This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.
Gas in and around galaxies in the Simba simulations

Sarah Ceridwen Appleby

Doctor of Philosophy
The University of Edinburgh
January 2023
Galaxy evolution is an interplay of physical processes across a wide range of cosmological size scales, from star formation and black hole growth on sub-pc scales, to interactions with their environments on Mpc scales. The region surrounding individual galaxies within their host halos is known as the Circumgalactic Medium (CGM), and connects galaxies to the wider Intergalactic Medium (IGM). The CGM acts as both a reservoir of inflowing gas and a repository for outflowing gas as it is transported outwards via galactic feedback. Quasar absorption line surveys have revealed the CGM to be a complex, multiphase environment in which ions tracing cool and hot gas are present within the same absorber systems.

Cosmological simulations are a powerful tool for building a theoretical understanding of these interconnected processes and resulting observations. They provide a 3D view of galaxies and their host halos, in which we intrinsically know the physical conditions of the constituent dark matter and baryons. As such, simulations are useful both for testing galaxy evolution models against the real universe, and for interpreting complex observations in a physical context.

In this thesis I present detailed comparisons of simulations to observations of star formation and quenching in galaxies, and of absorbers in galactic halos in an effort to constrain physical models. I use SIMBA, a suite of state of the art cosmological simulations with realistic sub-grid physical models based on high resolution zoom simulations, including novel treatments of dust evolution and of black hole growth and feedback. I investigate the nature of gas in galaxies and their halos, and show comparisons with observations. In addition, I isolate the impacts of galactic feedback and ionising radiation on the results by varying elements of the input physical models.

In Chapter 3, I present radial profiles of star forming and green valley galaxies
in \textit{Simba} at $z = 0$ and compare with observations of MaNGA spatially resolved spectroscopy. My analysis shows strong central depressions in star formation rate (SFR), specific SFR (sSFR), and gas fraction in \textit{Simba}’s green valley galaxies and massive star-forming systems, qualitatively as observed, owing to Active Galactic Nuclei (AGN) X-ray feedback, which pushes central gas radially outwards. These effects are less pronounced at higher redshift, owing to less powerful AGN feedback. I also present a comparison of galaxy half light radii to observations at various redshifts, finding that predictions for star-forming galaxy sizes are accurate but that quenched galaxies are impacted by numerical heating.

In Chapter 4, I characterise the low redshift CGM in \textit{Simba} and examine the impact of differing feedback prescriptions. Low mass galaxies live in multiphase, diverse halos, whereas high mass galaxies live in halos dominated by hot gas. Halo baryon fractions are generally $\leq 50\%$ of the cosmic fraction due to stellar feedback at low masses, and ‘jet-mode’ AGN feedback at high masses. I present a comparison of absorption statistics to the COS-Halos and COS-Dwarfs surveys: \textit{Simba} reproduces HI absorption well around star forming galaxies and broadly reproduces absorption of selected metal lines, including the observed dichotomy in OVI absorption between star forming and quenched galaxies. These predictions of metal line absorption are sensitive to the choice of photo-ionising background. Varying the galactic feedback prescription shows that stellar feedback enriches the CGM, while AGN feedback sets the gas temperature phase.

In Chapter 5 I extend this analysis with a theoretical perspective of the CGM in \textit{Simba} using a sample of CGM absorbers from a representative sample of \textit{Simba} galaxies, selected across a range of stellar masses and sSFR. Absorbers are more abundant around low mass, star forming galaxies; the CGM of green valley galaxies is more similar to that of quenched galaxies. The absorber overdensity depends on the sSFR of the galaxy, while the absorber temperature is set by galaxy stellar mass. Absorption from low ions arises from cold, dense CGM gas, whilst absorption from higher ions traces hotter, more diffuse gas. I also examine the proportion of absorption arising from collisionally ionised gas, and from gas that is associated with satellite galaxies.

Finally, in Chapter 6 I develop a tree-based machine learning mapping between the observable properties and the underlying physical conditions of the absorbers in the CGM absorber sample from Chapter 5. Such a mapping has the potential to be used to interpret CGM observations within a full cosmological context, assuming the \textit{Simba} galaxy evolution model.
Galaxies are complex, evolving systems of dark matter, stars, gas, dust, black holes, and planet systems. The stars that we see in the night sky are formed from collapsing clouds of gas that are cold and dense, so in order to continue producing stars, galaxies require a supply of cold gas. If the gas supply runs out, is not cold or dense enough, or is removed from the galaxy then star formation will cease.

Galaxies can grow by pulling in new gas from their surroundings in a process known as accretion. Gas flows into the galaxy either from the spiderweb-shaped network of gas that connects galaxies, or they can pull gas out from less massive galaxies if they pass near one another. On the flip side, galaxies can also lose their gas as a result of energetic processes that occur within galaxies. The two main drivers of this outflowing gas (known as ‘feedback’) are 1) supernova explosions at the end of stars’ lives and 2) the release of energy from material falling onto a supermassive black hole (which are thought to live at the centers of all galaxies). Galaxy evolution can therefore be seen as a balance between the gas that is flowing towards and away from galaxies. When galaxies acquire new gas it can be used to form new stars, but the feedback from supernova and black holes can drive gas away, or heat it such that it cannot be used to form stars.

This thesis is concerned with questions about the gas inside galaxies (known as the interstellar medium, ISM) and gas that is immediately outside galaxies (the circumgalactic medium, CGM). The ISM is the site of star formation inside galaxies, while CGM is of great interest as any gas flows entering or exiting galaxies need to pass through it. The ISM and CGM both hold the evidence of past and current accretion and feedback. By understanding gas in and around galaxies, we can learn about the processes that are critical for galaxy evolution, and therefore explain how modern day galaxies such as our Milky Way evolved from the Big Bang to the present day.
Galaxy simulations are an important tool for understanding these phenomena as they allow us to trace the flows of gas in and out of galaxies over the history of the universe. Simulations are produced by writing computer programs for all of the important physical laws and processes of nature, dividing the raw ingredients from the beginning of the universe into small pieces, and running the physical laws forward in time until the present day to form galaxies like our own. These synthetic galaxies are compared to real ones to assess the accuracy of the model. Similarities represent successes, and discrepancies highlight areas where the models need improvement. Comparing the gas in and around the simulated galaxies to that of real galaxies is a key test of these models. Over the last few decades, the field has evolved through trial and error; nowadays simulations can produce galaxies with realistic properties and structure.

In this thesis I used SIMBA, a group of computer simulations of large portions of the universe, containing thousands of synthetic galaxies. The SIMBA simulations are the next generation of the MUFASA simulations and were run on the DiRAC supercomputer in Durham using nearly two million hours of computing time, a process which took two months. SIMBA has a method for modelling the growth and feedback from black holes which is unique amongst simulations of this type. The main SIMBA simulation represents a portion of the universe over a billion trillion kilometers across and contains over 55000 galaxies.

I compared the properties of the gas inside and outside of synthetic galaxies from SIMBA with observations of real galaxies, and tested the performance of the SIMBA galaxy formation model. An important advantage of simulations is the ability to turn different processes off and on, which allows us to isolate their individual effects on the galaxies. I used this technique to measure the effect of feedback from supernovae and black holes on the ISM and CGM gas. I investigated the properties of the gas that lives around galaxies, and develop a method of predicting the gas properties from real observations assuming the SIMBA model. I showed that the SIMBA model can reproduce the flows of gas in and around galaxies well, an important success of the model. Interesting discrepancies do however exist between the model and the real universe which point to exciting areas for next generation simulations to improve on.
Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

Chapters 3, 4, and 5 are based on work published in Appleby et al. (2020), Appleby et al. (2021), and Appleby et al. (2023a) respectively. Chapter 6 is based on work that has been submitted to the Monthly Notices of the Royal Astronomical Society, and is currently under review (Appleby et al., 2023b). In these chapters, I use ‘we’ to acknowledge that this work was performed under the guidance of my supervisor and with the help of collaborators. Nevertheless, all work stems from first-author papers.

(Sarah Ceridwen Appleby, January 2023)
Acknowledgements

This thesis has been made possible through the support and encouragement of many friends and collaborators, who have made my PhD a thoroughly enjoyable experience.

First and foremost I would like to thank my brilliant supervisor Romeel Davé for all of his encouragement, guidance and wisdom. You have been endlessly patient with me and provided me with so many incredible opportunities. Your contagious enthusiasm for the science never fails to renew my motivation. The last few weeks finishing up the thesis have been mostly stress-free, which I attribute to your excellent and thoughtful supervision.

I would also like to thank several wonderful mentors whose support has been invaluable: my second supervisor Daniele Sorini, for the scientific insights, guidance, and expertise; Daniel Anglés-Alcázar, for his generosity and enthusiasm (and for hosting us in New York); Katarina Kraljic, whose emotional and technical support was crucial for the first paper and in general; finally Britton Smith, who was a fantastic summer project supervisor and who was the first person who suggested that I pursue a PhD.

Next, I would like to thank my examiners Freeke van de Voort, Trent Dupoy, and Beth Biller for their careful reading of the thesis, their thoughtful questions, and for making the viva surprisingly relaxed and fun.

I have been privileged at Edinburgh to be surrounded by a cohort of delightful and accomplished fellow PhD students: Andrea, Emma, Freddy, Harry, James, Manika, Marios, Massi (my colour palette sounding board), Phil, Rob (who just carried my desktop up to the East Tower), Rokas and others. You create such a friendly atmosphere in the department and have made this experience a real pleasure.

Next, the esteemed Drs. Dubber, Kondapally, Robson and Thomas for making me laugh so much and keeping me going over the last 4 years. I love you all and thank you for fixing my many \LaTeX{} problems. Double mention for Sophie, my day-1 PhD buddy - the IfA has not been the same since you left.

To Dr. Luciana Sinpetru: thank you for being a supportive friend, helping me
have a good work-life balance, and repeatedly watching my dog.

To my sweet dog Josy, my best wee pal, who sat and watched me write this thesis: thank you for the cuddles and for forcing me to go outside at regular intervals.

Lastly and most importantly, my amazing partner Laura, who now knows more than she ever wanted to about galaxies. You have been so patient and caring with me over the last few months and the last 11 years. I could not ask for a more wonderful partner. I love you.
## Contents

Abstract .............................. ii

Lay Summary ......................... iv

Declaration .......................... vi

Acknowledgements ................... vii

Contents ............................ ix

List of Figures ....................... xiv

List of Tables ....................... xviii

### 1 Introduction

1.1 Galaxies ........................................... 1

1.1.1 How are galaxies made? ....................... 1

1.1.2 How do galaxies get their gas? .................. 3

1.1.3 Gas inside galaxies ......................... 6

1.1.4 How do galaxies lose their gas? ............... 7

1.1.5 The Baryon Cycle ......................... 11

1.2 Observing gaseous halos ................. 12

1.2.1 Emission mapping ....................... 12
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2.2 Absorption lines</td>
<td>12</td>
</tr>
<tr>
<td>1.2.3 The multiphase CGM</td>
<td>14</td>
</tr>
<tr>
<td>1.3 Simulating gas in and around galaxies</td>
<td>16</td>
</tr>
<tr>
<td>1.3.1 Methods of cosmological simulations</td>
<td>17</td>
</tr>
<tr>
<td>1.3.2 Testing simulations with the CGM</td>
<td>23</td>
</tr>
<tr>
<td>1.4 Machine learning in Astronomy</td>
<td>23</td>
</tr>
<tr>
<td>1.4.1 Machine learning methods</td>
<td>23</td>
</tr>
<tr>
<td>1.4.2 Applications to astronomical problems</td>
<td>25</td>
</tr>
<tr>
<td>1.5 Thesis Outline</td>
<td>26</td>
</tr>
<tr>
<td>2 The Simba Simulations</td>
<td>28</td>
</tr>
<tr>
<td>3 The Impact of Quenching on Galaxy Profiles in Simba</td>
<td>37</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>37</td>
</tr>
<tr>
<td>3.2 Kennicutt-Schmidt relation</td>
<td>41</td>
</tr>
<tr>
<td>3.3 Size-mass relation</td>
<td>43</td>
</tr>
<tr>
<td>3.4 Radial Profiles</td>
<td>49</td>
</tr>
<tr>
<td>3.4.1 Galaxy selection</td>
<td>49</td>
</tr>
<tr>
<td>3.4.2 Star-forming vs GV profiles</td>
<td>52</td>
</tr>
<tr>
<td>3.4.3 Comparison to observations</td>
<td>58</td>
</tr>
<tr>
<td>3.4.4 Black hole feedback dependence</td>
<td>60</td>
</tr>
<tr>
<td>3.4.5 Satellites vs. centrals</td>
<td>66</td>
</tr>
<tr>
<td>3.4.6 Redshift dependence</td>
<td>70</td>
</tr>
<tr>
<td>3.4.7 Numerical convergence</td>
<td>73</td>
</tr>
<tr>
<td>3.5 Summary</td>
<td>75</td>
</tr>
</tbody>
</table>
4 The Low Redshift Circumgalactic Medium in Simba

4.1 Introduction ............................................................................ 79
4.2 CGM Mass and Metal Budget...................................................... 82
  4.2.1 Classifying CGM Gas ....................................................... 82
  4.2.2 Mass Budget .................................................................. 84
  4.2.3 Metal Budget ................................................................. 89
4.3 Comparison to COS-Halos and COS-Dwarfs................................. 93
  4.3.1 Mocking the COS surveys.................................................. 93
  4.3.2 Comparison to COS-Halos and COS-Dwarfs ...................... 99
  4.3.3 Summary of observational comparisons................................. 108
4.4 Dependence on feedback models ................................................... 109
  4.4.1 Halo baryon fractions ....................................................... 110
  4.4.2 CGM metallicities............................................................ 111
  4.4.3 CGM absorption ............................................................. 113
4.5 Conclusions ............................................................................. 115
4.6 Physical impact parameter .......................................................... 118
4.7 Resolution and volume convergence............................................... 120

5 The Physical Nature of Circumgalactic Medium Absorbers in Simba 123
5.1 Introduction ............................................................................ 123
5.2 Absorber sample ....................................................................... 127
  5.2.1 Galaxy selection ............................................................ 127
  5.2.2 Spectral generation .......................................................... 128
  5.2.3 Spectrum fitting .............................................................. 133
6 Mapping Circumgalactic Medium Observations to Theory Using Machine Learning

6.1 Introduction ............................................................................ 172

6.2 Random Forest Methods............................................................. 175
  6.2.1 Absorber sample .............................................................. 175
  6.2.2 Input and target features................................................... 175
  6.2.3 Training ........................................................................ 176

6.3 Predictive accuracy ................................................................. 178

6.4 Feature importance ............................................................... 181
  6.4.1 RF model importance...................................................... 181
  6.4.2 Change in predictive accuracy ............................................ 184

6.5 Phase space ........................................................................... 185

6.6 Conclusions ........................................................................... 189
7 Conclusions and Future Work

7.1 Summary of Results

7.1.1 The Impact of Quenching on Galaxy Profiles in SIMBA

7.1.2 The Low Redshift Circumgalactic Medium in SIMBA

7.1.3 The Physical Nature of Circumgalactic Medium Absorbers in SIMBA

7.1.4 Mapping Circumgalactic Medium Observations to Theory Using Machine Learning

7.2 Future Prospects

7.2.1 This thesis

7.2.2 The field

Bibliography
List of Figures

1.1 Gas consumption timescales, showing the need for new accretion...... 4
1.2 Galaxy stellar mass functions, showing growth regulation by galactic feedback................................................................. 8
1.3 Multiphase nature of CGM revealed by absorption line observations.. 15
1.4 How simulations of different size scales inform one another.............. 17

2.1 SIMBA comparisons to selected observations ................................ 35

3.1 Surface density of gas as a function of star formation rate surface density for star forming and green valley galaxies in Simba .......... 42
3.2 Galaxy size-mass relation in SIMBA at z = 2,1,0, with comparison to observations................................................................. 44
3.3 Evolution of V band half-light radii for star forming and quenched galaxies.................................................................................. 47
3.4 Selection of star forming and green valley galaxies in SFR-$M_*$ space.. 50
3.5 Radial profiles of SFR, sSFR, H1, H2, SFE and $f_{H2}$ as a function of stellar mass, for star forming and green valley galaxies ............... 54
3.6 Comparison of sSFR radial profiles of star forming and green valley galaxies to MaNGA SDSS (Belfiore et al., 2018) data ................. 57
3.7 Effect of jet-mode AGN feedback on the sSFR radial profiles of star forming and green valley galaxies ........................................ 61
3.8 Effect of X-ray mode AGN feedback on the sSFR radial profiles of star forming and green valley galaxies ........................................ 62
### 3.9 Comparison of gas and star formation profiles for central and satellite galaxies

<table>
<thead>
<tr>
<th>3.10 Radial sSFR profiles for star forming galaxies at $z = 2$</th>
<th>71</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.11 Resolution dependence of sSFR radial profiles for star forming and green valley galaxies</td>
<td>73</td>
</tr>
</tbody>
</table>

### 4.1 Mass budget in various halo components at $z = 0$ as a function of stellar mass

<table>
<thead>
<tr>
<th>4.2 Metal budget in various halo components at $z = 0$ as a function of stellar mass</th>
<th>89</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.3 Metallicity of various halo components at $z = 0$, separated by galaxy sSFR</td>
<td>91</td>
</tr>
<tr>
<td>4.4 Selection of analogous COS-Halos (Tumlinson et al., 2013) and COS-Dwarfs (Bordoloi et al., 2014) galaxy samples within SIMBA</td>
<td>94</td>
</tr>
</tbody>
</table>

### 4.5 Comparison of equivalent width profiles of H$\alpha$ and metal lines around SIMBA galaxies with COS-Halos (Tumlinson et al., 2013) and COS-Dwarfs (Bordoloi et al., 2014) observations

<table>
<thead>
<tr>
<th>4.6 Comparison of $f_{\text{cov}}$ profiles of H$\alpha$ and metal lines around simba galaxies with observations</th>
<th>105</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.7 Comparison of total path absorption profiles of H$\alpha$ and metal lines around simba galaxies with observations</td>
<td>107</td>
</tr>
<tr>
<td>4.8 Effect of stellar and AGN feedback on halo mass budgets, as a function of halo mass</td>
<td>110</td>
</tr>
<tr>
<td>4.9 Effect of stellar and AGN feedback on the metallicities of halo components, as a function of halo mass</td>
<td>112</td>
</tr>
</tbody>
</table>

### 4.10 Effect of stellar and AGN feedback on total path absorption profiles in the CGM

<table>
<thead>
<tr>
<th>4.11 Comparison of equivalent width profiles of H$\alpha$ and metal lines in the CGM, on a physical kpc scale</th>
<th>119</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.12 Resolution dependence of equivalent width profiles of H$\alpha$ and metal lines</td>
<td>121</td>
</tr>
</tbody>
</table>

### 5.1 Selection of representative galaxy sample across $M_\star$-sSFR space

| 129 |
5.2 Example CGM absorption spectrum of a galaxy at \( z = 0 \) .................................. 132
5.3 CGM column density distribution function (CDDF) for H\(_{i}\) and metal lines, separated by galaxy sSFR .......................................................... 137
5.4 CGM CDDF for H\(_{i}\) and metal lines, separated by galaxy \( M_\star \) ........... 140
5.5 Histograms of CGM absorber overdensity (separated by galaxy sSFR) and temperature (separated by galaxy \( M_\star \)) ........................................ 143
5.6 Overdensity-temperature-column density phase space for CGM absorbers .......................................................... 146
5.7 Proportion of absorbers and absorption arising from different regions of phase space .......................................................... 149
5.8 Proportion of absorption arising from collisionally ionised gas, as a function of impact parameter .......................... 152
5.9 CGM absorber overdensity against column density, separated by galaxy sSFR .......................... 155
5.10 Column density-weighted median physical conditions of absorbing CGM gas .......................................................... 158
5.11 Redshift dependence of the CGM CDDF .............................................. 160
5.12 Redshift dependence of column density-weighted median conditions of CGM absorbers .......................................................... 162
5.13 Proportion of absorption arising from gas that is associated with satellite galaxies, as a function of impact parameter ............. 164
5.14 Comparing column densities derived from spectrum fitting and from summing along the LOS .......................................................... 169
5.15 Comparing equivalent widths derived from spectrum fitting and from summing along the LOS .......................................................... 170

6.1 Physical conditions for H\(_{i}\) absorbers, predicted by RF models .......... 178
6.2 Physical conditions for C\(_{\text{II}}\) absorbers, predicted by RF models ........ 180
6.3 Physical conditions for C\(_{\text{IV}}\) absorbers, predicted by RF models ...... 180

xvi
6.4 Importance values for input features in the RF models for H\textsubscript{i} physical conditions................................................................. 182

6.5 Importance values for input features in the RF models for C\textsubscript{II} physical conditions................................................................. 183

6.6 Importance values for input features in the RF models for C\textsubscript{IV} physical conditions................................................................. 183

6.7 Predictive accuracy gain from each input feature in the RF models... 184

6.8 Overdensity and temperature phase space for the predicted physical conditions................................................................. 186

6.9 Overdensity and temperature phase space for the predicted physical conditions, with normalising transform applied......................... 187

6.10 Histograms of overdensity and temperature for H\textsubscript{i} absorbers in the test dataset, showing the effect of the normalising transform .......... 188

6.11 Physical conditions for Mg\textsubscript{II} absorbers, predicted by RF models ...... 192

6.12 Physical conditions for Si\textsubscript{III} absorbers, predicted by RF models ...... 192

6.13 Physical conditions for O\textsubscript{VI} absorbers, predicted by RF models ...... 192

6.14 Overdensity and temperature phase space for the predicted physical conditions, with additional Gaussian noise........................................ 193
List of Tables

2.1 SIMBA runs used in this work .................................................. 29

3.1 Median global galaxy properties for star forming and green valley SIMBA galaxies ................................................................. 51

3.2 Median global galaxy properties for star forming and green valley SIMBA galaxies at high resolution ............................................. 72

4.1 Ion Properties and detection thresholds ....................................... 99

5.1 CGM absorber sample properties for examining physical conditions .. 136

6.1 CGM absorber sample properties for training random forest models .. 176
Chapter 1

Introduction

1.1 Galaxies

1.1.1 How are galaxies made?

The current accepted picture of the growth of structure in the Universe is known as the Λ Cold Dark Matter (ΛCDM) model. In this model, ‘baryonic’ matter (a category which here includes all elementary particles in the Standard Model of particle physics) constitutes just 5% of the energy density of the Universe. Baryonic matter (or simply ‘baryons’) makes up all of the stars, gas and planets that we can observe directly. A further 25% of the energy density consists of dark matter, which (in the ΛCDM model) is cold, collisionless, and has negligible random motions. Dark matter does not interact electromagnetically and thus cannot be observed directly; rather its existence is inferred from its gravitational influence on baryonic matter. Thus the proportion of the matter density in baryons is roughly one sixth:

$$\frac{\Omega_b}{\Omega_b + \Omega_d} = \sim \frac{1}{6}$$  \hspace{1cm} (1.1)

Where \( \Omega_b \) is the baryonic density parameter (the comoving baryon density normalised to the critical density today), and \( \Omega_d \) is the dark matter density parameter. The remaining 70% of Universe’s energy density is a cosmological constant \( \Lambda \), which provides vacuum energy and drives the accelerated expansion of the Universe. The ΛCDM model and its predictions have been tightly constrained by
temperature fluctuations in the cosmic microwave background (CMB, Hinshaw et al., 2013; Planck Collaboration et al., 2016, 2020), observations of distant Type 1a supernovae (Riess et al., 1998; Perlmutter et al., 1999), galaxy clustering (Allen et al., 2011), primordial nucleosynthesis (Cooke et al., 2018), and fast radio bursts (Macquart et al., 2020).

The classical ‘bottom-up’ model of galaxy formation and evolution was developed in the 70s and 80s. From the initial ingredients of baryonic and dark matter, gravitational instabilities at early times grew hierarchically via gravitational attraction, creating the potential wells in which galaxies form (White & Rees, 1978). In principle, the infalling material is composed of one sixth baryons and five sixths dark matter; at this stage gravitational attraction is the dominant mechanism and the dark matter and baryons have the same velocity dispersion. Infalling material is continually pulled into the potential wells via spherical collapse until the velocity dispersion balances the gravitational potential, forming virialised regions known as ‘halos’ (Rees & Ostriker, 1977; Silk, 1977). Gas within the halo shock heats to the virial temperature and is supported by thermal pressure; for Milky Way mass halos, the virial temperature is $T_{\text{vir}} \sim 10^6$K. At sufficiently high densities (within the cooling radius of the halo, $r_{\text{cool}}$), the baryon cooling time $t_{\text{cool}}$ is shorter than the dynamical time of the halo $t_{\text{dyn}}$. Thus the baryons lose pressure support and fall inwards. If the baryons retain some angular momentum as they collapse then they will form a star-forming disk galaxy within $r_{\text{cool}}$ (Fall & Efstathiou, 1980).

In the idealised scenario, subsequent mergers between galaxies disturb the disk structure and form quiescent elliptical galaxies (Toomre & Toomre, 1972), although several other pathways to forming an elliptical galaxy exist (see review by Naab & Ostriker, 2017). Amongst these are environmental processes such as tidal forces and galaxy fly-bys, misaligned gas accretion between the gaseous stream and the galaxy disk, and morphological transformation via heating due to the interplay between the disk and the central bar. Fast rotating elliptical systems are likely to arise from misaligned gas accretion, whilst slowly rotating systems are likely due to merger activity. In addition, quiescent galaxies need not necessarily be elliptical in shape, for example in the case of ‘red disks’ (Schawinski et al., 2010; Bundy et al., 2010).
1.1.2 How do galaxies get their gas?

After galaxies are initially formed, they continue to grow further over cosmic time through a combination of accreting new gas and incorporating the mass of other galaxies through merger events. In this section I will outline the observations and theoretical understanding of gas accretion.

Accretion of cold gas onto galaxies is necessary to sustain star formation over cosmic time. The observational case for accretion of cold gas is encapsulated in Figure 1.1, which shows depletion timescales for gas conversion into stars, taken from Péroux & Howk (2020). The depletion timescales are computed from the gas density $\rho_{\text{gas}}$ and the star formation rate (SFR) density $\dot{\rho}_*$:

$$\tau_{\text{dep}} = \frac{\rho_{\text{gas}}}{\dot{\rho}_*}. \quad (1.2)$$

The SFR density is taken from data compiled by Madau & Dickinson (2014). The cold (neutral and molecular) gas densities were compiled from observations of 21cm emission at $z < 0.4$, and Lyα absorption at higher redshifts. The molecular (star forming) gas densities are compiled from CO emission surveys; molecular $\tau_{\text{dep}}$ data from massive galaxies (black dot-dashed line) were taken from Tacconi et al. (2020). The cold gas $\tau_{\text{dep}}$ is represented by a green line and varies with redshift, but is always longer than the dynamical time (taken to be 10% of the Hubble time, shown as a blue dotted line). This comparison indicates that the cold gas will never be fully depleted, therefore there is sufficient cold gas in the Universe to sustain star formation at all times. The $\tau_{\text{dep}}$ for molecular gas is represented with a grey line and is roughly constant with redshift. At $z < 1$ however, the $\tau_{\text{dep}}$ for molecular gas is lower than the dynamical time, indicating that although there is abundant cold gas available, the supply of cold gas is not converted efficiently into molecular gas within galaxies. Thus, the available molecular gas within galaxies is insufficient to sustain star formation. Therefore, given the observed increase in stellar mass and cold gas density at late times (Péroux & Howk, 2020), accretion onto galaxies from the ionised intergalactic medium (IGM) and galaxy halos is necessary to sustain star formation, and the gaseous material must be cooled by some mechanism into star forming gas. Locally, the Milky Way is in this situation, with only $5 \times 10^9 M_\odot$ of star forming fuel in the disk (Zheng et al., 2019) and a current SFR of 1-3 $M_\odot\text{yr}^{-1}$ (Kennicutt & Evans, 2012).
Figure 1.1: Timescales for gas conversion into stars, compiled from the literature by Péroux & Howk (2020). The depletion timescale of molecular gas at $z < 1$ is lower than the dynamical time, indicating that additional accretion onto galaxies is necessary to maintain the rate of star formation. Plot taken from Péroux & Howk (2020, their Figure 6a).
In addition, more direct observational evidence for accretion onto galaxies exists. At low redshift, galaxy gas is characterised by high metallicities (e.g. Andrews & Martini, 2013) since stars enrich their surroundings with metals at the end of their lives. As such low metallicity gas indicates new inflows from more pristine sources, such as primordial cosmic web filaments (Ford et al., 2014; Anglés-Alcázar et al., 2017b; Hafen et al., 2019). Such low metallicity gas has been detected as extraplanar H\textsc{ii} regions in spiral galaxies (Howk et al., 2018) and in spatially resolved H\textsc{i} regions in star forming galaxies (Hwang et al., 2019).

The canonical bimodal model of accretion from the IGM is that galaxies can grow via cooling of shock heated gas, or via cold filamentary inflows (Katz et al., 2003; Birnboim & Dekel, 2003; Kereš et al., 2005; Dekel & Birnboim, 2006; Stern et al., 2019). The former (‘hot mode’ accretion, also called ‘cooling flows’) is the conventional picture of spherical collapse of hot gas onto halos outlined by Rees & Ostriker (1977) and Silk (1977). By contrast, if the cooling radius is larger than the virial radius, then inflowing filamentary gas from the cosmic web is never heated to the virial temperature before it accretes onto the galaxy (cold mode accretion, also called ‘cold flows’). At higher redshift, cold mode accretion is prevalent since the higher gas densities lead to shorter cooling times; lower mass halos ($M_{\text{halo}} \lesssim 10^{11} \text{ M}_\odot$) are also dominated by cold mode accretion since they have lower virial temperatures and radii. Thus the dominant form of gas accretion in simulations depends both on redshift and halo mass (Katz et al., 2003; Birnboim & Dekel, 2003; Kereš et al., 2005; Kereš et al., 2009; Ocvirk et al., 2008; Dekel et al., 2009; van de Voort et al., 2011; Joung et al., 2012b).

To complicate matters, both modes may occur in the same halo - particularly at high redshift where filaments can remain cold in the outskirts due to high densities, bringing cold material into the hot virialised halos of massive galaxies (Kereš et al., 2005; Dekel & Birnboim, 2006; Ocvirk et al., 2008; Kereš et al., 2009; Dekel et al., 2009). Outer regions of the filament can thus be heated due to compression from the surrounding halo medium (Kereš et al., 2005; Joung et al., 2012b; Mandelker et al., 2020) with cooler gas remaining in the inner regions (Putman et al., 2012). The details of hot vs cold mode accretion depend on the simulation code and resolution (e.g. Peeples et al., 2019; Hummels et al., 2019; van de Voort et al., 2019; Bennett & Sijacki, 2020); see §1.3 for more on cosmological simulations. Finally, to accrete onto the disk the gas needs to cool further, possibly forming clouds within the halo (Maller & Bullock, 2004) which could be disrupted before they can accrete (Joung et al., 2012a), or mixing with
the galaxy at the disk-halo interface (Putman et al., 2012).

Massive galaxies can also gain gas from satellite galaxies living in their halos, which lose their gas as they move through the gaseous halo through a combination of ram pressure stripping (Gunn & Gott, 1972), interactions with other galaxies (Moore et al., 1996) and tidal forces (Gnedin, 2003). These processes leave gas in the halos of massive galaxies that eventually will be accreted onto the disk. For example, the Magellanic Stream of gas is composed of material tidally stripped from the Large and Small Magellanic Clouds (For et al., 2014) and will eventually supply 30-50% of the Milky Way’s HI mass (Putman et al., 2009). Overdensities that remain in the halo may enable cold filamentary inflows. Alternatively, if the gas is stripped at large distances from the galactic disk then the gas is mixed into the halo before cooling onto the disk along with the rest of the halo gas (Putman et al., 2012). In general, gas accreted from satellites has a higher metallicity than pristine filamentary gas as it has been processed through star formation (Fernández et al., 2012), making it possible to determine the origin of the gas.

1.1.3 Gas inside galaxies

Stars form in galaxies from the dense, cold, molecular gas in the interstellar medium (ISM) (see reviews by Tacconi et al., 2020; Saintonge & Catinella, 2022), where around 1% of the available gas is converted into stars per free fall time (Bigiel et al., 2008; Krumholz et al., 2012; Leroy et al., 2013). More broadly, the ISM consists of dust, and molecular, neutral and ionised gas, and is also the repository for the enriched gas produced by old and massive stars; hence its composition evolves over cosmic time. It also contains energy and momentum output from the same old and massive stars, with gas phases coexisting in the ISM at different densities and temperatures (McKee & Ostriker, 1977).

Galaxies broadly fall into two classes: star-forming spiral galaxies, and quiescent elliptical galaxies. They occupy clearly distinct regions in the color-mass parameter space, the so-called ‘blue cloud’ and ‘red sequence’ (e.g. Strateva et al., 2001; Baldry et al., 2004; Balogh et al., 2004). In between there is the ‘green valley’ (GV), regarded as a transition zone since all galaxies must begin as star-forming while the most massive galaxies tend to be quiescent, which suggests that at some point blue galaxies must stop forming stars and become red and dead (e.g. Bell et al., 2004; Faber et al., 2007; Martin et al., 2007; Fang et al., 2012).
What physical driver(s) ‘quench’ galaxies, i.e. transform them from being star-forming to quiescent, is a longstanding yet poorly understood question in galaxy evolution. Since cold, molecular gas is the requirement for star formation, quenching is linked with the loss of star forming gas, either by removal from the galaxy, or by heating to outside the conditions for star formation.

1.1.4 How do galaxies lose their gas?

Galaxies can also lose their gas via energetic, galaxy-scale outflows. Outflowing winds are ubiquitous (Rubin et al., 2014) and are often composed of multiphase gas that extend for many kpc, for example in M82 (the Cigar Galaxy, Strickland & Heckman, 2009; Lopez et al., 2020), NGC 3079 (Hodges-Kluck et al., 2020), NGC 5128 (Centaurus A, Weiß et al., 2008; McKinley et al., 2022), MRK 231 (Veilleux et al., 2016), and in the Milky Way itself (Di Teodoro et al., 2020).

These energetic outflows are linked to the suppression of galaxy growth. The star formation density in the Universe peaked at around $z \sim 2$, and has been steadily declining for the last 10 Gyr (Madau & Dickinson, 2014). Star formation within galaxies is so inefficient that only 10% of the baryons in galactic halos today exist in stars (Péroux & Howk, 2020). As such, holistic models of galaxy evolution must not only account for how galaxies grow, but also how their growth is regulated.

The suppression of galaxy growth at both the low and high mass end is demonstrated in Figure 1.2, showing a compilation of galaxy stellar mass functions from both simulations and observations, taken from Naab & Ostriker (2017). The dashed black shows the stellar mass function if the one sixth of matter in baryons were converted efficiently into stars. In contrast, the observed stellar mass functions are lower than this prediction at all masses, indicating that the build up of mass in stars is being regulated. Early numerical simulations and semi-analytic models without any suppression mechanisms showed that in such a scenario, galaxies grow to unrealistic masses (Bower et al., 2006; Croton et al., 2006; Somerville et al., 2008). Hence there must be some process that prevents accretion, or ejects star forming gas from galaxies. The distinctive knee shape in the stellar mass function indicates that different suppression mechanisms dominate above and below $M_\star \sim 10^{11} M_\odot$; the main mechanisms thought to regulate growth are SN-driven feedback at low masses, and black hole-driven feedback at high masses.
Figure 1.2: Galaxy stellar mass functions in simulations and observations, compiled by Naab & Ostriker (2017). The stellar mass function is lower than the expectation from the cosmic baryon fraction, indicating that star formation is inefficient. Star formation is regulated by SN feedback at low masses and AGN feedback at high masses. Plot taken from Naab & Ostriker (2017, their Figure 4).
In each population of newly formed stars, roughly 1% of the stars will have a mass greater than 8 $M_\odot$, assuming a Chabrier (2003) Initial Mass Function (IMF). Such massive stars are relatively short lived and will end their lives in a supernova event (unless they are massive enough to form a black hole). Each supernova event enriches its surroundings with heavy metals and releases roughly $10^{51}$ ergs (Khokhlov et al., 1993). For a low mass galaxy ($M_{\text{halo}} < 10^{11} M_\odot$) the total collective energetic output exceeds the galaxy binding energy, driving galaxy-scale metal-enriched winds (referred to as ‘SN-driven feedback’) that remove star forming gas from the galaxy. These winds can enrich gas as far from the galaxy as the IGM (Simcoe et al., 2004); thus metals can act as tracer particles to constrain feedback models.

