at higher redshifts (panel D).

Conversely, the right panels of of Figure 4.1 allows for substantial mass growth in massive galaxies. In this scenario, the number of galaxies per unit co-moving volume of massive galaxies is decreasing with increasing redshift. This would imply that the high-mass end slope of the $M_{\rm star}$-$M_{\rm halo}$ relation should steepen less than the former scenario or even get shallower. In fact if, on average, massive galaxies and their host dark matter haloes are growing in such a way that their respective change in number density at fixed mass is the same, the high-mass end of the $M_{\rm star}$-$M_{\rm halo}$ relation will remain roughly constant in slope, at fixed intrinsic scatter.

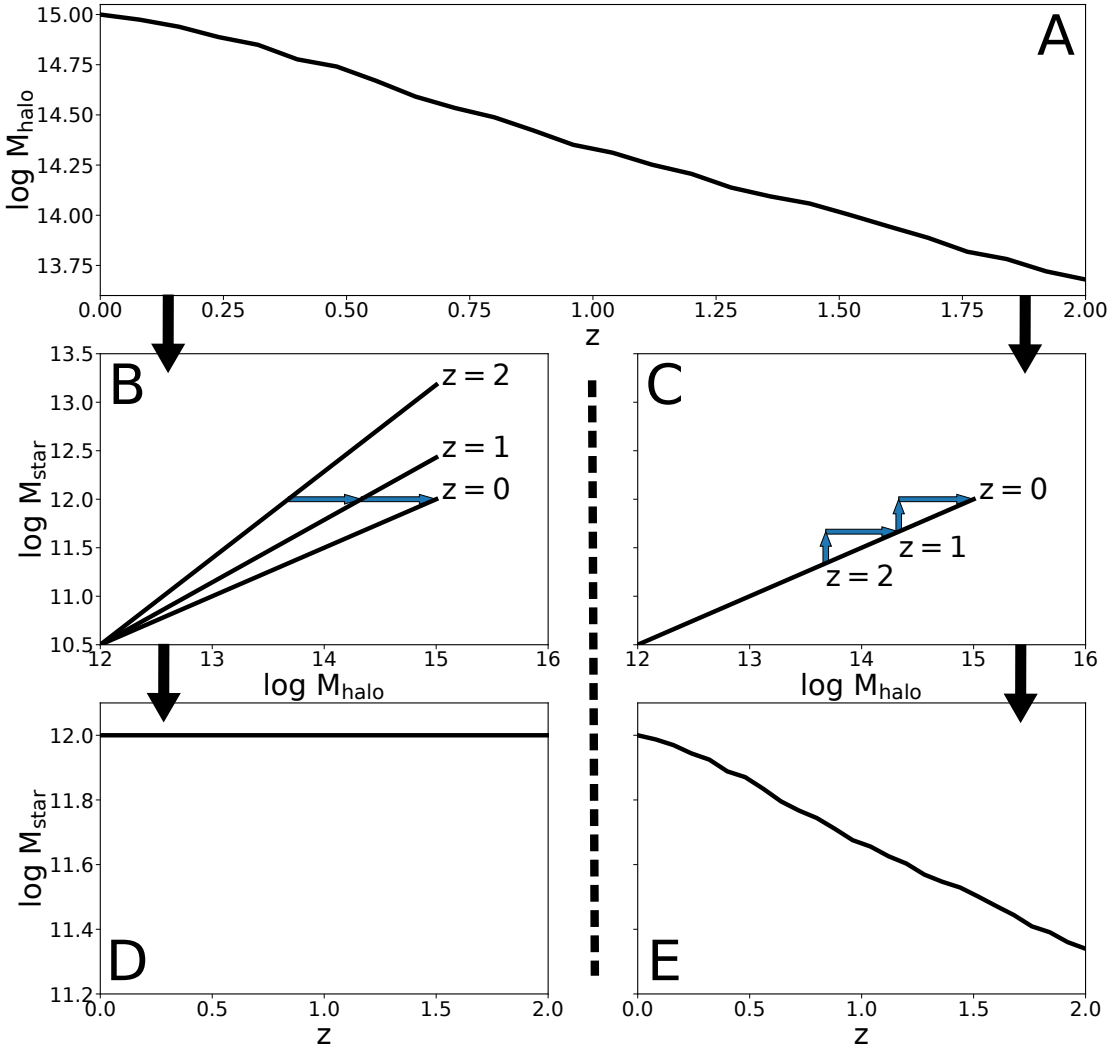



FIGURE 4.1: A cartoon sketching how two contrasting redshift evolutions in the $M_{\rm star}$-$M_{\rm halo}$ relation predicts predict different evolutions in $M_{\rm star}$. Panel A shows the mean evolution of a local log $M_{\rm halo}$= 15$[M_\odot]$ halo. Panels B and C show a gradually decreasing $M_{\rm star}$-$M_{\rm halo}$ relation and a static one respectively. The blue arrows on the relations show the evolution in $M_{\rm halo}$from panel A along with the corresponding evolutions in $M_{\rm star}$. Panels D and E explicitly show the corresponding evolutions in $M_{\rm star}$.

The above arguments imply that independent constraints on the shape and scatter of the $M_{\rm star}$-$M_{\rm halo}$ relation across time can reveal important clues on the average mass growth in individual galaxies. Unfortunately, an exact determination of the $M_{\rm star}$-$M_{\rm halo}$ relation and its evolution with redshift is severely hindered by observational selection effects and/or systematic uncertainties in, e.g., satisfactory light profile models, adequate stellar mass-to-light ratios, unavoidable field-to-field variations, unclear galaxy gradients, etc. Such issues are clearly at play even at relatively low redshifts $z \lesssim 0.5$ (e.g., Kravtsov et al., 2014; Shankar et al., 2014a; Bernardi et al., 2016b, 2017), but become progressively more severe at higher redshifts. This may be due to a variety of causes, from strong surface brightness dimming effects, changes in, e.g., galaxy morphologies and/or in the composition of the interstellar medium, and/or limitations in surveying sufficiently large and deep co-moving volumes (see Bernardi et al., 2017, and references therein).

Alternative methods for probing the relation between galaxies and their host dark matter haloes rely on reproducing the clustering of galaxies of given mass and redshit. These measurements can then be interpreted within the halo occupation distribution (HOD) formalism to derive the probability $P(N|M_{\rm halo})$ that a given halo of mass $M_{\rm halo}$ hosts a number $N$ of galaxies of the type considered (Yang et al., 2003; Tinker et al., 2005; van den Bosch et al., 2007; Zheng et al., 2007; Wake et al., 2011; Leauthaud et al., 2012). It is also possible to probe the high-mass end of the $M_{\rm star}$-$M_{\rm halo}$ relation directly for a small number of clusters by measuring the stellar mass of the central brightest cluster galaxy and the halo mass via, e.g., weak lensing or X-ray measurements of the hot gas (e.g., Finoguenov et al., 2007; Vikhlinin et al., 2009; George et al., 2011; Kravtsov et al., 2014).

Chapter 2 used a combination of abundance matching, galaxy clustering, and direct galaxy and halo mass measurements, to support the evidence of a high mass-end slope of the $M_{\rm star}$-$M_{\rm halo}$ relation significantly steeper than what was previously estimated, with a mean value and scatter of $\delta \gtrsim 0.4$ and $\sigma \sim 0.15dex$ (see Section 4.2.2), both weakly evolving with time up to at least $z \approx 0.5$. Note, however, that none of the techniques considered in Shankar et al. (2014b) are fully free of the observational biases and limitations discussed above. This is simply because Shankar et al. (2014b) compares AB results to an array of different observations at a range of redshifts, each with their own systematics and biases. It is therefore non-trivial to discern if the apparent redshift evolution is a result of redshift-dependant systematics in the observations or a 'real' evolution.

In this chapter, I devise a novel methodology that is largely independent of systematic effects, allowing for a more secure determination of the high mass-end of the $M_{\rm star}$-$M_{\rm halo}$ relation. The aim is to confirm and extend to higher redshifts, the preliminary findings

put forward by Shankar et al. (2014b), that the high-mass slope of the $M_{\rm star}$-$M_{\rm halo}$ relation is steeper than previously thought. In essence, the methodology starts from the abundance and parent halo mass distribution of massive central and satellite galaxies at a given redshift $z_{\rm obs}$, constrained via abundance matching and/or HOD measurements. The satellites were once centrals in their own discrete haloes at a $z > z_{\rm obs}$ (see the cartoon in Figure 4.3), obeying a specific $M_{\rm star}$-$M_{\rm halo}$ relation that, as detailed below, can be effectively constrained in both slope and scatter by the number and parent halo mass distribution at the redshift of observation $z_{\rm obs}$. This methodology has the unique advantage that it does not rely on any specific measurement of stellar masses at $z > z_{\rm obs}$, and as such it is virtually unaffected by systematic observational biases. In other words, we are able to constrain the evolution in the $M_{\rm star}$-$M_{\rm halo}$ relation simply based on observations carried out at the redshift of observation of the satellites $z_{\rm obs}$, without the need for any additional stellar mass and/or volume measurements at higher redshifts.

In Section 4.2 I discuss my full methodology for probing the high mass end of the $M_{\rm star}$-$M_{\rm halo}$ relation at $z = 0.5$ and $z > 0.5 \lesssim 1$. In Section 4.3 I present the results, and in Section 4.4 I discuss the implications these results have on galaxy evolution. In what follows we adopt a cosmology with $h = 0.7$, $\Omega_m = 0.3$, $\sigma_8 = 0.84$, $n_s = 1.0$ to be consistent with the parameters used by Bernardi et al. (2013) but note that none of the results presented in this chapter depend on the exact choice of cosmological parameters.

## 4.2 Overview and Methodology

The aim of this chapter is to probe the high-mass end slope and intrinsic scatter of the $M_{\rm star}$-$M_{\rm halo}$ relation at $z > 0.5$ by relying on *only* the statistics of central and satellite galaxies at $z = 0.5$, thus circumventing the need for higher-redshift stellar mass function measurements. Information on the redshift evolution of the $M_{\rm star}$-$M_{\rm halo}$ relation is vital to constrain the mass growth of the massive galaxies at $z \lesssim 1$, as detailed in Section 4.1.

My overall methodology can be summarised as follows:

1. At any redshift of interest I first create large catalogues of host dark matter haloes and (un-stripped) sub-haloes (those dark matter sub-structures inside the main halo that host satellite galaxies, and masses defined at their infall) extracted from the analytic halo mass functions of Tinker et al. (2008) and Giocoli et al. (2008), respectively. Working with an analytic mass functions allows for the production of halo catalogues with very large volumes without being affected by (the sometimes severe) resolution effects.

2. To each sub-halo is then randomly assigned a redshift of infall $z_{\rm inf}$. The full distribution of $z_{\rm inf}$ for all the satellites as a function of parent halo mass at $z = 0.5$ with $\log M_{\rm halo}/M_{\odot}h^{-1} \geq 12$ is extracted from the Bolshoi simulation (see Appendix B for details) and is shown in Figure 4.4.

3. I then populate the halo and sub-halo catalogues following a variable, redshift-dependent $M_{\rm star}$-$M_{\rm halo}$ relation to create mock galaxy catalogues at all redshifts of interest. The parameters used in this analytic relation are derived in the following steps.

4. In the initial "no-evolution" model, the $M_{\rm star}$-$M_{\rm halo}$ relation at $z > 0$ is derived via abundance matching between the halo and sub-halo mass functions and the $z = 0.1$ Bernardi et al. (2013) stellar mass function. The high-mass end slope $\delta$ and scatter $\sigma$ of the $M_{\rm star}$-$M_{\rm halo}$ relation are then allowed to vary, mimicking some possible redshift evolution in the number density of massive galaxies.

5. We constrain the degree of evolution in both $\delta$ and $\sigma$, particularly in the redshift range $0.5 < z < 1.2$, by fitting the inferred halo mass distributions for both central and satellite BOSS galaxies calibrated at $z = 0.5$ by Tinker et al. (2016).

In the rest of this section, I provide the full details of my methodology.

### 4.2.1 Dark matter halo catalogues

The foundation of the work presented in this chapter are dark matter mock catalogues of haloes and associated un-stripped sub-haloes. Firstly, to produce the catalogue of distinct parent dark matter haloes where the central galaxies will reside, I utilise the analytic halo mass function by Tinker et al. (2008). This halo mass function is well calibrated using a large set of N-body cosmological simulations ($< 5\%$ systematic uncertainty) over a large range of halo masses ($10^{11} \lesssim M_{\rm halo}/M_{\odot} \lesssim 10^{15}$). For a given volume, the halo mass function is integrated in such a way as to find the masses where the number of haloes is an integer, i.e., I perform the operation:

$$n = V \int_{M_{\rm halo}}^{\infty} \Phi(m) dm \quad n \in \mathbb{Z}\,. \tag{4.1}$$

Similarly to a random selection, the approach in equation 4.1 efficiently produces large catalogues of dark matter haloes. The volume, $V = 750^3 h^{-3} Mpc^3$ was chosen as it provides a good compromise between high number statistics and reasonable computing timescales.

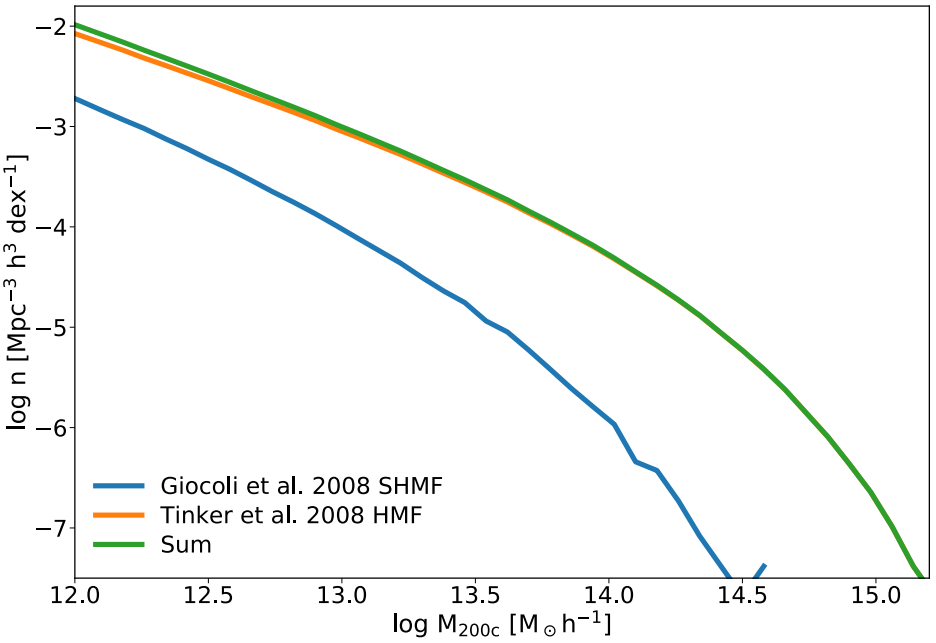



FIGURE 4.2: A plot showing the integrated halo mass function of Tinker et al. (2008) (orange), the integrated mass function of all subhaloes by George et al. (2011) (blue) and their sum (green). These are the principle mass functions used to build halo catalogues in this thesis.

Secondly, for each parent dark matter halo, I construct a modelled catalogue of sub-haloes extracted as in equation 4.1 from the Giocoli et al. (2008) modelled sub-halo mass function. The latter is characterized by a combination of a power-law mass distribution and an exponential cut-off. The power-law in the Giocoli et al. (2008) sub-halo mass function has a nominal value of $\beta = 0.8$ which is slightly reduce to $\beta = 0.65$ to fully match the more relevant sub-halo statistics in Behroozi et al. (2013b) and Tinker et al. (2016). The latter, in particular, is the halo catalogue adopted as a reference at $z = 0.5$. For reference, these two principle halo mass functions can be seen in Figure 4.2.

An essential component of the methodology is the infall redshift distribution $P(z_{\rm inf} \| M_{\rm halo})$. This is extracted from the Bolshoi simulation and then applied to the analytic mock catalogues. I trace each sub-halo with relevant host halo mass $\log M_{\rm halo}/M_\odot > 12$ from $z = 0.5$ to the time it becomes a distinct halo. To efficiently trace sub-haloes through the Bolshoi simulation I exploit a graph representation of the data using a Neo4j graph database[1]. The result is shown in Figure 4.4, which shows the mean $z_{\rm inf}$ for massive satellites is $z \sim 0.7$.

[1] *Full details of my querying algorithm is given in Appendix B.*

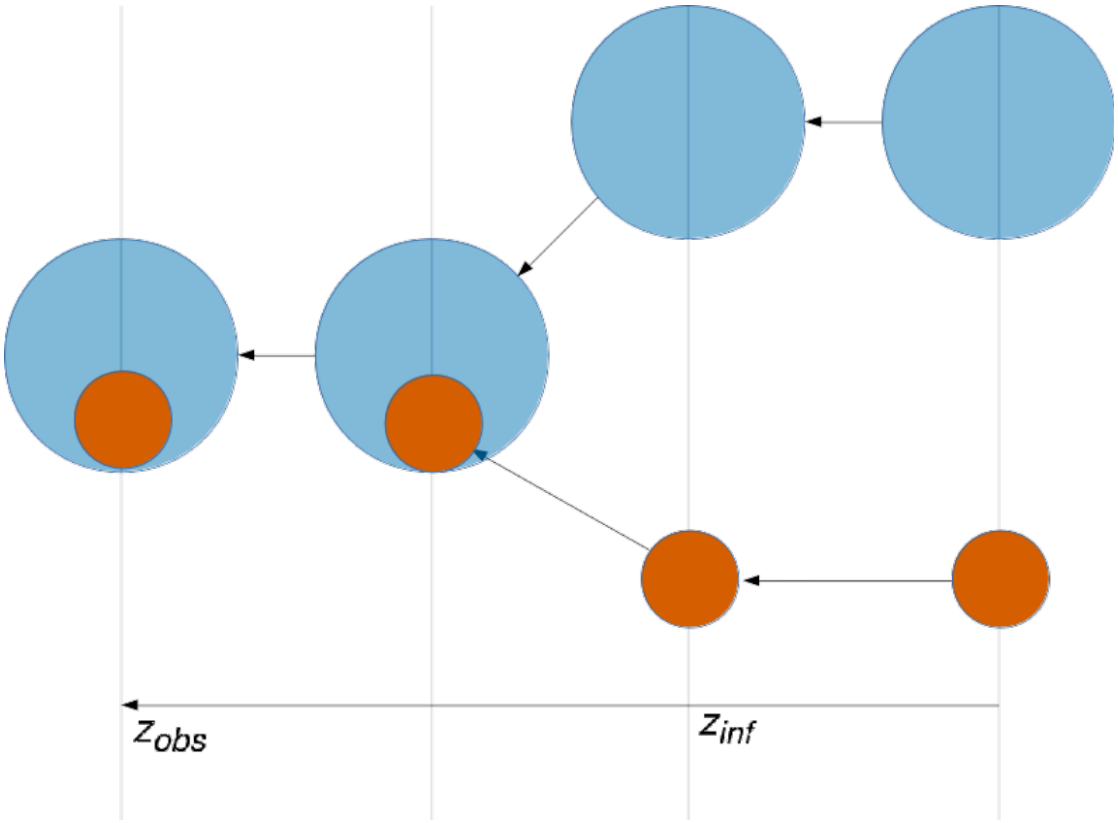



FIGURE 4.3: A cartoon showing how I find the redshift of infall for a satellite galaxy. Here the orange circles represent a halo which starts as a distinct halo at $z > z_{\rm inf}$and falls into a new host halo, represented in blue, at $z = z_{inf}$ and remains a satellite until $z_{obs}$.

### 4.2.2 Mock galaxy catalogues

To populate haloes and sub-haloes with galaxies, I adopt a double power-law $M_{\rm star}$-$M_{\rm halo}$ analytic relation as suggested by Moster et al. (2013)

$$\frac{M_{\rm star}}{M_{\rm halo}} = 2N \left[ \left( \frac{M_{\rm halo}}{M_1} \right)^{-\beta} + \left( \frac{M_{\rm halo}}{M_1} \right)^{\gamma} \right]^{-1} . \tag{4.2}$$

To constrain the parameters of the $M_{\rm star}$-$M_{\rm halo}$ relation up to $z \sim 1$, I follow a multi-step approach, building on the methodology outlined in Shankar et al. (2014b):

- I start by fitting the parameters of equation 4.2 via a cumulative abundance matching procedure between the halo and sub-halo mass functions and the Bernardi et al. (2013) stellar mass function. I assume zero intrinsic scatter in the $M_{\rm star}$-$M_{\rm halo}$ relation and that the stellar mass function has no evolution up to $z \sim 1$. Using a single stellar mass function and not including scatter makes fitting the parameters in the $M_{\rm star}$-$M_{\rm halo}$ relation much more simple and efficient. Selecting one of the

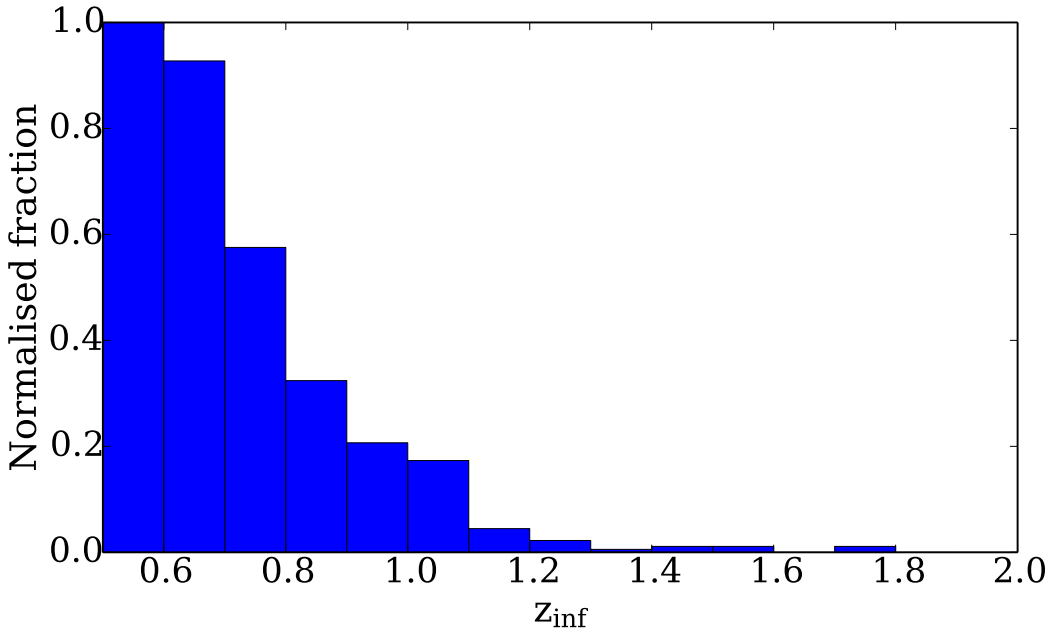



FIGURE 4.4: The distribution of redshift of infall, $z_{\rm inf}$, for all sub haloes in the Bolshoi simulation with $\log M_{\rm halo}/M_{\odot} > 12$ at $z = 0.5$. This distribution is randomly sampled to assign a $z_{\rm inf}$ to sub-haloes in the derived catalogue.

most recent stellar mass functions also somewhat minimises errors due to choosing different stellar mass functions with their own sets of systematics.

Given the strong redshift evolution in the halo mass function, preserving a fixed stellar mass function with imply a noticeable evolution in all the parameters of equation 4.2. Unless otherwise noted (see Section 4.3.2) I assume each redshift-dependent parameter $x_i$ in equation 4.2 to vary with redshift as

$$x_i = \bar{x}_i \left[\frac{1+z}{1+\bar{z}}\right]^{\alpha_i} , \tag{4.3}$$

where $\bar{x}_i$ is the value of $x_i$ at $\bar{z}$, and $\alpha_i$ controls the degree of redshift evolution characterizing parameter $x_i$. Using a weighted least squares regression, I simultaneously fit all the parameters in equation 4.2 using the Bernardi et al. (2013) stellar mass function and ten halo catalogues constructed at different redshifts, in steps of $\delta z = 0.1$ between $z = 0$ and $z = 1$.

Note that, observationally the stellar mass function does not change significantly at or below the knee in the redshift range $0 < z \lesssim 1.0$, but mostly above the knee (eg., Pérez-González et al., 2008; Muzzin et al., 2013). In other words, a decrease in the number density of massive galaxies (i.e., those more massive than the break in the Schechter function) at higher redshifts mostly tends to flatten out the high-mass end slope $\delta$ of the $M_{\rm star}$-$M_{\rm halo}$ relation, at fixed scatter. In what

follows it is safely assumed that all the redshift-dependent low-mass parameters in the $M_{\rm star}$-$M_{\rm halo}$ relation in equation 4.2 ($N$, $M_1$ and $\beta$) are not impacted by any possible variations in the number density of massive galaxies and/or in the scatter $\sigma$ up to $z \lesssim 1$.

- I then relax the assumptions of a strict zero scatter in the $M_{\rm star}$-$M_{\rm halo}$ relation. As I describe in Section 2.2.2, this modification will mostly affect the high-mass end slope $\delta = 1 - \gamma$ of the $M_{\rm star}$-$M_{\rm halo}$ relation in equation 4.2. In fact, as already recognised several times in the literature (Behroozi et al., 2010; Shankar et al., 2014b), increasing the input scatter in stellar mass at fixed halo mass can be accommodated by a shallower slope $\delta$, and vice versa.

- At any redshift of interest, I perform a Monte Carlo Markov Chain (MCMC)[2] in which I fix the "low-mass" parameters in equation 4.2 but allow for both the high-mass end slope $\delta$ and the intrinsic scatter $\sigma$, to vary to fit some sets of independent data. In the specific, I fit $\delta$ and $\sigma$, using an MCMC, to the $z = 0.5$ halo mass distributions for massive *central* BOSS galaxies derived by Tinker et al. (2016, see Section 4.3).

  In each iteration of the MCMC, I populate the mock halo catalogue constructed using the method described in Section 4.2.2, with a stellar mass using the $M_{\rm star}$-$M_{\rm halo}$ relation defined in Section 4.2.2 and add a random scatter to each galaxy with mean 0 and standard deviation $\sigma$. As detailed above, I then "freeze" the low-mass parameters to the best-fit values of the no-evolution model, but leave the high-mass end slope and intrinsic scatter as free parameters. I then extract the co-moving number density of host dark matter haloes of galaxies above a given stellar mass, $n_{\rm gal}(M_{\rm halo})$, and compare it to the one independently inferred by Tinker et al. (2016).

- In order to constrain the high mass-end abundance matching parameters of the $M_{\rm star}$-$M_{\rm halo}$ relation at $0.5 < z \lesssim 1$, the focus of this chapter, I assume an average value of the high-mass end slope $\delta_{z>0.5}$ and intrinsic scatter $\sigma_{z>0.5}$ (which are independent of the $z = 0.5$ values used for centrals). Varying these two parameters will clearly predict different abundances of satellite galaxies at $z = 0.5$. I thus constrain $\delta_{z>0.5}$ and $\sigma_{z>0.5}$ via a similar MCMC fitting routine as described for central galaxies but aimed at fitting the BOSS Tinker et al. (2016) *satellite* distribution of massive galaxies.

As, for the most part, current observations and models suggest the average number density of galaxies seems to gradually rise with cosmic time (e.g., Pérez-González et al.,

[2]Specifically, we use the emcee MCMC package in python (Foreman-Mackey et al., 2013) and plot the results using the ChainConsumer package (Hinton, 2016).

2008; Ilbert et al., 2013; Mortlock et al., 2013; Muzzin et al., 2013) one might infer the number densities of massive galaxies measured at higher redshifts to be lower than, or at the most comparable with, the abundances measured locally. In the strict limit of no redshift evolution in the Bernardi et al. (2013) stellar mass function, abundance matching would thus naturally predict a sort of "maximum slope" for $\delta$, at fixed scatter $\sigma$. This is because any decrease in the number density of massive galaxies at higher redshifts should correspond to a constant or even flatter $\delta$. I thus label this "maximum slope" model as the "no-evolution" model hereafter.

