Supercomputer Models of the Formation and Evolution of Galaxies

Simon Mutch

Presented in fulfillment of the requirements of the degree of Doctor of Philosophy

January 2013

Faculty of Information and Communication Technology
Swinburne University
When we remember we are all mad,
the mysteries disappear and life stands explained.

MARK TWAIN
Abstract

In this thesis, we explore the use of semi-analytic galaxy formation models and related techniques as a means to investigate the physics of galaxy formation and evolution. We begin by investigating the ability of a highly cited semi-analytic model to reproduce the evolution of the observed galaxy population over the last 7 billion years of cosmic time. This is achieved by carrying out a detailed statistical calibration of the model’s free parameters in order to simultaneously reproduce the observed galactic stellar mass function at $z=0$ and $z\approx0.8$, as well as the $z=0$ black–hole bulge relation. In order to be successful, we are required to push the parameters of the model associated with supernova feedback to implausibly high values, suggesting that the current implementation of this physical prescription may be inadequate. Additionally, we suggest that some extra mechanism is required to preferentially increase the efficiency of star formation in the most massive galaxies at high redshift.

In order to further explore the utility of semi-analytic models, we then present their novel use as a tool to investigate the current evolutionary status of the Milky Way and M31. The Milky Way is the most closely studied galaxy in the Universe. However, to understand the Milky Way’s place in the broader landscape of galaxy evolution, we require a baseline population of galaxies against which to compare. Using a sample of analogue galaxies drawn from both observational data and semi-analytic models we find that both the Milky Way and M31 may be “green valley” galaxies undergoing an important evolutionary transition. Furthermore, using the histories of our model analogue sample, we investigate the possible physical mechanisms which could be driving this evolutionary change.

Finally, we introduce a new, self-consistent model for connecting the growth of galaxies to the formation history of their host dark matter halos. This model dispenses with attempts to implement all of the dominant baryonic processes associated with galaxy evolution and replaces them with two simple, phenomenologically motivated equations that depend only on a single host halo property. We demonstrate the ability of our new formation history model to reproduce the observed red and blue stellar mass functions at $z=0$. Then, by adding a simple redshift dependence to the parameterisations, we show that it can also successfully match the observed global mass functions out to $z=4$. To conclude, we highlight the advantages of this model over the relevant alternatives, as well as highlight its general utility.
I have a lot of people to thank for helping me get to this point...

Of course, this begins with my supervisors Darren and Greg. You have both provided me with so much amazing guidance and support, for which I am so very grateful.

To Darren - Thank you for always helping me to see the big picture and not get too bogged down in the details. You have been a fantastic mentor and have taught me so much, not just about science but about how to think about research and understand the story. Also, thank you for taking me on as your first student; I hope I didn’t traumatise you too much!

To Greg - Where to begin? I can’t thank you enough for your patience and support throughout my PhD. You have willingly devoted so much of your time to helping me and have put up with all of my worrying and complaining over the last three years. Best of all, you have not only been my guide, but have also become a great and very close friend.

I also want to thank here Peder Norberg, without whose influence during my undergraduate degree, I would not have ended up even taking on a PhD in the first place.

To my friends Rob and Glenn - thanks for the Hawks games, the nights out, the squash and the banter. To Emily and Evelyn, I owe you both for making me feel welcome when I first came to Swinburne and for always being good for a chat round the odd Tim-Tam. To Alison and Amy, thank you for looking after Rach while I have been working and for being great, and very entertaining friends! To my friends back home as well - your random emails and occasional visits always bring a smile to my face. I hope more of you can make it out here soon!

To my Mum and Dad - Thank you for being such a constant and supportive presence in my life. None of this would have been possible without you. Also thank you to my brother and sister who have been giving me plenty of smiles with photos of my niece and nephew over the last year!

Last and by far most importantly: Thank you to my wife - without your constant love and support I would never have made it to where I am now. I can’t thank you enough for putting up with me these last couple of months. I love you so much. This thesis is dedicated to you... (not the most romantic of dedications I’ll grant you)...

Acknowledgments
Declaration

This thesis contains no material that has been accepted for the award of any other degree or diploma. To the best of my knowledge, this thesis contains no material previously published or written by another author, except where due reference is made in the text of the thesis. All work presented is primarily that of the author.

Chapter 2 has been published as:

**Mutch, Simon J.**, Poole, Gregory B., Croton, Darren J.
Constraining the last 7 billion years of galaxy evolution in semi-analytic models

Chapter 3 has been published as:

**Mutch, Simon J.**, Croton, Darren J., Poole, Gregory B.
The Mid-life Crisis of the Milky Way and M31
The Astrophysical Journal, Volume 736, Issue 2, article id. 84 (2011)

Minor alterations have been made to these works in order to maintain consistency of style. Chapter 2 also contains additional material not present in the associated paper.

Simon James Mutch
Abstract iii
List of Figures xiii
List of Tables xv

1 Introduction 1
  1.1 Structure formation ........................................ 1
    1.1.1 From quantum fluctuations to dark matter halos 2
    1.1.2 The cosmic web ........................................ 5
  1.2 N-Body dark matter simulations .......................... 6
  1.3 Creating a mock universe: Semi-analytic galaxy formation models .. 7
    1.3.1 Cooling ........................................ 9
    1.3.2 Star formation .................................... 11
    1.3.3 Supernova feedback ................................ 12
    1.3.4 AGN feedback .................................... 13
    1.3.5 The central–satellite dichotomy .................... 15
    1.3.6 Galaxy mergers .................................... 16
    1.3.7 Intra-cluster light ................................ 17
    1.3.8 Disk instabilities ................................ 18
    1.3.9 Galaxy luminosity and dust extinction ............ 19
  1.4 Alternative modelling techniques ....................... 20
    1.4.1 HODs .......................................... 20
    1.4.2 SHAM models .................................... 21
    1.4.3 Hydrodynamic simulations ........................ 21
  1.5 Thesis structure ....................................... 22

2 Constraining the evolution of semi-analytic models 25
  2.1 Introduction ........................................ 26
  2.2 Method ........................................... 29
    2.2.1 MCMC Implementation ............................ 29
    2.2.2 Principle Component Analysis ....................... 31
  2.3 The semi-analytical galaxy formation model ............ 32
    2.3.1 Star formation and supernova feedback ............ 34
## 3 The mid-life crisis of the Milky Way and M31

### 3.1 Introduction

### 3.2 Observational data

- 3.2.1 The Milky Way
- 3.2.2 M31
- 3.2.3 Galaxy Zoo control sample

### 3.3 Modeled data

- 3.3.1 A ‘Semi-analytic’ Model of Galaxy Formation
- 3.3.2 Selecting the Theoretical Milky Way Analogue Sample

### 3.4 Results

- 3.4.1 The position of the Milky Way and M31 on the colour–stellar mass diagram
- 3.4.2 How common are green valley Milky Way/M31-type galaxies?
- 3.4.3 Further constraints using the model analogue sample

### 3.5 Discussion: What makes a ‘green’ Milky Way?

- 3.5.1 Are Red Milky Way Analogue Simply Older than Blue?
- 3.5.2 Rapid or gradual shut-down of Milky Way star formation?
- 3.5.3 Cold gas depletion in spiral galaxies?
- 3.5.4 Cold gas heating in spiral galaxies?
- 3.5.5 Spiral galaxies on the edge

### 3.6 Conclusions


Contents

4 A formation history model of galaxy stellar mass growth 89
  4.1 Introduction ...................................... 90
  4.2 The simplest model of galaxy formation .............. 93
    4.2.1 The growth of structure .......................... 93
    4.2.2 The baryonic content of dark matter halos ........ 93
    4.2.3 Generating the galaxy population .................. 99
  4.3 Results ........................................... 102
    4.3.1 Redshift zero .................................... 102
    4.3.2 High redshift .................................... 106
    4.3.3 Incorporating a redshift evolution ................ 109
  4.4 Discussion ......................................... 111
  4.5 Conclusions ........................................ 114

5 Summary and forward look 117
  5.1 Summary .......................................... 117
  5.2 Future directions .................................. 119

Bibliography 123

Publications 135
List of Figures

1.1 Dark matter density profiles. .................................................. 4
1.2 The cosmic web of galaxies. .................................................... 5
1.3 A dark matter merger tree schematic. ....................................... 7
1.4 Populating the output of an N-body simulation with semi-analytic galaxies. 8
1.5 Normalised cooling function curves. ......................................... 10
1.6 The effect of AGN feedback on the galactic luminosity function. ....... 14

2.1 The observed $z=0$ (Baldry et al., 2008) and $z=0.83$ stellar mass functions. 39
2.2 1 & 2-D posterior probability distributions when constraining the model against the $z=0$ stellar mass function alone. .................... 43
2.3 The best fit $z=0$ stellar mass function when constraining against the observed $z=0$ stellar mass function alone. .......................... 45
2.4 The $f_{\text{BH}}-\alpha_{\text{SF}}$ posterior distribution when constraining the model against the $z=0$ black hole–bulge relation alone. ............. 46
2.5 Posterior probability distributions when constraining against the $z=0$ stellar mass function and black hole–bulge relation simultaneously. .... 47
2.6 The best fit $z=0$ model stellar mass function and black hole–bulge relation when constraining against both the $z=0$ stellar mass function and black hole–bulge relation simultaneously. ............. 48
2.7 The best fit $z=0.83$ stellar mass function when constraining against the observed $z=0.83$ stellar mass function alone. .................... 50
2.8 Posterior probability distributions for $\alpha_{\text{SF}}$ when constraining against the $z=0.83$ stellar mass function alone. ........................ 51
2.9 Posterior probability distributions when applying all constraints simultaneously. .......................................................... 53
2.10 The best fit $z=0$ and $z=0.83$ model stellar mass functions when applying all constraints simultaneously. ............................ 55

3.1 Colour vs. stellar mass diagram for the full Galaxy Zoo sample. ........... 71
3.2 Colour vs. stellar mass distribution of the observed Milky Way/M31 analogue sample. .......................................................... 73
3.3 Colour vs. stellar mass diagram for the model Milky Way/M31 analogue sample. .......................................................... 75
3.4 Model Milky Way/M31 star formation rate vs. stellar mass ............... 77
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.5</td>
<td>Normalized probability distributions of formation times for each colour sub-sample of the model Milky Way/M31 analogues.</td>
<td>78</td>
</tr>
<tr>
<td>3.6</td>
<td>The mean cold gas mass of each colour sub-sample of our model Milky Way/M31 analogues as a function of look-back time.</td>
<td>80</td>
</tr>
<tr>
<td>3.7</td>
<td>Normalized probability distributions of black hole masses for red, green and blue galaxies drawn from our model Milky Way/M31 analogue sample.</td>
<td>82</td>
</tr>
<tr>
<td>3.8</td>
<td>Illustration of the alternative evolution suggested by a green, passively evolving population of spiral galaxies compared against the standard paradigm.</td>
<td>84</td>
</tr>
<tr>
<td>4.1</td>
<td>The diversity of dark matter halo growth histories.</td>
<td>95</td>
</tr>
<tr>
<td>4.2</td>
<td>A comparison of dark matter halo and galaxy stellar mass functions.</td>
<td>97</td>
</tr>
<tr>
<td>4.3</td>
<td>$M_{\text{vir}}$ and $V_{\text{max}}$ halo growth histories.</td>
<td>98</td>
</tr>
<tr>
<td>4.4</td>
<td>The colour–stellar mass diagram of the $M_{\text{vir}}$ model.</td>
<td>103</td>
</tr>
<tr>
<td>4.5</td>
<td>The red and blue stellar mass functions produced by the non-evolving physics functions.</td>
<td>104</td>
</tr>
<tr>
<td>4.6</td>
<td>The high redshift stellar mass functions produced by the non-evolving physics functions.</td>
<td>107</td>
</tr>
<tr>
<td>4.7</td>
<td>The mean evolution of the stellar–halo mass relation for the non-evolving model.</td>
<td>108</td>
</tr>
<tr>
<td>4.8</td>
<td>The mean evolution of the stellar–halo mass relation for the evolving model.</td>
<td>109</td>
</tr>
<tr>
<td>4.9</td>
<td>The high redshift stellar mass functions produced by the evolving physics functions.</td>
<td>110</td>
</tr>
<tr>
<td>5.1</td>
<td>Preliminary results from of a new semi-analytic disk model.</td>
<td>120</td>
</tr>
</tbody>
</table>
List of Tables

2.1 The six free parameters of the semi-analytic model along with their original and most probable values. ............................. 33

3.1 Literature values for the relevant global properties of the Milky Way and M31. ........................................ 66

4.1 Fiducial physics function parameter values. ..................102
Theoretical models and simulations provide an important framework within which to test our current theories of galaxy formation and evolution. For example, what are the important processes that drive the statistical distributions of observed galaxy properties? How did particular classes of galaxies evolve to their current forms? How rare are such galaxies? In this thesis we explore some of these questions, particularly within the context of so called “semi-analytic” models.

As an introduction, this chapter presents a very brief overview of the emergence of structure in the Universe and the use of N-body techniques to simulate this growth. We then move on to introducing the dominant physical processes of galaxy formation and evolution within the context of semi-analytic galaxy formation models and their construction. We conclude by providing a brief overview of the most common alternative modelling techniques. Since each of the following chapters are self-contained, with their own targeted introductory material, the goal here is not to present a complete pedagogical review of any particular topic, but instead to provide a backdrop for the scientific material contained in this thesis.

1.1 Structure formation

Since the latter half of the 1990’s, Λ–Cold Dark Matter (ΛCDM) has emerged as the favoured cosmological model upon which we frame our understanding of the Universe. Although there exists a number of alternatives, no other model has proved to be as successful at simultaneously explaining such a broad range of observational signatures, such as the cosmic microwave background temperature anisotropy distribution, the matter density power spectrum and the accelerating expansion of the Universe. As the name suggests,
Chapter 1. Introduction

A key attribute of this model is the presence of a significant “dark matter” component in the Universe. This material is both non-baryonic and collisionless in nature and therefore its only interaction with the luminous matter is through gravitational interaction.

There is a great deal of observational evidence to support the presence of dark matter. Historically, its first detection can be traced back to the early work of Zwicky (1933), who attempted to estimate the dynamical mass of the Coma galaxy cluster using the virial theorem and found that there was far more mass present than could be accounted for by the luminous matter. Other supporting evidence for the existence of this unseen mass component comes from the flat rotation curves of disk galaxies (e.g. Rubin & Ford, 1970), the X-ray profiles of massive galaxy clusters (e.g. Allen et al., 2001) and gravitational lensing studies (e.g. Clowe et al., 2004). Modern observations are now able to constrain the parameters of the ΛCDM model to a high degree of precision, such that we can say with some certainty that only 5% of the mass-energy of the Universe is comprised of normal baryonic material, with 23% being dark matter (Komatsu et al., 2011). The remaining 73% represents the so called “vacuum” energy (the Λ of ΛCDM) which is thought to be driving the observed accelerating expansion of the Universe (Riess et al., 1998; Perlmutter et al., 1999).

Within the framework of this concordance cosmology, the remainder of this section provides a high-level overview of the formation and evolution of structure in the Universe, starting from the initial quantum density fluctuations and ending with the cosmic web of filamentary structures observed by large scale galaxy surveys. This material forms the conceptual foundation upon which the content of this thesis is built.

1.1.1 From quantum fluctuations to dark matter halos

The formation of structure can be traced all the way back to the very earliest stages of the Universe, when it was still less than approximately $10^{-40}$s old. At this time, quantum fluctuations seeded minute density inhomogeneities. These subsequently became amplified and stretched by a factor of around $10^{26}$ during an early stage called “inflation” (Guth, 1981): an extremely rapid phase of expansion lasting no longer than approximately $10^{-32}$s. The end of the inflationary period was marked by a phase transition, which allowed the formation of radiation and matter (in both baryonic and non-baryonic forms).

Within the framework of ΛCDM, the Universe was heavily diluted and almost completely homogeneous immediately after inflation, with only very minor density fluctuations that were random on all scales. In order to describe the magnitude of these fluctuations
1.1. Structure formation

at any given position in space, \( \rho(x) \), it is common to define the dimensionless value:

\[
\delta(x) = \frac{\rho(x)}{\rho_0} - 1,
\]

where \( \rho_0 \) is the average density. In the early Universe \( |\delta(x)| \ll 1 \) and the evolution of the initial overdensities can be tractably modelled using linear perturbation theory. Additionally, at these early times the ordinary matter was in the form of a plasma of free electrons, protons and alpha particles. After an initial phase where radiation dominated the dynamics, the Universe became matter dominated and the efficient scattering of photons by this plasma provided it with a source of pressure support against collapse. However, this was not the case for the dark matter component which feels only gravitational forces and thus quickly collapsed towards the shallow overdensities, causing them to grow. As the volume of causally connected space increased, this collapse proceeded on larger and larger scales.

Eventually local densities increased to the point where \( |\delta(x)| \approx 1 \) and structure formation transitioned to being a more complex, non-linear process. To gain insight into the ensuing evolution of the growing overdensities it is common to consider the idealised scenario of the collapse of an isolated and uniform spherical region of material (Gunn, 1977). The material which constitutes this sphere is imbued with kinetic energy by the general expansion of the Universe and will therefore grow in size until the time at which gravitational attraction takes over and it begins to collapse back in on itself. The time at which this turnaround from expansion to collapse occurs is inversely proportional to the magnitude of the initial overdensity, \( \delta(x) \), and hence the most overdense regions collapsed first (Binney & Tremaine, 2008).

The collapse of the overdensity will proceed until the point where it becomes virialised, entering an equilibrium state whereby the gravitational potential energy of the dark matter particles (\( U \)) approximately balances the associated kinetic energy (\( K \)) such that \( 2K + U = 0 \). This newly formed structure is commonly referred to as a dark matter “halo”. In a flat, matter dominated universe, the density contrast with respect to the global average at the time this virialisation occurs is \( \rho/\rho_0 \approx 200 \) (Peebles, 1980; Eke et al., 1996).\(^1\) The radius required to enclose an overdensity of this value is hence used as a common (but not unique) criterion to define the virial radius of a halo structure, \( R_{\text{vir}} \). The associated virial mass, \( M_{\text{vir}} \), is then simply given by the mass enclosed within this radius.

\(^1\)In fact \( \rho/\rho_0 \approx 178 \) in this case, however, this varies depending on the particular cosmological parameters and hence a value of 200 is used as a generally accepted approximation.

N-body simulations of dark matter structure formation (see §1.2 below) suggest that
Chapter 1. Introduction

Figure 1.1: Reproduction of two panels from figure 1 from Gao et al. (2008). This plot shows the stacked density profiles of approximately 400 dark matter halos extracted from the Millennium Simulation (black dots). The coloured lines in the left hand panel represent the results of a number of different NFW profile fits. There is a clear discrepancy between these fits and the true density profile, however, this is only true in the innermost regions of the halo. The right hand panel shows the results of fitting an Einasto profile, which almost completely eliminates this discrepancy.

Dark matter halos follow a near universal density profile which is well described by an NFW profile (e.g. Navarro et al., 1996, 1997) of the form:

$$\rho(r) = \frac{4r_s \rho_s}{r[1 + r/r_s]^2}, \quad (1.2)$$

where $\rho_s$ is the density at a characteristic scale radius from the centre, $r=r_s$. However, in practice this form fails to reproduce the shape of the profile in the inner most regions of the halo. Here, a more complicated Einasto profile (Einasto, 1965) has been shown to provide an excellent fit (Navarro et al., 2010; Gao et al., 2008, see also Fig. 1.1):

$$\ln \left( \frac{\rho(r)}{\rho_{-2}} \right) = -\frac{2}{\alpha} \left[ \left( \frac{r}{r_{-2}} \right)^\alpha - 1 \right], \quad (1.3)$$

where $\rho(r)$ is the density at a radius $r$ from the centre and $r_{-2}$ and $\rho_{-2}$ represent the radius and density respectively at the point where the logarithmic slope of the profile is equal to that of an isothermal density profile ($d \ln \rho(r)/d \ln r = -2$). The free parameter, $\alpha$, controls the rate of change of this slope.
1.1. Structure formation

1.1.2 The cosmic web

Of course, dark matter halos do not form in isolation as in our simple discussion above. In reality, multiple halos form simultaneously. These halos are gravitationally attracted to one another causing them to merge in a hierarchical fashion to form larger and larger conglomerations of matter. In tandem to this, they gain angular momentum through tidal torques (e.g. Peebles, 1969; White, 1984).

These growing dark matter structures are the sites of galaxy formation (White & Rees, 1978; White & Frenk, 1991). 300,000 years or so after the big bang, expansion allowed temperatures to drop enough for the ionised baryonic plasma to combine and form neutral atoms - the raw building blocks of stars, and hence galaxies. Photons were subsequently able to travel unhindered by scattering and the resulting lack of pressure support allowed the baryonic material to collapse into the potential wells provided by the dark matter halos. During the initial stages of the formation of the first galaxies, this primordial gas cooled and condensed down into the central regions of the halos to form rotationally supported disks (White & Rees, 1978). In this way, galaxies are a tracer of the central regions of dark matter halos and hence of the overall matter distribution.

Modern galaxy surveys such as the 2 degree Field Galaxy Redshift Survey (2dFGRS;
Colless et al., 2001) and Sloan Digital Sky Survey (SDSS; York et al., 2000) have given us a picture of the Universe in which galaxies are organised into a “cosmic web” of filamentary structures (see Fig. 1.2). This form of large scale structure is a natural result of continuing hierarchical growth. As the first dark matter halos began to collapse they left expanding and overlapping voids, directing the collapsing dark matter into sheets and filamentary structures (Binney & Tremaine, 2008). Ultimately, material was channeled to the intersection of these structures where galaxy clusters, the Universe’s largest bound structures, form. Only on scales larger than $\sim 100 \, h^{-1}\text{Mpc}$ does the self-similar nature of this structural pattern cease and the Universe can be considered homogeneous (Scrimgeour et al., 2012).

1.2 N-Body dark matter simulations

Although statistical techniques exist for constructing the mass distribution and merger histories of dark matter halos (e.g. Extended Press Schechter theory; Parkinson et al., 2008; Zhang et al., 2008), these are typically limited in both the precision and variety of the halo properties which they can describe. N-body dark matter simulations overcome these limitations by following the gravitational interaction of billions of dark matter particles in large cosmological volumes. Effects such as dynamical friction, violent relaxation and tidal torques are all naturally included. In addition, halo properties such as angular momentum and concentration can be calculated.

The raw output of N-body simulations is the mass, position and velocity of each individual particle. These are output at a number of predefined temporal snapshots. In order to identify individual dark matter halos in each snapshot, one must identify groups of gravitationally bound particles. There exists a number of different methodologies and dedicated codes to carry out this task (see Knebe et al., 2011, for an overview). However, the technique applied to generate the dark matter merger trees that are used throughout the work of this thesis is summarised below. For further details see Springel et al. (2005).

To begin with, a friends-of-friends (FOF) algorithm is employed to identify associations of particles. This algorithm works by grouping together all particles which are separated from one another by a carefully chosen “linking length” (Davis et al., 1985). When a halo merges with a more massive counterpart, it is not instantly destroyed but is rather stripped of its mass gradually over time within the potential well of its new parent. Such accreted halos are termed “subhalos” or “satellites”. In N-body simulations these satellites are identified by searching for regions of overdensity in the parent FOF halo. After identifying each satellite structure, they are subject to a gravitational unbinding procedure and only
1.3 Creating a mock universe: Semi-analytic galaxy formation models

Galaxy formation and evolution involves a complex interplay of processes, spanning a number of branches of astrophysics as well as a large dynamic range in scales. These include: cosmology, structure formation, stellar evolution, black hole physics and gas physics. Theoretical models developed to try and reproduce these elaborate systems have traditionally followed three distinct, but complimentary, methodologies (Neistein & Weinmann, 2010).

Statistical models, such as the halo occupation distributions (HOD; e.g. Peacock &...
Smith, 2000, see also §1.4.1), allow one to mimic a subset of the instantaneous properties of a large ensemble of galaxies. They are computationally inexpensive and are well suited to exploring clustering and other, purely statistical, quantities. However, their main drawback is that they contain no information about the physics which drives galaxy evolution. Additionally they are unable to track the history of a single galaxy over cosmic time.

At the other end of the spectrum, high-resolution hydrodynamical simulations attempt to fully model all of the important physics involved in the evolution of a galactic system. This is a computationally intensive procedure which naturally prohibits the simultaneous achievement of high spatial resolution in a cosmologically significant volume. Despite attempting to accurately follow all of the detailed processes as closely as possible, limits in the achievable resolution and in our knowledge of the “sub-grid” physics means that parameterised approximations are necessary for small scale processes such as star formation and black hole feedback.

Semi-analytic models (SAMs) exist somewhere between the extremes of these two methodologies. They were first introduced by White & Frenk (1991) and follow the general principles developed by White & Rees (1978) whereby baryonic gas collects in the deep potential wells provided by dark matter halos, before cooling to trigger a cascade of physical processes that ultimately lead to the diverse galaxy population observed to-
1.3. Creating a mock universe: Semi-analytic galaxy formation models

Semi-analytic models take, as their input, hierarchical dark matter halo merger trees. These trees may be calculated statistically using methods such as Monte Carlo realisations of Extended Press-Schechter theory (Press & Schechter, 1974; Bond et al., 1991; Parkinson et al., 2008). However, modern models more commonly obtain their trees from full N-body, dark matter only, cosmological simulations (a technique first carried out by Kauffmann et al., 1999). Due to the dominance of dark matter in the mass density of the universe, its large scale distribution (i.e. on scales larger than a single halo) is mostly unaffected by the presence of baryons. This allows us to pour a great deal of computational resources into generating accurate merger trees from N-body simulations of large, cosmologically significant volumes with excellent mass and time resolution. By post-processing these merger trees using a series of physically motivated analytic prescriptions to approximate the complex baryonic physics, semi-analytic models can then generate realistic galaxy populations without the need to ever rerun the expensive N-body calculations. These models thus provide both the ability to track the history of a single object (as in hydrodynamical simulations) and to simulate a large, cosmologically significant sample of galaxies with a comparatively small computational overhead (as in HOD models).

Modern semi-analytic models include a broad range of physical prescriptions that cover the dominant baryonic processes currently understood to shape the evolution of the galaxy population. In this section we provide a high-level overview of these processes, their effects on the evolution of galaxies and how they are typically implemented in a semi-analytic context. We try not focus on any particular published model, however, the topics discussed are largely motivated by the model of Croton et al. (2006) which is used extensively in chapters 3 & 4. For an excellent and more expansive treatment of semi-analytics the reader is referred to the review of Baugh (2006).

1.3.1 Cooling

In semi-analytic models it is assumed that each dark matter halo collapses with the universal baryon fraction in the form of gas. Initially, this gas is added to a “hot” reservoir that is shock heated to the virial temperature of the dark matter halo (Sutherland & Dopita, 1993):

\[ T_{\text{vir}} = \frac{1}{2} \frac{\mu m_p}{k} V_{\text{circ}}^2, \]  

(1.4)

where \( m_p \) is the proton mass, \( \mu m_p \) is the mean molecular weight of the gas, \( k \) is the Boltzmann constant and \( V_{\text{circ}} \) is the circular velocity at the virial radius of the halo.

If we further assume that this hot gas has an isothermal density distribution then the time required to cool and condense down onto the galaxy from a radius, \( r \), can be
Chapter 1. Introduction

Figure 1.5: Reproduction of the upper panel of figure 8 of Sutherland & Dopita (1993). This figure shows the normalised cooling function ($\Lambda(T, Z)$) curves for various metallicity values as a function of temperature.

Expressed as (White & Frenk, 1991):

$$t_{\text{cool}} = \frac{3}{2} \frac{kT}{\mu m_p \rho_{\text{gas}}(r) \Lambda(T_{\text{vir}}, Z)},$$

(1.5)

where $\rho_{\text{gas}}(r)$ is the density of hot gas contained within this radius. $\Lambda(T_{\text{vir}}, Z)$ is the cooling function (Sutherland & Dopita, 1993) which is dependent on both the temperature and metallicity of the gas (Fig. 1.5).

In order to calculate the mass of material which can cool in a time interval $\Delta t$, a cooling radius, $r_{\text{cool}}$, is defined at which the cooling time is equal to a physically motivated age of the halo. A discussion of the different choices for this timescale can be found in Lu et al. (2011a). However, a sample of some of the possibilities include the age of the Universe (e.g. Kauffmann et al., 1999), the free-fall time of the halo (e.g. Cole et al., 2000), the time since the last major merger (e.g. Somerville & Primack, 1999) or the halo dynamical time, $t_{\text{dyn}} = R_{\text{vir}}/V_{\text{circ}}$ (e.g. Springel et al., 2001).

For low mass halos, typically found at high redshift, $r_{\text{cool}}$ tends to be large. In cases where $r_{\text{cool}} > R_{\text{vir}}$, this indicates that there is not enough time for accreting material to reach hydrostatic equilibrium and hence the cooling rate simply becomes limited by the accretion rate onto the halo (White & Frenk, 1991). In high mass systems $r_{\text{cool}}$ lies well
1.3. Creating a mock universe: Semi-analytic galaxy formation models

within the virial radius and the time taken for the infalling gas to cool is long enough for it to become pressure supported and to form a hydrostatic hot atmosphere.

In cases where a such a hydrostatic hot halo is present, the cooling rate from this reservoir, $\dot{m}_{\text{cool}}$, can be calculated using a simple continuity equation for the mass flux across the evolving cooling radius (Croton et al., 2006):

$$\dot{m}_{\text{cool}} = 4\pi \rho_{\text{gas}} (r_{\text{cool}}) r_{\text{cool}}^2 \dot{r}_{\text{cool}}.$$  \hfill (1.6)

1.3.2 Star formation

Star formation in disk galaxies is generally observed to follow the Kennicutt–Schmidt relation whereby the star formation rate is simply proportional to the surface density of cold gas in the disk, $\Sigma_{\text{gas}}$, when averaged over scales $\gtrsim 10$ kpc (Kennicutt, 1998; Schmidt, 1959):

$$\Sigma_s \propto \Sigma_{\text{gas}} \Rightarrow \dot{M}_s \propto M_{\text{gas}}^n,$$  \hfill (1.7)

where $r_{\text{disk}}$ is the radius of the disk and $n \approx 1.4$.