Much work has been done using numerical simulations to study models of galactic winds (Hummels et al., 2013; Bird et al., 2015; Rosdahl et al., 2017; Pandya et al., 2021). Murray et al. (2005) outlined a model for momentum-driven winds, where the radiation pressure from supernovae drives the outflow by transferring momentum to gas particles. The input momentum per unit star formation is constant, with the mass outflow rate inversely proportional to the velocity of the wind:

$$\dot{M} \propto \frac{1}{v}$$  \hspace{1cm} (1.3)

Such winds are able to suppress early star formation and enrich the IGM due to their high mass outflow rates, but do not overheat the IGM due to their low velocities (Oppenheimer & Davé, 2006). An alternative to momentum-driven winds is energy-driven winds, where the input energy per unit star formation is constant (e.g. Bertone et al., 2007):

$$\dot{M} \propto \frac{1}{v^2}$$  \hspace{1cm} (1.4)

However in energy-driven winds, the energy can be radiated away before material driven away from the galaxy. Broadly, momentum-driven winds increase the velocity of gas particles, whereas energy-driven winds impart thermal energy. Momentum-driven winds were adopted for later work as they have been able to drive outflows and enrich the IGM far from host galaxies more successfully than energy-driven models (Murray et al., 2005; Oppenheimer et al., 2009).

At the high mass end, the energetic output from supernovae is insufficient to quench galaxies; instead, the dominant mechanism for suppressing star formation
is Active Galactic Nucleus (AGN) feedback. AGN are compact regions at the centers of galaxies that power luminous, energetic outflows of material, thought to be fuelled by accretion onto a central supermassive black hole at the center of the galaxy (see reviews by McNamara & Nulsen, 2007; Fabian, 2012; Kormendy & Ho, 2013; Heckman & Best, 2014). Recent studies propose that the apparent diversity in AGN arises from differences in viewing angle and propose that AGN can be classified into two fundamental types (Hardcastle et al., 2007; Heckman & Best, 2014) based on the accretion efficiency of the central black hole. These are ‘radiative’ (or ‘quasar’) mode at high black hole accretion efficiency, and jet mode (or ‘radio’ mode) at low accretion efficiency. Radiative mode AGN are able to efficiently convert the potential energy from accreted material into radiation. They have accretion disks which are geometrically thin and optically thick and can drive multiphase winds of molecular and warm ionised gas at velocities of \( \sim 10^3 \text{kms}^{-1} \). At low Eddington ratios, the output of the AGN is in mechanical energy, driving hot gas in two-sided collimated jets at \( \sim 10^4 \text{kms}^{-1} \), powered by accretion onto the black hole or from the spin energy of the black hole. These jets can inflate hot cavities and shock fronts in galactic halos. As such, the challenge for simulators is to include AGN feedback in a manner that represents the observed dichotomy in AGN properties.

The outflows of gas and radiation from AGN can significantly affect their host galaxies by heating the surrounding ISM and driving outflows that eject star forming fuel. The binding energy of a galaxy with velocity dispersion \( \sigma \) and stellar mass \( M_* \) is

\[
E_{\text{binding}} \approx M_* \sigma^2,
\]

and the energy released by the accreting black hole is

\[
E_{\text{bh}} = 0.1 M_{\text{bh}} c^2,
\]

where 0.1 is the radiative efficiency. The typical ratio of \( E_{\text{bh}} \) to \( E_{\text{binding}} \) is \( \sim 100 \) (Merritt & Ferrarese, 2001), so even if a small fraction of the radiated energy couples to gas, then all the surrounding gas will be driven out of the galaxy. The co-evolution of the black hole growth rate density and the star formation rate density strongly implies that the growth of black holes and subsequent AGN feedback are largely responsible for galaxy quenching (Kormendy & Ho, 2013). As well as driving winds, AGN feedback heats galactic halos and the IGM causing
long cooling times and ensuring galaxy quenching by preventing future accretion of gas (‘preventative feedback’, Lu et al., 2015, 2017; Davies et al., 2020). Finally, AGN feedback helps spread metals in the IGM; simulations with AGN feedback but zero-metallicity winds have more metal enrichment at $z < 3$ than simulations with enriched winds but no AGN feedback (Suresh et al., 2015).

### 1.1.5 The Baryon Cycle

In reality, inflowing and outflowing gas must both pass through the galaxy halo simultaneously and can interact; for example, accretion flows depend on the feedback prescription via its impact on radiative cooling (van de Voort et al., 2011; Nelson et al., 2015; Fielding et al., 2017; Dutta et al., 2022) and its disruptive effect on the halo (Muratov et al., 2015). Additionally, gas stripped from satellite galaxies is integrated into the diffuse halo (Putman et al., 2012; Tonnesen & Bryan, 2021; Hafen et al., 2019). Gas that has been previously ejected into the halo can also be recycled by radiatively cooling and re-accreting onto the galaxy. This idea is known as ‘baryon cycling’, or the ‘galactic fountain’ model (see reviews by Péroux & Howk, 2020; Donahue & Voit, 2022) of exchanges of mass and energy between galaxies and their surroundings.

Gas recycling has been directly observed as metal enriched inflowing gas (Rubin et al., 2012, 2022) and in HI emission around star forming galaxies (Oosterloo et al., 2007; Marasco et al., 2019). It has also been predicted by simulations, which show that at early times outflows can enrich the IGM, but at late times outflows do not escape the CGM and condense back on to the galaxy (Oppenheimer & Davé, 2008), providing a major source of fuel for new star formation (Oppenheimer et al., 2010). Simulations predict that low ionisation metals observed in galactic halos typically trace cold, dense enriched structures that will soon by re-accreted, while most high ionisation metals trace material that was ejected at earlier times (Ford et al., 2014). Recycled gas can even be exchanged between neighbouring galaxies in an intergalactic fountain (Anglés-Alcázar et al., 2017b).

The gaseous halo immediately surrounding a galaxy is also known as the Circumgalactic Medium (the CGM, Chen, 2017; Tumlinson et al., 2017). Given the interplay of accretion, galactic winds and recycling gas, the CGM is the site of much interest as both the reservoir of available gas and the repository for outflows. Studying the evolution of CGM gas over time and its dependence on host galaxy properties can provide constraints on models of galactic gas flows; as
such, understanding the CGM is crucial for understanding galaxy evolution as a whole. Definitions of the CGM vary; I will adopt a definition of gas within the galaxy virial radius but outside the main structure and ISM of the galaxy.

1.2 Observing gaseous halos

1.2.1 Emission mapping

Ideally, the CGM is observed directly via emission mapping, however the diffuse nature of halo gas makes detections challenging (Augustin et al., 2019; Wijers et al., 2019) since emission from radiative cooling scales as particle density squared. The relative proximity of the Milky Way’s halo makes it more feasible to map in emission (see review by Putman et al., 2012). Cold and warm ionised Galactic halo gas has been detected as high velocity clouds (HVCs) in 21cm HI emission (Stanimirović et al., 2008; Hsu et al., 2011; Putman et al., 2011, 2012) and H$\alpha$ emission (Tufte et al., 1998; Barger et al., 2017; Smart et al., 2019), respectively; additionally X-ray emission traces the hot ambient halo medium (e.g. Kaaret et al., 2020; Predehl et al., 2020; Bhattacharyya et al., 2022).

In the case of other galaxies, the halos of spiral galaxies have been mapped in 21 cm HI emission (the MHONGOOSE survey, Sardone et al., 2021), while OII emission has been used to study ionised outflows into the CGM (Rupke et al., 2019). The halos of individual massive galaxies have been studied in X-ray emission (Anderson et al., 2016), while for lower mass galaxies stacking observations is required to enhances the X-ray sensitivity (Bogdán et al., 2015).

1.2.2 Absorption lines

The most widespread technique for observing the CGM is through absorption lines in the spectra of background sources. As light from the background source passes through the intervening halo gas, some may be absorbed due to resonance transitions and re-emitted along a new path. Absorption from the HI and metals of a particular halo may be identified at the approximate redshift along the line of sight (LOS) of the main galaxy, providing a measure of the column density of gas along the LOS. Background sources are typically conveniently-located higher
redshift quasars, but other sources may be used, for example fast radio bursts (Prochaska & Zheng, 2019; Ocker et al., 2021), gamma ray bursts (Cucchiara et al., 2015; Bolmer et al., 2019), and other galaxies (Steidel et al., 2010; Cooke & O’Meara, 2015; Péroux et al., 2018; Zabl et al., 2020).

Targeted absorption surveys select background sources based on their proximity to specific structures or types of galaxies. For example, at low redshift such surveys have probed the CGM of the Milky Way (e.g. Sembach et al., 2003; Fox et al., 2013, 2014), L* galaxies (e.g. COS-Halos, Tumlinson et al. 2013, Werk et al. 2013, 2014, 2016, Peeples et al. 2014, Prochaska et al. 2017; CGM$^2$, Wilde et al. 2021, Tchernyshyov et al. 2022), sub-L* galaxies (e.g. COS-Dwarfs, Bordoloi et al. 2014; also Liang & Chen 2014, Johnson et al. 2015, 2017), AGN (COS-AGN, Berg et al., 2018), starburst galaxies (COS-burst, Heckman et al., 2017), massive quenched galaxies (COS-LRGs, Chen et al. 2018, 2019, Zahedy et al. 2019; the Red Dead Redemption Survey, Berg et al. 2019), and Lyman Limit System (CUBS, Chen et al., 2020; Zahedy et al., 2021; Cooper et al., 2021; Qu et al., 2022), as well as the ISM-CGM connection (COS-GASS, Borthakur et al., 2015, 2016) and the CGM-IGM connection (COS-IGrM, McCabe et al., 2021).

Absorption measurements have the important advantage of being able to probe more diffuse gas than emission mapping; additionally, the sensitivity is redshift independent, and as such absorption lines may be used to trace the cosmic evolution of the CGM. However, this technique also has its downsides. Firstly, each galaxy is typically limited to only one LOS, except for the Milky Way and nearby galaxies (e.g. M31, Lehner et al., 2020) or where the background source is gravitationally lensed (e.g. Lopez et al., 2018; Rubin et al., 2018; Mortensen et al., 2021). Individual LOS may however be aggregated into large absorber samples, which reduces cosmic variance effects and allows the typical CGM properties to be studied. Secondly, many of the strong transitions exist in the ultra-violet (UV) regime, necessitating space-based observations, or higher redshift studies whereby the lines have been redshifted into the visible regime (see Figure 6 of Tumlinson et al., 2017). A useful exception is the MgII doublet at 2796 Å, which traces dense and cold gas in the ISM and CGM and can be observed from the ground, making it one of the most widely observed transitions (e.g. Chen et al., 2010; Nielsen et al., 2013; Schroetter et al., 2016; Huang et al., 2016, 2020, 2021).

A related method using large spectroscopic surveys is to stack hundreds of spectra in an effort to raise the signal-to-noise (SNR) of the data and extract weaker absorption, with the downside of averaging out kinetic and ionization structure
in individual galaxies. The catalogs can be split by properties of the foreground galaxies such as mass, radius, SFR, colour, environment, and orientation to extract information about how these properties may affect the CGM (Bordoloi et al., 2011; Zhu & Ménard, 2013).

1.2.3 The multiphase CGM

Efforts to characterise the CGM using absorption lines have revealed a complex, multiphase structure. Figure 1.3 shows equivalent width of absorption profiles in the CGM for HI and various metal lines, compiled by Tumlinson et al. (2013). The measurements are taken from COS-Halos (Tumlinson et al., 2013; Werk et al., 2013), COS-Dwarfs (Bordoloi et al., 2014), MAGICAT (Nielsen et al., 2013; Liang & Chen, 2014), CaSBAH (Burchett et al., 2019), the X-ray study of Yao et al. (2012), KBSS (Rudie et al., 2012; Turner et al., 2015), the OVI study of Prochaska et al. (2011), and COS-GASS (Borthakur et al., 2015). The metal lines span an order of magnitude in terms of ionisation potential; as such these lines trace gas from a range of physical conditions. However, absorption across a range of underlying physical phases is detected in the same radial regions of the CGM.

Furthermore, absorption components from ions tracing different physical conditions can arise from the same individual absorber systems (e.g. Tripp et al., 2008; Johnson et al., 2017), even with resolved spectroscopy (Chen et al., 2020; Zahedy et al., 2019, 2021; Cooper et al., 2021; Qu et al., 2022). Where components tracing high and low ions are misaligned, this suggests that the gas is multiphase (e.g. Fox et al., 2013), whereas kinematically aligned components indicate that the absorbing cloud itself is multiphase (Tripp et al., 2011). Inferring the physical conditions of such absorbers from the relative ion abundances, gas density and metallicity in the cool CGM may vary by up to 1 dex within an individual halo (Zahedy et al., 2019). In the halo of the Milky Way, high velocity absorption lines tracing highly ionised metals lie near HVCs of neutral hydrogen, both spatially and kinematically (see Figures 1 and 2 of Putman et al., 2012). HVCs in the Milky Way also show a head-tail structure that suggests cold clouds moving through a warm medium (Putman et al., 2011). The observed multiphase nature of the CGM emphasises the mixture of gas present and the many possible physical origins of halo gas.
Figure 1.3: Equivalent width in the CGM against impact parameter for HI and various metal lines, for galaxies in COS-Halos (Tumlinson et al., 2013; Werk et al., 2013), COS-Dwarfs (Bordoloi et al., 2014), MAGICAT (Nielsen et al., 2013; Liang & Chen, 2014), CaSBAH (Burchett et al., 2019), the X-ray study of Yao et al. (2012), KBSS (Rudie et al., 2012; Turner et al., 2015), the OVI study of Prochaska et al. (2011), and COS-GASS (Borthakur et al., 2015). Plot taken from Tumlinson et al. (2017, their Figure 4).
1.3 Simulating gas in and around galaxies

Given the complexity of the observational picture and the variety of processes taking place in the CGM, numerical simulations are necessary to interpret the observations within a theoretical model. Such simulations solve the relevant physical equations for particles or cells that represent dark matter, gas, and stars, with the aim of reproducing the structure and kinematics of galaxies and their environments. With any numerical simulation, there is a trade off between computational cost and accuracy, which sets the resolution scale of the simulation and thus the level of detail that can realistically be captured. Since galaxy formation involves processes across a wide range of physical size scales, it is not currently feasible to directly simulate all relevant processes within a single cosmological simulation. Instead, individual projects focus on constraining smaller portions of the problem, with the results of small scale, high resolution simulations informing larger scale, lower resolution simulations.

Figure 1.4 illustrates the level of detail captured in simulations of different size scales and how these inform one another, focusing on gaseous processes. At the smallest scales, modellers simulate how individual cold clouds travel through a warm medium (Scannapieco & Brüggen, 2015; Armillotta et al., 2017; Gronke & Oh, 2018, 2020, 2022; Banda-Barragán et al., 2019; Sparre et al., 2019, 2020; Gronke et al., 2022), which inform ‘tall box’ simulations of patches of the star-forming ISM (Creasey et al., 2013; Walch et al., 2015; Seifried et al., 2017; Kim & Ostriker, 2017; Smith et al., 2020; Gutcke et al., 2021; Colman et al., 2022). On larger scales, idealised global simulations of galactic disks are typically used to study how SN winds propagate (Sarkar et al., 2015; Rosdahl et al., 2017; Fielding et al., 2018; Schneider & Robertson, 2018; Smith et al., 2018; Li et al., 2020; Schneider et al., 2020) while idealised global halo simulations are used to study the impact of black hole feedback (Choi et al., 2012; Costa et al., 2020; Prasad et al., 2015; Bourne & Sijacki, 2017; Weinberger et al., 2017b; Talbot et al., 2021; Husko & Lacey, 2022). ‘Zoom’ simulations re-model individual galaxies selected from cosmological boxes, at a higher resolution than the original simulation, e.g. FIRE and FIRE-2 (Hopkins et al., 2014, 2018), Auriga (Grand et al., 2017) and NIHAO (Wang et al., 2015), Agertz & Kravtsov (2016). Recent zoom projects have specifically focused numerical resources on the CGM (Peeples et al., 2019; Hummels et al., 2019; van de Voort et al., 2019). Parameterisations from higher resolution simulations, such as the bulk effects of galactic winds,
are then included in cosmological simulations of tens or hundreds of Mpc, e.g. EAGLE Schaye et al. (2015); Crain et al. (2015), Simba Davé et al. (2019), Illustris Vogelsberger et al. (2014), IllustrisTNG (Pillepich et al., 2018; Nelson et al., 2018a), Horizon-AGN (Dubois et al., 2016), FABLE (Henden et al., 2018), and Romulus25 (Tremmel et al., 2017). Whereas cosmological simulations lack the high resolution necessary to self-consistently model small scale processes, they capture the cosmic web of nodes, filaments and voids and thus contain large statistical samples of galaxies across a range of environments. The recent SMAUG project (Fielding et al., 2020b) combines these approaches by comparing results from idealised and cosmological simulations.

### 1.3.1 Methods of cosmological simulations

In a cosmological simulation, a typical volume of the universe is represented using particles or cells, the relevant physical forces on the particles are computed, and the system is evolved forward in discrete time steps. The expansion of the universe is accounted for by solving the equations in a comoving coordinate system, where
the expansion rate is computed from the Friedmann equation (Peacock, 1998; Mo et al., 2010), although gravity is treated as Newtonian. The boundaries of the box are periodic, with the assumption that the volume is sufficiently large that repeating the box approximates homogeneity and isotropy of the large scale structure. Here I outline the numerical techniques used to represent matter and compute the forces acting on each matter particle (see reviews by Somerville & Davé, 2015; Vogelsberger et al., 2020). Where relevant, I indicate which methods are used by EAGLE (Schaye et al., 2015; Crain et al., 2015) and IllustrisTNG (Pillepich et al., 2018), two of the current major cosmological simulation projects; the description of the SIMBA simulations that are used in this thesis is left for a separate discussion (see §2).

**Initial conditions**

First initial conditions must be generated to mimic the inhomogeneities in the early universe; these are generated in a similar manner between most modern cosmological simulations. The required initial conditions are the particle positions and velocities at high redshift (before structure formation becomes non-linear). To generate such conditions, the linear matter power spectrum is generated via a transfer function (Eisenstein & Hu, 1999), which is sampled randomly in a Gaussian manner for the simulation volume modes. The modes are evolved forward using the Zel’dovich approximation (Bertschinger, 1998). The baryon temperature is typically set to the microwave background temperature at that redshift. A popular choice of initial condition generator is MUSIC (Hahn & Abel, 2011).

**Gravity**

Numerical N-body methods calculate gravitational forces by solving Poisson’s equation using either a particle-based, mesh-based, or hybrid approach. Particle-based methods seek to compute the gravitational forces on each particle, where to avoid calculating contributions from all other particles (an approach which scales as $O(N^2)$), the contributions from distant particles are approximated based on their multipole moments (tree codes, Barnes & Hut, 1986). Gravitational interactions are softened on small scales with a kernel-smoothing technique to avoid two-body scattering, setting the resolution scale at the force softening.
length. Mesh-based methods, such as the particle-mesh method (PM Hockney & Eastwood, 1988), solve the gravitational potential on a grid with a Fast Fourier Transform of the density field, and particles are moved along gradients in the potential. The resolution scale is set by the grid size. Both tree codes and PM methods both scale as $O(N\log N)$, but PM methods are faster. The hybrid ‘Tree-PM’ approach (Bode & Ostriker, 2003), uses a tree code to compute the short range forces and a Fourier-based PM method for long range forces. This method is popular among modern cosmological simulations (e.g. GADGET-2/3, Springel 2005; AREPO, Springel 2010b; GIZMO, Hopkins 2017). The EAGLE simulation uses GADGET-3; IllustrisTNG uses AREPO.

**Hydrodynamics**

To model baryons, the simulation must also include hydrodynamics by solving the Euler equations that represent the conservation of mass, momentum and energy, assuming an ideal gas. Numerical hydrodynamics solvers use either Eulerian, Lagrangian or Lagrangian-Eulerian techniques. Eulerian methods treat the problem as flows of gas through cell boundaries, whereas Lagrangian methods model parcels (or ‘elements’) of gas that move through space; see Hu et al. (2022) for a recent comparison of these methods.

The most common Eulerian method is the high-order Godunov scheme, whereby a Riemann problem is solved at cell faces for the flux across the cell; for uniform cells, this is known as a first-order Godunov solver. Parabolic interpolation can be included for a more accurate solution (the Piecewise Parabolic Method, Colella & Woodward, 1984). Cells can be split into sub-cells to improve resolution, usually based on the local mass distribution, in a technique known as Adaptive Mesh Refinement (AMR), which is popular in modern astrophysical codes (e.g. Enzo, Bryan et al. 2014; RAMSES, Teyssier 2010).

The most common Lagrangian method is Smoothed Particle Hydrodynamics (SPH, see reviews by Monaghan, 1992; Springel, 2010a). In SPH, the local density field is represented using a kernel-weighted sum over the neighbouring particles within a smoothing length. Linear momentum, angular momentum, and mass are simultaneously conserved in ‘classic’ SPH, while entropy-conserving (EC-)SPH (used in GADGET-2, Springel, 2005) was developed to also conserve energy and entropy.
Lagrangian methods are naturally resolution adaptive and easily implemented in 3D, whereas Eulerian methods handle shocks and discontinuities better. Lagrangian-Eulerian codes seek to combine the advantages of both approaches. For example the AREPO code (Springel, 2010b), used for IllustrisTNG, uses Voronoi tessellation to create a mesh around particles, which is re-generated with each timestep to follow the flow of the fluid. A similar Lagrangian-Eulerian method is Meshless Finite-Mass (MFM, Hopkins, 2015), implemented in the GIZMO code (Hopkins, 2015, 2017); in this case the ‘mesh’ is created with a deformable kernel that is re-generated with each timestep in such a way that the mass within each cell is constant.

**Heating and cooling**

Simulations that include baryons must account for how baryons dissipate their energy via radiative cooling and gain thermal energy via photoionisation heating. As such most modern cosmological simulations include radiative cooling and photoionisation heating, as well as metal line cooling, which dominates for gas with $10^5 \text{K} < T < 10^7 \text{K}$. The cooling processes are included in the energy equation using cooling functions, for example those tabulated in the GRACKLE cooling library (Smith et al., 2017). The redshift-dependent metagalactic radiation field (e.g. Haardt & Madau, 2012; Faucher-Giguère et al., 2009; Faucher-Giguère, 2020; Khaire & Srianand, 2019) is included as a source of heating, assuming spatial uniformity.

**Sub-grid processes**

Due to the finite resolution of a cosmological simulation, sub-grid algorithms are required to emulate the bulk effect of small scale processes that cannot be solved self-consistently during the simulation. Sub-grid processes are typically parameterisations of processes that are not modelled directly, where the parameterisations are tuned to match a specific set of observations. As such, much recent work has focused on testing the sub-grid implementation by examining the results of such models within and across different simulation projects, and comparing these to different sets of observations (e.g. Rosdahl et al., 2017; Smith et al., 2018; Habouzit et al., 2021; Oppenheimer et al., 2021).
**Star formation:** Stars form from cold, dense gas; as such gas particles are assigned a star formation rate (SFR) based on the Schmidt (1959) law, often calibrated to match the Kennicutt (1998) relation between gas surface density and star formation surface density. Star formation is typically limited to gas particles above a certain threshold in density. Based on the SFR, a portion of the gas particle is transformed into a star particle while conserving mass. Star particles are collisionless and modelled as single-metallicity stellar populations assuming a universal IMF. A secondary sub-grid model is necessary to artificially overpressurise the ISM to prevent gas from collapsing to the Jeans scale.

**Stellar feedback:** Star formation leads to feedback which regulates further star formation by driving galactic-scale winds. Implementations for stellar feedback differ in how the injected energy and momentum are coupled to the surrounding gas. Energy may be deposited either in thermal or kinetic form. In thermal form, the energy may be radiated away excessively, losing the intended impact; some implementations disable radiative cooling temporarily to counteract this. In kinetic form, though the energy is not as easily radiated away, hydrodynamically decoupling gas particles into separate wind particles can still be necessary to achieve galactic-scale outflows. The EAGLE simulations (Schaye et al., 2015; Crain et al., 2015) treat stellar feedback thermally with stochastic energy deposition, where a free parameter sets the probability of neighbouring gas particles being heated, which results in increased cooling times. The IllustrisTNG simulations (Pillepich et al., 2018) use a kinetic scheme, whereby wind particles are launched stochastically and hydrodynamically decoupled until they leave the ISM.

**Black hole growth:** Black holes are typically seeded in galaxies above a threshold in stellar mass, where the black hole mass is either fixed or calibrated to the $M_{bh} - \sigma$ relation. Black holes are predominantly grown in simulations via spherical Bondi-Hoyle-Lyttleton accretion (Hoyle & Lyttleton, 1939; Bondi & Hoyle, 1944; Bondi, 1952); both EAGLE and IllustrisTNG use this approach. An alternative considering the accretion of gas with angular momentum is torque-limited accretion (Anglés-Alcázar et al., 2015).

**Black hole feedback:** As with stellar feedback, the energy from AGN feedback may be deposited as either thermal or kinetic. Some simulations that include AGN feedback aim to mimic the observed dichotomy in AGN between jet and radiative mode feedback. In IllustrisTNG for example, a fraction of the radiative feedback energy from the sub-grid accretion disk couples to nearby gas as thermal
energy, while jets are modelled as pure kinetic feedback, imparting momentum to the surrounding gas in a random direction (Weinberger et al., 2017a). In EAGLE however, there is a single mode of thermal AGN feedback, where energy is added to a ‘reservoir’ around the black hole at each time step; when the reservoir has sufficient energy to heat a certain threshold number of gas particles, the thermal energy is added stochastically to the surrounding particles (Schaye et al., 2015).

**Chemical enrichment**

Most codes track the enrichment of gas with a number of heavy elements in order to predict metallicities and inform metal cooling calculations (e.g. Oppenheimer & Davé, 2008). Metal yields from individual star particles are computed assuming a particular IMF and thus a mass of stars in the requisite mass ranges for Asymptotic Giant Branch (AGB) stars, Type II supernovae, and Type 1a supernovae, which are responsible for the production of heavy metals. Chemical enrichment is not a sub-grid process since it is enacted directly from star particles; however uncertainties arise from the input metal yield models.

**Current challenges**

A number of important but technically challenging astrophysical processes are not commonly included in cosmological simulations due to the computational cost. Firstly, the turbulent mixing of gas layers is likely a significant factor for creating multiphase gas in the CGM (Fielding et al., 2020a; Tan et al., 2021; Rennehan, 2021). Secondly, magnetic fields are thought to be ubiquitous in astrophysics, both in the ISM and on extragalactic scales (McKee & Ostriker, 2007; Beck & Wielebinski, 2013; Han, 2017), but are not normally included in cosmological simulations. By comparing zoom simulations with and without magnetic fields, van de Voort et al. (2021) showed that magnetic fields strongly influence the flow of gas in galaxy halos, allowing less gas to leave the halo and resulting in less well-mixed gas. IllustrisTNG does include magnetic fields as ideal magnetohydrodynamics (Marinacci et al., 2018), which assumes the plasma conducts perfectly and that relativistic effects are negligible. Thirdly, cosmic rays are driven by magnetic fields and contribute to galactic outflows (Simpson et al., 2016) and the survival of cold gas in the CGM (Butsky et al., 2020). Finally, radiative transfer can directly heat gas and alter the ionisation state, but
is typically only included in cosmological simulations of the Epoch of Reionisation (e.g. Kannan et al., 2022). Including these effects in cosmological simulations down to low redshift represents the next phase in galaxy modelling.

### 1.3.2 Testing simulations with the CGM

Given the variety of numerical approaches in modern cosmological simulations, close comparisons between observations and numerical predictions can provide useful constraints on the underlying galaxy formation models. In recent years there has been much work on this topic; I focus here on work at low redshift. The cool CGM in IllustrisTNG has been probed using MgII emission (Nelson et al., 2021) and absorption (DeFelippis et al., 2021). More broadly, a range of UV metal lines have been used to probe the cool and warm ionised CGM in simulations, testing specific stellar wind implementations (Ford et al., 2013, 2014, 2016; Hummels et al., 2013), the NIHAO simulation suite (Gutcke et al., 2017), EAGLE (Oppenheimer et al., 2016, 2018b), IllustrisTNG (Nelson et al., 2020; DeFelippis et al., 2021), and FIRE-2 (Li et al., 2021). The warm hot phase has been probed with X-ray absorption and emission in EAGLE (Wijers et al., 2019, 2020; Wijers & Schaye, 2022; Oppenheimer et al., 2020b) and IllustrisTNG (Nelson et al., 2018b; Oppenheimer et al., 2020b; Truong et al., 2020; Pillepich et al., 2021). In general such comparisons show that despite their modest resolution, cosmological simulations have had success in reproducing observational metrics of the CGM and correlations with galaxy properties.

### 1.4 Machine learning in Astronomy

Modern astronomical research is data-intensive, often measuring on the terabyte/petabyte scale (Fluke & Jacobs, 2020). Machine learning (ML) is a class of automated processes for learning relationships in data. ML is now widely used in Astronomy to cope with the data-intensive nature of the research.

#### 1.4.1 Machine learning methods

ML algorithms can assist with a variety of problems, chiefly classification (category labels are applied to new data based on a training dataset) or regression
Numerical values are predicted based on relationships between variables. The choice between a classification or regression algorithm depends on the problem at hand - classification algorithms are most useful where there are a small number of discrete categories, whereas regression is more appropriate for continuous data. ML algorithms are also either ‘supervised’, meaning that the machine learns by example with data designated as input and output, or ‘unsupervised’, where the machine learns freely from the dataset. In supervised ML, the data is randomly split into two portions, one for training the model and the other for testing the model on new data to avoid over-fitting. Here I describe two of the most widely used methods in Astronomy (Fluke & Jacobs, 2020): Neural Networks and Random Forests.

**Neural Networks**

Artificial Neural Networks (ANNs or NNs, McCulloch & Pitts, 1943; Fukushima, 1980; Rumelhart et al., 1986) are a class of ML algorithms inspired by neurons in the brain. ANNs may be either supervised or unsupervised. They are composed of interconnected nodes (artificial neurons) which are arranged in layers. Each node receives an input and outputs a signal to one or more other neurons. The output of each node is a function of the weighted sum of its inputs; the weights can be tuned to improve the accuracy of the output in either predicting an outcome or a classification.

NNs have the advantages of being accurate, efficient, capable of processing unorganised data, and once trained, do not need a complete input dataset in order to predict a result. Convolutional Neural Networks (CNNs, deep learning) are a specialised type of ANN that use convolution in at least one of their layers. CNNs are specifically designed to process pixel data and are therefore particularly useful in Astronomy for image processing. However, the downside of NNs in a scientific context is that the learned model is often opaque and hard to interpret, hence NN algorithms are often thought of as black boxes.

**Random Forest Regression**

Random Forest (RF) (Breiman, 2001) is a supervised, decision tree-based, ensemble method of machine learning. The term ‘ensemble method’ refers to the process of combining predictions from several machine learning runs (in this
case, individual decision trees) in order to more accurately predict the output. A decision tree splits the input dataset in a top-down manner, in which the best split for the data is found by minimising a cost function. They have the advantage of being easy to interpret and have low bias in their predictions for the training data. However, individual decision trees are prone to over-fitting to the training data, therefore their predictions for new data have high variance.

The RF algorithm counteracts this effect by constructing many decision trees, each trained on a subset of the data. The training data subsets are randomly chosen with replacement, and their outputs averaged for an overall prediction in a process known as bootstrap aggregation (‘bagging’, see Breiman, 1996). In this way, RF models retain the low bias of a decision tree, while also minimising the variance on predictions for new data.

1.4.2 Applications to astronomical problems

Using NNs, galaxy morphologies have been classified using citizen science labelled data in the Galaxy Zoo project (Banerji et al., 2010). More recently, CNN pipelines have been developed to classify galaxy images (e.g. Lukic et al., 2018; Ferreira et al., 2020; Cheng et al., 2020; Cavanagh & Bekki, 2020; Bickley et al., 2021), cross-match radio galaxies with their optical/infrared counterparts (Alegre et al., 2022), and predict cosmological parameters from simulations and mock observations (Hassan et al., 2020; Ntampaka et al., 2020; Villaescusa-Navarro et al., 2021).

RF models are widely used in a range of astronomical applications, and have been remarkably successful given the relative simplicity of the approach. The advantage of RF models over NNs is in the interpretability of the output models, as they indicate the relative importance of the input variables in reaching a prediction. In galaxy formation, RFs (and related tree-based methods) have been used for regression problems using both simulation and observational data, for example in predicting the properties of large scale structure (Lucie-Smith et al., 2018; Lovell et al., 2021; Li et al., 2022) and the properties of galaxies and haloes (Ucci et al., 2017; Nadler et al., 2018; Rafieferantsoa et al., 2018; Li et al., 2019; Cohn & Battaglia, 2020; Moews et al., 2021; Mucesh et al., 2021; Delgado et al., 2022; McGibbon & Khochfar, 2022). RFs are also useful for classifying galaxies and their components (e.g. Goulding et al., 2018; Bluck et al., 2022; Kwik et al., 2022).
1.5 Thesis Outline

In this thesis, I use the SIMBA simulations (Davé et al., 2019) to investigate the nature of gas inside galaxies and in the CGM at low redshift, and compare the predictions from SIMBA to relevant observations. The SIMBA simulations are unique among cosmological simulations as they contain a novel treatment of black hole growth via torque-limited accretion, which allows for a physically motivated treatment of black hole feedback. The simulations also include an on-the-fly treatment of the production, growth and destruction of dust particles, and a number of other updates on the preceding simulation, MUFASA (Davé et al., 2016). Chapter 2 is devoted to presenting the SIMBA simulations and detailing the physical models.

The science results of this thesis are organised into four chapters. Below I outline the science questions of each chapter.

**Chapter 3: The impact of quenching on galaxy profiles in SIMBA**

1. Can the SIMBA model reproduce observations of the size-mass relation (and its evolution) in galaxies of different types?
2. Where does the star forming gas live inside SIMBA galaxies?
3. How does star formation quenching proceed inside SIMBA galaxies? What physical mechanisms are responsible?

**Chapter 4: The Low Redshift Circumgalactic Medium in SIMBA**

1. What do the mass and metal budgets of halos look like in SIMBA? What is the metallicity of the halo components?
2. Can SIMBA reproduce observations of CGM absorption properties of star forming and quenched galaxies, of different masses?
3. How do the composition and absorption properties of the CGM depend on the feedback mechanisms?

**Chapter 5: The Physical Nature of Circumgalactic Medium Absorbers in SIMBA**

26
1. What is the distribution of column densities of absorbers in the CGM for galaxies of different types?

2. What physical conditions do H i and metal absorbers trace in the CGM?

3. How much of the absorption arises from photoionised versus collisionally ionised gas? How much arises from satellite galaxies?

**Chapter 6**: Mapping Circumgalactic Medium observations to theory using machine learning

1. Is it possible to use machine learning to predict the physical CGM conditions from absorption properties?

2. What quantities provide the most useful information for predicting physical conditions?

3. Can such models reproduce the phase space structure of the CGM?

Throughout this work, we assume a flat \( \Lambda \)CDM cosmology with \( \Omega_M = 0.3 \), \( \Omega_\Lambda = 0.7 \), \( \Omega_b = 0.048 \), \( H_0 = 68 \) km s\(^{-1}\)Mpc\(^{-1}\), \( \sigma_8 = 0.82 \) and \( n_s = 0.97 \). All logarithms assume base 10.
Chapter 2

The Simba Simulations

Gravity and hydrodynamics

Simba (Davé et al., 2019) is a suite of state-of-the-art cosmological simulations and is the successor to the earlier Mufasa simulations (Davé et al., 2016). Simba is run using a modified version of the gravity plus hydrodynamics code Gizmo (Hopkins, 2015) which uses the Gadget-3 tree-particle-mesh gravity solver (Springel, 2005) and a Meshless Finite Mass (MFM) solver for hydrodynamics. The fiducial Simba simulation models a 100 $h^{-1}$Mpc comoving volume from $z = 249 \rightarrow 0$, with $1024^3$ gas elements and $1024^3$ dark matter particles. We also vary the sub-grid AGN feedback prescription using 50 $h^{-1}$Mpc comoving volumes, with $512^3$ gas elements and $512^3$ dark matter particles. These are the No-Xray (X-ray feedback turned off), No-jet (X-ray and jet feedback turned off), No-AGN (all black hole feedback turned off) and No-feedback (no black hole or stellar feedback) runs; the details of these feedback modes are described below. In all cases, the mass resolution is $1.82 \times 10^7$ M$_\odot$ for gas elements and $9.6 \times 10^7$ M$_\odot$ for dark matter particles, and the minimum adaptive softening length is $\epsilon_{\text{min}} = 0.5 \, h^{-1}\text{kpc}$, which corresponds to 0.5% of the mean interparticle spacing between dark matter particles. We test resolution convergence using 25 $h^{-1}$Mpc comoving volumes: the first with $256^3$ gas elements and dark matter particles (the same mass resolution as the larger volumes), and a higher resolution run with $512^3$ gas elements and dark matter particles. The higher resolution run has 8 times the mass resolution of the fiducial volume ($2.3 \times 10^6$ M$_\odot$ and $1.2 \times 10^7$ M$_\odot$ respectively), and twice the effective spatial resolution of the main
<table>
<thead>
<tr>
<th>Simulation</th>
<th>Box size ($h^{-1}$Mpc)</th>
<th>No. of particles</th>
<th>Feedback model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simba 100</td>
<td>100</td>
<td>$2 \times 1024^3$</td>
<td>Full feedback (Stellar winds, AGN winds, jets and X-ray heating)</td>
</tr>
<tr>
<td>Simba 50</td>
<td>50</td>
<td>$2 \times 512^3$</td>
<td>Full feedback</td>
</tr>
<tr>
<td>Simba 25</td>
<td>25</td>
<td>$2 \times 256^3$</td>
<td>Full feedback</td>
</tr>
<tr>
<td>Simba 25 (High resolution)</td>
<td>25</td>
<td>$2 \times 512^3$</td>
<td>Full feedback</td>
</tr>
<tr>
<td>No-Xray</td>
<td>50</td>
<td>$2 \times 512^3$</td>
<td>SW, AGN, jets</td>
</tr>
<tr>
<td>No-jet</td>
<td>50</td>
<td>$2 \times 512^3$</td>
<td>SW, AGN</td>
</tr>
<tr>
<td>No-AGN</td>
<td>50</td>
<td>$2 \times 512^3$</td>
<td>SW</td>
</tr>
<tr>
<td>No-feedback</td>
<td>50</td>
<td>$2 \times 512^3$</td>
<td>None</td>
</tr>
</tbody>
</table>

Table 2.1: Simba runs used in this work.