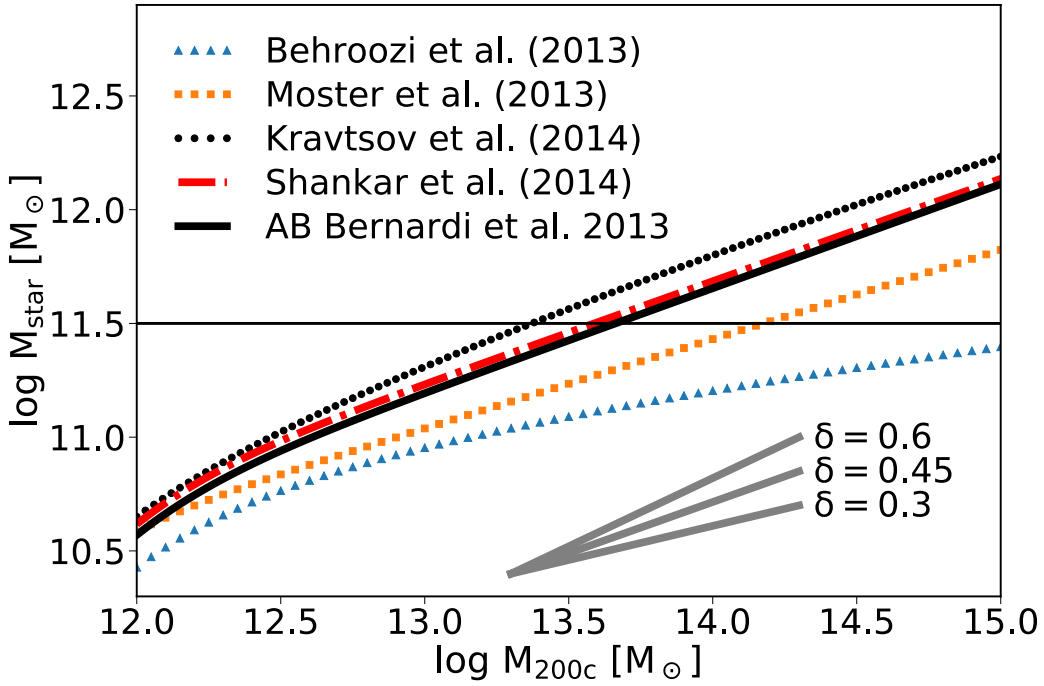



FIGURE 4.5: A comparison among estimates of the stellar $M_{\rm star}$-$M_{\rm halo}$ relations from the literature at $z = 0.1$. The black solid line shows the 'no evolution' relation which I define in section 4.2.2, in which I use the analytic relation by Moster et al. (2013) changing the high mass slope and the scatter to fit the Bernardi et al. (2013) stellar mass function. Additionally, I show three reference slopes with $\delta = 0.6$, 0.45 and 0.3.

## 4.3 Probing the high-mass end slope $\delta$ and intrinsic scatter $\sigma$ with BOSS massive galaxies.

In this section, I lay out the first constraints on the high mass-end slope $\delta$ and the intrinsic scatter $\sigma$ of the $M_{\rm star}$-$M_{\rm halo}$ relation at $z = 0.5$, which will represent the benchmark for setting constraints on the same quantities at higher redshifts. As described in Section 4.2.2, I first fit the low-mass parameters using the Bernardi et al. (2013) stellar mass function. I then run an MCMC to fit the differential halo mass function

distributions $n_{\rm gal}(M_{\rm halo})$ of massive central[3] galaxies as derived by Tinker et al. (2016), only vary the high-mass end slope $\delta$ and intrinsic scatter $\sigma$. Specifically, Tinker et al. (2016) adopted abundance matching techniques to create the galaxy mocks that best reproduce the stellar mass function and the large-scale clustering of the massive galaxies in a complete sample from the Baryon Oscillation Spectroscopic Survey (BOSS). Using this approach, they find a $M_{\rm star}$-$M_{\rm halo}$ relation with a high-mass end slope $\delta \gtrsim 0.4$ and a scatter of $\sigma \lesssim 0.18 dex$ in the range $13 \lesssim \log M_{\rm halo}/M_{\odot} \lesssim 14$ (see, e.g., their Figure 9).

In what follows, broadly following Shankar et al. (2014b), I will specifically focus on the sub-sample of the Tinker et al. (2016) mock galaxy catalogue with $\log M_{\odot}/M_{\odot} > 11.5$ (Kroupa IMF)[4]. As evident from Figure 4.6, this cut is chosen to maximise the completeness of the sample ($\gtrsim 90\%$), but still allowing for a sensible number of satellite galaxies, a vital component of my modelling. Moreover, this mass threshold allows to probe the very massive end of the $M_{\rm star}$-$M_{\rm halo}$ relation, which is most sensitive to changes in slopes and scatters.

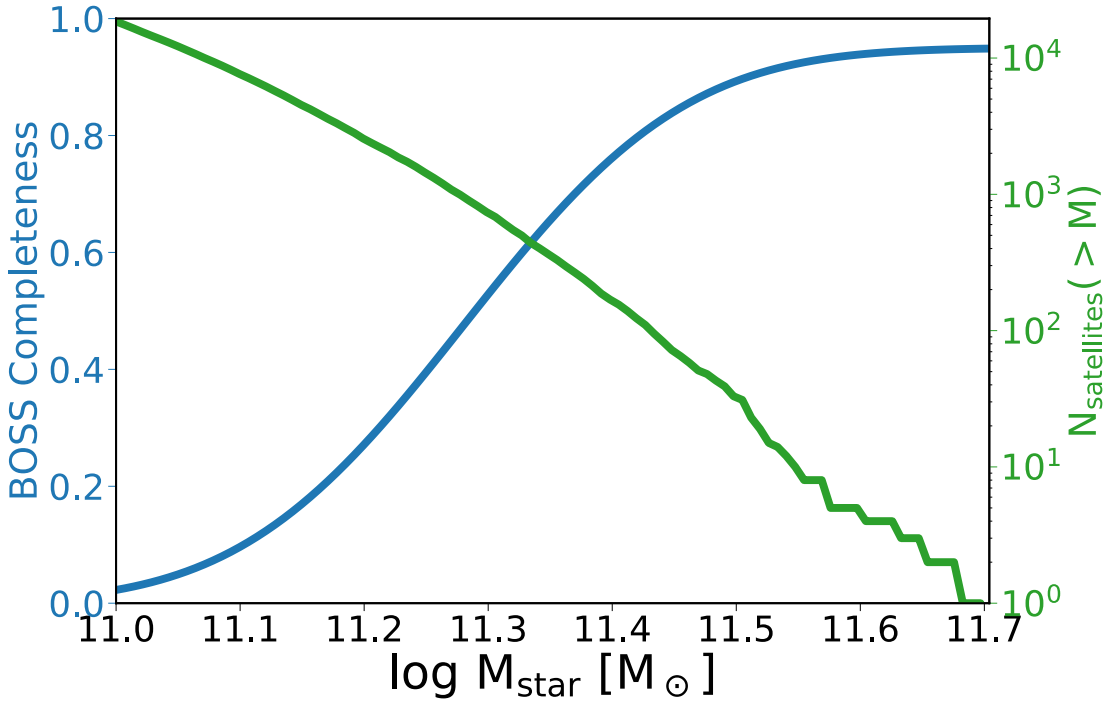



FIGURE 4.6: Blue: the boss incompleteness at $z = 0.5$ from Leauthaud et al. (2016). Green: the expected communicative number of satellite galaxies that are in all the galaxy clusters in the $750^3 [Mpc^3/h^3]$ box I use in this chapter (assuming the Moster et al. (2013) $M_{\rm star}$-$M_{\rm halo}$ relation) as a function of the satellite galaxy's mass.

[3]When comparing with the Tinker et al. (2016) sample I subtract $0.05 dex$ to convert from their adopted Kroupa (2001) IMF, to the reference Chabrier (2003) IMF, following Bernardi et al. (e.g., 2010).

[4]Note that Tinker et al. (2016) generates halo catalogues from the Multidark simulation (Riebe et al., 2011) which adopts slightly different cosmological parameters than the reference cosmology used in this work. However, I find my results are not affected by this.

### 4.3.1 Constraints at $z = 0.5$

The grey bell-shaped curves[5] shown in Figure 4.7 show the $n_{\rm gal}(M_{\rm halo})$ distribution for the Tinker et al. (2016) central galaxies[6] with $\log M_{\rm star}/M_{\odot} > 11.5$. Before showing the results of the MCMC, it is instructive to qualitatively discuss the effects on the $n_{\rm gal}(M_{\rm halo})$ halo distributions of changing either $\delta$ (top panel) or $\sigma$ (bottom panel), while keeping all the other parameters of the $M_{\rm star}$-$M_{\rm halo}$ relation fixed. A steeper high-mass end slope $\delta$, corresponds to a higher number density of massive galaxies which are mapped via abundance matching to less massive and more numerous host haloes. This in turn has the effect to increase the $n_{\rm gal}(M_{\rm halo})$ distributions (blue dot-dashed lines) and to naturally shift the peak of the distributions to lower host halo masses. Similarly, increasing the intrinsic scatter at fixed $\delta$, will clearly increase and broaden the $n_{\rm gal}(M_{\rm halo})$ distributions, as more numerous and less massive haloes will be mapped to galaxies of the same stellar mass. Just the opposite effects will occur by decreasing $\delta$ or $\sigma$ (dashed, orange lines). It is already evident from Figure 4.7 that current data tend to prefer very small scatters $\sigma \sim 0.16$ and steeper slopes $\delta \sim 0.46$ than the usual values $\delta < 0.30$ quoted in the literature at these redshifts (e.g., Behroozi et al., 2013b; Moster et al., 2013).

It is of particular relevance to note that Tinker et al. (2016) use PCA mass estimates from BOSS (Chen et al., 2012), which are roughly consistent with those used in Bernardi et al. (2013) (see Bernardi et al., 2016b, for details). On the other hand, I checked that the stellar masses used by Tinker et al. (2016) are on average $\lesssim 0.25dex$ higher than those adopted by Guo et al. (2014) (see also Shankar et al., 2014b). Nevertheless, I find that that a steeper slope $\delta \gtrsim 0.4$ and a small scatter $\delta \sim 0.15dex$, remain broadly valid at $z = 0.5$ even when switching to the Guo et al. (2014) BOSS galaxy sample. Appendix C explains my choice in using the Tinker et al. (2016) $n_{\rm gal}(M_{\rm halo})$ distributions as opposed to the distribution from Guo et al. (2014).

Figure 4.8 shows the results of the MCMC. Within the top three panels, the lower left panel reports the contour plots for the two parameters that are varied, the scatter $\sigma$ and slope $\delta$, while the lower right and upper panels show their relative distributions. The best-fit value for the scatter is $\sigma = 0.165^{+0.004}_{-0.006}$, which is consistent with what inferred at low redshifts, and by other independent analyses carried out at the same redshift (Shankar et al., 2014b; Tinker et al., 2016). The constrains on the slope are extremely tight, with best fit value $\delta \sim 0.458 \pm 0.004$. The latter appears to be highly

[5]These grey contours include statistical errors on the Tinker et al. (2016) $n_{\rm gal}(M_{\rm halo})$ distributions due to the variation caused by changing the scatter to $\sigma = 0.16$ and $\sigma = 0.20$.

[6]It is worth mentioning that I have converted halo masses in Tinker et al. (2016), defined to be 200 times the background density of the universe, to halo masses defined as 200 times the critical density. The conversion is obtained via the Hu & Kravtsov (2003) formalism, coupled to the concentration-halo mass relation from Diemer & Kravtsov (2015), inclusive of a scatter of $0.16dex$.

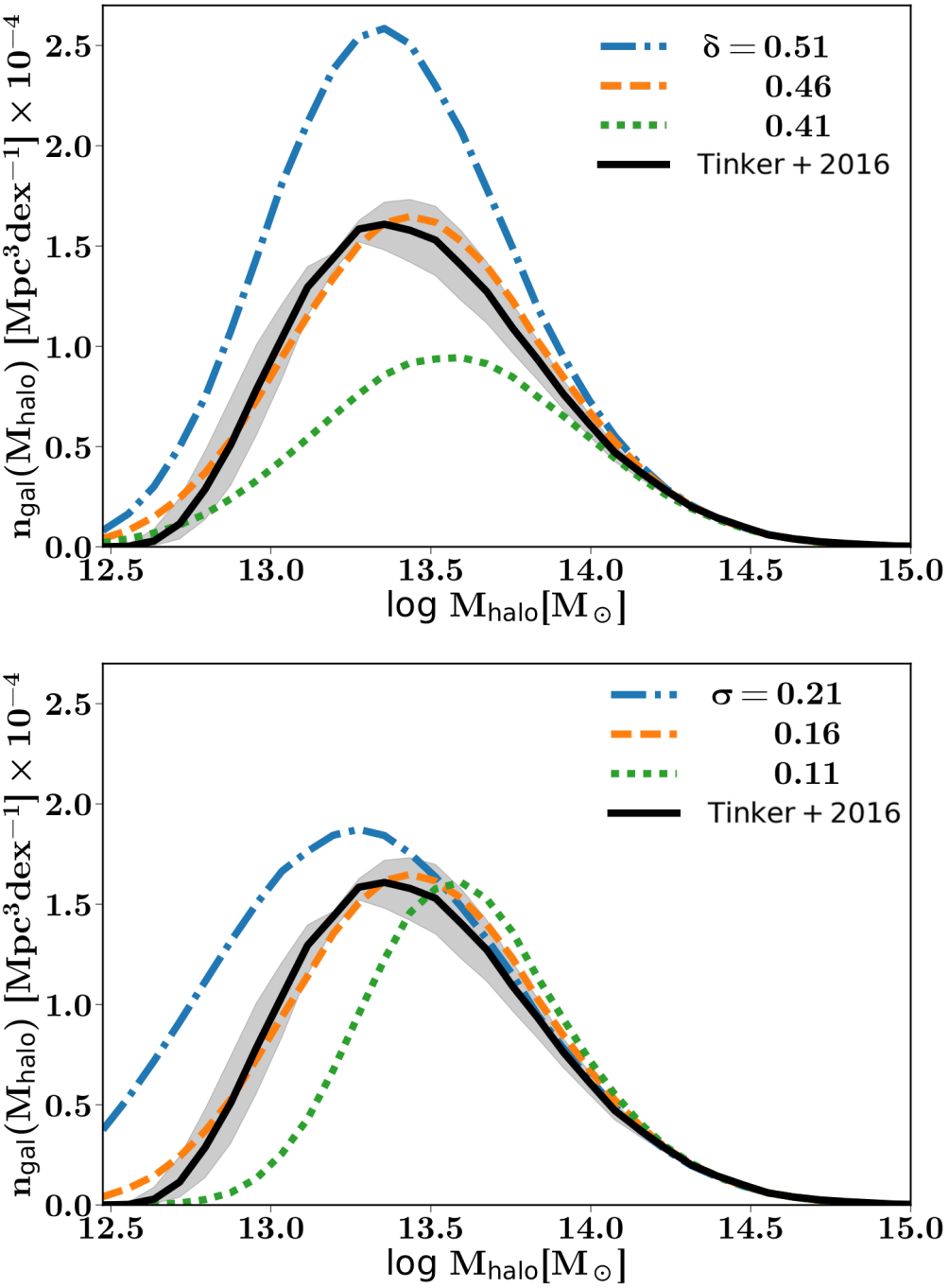



FIGURE 4.7: The galaxy halo mass distributions $n_{\rm gal}(M_{\rm halo})$ of massive galaxies with stellar mass $\log M_{\rm star}/M_\odot > 11.5$. I show the effects of changing the high-mass end slope $\delta$ (top) and the intrinsic scatter $\sigma$ (bottom).

inconsistent with previous results. For example, the slope of the Moster et al. (2013) relation (vertical, dashed red line) is $\delta = 0.28$ at $z = 0.5$, which is more than 40 standard deviations shallower than the best-fit value. This discrepancy can be largely ascribed to the outdated stellar mass functions adopted by Moster et al. (2013, see discussions in, e.g., Kravtsov et al. 2014; Shankar et al. 2014b; Bernardi et al. 2017; Shankar et al.

2017).

More interestingly, the best-fit value is more than $10\sigma$ lower than what suggested by a strictly no-evolution model (vertical solid blue line). The latter is derived via abundance matching between the Bernardi et al. (2013) stellar mass function and the $z = 0.5$ (sub)halo mass function assuming the same intrinsic scatter of $\sigma = 0.165$. Some evolution in $\delta$ is indeed expected at $z = 0.5$. As evident in the bottom panel of Figure 4.8, at face value the Tinker et al. (2016) stellar mass function (dot-dashed, orange line) suggests in fact a slightly lower number density of very massive galaxies with $\log M_{\rm star}/M_{\odot} > 11.5$, with respect to the one predicted in the local Universe by the Bernardi et al. (2013) stellar mass function (solid blue line). This in turn inevitably induces flatter slopes $\delta$ at $z = 0.5$ as massive galaxies will now be mapped to less numerous and more massive host haloes.

At this level of the analysis, it is however unclear if this apparent evolution in $\delta$ from $z = 0.1$ to $z = 0.5$ is a simple bi-product of systematics in stellar mass estimates between the Tinker et al. (2016) and Bernardi et al. (2013) galaxy samples (see, e.g., discussions in Bernardi et al., 2016b). What is more relevant to emphasize here is that the methodology, based on fitting the $n_{\rm gal}(M_{\rm halo})$ host halo mass distributions of galaxies at a given redshift, is exceptionally efficient in detecting even tiny variations in the number densities of massive galaxies, and thus in $\delta$ and/or $\sigma$, over cosmic time. I now move on to set constraints on $\delta$ and $\sigma$ at $z > 0.5$, the main aim of this chapter.

### 4.3.2 The "frozen" model

In order to set constraints on the high-mass end slope $\delta$ and scatter $\sigma$ of the $M_{\rm star}$-$M_{\rm halo}$ relation at $z > 0.5$, I now focus on the $n_{\rm gal}(M_{\rm halo})$ host halo mass distributions of massive *satellite* galaxies at $z = 0.5$. As outlined in Section 4.2, satellites at $z = 0.5$ were in fact centrals in their own respective dark matter haloes at some redshift $z > 0.5$, and followed a specific $M_{\rm star}$-$M_{\rm halo}$ relation applicable to central galaxies at those redshifts. Varying the high-$z$ $M_{\rm star}$-$M_{\rm halo}$ relation, convolved with different stellar mass growth histories in the satellites after infall, will thus substantially impact the shape of the $z = 0.5$ $n_{\rm gal}(M_{\rm halo})$ distributions of satellite galaxies, as detailed below. The match to the Tinker et al. (2016) distributions will then constrain the viable models.

As I did for central galaxies in Figure 4.7, Figure 4.9 shows an example of a comparison between the Tinker et al. (2016) $n_{\rm gal}(M_{\rm halo})$ parent halo mass distributions for massive satellite galaxies with $\log M_{\rm star}/M_{\odot} > 11.5$ at $z = 0.5$, and three $z = 0.5$ model realisations characterized by different high-mass end slopes $\delta$ (top) and scatters $\sigma$ (bottom). Satellites are substantially less numerous and of course inhabit more massive haloes with

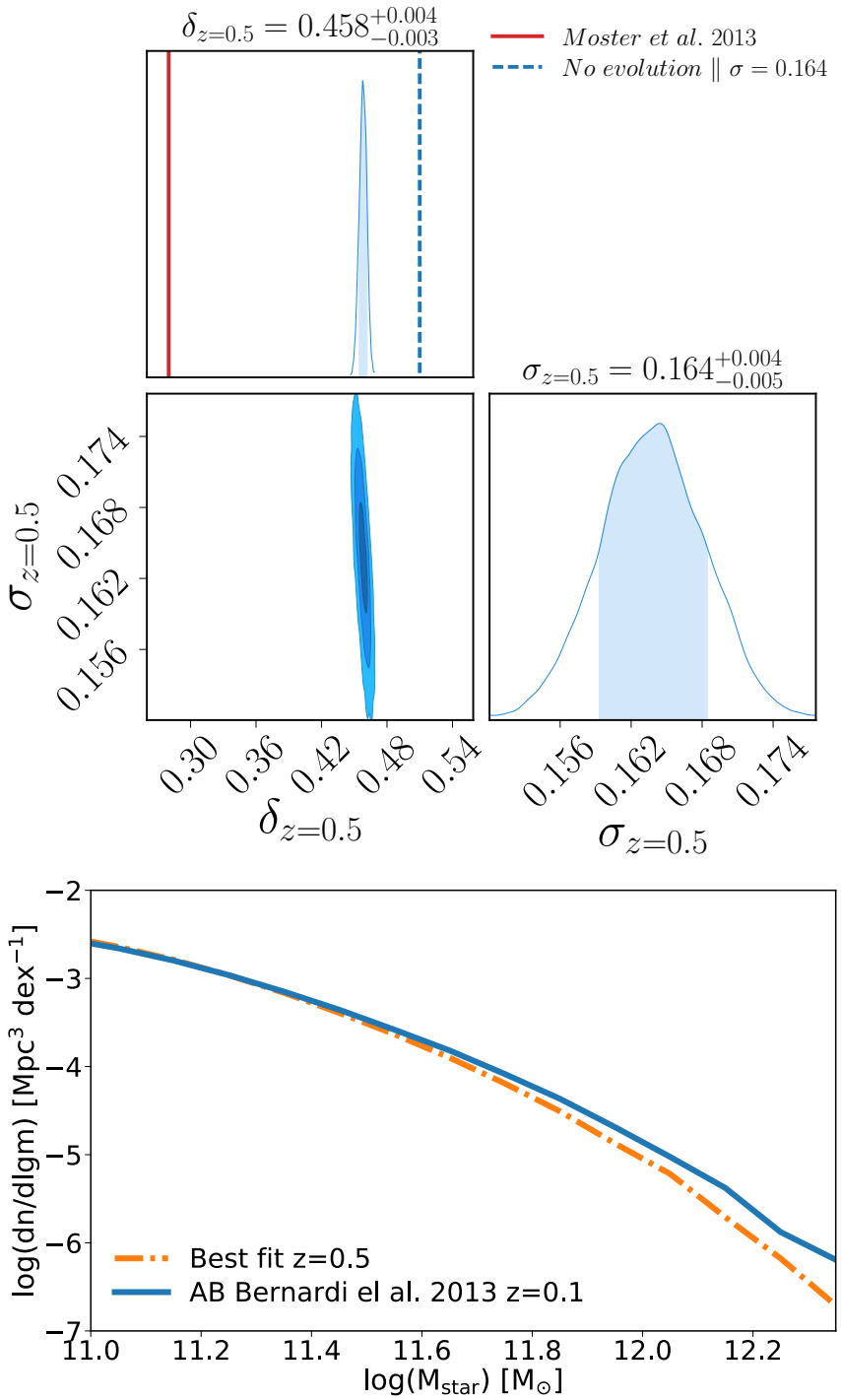



FIGURE 4.8: (top) The results of the MCMC for central galaxies at $z = 0.5$. The histograms show the probability distributions of the high mass slope, $\delta$ and intrinsic scatter, $\sigma$ at $z = 0.5$. The shaded areas shows the 1, 2 and 3 standard deviation confidence regions. The number quoted on top of each histogram is the maximum likelihood of that parameter with its associated error. Additionally, I show the value Moster et al. (2013) high mass slope (red dashed line) and the slope that would imply no evolution in stellar masses between $z = 0.5$ and $z = 0$ (solid blue line) in the top left histogram. (Bottom) The stellar mass function taken from the maximum likelihood parameters of the MCMC as compared to the $z = 0$ stellar mass function from Bernardi et al. (2013).

respect to centrals of similar stellar mass at the same redshift. Nevertheless, Figure 4.9 shows that the method yields very similar results to those obtained for central galaxies, with preferred values of $\delta \gtrsim 0.41$ and $\sigma \sim 0.16dex$. Adopting satellite distributions alone is thus as effective as central galaxies in constraining the parameters of the $M_{\rm star}$-$M_{\rm halo}$ relation.

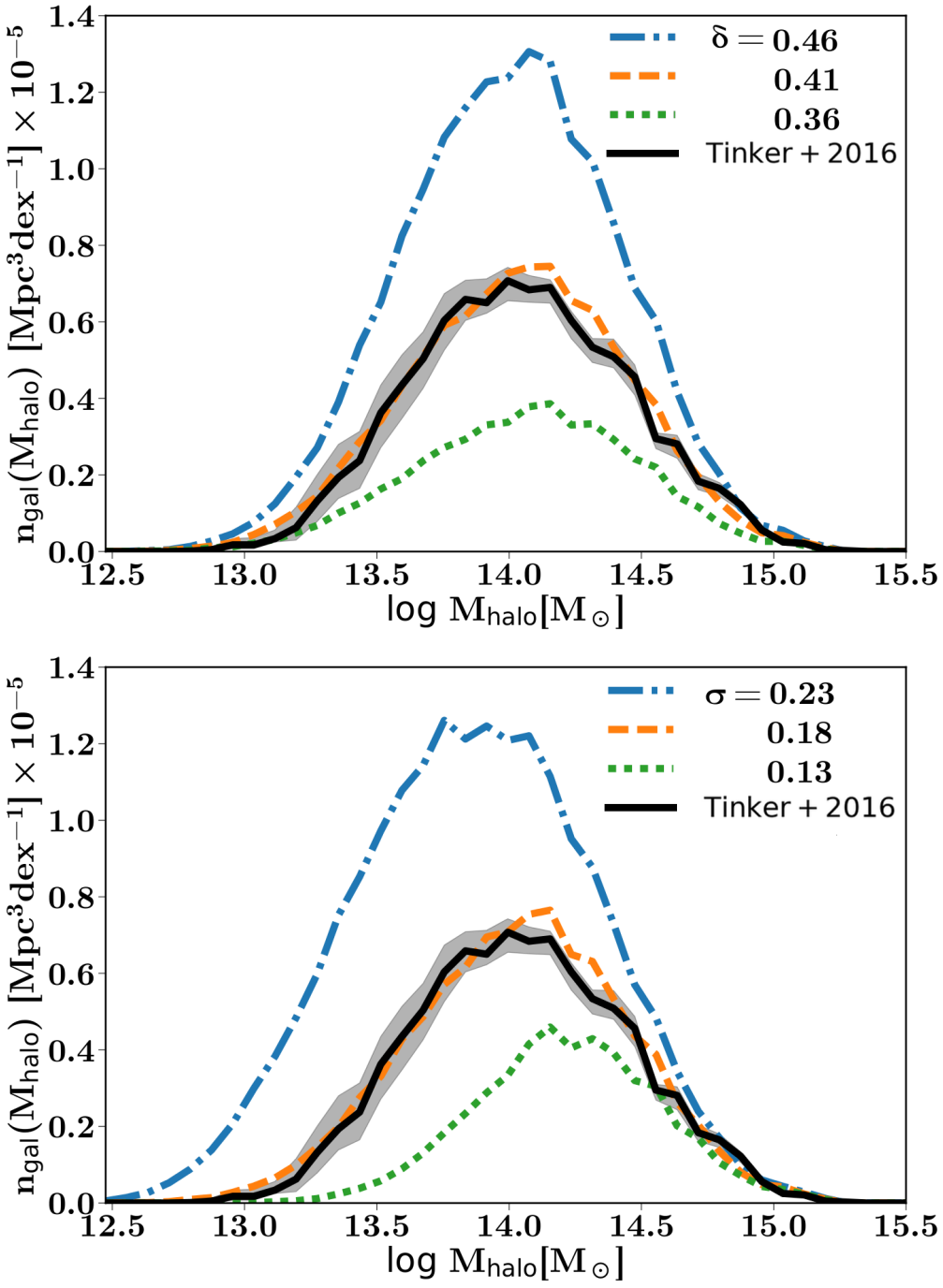



FIGURE 4.9: Same format as Figure 4.7, showing the $n_{\rm gal}(M_{\rm halo})$ parent halo mass distributions for *satellite* galaxies with stellar mass $\log M_{\rm star}/M_{\odot} > 11.5$.