As demonstrated by Kennicutt (1998), Eqn. 1.7 can be equivalently expressed as:

$$\dot{M}_s = \epsilon_{\text{SF}} \frac{M_{\text{gas}}}{t_{\text{dyn}}},$$  \hfill (1.8)

where $\epsilon_{\text{SF}}$ is a unitless normalising factor and $t_{\text{dyn}}$ is the dynamical time of the disk which can be expressed as $t_{\text{dyn}} = r_{\text{disk}}/V_{\text{circ}}$. Hence, with knowledge of both the disk mass and its radial extent, this relation can be used to provide an in-situ star formation rate. This form of the star formation law is, however, by no means unique (Somerville & Primack, 1999). For example, in the model of Croton et al. (2006), $M_{\text{gas}}$ is replaced by the excess mass above a given threshold that is dependent on the properties of the disk (Kauffmann, 1996). In other models $\epsilon_{\text{SF}}$ is often replaced by a function of the disk circular velocity (e.g. Cole et al., 2000).

Regardless of the particular form of Eqn. 1.8, the use of $t_{\text{dyn}}$ requires the ability to assign a realistic radial extent to the disk. Under the assumption of full conservation of specific angular momentum, the scale length of the disk, $r_s$, formed by cooling gas can be approximated using just the spin, $\lambda$, of the host dark matter halo (Mo et al., 1998):

$$r_s = \frac{1}{\sqrt{2}} \lambda R_{\text{vir}}.$$  \hfill (1.9)
Chapter 1. Introduction

For a typical dark matter halo with a spin parameter of $\lambda=0.05$ (Bullock et al., 2001), $r_s \approx 0.04R_{\text{vir}}$. Although there are a number of other assumptions involved in the derivation of Eqn. 1.9 (see Mo et al., 1998), the success of semi-analytic models at reproducing the star formation rate and stellar mass distributions of galaxies suggests that these approximations are reasonable when averaged over large samples of galaxies and when including all of the potential complicating factors such as supernova and AGN feedback.

When calculating the contribution to the stellar mass of a galaxy that results from a single star formation episode, the finite lives of stars must be taken into account. Although this can be treated self-consistently (Hatton et al., 2003; Benson & Bower, 2010), a standard approximation made in semi-analytic modelling is that of “instantaneous recycling” (Cole et al., 2000) whereby some fraction of the newly formed stars is assumed to be immediately recycled back into the cold gas reservoir of the galaxy. The value of this fraction is dependent on the assumed initial mass function (IMF), however, typical values are on the order of a few tens of percent.

As well as adding mass in the form of cold gas, these dying stars also enrich the ISM with new metals. A common method for modelling this is to assume that every star formation episode results in a fixed fraction of metals being returned to the ISM or hot halo. However, in some cases this fraction is taken to be a function of properties such as the instantaneous stellar metallicity (e.g. Benson & Bower, 2010).

1.3.3 Supernova feedback

For each episode of star formation that occurs, some fraction of the stars will end their lives as powerful supernovae. These events can deposit significant amounts of energy into the interstellar medium (ISM), reheating and possibly even ejecting large amounts of cold gas from the disk. Unfortunately, there is a broad variety of ways in which different semi-analytic models parametrise these effects. We therefore limit ourselves to the discussion of the particular implementation of Croton et al. (2006) which is the model we later use in Chapters 2 and 3.

When a star formation episode occurs, an average energy injection to the ISM is calculated:

$$\Delta E_{\text{SN}} = \epsilon_{\text{SN}}0.5V_{\text{SN}}^2\Delta M_*, \quad (1.10)$$

where $\Delta M_*$ is the mass of stars formed, $0.5V_{\text{SN}}^2$ is the mean energy of supernova output per unit of this mass and $\epsilon_{\text{SN}}$ is a free normalising parameter (Croton et al., 2006). This energy is then used to reheat a mass of gas back into the hot halo component (or some equivalent). If any energy is left over from this reheating event, it may be used to eject
some fraction of mass from the system entirely. Both reheated and ejected gas is typically assumed to carry with it the average abundance of metals contained in the disk. These metals play an important role in the further cooling of gas from the hot halo by providing metal lines as a mechanism to radiate energy and hence increase the cooling efficiency (Sutherland & Dopita, 1993).

Often, the material which is fully ejected from the galaxy/halo system is added to a separately tracked reservoir which may be allowed to later re-cool into the hot halo component. The rate and mechanism by which ejected gas is recaptured by the halo is an extremely uncertain component of semi-analytic models, depending on a number of possible factors, some of which may be external to the galaxy/halo system. For example, how far does this ejected material make it from the halo before it is re-accreted? What is to stop it from being attracted into the potential well of another nearby companion? Does the halo have to increase the depth of its potential well before re-accretion?

When cold gas is ejected from the disk, it becomes unable to participate in star formation until it is later re-accreted. In this manner, supernova feedback is able to regulate the in-situ stellar mass growth of a galaxy. However, due to the increasing depth of the potential well, it takes more energy to eject material from the disks embedded in more massive halos. This results in supernova feedback being most efficient in lower mass systems and is an important mechanism by which semi-analytic models are able to match the observed faint-end slope of the galactic luminosity function. Indeed, this was the initial motivation behind including supernova feedback in the models in the first place (Baugh, 2006).

Unfortunately, realistic levels of supernova feedback alone is often unable to fully account for the inefficiency of star formation required to reproduce the low baryon fractions of dwarf galaxies. Motivated by this, modern semi-analytic models also include the effects of the heating of accreting gas by ionising photons from the cosmological UV background. This can be efficient in reducing the levels of cooling in low mass halos and therefore allows lower, more physically plausible supernova feedback efficiencies to be used in order to regulate the stellar mass growth in these systems (Benson et al., 2002).

1.3.4 AGN feedback

It is generally thought that almost every galaxy in the Universe harbours a massive black hole at its centre. Although it is currently unclear as to what the initial seed was for the formation of these objects, their ubiquitous presence indicates that they likely share a common formation mechanism in all galaxies. The mass of the central black hole has been
Figure 1.6: Reproduction of figure 8 of Croton et al. (2006). The left hand panel displays the $K$-band galactic luminosity function produced by the fiducial Croton et al. (2006) model (solid black line) against observed data (blue error bars). The black dashed line shows effect of turning off AGN feedback. Without this feedback there is a clear over-prediction of the number of bright galaxies. The right hand panel shows the same result, this time for the $b_J$-band luminosity function.

Shown to be closely correlated with a number of galactic bulge properties such as mass, luminosity, concentration and velocity dispersion (e.g. Häring & Rix, 2004; Magorrian et al., 1998; Graham et al., 2001; Ferrarese & Merritt, 2000). In addition, suggested trends with other galactic properties such as total mass (e.g. Peng, 2007) and globular cluster number (Burkert & Tremaine, 2010) leads to a consistent picture in which the growth and evolution of a galaxy is intrinsically linked to that of its central black hole.

An important growth mechanism for central black holes is thought to be galaxy mergers. These mergers will typically result in the coalescence of the central black holes. Gas rich mergers are additionally expected to drive significant amounts of material towards the central regions of a galaxy where it is then captured and accreted on to the black hole. During such an event, the inspiraling material forms an accretion disk which becomes super-heated due to immense frictional forces. In some cases, accreting black holes may also launch massive collimated jets, extending to sizes exceeding that of the galaxy itself. These active galactic nuclei (AGN) are some of the most energetic objects in the Universe. The energy they inject into the interstellar medium of their host galaxies can be enough to heat or even completely eject large amounts of gas (e.g. through radiation or momentum driven winds). This can lead to a temporary reduction or even complete cessation of star
formation. Moreover, AGN are thought to be capable of preventing fresh incoming gas from cooling and condensing down onto the galaxy.

For the first decade of their existence, a common problem for semi-analytic models was the difficulty they had in reproducing the steep slope of the bright end of the luminosity function. As discussed in the previous section, supernova feedback is only efficient in reducing the level of cooling and star formation in lower mass halos. For more massive halos, this form of feedback simply is not powerful enough to lift a significant amount of material out of the deep potential wells. Several proposals were made for how to overcome this problem, all of which allowed the relevant authors to reproduce the observed slope of the luminosity function, but most of which were ad-hoc in terms of their physical motivation (see section 3.3 of Baugh, 2006, for an excellent overview).

Although semi-analytic models including the growth of central super-massive black holes had been investigated previously (e.g. Kauffmann & Haehnelt, 2000), it wasn’t until the works of Croton et al. (2006) and Bower et al. (2006) (amongst others) that these objects were successfully utilised to stem the growth of the most massive galaxies. In these models black holes grow primarily via major merger events which both result in the coalescence of the progenitor black holes and the potential accretion of large amounts of cold gas. However, their major effect on cooling and star formation comes from a quiescent, or “radio-mode” feedback mechanism. Here some fraction of the hot gas which condenses out of a hydrostatic hot halo is assumed to be accreted onto the central black hole, causing AGN activity that can slow, or even completely prevent, any further cooling. Radio-mode AGN can thus limit the fresh supply of cold gas in the disk that can fuel further star formation. However, since this feedback mechanism requires both a well established hydrostatic hot halo and a central black hole massive enough to prevent significant cooling, conversely to supernova feedback, it is only really effective in the most massive galaxy/halo systems (see Fig. 1.6).

### 1.3.5 The central–satellite dichotomy

Observations of galaxy clusters reveal a clear dichotomy in the way that the central massive galaxy and the smaller orbiting members evolve and interact with their surroundings. In particular, satellite galaxies often have redder colours and lower gas fractions with decreasing radius from the cluster centre (e.g. Butcher & Oemler, 1984). This is thought to be due to a number of environmental processes such as ram pressure stripping and gas strangulation (Gunn & Gott, 1972; Balogh et al., 2000, e.g.) as these objects sink deeper into the massive potential well of the cluster halo.
In order to mimic these processes, semi-analytic models often treat the galaxies of satellite halos differently from their central counterparts. Again, there can be a great deal of variation from model to model, however, a common approach is to assume that the hot halo of a satellite is stripped away upon infall and added to the hot component of the central. This prevents further cooling of gas onto the infalling galaxy, causing it to rapidly redden as it runs out of fuel for star formation (Cole et al., 2000; Croton et al., 2006). In addition, more advanced recipes of stripping have been explored which allows this hot gas to be removed gradually (e.g. Benson et al., 2002; Font et al., 2008; Guo et al., 2011), more in keeping with the results of hydrodynamic situations (McCarthy et al., 2008).

Whilst some semi-analytic implementations treat all satellite galaxies equally, others take a more complicated approach. For example, Somerville & Primack (1999) allow mergers between two satellites as they orbit in the potential of a central, while Benson & Bower (2010) implement an arbitrarily deep nesting of satellites within satellites to similar effect. Guo et al. (2011) take a slightly different approach, using only a single satellite/central definition but allowing satellites which manage to pass outside the virial radius of their central halo to once more become centrals themselves.

1.3.6 Galaxy mergers

As halos coalesce and merge hierarchically, this inevitably leads to the merging of the galaxies which they host. In semi-analytic models, merger events occur between satellite galaxies and their associated parent centrals. After crossing over the virial radius of the parent halo, the time taken for a satellite to infall to the centre and merge can be modelled using dynamical friction arguments (see section 8.1 of Binney & Tremaine, 2008). Different models include varying amounts of complexity when calculating the infall timescale, some using formulae calibrated from analytic arguments (Lacey & Cole, 1993) or simulations (e.g. Jiang et al., 2008) that include variations in the angular momentum of the satellite orbits and possible mass stripping (e.g. Benson & Bower, 2010).

Both due to the use of increasingly higher resolution simulations, and improvements in our techniques for identifying and tracking halo structures, infalling sub-halos can be followed in modern N-body simulations for a significant amount of time after accretion. Once tidal stripping has reduced the mass of the sub-halo below the resolution limit of the merger trees (typically around 20 particles) its position can still be represented by the most bound particle at the last snapshot it was identified. Again, dynamical friction arguments can then be used to determine how long the final stages of infall will last. Since the satellite galaxy is typically located deep in the potential well of its parent by the time
Creating a mock universe: Semi-analytic galaxy formation models

1.3. Creating a mock universe: Semi-analytic galaxy formation models

this merger timescale is calculated, the effects of tidal stripping and orbital variations can be largely ignored. If one further approximates the density profile interior to the satellite radius as isothermal, the dynamical friction timescale, \( t_{\text{fric}} \), can be expressed as (Binney & Tremaine, 2008):

\[
t_{\text{fric}} = \frac{1.17 V_{\text{circ}}}{G \ln \Lambda M_{\text{sat}}} r^2,
\]

where \( G \) is the gravitational constant, \( V_{\text{circ}} \) is the circular speed of the central halo at \( r \) (\( \equiv V_{\text{vir}} \) under the assumption the inner regions of the dark matter halo are isothermal), and \( M_{\text{sat}} \) is the mass of the infalling satellite. The quantity \( \ln \Lambda \) is a Coloumb logarithm which can be approximated using various combinations of the central and satellite masses. Croton et al. (2006) utilise the following relation:

\[
\ln \Lambda = \ln(1 + \frac{M_{\text{vir}}}{M_{\text{sat}}})
\] (1.12)

When the galaxy merger does eventually occur, hydrodynamical simulations suggest that the typical outcome is strongly dependent on the baryonic mass ratio of the two galaxies (Cox et al., 2008). For this reason, merger events are commonly split into two categories: major mergers, where the mass ratio of the progenitor galaxies is greater than \( f_{\text{merger}} \); and minor mergers where the mass ratio is smaller than this. Typically the value of \( f_{\text{merger}} \) is 25% to 33%; however, the exact outcome of a merger event is also dependent on a whole host of other properties, some of which are not followed in semi-analytic models (Cox et al., 2006). The precise merger ratio used to separate these two categories is therefore typically treated as a variable parameter.

Major merger events usually result in the complete destruction of the disk components of both galaxies and may also trigger a significant star formation episode. The resulting remnant galaxy is hence a gas poor elliptical. Minor mergers have a less profound effect on the morphology of the central galaxy, and typically leave its disk component intact. Although there is a great deal of variation in exactly how these merger events are treated from model to model in the literature, they often result in a burst of star formation and growth of the central galaxy bulge (e.g. Cole et al., 2000; Croton et al., 2006).

1.3.7 Intra-cluster light

Hydrodynamical simulations suggest that a significant fraction of stellar mass can be ejected during a galaxy merger and the subsequent relaxation of the system (Murante et al., 2007). A smaller fraction may additionally be stripped from the infalling satellite by ram-pressure and tidal forces if it falls towards a massive cluster dominant galaxy.
These lost stars end up in a diffuse halo component, or “intra-cluster light” (ICL). It has been suggested that significant fractions of the total stellar mass content of an infalling satellite may be lost to the ICL in the most massive halos (Monaco et al., 2006; Conroy et al., 2007; Behroozi et al., 2012).

The inclusion of an ICL component in semi-analytic models is a relatively recent development (Benson & Bower, 2010; Guo et al., 2011). However, they may prove to be an extremely important addition. A common issue with most models is a difficulty in matching the growth of the high mass end of the galactic stellar mass function (see Chap. 2 for more details). Hierarchical growth ensures that the most massive halos form late, however, observations suggest that the high mass end of the stellar mass function is already in place by $z=1$ (Fontana et al., 2004; Bundy et al., 2006; Brown et al., 2007). This observational trend is an important manifestation of mass “downsizing” (Cowie et al., 1996; Heavens et al., 2004; Conroy & Wechsler, 2009). The addition of AGN feedback effects have greatly helped to resolve this issue by curtailing in-situ star formation in massive galaxies, however, it does not reduce the build up of stellar mass brought in by merging satellite galaxies. By including satellite stripping to an ICL component, this ex-situ growth can also be reduced, allowing the models to more easily reproduce the lack of evolution at the high mass end of the stellar mass function. In addition, predictions of the amount of light contained in the ICL can be used as another important constraint on semi-analytic models in general (Benson & Bower, 2010).

1.3.8 Disk instabilities

Disk instabilities, such as bars or spiral arms, may enhance the amount of in-situ star formation by compressing the gas and hence aiding the process of gravitational collapse. These instabilities may be brought about naturally by perturbations introduced by star formation or gas inflow. Alternatively they may be instigated by dynamical interactions with other nearby galaxies and dark matter halos (e.g. Byrd et al., 1986; Moore et al., 1998).

As these instabilities drive shocks through the gas, the excess energy is dissipated away causing gas to flow inwards towards the central regions of the galaxy. In addition, strong bars also have the ability to gravitationally alter the orbits of stars, causing them to fall inwards towards the central regions. The resulting random orbital encounters these stars have with each other can lead to the formation (or growth) of a dynamically supported bulge component (e.g. Kormendy & Kennicutt, 2004).

Early N-body simulations suggest that the stability of a galactic disk to the formation
of such instabilities can be described by the following relation (Efstathiou et al., 1982):

\[ \epsilon_{\text{instab}} = \frac{V_{\text{max}}}{\sqrt{GM_{\text{disk}}/R_{\text{disk}}}}, \]  

(1.13)

where \( V_{\text{max}} \) is the maximum circular velocity of the system, \( M_{\text{disk}} \) is the stellar mass of the disk and \( R_{\text{disk}} \) is the exponential scale length.

If \( \epsilon_{\text{instab}} \lesssim 1 \) then the disk becomes unstable under its own self-gravity (Mo et al., 1998). In this situation, semi-analytic models tend to assume the formation of a transient bar structure which transfers enough gas and/or stellar mass to the bulge component to return stability. In addition the formation of the bar may also induce a burst of star formation through the production of shocks in the disk.

### 1.3.9 Galaxy luminosity and dust extinction

The intrinsic observable quantity of all galaxies is photon flux. This can be computed in semi-analytic models by assuming an IMF for each star formation burst and then using a stellar population synthesis (SPS) model to calculate what the spectrum or broad band flux of this burst would look like to an observer. The particular choice of SPS model can have significant consequences for the galaxy magnitudes predicted by the models (Henriques et al., 2012). For example, the inclusion of the effects of thermally pulsating asymptotic giant branch stars can alter the predicted \( K \)-band magnitude of a galaxy by up to 1 magnitude at \( z=1 \) (e.g. Tonini et al., 2011).

In addition to the choice of SPS model, one must also consider the effects of internal dust extinction. As semi-analytic models do not typically have any information about the spatial distribution of stars or gas in each galaxy, this can only be modelled approximately. However, there is again significant variation in the precise treatments of different models. Some simply assume the dust of a galaxy to be distributed in a uniform “plane-parallel slab” (Kauffmann et al., 1999). Combined with a random inclination angle for the galaxy relative to the observer, this dust model can provide reasonable first-order corrections for the effects of dust extinction (Croton et al., 2006; Guo et al., 2011). Alternatively, some models apply corrections with more realistic dust distributions that take into account the clumpiness of the ISM (Cole et al., 2000; Benson & Bower, 2010). Finally, there have been a number of works which have directly coupled semi-analytic models to detailed radiative transfer codes (e.g. Granato et al., 2000) or fitted libraries created from these codes (e.g. Fontanot et al., 2009b; Almeida et al., 2010). Such methods have the advantage of being more accurate than the simple approximations described above, however, they are also
considerably more computationally demanding.

1.4 Alternative modelling techniques

Having introduced semi-analytic models and their implementation in §1.3 above, we now provide a brief discussion of some of the alternative techniques for modelling galaxy formation. The aim here is to place semi-analytics in the context of these alternatives, thus highlighting its relative strengths and weaknesses.

1.4.1 HODs

The halo model is a popular tool for statistically describing the spatial distribution of particular galaxy populations (e.g. Seljak, 2000; Cooray & Sheth, 2002; Berlind et al., 2003; Tinker et al., 2008; Blake et al., 2008). Within this model, the clustering of galaxies is expressed using two separate components: the ‘1-halo term’ which describes the contribution from pairs of galaxies within a single halo, and the ‘2-halo term’ which similarly describes the contribution due to galaxies in different halos.

An important ingredient of this framework is the Halo Occupation Distribution (HOD) which describes the number of galaxies within a halo of a given mass, $M$. The HOD has a number of free parameters which control, for example, the minimum halo mass which will contain a single galaxy, and the concentration of the dark matter density profile. By calibrating these parameters to reproduce the observed clustering of galaxies, one can produce realistic mock catalogues. In addition, one can use observations of particular galaxy classes (e.g. luminous red-galaxies; Blake et al., 2008) or as a function of luminosity (e.g. Zehavi et al., 2011) to probe the way in which these galaxies populate dark matter halos. This can also allow the calculation of quantities such as mass–light ratios (Tinker et al., 2005). The dark matter halo properties may be generated statistically (e.g. use extended Press–Schechter techniques), or they may be taken from the merger trees of N-body simulations (Benson et al., 2000; Zheng et al., 2005).

An important difference between HOD and semi-analytic models is that HODs require no information about the formation and evolution of galaxies. The results are statistically constrained to reproduce the correct galaxy clustering at individual redshifts. However, this limits the ability of these models to be used to understand galaxy formation as there is no way to self-consistently connect individual galaxies to their progenitors at earlier times or descendants at later times.
1.4.2 SHAM models

Sub-Halo Abundance Matching (SHAM) models take an alternative approach to HODs. Here galaxies are drawn from an observationally determined luminosity or stellar mass function and then assigned to a dark matter subhalo taken from the merger trees of N-body simulations. In the simplest case, this assignment is made by rank ordering the subhalos and galaxies in terms of mass and then monotonically assigning the most massive galaxy to the most massive halo and so on (Conroy et al., 2006). In theory this methodology requires no free parameters (other than that of the cosmological model). However, in practice this is not the case as one must include effects such as the statistical scatter in the galaxy–halo assignment due to differences in halo formation histories (Moster et al., 2012).

Since the dark matter subhalos are drawn from hierarchical merger trees, one knows the formation history of each subhalo. Hence, by averaging over large samples, these models have had considerable success in constraining the mean stellar–halo mass relation as a function of redshift as well as the importance of an intra-cluster light component in the accounting of galaxy stellar mass (e.g. Conroy et al., 2007; Behroozi et al., 2012; Moster et al., 2012). However, like HOD models, SHAM models are applied at individual redshifts separately, therefore there is no way to self-consistently track the evolution of a single galaxy over time. This is a key disadvantage when compared to semi-analytic models.

1.4.3 Hydrodynamic simulations

The final category of galaxy modelling which we consider is that of hydrodynamical (hydro) simulations. These provide a much more detailed treatment of galaxy formation, modelling the physics of typically millions of particles that represent both dark and baryonic matter. Such a complex undertaking results in these simulations being extremely computationally intensive, often taking weeks or months to run on a modern supercomputer with typical resources. This prohibits the ability to simultaneously model cosmological volumes whilst having enough resolution to follow the collapse of even the most massive star forming regions. To deal with this, one must fall back on so called “sub-grid” prescriptions to model the processes of star formation, supernova feedback, black hole growth, etc. In this respect, hydro simulations suffer from many of the same uncertainties as semi-analytic models.

On larger scales which can be fully resolved, hydro simulations model the physics of gas cooling, shocks and other complex dissipative processes. However, due to the
discretized nature of the mass and temporal resolution, there are a number of, often poorly understood, numerical effects which one must contend with. A number of different numerical schemes may be used to solve the hydrodynamic equations (e.g. grid and smooth-particle methods) and there can be significant differences between the resulting galaxies produced by each (Frenk et al., 1999; Kereš et al., 2012). Additionally, even with sub-grid prescriptions, current hydro simulations are still unable to describe the large volumes probed by most modern galaxy surveys. A final issue is the difficulty that many simulations have in matching a number of key observed statistical properties of the galaxy population such as the stellar mass function (e.g. Crain et al., 2009).

Despite all of these issues, hydro simulations provide a level of detail and spatial information that semi-analytic models can never achieve. They also provide the only method to simulate a number of key physical processes associated with galaxy evolution such as gas cooling, galaxy mergers and disk instabilities. For this reason, they are used extensively to inform the prescriptions employed by modern semi-analytic models. Hydrodynamic simulations and semi-analytics are thus complementary, with both being required together, along with HOD and SHAM models, to obtain the most complete theoretical picture of galaxy formation and evolution.

1.5 Thesis structure

The aim of this thesis is to explore current semi-analytic techniques for modelling the formation and evolution of galaxies. Specifically, we investigate their strengths, weaknesses and uses as well as exploring phenomenologically motivated alternatives. To this end, we:

1. apply Monte Carlo Markov Chains as a statistical technique to rigorously constrain the free parameters of a semi-analytic galaxy formation model at $z=0$ and $\sim1$ simultaneously (which has never previously been done);

2. apply semi-analytic modelling techniques in a novel manner to gain insight into the current state and future evolution of the Milky Way and M31; and

3. present a new modelling technique that possesses many of the advantages of both semi-analytic and HOD models, but which circumvents a number of their drawbacks, opening up new possibilities for exploring galaxy evolution to high redshift.

A brief overview of each chapter now follows.

In Chapter 2 we present a detailed statistical analysis of the Croton et al. (2006) semi-analytic model using Monte-Carlo Markov Chains. In particular we focus on the
1.5. Thesis structure

ability of this model to reproduce the observed stellar mass function and black–hole bulge relation at $z=0$ and 0.8. We find that the model is capable of providing an excellent reproduction of the relevant observations at each redshift independently. However, when constraining against both redshifts simultaneously tensions in the physical prescriptions are found. In particular, our results highlight potential deficiencies in the way in which star formation and supernova feedback are treated.

In Chapter 3 the same semi-analytic model is used, in combination with Sloan Digital Sky Survey (SDSS) data, to investigate the evolutionary state of the Milky Way and M31 in a cosmological context. By comparing the observed global colours and star formation rates of these two galaxies to a sample of analogues drawn from either the semi-analytic model or SDSS data, we suggest that both the Galaxy and M31 may be “green valley” objects, undergoing an important evolutionary transition as they quiescently evolve towards being passive red spirals. This somewhat novel use of the semi-analytic model thus reveals the possible fate of our own galaxy and nearest massive neighbour. Using the histories of our simulated analogue galaxies, we also investigate the likely drivers of this evolution, finding the regulation of cold gas supply by the central super-massive black holes to be a possible cause.

In Chapter 4 we take a step back from detailed semi-analytic models and introduce a new “formation history” model for galaxy evolution. This new model dispenses with the implementation of multiple prescriptions to describe the various baryonic processes associated with galaxy formation. Instead, it replaces them with two simple, phenomenologically motivated equations that directly tie the stellar mass growth of a galaxy to the formation history of its host dark matter halo. We demonstrate the ability of this model to successfully reproduce the red and blue stellar mass functions at $z=0$ and investigate potential additions that allow it to also match the observed global stellar mass function all the way out to at least $z=4$. We conclude by outlining the general utility of this new model and its advantages over other galaxy modelling methodologies.

Finally, in Chapter 5, we draw together and summarise the findings of each chapter before moving on to describe the future prospects and possible extensions of this work.
Constraining the last 7 billion years of galaxy evolution in semi-analytic models

We investigate the ability of the Croton et al. (2006) semi-analytic model to reproduce the evolution of observed galaxies across the final 7 billion years of cosmic history. Using Monte-Carlo Markov Chain techniques we explore the available parameter space to produce a model which attempts to achieve a statistically accurate fit to the observed stellar mass function at \( z=0 \) and \( z\approx0.8 \), as well as the local black hole–bulge relation. We find that in order to be successful we are required to push supernova feedback and gas re-incorporation efficiencies to extreme limits which are, in some cases, unjustified by current observations. This leads us to the conclusion that the current model may be incomplete. Using the posterior probability distributions provided by our fitting, as well as the qualitative details of our produced stellar mass functions, we suggest that any future model improvements must act to preferentially bolster star formation efficiency in the most massive halos at high redshift.

Chapter 2. Constraining the evolution of semi-analytic models

2.1 Introduction

Modern semi-analytic galaxy formation models are a commonly used tool to aid in interpreting the statistical properties of large galaxy samples (e.g. Kauffmann et al., 1999; Hatton et al., 2003; Croton et al., 2006; Bower et al., 2006; De Lucia & Blaizot, 2007; Somerville et al., 2008; Guo et al., 2011; Benson, 2012). In a ΛCDM universe, the physical properties of galaxies are largely determined by the attributes of the halos in which they form, such as their mass and merger history (Mo et al., 1998). Semi-analytic models attempt to capture this dependence, as well as the complex baryonic processes involved in galaxy evolution, through a series of time evolving differential equations. Free parameters in the equations allow us to account for missing details in our understanding and/or implementations of the relevant physics.

Traditionally, these parameters are ‘hand tuned’ to accurately reproduce a small subset of important observations, as well as achieve a reasonable level of agreement with a larger number of other observed quantities. In this way, semi-analytic models have had considerable success in reproducing many of the most basic statistical quantities of the local Universe such as the galactic stellar mass function, the black hole–bulge relation, luminosity functions, colour–stellar mass relations, Tully-Fisher relations and correlation functions.

The procedure of manually calibrating model parameters can be extremely useful in developing an intuition for the importance of each of the component physical prescriptions and how they connect together. However, it is often a challenging and time-intensive task. The quality of fit is usually assessed visually, without providing a statistical measure of success. Hence there is no way to confirm that the chosen parameter values do truly provide the best possible reproduction of the data, or indeed that they are unique. Also, as the models become more sophisticated the number of free parameters naturally grows, as does the range of constraining observations. These parameters can have complex and highly degenerate interdependencies and, although the physically motivated parametrisations give us a broad idea of what the major effects of each parameter should be, it is extremely difficult to predict the exact consequences of any changes on the full range of galaxy properties produced. This problem is ubiquitous in any flavour of galaxy formation simulation.

Fortunately, semi-analytic models are relatively computationally inexpensive, especially when compared to full hydrodynamical galaxy formation simulations, and thus can be run quickly. This provides us with the ability to explore the parameter space of these models in a sensible time frame, allowing us to not only find the precise parameter values
that produce the best match to the observable Universe, but also understand the complex interplay between the included physical processes. As a result, there have been a number of attempts to automatically calibrate semi-analytic models using Bayesian statistical tools such as Monte-Carlo Markov Chains (MCMC; e.g. Henriques et al., 2009; Lu et al., 2011b, 2012).

MCMC techniques have only recently been applied to the task of calibrating galaxy formation models, despite having been used extensively for a number of years in other areas of astronomy such as cosmological parameter estimation (e.g. Lewis & Bridle, 2002). Kampakoglou et al. (2008) was the first, constraining a fully analytic model of star formation against a number of observations. These included the cosmic star formation history and type-II supernova rate out to high redshift. They also applied a novel Bayesian procedure to account for unknown systematics in their observational datasets.