Simba volume. Cosmological initial conditions are generated using MUSIC (Hahn & Abel, 2011) assuming a cosmology consistent with Planck Collaboration et al. (2016): $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$, $\Omega_b = 0.048$, $H_0 = 68$ km s$^{-1}$Mpc$^{-1}$, $\sigma_8 = 0.82$ and $n_s = 0.97$. Table 2.1 summarises the simulations used in this work.

Heating and cooling

Radiative cooling and photoionisation heating are implemented using the Grackle-3.1 library (Smith et al., 2017) in non-equilibrium mode, which offers two main advantages over the Grackle-2.1 library used in Mufasa that improve the accuracy of the baryonic thermal evolution. First, adiabatic and radiative terms are evolved together during the cooling time-sub-step, resulting in a more accurate and stable thermal evolution. Secondly, self-shielding is included self-consistently during the simulation run following the Rahmati et al. (2013) prescription, in which the metagalactic ionising flux is attenuated depending on gas density and assuming a Haardt & Madau (2012) spatially uniform ionising background.

Star formation

Star formation is modelled using an $H_2$-based Schmidt (1959) relation, where the $H_2$ fraction is computed using the sub-grid prescription of Krumholz & Gnedin (2011) based on metallicity and local column density, modified to account for variations in resolution (Dave et al., 2016). The star formation rate (SFR) is
thus calculated from the density of molecular gas $\rho_{H_2}$ and the dynamical time $t_{\text{dyn}}$ via SFR = $\epsilon_\ast \rho_{H_2}/t_{\text{dyn}}$, where $\epsilon_\ast = 0.02$ (Kennicutt, 1998). The neutral hydrogen content of gas particles is computed on the fly using the Rahmati et al. (2013) self-shielding prescription described above. This gives the total shielded gas; subtracting off the molecular hydrogen fraction gives the fraction of gas in H\textsc{i}.

We apply artificial ISM pressurisation at a minimum level required to resolve the Jeans mass in star-forming gas (Davé et al., 2016), applied above a hydrogen number density $n_H$ of $n_{\text{th}} > 0.13 \text{ cm}^{-3}$, such that the temperature of this gas has a floor of

$$\log(T/\text{K}) = 4 + \frac{1}{3} \log_{10} \frac{n_H}{n_{\text{th}}}. \quad (2.1)$$

We define ISM gas eligible for star formation as having a temperature less than 0.5 dex above this temperature floor, along with $n_H > n_{\text{th}}$. Note that not all ISM gas is actively star-forming, as it must also have $H_2$; at low metallicity, the threshold to form $H_2$ can be at densities well above $n_{\text{th}}$.

**Chemical enrichment**

The chemical enrichment model tracks 11 elements (H, He, C, N, O, Ne, Mg, Si, S, Ca, Fe) during the simulation, from Type II supernovae (SNII), Type Ia supernovae (SNIa), and Asymptotic Giant Branch (AGB) stars, as described in Oppenheimer & Davé (2006). For SNII we use the Nomoto et al. (2006) yields, interpolated to the metallicity of the gas being enriched. For SNIa we use the Iwamoto et al. (1999) yields, assuming that each supernova yields 1.4 M\text{\odot} of metals. For AGB stars, we follow Oppenheimer & Davé (2008), using lookup tables as a function of age and metallicity, assuming a helium fraction of 36%, a nitrogen yield of 0.00118, and a mass-loss rate computed from the Chabrier (2003) IMF. We assume solar abundances from Asplund et al. (2009) to convert metallicities into solar units.

**Star formation-driven winds**

Star formation-driven winds are an important mechanism for enriching the CGM, so we describe them and their chemical enrichment more fully. Massive stars drive
galactic outflows through a combination of SNII, radiation pressure and stellar winds; we use a subgrid model to represent the net effect of these phenomena via two-phase kinetic feedback. Wind particles with velocity $\mathbf{v}$ and acceleration $\mathbf{a}$ are ejected in the direction $\pm \mathbf{v} \times \mathbf{a}$ (within the smoothing length), characterised by two free parameters: the mass loading factor $\eta$, and the wind speed $v_w$. The scaling of mass loading factor with stellar mass is based on mass outflow rates from star forming regions in the Feedback In Realistic Environments (FIRE, Hopkins et al., 2014) simulations, computed by tracking individual particles in Anglés-Alcázar et al. (2017b). The scaling is described by a broken power law at $M_0 = 5.2 \times 10^9 \, M_\odot$:

$$
\eta(M_\star) \propto \begin{cases} 
9 \left( \frac{M_\star}{M_0} \right)^{-0.317}, & \text{if } M_\star < M_0 \\
9 \left( \frac{M_\star}{M_0} \right)^{-0.761}, & \text{if } M_\star > M_0 
\end{cases}
$$

(2.2)

with a flat $\eta(M_\star)$ for objects with fewer than 16 star particles, to allow growth in poorly resolved galaxies that would otherwise struggle to grow due to excessive feedback. $M_\star$ is obtained via an on-the-fly friends-of-friends finder applied to ISM gas and stars.

The wind velocity scaling with galaxy properties is also based on the FIRE simulations (Muratov et al., 2015), with an extra velocity kick $\Delta v(0.25R_{\text{vir}})$ to account for the difference in radius between the wind launch site (typically well within the ISM) and the larger radius of $0.25R_{\text{vir}}$ used in the FIRE scalings and further increased to account for hydrodynamic slowing:

$$
v_w = 1.6 \left( \frac{v_{\text{circ}}}{200 \, \text{km s}^{-1}} \right)^{0.12} v_{\text{circ}} + \Delta v(0.25R_{\text{vir}}).
$$

(2.3)

The circular velocity $v_{\text{circ}}$ is obtained using a baryonic Tully-Fisher-based scaling relation from $M_\star$. $\Delta v(0.25R_{\text{vir}})$ is computed based on the gravitational potential difference of a wind particle’s launch location versus $0.25R_{\text{vir}}$, as described in Davé et al. (2016).

To model the observed two-phase nature of galactic winds, 30% of wind particles are ejected hot, with a temperature set by the supernova energy ($u_{\text{SN}} = 5.165 \times 10^{15} \, \text{erg g}^{-1}$) minus kinetic energy, while the remaining particles are ejected at $T \approx 10^3 \, \text{K}$. Ejected wind particles are hydrodynamically decoupled to avoid numerical inaccuracies arising from gas elements with high Mach numbers relative to their surroundings, and cooling is shut off to allow hot winds to deposit their
thermal energy into the CGM. Outflowing wind particles are recoupled when 1) the particle’s velocity is similar to its surroundings and its density is less than that of the ISM, or 2) a density limit of less than 1% of the ISM threshold density is reached, or 3) a time limit of 2% of the Hubble time at launch has passed.

Metal loading by SNII is modelled by allowing wind particles to extract some metals from their surroundings at launch time, enabling winds to transport additional metals into the CGM. The metallicity added to the wind particles is given by:

\[ dZ = f_{\text{SNII}} y_{\text{SNII}}(Z)/\text{MAX}(\eta, 1), \]

where \( f_{\text{SNII}} = 0.18 \) is the mass loss fraction due to SNII, \( y_{\text{SNII}}(Z) \) is the metallicity-dependent yield for each species, and \( \eta \) is the mass loading factor. The metal mass is subtracted from the surrounding gas in a kernel-weighted manner; at all times metal mass is conserved. Dust in wind particles is assumed to be destroyed if the particle is launched hot, but remains fully intact if launched cool; this carries some dust into the CGM.

**Dust**

SIMBA tracks cosmic dust using a sub-resolution prescription, as a fraction of each gas element’s metal budget that is passively advected along with gas particles, as described in Li et al. (2019). The prescription broadly follows that in McKinnon et al. (2016), with some improvements. Dust grains grow via condensation following Dwek (1998) but with updated condensation efficiencies, as well as accretion of gas-phase metals via two-body collisions. Dust is destroyed by collisions with thermally excited gas following the analytic approximation of dust growth rate from Tsai & Mathews (1995). A mechanism for dust destruction via SN shocks (which enhance inertia and thermal sputtering of dust grains) is implemented following McKinnon et al. (2016). Dust is also instantaneously destroyed (dust mass and metals transformed into gas particles) in hot winds, during star formation, and in gas impacted by AGN feedback, except in cold star forming winds and radiative-mode Eddington AGN feedback to allow these winds to transport dust out of the galaxy. Dust that is destroyed is returned back to the gaseous metal phase. Li et al. (2019) showed that SIMBA predicts global galaxy dust properties in reasonable agreement with observations across cosmic time.
**Black hole growth and feedback**

SIMBA’s main improvement on MUFASA is the addition of black hole growth via torque-limited accretion (Hopkins & Quataert, 2011; Anglés-Alcázar et al., 2013, 2015) and AGN feedback via bipolar kinetic outflows. Black holes are seeded with $M_{\text{seed}} = 10^4 \, M_\odot$ into galaxies with a minimum stellar mass of $\gtrsim 10^{9.5} \, M_\odot$ and grown self-consistently during the simulation via a two-mode accretion prescription. For cold gas ($T < 10^5 \, \text{K}$) black hole growth is implemented following the torque limited accretion model of Anglés-Alcázar et al. (2017a) which is based on Hopkins & Quataert (2011), while for hot gas ($T > 10^5 \, \text{K}$) Bondi accretion (Bondi, 1952) is adopted. Unlike Bondi accretion, torque-limited accretion does not require the black hole to self-regulate its own growth (Anglés-Alcázar et al., 2015), which allows for a more physical AGN feedback model.

The AGN feedback implementation in SIMBA is designed to mimic the observed dichotomy in black hole growth modes seen in real AGN (e.g. Heckman & Best, 2014): a ‘radiative’ mode at high Eddington ratios ($f_{\text{Edd}}$) characterised by mass-loaded radiatively-driven winds, and a ‘jet’ mode at low $f_{\text{Edd}}$, characterised by high velocity jets of $\sim 10^4 \, \text{km s}^{-1}$. SIMBA’s AGN feedback model has three mechanisms: 1) a radiative mode at high Eddington ratios ($f_{\text{Edd}}$), 2) jet mode at low $f_{\text{Edd}}$ and 3) X-ray feedback. Radiative and jet mode feedback are implemented kinetically, with outflows ejected in a direction $\pm$ the angular momentum of the inner disk and with zero opening angle.

Radiative mode winds are ejected at the temperature of the ISM with a black hole mass-dependent outflow velocity, calibrated using X-ray detected AGN in SDSS (approximately following Perna et al., 2017). The transition to jet mode begins when $f_{\text{Edd}} < 0.2$, where outflows are ejected at the halo virial temperature and receive a velocity boost that increases as $f_{\text{Edd}}$ decreases, up to a maximum of 7000 km s$^{-1}$ above the radiative mode speed at $f_{\text{Edd}} = 0.02$. In our simulations, jet mode feedback occurs only for galaxies with $M_{\text{BH}} > 10^{7.5} \, M_\odot$, to prevent small black holes with temporarily small accretion rates from driving high-powered jets. The kinetic AGN feedback velocity is thus

$$v_w = 500 + 500(\log M_{\text{BH}} - 6)/3 + 7000 \log(f_{\text{Edd}}/0.2) \, \text{km s}^{-1}.$$  \hspace{1cm} (2.5)

These outflows are decoupled for $10^{-4} t_H$, so they explicitly do not impact the ISM; at full jet velocity, they typically go several tens of kpc before recoupling. The outflow mass loading factor is allowed to vary such that the momentum output...
of the black hole is always $20L/c$, where $L$ is the bolometric AGN luminosity. We include feedback due to X-ray heating from the accretion disk following Choi et al. (2012), for galaxies with $f_{\text{gas}} < 0.2$ and black holes with full-velocity jets. Our X-ray feedback implementation works in two ways: for non-ISM gas ($n_H < 0.13 \text{ cm}^3$), we directly increase the temperature of the gas, while for ISM gas half of the X-ray energy is used to give the gas particles a radial outwards kick, and the rest is added as heat.

These black hole accretion and feedback subgrid models reproduce the observed quenched galaxy fractions (Davé et al., 2019), galaxy–black hole correlations (Thomas et al., 2019), and the radio galaxy population (Thomas & Davé, 2022) very well. Globally, the jet mode is primarily responsible for quenching galaxies, while the X-ray feedback has a small but important role in suppressing residual star formation, and radiative AGN feedback has little impact on galaxy properties.

**Halo and galaxy identification**

Halos are identified during the simulation using Subfind, which is implemented in GIZMO. We do not identify sub-halos, but instead identify galaxies in post processing using a 6D friends-of-friends galaxy finder, using all stars, black holes, and gas elements with $n_H > n_{\text{th}}$. We assume a spatial linking length of $0.0056$ times the mean interparticle spacing (equivalent to twice the minimum softening length), and the velocity linking length is set by the local dispersion. Black holes and H I gas are assigned to the galaxy to which they are most gravitationally bound; we take the most massive black hole particle as the central black hole. Halos and galaxies are cross-matched using the yt-based (Turk et al., 2011) python package CAESAR. CAESAR also pre-computes a number of global galaxy properties, such as stellar mass and the instantaneous star formation rate (SFR). The SFR is taken to be the sum of the instantaneous SFRs of all the gas particles in the galaxy. Particle data is read using PyGadgetReader.

**Calibration**

The Simba model is primarily tuned to reproduce the galaxy stellar mass function at $z = 0$ and $z = 2$, and the $M_\star - \sigma$ relation at $z = 0$. For the former, the main free

---

1https://caesar.readthedocs.io

2https://github.com/jveitchmichaelis/pygadgetreader
parameters are the SF and AGN outflow wind speeds, which set the behaviour of the galaxy stellar mass function at the low and high mass ends respectively. For the latter, the main free parameter is the black hole accretion efficiency which is set to 10% and is tested by Simba’s ability to reproduce the $M_\star - M_{BH}$ relation.

**Successes of the Simba model**

The Simba model produces galaxies that are in good agreement with a range of observations. At low redshift, these include the star forming main sequence, quenched galaxy fractions, gas phase and stellar metallicities (Davé et al., 2019), black hole properties and coevolution with galaxy properties (Thomas et al., 2019), dust properties (Li et al., 2019), H$\text{I}$ and H$_2$ mass functions (Davé et al., 2020), the baryonic Tully-Fisher relation (Glowacki et al., 2020), and radio galaxy populations (Thomas et al., 2021). On larger mass scales, Simba reproduces X-ray scaling relations for massive halos (Robson & Davé, 2020) and low redshift Ly$\alpha$ absorption statistics along random lines of sight (Christiansen et al., 2020). Figure 2.1 shows selected examples of comparisons between Simba and observations: the star forming main sequence (Davé et al., 2019, top left),
Lyα absorption statistics from the IGM (Christiansen et al., 2020, top right), the galaxy H I mass function (Davé et al., 2020, bottom left), and the H I mass fraction in galaxies (Davé et al., 2019, bottom right). These examples demonstrate that SIMBA is able to accurately reproduce populations of star forming and quenched galaxies, as well as observations of H I gas content on small and large scales, both of which are important considerations for the ISM and CGM.
Chapter 3

The Impact of Quenching on Galaxy Profiles in Simba

3.1 Introduction

Modern galaxy formation models generally invoke feedback mechanisms associated with active galactic nuclei (AGN) to quench galaxies (see e.g. Somerville & Davé, 2015, and references therein). Beyond this general notion, there remains much uncertainty regarding the physical mechanisms by which such AGN feedback operates, what triggers such feedback, and with which galaxy and/or halo properties such feedback most strongly correlates.

Generally, quenching mechanisms fall into two broad categories. In merger quenching, major mergers are responsible for generating a starburst that evacuates the gas due to strong stellar and AGN feedback, leaving a dispersion-supported galaxy with little cold gas left to form stars (Springel, 2005; Hopkins et al., 2008). In halo quenching, feedback associated with AGN causes the halo gas around the galaxy to be heated, which starves the central galaxy of further accretion, eventually causing a cessation of star formation (Bower et al., 2006; Croton et al., 2006; Somerville et al., 2008; Gabor & Davé, 2015; Peng et al., 2015). Both models have observational support: for merger-driven quenching, observations clearly connect mergers with starbursts and AGN activity (e.g. Sanders & Mirabel, 1996), while for halo-driven quenching, bubbles seen in X-ray emission of galaxy clusters could potentially provide sufficient $PdV$ work to offset
gas cooling (McNamara & Nulsen, 2007). Many galaxy formation models, both semi-analytic and hydrodynamical, have implemented one or both of these forms of quenching in heuristic ways, and are thereby able to broadly reproduce the observed population of quenched galaxies.

A different set of constraints on quenching is provided by the bimodality in galaxy morphologies. At face value, merger quenching is attractive because it combines the quenching of star formation with a nearly concurrent transformation from spiral into elliptical. However, the existence of numerous “red disks” (Schawinski et al., 2010; Bundy et al., 2010) with late-type morphologies but little or no star formation suggest that the morphological transformation and quenching are not necessarily coeval. Meanwhile, simulations suggest that after halo quenching causes starvation, the typically denser environment can result in minor mergers or galaxy harassment that can transform morphologies without the need for a major merger (Oser et al., 2012; Gabor & Davé, 2012). However, the existence of rapidly-quenched systems such as post-starburst galaxies (e.g. Zabludoff et al., 1996; Wild et al., 2010) suggest that such a slow mechanism as starvation may not be sufficient to explain all quenched systems. Alternatively, it was shown that while mergers can lead to the formation of ellipticals, triggered AGN-regulated quenching is needed in order to freeze the post-merger morphology of a galaxy and prevent the disk re-formation (Gabor & Davé, 2012; Dubois et al., 2016). Hence it is likely that both quenching mechanisms are at play, with variations in importance that depend on galaxy mass, merger history, cosmic epoch, and environment.

To shed more light on galaxy quenching mechanisms, it is interesting to examine whether quenching occurs inside-out or outside-in, i.e. whether the bulge region drops in star formation rate prior to the disk, or vice versa. Inside-out quenching could indicate some internal process is responsible for evacuating or heating the star-forming gas in the central region. Inside-out quenching can also be associated with ‘wet compaction’ events due to minor mergers or tidal streams, leading to a ring of star-forming gas around the centre (Tacchella et al., 2016). Outside-in quenching might occur in particular if environmental processes such as gas stripping in the outskirts are the dominant quenching mechanisms. A process such as starvation may slowly affect the entire disk, causing an overall drop in star formation everywhere (van den Burgh, 1991; Elmegreen et al., 2002). Thus by measuring the star formation rate and gas profiles of galaxies that are transitioning to being quenched, it may be possible to discriminate between
Improving surveys can now measure the rate of galaxy growth via star formation as a function of galaxy radius, in massive galaxies that are likely to be on their way to being quenched. Recently Belfiore et al. (2018) used spatially resolved spectroscopy from the Mapping Nearby Galaxies at APO (MaNGA) Sloan Digital Sky Survey (SDSS) (Bundy et al., 2015) to derive star formation rates from Hα flux and compute radial profiles of sSFR for star forming (SF) and GV galaxies. They find that at low stellar masses, the SF galaxies have flat radial sSFR profiles, but with increasing stellar mass galaxies show more centrally suppressed star formation. In particular, GV galaxies of all masses have sSFR profiles that are suppressed at all radii, as is expected from galaxies that are on their way to being quenched, and also show much stronger central suppression, particularly for galaxies with log(M*/M⊙) \gtrsim 10.0. In addition, decreasing SFR at the centre indicates that the suppression is not merely due to the increasing mass of the stellar bulge component, but is evidence for inside-out quenching. Similar findings in the literature show that transition galaxies with high stellar mass typically have central suppression in their sSFR profiles (González Delgado et al., 2016; Coenda et al., 2018; Ellison et al., 2018; Liu et al., 2018; Sánchez et al., 2018; Spindler et al., 2018; Quai et al., 2019). Moreover, the fraction of inside-out quenching increases with stellar mass (Lin et al., 2019), suggesting that the fraction of inside-out quenching is higher than the fraction of outside-in quenching at a given stellar mass and environment.

At higher redshifts, SF galaxies can already be seen to develop central depressions in their SFR profiles as they begin their GV transition phase. At z \approx 1, SF galaxies with high mass (10.5 < log(M*/M⊙) < 11.0) show an enhancement in Hα, whereas less star forming galaxies of the same mass show central suppression of Hα and inferred sSFR (Nelson et al., 2016). These observations are reproduced in the high-resolution Feedback In Realistic Environments (FIRE) zoom simulations, as a consequence of bursty star formation (Orr et al., 2017). At z \approx 2, dust corrected sSFR profiles are found to be flat for galaxies with log(M*/M⊙) < 11.0, while for galaxies with log(M*/M⊙) \approx 11.0, the sSFR profiles are centrally suppressed by a factor of \sim 1 dex relative to the outskirts (Tacchella et al., 2018), demonstrating that inside-out quenching is already beginning in most massive star forming galaxies by z \sim 2. Inside-out quenching has also been independently observed via molecular gas profiles at z \sim 2 (Spilker et al., 2019). These observations support an inside-out quenching scenario, that
is, the fractional rate of new star formation is higher in the outskirts than in the bulge region.

These observations of star formation distribution within galaxies provide strong constraints on galaxy formation models. Modern cosmological simulations are able to reproduce a variety of observational galaxy trends despite substantial differences in their prescriptions for sub-grid processes such as AGN feedback, motivating new tests by which to assess models. Recently Starkenburg et al. (2019) presented radial profiles of sSFR from the Evolution and Assembly of GaLaxies and their Environments (EAGLE, Crain et al., 2015; Schaye et al., 2015) and Illustris (Genel et al., 2014; Vogelsberger et al., 2014) cosmological simulations, both of which are able to quench galaxies in broad agreement with observations. They demonstrate that while the profiles of simulated SF galaxies are in reasonable agreement with observations (Belfiore et al., 2018), both simulations produce GV galaxies that have centrally concentrated star formation at all stellar masses, in direct contrast to observations. This suggests that galaxies in cosmological simulations are predominantly quenching from outside-in putatively owing to halo heating, and that current cosmological models have difficulty reproducing the observed inside-out quenching. This discrepancy between state-of-the-art cosmological simulations and observations identifies sSFR profiles as a key test for galaxy formation simulations.

In this Chapter, we examine the profiles of star formation rate and gas content, relative to stellar mass profiles, within the Simba simulation (Davé et al., 2019). Simba produces galaxies that are in good agreement with observations for a range of probes including stellar mass, star formation rate, neutral and molecular gas properties, black hole properties (Thomas et al., 2019), and dust properties (Li et al., 2019). Most relevant for this work is that Simba yields a quenched fraction as a function of stellar mass that is in good agreement with observations (Davé et al., 2019), hence it provides a useful platform to study how quenching proceeds within these simulated galaxies. Simba includes three forms of AGN feedback, which heuristically describe radiative winds, bipolar jets, and X-ray radiation pressure, hence by running variants with these modules turned on and off, it becomes possible to examine which aspects of AGN feedback are responsible for quenching.

This chapter is organised as follows. §3.3 presents the size-mass relation and its redshift evolution for simulated galaxies compared to observations. In §3.4 we show radial profiles for star-forming and GV galaxies, compare with the observed
sSFR profiles, study the impact of different black hole feedback prescriptions on the radial profiles, study the differences in radial profiles between centrals and satellites, and examine the redshift evolution of radial profiles. Finally, in §3.5 we conclude and summarize.

3.2 Kennicutt-Schmidt relation

Since we are concerned with gas properties and the quenching of star formation, let us first verify that the star particles in Simba form from physically reasonable gas. Simba uses a volumetric surface density relation for star formation but does not employ the Kennicutt-Schmidt (K-S) relation directly. In contrast, the EAGLE project (Schaye et al., 2015; Crain et al., 2015) does directly implement a K-S relation for star formation. Since our simulations are not explicitly calibrated to reproduce the K-S relation, this forms a test for the Simba model.

Figure 3.1 shows the K-S relation of SFR surface density versus H I + H 2 surface density for Simba galaxies from the fiducial simulation with $M_\star > 10^{10} M_\odot$ at $z = 0$. The points are colour-coded by sSFR. We show a running mean of this for all galaxies that have sSFR > $10^{-1.8}$ Gyr$^{-1}$ and non-zero gas surface density as the blue dashed line. There are very few galaxies at low gas and SFR surface densities, so we display the mean values only if there are at least 5 galaxies in the given bin. We show the observed Kennicutt (1998) relation for star-forming galaxies (black dashed line), as well as the resolved relation from Bigiel et al. (2008) as the contours. Simba shows a reasonable agreement with the K-S relation, albeit slightly low in amplitude; this could be adjusted by increasing $\epsilon_\star$. Furthermore, it is seen that lower sSFR galaxies tend to lie below the K-S relation, which is consistent with observed early-type galaxies having lower star formation efficiencies (e.g. Davis et al., 2016). The magenta dashed line shows a running median using only H 2, which shows a roughly linear relation between $\Sigma_{SFR}$ and $\Sigma_{H2}$, and highlights how the turn-down in the K-S relation at low gas surface densities owes to an increase in the non-star forming H 1 content.

Overall, Simba reproduces the K-S relation reasonably well, which shows that the relationship between gas and SFR surface density, central to the analysis in this Chapter, is adequately represented.
Figure 3.1: Surface density of gas ($\Sigma_{H_1+H_2}$) as a function of surface density of SFR ($\Sigma_{\text{SFR}}$) for SF and GV galaxies in the $100h^{-1}$ Mpc SIMBA box, colour coded by their sSFR. $\Sigma_{H_1+H_2}$ and $\Sigma_{\text{SFR}}$ are computed within the half-light radius of each galaxy (see §3.3 for details on how sizes are computed). Galaxies with $\Sigma_{\text{SFR}} < 10^{-4.5}$ $M_\odot$ yr$^{-1}$ kpc$^{-2}$ have been plotted at that value for visibility. A running mean for the non-quenched galaxies ($\text{sSFR} > 10^{-1.8}$ Gyr$^{-1}$) is shown as the blue dashed line, and a running mean using only the molecular gas is shown as the magenta dashed line. The black dashed line is the best-fit relation to local spirals from Kennicutt (1998). The black contours are resolved galaxy observations from Bigiel et al. (2008). The observations have been scaled to a Chabrier IMF as assumed in SIMBA.
3.3 Size-mass relation

Since we will scale our profiles by galaxy half-light radius, it is important to first check whether SIMBA yields sizes that are in reasonable agreement with observations. For completeness, we do this at $z = 0 \rightarrow 2$, for star-forming and quenched systems, even though for the rest of this Chapter we will primarily focus on $z = 0$ non-quenched galaxies.

For each galaxy, we find $R_{\text{half}}$ by computing the half-luminosity radius in a particular band from individual stellar spectra of star particles using Pyloser\textsuperscript{1} (PYthon Line Of Sight Extinction by Ray-tracing). Pyloser generates a single stellar population (SSP) model using Flexible Stellar Population Synthesis\textsuperscript{2} (FSPS; Conroy et al., 2009; Conroy & Gunn, 2010) and uses this to compute a spectrum for each star particle, interpolated to its age and metallicity and assuming a Chabrier (2003) IMF. Convolving the spectrum with a given bandpass gives the magnitude in a particular band. Pyloser accounts for dust attenuation by computing the extinction to each star particle based on the kernel-smoothed line of sight dust column density, converted to $A_V$ assuming Milky Way scalings (Watson, 2011). Given $A_V$, we attenuate each star’s spectrum assuming a Calzetti et al. (2000) dust attenuation law for stars in galaxies with $\log \text{SFR}/\text{Gyr}^{-1} > 0$, a Milky Way dust extinction law (Fitzpatrick & Massa, 2007) for $\log \text{SFR}/\text{Gyr}^{-1} < -1$, and a linear combination of these in between; see Salim & Narayanan (2020) for a recent review of dust attenuation laws. Given each star’s (extincted) luminosity, we compute the half-light radius of every galaxy.

At $z = 0$, we compute $R_{\text{half}}$ in the SDSS $r$ band to compare with SDSS data, and for higher redshifts we choose the V band to compare with that quoted from $k$-corrected CANDELS (Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey) data. The radius of each galaxy is found by averaging the three 2D projections along the $x$, $y$ and $z$ axes, i.e. the sizes are computed along axes with random orientation with respect to the galaxies.

Figure 3.2 shows $R_{\text{half}}$ for all $M_*>10^{10}M_\odot$ galaxies in the $100h^{-1}$ Mpc SIMBA volume at $z = 2, 1, 0$ (left to right), colour coded by sSFR. We compare to van der Wel et al. (2014) and Allen et al. (2017) at $z = 2$, van der Wel et al. (2014) at $z = 1$, and Zhang & Yang (2019) at $z = 0$, separated into star forming and

---

\textsuperscript{1}https://pyloser.readthedocs.io/en/latest/

\textsuperscript{2}http://dfm.io/python-fspss/current/
Figure 3.2: Half-light radius as a function of stellar mass for $z = 2, 1, 0$ (from left to right) for galaxies in the $100h^{-1}$ Mpc SIMBA box, colour coded by their sSFR. The half-light radii are computed in the V band for $z = 2, 1$, and in the R band for $z = 0$. The dark blue and magenta lines are the running medians for the star forming and passive galaxies respectively; the solid lines represent the main SIMBA volume, and the dashed lines represent the high resolution volume. The horizontal dashed black lines show the effective size resolution limit of the fiducial simulation, below which galaxies are not well resolved. Observations are shown from van der Wel et al. (2014) and Allen et al. (2017) at $z=2$, van der Wel et al. (2014) at $z=1$ and Zhang & Yang (2019) at $z=0$. The sizes of the star forming galaxies are in broad agreement with the observations at $z = 0$, while at higher redshifts they are smaller than observed. Passive galaxies have sizes in a good agreement with observations at $z = 0$, while at higher redshifts they are larger than their star forming counterparts at $z = 0$. Galaxies in the high resolution volume are consistent with the fiducial volume at all redshifts except for passive galaxies at $z = 0$. 
quiescent. At each redshift we show separate running medians for the SF and passive galaxy populations (magenta and blue lines), defining star forming as \( \text{sSFR} > 10^{-1.8+0.3z} \text{Gyr}^{-1} \) as in Davé et al. (2019). The effective spatial resolution is indicated by the dashed black lines in each panel. This is the radius out to which the gravitational force is softened, given by the minimum Plummer softening scale \((0.5h^{-1}\text{kpc}, \text{comoving})\) multiplied by a factor of 2.8 for our assumed cubic spline kernel (Springel, 2005).

At \( z = 0 \), the sizes of the SF galaxies are in good agreement with the observations. The good agreement with data for the SF galaxies is an important success for Simba; there was no tuning done to obtain this agreement. In contrast, passive galaxies have sizes that are significantly larger than the observations at \( M_\ast \lesssim 10^{11.5} M_\odot \). In fact, the passive galaxies are slightly larger than the SF galaxies at all masses, which is the opposite of what is seen for real galaxies. This indicates that we are not reproducing the compact nature of the stellar distribution in passive galaxies, particularly at low masses. This was already noted at \( z = 0 \) in Davé et al. (2019). At \( z = 1 \) and \( z = 2 \), we see that passive galaxy sizes are in better agreement with observations, but here the star-forming galaxies are too small. We have checked that we see the same trends in the simulations without X-ray and/or jet feedback, showing that this does not owe to the AGN feedback model in Simba.

By examining stellar surface density images of Simba galaxies, we have noticed that our SF galaxies do not have the extended thin stellar disks that are common to real SF galaxies. They typically have a gas component that has settled into a thinner disk, but a much puffier thick disk or even spheroidal stellar distribution. Unlike stars, gas particles in the simulation are able to dissipate energy through hydrodynamic interactions, allowing them to settle into disks more easily than the stellar component. Since the \( r \) or \( V \) band half-light radii of the galaxies generally trace the stellar component, this indicates that something is puffing out stellar orbits. We have checked that newly-formed stars lie in a thin disk.

One possibility is numerical resolution, as older stars in present-day galaxies have undergone dozens of orbits where two-body effects and other dynamical noise can artificially heat the orbits. We investigate this by looking at the higher resolution \( 25h^{-1}\text{Mpc}, 2 \times 512^3 \) particle Simba volume. The dashed lines in Figure 3.2 show that at higher resolution, the sizes of star forming galaxies and the high redshift passive galaxies agree with the sizes of the lower resolution box, showing that these sizes are numerically converged. This is not the case, however, for the
passive galaxies at $z = 0$, where the increased resolution has decreased the sizes, particularly at high stellar masses. This indicates that the large sizes of passive galaxies in the main simulation is likely due to numerical heating of stellar orbits, since these galaxies are composed almost entirely of star particles, with little gas. The fact that this only appears at late epochs is consistent with the idea that it is an effect that happens over many orbital periods. Ludlow et al. (2019) pointed out that overly small softening values can actually decrease resolution owing to two-body scattering effects, so SIMBA’s adaptive gravitational softening may exacerbate this issue. It could also be that there is some missing physics in SIMBA that compacts low mass galaxies during quenching (Tacchella et al., 2017), but given that SIMBA does produce quite compact galaxies at $z \sim 2$, we favor the explanation of numerical heating. We note that star-forming galaxies will not suffer from this heating as much, because its stars were formed more recently in a thin disk.

Looking at the higher redshifts, at $z = 2$ the SF galaxies are significantly smaller than the observations, by a factor of $\sim 1.5$. The passive galaxies are in good agreement with the observations at this redshift, however the SF galaxies represent the majority of the population. By $z = 1$, there is a larger population of passive galaxies which are in broad agreement with the observations, and SF galaxies are still smaller than their observational counterparts. The small sizes of the high redshift galaxies indicate that the SIMBA galaxies grow more rapidly since $z = 2$ than the real galaxies, suggesting that the growth modes for SIMBA galaxies differ from that in real galaxies.

Figure 3.3 quantifies the median size growth rate, showing the redshift evolution of $R_{\text{half}}$ in the V band for central galaxies with $M_*$ within 5% of $5 \times 10^{10} M_\odot$. The galaxies are separated into star forming and passive using the same sSFR $> 10^{-1.8+0.3(z)} \text{Gyr}^{-1}$ cut as before. We choose this mass range to compare to the evolution of galaxies with $5 \times 10^{10} M_\odot$ in van der Wel et al. (2014), which are shown in the figure. We also show the $z = 0$ size measurements for this mass from SDSS Zhang & Yang (2019). By focusing on a particular mass, this plot allows us to examine the redshift dependence of the galaxy sizes.

The best fits for the evolution of the median sizes in SIMBA give sizes that scale as $(1 + z)^{-0.78}$ for the SF galaxies, consistent with observations showing a scaling of $(1 + z)^{-0.75}$ for star-forming systems (van der Wel et al., 2014). For passive galaxies, the median sizes scale as $(1 + z)^{-1.06}$, which is essentially a $(1 + z)^{-1}$ evolution for the passive galaxy populations. This is consistent with expectations
Figure 3.3: Evolution of the V band half-light radii of star forming (light blue) and passive (magenta) central galaxies at $M_\star \sim 5 \times 10^{10} M_\odot$. The solid lines show the running medians at each redshift and the shaded regions enclose 50% of the data. The best fits to the evolution of the median sizes are $5.2 (1+z)^{-0.78}$ and $5.9 (1+z)^{-1.06}$ for star-forming and passive centrals respectively. The dashed lines show the corresponding redshift evolution for observations of V band galaxies sizes from van der Wel et al. (2014), separated into star forming (blue) and quiescent (red) galaxies. The squares at $z = 0$ are the corresponding R band half-light radii from SDSS (Zhang & Yang, 2019), offset by $\pm 0.01$ for clarity.
for a simple disk formation model (Mo et al., 1998), and that passive galaxies do not undergo any compaction when they quench out of the star-forming sequence. However, this is inconsistent with observations showing a scaling of \((1+z)^{-1.48}\) for passive systems (van der Wel et al., 2014). We note that our star-forming galaxy amplitude appears too low when compared to van der Wel et al. (2014), but their fitting function at \(z = 0\) also lies noticeably above the SDSS measurement from Zhang & Yang (2019); it is beyond the scope here to examine why these two observational results disagree, albeit mildly. Finally, we note that our spatial resolution is fixed in comoving coordinates, which means that it scales as \((1+z)^{-1}\) in physical coordinates. Hence the scalings of passive galaxies are consistent with an evolution in the numerical softening length, although our actual softening values are nominally smaller than galaxy sizes.