I now move on to the full modelling and add satellites to the MCMC fitting routine

to extend the constraints on the $M_{\rm star}$-$M_{\rm halo}$ relation to $z > 0.5$. To this purpose I simultaneously fit both Tinker et al. (2016) $n_{\rm gal}(M_{\rm halo})$ central and satellite $z = 0.5$ halo mass distributions, assuming *average* values for $\delta$ and $\sigma$ at $z > 0.5$ (independent of those at $z = 0.5$). For the latter, I have in fact checked that a full redshift parametrization as given in equation 4.3 cannot be efficiently constrained by the present data.

I initially consider all satellites to strictly preserve their stellar mass after infall until $z = 0.5$. This could be either interpreted as having minimal star formation and stellar stripping, or simply that the two (or more) processes influencing massive satellites evolution balance each other. I consider this as the reference model, and label it the "frozen" model. I will explore the impact of relaxing the assumption of no evolution after infall in the next Sections.

The best-fit values for the frozen are shown in Figure. 4.10, with a high-redshift high-mass end slope $\delta_{z>0.5} = 0.38^{+0.04}_{-0.03}$ and scatter $\sigma_{z>0.5} = 0.19^{+0.03}_{-0.03}$. At the same time I constrain the values of the slopes and scatter at $z = 0.5$ to be $\delta_{z=0.5} \sim 0.46$ and $\sigma_{z=0.5} \sim 0.16dex$, fully consistent with those obtained in Section 4.3 when comparing with only the $n_{\rm gal}(M_{\rm halo})$ for centrals. Adding the satellites, as expected, does not alter the constraints on the $M_{\rm star}$-$M_{\rm halo}$ relation at $z = 0.5$.

All in all, I find evidence for the slope $\delta$ to decrease beyond redshift $z = 0.5$, while the scatter to remain roughly constant. This inevitably implies evolution (reduction) in the number density of massive galaxies at earlier cosmic epochs, as seen in the bottom panel of Figure 4.10. Our reconstructed stellar mass function shows in fact a decrease of a factor of three up to an order of magnitude for galaxies with stellar mass $\log M_{\rm star}/M_{\odot} \gtrsim 12$ from $z \sim 0.5$ and $z \sim 0.1$ (dot-dashed, orange line and solid, blue line, respectively) to higher redshifts (green, dotted line).

### 4.3.3 Models with tidal stellar stripping

In the previous section I have explored a model where the stellar mass of satellites after infall remains strictly constant, which may well apply in situations of perfect balance between stellar mass losses due to, e.g., tidal stellar stripping, and stellar mass regrowth from residual cold gas in the satellite galaxy. In this section, I explore the possibility of relaxing the assumption of equilibrium, allowing for satellites to specifically undergo substantial amount of stellar stripping of 30%, of the order of (in fact a little less than) the usual amounts inferred from simulations and semi-empirical cosmological models (e.g., Klimentowski et al., 2009; Cattaneo et al., 2011). I assume that the stars that lost during the mass loss are contributing to the intracluster light rather than merging with the central or being expelled from the halo.

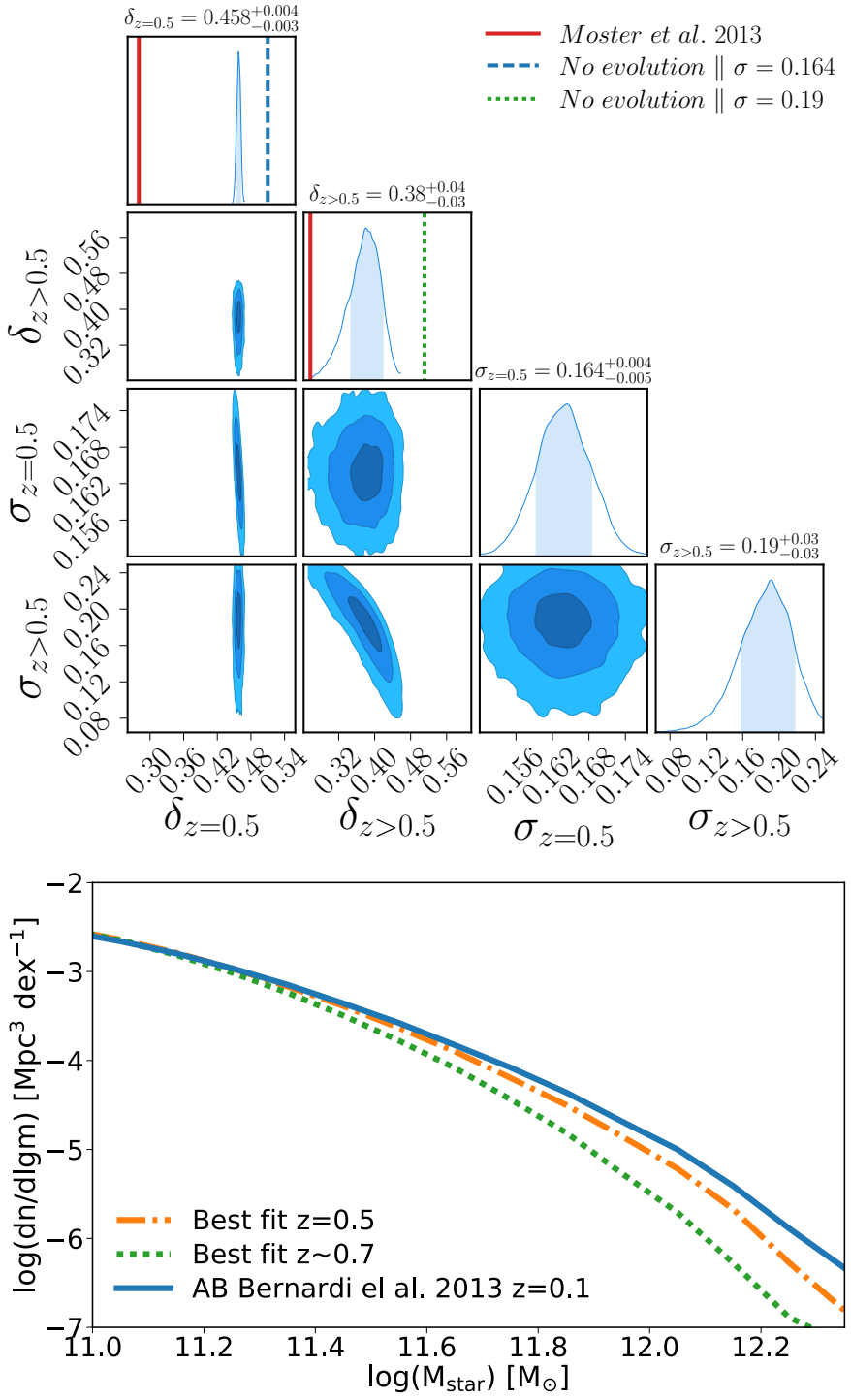



FIGURE 4.10: Similar format to Figure 4.8. Here I have added the constraints on the $n_{\rm gal}(M_{\rm halo})$ of massive satellite galaxies to probe higher redshifts $z > 0.5$ (with mean redshift of infall $z_{\rm inf}$= 0.7). I assume a "frozen" model, i.e., satellites maintain constant stellar mass after infall. I have also added for comparison the values for $\delta$ from Moster et al. (2013, red, long-dashed line) at $z > 0.5$, and the no-evolution model (solid, vertical lines) with two values of the scatter, as labelled. In the bottom panel we additionally show the implied stellar mass function derived from the maximum likelihood parameters of the $M_{\rm star}$-$M_{\rm halo}$ relation at $z = 0.5$ and at $z > 0.5$ (orange, dot-dashed and green, dotted lines, respectively).

As the methodology cannot currently effectively probe any time- and/or mass-dependent, second-order dependences, just as for the constant $z > 0.5$ slope, scatter $\delta_{z>0.5}$, and $\sigma_{z>0.5}$, I include in the modelling a constant stellar stripping, independent of infall time and/or the potential well of the host halo. In practice, I increase the stellar mass threshold $\log M_{\rm star}/M_{\odot} = 11.5$ by the percentage of assumed stellar stripping, and then re-run the MCMC fitting routine.

The results are reported in Figure 4.11. The best-fit value for the high-redshift slope and scatter are now $\delta_{z>0.5} = 0.49^{+0.03}_{-0.04}$ and $\sigma_{z>0.5} = 0.23^{+0.02}_{-0.04}$, both larger than what found for the frozen model, especially in terms of slope. The latter is of course not unexpected, as a steeper slope implies more numerous massive galaxies, which then gradually decrease with cosmic time due to stripping. What is less expected is that the best-fit slope (and scatter) implies, as shown in the bottom panel of Figure 4.11, a number density of massive galaxies at $z \sim 0.7$ (dotted, green line) even higher than what inferred locally at $z \sim 0.1$ (solid, blue line). Such a large number density of massive galaxies at $z > 0.5$, and subsequent non-linear evolution at lower redshifts, are disfavoured by almost all deep surveys, which suggest instead a gradual build-up of the stellar mass function, as also implied by detailed continuity equation arguments (e.g., Peng et al., 2010; Lee & Yi, 2013; Aversa et al., 2015).

It is important to note that the average increase in $\delta$ between $z = 0.5$ and $z \sim 0.7$ in this model with stellar stripping is not a direct result of the stripping. Instead, it is a consequence of the *central* galaxy population being, on average, more massive at $z \sim 0.7$ than at $z = 0.5$.

If the slope $\delta$ is truly much steeper at $z > 0.5$ than at $z = 0.5$ (and in fact, even steeper than the no-evolution slope), then individual massive galaxies have either lost substantial amounts of stellar mass just below $z \sim 0.7$, and/or have undergone a substantial number of mergers, depleting the population, which is not supported by observational data (e.g., Mundy et al., 2017, and references therein). The much more likely scenario is that tidal stripping is possibly limited to $\lesssim 10\%$, at least above $\log M_{\rm star}/M_{\odot} \gtrsim 11.5$, which would imply that intracluster light should be originating from tidal stripping of lower mass and younger galaxies, as supported by some independent observations. DeMaio et al. (2015), for example, analysed four cluster galaxies in the CLASH sample at $0.44 \leq z \leq 0.57$, finding that the ICL colour and colour gradient could arise from $L^*$ galaxies or even dwarfs, but definitely not from much more massive galaxies. An overall low level of stellar stripping would also nicely reconcile with the need for "healty" satellites, which could better promote efficient size and mass growth via (dry) mergers in massive central galaxies (e.g., Taranu et al., 2013; Shankar et al., 2015).

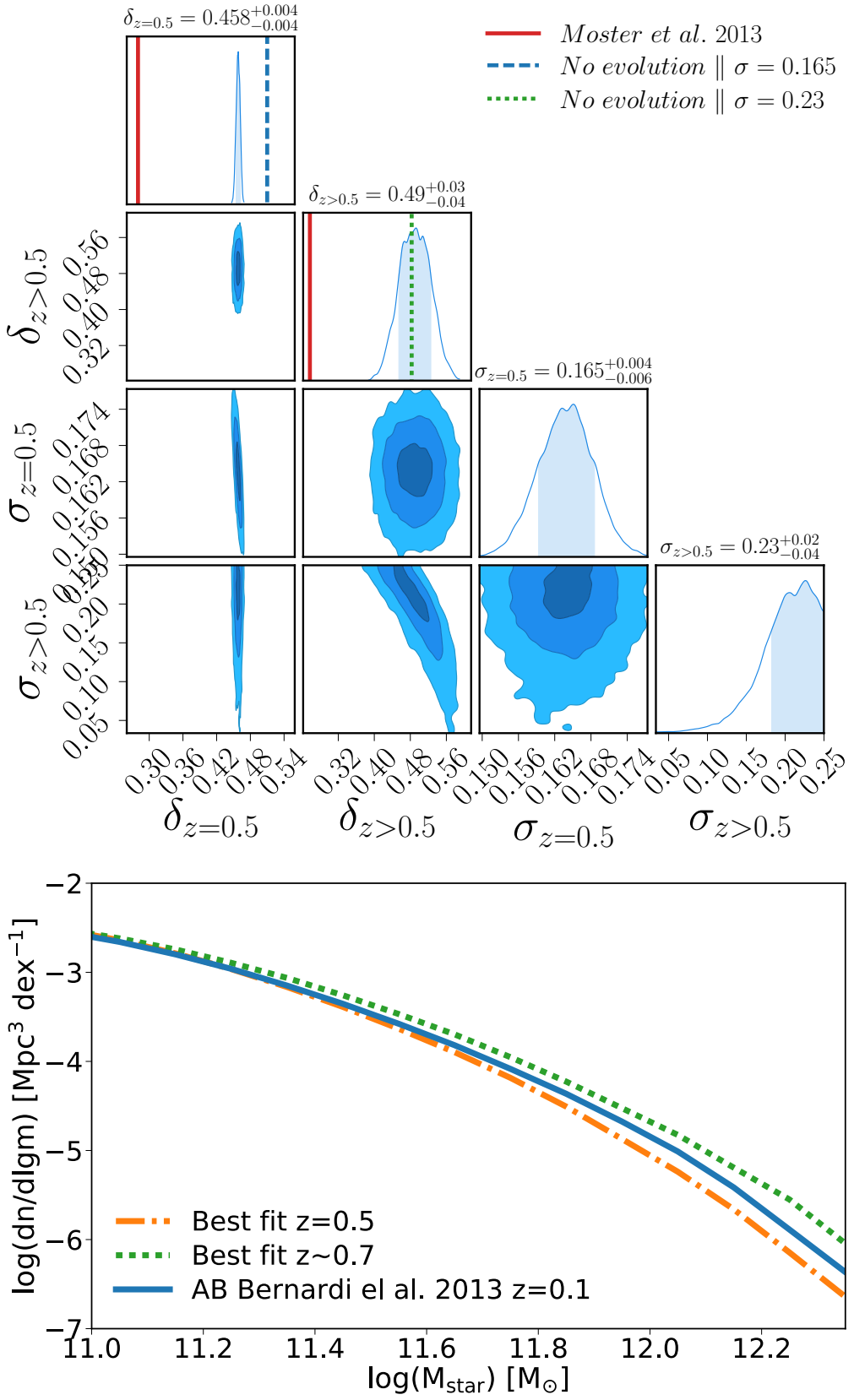



FIGURE 4.11: Same as Figure 4.8 but now assuming that all satellites loose 30% of their stellar mass at $z_{\rm inf}$.

### 4.3.4 Stellar mass loss and star formation

To the models with 30% stellar stripping, I now also include star formation occurring in satellite galaxies after infall. To assess the (maximum) impact that star formation may have on the best-fit $M_{\rm star}$-$M_{\rm halo}$ relation, I allow for massive satellites to continuously form stars between redshift of infall $z_{\rm inf}$ (when they have their stellar masses assigned) and $z = 0.5$ (when they are observed) using a redshift and stellar mass-dependant rate, as specified by the empirical star formation rate (SFR) by Tomczak et al. (2016). As shown in Figure 4.12, I checked the Tomczak et al. (2016) SFR relation to be representative of the (specific) SFR in massive galaxies in the redshift range $0 < z < 1$, broadly matching the latest data by (e.g., Peeples & Somerville, 2013; McDonald et al., 2016, and references therein). I also self-consistently include stellar mass loss rate (for a Chabrier IMF) as parametrized by Behroozi et al. (2013b).

To include SFR in the modelling, I follow a statistical approach. I first derive the redshift-dependent lower limit in stellar mass of a galaxy that at any redshift of infall $z_{\rm inf} > 0.5$ would end up by $z = 0.5$ having a stellar mass equal to the selected mass cut of $\log M_{\rm star}/M_{\odot} \geq 11.5$(+any stripping). I then select all the satellite galaxies with infall stellar mass above this time-dependent threshold mass. If a satellite is more massive than this limit at $z_{\rm inf}$, it is in fact expected to become more massive than $\log M_{\rm star}/M_{\odot} = 11.5$(+ stripping) by $z = 0.5$, and therefore to be included in the chosen stellar mass cut.

Figure 4.13 shows the results of the MCMC for a model with both stellar stripping and star formation. It is clear that the best-fit values for the high-redshift slope and scatter $\delta_{z>0.5} = 0.48^{+0.04}_{-0.03}$ and $\sigma_{z>0.5} = 0.22^{+0.03}_{-0.03}$ are nearly identical to those without star formation (cfr. Figure 4.11). Including star formation, at the observed rate typical of massive galaxies at $z \lesssim 1$, does not therefore alter the overall conclusions disfavouring a strictly no-evolution model since $z \sim 1$.

## 4.4 Discussion

### 4.4.1 Implications for galaxy evolution since $z \sim 1$

By comparing the (large) stellar masses in central galaxies in massive haloes today, with the (relatively low) baryon fraction in their halo progenitors at $z > 2$, the previous chapter (Buchan & Shankar, 2016) argued that the formation and evolution of very massive galaxies should fall into one of the following two scenarios: 1) They are born at $z > 2$ already very massive and extended, with strictly passive evolution thereafter

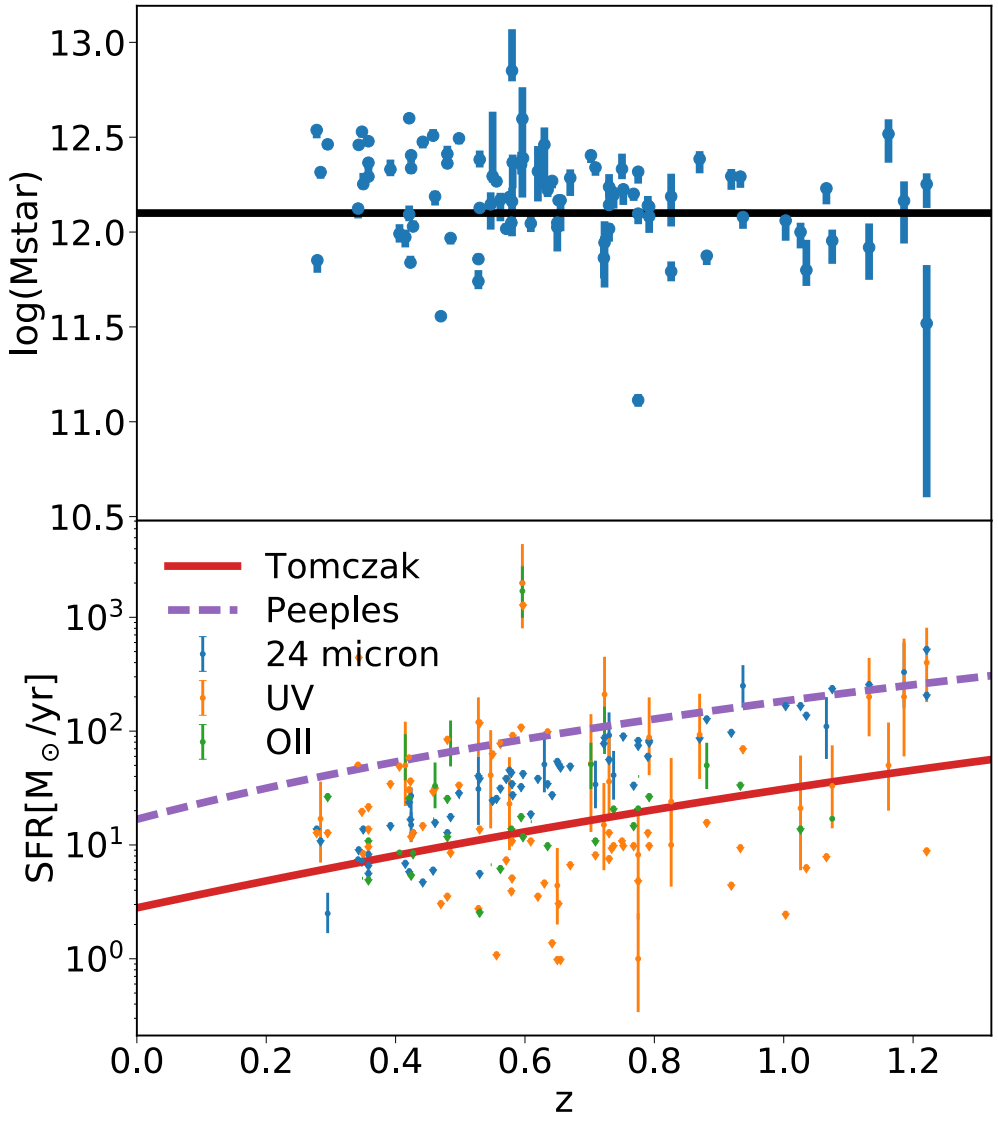



FIGURE 4.12: Top: the blue data points show the stellar masses of BCGs in the McDonald et al. (2016) sample. The horizontal line shows that log$M_{\rm star}$= 12.1[$M_\odot$] is representative of the stellar masses of the sample. Bottom: A comparison between the star formation rates of the BCGs shown in the top panel (measured through the labelled technique) as well the empirical star formation rates by Tomczak et al. (2016) and Peeples & Somerville (2013) using a stellar mass of log$M_{\rm star}$= 12.1[$M_\odot$] at all redshifts.

and no efficient, not even adiabatic, size expansion (Fan et al., 2008; Damjanov et al., 2009; Lapi et al., 2011; Ragone-Figueroa & Granato, 2011; Chiosi et al., 2012; Merlin et al., 2012; Posti et al., 2014). 2) They form only a relatively moderate fraction of their final stellar mass at their formation epoch, and then substantially grow in mass and size at later epochs, predominantly via minor mergers (Hopkins et al., 2009; Naab et al., 2009; Shankar et al., 2010a,b; Lidman et al., 2012, 2013; Shankar et al., 2013; Bai

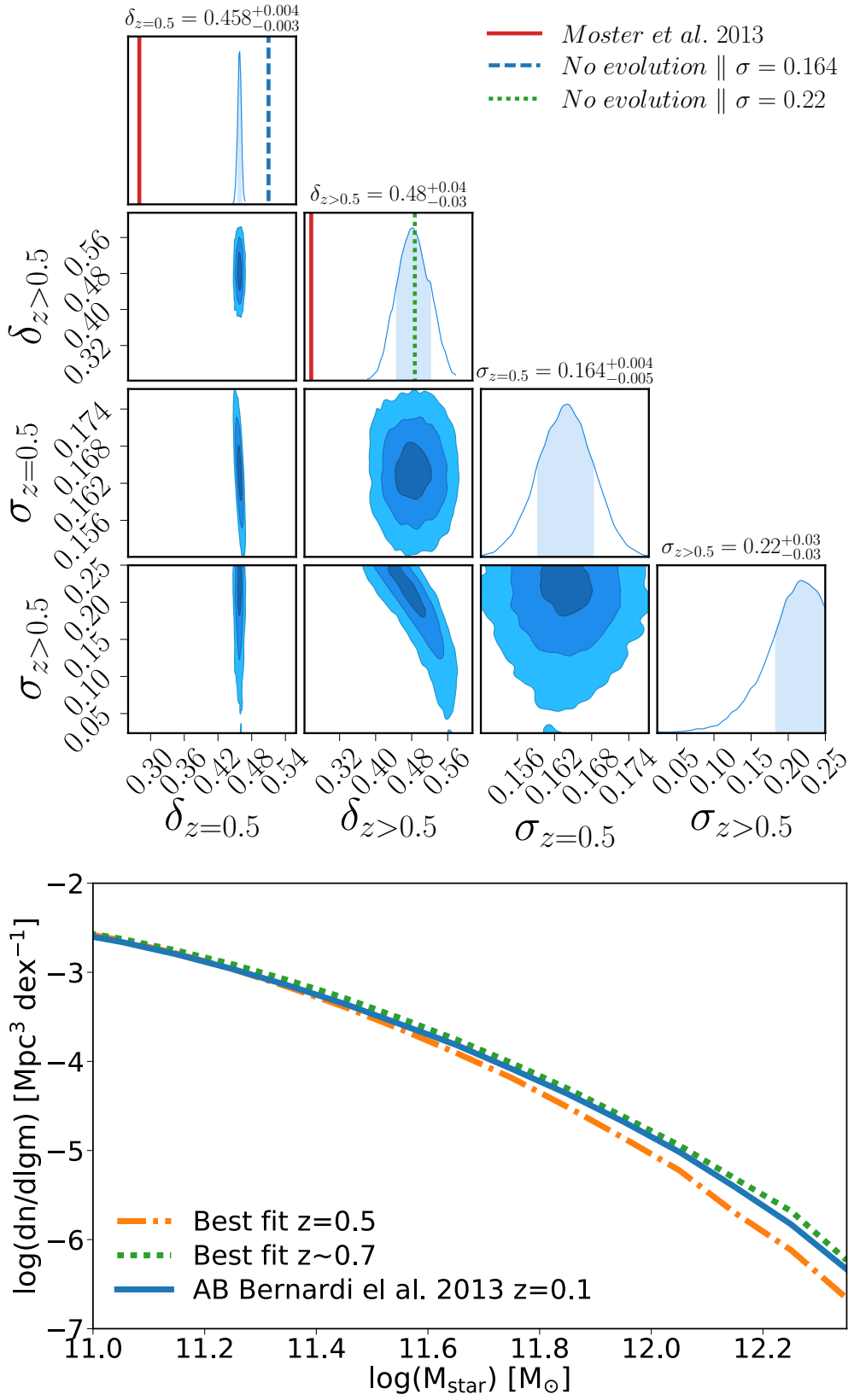



FIGURE 4.13: Same as Figure 4.11 but now also including star formation in satellite galaxies between their $z_{\rm inf}$and $z = 0.5$ at a rate given by Tomczak et al. (2016) and additional mass loss due to stellar evolution given by Behroozi et al. (2013b).

et al., 2014; Shankar et al., 2014a; Peralta de Arriba et al., 2015; Shankar et al., 2015; Bellstedt et al., 2016).

In this work I have presented an efficient methodology to break the above degeneracy by setting constraints on the assembly history of very massive, central galaxies at $z \leq 1$.

In the previous sections I have in fact shown that the inferred satellite host halo mass distribution $n_{\rm gal}(M_{\rm halo})$ at $z = 0.5$ is consistent with negligible evolution in the high-mass end of the $M_{\rm star}$-$M_{\rm halo}$ slope and scatter at $z > 0.5$, with reference values of $\delta \sim 0.45$ and $\sigma \sim 0.16dex$. This behaviour can be more easily understood in terms of the satellite fraction:

$$f_{sat} = \frac{\int n_{gal,sat}(M_{host})\ dM_{host}}{\int \left[n_{gal,cen}(M_{host}) + n_{gal,sat}(M_{host})\right] dM_{host}}\,. \tag{4.4}$$

At fixed scatter, a too steep slope in the $M_{\rm star}$-$M_{\rm halo}$ relation at $z > 0.5$ of $\delta \sim 0.5$, comparable to assuming a strict no evolution in the stellar mass function up to $z \sim 1$, would imply too many massive galaxies at $z > 0.5$, and thus an $f_{sat}$ that is too high at $z = 0.5$, with respect to the clustering data, inferred at those redshifts. Conversely, a too flat high-mass end slope, (e.g., $\delta \lesssim 0.30$; as put forward in previous works Behroozi et al. 2013b) would imply too few massive galaxies at $z > 0.5$ to satisfy the constraints on the $f_{sat}$, inferred at $z = 0.5$.

I verified that the main conclusion on the weak evolution of slope and scatter in the $M_{\rm star}$-$M_{\rm halo}$ relation holds true even allowing for further post-processing of the satellites after accretion. In particular, even allowing after infall for relatively moderate levels of stellar stripping ($\sim 10\%$) and/or of specific star formation rate at the level constrained at $z \lesssim 1$ for massive galaxies (Tomczak et al., 2016), induces comparable results. Higher levels of stellar stripping may be possible, but it would create a very unlikely non-linear evolution in the number density of massive galaxies as a function of redshift. The constraint I place on stellar stripping in massive galaxies is in nice agreement with independent, recent work by Moster et al. (2017) also based on semi-empirical routines. It is also interesting to note that low levels of tidal stripping in very massive galaxies suggests the ICL must have been formed by lower-mass galaxies at $z < 1$ (DeMaio et al., 2015).