In parallel to this work, Henriques et al. (2009) investigated the De Lucia & Blaizot (2007) semi-analytic galaxy formation model by calibrating it against the redshift zero $K$-band luminosity function, colour–stellar mass relation, and black hole–bulge relation. Using their results, they were able to draw conclusions about the interplay of the different parameters in their model, as well as highlight some potential tensions in simultaneously matching both the black hole–bulge relation and $K$-band luminosity function. Following on from this, Lu et al. (2012) calibrated a generic semi-analytic model (Lu et al., 2011b), again against the $z=0$ $K$-band luminosity function. The large number of free parameters and general construction of their model allowed them to mimic the implementations of a number of different previously published models, and thus to make more wide-reaching arguments about the success of semi-analytics in general when attempting to replicate the observed Universe. Using a method outlined in Lu et al. (2011b), the authors also used the parameter probability distribution to place uncertainties on a number of predictive quantities, both in the local Universe and out to higher redshifts.

Rather than also implementing MCMC methods, Bower et al. (2010) introduced the novel Bayesian technique of model emulation to calibrate the Bower et al. (2006) semi-analytic model. Their constraints were the $z=0$ $K$ and $b_J$-band luminosity functions. While model emulation provides a significantly better scaling with large numbers of parameters than MCMC methods, the details of its application are relatively complex. In a companion paper, Benson & Bower (2010) implemented this technique to aid in calibrating a new and updated version of the Bower et al. (2006) model against a large range of 21 different observations and across multiple redshifts. They also utilised 29 free parameters. Their aim, however, was not to provide an accurate statistical reproduction of each
of these observations, but to provide a final model which did a reasonable job of qualitatively matching as many of them as possible. Each of the observations was therefore given an arbitrary weighting in the total likelihood calculation. In addition, a last manual adjustment of the parameters was made to provide their fiducial model.

In this chapter, we use Monte-Carlo Markov Chain techniques to statistically calibrate the Croton et al. (2006) semi-analytic model. In particular we investigate whether the published version of this model is capable of replicating not only the present day Universe, but also the time evolution of the full galaxy population. To achieve this, we extend these previous works by considering the evolution of the galactic stellar mass function between $z=0$ and $z=0.8$, but with the restriction that we must simultaneously match the $z=0$ black hole bulge relation. Our work can most closely be compared with that of Henriques et al. (2012), as we use a similar model and one which is also run on the merger trees constructed directly from N-body simulations. However, comparisons can also be made with the works of Lu et al. (2012, 2011b) and Bower et al. (2010), given the similarities in the utilised modelling and analysis techniques.

We emphasise that in all of the aforementioned studies to which our work can be compared, the relevant models have only been constrained to match observations of the local Universe. While Lu et al. (2012) provides predictive quantities out to high redshift to draw valuable conclusions about the validity of their model across time, the work of this chapter constitutes the first time that a robust statistical calibration has been carried out at two redshifts simultaneously. This allows us to definitively test the ability of our model to reproduce the observed growth of stellar mass in the Universe at $z \lesssim 1$.

This chapter is laid out as follows: In §2.2 we introduce Monte-Carlo Markov Chain methods and discuss some of the details of our particular implementation. In §2.3 we provide a brief overview of the Croton et al. (2006) semi-analytic model, focussing on the physical prescriptions that are of particular relevance to this chapter. We then move on to describing the observational quantities we use to constrain our model in §2.4. Our results and analysis are presented in §2.5, with a detailed discussion of their significance found in §2.6. Finally, we conclude by summarising our main results in §2.7.

A standard ΛCDM cosmology with $\Omega_m=0.25$, $\Omega_\Lambda=0.75$, $\Omega_b=0.045$ is utilised throughout this chapter. All results are quoted with a Hubble constant of $h=0.7$ (where $h=H_0/100\text{km}s^{-1}\text{Mpc}^{-1}$) unless otherwise indicated.
2.2 Method

In general, we wish to find the model parameter set with the highest statistical likelihood as well as its uncertainty, given various observational constraints. In theory, the most straightforward way of achieving this is to invoke the Law of Large Numbers and draw independent samples from the joint posterior probability distribution function (PDF) of the model parameters; this is the probability of each parameter combination, given the set of constraining observations. Unfortunately, the presence of complex interdependencies between the parameters means we often do not know the form of the complicated joint posterior a-priori.

One way to overcome this problem is to implement Monte-Carlo Markov Chain methods. This is a Bayesian statistical technique for probing complex, highly degenerate probability distributions. Specifically, we employ the commonly used Metropolis–Hastings algorithm (Metropolis et al., 1953; Hastings, 1970). In this section, we provide a brief outline of the methodology of this algorithm as well as detail our particular implementation. A more general overview of MCMC and Bayesian techniques can be found in other works (e.g. Lewis & Bridle, 2002; Trotta, 2008, and references therein).

2.2.1 MCMC Implementation

We begin by selecting a fixed point in parameter space, \((\theta_i)\) and calculating the likelihood for this particular parameter set given our chosen constraints. Next we select another random position \((\theta_{i+1})\) a distance \(l\) away, where \(l\) is the ‘linking length’ of the chain, and calculate the associated likelihood. If this likelihood is higher than that of the previous parameter set then we accept this new proposition and store it in our chain. The cycle is then repeated.

One could imagine that, given enough iterations, this method would eventually lead the chain to settle at the most likely parameter values and remain there. However, the aim is to map out the posterior distribution and not just find the position of maximum likelihood. The chain is hence allowed to accept some propositions with a lower likelihood than that of the current value. The probability of this occurring is calculated from the ratio of the proposal \((i_{th}+1\) state) and current \((i^{th}\) state) likelihoods:

\[
P(\theta_{i+1}) = \min \left(1, \frac{P(\theta_{i+1})}{P(\theta_i)} \right)
\]

(2.1)

In this manner, the chain will map out the desired model parameter joint posterior distribution, with the number of accepted propositions in any one region of parameter
space being proportional to the likelihood of that region.

The size of the chain linking length is important to achieving an efficient and well
mixed MCMC chain, and must be set appropriately depending on the topology of the
target distribution. In some cases the posterior distribution may have a number of local
probability maxima which may be separated by large regions of low probability. If the
linking length is too small then a large number of consecutive chain steps to lower likelihood
regions will be required to bridge any gaps between the likelihood peaks. If it is too large,
then almost all propositions will be rejected in the situation where the chain happens
upon a highly peaked region of probability space. In either of these cases, the efficiency
with which the MCMC chain will converge to the target probability distribution will be
reduced.

The efficiency of the Metropolis-Hastings algorithm can be further maximised by using
a suitable proposal distribution $q(\theta)$ when selecting each new proposition for the chain. If
this proposal distribution is correctly chosen then the chain propositions can be preferen-
tially scattered along the planes of degeneracies, thus allowing the use of a larger linking
length. To ensure that we still converge to the target distribution, equation 2.1 must be
modified to include $q(\theta)$:

$$P(\theta_{i+1}) = \min\left(1, \frac{P(\theta_{i+1}) q(\theta_{i+1} | \theta_i)}{P(\theta_i) q(\theta_i | \theta_{i+1})}\right)$$

(2.2)

With these efficiency considerations in mind, we carry out an auto-tuning phase at the
start of every MCMC run. To begin with the linking length is iteratively modified until a
proposition acceptance rate of $35 \pm 5\%$ is achieved. Next, we construct a covariance matrix
and rotate it via a Choleski decomposition to define our proposal distribution $q(\theta)$. Again,
this procedure is carried out iteratively until an acceptance rate of $35 \pm 5\%$ is achieved.
Finally, we implement no fewer than 2 more linking length and covariance matrix tuning
phases before then fixing both of these for the remainder of the run.

Depending on the starting parameter values, a subset of the first accepted propositions
in the chain may not map out the target distribution in the desired manner. This is due
to the fact that the chain may first need to wander out of a region of very low probability.
In order to accommodate this, the chain is initially allowed to run to stable configuration,
in what is known as a ‘burn-in’ period. None of the propositions in this phase are used in
the final analysis.

Although far more efficient than simply probing the entire $N$–dimensional parameter
space with a regular grid, a MCMC chain still typically requires many tens of thousands of
propositions to fully sample the posterior. This necessitates short run times, in the order of
a few seconds or less, for a single realisation of the semi-analytic model. For this reason, we cannot run each model iteration on the full dark matter merger trees of the entire input Millennium simulation. Instead we restrict ourselves to running on 1/512 of the full 0.125 \( h^{-3} \) Gpc\(^3\) volume. This is equivalent to a comoving volume of \( 2.44 \times 10^5 \ h^{-3} \) Mpc\(^3\).

Rather than choosing a contiguous sub-volume of the simulation to form our input merger tree set, we randomly subsample an equivalent fraction of the total number of merger trees. This moderates the effects of cosmic variance and also allows us to fully probe the halo mass function up to some maximum limiting mass. Note that we use the same merger tree sample for every MCMC chain and figure presented in this chapter. After making a number of technical changes to the code base, we reduce the run time for a single input file from approximately 1.5 minutes to a just a few seconds with the Croton et al. (2006) model running on 64 cores of the Swinburne University of Technology’s Green Machine\(^1\). These changes include load-balancing the input dark matter merger trees from each individual file across multiple CPU cores, as well as removing costly magnitude calculations.

For all of the results presented below, we combine two fully independent MCMC chains, each with 100 000 model calls in their integration phases. This is typically adequate to achieve well mixed and converged results except for explicit cases which we specifically highlight in the text. In order to assess this we implement the Rubin–Gelman statistic (assuming \( \hat{R} \leq 1.03 \) indicates convergence; Gelman & Rubin, 1992) as well as visually inspect the chain traces. We also run several other shorter test chains, all with different random starting positions, in order to ensure that we are not missing any discrete regions of high probability in our analysis.

2.2.2 Principle Component Analysis

The primary product of an MCMC analysis is the posterior probability distribution of the \( N \)-dimensional parameter space of the model, constrained against the relevant observables. This distribution contains a wealth of information, not only about the highest likelihood parameter combination, but also about the level to which each parameter is constrained and the degeneracies that exist between them. In order to aid with our interpretation of the posterior distributions, we carry out a principal component analysis (PCA) when appropriate. This method compresses the information contained in the PDF into as few basis vectors as possible. In practice, the problem reduces to an eigenvector decomposition of the covariance matrix, where the eigenvectors are the principal components and the

\(^1\)See http://www.astronomy.swin.edu.au/supercomputing/ for further details
corresponding eigenvalues provide a measure of the amount of variance they describe. By carrying out such an analysis on a MCMC chain, we are able to identify which parameters are responsible for describing the bulk of the scatter in the posterior probability distributions. Parameters which provide almost no variance in any of the principal components can thus be interpreted as being well constrained by the relevant observations.

There are underlying assumptions and limitations associated with PCA that necessitate care in its interpretation. This is particularly true in the case of pathological PDFs exhibiting multiple discrete probability peaks or highly non-linear degeneracies. We must therefore be wary of placing strong emphasis on the precise values obtained from such an analysis. However, PCA does provide a valuable tool for gaining a qualitative insight into which physical prescriptions of the model are most important for matching particular observations. In some cases, a visual inspection of the PDF may indicate that a parameter is well constrained by the relevant observations, however, a PCA analysis could indicate that it is in fact small variations in the value of this parameter that drives larger changes to the other parameters. Also, if by adding a new observational constraint the number of principal components decreases, this indicates that the new constraint adds information that successfully reduces the model freedom.

In order to carry out a principal component analysis of a posterior distribution, the following steps are followed: First, we take the integration phase of the MCMC chain and calculate the mean value for each model parameter. This value is then subtracted from all of the proposition vectors. Next, a covariance matrix is constructed and an eigenvector decomposition of this matrix carried out. Finally, the resulting eigenvectors are ranked in order of decreasing eigenvalue. As discussed above, the eigenvalues are a measure of the amount of variance described by each eigenvector. Deciding how many of the top ranked eigenvectors form the principal component set is arbitrary; however, we follow the common practice of including increasingly lower ranked vectors until we have recovered 90% or more of the total variance in our final set.

2.3 The semi-analytical galaxy formation model

In this section we describe the aspects of the Croton et al. (2006) semi-analytic model relevant to it’s use in this chapter. This model has a number of free parameters which regulate a broad range of physical processes from black-hole accretion and feedback, to the effects of cosmic re-ionisation. However, as in Henriques et al. (2009), we focus only on the six main parameters which regulate star formation, super-nova feedback and black hole growth (Table 2.1). These are less well constrained by observation or theory than
2.3. The semi-analytical galaxy formation model

<table>
<thead>
<tr>
<th>Parameter name</th>
<th>Physical prescription</th>
<th>Original value</th>
<th>$z = 0$</th>
<th>$z = 0.83$</th>
<th>$z = 0 + 0.83$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_{SF}$</td>
<td>In-situ star formation</td>
<td>0.07</td>
<td>0.019$^{+0.003}_{-0.003}$</td>
<td>0.044$^{+0.033}_{-0.019}$</td>
<td>0.055$^{+0.011}_{-0.016}$</td>
</tr>
<tr>
<td>$\epsilon_{disk}$</td>
<td>SN feedback</td>
<td>3.5</td>
<td>5.14$^{+1.22}_{-0.76}$</td>
<td>4.79$^{+1.87}_{-0.94}$</td>
<td>13.8$^{+4.1}_{-2.2}$</td>
</tr>
<tr>
<td>$\epsilon_{halo}$</td>
<td>SN feedback</td>
<td>0.35</td>
<td>0.26$^{+0.06}_{-0.03}$</td>
<td>0.41$^{+0.16}_{-0.06}$</td>
<td>1.18$^{+0.38}_{-0.20}$</td>
</tr>
<tr>
<td>$\gamma_{ej}$</td>
<td>Gas Reincorporation</td>
<td>0.5</td>
<td>$7.1^{+4.9}_{-4.8} \times 10^{-3}$</td>
<td>$7.1^{+1.0}_{-4.2} \times 10^{-3}$</td>
<td>$1.13^{+0.30}_{-0.24}$</td>
</tr>
<tr>
<td>$f_{BH}$</td>
<td>Black hole growth</td>
<td>0.03</td>
<td>0.015$^{+0.002}_{-0.003}$</td>
<td>0.015$^{+0.035}_{-0.010}$</td>
<td>0.025$^{+0.007}_{-0.007}$</td>
</tr>
<tr>
<td>$\kappa_{AGN}$</td>
<td>Black hole growth</td>
<td>$5.89 \times 10^{-4}$</td>
<td>$1.90^{+0.39}_{-0.33} \times 10^{-4}$</td>
<td>$1.71^{+1.13}_{-1.21} \times 10^{-4}$</td>
<td>$1.47^{+0.24}_{-0.67} \times 10^{-4}$</td>
</tr>
</tbody>
</table>

Table 2.1: The six free parameters of the semi-analytic model which we focus on in this chapter. The original values of Croton et al. (2006) are listed along with the best parameter values calibrated at $z=0$ (stellar mass function + black hole–bulge relation), $z=0.83$ (stellar mass function), and $z=0$ and 0.83 simultaneously. The quoted uncertainties represent the 68% confidence limits of the marginalised posterior distributions where appropriate. See §2.3 for a description of the role played by each parameter.
Chapter 2. Constraining the evolution of semi-analytic models

many of the other model parameters (c.f. Croton et al., 2006, Table 1), and are strongly
dependant on the particular implementations of the physical processes.

For each parameter we employ a flat prior in log_{10} space and our MCMC chains explore
the probability distributions for each parameter in this space. By making this choice we are
able to efficiently deal with those parameters which have plausible values spanning more
than an order-of-magnitude. The allowed range in values may vary between the different
constraint combinations presented below, however, we have ensured that the employed
limits always encompass the full range of likely values.

The remainder of this section is devoted to outlining the role that each of these six free
parameters play in shaping the properties of the galaxy population. For a more detailed
description that includes all of the physical prescriptions present in the model, the reader
is referred to Croton et al. (2006). Those already familiar with the model can forgo this
section.

2.3.1 Star formation and supernova feedback

In accordance with the work of Kennicutt (1998), star formation is regulated by a critical
surface density of cold gas. This is in turn related to the radius of the galaxy disk using the
empirical relation of Kauffmann (1996). Whenever the mass of cold gas (m_{cold}) exceeds
the critical mass suggested by this relation (m_{crit}), a burst of star formation occurs. The
star formation rate (\dot{m}_*) is then given by:

$$\dot{m}_* = \alpha_{SF}(m_{cold} - m_{crit})/t_{dyn,disk},$$

(2.3)

where \(t_{dyn,disk}\) is the dynamical time of the disk and \(\alpha_{SF}\) is a free parameter controlling
the efficiency at which the excess cold gas is converted into stars over this timescale.

With each new star formation episode, the highest mass stars will rapidly evolve and
end their lives as energetic supernovae. The injection of this energy into the galaxy
interstellar medium will heat up a fraction of the cold gas, expelling it from the plane of
the disk and into the surrounding hot halo component. The amount of cold gas reheated
in this way follows:

$$\Delta m_{reheated} = \epsilon_{disk} \Delta m_*.$$  (2.4)

The parameter \(\epsilon_{disk}\) is equivalent to the supernova wind mass loading factor with
Croton et al. (2006) fixing its value to be 3.5 based on the observations of Martin (1999).

The amount of energy released per unit mass over the relevant time interval is approx-
2.3. The semi-analytical galaxy formation model

imated by:

$$\Delta E_{\text{SN}} = 0.5\epsilon_{\text{halo}} V_{\text{SN}}^2 \Delta m^* ,$$  \hspace{1cm} (2.5)

where $0.5V_{\text{SN}}^2$ is the mean energy injected by supernova per unit mass of star formation and the parameter $\epsilon_{\text{halo}}$ controls the efficiency with which this energy can actually reheat the disk gas.

The amount of energy required to adiabatically reheat $\Delta m_{\text{reheated}}$ of cold gas and add it to the hot halo reservoir is given by:

$$\Delta E_{\text{hot}} = 0.5 \Delta m_{\text{reheated}} V_{\text{vir}}^2 ,$$  \hspace{1cm} (2.6)

where $V_{\text{vir}}$ is the virial velocity of the host dark matter halo and $0.5V_{\text{vir}}^2$ is the thermal energy per unit mass of the hot halo component. If $\Delta E_{\text{excess}} = \Delta E_{\text{SN}} - \Delta E_{\text{hot}}$ is positive then enough energy is provided to physically eject some fraction of the mass from the system entirely:

$$\Delta m_{\text{ejected}} = \frac{\Delta E_{\text{excess}}}{E_{\text{hot}}} m_{\text{hot}} = \left( \epsilon_{\text{halo}} \frac{V_{\text{SN}}^2}{V_{\text{vir}}^2} - \epsilon_{\text{disk}} \right) \Delta m^* .$$  \hspace{1cm} (2.7)

This ejected gas is added to an external reservoir of material from where it plays no further role in the current heating/cooling cycle. As the dark matter halo grows, some of this ejected material may fall back into the deepening potential well and will be added back into the hot halo component. The fraction of ejected material that is re-incorporated per halo dynamical time is controlled by the parameter $\gamma_{\text{ej}}$:

$$m_{\text{ejected}} = -\gamma_{\text{ej}} m_{\text{ejected}} / t_{\text{dynamical}} .$$  \hspace{1cm} (2.8)

2.3.2 Supermassive black hole growth and feedback

As discussed by Croton et al. (2006), eqn. 2.7 implies that for galaxies in halos with $V_{\text{vir}}^2 > \epsilon_{\text{halo}} / \epsilon_{\text{disk}} V_{\text{SN}}^2$, supernova feedback processes are unable to eject any material from the galaxy–halo system. For their choice of parameters, this corresponds to dark matter halos with $V_{\text{vir}} \gtrsim 200 \text{ km s}^{-1}$. In systems more massive than this supernova feedback becomes inefficient at suppressing the long term cooling of gas and associated star formation. The result is an over prediction of the number of high mass galaxies in the Universe. The inclusion of feedback effects from super massive central black holes provides a well motivated and physically plausible mechanism for further regulating the cooling of gas in these high mass systems.
Central black holes grow via two mechanisms in our model. The first is the ‘quasar’
mode which results from galaxy merger events. During such an event, the progenitor black
holes are assumed to coalesce with no loss of mass due to dissipative processes. A fraction
of the cold gas of the progenitor galaxies is also accreted by the newly formed central black
hole, increasing its mass further:

\[ \Delta m_{\text{BH,quasar}} = \frac{f_{\text{BH}} m_{\text{cold}} m_{\text{sat}}}{m_{\text{central}}} \frac{1}{1 + (280 \text{ km s}^{-1}/V_{\text{vir}})^2}, \tag{2.9} \]

where \( f_{\text{BH}} \) is a free parameter and the normalisation constant of 280 km s\(^{-1}\) was originally
motivated by the desire to reproduce the observed black hole–bulge relation (Kauffmann
& Haehnelt, 2000). This is the dominant growth mechanism for black holes in our model,
although it is important to note that this growth is not accompanied by the injection of
energy into the inter-stellar medium.

Black holes are also allowed to grow quiescently through the continual accretion of hot
gas in what is called the ‘radio’ mode. This is characterised by the following simple model:

\[ \dot{m}_{\text{BH,radio}} = \kappa_{\text{AGN}} \left( \frac{m_{\text{BH}}}{10^8 M_\odot} \right) \left( \frac{f_{\text{hot}}}{0.1} \right) \left( \frac{V_{\text{vir}}}{200 \text{ km s}^{-1}} \right)^3, \tag{2.10} \]

where \( \kappa_{\text{AGN}} \) is our last free parameter and \( f_{\text{hot}} \) is the fraction of the dark matter halo
mass in the hot component. In contrast to the quasar mode, here material accreted by
the black hole results in the injection of energy directly into the interstellar medium:

\[ L_{\text{BH}} = \eta \dot{m}_{\text{BH}} c^2. \tag{2.11} \]

The effect is a reduction, or even complete cessation, of cooling onto the disk. By regulating
the availability of cold gas in massive galaxies, this feedback mechanism is able to efficiently
reduce the normalisation of the massive end of the stellar mass function (c.f. Croton et al.,
2006, figure 8).

### 2.4 Observational Constraints

In this section we provide an overview of the observational constraints used in our analysis:
the stellar mass function and black hole–bulge relation. We also discuss the statistical
tests we employ to assess the quality of the reproduction achieved by our model. We
implement the stellar mass function constraint at both \( z=0 \) and \( z=0.83 \), and the black
hole–bulge relation constraint at \( z=0 \) alone. This allows us to test if our model can not
only reproduce the local and high redshift universes independently, but also if it can be successful at constraining the late time evolution of the galaxy population between these two epochs.

2.4.1 The Stellar Mass Function

The stellar mass function is a fundamental observable in the study of galaxy formation and evolution. It provides one of the most basic statistical descriptions of the galaxy population – the number of galaxies per unit stellar mass, per unit volume \((\phi)\) as a function of stellar mass \((M_*)\) – and is directly influenced by the full range of physical processes associated with the evolution of the galaxy population. It is therefore important for any successful galaxy formation model to be able to provide a realistic reproduction of this quantity.

For both our low and high redshift stellar mass functions, we have invested a great deal of effort to use the most suitable observations that permit the use of accurate uncertainties in our analysis. In order to fairly judge the ability of a model to reproduce any observational constraint, it is extremely important that the observations have realistic uncertainties. If these are under estimated, then the model likelihood will be unfairly punished for predicting slight deviations; if they are overestimated then the constraints on the model parameters will be poor.

Low redshift

There are a large number of local measurements of the galaxy stellar mass function available in the literature (e.g. Cole et al., 2001; Baldry et al., 2004b; Panter et al., 2007). Typically, stellar masses are inferred in these works through the use of empirically determined stellar mass–light ratios. Unfortunately, masses estimated in this way require the use of a number of implicit assumptions regarding the stellar initial mass function (IMF), star formation histories, and the integrated effects of dust extinction. As a result, these masses can often suffer from large systematic uncertainties (Conroy et al., 2009) which can be difficult to quantify and are often not included in published stellar mass functions.

In this chapter, we utilize the \(z=0\) stellar mass function of Baldry et al. (2008). The main advantage of this particular work, for our purposes, is that the quoted uncertainties include an estimate of the systematic contributions associated with the use of colour dependent mass-light ratios, as discussed above, as well as the usually considered Poisson uncertainties. This was achieved by considering the mass function produced using a range of different stellar mass determinations, aggregated from five independent works, of matching galaxies drawn from the Sloan Digital Sky Survey York et al. (SDSS 2000) New
In order to directly compare our model to these observations, we convert the averaged Kroupa (2001) and Chabrier (2003b) IMF used by Baldry et al. (2008) to the Salpeter (1955) IMF assumed by our model. This is done by applying a systematic shift of +0.26 dex to the stellar mass values of the observed stellar mass function. To be completely accurate we should vary the magnitude of this shift as a function of stellar mass (e.g. Wilkins et al., 2008), however, we expect the omission of this second order correction to have a limited effect on our results.

The particle mass of the Millennium simulation, from which our input dark matter merger trees are generated, is $8.6 \times 10^8 \, M_\odot h^{-1}$. Typically, $\sim 100$ particles are required to attain well resolved, non-stochastic merger histories for a dark matter halo. Using the default published model of Croton et al. (2006), this corresponds to galaxies with stellar masses of $\log_{10}(h^{-2} M_\odot) \approx 9.5$. Since we are using only 1/512 of the full simulation volume in our analysis, we are unable to fully average out the stochastic nature of the properties of galaxies below this mass and thus we use this as a conservative lower limit on the reliability of stellar masses generated by the model. We reflect this in our analysis by cutting our constraining observations to only include stellar masses above this lower limit. Similarly, our use of a smaller sub-volume of the full simulation means that we do not accurately sample the high mass end of the halo mass function and therefore the highest stellar masses. For this reason we also employ an upper limit on the stellar mass range over which we constrain the model of $\log_{10}(h^{-2} M_\odot) \approx 11.65$.

**High redshift**

In this chapter we also constrain the model using MCMC at redshifts greater than zero. Unfortunately, it is extremely challenging to measure the observed stellar mass function at high redshift. To fully sample both the low and high mass tails of the distribution simultaneously one requires a survey sample of both high depth and large volume. In addition to this, the systematic uncertainties associated with assumptions such as star formation histories become even larger at increasingly higher look-back times, and again, these systematics are often excluded from any quantitative analysis in the literature. It is therefore unsurprising that many published $z \gtrsim 0.5$ stellar mass functions display significant disagreement, sometimes to the extent that their $2\sigma$ uncertainties do not overlap.

In order to obtain a single $z \approx 0.8$ stellar mass function which provides a reasonable estimate of the systematics, we create a weighted average from a number of recently published Schechter function fits in the literature: Drory et al. (2009)($z=0.8-1.0$), Ilbert et al. (2008)($z=0.4-1.0$), and Ilbert et al. (2010)($z=0.4-1.0$).
2.4. Observational Constraints

Figure 2.1: The observed $z=0$ (red circles; Baldry et al., 2008) and $z=0.83$ (blue squares) stellar mass functions with 68% confidence limits employed as model constraints in this chapter. For clarity, the $z=0.83$ data has been shifted by $-0.02$ dex in stellar mass. The $z=0.83$ values and associated uncertainties are a result of homogenising and combining several published Schechter functions (see §2.4.1 for details).

Pozzetti et al. (2007) $(z=0.7–9.0)$. Each was converted from a Chabrier (2003b) IMF to a Salpeter (1955) IMF using a constant offset of $+0.24 \log_{10}(M_*/M_\odot)$ in stellar mass. All values were also homogenised to $h=1$ and the mass functions cut at the relevant mass completeness limits of each sample. Pozzetti et al. (2007) provides four mass function fits calculated from the VVDS (VIMOS-VLT Deep Survey; Le Fèvre et al., 2005) survey using two different sample selection criteria and two alternative star formation history models. In total, we therefore employ six observed stellar mass functions covering a redshift range of $z=0.7–1.0$.

Our final result was calculated by averaging these homogenised observations to provide a single mass function at a mean redshift of $z=0.83$. This was done as follows: All of the utilised mass functions provide $\pm 1\sigma$ uncertainties on the best fitting Schechter function parameter values. We incorporate these by taking each Schechter function in turn and generating 1000 realisations with parameter values randomly sampled from appropriate probability distributions. Ideally these distributions would account for the covariance which exists between the different fitted parameters, but this information was not pro-
vided in the relevant publications. Instead we sampled Gaussian (or skewed Gaussian) distributions centered on the best fit values and with their quoted standard deviations. The mean and 1σ uncertainties of \( \phi \) in each stellar mass bin are then calculated using the random realisations from all six of the observed input functions. To ensure consistency with the \( z=0 \) stellar mass function of Baldry et al. (2008) we demand that \( \phi \) and its upper uncertainty is less than or equal to the respective \( z=0 \) values at all stellar masses. Such a restriction is well justified given that the observed total stellar mass density is found to decrease by approximately a factor of a half between \( z=0–1 \) (Drory et al., 2009).

Our final aggregated \( z=0.83 \) stellar mass function is shown in Fig. 2.1. As a simple first order check of its validity, we confirm that the integrated stellar mass density, over the range of stellar masses present, is \( 0.64^{+0.21}_{-0.19} \) times that of the \( z=0 \) value. The upper and lower bounds here account for the uncertainty in the mass functions at both redshifts. This shows broad agreement with observational results (e.g. Marchesini et al., 2009). For comparison, Fig. 2.1 also displays the constraining \( z=0 \) observations of Baldry et al. (2008).

In order to calculate the likelihood of the model stellar mass functions, given the observational data, we use a simple \( \chi^2 \) statistic. For a single stellar mass bin:

\[
L_{\text{SMF}}(\theta) \propto \exp(-\chi^2(\theta)) = \exp\left(-\frac{(\phi_{\text{obs}} - \phi_{\text{mod}}(\theta))^2}{\sigma_{\text{obs}}^2 + \sigma_{\text{mod}}^2(\theta)}\right), \tag{2.12}
\]

where \( \theta \) is the set of model parameters used and \( \sigma \) represents the associated uncertainties in each measurement. We estimate \( \sigma_{\text{mod},i} \) using Poisson statistics to be \( \sqrt{n_i} h^3\text{Mpc}^{-3}\text{dex}^{-1} \), where \( n_i \) is the number of model galaxies in bin \( i \).