Interestingly, including dust extinction in our computation of \(R_{\text{half}}\) has a substantial effect on the star forming galaxies. Without dust extinction, the size evolution goes as \((1+z)^{-1.16}\) and \((1+z)^{-1.11}\) for star forming and passive galaxies, respectively, making the evolution of star forming and passive galaxy evolution essentially the same. The substantial change in size evolution of the star forming galaxies is due to the increase in size at high redshift due to dust. Dust attenuation obscures light preferentially at the centres of galaxies, increasing the sizes, thus bringing the sizes into closer agreement with observations than without dust. This is particularly true for high redshift star forming galaxies as these objects contain the most dust (Li et al., 2019), while for passive galaxies with little dust the effect is weak.

Other simulation projects have had varying levels of success in reproducing the galaxy sizes. Illustris TNG is able to reproduce the mass-size relation of both SF and quenched galaxies at \(z = 0\) (Genel et al., 2018), showing good agreement with SDSS (Shen et al., 2003; Bernardi et al., 2014) and van der Wel et al. (2014) extrapolated to \(z = 0\). They are able to do this in part because they tune their simulation to match the \(z = 0\) mass-size relation. However, the success of having quenched galaxies smaller than star-forming is something SIMBA fails, potentially in part because TNG has \(\approx 20\times\) better mass resolution than SIMBA. Likewise, EAGLE (Crain et al., 2015; Schaye et al., 2015) has demonstrated that they match the low redshift Shen et al. (2003) SDSS measurements of star-forming galaxy sizes, though once again they tune their simulation to do so. They are also able to get quenched galaxies smaller than star-forming, with a mass resolution \(\approx 10\times\) better than SIMBA’s. Horizon-AGN (Dubois et al., 2014) has shown that
their disk-dominated galaxy sizes are in agreement with van der Wel et al. (2014) at $z = 0.25$ to within a factor of $\sim 2$, but similar to SIMBA their elliptical galaxies are less compact than their disk galaxies (Dubois et al., 2016). They attribute this discrepancy to their limited spatial resolution, which is comparable to the resolution in SIMBA.

In summary, SIMBA produces low-$z$ SF galaxy half-light sizes in good agreement with observations. These constitute the galaxies we are most concerned with for the profiles in this work. However, SIMBA predicts that quenched galaxies have slightly larger sizes than star-forming systems at $z = 0$ which is opposite to what is observed. For the rest of this Chapter, we will investigate the radial profiles of star-forming or GV galaxies in various physical quantities, scaled by the half-light radii, primarily at $z = 0$. We will not consider galaxies that are fully quenched. Hence while it is a notable discrepancy that SIMBA does not reproduce the sizes of today's quenched galaxies, this is not critical for the results in the remainder of this work.

3.4 Radial Profiles

3.4.1 Galaxy selection

We now examine galaxy radial profiles in SIMBA, focusing mainly on star-forming and GV systems at $z = 0$, though we will look at redshift evolution in §3.4.6. We will focus on massive galaxies with $M_\star \geq 10^{10}$ $M_\odot$, corresponding to $\geq 550$ star particles, in order to ensure we can get sufficient resolution for robust profile measurements, and also because this is where observations for comparison are most abundant and secure. We separate star-forming and GV galaxies via a cut in SFR($M_\star$), described below. We will not consider quenched galaxies further.

Figure 3.4 illustrates the way we select our simulated galaxy sample for this work. This shows the star-forming main sequence, $M_\star$ vs. SFR, for all galaxies with $M_\star > 10^9$ $M_\odot$ in the $100h^{-1}$ Mpc SIMBA box at $z = 0$. Points are colour coded by their H I mass to stellar mass ratio, $f_{HI}$. Vertical dotted lines denote the low, intermediate, and high mass bins that will be used for this work; all galaxies above $M_\star > 10^{11}$ $M_\odot$ are grouped into the high mass bin. We will not consider galaxies with $M_\star < 10^{10}$ $M_\odot$ further. The magenta lines demarcate the GV, as we discuss below.
Figure 3.4: Star formation rate–stellar mass relation for all galaxies in the SIMBA 100 $h^{-1}$Mpc box at $z = 0$, colour coded by their H I mass to stellar mass ratio, $f_{\text{HI}}$. The magenta dashed lines show the selection of the GV galaxies from Belfiore et al. (2018), with the upper and lower lines corresponding to their SFMS and SFMS minus 1 dex, respectively. The black dotted lines show the three considered stellar mass bins labelled low ($10^{10} < M_* < 10^{10.5}M_\odot$), intermediate ($10^{10.5} < M_* < 10^{11}M_\odot$) and high mass ($M_* > 10^{11}M_\odot$).
Table 3.1: Median global galaxy properties and number of galaxies in the three mass bins for SF and GV galaxies.

<table>
<thead>
<tr>
<th>Median</th>
<th>Star forming</th>
<th>GV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>Int</td>
</tr>
<tr>
<td>$N_{\text{gal}}$</td>
<td>1,767</td>
<td>603</td>
</tr>
<tr>
<td>log($M_*/M_\odot$)</td>
<td>10.2</td>
<td>10.7</td>
</tr>
<tr>
<td>SFR ($M_\odot$ yr$^{-1}$)</td>
<td>1.56</td>
<td>3.64</td>
</tr>
<tr>
<td>log (sSFR / yr$^{-1}$)</td>
<td>-9.98</td>
<td>-10.16</td>
</tr>
<tr>
<td>log($M_{\text{HI}}/M_\odot$)</td>
<td>9.8</td>
<td>9.9</td>
</tr>
<tr>
<td>log($M_{\text{H}<em>2}/M</em>\odot$)</td>
<td>9.7</td>
<td>10.1</td>
</tr>
<tr>
<td>$R_{\text{half}}$ (kpc)</td>
<td>3.8</td>
<td>5.4</td>
</tr>
</tbody>
</table>

This figure shows the usual structure of the main sequence. There is a blue cloud of star-forming galaxies extending towards low masses. As $M_*$ increases there are more quenched galaxies with low SFR, primarily as a result of AGN jet feedback (Davé et al., 2019), resulting in the development of a red sequence. As galaxies quench from the SF blue cloud to the red sequence, they move across $M_*$–SFR space in the transitional GV region. There are occasional rejuvenations in the other direction, but they are quite rare in Simba, and additionally galaxies do not tend to loiter in the GV. It turns out the timescales to quench are bimodal, with slow quenching times of $0.1 \times t_{\text{Hubble}}$ and fast quenching times of $0.01 \times t_{\text{Hubble}}$ (see Rodríguez Montero et al., 2019, for a full analysis of quenching times and rejuvenations in Simba). Thus in Simba the vast majority of GV galaxies are on their way to being quenched eventually.

To demarcate the GV, we follow the definition of Belfiore et al. (2018). Their star forming main sequence (SFMS) is given by:

$$\log(\text{SFR} / M_\odot \text{ yr}^{-1}) = 0.73 \log(M_*/M_\odot) - 7.33,$$

with a scatter of 0.39 dex. The upper magenta dashed line in Figure 3.4 is the lower boundary of their star SF galaxies which corresponds to $1\sigma$ below the SFMS. They define GV galaxies as those with with SFR down to 1 dex below this, which is indicated by the lower magenta line. These demarcations have a sub-unity slope, so they do not directly correspond to fixed cuts in sSFR, but for our mass range, star-forming galaxies defined this way typically have sSFR $\gtrsim 10^{-10.5}$ yr$^{-1}$.

With this selection, in Simba at $z = 0$ we obtain 1,767 star-forming galaxies in the low-mass bin ($10^{10} M_\odot < M_* < 10^{10.5} M_\odot$), 603 in the intermediate-mass bin
(10^{10.5} \ M_{\odot} < M_\star < 10^{11} \ M_{\odot})$, and 105 in the high-mass bin ($M_\star > 10^{11} \ M_{\odot}$). For GV galaxies, these mass bins contain 1465, 373, and 53 galaxies, respectively. Table 3.1 shows median values for various galaxy properties in our star-forming and GV samples.

### 3.4.2 Star-forming vs GV profiles

We now examine the profiles in various physical quantities of the star-forming and GV samples defined as above. To generate profiles, we first rotate each galaxy such that it is face on, aligned with the angular momentum vector of all its cold gas and stars. We compute individual SFR and $M_\star$ surface density radial profiles from the gas and star particles, and use these to compute an sSFR profile for each galaxy:

$$sSFR(R) = \frac{\Sigma_{\text{SFR}}(R)}{\Sigma_{M_\star}(R)},$$

where the $\Sigma$ represents the surface density within an annulus centred at radius $R$ in the subscripted quantity. Where radial bins contain no gas, we take the SFR to be zero. Changes in sSFR can be due to a change in molecular gas fraction ($f_{H_2}$) or the star formation efficiency (SFE):

$$sSFR = \frac{SFR}{M_{H_2}} = \frac{M_{H_2}}{M_\star} = SFE \times f_{H_2}.$$  

(3.3)

To isolate which of these is responsible for any trends in sSFR, we decompose our sSFR profiles into profiles of $f_{H_2}$ and SFE. These profiles are computed from the profiles of SFR, $M_\star$ and $H_2$ surface density:

$$SFE(R) = \frac{\Sigma_{\text{SFR}}(R)}{\Sigma_{H_2}(R)}, \quad f_{H_2}(R) = \frac{\Sigma_{H_2}(R)}{\Sigma_{M_\star}(R)}.$$  

(3.4)

We centre profiles on the position of the galaxy’s central black hole (which in SIMBA is tied to the location of the lowest potential), or in rare cases where there is no black hole then we choose the centre of mass of the star particles; our results are essentially unchanged if we always just use the centre of mass. All profiles are normalised to the half-light radius $R_{\text{half}}$, that we compute as in §3.3 using the extinguished SDSS $r$ band light at $z = 0$ for comparison to low-$z$ SDSS data, and the extinguished rest-frame $V$ band for higher redshift comparisons, in order to mimic the band typically quoted from observations. Our results are not sensitive to this
choice within rest-frame optical bands.

Figure 3.5 shows radial surface density profiles scaled by $R_{\text{half}}$ for SFR, sSFR, H$_1$, H$_2$, SFE and $f_{H2}$, separated into the mass bins shown in Figure 3.4. The blue→green lines show the SF galaxies; the purple→orange lines show the GV galaxies. In each mass bin, the overall profile is the Tukey biweight of the individual profiles. The Tukey biweight is a robust estimator for the mean that ignores outlying points (see e.g. Belfiore et al., 2018; Starkenburg et al., 2019). It is qualitatively similar to a median and also gives a robust estimation of the sample standard deviation (referred to as the biweight scale estimator). We also compute the error coming from cosmic variance via jackknife resampling over the 8 simulation sub-octants. We add this error in quadrature with the Tukey biweight scale estimator divided by $\sqrt{N}$ (where $N$ is the number of galaxies in the mass bin), to obtain the shaded region around each line. The vertical dotted lines show the spatial resolution in units of the median $R_{\text{half}}$ for each mass bin, computed separately for SF and GV galaxies. Predictions on scales smaller than these lines are likely to be compromised by resolution effects.

These profiles illustrate the stark structural differences between SF and GV galaxies in their star formation and gas profiles. Overall, the GV galaxies have lower star formation and cold gas content at almost every radius compared to their SF counterpart at same mass. This is not unexpected, since these galaxies are on their way to being quenched, and thus will end up with very low SF and cold gas contents. For both SF and GV galaxies, the total star formation rate increases with stellar mass for both galaxy types, as quantified in Table 3.1. The SFR profiles show that star formation activity drops at a similar rate for all masses beyond $R \gtrsim 0.5R_{\text{half}}$, but within this radius the profiles are mass-dependent. For the SF galaxies, the highest mass galaxies show the strongest reduction of SFR in their centres, whereas for GV galaxies SFR drops towards zero in the centre for all mass bins similarly. The SF galaxies show more extended star formation, dropping to near zero SFR around $3R_{\text{half}}$, as opposed to $1.25R_{\text{half}}$ for the GV galaxies.

The sSFR profiles (upper right panel) show the same trend as the SFR; in general the galaxies have undergone quenching at both their centres and their outskirts, with a band of star formation occurring between $\sim 0.25 - 1R_{\text{half}}$ for the GV galaxies and between $\sim 0.5 - 2.5R_{\text{half}}$ for the SF galaxies. A decrease in sSFR may be due to either an increase in stellar mass or a decrease in SFR (Spindler et al., 2018); the suppression of the SFR shows that the sSFR suppression is not
Figure 3.5: Radial profiles of SFR, sSFR, H1, H2, SFE and fH2 as a function of stellar mass. The green/blue lines show the SF galaxy profiles, and the orange/purple lines show the GV galaxy profiles, in low, intermediate, and high mass bins as indicated. The displayed radial profiles are Tukey biweights of the individual galaxy profiles in each mass bin. The light shaded regions around each line show the Tukey biweight scale estimator divided by \( \sqrt{N} \), combined with cosmic variance uncertainties from jackknife resampling over the 8 simulation sub-octants. The vertical dotted lines show the spatial resolution in units of the median half-light radius for each mass bin, computed separately for SF and GV galaxies. Overall, GV galaxies show substantially different profiles than SF galaxies, with globally lower star formation and gas content, with the star formation confined to within \( R_{\text{half}} \) as opposed to out to several half-light radii, and with a strong central dip in the SFR and sSFR profiles. This points towards two distinct mechanisms: inside-out quenching causing the suppression of star-forming gas in the central regions, and outside-in quenching suppressing the star formation efficiency in the outskirts.
simply due to the large stellar mass of the central galactic bulge. There is also a mass dependence – a higher stellar mass translates to a lower overall level of sSFR, as expected from the sub-linear slope of the star-forming main sequence. The highest mass galaxies generally have the largest black holes and the most powerful AGN feedback, so they are expected to be in the process of quenching. Indeed, high-mass SF galaxies also show some sSFR (and SFR) suppression in the centre, indicating that these galaxies are likely affected by the same mechanism(s) as the GV galaxies. This suggests that massive SF galaxies are in the early stages of the quenching process that has more substantially affected the GV galaxies, which is consistent with the idea of slow transition to the red sequence in massive SF galaxies inferred from observations (Schawinski et al., 2014).

The middle panels of Figure 3.5 show the cold gas (\(\text{H}^\text{i}\) and \(\text{H}_2\)) mass density profiles. For star-forming galaxies, similar to the \(\Sigma_{\text{SFR}}\) profiles, the \(\Sigma_{\text{HI}}\) profiles show that \(\text{H}^\text{i}\) and \(\text{H}_2\) surface densities decrease with increasing radial distance, but are suppressed in the centre of the galaxy. There is only a weak dependence on mass, as the \(\text{H}^\text{i}\) profile is fairly universal with the only difference being a somewhat more rapid drop in the outskirts in more massive galaxies, and the \(\text{H}_2\) profile being mildly lower for the lowest mass galaxies. Observationally, it has been noted that SF galaxies exhibit a near-universal cold gas surface density profile when scaled by size (Bigiel & Blitz, 2012), which is qualitatively consistent with what we find. Essentially, the increase in overall galaxy size yielding a larger gas content is offset by the decrease in cold gas content in more massive galaxies, resulting in profiles that are broadly independent of mass.

Comparing the SF galaxies to the GV galaxies shows marked differences. In \(\text{H}^\text{i}\), the gas content has dropped sharply in the outskirts relative to SF galaxies. This could owe to starvation not replenishing cold gas in the outskirts as it moves inwards to form stars, evaporation by a growing hot halo that is more prevalent around GV galaxies than SF systems at the same mass, and/or environmental processes where interactions with nearby satellite galaxies have dynamically heated the cold gas. We will examine satellite vs. central profiles in §3.4.5.

For \(\text{H}_2\), the situation is more curious. In SF galaxies, the \(\text{H}_2\) surface density peaks at a higher value, shows a greater drop in the middle, and drops off more quickly in the outskirts than the \(\text{H}^\text{i}\). All these trends are consistent with a higher \(\text{H}_2\) fraction in denser gas, with the exception of the central dip; in this case, it could be that extraplanar gas in the foreground of our face-on galaxies is predominantly in \(\text{H}^\text{i}\) form, and so fills in the central region in projection.
For the GV galaxies, there is a drop in the H$_2$ content in the inner parts relative to SF galaxies, but in the outskirts, the H$_2$ profile is actually shallower in the outskirts, particularly for more massive galaxies. Yet despite the H$_2$ having quite an extended profile, the SFR surface density remains well confined to the central region. In other words, there is still substantial H$_2$ in the outskirts of GV galaxies, but it is not forming stars. This could owe to the fact that the physical densities are substantially lower in the outskirts, and since in our simulations SFR$\propto \rho^{1.5}$, this can cause a strong drop in SFR even if the projected $\Sigma_{H_2}$ remains high. Interestingly, if one postulates a threshold H$_2$ surface density of $\log \Sigma_{H_2} = 1.1$ in order to have sufficient physical density for star formation (dotted horizontal line), then this would truncate star formation at $\sim 3R_{\text{half}}$ in SF galaxies and $\sim R_{\text{half}}$ in GV galaxies, which is essentially what is seen in the $\Sigma_{SFR}$ plot in the upper right. Such a surface density threshold is approximately coincident with the turn-down in $\Sigma_{SFR}$ seen in the K-S at low gas surface densities (Fig. 3.1).

The bottom two panels of Figure 3.5 quantify the connection between star formation and dense gas more clearly. Since $s\text{SFR} = \text{SFE} \times f_{H_2}$, we can subdivide the sSFR profile into profiles for SFE (left) and $f_{H_2}$ (right) to better understand why GV galaxies have suppressed star formation. In essence, the plot in the upper right is convolution of the two bottom plots.

It is clear that the H$_2$ fractions (bottom right), while lower for GV galaxies, have a similar radial trend between the SF and GV galaxies: the gas fraction is relatively flat except in the central region where it drops. Meanwhile, the SFE shows a rapid decline with radius, and no central drop. We thus see that the there are two separate effects which conspire to take the sSFR profiles from SF to GV: In the outer regions, the sSFR is suppressed owing to a rapid SFE decline in the outskirts; this is primarily governed by the physical density of the star-forming gas. In contrast, in the innermost region the H$_2$ fraction drops quickly, and hence the central hole in sSFR in GV and massive SF galaxies primarily owes to molecular gas being removed either by heating or expulsion. Gas in the centre is forming stars at a similar efficiency as gas at the peak in the SFR profile, but there is simply much less of it.

In summary, GV galaxies show substantially different profiles than SF galaxies, with overall lower star formation and gas content, and the star formation being confined to within $R_{\text{half}}$ as opposed to out to several half-light radii. There is a strong central dip in the SFR as well as sSFR profiles, which occurs in all GV galaxies along with massive SF galaxies. The SFR drop in GV systems thus
Figure 3.6: The sSFR radial profiles for all SF (left) and GV (right) galaxies for the full Simba run, with increasing $M_*$ bins shown in green to blue. The solid lines show the profiles in the frame rotated such that the galaxies are face-on, the dashed line for the highest mass bin shows the non-rotated profiles (i.e. the randomly orientated case). Observations of sSFR profiles of galaxies from MaNGA SDSS (Belfiore et al., 2018) are shown as the pink/purple lines. The radial profiles are Tukey biweights of the individual galaxy profiles in each mass bin. The light shaded regions correspond to the Tukey biweight scale estimator divided by $\sqrt{N}$, combined with cosmic variance uncertainties from jackknife resampling over the 8 simulation sub-octants. The vertical dotted lines show the spatial resolution in units of the median half-light radius for each mass bin, computed separately for SF and GV galaxies.

SF galaxies show a reasonable good agreement with the MaNGA data at low mass, but the inside-out quenching occurring in massive star-forming galaxies appears to be too strong. For the GV galaxies, Simba reproduces a drop in sSFR in the central regions seen in the data, however in the outskirts the sSFR is also suppressed, in conflict with the data.

appears to be driven by two different effects in the inner and outer regions. In the central region, the amount of star-forming gas is suppressed, while in the outskirts, molecular gas is still present but has a suppressed efficiency of forming into stars. Thus it appears that quenching in Simba galaxies occurs both inside-out and outside-in, and may indicate two distinct physical mechanisms. We will examine which AGN feedback mechanisms are responsible for these effects in §3.4.4. Next we conduct a more careful comparison to observations of SF vs. GV galaxy profiles.
3.4.3 Comparison to observations

The sSFR profiles of star-forming versus GV galaxies has been measured in the SDSS MaNGA Survey by Belfiore et al. (2018). They showed that the sSFR profiles of GV galaxies tend to be strongly suppressed in the centres relative to SF galaxies. Starkenburg et al. (2019) examined these trends in the Illustris and EAGLE simulations, and surprisingly found that despite these models quenching galaxies globally as observed, the sSFR profiles of similarly-selected GV galaxies did not show any strong central suppression in either simulation, but rather a centrally concentrated sSFR profile that was qualitatively similar to that in SF galaxies. They thus highlighted this comparison as a key test of how galaxies quench radially in models, one that some state of the art simulations fail to satisfy.

In the more recent Illustris TNG simulation, radial profiles show some central suppression of star formation, but only in the most massive galaxies (Nelson et al., 2019). In this section we undertake the Starkenburg et al. (2019) comparison in Simba.

Figure 3.6 shows the comparison of sSFR profiles for SF galaxies (left panel) and GV galaxies (right panel) in Simba, with increasing \( M_\star \) bins shown in green to blue, versus profile in the same \( M_\star \) bins from Belfiore et al. (2018, orange to purple). All profiles are scaled to the \( r \)-band half-light radius \( R_{\text{half}} \), and are computed over all profiles in each mass bin using the Tukey biweight estimator as done in Belfiore et al. (2018). These Simba profiles are exactly as plotted in Figure 3.5, but here we zoom in on the central region (\( R/R_{\text{half}} < 2 \)), split SF and GV into separate panels, and show observations overlaid.

For the SF galaxies (left panel), Simba predicts sSFR profiles that are qualitatively consistent with observations. From the centre outwards, the profiles show a rise in the central region, and then a mostly flat profile in the outskirts. The rise is faster in higher mass galaxies, indicating greater suppression of central SFR in these systems. However, there are some clear discrepancies versus data, particularly for higher-mass SF galaxies. First, the sSFR values peak at smaller radii in Simba (\( R_{\text{peak}} \sim 0.5R_{\text{half}} \)) versus observations (\( R_{\text{peak}} \sim 1 - 1.5R_{\text{half}} \)). A more blatant discrepancy is seen in the inner region, where the drop seen in the observations is not nearly as abrupt as that predicted in Simba for \( M_\star > 10^{10.5} \, M_\odot \) galaxies. Hence the agreement between Simba and the MaNGA data is reasonable for lower-mass SF galaxies, but the inside-out quenching already occurring in massive star-forming galaxies appears to be too severe.
For the GV galaxies (right panel), it is clear that Simba produces a drop in the central sSFR. This is in good agreement with the Belfiore et al. (2018) data, at least better than other current simulations (Starkenburg et al., 2019). The sSFR starts at similar values at its peak in Simba and in the data \( \text{sSFR} \approx 10^{-11} \text{yr}^{-1} \), and drops by an order of magnitude or more towards the middle. The main difference is that the decline is more gradual in the data, starting at around \( \sim R_{\text{half}} \), while in the simulations it begins dropping inside \( \sim 0.5R_{\text{half}} \). We will discuss this further in §3.4.4.

Now examining the outer parts of the sSFR profile \( (R/R_{\text{half}} \gtrsim 1) \), we see that the sSFR in GV galaxies is also suppressed relative to SF galaxies in Simba at all masses in the outskirts. This is clearly in conflict with the data. Interestingly, EAGLE and Illustris likewise produce sSFR profiles that drop rapidly more quickly than observed in GV galaxies, so while Simba yields a central hole in sSFR in better agreement with data than those simulations, in the outskirts Simba is similar to other simulations. The discrepancy in the outskirts could be due in part to the conversion of H\( \alpha \) to SFR in Belfiore et al. (2018), which could have some contribution from non-star forming diffuse ionised gas. The contribution of dust-scattering to the H\( \alpha \) emission of diffuse ionised gas in simulations is estimated to be at most 50% of the total emission (Ascasibar et al., 2016). Depending on how radially-dependent this contribution is, it could make a difference to the outskirts of the observed radial profiles.

We note that the Belfiore et al. (2018) analysis computes radial profiles in elliptical apertures to account for inclination, using elliptical Petrosian effective radii \( R_e \) and inclinations from the NASA-Sloan catalogue (NSA v1.0.1\(^3\), Blanton et al. 2011) to construct de-projected radial profiles with elliptical annuli of semimajor axis 0.15\( R_e \). We have approximated this process by first making all our galaxies face-on before calculating profiles. These methods should be identical in the case of a perfectly thin disk, but in our simulations, the stellar disks are not particularly thin. A similar procedure would likely blur out the central region, and thus potentially mitigate the differences with the SF population. We demonstrate how large an effect this may have by showing non-rotated profiles for the highest mass bin in Figure 3.6 – the rotated and non-rotated profiles show almost no difference beyond 0.5\( R_{\text{half}} \), but in the central region the rotated profiles have much lower sSFR. We see that de-projecting the profiles emphasises the central suppression in star formation.

\(^3\)https://www.sdss.org/dr13/manga/manga-target-selection/nsa/
Also, it is worth noting that the Belfiore et al. (2018) data explicitly removes galaxies with Seyfert-like line ratios. Seyferts are typically large star-forming disks with strong AGN activity. We are not currently able to identify Seyferts via line ratios in our simulation, so we have not mimicked this selection. It may be possible that such galaxies would have SFR profiles that drop rapidly in the middle owing to the putative nuclear AGN feedback, which would make the profiles of SF galaxies drop more quickly in the centres. Unfortunately, it is difficult to measure inner SFRs in Seyferts owing to AGN contamination, which is precisely why these were excluded in observations. Seyferts make up a relatively small fraction (∼10%) of the overall disk population, but may contribute more to the most massive bins. Hence this may explain part of the massive SF galaxy discrepancy.

Recall that SIMBA has two separate effects going on to suppress star formation in GV galaxies (§3.4): In the inner parts, this owes to removal of star-forming gas which lowers the gas fraction, while in the outer parts it owes to a lower star-forming efficiency. It appears that the physics driving the inner suppression is roughly consistent with observations for GV galaxies, but that driving the outer suppression via a drop in the SFE is not. It also appears that the onset of the inside-out quenching occurs in massive star-forming galaxies in SIMBA is quite strong, whereas such galaxies in observations show only a mild central suppression. Hence while reproducing the central sSFR drop in GV galaxies is a qualitative success of SIMBA, there remain substantial discrepancies in galaxy sSFR profiles. To explore the physical drivers of these various effects, we now examine which feedback mechanisms are responsible for these trends.

### 3.4.4 Black hole feedback dependence

In SIMBA, AGN feedback is primarily responsible for quenching galaxies. Of the three forms of AGN feedback implemented in SIMBA, the jet mode feedback is most directly responsible for quenching, the X-ray feedback by itself does not quench but is nonetheless important for fully quenching galaxies, and the radiative mode feedback is essentially irrelevant for quenching (Davé et al., 2019). The question is then, how do these various AGN feedback forms impact the profiles of quenching galaxies? To answer this, here we compare our profiles in a full-physics SIMBA run versus two other runs: No-jet, where we turn off both X-rays and jets, and no-Xray, where we turn off only the X-ray feedback but leave jets on. These
Figure 3.7: The sSFR radial profiles for all SF (left) and GV (right) galaxies for the Simba without X-ray or jet feedback (pink/purple lines) and the full Simba (blue/green lines) $50h^{-1}\text{Mpc}$ runs. The displayed radial profiles are Tukey biweights of the individual galaxy profiles in each mass bin. The light shaded regions around each line show the Tukey biweight scale estimator divided by $\sqrt{N}$, combined with cosmic variance uncertainties from jackknife resampling over the 8 simulation sub-octants. The vertical dotted lines show the spatial resolution in units of the median half-light radius for each mass bin, computed separately for SF and GV galaxies. The most notable difference in the sSFR profiles of SF galaxies occurs in the central region of more massive SF galaxies, where the No-jet run produces no dip, suggesting that it is a direct result of either jet or X-ray feedback. The differences for the GV galaxies are much more striking, showing a strongly centrally concentrated star formation, with the dropoff in the sSFR profile being stronger than in the full feedback case. Hence AGN feedback seems crucial for redistributing the star formation in GV galaxies at all radii.
Figure 3.8: The sSFR radial profiles for all SF (left) and GV (right) galaxies for the Simba without X-ray feedback (pink/purple lines) and the full Simba (blue/green lines) 50h^{-1}Mpc runs. The displayed radial profiles are Tukey biweights of the individual galaxy profiles in each mass bin. The light shaded regions around each line show the Tukey biweight scale estimator divided by \sqrt{N}, combined with cosmic variance uncertainties from jackknife resampling over the 8 simulation sub-octants. The vertical dotted lines show the spatial resolution in units of the median half-light radius for each mass bin, computed separately for SF and GV galaxies. Turning on jets and leaving the X-rays off leads to even more centrally concentrated sSFR profiles that are suppressed in the outskirts for both SF and GV galaxies compared to the full Simba and No-jet runs. Jet feedback appears to have the overall effect of slightly suppressing star formation in the outskirts, causing that gas to move inwards in order to form stars in a more centrally concentrated manner. Hence jet feedback as implemented in Simba is not responsible for the central sSFR suppression observed by Belfiore et al. (2018).
are done using $50h^{-1}\text{Mpc}$, $2 \times 512^3$ particle runs, but the numerical resolution and input physics are otherwise identical to the full SIMBA run.

Figure 3.7 shows the sSFR profiles for galaxies in the No-jet SIMBA run (i.e. without X-ray or jet mode feedback) shown in purple to orange. As in Figure 3.6, the left panel shows the SF galaxies, while the right panel shows the GV galaxies. We reproduce the results from the full SIMBA simulation for comparison (in blue to green), but leave off the Belfiore et al. (2018) observations to avoid confusion. Note that in the No-jet simulation, massive galaxies do not quench, but typically end up as star-forming or GV galaxies (Davé et al., 2019).

The outer SF galaxy profiles are not markedly different over most radii with jet and X-ray feedback off versus in the full SIMBA model. The sSFR values are only mildly higher in the No-jet case at all radii beyond $\gtrsim 0.5R_{\text{half}}$. The profiles are also nearly identical for SF galaxies regardless of $M_\star$, whereas the full SIMBA model yields a stronger mass dependence, in better agreement with the observations. The most notable difference occurs in the central region of more massive SF galaxies, where the No-jet run produces no sSFR dip in the central region. Hence we infer that this dip is a direct result of either jet or X-ray feedback. Although we do not show the observations from Belfiore et al. (2018) overlaid, the core sSFR is higher in the No-jet run compared to observations, particularly for the more massive galaxies. Hence the Belfiore et al. (2018) data seems to require some suppression of core SF in massive galaxies, but not as much as in the full SIMBA run. Outside the central region, the profiles are in good agreement with data. This shows that AGN feedback has a modest but non-trivial impact on even star-forming galaxies, and at least some AGN feedback is already required to produce a central sSFR depression in massive SF systems.

Turning to the GV galaxies (right panel), the differences are much more striking. The No-jet sSFR profiles show strongly centrally concentrated star formation, with more compact extent (relative to $R_{\text{half}}$) for more massive systems. Also, the GV profiles are quite a bit steeper than the SF galaxy profiles. This shows that the reason these galaxies are in the GV is that star formation has been eroded in the outskirts, likely because these galaxies live in shock-heated hot halos (Kereš et al., 2005) that is starving these systems of fresh gas. In other words, without significant AGN feedback, suppression of star formation occurs outside-in. This is exactly opposite to the way that GV galaxies are observed to be quenching, from the inside-out.
Interestingly, the *No-jet* GV profiles show an even stronger dropoff in the sSFR than in the full feedback case; thus if anything, AGN feedback appears to be puffing out the star forming gas relative to the stellar mass (or $r$-band light). Hence it is not possible to solve the discrepancy between *SIMBA* and the *Belfiore et al.* (2018) data in the outskirts by simply saying that AGN should have no effect there. Indeed, AGN appear crucial at all radii for redistributing the star formation in GV galaxies in a manner consistent with observations, for all galaxy masses probed here.

Given that the combination of jet and X-ray feedback appears to be crucial for altering the GV sSFR profiles to be in better qualitative agreement with data, it is interesting to ask which of these two modes is most responsible. To examine this, we now examine *no-Xray* where we turn on the jet feedback, but leave the X-ray feedback off.

Figure 3.8 shows the resulting profiles from the *no-Xray* run, analogous to Figure 3.7. The *no-Xray* run does produce some quenched galaxies, but generally they do not have as low sSFR as observed. Hence a histogram of sSFR’s from this model does not agree with observations as it does for the full *SIMBA* run (Davé et al., 2019).

Remarkably, turning on jets and leaving the X-rays off actually leads to even more concentrated sSFR profiles. Now, even the SF galaxy profiles are clearly wrong – they are centrally peaked, and are suppressed in the outskirts, relative to the full *SIMBA* and *No-jet* runs. It appears that jet feedback has the overall effect of slightly suppressing star formation in the outskirts, causing that gas to move inwards in order to form stars in a more centrally concentrated fashion.

Recall that in *SIMBA* our jets are purely bipolar, and explicitly do not interact with surrounding gas until they are outside the ISM. Hence it is not surprising that they do not suppress the central SF, but it is curious that they indirectly cause an enhancement, at least relative to the stellar mass distribution.

Moving to the GV galaxies (right panel), the central concentration of sSFR is now even more apparent than in the *No-jet* case. In GV galaxies, the jet feedback is strongly suppressing the star formation in the outer regions leading to much steeper profiles relative to *SIMBA*, but the innermost sSFR is essentially unchanged from the SF galaxies in the left panel. Clearly, jet feedback as implemented in *SIMBA* is not responsible for the central sSFR hole that is observed by *Belfiore et al.* (2018).
It is only when we turn on X-ray feedback as in the full SIMBA run that we produce central suppression in the GV (as well as massive SF) population. We conclude therefore that it is SIMBA’s X-ray feedback that is responsible for creating the central depression in sSFR as observed by Belfiore et al. (2018) and others.

How does X-ray feedback as implemented in SIMBA cause inside-out quenching? Our implementation of X-ray feedback represents a sub-resolution model for momentum input from the X-ray photons generated in the accretion disk. In it, a kick is applied outwards from the black hole onto gas within the black hole accretion kernel, based on the X-ray radiative momentum input coupling to hydrogen gas as outlined in Choi et al. (2012). Since the momentum input drops as $1/r^2$ from the black hole, the gas closest to the black hole is most strongly kicked, which creates a hole in the cold gas and hence in the star formation.

Importantly, SIMBA implements a 2 kpc maximum radius for the black hole accretion kernel (or 256 nearest neighbours, whichever is smaller), and thus for the extent of direct X-ray feedback kicks. In principle there is no reason why X-ray photon pressure should be limited to this radius; this was done purely for computational convenience. Since there was already a neighboring particle list identified for the black hole accretion module out to (up to) 2 kpc, it was most straightforward to implement a kick on these pre-identified particles. The inadvertent result of this is that X-ray feedback is only immediately felt out to $\lesssim 2$ kpc. This may explain why our full SIMBA profiles rise quickly out to a maximum at a $2−3$ kpc ($\sim 0.5 R_{\text{half}}$); had we allowed our X-ray feedback to operate to larger radii, it is possible we would have generated a more gradual rise in sSFR out to larger radii, which would qualitatively be in better agreement with the observed profiles of Belfiore et al. (2018).

The X-ray feedback does not fix the too-rapid dropoff in sSFR at $R \gtrsim R_{\text{half}}$ in GV galaxies, seen at all masses and even with AGN feedback mostly off. This remains something of an enigma. This rapid dropoff also appears in GV galaxy profiles in EAGLE and Illustris (Starkenburg et al., 2019), hence it seems to be a fairly generic outcome of current galaxy formation models: galaxies that have depressed overall star formation tend to have it particularly depressed in their outskirts, in clear disagreement with observations. One commonality between all these simulations is that their subgrid AGN feedback models quench galaxies primarily by keeping the surrounding circum-galactic gas hot (e.g. Davies et al., 2020), albeit via differing mechanisms. This is long known to be a successful approach to quenching (Croton et al., 2006; Bower et al., 2006; Gabor & Davé, 2020).
It is possible, however, that such preventive feedback preferentially suppresses star formation in the outskirts of galaxies by shutting off the accretion that would otherwise replenish an extended cold gas reservoir. This appears to be in contradiction with observations. One perhaps relevant point in Simba is that there is significant molecular gas in the outskirts, but it has low star formation efficiency (Fig. 3.5, bottom panels). Adjusting the star formation prescription to have that gas continue to form stars would yield better agreement for sSFR profiles in the outskirts. We leave such explorations for future work. In any case, it appears to be a significant challenge for models to quench galaxies as observed globally, while retaining active star formation out to several half-light radii in transitional GV galaxies.