One interesting question now to ask is what the inferred weak redshift evolution in the (steep) high mass-end slope of the $M_{\rm star}$-$M_{\rm halo}$ relation actually implies for the mass growth of individual galaxies at $z \lesssim 1$ (see discussion and sketch in Section 4.1 and Figure 4.1). To this purpose, following the methodology outlined in Buchan & Shankar (2016), at each redshift of interest in the range $0 < z < 1$, I assign galaxies to host dark matter haloes by using the redshift-dependent Moster et al. (2013) relation with a constant scatter of $0.18dex$. I then select at $z = 0.1$ the "typical" massive galaxy in the

sample with $\log M_{\rm star}/M_{\odot} = 11.7 \pm 0.1$, and trace its putative progenitors back in time by following the evolutionary mass growth of their "typical" host dark matter haloes. I repeat this exercise for the best-fit $M_{\rm star}$-$M_{\rm halo}$ relations and intrinsic scatters at $z = 0.5$ and $z = 0.7$. All stellar evolutionary tracks are normalised at $z = 0.5$, which is the redshift of reference and starting point in this chapter.

The results are shown in Figure 4.14 The Moster et al. (2013) model (dashed, blue line) predicts a broad evolution of a factor of $\sim 2$ in stellar mass growth since $z \sim 1$. This trend agrees well with the constraints at $z \gtrsim 0.5$ (green circles with error bars), which also imply non-negligible growth of $\sim 50\%$ in the redshift range $0.5 < z < 1.0$. For completeness, Figure 4.14 also reports the results of the semi-empirical model put forward by Shankar et al. (2015) (dotted, orange line), which also consistently predicts similar degree of stellar mass growth in the same redshift range. It is interesting to note that Shankar et al. (2015) predict an average stellar mass growth less prominent at $z < 0.35$ than within $0.35 < z < 1.0$ (about half of the growth for the same amount of cosmic time). This possible slowdown has been reported in the recent literature (Oliva-Altamirano et al., 2015).

### 4.4.2 Comparison with previous works

The literature on the evolution of the most massive, central galaxies in the local Universe is extremely rich. Here I do not attempt to provide a comprehensive review of previous papers on this topic (relevant reviews may be found in, e.g., Conselice 2014), but rather to make contact with just a few observational and theoretical results appeared in recent years of direct interest to the present work.

From the observational point of view Vulcani et al. (2016) took advantage of a SAM and UltraVISTA catalogue confirming that the growth history of massive galaxies is in fact dominated by in-situ star formation, which however becomes gradually less dominant at lower redshifts, being replaced by mergers at $z < 1$.

Bernardi et al. (2016b, and references therein) measured the Sérsic-based stellar mass function of CMASS galaxies at $z \sim 0.55$ found very weak evolution in the number density of the most massive galaxies. However, in line with our results, they also emphasize that evolution in the clustering amplitude of SDSS and CMASS galaxies samples matched in co-moving number density, tends to rule out a strictly passive evolution or any minor merger scenarios which preserve the rank ordering in stellar mass of the population.

Other recent observational work by Bellstedt et al. (2016) suggested that the stellar mass growth of the most massive central galaxies since $z \sim 1$ may be up to a factor

of three. Contini et al. (2017) stressed, however, that the overall inferred evolution of the stellar mass function and star formation rate density are hindered by the systematic uncertainties both in the stellar mass and star formation rate estimates. Gargiulo et al. (2016) derived the number density $\rho$ at $1 < z < 1.6$ from the MUNICS and GOODS-South surveys, and at $0.2 < z < 1.0$ from the COSMOS spectroscopic survey. They suggested that ultramassive dense galaxies missing in the local Universe could have joined, in the last 9 Gyr, the non-dense early-type galaxy population through minor mergers.

On the theoretical side, Rodriguez-Gomez et al. (2016) found from the Illustris simulation that the "two-phase" model for galaxy formation, with both substantial in-situ star formation as well as late ex-situ assembly of stellar mass, is a good approximation for especially the most massive galaxies. For the latter, although with large variations depending on environment and age (see also, e.g., Keating et al., 2015; Saito et al., 2016; Wellons et al., 2016), over 80% of their final stellar mass could be gained via mergers (see also, e.g., Gonzalez et al., 2011; Shankar et al., 2013). They also claim that $\sim 50\%$ of the ex-situ stellar mass comes from major mergers, $\sim 20$ per cent from minor mergers ($1/10 < \mu < 1/4$), and $\sim 20$ per cent from very minor mergers ($\mu < 1/10$).

Also, Ownsworth et al. (2014) use number density conservation to find that $51 \pm 20\%$ of the stellar mass of galaxies with $\log M_{\rm star}/M_{\odot} > 11.24$ at $z = 0.3$ comes from major and minor mergers and $24 \pm 8\%$ from in-situ starformation since $z = 3$. They therefore agree that there is substantial growth in stellar mass.

Gu et al. (2016) specifically focused on the evolution of the intrinsic scatter of the high-mass end of the $M_{\rm star}$-$M_{\rm halo}$ relation since $z \sim 2$. They claimed that hierarchical assembly of dark matter haloes and their central galaxies is the key towards a tight scatter of $\sigma \sim 0.16dex$, with little dependence on host halo mass and redshift, in excellent agreement with the findings presented in this chapter.

## 4.5 Summary

In this chapter, I have outlined a promising framework to set valuable constraints on the slope and scatter of the high-mass end of the $M_{\rm star}$-$M_{\rm halo}$ relation. The methodology is specifically designed to bypass the still substantial and unavoidable systematics in galaxy stellar mass and number density measurements, gradually more prominent with increasing redshift. To this purpose, I utilise the halo mass distribution of massive, $\log M_{\rm star}/M_{\odot} > 11.5$ BOSS *satellite* galaxies at $z = 0.5$, as independently inferred by Tinker et al. (2016) from abundance matching and clustering techniques, to constrain

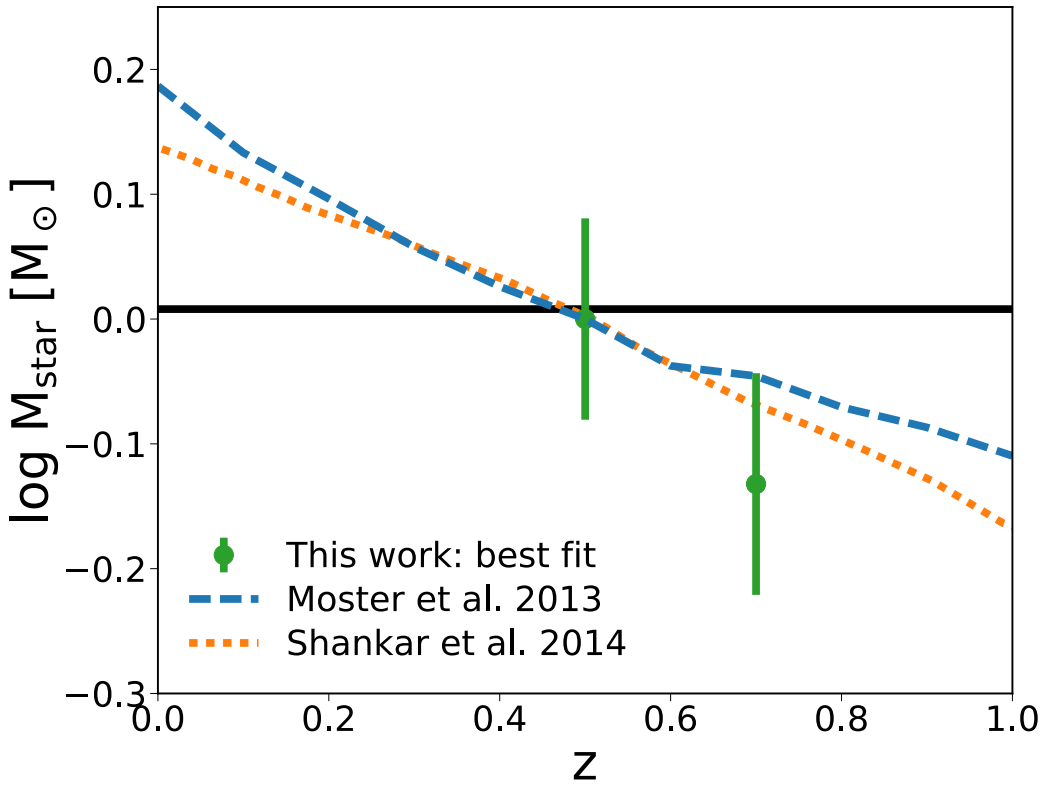



FIGURE 4.14: The predicted mean evolution of stellar mass in individual galaxies, anchored at z=0.5. I have selected haloes in the Bolshoi simulation at $z = 0$ that have log$M_{\rm star}\sim$ 11.7 using the Moster et al. (2013) $M_{\rm star}$-$M_{\rm halo}$ (blue line) and abundance matching to the Bernardi et al. (2013) stellar mass function (errorbars). I trace the progenitors of these haloes to higher redshift. In the blue line shows the median stellar mass of the progenitors when populated using the Moster et al. (2013) $M_{\rm star}$-$M_{\rm halo}$ relation. The error bars show the mean stellar mass when populated using the maximum likelihood $M_{\rm star}$-$M_{\rm halo}$relation assuming the frozen model. The horizontal line shows a constant stellar mass. The orange line shows the results of a full semi-empirical model from Shankar et al. (2015).

the $M_{\rm star}$-$M_{\rm halo}$ relation at $0.5 < z \lesssim 1.2$. My routine is to populate satellite haloes with stellar masses at their respective infall redshifts with $M_{\rm star}$-$M_{\rm halo}$ relations with varying high-mass end slopes $\delta$ and scatter $\sigma$. After infall, satellites are left frozen in stellar mass, or are allowed to evolve in stellar mass via stripping with and without star formation. I utilise an MCMC routine to fit the high-mass end slope and scatter in each respective model to the $n_{\rm gal}(M_{\rm halo})$ distribution.

The main results in this chapter can be summarised as follows:

1. When assuming satellites to be strictly frozen after infall, the best-fit values are $\delta_{z>0.5} \sim 0.38$ and $\sigma_{z>0.5} \sim 0.19dex$. The former implies a slope much steeper than what previously inferred, significantly steeper than what previously suggested in the literature at these redshifts ($\delta \lesssim 0.25$; Behroozi et al. 2013b; Moster et al.

2013). Still, the inferred slopes are significantly lower than what expected from a pure no-evolution scenario in the stellar mass function, implying some stellar mass growth in massive galaxies since $z \sim 1$ of $\sim 50\%$ in the redshift range $0.5 < z < 1.0$. A nearly constant and small scatter is in line with what expected from a hierarchical assembly scenario.

2. Stellar stripping of massive satellites after infall must have been relatively small, $\lesssim 10\%$. Larger amounts of stripping would have implied unrealistically high-$z$ number densities of massive galaxies. This in turn implies that intracluster light should have originated from lower-mass and younger satellites.

3. Including star formation in infalling satellites, at the rate observed in massive galaxies at $z < 1$, does not alter any of the above conclusions.

The next generation of galaxy surveys (e.g., Euclid, LSST) will allow to measure precise clustering of massive galaxies up to $z \sim 1$ thus allowing the methodology presented in this chapter to be extended closer to the peak of the star formation rate ($z \sim 2$) and possibly even beyond.

# Chapter 5

# Testing the mass and size growth of massive galaxies using numerical simulations.

## 5.1 Introduction

The key focus thus far has been on setting more secure constrains on the evolution of the most massive galaxies with the aid of semi-empirical models, specifically designed to minimise the number of assumptions and parametrisations. Now, it is informative to check the results using state-of-the-art numerical simulations. In this chapter, I discuss the results of two projects that utilise the latter technique, tailored to probe the evolution of the most massive galaxies, found at the centres of cluster-sized dark matter haloes.

Thanks to the vast increase in computing power in the past few decades, it has been possible to tackle many of the puzzles in galaxy evolution using increasingly sophisticated numerical simulations. The specific numerical set-up of these simulations depends on exactly what physics they are trying to probe. In general, there are three broad tiers of processes that are involved in galaxy evolution.

The first one represents the gravitational forces that govern dark matter structure formation and, to first order, galaxy dynamics. Simulations that only explore gravitational effects usually utilise a large number of collisionless particles. The force on the $i^{\rm th}$ particle, $\vec{F_i}$, is determined by the gravitational potential, $\nabla\phi$, caused by all the other particles in the simulation:

$$\vec{F_i} = -\nabla\phi|_i \,. \tag{5.1}$$

The second tier of processes are the more complex gas dynamics which are probed by numerically solving the hydrodynamical equations. There are three common methodologies for solving these hydrodynamical equations.

- The first is to use the Lagrange formalism where the coordinates are co-moving with respect to the fluid element. In this framework, it is possible to treat each fluid element as a particle in so-called smooth-particle hydrodynamics (SPH; e.g. GADGET; Springel 2005a). The global properties of the gas are found by smoothing over the particles with a kernel.
- The Second method is to use the Eulerian formalism where space is divided into a Cartesian grid. The hydrodynamical equations are solved by considering the net flux of mass, momentum and energy across each cell boundary. In Adaptive mesh refinement (AMR) codes (e.g. RAMSES; Teyssier 2002), the cells are allowed to divide, providing a high resolution with little computational overhead.
- The last commonly used method is a combination of the previous two. Space is divided into a mesh using a Voronoi tessellation. In so-called moving-mesh codes (e.g. AREPO; Springel 2010), each cell is allowed to 'move' following the velocity of the local flow and thus solving some key numerical issues in the above two methods.

The final tier include the physics of radiative transfer and magnetic processes. These are important in small-scale structures and in high-energy systems. In simulations of galaxy evolution, these second order effects are accounted for in 'sub-grid physics' which encompass for the processes below the resolution limit.

In this chapter, I first utilise N-body simulations to explore if quasi-adiabatic expansion (as described in Section 1.6.2.1) can explain the size evolution in massive galaxies. Secondly, I explore the evolution of massive galaxies in tailored, cosmological zoom-in simulations. I present my re-analysis of simulations run by Martizzi et al. (2014) as well as one simulation I ran.

## 5.2 Quasi-adiabatic expansion of an isolated galaxy

One of the key results of the work presented in Chapter 3 is that the observation of whether or not massive galaxies at $z \gtrsim 2$ already follow the $z = 0$ size-mass relation is crucial in discerning between between the monolithic and hierarchical assembly scenarios (as described in Section 1.6.2.1). This is because I showed that the stellar mass of $z = 0$

massive central galaxies is, on average, higher than the total baryonic mass of their progenitor proto-clusters at $z \gtrsim 2$. This implies that either baryons can be converted to stars with 100% efficency at these early epochs, or that they assemble later. In the former scenario, it is challenging to devise a model in which galaxies are born with their full stellar mass and while still allowing for substantial size evolution, or vice-versa (e.g. Hopkins et al., 2010).

Strong AGN feedback at early epochs might be one avenue for dramatic size evolution. In the so-called quasi-adiabatic expansion model whereby substantial fractions of the baryons are expelled from inside the galaxy to outside the virial radius (Fan et al., 2008). As a reaction to the central gravitational potential lowering, the galaxy will dynamically expand. Physically, this is because the potential energy of the galaxy has dropped rapidly and so no-longer obeys the virial equation: $2 \times$ kinetic energy + potential energy $= 0$. Because the galaxy is no-longer in equilibrium, it will expand and in doing so lower the mean kinetic energy. The expansion will stop when the kinetic energy is lowered enough to re-obey the virial equation.

Groups such as Ragone-Figueroa & Granato (2011) have run numerical experiments on removing baryons from simulated elliptical galaxies, mimicking the effect of AGN feedback. They showed that the amount of expansion is roughly proportional to the amount of mass lost: $R \times M = constant$, that is, if the galaxy looses half its baryons, it will roughly double in size.

If some evolution from AGN feedback is ensuing, it should leave traces in observations of the size-mass relation. Barro et al. (2013) however, showed that compact high-redshift galaxies move from a "blue nugget" to "red nugget" phase, possibly due to AGN feedback, with no apparent change in their sizes. Could morphology, e.g. the presence of a disc, prevent size expansion?

To answer this question, I ran a series of numerical experiments whereby an initially over-massive (with respect to abundance matching relations), three component (stellar bulge, stellar disc and dark matter halo) galaxy loses a substantial fraction of its baryonic mass. In the following sections, I detail the numerical set-up and the results of the simulations.

### 5.2.1 Numerical set-up

Firstly, an isolated, dynamically stable, over-massive galaxy is generated using the GALIC code (Yurin & Springel, 2014). Next, a fraction of baryons are instantaneously removed to bring the galaxy down to a 'normal stellar mass' given by abundance matching relations. For simplicity, the N-body code, GADGET2 (Springel, 2005b) was chosen.

The baryonic mass of the galaxy is artificially lowered by randomly removing particles from the simulation then the resulting galaxy is allowed to dynamically relax. Note that the same exercise is significantly more difficult with a grid-based code as changing the mass in each cell may lead to large numerical artefacts. In addition, GADGET2 is freely available, well documented and has pre-written initial conditions generators and analysis tool kits.

In the work presented here, there is no gas or sub-grid physics in the simulations. This is because quasi-adiabatic expansion is purely a dynamical effect, and thus the details of galaxy composition are not be important. Dark matter and star particles can be treated as collisionless. Dark matter only interacts via gravity and the average distance between stars is much larger than a stellar radius making them effectively collisionless. A simulation that only includes collisionless particles is by far simpler to run and analyse than a full hydrodynamic simulation.

Furthermore, I randomly remove particles with equal probability from the galaxy rather than weighting them by their position. A more realistic model would be to remove particles preferentially from the centre of the galaxy (simulating kinetic/radio mode feedback). However, this first order approximation is a good initial step.

The initial properties of the galaxy are given as follows:

- I generate an isolated, massive galaxy with halo mass, $M_{\rm halo}$, at some redshift $0 \leq z < z_{\rm form}$.
- The virial radius of the halo is given by:

$$r_{\rm vir} = \left[\frac{200}{\Delta_{\rm vir}\Omega_m(t)}\frac{GM_{\rm vir}}{10H(t)}\right]^{\frac{1}{3}} \tag{5.2}$$

$$\Delta_{\rm vir} \approx \frac{18\pi^2 + 82x - 39x^2}{\Omega_m(t)}$$

$$x = \Omega_m(t) - 1$$

- The final stellar mass of the galaxy, $M_{\rm star,f}(z)$ is given by the Moster et al. (2010) abundance matching relation. The initial stellar mass (before mass is removed) is:

$$M_{\rm star,i} = \frac{M_{\rm star,f}}{(1-f)}\,, \tag{5.3}$$

  where $f$ is the fraction of the total stellar mass lost.

- The halo concentration is given by Dutton & Macciò (2014) with no scatter.
- In the work presented here, particles are only removed from the bulge. The final stellar mass, $M_{\rm star,f}$ is split equally between bulge and disc meaning the initial disc mass is $M_{\rm star,f}/2$ and bulge mass is $M_{\rm star,f}(f + 1/2)$ (see equation 5.3).
- The disc is assumed to follow an exponential profile with an initial half mass radius given by the Shen et al. (2003) empirical size-mass relation.
- The bulge-half mass radius is given by the redshift-dependant parametrisations by van Dokkum et al. (2010), normalised to match the relation for $z = 0$ local SDSS galaxies (Bernardi et al., 2013):

$$R_{0.5m} = \frac{10.8}{(1+z)^{1.35}} \times \frac{M_{\rm star}}{10^{11}}\ [kpc]\,. \tag{5.4}$$

The simulations I present in this section have $M_{\rm halo}= 3 \times 10^{13} M_{\odot}$, concentration $c = 4$, $R_{\rm vir} \sim 500\ kpc$, $M_{\rm star,f} = 2 \times 10^{11} M_{\odot}$. The radii are dependent on $M_{\rm star,i}$. In total, $5 \times 10^5$ dark matter particles and $5 \times 10^5$ star particles are used. A key parameter when using particle based simulations is the gravitational softening. This essentially defines the spatial resolution and prevents the particles becoming effectively collisional. The softening lengths used here are $0.01kpc$ and $0.001kpc$ for the dark matter and stars respectively.

### 5.2.2 Results

#### 5.2.2.1 Expansion due to mass-loss

In this section, I present the results of six simulations using three different mass-loss fractions, $f$, and two values of disc height (the disc thickness-to-radius ratio; $DH = h_z/R_{0.5m}$). The simulations are initialised using the methodology detailed in Section 5.2.1. Specifically, I use $f = 0.5$, 0.33 and 0.2 and $DH = 0.2$ and 0.4. A thick disc has a typical value of $DH = 0.2$ but I also include a very extreme disc geometry to test if it is at all possible for the disc to 'block' quasi-adiabatic expansion.

The simulations are then run for one course timestep using the GADGET2 code, after which a fraction of the stellar mass, $f$ is removed from the bulge. The simulation is allowed to run for a further $1Gyr$ without additional intervention. Snapshots of the simulation are saved every $10Myr$. The half-mass radius of the stellar component is extracted from each snapshot by finding the 3D sphere that contains 50% of the stellar particles.

Figure 5.1 shows the expansion of the galaxies in the six simulations. The y-axis shows the fractional increase in half-mass radius and the x-axis shows the time since mass-loss. I firstly find that after the particles are removed, the galaxy indeed does expand. The amount of expansion is directly proportional to the amount of mass that was removed.

It can also be seen in Figure 5.1 that the galaxy's expansion oscillates. This is most likely a simple dynamical effect. At first, the galaxy rapidly expands due to the loss in potential energy which was binding the galaxy together. As it expands, the mean kinetic energy will drop, slowing the expansion. Since the system is only lightly damped (possibly because the particles making up the galaxy are collision-less) meaning the galaxy's size will oscillate about its new equilibrium.

In addition, Figure 5.1 shows that the stellar mass the disc geometry does not inhibit expansion but acts to increase the dampening in the oscillations as the galaxy relaxes.

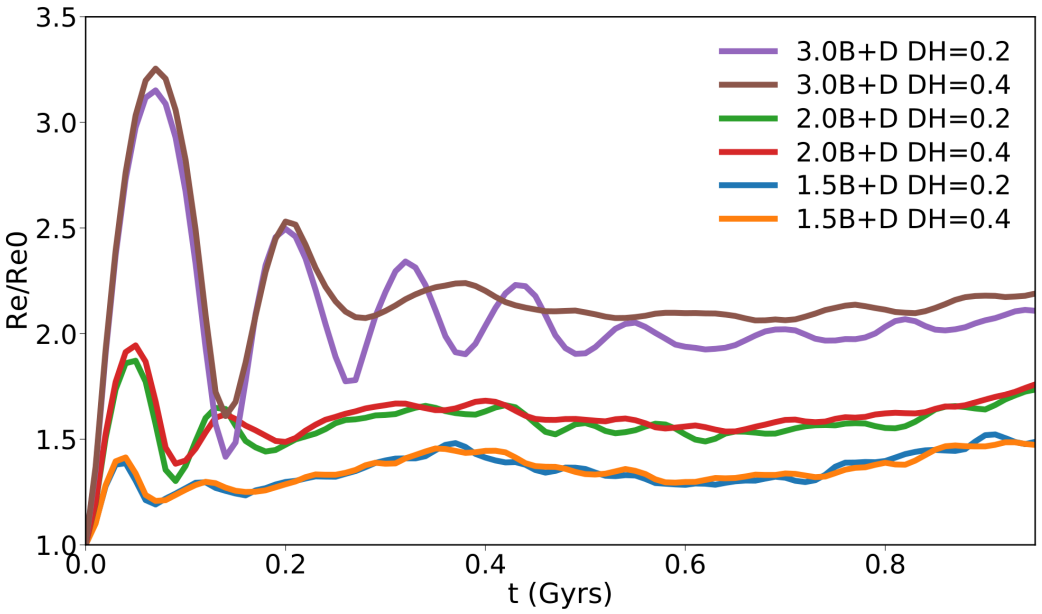



FIGURE 5.1: The fractional expansion of a galaxy after a portion of its mass has been removed. The blue, green and purple lines show when the stellar mass starts as 3, 2 and 1.5 times the final bulge mass (plus the disk mass). In other words, when 20%, 33% and 50% of the stellar mass is removed. The orange, red and brown lines show the same but for a thicker disc geometry (the disk height, $DH = h_z/R_{05m}$). Time starts when the particles are removed.

#### 5.2.2.2 Testing the significance of this expansion

As demonstrated in Section 5.2.2.1, a bulge-dominated galaxy will expand if substantial amounts of baryons are removed. In the following section, I will discuss an important test as well as the wider implications of this result.

One important test to verify the significance of the expatiation is to test stability of the final galaxy without any mass loss. I initialized an isolated galaxy with the same properties as the ones above after the particles have been removed. This means that the initial conditions generator should create a galaxy which is stable and thus not expand. This galaxy was then simulated for 0.5 *Gyr*, without any mass-loss. Figure 5.2 shows the initial density profile of the galaxy (blue dashed line) and after 0.5 *Gyr* (orange solid line), calculated by finding the mass contained in thin shells distance $r$ from the centre of the galaxy. I also show their respective half-mass radii (vertical dashed and solid lines). It is clear that there is a numerical effect (possibly due to the choice of softening length or exact seed galaxy) which is causing particles to diffuse from the central regions, flattening the density profile and thus quasi-expanding the galaxy.

I find that at 0.5 *Gyr* (the time at which all the galaxies which had particles removed had relaxed), the isolated galaxy in this test had expanded by a factor $\sim 1.5$, an amount comparable to the 20% mass-loss (blue and orange lines in Figure 5.1). This test has possibly shown that numerical effects could contribute to the observed expansion in the simulated galaxies. However, it is clear that when substantial amounts of baryons are lost ($\gtrsim 33\%$), the observed increase in half-mass radius is mostly due to dynamical quasi-adiabatic expansion. In either case, the results presented in this section should be viewed as upper-limits on the amount of expansion.

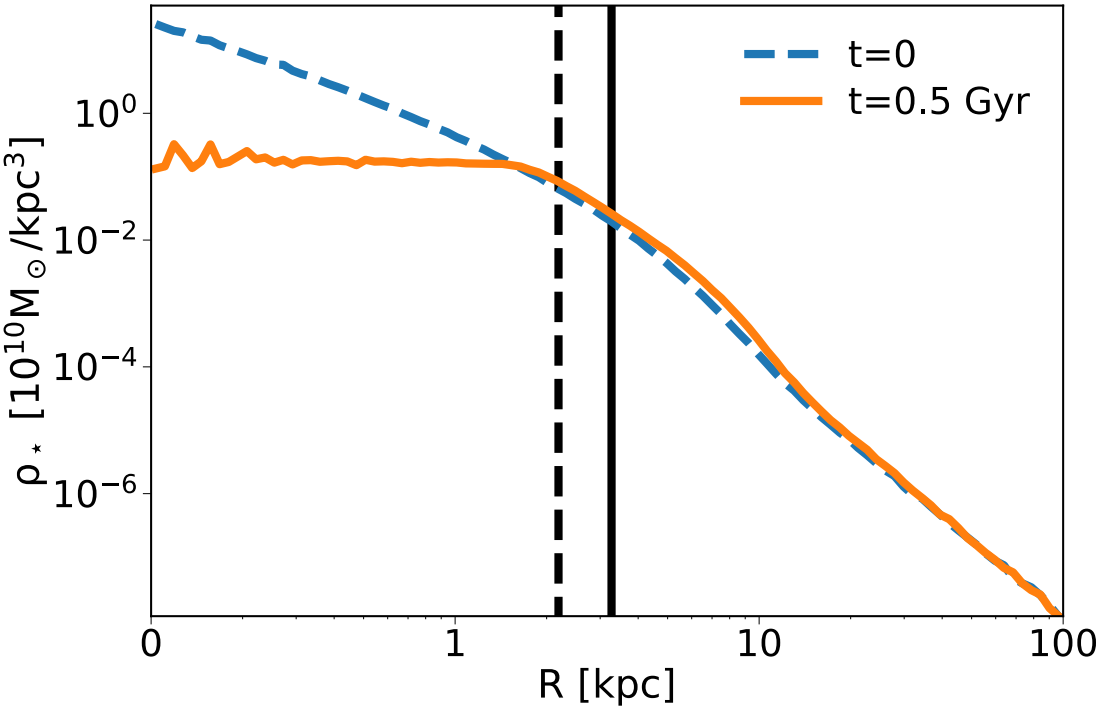



FIGURE 5.2: The 3D density profile of an over-massive galaxy with the initial conditions described in Section 5.2.1 immediately after its initialization (blue line) and after 0.5*Gyr* of time (orange line). The vertical lines show their respective half-mass radii.