We note that a number of previous works which have calibrated semi-analytic models using MCMC techniques have tended to favor the use of the \( K \)–band luminosity function as their primary constraint instead of the stellar mass function (Henriques et al., 2009; Lu et al., 2012). As the \( K \)–band is well known to be a good proxy for stellar mass, both quantities provide comparable constraints on the galaxy population. As discussed above, it can be difficult to derive accurate stellar masses for observed galaxies due to the degeneracies and poorly understood systematics of dust attenuation, mass–light ratios and IMFs. Luminosity functions are, however, directly observable and it is for this reason that they have been adopted by previous works. Unfortunately, producing a luminosity function from a semi-analytic model involves many of the same poorly understood physics and systematic uncertainties. Specifically, we must include assumptions about dust attenuation and stellar population synthesis models, in order to convert model stellar masses to luminosities.
2.4. Observational Constraints

As discussed previously, having realistic estimates of the relevant uncertainties is important for our MCMC analysis. Thus, we prefer to implement the stellar mass function as the primary constraint, due to the availability of a number of works which provide a quantitative analysis of some of the uncertainties associated with measuring a stellar mass function at various redshifts (e.g. Baldry et al., 2008; Pozzetti et al., 2007; Marchesini et al., 2009). Although it is true that these uncertainties may still be underestimated, a similar estimate of the systematics associated with a model derived luminosity function is beyond the scope of this chapter.

2.4.2 The Black Hole–Bulge Relation

It is well established that the masses of central super-massive black holes show direct correlations with the properties of their hosts’ bulges (e.g. Magorrian et al., 1998; Häring & Rix, 2004; Sani et al., 2011). This suggests a physical connection between the mass growth of these two components. Given the importance of AGN feedback in shaping the galaxy population, especially for high galaxy masses at \( z < 1 \), it is important that our model be able to reproduce this observed relation. This is especially so if we wish to use the model to make any predictions for how black holes of different masses populate different galaxy types.

Similarly to Henriques et al. (2009), we implement the observations of Häring & Rix (2004) as our constraint for the \( z = 0 \) black hole–bulge relation. Their sample is comprised of 30 nearby galaxies (the majority of objects being \( \lesssim 42 h^{-1} \text{Mpc} \) away) with the bulge and black hole masses sourced from a number of different works.

Observationally, it is still unclear whether or not there is a significant evolution in the black hole–bulge relation between \( z = 0 \) and \( z = 1 \). In general, an evolution is predicted by theory (e.g. Croton, 2006), and is tentatively measured by a number of authors (e.g. McLure et al., 2006; Merloni et al., 2010). Unfortunately, observations of \( z > 0 \) black hole–bulge relations are generally hampered by systematic uncertainties which, when included, make the significance of a deviation from the null hypothesis of no evolution much less certain (Schulze & Wisotzki, 2011). For this reason, we choose not to implement a black hole–bulge relation constraint at \( z = 0.83 \).

To assess the likelihood of our model fit to the data, we implement the same likelihood calculation as that of Henriques et al. (2009). First, the galaxy sample is segregated into two bins defined by three lines perpendicular to the best fit relation of Häring & Rix (2004):

\[
\log_{10}(M_{\text{BH}}) = -0.89(\log_{10}(M_{\text{bulge}}/M_\odot) - 11) + \text{offset}
\]  
(2.13)
where offset=[5.39, 8.2, 12.23]. The binomial probability theorem is then used to calculate what the likelihood is of finding the ratio of observed galaxies above and below the H"{a}ring & Rix (2004) best fit line in each bin:

\[
L_{\text{BHBR}} = \begin{cases} 
2I_p(k, n - k + 1) & I_p \leq 0.5 \\
2(1 - I_p(k, n - k + 1)) & I_p > 0.5 
\end{cases}
\]

(2.14)

where \( k \) is the number of observed galaxies above the best fit line in each bin, \( n \) is the total number of galaxies in the bin, and \( p(\theta) \) is the fraction of galaxies above the best fit line from the model. \( I_p(x, y) \) is the regularised incomplete gamma function. As described in Henriques et al. (2009), the reason for using two formulae with conditions is to ensure that any excess of galaxies both above and below the best fit line results in a low likelihood (i.e. both tails of the distribution).

### 2.5 Analysis

In this section we present our main analysis of this chapter. First, we investigate the restrictions placed on the model parameters by the individual observations at \( z=0 \) and \( z=0.83 \). This allows us to test which parameters are most strongly constrained by each observation as well as identify any tensions between these constraints. The findings are then used to guide our interpretation when calibrating the model against all three of our constraints at both redshifts simultaneously (§2.5.3).

#### 2.5.1 Redshift zero

**The Stellar Mass Function**

We begin by considering the \( z=0 \) stellar mass function and investigate the restrictions placed on the model parameters by this constraint alone. The histograms on the diagonal panels of Fig. 2.2 display the 1-dimensional marginalised posterior distributions for each of the six free parameters. The highly peaked, Gaussian-like distributions of the star formation (\( \alpha_{\text{SF}} \)), supernova halo gas ejection (\( \epsilon_{\text{halo}} \)), and supernova cold gas reheating (\( \epsilon_{\text{disk}} \)) efficiencies indicate that these are well constrained by the observed \( z=0 \) stellar mass function alone. Conversely, the wide and relatively flat distributions of the merger driven black hole growth (\( f_{\text{BH}} \)) and radio mode AGN heating (\( \kappa_{\text{AGN}} \)) efficiencies, suggest that their precise values are not particularly well constrained. The remaining off-diagonal panels of Fig. 2.2 show the 2-dimensional posterior probability distributions for all combinations
2.5. Analysis

![Figure 2.2: 2-D posterior probability distributions (off-diagonal panels) for all combinations of the log of the six free model parameters when constraining the model against the z=0 stellar mass function alone. Black lines indicate the 1 and 2-σ confidence contours. The limits of each panel indicate the prior ranges. Orange circles show the location of the marginalised best values of each parameter. The histograms in the diagonal panels represent the 1-D marginalised probability distributions, with the 1 and 2σ confidence intervals shown by the dark and light shaded regions respectively.](image)

Although the z=0 stellar mass function alone does not allow us to say what the precise values of the merger driven black hole growth efficiency ($f_{BH}$) and radio mode AGN heating efficiency ($\kappa_{AGN}$) must be, it does place a strong constraint on their ratio. This is indicated by the 2-D posterior distribution of $f_{BH}$ vs. $\kappa_{AGN}$ which shows a strong correlation between the allowed values of these two parameters. This is a direct consequence of the degeneracy between central black hole mass (which is dominated by quasar mode growth and thus $f_{BH}$) and the value of $\kappa_{AGN}$ in determining the level of radio mode heating (c.f. Eqn. 2.10). This heating plays a key role in shaping the high mass end of the stellar mass function...
where supernova feedback becomes ineffective at regulating the availability of cold gas. A similar degeneracy was also noted by Henriques et al. (2009) when constraining their model against the observed K-band luminosity function.

A principal component analysis of the joint posterior suggests that its variance can be understood predominantly through the combination of two equally weighted principal components. The star formation efficiency \( \alpha_{\text{SF}} \) and supernova halo gas ejection efficiency \( \epsilon_{\text{halo}} \) provide almost no contribution to the variance in either component, indicating that both are truly well constrained by the stellar mass function. On the other hand, the value of the ejected gas reincorporation rate parameter \( \gamma_{\text{ej}} \) does contribute significantly to both components. Interestingly, the supernova cold gas reheating efficiency \( \epsilon_{\text{disk}} \) also makes a dominant contribution to one of the principal components, suggesting that, although it appears well constrained in Fig. 2.2, small variations about the mean can be accommodated by a combination of changes to the remaining parameters controlling black hole growth and AGN radio mode feedback \( f_{\text{BH}} \) and \( \kappa_{\text{AGN}} \).

As well as investigating the parameter constraints and degeneracies, we also wish to know what single set of parameters provides us with the best overall reproduction of the relevant observations. The orange points in Fig. 2.2 indicate the marginalised best parameter values. This is the parameter set around which there was the largest number of accepted propositions in the MCMC chain. These values, along with their 68% confidence limits, are presented in Table 2.1.

In Fig. 2.3 we show the stellar mass function produced by the model using these best fit parameters, as well as the constraining observations of Baldry et al. (2008) and the model prediction using the default Croton et al. (2006) parameters. The orange shaded region encompassing the best fit line indicates the associated 95% confidence limits. These are calculated using all of the mass functions produced during the integration phase of the MCMC chain. When compared to the original Croton et al. (2006) results, the best fit model more accurately reproduces the distribution over the full range of masses - in particular the dip and subsequent rise in galaxy counts that occurs around \( 10^{10} h^2M_\odot^{-1} \).

### The Black Hole–Bulge Relation

In order to break the above degeneracy between the merger driven black hole growth and radio mode AGN heating efficiencies \( f_{\text{BH}} \) and \( \kappa_{\text{AGN}} \), we require the addition of another constraint which directly ties the properties of the central black holes to those of the galaxies in which they form. Following Henriques et al. (2009), we turn to the observed black hole–bulge mass relation for this purpose. Unlike the stellar mass function, which
2.5. Analysis

Figure 2.3: The $z=0$ stellar mass function resulting from the best fit parameter values, as determined by constraining the model against the observed $z=0$ stellar mass function alone. The solid line with shaded region shows the model result, along with the 95% confidence limits calculated from the posterior distribution. Blue error bars indicate the constraining observations and 68% confidence regions of Baldry et al. (2008). The default Croton et al. (2006) prediction is shown by the red dotted line. Only stellar masses in the unshaded region of the plot were used to constrain the model. The model prediction when using the best fit parameters constrained against the $z=0.83$ stellar mass function is also shown for comparison (black dashed line; §2.5.2).

provides strong constraints in a number of parameter planes, the black hole–bulge relation only constrains the $f_{\text{BH}}-\alpha_{\text{SF}}$ (star formation efficiency) plane.

The utility of this particular constraint can be traced to the fact that it provides a relation between the mass of the central black hole and spheroidal component of a galaxy. Bulges can grow in the model via two different mechanisms. The first is through merger events. However, none of the six free model parameters directly influences the strength of this mechanism. The second method of bulge growth is via disk instabilities. We treat this using a modified version of the simple, physically motivated prescription of Mo et al. (1998) whereby, once the surface density of stellar mass in the disk of a galaxy becomes too great, the disk becomes unstable. In this situation, a fraction of the disk stellar mass is transferred to the bulge component in order to restore stability. Hence, bulge growth via this mechanism is regulated by the amount of stars already present in the disk as well
as the mass of new stars forming at any given time. These, in turn, are modulated by the efficiency of star formation ($\alpha_{SF}$). Black holes, on the other hand, gain the majority of their mass via the merger driven quasar mode which is regulated by $f_{BH}$.

The posterior probability distribution for the black hole–bulge relation constraint alone is shown in Fig. 2.4. As expected, increasing the efficiency of star formation ($\alpha_{SF}$), and therefore the growth of bulges through disk instabilities, requires an increase in the efficiency of black hole growth ($f_{BH}$). Although omitted here for brevity, the constraints provided by this observation on the three parameters which modulate star formation and supernova feedback, are extremely weak. However, in the case of the supernova halo gas ejection efficiency parameter ($\epsilon_{\text{halo}}$), the marginalised posterior distribution only overlaps with those of the stellar mass function constraint to within $2\sigma$. In other words, there is a slight tension between the parameter sets favoured by the black hole–bulge relation and stellar mass function.

In Fig. 2.5 we show the marginalised posterior distributions for the six model parame-
Figure 2.5: 2-D posterior probability distributions (off-diagonal panels) for all combinations of the log of the six free model parameters when constraining the model against the $z=0$ stellar mass function and black hole–bulge relation simultaneously. Black lines indicate the 1 and 2-$\sigma$ confidence contours. The limits of each panel indicate the prior ranges. Orange circles show the location of the marginalised best values of each parameter. The histograms in the diagonal panels represent the 1-D marginalised probability distributions, with the 1 and 2-$\sigma$ confidence intervals indicated by the dark and light shaded regions respectively. This figure can be directly compared with Fig. 2.2.

As expected, the distributions look similar to those of Fig. 2.2, with the exception that we now also tightly constrain the values of the merger driven black hole growth and radio

$$L(\theta) = L_{\text{SMF}}(\theta) \cdot L_{\text{BHBR}}(\theta)$$

(2.15)
Chapter 2. Constraining the evolution of semi-analytic models

Figure 2.6: The $z=0$ model stellar mass function (left) and black hole–bulge relation (right) obtained using the marginalised best parameter values found when constraining against both the $z=0$ stellar mass function and black hole–bulge relation simultaneously. For the stellar mass function, the solid line with shaded region shows the model result along with the 95% confidence region calculated from the posterior distribution. Blue error bars show the constraining observations of Baldry et al. (2008). In the right hand panel, the model galaxies are indicated by the shaded grey hexbins, while the observational constraint of Häring & Rix (2004) is shown as the red crosses along with the published best fit line.

mode AGN feedback efficiencies ($f_{\text{BH}}$ and $\kappa_{\text{AGN}}$) through the addition of the information contained in Fig. 2.4. Unfortunately, we are still only able to provide an upper limit on the value of the re-incorporation efficiency parameter, $\gamma_{\text{ej}}$. However, this does tell us that the model prefers re-incorporation of ejected gas to occur on a timescale longer than approximately 10 halo dynamical times (roughly equivalent to the Hubble time).

A principal component analysis indicates that the addition of the black hole–bulge relation constraint has reduced the number of principal components from two to one. This confirms that we have successfully reduced the freedom of the parameter values with respect to one another. Again we find that star formation efficiency, $\alpha_{\text{SF}}$, is fully constrained. However, we now find that supernova cold gas reheating efficiency, $\epsilon_{\text{disk}}$, also contributes practically no information on the variance of the joint posterior distribution. This is due to the fact that we have now restricted the allowed values of the black hole growth and radio mode AGN heating efficiencies ($f_{\text{BH}}$ and $\kappa_{\text{AGN}}$), thus preventing them from compensating for any small shift to $\epsilon_{\text{disk}}$, as was allowed when constraining against the $z=0$ stellar mass function alone.

In Fig. 2.6 we show the $z=0$ stellar mass function and black hole–bulge relation obtained using the marginalised best parameters. These parameter values form our fiducial $z=0$ set and are listed in Table 2.1.
A comparison with Fig. 2.3 indicates that our reproduction of the observed stellar mass function remains excellent. However, we note that the likelihood of the black hole–bulge relation is only 0.2 when including the stellar mass function constraint. This is caused by a slight tension between the preferred parameter values of these two constraining observations. A similar result was found by Henriques et al. (2009) when calibrating their model against the $K$-band luminosity function and black hole–bulge relation.

Despite this drop in likelihood, statistical agreement within $2\sigma$ is still achieved and the resulting black hole–bulge relation remains cosmetically acceptable. Finally, we note that a great deal of observational uncertainty remains in the precise normalisation and slope of the black hole–bulge relation. It is therefore possible that future black hole–bulge relation measurements will result in a relation that is more easily reconciled with the stellar mass function in our model.

### 2.5.2 High redshift

In the previous section we investigated the constraining power of the $z=0$ stellar mass function and black hole–bulge relation. These observational quantities allowed us to place strong restrictions on the values of all but one of the free model parameters. In this section we investigate the resulting $z>0$ model predictions and also implement our $z=0.83$ stellar mass function constraint on its own (§2.4.1) in order to test the restrictions it imposes on the model parameters.

In Fig. 2.7 we show the prediction of the default Croton et al. (2006) model parameter values (red dotted line), compared against the observed stellar mass function at $z=0.83$ (blue squares; see §2.4.1 for details). Similarly to the redshift zero case, the default model over-predicts the number of galaxies at low masses. However, it now also predicts a steeper slope to the high mass end than is observed.

Also shown in Fig. 2.7 is the result obtained using the fiducial $z=0$ parameters of the previous section. This model predicts an unrealistic build up of galaxies around $\log_{10}(M [h^2/M_\odot])=10.25$ which constitutes the population that will later evolve to fill the high mass end of the distribution at $z=0$. This over-density is a direct result of under-efficient supernova feedback, allowing lower mass galaxies to hold on to too much of their cold gas which is subsequently converted into stellar mass. Overall, the fiducial $z=0$ parameters appear to do a worse job of reproducing the $z=0.83$ observations than the original Croton et al. (2006) values.

The solid black line and shaded region of Fig. 2.7 indicate the best fit mass function and associated 95% confidence regions found by constraining against the observed $z=0.83$
Chapter 2. Constraining the evolution of semi-analytic models

Figure 2.7: The resulting $z=0.83$ stellar mass function and 95% confidence regions (black line and shaded region) produced by constraining the model against the observed $z=0.83$ stellar mass function alone. The constraining observations and 1σ uncertainties are shown as blue error bars (c.f. §2.4.1). Also shown for comparison are the $z=0.83$ stellar mass functions produced by the default Croton et al. (2006) (red dotted line) and $z=0$ fiducial (black dashed line) parameter values.

stellar mass function alone. An excellent agreement can be seen across all masses. The marginalised best parameter values and 68% uncertainties are listed in Table 2.1. A subset of the posterior probability distributions are presented in Fig. 2.8. For comparison, the equivalent 1 and 2σ confidence regions of the $z=0$ stellar mass function+black hole–bulge relation constraint from Fig. 2.5 are also indicated with grey contour lines.

As was the case when constraining the model against the $z=0$ stellar mass function alone (§2.5.1), there is little restriction on the values of black hole growth and radio mode AGN feedback efficiencies ($f_{\text{BH}}$ an $\kappa_{\text{AGN}}$). This is due to the fact that we have no black hole–bulge relation constraint to break the degeneracy between these two parameters. The star formation efficiency ($\alpha_{\text{SF}}$) confidence region is also significantly larger at $z=0.83$ compared to $z=0$. This is primarily a reflection of the larger observational uncertainties across all stellar masses at this redshift. Interestingly, we also find that the most likely value of the supernova cold gas reheating parameter, $\epsilon_{\text{disk}}$, shows little evolution between redshifts, despite the need to increase the upper prior limit on this parameter to account...
2.5. Analysis

Figure 2.8: 2-D posterior probability distributions for the star formation efficiency, $\log_{10}(\alpha_{\text{SF}})$, against the five other free model parameters when constraining the model against the $z=0.83$ stellar mass function alone. The limits of each panel are equivalent to the MCMC prior ranges. Orange circles indicate the location of the marginalised best values of each parameter. Black lines represent the 1 and 2σ confidence contours with grey lines indicating the equivalent results from Fig. 2.5. The histogram in the top panel displays the 1-D marginalised probability distribution for $\alpha_{\text{SF}}$, with the 1 and 2σ confidence intervals shown by the dark and light shaded regions respectively.
for a slightly extended probability tail extending past our original upper limit of $\epsilon_{\text{disk}}=10$.

The parameters controlling star formation and supernova halo gas ejection efficiencies ($\alpha_{\text{SF}}$ and $\epsilon_{\text{halo}}$) display the largest differences in posterior distributions with respect to $z=0$. In fact, a principal component analysis suggests that the only parameter which is truly well constrained by the $z=0.83$ stellar mass function is $\alpha_{\text{SF}}$. Its value is approximately 2.5 times higher than was the case at $z=0$ which is driven by the need to form high mass galaxies more rapidly in order to achieve a better match to the massive end of the observed stellar mass function. However, a side effect is the further build up of galaxies at intermediate masses ($\log_{10}(M[h^2/M_\odot])=10.0-10.5$) which must then be alleviated by increasing the strength of the supernova cold gas ejection efficiency, $\epsilon_{\text{halo}}$.

From Eqn. 2.7, we see that the amount of gas ejected from the dark matter halo entirely by supernova feedback equals zero for $V_{\text{vir}}=V_{\text{vir}}^{\text{cutoff}}=V_{\text{SN}}(\epsilon_{\text{halo}}/\epsilon_{\text{disk}})^{1/2}$. Hence by increasing $\epsilon_{\text{halo}}$ (the halo hot gas ejection efficiency), we increase the characteristic halo mass at which supernova feedback becomes ineffective at ejecting gas from the system. The net result is a reduction of star formation in more massive galaxies (which preferentially populate dark matter halos with higher masses, and hence higher virial velocities) due to a reduced availability of hot gas which can then cool to fuel star formation.

Given that $V_{\text{vir}}^{\text{cutoff}}$ depends on the ratio of the supernova ejection and reheating parameters ($\epsilon_{\text{halo}}$ and $\epsilon_{\text{disk}}$), why does the $z=0.83$ stellar mass function preferentially modify $\epsilon_{\text{halo}}$ from its $z=0$ fiducial value instead of $\epsilon_{\text{disk}}$? Again from Eqn. 2.7, we see that a change in $\epsilon_{\text{disk}}$ results in a proportional change to the ejected mass. However, this is independent of the host halo properties. On the other hand, modifying $\epsilon_{\text{halo}}$ results in a change to the ejected mass with a magnitude that is inversely proportional to $V_{\text{vir}}^2$. Hence increasing $\epsilon_{\text{halo}}$ results in both an increase in the value of $V_{\text{vir}}^{\text{cutoff}}$, as well as preventing the build-up of excess star forming galaxies just above this halo velocity where radio mode black hole feedback is still inefficient.

### 2.5.3 Combined redshifts

Having presented the results of constraining the model to match observations at $z=0$ and 0.83 individually, we now investigate if it is possible to achieve a satisfactory result at both redshifts simultaneously. As shown in Fig. 2.8, there is some tension between the marginalised posterior distributions at each redshift. However, it is possible that a parameter configuration may exist which, although not achieving the best possible reproduction at either redshift, will still provide a satisfactory combined result.

Fig. 2.9 shows the 1 and 2-D posterior distributions for the model when constrained
2.5. Analysis

Figure 2.9: 2-D posterior probability distributions (off-diagonal panels) for all combinations of the log of the six free model parameters when constraining against the $z=0$ and $z=0.83$ stellar mass functions, and the $z=0$ black hole–bulge relation, all simultaneously. Black lines indicate the 1 and 2-$\sigma$ confidence contours. The limits of each panel indicate our prior ranges. Orange circles show the location of the marginalised best values of each parameter, while red diamonds indicate the values from the single parameter set which produced the best reproduction of the observations (maximum likelihood parameters). The histograms in the diagonal panels represent the 1-D marginalised probability distributions, with the 1 and 2-$\sigma$ confidence intervals indicated by the dark and light shaded regions respectively. The 1-D maximum likelihood distributions are also shown for comparison (red dashed lines). This figure can be directly compared with Figs. 2.2 and 2.5.

simultaneously against the $z=0$ stellar mass function and black hole–bulge relation, as well as the $z=0.83$ stellar mass function. The 1-D histograms indicate that this constraint combination places strong restrictions on all of the free model parameters, including the ejected gas reincorporation rate parameter, $\gamma_{ej}$. For all previously investigated constraint combinations, the value of $\gamma_{ej}$ has made little impact on the ability of the model to reproduce the relevant observations. This has been true for values spanning several orders of
magnitude. However, when constraining the model to reproduce two redshifts simultaneously, gas reincorporation rate plays a key role.

As discussed in §2.5.2, the individual redshift constraints provide strong, but irreconcilable, restrictions on the value of the star formation efficiency, $\alpha_{\text{SF}}$. When these constraints are combined, the model is therefore forced to pick one of the preferred $\alpha_{\text{SF}}$ values and use the freedom in the other parameters, including $\gamma_{\text{ej}}$, to maximise the joint likelihood.

The main effect of altering the ejected gas reincorporation efficiency, $\gamma_{\text{ej}}$, occurs at the low mass end of the stellar mass function. It’s only here that supernova feedback is efficient at expelling gas and galaxies have a significant amount of material in their ejected reservoirs. In addition, in our model the ejected mass reservoirs of in-falling satellite galaxies are immediately incorporated into the hot halo components of their more massive parents and hence no extra material is added to the ejected component of these larger systems. By increasing the value of $\gamma_{\text{ej}}$, the timescale over which expelled gas makes its way back into the heating/cooling cycle is decreased. This results in more cold gas being available for forming stars in the lowest mass galaxies, with the net effect being a raising of the low mass end of the stellar mass function.

As shown in Fig. 2.7, the fiducial $z=0$ parameter set already produces a stellar mass function at $z=0.83$ which is overabundant in low mass galaxies. In order for $\gamma_{\text{ej}}$ to help to alleviate this, its value needs to be reduced, thus reincorporating less ejected gas into lower mass systems. Unfortunately however, the marginalised best value of $\gamma_{\text{ej}}$ using the $z=0$ constraints is already extremely low ($1.7 \times 10^{-3}$) and reducing it even further has a negligible effect. The model is therefore unable to utilise this parameter to maximise the joint-redshift likelihood when using the fiducial $z=0$ star formation efficiency. Fortunately however, $\gamma_{\text{ej}}$ can be used to maximise the likelihood achieved with the $z=0.83$ preferred parameters.

Compared to the $z=0$ case, the $z=0.83$ marginalised best parameters have a higher star formation efficiency, $\alpha_{\text{SF}}$, resulting in a more rapid transition of galaxies from low to high masses. As shown in Fig. 2.3, the effect on the stellar mass function at $z=0$ is an over-abundance at high stellar masses and a corresponding under-abundance at low masses. Increasing the low value of the gas reincorporation rate parameter ($\gamma_{\text{ej}}$) can have a significant effect here by increasing the number of galaxies below the knee of the mass function. By also increasing the values of the supernova feedback gas reheating and ejection parameters ($\epsilon_{\text{disk}}$ and $\epsilon_{\text{halo}}$) the model can achieve the correct overall shape whilst moving the position of the knee by only a small amount. This allows a better reproduction of the $z=0$ mass function to be achieved.
2.5. Analysis

Figure 2.10: The $z=0$ (left) and $z=0.83$ (right) model stellar mass functions obtained using the highest likelihood model parameters when constraining against the $z=0$ and $z=0.83$ stellar mass functions, and the $z=0$ black hole–bulge relation, simultaneously. The solid line with shaded region shows the model result along with the 95% confidence region calculated from the posterior distributions. Blue error bars show the relevant constraining observations.

The marginalised best values and their uncertainties are presented in Table 2.1. Fig. 2.10 displays the resulting $z=0$ and $0.83$ stellar mass functions. Even though the star formation efficiency parameter is close to the preferred $z=0.83$ value, the changes made to the other parameters have resulted in a visually poorer reproduction of $z=0.83$ mass function. However, a reduced $\chi^2$ of 1.28 (with 14 degrees of freedom) indicates that the fit of the highest likelihood line is still statistically reasonable\(^2\). In order to achieve this level of agreement at both redshifts simultaneously, we note that we have been required to push the parameters associated with supernova feedback and reincorporation ($\epsilon_{\text{disk}}$, $\epsilon_{\text{halo}}$ and $\gamma_{ej}$) to values that are perhaps unrealistic. We discuss the interpretation and possible physical implications of this outcome in the next section.

Finally we note that we have increased the lower limit on the reincorporation efficiency ($\gamma_{ej}$) prior to 0.1 when applying our joint redshift constraints. In testing, we allowed this parameter to go as low as $10^{-3}$, however, we found that this introduced a large peak in the marginalised posterior distributions, corresponding to the alternative, but lower probability, $z=0$ preferred star formation efficiency ($\alpha_{\text{SF}}$). As discussed above, in order to maximise the likelihood achieved using this $\alpha_{\text{SF}}$ value, the reincorporation efficiency must be lowered as much as possible. However, all values of $\gamma_{ej}\lesssim 0.1$ produce approximately identical joint likelihoods as any sufficiently small value results in almost no mass of

\(^2\)A reduced $\chi^2$ of 1.28 with 14 degrees of freedom corresponds to a P-value of 0.21. This indicates that, given the model and observational uncertainties, and assuming the null hypothesis that the model does indeed replicate the observations, there is a 21% chance of randomly achieving this $\chi^2$ or worse.
ejected material being reincorporated into low mass halos. When these low mass halos grow sufficiently, they will eventually reincorporate the material. However, they will also have set up a hydro-static hot halo, meaning the effect of recapturing this mass on the central galaxy will be minimal.

When we allowed the reincorporation efficiency to go to lower values the MCMC chain spent a large number of successful proposals mapping out the extensive volume provided by this flat low likelihood feature. Each of these propositions contributed to a marginalised peak coinciding with the lower likelihood $z=0$ preferred star formation efficiency. This feature of the posterior distribution highlights the importance of selecting suitable priors which encompass the full range of physically plausible parameter values, but which do not include large areas of parameter space which are indistinct from each other due to the details of the model.

If we were to have allowed values of $\gamma_{ej} \lesssim 0.1$ in our presented analysis, we would have unfairly biased our posterior distributions towards a region of lower likelihood. It is important to note however, that given a suitably long chain, the MCMC would eventually still have converged to provide us with the same posterior distribution peaks as presented in Fig. 2.9. Our poor choice of lower prior for $\gamma_{ej}$ would simply have meant that such a convergence would have required an unfeasibly long chain.

2.6 Discussion

2.6.1 Interpreting the joint redshift constraint results

The fiducial parameter values for our joint redshift constraints (c.f. table 2.1) provide us with a valuable insight into exactly where the tensions lie within the model when trying to successfully reproduce the late time growth of stellar mass in the Universe. The parameters associated with supernova feedback have been pushed to their limits, and possibly to unrealistic values. Using our MCMC analysis of the model when constrained against each redshift individually allows us to understand the cause of this as follows:

- The values of all of the parameters are driven by the need to put the high mass end of the stellar mass function in place as early as possible. This is illustrated in Fig. 2.7 where we see that the high redshift stellar mass function produced by the $z=0$ fiducial parameter values (dashed line) underpredicts the number density of high mass galaxies. In order to alleviate this discrepancy and provide the best possible match to the observations at $z=0.83$ alone, a relatively high star formation efficiency is required (see Fig. 2.8).
• Unfortunately, as shown in Fig. 2.3, this high star formation efficiency causes an under-prediction of the number of low mass galaxies at z=0, due to their rapid growth. In order to counteract this and to provide the best possible result at both redshifts simultaneously, the preferred re-incorporation rate of ejected material (γej) must be increased. This efficiently increases the number density of galaxies with stellar masses below the knee of the mass function whilst leaving the high mass distribution unchanged. Although not implausible, it is worth noting that such a high value of γej requires the presence of some mechanism to rapidly return ejected material into the heating/cooling cycle over timescales close to, or less than, the dynamical time of the host dark matter halo.