To summarise, Simba shows low-mass star-forming galaxy sSFR profiles and GV sSFR central depressions that are broadly in agreement with observations. The central depressions owe specifically to X-ray feedback as implemented in Simba, which imparts outwards momentum to the gas surrounding the black hole. Other simulations such as EAGLE and Illustris, which do not have such a mechanism, fail to match this. It is possible that Illustris-TNG may fare better, because although they do not implement X-ray feedback as in Simba, they tend to sphericalise the jet energy input by randomly re-orienting the jet at every timestep, and they do not shut off hydrodynamic interactions between the jet and the ISM as in Simba (Weinberger et al., 2018). This could result in a qualitatively similar outward momentum injection. One could envision that simply heating the ISM near the black hole might also be sufficient to create an sSFR hole, but this is essentially how the EAGLE AGN feedback model operates, and it does not succeed. More generally, our results imply that current observations require some mechanism that evacuates gas from the central regions of galaxies during the quenching process, in a manner that operates approximately like Simba’s X-ray feedback module. This inside-out quenching, along with the unresolved discrepancies in the outskirts of GV sSFR profiles, represent key constraints on quenching prescriptions in current galaxy formation models.

3.4.5 Satellites vs. centrals

So far, we have not distinguished centrals versus satellites, and simply considered all galaxies within our specified cuts. However, satellites can experience environmental quenching processes that could in principle impact their profiles
differently than internal processes such as AGN feedback. For instance, ram pressure and tidal stripping might remove gas preferentially from the outskirts, which would result in more compressed H I profiles as seen in GV galaxies, but for reasons that do not involve AGN feedback. Alternatively, they could be starved of gas infall owing to living within a hot halo, and thus have their SFR suppressed. Spindler et al. (2018) examined satellite vs. central profiles in MaNGA and found that satellites have overall lower sSFR at most radii vs mass-matched centrals, but that the suppression in the central region is similar. They interpret this to suggest that satellites have lower sSFR overall owing to strangulation that cuts off their broader gas supply, but that the core sSFR suppression is a separate internal process. Here we examine the profiles of satellite versus central galaxies in Simba to better understand how they are impacted by satellite-specific processes.

Figure 3.9 shows radial profiles of $\Sigma_{\text{sSFR}}$ and $\Sigma_{\text{H I}}$ for galaxies separated into satellites and centrals (top row), and the logarithmic difference of the radial profiles (bottom row). Central galaxies are identified as the most baryonically massive galaxies in their halos, and satellites are all others. We bin galaxies into two bins following Spindler et al. (2018), namely $10 \leq \log(M_\star / M_\odot) \leq 10.6$ (turquoise, 1591 centrals and 349 satellites) and $\log(M_\star / M_\odot) > 10.6$ (blue, 349 centrals and 74 satellites). They also have a bin to lower masses but we eschew this owing to numerical resolution concerns. The Spindler et al. (2018) data in those bins is shown in orange and purple, respectively.

Qualitatively, Simba reproduces the trends seen in the observations. Beyond the core, satellites in Simba have lower sSFR than centrals, which is consistent with the data. The magnitude of the difference in the outer regions is similar to what is observed. Meanwhile, for $R \leq 0.5R_{\text{half}}$, the trend reverses. This is broadly seen in the observations as well, though the trend does not fully reverse for the low-mass bin. This qualitative agreement suggests that Simba accounts for both internal and external quenching processes in satellites to yield rough agreement with observations.

Quantitatively, there are some significant differences. The bottom left panel shows that in Simba, centrals have a significantly larger sSFR suppression than satellites at the same mass, whereas the effect is relatively weak in the observations. This would suggest that X-ray AGN feedback is much weaker in satellites as in centrals, whereas it should be closer in order to match the data. It is not immediately evident why X-ray feedback is weaker in satellites. Part of the difference could owe to differences in the way centrals and satellites are identified in observations.
Figure 3.9: *Top*: Radial profiles of $\Sigma_{\text{sSFR}}$ (left) and $\Sigma_{\text{HI}}$ (right) for SF galaxies, split into centrals (dashed lines) and satellites (dotted lines), shown for two mass bins as green and blue lines. *Bottom*: Ratio of sSFR (left) and $\Sigma_{\text{HI}}$ (right) radial profiles for satellite and central SF galaxies. Observations from Spindler et al. (2018) are shown as orange and purple lines. The horizontal black dashed line shows where the profiles for satellites and centrals are equal. For all panels, the displayed radial profiles are the ratio of the Tukey biweights of the satellite and central samples. The light shaded regions show the Tukey biweight scale estimator divided by $\sqrt{N}$, combined with cosmic variance uncertainties from jackknife resampling over the 8 simulation sub-octants. The vertical dotted lines show the spatial resolution in units of the median half-light radius for each mass bin. Qualitatively, SIMBA reproduces the trends seen in the observations, suggesting that it accounts for both internal and external quenching processes in satellites. Quantitatively, centrals in SIMBA have a significantly larger sSFR suppression than satellites at the same mass compared to observations, suggesting that X-ray AGN feedback is much weaker in satellites than in centrals.
versus our simulations; another possible explanation is that satellites may have preferentially lower mass black holes, which may make the transition to X-ray feedback less likely for these galaxies. We will explore this further in future work. The MaNGA data only probes out to $1.5R_{\text{half}}$, but SimBA can examine these trends to larger radii, and predicts that the satellites truncate their sSFR at $\sim 2R_{\text{half}}$ as opposed to centrals which generally extend out to $\gtrsim 3R_{\text{half}}$.

Examining $\Sigma_{\text{HI}}(R)$ in the right panel, we see some differences between centrals and satellites but they are generally much reduced relative to that seen for sSFR. Satellites clearly show a steeper profile in the outskirts than centrals, but interestingly at $R \lesssim R_{\text{half}}$ they actually have higher HI surface densities. A mild central depression is seen in HI for the more massive galaxies, and this drop is identical in the centrals and satellites. Hence HI profiles are not quite as dramatically sensitive to environment as sSFR profiles, but nonetheless show a clear impact of gas suppression processes.

These trends are consistent with various potential environmental processes acting on satellite galaxies relative to centrals. As for why the outskirts of the centrals and satellites differ, satellites are more adversely affected by environmental processes as their lower masses leave outskirts more vulnerable. These processes include ram pressure stripping (removal of gas due to heating in the intracluster medium (Gunn & Gott, 1972), galaxy harassment (gas removal due to frequent high speed galaxy encounters (Moore et al., 1996), mergers (collisions between galaxies (Toomre & Toomre, 1972) and strangulation (galaxies are unable to replenish their gas supply (Larson et al., 1980; Peng et al., 2015). These processes leave different observational signatures. Strangulation should deplete gas uniformly across the galaxy (‘anemic galaxies’, e.g. van den Burgh, 1991; Elmegreen et al., 2002; Spindler et al., 2018), whereas stripping removes gas preferentially from the outskirts and could lead to enhanced star formation confined to the galaxy centre (Spindler et al., 2018). In our case we see that the inner regions of the satellites are enhanced in both SFR and HI, while the outskirts are more depleted than the centrals, which is broadly consistent with a ram pressure stripping scenario (Cunnama et al., 2014; Rafieferantsoa et al., 2019).
3.4.6 Redshift dependence

We have shown that both inside-out and outside-in quenching occurs in SIMBA in $z = 0$ GV galaxies, and that the driving physical mechanism within SIMBA appears to be its X-ray AGN feedback implementation. An interesting question is, when do these quenching mechanisms become apparent? At higher redshifts, it becomes more difficult to select GV galaxies, owing to overall younger stellar populations and an increased prevalence of dusty galaxies. However, it is still possible to examine massive galaxies, which should have a higher fraction of galaxies in the process of quenching than at lower masses. Here we examine the radial sSFR profiles of SF galaxies at $z = 2$ in SIMBA, as a function of $M_*$, and compare to selected observations.

Figure 3.10 shows the radial surface density profiles of sSFR for SF galaxies at $z = 2$. We select SF galaxies as having log (sSFR/yr$^{-1}$) > -9.5, to compare with SINFONI observations at $z = 2$ (Tacchella et al., 2018), shown as orange/purple lines. We obtain 988 galaxies in the low-mass bin ($10^{10} M_\odot < M_* < 10^{10.5} M_\odot$), 318 in the intermediate-mass bin ($10^{10.5} M_\odot < M_* < 10^{11} M_\odot$), and 49 in the high-mass bin ($M_* > 10^{11} M_\odot$).

In general the sSFR profiles steadily increase towards the centre. At $z = 2$, the profiles for low- and intermediate-mass bins are nearly identical, whereas the profile for the highest mass bin shows lower sSFR at all radii. The profile for the highest mass bin flattens at the centre, and the other masses show a slight decrease.

Comparing with what we have already seen at low redshift, first we see that at $z = 2$ there is active star formation at all radii across the galaxies, without a sharp decrease to zero at any point, whereas at low redshift the sSFR of SF galaxies drops to zero at $\sim 3R_{\text{half}}$. Overall the level of star formation across the galaxies is higher, as is expected at high redshift. We find that the level of star formation is sensitive to the exact sSFR cut used to select the galaxies, but that the trends remain unchanged with different sSFR cuts, so in this case the shape of the profiles is more important than the exact level of star formation.

The trend of increasing star formation in the centre is similar to the No-jet and no-Xray models described in §3.4.4, so it appears that the high redshift galaxies have not been affected by AGN feedback nearly to the same degree as the low redshift galaxies. This is not surprising since at high redshift SIMBA’s AGN feedback has
Figure 3.10: Radial sSFR profiles for SF galaxies at $z = 2$. The displayed radial profiles are Tukey biweights of the individual galaxy profiles in each mass bin. The light shaded regions around each line show the Tukey biweight scale estimator divided by $\sqrt{N}$, combined with cosmic variance uncertainties from jackknife resampling over the 8 simulation sub-octants. The vertical dotted lines show the spatial resolution in units of the median half-light radius for each mass bin. Observations of sSFR profiles of SF galaxies from SINFONI (Tacchella et al., 2018) at $z = 2$ are shown as the pink/purple lines. Active star formation at all radii across the galaxies suggests that neither the mechanisms for outside-in nor inside-out quenching have had a substantial impact.
Table 3.2: Median global galaxy properties and number of galaxies for the two mass bins for SF and GV galaxies in the high resolution simulation.

<table>
<thead>
<tr>
<th>Median</th>
<th>Star forming</th>
<th>GV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>Int</td>
</tr>
<tr>
<td>( N_{\text{gal}} )</td>
<td>61</td>
<td>31</td>
</tr>
<tr>
<td>( \log(M_*/M_\odot) )</td>
<td>10.2</td>
<td>10.9</td>
</tr>
<tr>
<td>SFR (M_\odot yr^{-1})</td>
<td>1.53</td>
<td>5.48</td>
</tr>
<tr>
<td>( \log(\text{sSFR} / \text{yr}^{-1}) )</td>
<td>-10.06</td>
<td>-10.25</td>
</tr>
<tr>
<td>( \log(M_{\text{HI}}/M_\odot) )</td>
<td>10.2</td>
<td>10.4</td>
</tr>
<tr>
<td>( \log(M_{\text{H}<em>2}/M</em>\odot) )</td>
<td>9.4</td>
<td>9.9</td>
</tr>
<tr>
<td>( R_{\text{half}} ) (kpc)</td>
<td>4.2</td>
<td>6.3</td>
</tr>
</tbody>
</table>

little impact in general. This is because the black holes at high redshift tend to have higher accretion rates and are thus emitting feedback radiatively (not in jet or X-ray form), which as shown in Davé et al. (2019) has little impact on galaxy properties. Similarly, at \( z = 2 \) the environmental effects that cause the outside-in quenching (particularly in satellite galaxies, see section 3.4.5) have not had a large impact, hence there is active star formation into the outskirts. It thus appears that at high redshift neither the mechanisms for outside-in nor inside-out quenching have had a substantial impact. In short, the AGN feedback modules responsible for quenching in SIMBA are not yet much in place at \( z \sim 2 \).

In contrast, Figure 3.10 shows that the observations indicate the more massive SF galaxies have clear central depressions in star formation even as early as \( z = 2 \) (Tacchella et al., 2018). The observed sSFR profiles are flat for galaxies in the low mass bin, and the higher mass bin shows some suppression at the centre. It is clear that SIMBA does not fully reproduce the observed trends at high redshift, suggesting that some aspects of the AGN feedback implementation is unrealistic at early epochs. In particular, if X-ray feedback is responsible for creating central holes as seen at low redshift, then it may be that the X-ray feedback is insufficiently effective at higher \( z \) in SIMBA. Inside-out quenching can also be associated with non-AGN related mechanisms, such as a wet compaction event due to minor mergers or tidal streams, which leads to a ring of star-forming gas around the centre (Tacchella et al., 2016). Reproducing the central suppression in star formation at early epochs is thus another test that can be used to constrain AGN feedback models.
Figure 3.11: Radial sSFR profiles for SF galaxies (left) and GV galaxies (right) in the high resolution SIMBA run (orange/purple lines) and the SIMBA run at fiducial resolution (green/blue lines). The vertical dotted lines show the spatial resolution in units of the median half-light radius for each mass bin, computed separately for SF and GV galaxies. The displayed radial profiles are Tukey biweights of the individual galaxy profiles in each mass bin. The light shaded regions around each line show the Tukey biweight scale estimator divided by $\sqrt{N}$, combined with cosmic variance uncertainties from jackknife resampling over the 8 simulation sub-octants. Profiles of the high resolution galaxies are qualitatively similar to those from the main fiducial volume.

### 3.4.7 Numerical convergence

To test for numerical convergence, particularly in the most central radial bins, we compare to the higher resolution SIMBA volume of $25h^{-1}$Mpc with $2 \times 512^3$ particles, which we denote SIMBA-hires. This simulation implements the same physical subgrid models as the main fiducial $100h^{-1}$Mpc simulation, but with 8 times better mass resolution and twice the spatial resolution.

We compute radial profiles of sSFR for SF and GV galaxies as before, using the same GV definition from Belfiore et al. (2018) as in section 3.4.1. At $z = 0$, we obtain 61 SF galaxies in the low-mass bin ($10^{10} M_\odot < M_\star < 10^{10.5} M_\odot$) and 31 in the high-mass bin ($M_\star > 10^{10.5} M_\odot$). In the same mass bins we find 35 and 11 GV galaxies respectively. With the small volume and high resolution, this simulation primarily probes galaxies of lower stellar mass, so the two highest mass bins are grouped together as there are few galaxies in these bins. Table 3.2 shows the median properties of SF and GV galaxies in this simulation.

Figure 3.11 shows radial surface density profiles of sSFR for SF and GV galaxies in SIMBA-hires. Owing to the low numbers, the profiles for the high resolution simulation are noisy, with a large range of sSFR per radial bin especially for high
masses. In particular, there is a lot of variation in the centres of the SF galaxies and the outskirts of the GV galaxies.

SimBa-hires produces radial profiles that are qualitatively similar to those from the main SimBa simulation. Importantly, the central suppression of star formation is still present in the high resolution case. The profiles for the high resolution GV galaxies show that the high mass galaxies show qualitatively the same level of star formation suppression as in the low resolution case. The SF galaxy profiles have the same shape as in the fiducial simulation, indicating that the quenching processes occur over the same radial scales as before. The broad agreement between the two simulations is an important result which reinforces the main trends described in this work and indicates that the qualitative trends are not sensitive to numerical resolution.

Looking in more detail, there are some interesting discrepancies between the two simulations that point to the effects of numerical resolution. For the GV galaxies, inside-out quenching in low mass galaxies is not as strong in SimBa-hires, indicating that these central radial bins were not well resolved in the low resolution case. The difference between the centres of the high and low resolution profiles is much more pronounced for the low galaxy mass bins, which intuitively makes sense as lower mass objects should be less well resolved to begin with. When analysing radial profiles elsewhere in this work, we should remember that the centres of the low mass GV galaxies are less well resolved. The low mass SF galaxies are more similar in terms of shape between the two simulations.

In §3.3 we noted that the sizes of passive galaxies are significantly smaller in SimBa-hires, whereas the star forming galaxies were more numerically converged between the two simulations. The passive objects have few gas elements and are impacted by long-term numerical heating in the orbits of star particles that cannot dissipate their energy. Hence star forming galaxies are expected to be better resolved generally. Following from this, profiles from star forming galaxies should be more robust than the green valley profiles as GV galaxies have lower gas fractions (see §3.4). This could explain why the inner regions of GV galaxies are less well converged than for SF galaxies, at least in terms of profile shape.

For the GV galaxies, star formation extends further to the outskirts than for the low resolution profiles. The decrease in the outskirts is also more gradual than in the low resolution case – outside-in quenching is somewhat weaker in the high resolution simulation. This could indicate that strong outside-in quenching is
linked to low resolution. However, more likely it arises as a result of the small volume of the simulation. In the $25h^{-1}\text{Mpc}$ box, we do not produce the most massive galaxies with the strongest AGN feedback, potentially weakening the effect of environmental outside-in quenching.

Looking at the SF galaxies, the high resolution case shows overall lower sSFR for both mass bins. This is a result of high galaxy stellar masses in the high resolution simulation owing to an abundance of lower mass star particles. The galaxy stellar mass function for the high resolution case does not match observations as well as the low resolution case. We have checked that profiles of SFR in the high resolution volume show that star formation activity is consistent with the main simulation.

Overall, the results in this Chapter are qualitatively unchanged when examined at 8x better mass resolution. While this is still a limited dynamic range, it suggests that key results such as the suppression of gas and star formation in the central regions of GV galaxies is a robust outcome of the physics in SIMBA, rather than a numerical artifact. Nonetheless, there is some non-trivial sensitivity to the results particularly for lower-mass galaxies that suggests caution when using such profiles to quantitatively constrain the input physics in SIMBA or simulations with similar resolution.

### 3.5 Summary

We have examined the profiles of star-forming (SF) and green valley (GV) galaxies in the SIMBA simulation, a state of the art cosmological hydrodynamic simulation that yields a population of quenched galaxies in good agreement with observations. We have examined the redshift evolution of half-light radii of SF and passive galaxies. We separate SF and GV galaxies via a nonlinear cut in SFR–$M_\star$ space following Belfiore et al. (2018), and focus on relatively well-resolved $M_\star > 10^{10}M_\odot$ galaxies. We examine sSFR profiles, but also study profiles in SFR, gas surface density, star formation efficiency (or depletion time), and gas fraction. We further examine differences in the profiles of central vs. satellite galaxies, and the evolution of sSFR profiles out to $z = 2$. We compare to $z = 0$ observations of Belfiore et al. (2018) and Spindler et al. (2018), and $z \sim 1−2$ data from Tacchella et al. (2018).
Our main conclusions are as follows:

- Simba reproduces $z = 0$ star-forming galaxy sizes well, but yields quenched galaxies sizes that are too large at low $z$. The evolution of star forming sizes is also well reproduced, $R_{\text{half}} \propto (1 + z)^{-0.78}$. However passive galaxies are of comparable size to star forming galaxies at $z = 0$ and have too shallow a redshift evolution, $\propto (1 + z)^{-1.06}$. This is consistent with the evolution of the numerical softening length, suggesting passive galaxy sizes are impacted by numerical resolution. For the majority of this Chapter, we only employ $z = 0$ star-forming galaxy sizes.

- Examining $z = 0$ galaxy profiles, we see that the surface density of star formation of all galaxies with $M_* > 10^{10}M_\odot$ in Simba shows a peak at $R \sim 0.5 - 1R_{\text{half}}$, where $R_{\text{half}}$ is the $r$-band half-light radius, an exponential dropoff to large radii, and a central $\Sigma_{\text{SFR}}$ depression in high-mass star forming galaxies and all GV galaxies. These trends at low radii are seen above the scale of the effective spatial resolution and broadly reproduced in a higher-resolution run, except for the case of the lowest mass galaxies.

- The sSFR profile shows a qualitatively similar trend as the $\Sigma_{\text{SFR}}$ profile, with a sharper cutoff at $\sim 3R_{\text{half}}$. Together, this shows that the central depression in the sSFR profile is a consequence of a lack of star formation in the core, not an excess of bulge stellar mass.

- GV galaxies show lower overall $\Sigma_{\text{SFR}}$ and sSFR at all radii, and have profiles with much larger central depressions and rapid truncation at $R \gtrsim R_{\text{half}}$, than typical star-forming galaxies.

- The H\textsc{i} surface density profiles for galaxies are virtually identical at $0.5 \lesssim R/R_{\text{half}} \lesssim 3$ for all star-forming galaxies, but more massive systems show less $\Sigma_{\text{H}1}$ in the core and outskirts. GV $\Sigma_{\text{H}1}$ profiles are similar to SF profiles at $R \lesssim R_{\text{half}}$, but show a much more rapid decline beyond the this.

- The molecular gas profiles show considerably larger extent than the SFR or sSFR profiles, for both SF and GV galaxies. This is reflected in the H\textsubscript{2} fraction ($f_{\text{H}2} = \Sigma_{\text{H}2}/\Sigma_*$) profiles, which are fairly flat for $R \gtrsim 0.5R_{\text{half}}$, for both galaxy types. In the core region, $f_{\text{H}2}$ drops rapidly, showing that the central depression is caused by an evacuation of dense star-forming gas. In the outer region, there is still substantial amounts of H\textsubscript{2}, but it is evidently not forming stars.
• This is corroborated by examining the SF efficiency \( \text{SFE} = \Sigma_{\text{SFR}} / \Sigma_{H2} \) profiles, which show a rapid decline in the outskirts but no central depression. The change in GV profiles in the outskirts is thus entirely driven by a dropping SFE. A simple scenario in which gas with \( \Sigma_{H2} \lesssim 1.1 M_\odot \, \text{pc}^{-2} \) doesn’t form stars roughly reproduces the mean SFR truncation radius in SF vs. GV galaxies.

• We compare our SF and GV sSFR profiles to observations of Belfiore et al. (2018). Simba yields a central depression in sSFR in qualitative agreement with data, unlike the EAGLE and Illustris simulations (Starkenburg et al., 2019). Quantitatively however, massive \( (M_\star > 10^{10.5} M_\odot) \) SF profiles show too large a central depression, and the shape of the central depression in the GV galaxies is not in perfect agreement with observations.

• In contrast with observations, Simba also yields a strong truncation in the sSFR profiles at \( R \gtrsim R_{\text{half}} \) for GV galaxies. This is present in all AGN feedback variants, so it is not associated in particular with quenching. In other words, Simba galaxies quench inside-out as observed, but some other physics may be incorrect which results in outside-in quenching; this may be related to the star formation prescription.

• Using test simulations with various AGN feedback modules turned on and off, we demonstrate that it is specifically Simba’s implementation of X-ray AGN feedback that is responsible for creating the central depression. Turning this off results in steeply rising profiles for GV galaxies, as also seen in EAGLE and Illustris. While Simba’s X-ray feedback is quite heuristic (and was included mainly because of the physical motivation outlined in Choi et al. 2012), this demonstrates that some internal feedback process that generates outwards momentum deposition seems to be required in order to generate GV galaxy profiles as observed.

• Satellite galaxies show depressed sSFR relative to centrals at \( R \gtrsim R_{\text{half}} \), but an enhancement within (though still depressed overall). This is qualitatively consistent with observations from Spindler et al. (2018), though in Simba the core enhancement is larger than in the data. Simba also shows a smaller radial extent of star formation in satellites vs. centrals, at radii beyond the range that is probed by the Spindler et al. (2018) data.

• The H\text{I} surface density profile of satellites is likewise enhanced in the inner regions and depressed in the outer regions relative to centrals, but the effects
are more modest than in the sSFR profiles.

- At \( z = 2 \), SIMBA galaxies do not show central sSFR depressions in galaxies of any mass. This is understandable because X-ray feedback in SIMBA is tied to AGN jet feedback, which become widespread only at \( z \lesssim 1 \). However, this prediction is in contrast to observed sSFR profiles from Tacchella et al. (2018) at \( z \sim 2 \), which do show central depressions in the most massive SF galaxies. Hence it appears that SIMBA’s assumption of tying X-ray feedback to jet feedback may need to be revisited.

- The central star formation suppression is also produced in a SIMBA run with 8x better mass resolution. GV galaxies in the high resolution volume show qualitatively similar levels of inside-out quenching, except in the inner radial bins of the lowest mass galaxies. Radial profiles of SF galaxies in the two simulations have the same shape, but the normalisation of the profile is low in the high resolution box, owing to more abundant low mass star particles.

- Green valley galaxies in the high resolution simulation have more extended star formation. This is probably an outcome of the low volume of this simulation; the absence of the highest mass galaxies results in less heating of surrounding gas due to AGN feedback, and thus outside-in quenching due to the environment is weaker.

Overall, our results demonstrate the valuable constraints provided by sSFR (and other) profiles of galaxies as they move from the star-forming to the quenched regime. While it is encouraging that SIMBA’s X-ray feedback reproduces the observed central sSFR depressions in GV galaxies, the various other discrepancies highlight that there are aspects of quenching in state of the art models that require further improvement. Simulations such as SIMBA with sufficient resolution to examine the internal structure of galaxies, albeit coarsely, while still modeling a representative galaxy population, can now take advantage of these structural constraints to better understand the physical mechanisms by which galaxies quench.
4.1 Introduction

Galaxies interact dynamically with their environment. They accrete new material for star formation, either pristine gas from the intergalactic medium (IGM; Kereš et al., 2005; Dekel et al., 2009), or recycled gas that has been processed through star formation (Oppenheimer et al., 2010; van de Voort et al., 2011). Galaxies also eject material via energetic outflows from supernova events and/or black hole activity (Veilleux et al., 2005). Additionally, they can lose material via mergers or due to environmental processes such as ram pressure stripping. These processes, collectively referred to as the baryon cycle, are believed to be a primary regulator of how galaxies form and evolve. Any material entering or exiting a galaxy halo must pass through the circumgalactic medium (CGM). Therefore the CGM is a potentially rich source of information about galaxy evolution.

In the local universe, the CGM of star-forming galaxies is typically probed via absorption lines in the spectra of background quasars passing at small impact parameter from the foreground galaxies. Such observations show that the CGM is a multiphase, complex environment with many physical origins (Chen, 2017; Tumlinson et al., 2017). More massive quenched galaxies are often associated with X-ray halos (Mulchaey, 2000; Rosati et al., 2002), showing that their CGM is predominant in the form of hot gas. Nonetheless, absorption line observations
around massive galaxies show the presence of cool gas (Thom et al., 2012; Zahedy et al., 2019), indicating that the CGM is multi-phase at all halo mass scales. Given that absorption lines can typically probe to much lower gas densities than emission measures such as X-ray and resonant line maps using integral field units, they hold great promise in being able to elucidate the physical and dynamical conditions in the CGM across a wide range of galaxy masses and types.

The primary drawback of absorption-line measurements is that typically each galaxy has only one or at most a few lines of sight by which to extract information. Current absorption line samples also remain relatively small because many of the key transitions are in the ultraviolet, requiring space-based observations at low redshifts or else optical observations of fainter distant sources at high redshifts. Unfortunately, the intrinsic variations in the CGM can be large even within a narrow range of galaxy or halo masses (Hani et al., 2019), which means that fully characterising the diversity of CGM in absorption is observationally daunting.

One way to augment such observational studies towards a more physical picture is to look at the CGM in cosmological hydrodynamic simulations (Ford et al., 2013; Hummels et al., 2013; Ford et al., 2014, 2016; Nelson et al., 2018b; Oppenheimer et al., 2018b; Hafen et al., 2019; van de Voort et al., 2019; Peeples et al., 2019; Hafen et al., 2020). If such simulations reasonably reproduce the observed absorption line statistics and other galaxy properties, they could potentially offer a full 3-D view of the CGM, and enable us to interpret the physical and dynamical state of the absorbing gas, at least within a context of a given model. Moreover, simulations used as numerical experiments allow us to test the relationship between CGM observables and galactic feedback processes, in particular star formation feedback at lower mass scales and AGN feedback in \( L^* \) galaxies. As such, simulations have become an important tool in understanding the baryon cycle via CGM absorption line data (Tumlinson et al., 2017).

At low redshifts, Hubble’s Cosmic Origins Spectrograph (COS) provides the main datasets for constraining models. Among key ideas emerging from low-\( z \) CGM simulations are that low ions arise in cool, dense gas located closer to galaxies, while high ions like O VI arise more in extended hotter gas that can be heated from virialisation as well as feedback (Ford et al., 2013; Hummels et al., 2013). Ford et al. (2014) used particle tracking to show that high ions trace metals deposited long ago via outflows, while low ions trace inflowing/outflowing gas within the past few Gyr. Comparing these same simulations to the COS-Halos sample, Ford et al. (2016) found good agreement for H I and some metal lines
and an underproduction of Si\textsc{iii} and O\textsc{vi}. Gutcke et al. (2017) used H\textsc{i} and O\textsc{vi} absorption to trace cold and hot CGM gas in the NIHAO simulation suite (Wang et al., 2015), finding that the inner CGM is dominated by extended gas disks. The EAGLE simulation (Crain et al., 2015; Schaye et al., 2015) reproduces COS-Halos observations of low metal ion absorption, arising from cool clouds in a hot ambient medium (Oppenheimer et al., 2018b). When including time-variable AGN radiation in EAGLE to mimic overionisation in the CGM, O\textsc{vi} absorption can increase to the level needed to match COS-Halos (Oppenheimer et al., 2018a). IllustrisTNG (Pillepich et al., 2018) predicts O\textsc{vi} absorption in good agreement with COS-Halos and reproduces the dichotomy of O\textsc{vi} absorption seen between star forming and quenched galaxies (Nelson et al., 2018b). In the Romulus25 simulations (Tremmel et al., 2017) O\textsc{vi} absorption traces black hole mass and accretion history (Sanchez et al., 2019). These examples illustrate how comparisons to low-\(z\) CGM absorbers, primarily from COS, provide valuable insights on the physical state of CGM gas along with constraints on galaxy formation simulations.

Here we use the SIMBA cosmological hydrodynamic simulation to explore the CGM around low-redshift galaxies. SIMBA has been shown to reproduce a wide variety of galaxy observations at low redshifts (Chapter 3 of this thesis, also Davé et al., 2019; Li et al., 2019; Thomas et al., 2019; Davé et al., 2020; Thomas & Davé, 2022), as well as low-redshift Ly\(\alpha\) statistics along random lines of sight (Christiansen et al., 2020) particularly when employing the recent Faucher-Giguère (2020) model of the ionising background. Being a cosmological simulation, SIMBA has \(\sim\)kpc resolution within dense regions, degrading in the CGM owing to the adaptive gas smoothing scale. This means it does not have as high resolution as zoom simulations, particularly ones that target the CGM, which is likely to most impact the statistics of low-ionisation absorbers that come from the densest CGM gas (van de Voort et al., 2019; Peeples et al., 2019) while having lower impact on mid- and high-ionisation absorbers that are most commonly detected in the CGM. On the other hand, a major advantage is that SIMBA models a representative cosmological volume of 100 \(h^{-1}\)Mpc, enabling statistical comparisons against observations, including being able to mimic selection criteria on host galaxy properties, while at the same time consistently reproducing IGM statistics (see e.g. Sorini et al., 2020).

In this chapter, we compare SIMBA CGM absorption directly to the COS-Halos (Tumlinson et al., 2013) and COS-Dwarfs (Bordoloi et al., 2014) samples.
These samples span three orders of magnitude in stellar mass, including both star-forming and quenched systems. By performing a careful comparison to these data using mock absorption spectra around similarly-selected galaxies, we can examine the CGM around a wide range of galaxy types, providing a detailed test of SIMBA’s CGM predictions as well as insights into the physical state of the CGM gas. Moreover, by using SIMBA variants with individual feedback modules turned on and off, we can perform numerical experiments to understand the key physical drivers of CGM absorption within SIMBA, as well as test the impact of the assumed photoionising background among various recent determinations thereof. This chapter builds and extends that of Ford et al. (2016) in three main ways: (i) we study the CGM around star-forming and quenched galaxies separately, enabled by the observationally-concordant dichotomy in the galaxy population from SIMBA; (ii) we study the correlation between input physics and CGM properties using SIMBA’s variant runs with individual feedback modules turned on and off; and (iii) we examine the effect of assuming several different meta-galactic photo-ionising backgrounds.

This chapter is organised as follows. In §4.2 we present the mass and metal budgets of the $z = 0$ galaxy halos in SIMBA. In §4.3 we compare H I and metal line absorption in SIMBA to the COS-Halos and COS-Dwarfs surveys. In §4.4 we examine the effect of stellar and AGN feedback on the CGM. Finally in §4.5 we conclude and summarise. Unless explicitly stated otherwise, distance units are expressed in co-moving coordinates.

4.2 CGM Mass and Metal Budget

4.2.1 Classifying CGM Gas

To begin, we examine the mass and metal budget of CGM gas in various phases within SIMBA around different galaxy types. Such mass and metal budgets are key physical properties that observations aim to quantify using CGM absorption (Peeples et al., 2014). This will set the stage for understanding the absorption line statistics that probe these different CGM phases, and how these are impacted by various modeled physical processes. We focus on $z = 0$, but these CGM budgets do not vary significantly over the range of redshifts explored by COS-Halos and COS-Dwarfs.
We classify host galaxies and CGM gas as follows. We consider only central galaxies, as the CGM of satellite galaxies is not easily distinguishable from the CGM of their central galaxies. Instead of treating satellite galaxies separately, we simply include contributions from satellite galaxies in the mass and metal budget of central galaxies, as would likely be done observationally. We consider the extent of the CGM to be within the galaxy’s FOF halo (but see e.g. Prochaska et al., 2011; Wilde et al., 2021, for alternative definitions). We have verified that defining the CGM to be within R_{200} instead results in very minor differences to the results below.

For each central galaxy, we compute the baryon mass and metal mass of each component within its CGM: stars, ISM gas, dust, wind particles, and cool, warm, and hot CGM gas. We define the wind phase as particles that are currently hydrodynamically decoupled from the surrounding gas (described in §2). As mentioned in §2 ISM gas has hydrogen number density $n > n_{th}$; this includes all star-forming gas. All remaining gas elements within the halo are considered part of the CGM, which is split into 3 temperature ranges: ‘cool’ gas dominated by photoionisation $T < T_{\text{photo}}$, set at $10^{4.5} \text{ K}$; ‘warm’ gas above the photoionisation threshold but below half the virial temperature of the halo ($T_{\text{photo}} < T < 0.5T_{\text{vir}}$); and ‘hot’ gas above half the virial temperature, $T > 0.5T_{\text{vir}}$. $T_{\text{vir}}$ is calculated according to equation 4 of Mo & White (2002):

$$T_{\text{vir}} = 3.6 \times 10^5 \left( \frac{V_c}{100 \text{ km s}^{-1}} \right)^2 \text{ K},$$

(4.1)

where $V_c$ is the circular velocity of the halo. We separate the hot and warm gas in terms of the virial temperature in order to draw a distinction between thermally stable virialised gas, and thermally unstable warm gas in a transition phase between thermalisation and the photoionised regime. We note that previous studies have typically taken this separation at a fixed temperature such as $10^{4.5} \text{ K}$ or $10^5 \text{ K}$ (e.g. Ford et al., 2013; Peeples et al., 2014), but given that we will be examining trends over a fairly wide range of halo masses, this partitioning better captures the physical demarcations between thermally stable and unstable gas over a range of halo masses. Equation 4.1 drops below $T_{\text{photo}}$ for $M_{\text{halo}} \lesssim 10^{10.6} M_\odot$, but in SIMBA these halos are below our resolution limit so do not factor into our analysis.
4.2.2 Mass Budget

Using these galaxy and phase classifications, we now examine the CGM mass budgets around central galaxies in Simba. The upper panels in Figure 4.1 show the median total mass of each halo baryonic component (hot CGM in magenta; warm CGM in peach; cool CGM in yellow; winds in light blue; dust in teal; ISM in green; stars in dark blue) for the halos of $z = 0$ central galaxies from the fiducial Simba simulation, as a function of central galaxy stellar mass. The lower panels show the median mass as a fraction of the baryon mass expected if the halo has retained its cosmic share of baryons, defined as:

\[ f_\Omega = \frac{M}{M_{\text{halo}}} \frac{\Omega_b}{\Omega_M}. \] (4.2)

The middle and right panels show the sample separated into star forming and quenched galaxies using a specific star formation rate (sSFR) cut of $\log(sSFR / \text{Gyr}^{-1}) > -1.8 + 0.3z$ as in Davé et al. (2019). The error bars in the upper panels span from the 25th to 75th percentiles of the data.

In general each component mass increases as a function of stellar mass. However relative contributions to the total mass budget vary substantially depending on the dominant physical mechanisms at play. At the high mass end, the trends are driven by galaxies that are quenched, primarily due to AGN jet feedback from massive black holes (Davé et al., 2019), whereas at the low mass end galaxies are predominantly star forming since their AGN feedback is comparatively weak. We examine this in more detail in §4.4.