### 5.2.3 Conclusions

In this section I explored a model whereby an initially over-massive galaxy has a large fraction of its baryons removed, simulating strong AGN feedback and observed the effect this has on the galaxy's size. I found that expansion of the galaxy is roughly proportional to the fraction of baryons lost, in line with the findings of Ragone-Figueroa & Granato (2011). In addition, this test shows that morphology and/or dynamics of the galaxy does not change the end product of the quasi-adiabatic expansion as put forward by Fan et al. (2008). Indeed, this means that strong kinetic/radio mode feedback cannot be the mechanism to transform compact high-redshift galaxies from a "blue nugget" to "red nugget" phase. More work is needed to verify this effect, starting from different seed galaxies, including hydrodynamics and a more realistic radio-mode feedback prescription.

## 5.3 A cosmological zoom-in with RAMSES

At this stage, it is informative to compare the results presented thus far, derived from semi-empirical models, to state-of-the-art cosmological zoom-in simulations. To this end, I have re-analysed eight simulations by Martizzi et al. (2014) as well as (starting to run) my own tailored simulation.

### 5.3.1 Numerical set-up

In this section, I discuss the numerical set-up of the simulation I ran for this work. The simulations run by Martizzi et al. (2014) use the same methodology but with different initial conditions and slightly different sub-grid physics. In both cases, the RAMSES code (Teyssier, 2002) was used. RAMSES treats stars and dark matter as collisionless particles but uses grids for the gas. Adaptive mesh refinement is implemented to achieve a high spatial resolution in high density regions without wasting computing power on low-density space. Furthermore, zoom-in simulations take this methodology one step further by only allowing one region of the cosmological box to be refined to higher resolution. The latter technique allows for extremely high resolution simulations (e.g. Hirschmann et al., 2013; Dubois et al., 2015), while still accounting for the large scale inflows and outflows of gas into the halo and ultimately onto the galaxy.

In general, cosmological zoom-in simulations are run as follows:

- Initial conditions are generated with the MUSIC code (Hahn & Abel, 2011) at $z = 50$.

- The cosmological box is first simulated only with the dark matter and with a low, uniform resolution.
- At $z = 0$, the halo of interest is selected. The particles from this halo (plus a small border) are traced back to the initial conditions. The surface enclosing the high redshift particles defines the co-moving volume which will be simulated at higher resolution.
- The simulation is then re-run. Outside the high resolution area, the simulation uses the low, uniform resolution as before. However, the mesh and particles in the area of interest are allowed to adaptively refine (up to a limit), giving it a higher resolution than the rest of the box. Snapshots are saved at pre-determined redshifts.
- The simulation can then be analysed.

I initially simulated a $(100/h\ Mpc)^3$ cosmological box containing only dark matter to $z = 0$. I selected a massive halo which is towards the outskirts of a high density region to re-simulate at higher resolution. The zoom region was $3 \times R_{vir}$ and the peak resolution inside this region was $\sim 1kpc$. In my simulation, I utilise the sub-grid physics and physical parameters as used in the Horizon-AGN simulation (Dubois et al., 2014) which were tuned to broadly reproduce the statistical properties of local galaxies. I aimed to run my simulation to $z = 0.5$ to connect with Chapter 4. However, due to technical issues[1], I only ran it to $z \sim 4.3$. In the following section, I discuss the results of my simulation, which I have complemented with my analysis of the Martizzi et al. (2014) simulations which have been run to $z = 0$.

### 5.3.2 Analysis and results

The following analysis applies to both the simulation I ran for this work and to the Martizzi et al. (2014) simulations. Firstly, each snapshot gets analysed with the HOP halo finding algorithm (Eisenstein & Hut, 1998, See Section A.2 for brief discussion on halo finders). The next stages of the analysis heavily utilise the pynbody python package (Pontzen et al., 2013). Each snapshot is centred on the halo of interest using the centroid determined from the halo finder, after which it is re-centred using the more accurate shrinking-sphere method (Power et al., 2003).

[1]These issues are set to be resolved in future work so that the simulation can be completed along with a detailed analysis (Grylls et al. in prep).

#### 5.3.2.1 Global properties

After the halo has been centred, some initial key properties can be gathered. Firstly, I measure the halo radius and mass ($R_{200c}$ and $M_{200c}$ respectively) by finding the sphere in which the mean density is 200 times the critical density of the Universe at that redshift. Additionally, I find the stellar and gas masses enclosed in that sphere.

Figure 5.3 shows an image of the projected 2D density of the dark matter halo I simulated at $z \sim 4.3$ within a cube with sides $6 \times R_{200c}$ (where darker regions show higher density). The orange circle in the image shows $R_{200c}$.

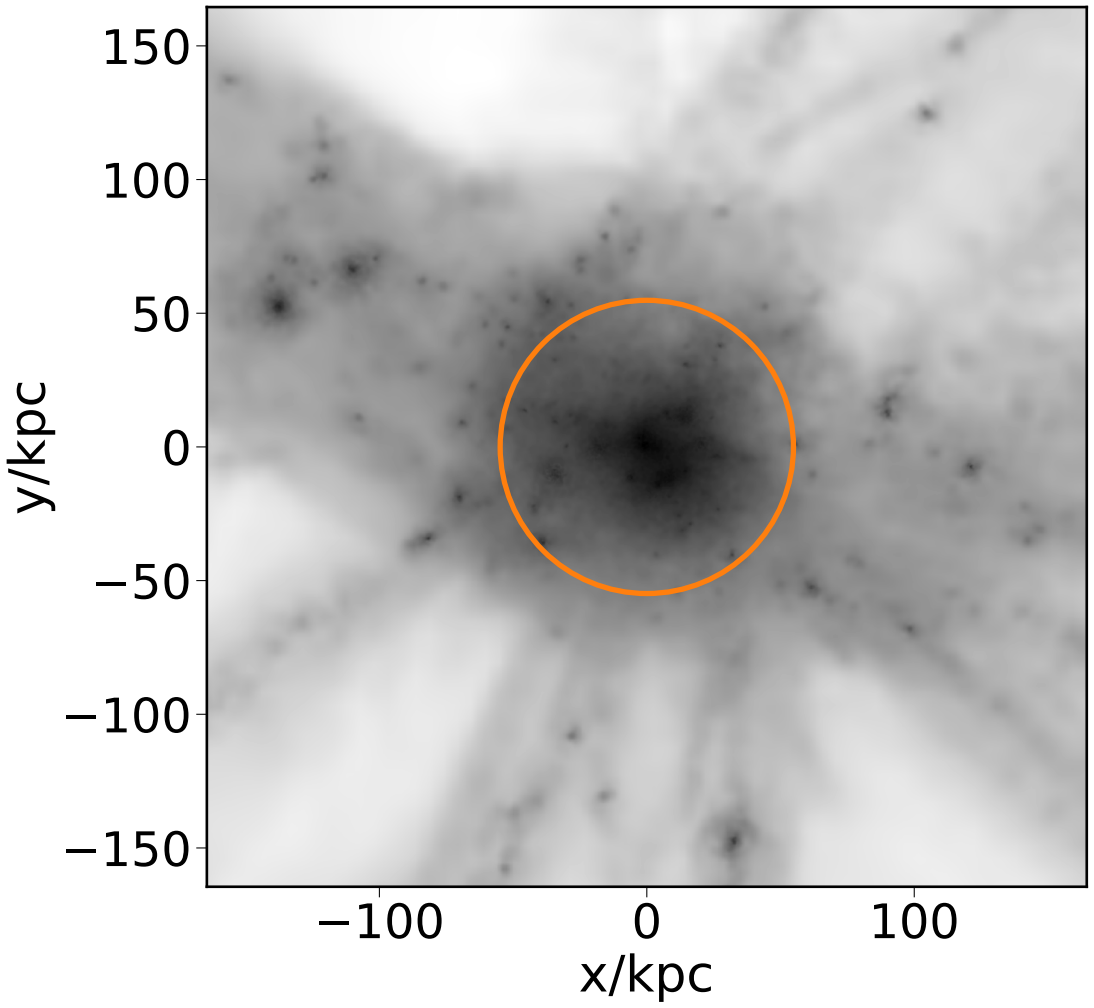



FIGURE 5.3: An image of the projected 2D density of the dark matter halo I simulated at $z \sim 4.3$ within a cube with sides $6 \times R_{200c}$. Darker regions show higher densities. The orange circle shows $R_{200c}$.

The lines in Figure 5.4 show the evolution of $M_{200c}$ in each of the chosen haloes in the zoomed-in simulations by Martizzi et al. (2014). The star shows the mass of the halo in my simulation at $z \sim 4.3$. It is clear that each halo has substantial growth in mass

over cosmic time, in agreement with the structure formation framework discussed in Section 1.3, and they all end up as massive clusters at $z = 0$.

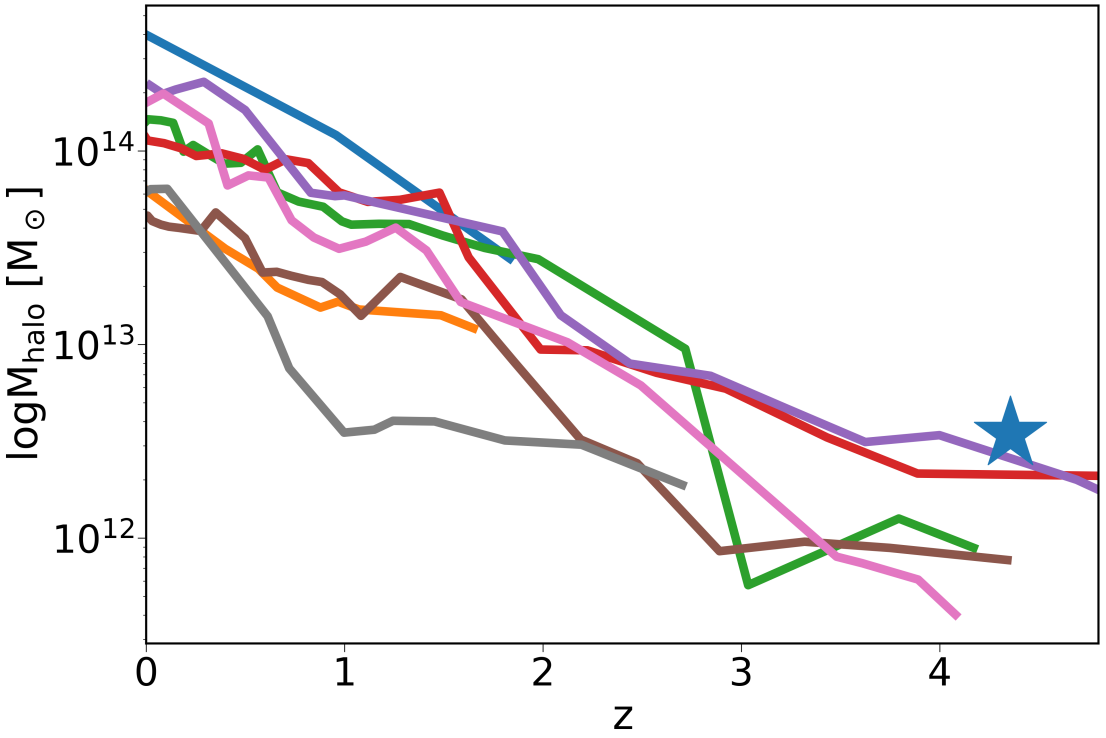



FIGURE 5.4: The evolution of $M_{\rm halo}$ with redshift at the centre of each of Martizzi et al. (2014) zoom-in simulations (coloured lines). Additionally, I show $M_{\rm halo}$ of my simulated halo at $z \sim 4.3$ (blue star).

#### 5.3.2.2 Estimating stellar mass

Having probed the halo properties in each zoom-in simulation, I now turn to the central galaxy. There are in general, two ways to measure the stellar mass of massive simulated galaxies (Puchwein et al., 2010). The first is to define a 'maximum radius' which specifies the outer edge of the galaxy, normally based on empirical estimates of a galaxy's size based on its halo mass (from, e.g. Bernardi et al., 2014). The other method is to fit 2D or 3D density profiles with a model template. The model density profile can then be integrated to give the total mass of the galaxy. Importantly, the massive elliptical galaxies at the centres of clusters have an excess of stars at large radii with respect to lower mass elliptical galaxies (known as the intra-cluster light; ICL Schombert, 1986). To account for the ICL, I use the latter method for measuring the central galaxy's mass. Specifically, I assume that the 3D density profile of the central galaxy and ICL can each be well described by Hernquist (1990) profiles[2] with the analytic form:

[2]In principle, it could be better to fit a Sérsic+exponential profile or even two Sérsic profiles. (Graham & Driver, 2007). However, as discussed by Puchwein et al. (2010), the increase in degrees of freedom with these models makes it difficult to unambiguously de-blend the ICL from the central galaxy.

$$\rho(r) = \frac{M}{2\pi}\frac{a}{r}\frac{1}{(a+r)^3}\,, \tag{5.5}$$

where $M$ is the total stellar mass and $a$ is the scale radius. Equation 5.5 leads to the cumulative mass distribution:

$$M(<r) = M\frac{r^2}{(r+a)^2}\,. \tag{5.6}$$

Note that the half-mass radius, $R_{0.5M} = (1+\sqrt{2})a$. This profile closely reflects the surface brightness profile, $I(r) \propto r^{\frac{1}{4}}$, as observed in elliptical galaxies (Uson et al., 1991; Scheick & Kuhn, 1994; Gonzalez et al., 2000).

To measure the 3D stellar density profile, I use the profile tool in the pynbody python package, which finds the number of star particles within thin shells, distance $r$ from the galaxy centre. However, as seen Figure 5.3, there are many sub-haloes inside the inner regions of the host halo. Each of these sub-haloes will contain a satellite galaxy which will heavily disturb the density profile of the central galaxy. I therefore iteratively mask stellar over-densities within $0.5R_{200c}$ until the smoothed galaxy profile always decreases. The mass and the volume of these masked regions are subtracted from the density calculation and thus do not bias the mean density calculation.

For every snapshot in each simulation, I calculate the 3D density profile, to which I fit both one and two Hernquist profiles using a least-squares regression. The purple line in Figure 5.5 shows the 3D density profile for the galaxy at the centre of simulation run in this work at $z \sim 4.3$. The red line shows the best fit single Hernquist profile. Instead, the blue line shows the best fit double Hernquist profile where the first 'galaxy' component is shown as orange and the second 'ICL' component is shown in green.

Figure 5.6 shows the evolution of stellar mass of the galaxy at the centre of the zoomed-in halo in each of the Martizzi et al. (2014) simulations. Furthermore, the blue star shows the stellar mass of the galaxy in my simulation at $z \sim 4.3$. The top panel shows the best fit integrated stellar mass if only one Hernquist profile is fitted. Instead, the middle and bottom panels show the results when two Hernquist profiles are jointly fitted, with the middle panel just showing the mass from the first galaxy profile and the bottom panel showing the sum of the two profiles. It is clear in all cases that individual, central, massive galaxies have substantial stellar mass evolution since their formation, in-line with the conclusions drawn throughout this thesis.

I next test the methodology used in Chapter 3. In that chapter, I compared the empirical estimates of the stellar mass of local central galaxies to the total baryonic mass available

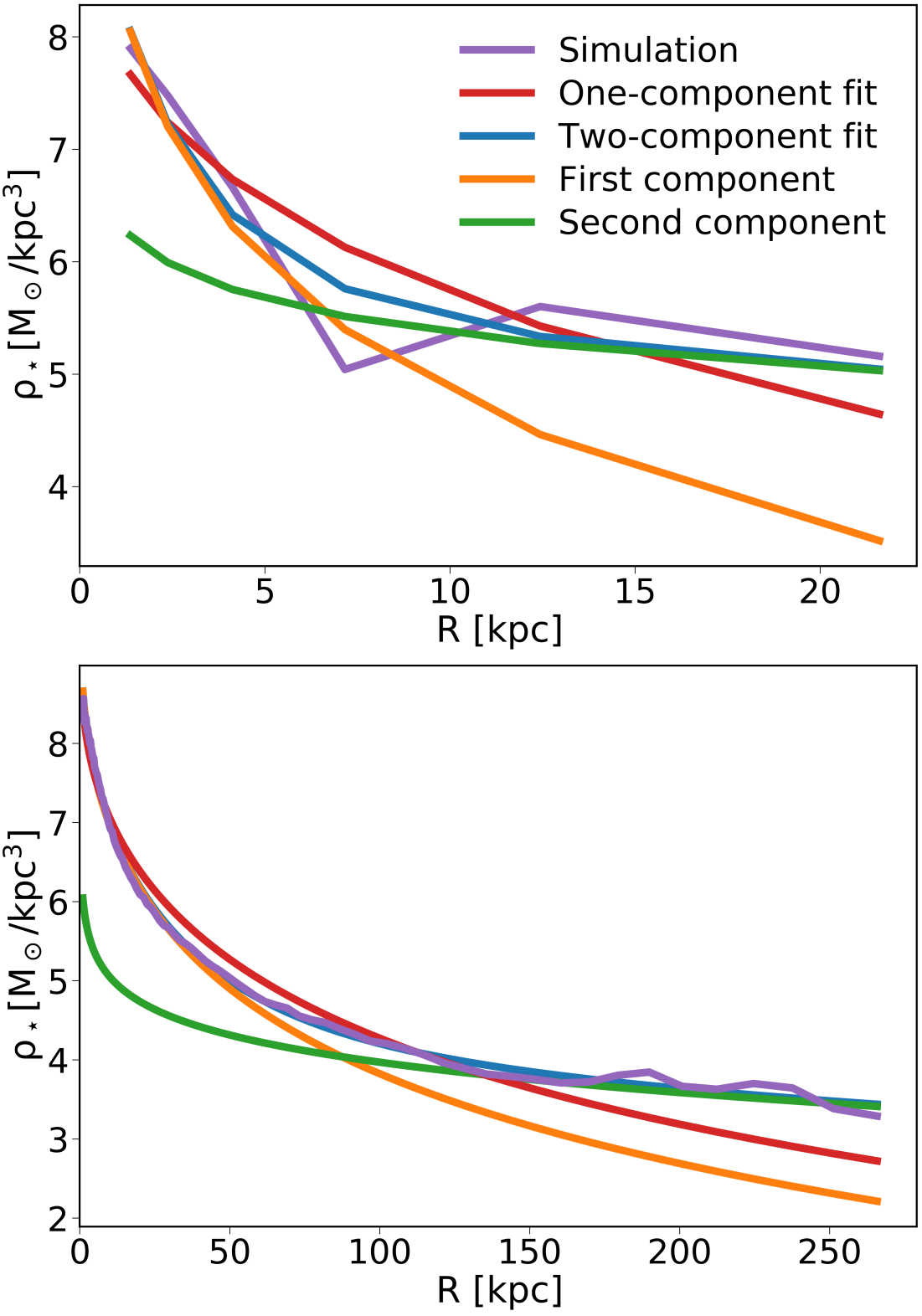



FIGURE 5.5: The purple line shows the 3D density profile of the central galaxy of the simulation run in this work at $z \sim 4.3$ (top) and one example fit from the Martizzi et al. (2014) simulations at $z = 0$ (bottom). The red line shows the best fit single Hernquist profile. Instead, the blue line shows the best fit double Hernquist profile, where the orange line shows the first 'galaxy' profile and the green shows the second 'ICL' profile.

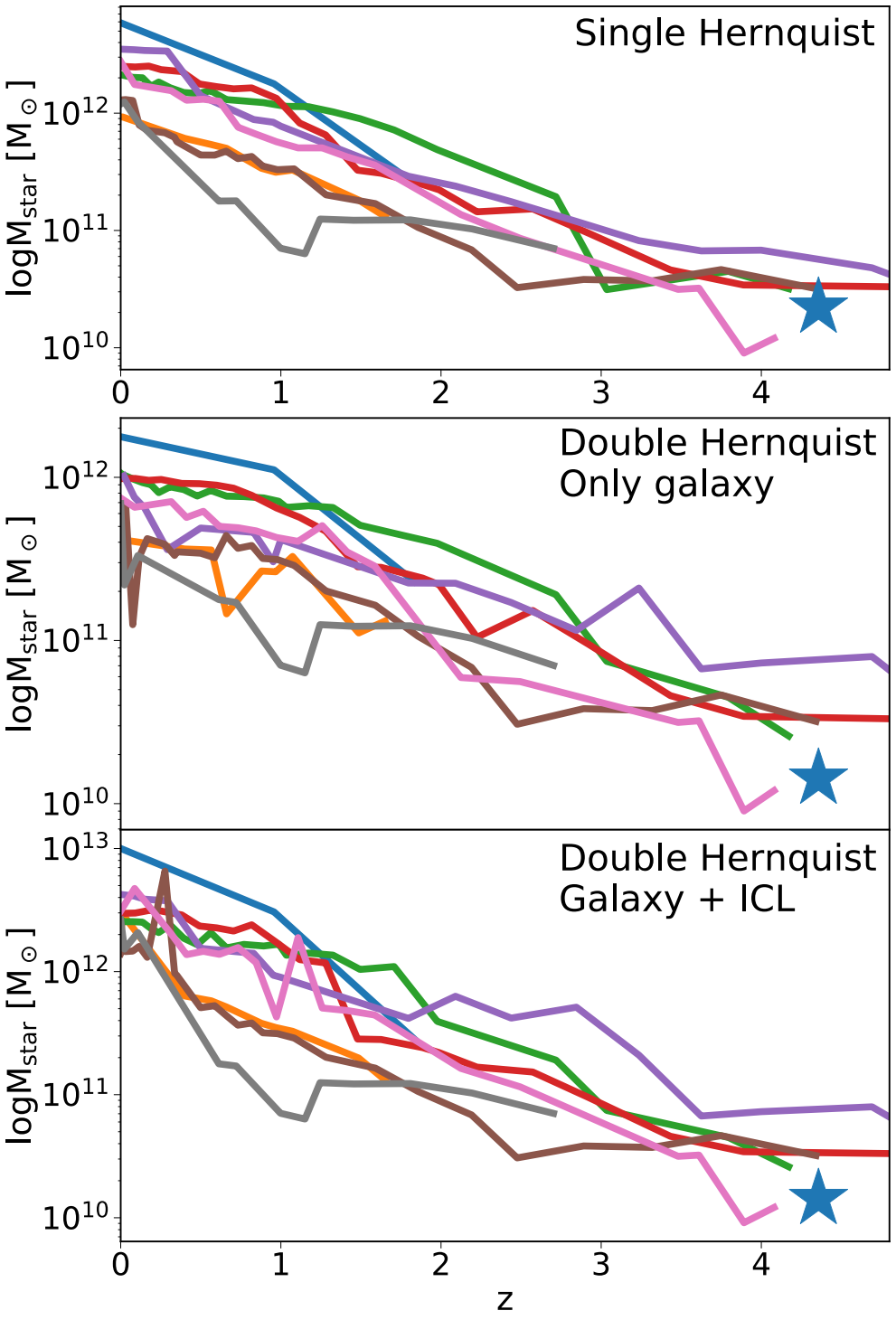



FIGURE 5.6: The coloured lines show the evolution of $M_{\rm star}$ in each of the central galaxies of the zoomed-in dark matter halo in the Martizzi et al. (2014) simulations. The blue star shows the $M_{\rm star}$ of the central galaxy of the zoomed-in dark matter halo in my simulation at $z \sim 4.3$. The top panel shows $M_{\rm star}$ when only a single Hernquist profile is used. The middle panel shows $M_{\rm star}$ measured using the first 'galaxy' profile and the bottom panel shows the sum of the 'galaxy' and 'ICL' profiles when two Hernquist profiles are fitted.

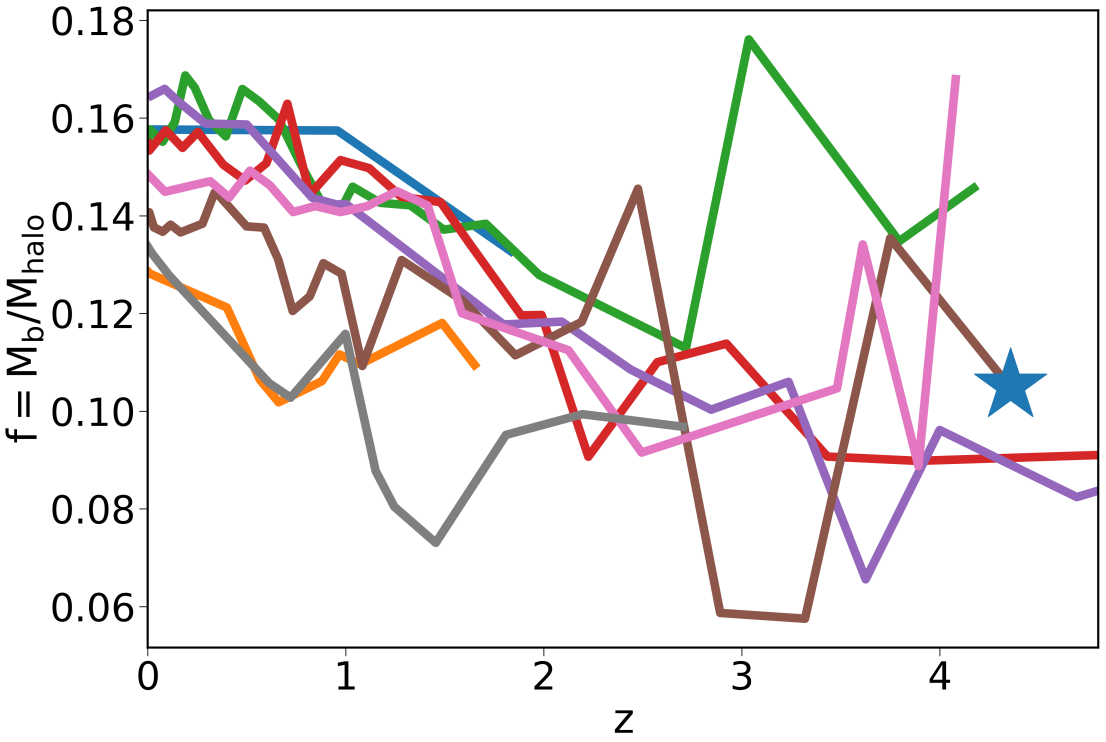



FIGURE 5.7: The mean ratio of the baryonic (stars plus gas) mass to the dark matter mass within $R_{200c}$ in the zoomed-in dark halo in the Martizzi et al. (2014) simulations (coloured lines) and of this work (blue star).

for galaxy formation inside their progenitor halo at high redshift. I found that the local stellar mass was equal to, if not greater than, the total baryonic mass at high redshift and thus concluded that not all of the local stellar mass could have been formed at those early epochs.

Firstly, I test the results of Crain et al. (2007) who find that the ratio of baryons to dark matter in haloes is $\sim$ 16%, independent of halo mass and redshift. This is an important input into my analysis in Chapter 3. Figure 5.7 shows the ratio of baryons to dark matter inside $R_{200c}$, as a function of redshift, in each of the Martizzi et al. (2014) simulations. The star shows the ratio inside the zoomed-in halo in my simulation. I find a slightly lower baryon fraction than Crain et al. (2007) which would, in fact, strengthen the conclusions in Chapter 3.

Figure 5.8 directly shows the total baryonic mass inside $R_{200c}$ for each simulation as a function of redshift, and the stellar mass of the central galaxy from the two component fit (the most conservative of the three measurements). It is again clear that the baryonic mass inside the halo at $z \gtrsim 3$ is lower than the stellar mass of the central galaxy at $z = 0$.