• However, the rise in the number of very low stellar mass galaxies that results from such a high gas re-incorporation efficiency is extremely large and leads to an over-estimation of their number density at both redshifts. To compensate, the preferred supernova halo gas ejection efficiency (ϵhalo) is forced to values greater than one, implying that either the mean kinetic energy of supernova explosions per unit mass is too low, or that some other physical mechanism exists to enhance the deposition of this energy into the ISM.

• Increasing the value of the supernova ejection efficiency to such high values allows for the efficient regulation of star formation in larger and larger systems, thus pushing the knee of the z=0 stellar mass function to higher masses. In order to return the knee to its correct position the model is forced to equivalently increase the supernova feedback mass loading factor (ϵdisk) to ≈14. Unfortunately, such a high mass loading factor is difficult to reconcile with current observational estimates (e.g. Martin, 1999; Rupke et al., 2002; Martin, 2006) and may suggest the need for an additional halo mass dependence for this parameter (e.g. Oppenheimer & Davé, 2006; Hopkins et al., 2012).

The high values preferred by these parameters suggests that the model prescriptions for supernova feedback, and possibly gas re-incorporation, are insufficient. Similar assertions have been made by other works, most commonly based upon calibrating semi-analytic models to reproduce the z=0 stellar mass or luminosity functions and then investigating the z>0 predictions (e.g. Guo et al., 2011; Lu et al., 2012). However, we confirm this finding for the first time through explicitly attempting to match the observed stellar mass functions at two redshifts simultaneously. This allows us to exclude the possibility that a physically acceptable parameter combination exists within the framework of our current
physical prescriptions.

2.6.2 Implications

In Fig. 2.10 we show the highest likelihood stellar mass functions produced by the model when simultaneously constrained against the observed $z=0$ and 0.83 stellar mass functions and the $z=0$ black hole–bulge relation. Although we do manage to achieve statistically reasonable fits to the stellar mass function at both epochs, there are some clear tensions at high redshift. We now discuss the possible implications for our semi-analytic model as well as for the growth of stellar mass in the Universe.

Despite large observational uncertainties associated with measuring the number density of the most massive galaxies at $z \gg 0$, the phenomenon of galaxy ‘downsizing’, whereby the most massive galaxies in the Universe are in place at early times, is well established (Heavens et al., 2004; Neistein et al., 2006). Our results extend those of previous works in suggesting that current galaxy formation models built upon the hierarchical growth of structure find it difficult to reproduce the quantitative details of this phenomenon (e.g. De Lucia et al., 2006; Kitzbichler & White, 2007; Guo et al., 2011; Zehavi et al., 2012). In particular, a comparison of the stellar mass functions produced when constraining against each redshift individually demonstrates that the model struggles to successfully put the highest mass galaxies in place early on (black dashed line of Fig. 2.7) without also under-predicting the number of low mass galaxies at $z=0$ and equivalently over-predicting the number density of the most massive galaxies (black dashed line of Fig. 2.3).

It is possible that the model’s under-prediction at high masses may be partially alleviated by convolving the $z=0.83$ model mass function with a normal distribution of a suitable width in order to account for systematic uncertainties in the observed stellar masses (Kitzbichler & White, 2007; Guo et al., 2011). We do not carry out such a procedure here, as at least some fraction of this uncertainty should be included in our constraining observations and we do not wish to add a further add-hoc correction without justification. However, it may prove impossible for the model to self-consistently reproduce the observed stellar mass function at multiple redshifts without including these additional observational uncertainties (Moster et al., 2012).

In our model, star formation, supernova feedback and gas re-incorporation are assumed to proceed with a constant efficiency as a function of halo mass across the full age of the Universe. However, our findings could be interpreted as suggesting the need to incorporate an explicit time dependence to these efficiencies; in particular to provide a preferential increase to the rate of stellar mass growth in massive halos in the early Universe. This
would help establish the high mass end of the stellar mass function early on without over-producing the number of lower mass galaxies at late times.

By extending our Bayesian analysis to include model selection, we could formally test if this addition to the physics is statistically warranted. Unfortunately though, adding further layers of parametrisation to current processes, such as an explicit time dependence, makes the interpretation of model results increasingly difficult. This is especially so when attempting to uncover the relative importance of physical processes that shape the evolution of different galaxy populations. To combat this, it is important to ensure that all new additions to a semi-analytic model have a strong physical motivation, even if they do improve the results.

Having said that, modifications to the rate of growth of stellar mass in the early Universe is supported by other studies. In particular, it has been suggested that the star formation efficiency of galaxies must peak at earlier times for more massive galaxies (e.g. Moster et al., 2012). This could possibly represent a number of physical processes such as a metallicity dependent star formation efficiency (Krumholz & Dekel, 2012) or a rapid phase of stellar mass growth due to high redshift cold flows (Dekel et al., 2009b). Alternatively, the apparent need to incorporate an explicit time dependence to the star formation and feedback efficiencies could signify an overestimation of the merger timescales in the model at early times (Weinmann et al., 2011). Also, we note that star formation proceeds in our model following a simple gas surface density threshold argument. It may be the case that we are over predicting the size of the most massive galaxies at early redshift and therefore under-predicting the level of star formation in these objects. We do not specifically track the build up of angular momentum in our simulated galaxies, instead relying on the spin of the parent dark matter halo as being a good proxy. It is unlikely that this assumption is valid at high redshift (e.g. Dutton & van den Bosch, 2009; Kimm et al., 2011; Brook et al., 2011) and, even if it were, the low time resolution of our input simulation coupled with the low number of particles in halos at these times might mean that the halo spin values are systematically unreliable here.

Rather than suggesting the need for a time dependent star formation efficiency, an additional interpretation of our findings could be the need to include an intra-cluster light component (ICL) in the model (Gallagher & Ostriker, 1972; Purcell et al., 2007; Guo et al., 2011). Sub-halo abundance matching studies have suggested that mergers involving massive galaxies may result in significant fractions ($\gtrsim 80\%$) of the in-falling satellite mass being deposited in to this diffuse component rather than ending up in the central galaxy (Conroy et al., 2007; Behroozi et al., 2012). Since the majority of late time growth of
the most massive galaxies is heavily dominated by mergers, the inclusion of such a stellar mass reservoir may allow the effective suppression of the growth of these massive objects, reducing the need to push the supernova feedback parameters to such extreme values.

Finally, the majority of modern galaxy formation models are now able to reproduce many of the most important observational quantities of the \( z=0 \) Universe. Any changes to the underlying cosmology or input merger tree construction can typically be accounted for by varying the free model parameters. However, achieving the correct time evolution of the full galaxy population is a more difficult task and makes us far more dependant on the details of dark matter structure growth. If this growth does not correctly match the real Universe then this is something that the models will try to counteract, possibly leading us to conclusions about the baryonic physics which could be incorrect. To fully assess the level to which missing or poorly understood baryonic physics are responsible for discrepancies from observed galaxy evolution, a detailed analysis of the effects of cosmology, dark matter halo finding, and merger tree construction on the output of a single galaxy formation model is needed.

2.7 Conclusions

In this chapter we investigate the ability of the Croton et al. (2006) semi-analytic galaxy formation model to reproduce the late time evolution of the growth of galaxies from \( z \approx 0.8 \) to the present day. In particular we focus on matching the \( z=0 \) and \( z=0.83 \) stellar mass functions as well as the \( z=0 \) black hole–bulge relation. To achieve this we utilise Monte-Carlo Markov Chain techniques, allowing us to both statistically calibrate the model against the relevant observations and to investigate the degeneracies and tensions between different free parameters.

Our main results can be summarised as follows:

- The Croton et al. (2006) model is able to provide a good agreement with the stellar mass function and black hole–bulge relation at \( z=0 \) (§2.5.1; Figs. 2.2, 2.3). However, when attempting to match both simultaneously there are some minor tensions between the favoured parameter values (§2.5.1; Figs. 2.4, 2.6).

- The model is also able to independently provide a good agreement with the observed stellar mass function at \( z=0.83 \). In order to achieve this, a higher star formation efficiency is necessary than was preferred to match the \( z=0 \) constraints. This is to ensure that the massive end of the stellar mass function is entirely in place by this early time, as required by our constraining observations (§2.5.2; Figs. 2.7, 2.8).
2.7. Conclusions

• When attempting to reproduce the observations at both redshifts simultaneously, a number of tensions in the model’s physical prescriptions are highlighted. In particular, the struggle to reconcile the high star formation efficiency required to reproduce the high mass end of the stellar mass function at \( z=0.83 \) with the observed increase in normalisation of the low mass end at \( z=0 \) (§2.5.3; Fig. 2.10).

• Our attempts to model the evolution of galaxies at \( z\lesssim0.8 \) suggest that the supernova feedback prescriptions of the model may be incomplete, possibly requiring the addition of extra processes that preferentially enhance star formation in the most massive galaxies at \( z>1 \) (§2.6).

This is the first time that a full semi-analytic model, based on the input of N-body dark matter merger trees, has been statistically calibrated to try and reproduce a small focussed set of observations at multiple redshifts simultaneously. Only by carrying out this procedure, and fully exploring the available parameter space of our particular model, are we able to conclusively demonstrate that the model struggles to match the late time growth of the galaxy stellar mass function. Our analysis also provides us with important insights as to what changes may be necessary to alleviate these tensions. Having said that, despite requiring somewhat unlikely parameter values, we do achieve a statistically reasonable fit to the observations at both redshifts simultaneously. For the purpose of producing mock galaxy catalogues at \( z\lesssim0.8 \), this best fit model is generally adequate.

For this work, we have only considered the stellar mass function and black hole–bulge relation. However, it is likely that the addition of extra observational constraints will further help to isolate the parts of the model which require particular attention. For example, the gas mass–metallicity relation (Tremonti et al., 2004) is particularly sensitive to the re-incorporation efficiency due to its ability to regulate the dilution of a galaxy’s gas component with low metallicity material ejected at early times. In future work we will extend our analysis to redshifts greater than one and investigate the constraints provided by other quantities such as the mass–metallicity relation as well as the baryonic Tully-Fisher relation, galaxy colour distribution and stellar mass density evolution.

Finally, we also stress that our results and conclusions are sensitive to the magnitude of the uncertainties associated with our observational constraints. Although we have endeavoured to ensure their accuracy, it is likely that they may still be underestimated. For example, the high mass end of \( z\gg0 \) stellar mass functions are heavily susceptible to cosmic variance effects due to the deep observations required to simultaneously resolve galaxies at lower mass scales (typically done in smaller fields). Some works have also suggested systematic uncertainties of 0.3 dex or more when estimating stellar masses via
broad-band photometry (Conroy et al., 2009). Furthermore, the magnitude of many of these uncertainties increases significantly with redshift, and can result in errors of up to 0.8 dex around the knee of the measured stellar mass functions at $z = 1.3 - 2$ (Marchesini et al., 2009). More detailed comparisons between state-of-the-art semi-analytic models and high redshift observations will therefore require us to greatly improve our measurements at these redshifts.
The mid-life crisis of the Milky Way and M31

Upcoming next generation galactic surveys, such as Gaia and HERMES, will deliver unprecedented detail about the structure and make-up of our Galaxy, the Milky Way, and promise to radically improve our understanding of it. However, to benefit our broader knowledge of galaxy formation and evolution we first need to quantify how typical the Galaxy is with respect to other galaxies of its type. Through modelling and comparison with a large sample of galaxies drawn from the Sloan Digital Sky Survey and Galaxy Zoo, we provide tentative yet tantalizing evidence to show that both the Milky Way and nearby M31 are undergoing a critical transformation of their global properties. Both appear to possess attributes that are consistent with galaxies midway between the distinct blue and red bimodal colour populations. In extragalactic surveys, such ‘green valley’ galaxies are transition objects whose star formation typically will have all but extinguished in less than 5 Gyr. This finding reveals the possible future of our own galactic home and opens a new window of opportunity to study such galactic transformations up close.

Chapter 3. The mid-life crisis of the Milky Way and M31

3.1 Introduction

It is a natural extension of the Copernican principle that our Galaxy should be typical when compared to other galaxies of its type. As such, the Milky Way is commonly employed as the template for an archetypal spiral galaxy, and is frequently used to interpret both observations and simulation data. However, despite being the most closely studied galaxy in the universe, some of the Milky Way’s most basic global properties, such as its colour and star formation rate, remain relatively poorly constrained. This is due, in large part, to our location within the disk and the difficulty in taking such measurements, especially at optical wavelengths where obscuration by dust can be particularly problematic.

Next generation galactic survey instruments, such as Gaia (Lindegren et al., 2008) and HERMES (Wylie-de Boer & Freeman, 2010), will greatly improve our knowledge of the Milky Way and its global properties. The resulting benefit to our broader knowledge of galaxy formation will, however, require us to know if the Milky Way really is a typical $L^*$ blue spiral galaxy, and if not, in what way(s) it differs?

Important insights into this question have already been made. For example, it has been suggested that the probability of the Milky Way possessing two satellites of similar luminosity to the Magellanic Clouds is small ($\sim$3.5%; Liu et al., 2011). It also appears that the Milky Way is under-luminous by approximately 1σ with respect to the main locus of the Tully–Fisher relation (Flynn et al., 2006). This finding is confirmed by Hammer et al. (2007), who further note that there are other Galactic properties with significant deficiencies, such as its stellar mass and disk radius. From their results they estimate that only $\sim 7\%$ of spirals in the local universe are similar to the Milky Way with respect to $K$-band magnitude, circular velocity, and disk scale length. Interestingly, based on the same properties they find that M31 is a much better template for a typical spiral. Using low chemical abundances in the Milky Way’s outskirts as evidence, it is suggested that a likely cause for the Milky Way’s disparity is an extremely quiet merger history with no accretion of objects more massive than $\sim 10^9 M_\odot$ in the last 10 Gyr. This contrasts with the majority of other spirals, such as M31, which have had more continuous episodes of mass accretion.

Further support for an uncharacteristically quiet accretion history of the Milky Way comes from the age of the Galaxy’s thin disk ($\simeq 10$ Gyr). Stewart et al. (2008) use simulations to show that around 95% of present day Milky Way size dark matter halos have undergone at least one merger with an object more massive than the current disk in the last 10 Gyr. Presumably such a merger would have, at a minimum, efficiently heated the thin disk, which is inconsistent with current observations.
Despite these exceptional properties there is also evidence to suggest that in many other respects the Milky Way is unremarkably normal. Using a simulated sample of Milky Way type galaxies, de Rossi et al. (2009) find that the Galaxy is standard in terms of its gas fraction, luminosity weighted stellar age, and stellar mass. They did, however, find a large dispersion in these properties, stemming from a wide range of different assembly histories. In this sense, it is important to recognize that the characterization of ‘typical’ for any galaxy may primarily depend on the property of interest.

One global property of the Milky Way which has largely escaped scrutiny is its colour index. A galaxy’s colour index provides an important insight into its current evolutionary state and is closely linked to its past star formation history. In this chapter we compare the $B-V$ colour index of the Milky Way and M31 with galaxies drawn from the Sloan Digital Sky Survey Galaxy Zoo morphological database (Lintott et al., 2008). We then confirm and interpret these results using a theoretical Milky Way galaxy catalog constructed from a semi-analytical galaxy formation model, for which we know the full star formation and colour histories of each galaxy. Using colour as a probe of the evolutionary state of the Milky Way and M31, we find evidence that both galaxies may be undergoing a transition on to the red sequence.

This chapter is laid out as follows. In Section 3.2, we introduce our observational data and discuss the relevant properties of the Milky Way and M31. In Section 3.3, we describe the theoretical model we use to interpret our findings. Our main results are presented in Section 3.4 — in particular we show that both the Milky Way and M31 would probably be classified as transition galaxies if observed from the vantage point of an extragalactic survey. We discuss our results and highlight some of the possible physical mechanisms that may explain them in Section 3.5. Finally, a summary of our findings and conclusions are presented in Section 3.6.

Where relevant, we adopt the following ΛCDM cosmology in this chapter: $\Omega_m = 0.25$, $\Omega_\Lambda = 0.75$, $\Omega_b = 0.045$. All results are quoted with a Hubble constant of $h = 0.7$ (where $h \equiv H_0/100 \text{ km s}^{-1} \text{Mpc}^{-1}$) unless otherwise stated. The Vega magnitude system and standard optical BVRI filters are employed throughout, with the exception of Sloan Digital Sky Survey (SDSS) $u-r$ colours, which are given in the AB magnitude system.

### 3.2 Observational data

In Table 3.1, we present a summary of the three key properties of the Milky Way and M31 to be studied in this chapter: $B-V$ integrated colour, stellar mass, and star formation rate. Central to our work in this chapter is a comparison with a larger, cosmologically
Chapter 3. The mid-life crisis of the Milky Way and M31

### Table 3.1: Literature values for the relevant global properties of the Milky Way and M31.

<table>
<thead>
<tr>
<th>Galaxy</th>
<th>Property</th>
<th>Value</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milky Way</td>
<td>$B-V$</td>
<td>$0.83 \pm 0.15$</td>
<td>van der Kruit (1986)</td>
</tr>
<tr>
<td></td>
<td>$M_\star$</td>
<td>$(5.18 \pm 0.5) \times 10^{10} M_\odot$</td>
<td>Flynn et al. (2006)</td>
</tr>
<tr>
<td></td>
<td>SFR</td>
<td>$0.68 - 1.45 , M_\odot , yr^{-1}$</td>
<td>Robitaille &amp; Whitney (2010)</td>
</tr>
<tr>
<td>M31</td>
<td>$B-V$</td>
<td>$0.86 \pm 0.04^a$</td>
<td>de Vaucouleurs et al. (1991)</td>
</tr>
<tr>
<td></td>
<td>$M_\star$</td>
<td>$(10.4 \pm 0.5) \times 10^{10} M_\odot$</td>
<td>Geehan et al. (2006)</td>
</tr>
<tr>
<td></td>
<td>SFR</td>
<td>$0.41 - 0.59 , M_\odot , yr^{-1}$</td>
<td>Kang et al. (2009)</td>
</tr>
</tbody>
</table>

*a*Includes foreground extinction correction (Schlegel et al., 1998).

selected local ‘control’ population, drawn from the SDSS and Galaxy Zoo data sets. To enable this we have scoured the literature to find the most current values for the Milky Way and M31.

#### 3.2.1 The Milky Way

There is much confusion and a great number of disparate results in the literature for many of the Milky Way’s properties. Much of this is due to the inherent difficulties associated with observationally determining the properties of a spiral galaxy which is not seen face on and is extended on the sky. This problem is compounded for the Milky Way due to our position within the Galaxy itself.

Take, for example, the commonly quoted but little discussed $B-V$ colour value of $B-V = 0.83 \pm 0.15$ (van der Kruit, 1986). It was derived by comparing the Galactic background light, as observed by the Pioneer 10 probe (launched in 1972), to models of the Galaxy’s stellar distribution. It is currently the most commonly cited estimate for this property, despite its age, and is found in a number of more recent publications (Yin et al., 2009; Flynn et al., 2006). Although there are a small number of other estimates in the literature, they are often conflicting and many are not sourced from refereed papers. It is out with the scope of this chapter to discuss all of these values in detail, however, we note that, taking into account the large systematic errors associated with making such a measurement, they are all in approximate statistical agreement.

There are a similarly small number of estimates for the total stellar mass of the Milky Way. The value of $M_\star = (5.18 \pm 0.33) \times 10^{10} M_\odot$ comes from extrapolating local stellar surface density measurements to the full disk and adding a bulge mass estimated using
3.2. Observational data

dynamical constraints (Flynn et al., 2006). The quoted uncertainty results from considering a realistic range of assumed disk scale lengths. The true uncertainty on this value is likely considerably larger, however, with no real understanding of the underlying systematics we choose to simply round the quoted uncertainty up to \( \pm 0.5 \times 10^{10} M_\odot \) for this work. Similar stellar mass values for the Milky Way are obtained by converting \( K \)-band measurements to a stellar mass value via colour dependent \( M_\odot/L_K \) ratios (Hammer et al., 2007). It should be noted, however, that a slightly bluer colour value for the Galaxy of \( B-V=0.79 \) has been utilized during the derivation of this result (note that this is still in good statistical agreement with our fiducial value of \( B-V = 0.83 \pm 0.15 \)).

The most up-to-date determination of the Milky Way’s current star formation rate is \( \dot{M}_* = 0.68 - 1.45 M_\odot\text{yr}^{-1} \) (Robitaille & Whitney, 2010). This is calculated by comparing the number of Spitzer/IRAC GLIMPSE observed pre-main-sequence young stellar objects to population synthesis models. The derived value is slightly lower than, but still broadly consistent with, previous estimates (Rana, 1991; Boissier & Prantzos, 1999; Fraternali, 2009, e.g.,). However, Robitaille & Whitney (2010) argue that this difference is largely due to the assumed initial mass function (IMF), with their choice being more realistic, especially at low stellar masses. This value is also directly derived from actual statistics of pre-main-sequence objects rather than simply from global observables.

Finally, we assume the Hubble type of the Milky Way to be approximately Sh/c (Kennicutt, 2001; Hodge, 1983).

3.2.2 M31

Due to the benefit of its large angular size on the sky, the majority of authors spatially decompose the properties of M31 rather than use its integrated properties. The best global \( B-V \) colour estimate of \( 0.92 \pm 0.02 \) comes from the Third Reference Catalogue of Bright Galaxies (RC3; de Vaucouleurs et al., 1991). This value has not been corrected for either internal or Galactic extinction and so we apply a foreground extinction correction of \( E(B-V)=0.062 \) (Schlegel et al., 1998)\(^1\) to arrive at the final colour listed in Table 3.1: \( B-V = 0.86 \pm 0.04 \). We note that the RC3 catalog value is in excellent agreement with a number of independent works (de Vaucouleurs, 1958; Walterbos & Kennicutt, 1987) and hence we are confident in its fidelity.

Our utilized measurement for the global stellar mass of M31 was determined through the summation of the disk and bulge component from a semi-analytic mass model (Geehan

\(^1\)Due to the unavailability of a resolved temperature structure map for M31, Schlegel et al. (1998) estimated the internal dust extinction towards this object using the median value calculated from a surrounding annulus (cf. Appendix C of Schlegel et al., 1998).
et al., 2006). The resulting value of $M_\star = 10.4 \times 10^{10} M_\odot$ is consistent with the findings of a number of other recent works (e.g., Hammer et al., 2007; Barnby et al., 2006). Considering the range of values quoted in the literature, we choose to estimate an uncertainty of $\pm 0.5 \times 10^{10} M_\odot$.

The star formation rate of M31 is much better established than that of the Milky Way. The range of allowed values is $\dot{M}_* = 0.41 - 0.59 M_\odot \text{yr}^{-1}$ and is derived from combined UV and IR fluxes in star formation regions with an age of less than approximately 0.4 Myr (Kang et al., 2009). However, we note that this range of star formation rate values is identical, to within $\sim 0.1 M_\odot \text{yr}^{-1}$, to the value when averaged over the last 400 Myr. The mid point quoted in Table 3.1 is for an assumed metallicity of $Z = 0.02$, with the upper and lower boundaries corresponding to $Z = 0.05$ and $Z = 0.008$, respectively.

Following Tully & Pierce (2000), we take the Hubble type of M31 to be approximately Sb.

### 3.2.3 Galaxy Zoo control sample

To fairly interpret the properties of the Milky Way and M31 we employ a large control sample of $z \approx 0$ galaxies from Bamford et al. (2009); a luminosity-limited catalogue of $\sim 130,000$ galaxies taken from SDSS (DR6; Adelman-McCarthy et al., 2008) data and cross correlated with Galaxy Zoo (data release 1; Lintott et al., 2011) visual morphologies. The sample has been selected to contain galaxies with spectroscopic redshifts in the range $0.01 < z < 0.085$, corresponding to an $r$-band absolute Petrosian magnitude limit of $M_r < -20.17$ and is carefully sub-sampled in an effort to remove any redshift dependent selection biases. All magnitudes are in the AB system and are $k$-corrected to $z = 0$.

Stellar masses are derived using the method of fitted colour dependent mass-to-light ratios discussed in Baldry et al. (2006). We refer the reader to Bamford et al. (2009) for further details. We henceforth refer to this sample as the full Galaxy Zoo sample.

In order to produce a sample of Milky Way/M31 analog galaxies we select only galaxies which:

- are visually classified spirals with de-biased morphological likelihoods $p_{\text{spiral}} \geq 0.8$ (c.f. Bamford et al., 2009),

- have stellar masses in the range $10.66 < \log_{10}(M_\star [M_\odot]) < 11.12$ (c.f. $\log_{10}(M_\star_{\text{MW}} / M_\odot) = 10.71$; $\log_{10}(M_\star_{\text{M31}} / M_\odot) = 11.02$),

- possess an SDSS $\text{fracdev}$ value in the range 0.1–0.5 (i.e., have an approximate Sb/c morphology; c.f. Masters et al., 2010b),
• and are face-on with $\log_{10}(a/b) < 0.2$, where $a/b$ is the major to minor axis ratio as observed on the sky.

The $\text{fracdev}$ parameter measures the relative fraction of the best fitting light profile that comes from a de Vaucouleurs fit. Following Masters et al. (2010b), we use this value to approximate a galaxy’s bulge-total mass ratio and hence its morphology. This selection will, in reality, encompass a range of morphological types from $\sim$Sb-Sc. We place no direct constraints on the visibility of spiral arms in our sample. By selecting only face-on galaxies, we minimize, as much as possible, the systematic reddening effects of intrinsic interstellar absorption which could cause otherwise blue Sb/c galaxies to masquerade as green valley or red sequence members. We also note that, although we have chosen to utilize a single comparison set for both the Milky Way and M31, we have confirmed that splitting our selection criteria into two equal stellar mass ranges makes no qualitative difference to our findings.

Our resulting Galaxy Zoo Milky Way/M31 analog sample contains 997 objects.

### 3.3 Modeled data

The observational data presented in §3.2 provides a reasonably precise snapshot of the current state of the Milky Way and M31, as well as a census of the broader local galaxy population. However, it does not give us any insight into the range of evolutionary histories that may have led to this final state, or of the physical influences that were pertinent. To obtain this we employ a model of galaxy evolution that explicitly tracks the full history of $L^*$ galaxy growth and the physics that shapes it.

#### 3.3.1 A ‘Semi-analytic’ Model of Galaxy Formation

The semi-analytic method of modeling galaxy formation was first introduced by White & Frenk (1991) and follows a two-staged approach:

• First, a numerical simulation of the growth of the dark matter structure of the Universe is run to establish the potential formation sites and gravitational evolution of galaxies, galaxy groups and clusters. For our work we use the Millennium Simulation (Springel et al., 2005), a cosmological simulation that evolves $\sim$10 billion dark matter particles in a $1.24 \times 10^8 \text{ Mpc}^3 h^{-3}$ representative volume. The Millennium Simulation accurately tracks the hierarchical growth of structure from two orders-of-magnitude smaller than a Milky Way system (i.e. dwarfs), to three orders-of-magnitude larger (i.e. clusters).
Second, analytic prescriptions that describe the physics of galaxy formation are coupled to the dark matter simulation to predict the evolution of baryons in the evolving dark matter structure. For this chapter we use the commonly utilized model of Croton et al. (2006). This model is calibrated to produce a good global match to the local galaxy population in terms of its stellar mass and luminosity functions, galaxy colours, morphologies, and clustering (De Lucia & Blaizot, 2007; Croton et al., 2007; Kitzbichler & White, 2008). To achieve this the model employs prescriptions that describe gas accretion and cooling in dark matter halos, galaxy disk formation, star formation and dust, supernova feedback, the production and evolution of metals, galaxy-galaxy mergers and morphological transformations, black hole growth and active galactic nuclei feedback.

See §2.3, as well as Croton et al. (2006), for a more detailed description of each component of the model and its construction.

### 3.3.2 Selecting the Theoretical Milky Way Analogue Sample

Once the model is run we have on hand a large number of simulated galaxies at $z = 0$ ($\sim 25$ million) with well defined properties (masses, luminosities, colours, star formation rates, ...) and complete evolutionary histories. From these galaxies we sub-select Milky Way/M31 analogues to be used to interpret the observed state of the Milky Way and M31.

Following a set of criteria which are as similar as possible to that of the Galaxy Zoo sample (§3.2.3), we select Milky Way/M31 analogues as all those galaxies from the model with the following well defined properties\(^2\). They must:

- be the most massive (i.e. central) galaxy of their dark matter halo,
- possess a stellar mass in the range $10.66 < \log_{10}(M_\star [M_\odot]) < 11.12$ (identical to our face-on Galaxy Zoo Milky Way/M31 analogue sample),
- possess an approximate Sb/c morphology with a bulge-total luminosity ratio of $1.5 < M_B^{\text{bulge}} - M_B^{\text{total}} < 2.6$ (Simien & de Vaucouleurs, 1986), and
- be face-on with $\log_{10}(a/b) < 0.2$ for an observer at a random position in our simulated volume (reducing the effects of internal dust extinction and again mimicking our face-on Galaxy Zoo Milky Way/M31 analogue sample criteria).

\(^2\)Note that we include the effects of dust extinction on galaxy luminosity through a simple ‘plane-parallel slab’ model (Kauffmann et al., 1999), and assume a Chabrier initial mass function when calculating galaxy masses (Chabrier, 2003a).
3.4. Results

![Contour plot of u-r colour vs. stellar mass for the full Galaxy Zoo sample. The contours are linearly spaced. The stellar mass corresponding to the absolute magnitude limit of the sample is plotted as a dotted line. The solid line indicates the location of an independently derived best fit bimodal colour population division (Baldry et al., 2006). The bounding dashed lines are ±0.1 (u-r) of this division and delineate our definition of the green valley region. The locations of the Milky Way and M31 are indicated using the values listed in Table 3.1 after converting the B-V colour values to u-r (see §3.4.1). It is apparent that M31 is a candidate green valley member, however, the large uncertainty on the colour of the Milky Way precludes us from making a similar statement in its case. These criteria result in a sample of 28 439 Milky Way-type galaxies which we refer to as the theoretical Milky Way/M31 analogue sample. Note the larger number of galaxies in this sample when compared to the Galaxy Zoo analogue sample. This is predominantly due to the increased volume probed by the simulation. We also note that the use of the above bulge-to-total luminosity ratio to define galaxy morphology is again an approximation which will result in the selection of galaxies of ~Sb-Sc morphology.](image-url)
3.4 Results

3.4.1 The position of the Milky Way and M31 on the colour–stellar mass diagram

Galaxy colour is commonly used to identify galaxies at broadly different evolutionary stages. As a galaxy ages, its cold gas reserves – the raw material for star formation – are depleted, causing star formation to subside and the primary stellar population to age and redden. As a result, and in the absence of complications due to dust obscuration, we find that red galaxies are older and no longer forming stars, while blue galaxies are younger with active star formation and larger reserves of cold gas.