Broadly, across all stellar masses the largest contributions to the mass budget come from stars, hot CGM, warm CGM, ISM gas, and cool CGM, while winds and dust contribute small amounts (see upper left panel). At low masses, the various CGM phases have comparable mass, growing with $M_*$ more slowly than the stellar mass. At $M_* \gtrsim 10^{10.5} M_\odot$, the hot and warm CGM start increasing more quickly, while the cool CGM and ISM continue a nearly flat trend with $M_*$. At $M_* \gtrsim 10^{11} M_\odot$, the hot CGM mass begins to dominate over the stellar mass, leading to significant X-ray emission (e.g. Robson & Davé, 2020). The low wind phase mass reassures us that measurements of metal-line absorption in our simulations are not significantly impacted by including hydrodynamically decoupled particles, except perhaps in the very lowest mass galaxies that have large mass loading factors. The dust mass contributes negligibly to the overall
Figure 4.1: Top panels: Median mass budget of different halo components at $z = 0$ (colour coded as in the legend inside the top-left panel), as a function of the stellar mass of the respective central galaxy. For every halo component, the error bars span the 25th-75th percentile intervals of the mass distribution within each stellar mass bin. The left panel includes all galaxies, whereas the central and right panels are restricted to star forming and quenched galaxies, respectively. Bottom panels: Median mass fraction in each halo component (colour coded as in the upper panels), relative to the expected mass of cosmological baryons in each halo, as a function of central galaxy stellar mass. The three panels refer to all, star forming and quenched galaxies, as above. The baryon fraction is below unity at all masses, particularly for low mass quenched galaxies; quenched galaxy halos contain mostly hot gas while star forming galaxy halos are more diverse in gas phase.
CGM mass budget, but may have interesting consequences for background source reddening (Peek et al., 2015); we will explore this in future work.

These trends can be better understood by separating the galaxy population into star-forming and quenched galaxies (see the middle and right panels of Figure 4.1). For star-forming galaxies, the hot CGM increases commensurately with the stellar mass, and is always comparable to it. At low masses this likely owes to hot winds carrying hot material into the CGM, while at higher masses this owes to the onset of a shock-heated gaseous halo (Birnboim & Dekel, 2003; Kereš et al., 2005), leading to a coincidentally steady trend with \( M_* \). The ISM increases more slowly with \( M_* \), leading to lower galaxy gas fractions at higher masses (Davé et al., 2020). The cool and warm CGM masses roughly track the ISM mass and are broadly comparable to it. We note that the scaling of hot CGM with stellar mass owes to our choice of a halo mass-dependent division between hot, warm, and cool CGM gas; previous results using a constant temperature threshold tended to find that the hot component becomes small at low masses (Kereš et al., 2009; Faucher-Giguère et al., 2011; Gabor & Davé, 2015), but this owes to the fact that the virial temperature becomes comparable or below what is canonically considered hot gas.

The quenched galaxies (rightmost panels), on the other hand, show significantly different trends particularly for the warmer CGM phases. The hot and warm CGM masses increase dramatically with \( M_* \), from being nonexistent at \( M_* \lesssim 10^{10.5} M_\odot \) to dominating the overall halo baryonic mass at \( M_* \gtrsim 10^{11} M_\odot \). This primarily owes to virial shock heating, exacerbated by the onset of halo quenching owing primarily to jet AGN feedback in Simba, which enacts quenching by heating halo gas (e.g. Gabor & Davé, 2015). The warm gas increases in tandem because as hot gas accumulates more of it is able to cool through the thermally unstable warm regime. Interestingly, the quenching does not fully remove the gas from the ISM and cool CGM phases; these continue to grow in mass even in hot halos at a similar rate as seen in star-forming galaxies. It may be surprising to see that the halos of massive quenched galaxies contain a significant mass of cool CGM gas, hinting that the CGM of these quenched galaxies is multi-phase. To a lesser extent this is also true of ISM gas, although this gas is likely to be associated with satellite galaxies’ ISM.

The wind mass for star-forming galaxies is roughly constant with stellar mass, reflecting a transition in the dominant wind outflow mechanism. The majority of the wind mass is linked to star-formation driven feedback, which has high mass
and metal loading factors and long decoupling times to enable metal deposition in the CGM. Star-formation driven winds are most effective in low mass galaxies, resulting in high wind masses. Conversely, there is little mass in AGN-driven winds due to their relatively lower mass and metal loading factors and short decoupling times, leading to lower wind masses around quenched systems.

The bottom panels of Figure 4.1 show that the overall baryon fraction is well below unity at all stellar masses, indicating that the halos have less mass in baryons than their cosmic share. This is a non-trivial result; in the simplest scenario, one expects that baryons would cool and condense into the central region of the halo, and the reduced pressure support would result in this fraction exceeding unity, which indeed happens in simulations without any feedback (Davé, 2009). The introduction of feedback results in significant fractions of the baryons being lost to the IGM. As noted in Borrow et al. (2020) and Robson & Davé (2020), SIMBA halos around \( L^* \) galaxies struggle to retain more than 30% of their baryons.

Star forming galaxies (lower middle panel) are able to retain a higher fraction of their baryons, but even without the strong AGN feedback that would enact quenching, their halos still only retain around 40-50% of their baryons. Examining a SIMBA run without jet-mode feedback shows that the halos of star forming galaxies retain a much higher fraction of their baryons, between \( \sim 60\% \) for \( M_\star \sim 10^{9.5} \) M\(_\odot\) and \( \geq 90\% \) for \( M_\star > 10^{11} \) M\(_\odot\) (see §4.4). The stellar fraction continues to grow to the highest masses, consistent with observations of star-forming galaxies above \( L^* \) (Posti et al., 2019). Cui et al. (2021) has shown that in SIMBA, red and blue galaxies follow different stellar-to-halo mass relations that are individually in good agreement with data. This indicates that AGN jet mode feedback has two effects – both evacuating baryons from their own host halos, and also reducing the overall fraction of baryons contained in any halo. The overall baryon fraction in SIMBA follows a similar trend to IllustrisTNG, with a clear minimum around \( L^* \) systems, while EAGLE’s baryon fractions stay more depressed to the lowest masses (Davies et al., 2019).

Quenched galaxies (lower right panel) with \( M_\star \lesssim 10^{11} \) M\(_\odot\) only contain \( \sim 20\% \) of their share of baryons, indicating that feedback is highly effective at removing baryons and thus preventing accretion. Feedback is most efficient in this mass range since the halos lack the gravitational potential to retain their baryons against the energetic input from AGN feedback. The jets could be quenching star-forming galaxies by heating and ‘boiling off’ their gas supply, or by limiting accretion of new gas via preventive feedback (e.g. Lu et al., 2015, 2017; van de
The roughly constant fraction of baryons with stellar mass hints that it is the latter; if halos were losing gas due to heating we would expect an increase in baryon fraction in halos with higher gravitational potential. Galaxy clustering may compound this effect on higher mass galaxies (explaining the downturn at high mass), since these galaxies are likely to exist in high density regions (Gabor & Davé, 2015; Cucciati et al., 2017; Kraljic et al., 2020a) and hence are more likely to have neighbours with strong jet-mode AGN feedback. We note that Robson & Davé (2020) found good agreement with observed hot gas fractions as a function of halo mass at galaxy group scales in SIMBA, and demonstrated that this owes to jet feedback in SIMBA.

Low-mass ($M_\star \lesssim 10^{10} M_\odot$) quenched centrals could be influenced by environmental processes associated with large-scale structure or AGN feedback from nearby massive galaxies; SIMBA reproduces the numbers of such low-mass central galaxies quite well (see figs. 4 and 5 in Dickey et al., 2021). Finally, we note the strong drop in the overall halo baryon fractions at $M_\star \sim 10^{10} M_\odot$, which occurs because the black holes here typically grow large enough to turn on jet feedback, resulting in the appearance of quenched galaxies (Davé et al., 2019). The drop is not evident in the star forming and quenched galaxy samples, but happens in the combined sample because the fraction of quenched galaxies jumps (perhaps too) abruptly at this mass.

In summary, the interplay of various feedback mechanisms results in a complex dependence of the halo baryon fraction on stellar mass in SIMBA, being most suppressed between $10^{10} \lesssim M_\star \lesssim 10^{11} M_\odot$ and higher outside these masses. This is particularly evident when looking at quenched galaxies, which have a total halo baryon fraction of $\sim 20\%$ for $M_\star \lesssim 10^{11} M_\odot$, indicating widespread evacuation of halos by feedback that enacts rapid quenching (Rodríguez Montero et al., 2019). Oppenheimer et al. (2020a) noted that CGM evacuation is also the cause of quenching in the EAGLE simulation. Moreover, quenched galaxies’ CGM are dominated by hot gas, although cool CGM gas has not completely vanished. Star-forming galaxies meanwhile have a multi-phase CGM, and show a fairly constant baryon fraction of $\sim 40 - 50\%$ at all masses. Interestingly, there is always a substantial hot CGM component of $\sim 30 - 40\%$ of the total halo baryons, which indicates that virial-temperature gas is omnipresent even at low $M_\star$; the CGM of such galaxies at $z = 0$ is not completely dominated by cold infalling filaments as has been argued from previous simulations (Kereš et al., 2005; Dekel et al., 2009).
Figure 4.2: Top panels: As in Figure 4.1, except that we now show the median metal mass of each halo component. Bottom panels: Median metal mass fraction of each component relative to total available metal mass, following the same colour coding and panel structure as in Figure 4.1. Quenched galaxies have most of their metals locked into stars, whereas in star forming galaxies no single component dominates the metal mass fraction.

4.2.3 Metal Budget

Since CGM tracers are predominantly metal absorption lines, it is instructive to examine the metal content of halos within the various CGM phases and galaxy types. Figure 4.2 shows the median metal mass (upper) and metal mass fraction (lower) for each halo component (colours same as Figure 4.1), as a function of central galaxy stellar mass. The error bars in the upper panels indicate the 25th and 75th percentiles of the data. Galaxies are separated into star forming and quenched using the same sSFR cut as in Figure 4.1.

At the low mass end, metals are evenly distributed across stars, ISM gas, wind particles, dust and cool CGM gas, with ~ 10% of metals in warm and hot CGM gas. Compared with their overall mass, warm and hot CGM gas contributes relatively little to the metal budget. At intermediate masses, metals are more preferentially locked into stars, and at the high mass end the metal mass of the hot CGM gas increases to ~ 35% of the available metals; both of these effects are due to the increasing mass of these components overall (see Fig. 4.1). Dust makes
a significant contribution to the metal mass, comparable to the ISM as found in Li et al. (2019). The large majority of dust is in the ISM; the dust ejected in (cool) outflows is highly uncertain since the production and destruction mechanisms under such conditions are not well constrained. The metal mass fraction in winds is strongly dependent on stellar mass: in galaxies with \( M_* > 10^{10} \, M_\odot \) winds compose a negligible proportion of the metals, whereas low mass galaxies have \( \sim 20% \) of their metal mass in wind particles, owing to the increase mass in winds and more substantive metal loading.

Star forming galaxies have a higher fractions overall of metals in dust, the ISM and cool CGM gas. The metal mass in stars increases steeply with stellar mass such that at the high mass end most metals are locked into stars, with the remaining metal budget composed equally of ISM gas, dust, and cool, warm and hot CGM gas. At low masses, star forming galaxies have a metal mass in dust that is comparable to that in stars, owing to the high mass overall of metal-loaded wind particles. Wind metal mass is roughly constant with stellar mass, such that the fraction at the high mass end is negligible. Overall, no single halo gas phase dominates the metal budget in star-forming systems.

Quenched galaxies have most of their metals locked into stars at moderate stellar masses. The higher fraction of metals in stars is due to the lower metal mass in other components as compared to star-forming galaxies; the metal mass locked into stars is similar at a given \( M_* \). The significant proportion of metals in warm and hot CGM gas is in line with the overall gas masses of these components. Metals could end up in these hotter phases by direct transportation of heated gas from the galaxies via AGN jets, or via heating by e.g. virialization or hot jet outflows in the CGM following ejection of cooler star formation driven winds; an investigation into how metals arrive in this phase is left for future work utilising particle tracking.

Peeples et al. (2014) conducted a halo metal census inferred from ionisation modeling of COS-Halos metal lines and other data. Our results on the relative amounts of metals in various phases are in fairly good agreement with this. Generally, Peeples et al. (2014) found that the fraction of metals in stars increases with \( M_* \), in ISM gas decreases with \( M_* \), and in dust is fairly constant; these qualitatively match SIMBA predictions. The CGM metal fractions were less well constrained, but overall they found that it increased to lower masses, which SIMBA likewise predicts. SIMBA results are thus broadly consistent with this more model-independent metal census.
Figure 4.3: Median metallicity of each halo component as a function of central galaxy stellar mass, separated into all (left), star forming (middle) and quenched (right) galaxies. The meaning of the error bars and the colour coding is the same as in Figure 4.1. For quenched systems the CGM metallicity is roughly constant with $M_*$, while for star forming systems it increases with $M_*$. While the metal mass is useful for budgeting, the metallicity is an indicator of the physical origins of the gas; low metallicity, ‘pristine’ gas is likely to have been accreted from the IGM, whereas higher metallicity gas has been enriched by star formation in galaxies, hence the mass-metallicity relation places constraints on galactic outflows (Finlator & Davé, 2008). Davé et al. (2019) showed that SIMBA generally reproduces the observed galaxy gas-phase metallicity–stellar mass relation at both $z = 0$ and $z \approx 2$; here we examine the metallicity of gas that has ended up outside the ISM.

Figure 4.3 shows the median metallicity of each halo component as a function of central galaxy stellar mass, separated into star forming and quenched galaxies. Metallicity is computed as the total metal mass fraction of each component $i$:

$$Z_i = \frac{M_{Zi}}{M_{\text{gas } i}},$$

scaled by the solar metallicity $Z_\odot = 0.0134$ (Asplund et al., 2009). Dust metallicity is not included in this figure since by definition dust is composed 100% of metals. The trends in stellar and ISM gas-phase metallicity with stellar mass represent the well-known mass-metallicity relation (e.g. Tremonti et al., 2004); see Davé et al. (2019) for a detailed discussion of these relations in SIMBA. Here we go further by decomposing the total gas into distinct CGM components.

The ISM and CGM components tend to show a rise in metallicity at low masses ($M_* \lesssim 10^{10.5} M_\odot$), and then a flat or even downwards trend towards higher masses,
in contrast to stellar metallicities which are continually increasing. In Finlator & Davé (2008), the galaxy metallicity scales as the yield divided by \((1 + \eta)\), where \(\eta\) is the mass loading factor. Thus for the ISM, this flattening trend can be understood as \(\eta\) approaching unity at \(M_\star \gtrsim 10^{10.5} M_\odot\). The CGM components, interestingly, reflect this ISM trend, which indicates that the CGM metals are enriched by the winds from the galaxy.

The hot CGM shows the lowest metallicity among all the halo gas components, indicating the highest amount of less enriched infall diluting the CGM contributions from outflows. The hot gas that is enriched is more likely to suffer metal line cooling and move to lower-temperature phases, which selects the lowest metallicity gas to remain hot. The increased halo baryon mass fraction at the highest masses along with the dropping metallicity suggests that at high \(M_\star\), the galaxies’ halos are preferentially able to retain their cosmic inflow against feedback, which is as expected owing to their larger potential wells.

Examining the star forming and quenched populations separately reveals an interesting dichotomy. For star-forming galaxies, the CGM uniformly increases in metallicity in all phases, except for the few very most massive systems, indicating continual enrichment up to the highest masses. In contrast, the quenched galaxy CGM metallicities are remarkably constant, though they show a mild peak at \(\sim 10^{10.3} M_\odot\). These galaxies are not forming new metals, so the CGM of these systems must have been enriched during the star-forming phase, or by accreting metals from other galaxies (Anglés-Alcázar et al., 2017a). The hot phase metallicity is \(\sim 1/4\) solar, which is broadly as observed in X-ray gas in groups and clusters, but shows a trend of increasing metallicity towards lower \(M_\star\) (and thus lower \(M_{\text{halo}}\)) systems that is an interesting prediction. The other components tend to be around solar to half-solar.

Overall, the metal mass budget of the CGM partly traces the gas mass budget, but is impacted by the different origin of gas around star-forming and quenched systems. Stars generally dominate the overall budget of metals remaining in the halo, except at the lowest masses probed here. The metals are generally multi-phase (like the mass) in star-forming systems, but in quenched systems the cooler components are important at low masses but the hot gas dominates the metal budget at \(M_\star \gtrsim 10^{11} M_\odot\). Many of these metals were likely accreted from the IGM at late epochs (Oppenheimer et al., 2012), and some may have been deposited by satellites via stripping or outflows; we leave an investigation on the origin of these metals for future work. The metallicities of these components also
show stark differences between star-forming and quenched systems, with star-forming galaxies having increasing metallicities with $M_*$ in all components, while quenched systems have approximately constant metallicities at all $M_*$, apart from in the stellar component. The low metallicity ($\sim 10\%$ solar) of the hot CGM is a reflection of both less enriched infall and metal cooling. These trends are predictions from SIMBA that should in principle be decipherable from a suite of well-characterised CGM absorbers.

4.3 Comparison to COS-Halos and COS-Dwarfs

We now examine the CGM in absorption in SIMBA. We focus on recent absorption-line studies using the Cosmic Origins Spectrograph (COS) probing the CGM of $L^*$ galaxies (COS-Halos, Werk et al., 2013, 2014; Tumlinson et al., 2013) and sub-$L^*$ dwarf galaxies (COS-Dwarfs, Bordoloi et al., 2014). In order to test how well the CGM might be represented in SIMBA, we perform a close comparison to these observations by examining the CGM in SIMBA in terms of observational absorption statistics. We conduct analogous absorption-line surveys within the SIMBA volume that mimic the selection properties and instrumental characteristics of these observational surveys.

4.3.1 Mocking the COS surveys

Galaxy selection

To compare to these observations of the CGM in absorption, we choose galaxies from the fiducial SIMBA run whose properties are similar to those of the surveyed galaxies. Overall, our procedure for this closely follows that in Ford et al. (2016). Other CGM absorption studies have compared specifically to COS-Halos, either by similarly creating matched galaxy samples (Oppenheimer et al., 2016; Nelson et al., 2018b) or neglecting this extra step of observational realism (e.g. Hummels et al., 2013; Ford et al., 2016).

For each galaxy in the observational sample, we select the 5 galaxies in SIMBA with the closest match in terms of stellar mass and sSFR. We scale stellar masses in COS-Dwarfs and COS-Halos to the Chabrier IMF assumed in SIMBA by dividing by a factor of 1.6 (Madau & Dickinson, 2014). We impose an isolation criteria
Figure 4.4: Specific star formation rate (sSFR) versus stellar mass ($M_*$) for the SIMBA galaxies in our COS-Dwarfs and COS-Halos samples. Points are colour-coded by their $r_{200}$-scaled impact parameter. Dark grey circles and light grey triangles represent galaxies in COS-Halos and COS-Dwarfs, respectively. Where sSFR is a lower-limit for these galaxies, this is indicated by a downward arrow. Where simulated galaxies have little or no star formation we set their sSFR to $10^{-11.5}\text{yr}^{-1}$ for plotting purposes, also indicated by downward arrows.
to mimic that of the COS surveys, requiring that no other central galaxies are within 1 $h^{-1}$Mpc of each selected galaxy.

For the COS-Dwarfs and COS-Halos samples we select galaxies from the $z = 0$ and $z = 0.25$ snapshots, respectively, closest to the typical redshifts of these surveys. We consider only galaxies with $M_* > 5.8 \times 10^8$ $M_\odot$, which is SIMBA’s nominal mass resolution limit, meaning we exclude the lowest mass galaxies in COS-Dwarfs. As in §4.2, we then separate star forming and quenched populations using a specific star formation rate cut of $\log(sSFR/\text{Gyr})^{-1} > -1.8 + 0.3z$. For COS-Dwarfs this gives 29 star forming and 8 quenched galaxies in the original sample, and 142 star forming and 43 quenched analogous L$^*$ galaxies in SIMBA (where the non-exact split is due to a COS-Dwarfs galaxy that is near to our sSFR boundary). For COS-Halos, this gives 27 star forming and 17 quenched galaxies in the original sample, and 135 star forming and 85 quenched analogous sub-L$^*$ galaxies in SIMBA.

Figure 4.4 shows stellar mass against sSFR, colour-coded by impact parameter $\rho$ scaled to $r_{200}$ of the host galaxy’s halo, for galaxies in our mock COS-Halos and COS-Dwarfs samples. SIMBA galaxies with little or no star formation are plotted as downward arrows at an sSFR of $10^{-11.5} \text{yr}^{-1}$; for matching purposes, these are treated at having an sSFR of $10^{-11.5} \text{yr}^{-1}$. The original survey galaxies are also shown as dark grey circles (COS-Halos) and light grey triangles (COS-Dwarfs). Grey arrows indicate lower-limit sSFRs for some observed galaxies. This shows that our simulated galaxies are a good representation of the COS samples.

Spectral generation

For each galaxy in our sample, we select lines of sight (LOS) through the simulation to probe its CGM at the impact parameter $\rho$ of the corresponding COS galaxy-quasar pair. We choose 8 LOS per galaxy with impact parameter $\rho$, starting from $(x_{\text{gal}} + \rho, y_{\text{gal}})$ and proceeding anti-clockwise, selecting a point every 45 degrees. All LOS are parallel to the $z$-axis of the simulation box. We generate absorption spectra for a selection of ions which probe a range of ionisation potentials that are covered by the COS surveys: H$^\text{i}$ 1215Å and C$^\text{IV}$ 1548Å for the COS-Dwarfs sample; H$^\text{i}$ 1215Å, Mg$^\text{II}$ 2796Å, Si$^\text{III}$ 1206Å, and O$^\text{VI}$ 1031Å for the COS-Halos sample. These are among the most commonly seen lines in these surveys, offering the largest statistical samples for comparison to SIMBA. When comparing H$^\text{i}$ absorption we ignore one COS-Dwarfs galaxy.
with an intervening Lyman limit system.

Mock absorption spectra are generated through the simulation volume using the spectrum generation tools in the PyGAD\textsuperscript{1} analysis package (Röttgers et al., 2020). The spectra are computed from gas particles whose smoothing lengths intersect with the LOS, so spectrum generation is not dependent on our halo-finding procedure. The ionisation fractions for each species are found using look up tables in density and temperature generated using version 17.01 of CLOUDY (Ferland et al., 2017). Self-shielding is applied for metal lines via an attenuation of the overall background following Rahmati et al. (2013) prescription\textsuperscript{2}. The particle densities are multiplied by the ion fraction for each species, assuming a $\approx 76\%$ mass fraction for H, and the individually-tracked mass fractions within Simba for each metal. The resulting ion densities are smoothed along the LOS using the same spline kernel used in the GIZMO simulation code, employing the smoothing length and (for metal ions) the relevant metal content of each particle, into $6 \, \text{km s}^{-1}$ width pixels (approximately corresponding to COS pixels) along the LOS. The resulting ion column densities for each pixel are converted into an optical depth using the oscillator strength of each species. See Röttgers et al. (2020) for further details on PyGAD.

We add effects to the spectra to approximately mimic the COS observations: i) spectra are convolved with the line spread function for the COS G130M grating (apart from the Mg\textsc{ii} spectra, whose observations were taken with Keck; Werk et al. 2013); ii) random noise is added with a signal-to-noise ratio (SNR) per pixel of 12, corresponding to SNR per COS resolution element of $\approx 20$; iii) we mock a continuum fitting procedure broadly following Danforth et al. (2016) by iteratively removing pixels that are $> 2\sigma$ below a polynomial best-fit to the spectrum until the continuum fit changes fractionally by less than $10^{-4}$.

**Choice of photo-ionising background**

Computing the ionisation fractions in CLOUDY requires assuming a spectrum of incident photons, which is critical since most of the CGM UV absorption lines detected are typically photo-ionised (Ford et al., 2013), potentially even

\textsuperscript{1}\url{https://bitbucket.org/broett/pygad}

\textsuperscript{2}This is applied as an overall reduction factor to the entire ionising background spectrum in the optically thin limit; this is not correct in detail since self-shielding also changes the spectral shape, but it is the best we can do in lieu of an expensive multi-frequency radiative transfer calculation.
O VI (Oppenheimer & Davé, 2009). We find that our choice of photo-ionising background can make a significant difference to the metal line absorption statistics; unfortunately this choice remains rather poorly constrained by direct observations.

We will thus compare results for three different ionising backgrounds: (i) Faucher-Giguère (2020) (hereafter FG20), an updated version of the (Faucher-Giguère et al., 2009) background; (ii) Haardt & Madau (2001) (hereafter HM01), an older determination utilised in many previous studies including Ford et al. (2013); and (iii) Haardt & Madau (2012), with the entire spectrum increased in amplitude by a factor of 2 (hereafter HM12x2).

The low-redshift H I photoionisation rate can be constrained by comparing to the observed H I column density distribution (e.g. Kollmeier et al., 2014) or mean flux decrement (e.g. Christiansen et al., 2020). For Simba, the mean flux decrement is reasonably well-matched to data assuming FG20, whereas the HM12 model requires an increase of a factor of 2. This gives a $z = 0$ H I photoionisation rate of $\Gamma_{HI} = 4.7 \times 10^{-14}$ and $\Gamma_{HI} = 4.5 \times 10^{-14}$s$^{-1}$ for our two backgrounds, respectively, so differences in absorption properties will reflect variations in the shape of the background above 1 Ry. HM01 is higher in terms of the H I photoionisation rate ($\Gamma_{HI} = 8.3 \times 10^{-14}$s$^{-1}$ at $z = 0$), and its shape is substantially different owing to a larger assumed galaxy escape fraction and not including a soft X-ray background; we include it here for comparison to previous results.

We note that Simba was run with the original Haardt & Madau (2012) background. Therefore none of the choices here are self-consistent with the simulation run. Fortunately, the ionising background makes a negligible difference in the hydrodynamical forces in cosmological simulations, so applying the correction in post-processing is a very good approximation (Weinberg et al., 1997).

**Absorption statistics**

From the synthetic spectra we compute 3 observables in order to compare with the COS surveys: (i) the equivalent width EW, computed within a $\pm 300$ km s$^{-1}$ window centered on each galaxy; (ii) the covering fraction $f_{cov}$, which is the fraction of lines of sight where the equivalent width exceeds the detection threshold; and (iii) the path absorption $dEW/dz$, which is the cumulative equivalent width in $\pm 300$ km s$^{-1}$ per unit redshift.
The EW is the most basic statistic for unresolved or marginally resolved absorbers, which is typically the case for metal lines from cooler gas observed with COS. For $f_{\text{cov}}$, we employ a detection threshold of 0.2 Å for H\textsc{i} and 0.1 Å for all metal ions, following Werk et al. (2013) and Ford et al. (2016). The covering fraction constrains the spatial structure and extent of the CGM absorption in each ion. The path absorption is effectively a stacked measure of circumgalactic absorption, computed directly from the spectra without any consideration of individual lines or detection thresholds. These measures thus provide differing quantifications of CGM absorption.

To compare to COS-Halos, we access the data from the PyIGM\textsuperscript{3} repository, which includes the equivalent widths, column densities and Doppler parameters for each metal line component of the absorption spectra. This repository compiles the data from the COS-Halos and COS-Dwarfs surveys with Hubble, along with ancillary data including Mg\textsc{ii} from Keck spectra and galaxy properties from spectroscopic follow-up. The data are reduced and analysed as described in Tumlinson et al. (2013); Werk et al. (2014), and the analysis tools are available on the repository (though here we rely on the reduced data products). For COS-Dwarfs, the H\textsc{i} and C\textsc{iv} data as a function of impact parameter was kindly provided to us by survey PI R. Bordoloi.

We focus on the EWs here, since in many cases the Doppler parameter cannot be directly determined from the spectra, and the PyIGM column densities therefore assume a Doppler parameter that may not be correct (Röttgers et al., 2020). The EW in PyIGM were measured in ±600 km s\textsuperscript{-1} windows around the galaxy redshift, after continuum fitting. In our case, we measure EW in a smaller window, but have checked that the contribution from ±300–600 km s\textsuperscript{-1} is negligible, as also found in Ford et al. (2016). We consider all gas along the line of sight through the entire volume that contribute within this velocity window; some of it may lie outside of halos and be projected within in redshift space (Rahmati et al., 2015; Nelson et al., 2018b).

As in Ford et al. (2016), we set a fixed EW detection threshold for each ion, motivated by the sensitivity limits of the COS surveys. The detection limits are listed in Table 4.1, and for reference the ionisation potentials are listed to show the range from low ions (Mg\textsc{ii}) to high (O\textsc{vi}). We apply the same limits to the simulated absorbers. Upper limits are treated as detections at that limit, but

\textsuperscript{3}https://pyigm.readthedocs.io/en/latest/
Table 4.1: Ion Properties and detection thresholds

<table>
<thead>
<tr>
<th>Ion</th>
<th>Ionisation Energy (Ry)</th>
<th>Detection Threshold (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H\textsubscript{i}</td>
<td>1.0</td>
<td>0.2</td>
</tr>
<tr>
<td>Mg\textsubscript{ii}</td>
<td>1.11</td>
<td>0.1</td>
</tr>
<tr>
<td>Si\textsubscript{iii}</td>
<td>2.46</td>
<td>0.1</td>
</tr>
<tr>
<td>C\textsubscript{iv}</td>
<td>4.71</td>
<td>0.1</td>
</tr>
<tr>
<td>O\textsubscript{vi}</td>
<td>10.15</td>
<td>0.1</td>
</tr>
</tbody>
</table>

since we consider medians here, this only affects the results when there are very few detections.

We choose to compare all spatial statistics with impact parameters normalised to \( r_{200} \), defined as the radius enclosing 200 times the critical density. We do this in order to be able to meaningfully compare profiles across the large mass range spanned by both COS-Halos and COS-Dwarfs, under the expectation that the influence of the baryon cycle around each galaxy is approximately proportional to the virial radius. There is evidence for such self-similarity at least in terms of Mg\textsubscript{ii} absorbers (Churchill et al., 2013). In §4.6 we show the EW versus physical impact parameter instead, but this choice does not impact our results; it does however make some of the radial trends less evident, as also noted in Ford et al. (2016).

For Simba we obtain \( r_{200} \) directly from the identified halos for each (central) galaxy. For the observations, \( r_{200} \) was computed in Werk et al. (2013) from the stellar mass based on the \( M_* - M_{\text{halo}} \) relation of Moster et al. (2010). We note that our simulations are in good agreement with the observed \( M_* - M_{\text{halo}} \) relation for massive halos where this is observed via e.g. weak lensing, even when considering quenched and star-forming galaxies separately (Cui et al., 2021).

### 4.3.2 Comparison to COS-Halos and COS-Dwarfs

We now compare Simba to COS-Halos and COS-Dwarfs using the absorption spectra and statistics described above. This section represents the main results of this Chapter.
Figure 4.5: Equivalent width of H\textsc{i} and selected metal lines against (as indicated in each panel) $r_{200}$-scaled impact parameter for the COS-Halos and COS-Dwarfs galaxy samples, using the FG20 (solid), HM12x2 (dashed) and HM01 (dotted) ionising backgrounds. Light blue and light pink lines represent star forming and quenched galaxies in the Simba sample, respectively; shaded regions show indicative cosmic variance uncertainties around the FG20 results. Dark blue and magenta points represent star forming and quenched galaxies in the COS samples (circles for COS-Halos, triangles for COS-Dwarfs), respectively; vertical error bars represent the 25th and 75th percentiles of the equivalent width distribution in the data, while horizontal error bars indicate the width of the bins. Black horizontal dotted lines indicate the detection threshold of each line under the COS survey conditions. CGM absorption is higher around star forming galaxies than quenched galaxies. Simba is in reasonable agreement with observations; H\textsc{i} absorption around star forming galaxies is well-reproduced, although metal line absorption is sensitive to the assumed ionising background.
Figure 4.5 shows median equivalent width of the selected ions, against $r_{200}$-scaled impact parameter for the COS-Halos and COS-Dwarfs galaxy samples, using the FG20 (solid), HM12x2 (dashed) and HM01 (dotted) ionising backgrounds. SIMBA star forming and quenched galaxies are represented by light blue and light pink lines, respectively, and shaded regions indicate the cosmic variance within eight sub-quadrants over the simulation volume for the FG20 results (which typically dominates over Poisson variance). Dark blue and magenta points represent the actual COS-Dwarfs/COS-Halos data, respectively; the error bars are the 25th and 75th percentiles of the data. The varying bin widths for the observational data are chosen such that each bin contains $\sim 8$ galaxies. The detection thresholds for each ion are indicated by horizontal dotted lines.

In general, for each ion and for both star forming and quenched populations, the equivalent width of absorption is higher around star forming galaxies, and decreases with impact parameter. This latter feature is enhanced by scaling impact parameter by $r_{200}$, as can be seen by comparing to Figure 4.11 in §4.6, suggesting that CGM properties correlate better with $r_{200}$-scaled radii than physical radii. H$\text{I}$ absorption is $\sim 10$ and $\sim 3 - 4$ times higher in SIMBA star forming galaxies than in quenched galaxies, for the mock COS-Halos and COS-Dwarfs samples respectively (upper left and upper middle panels). The predicted H$\text{I}$ absorption around star forming galaxies in SIMBA is in good agreement with both observational surveys. The accurate prediction of H$\text{I}$ absorption around star-forming galaxies is a non-trivial success that did not require any specific tuning.

In contrast, SIMBA’s quenched galaxies show significantly less H$\text{I}$ absorption compared with the observations. Quenched galaxies in SIMBA contain little cold gas in their halos (see Figure 4.1) since the quenching mechanism heats and removes the cold gas supply and prevents further accretion via jet feedback. Meanwhile the COS-Halos observations suggest that real galaxies are quenched in such a way that they retain substantial cool gas in their CGM. The presence of cool gas around quenched galaxies remains an interesting and non-trivial quandary (Thom et al., 2012; Tumlinson et al., 2017) to constrain quenching models, though more data is needed to increase the statistical significance.

Comparing COS-Halos H$\text{I}$ (upper left) with COS-Dwarfs H$\text{I}$ (upper middle), we see that there is substantially less separation between the star-forming and
quenched absorption in the lower-mass sample. This suggests that such lower mass galaxies are perhaps quenched by a different physical process. In Simba, the AGN jet feedback that quenches galaxies only turns on at $M_\star \gtrsim 10^{10}\, M_\odot$, so most COS-Dwarfs galaxies would not have this. They may instead be ejected satellites or “neighborhood quenched” galaxies (Gabor & Davé, 2015), although the isolation criterion used in observations that we mimicked tried to avoid this. We note that Dickey et al. (2021) examined the quenched fraction of dwarf galaxies in Simba compared to observations, and although there is an excess of quenched systems at the very lowest masses, in the COS-Dwarfs mass range Simba well reproduces the observed fraction of quenched dwarfs.

Choice of photo-ionising background has little impact on our H\textsc{i} absorption results. Assuming instead the HM12x2 background results in almost identical H\textsc{i} absorption around star forming galaxies in the COS-Halos and COS-Dwarfs samples, while the HM01 background results in slightly lower H\textsc{i} absorption. Around quenched galaxies in the COS-Halos and COS-Dwarfs samples, the HM12x2 and HM01 backgrounds both result in slightly higher H\textsc{i} absorption than the FG20 background. These are small differences within the uncertainties; H\textsc{i} absorption is a less sensitive probe of ionisation variations due to the saturation of absorption features which places them on the logarithmic part of the curve of growth.

Cool CGM gas can also be explored via low ionisation lines such as Mg\textsc{ii} (upper right panel) and Si\textsc{iii} (lower left panel). In this case, the story is less conclusive. For the FG20 and HM12x2 backgrounds used here, Simba underpredicts the EWs of these low ionisation lines around star-forming galaxies by $\sim \times 3$. However, assuming HM01, Simba shows much better agreement. This highlights the sensitivity of the assumed ionising background shape in comparing to metal lines. Furthermore, there are uncertainties in the Mg and Si yields assumed in Simba; we note that Simba reproduces the observed $z = 0$ stellar mass – gas-phase metallicity relation fairly well (Davé et al., 2019), although this is primarily driven by the oxygen abundance. Meanwhile, Simba’s quenched galaxies are much less sensitive to ionising background and are not obviously in disagreement with the corresponding COS-Halos observations, modulo the large uncertainties in the data, and the fact that most of the absorbers are close to the detection limit. Larger samples of Mg\textsc{ii} absorbers exist around more massive quenched galaxies (e.g. Zahedy et al., 2019), and we plan to compare to such samples in the future. Similarly, for Si\textsc{iii} around quenched galaxies, Simba predicts absorption
that is within the range of the large uncertainties. This indicates that SIMBA quenched halos do contain at least some cool absorbing gas, whose origin is an interesting question that we will investigate in future work.