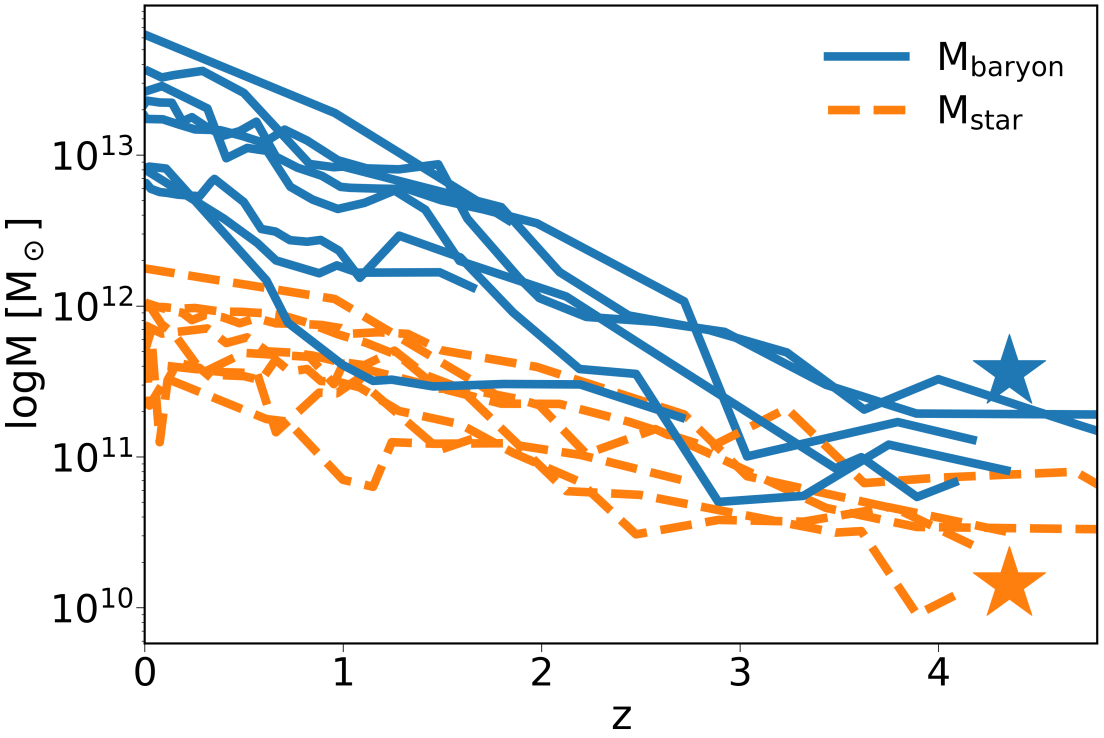



FIGURE 5.8: The evolution in the stellar mass of the central galaxies in the zoomed-in halo in the Martizzi et al. (2014) simulations (dashed orange lines) and of this work (orange star). Each line represents a different simulation. Additionally, I show the evolution of the total baryonic (stellar and gas) mass within $R_{200c}$ (blue solid lines and blue star).

#### 5.3.2.3 Estimating the half-mass radius

In addition to extracting the stellar mass of the central galaxy, the fits to the density profile also provide the half-mass radius. Figure 5.9 shows evolution of the half-mass radius, $R_{0.5M}$ of the central galaxy in the zoomed-in haloes of the Martizzi et al. (2014) simulations (lines) and mine (star). It is clear that there is substantial size evolution when only one Hernquist profile is fitted (top panel). However, it is less clear if the size of the central galaxy evolves when it is de-blended from the ICL (middle and bottom panels).

### 5.3.3 Conclusions

The results of this section clearly indicate that the stellar mass of the central galaxy evolves with redshift. There are two avenues to further explore the mechanisms that contribute to the stellar mass growth. The first is to trace the star particles back to their formation to test if they were created within the central galaxy, or ex-situ and bought in via mergers. The second is to probe the growth of the ICL by observing where the

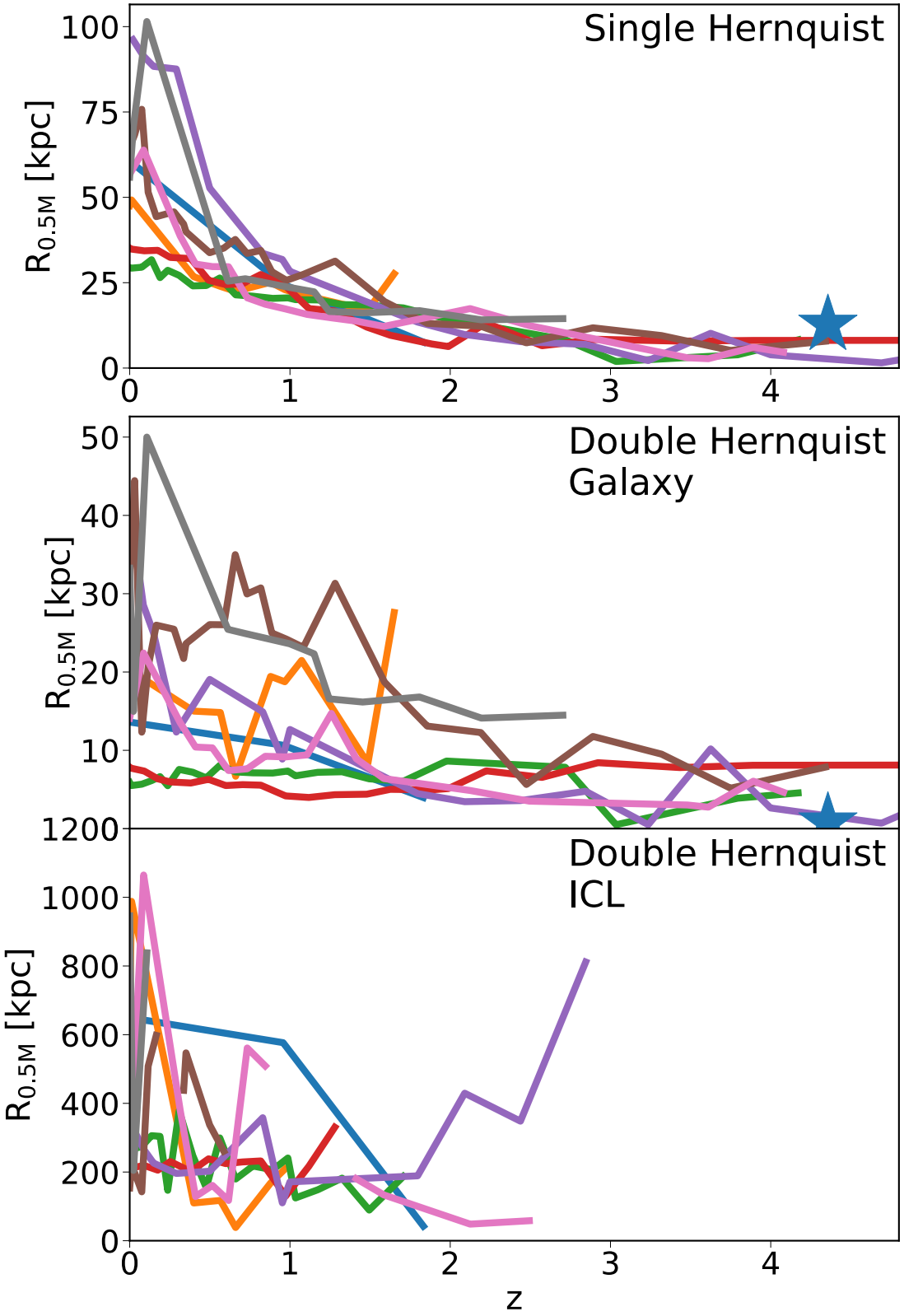



FIGURE 5.9: Same as Figure 5.6, but showing the evolution the half-mass radius, $R_{0.5M}$.

star particles that initially started in satellites end up, as a function of the stellar mass of the satellite and of their orbit.

The question whether or not these simulations support substantial size evolution in massive galaxies depends on how exactly the galaxy is de-blended into the galaxy and ICL components. This possibly needs higher spatial and temporal resolution simulations since high redshift, so that either more complex template profiles can be used or so that the two components can be probed via their dynamics.

Overall, the results of these cosmological zoom-in simulations seem to confirm the conclusions I drew throughout this thesis that the most massive galaxies have substantial growth since their formation. More detailed simulations will be run in the near future once technical issues with the RAMSES code have fixed, providing a more statistically significant sample of massive galaxies to draw conclusions from.

# Chapter 6

# Conclusion.

This thesis aimed to set more stringent constraints on the still debated formation and evolution scenarios for the most massive galaxies. To achieve this, I developed a series of semi-empirical models to answer the following questions:

1. On average, how much do massive galaxies grow in stellar mass since their formation epoch?
2. Can mergers alone account for such growth in size and stellar mass?
3. Can a basic monolithic collapse scenario account for the observations?

In the following section, I detail my findings, and use them to answer the questions posed above.

## 6.1 Main conclusions

### 6.1.1 Chapter 2

In Chapter 2, I first showed that by comparing basic abundance matching predictions to observations of massive galaxies, the high mass slope of the $M_{\rm star}$-$M_{\rm halo}$ relation appears to be substantially steeper than the commonly used relations in the literature, and the scatter small ($\sim 0.15$ *dex*). Secondly, I showed that there are signs of substantial growth in stellar mass $z \lesssim 1$. However, this latter conclusion rests on the accuracy of stellar masses measurements and the estimates of co-moving volumes at $z \sim 1$.

I also developed a state-of-the-art, semi-empirical model from scratch, following the general methodology laid out in Shankar et al. (2014a). By construction, this model has

a minimum number of input assumptions and parametrisations. The model takes dark matter merger trees as input and gives the stellar mass, gas fraction and effective radius as functions of redshift as output.

By inputting merger trees which are drawn from a cosmologically significant sample, I showed that massive galaxies can grow by a factor two in stellar mass between $z = 1$ and $z = 0$, consistent with other studies. Additionally, massive galaxies can grow in size by (up to) a factor of four over the same redshift range, providing that mergers are efficient (satellites have parabolic orbits). Finally, massive galaxies have $\sim 1$ major mergers between $z = 1$ and $z = 0$ and $\sim 8$ minor mergers.

### 6.1.2 Chapter 3

In Chapter 3, I set tighter constrains on the assembly and evolution of massive, central galaxies. I utilised a catalogue of dark matter haloes created from the Bolshoi simulation. I populated these haloes with stellar mass using the stellar mass to halo mass relation by Kravtsov et al. (2014) and Shankar et al. (2014b) at $z = 0$, and selected haloes with log $(M_{\rm star}) > 11.5\ M_{\odot}$. I then traced host haloes back to the putative formation epoch, $z_{\rm form}= 2 - 4$, as inferred from the stellar ages of massive ETGs. At this epoch, I estimated the total mass in baryons within the halo from the baryon fraction. I found that the stellar mass of early-eype galaxies in the local universe is comparable to, if not higher than, the total baryonic mass contained within the progenitor halo. From this comparison, I drew the following important conclusions.

1. In-situ formation: For these massive galaxies to have fully assembled at the formation epoch, the efficiency of converting baryons into stars needs to be extremely high, if not 100%. I also show that this assembly scenario would lead to all ETGs being extreme outliers with respect to what is predicted by abundance matching at $z_{\rm form}$.

2. Size: Even when assuming an extremely efficient star formation at $z_{\rm form}$, the galaxy would not be allowed any size growth since the formation epoch. Even an in-situ expansion would in fact require a mass loss of $\geq 70\%$ of the initial baryon content to be sufficiently efficient. Thus, in a strictly monolithic scenario, progenitors of massive galaxies should already be extended systems at their formation epoch. Measurements of the structure of massive galaxies in massive haloes will be critical to assess this possibility.

3. Late assembly: Star formation could contribute to the stellar mass growth of the progenitors of massive galaxies, but cannot explain their full evolution. I show

through state-of-the-art, cosmological, semi-empirical models that massive galaxies could have indeed assembled most of their final mass via late mergers and be consistent with available data on their size evolution. It remains to be seen the impact of mergers on other (tight) galaxy scaling relations involving velocity dispersion (e.g., Bernardi et al., 2011a,b; Shankar et al., 2016, and references therein).

### 6.1.3 Chapter 4

In Chapter 4, I outlined a promising framework to set valuable constraints on the slope and scatter of the high-mass end of the $M_{\rm star}$-$M_{\rm halo}$ relation. The methodology was specifically designed to bypass the substantial and unavoidable systematics in galaxy stellar mass and number density measurements which become gradually more prominent with increasing redshift. To this purpose, I utilised the halo mass distribution of massive, $\log M_{\rm star}/M_{\odot} > 11.5$ BOSS *satellite* galaxies at $z = 0.5$, as independently inferred by Tinker et al. (2016) from abundance matching and clustering techniques, to constrain the $M_{\rm star}$-$M_{\rm halo}$ relation at $0.5 < z \lesssim 1.2$. My method was to populate satellite haloes with stellar masses at their respective infall redshifts with $M_{\rm star}$-$M_{\rm halo}$ relations with varying high-mass end slopes $\delta$ and scatter $\sigma$. After infall, satellites are left frozen in stellar mass, or are allowed to evolve in stellar mass via stripping with and without star formation. I utilised an MCMC routine to fit the high-mass end slope and scatter in each respective model to the $n_{\rm gal}(M_{\rm halo})$ distribution.

The main results in Chapter 4 can be summarised as follows:

1. When assuming satellites to be strictly frozen after infall, the best-fit values were $\delta_{z>0.5} \sim 0.38$ and $\sigma_{z>0.5} \sim 0.19$ dex. The former implied a slope much steeper than what previously suggested in the literature at these redshifts ($\delta \lesssim 0.25$; Behroozi et al. 2013b; Moster et al. 2013). Still, the inferred slopes were significantly lower than what expected from a pure no-evolution scenario in the stellar mass function, implying some stellar mass growth in massive galaxies in the redshift range $0.5 < z < 1.0$. A small and nearly constant scatter is in line with what expected from a hierarchical assembly scenario.

2. Stellar stripping of massive satellites after infall must have been relatively small, $\lesssim 10\%$. Larger amounts of stripping would have implied unrealistically high-$z$ number densities of massive galaxies. This in turn implies that intracluster light should have originated from lower-mass and younger satellites.

3. Including star formation in infalling satellites, at the rate observed in massive galaxies at $z < 1$, does not alter any of the above conclusions.

### 6.1.4 Chapter 5

In Chapter 4, I showed the results of two projects involving numerical simulations that were designed to enhance and check the results presented in the previous chapters. First I utilised N-body simulations to explore if quasi-adiabatic expansion (as described in Section 1.6.2.1) can explain the size evolution in massive galaxies. Secondly, I explore the evolution of massive galaxies in tailored, cosmological zoom-in simulations.

I first explored if the morphologies of massive galaxies can effect the dynamical quasi-adiabatic expansion caused by strong mass-loss. To this end, I simulated an initially over-massive (with respect to abundance matching relations), isolated galaxy with the N-body code, GADGET2. After one simulated time step, I randomly removed a fraction of the particles such that it now follows abundance matching relations. I found that the galaxy does indeed expand, where the fractional change in radius was roughly proportional to the fractional change in mass ($M \times R = constant$). Additionally, I found this is independent of the exact morphology of the galaxy.

Furthermore, I ran a test to check the significance of the observed expansion by simulating an isolated galaxy with similar properties to those which I removed mass from. I found that numerical effects may play a moderate role in the expansion when only a modest fraction of baryons are removed but will not impact the conclusions when I remove a more extreme fraction of the baryons.

Secondly, I conducted a complementary project to verify the results presented in the previous chapters utilising cosmological zoom-in simulations. Specifically, I re-analysed eight simulations by Martizzi et al. (2014) as well as (starting to run) my own tailored simulation.

The conclusions of the project can be summarised as follows:

- The results of the simulations clearly indicated that the stellar mass of the central galaxy evolves with redshift, both when the galaxy is and is not de-blended from the ICL. However, I did not explore the exact mechanism for stellar mass growth.
- I verified the conclusions of Chapter 3 that the total mass in baryons is $\lesssim 16\%$ of the halo mass. Additionally, I showed that the the stellar mass of giant elliptical galaxies at the centres of clusters is higher than the total mass in baryons at $z_{\rm form} \gtrsim 2$.

- Whether or not these simulations support substantial size evolution in massive galaxies depends on how exactly the galaxy is de-blended from the ICL.

### 6.1.5 Answering the key questions

Here I answer the three questions posed at the start of this thesis.

1. It is highly probably that massive galaxies assemble a substantial proportion of their stellar mass at later epochs.
2. If the satellite galaxies have, on average, parabolic orbits, they can fully account for the increase in stellar mass and size of massive early type-galaxies between the epoch of formation and the present day.
3. A simple monolithic collapse scenario could still be consistent with observations so long as they are born extended. However, numerical simulations do not support this picture.

## 6.2 Future work

Work has already been undertaken by our group to progress the research presented in this thesis. Here I detail some extensions to my analysis that will be submitted this year.

Velocity dispersion has been added to the semi-empirical model I presented in Chapter 2 which allows us to probe galaxy formation using the fundamental plane (Grylls et al. in prep.). In addition, black holes will be added to the model so we can probe their growth, thus giving a more complete view of massive galaxy formation.

The cosmological zoom-in simulation I presented in Chapter 5 will be completed, allowing us to probe the formation of massive galaxies in more detail. We will isolate the stars formed in satellite galaxies to test how efficiently they merge with the central galaxy, and how much they contribute to the ICL.

The methodologies I presented in this thesis have provided more secure constraints on massive galaxy formation and evolution. The next generation of great observatories (e.g., Euclid, LSST, E-ELT JWST) will allow for more precise and statistically relevant measurements of the stellar mass, structure and clustering properties of high redshift, massive proto-galaxies. This will be the key to discern between merger scenarios and extremely efficient starbursts events.

# Appendix A

# Generating merger trees.

## A.1 Analytic trees

The first method to generate merger trees is to use Press-Schechtor formalism (Press & Schechter, 1974) to construct the halo mass function at each redshift. It is also possible to construct a conditional mass function of progenitor haloes. In more detail, this is the fraction of trajectories that an object at time $t$, with mass between $M$ and $M + dm$ that make their first uncrossing through the threshold $\omega = \delta_c(t)$ with variance in the range $S$ and $S + dS$. These variances can be converted to masses such that this conditional mass function becomes

$$f(M_1, z_1|M_0, z_0) = f_{sc}(S_1, \omega_1|S_0, \omega_0)|\frac{dS_1}{dM_1} \tag{A.1}$$

$$f_{sc}(S_1, \omega_1|S_0, \omega_0) = \frac{1}{\sqrt{2\pi}}\frac{\Delta\omega}{\Delta S^{3/2}} exp\left(-\frac{\Delta\omega^2}{2\Delta S}\right) \tag{A.2}$$

where $f(M_1, z_1|M_0, z_0)$ is the fraction of mass represents the fraction of mass from haloes of mass $M_0$ at redshift $z_0$ that is contained in progenitor haloes of mass $M_1$ at an earlier redshift $z_1$.

The first useful quantity to define is the average number of objects with mass $M_1$ that an object with mass $M_2$ 'fragments into' in a time step of $dt_1$. This is given by

$$\frac{dN}{dM_1} = \frac{df_{sc}}{dt}\frac{M_2}{M_1}dt \tag{A.3}$$

It then follows that the mean number of haloes that fragment from a halo with mass $M_2$, given a resolution limit ($M_{res}$), is given by:

$$P = \int_{M_{res}}^{M_2/2} \frac{dN}{dM_1} dM_1 \tag{A.4}$$

Additionally, the mass that gets accreted onto halo over this $dz$

$$P = \int_{0}^{M_{res}} M_1 \frac{dN}{dM_1} dM_1 \tag{A.5}$$

The first, most simple algorithm to generate merger trees is Cole et al. (2000) which works as follows. Starting with a halo with mass $M_2$ at $z$, a random number, R is drawn. If $R > P$ then the halo does not split in that time step. However, the accreted mass, given by Equation A.5, is still deducted from the main halo so that the new main progenitor has mass $M_2(1 - F)$. However, if $R < P$, the halo splits. In this case, a random value $M_1$ is generated from the conditional mass function at that redshift, in the range $M_{res} < M_1 < M_2/2$. Accreted mass is also deducted such that the main progenitor has mass $M_2(1 - F) - M_1$. This is then looped until the main progenitor reaches a given value or time reaches a given epoch.

The second, more complex algorithm is by Somerville & Kolatt (1999) who allow for multiple mergers per time step. The algorithm works by drawing a random number which is converted to a mass $M_1$ in the range $0 \leq M_1 \leq M_2$ using equation A.5. If $M_1 < M_{res}$ then $M_1$ is added to the accreted mass. If $M_1 > M_{res}$ it is a progenitor. If the remaining mass $\Delta M = M_2 - M_1$ is less than $M_{res}$ it is added to the accreted mass. If, however, $\Delta M > M_{res}$ then it will contain another progenitor. In this case, another random mass, $M_3$ is drawn from the mass weighted probability distribution. If $M_3 > \Delta M$ then it is thrown away and another mass is drawn. If $M_3 < M_{res}$ it is added to the accreted mass and if $\Delta M > M_3 > M_{res}$ it is another progenitor. In the latter case, $\Delta M = M_1 - (M_2 + M_3)$. This is repeated until $\Delta M < M_{res}$.

The last, most complex algorithm by Parkinson et al. (2008) claims to be more faithful to numerical simulations. To achieve this, Parkinson et al. (2008) introduce the perturbation function to equation A.3

$$\frac{dn}{dM_1} \rightarrow \frac{dn}{dM_1} G(\sigma(M_1)/\sigma(M_2), \omega_2/\sigma_2) \tag{A.6}$$

where $G(\sigma(M_1)/\sigma(M_2), \omega_2/\sigma_2)$ is chosen such that it is of order unity over most of masses, mass ratios and redshifts of interest. Specifically, Parkinson et al. (2008) choose to use:

$$G(\sigma(M_1)/\sigma(M_2), \omega_2/\sigma_2) = G_0 \left(\frac{\sigma_1}{\sigma_2}\right)^{\gamma_1} \left(\frac{\omega_1}{\sigma_2}\right)^{\gamma_2} \tag{A.7}$$

This has the advantage that the algorithm maintains the speed of the Cole et al. (2000) algorithm but the merger rate can be fitted to numerical simulations. The values of $G_0, \gamma_1$ & $\gamma_2$ were fitted by Parkinson et al. (2008) such that the merger trees their algorithm produce match statistical properties of merger trees produced by Cole et al. (2008) from the Millennium simulation. Throughout this work we adopt their best fir parameters of $G_0 = 0.57, \gamma_1 = 0.19, \gamma_2 = -0.005$. However, in principle these values could be refit to match trees produced from any numerical simulation with any reasonable ΛCDM cosmology.

## A.2 Numerical trees

In this Appendix Chapter, I compare different extended-Press-Schechter based algorithms for generating merger trees, along with numerical merger trees, with the aim of selecting the method for use in my semi-empirical model. An increasingly popular method for generating merger trees is to directly use the results of large volume, dark matter-only n-body simulations (eg. Millennium Springel 2005a, Bolshoi Klypin et al. 2011). Over the past three decades, increasing computing power has allowed astronomers to run large volumes and higher resolutions to understand exactly how dark matter structures grow. The exact details of these simulations and the codes vary greatly, however, generally they work either by using collisionless particles or using an adaptive mesh refinement code. In either case, the simulation is initialised as a 3D, random density field following a power spectrum defined by a set of cosmological parameters at high redshift. This approximates the mass distribution of dark matter soon after inflation has ended. In all cases, they use periodic boundary conditions to avoid edge effects.

This random density field is then allowed to evolve under gravity to the present day forming collapsed structures. In principle, this can be done by finding the gravitational acceleration on every particle (cell) due to every other particle (cell). However, because this is scales as the number of particles, $n_p$, as $n_p(n_p - 1)$ more sophisticated algorithms (e.g., tree particle-mesh (eg, GADGET Springel, 2005b), multigrid relaxation (e.g., RAMSES Teyssier, 2002), fast multipole (e.g., PKDGRAV3 Potter et al., 2016), etc) are implemented to calculate forces.

At each snapshot of the simulation, so called 'halo finders' are run to identify these bound structures. These generally work by finding groups of particles in close proximity (FoF algorithms) or by looking for peaks and troughs in the density field (density peak finders). More complex algorithms such as (to name but a few) bound density maxima (BDM Klypin & Holtzman, 1997), AMIGA halo finder (AHF Knollmann & Knebe, 2009), SUBFIND (Springel et al., 2001) and ROCKSTAR (Behroozi et al., 2013a) build on these two techniques.

Finally, haloes are linked together between snapshots by tagging all the particles in each halo and following them between snapshots. Haloes in different snapshots are linked together if they share a significant number of particles. Mergers occur when two haloes are no longer distinguishable.

## A.3 Comparison

Dark-matter-only simulations have the strong advantage that they do not make assumptions about the collapse of dark matter haloes and the thus haloes are more representative of real haloes. However, this realism makes defining mergers very difficult during messy interactions. The semi-empirical model is very sensitive to fluctuations in halo mass because of the method used to keep main sequence galaxies following empirical relationships. Simulations are also limited by their modest volume and resolutions making it more difficult to probe very massive galaxies. Therefore, it is advantageous to use analytic merger trees for their simplicity providing that they faithfully reproduce the evolution predicted by numerical simulations.

In this section, I compare merger trees from SK99, C00 and P08 to the Bolshoi simulation. To this end, we assume that the Bolshoi simulation (Klypin et al., 2011, the reference simulation used throughout this thesis) accurately predicts the true evolution of dark matter haloes and will find the best analytic algorithm that matches it. The first important property of merger trees is that the main progenitor is growing at the correct rate. This is important for the semi-empirical model since the main progenitor is populated with a galaxy at high redshift using an $M_{\rm star}$-$M_{\rm halo}$ relation. It is thus important to get the correct progenitor halo mass to ensure the seed galaxy mass is correct. Figure A.1 shows the median mass growth of the main progenitor and its $1\sigma$ dispersion produced using the three algorithms and from the Bolshoi simulation (the four panels as labelled) and for four initial $z = 0$ halo masses (four colours as labelled). This figure shows the well-documented trend that the more massive haloes grow at a faster rate than low mass haloesMcBride et al. (2009); van den Bosch et al. (2014); Correa et al. (2015). Instead, Figure A.2 shows difference the in mass growth between

the three analytic algorithms (coloured lines) and the trees from the Bolshoi simulation for four different initial halo masses (four panels as labelled). I find that the Parkinson et al. (2008) algorithm most closely matches the simulation.

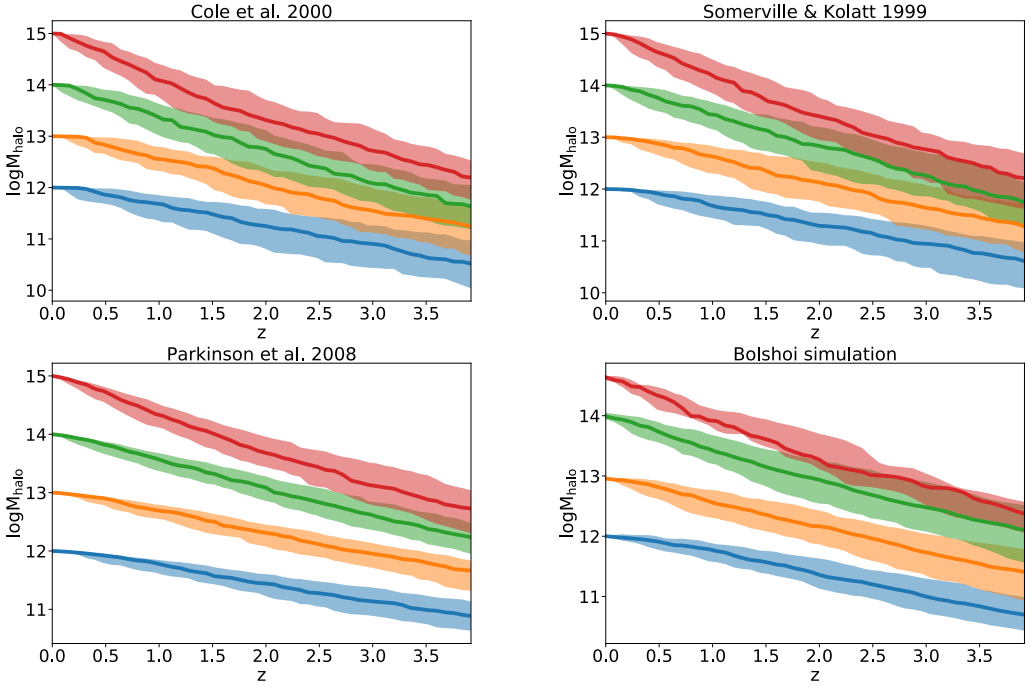



FIGURE A.1: The median and one standard deviation mass accretion histories of main progenitor haloes with $z = 0$ masses: $\log M_{halo} = 12, 13, 14$ & 15. The bottom right panel is constructed using haloes in the Bolshoi simulation. The other panels are constructed using algorithms based on extended Press-Schector theory. In all cases except $\log M_{halo} = 15$ in the Bolshoi simulation (due to volume limitations), 100 haloes are simulated.