It has been well established for some time that these blue and red galaxies form distinct populations on a colour vs. magnitude (and hence colour vs. stellar mass) diagram (Strateva et al., 2001; Baldry et al., 2006). They are commonly referred to as the ‘red sequence’ and ‘blue cloud’, with the sparsely occupied region between them dubbed the ‘green valley’.

In Fig. 3.1 we show the full Galaxy Zoo observational sample (§3.2.3) plotted in the $u$-$r$ colour magnitude vs. stellar mass plane with linearly spaced contours. The dotted lines indicate the minimum stellar mass sample selection limit as a function of colour, corresponding to the magnitude limit of the sample.

Both the red sequence and blue cloud are clearly visible in Fig. 3.1, and highlight the green valley as an under-densely populated region between the two. We have overlaid the independently derived best fit bimodal colour population division of Baldry et al. (2006) (solid curved line). The green valley region is defined to be $\pm 0.1(u-r)$ of this division (bounded by dashed curved lines). We have checked that making minor variations to the assumed width of the valley region results in no qualitative change to the conclusions of this chapter.

We frame the Milky Way and M31 against the backdrop of the large Zoo sample by over-plotting their colours in Fig. 3.1, after converting from $B$-$V$ (Table 3.1) to $u$-$r$ (SDSS model) using the following formula:

$$ (u - r) = (1/0.3116) \left[ (B - V) - 0.1085 \right]. $$

This was derived from the magnitude system conversions provided by Robert Lupton and published on the SDSS website\(^3\).

\(^3\)http://www.sdss.org/dr7/algorithms/sdssUBVRITransform.html
3.4. Results

Figure 3.2: Plot of $B-V$ colour as a function stellar mass for the Galaxy Zoo sample. Grey points represent a random sampling of the full sample population. Solid and dashed lines are as in Fig. 3.1. Green shaded contours are logarithmically spaced and show the distribution of Milky Way/M31 analogue galaxies in colour space for varying selection criteria. Panel (a) shows distribution for our fiducial criteria, described in §3.2.3. Panel (b) shows how the relative colour fractions change if we include galaxies of all inclinations. Finally, panel (c) demonstrates the effects of relaxing our criteria to include all spiral types (but maintaining the inclusion of only face-on galaxies).

The stellar mass and resulting $u-r$ colour of M31 is consistent with the green valley region of the full Zoo sample. However, the size of the uncertainty on the colour of the Milky Way is too large to allow us to place any reasonable constraints on its colour classification. Further evidence is required to determine how statistically plausible it is that these two galaxies are indeed ‘green’.

3.4.2 How common are green valley Milky Way/M31-type galaxies?

We now look for evidence to support the location of the Milky Way and M31 in the green valley, as shown in Fig. 3.1. The Galaxy Zoo data set provides us with a statistically significant sample of Milky Way-like galaxies with measured morphologies, stellar masses
and colours (§3.2.3). With it we can investigate the statistical likelihood of Sb/c spiral galaxies residing in the green valley or on the red sequence.

In Fig. 3.2(a) we show the $B-V$ colour vs. stellar mass distribution of the Galaxy Zoo Milky Way/M31 analogue sample (green contours) superimposed against the full Zoo sample used previously in Fig. 3.1 (randomly sampled 2%; grey points). The Johnson $B-V$ colour values were converted from the SDSS AB model magnitude $u-r$ colours using the inverse of the transformation provided in Eqn. 3.1. The thick solid line again indicates the optimum red/blue colour division derived by Baldry et al. (2006). The green valley division lines of Fig. 3.1 (again converted to $B-V$ using Eqn. 3.1) have also been over-plotted.

Our colour classification results in 3% of our Milky Way/M31 analogues being classified as red, 13% as green and 84% as blue. In other words, from our Galaxy Zoo data we find that approximately 1 in 6 local Milky Way-like galaxies lie red-ward of the blue cloud. This suggests that it is statistically plausible for both the Milky Way and M31 to be candidate green valley members.

Fig. 3.2(b) is identical to Fig. 3.2(a), however, here we place no restriction that galaxies must be face-on. By removing this constraint we find a considerably higher fraction of galaxies (~45%) lying redward of the blue cloud. This indicates that many of the galaxies in the Galaxy Zoo Milky Way/M31 analogue sample likely suffer from significant reddening due to internal dust extinction and therefore do not have truly passive stellar populations. By sub-selecting only face-on galaxies in our analysis, we have reduced this effect considerably, allowing us to more realistically estimate the relative fractions of galaxies redder than the blue cloud (Masters et al., 2010a). However, we do note that it is still possible for the colours of our face-on sample to be contaminated by dust to some non-negligible extent (Driver et al., 2007). Although we make no attempt to quantify the magnitude of this effect, a simple investigation using a sample of face-on Galaxy Zoo spirals was carried out by Masters et al. (2010b). They found little evidence in their sample to suggest that red face-on spirals have a significantly larger dust content, and hence suffer from a greater reddening effect, than their blue counterparts at a given stellar mass. They hence concluded that dust reddening in their face-on sample was not a major cause for creating a red spiral population.

Displayed in Fig. 3.2(c), is the distribution of galaxies from the face-on Galaxy Zoo Milky Way/M31 analogue sample, but this time selecting all spirals, regardless of the morphology estimated by their fracdev values. Again, we find an increased fraction of galaxies redward of the blue cloud (~50%) when compared to our fiducial Sb/c morphology sample (Fig. 3.2(a)). As all of the galaxies in this plot are still face-on, this increased
3.4. Results

Figure 3.3: $B-V$ colour as a function stellar mass for the theoretical model galaxies. Grey points represent a random sampling of the full model output at $z=0$. Green shaded contours are logarithmically spaced and show the distribution of face-on model Milky Way/M31 analogue galaxies in colour space. The solid and dashed curved lines are as described in Fig. 3.1. The green valley region has been both raised and widened by 0.01 $B-V$ compared to the observational sample (Figs. 3.1 and 3.2) in order to more accurately reflect the location of the green valley in the model data. Our qualitative results are, however, insensitive to the precise magnitude of these changes. There is good qualitative and quantitative agreement between the model Milky Way analogue data set, shown here, and the observational Galaxy Zoo sample, shown in Fig. 3.2.

The high fraction of reddened Milky Way-type galaxies found when we relax our selection criteria demonstrates the conservative nature of our estimate that at least 1 in 6 local Milky Way/M31 analogues lie in the green valley or on the red sequence (dependent on how tightly we constrain the sample classification).
Chapter 3. The mid-life crisis of the Milky Way and M31

3.4.3 Further constraints using the model analogue sample

In Fig. 3.3, we show an equivalent colour vs. stellar mass plot for our theoretical Milky Way/M31 analogue sample at $z = 0$, constructed from the galaxy formation model described in §3.3.1. In this case, the grey points represent a randomly sampled 0.03% of the full model output and the green contours represent our model Milky Way/M31 analogues (§3.3.2). The green valley region has been raised by 0.01 $B-V$, and similarly widened by the same amount, with respect to our observational valley region so as to more accurately reflect the location of the green valley in the model data. We confirm that making sensible variations to both the offset and width of the defined green valley region results in no difference to our qualitative conclusions. Although the precise distribution of galaxies is not identical, the same colour bimodality present in the observational data (Fig. 3.1) is clearly visible. We find 24422 blue, 1623 green and 2394 red model Milky Way-like galaxies, corresponding to 86%, 6% and 8% of the sample respectively. Hence 14% of model Milky Way galaxies lie red-ward of the blue cloud (c.f. 16% in the observational Zoo sample (§3.4.2)).

The general agreement between the model and Galaxy Zoo data, for both the colour–stellar mass distribution and, in particular, the fraction of Sb/c spirals as a function of colour, gives us confidence that our simulation is a reasonable tool with which to extend our analysis. We now look for additional correlations using stellar mass and star formation rate as a function of galaxy colour, and then compare these to the observed masses and star formation rates of the Milky Way and M31.

In the left-hand panel of Fig. 3.4 we plot the star formation rate of our model galaxies against their stellar mass. The grey points are again a randomly sampled selection of the entire model output. These points form a locus which displays a trend of increasing star formation rate with increasing stellar mass and which can largely be associated with the blue cloud galaxies on the colour vs. stellar mass plot of Fig. 3.3 (Noeske et al., 2007). The red, green and blue points show the location of our theoretical Milky Way/M31 analogue sample in this space, broken down by colour as classified previously in Fig. 3.3. As expected, we find that blue galaxies have the highest star formation rates, and red galaxies the lowest. The actual observed positions of the Milky Way and M31 have been over-plotted using the values listed in Table 3.1. Both of these galaxies have star formation rates which place them below the main locus of the model star forming galaxies.

The right-hand panel of Fig. 3.4 shows the normalized star formation rate probability distribution (in log space) for our theoretical Milky Way/M31 analogue sample, split by colour. The observed star formation rate values of both the Milky Way and M31 are
3.4. Results

Figure 3.4: [Left] Plot of star formation rate (SFR) as a function stellar mass for our model galaxies. Grey points represent a random sampling of the full model output. Blue upward triangles, green circles and red downward triangles show the distribution of face-on Milky Way/M31 analogue galaxies in this simulated sample. The real, observed locations of both the Milky Way and M31 in this parameter space are indicated using the values listed in Table 3.1. The uncertainties in the stellar masses of both of these galaxies are less than, or equal to, the width of the symbols and so they have been omitted for clarity. [Right] Normalized probability distribution of log_{10}(SFR) for each Milky Way/M31 analogue colour sub-sample. The normalization is such that the probability of belonging to each colour sample at any given SFR may be directly compared. The SFRs of the Milky Way and M31 are shown as horizontal shaded regions. Against the backdrop of the galaxy formation model, they are approximately 2–5 and 3–4 times more likely to be associated with the green valley than the blue cloud respectively.

overlaid as grey shaded regions. When compared with the model star formation rate distributions, the observed star formation rates of both the Milky Way and M31 support their previous classification as green valley members in the colour–stellar mass diagram. In the case of the Milky Way, we find the probability of it being associated with the green valley, based on its star formation rate, to be \( \sim 3–5 \) times greater than the probability of it being associated with the blue cloud (relative to the model population). Similarly, the star formation rate of M31 makes it \( \sim 3–4 \) times more likely to be associated with the green valley. In fact, by this diagnostic, M31 is more likely to be red rather than green or
Chapter 3. The mid-life crisis of the Milky Way and M31

Figure 3.5: Normalized probability distributions of formation times for each colour sub-sample of the model Milky Way/M31 analogues. This time corresponds to the last snapshot in the simulation at which a galaxy possessed half of its $z=0$ stellar mass. All three colour sub-samples have similar formation time distributions indicating that age has little effect in determining the $z=0$ colour distributions of our analogue galaxies.

3.5 Discussion: What makes a ‘green’ Milky Way?

Gathering together the main results of §3.4 suggests that both the Milky Way and M31 are possible green valley members. In the generalized framework of galaxy evolution such objects are thought to represent those galaxies which are in the process of maturing from blue and star forming to ‘red and dead.’ How long does this transitional phase last? What is the mechanism which triggers a galaxy’s departure from the blue cloud? What is the significance of a ‘green Milky Way’? Our theoretical Milky Way/M31 analogue sample provides us with a powerful tool with which to gain important insights into these complex questions.
3.5. Discussion: What makes a ‘green’ Milky Way?

3.5.1 Are Red Milky Way Analogues Simply Older than Blue?

It seems reasonable to postulate that red sequence members are simply older than their blue counterparts, and so have progressed further in their colour evolution. In this scenario, green valley members would simply represent an intermediate population in terms of age.

Fig. 3.5 shows the distribution of formation times for red, green and blue model Milky Way/M31 analogues, defined to be the time (in Gyr) when the galaxy had reached half its final stellar mass (Croton et al., 2007). Immediately obvious is the similarity in the formation time distributions of all three colour sub-samples. Qualitatively similar results are found when we define formation time to be 10% or 90% of the final stellar mass, or use halo mass instead. This suggests that red or green Milky Way-type galaxies are not simply older versions of their blue counterparts. Based on the histories of our model galaxies, we can hence discount formation time as being a critical property on which spiral galaxy colour depends.

3.5.2 Rapid or gradual shut-down of Milky Way star formation?

The fractional abundance of Galaxy Zoo Milky Way-like green valley or red sequence galaxies is approximately 15%. For those model analogues which have crossed on to the red sequence in the last \( \sim 10 \) Gyr, we find the average time spent in the green valley to be \( \sim 1.5 \) Gyr.

The time-scale for passive evolution across the green valley after an instantaneous shut down of star formation is typically around a Gyr, somewhat shorter than the transit time found in the model (as measured from the same start and end colour points). This suggests that it is unlikely that star formation is simply ‘switched off’, but instead undergoes a somewhat more gradual decline (Baldry et al., 2004a; Balogh et al., 2011). This is also in qualitative agreement with Masters et al. (2010a) who similarly suggest a gradual shutdown of star formation associated with their sample of red passive spirals drawn from Galaxy Zoo.

3.5.3 Cold gas depletion in spiral galaxies?

Due to their typically quiescent merger histories, it is more likely that in-situ processes, rather than merger related processes, are responsible for stifling star formation in Milky Way-like galaxies. In-situ star formation in the model is directly driven by the amount of cold gas in, and the dynamical time-scale of, the galactic disk (Kennicutt, 1998). Simply put, the absence of cold gas implies the absence of star formation. Hence, we focus here
Chapter 3. The mid-life crisis of the Milky Way and M31

Figure 3.6: Plot of the mean cold gas mass of each colour sub-sample of our model Milky Way/M31 analogues as a function of look-back time. There is an obvious trend for red galaxies to have less cold gas than their bluer counterparts at $z=0$. All colour samples have an approximately equal peak mean cold gas mass, with the difference at $z=0$ being due to the time at which the mean cold gas content began its decline (as indicated by the arrows).

Cold gas in galaxies is regulated by five primary mechanisms: cooling flows from the halo, star formation, supernova feedback, AGN heating, and galaxy-galaxy mergers/interactions. The complex interplay between each is non-trivial to predict \textit{a priori}, requiring the use of models and simulations.

In Fig. 3.6 we show the mean cold gas mass for red, green and blue model Milky Way-like galaxies as a function of look-back time. The evolution of cold gas shows a consistent trend across all three colour bins: a growth in cold gas content from high redshift which peaks at approximately the same mass, followed by a decline at lower redshift as the galaxy moves off the blue cloud. Importantly, this turn-over occurs at earlier times for redder galaxies (black arrows). Notice the clear deficit of cold gas associated with local red spirals - a factor of two lower than blue spirals on average.

In general, cooling flows are the dominant source of fresh cold gas in the Milky Way/M31 analogue sample in our model. Any mechanism which leads to the net re-
duction of cold gas mass will thus need to remove the gas at a faster rate than it is being replenished from the surrounding halo as the system grows with time.

3.5.4 Cold gas heating in spiral galaxies?

We now identify the three primary mechanisms thought to be responsible for reducing the abundance of disk gas in massive, gravitationally dominant spiral galaxies, and discuss each of these in turn.

First is the depletion of gas due to galaxy-galaxy mergers, which rapidly transforms disk gas into stars via starbursts. We have already argued that major mergers are not prevalent in our sample of Sb/c galaxies (by selection) and hence do not play a significant role here. We also find that minor merger bursts only alter the gas mass, on average, by at most 5% at $z = 0$. The trend seen in Fig. 3.6 is therefore not due to mergers.

Second is the heating associated with supernova feedback following episodes of star formation. Supernova feedback acts to remove gas from the disk and return it to the surrounding halo medium. Supernova are certainly acting in our model in galaxies in this mass range. However, the energy available to remove gas from an $L^*$ galaxy disk is significantly smaller, on average, than that needed to offset the cooling rate from an $L^*$ host halo. In addition, star formation is needed to produce supernova and hence supernova feedback is not a plausible path to a population of long-term star formation deficient (i.e. red) galaxies.

The third mechanism to reduce the abundance of disk gas, and hence star formation, is some form of active black hole heating. Black holes provide a plausible way of regulating the supply of halo gas into the disk through periodic heating cycles of low luminosity feedback (Croton et al., 2006). Their inclusion in galaxy formation models has provided a solution to a number of longstanding problems in galaxy formation theory, such as the overcooling problem in cluster halos, and the turn-over at the bright end of the galaxy luminosity and stellar mass functions.

In Fig. 3.7 the distribution of black hole masses in our model analogue galaxies is shown. As found in several works (e.g. Croton et al., 2006, 2007), such masses are consistent with the observed local $M_{\text{BH}} - M_{\text{bulge}}$ relation. Even so, Fig. 3.7 reveals that red sequence Milky Way/M31 analogues tend to have a mean central black hole mass that is approximately two times larger than that of their blue cloud counterparts. When such galaxies become active, the heating is similarly larger by a comparable factor. In the model, this extra heating provides enough energy to suppress cooling and stifle the supply of fresh gas in a
Chapter 3. The mid-life crisis of the Milky Way and M31

Figure 3.7: Normalized probability distributions of black hole masses for red, green and blue galaxies drawn from our model Milky Way/M31 analogue sample. Red galaxies are, on average, a factor of two more massive than blue counterparts. There is also a notable trend for the spread in black hole masses to decrease for redder galaxies.

way that smaller holed spirals cannot.

3.5.5 Spiral galaxies on the edge

AGN heating is the dominant physical process responsible for the decline in star formation rates in our model Milky Way/M31 analogue sample, and hence explains the green and red Sb/c model galaxy subpopulations. There is some observational support for such a mechanism acting in real Milky Way-like objects. For example, the fractional abundance of galaxies hosting active black holes is concentrated towards the green valley, supporting its theoretical use as a shut-down mechanism. Schawinski et al. (2010) suggest that the Milky Way should lie in the highest duty cycle region of the colour–stellar mass plane of late-type galaxies. Furthermore, Masters et al. (2010a) select a sample of face-on spirals with passive disks from Galaxy Zoo and find an increased fraction of Seyfert+LINER galaxies; a commonly suggested source of LINER emission in galaxies is low-luminosity

\[^4\]Remember that \(L^*\) galaxies straddle an observed tipping point between star forming and passive, and the model is constrained to mimic this observation.
AGN. Finally, several studies (Giordano et al., 2010; Masters et al., 2011, e.g.,) find an increased prevalence of optical bars in redder spiral galaxies, with Masters et al. (2010a) also noting an increased bar fraction in their face-on, passive sample. Such bars have long been favored as a mechanism to channel gas onto a black hole through instabilities, and may be relevant to our understanding of the Milky Way, which is purported to have a significant bar (Blitz & Spergel, 1991).

Tentative support also exists to suggest that Sgr A* (the central black hole of the Milky Way) itself, may have been considerably more active than its current state in the recent past. For example, the observed luminosity of the Milky Way’s black hole may have been almost 5 orders-of-magnitude brighter as little as 100 years ago (Terrier et al., 2010), and there have been recent claims of observed gamma ray bubbles seen above and below the galactic plane, supposedly as a result of an energetic event in the Galaxy’s central region dated to within the past 10 Myr (Su et al., 2010). Recent research has suggested that this energetic event is most likely due to past AGN activity (Guo & Mathews, 2012). Still, we stress that we are not aware of any definitive evidence to suggest that either the Milky Way or M31 would be classified as active galaxies if observed from a more privileged position in time or space. In particular, although the mass of the central black hole of M31 (approximately 1.4 × 10^8 M_☉; Bender et al., 2005) is consistent with the red/green model analogue populations of Fig. 3.7, the mass of Sgr A* is only thought to be approximately 4.3 × 10^6 M_☉ (Gillessen et al., 2009). Additionally, if it proves that no recent AGN activity has indeed occurred in these two galaxies then for AGN to remain as a viable mechanism responsible for causing their reddening, a large time delay between any AGN event and the associated suppression of star formation would be required.

One thing we can say for certain is that L* galaxies, such as the Milky Way, occupy a unique position in the hierarchy of galaxy evolution that borders both star forming and passive galaxies on average. Their position at the ‘knee’ of the galaxy luminosity function corresponds to the point where star formation quenching starts to become efficient (Benson et al., 2003) and also to the transition valley region of the D_n(4000)–stellar mass plane (see figure 1 of Kauffmann et al. 2003). Thus, the significance of the separation in black hole mass between red and blue spirals seen in Fig. 3.7 appears to simply be a consequence of our model Milky Way/M31 analogues straddling this observationally required tipping point; the more massive holed spirals are forced onto the red sequence (in our model through quenching via AGN heating) and the less massive holed spirals remain fixed in the blue cloud. Our model green valley Milky Way-like galaxies may simply represent the mid-point of such evolution.
Chapter 3. The mid-life crisis of the Milky Way and M31

Figure 3.8: [Upper] Illustration of the standard paradigm for galaxy evolution in a hierarchical universe. [Lower] The alternative evolution suggested by a green, passively evolving population of spiral galaxies. The main difference is the lack of a quasar mode phase in the alternative scenario. This is due to the depletion of cold gas, and the associated reddening of the stellar populations, which has already occurred in both galaxies before the merger event. With not enough cold gas to act as fuel, no powerful quasar mode burst occurs. See the §3.5.5 for a detailed discussion.

Although AGN heating is the physical mechanism responsible for truncating star formation in green and red Sb/c galaxies in the model, we note that we currently have few clear ways to distinguish this quenching mechanism from others that share similar scaling relations (at least given the current observational data). The form of the invoked heating is something which has primarily been motivated by galaxies in more massive halos (e.g. clusters), and it is natural to extend this down to L* objects as a first approximation. We are thus provided with a valuable opportunity for both theorists and observers. Studies of a wider range of physically motivated quenching mechanisms, especially in lower mass L*-scale dark matter halos, should allow modelers to find consistency with a wider range of observed galaxy populations. In harmony with this, programs and upcoming survey instruments (such as Gaia and HERMES), will allow observers to accurately quantify the nature of the colour and star formation rate transformation that our own galaxy may be currently undergoing.
3.6. Conclusions

No matter what the cause, if the Milky Way and M31 truly are green valley galaxies and are transitioning on to the red sequence (perhaps slowly), as suggested here by their global colours and star formation rates, then this suggests a new paradigm of massive galaxy death which we outline in Fig. 3.8 and which is somewhat different to the conventional picture. In the traditional, hierarchical paradigm of galaxy evolution (upper), a galaxy’s life on the blue cloud \((a)\) ends in a gas rich major merger with another galaxy. This triggers the rapid accretion of cold gas on to the central black hole, seen observationally as a quasar \((b)\). The quasar rapidly fades (over a few 100 Myr) to much lower luminosities (Hopkins et al., 2006), and heating from this low luminosity AGN then suppresses the further cooling of gas onto the galaxy over much longer timescales \((c)\). This limits both the reformation of a new galactic disk, locking in the spheroidal morphology that resulted from the initial merger, and future star formation. The galaxy finally retires onto the red sequence as a passive ‘red and dead’ elliptical \((d)\). It may still grow through subsequent dissipationless ‘dry’ mergers, but these only serve to maintain its colour and morphology (Faber et al., 2007).

Our own Local Group galaxies paint an alternative picture (lower panel of Fig. 3.8). In this case, both the Milky Way and M31 are expected to merge in the next \(\sim 5 \text{ Gyr}\) (Cox & Loeb, 2008). If our current understanding of the transit time across the green valley is correct (a few Gyr), and if both galaxies are not ‘special’ long-term residents of this region, then they should have already quiescently evolved on to the red sequence before the merger event. The eventual collision will thus be gas poor and it is likely that no quasar of significance will occur. The fading disks will still be destroyed, again resulting in a remnant elliptical galaxy with a mass approximately \(2L^*\). Subsequent evolution through dry mergers should then re-establish the standard paradigm.

This alternative picture is supported by a number of other observations of external galaxies. For example, Darg et al. (2010) found a small merger fraction of \(\sim 5\%\) in the local universe, thus suggesting that mergers may be less important to current galaxy evolution than previously considered. Also it has been demonstrated that in the Galaxy Zoo sample, the rate of change of galaxy colour from blue to red with environmental density is faster than the associated rate of change from spiral to elliptical (Bamford et al., 2009; Skibba et al., 2009).

3.6 Conclusions

Using the current best estimates of stellar masses, colour indices and star formation rates for the Milky Way and M31, we investigate the current evolutionary position of these
two galaxies through a comparison with an observational sample of galaxies drawn from SDSS / Galaxy Zoo1 data (Bamford et al., 2009). Our key observational results can be summarized as follows:

- M31 is broadly consistent with being a member of the green valley population of the colour vs. stellar mass diagram. The large uncertainty in the global colour of the Milky Way makes a similar conclusion regarding its colour classification - based solely on this quantity - impossible. (§3.4.1).

- At least 1 in 6 Milky Way-type galaxies in the local Universe lie red-ward of the blue cloud (assuming our most conservative selection criteria including galaxy orientation and Hubble classification) (§3.4.3).

We then utilize the semi-analytic galaxy formation model of Croton et al. (2006) to generate a sample of theoretical Milky Way/M31 analogues. After showing that this model is able to qualitatively reproduce the observed bimodality in the colour vs. stellar mass plane for Milky Way-type galaxies, we use it to support our findings and elucidate the dominant physical mechanisms which may be at play:

- The observed star formation rates of the Milky Way and M31 are consistent with both galaxies being categorized as green valley members when compared to the distribution of model Milky Way/M31 analogue star formation rates in different colour bins (§3.4.3).

- There is no significant correlation of formation times with colour in our model Milky Way/M31 analogue galaxies, suggesting that the red sequence analogues are not simply older versions of their blue counterparts (§3.5.1). Also, red model Milky Way/M31 galaxies transition across the green valley somewhat slower than what would be expected if star formation was instantaneously truncated (§3.5.2).

- The reduction of the availability of cold gas is the main cause for the decline of star formation in green and red model analogue galaxies. In the model this is a result of heating from active super-massive black holes. Green model Milky Way/M31 analogues straddle the tipping point between efficient and inefficient (or no) AGN feedback (§3.5.3,§3.5.5).

- As the Milky Way and M31 are generally not considered to be active galaxies, our results suggest that alternative processes may be acting (although their precise nature is not currently obvious), to explain the position of these two galaxies in the
3.6. Conclusions

...colour–stellar mass plane. If low level AGN activity is responsible then it would require the presence of a considerable time delay between the feedback event and the truncation of star formation in such galaxies (§3.5.5).

Our key result, that both the Milky Way and M31 are possibly ‘green’, has a number of important implications. Although these galaxies are by no means atypical, they may not be examples of archetypal star forming spirals. Instead, they possibly represent a population of galaxies in the midst of a transitional process. Recent observations reveal that as many as 60% of all galaxies which transition on to the red sequence have undergone a quiescent disk phase (Bundy et al., 2010) and red sequence spiral galaxies are not uncommon (Masters et al., 2010a, e.g.). Care must therefore be exercised when using the Milky Way as a benchmark in studies of external galaxies.

The standard paradigm of galaxy evolution is for merger events to drive major changes in star formation and hence the bulk movement of objects on the colour–stellar mass diagram. The presence of high stellar mass spiral galaxies in the green valley and red sequence regions is somewhat contrary to this picture (Fig. 3.8). The fact that both our own galaxy and our nearest neighbour may be actively taking part in this alternative evolutionary scenario reveals their possible eventual fate. The consequences for our understanding of galaxy evolution are significant (Peebles & Nusser, 2010) and serve to highlight the wide-reaching importance of Galactic experiments such as HERMES and Gaia, as well as APOGEE (Allende Prieto et al., 2008) and upcoming cold gas surveys with ASKAP and ALMA (Johnston et al., 2008; Minniti et al., 2010). The Milky Way may provide us with a rare opportunity to examine, in exquisite detail, a galaxy actively experiencing one of the most important global transformations in galactic evolution.
A formation history model of galaxy stellar mass growth

We introduce a new, simple model to self-consistently connect the growth of galaxies to the formation history of their host dark matter halo. Our model is defined by two simple functions: The “baryonic growth function” which controls the rate at which new baryonic material is made available for star formation, and the “physics function” which controls the efficiency with which this material is converted into stars. Using simple, phenomenologically motivated forms for both functions that depend only on a single halo property, we demonstrate the model’s ability to reproduce the $z=0$ red and blue stellar mass functions. By adding redshift as a second input variable to the physics function we show that the global stellar mass function can also be reproduced out to $z=4$. We conclude by discussing the general utility of our new model, highlighting its usefulness for creating mock galaxy samples which have a number of key advantages over those generated by other techniques.

*This chapter is presented as an advanced draft of a journal submission.*
4.1 Introduction

Theoretical models are an important and commonly used tool for interpreting and furthering our understanding of observed galaxy populations. Typically, these models are used to generate mock galaxy catalogues which are then compared to equivalent samples drawn from the real Universe. If the models are successful in reproducing these observations then their construction can often be used to make important inferences about the physics of galaxy formation and evolution.

There are a number of different methods for generating mock galaxy samples for comparison with observations. At the most advanced and complex end of the scale are full hydrodynamic simulations. These simulations attempt to solve the physics of galaxy formation from first principles, directly modelling complex baryonic processes such as cooling and shocks in tandem with the dissipationless growth of dark matter structure. Unfortunately, the associated high computational cost prohibits the resolution of small scale physical processes such as star formation and black hole feedback in a volume large enough to provide a cosmologically significant sample of galaxies. Hydrodynamic simulations are therefore typically required to resort to parameterised approximations in order to deal with these unresolved “sub-grid” processes. In addition they must also deal with the complex and often poorly understood numerical effects that come with modelling dissipational physics using a finite physical and temporal resolution.

Semi-analytic galaxy formation models attempt to overcome the computational costs associated with hydrodynamic simulations by separating the baryonic physics of galaxy formation from the dark matter dominated growth of structure (Kauffmann et al., 1999). This is achieved by taking pre-generated dark matter halo merger trees and post-processing them with a series of physically motivated parametrisations that attempt to capture the mean behaviour of the dominant baryonic processes involved in galaxy formation. The resulting speed means that these models can be used to generate cosmologically significant samples of galaxies using only modest computing resources. However, semi-analytic models typically require a large number of free parameters, many of which are often not well constrained by theory or observation (Neistein & Weinmann, 2010). The complicated and interlaced nature of the different physical prescriptions also means that the effects of these parameters on the final galaxy population are often highly degenerate (Chapter 2, see also Lu et al. 2012). This can make it difficult to disentangle the relative contributions of each physical process in shaping the predicted galaxy population, thus complicating the interpretation of these models. Additionally, our relatively poor understanding of high redshift galaxy formation means that at least some of the parameterisations used may not
be appropriate at these early times (see §2.6).