$C_\text{IV}$ (1548,1550Å) is more easily traceable in COS-Dwarfs owing to its $z \sim 0$ sample, while at COS-Halos redshifts it moves beyond the COS/G160M grating’s coverage. This is an interesting mass regime because for a $10^{9.5}$ M$_\odot$ galaxy, the typical virial temperature of its halo is $\sim 10^{5.3}$K, which is near the collisional ionisation peak of $C_\text{IV}$. Hence $C_\text{IV}$ can trace both photo-ionised and collisionally ionised gas in COS-Dwarfs, since as Figure 4.1 shows, both star forming and quenched galaxies have substantial gas in the hot phase.

Comparing to observations, the equivalent widths of $C_\text{IV}$ absorbers around star forming galaxies agree in the central region, but SIMBA over-predicts them in the outskirts. Around quenched galaxies, SIMBA over-predicts $C_\text{IV}$ absorption at a similar level, though the statistical significance is lower. Assuming a different ionising background does not result in any significant change in absorption, suggesting that the ionising background has less impact at these lower masses for $C_\text{IV}$. Generally, SIMBA doesn’t reproduce as strong a radial gradient in EWs as seen in the COS-Dwarfs data. This is also mildly seen in the radial trend for H$\text{I}$. This could indicate that the metals in SIMBA are being distributed too far via outflows around low-mass galaxies, although this issue is not evident around the higher-mass COS-Dwarfs galaxies. Alternatively, it could be that the ionisation conditions are incorrect in the outskirts of SIMBA dwarf halos.

Turning to O$\text{VI}$ in COS-Halos (lower right), assuming either an FG20 or HM12x2 background SIMBA slightly over-predicts O$\text{VI}$ absorption around star forming galaxies by a factor of $\sim 2$. Around quenched galaxies, O$\text{VI}$ absorption is in reasonable agreement within the uncertainties, and reflects the dichotomy in O$\text{VI}$ absorption around quenched vs. star-forming galaxies owing to oxygen moving into higher ionisation states in the hot gas around quenched systems (Oppenheimer et al., 2020a). Strong O$\text{VI}$ absorption in the CGM has been associated with high star formation activity (e.g. Tumlinson et al., 2011; Zahid et al., 2012; Suresh et al., 2017) since oxygen is produced in Type II SNe events, but simulations suggest that the oxygen in today’s CGM was deposited many Gyr ago (Ford et al., 2014). Interestingly, previous studies of metal-line absorption in simulations (e.g. Hummels et al., 2013; Ford et al., 2016; Liang et al., 2016; Oppenheimer et al., 2016; Gutcke et al., 2017; Suresh et al., 2017) with the exception of IllustrisTNG (Nelson et al., 2018b) have found an under-prediction
of O\textsc{vi} around star forming systems by a factor of a few – SIMBA is much closer in agreement to the observations on this front. The increase seen in SIMBA owes in part to our choice of ionising background – assuming instead the HM01 background as was done in Ford et al. (2016) results in an order of magnitude lower level of O\textsc{vi} absorption, as seen by the dotted line. This highlights that comparisons of O\textsc{vi} absorption between models can be sensitive to the ionising background assumed, and without better empirical constraints on the strength of the extreme UV background, it is difficult to make robust comparisons to data. In contrast, as with the lower ions, O\textsc{vi} absorption around quenched systems is much less sensitive to ionising background. Another factor may be that SIMBA’s outflows are ejected 30\% in SN-heated gas, while the Ford et al. (2016) simulations ejected winds solely in a cool phase.

Overall, depending on tracer, distance, and assumed UVB, SIMBA either reasonably reproduces the observed H\textsc{i} and metal line absorption in the CGM, under-predicts the median EW by up to an order magnitude, or over-predicts EW by a factor of a few. Particularly interesting discrepancies include the deficit of H\textsc{i} absorption around quenched galaxies, an under-prediction of Si\textsc{iii}, and a too-weak radial gradient in C\textsc{iv}. The choice of assumed ionising background is an important uncertainty in comparisons to observations, although the results for more modern determinations (HM12x2 and FG20) are more similar.

**Covering fractions**

Figure 4.6 shows $f_{cov}$ for each ion against $r_{200}$-scaled impact parameter for the COS-Halos and COS-Dwarfs galaxy samples, using the FG20 (solid), HM12x2 (dashed) and HM01 (dotted) ionising backgrounds. SIMBA star forming and quenched galaxies are represented by light blue and light pink lines, respectively; shaded regions show the typical Poisson errors. Star forming and quenched COS-Halos and COS-Dwarfs observations are represented by dark blue and magenta points, respectively. Horizontal error bars on the observations indicate the width of the bins, and vertical error bars are Poisson uncertainties. We note that we have applied the same detection thresholds (listed in Table 4.1) to the observations sample to ensure a fair comparison.

In the observations and in SIMBA, the star forming galaxies show a higher $f_{cov}$ than quenched galaxies. Also, in general the SIMBA galaxies show a radial decrease in $f_{cov}$ for both the COS-Halos and COS-Dwarfs samples, but this trend
Figure 4.6: $f_{\text{cov}}$ for H\textsc{i} and selected metal lines (as indicated in each panel) against $r_{200}$-scaled impact parameter for the COS-Halos and COS-Dwarfs galaxy samples, using the FG20 (solid), HM12x2 (dashed) and HM01 (dotted) ionising backgrounds. Light blue and light pink lines represent star forming and quenched galaxies in the Simba sample. The shaded regions indicate typical Poisson errors. Dark blue and magenta points represent star forming and quenched galaxies in the COS samples (circles for COS-Halos, triangles for COS-Dwarfs); horizontal error bars indicate the width of the bins and vertical error bars are Poisson errors. With a few exceptions, Simba reasonably reproduces observed $f_{\text{cov}}$ for most species, particularly the dichotomy between star forming and quenched systems seen in O\textsc{vi}.

is less clear in the data given the errorbars. These trends broadly mimic those seen for the observed EWs.

For H\textsc{i} around star-forming systems, Simba shows near unity covering fraction around the larger galaxies in COS-Halos, but a strong negative gradient in $f_{\text{cov}}$ for the dwarf sample. These trends impressively follow those seen in the data, suggesting that Simba reasonably represents the trend in cool halo gas versus stellar mass. In contrast, the quenched galaxies show more H\textsc{i} in the outskirts than seen in Simba. The extremely high $f_{\text{cov}}$ at large radii in COS-Halos is curious and highly discrepant from Simba, though it may be a product of small number statistics.

The trends for the metal ions broadly follow those seen for the EWs. Mg\textsc{ii}
is in good agreement with observations, with a higher covering fraction around star forming galaxies than quenched galaxies as also observed in larger samples of Mg\textsc{ii} absorbers (e.g. Lan et al., 2014; Lan, 2020). Si\textsc{iii} around star-forming galaxies in SIMBA is low by \(~30\%\) for FG20 and HM12×2, but in good agreement if HM01 is assumed. The discrepancy in Si\textsc{iii} around quenched galaxies is more pronounced. $f_{\text{cov}}$ of C\textsc{iv} is \(~20\%\) too low at all impact parameters. Perhaps the most impressive agreement is for O\textsc{vi}, for which a large dichotomy between the star-forming and quenched populations is seen, and nicely reproduced in SIMBA; we note that this arises without any specific tuning, although there is minor variation in this owing to the choice of ionising background. With modest exceptions, SIMBA generally reproduces the observed covering fractions around these galaxy populations.

**Path absorption**

Figure 4.7 shows the path absorption $dE/W/dz$ for each ion against $r_{200}$-scaled impact parameter for the COS-Halos and COS-Dwarfs galaxy samples, using the FG20 (solid), HM12×2 (dashed) and HM01 (dotted) ionising backgrounds. As in the previous figures, SIMBA star forming and quenched galaxies are represented by light blue and light pink lines respectively; shaded regions show the typical cosmic variance uncertainties. Dark blue and magenta points represent the COS-Dwarfs and COS-Halos observations; horizontal error bars indicate the width of the bins and vertical error bars are the standard variation within each bin. For the data, we compute these values by summing the equivalent widths of all the lines above the detection thresholds as indicated in Figure 4.5, and then dividing by the $\Delta z$. We cannot fully mimic the procedure that we do in the simulations of adding up all the equivalent widths directly from the spectra because this would involve reducing the spectra and associating absorption features with specific ions, which is an involved process that is beyond the scope of this work. Nonetheless, at least in SIMBA, the path absorption around star-forming systems is dominated by lines above these detection thresholds, so it is reasonable to compare these results directly to observations since the uncounted equivalent width below the detection limit in the data will not dominate. In contrast, for the quenched galaxies, this is typically not the case; hence this comparison should be viewed with some caution, nonetheless we show it as the red lines. Our predicted path absorption could be
Figure 4.7: Total path absorption of H\textsc{i} and selected metal lines (as indicated in each panel) against $r_{200}$-scaled impact parameter for the COS-Halos and COS-Dwarfs galaxy samples, using the FG20 (solid), HM12x2 (dashed) and HM01 (dotted) ionising backgrounds. Light blue and light pink lines represent star forming and quenched galaxies in the SIMBA sample; shaded regions show typical cosmic variance uncertainties. Dark blue and magenta points represent star forming and quenched galaxies in the COS samples (circles for COS-Halos, triangles for COS-Dwarfs); horizontal error bars indicate the width of the bins while vertical error bars are the standard deviations from computing path absorption individually for each galaxy. SIMBA closely reproduces total path absorption of H\textsc{i} around star forming galaxies; for metal lines in the COS-Halos sample, our results are sensitive to ionising background.

Compared to data of any quality, so it provides a more general prediction for future spectral stacking experiments to quantify weak CGM absorption.

Overall, the results for total path absorption closely resemble that of equivalent width. As seen previously with equivalent width and covering fractions, total path absorption is higher in star forming galaxies and decreases as a function of impact parameter. SIMBA closely reproduces the total path absorption of H\textsc{i} around star forming L$^\ast$ galaxies and broadly produces metal line total path absorption that generally agrees with the COS-Halos observations, however Mg\textsc{ii} and Si\textsc{iii} (O\textsc{vi}) are slightly under (over) predicted. SIMBA also closely reproduces the total path absorption of H\textsc{i} and C\textsc{iv} around sub-L$^\ast$ star forming galaxies, with slight over-prediction of C\textsc{iv} in the outskirts. Thus in general, the path absorption tells a...
similar story to the other measures, although with improved statistics it could potentially highlight important differences in the weak absorber population.

### 4.3.3 Summary of observational comparisons

SIMBA broadly reproduces the COS-Halos and COS-Dwarfs observations of H\emph{i} and metal CGM absorbers, but there are some significant discrepancies as well. A notable success is that the simulations reproduce the near-unity H\emph{i} covering factors around star forming L\emph{*} galaxies, together with the strong radial gradient around dwarfs. In contrast, SIMBA under-produces H\emph{i} around quenched galaxies. \textcite{van2019} and \textcite{hummels2019} have pointed out using CGM-focused zoom simulations that the H\emph{i} content of L\emph{*} halos, which typically contain substantial hot gas, is quite sensitive to numerical resolution (though see \textcite{suresh2019}, who found little sensitivity due to the already high resolution). It could be that SIMBA simply lacks the required resolution to model condensations of H\emph{i} within hot halos, even while reproducing the H\emph{i} in halos that have substantial cool gas.

For metal lines, the low ions Mg\emph{II} and Si\emph{III} are somewhat under-produced around star-forming galaxies. This might owe to these ions tracing denser CGM gas for which our predictions might be particularly sensitive to resolution effects. However, in §4.7 we show that the predictions are not strongly sensitive to resolution, for the range of resolutions we can probe among our runs. Meanwhile, the low ion predictions are quite sensitive to the ionising background; switching to HM01 yields quite good agreement, consistent with \textcite{ford2016}. Hence it is difficult to interpret this disagreement without independent constraints on the strength and shape of the ionising background.

O\emph{VI} present a significant success for SIMBA. Previous works highlighted the challenge for cosmological models to produce enough O\emph{VI} absorption around star-forming galaxies, but this is partially mitigated in SIMBA even when using the same ionising background (HM01), and disappears completely using a more recent background (HM12 or FG20). Moreover, SIMBA nicely reproduces the dichotomy in O\emph{VI} absorption between star forming and quenched systems, as also seen in IllustrisTNG and EAGLE (\textcite{nelson2018b, oppenheimer2020a}). This provides circumstantial evidence that sufficient O\emph{VI} arises from a volume filling warm-hot phase (\textcite{ford2013}) rather than interfaces of cold clouds in a hot medium (\textcite{heckman2002}) that would be unresolved in SIMBA.
Numerical convergence is nonetheless an important systematic in these comparisons. In §4.7 we examine the convergence with respect to both simulation resolution and volume in the COS-Halos and COS-Dwarfs comparisons above. In general, higher resolution produces more absorption overall, as does smaller volume; the latter owes to less high-mass objects with jet-mode AGN feedback heating. These variations are at the $\sim 0.1 - 0.2$ dex level for most ions, though for $\text{C} \text{iv}$ in the COS-Dwarfs sample it is higher. These systematics, comparable to those from varying the ionising background, must be kept in mind when assessing (dis)agreement with the observations.

These results encouragingly suggest that Simba is generally populating the CGM with cool gas and metals in accord with absorption line observations, but also highlight some of the challenges in interpreting such observations given current knowledge. One way forward is to attempt to connect CGM physical and absorption line properties with feedback processes, to understand how the CGM provides direct constraints on baryon cycling. This is what we do next.

### 4.4 Dependence on feedback models

Feedback plays a critical role in setting the properties of galaxies (Somerville & Davé, 2015; Naab & Ostriker, 2017), but less is understood about how such processes impact the CGM. To study the impact of feedback on the CGM, we make use of the following suite of feedback variant Simba simulations: No-Xray (no X-ray AGN feedback); No-jet (no jet-mode or X-ray AGN feedback); No-AGN (no jet-mode or X-ray or radiative AGN feedback); and No-feedback (No AGN or SF driven feedback). These are run in $50h^{-1}\text{Mpc}$ volumes with $512^3$ gas elements and dark matter particles, each with identical initial conditions. We note that these variants fail to reproduce key galaxy observations (Davé et al., 2019) – for instance without jet feedback we get essentially no quenched galaxies – but instead are intended as numerical experiments to connect specific feedback processes to their impact on the CGM.

By comparing between these variants, we can isolate how SF feedback and Simba’s AGN feedback modes impact the physical and observable properties of the CGM. We first study the composition of the CGM in the AGN variant volumes by focusing on mass fractions and metallicities of the halo components for each galaxy. We then examine how these variants impact absorber statistics,
Figure 4.8: Median mass fraction in each halo component relative to $M_{\Omega b}$ as a function of total halo mass for the AGN feedback variant simulations (from left to right: full SIMBA model, no X-ray AGN feedback, no jet-mode or X-ray AGN feedback, and no AGN or SN feedback). The colour coding of each component is the same as in Figure 4.1. We omit the model with SN feedback and no AGN feedback from this plot since there is little difference between this and the no-jet model. Stellar feedback reduces baryon fraction at the low $M_*$ end, while jet-mode AGN feedback heats CGM and ISM gas and strongly suppresses baryon fraction at the high $M_{\text{halo}}$ end.

in particular their path absorption.

4.4.1 Halo baryon fractions

Figure 4.8 shows the median mass fraction $f_\Omega$ in each component versus halo mass, scaled by the baryon mass expected if the halo has retained it cosmic share of baryons (i.e. $f_\Omega \equiv M_{\text{halo}} \frac{\Omega b}{\Omega M}$) for each of the AGN feedback variant simulations. We plot points only where there are >10 galaxies in a halo mass bin, and show results for the total galaxy population as there are too few quenched galaxies in the absence of jet-mode feedback. We compare at fixed halo (not stellar) mass in this case because the different variants have quite different efficiencies of star formation, while the halo masses are fairly consistent among these runs (Sorini et al., in prep.). The $50h^{-1}$Mpc volume reproduces the halo baryon fractions seen previously in the $100h^{-1}$Mpc volume (§4.2.2), except that the smaller volume does not probe the highest halo masses; this shows good volume convergence for the halo baryon fractions in the various phases.

Starting with the no-feedback run in the rightmost panel, we see that in the
absence of feedback, all but the smallest halos retain close to their cosmic share of baryons. The stellar fractions are very high, demonstrating overcooling (Davé et al., 2001; Balogh et al., 2001). While all halos have virial-temperature gas, the fraction of it increases steadily with halo mass. These establish the trends predicted without any baryon cycling.

Introducing SF feedback (i.e. the No-jet model) results in a significant suppression of baryons in low-mass halos. This follows the trend assumed in SIMBA that the mass loading factor in outflows scales inversely with stellar mass, and that the smaller potential wells enable easier escape from low-mass halos. Small halos are less able to retain their outflows, creating a baryon deficit.

The most dramatic change occurs when we turn on AGN jet feedback (No-Xray), causing a strong reduction in the halo baryon fraction around all but the smallest galaxies. This demonstrates that the jet-mode of our AGN model is responsible for the reduction in halo baryon fraction. In contrast, by comparing panels it is clear that that X-ray and radiative AGN feedback have comparatively little effect on the overall baryon fraction contained in halos. The reduction in baryon fraction for galaxies with $M_\star \gtrsim 10^{10}$ M$_\odot$ occurs when jet feedback begins turning on, and becomes stronger around higher mass galaxies. In the extreme case, at $M_\star \sim 10^{11.5}$ M$_\odot$ the fraction decreases from unity to 0.4 in the presence of jet mode feedback.

Jet feedback also has the most substantive effect on the CGM phases. It heats the halo gas, decreasing the relative fractions of ISM and cool and warm CGM gas, and decreasing the relative fraction in stars at the high mass end due to the overall decrease in star formation in the most massive galaxies. Adding X-ray AGN feedback to get the full SIMBA physics, we see only modest changes in the CGM mass budgets. The most dramatic appears to be the increase in hot gas fraction in the most massive halos, owing to X-ray feedback adding heat directly to surrounding gas if it is not star-forming; however, this may be subject to small number statistics in these most massive systems.

4.4.2 CGM metallicities

Next we present the effect of feedback on the metallicity of the halo gas components. Figure 4.9 shows the difference in median log metallicity between SIMBA and each of the AGN feedback variant simulations, as a function of halo
Figure 4.9: Difference in log metallicity between the fiducial Simba run and the variant without X-ray heating (yellow dashed line), without X-ray heating or AGN jets (red solid line), with stellar feedback only (magenta dotted line), and with no feedback prescription at all (dark blue dot-dashed line). Each panel refers to a different gas phase, as reported inside the plots. The error bars indicate the 25th and 75th percentiles of the data; for simplicity we show one representative error bar per line. Very low $\Delta \log Z$ in the warm gas component are plotted at -2. Points are slightly offset horizontally for easier viewing. Stellar feedback increases the metallicity of the warm and hot CGM components, while there is little impact on the cool CGM and ISM gas.

A value above (below) zero indicates that galaxies in the AGN variant simulation have higher (lower) metallicity than in the main Simba volume. We show one representative error bar per line; the error bars indicate the 25th and 75th percentiles of the data. As before we plot points only where there are >10 galaxies in a halo mass bin, and show results for the total galaxy population.

In the no feedback simulation metals are primarily locked into stars, resulting in a high stellar metallicity and low metallicity in the hot and warm phase CGM (lower than in the full Simba model by 1.5-0.5 dex, from the low $M_*$ end to high $M_*$ end.). Within the uncertainties, the metallicities of the cool CGM and the ISM are not affected by the AGN and SF feedback model.

Turning on SF feedback raises the metallicity of the warm and hot CGM, seen in the increase in $\Delta \log Z$ between the No-feedback and No-AGN models. This is especially true at the low mass end where SF feedback is most effective. Despite 70% of SN wind particles being ejected cold at $T \sim 10^3$K, the lack of a change in metallicity in the cool CGM phase indicates that these ejected metals do not remain cold. Instead the increase seen in the warm and hot components indicates that the metal-enriched gas is heated towards the halo virial temperature.
Radiative-mode AGN feedback has little impact on the hot CGM. However, introducing jet feedback raises this metallicity by 0.5 dex, possibly because metal-rich gas that was previously in the cool or warm phases is heated by the jets to the virial temperature.

### 4.4.3 CGM absorption

Finally, we examine the impact of feedback on CGM absorption, by repeating the COS-Halos and COS-Dwarfs comparison using the feedback variant simulations. We use the same galaxy selection procedure as section 4.3.1 to select suitable galaxies from the $50h^{-1}$Mpc SIMBA volume, in this case selecting 4 SIMBA galaxies per COS galaxy to account for reduced volume. We then identify matching halos from the AGN variant simulations to probe the effect of AGN feedback on specific halos, where matching halos are defined as having the most number of dark matter particles in common. This is possible because all the feedback variants are run from identical initial conditions, so the dark matter particles have the same IDs. We ignore some high mass galaxies from the COS-Halos sample that have no analogs in the small $50h^{-1}$Mpc SIMBA volume.

Figure 4.10 shows the total path absorption against $r_{200}$-scaled impact parameter in the COS-Halos and COS-Dwarfs SIMBA galaxy samples, using the full SIMBA model (solid), the No-Xray model (dashed), the No-jet model (dotted) and the No-feedback model (dot-dashed). We assume an FG20 ionising background. Light blue and light pink lines represent star forming and quenched galaxies; error bars show the cosmic variance uncertainties. The quenched galaxies in the No-jet and No-feedback models are not shown due to small number statistics, since jet feedback is the dominant quenching mechanism in SIMBA. We also choose not to show the No-AGN results as we find these absorption statistics are similar to that of the No-jet simulation, i.e. radiative-mode feedback has little effect on the CGM in absorption in our simulations.

In the absence of SF driven feedback to enrich the CGM, we see reduced CGM absorption for all ions, particularly for SiIII and OVI. This is not due to a change in the multiphase structure of the CGM, since Figure 4.8 shows that SF feedback does not significantly change the temperature of the CGM. Instead SF feedback enriches the CGM by spreading metals away from central galaxies across a range of ionisation potentials. Hence SF feedback is primarily responsible for enriching the CGM.
Figure 4.10: Total path absorption for different AGN models (solid: full SIMBA; dashed: No-Xray; dotted: No-jet; dot-dashed: No-feedback) for H\(_\text{i}\) and selected metal lines (as indicated in each panel), against \(r_{200}\)-scaled impact parameter in the COS-Halos and COS-Dwarfs SIMBA galaxy samples. We assume an FG20 ionising background. Light blue and light pink lines represent star forming and quenched galaxies, respectively; shaded areas show the cosmic variance uncertainties. Stellar feedback enriches the CGM leading to increased absorption, while jet-mode AGN feedback heats the CGM and leads to decreased absorption around L* galaxies.
In the $L^*$ galaxy COS-Halos sample, absorption around star forming galaxies increases in the absence of X-ray and jet feedback. Turning on jet feedback leads to overall heating of galaxy halos and the IGM (Christiansen et al., 2020), resulting in reduced absorption for all ion species considered here. The decrease is largest for low ions, which trace cool gas, and at the outskirts of the halos, where gas is additionally heated due to jet feedback from neighbouring galaxies. The decrease in absorption around $L^*$ galaxies from turning on X-ray feedback is due to the direct heating of non-ISM gas, which scales with X-ray flux. X-ray feedback has no effect on the sub-$L^*$ galaxies in the COS-Dwarfs sample, which typically do not have massive black holes with the $f_{\text{Edd}} < 0.2$ required for jet and X-ray feedback. Jet feedback only reduces absorption further away from these galaxies at $\lesssim r_{200}$, where the IGM is heated by jet feedback from nearby galaxies. These results show that SF feedback and jet-mode AGN feedback have the largest impact on CGM absorption properties, the former providing the metals, and the latter important for setting the gas phase. Thus, observational surveys such as COS-Halos and COS-Dwarfs may provide constraints on the amount of metals transported in stellar winds, and on AGN jet velocity.

4.5 Conclusions

We have examined the physical and observable properties of CGM gas around galaxies of a wide range in mass and specific SFR in the SIMBA cosmological hydrodynamic simulation at low redshifts. We examine halo baryon and metal contents as a function of mass, and conduct a detailed comparison to H\textsuperscript{i} and metal absorption line observations in the COS-Halos and COS-Dwarfs databases. Finally, we examine the impact of various star formation and AGN feedback modules in SIMBA on CGM properties. Our main results are as follows:

- Baryon fractions in halos across the entire mass range probed are well below the cosmic fraction. This indicates that missing halo baryons are truly redistributed beyond $r_{200}$, not just undetected. The CGM of $L^*$ galaxies has largely been evacuated, down to as little as $\sim 10\%$ of the global baryon fraction in CGM gas. Star-forming systems, even at the high-mass end, tend to have a CGM with a comparable mix of cool $\sim 10^4\text{K}$ gas and gas near the virial temperature, while for quenched systems, even at the low-mass end, the CGM gas is near the virial temperature. Thus the distinction between
galaxies having hot vs. multi-phase CGM is not so much of a division in mass but rather one in sSFR.

- For star-forming systems, metals are increasingly locked into galactic components (stars, dust, ISM) towards higher $M_*$, reflecting the increased efficiency of star-formation up to the highest masses. For the CGM, metals are distributed among the phases, with the hotter phases having a larger share at higher $M_*$. In quenched systems, in contrast, CGM metals are mostly hot, reflecting the overall mass phase budget, and there is no trend with mass except a mild increase at high masses where the star formation efficiency drops.

- CGM metallicities show strong trends with temperature, as the virial-temperature CGM gas has $\sim 0.5 - 1$ lower metallicity than cooler phases. The cool and warm CGM, in contrast, have metallicities close to the ISM, which itself generally tracks stellar metallicities. The lower hot phase metallicity likely reflects less metal enriched halo inflow that is virialised, mixing with metals deposited via outflows.

- Absorption line statistics around a matched sample of SIMBA galaxies are broadly in agreement with the COS observations. Equivalent width, covering fraction and total path absorption decrease with increasing $r_{200}$-scaled impact parameter for each ion species. CGM absorption is higher around star forming galaxies than around quenched galaxies; the difference is smaller around sub-$L^*$ galaxies, suggesting lower-mass galaxies are quenched via a different mechanism.

- SIMBA predictions of H I equivalent width and total path absorption around star forming galaxies are in good agreement with COS-Halos and COS-Dwarfs observations, while covering fractions are under-predicted around L$^*$ star forming galaxies and too extended around sub-L$^*$ galaxies. SIMBA under-predicts H I EW and $f_{cov}$ around quenched galaxies.

- Metal line absorption around L$^*$ galaxies is sensitive to choice of photoionising background. When assuming a FG20 background, SIMBA under-predicts EW, $f_{cov}$ and total path absorption of low ions Mg II and Si III. Assuming instead a HM01 brings the predictions into closer agreement with COS-Halos observations. Around quenched galaxies, SIMBA does not reproduce as strong a radial gradient of low ion absorption seen in COS-Halos.
• SIMBA reproduces observed O\textsc{vi} absorption separately in star-forming and quenched systems, when employing more recent ionising background determinations. Reproducing the observed dichotomy is thus an important success of SIMBA’s model for halo gas heating associated with galaxy quenching, and supports the scenario that O\textsc{vi} mostly arises from a smoother volume-filling warm phase (Ford et al., 2013) that gets ionised to higher states around quenched galaxies (Oppenheimer et al., 2016).

• Around sub-L\textsuperscript{*} galaxies, EW, \( f_{\text{cov}} \) and total path absorption of C\textsc{iv} around star forming galaxies are in good agreement with COS-Dwarfs observations in the central region, but SIMBA over-predicts in the outskirts. Around quenched galaxies, SIMBA over-predicts C\textsc{iv} at all impact parameters.

• Feedback strongly lowers the mass fraction in the CGM. With no explicit feedback, halos contain roughly their cosmic share of baryons. Star-formation feedback increasingly evacuates the CGM below L\textsuperscript{*}, while jet AGN feedback has a dramatic impact at \( \gtrsim L\textsuperscript{*} \). Radiative and X-ray AGN feedback have minimal effects. Thus in SIMBA, the same jets that enact galaxy quenching are largely responsible for setting the CGM contents around non-dwarf galaxies.

• Feedback has little impact on the metallicity of the ISM and cool CGM at a given \( M_\star \). For warmer phases, in contrast, star formation feedback has the greatest impact. This indicates that hot halo gas is primarily enriched via galactic winds, likely at earlier epochs (Oppenheimer et al., 2012; Ford et al., 2014). The AGN jets have a noticeable impact as well, mildly increasing the hot CGM metallicity.

• The above trends among feedback variants are reflected in CGM absorption, with the enrichment of the CGM by star formation driven feedback having the most dramatic impact at raising absorption in all ions including H\textsc{i}. AGN jet feedback sets the gas phase of the CGM and tends to counteract this to reduce absorption in the lower ions, but has little impact on O\textsc{vi}, and essentially no impact on the COS-Dwarfs results since jets are not active in such galaxies.

Our results highlight the complex interplay between galaxy formation, large-scale structure evolution, and feedback processes in setting the physical and observable properties of the CGM. SIMBA provides a self-consistent cosmological description.
of the CGM that is observationally concordant within current systematics at the \(\lesssim 2\sigma\) level for a given ionising background, while predicting interesting results such as a dramatic evacuation of CGM gas, a strongly lower metallicity in the hot CGM phase, and much different trends around star-forming and quenched galaxies regardless of mass, a trend also seen in IllustrisTNG (Nelson et al., 2018b).

Current CGM absorption measures at low-\(z\) do not tightly constrain models, owing not only to small numbers but also systematic uncertainties such as a significant dependence of metal absorption on the shape of the ionising background. In our case, we have only considered spatially-uniform backgrounds, but low ions arising close to galaxies could further be impacted by the escape of ionising photons from the host galaxy. On the modeling side, numerics remain an important systematic, as idealised CGM models suggest extremely small scale structures that are unresolvable in any cosmological setting, particularly related to how outflows interact with ambient gas (e.g. Fielding et al., 2020a); new subgrid prescriptions such as Physically Evolved Winds (PhEW; Huang et al., 2020) aim to better account for such effects. Simba’s resolution is hopelessly far from that required to directly model these small-scale interactions, but it remains to be seen how much such interactions drive bulk CGM properties. With current modeling capabilities, statistically comparing to absorption line surveys spanning a range of galaxy masses and environments as we have done in this work is only possible at cosmological resolutions. Clearly, much work remains to connect large-scale models such as Simba with small-scale work illuminating the detailed physics of the CGM in order to provide a more robust description of the CGM and its observable properties.

### 4.6 Physical impact parameter

Previous work using the COS-Halos and COS-Dwarfs data have typically displayed spatial absorption dependence in terms of impact parameter in physical kpc units (e.g. Werk et al., 2013; Tumlinson et al., 2013; Bordoloi et al., 2014; Ford et al., 2016). Here we provide the equivalent widths of Figure 4.5 as a function of physical impact parameter as these units may be more familiar to readers. Figure 4.11 shows median equivalent width of selected ions against impact parameter for our COS-Halos and COS-Dwarfs galaxy samples, using the FG20 (solid), HM12x2 (dashed) and HM01 (dotted) ionising backgrounds. As before Simba star forming...
and quenched galaxies are represented by light blue and light pink lines, and shaded regions indicate the cosmic variance within eight sub-quadrants over the simulation volume for the FG20 results. Dark blue and magenta points represent the actual COS-Dwarfs/COS-Halos data; the error bars are the 25th and 75th percentiles of the data. The varying bin widths are chosen such that each bin contains \( \sim 8 \) galaxies. The detection thresholds for each ion are indicated by horizontal dotted lines.

Comparison with Figure 4.5 demonstrates that aligning the galaxies on a common scale results in stronger radial trends, although the overall trends do not change significantly when using physical kpc units. The O\textsc{vi} absorption is in slightly better agreement with observations using the physical kpc units, although the difference is within a factor of 2. This illustrates the level of systematic uncertainty on comparing to observations using different radial measures.
4.7 Resolution and volume convergence

To assess the impact of numerical limitations on our results, we perform a convergence test using a higher resolution $25 h^{-1}\text{Mpc}$ volume with the full SIMBA physics and $512^3$ gas and dark matter particles. The higher resolution run has 8 times the mass resolution of the fiducial volume, however the smaller volume does not properly probe galaxies at the high $M_\ast$ end with the most massive black holes seen in the $100 h^{-1}\text{Mpc}$ box. Hence heating and preventative feedback due to AGN is weaker in smaller volume simulations (Christiansen et al., 2020). We quantify the volume convergence by examining the $50 h^{-1}\text{Mpc}$ volume with $512^3$ gas and dark matter particles (the same run as in section 4.4), and a second $25 h^{-1}\text{Mpc}$ volume with $256^3$ gas and dark matter particles. These runs have the same mass resolution as the fiducial volume.

We select a sample of galaxies matching the COS-Halos and COS-Dwarfs criteria using the procedure outlined in section 4.3.1. For the $50 h^{-1}\text{Mpc}/512^3$ volume, we use the same galaxy sample as in section 4.4. Due to the low volume of the $25 h^{-1}\text{Mpc}$ runs, there are insufficient galaxies in these simulations to match every COS survey galaxy, particularly at the high mass end. To account for this we select fewer analogous SIMBA galaxies (4 per COS galaxy) and omit COS galaxies from our sample that have insufficient analogous SIMBA galaxies. Additionally, we do not impose the isolation criteria (no other central galaxies with $1 h^{-1}\text{Mpc}$) for the $25 h^{-1}\text{Mpc}$ samples.

Figure 4.12 shows median equivalent width of selected ions against $r_{200}$-scaled impact parameter for the COS-Halos and COS-Dwarfs SIMBA galaxy samples using different resolution SIMBA runs: fiducial $100 h^{-1}\text{Mpc}$ volume (solid); $50 h^{-1}\text{Mpc}$ volume with $512^3$ particles (dot-dashed); $25 h^{-1}\text{Mpc}$ volume with $256^3$ particles (dotted); higher resolution $25 h^{-1}\text{Mpc}$ volume with $512^3$ particles (dashed). We assume an FG20 ionising background. Light blue lines represent star forming galaxies in the SIMBA sample; shaded regions show the cosmic variance uncertainties. Dark blue points represent star forming galaxies in the COS samples; vertical error bars represent the 25th and 75th percentiles of the data, while horizontal error bars indicate the width of the bins. Black horizontal dotted lines indicate the detection threshold of each line under the COS survey conditions.

Comparing the $100$ and $50 h^{-1}\text{Mpc}$ runs, the closeness of the results from the
Figure 4.12: Equivalent width of H\textsc{i} and selected metal lines (as indicated in each panel) against $r_{200}$-scaled impact parameter for the star forming COS-Halos and COS-Dwarfs galaxy samples (circles and triangles, respectively) in different resolution Simba runs: fiducial 100 $h^{-1}$Mpc volume (solid); 50 $h^{-1}$Mpc volume with 512$^3$ particles (dot-dashed); 25 $h^{-1}$Mpc volume with 256$^3$ particles (dashed); higher resolution 25 $h^{-1}$Mpc volume with 512$^3$ particles (dotted). Shaded regions show the cosmic variance uncertainties. Dark blue points represent star forming galaxies in the COS samples; vertical error bars represent the 25th and 75th percentiles of the data, while horizontal error bars indicate the width of the bins. Black horizontal dotted lines indicate the detection threshold of each line under the COS survey conditions. Both decreased volume and increased resolution separately lead to increased absorption in the CGM.
COS-Halos sample shows that these simulations are converged in terms of volume. Hence we are probing a large enough scale to capture the processes that suppress absorption in the CGM of L* galaxies. For the COS-Dwarfs sample there is some difference between the 100 and 50 $h^{-1}$Mpc runs; the lack of heating by the strongest AGN in the lower volume simulations leads to an increase in absorption around sub-L* star forming galaxies. Comparing the 50 and 25 $h^{-1}$Mpc runs, we see that absorption around star forming in the 25 $h^{-1}$Mpc volume with 256$^3$ particles is increased in the 25 $h^{-1}$Mpc run, for all ion species and in both COS-Halos and COS-Dwarfs samples. Hence a larger, more representative volume leads to less absorption, indicating a lack of volume convergence in the lower volume simulations.

Meanwhile, increasing the resolution from 256$^3$ to 512$^3$ particles leads to a further increase in absorption, indicating that our simulations are not ideally converged, although the increase is typically $\lesssim \times 2$ (at a given box size) which is comparable to other systematics in the comparison. In the COS-Halos sample, H\textsc{i} and O\textsc{vi} show a moderate absorption increase at high impact parameters due to resolution. However, at low impact parameters the increase due to resolution is small for galaxies in the COS-Halos sample, suggesting that these results are close to resolution convergence. Galaxies in the COS-Dwarfs sample see a particularly large increase in absorption across the entire range of impact parameters, perhaps since these low mass galaxies are more poorly resolved in the fiducial simulation. In summary, both reduced volume and increased resolution lead to significant changes in the absorption structure of the CGM, suggesting that both high resolution and sufficient volume to capture AGN heating sources are required to properly reproduce CGM statistics in simulations, which is numerically challenging.
Chapter 5

The Physical Nature of Circumgalactic Medium Absorbers in SIMBA

5.1 Introduction

The circumgalactic medium (CGM) is broadly defined as the reservoir of baryonic material surrounding a galaxy (see Tumlinson et al., 2017, for a review). It holds material that serves as fuel for future star and black hole growth, as well as the by-products of feedback processes from star formation and active galactic nuclei (AGN) (e.g. Hafen et al., 2019). A significant fraction of the metal budgets of \( L^\star \) galaxies lies in the CGM (Peeples et al., 2014). Additionally, the loss of gas from the CGM via AGN-driven ‘preventative’ feedback has been linked to galaxy quenching and morphological transformation (Davies et al., 2020; Oppenheimer et al., 2020a). The CGM is regarded as a key for understanding the cycle of baryons between galaxies and their surrounding environments (Hafen et al., 2020) that is responsible for setting the physical properties of galaxies and their evolution over time (see Péroux & Howk, 2020, for a review).