Next, I compare the sub-halo mass functions produced from he different algorithms. This is effectively a measure of the merger rate and the average merger ratio of dark matter haloes. This is important for the semi-empirical model because it directly determines the rate of mergers of satellite galaxies with the main progenitor and the merger ratio. Figure A.3 shows a comparison between the subhalo mass function produced by creating merger trees from each analytic algorithm down to a lower resolution. The top panels show a comparison the sub-halo mass functions with the Giocoli et al. (2008) subhalo mass function, found from numerical simulations for two stating halo masses. The bottom panels show the difference between the subhalo mass functions found from merger trees and the one by Giocoli et al. (2008). I find that the Parkinson et al. (2008) algorithm matches much more closely. Because of these two tests, I conclude that Parkinson et al. (2008) is the best algorithm.

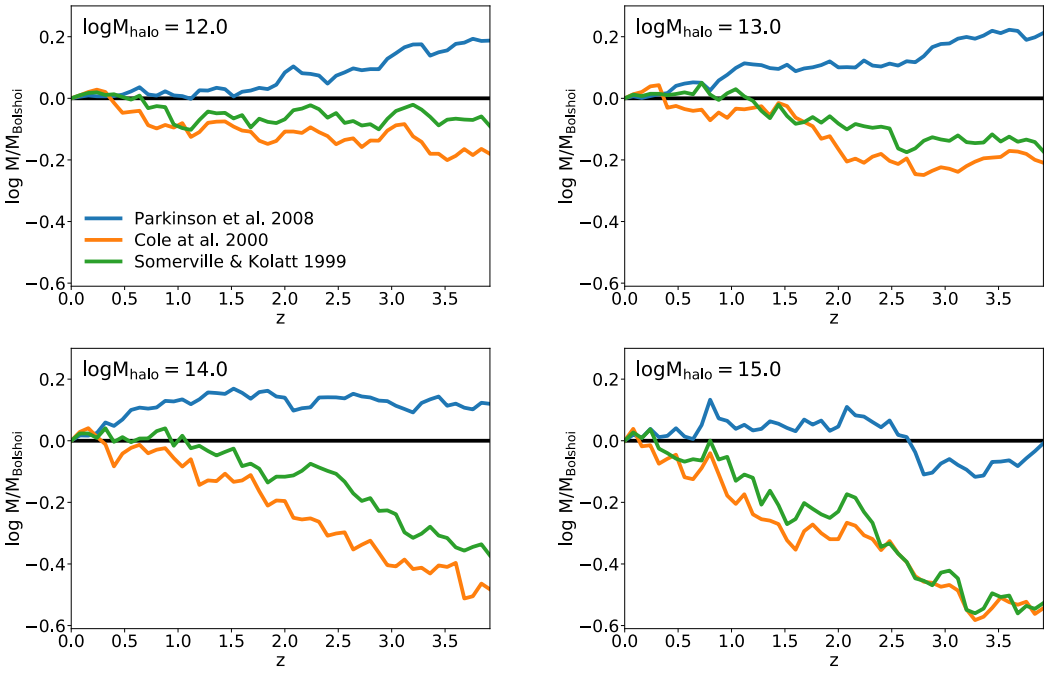



FIGURE A.2: A comparison between the mass accretion histories of three extended Press-Schector bases algorithms to that of haloes in the Bolshoi simulations.

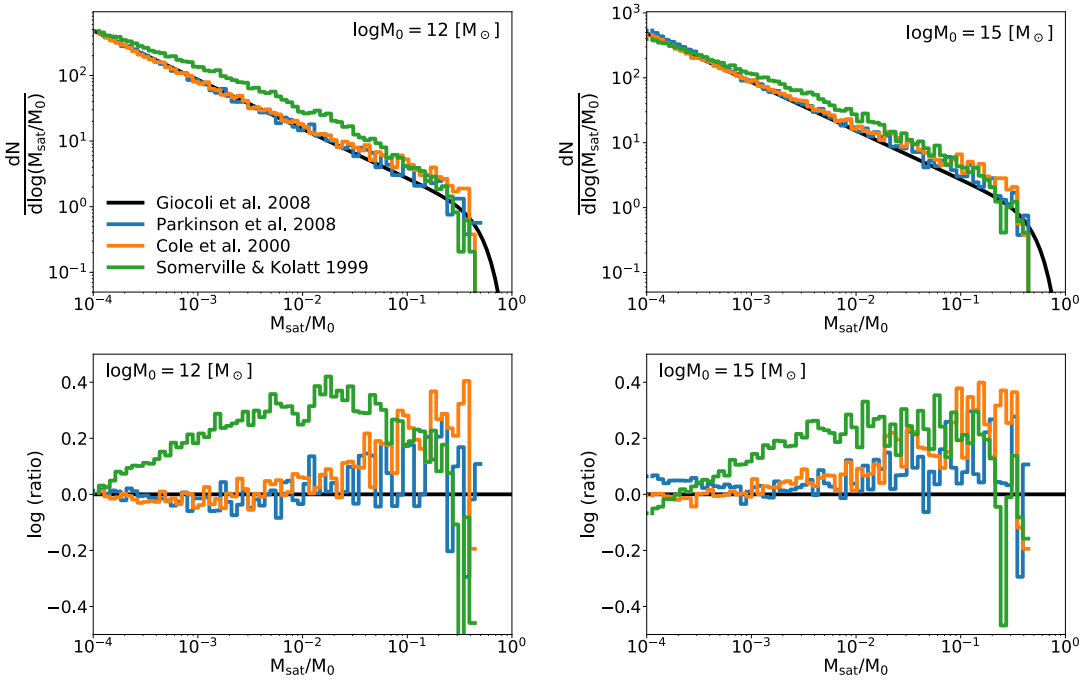



FIGURE A.3: A comparison between the mean $1^{st}$ order subhalo mass functions predicted with three extended Press-Schector bases algorithms for producing merger trees to analytic relation of Giocoli et al. (2008) for two initial halo masses.

# Appendix B

# Modelling the $z_{\text{inf}}$ graph structure with Neo4j

A critical component of the methodology incorporated in Chapter 4 is knowing the infall redshift distribution, $P(z_{\text{inf}} \| M_{\text{halo}})$. This is estimated using the Bolshoi simulation Klypin et al. (2011) which simulates the growth of dark matter structures over cosmological time. The data comprises of lists of dark matter haloes that exist at each redshift step and their relationship to past haloes. For many years this has been thought of as a tree structure (Lacey & Cole, 1993). To construct the infall redshift distribution it is necessary to trace the evolution of individual sub-haloes back through time until they are an independent halo. The size of the dataset and the requirement to query the tree structure within the data makes this very computationally inefficient if the data is stored as flat files or within a classic relational database (e.g. SQL). Consequently, a graph database was built to enable efficient querying and exploration of the data; the Neo4j[1] graph database was chosen as it is widely used within industry and has a community, open-source version (Robinson et al., 2015).

A Neo4j graph database comprises of two fundamental object types. The first are nodes that represent an entity of some variety; any node may have individual attributes and labels that specify its characteristics. The second object type are relationships that describe the connection between two nodes; these too may have labels and attributes. Within out representation of the Bolshoi simulation I constructed the following objects:

- Node: Halo nodes, each halo and sub-halo within the dataset was represented by a unique node.
- Node: Redshift nodes, each redshift step was given it's own unique node.

[1] Neo4j Open Source project

TABLE B.1: Properties of the Neo4j database: the number of nodes and relationships broken down by label.

| Label | # |
|---|---|
| Halo nodes | $10^9$ |
| Redshift nodes | 62 |
| IS_SUB_OF | $2 \times 10^8$ |
| IS_PARENT_OF | $10^9$ |
| AT_REDSHIFT | $10^9$ |

- Relationship: IS_SUB_OF, connected sub-haloes to its host halo.
- Relationship: IS_PARENT_OF, connected haloes to its parent halo at previous redshift step.
- Relationship: AT_REDSHIFT, connected haloes to the redshift node at which it exists.

A common way to describe example graph structure is to use ASCII art where nodes are indicated by `( )`, and relationships by `[ ]`. All relationships within Neo4j need to also have a direction and this is indicated by the `>`. Therefore an ASCII art representation of the sort of structure that would exist within the graph model of the data is:

```
(z:  z=1.66)<-[:IS_AT_REDSHIFT]-(HALO: id=34)-\\
[:IS_PARENT_OF]->(HALO: id=235)-[:IS_SUB_OF]->\\
(HALO: id=456)-[:IS_AT_REDSHIFT]->(Z: z=2)
```

As the dataset was static the fastest way to build the Neo4j database was to use the NEO4J-IMPORT tool. As a pre-processing step this required loading all of the breadth-first-search Bolshoi data files and extracting out unique lists of haloes and their associated relationships. These lists were then used to build the final graph representation of the database that was subsequently indexed on the redshift nodes. The properties of the final database are given in Table B.1. The size of the final database was 130GB compared to the ∼1TB of original Bolshoi simulation data files; however, it should be noted that other than the halo ids and redshifts the only additional attributes from the original data that were stored were the M200b and M200c values (these correspond to the mass enclosed within the over-density $200 \times \rho_b$ and $200 \times \rho_{crit}$ respectively; see Klypin et al. (2011) for more details). A desktop computer was used throughout but it required additional memory when constructing the graph database.

From storing the simulation data in this manner makes it then very easy and efficient to query the database to get the information required to estimate the infall redshift distribution, $P(z_{\rm inf} \| M_{\rm halo})$. This is because the merger trees are natural trees structures which is extremely efficient to traversing using Neo4j (see Robinson et al. (2015) for

comparisons between Neo4j with SQL when traversing data). In more detail, I search for all sub-haloes that exist at redshift 0.5 and then trace their "heritage" back in time until they are an independent halo. Neo4j databases are queried with the CYPHER query language that is designed to be intuitive as related to graph structures (Robinson et al., 2015). The exact query used to generate the data to estimate $P(z_{\rm inf}\|M_{\rm halo})$ as shown in Figure 4.4 of Section 4.2.1 is:

```
MATCH (r:redshift {z:0.5}) <-[:AT_REDSHIFT]-(h1:Halos)-\\
[:IS_SUB_OF]->(h2:Halos)
WITH h1 as myhalo
MATCH p=myhalo -[:IS_PARENT_OF*]->(h3:Halos)
WHERE NOT (h3)-[:SUB_OF]->() AND (h3)<-[:IS_PARENT_OF]-()-[:SUB_OF]->()
MATCH (h3)-[:AT_REDSHIFT]->(zinf:scale)
RETURN myhalo,h3,zinf
```

Additionally, this query can be run for fine bins of host halo mass. Figure B.1 shows how the mean $z_{\rm inf}$ depends on the host halo mass and the mass of the sub-halo at infall. I find that there is only a weak trend with host halo mass. This shows that it is safe to draw a random $z_{\rm inf}$ from the total distribution rather than needing to fit a specific $z_{\rm inf}(M_{\rm host}, M_{\rm inf})$ function.

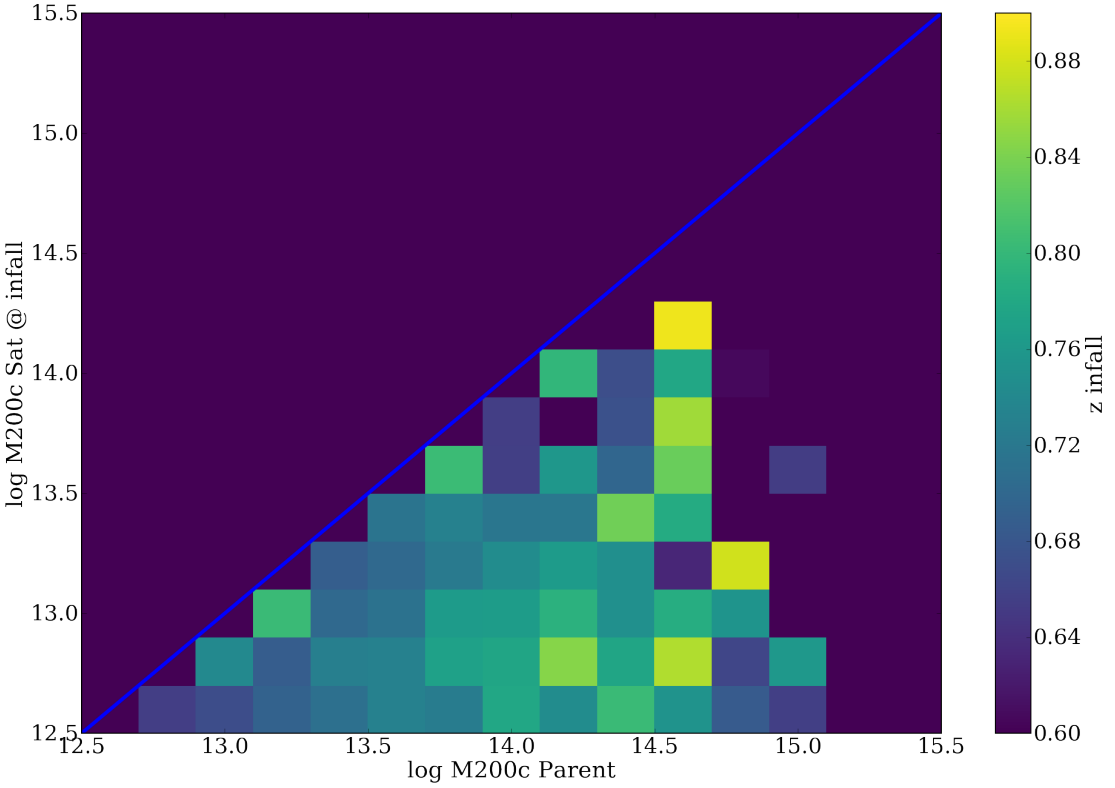



FIGURE B.1: The mean $z_{\rm inf}$ for different host halo masses and sub-halo mass at infall. I fond that there is only a weak trend with host halo mass.

# Appendix C

# Test the Guo et al. (2014) and Tinker et al. (2016) $dP/d\log M_{\rm halo}$ distributions

In this Appendix Chapter, I compare the central and satellite $dP/d\log M_{\rm halo}$ distributions from Guo et al. (2011) and Tinker et al. (2016) for use in Chapter 4. Initially, I had planned on using the halo occupations distribution (HOD) results from Guo et al. (2014) in Chapter 4. However, when using that data, the fitting routine would always fit a very steep slope with zero intrinsic scatter. Additionally, when using the same parameters as the centrals, the $dP/d\log M_{\rm halo}$ distribution would have roughly the correct normalisation but be systematically offset towards lower host halo masses.

To evaluate if there is an issue with the satellite HOD results of Guo et al. (2014), I compared them to the AB+HOD results of Tinker et al. (2016). They use different methods to measure mass which creates a 0.28dex systematic offset between the two. After correcting for this, I find that their $dP/d\log M_{\rm halo}$ distributions are in close agreement (shown in the top panel of Figure C.1). Conversely, I find that there is an offset in their respective satellite distributions (as seen in the bottom panel).

To test which satellite is correct, I devised a self consistency check. Guo et al. (2014) additionally have HOD results at $z = 0.7$ which is coincidently the mean redshift of infall for satellites at $z = 0.5$. Furthermore, the two-halo term in the HOD model gives the average number of central galaxies within the sample that have a halo mass $M_{\rm halo}$. Flipping this logic, since a halo always contains one central, it is the probability that halo of mass $M_{\rm halo}$ contains a central galaxy that is in the sample. The $z = 0.5$ and $z = 0.7$ probability distributions are are shown in Figure C.2 in blue and orange respectively.

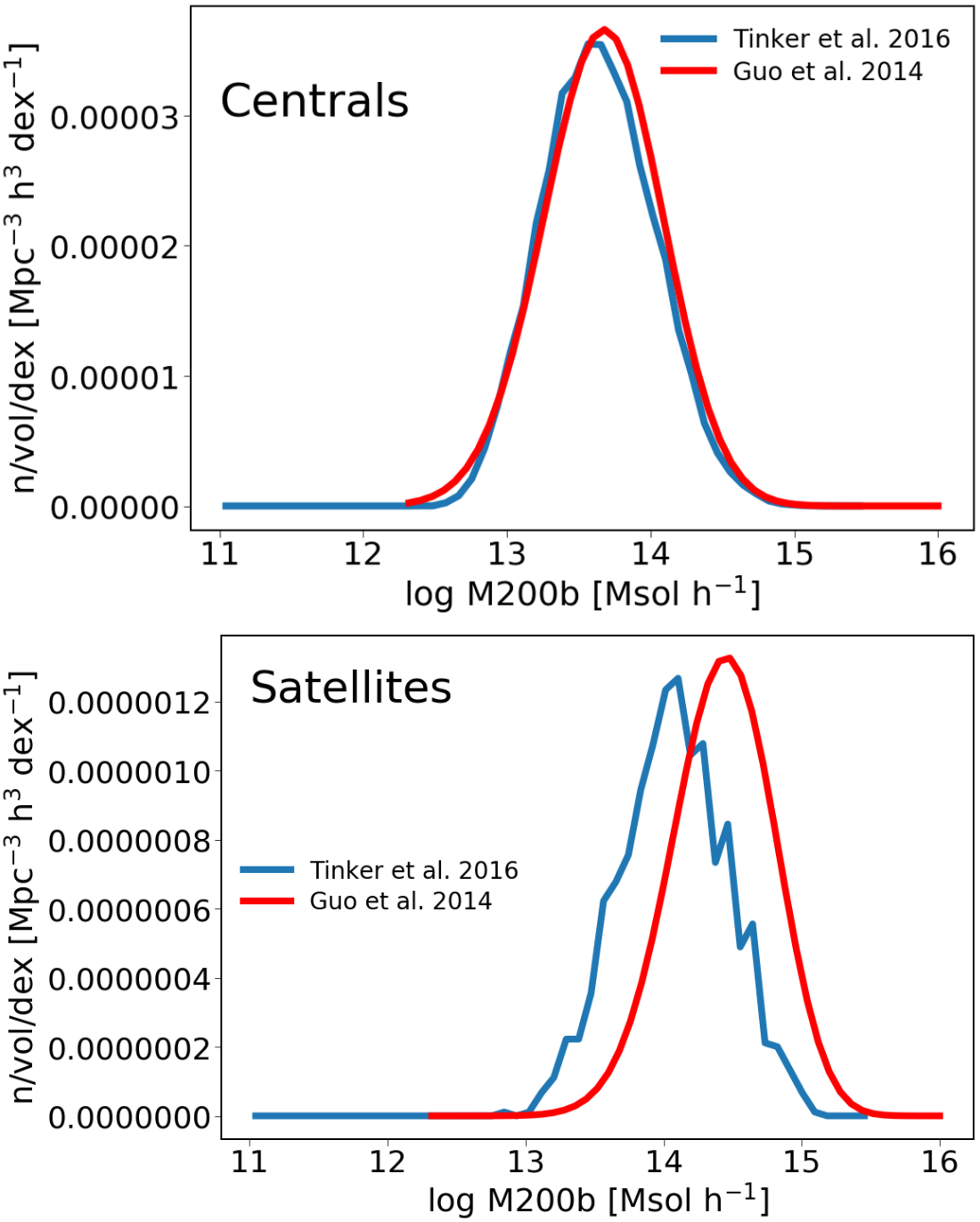



FIGURE C.1: A comparison between the central (top) and satellite (bottom) $dP/d\log M_{\rm halo}$ distributions from Tinker et al. (2016, Blue) and Guo et al. (2014, red).

For each sub-halo, I draw a random number, $R$ for a uniform distribution in the range $0 \leq R \leq 1$. If $R$ is less than the $P(\mathrm{LRG}|M_{\rm halo})$ the sub-halo contains a massive galaxy that is included in the sample.A comparison the satellite $dP/d\log M_{\rm halo}$ distributions with this self consistency check is shown in Figure C.3. It is clear that Tinker et al. (2016) is self-consistent whereas Guo et al. (2014) is not. Therefore, In Chapter 4, I utilise the central distributions from both Tinker et al. (2016) and Guo et al. (2014) but

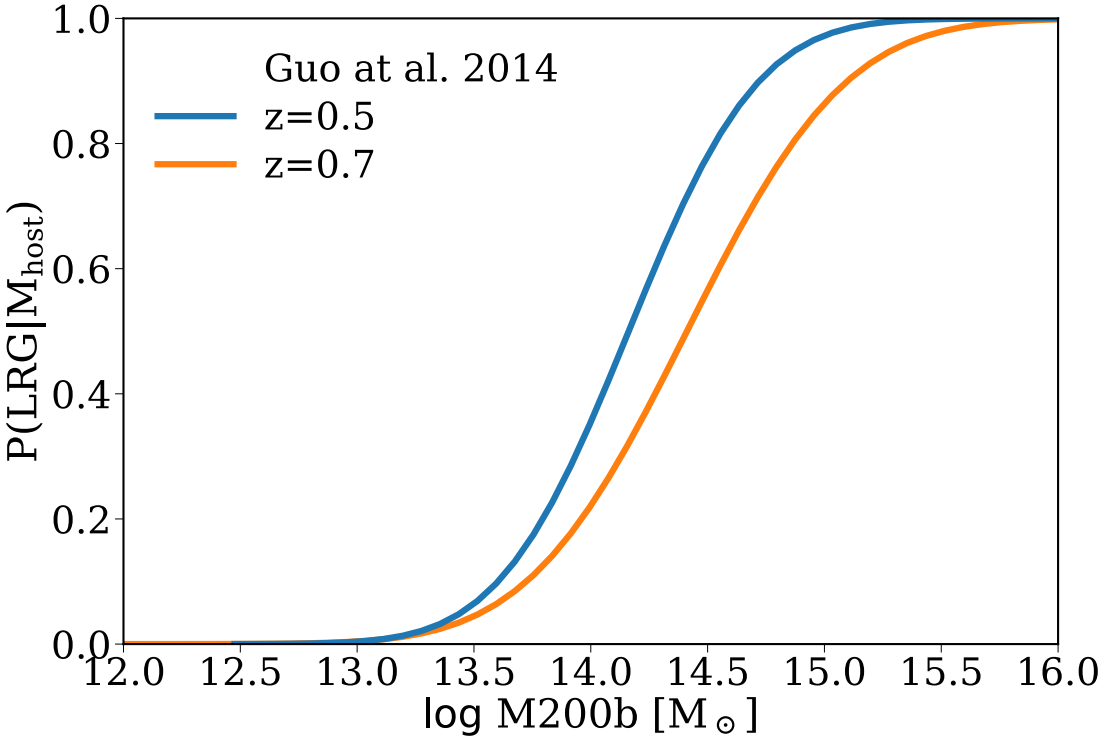



FIGURE C.2: The probability that a halo with a given mass contains a central galaxy with a mass $\log M_{\rm star} > 11.5[M_\odot]$ at $z = 0.5$ (blue) and $z = 0.7$ (orange) from Guo et al. (2014).

I only utilise the satellite distribution from Tinker et al. (2016).

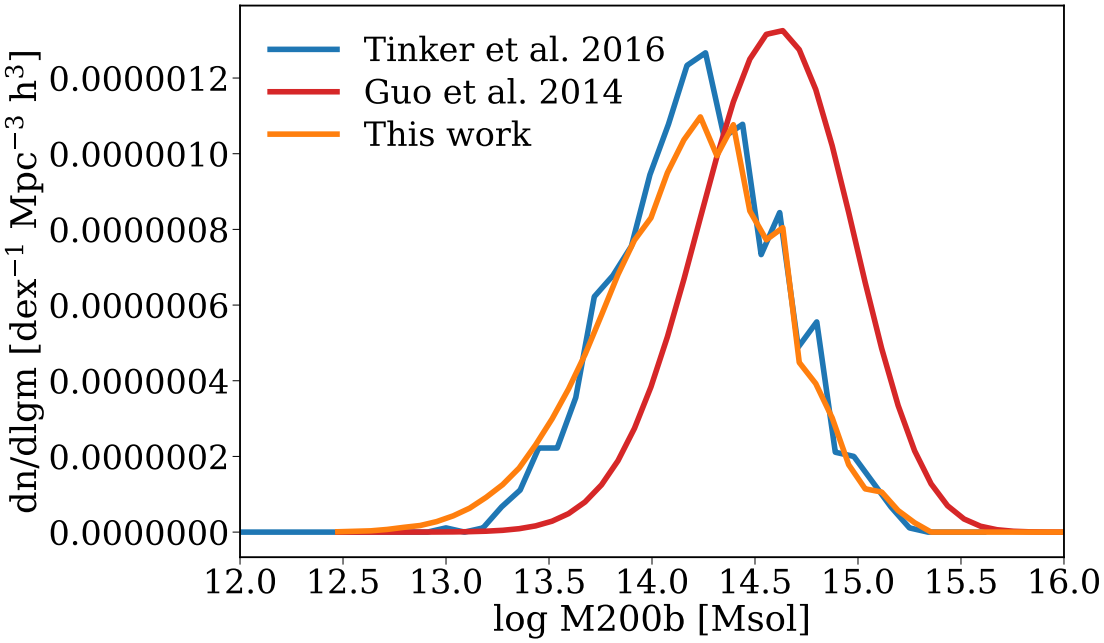



FIGURE C.3: A comparison between the satellite $dP/d\log M_{\rm halo}$distributions from Tinker et al. (2016, Blue) and Guo et al. (2014, red) with the self consistency test described in this Appendix.