In some cases our goal is not to try and produce a complete galaxy population that exactly matches the real Universe as closely as possible. For many science questions and applications it is sufficient to construct much simplified “toy” models. These typically build an average population of galaxies using vastly simplified approximations, and are often designed to test new ideas and interpretations or to allow the investigation of particular trends or features found in observational data. An example is the “reservoir” model of Bouché et al. (2010). Here, averaged dark matter halo growth histories are used to track the typical build up of cold gas in galaxies. In their fiducial formalism, the accretion of baryons onto a galaxy is only allowed to occur when the host halo lies in a fixed mass range. However, within this range the accretion is modeled as a simple fraction of the halo growth rate. Using a standard Kennicutt–Schmidt law (Kennicutt, 1998) for star formation, this simple model is able to reproduce the observed scaling behaviours of the star forming main sequence and Tully-Fisher relations. Expanding upon this framework, Krumholz & Dekel (2012) also introduced a metallicity dependent star formation efficiency, allowing them to straightforwardly investigate the associated effects on the star formation histories of galaxies.

Rather than attempting to generate galaxy populations based on our theories of the relevant physics, an alternative method of generating mock galaxy samples is to use purely statistical methods. Halo Occupation Distribution (HOD) models use observed galaxy clustering measurements to constrain the number of galaxies of a given type within a dark matter halo of a any given mass (Peacock & Smith, 2000; Zheng et al., 2005). For the purposes of constructing mock galaxy catalogues, such a methodology has the advantage that it requires no knowledge of the how each galaxy forms and is also statistically constrained to produce the correct result. However, this limits our ability to learn about the physics of galaxy formation and their evolution, as one has no way to self-consistently connect galaxies at any given redshift to their progenitors or decedents at other times.

A similar method to HODs for creating purely statistical mock catalogues is subhalo abundance matching (SHAM; e.g. Conroy et al., 2006). SHAM models are typically constructed by generating a sample of galaxies of varying masses, drawn from an observationally determined stellar mass function. Each galaxy is then assigned to a dark matter halo, taken from a halo mass function generated using an N-body simulation. This assignment is made such that the most massive galaxy is placed in the most massive halo and so on, proceeding to lower and lower mass galaxies. Due to possible differences in the formation histories of halos of any given mass, an artificial scatter is often added during
Chapter 4. A formation history model of galaxy stellar mass growth

this assignment procedure (e.g. Conroy et al., 2007; Behroozi et al., 2012; Moster et al., 2012).

By leveraging the use of dark matter merger trees as the source of the halo samples at each redshift, both HOD and SHAM studies have been able to provide important constraints on the average build up of stellar mass in the Universe (Zheng et al., 2007; Conroy & Wechsler, 2009; Moster et al., 2012; Behroozi et al., 2012). This has allowed these studies to also draw valuable conclusions about topics such as the efficiency of star formation as a function of halo mass and the role of intra-cluster light (ICL) in our accounting of the stellar mass content of galaxies. However, both HOD and SHAM models are applied independently at individual redshifts and do not self-consistently track the growth history of individual galaxies. This limits the remit of these models to considering only the averaged evolution of certain properties over large samples.

Our goal in this chapter is to present a new class of galaxy formation model which allows us to achieve the “best of both worlds”: providing both a self-consistent growth history of each individual galaxy, whilst also minimising any assumptions about the physics which drives this growth. The model we present is closely related to the simple “accretion floor” model of Bouché et al. (2010), but with several important generalisations and extensions that greatly increase its utility whilst still maintaining a high level of transparency and simplicity. This is achieved by tying star formation (and hence the growth of stellar mass) to the growth of the host dark matter halo in N-body dark matter merger trees using a simple but well motivated parameterisation that depends only on the properties of the halo itself. In this way, we are able to provide a complete formation history for every galaxy.

This chapter is laid out as follows: In §4.2 we introduce the framework of our new model. In particular §4.2.2 focusses on how we build up the baryonic content of dark matter halos with the practical details of the model’s application outlined in §4.2.3. In §4.3 we present some basic results, in particular investigating the model’s ability to reproduce the observed galaxy stellar mass function at multiple redshifts. In §4.4 we discuss our findings as well as outline the general utility of the model and a number of possible ways in which it could be extended. Finally, we present a summary of our conclusions in §4.5.

A WMAP1 (Spergel et al., 2003) ΛCDM cosmology with \( \Omega_m = 0.25, \Omega_\Lambda = 0.75, \Omega_b = 0.045 \) is utilised throughout this chapter. In order to ease comparison with the observational datasets we employ, all results are quoted with a Hubble constant of \( h = 0.7 \) (where \( h = H_0 / 100 \text{km s}^{-1} \text{Mpc}^{-1} \)) unless otherwise indicated. Magnitudes are presented using the Vega photometric system.
4.2 The simplest model of galaxy formation

4.2.1 The growth of structure

The aim of the model presented in this chapter is to self-consistently tie the growth of galaxy stellar mass to that of the host dark matter halos in as simple a way as possible. In order to achieve this we require knowledge of the properties and associated histories of a cosmological sample of dark matter halos spanning the full breadth of cosmic history. We obtain this in the form of merger trees constructed from the output of the N-body dark matter Millennium Simulation (Springel et al., 2005).

Using the evolution of over $10^{10}$ particles in a cubic volume with sides of length 714 Mpc, the Millennium Simulation merger trees track the build up of dark matter halos larger than approximately $2.9 \times 10^{10} M_\odot$ over 64 temporal snapshots. These snapshots are logarithmically spaced in expansion factor between redshifts 127–0 with an average separation of around 200 Myr. Each individual dark matter structure was identified using a Friends-of-Friends linking algorithm with further substructures (sub-halos) identified using the SUBFIND algorithm of (Springel et al., 2001). The simulation was run using a concordance ΛCDM cosmology with first year WMAP (Spergel et al., 2003) parameters: $(\Omega_m, \Omega_\Lambda, \sigma_8, h_0) = (0.25, 0.75, 0.9, 0.73)$.

4.2.2 The baryonic content of dark matter halos

The maximum star formation rate of a galaxy is regulated by the availability of baryonic material that can act as fuel. In our formation history model we assume that every dark matter halo carries with it the universal fraction of baryonic material, $f_b = 0.17$ (Spergel et al., 2003). However, some of these baryons will already be locked up in stars or contained in reservoirs of material that are unable to participate in star formation. Therefore we parametrise the dependence of the amount of newly accreted baryonic material which is available for star formation on the properties of the host dark matter halo using a baryonic growth function, $F_{\text{growth}}$.

In practice, only some fraction of this available material will actually make its way in to the galaxy, with an even smaller amount then successfully condensing to form stars in a suitably short time interval. The efficiency with which this occurs depends on a complex interplay of non-conservative baryonic processes both internal and external to the galaxy–halo system. A number of important examples include shock heating, feedback from supernova and active galactic nuclei (AGN), as well as environmental processes such as galaxy mergers and tidal stripping. Here we assume all of these complicated and
A formation history model of galaxy stellar mass growth

interwined mechanisms can be distilled down into a single, arbitrarily complex physics function, $F_{\text{phys}}$.

Combining all of this together, we can write down a deceptively simple equation for the growth of stellar mass in the universe:

$$\text{SFR} \equiv \dot{M}_* = F_{\text{growth}} \cdot F_{\text{phys}}. \quad (4.1)$$

In the following sections, we discuss the form we use for each of the baryonic growth and physics functions in turn.

The baryonic growth function

In order to explore the simplest form of our formation history model, we begin by assuming that as a dark matter halo grows, all of the fresh baryonic material it brings with it is immediately available for star formation. This corresponds to a baryonic growth function which is simply given by the rate of growth of the host dark matter halo:

$$F_{\text{growth}} = f_b \frac{dM_{\text{vir}}}{dt}. \quad (4.2)$$

In practice halos of the same $z=0$ mass may show a diverse range of growth histories, all of which are captured by our model. In Fig. 4.1 we demonstrate this by showing the individual growth histories of a random sample of dark matter halos selected from two narrow mass bins. From this figure we see that there can be variations in the time at which similar halos at redshift zero reach a given mass. For example, whilst some halos reach $10^{12} M_\odot$ at $z=5$, others do not reach this value until $z=2$. In addition, some halos may have complex growth histories, achieving their maximum mass at $z>0$. This may be caused by various effects such as stripping during dynamical encounters with other halos. Since the baryonic growth function maps the formation history of each individual dark matter halo to the stellar mass growth of its galaxy, this diversity in growth histories is fully captured, propagating through to be reflected in the predicted galaxy populations at all redshifts. This is an important attribute of the model which sets it apart from other statistical-based methods which merely map the properties of galaxies to the instantaneous or mean properties of halos, independently of their histories (e.g. HOD and SHAM models). These methods typically have to add artificial scatter to approximate the effects of variations in the halo histories, whereas this variation is a self-consistent input to our formation history model.
4.2. The simplest model of galaxy formation

Figure 4.1: This figure demonstrates the large variation in the possible growth histories of halos which all have approximately equal masses by redshift zero. The blue and red lines represent 30 randomly selected growth histories for halos with final $M_{\text{vir}}$ values of approximately $10^{12}$ and $10^{13} M_\odot$ respectively. Variations of 3–4 Gyr in the time at which these halos reach a given mass is common. Unlike statistical techniques for tying galaxy properties to their host halos, our formation history model implicitly includes the full range of different halo growth histories and their effects on the predicted galaxy population.

The physics function

The physics function describes the efficiency with which baryons are converted into stars in halos of a given mass. The form of this function may be arbitrarily complex, however, the goal of this chapter is to find the simplest model which successfully ties the growth of galaxy stellar mass to the properties of the host dark matter halos. The physics function is unlikely to provide an accurate reproduction of the details of the full input physics, but should be successful in reproducing the combined effects on the growth of stellar mass in the universe. In this spirit, we begin by assuming that there is only one input variable: the instantaneous virial mass of the halo, $M_{\text{vir}}$.

Although still not understood in detail, the observed relationship between dark matter halo mass and galaxy stellar mass is well documented (e.g. Zheng et al., 2007; Wang et al., 2012; Yang et al., 2012). Assuming the favoured ΛCDM cosmology, a comparison of the observationally determined galactic stellar mass function to the theoretically determined
halo mass function indicates that the averaged efficiency of stellar mass growth varies strongly as a function of halo mass. In Fig. 4.2, we compare a Schechter function fit of the observed redshift zero stellar mass function (Bell et al., 2003, solid blue line) against the dark matter halo mass function of the Millennium Simulation (red dashed line). The halo mass function has been multiplied by $f_b$ in order to approximate the total amount of baryons available for star formation in a halo of any given mass.

The increased discrepancy between the stellar mass function and halo mass functions at both low and high masses indicates that the efficiency of star formation is reduced in these regimes. It is commonly held that at low masses the shallow gravitational potential provided by the dark matter halos allows supernova feedback to efficiently eject gas and dust from the galaxy. This reduces the availability of this material to fuel further star formation episodes and hence temporarily stalling in-situ stellar mass growth. Other processes such as the photo-ionisation heating of the intergalactic medium may also play an important role in reducing the efficiency of star formation in this low mass regime (Benson et al., 2002, and references therein). At high halo masses, it is thought that inefficient cooling coupled with strong central black hole feedback also leads to a quenching of star formation (e.g. Croton et al., 2006). Therefore, it is only between these two extremes, around the knee of the galactic stellar mass function, that stellar mass growth reaches its highest average efficiency.

We begin by parametrising the physics function as a simple log-normal distribution centered around a halo virial mass $M_{\text{peak}}$, and with a standard deviation $\sigma_{M_{\text{vir}}}$:

$$F_{\text{phys}}\left(\frac{M_{\text{vir}}}{M_\odot}\right) = E_{M_{\text{vir}}} \exp\left(-\left(\frac{\Delta M_{\text{vir}}}{\sigma_{M_{\text{vir}}}}\right)^2\right)$$

(4.3)

where $\Delta M_{\text{vir}} = \log_{10}\left(\frac{M_{\text{vir}}}{M_\odot}\right) - \log_{10}\left(\frac{M_{\text{peak}}}{M_\odot}\right)$ and the parameter $E_{M_{\text{vir}}}$ represents the maximum possible efficiency for converting in-falling baryonic material into stellar mass, achieved when $M_{\text{vir}}=M_{\text{peak}}$. Such a distribution was found by the SHAM study of Conroy et al. (2009) to provide a good match to the derived star formation rates as a function of halo mass for $z<1$.

This simple form of the physics function provides a number of desirable properties. In Fig. 4.3, we present the average growth histories of five samples of dark matter halos chosen from the Millennium Simulation merger trees by their final redshift zero masses (solid blue lines). For clarity, we only plot these histories out to redshifts where more than 80% of the halos in each sample have masses which are twice the resolution limit of the input merger trees. The grey shaded region indicates the amplitude of the physics function
4.2. The simplest model of galaxy formation

The simplest model of galaxy formation is defined by Eqn 4.3 when using our fiducial parameter values (see §4.3.1 for details). As the halos grow, they pass through the region of efficient star formation at different times depending on their final masses. Galaxies hosted by the most massive $z=0$ halos form the majority of their in-situ stellar mass at earlier times whereas those in the lowest mass halos are still to reach the peak of their growth. In addition, lower mass halos tend to spend a longer time in the efficient star forming regime compared to their high mass counterparts. These trends qualitatively agree with the observed phenomenon of galaxy downsizing (e.g. Cowie et al., 1996; Cattaneo et al., 2008).

Sub-halo abundance matching studies have suggested that $V_{\text{max}}$ may be more tightly

![Figure 4.2: A comparison of the observed galactic stellar mass function (blue solid line) compared with the halo mass function of the Millennium Simulation (red dashed line). The halo mass function has been multiplied by the universal baryon fraction, in order to demonstrate the maximum possible stellar mass content as a function of halo mass. The closer the stellar mass function is to this line, the more efficient star formation is in halos of the corresponding mass. If galaxies were to form stars with a fixed efficiency at all halo masses then the slope of the stellar mass function would be identical to that of the halo mass function. The altered slope at both high and low masses indicates that star formation (as a function of halo mass) is less efficient in these regimes. At low masses, this is commonly attributed to efficient gas ejection due to supernova feedback, whereas at high masses energy injection from central super-massive black holes is thought to be able to effectively reduce the efficiency of gas cooling. However, many other physical processes may also contribute to these effects in both regimes.](image)
Chapter 4. A formation history model of galaxy stellar mass growth

Figure 4.3: The mean virial mass \( M_{\text{vir}} \) growth histories for five samples of dark matter halos with varying final masses (blue solid lines). The histories are only plotted until the redshift when a fifth of the halos have masses less than twice the 20 particle lower limit in order to avoid resolution artefacts. The grey shaded region indicates the amplitude of the physics function in Eqn. 4.3. This is also illustrated by the log-normal curve on the right-hand side (orange shaded region). Galaxies with halo masses within the peaked region form stars efficiently; outside this mass range, the amount of star formation is negligible. All galaxies of sufficient mass at \( z=0 \) will cross the efficient star formation mass range at some point in their history, with this period typically coming earlier for more massive halos. Also shown is the mean maximum circular velocity \( V_{\text{max}} \) evolution for halos of varying final velocity (red dashed lines). These samples have been chosen to have \( z=0 \) maximum circular velocity values similar to the mean values of the five mass selected samples. There are clear differences in the evolution of \( V_{\text{max}} \) and \( M_{\text{vir}} \) which will in turn result in differences between the produced galaxy populations.

coupled to the stellar mass growth of galaxies than \( M_{\text{vir}} \) (e.g. Reddick et al., 2012). This makes intuitive sense as \( V_{\text{max}} \) is closely related to the concentration of the inner regions of the host halo, where galaxy formation occurs. Therefore, as well as virial mass we also consider the case of a physics function where the dependent variable is the instantaneous
maximum circular velocity of the host halo, $V_{\text{max}}$:

$$F_{\text{phys}}(V_{\text{max}}/(\text{km s}^{-1})) = \mathcal{E}V_{\text{max}} \exp\left(-\left(\frac{\Delta V_{\text{max}}}{\sigma_{V_{\text{max}}}}\right)^2\right)$$  \hspace{1cm} (4.4)

where $\Delta V_{\text{max}} = \log_{10}(V_{\text{max}}/(\text{km s}^{-1})) - \log_{10}(V_{\text{peak}}/(\text{km s}^{-1}))$. To avoid confusion, from now on we will refer to the formation history model constructed using this physics function as the "$V_{\text{max}}$ model". Similarly, we will refer to the case of $F_{\text{phys}}(M_{\text{vir}})$ as the "$M_{\text{vir}}$ model".

In Fig. 4.3 we show the average $V_{\text{max}}$ growth histories for a number of different $z=0$ selected samples (red dashed lines). The y-axis has been scaled such that the grey band also correctly depicts the changing amplitude of the $V_{\text{max}}$ physics function as well as its $M_{\text{vir}}$ counterpart. Additionally, each of the $V_{\text{max}}$ samples in Fig. 4.3 (red dashed lines) are chosen to have mean $z=0$ values close to that of the five $M_{\text{vir}}$ samples (blue lines). However, there are clear differences between the growth histories of these two halo properties. In particular, the evolution of $V_{\text{max}}$ is slightly flatter, resulting in halos transitioning out of the efficient star forming region at an earlier time than the equivalent $M_{\text{vir}}$ sample. Such differences will have important consequences for the time evolution of the galaxy populations generated by each of the two physics functions and we highlight some of these in §4.3 below.

By combining the baryonic growth function with a physics function of the forms presented here, our resulting model may be thought of as a simplified and extended version of that presented by Bouché et al. (2010). Unlike their model, the scaling of gas accretion efficiency with halo mass, and the dependence of star formation on previously accreted material, is implicitly contained within our physics function. Most importantly though, Bouché et al. (2010) use statistically generated halo growth histories instead of simulated merger trees. Hence their model contains no information about the scatter due to variations in halo formation histories as well as no self-consistent stellar mass growth due to mergers.

### 4.2.3 Generating the galaxy population

Armed with the forms of our baryonic growth (Eqn. 4.2) and physics (Eqn. 4.3) functions, we now discuss the method we employ to generate the galaxy population from the input dark matter merger trees.

The change in dark matter halo mass, coupled with the time between each merger tree snapshot, provides us with the value of $dM_{\text{vir}}/dt$. This change in mass, $dM$, naturally includes growth due to both smooth accretion and merger events. Combined with the
instantaneous value of $M_{\text{vir}}$ or $V_{\text{max}}$, the rate of halo mass growth is then used to calculate a star formation rate for the occupying galaxy following Eqn. 4.1.

Some fraction of the mass formed by each new star formation episode will be contained within massive stars. The lives of these stars will be relatively short and therefore they will not contribute to the measured total stellar mass content of the galaxy. In order to model this effect we invoke the “instantaneous recycling” approximation (Cole et al., 2000), whereby some fraction of the mass of newly formed stars is assumed to be instantly returned to the galaxy ISM. Based on Croton et al. (2006) we take this fraction to be 30%, however, we note that changes to this value can be trivially taken into account by appropriately scaling the value of $E$ in the physics function.

By tracking the growth of stellar mass for each one of the available dark matter merger trees of the Millennium Simulation using the above methodology, we are able to generate a $z=0$ sample of approximately 25 million galaxies, each with self-consistent growth histories spanning the full breadth of cosmic history.

Although well motivated and conceptually simple, our use of $dM_{\text{vir}}/dt$ as the form of the baryonic growth equation (Eqn. 4.2) does introduce some practical considerations. For example, the change in halo mass from snapshot-to-snapshot in the input dark matter merger trees can be stochastic in nature, especially for the case of low mass or diffuse halos identified in regions of high density. Also, as satellite galaxies infall into larger systems, their halos are tidally stripped, leading to a negative change in halo mass and thus a reduction in stellar mass according to Eqn. 4.1. In the real universe, we expect that the galaxy itself is located deep within the potential well of its host halo and is therefore largely protected from the earliest stripping effects suffered by the dark matter (Peñarrubia et al., 2010). We must therefore decide when, if at all, to allow stellar mass loss when using this formalism. For simplicity, we address this by setting the star formation rate of satellite galaxies to be zero at all times; in other words fixing their stellar mass upon infall. This is unlikely to be true in the real universe across all mass and environment scales (Weinmann et al., 2006), however, the assumption of little or no star formation in satellite galaxies is a reasonable approximation and is relatively common in analytic galaxy formation models (e.g. Kauffmann et al., 1999; Cole et al., 2000; Bower et al., 2006; Croton et al., 2006). In the spirit of pursuing the simplest model, we therefore deem this to be a satisfactory approximation.

The form of the baryonic growth function presented in Eqn. 4.2 above is only one of a number of possibilities. As an example, one could use the instantaneous halo mass divided by its dynamical time, $M_{\text{vir}}/t_{\text{dyn}}$. This quantity grows more smoothly over the
4.2. The simplest model of galaxy formation

The simplest model of galaxy formation assumes a constant baryonic growth function that is never negative. Additionally, one may speculate this is a better representation of the link between stellar and halo mass build up. However, for simplicity, we do not investigate alternative forms of the baryonic growth function, but leave this to future work.

Satellite galaxies are explicitly tracked in the input merger trees until their host sub-halos can no longer be identified or fall below the imposed resolution limit of 20 particles. At this point, their position is approximated by the location of the most bound particle at the last snapshot the halo was identified. We then follow Croton et al. (2006) in assuming that the associated satellite galaxy will merge with the central galaxy of the parent halo after a timescale motivated by dynamical friction arguments (Binney & Tremaine, 2008):

$$t_{\text{merge}} = \frac{1.17}{G} \frac{V_{\text{vir}}^2}{m_{\text{sat}}} \ln(1 + M_{\text{vir}}/m_{\text{sat}}),$$

(4.5)

where $V_{\text{vir}}$ and $M_{\text{vir}}$ are the virial velocity and mass of the parent dark matter halo in km s$^{-1}$ and M$_\odot$ respectively, $r_{\text{sat}}$ is the current radius of the satellite halo in kpc, and $m_{\text{sat}}$ is the mass of the satellite in M$_\odot$. In these units, the gravitational constant, $G$, is given by $4.40 \times 10^{-9}$ kpc$^2$ km s$^{-1}$ M$_\odot^{-1}$ Myr$^{-1}$.

The ensuing merger event results in a galaxy with a stellar mass equal to the direct sum of the progenitor masses. Also, as in-falling satellite halos are stripped of their dark matter, this mass is typically added to the parent halo, contributing to its $dM_{\text{vir}}/dt$ value and thus the amount of star formation in the central galaxy. Our model therefore implicitly includes merger driven star-bursts under the assumption that these proceed with the same efficiency as in-situ star formation.

Knowledge of the star formation rate of each galaxy also allows us to calculate luminosities. For this purpose we use the simple stellar population models of Bruzual & Charlot (2003) and assume a Salpeter (1955) initial mass function. In the real universe supernova ejecta enriches the intra galactic medium, altering the chemical composition of the next generation of stars and the spectrum of the light they emit. As we do not track the amount of gas or metals in our simple model, (in the way in which a semi-analytic model would) we assume all stars are of 1/3 solar metallicity. This is a common assumption when no metallicity information is available. Finally, a simple “plane-parallel slab” dust model (Kauffmann et al., 1999) is applied to the luminosity of each galaxy in order to provide approximate dust extincted magnitudes. These magnitudes are used below to augment our analysis by allowing us to calculate the $B-V$ colour for each galaxy at $z=0$. However, our main focus will remain on stellar masses as these are a direct model
Chapter 4. A formation history model of galaxy stellar mass growth

<table>
<thead>
<tr>
<th>$M_{\text{vir}}$ model</th>
<th>$\log_{10}(M_{\text{peak}}/M_\odot)$</th>
<th>$\sigma_{M_{\text{vir}}}$</th>
<th>$\varepsilon_{M_{\text{vir}}}$</th>
<th>$\alpha_{M_{\text{vir}}}$</th>
<th>$\beta_{M_{\text{vir}}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>static ($\S$4.3.1)</td>
<td>11.7</td>
<td>0.8</td>
<td>0.6</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>evolving ($\S$4.3.3)</td>
<td>11.7</td>
<td>0.8</td>
<td>0.7</td>
<td>-0.5</td>
<td>-0.03</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$V_{\text{max}}$ model</th>
<th>$\log_{10}(V_{\text{peak}}/(\text{km s}^{-1}))$</th>
<th>$\sigma_{V_{\text{max}}}$</th>
<th>$\varepsilon_{V_{\text{max}}}$</th>
<th>$\alpha_{V_{\text{max}}}$</th>
<th>$\beta_{V_{\text{max}}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>static ($\S$4.3.1)</td>
<td>2.1</td>
<td>0.2</td>
<td>0.6</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>evolving ($\S$4.3.3)</td>
<td>2.1</td>
<td>0.2</td>
<td>0.7</td>
<td>-0.5</td>
<td>-0.09</td>
</tr>
</tbody>
</table>

Table 4.1: The fiducial parameter values of the physics function when using either $M_{\text{vir}}$ or $V_{\text{max}}$ as the input variable. Values are presented for both the non-evolving (see $\S$4.3.1) and evolving (see $\S$4.3.3) form. For the non-evolving form the values were chosen to provide the best reproduction of the observed $z=0$ colour-split stellar mass function. For the evolving form, the values were chosen to reproduce the evolution of the peak stellar–halo mass relation of Moster et al. (2012).

4.3 Results

Having outlined the construction of our simple formation history model, we now present some basic results which showcase its ability to recreate the observed distribution of galactic stellar masses. We begin by considering redshift zero alone, before moving on to investigate the results at higher redshifts. Throughout, we contrast and compare the variations between the predicted galaxy populations when using $M_{\text{vir}}$ or $V_{\text{max}}$ as the input halo property to the physics function (Eqns. 4.3 & 4.4).

4.3.1 Redshift zero

In order to determine the “best” parameters for the $M_{\text{vir}}$ form of the physics function we ran a grid of models covering the full plausible range of values. For each model realisation, we divided the $z=0$ galaxies into red and blue populations using a mass independent colour cut of $B-V = 0.8$. This is equivalent to the colour division found by the 2dF Galaxy Redshift Survey (Cole et al., 2005). From these, we selected the parameters which provided the best visual reproduction of the observed colour split stellar mass functions of Bell et al. (2003). These observed mass functions are constructed from a $g$-band limited sample taken from a combination of Sloan Digital Sky Survey (SDSS)
4.3. Results

In Fig. 4.4 we show the colour–stellar mass diagram produced using the best parameters of our $M_{\text{vir}}$ model. The black dashed line indicates the colour split used to divide the galaxies into red and blue populations. Although there is a lack of a clear colour bimodality as seen in observational data (e.g. Baldry et al., 2004b), there is an overabundance of galaxies with $B-V \approx 0.87$ corresponding to the observed “red sequence”. The presence of this feature at approximately the correct position in colour space (Cole et al., 2005) is a notable success for such a simple model.

In the upper panel of Fig. 4.5 we show the resulting red and blue galaxy stellar mass functions (solid lines) against the corresponding constraining observations (error bars). Despite its simplicity, the chosen form of the physics function produces a good reproduction of the observed mass functions.
Figure 4.5: The red and blue $z=0$ stellar mass functions produced by the formation history model (solid lines) when using $M_{\text{vir}}$ (top panel) and $V_{\text{max}}$ (bottom panel) as the input variable to the physics function (Eqns. 4.3 & 4.4). Galaxy colour is classified using a mass-independent colour cut of $B-V=0.8$. The observed colour split stellar mass functions of Bell et al. (2003) are shown for comparison. The parameters of the physics functions using both input variables have been individually chosen to provide the best visual reproduction of these observations. However, the fact that such an agreement can be achieved at all is an important success given the simplicity of the formation history model.
to the observed red and blue stellar mass functions. This is true across a wide range in stellar mass, indicating that the model is capable of successfully matching the integrated time evolution of stellar mass growth as a function of halo mass at \( z=0 \). Also, since blue galaxies preferentially trace those objects which have undergone significant recent star formation, the model’s reproduction of the observed blue mass function suggests that the rate of star formation as a function of stellar mass near this redshift is also in broad agreement with the observed universe.

The fact that such an agreement is attainable with this simple model should be viewed as a key success of the methodology and a validation of the general form we have chosen for the physics function, \( F_{\text{phys}}(M_{\text{vir}}) \). Having said this, there are some differences in the upper panel of Fig. 4.5 worth noting. In particular, there is an over-prediction in the number density of the most massive red galaxies. We expect that this discrepancy could be lessened by refining the physics function parameter values with a more sophisticated calibration method such as a Monte-Carlo Markov Chain analysis. This is discussed further in §4.4.

Having established that a physics function constructed using \( M_{\text{vir}} \) as the single input variable can successfully provide a good match to the observed \( z=0 \) red and blue stellar mass functions, we now turn our attention to the results of using \( V_{\text{max}} \) as the input property (Eqn. 4.4).

In the lower panel of Fig. 4.5 we present the colour split stellar mass functions for the \( V_{\text{max}} \) model. Again, a fixed colour division of \( B-V=0.8 \) is used to define the two colour populations and we have chosen the physics function parameter values to provide the best visual reproduction of the Bell et al. (2003) data. The resulting parameter values are presented in Table 4.1. Unsurprisingly, a comparison with the equivalent values of the \( M_{\text{vir}} \) model indicates that the peak efficiency of converting fresh baryonic material into stars in a single time step remains unchanged (i.e. \( \mathcal{E}_{V_{\text{max}}} = \mathcal{E}_{M_{\text{vir}}} \)).