# Bibliography


Alam S., et al., 2016, preprint, (arXiv:1607.03155)

Ascaso B., Lemaux B. C., Lubin L. M., Gal R. R., Kocevski D. D., Rumbaugh N., Squires G., 2014, MNRAS, 442, 589

Auger M. W., Treu T., Bolton A. S., Gavazzi R., Koopmans L. V. E., Marshall P. J., Moustakas L. A., Burles S., 2010, ApJ, 724, 511

Aversa R., Lapi A., de Zotti G., Shankar F., Danese L., 2015, ApJ, 810, 74

Babcock H. W., 1939, Lick Observatory Bulletin, 19, 41

Bai L., et al., 2014, ApJ, 789, 134

Baldry I. K., Balogh M. L., Bower R. G., Glazebrook K., Nichol R. C., Bamford S. P., Budavari T., 2006, MNRAS, 373, 469

Barnes J. E., 1988, ApJ, 331, 699

Barro G., et al., 2013, ApJ, 765, 104

Baugh C. M., 2006, Reports on Progress in Physics, 69, 3101

Behroozi P. S., Conroy C., Wechsler R. H., 2010, ApJ, 717, 379

Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013a, ApJ, 762, 109

Behroozi P. S., Wechsler R. H., Conroy C., 2013b, ApJ, 770, 57

Bell E. F., et al., 2006, ApJ, 640, 241

Bellstedt S., et al., 2016, MNRAS, 460, 2862

Bennett C. L., et al., 2013, ApJS, 208, 20

Benson A. J., 2012, New A, 17, 175

Bernardi M., et al., 2003, AJ, 125, 1866

Bernardi M., Shankar F., Hyde J. B., Mei S., Marulli F., Sheth R. K., 2010, MNRAS, 404, 2087

Bernardi M., Roche N., Shankar F., Sheth R. K., 2011a, MNRAS, 412, L6

Bernardi M., Roche N., Shankar F., Sheth R. K., 2011b, MNRAS, 412, 684

Bernardi M., Meert A., Sheth R. K., Vikram V., Huertas-Company M., Mei S., Shankar F., 2013, MNRAS, 436, 697

Bernardi M., Meert A., Vikram V., Huertas-Company M., Mei S., Shankar F., Sheth R. K., 2014, MNRAS, 443, 874

Bernardi M., Meert A., Sheth R. K., Fischer J.-L., Huertas-Company M., Maraston C., Shankar F., Vikram V., 2016a, preprint, (`arXiv:1604.01036`)

Bernardi M., Meert A., Sheth R. K., Huertas-Company M., Maraston C., Shankar F., Vikram V., 2016b, MNRAS, 455, 4122

Bernardi M., Meert A., Sheth R. K., Fischer J.-L., Huertas-Company M., Maraston C., Shankar F., Vikram V., 2017, MNRAS,

Blanton M. R., et al., 2003, ApJ, 592, 819

Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440

Boselli A., Gavazzi G., 2006, PASP, 118, 517

Bournaud F., Dekel A., Teyssier R., Cacciato M., Daddi E., Juneau S., Shankar F., 2011, ApJ, 741, L33

Bouwens R. J., et al., 2010, ApJ, 709, L133

Bower R. G., 1991, MNRAS, 248, 332

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

Bruzual G., Charlot S., 2003, MNRAS, 344, 1000

Buchan S., Shankar F., 2016, MNRAS, 462, 2001

Buitrago F., Trujillo I., Conselice C. J., Bouwens R. J., Dickinson M., Yan H., 2008, ApJ, 687, L61

Buitrago F., Trujillo I., Conselice C. J., Häußler B., 2013, MNRAS, 428, 1460

Bullock J. S., 2010, preprint, (`arXiv:1009.4505`)

Cappellari M., et al., 2013, MNRAS, 432, 1862

Carollo C. M., et al., 2013, ApJ, 773, 112

Cattaneo A., Dekel A., Devriendt J., Guiderdoni B., Blaizot J., 2006, MNRAS, 370, 1651

Cattaneo A., Mamon G. A., Warnick K., Knebe A., 2011, A&A, 533, A5

Chabrier G., 2003, PASP, 115, 763

Chapman S. C., Blain A. W., Smail I., Ivison R. J., 2005, ApJ, 622, 772

Chen Y.-M., et al., 2012, MNRAS, 421, 314

Chiosi C., Merlin E., Piovan L., 2012, preprint, (`arXiv:1206.2532`)

Cid Fernandes R., Mateus A., Sodré L., Stasińska G., Gomes J. M., 2005, MNRAS, 358, 363

Cimatti A., et al., 2008, A&A, 482, 21

Cirasuolo M., Shankar F., Granato G. L., De Zotti G., Danese L., 2005, ApJ, 629, 816

Citro A., Pozzetti L., Moresco M., Cimatti A., 2016, preprint, (`arXiv:1604.07826`)

Clauwens B., Franx M., Schaye J., 2016, preprint, (`arXiv:1605.00009`)

Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168

Cole S., Helly J., Frenk C. S., Parkinson H., 2008, MNRAS, 383, 546

Colín P., Avila-Reese V., Valenzuela O., 2000, ApJ, 542, 622

Colless M., Saglia R. P., Burstein D., Davies R. L., McMahan R. K., Wegner G., 2001, MNRAS, 321, 277

Condon J. J., 1992, ARA&A, 30, 575

Conroy C., Gunn J. E., White M., 2009, ApJ, 699, 486

Conroy C., Graves G. J., van Dokkum P. G., 2014, ApJ, 780, 33

Conselice C. J., 2014, ARA&A, 52, 291

Contini E., Kang X., Romeo A. D., Xia Q., 2017, ApJ, 837, 27

Correa C. A., Wyithe J. S. B., Schaye J., Duffy A. R., 2015, MNRAS, 450, 1514

Cowie L. L., Hu E. M., Songaila A., Egami E., 1997, ApJ, 481, L9

Cox T. J., Dutta S. N., Di Matteo T., Hernquist L., Hopkins P. F., Robertson B., Springel V., 2006, ApJ, 650, 791

Crain R. A., Eke V. R., Frenk C. S., Jenkins A., McCarthy I. G., Navarro J. F., Pearce F. R., 2007, MNRAS, 377, 41

D'Souza R., Vegetti S., Kauffmann G., 2015, MNRAS, 454, 4027

Daddi E., Cimatti A., Renzini A., Fontana A., Mignoli M., Pozzetti L., Tozzi P., Zamorani G., 2004, ApJ, 617, 746

Daddi E., et al., 2005, ApJ, 626, 680

Damjanov I., et al., 2009, ApJ, 695, 101

De Lucia G., Boylan-Kolchin M., Benson A. J., Fontanot F., Monaco P., 2010, MNRAS, 406, 1533

De Lucia G., Fontanot F., Wilman D., Monaco P., 2011, MNRAS, 414, 1439

DeMaio T., Gonzalez A. H., Zabludoff A., Zaritsky D., Bradač M., 2015, MNRAS, 448, 1162

Dekel A., Birnboim Y., 2006, MNRAS, 368, 2

Dekel A., Mandelker N., 2014, MNRAS, 444, 2071

Dekel A., et al., 2009, nat, 457, 451

Delaye L., et al., 2014, MNRAS, 441, 203

Diemer B., Kravtsov A. V., 2015, ApJ, 799, 108

Drory N., Salvato M., Gabasch A., Bender R., Hopp U., Feulner G., Pannella M., 2005, ApJ, 619, L131

Dubois Y., Gavazzi R., Peirani S., Silk J., 2013, MNRAS, 433, 3297

Dubois Y., et al., 2014, MNRAS, 444, 1453

Dubois Y., Volonteri M., Silk J., Devriendt J., Slyz A., Teyssier R., 2015, MNRAS, 452, 1502

Duc P.-A., Cuillandre J.-C., Karabal 2015, MNRAS, 446, 120

Dutton A. A., Macciò A. V., 2014, MNRAS, 441, 3359

Dye S., et al., 2014, MNRAS, 440, 2013

Efstathiou G., Frenk C. S., White S. D. M., Davis M., 1988, MNRAS, 235, 715

Eisenstein D. J., Hut P., 1998, ApJ, 498, 137

Eisenstein D. J., et al., 2011, AJ, 142, 72

Elsner F., Feulner G., Hopp U., 2008, A&A, 477, 503

Fall S. M., Efstathiou G., 1980, MNRAS, 193, 189

Fan L., Lapi A., De Zotti G., Danese L., 2008, ApJ, 689, L101

Fan L., Lapi A., Bressan A., Bernardi M., De Zotti G., Danese L., 2010, ApJ, 718, 1460

Farouki R. T., Shapiro S. L., 1982, ApJ, 259, 103

Finoguenov A., et al., 2007, ApJS, 172, 182

Fontana A., et al., 2004, A&A, 424, 23

Fontana A., et al., 2006, A&A, 459, 745

Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306

Franx M., et al., 2003, ApJ, 587, L79

Fukugita M., Hogan C. J., Peebles P. J. E., 1998, ApJ, 503, 518

Gao L., Loeb A., Peebles P. J. E., White S. D. M., Jenkins A., 2004, ApJ, 614, 17

Gargiulo A., Saracco P., Tamburri S., Lonoce I., Ciocca F., 2016, A&A, 592, A132

Genel S., et al., 2012, ApJ, 745, 11

George M. R., et al., 2011, ApJ, 742, 125

Gerhard O. E., 1981, MNRAS, 197, 179

Giocoli C., Pieri L., Tormen G., 2008, MNRAS, 387, 689

Gonzalez A. H., Zabludoff A. I., Zaritsky D., Dalcanton J. J., 2000, ApJ, 536, 561

Gonzalez J. E., Lacey C. G., Baugh C. M., Frenk C. S., 2011, MNRAS, 413, 749

Gonzalez A. H., Sivanandam S., Zabludoff A. I., Zaritsky D., 2013, ApJ, 778, 14

Governato F., Babul A., Quinn T., Tozzi P., Baugh C. M., Katz N., Lake G., 1999, MNRAS, 307, 949

Graham A. W., Driver S. P., 2007, ApJ, 655, 77

Granato G. L., De Zotti G., Silva L., Bressan A., Danese L., 2004, ApJ, 600, 580

Granato G. L., Silva L., Lapi A., Shankar F., De Zotti G., Danese L., 2006, MNRAS, 368, L72

Grogin N. A., et al., 2011, ApJS, 197, 35

Gross M. A. K., Somerville R. S., Primack J. R., Holtzman J., Klypin A., 1998, MNRAS, 301, 81

Gu M., Conroy C., Behroozi P., 2016, ApJ, 833, 2

Guo Q., et al., 2011, MNRAS, 413, 101

Guo H., et al., 2014, MNRAS, 441, 2398

Hahn O., Abel T., 2011, MNRAS, 415, 2101

Hammer F., Flores H., Puech M., Yang Y. B., Athanassoula E., Rodrigues M., Delgado R., 2009, A&A, 507, 1313

Heitmann K., et al., 2014, Computing in Science Engineering, 16, 14

Hernquist L., 1990, ApJ, 356, 359

Hinton S., 2016, JOSS, 1

Hirschmann M., Naab T., Somerville R. S., Burkert A., Oser L., 2012, MNRAS, 419, 3200

Hirschmann M., et al., 2013, MNRAS, 436, 2929

Hoekstra H., Hsieh B. C., Yee H. K. C., Lin H., Gladders M. D., 2005, ApJ, 635, 73

Hogg D. W., et al., 2002, AJ, 124, 646

Hopkins P. F., Hernquist L., Cox T. J., Kereš D., 2008, ApJS, 175, 356

Hopkins P. F., Hernquist L., Cox T. J., Keres D., Wuyts S., 2009, ApJ, 691, 1424

Hopkins P. F., Bundy K., Hernquist L., Wuyts S., Cox T. J., 2010, MNRAS, 401, 1099

Hopkins P. F., Kereš D., Oñorbe J., Faucher-Giguère C.-A., Quataert E., Murray N., Bullock J. S., 2014, MNRAS, 445, 581

Hu W., Kravtsov A. V., 2003, ApJ, 584, 702

Hubble E., 1929, Proceedings of the National Academy of Science, 15, 168

Hubble E. P., 1936, Realm of the Nebulae

Hubble E., Humason M. L., 1931, ApJ, 74, 43

Huertas-Company M., et al., 2013, MNRAS, 428, 1715

Huertas-Company M., et al., 2015, ApJ, 809, 95

Ilbert O., et al., 2013, A&A, 556, A55

Israel F. P., 1997, A&A, 328, 471

Jauzac M., et al., 2015, MNRAS, 452, 1437

Jedamzik K., 1995, ApJ, 448, 1

Jiang F., van den Bosch F. C., 2014, MNRAS, 440, 193

Jorgensen I., Franx M., Kjaergaard P., 1996, MNRAS, 280, 167

Kapteyn J. C., 1922, ApJ, 55, 302

Kartaltepe J. S., et al., 2015, ApJS, 221, 11

Katz N., 1992, ApJ, 391, 502

Katz N., Gunn J. E., 1991, ApJ, 377, 365

Kauffmann G., White S. D. M., Heckman T. M., Ménard B., Brinchmann J., Charlot S., Tremonti C., Brinkmann J., 2004, MNRAS, 353, 713

Keating S. K., Abraham R. G., Schiavon R., Graves G., Damjanov I., Yan R., Newman J., Simard L., 2015, ApJ, 798, 26

Kennicutt Jr. R. C., 1983, ApJ, 272, 54

Kennicutt Jr. R. C., 1998, ApJ, 498, 541

Khochfar S., Burkert A., 2005, MNRAS, 359, 1379

Klimentowski J., Łokas E. L., Kazantzidis S., Mayer L., Mamon G. A., 2009, MNRAS, 397, 2015

Klypin A., Holtzman J., 1997, ArXiv Astrophysics e-prints,

Klypin A., Gottlöber S., Kravtsov A. V., Khokhlov A. M., 1999, ApJ, 516, 530

Klypin A. A., Trujillo-Gomez S., Primack J., 2011, ApJ, 740, 102

Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, MNRAS, 457, 4340

Knollmann S. R., Knebe A., 2009, ApJS, 182, 608

Koekemoer A. M., et al., 2011, ApJS, 197, 36

Kosowsky A., Turner M. S., 1995, Phys. Rev. D, 52, R1739

Kowalski M., et al., 2008, ApJ, 686, 749

Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, ApJ, 609, 35

Kravtsov A., Vikhlinin A., Meshscheryakov A., 2014, preprint, (`arXiv:1401.7329`)

Kroupa P., 2001, MNRAS, 322, 231

Lacey C., Cole S., 1993, MNRAS, 262, 627

Lacey C., Cole S., 1994, MNRAS, 271, 676

Lacey C. G., et al., 2015, preprint, (`arXiv:1509.08473`)

Landy S. D., Szalay A. S., 1993, ApJ, 412, 64

Lapi A., Gonzalez Nuevo J., Fan L., 2011, ApJ, 742, 24

Lapi A., Salucci P., Danese L., 2013, ApJ, 772, 85

Laporte C. F. P., White S. D. M., 2015, MNRAS, 451, 1177

Larson R. B., 1975, MNRAS, 173, 671

Le Fèvre O., et al., 2005, A&A, 439, 845

Leauthaud A., et al., 2012, ApJ, 744, 159

Leauthaud A., et al., 2016, MNRAS, 457, 4021

Leavitt H. S., 1908, Annals of Harvard College Observatory, 60, 87

Leavitt H. S., Pickering E. C., 1912, Harvard College Observatory Circular, 173, 1

Lee J., Yi S. K., 2013, ApJ, 766, 38

Lemaître G., 1933, Annales de la Société Scientifique de Bruxelles, 53

Lidman C., et al., 2012, MNRAS, 427, 550

Lidman C., et al., 2013, MNRAS, 433, 825

Lilly S. J., Carollo C. M., Pipino A., Renzini A., Peng Y., 2013, ApJ, 772, 119

Liu F. S., Lei F. J., Meng X. M., Jiang D. F., 2015, MNRAS, 447, 1491

Lotz J. M., Madau P., Giavalisco M., Primack J., Ferguson H. C., 2006, ApJ, 636, 592

Lotz J. M., Jonsson P., Cox T. J., Primack J. R., 2008, MNRAS, 391, 1137

Mandelbaum R., Seljak U., Kauffmann G., Hirata C. M., Brinkmann J., 2006, MNRAS, 368, 715

Maraston C., 2005, MNRAS, 362, 799

Maraston C., et al., 2013, MNRAS, 435, 2764

Marchesini D., van Dokkum P. G., Förster Schreiber N. M., Franx M., Labbé I., Wuyts S., 2009, ApJ, 701, 1765

Marchesini D., et al., 2014, ApJ, 794, 65

Martizzi D., Jimmy Teyssier R., Moore B., 2014, MNRAS, 443, 1500

Marulli F., Bonoli S., Branchini E., Moscardini L., Springel V., 2008, MNRAS, 385, 1846

McBride J., Fakhouri O., Ma C.-P., 2009, MNRAS, 398, 1858

McCavana T., Micic M., Lewis G. F., Sinha M., Sharma S., Holley-Bockelmann K., Bland-Hawthorn J., 2012, MNRAS, 424, 361

McCracken H. J., et al., 2012, A&A, 544, A156

McDermid R. M., Alatalo K., Blitz L., 2015, MNRAS, 448, 3484

McDonald M., et al., 2016, ApJ, 817, 86

Menci N., Fontana A., Giallongo E., Grazian A., Salimbeni S., 2006, ApJ, 647, 753

Mendel J. T., et al., 2015, ApJ, 804, L4

Merlin E., Chiosi C., Piovan L., Grassi T., Buonomo U., Barbera F. L., 2012, MNRAS, 427, 1530

Mihos J. C., Hernquist L., 1994, ApJ, 431, L9

Mihos J. C., Hernquist L., 1996, ApJ, 464, 641

Mo H. J., Mao S., White S. D. M., 1998, MNRAS, 295, 319

Mo H., van den Bosch F. C., White S., 2010, Galaxy Formation and Evolution

Montes M., Trujillo I., Prieto M. A., Acosta-Pulido J. A., 2014, MNRAS, 439, 990

Mortlock A., et al., 2013, MNRAS, 433, 1185

Moster B. P., Somerville R. S., Maulbetsch C., van den Bosch F. C., Macciò A. V., Naab T., Oser L., 2010, ApJ, 710, 903

Moster B. P., Naab T., White S. D. M., 2013, MNRAS, 428, 3121

Moster B. P., Naab T., White S. D. M., 2017, preprint, (`arXiv:1705.05373`)

Mundy C. J., Conselice C. J., Duncan K. J., Almaini O., Häußler B., Hartley W. G., 2017, preprint, (`arXiv:1705.07986`)

Muzzin A., et al., 2009, ApJ, 698, 1934

Muzzin A., et al., 2013, ApJ, 777, 18

Naab T., Burkert A., 2003, ApJ, 597, 893

Naab T., Khochfar S., Burkert A., 2006, ApJ, 636, L81

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

Nagashima M., 2001, ApJ, 562, 7

Nair P., van den Bergh S., Abraham R. G., 2011, ApJ, 734, L31

Negroponte J., White S. D. M., 1983, MNRAS, 205, 1009

Newman A. B., Ellis R. S., Bundy K., Treu T., 2012, ApJ, 746, 162

Newman A. B., Ellis R. S., Andreon S., Treu T., Raichoor A., Trinchieri G., 2014, ApJ, 788, 51

Nipoti C., Treu T., Auger M. W., Bolton A. S., 2009, ApJ, 706, L86

Ocvirk P., Pichon C., Lançon A., Thiébaut E., 2006, MNRAS, 365, 46

Oliva-Altamirano P., Brough S., Jimmy Kim-Vy T., Couch W. J., McDermid R. M., Lidman C., von der Linden A., Sharp R., 2015, MNRAS, 449, 3347

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

Ownsworth J. R., Conselice C. J., Mortlock A., Hartley W. G., Almaini O., Duncan K., Mundy C. J., 2014, MNRAS, 445, 2198

Pahre M. A., Djorgovski S. G., de Carvalho R. R., 1998, AJ, 116, 1591

Parkinson H., Cole S., Helly J., 2008, MNRAS, 383, 557

Parkinson D., et al., 2012, Phys. Rev. D, 86, 103518

Partridge R. B., Peebles P. J. E., 1967, ApJ, 147, 868

Peacock J. A., Heavens A. F., 1990, MNRAS, 243, 133

Peeples M. S., Somerville R. S., 2013, MNRAS, 428, 1766

Peng Y.-j., Lilly S. J., Kovač K., 2010, ApJ, 721, 193

Peralta de Arriba L., et al., 2015, MNRAS, 453, 704

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

Pettini M., Shapley A. E., Steidel C. C., Cuby J.-G., Dickinson M., Moorwood A. F. M., Adelberger K. L., Giavalisco M., 2001, ApJ, 554, 981

Pipino A., Devriendt J. E. G., Thomas D., Silk J., Kaviraj S., 2009, A&A, 505, 1075

Planck Collaboration et al., 2014, A&A, 571, A16

Poggianti B. M., et al., 2006, ApJ, 642, 188

Pontzen A., Roškar R., Stinson G. S., Woods R., Reed D. M., Coles J., Quinn T. R., 2013, pynbody: Astrophysics Simulation Analysis for Python

Pontzen A., Tremmel M., Roth N., Peiris H. V., Saintonge A., Volonteri M., Quinn T., Governato F., 2017, MNRAS, 465, 547

Posti L., Nipoti C., Stiavelli M., Ciotti L., 2014, MNRAS, 440, 610

Potter D., Stadel J., Teyssier R., 2016, preprint, (`arXiv:1609.08621`)

Power C., Navarro J. F., Jenkins A., Frenk C. S., White S. D. M., Springel V., Stadel J., Quinn T., 2003, MNRAS, 338, 14

Pozzetti L., et al., 2007, A&A, 474, 443

Press W. H., Schechter P., 1974, ApJ, 187, 425

Puchwein E., Springel V., Sijacki D., Dolag K., 2010, MNRAS, 406, 936

Ragone-Figueroa C., Granato G. L., 2011, MNRAS, 414, 3690

Raichoor A., et al., 2011, ApJ, 732, 12

Ravindranath S., et al., 2006, ApJ, 652, 963

Riebe K., et al., 2011, preprint, (`arXiv:1109.0003`)

Riess A. G., et al., 1998, AJ, 116, 1009

Robinson I., Webber J., Eifrem E., 2015, Graph Databases, 2 edn. O'Reilly Media

Rodriguez-Gomez V., et al., 2016, MNRAS, 458, 2371

Romanowsky A. J., Strader J., Spitler L. R., Johnson R., Brodie J. P., Forbes D. A., Ponman T., 2009, AJ, 137, 4956

Rota S., et al., 2017, A&A, 601, A144

SDSS Collaboration et al., 2016, preprint, (arXiv:1608.02013)

Saglia R. P., et al., 2010, A&A, 524, A6

Saito S., et al., 2016, MNRAS, 460, 1457

Salpeter E. E., 1955, ApJ, 121, 161

Sanders D. B., Soifer B. T., Elias J. H., Madore B. F., Matthews K., Neugebauer G., Scoville N. Z., 1988, ApJ, 325, 74

Schawinski K., et al., 2014, MNRAS, 440, 889

Schaye J., et al., 2015, MNRAS, 446, 521

Schechter P., 1976, ApJ, 203, 297

Scheick X., Kuhn J. R., 1994, ApJ, 423, 566

Schombert J. M., 1986, ApJS, 60, 603

Scoville N., et al., 2007, ApJS, 172, 1

Sérsic J. L., 1963, Boletin de la Asociacion Argentina de Astronomia La Plata Argentina, 6, 41

Shankar F., Bernardi M., 2009, MNRAS, 396, L76

Shankar F., Lapi A., Salucci P., De Zotti G., Danese L., 2006, ApJ, 643, 14

Shankar F., Marulli F., Bernardi M., Dai X., Hyde J. B., Sheth R. K., 2010a, MNRAS, 403, 117

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

Shankar F., Marulli F., Bernardi M., Mei S., Meert A., Vikram V., 2013, MNRAS, 428, 109

Shankar F., et al., 2014a, MNRAS, 439, 3189

Shankar F., Guo H., Bouillot 2014b, ApJ, 797, L27

Shankar F., et al., 2015, ApJ, 802, 73

Shankar F., et al., 2016, MNRAS,

Shankar F., et al., 2017, ApJ, 840, 34

Shapley H., Curtis H. D., 1921, Bulletin of the National Research Council, Vol. 2, Part 3, No. 11, p. 171-217, 2, 171

Shapley A. E., Steidel C. C., Pettini M., Adelberger K. L., 2003, ApJ, 588, 65

Shen S., Mo H. J., White S. D. M., Blanton M. R., Kauffmann G., Voges W., Brinkmann J., Csabai I., 2003, MNRAS, 343, 978

Slipher V. M., 1917, Proceedings of the American Philosophical Society, 56, 403

Smail I., Ivison R. J., Blain A. W., 1997, ApJ, 490, L5

Smail I., Chapman S. C., Ivison R. J., Blain A. W., Takata T., Heckman T. M., Dunlop J. S., Sekiguchi K., 2003, MNRAS, 342, 1185

Sofue Y., Rubin V., 2001, ARA&A, 39, 137

Somerville R. S., Kolatt T. S., 1999, MNRAS, 305, 1

Somerville R. S., Lemson G., Kolatt T. S., Dekel A., 2000, MNRAS, 316, 479

Somerville R. S., Primack J. R., Faber S. M., 2001, MNRAS, 320, 504

Somerville R. S., Hopkins P. F., Cox T. J., Robertson B. E., Hernquist L., 2008, MNRAS, 391, 481

Sonnenfeld A., Gavazzi R., Suyu S. H., Treu T., Marshall P. J., 2013, ApJ, 777, 97

Springel V., 2005a, MNRAS, 364, 1105

Springel V., 2005b, MNRAS, 364, 1105

Springel V., 2010, MNRAS, 401, 791

Springel V., Yoshida N., White S. D. M., 2001, New A, 6, 79

Steidel C. C., Giavalisco M., Dickinson M., Adelberger K. L., 1996, AJ, 112, 352

Steidel C. C., Adelberger K. L., Shapley A. E., Pettini M., Dickinson M., Giavalisco M., 2003, ApJ, 592, 728

Steidel C. C., Erb D. K., Shapley A. E., Pettini M., Reddy N., Bogosavljević M., Rudie G. C., Rakic O., 2010, ApJ, 717, 289

Stewart K. R., Bullock J. S., Wechsler R. H., Maller A. H., 2009, ApJ, 702, 307

Strateva I., et al., 2001, AJ, 122, 1861

Strazzullo V., et al., 2010, A&A, 524, A17

Strazzullo V., et al., 2013, ApJ, 772, 118

Stringer M., Trujillo I., Dalla Vecchia C., Martinez-Valpuesta I., 2015, MNRAS, 449, 2396

Sullivan M., et al., 2011, ApJ, 737, 102

Swinbank A. M., et al., 2005, MNRAS, 359, 401

Taranu D. S., Dubinski J., Yee H. K. C., 2013, ApJ, 778, 61

Teyssier R., 2002, A&A, 385, 337

Thanjavur K., Simard L., Bluck A. F. L., Mendel T., 2016, MNRAS, 459, 44

Thomas D., Maraston C., Bender R., Mendes de Oliveira C., 2005, ApJ, 621, 673

Tinker J. L., Weinberg D. H., Zheng Z., Zehavi I., 2005, ApJ, 631, 41

Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, ApJ, 688, 709

Tinker J. L., et al., 2016, preprint, (`arXiv:1607.04678`)

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

Tomczak A. R., et al., 2016, ApJ, 817, 118

Toomre A., 1964, ApJ, 139, 1217

Toomre A., 1977, in Tinsley B. M., Larson D. Campbell R. B. G., eds, Evolution of Galaxies and Stellar Populations. p. 401

Torrey P., et al., 2015, MNRAS, 454, 2770

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

Tyson J. A., Valdes F., Jarvis J. F., Mills Jr. A. P., 1984, ApJ, 281, L59

Uson J. M., Boughn S. P., Kuhn J. R., 1991, ApJ, 369, 46

Vale A., Ostriker J. P., 2004, MNRAS, 353, 189

Vikhlinin A., Kravtsov A., Forman W., Jones C., Markevitch M., Murray S. S., Van Speybroeck L., 2006, ApJ, 640, 691

Vikhlinin A., et al., 2009, ApJ, 692, 1033

Vincoletto L., Matteucci F., Calura F., Silva L., Granato G., 2012, MNRAS, 421, 3116

Vulcani B., et al., 2016, ApJ, 816, 86

Wake D. A., et al., 2008, MNRAS, 387, 1045

Wake D. A., et al., 2011, ApJ, 728, 46

Weinmann S. M., van den Bosch F. C., Yang X., Mo H. J., 2006, MNRAS, 366, 2

Welker C., Dubois Y., Devriendt J., Pichon C., Kaviraj S., Peirani S., 2015, preprint, (`arXiv:1502.05053`)

Wellons S., et al., 2016, MNRAS, 456, 1030

White S. D. M., Frenk C. S., 1991, ApJ, 379, 52

White M., Martini P., Cohn J. D., 2008, MNRAS, 390, 1179

Wilman D. J., Fontanot F., De Lucia G., Erwin P., Monaco P., 2013, MNRAS, 433, 2986

Yang X., Mo H. J., van den Bosch F. C., 2003, MNRAS, 339, 1057

Yang X., Mo H. J., Zhang Y., van den Bosch F. C., 2011, ApJ, 741, 13

Yano T., Nagashima M., Gouda N., 1996, ApJ, 466, 1

Yurin D., Springel V., 2014, MNRAS, 444, 62

Zheng Z., et al., 2005, ApJ, 633, 791

Zheng Z., Coil A. L., Zehavi I., 2007, ApJ, 667, 760

Zolotov A., et al., 2015, MNRAS, 450, 2327

Zwicky F., 1933, Helvetica Physica Acta, 6, 110

van Dokkum P. G., Franx M., 1996, MNRAS, 281, 985

van Dokkum P. G., et al., 2003, ApJ, 587, L83

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

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

van Uitert E., et al., 2016, MNRAS, 459, 3251

van den Bosch F. C., et al., 2007, MNRAS, 376, 841

van den Bosch F. C., Jiang F., Hearin A., Campbell D., Watson D., Padmanabhan N., 2014, MNRAS, 445, 1713

van der Burg R. F. J., et al., 2013, A&A, 557, A15

van der Wel A., Franx M., van Dokkum 2014, ApJ, 788, 28