As was the case for the \( M_{\text{vir}} \) model (upper panel of Fig. 4.5), an excellent reproduction of the observations is attainable when using \( V_{\text{max}} \) as the input parameter to the physics function. However, there are some notable differences. For example, the over-prediction of high mass red galaxies has been alleviated, although at the cost of now somewhat over-predicting the number density at the low mass end of this colour population. However, for the purposes of this section, we simply wish to highlight that although there are clear differences in the time evolution of \( M_{\text{vir}} \) and \( V_{\text{max}} \) (cf. Fig. 4.3), both properties can produce a good match to the distribution and late time growth of stellar mass at \( z=0 \) given a suitable choice for values of the free parameters of the physics function.
4.3.2 High redshift

In the previous section we demonstrated that our simple formation history model is capable of reproducing the observed red and blue galaxy stellar mass functions of the local Universe. We also showed that this is true independent of whether we utilise $M_{\text{vir}}$ or $V_{\text{max}}$ as the input variable of the physics function (Eqns. 4.3 & 4.4). However, as seen in Fig. 4.3, there are important differences in the time evolution of these halo properties. This suggests that we should see corresponding differences in the galaxy populations predicted at higher redshifts.

In Fig. 4.6 we present the high redshift stellar mass functions of both the $M_{\text{vir}}$ and $V_{\text{max}}$ models (dashed lines) against the observational data of Pérez-González et al. (2008) and Caputi et al. (2011) (data points). The solid lines represent the formation history model results after a convolution with a normally distributed random error of dispersion 0.3 dex for $z<3$ and 0.45 dex for redshifts greater than this value (Moster et al., 2012). Such a convolution is common practice and approximates the missing uncertainties in the observational data due to the full range of systematics involved with producing stellar mass estimates from high redshift galaxy observables (e.g. Fontanot et al., 2009a; Guo et al., 2011; Santini et al., 2012).

There are clear quantitative differences between the stellar mass functions produced by the two models, which become more pronounced as we move to higher and higher redshifts. At $z\approx4$ (lower right panel), the $V_{\text{max}}$ model predicts a sharp fall off in the number density of galaxies with stellar masses greater than $10^{10.2} M_\odot$. When using $M_{\text{vir}}$ to define the physics function, this drop off is far more gradual, resulting in a differing prediction in the number density of these galaxies by up to 2 orders of magnitude. Despite this, both versions of the physics function predict $z>0$ stellar mass functions which are too steep at high masses, although the addition of the random uncertainties (solid lines) largely alleviates this problem. The differences below the knee of the mass functions remain though. At these low masses, both models predict an increasing number of galaxies with decreasing redshift which is primarily attributed to growth through in-situ star formation. However, the $V_{\text{max}}$ model predicts that a large fraction of galaxies within the range $10^9$–$10^{10.2} M_\odot$ are already in place by $z=4$, with a correspondingly slower evolution to $z=0$.

Many of these differing qualitative predictions can be understood by considering the variations in the time evolution of $M_{\text{vir}}$ and $V_{\text{max}}$ as shown in Fig. 4.3. For example, the deficit of high stellar mass galaxies in the $V_{\text{max}}$ model at $z\approx4$ is due to the fact that the halos hosting these objects (those with the highest $M_{\text{vir}}$ and $V_{\text{max}}$ values at $z=0$) are first reliably identified in the merger trees with large $V_{\text{max}}$ values that already place
4.3. Results

Figure 4.6: The \( z \approx 0.9, 1.8, 3.3 \) and 3.9 stellar mass functions predicted by the formation history model when using \( M_{\text{vir}} \) (dashed black line) and \( V_{\text{max}} \) (dashed grey line) as the input halo property to the physics function (Eqns. 4.3, 4.4). Observational data from Pérez-González et al. (2008, PG08) and Caputi et al. (2011, Cap11) are shown for comparison. The solid lines show the results of convolving the model stellar masses with a normally distributed random uncertainty of 0.3 or 0.45 dex (for redshifts less than/greater than 0.3 respectively) in order to mimic the systematic uncertainties associated with the observed masses. Despite their close agreement at \( z = 0 \) (see Fig. 4.5) there are clear differences between the \( M_{\text{vir}} \) and \( V_{\text{max}} \) models at higher redshifts, associated with the different time evolution of these two halo properties (see Fig. 4.3).

them above the peak of the star forming band. The result is a reduced amount of in-situ star formation at early times, with effects that carry all the way through to \( z = 0 \) as these galaxies grow, predominantly through merging. We can similarly understand the cause of the larger predicted number density of high redshift intermediate mass galaxies in the \( V_{\text{max}} \) model. In this case, the lowest mass halos present at high redshifts have spent a longer time close to the peak of the efficient star forming band. This results in these halos already hosting significant amounts of stellar mass by \( z = 4 \) and is also the cause of the turnover at low galaxy masses.

To illustrate this further, in Fig. 4.7 we show the evolution of the mean stellar–halo mass relation for both models. The grey shaded regions represent the efficiencies predicted by the sub-halo abundance matching model of Moster et al. (2012). We have specifically
chosen to compare our results against the work of Moster et al. (2012), as they take their halo masses from the same dark matter merger trees as used in this chapter (as well as the higher resolution Millennium-II simulation) and also construct their model to match the same high-redshift stellar mass functions of Pérez-González et al. (2008). Hence, the grey shaded regions of Fig. 4.7 shows the evolution in the integrated stellar mass growth efficiency which our model must achieve in order to successfully replicate the observed stellar mass functions of Fig. 4.6.

By construction both the $M_{\text{vir}}$ and $V_{\text{max}}$ models produce extremely similar relations at $z=0$, but with clear differences at higher redshifts. It is these variations in the typical amount of stars formed within halos of a given mass that drives the different predictions for the evolution of the stellar mass function seen in Fig. 4.6. For example, the much higher average stellar mass content of low mass halos at $z=3$ when using the $V_{\text{max}}$ defined physics function is the cause of the increased normalisation of the low mass end of the relevant stellar mass function in Fig. 4.6.

Importantly, it can be seen from Fig. 4.7 that neither $M_{\text{vir}}$ or $V_{\text{max}}$ models reproduce the evolution of the stellar–halo mass relation found by Moster et al. (2012); in particular with respect to the position and normalisation of the peak value. The use of a redshift independent virial mass to define the peak in-situ star formation efficiency of the $M_{\text{vir}}$ model means that, by construction, there is no change in the position of the peak stellar–halo mass relation. However, due to the evolving $M_{\text{vir}}-V_{\text{max}}$ relationship, the position of the peak efficiency for the $V_{\text{max}}$ model does evolve, but in the direction opposite to that required. Based purely on this inability to reproduce the required evolution in the stellar–
4.3. Results

Figure 4.8: The evolution of the stellar–halo mass relation of central galaxies for a redshift evolving physics function defined in terms of $M_{\text{vir}}$ (Eqn. 4.3; black solid line) and $V_{\text{max}}$ (Eqn. 4.4; grey dashed line). Orange shaded regions indicated the subhalo abundance matching results of Moster et al. (2012). A comparison with Fig. 4.7 indicates that by shifting both the normalisation ($E_{M_{\text{vir}}}$, $E_{V_{\text{max}}}$) and position ($M_{\text{peak}}$, $V_{\text{peak}}$) of the physics functions, we are able to reproduce the correct evolution of the stellar–halo mass relation at high redshifts. This leads to a much better agreement between the observed and predicted stellar mass functions at $z>0$. (see Fig. 4.9).

halo mass relation, it is clear that there is little chance for our current, non-evolving physics functions to match the observed distribution of stellar masses in both the low and high redshift universe simultaneously, irrespective of the values of the available parameters.

4.3.3 Incorporating a redshift evolution

Although capable of reproducing the observed red and blue stellar mass functions at $z=0$, we showed in §4.3.2 that our simple formation history model struggles to reproduce the high redshift distribution of stellar masses. Importantly, we also concluded that it is difficult to imagine a combination of physics function parameters (see Eqns. 4.3 & 4.4) which could alleviate this discrepancy. In this section we therefore look to extend our simple model by introducing a redshift dependence into the physics function. In effect, this translates to the introduction of an evolution of the star formation efficiency with time for a fixed halo mass/maximum circular velocity. Such an evolution is well motivated by both theory and observations which suggest the presence of alternative/additional star formation mechanisms at high redshift when compared to those of the local Universe. For example, so called “cold-mode” accretion (Kereš et al., 2005; Brooks et al., 2009) is thought to be able to effectively fuel galaxies of massive halos at high redshift, allowing for increased star formation. In addition, the early universe was also a more dynamic place, with an enhanced prevalence of gas rich galaxy mergers and turbulence driven star formation due to violent disk instabilities (e.g. Dekel et al., 2009a; Wisnioski et al., 2011).
Motivated by the need to reproduce the evolving location and height of the stellar–halo mass relation found by Moster et al. (2012), we modify the physics function, $F_{\text{phys}}(M_{\text{vir}})$, of Eqn. 4.3 by introducing a simple power law dependence on redshift to the star formation efficiency normalization ($\mathcal{E}_{M_{\text{vir}}}$) and peak location ($M_{\text{peak}}$):

$$
\mathcal{E}_{M_{\text{vir}}}(z) = \mathcal{E}_{M_{\text{vir}}} \cdot (1+z)^{\alpha_{M_{\text{vir}}}}, \quad (4.6)
$$

$$
\log_{10}(M_{\text{peak}}(z)) = \log_{10}(M_{\text{peak}}) \cdot (1+z)^{\beta_{M_{\text{vir}}}}. \quad (4.7)
$$

The precise values of the redshift scalings have been chosen to provide the best reproduction of the Moster et al. (2012) stellar–halo mass relation and are given by $\alpha_{M_{\text{vir}}} = -0.5$ and $\beta_{M_{\text{vir}}} = -0.03$. We note that to achieve this we do not find it necessary to also vary the
remaining free physics function parameter, $\sigma_{M_{\text{vir}}}$. Also, due to the lower star formation efficiency with increasing look-back time produced by the above redshift dependencies, we find it necessary to increase the value of $E_{M_{\text{vir}}}$ slightly from its previous value of 0.6 to 0.7. This allows the model to correctly reproduce the normalisation of the blue stellar mass function at $z=0$.

We similarly modify the $V_{\text{max}}$ physics function, $F_{\text{phys}}(V_{\text{max}})$, by adding $1+z$ terms:

$$E_{V_{\text{max}}}(z) = E_{V_{\text{max}}} \cdot (1+z)^{\alpha_{V_{\text{max}}}}, \quad (4.8)$$

$$\log_{10}(V_{\text{peak}}(z)) = \log_{10}(V_{\text{peak}}) \cdot (1+z)^{\beta_{V_{\text{max}}}}. \quad (4.9)$$

The corresponding values of $\alpha_{V_{\text{max}}}$ and $\beta_{V_{\text{max}}}$ are -0.5 and 0.09 respectively (see Table 4.1). We note the need for a higher value of $\beta_{V_{\text{max}}}$, when compared to the virial mass equivalent. This is necessary to overcome the incorrect evolution of the stellar–halo mass relation peak which was already present with the redshift independent version of this function (see Fig. 4.7).

In Fig. 4.8, we present the stellar–halo mass relations of the $M_{\text{vir}}$ and $V_{\text{max}}$ models when incorporating the redshift dependent parameters. The grey shaded regions again indicate the results of Moster et al. (2012). By incorporating the redshift dependence we are now able to successfully reproduce the qualitative evolution of both the normalisation and peak position of the stellar–halo mass relation required at $z\geq1$. The effects of this on the predicted high redshift stellar mass functions of both the $M_{\text{vir}}$ and $V_{\text{max}}$ models can be seen in Fig. 4.9 where, as expected, we now find a far better agreement with the observations than was previously possible for a non-evolving physics function (c.f. Fig. 4.6). This is especially so if we consider only the observational data of P´erez-Gonz´alez et al. (2008) and also convolve our model with a random uncertainty as was done in §4.3.1.

### 4.4 Discussion

In §4.3.1, we demonstrated that our most basic, non-evolving form of the physics function is able to successfully reproduce the observed red and blue stellar mass functions of the local Universe (Fig. 4.5). This key result highlights the validity of our basic methodology and model construction. In addition, it reinforces the commonly held belief that the growth of galaxies is intrinsically linked to the growth of their host dark matter halos (White & Rees, 1978).

Although the level of agreement achieved with the observed $z=0$ colour split stellar mass functions is generally very good, there are some discrepancies. Most notably, there is
an over-prediction in the number density of the most massive red galaxies when using $M_{\text{vir}}$ as the input property to the physics function (left panel of Fig. 4.5). Our $z>0$ analysis suggests that this is at least partially due to an incorrect evolution of the stellar–halo mass relation with time (see Fig. 4.7). However, we also note that this excess is also a common feature of traditional semi-analytic galaxy formation models which similarly tie the evolution of galaxies to the masses of their host dark matter halos. In such models, a common method for overcoming this discrepancy is to evoke efficient feedback from active galactic nuclei in the most massive halos (Croton et al., 2006). This is already mimicked within the framework of our formation history model through the turnover at the high $M_{\text{vir}}/V_{\text{max}}$ end of the physics function, however a sharper cutoff may be required.

Other scenarios to explain the over-prediction in the number density of high mass galaxies include over-merging of satellites in the most massive halos (Klypin et al., 1999). Additionally, it has been suggested that galaxy mergers may result in a significant fraction of the in-falling satellite stellar mass being added to a diffuse intra-cluster light component instead of to the newly formed galaxy (Monaco et al., 2006; Conroy et al., 2007). The strength of this effect is expected to increase significantly with increasing halo mass (Behroozi et al., 2012). By simply adding a mass dependent amount of stellar mass to an ICL component during a merger event, our simple formation history model can again easily be adapted to explore these scenarios.

In this chapter we have deliberately restricted ourselves to considering only a very simple form of the physics function. This has allowed us to take advantage of the resulting transparency when interpreting our findings. However, we stress that the model can easily be extended to include arbitrary levels of physically motivated complexity.

For example, we have chosen to use a log-normal distribution to define the form of the physics function. Although being conceptually simple, the symmetric nature of this formalism implicitly assumes that the physical mechanisms responsible for quenching star formation in both low and high mass halos scales identically with halo mass (Fig. 4.3). This assumption has little physical justification and in order to provide the best results, it may be necessary to independently adjust the slope of the function at both low and high masses, and perhaps even as a function of redshift. In future work, we will address this issue by carrying out a full statistical analysis aimed at testing a number of different functional forms for both the physics and baryonic growth functions.

However, even within this reduced scope, we have learned a great deal from simply examining the high redshift stellar mass function predictions of the formation history model. In particular, we have highlighted the need for the physics function to produce an
evolution in the stellar–halo mass relation as a function of redshift in order to match the observed space density of massive galaxies at early times. For example, using \( V_{\text{max}} \) as the input property to the function introduces such an evolution, but in the wrong direction. Future improvements to the model could focus on finding a halo property that does evolve correctly with time and would thus be a more natural anchor of the physics function. This would avoid the need for artificially introducing an evolution to match the observations, as we have done here.

The need for an evolving stellar–halo mass relation is well established in the literature, however, the precise form with which this evolution manifests itself is less clear. The results of subhalo abundance matching studies, such as that of Moster et al. (2012) which we compare to in this chapter, are quite sensitive to the choice of input datasets and the technical aspects of the methodology. For example, Moster et al. (2012) find that the peak stellar–halo mass ratio increases from just 0.15% at \( z = 4 \) to 4% at \( z = 0 \), with a corresponding shift in position from a halo mass of \( 10^{12.5} \, M_\odot \) to \( 10^{11.8} \, M_\odot \). In contrast, an alternative study carried out by Behroozi et al. (2012) finds very little change in either the normalisation or peak location over a broad range in redshift. However, they do note a marked drop in the relation for the most massive halos at \( z \lesssim 2.5 \). This results in a qualitatively different prediction for the evolution of these massive halos, such that their efficiency of converting baryons into stars is higher as a function of look back time (the opposite trend to that found by Moster et al. (2012)).

The required evolution of the stellar–halo mass relation also agrees with our findings in Chapter 2. Here it was suggested that the semi-analytic galaxy formation model of Croton et al. (2006) requires a preferential boost to the star formation efficiency of high mass galaxies at early times in order to replicate the observed redshift evolution of the stellar mass function. Such a boost is analogous to shifting the peak of the stellar–halo mass relation to higher halo masses, as we have found to be necessary for our simple formation history model. This further demonstrates the utility of this model, which allows us to arrive at the same qualitative results as the more complex semi-analytics, but with fewer assumptions about poorly understood physics and fewer free parameters.

Finally, we note that the simplicity of our formation history model results in it being extremely fast and computationally inexpensive when compared to traditional semi-analytic models. This allows it to be straightforwardly calibrated against a number of observed relations using statistical techniques such as the Monte-Carlo Markov Chain (MCMC) methods. As demonstrated in Chapter 2, this would allow us to assess the degeneracies between the free model parameters as well as quantitatively evaluate its ability to repli-
cate the relevant observations and place uncertainties on all of the predictive quantities it produces. Most importantly though, by carrying out this procedure the model can then be used to provide statistically accurate (against select observations) mock catalogues for use by large surveys. Further to what can be achieved using current Halo Occupation Distribution or sub-halo abundance matching methods, catalogues produced using our model include both full growth histories and star formation rate information for each individual galaxy, with no need to add any artificial scatter to approximate variations in formation histories. In addition, the direct and clear dependence of the model on the halo properties of the input dark matter merger trees, makes it an ideal tool for investigating a number of additional topics. Examples include: comparing the effects of variations between different N-body simulations and halo finders on the physics of galaxy formation and evolution; investigating the predictions of simple monolithic collapse scenarios; contrasting various mass dependent merger star burst models; and exploring the ramifications of N-body simulations run with alternative theories of gravity.

4.5 Conclusions

In this chapter we introduce a simple model for self-consistently connecting the growth of galaxies to the formation history of their host dark matter halos. This is achieved by directly tying the time averaged change in mass of a halo to the star formation rate of its galaxy via two simple functions: the “baryonic growth function” and the “physics function” (Eqns. 4.2,4.3). Closely related to the model of Bouche et al. (2010) in terms of its basic methodology, our model has a number of innovative features. Firstly, we utilise N-body dark matter merger trees to provide self consistent growth histories of individual halos that naturally includes scatter due to varying formation histories. Secondly, we implement a single, unified physics function which encapsulates the effects of all of the intertwined baryonic processes associated with galactic star formation and condenses them down into a simple mapping between star formation efficiency and dark matter halo mass growth.

As well as introducing this new model, we demonstrate its ability to replicate important observed relations such as the galactic stellar mass function, as well as illustrate some examples of its potential for investigating different theories of galaxy formation and evolution. Our main results can be summarised as follows:

• Motivated by the observed suppression of star formation efficiency in both the most massive and least massive dark matter halos we begin by parametrising the physics
4.5. **Conclusions**

function as a simple, non-evolving, log-normal distribution with a single independent variable of either halo virial mass, $M_{\text{vir}}$, or maximum circular velocity, $V_{\text{max}}$ (Fig. 4.3).

- Our use of N-body merger trees provides the model with the ability to calculate self-consistent star formation rates for individual objects, and thus provide predictions for secondary properties such as galaxy colour. With just three free parameters controlling the position, normalisation and dispersion of the peak star formation efficiency, we show that our simple model can successfully reproduce the observed red and blue stellar mass functions at redshift zero. Assuming a suitable choice of the parameters, this result is independent of the use of $M_{\text{vir}}$ or $V_{\text{max}}$ as the input property to the physics function (Fig. 4.5).

- For the purposes of replicating the stellar mass functions at high redshift, we find our simple model to be inadequate. This is due to an inability to produce the correct evolution of the stellar–halo mass relation with time (Figs. 4.6 & 4.7).

- We therefore investigate the use of redshift as a second input variable to the physics function in order to control the position and normalisation of the peak star formation efficiency with time. Using this simple adaption alone, the formation history model is able to correctly recover the observed high redshift stellar mass functions (Figs. 4.8 & 4.9).

In order to demonstrate its construction and utility we have presented one of the simplest forms of this new formation history model. However, a fundamental strength of the model is that it can be easily extended to arbitrary levels of complexity in order to investigate a whole host of physical processes associated with galaxy formation and evolution, just a few of which we have outlined in §4.4. In future work we will investigate the predictions made when using alternative forms of the baryonic growth and physics functions. We will also statistically calibrate the free parameters to simultaneously reproduce a host of observed relations at multiple epochs.
5

Summary and forward look

5.1 Summary

A criticism often leveled at semi-analytic models is that they possess too many free parameters, some of which may be poorly constrained by observation or theory. However, in reality, the number and freedom of these parameters is merely a reflection of the complexity of the many included processes and our incomplete knowledge of the relevant physics. In fact, by finding the parameter values that provide the most accurate reproduction of the observed universe, we can learn a great deal about galaxy formation. In Chapter 2 we rigorously explore the available parameter space of the Croton et al. (2006) semi-analytic model using statistical Monte-Carlo Markov Chain methods. By demanding that the model reproduce both the $z=0$ stellar mass function and black hole–bulge relation we are able to put strong constraints on the allowed values of all but one of the free parameters of interest. However, when further constraining the model against the observed stellar mass function at $z=0.83$, we find that a number of tensions in the model prescriptions emerge. In particular, it is suggested that the implementation of supernova feedback may be incomplete and that the addition of some mechanism for preferentially enhancing star formation in the most massive galaxies at high redshift is necessary.

The ability of semi-analytic models to provide large, cosmologically significant samples of galaxies with individual growth histories has meant that these models are well suited to exploring the evolution of statistical properties, especially when applied to the analysis of survey science data. However, in Chapter 3 we demonstrate an alternative, novel use for semi-analytic models as a tool to understand the cosmological context of individual galaxies. In particular we focus on the current evolutionary status of the Milky Way and M31, as measured by their global colours. Using the best available colour measurements
in the literature for these two objects, along with SDSS and Galaxy Zoo observational samples, we show that both may be classified as “green valley” members. Such galaxies are commonly thought to be undergoing an important evolutionary transition from typical blue star forming spirals to red, quiescent ellipticals. We further reinforce this classification by comparing the star formation rates of these two galaxies to a sample of analogues drawn from the semi-analytic model. Finally, using the growth histories of our analogue galaxies we find that, in the model, the cessation of star formation is driven by a regulation of the supply of fresh cold gas due to low level feedback from the central massive black holes. Such processes could be the target of future observing programs (see, for example, Guo & Mathews, 2012). Given the typical timescales for galaxies to pass through the green valley region we also suggest that both the Milky Way and M31 will be largely devoid of active star formation well before the time when they are eventually expected to merge together. Our finding that both our own galaxy and nearest massive neighbour may be classified as these relatively rare transition objects has clear implications for their common use as templates with which to interpret our observations of other galaxies.

The relative simplicity of semi-analytics is a key attribute which greatly aids our analysis in Chapter 3, allowing us to easily identify the particular physical mechanisms which are most important in reducing the level of star formation in our Milky Way analogue galaxies. However, the current trend for semi-analytic models is to become increasingly more complex as we include more physical processes and refine our knowledge of current ones. While this may be viewed as a natural evolution, the addition of further parameterisations and complexity can confuse the interpretation of the models. Hence there is much to be gained by stepping back and keeping the models as simple as possible. Motivated by this, in Chapter 4 we introduce a new “formation history” model for galaxy evolution. Instead of attempting to treat each individual baryonic process in a physically motivated fashion, here we self-consistently tie the growth of galaxy stellar mass to the growth history of its host halo using two simple, phenomenologically motivated equations. The key advantage over other related methodologies such as HOD or SHAM models is the ability to self-consistently track the growth histories of individual galaxies. A simple calibration by eye demonstrates that our formation history model is able to reproduce the observed red and blue stellar mass functions at $z=0$. We highlight that this success is largely independent of whether virial mass or maximum circular velocity is used to describe host halo growth. However, we show that in order to recover the form of the global stellar mass function out to at least $z=4$, we must introduce a simple redshift evolution to the model parameterisations that shifts the peak of the star formation efficiency to higher halo
masses at high redshift. This result reinforces our findings in Chapter 2. Finally, we note that the transparency and extensibility afforded by our formation history model makes it ideally suited to exploring different evolutionary scenarios of galaxy growth out to high redshift. A number of examples are outlined in the discussion section of Chapter 4, however, further detail on one particular extension of the model to investigate the evolution of the quasar luminosity function is given below.

5.2 Future directions

The work carried out in this thesis opens up a number of possible avenues for future exploration. In this final section we discuss a small selection of these, as well as highlight some key areas of future research for semi-analytic models in general.

Statistically calibrating semi-analytic models

In Chapter 2 we used Monte-Carlo Markov Chains to statistically calibrate the semi-analytic model of Croton et al. (2006). In future this technique will likely become a standard tool for all semi-analytic models. Not only does it allow one to be sure that the best possible results are being achieved, but by sampling the posterior distribution of the fitting analysis one can also provide statistical uncertainties on all of the predictive quantities produced.

Although our work constitutes the first time a semi-analytic model has been robustly constrained at multiple epochs simultaneously, we only utilise two observational constraints: the stellar mass function and the black hole–bulge relation. An obvious next step will be to attempt to constrain against a larger number of observations. Possibilities include the Tully–Fisher relation, star formation rate density evolution, colour fractions as a function of stellar mass and galaxy clustering. As we add more and more constraints the models will be pushed to their limits, allowing us to identify further ways in which to improve their physical prescriptions.

The evolution of the quasar luminosity function

In Chapter 4 we introduced our new “formation history” model of galaxy evolution. In order to demonstrate the construction and general utility of this new model, we focused only on its most basic form. However, there are many ways in which it can be extended to investigate a broad variety of topics in galaxy evolution. One such example which we are actively pursuing is the evolution of the quasar luminosity function out to high redshift.
Figure 5.1: Preliminary results from of a new semi-analytic disk model which we are currently developing. Each panel shows the surface density profiles of (from top-to-bottom) stellar mass, atomic cold gas, molecular cold gas and total cold gas. Note that bulges are not included in these profiles. The thick grey lines with error bars show the nearby spiral galaxies from the observational sample of (Leroy et al., 2008) with $200 \leq V_{\text{vir}}/(\text{km s}^{-1}) \leq 235$. The red lines show the mean results from a similarly selected sample of galaxies in the new, and as yet uncalibrated, disk model. The blue shaded region indicates the corresponding $1\sigma$ uncertainty of the model results. The thin grey vertical lines show the disk annuli outer edges for the largest of the averaged model galaxies in the sample.

Recent observational studies have suggested that there is a direct correlation between quasar luminosity and the star formation rate of its host galaxy (Mullaney et al., 2012). By coupling this relation to a simple quasar light curve model, we will be able to produce quasar luminosity functions using the formation history model. After statistically calibrating the model using our MCMC tools to match the observed stellar mass/luminosity functions out to high redshift, we will then be able to make full predictions, with appropriate uncertainties, for the evolution of the quasar luminosity function.

Spatially decomposed galaxy properties

One of the shortcomings of current semi-analytic models is that they generally do not provide any 3-dimensional information for galaxies, other than their real-space positions and
velocities. However, recent attempts have been made to incorporate simple disk models (Fu et al., 2010), thus allowing the calculation of mass and light profiles for each galaxy. We are currently developing a new radial disk model that builds upon these previous works and includes a number of technical improvements to directly address their shortcomings. Such a model opens up previously unexplored possibilities for semi-analytics to test theories for new categories of galaxy properties, as well as to make predictions for next generation surveys. For example, combined with a physical prescription for determining the fraction of cold gas in different components, it could be used to make self-consistent predictions for atomic and molecular gas fractions, as well as equivalent source sizes and radial profiles (see Fig. 5.1), in order to address the science goals of upcoming radio survey instruments such as the Square Kilometre Array (SKA) and it’s pathfinders. Additionally, these radial mass profiles can be used to aid in interpreting the results of new instruments such as the Sydney-AAO Multi-object Integral Field Spectrograph (SAMI) that will survey thousands of galaxies with spatial decomposed properties.

The high redshift Universe

New galaxy surveys such as the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS) are now allowing us to obtain a statistically complete census of galaxies down to $10^9 M_\odot$ at redshifts extending out to $z \sim 2$ (Grogin et al., 2011). Being able to accurately model the growth and evolution of these high redshift galaxies will be an important challenge for semi-analytic models in the coming years. The parameterisations of many of the implemented physical processes in current generations of models tend to be calibrated to the low redshift Universe, with the assumption that these scalings also hold true at early times. A prime example is the star formation prescription which often assumes an axisymmetric, exponential thin disk for which a Kennicutt–Schmidt type star formation law holds. This is unlikely to be true in the early Universe due to the dynamic and evolving nature of this epoch, where disturbed morphologies and multiple galaxy interactions were common. We may find that in order to be successful at reproducing the properties and distributions of high redshift galaxies, we must include new physical prescriptions or make non-trivial adaptations to current ones.

Although faithfully reproducing the high redshift galaxy population will be a challenge, the potential insights which semi-analytic models can provide into the evolution of the early Universe, and the resulting effects at later times, is significant. For example, semi-analytics have already been used to help understand how the reionisation of the inter-galactic medium by early galaxies alters the galactic luminosity function at $z=0$ (e.g.
Benson et al., 2002). They have also been used to investigate the structure of reionisation and its dependence on the strength of supernova feedback (Kim et al., 2012). As a member of the new Dark-ages Reionisation And Galaxy-formation Simulation (DRAGONS) project, I will be furthering this work by directly coupling specially adapted semi-analytic models with semi-numerical schemes for simulating the ionisation structure of the early Universe. This new hybrid model will run on the dark matter merger trees constructed from specially designed, high temporal resolution N-body simulations, and will possess new physical prescriptions that are directly informed by detailed hydrodynamical simulations of individual high redshift galaxies.
Allende Prieto C. et al., 2008, Astronomische Nachrichten, 329, 1018
Baugh C. M., 2006, Reports on Progress in Physics, 69, 3101


Dekel A. et al., 2009a, Nature, 457, 451
Einasto J., 1965, Trudy Astrofizicheskogo Instituta Alma-Ata, 5, 87


Hastings W. K., 1970, Biometrika, 57, 97
Johnston S. et al., 2008, Experimental Astronomy, 22, 151
Komatsu E. et al., 2011, ApJS, 192, 18
Minniti D. et al., 2010, New A, 15, 433


Peebles P. J. E., 1980, The large-scale structure of the universe


Springel V. et al., 2005, Nature, 435, 629
Bibliography


Trotta R., 2008, Contemporary Physics, 49, 71


Zwicky F., 1933, Helvetica Physica Acta, 6, 110
The following publications contain material presented in this thesis:

**Mutch, Simon J., Croton, Darren J., Poole, Gregory B.**

*The Mid-life Crisis of the Milky Way and M31*

The Astrophysical Journal, Volume 736, Issue 2, article id. 84 (2011)

**Mutch, Simon J., Poole, Gregory B., Croton, Darren J.**

*Constraining the last 7 billion years of galaxy evolution in semi-analytic models*