The CGM can be probed via its absorption or emission properties. Owing to its diffuse and multi-phase nature, the CGM is typically very faint in emission (Putman et al., 2012) except in the most massive halos where hot X-ray emitting gas is present (Werner & Mernier, 2020). For more typical halos, absorption...
lines present the best approach to measuring CGM gas properties as they are sensitive to lower column densities. Absorption line studies require finding a coincident bright source (usually a quasar) illuminating a foreground CGM (e.g. Tumlinson et al., 2013; Bordoloi et al., 2014; Borthakur et al., 2015; Berg et al., 2018), or alternatively quantifying the galaxy population along a given line of sight (LOS), and then measuring the desired atomic transitions at the redshift of the intervening CGM (e.g. Rudie et al., 2012). While this technique can probe fairly diffuse gas in multiple elements and ionic states, it is limited by having typically a single or small number of LOS probing a given CGM, and only in select transitions that are strong and in the observable wavelength window. Since many of the strongest CGM transitions are in the ultraviolet (UV), at redshifts $z \lesssim 1$ this typically requires space-based UV observations, adding to the challenge.

Hydrodynamic simulations provide a holistic approach to model the CGM of galaxies and help interpret absorption line observations. By extracting spectra through the simulation volume, it is possible to compile statistics of simulated absorption lines and investigate the origin and underlying physical conditions of CGM absorbers. Hani et al. (2019) employed this approach using the Auriga simulations (Grand et al., 2017) to highlight the wide diversity of physical conditions in the CGM of L* galaxies. Such studies have revealed that galactic outflows are responsible for supplying the CGM with the enriched gas that is observed via metal absorption lines (Oppenheimer et al., 2012; Ford et al., 2013, 2014; Hummels et al., 2013), which is reflected in an angular dependence of CGM metallicity (Péroux & Howk, 2020). Absorption from low metal ions arises from recently enriched, dense regions close to galaxies that will be re-accreted onto galaxies within a few Gyr; in contrast, high metal ions trace more diffuse gas that was enriched via outflows many Gyr ago (Oppenheimer & Davé, 2009; Ford et al., 2013, 2014; Oppenheimer et al., 2012). The neutral hydrogen (H1) absorption of moderate column density arises from gas that is metal poor, whereas strong H1 absorbers trace cool, enriched gas that is typically inflowing (Ford et al., 2014). Oppenheimer et al. (2018b) compiled this picture into a two-phase model of the CGM around L* galaxies in which halo gas within $0.5r_{200}$ is composed of cool, low metal line clouds within a hot ambient medium traced by higher metal lines, while halo gas outwith $0.5r_{200}$ is heated to the virial temperature. Such cool gas clouds are thought to cool from initial local overdensities which trigger thermal instabilities (Nelson et al., 2020).

Careful comparisons between full cosmological galaxy evolution simulations and
observations can provide constraints on models and highlight discrepancies between model predictions and the real universe. Early cosmological simulations used statistics of Ly-α absorption in the intergalactic medium (IGM) to constrain cosmologies (e.g. Davé et al., 1999). More recent numerical studies have used contemporary IGM/CGM absorption observations as benchmarks to test the success of their simulations (e.g. Davé et al., 2010; Oppenheimer et al., 2012; Péroux & Howk, 2020) and investigate the impact of physical processes such as turbulence (Oppenheimer & Davé, 2009) and stellar feedback Hummels et al. (2013). Gutcke et al. (2017) presented a comparison of the NIHAO simulation suite (Wang et al., 2015) to H I and O VI absorption statistics, finding good agreement with observations. Ford et al. (2016) tested models for stellar winds by comparing their simulations to the COS-Halos survey of L* galaxies (Tumlinson et al., 2013; Werk et al., 2014), finding that strong winds are necessary to reproduce CGM metal absorption. The EAGLE simulations (Crain et al., 2015; Schaye et al., 2015) have also been compared with COS-Halos (Oppenheimer et al., 2016, 2018b), showing that EAGLE reproduces the correlation between O VI column density and star formation rate and is in good agreement with the low metal line statistics. The IllustrisTNG simulations (Pillepich et al., 2018) have likewise been shown to reproduce the observed dichotomy in O VI absorption between star forming and quenched galaxies seen in COS-Halos (Nelson et al., 2018b), as well as reasonable Mg II absorber populations (DeFelippis et al., 2021).

In Chapter 4 we explored the CGM as seen in absorption lines in the Simba simulation (Davé et al., 2019). In particular, we tailored the extracted spectra to mimic the COS-Halos (Tumlinson et al., 2013; Werk et al., 2014) and COS-Dwarfs (Bordoloi et al., 2014) surveys, spanning three orders of magnitude in galaxy and halo mass. We showed that the absorption line properties of H I and selected metal ions (Mg II, Si III, C IV, O VI) around star forming galaxies are in good agreement with the observations, depending on the assumed photoionising background. In particular, Simba reproduces H I absorption very well around star forming galaxies. We also examined the mass and metal budgets of halos in our simulations, and found that halo baryon fractions are generally less than half of the cosmic fraction, due to stellar feedback at low stellar masses and jet-mode AGN feedback at high stellar masses. Around quenched galaxies, the CGM baryons that remain are dominated by hot gas with \( T > 0.5T_{\text{vir}} \), while the CGM of star-forming galaxies is more multi-phase.

Cosmological simulations enable the investigation of the CGM around a wide
range of galaxy types. However, they lack the resolution that modern zoom re-
simulations can achieve, specifically in the diffuse CGM that is coarsely sampled
by Lagrangian fluid elements. Recently, several groups have presented “CGM
zoom” simulations, where the resolving power is focused within the CGM rather
than the galaxy. Such simulations have found that resolution is an important
consideration, although their galaxy sample sizes are low due to the computational
cost. For example, Hummels et al. (2019) demonstrated that increasing resolution
in the halo leads to diminished warm/hot gas content and enhanced cool gas
content, where the cool structures are able to survive for longer timescales
and at smaller masses. This results in an increase in absorption of low ions
such as H$_1$, and a decrease in absorption of high ions such as O$_{vi}$. Similarly,
Peeples et al. (2019) and van de Voort et al. (2019) both find in their CGM
zoom simulations that although the average CGM properties are unaffected, the
distribution of physical conditions in the CGM are sensitive to resolution, which
has consequences for H$_1$ and metal absorption. van de Voort et al. (2019) find
that this leads to enhanced H$_1$ column densities in the outer CGM. Moreover,
the typical CGM structures in simulations with enhanced resolution have lower
masses down to $< 10^4$ M$_\odot$, which is far below SIMBA’s resolution limit. In
the IllustrisTNG cosmological simulations, increased resolution has also been
shown to lead to an increase in the formation of cold, small-scale CGM clouds
(Nelson et al., 2020). Numerical resolution clearly makes a difference for e.g.
very high column density H$_1$ absorbers. However, Suresh et al. (2019) found
using IllustrisTNG that metal absorption in the CGM was relatively insensitive
to resolution. Although the precise, quantitative impact of resolution on the
current work is unclear, it is nonetheless an important consideration.

SIMBA’s resolution is even lower than IllustrisTNG, which is a limitation of the
current work. On the flip side, SIMBA provides large galaxy sample with a unique
galaxy formation model that reproduces a wide variety of galaxy observations at
low redshifts (Chapter 3 of this thesis; also Davé et al., 2019; Li et al., 2019;
Thomas et al., 2019; Davé et al., 2020; Thomas et al., 2021; Glowacki et al.,
2020; Cui et al., 2021) and low-redshift Ly$\alpha$ statistics along random lines of
sight (Christiansen et al., 2020), which make it valuable to examine despite its
modest resolution. In Chapter 4 we showed that increased resolution in SIMBA
leads to enhanced absorption around star forming galaxies, for all ions probed
across a range of ionisation energies. Hence it is likely that the absorber sample
presented in this work represent minimum column densities and absorber counts.
However, the same numerical tests also showed that simulation volume has an
effect: the smaller boxes contain fewer of the most massive galaxies, leading to weaker preventative AGN feedback and enhanced CGM absorption. As such, the cosmological volumes offered by simulations such as Simba provide an important large-scale context for studying the CGM.

In this Chapter, we follow on our previous work to investigate the underlying physical nature of the population of CGM absorbers within Simba. We construct a new carefully-selected sample of synthetic spectra for a range of key ions as a function of both galaxy stellar mass and specific star formation rate (sSFR). In particular, we present counting statistics of the CGM absorber population and study properties of the absorbing CGM gas by examining its distribution within cosmic physical phase space. We study the relationship between physical density and column density of CGM absorbers, and investigate the contribution of collisional ionisation to the absorber statistics. Lastly, we quantify the contribution to the LOS spectra from satellite galaxies living in the halos of other galaxies.

This chapter is organised as follows. In §5.2 we describe the galaxy selection and the spectrum generation and fitting processes. In §5.3 we present the CGM counting statistics and physical underlying conditions. In §5.4 we discuss the redshift dependence of our results. Finally in §5.5 we present the contributions of satellite galaxies to the original absorber sample.

5.2 Absorber sample

5.2.1 Galaxy selection

A representative sample of galaxies is obtained across a range of physical properties within each of the $z = 0$, 0.25, 0.5 and 1 snapshots, by defining regions in $M_\star$-sSFR space and selecting galaxies within each region. We define three categories of galaxies within each mass bin: 1) star-forming galaxies, with $\log(sSFR/\text{Gyr}^{-1}) > -1.8 + 0.3z$ for consistency with previous work with the Simba simulation (e.g. Thomas et al., 2019); 2) green valley galaxies, with sSFR within 1 dex below the star-forming galaxy threshold; 3) quenched galaxies, with SFR $= 0 \text{ M}_\odot \text{yr}^{-1}$ to ensure that these galaxies are fully quenched. We impose a lower stellar mass limit of $M_\star > 10^{10} \text{ M}_\odot$ and require that the galaxies are the central galaxy in their halos. We select 12 galaxies in each $M_\star$-SFR region (apart from in
the highest mass green valley region, in which there are only 8 galaxies matching these criteria at $z = 0$). Subsamples of 12 were chosen to balance a reasonable sample size with the computational cost of generating spectra and the size of the smallest $M_\ast$-SFR bin. All other bins contain more than 12 galaxies; in these bins we randomly select galaxies using Numpy’s `random.choice` function with a seed of 1 and assuming a uniform distribution. We select a constant number of galaxies per bin to ensure that any surplus or deficit of CGM absorbers around specific types of galaxies is not due to the over-representation of such galaxies in our initial sample. We have verified that the galaxies in our sample are randomly orientated with respect to the line of sight and thus our results are not biased due to galaxy inclination.

Figure 5.1 shows sSFR against $M_\ast$ for our $z = 0$ galaxy sample. Horizontal lines of constant sSFR show our boundaries between star-forming, green valley and quenched galaxies. The vertical lines indicate the boundaries between our stellar mass bins. The color scale shows the mass-weighted average temperature of gas elements in each galaxy’s host halo. We show how this sample relates to the overall SIMBA population with a 2D log-scale histogram of SIMBA galaxies in greyscale. Galaxies with zero star formation are plotted at sSFR = $10^{-3.8}$ Gyr. A comparison with the observed star-forming main sequence of galaxies (sSFR > $10^{-1.8}$ Gyr) from the GALEX-SDSS-WISE Legacy Catalog (GSWLC, Salim et al., 2016, 2018) is shown as a running median in black squares. The error bars are the standard deviation within each bin. The overall population of star forming SIMBA galaxies at low redshift reproduces the observed main sequence well (Davé et al., 2019).

### 5.2.2 Spectral generation

For each galaxy in our sample, we probe its CGM by generating mock spectra for multiple lines of sight (LOS) through the simulation box. We select impact parameters ($r_\perp$) for each galaxy in multiples of the halo $r_{200}$, in order to probe the impact of radial distance from the galaxy whilst placing galaxies of varying masses on a common scale. For each impact parameter, we select 8 equally-spaced LOS in a circle around the galaxy with radius $r_\perp$, starting from $(x_{\text{gal}} + r_\perp, y_{\text{gal}})$ and proceeding anti-clockwise. This results in 8480 LOS spectra per line transition. Absorption spectra are generated for a selection of commonly observed ions spanning a range of excitation energies: H\textsc{i} 1215Å, Mg\textsc{ii} 2796Å, C\textsc{ii} 1334Å, Si\textsc{iii} 1206Å, C\textsc{iv} 1548Å and O\textsc{vi} 1031Å.
Figure 5.1: Specific star formation rate (sSFR) against stellar mass ($M_\star$) for our $z = 0$ galaxy sample, colour-coded by the mass-weighted average halo temperature. Horizontal lines of constant sSFR indicate the partition between star-forming, green valley and quenched galaxies. Vertical lines show the boundaries between $M_\star$ bins. The observed star-forming main sequence from GSWLC is shown as black squares. The SIMBA galaxy population as a whole are shown as a log-scale 2D histogram in greyscale.
Mock spectra are generated following the same procedure as in Chapter 4. The python analysis package PyGAD (Röttgers et al., 2020) is used to generate absorption spectra through the simulation volume in the z-axis direction by identifying the gas elements whose smoothing lengths intersect with the given LOS. For each gas element, the ionisation fractions are found using look up tables generated with version 17.01 of the CLOUDY cloud simulation code (Ferland et al., 2017) using Cloudy Cooling Tools\(^1\). Computing ionisation fractions with CLOUDY requires an assumption for the incident photoionising background spectrum; indeed in Chapter 4 we demonstrated that the choice of photoionising background has a significant impact on the metal line absorption statistics. For consistency with earlier work on the CGM and IGM in SIMBA (Chapter 4 of this thesis; also Christiansen et al., 2020; Bradley et al., 2022), we assume a Faucher-Giguère (2020) photoionising UV background. For neutral H\(_1\), self-shielding is already applied directly during the simulation run; for metal lines, self-shielding is applied following the Rahmati et al. (2013) prescription, whereby an overall attenuation factor is applied to the entire ionising background spectrum in the optically thin limit.

The ion densities for each gas element are found by multiplying the gas densities by the ionisation fractions for each species, assuming a \(\sim 76\%\) mass fraction for hydrogen and the individually-tracked mass fractions within SIMBA for the metals. The ion densities are smoothed along the LOS into pixels using the gas elements’ individual smoothing lengths and metal masses (for metal lines), and using the same spline kernel used in the Gizmo simulation code. The spectral pixel scale is \(2.5\text{ km s}^{-1}\). The optical depths for each pixel are computed from the resulting ion column densities using the oscillator strength for each species. Column density-weighted physical density, temperature, metallicity, and peculiar velocity are also computed in the same manner within the LOS pixels. For further details on PyGAD, see Röttgers et al. (2020). We exclude wind particles from the spectral generation since these gas elements are hydrodynamically decoupled from the surrounding gas; these represent a very small fraction of CGM mass in the halos of the galaxies we consider (see §4.2.2).

We apply the following effects to the resulting synthetic spectra to mimic observations gathered with Hubble Space Telescope’s Space Telescope Imaging Spectrograph:

\(^1\)https://github.com/brittonsmith/cloudy_cooling_tools
1. The spectra are convolved with the STIS E140M line spread function (LSF), which is the medium resolution Multi-Anode Microchannel Array (MAMA) Echelle spectrograph LSF, centered on 140nm. The typical FWHM is around 6km s\(^{-1}\), while the pixel scale is typically 2.5km s\(^{-1}\). For Mg\(\text{II}\), we convolve with a Gaussian with FWHM of 6km s\(^{-1}\) to mimic Keck spectra.

2. Gaussian random noise is added to the spectra with a signal to noise ratio (SNR) of 30.

3. A mock continuum fitting procedure is applied to the spectra as in Chapter 4 (iteratively removing pixels that are \(> 2\sigma\) below a polynomial best-fit to the spectrum until the fractional change between continuum fits is less than \(10^{-4}\)).

While our goal is not to compare with observations, this choice allows us to examine absorber properties at the highest resolution among currently available UV spectrographs capable of obtaining significant samples of CGM absorbers. The resolution of the COS spectrograph, in contrast, is often insufficient to resolve metal lines arising in cooler gas, which would make Voigt profile fitting parameters (described in the next section) less robust.

Figure 5.2 shows example synthetic CGM absorption spectra from our sample (grey lines). The central galaxy is a star forming \(M_*=10^{10.8}\, M_\odot\), SFR = 14.2 \(M_\odot\) yr\(^{-1}\) galaxy at \(z=0\). Lower mass, star forming galaxies such as this are over-represented in our absorber sample since their CGM contain an over-abundance of absorbers compared with the population as a whole (§5.3.1). The impact parameter of the line of sight is \(r_\perp=0.25r_{200}=155\, h^{-1}\)kpc and the spectrum is centered on the galaxy systematic velocity. Each panel shows the spectrum for a different ion. The shapes of these spectra are representative of our spectrum sample: the H\(\text{I}\) spectrum shows strong, saturated absorption with broad wings, while the metal line spectra also show strong absorption, with Mg\(\text{II}\), C\(\text{II}\) and Si\(\text{III}\) also saturating in places. The Mg\(\text{II}\) spectrum is composed of many narrow lines, which we find is typical of Mg\(\text{II}\) absorption in Simba. The H\(\text{I}\) and metal absorption lines appear aligned with one another in velocity space, despite the range of physical conditions probed with these lines.
Figure 5.2: Example synthetic spectra (grey lines) of the CGM of a star forming galaxy from our $z = 0$ sample with $M_* = 10^{10.8} \, M_\odot$ and SFR = 14.2 $M_\odot$ yr$^{-1}$. The line of sight is $r_\perp = 0.25r_{200} = 155 \, h^{-1}$kpc from the galaxy. The spectra are centered on the galaxy’s systematic velocity. The detected absorption regions are shaded in grey and numbered at the top of each plot. The solid pink lines show the overall fitted model; where models consist of multiple lines, these are shown as dashed pink lines. The $\chi_r^2$ for each region is given in the legend on the lower left of each panel.
5.2.3 Spectrum fitting

We extract observables from the spectra within a ±600 km s$^{-1}$ window centered on each galaxy. Fitting Voigt profiles to the spectra provides the column density $N$, the Doppler $b$ parameter, the wavelength (or velocity) location along the LOS, and equivalent width (EW) for each line in the absorption system. We also compute the EW directly from the flux for comparison.

The Voigt profile fitting procedure broadly follows that of AutoVP (Davé et al., 1997), but reformulated into Python and integrated into PYGAD. To simplify the fitting procedure, we first identify absorption regions by computing the detection significance ratio of each pixel, which is defined as the Gaussian-smoothed flux equivalent width divided by the Gaussian-smoothed noise equivalent width. Absorption regions are identified where the flux drops below the continuum level and the ratio is $> 5\sigma$. We ensure that the region edges begin at continuum level, and merge together nearby regions within 2 pixels of one another. We impose a region length limit of $> 2$ pixels to avoid spurious absorbers. Table 5.1 gives the number of LOS (out of the 8480 LOS per species) that have no detected absorption regions.

Each absorption region is fitted individually. In the first instance a line is added at the position of lowest flux, and an initial estimate for $N$ and $b$ is made based on the depth and velocity width of the local flux minimum, respectively. For saturated lines (those with a residual flux below the noise level), a line is placed in the middle of the saturated trough and the initial estimate for $N$ and $b$ is made by finding the values that give the lowest reduced $\chi^2$ in a coarse grid of log$N$-log$b$ space. The best-fit Voigt profile solution is then found using the `minimize` routine from the `scipy.optimize` subpackage\(^2\) to search parameter space for the line parameters which minimise $\chi^2$. We supply prior bounds of $12 < \log(N/\text{cm}^{-2}) < 19$ for H$\text{i}$ and $11 < \log(N/\text{cm}^{-2}) < 17$ for metal lines. Bounds on $b$ are computed based on the thermal line width at $10^4\text{K}$ and $10^7\text{K}$. To avoid overfitting that occurs with steeply saturated absorbers, the $\chi^2$ is asymmetrically penalised further where the model is $> 3\sigma$ below the data.

If $\chi^2 < 2.5$ is reached then the solution is accepted, as we find that this gives a reasonable visual fit to the data in most cases. Otherwise, the model is subtracted from the data to find a residual flux and a line is placed at the residual minimum.

\(^2\)https://docs.scipy.org/doc/scipy/reference/optimize.html
and the process is repeated until an acceptable fit is achieved, up to a maximum of 10 lines. New lines must improve the $\chi^2_r$ by at least 5%, and if the model does not improve sufficiently after 2 additional lines then the process is halted, and the additional lines are ignored. If after 10 lines an acceptable solution is not found then we revert to the number of lines that performed best. Our routine then attempts to find the solution with the fewest lines by iteratively re-fitting with each line removed one at a time. If the $\chi^2_r$ acceptance threshold is reached, or the $\chi^2_r$ increases by less than 5% then the line is removed from the solution. Formal errors in $N$ and $b$ are found by computing the Hessian covariance matrix. In cases where an acceptable fit is not found, the routine flags these for possible manual fitting, although for this work we do not revisit these fits. Table 5.1 gives the number and fraction of fitted LOS that are flagged as having at least one absorption region with an unacceptable fit ($\chi^2_r > 50$). We verify that the EWs resulting from these fits closely correspond to the EWs computed directly from the spectra, broadly following a 1:1 relation.

Although we adopt a $\chi^2_r < 2.5$ acceptance limit during the fitting process, the solution for a LOS does not always reach this threshold. To cope with such cases, we adopt $\chi^2_r$ acceptance thresholds for our absorber sample such that we recover 90% of the total EW across all LOS for each species. While the Voigt profile fitter occasionally overestimates the EW for individual lines of sight, we have checked that the total EW computed directly by summing all lines of sight is always higher than the total EW obtained by fitting the spectra. The upper limits ($\chi^2_{r,90}$) are given in Table 5.1, along with the sample size and $N$ completeness limit for each species. In practice the quality of the fit is much better than these limits in most cases; the median $\chi^2_r$ of absorption lines in our sample is also given in Table 5.1. We are reluctant to adopt a stringent fixed $\chi^2_r$ for all species as this leads to an incomplete sample of absorbers. We show a comparison of individual LOS EW and $N$ obtained from our Voigt profile fitting routine to those obtained directly from the spectra in §5.7.

Figure 5.2 shows the Voigt profile fits for the example spectrum. The detected regions of absorption are shaded in grey and numbered along the top of each panel. The individual fitted absorber components are plotted as dashed pink lines, with the overall best-fit model plotted in solid pink. The $\chi^2_r$ for each absorption region is displayed in the lower left of each panel. Overall, the quality of the Voigt profile fits to the data are good considering the fits are produced automatically with no manual intervention. Inspecting the fits by eye, the models clearly capture the
overall shape of the original spectra.

The Mg\textsc{ii} spectrum contains an example of a poor Voigt profile fit (absorption region 2 in the Mg\textsc{ii} panel, with $\chi_r^2 = 46.22$). In this case the fitting routine has not succeeded in finding an accurate solution. However, by eye the fit is not completely unreasonable: the fit has two strong absorption features at the position of the two main absorption features in the data, but the fit has not captured the fine detail of the many small absorption lines within these features. This is typical of the high $\chi_r^2$ Mg\textsc{ii} and Si\textsc{iii} lines in our sample.

### 5.3 CGM absorbers

#### 5.3.1 Column density distribution functions

We begin our analysis with counting statistics of the CGM absorber sample using the column density distribution function (CDDF), which gives the column density probability distribution normalised by the comoving path length along the LOS. The CDDF is defined as:

$$ f(N, z) = \frac{\partial^2 n}{\partial N \partial X}, \quad (5.1) $$

where $n$ is the number of absorbers in each bin of log$N$, and $\partial X$ is the comoving path length, designed to account for cosmological evolution (e.g. Wijers et al., 2019):

$$ \partial X = dz \frac{H_0}{H(z)} (1 + z)^2, \quad (5.2) $$

where $dz = 600 \text{ km s}^{-1} \times N_{\text{spec}} / c$ is the redshift path length (with $N_{\text{spec}}$ being the number of spectra generated) and $H(z)$ is the Hubble constant at $z$.

Figure 5.3 shows Simba’s predictions for the CDDF in the CGM for each species. Grey lines show the CDDF for all galaxies, while the blue, green, and red lines show the galaxies separated into star forming, green valley, and quenched sub-samples, respectively. The smaller panels show the CDDF fraction with respect to the case for all galaxies and at all impact parameters. Horizontal error bars show the width of the column density bins, where we have increased the bin width at the high column density end to account for low numbers. Vertical error bars
Table 5.1: Absorber sample properties: \( E \), the excitation energy of the species; \( n_{\text{LOS no-abs}} \), the number of LOS with no detected absorption regions; \( n_{\text{reject}} \) and \( f_{\text{reject}} \), the number and fraction of fitted LOS flagged as having at least one unacceptable absorption region; \( n_{\text{absorbers}} \), the resulting number of fitted absorbers below the \( \chi_r^2 \) limit; \( \log(N_{\text{min}}/\text{cm}^{-2}) \), the column density completeness limit; \( \chi_{r}^{2,90} \), the \( \chi_r^2 \) below which we recover 90% of the total EW; Median \( \chi_r^2 \), the median \( \chi_r^2 \) of all absorbers.
Figure 5.3: The $z = 0$ CDDF for each ion considered in this work. Colours represent galaxy types: all galaxies (grey), star forming (blue), green valley (green) and quenched (red). The smaller panels show the fraction with respect to the CDDF for all galaxies and at all impact parameters. Horizontal error bars show the width of the column density bins. Vertical error bars on show representative cosmic variance uncertainties, computed using all absorbers. Vertical lines indicate the completeness limit in each case.
show representative cosmic variance uncertainties, combined with the Poisson error within each bin, computed using all galaxies and all LOS. Error bars in the smaller panels are slightly offset for easier viewing. To determine the cosmic variance error, the $x$ and $y$ sides of the simulation box are each split into quarters, forming 16 equal-sized cells across the face of the simulation. The absorbers are binned into these cells based on the $x-y$ positions of their sightlines. The cosmic variance uncertainties are computed by jackknife resampling: we iterate over the cells and omit the absorbers from each cell in turn, and compute the CDDF from the remaining absorbers. The cosmic variance error is the variance of the 16 estimates of the CDDF. The uncertainties in the smaller panels are those of from all galaxies, added in quadrature with the uncertainties from star-forming, green valley and quenched galaxies. Where the uncertainties are large (e.g. at the high column density end for C$\text{\textsc{ii}}$ and Si$\text{\textsc{iii}}$), this arises from low numbers of such absorbers.

The overall shape of the CDDF in each case is that of a power law at high column densities with a turnover at the low column end around $\log(N/cm^{-2}) > 12 - 13$ (the exact turnover column density varies for each ion). The turnover may be due to incompleteness in our sample of low column density absorbers, or a genuine physical decrease in absorbers of this strength. We have conducted completeness tests on our sample by re-generating our LOS sample using an SNR of 100 to see whether this uncovers low column density absorbers that otherwise would be below the noise limit. We see little difference in the CDDFs from this high SNR sample, suggesting that such low column density absorbers genuinely do not exist in our simulation, likely due to numerical resolution. We compute lower completeness limits on our sample by fitting the power law portion of each CDDF and identifying where the CDDF falls below 50% of the expectation at low column densities. The completeness limits for each species are shown as dashed vertical lines in Figure 5.3 and given in Table 5.1. For the remainder of this work we will only discuss results for absorbers above these limits.

Considering the CDDFs for absorbers from all galaxies in our sample, we find abundant H$\text{\textsc{i}}$ absorption in our CGM sample, even out to high column densities, resulting in a shallow power law slope for H$\text{\textsc{i}}$. Mg$\text{\textsc{ii}}$ absorbers are the least frequently occuring ion in our sample, and typically have low column density. Hence the CDDF for Mg$\text{\textsc{ii}}$ barely probes the low column density turnover. Our absorber sample does not contain any O$\text{\textsc{vi}}$ absorbers outside of a narrow range of column densities ($12 < \log(N/cm^{-2}) < 15$), resulting in a CDDF which
probes only the turnover and a limited portion of the power law. This reflects our spectral generation approach, which provides limited sampling at very high column densities. In contrast, Bradley et al. (2022) also used SIMBA to examine global O\textsc{iv} statistics along random lines of sight using projected column density maps through the simulation volume, which allowed many more lines of sight thus enabling the O\textsc{iv} CDDF to be measured to higher column densities; they showed that the resulting CDDF is in very good agreement with observations at $z \lesssim 0.7$.

Splitting the absorbers by the star formation activity of the central galaxy, we can begin to see the effect of star formation on the CGM. For each ion, absorbers from around star forming galaxies have higher number counts at all column densities. Thus star forming galaxies contribute more to the overall CDDF. For Si\textsc{iii} and C\textsc{iv}, the high column density absorbers only occur in the CGM of star forming galaxies. The difference plots show the relative contribution of each galaxy subsample to the total CDDF; for each species, the difference plots show a positive (negative) slope for star forming (green valley and quenched) galaxies.

Examining the absorber statistics from the green valley sample, it is clear that the population of CGM absorbers around green valley galaxies more closely resembles that of quenched galaxies, rather than the star forming galaxies. This is particularly true for low ions, which (as we will later show) trace cooler, denser gas. Although these green valley galaxies have some residual star formation, their CGM gas has already dropped out of the necessary conditions for abundant and high column density absorption. This suggests that the CGM of green valley galaxies ‘quenches’ (that is, resembles that of a quenched galaxy) before star formation is fully halted in the central galaxy, at least in terms of their cool, dense gas content.

Having examined the dependence on star formation activity, we now investigate the stellar mass dependence of SIMBA’s absorber counting statistics. Figure 5.4 again shows SIMBA’s predictions for the CDDFs in the CGM; in this case the grey lines still represent the CDDF for all galaxies, while the blue, purple, and orange lines show the galaxies separated into broad stellar mass bins. As before, the smaller panels show the CDDF fraction with respect to the case for all galaxies at all impact parameters. The errorbars are computed in the same manner as Figure 5.3. The vertical lines in each panel indicate the completeness limits.

In general, there is little difference in the CDDF from galaxies in our low and intermediate mass bins, and these CDDFs essentially follow the shape and
Figure 5.4: As in Figure 5.3, the z = 0 CDDF for each ion considered in this work. In this case the colours represent broad stellar mass bins. The smaller panels show the fraction with respect to the CDDF for all galaxies and at all impact parameters. Horizontal error bars show the width of the column density bins. Vertical error bars on show representative cosmic variance uncertainties, computed using all absorbers. Vertical lines indicate the completeness limit in each case.
normalisation of the CDDF for the full galaxy sample. Where the CDDF from these stellar mass bins is substantially higher than the CDDF for all galaxies, this generally occurs below the completeness limit. Thus, the counting statistics of galaxies with $10 < \log(M_\star / M_\odot) < 11$ are largely independence of stellar mass. In Chapter 4 we examined the mass and metal budgets of halos in SIMBA, and found that (when combining systems regardless of star formation activity) galaxies in this mass range have a multiphase CGM with roughly equal proportions of metal mass in cool, warm and hot CGM, as well as 10-20% of metal mass gas in the ISM. Hence these galaxies have an abundance of metal enriched gas in their halos that is capable of producing measurable absorption.

However, galaxies in the highest mass bin ($11 < \log(M_\star / M_\odot) < 11.5$) have significantly fewer absorbers at all column densities. The CDDF fraction from high mass galaxies is roughly constant with column density (above the completeness limit). The CDDF of HI absorbers is at the level of $\sim 70\%$ of the overall CDDF; for metal absorption it is slightly lower at $\sim 50\%$. In this stellar mass range, halos in SIMBA become less multiphase in nature and more homogeneous, with a growing proportion of mass and metals in the hot phase of the CGM, largely due to stronger AGN feedback at the high mass end (see §4.4). The lower CDDF of high mass galaxies compared with the overall CDDF demonstrates that this leads to a reduction in number of these CGM absorbers, as CGM gas is heated into higher ionisation states such as O vii that are not traced in the ultraviolet (see Bradley et al., 2022).

In summary, HI is the most abundant species in the CGM of SIMBA galaxies, with a large absorber population out to the high column density end. Star formation activity and stellar mass of the central galaxy both play a role in the absorber statistics of the CGM. The CDDF for each species is lower for quenched galaxies than star-forming galaxies, while the CDDFs for green valley galaxies is more similar to quenched than star-forming galaxies, suggesting that the effects of quenching are apparent in the CGM before star formation fully halts in the central galaxy. High mass galaxies also have lower absorber CDDFs due to the growing proportion of the CGM that is made up of hot gas. Numerical convergence tests conducted in the previous Chapter indicate that increased particle resolution results in a moderate increase in equivalent width, particularly for HI and O vi absorption in the outskirts of halos, suggesting that these CDDFs (produced from the fiducial SIMBA volume) represent minimum CGM absorber counts, particularly at the low column density end (§4.7). In addition, as can be seen in
Figure 5.14 of §5.7, the Voigt profile fitting routine can overestimate H\textsc{i} column densities in Lyman Limit Systems due to the saturated nature of the absorption. As such, the CDDFs presented here for H\textsc{i} are likely to be overestimated at the high column density end.

### 5.3.2 Physical conditions of absorber sample

We now turn to the underlying physical conditions that give rise to H\textsc{i} and metal absorption in the CGM by examining the distribution of CGM absorbers in overdensity-temperature phase space. We obtain underlying physical conditions for each absorber by selecting column density-weighted properties (computed during the spectral generation process, see §5.2.2) at the pixel nearest to the absorber’s fitted LOS wavelength. Thus the results presented here represent the gas conditions that are contributing most to the absorption. Figure 5.5 shows histograms of overdensity (upper panels) and temperature (lower panels) of the absorbers for the six ions we consider. Overdensity is defined as $\delta = \rho_m / \bar{\rho}_m$, where $\rho_m$ is the matter density, and $\bar{\rho}_m$ is the mean matter density of the universe. Each histogram is normalised such that the area sums to unity. Overdensity histograms are split into absorbers from star forming (blue), green valley (green) and quenched (red) galaxies. Temperature histograms are split by the stellar mass of the central galaxy: blue (low), purple (intermediate), and orange (high). In addition, vertical lines at the top of each panel indicate the medians of each galaxy sub-sample, when split into inner ($r_\perp / r_{200} = 0.25$ and 0.5 exactly; dashed lines) and outer ($r_\perp / r_{200} = 0.75, 1$ and 1.25 exactly; shorter dotted lines) CGM LOS, following the inner and outer CGM definitions of Oppenheimer et al. (2018b).

We include lines at $r_\perp / r_{200} = 1.25$ to reflect the fact that the outer boundary of the CGM is not well-defined.

The H\textsc{i} absorbers trace gas across a broad range of physical conditions, both in terms of overdensity and temperature. Neutral H\textsc{i} absorption can occur in a wide range of conditions due to the high oscillator strength of the H\textsc{i} transition, though it traces hotter gas less well owing to collisional ionisation effects. Additionally, the presence of H\textsc{i} does not rely on the metal enrichment of the gas, thus less enriched lower-density regions farther from the galaxy can still yield strong absorption.

By comparison, the overdensity and temperature distributions for the metal lines are generally narrower, because individual metal ions tend to trace specific phase
Figure 5.5: Histograms of overdensity (top) and temperature (bottom) at the fitted positions of CGM absorbers for each of the ions we consider. Overdensity histograms are split by the type of central galaxy: star forming (blue), green valley (green) and quenched (red). Temperature histograms are split by the stellar mass of the central galaxy: blue (low), purple (intermediate), and orange (high). Vertical lines in the upper portion of each panel indicate the medians for each galaxy sample when split by inner (dashed) and outer (dotted) CGM LOS.
space conditions. The median overdensity for metal line absorbers moves to lower overdensities and higher temperatures for the higher ionisation species; we will examine this relationship with ionisation potential in more detail in section 5.3.6. This is expected since atoms tend to lie in lower ionisation states for denser, cooler gas. Additionally, chemical enrichment tends to be strongest close to the galaxy where the CGM is densest, yielding more favorable conditions for low ion absorption. Hence e.g. Mg II peaks at $\delta \sim 10^3$, while O VI peaks at $\delta \sim 10^2$, similar to H I. Meanwhile, the temperature histograms show low ions peaking at $T \lesssim 10^{4.5}$K, while O VI is most commonly found near its collisional ionisation peak temperature of $10^{5.5}$K, albeit with a significant tail to lower temperatures representing photoionised gas. We will examine the relative contribution to the metal absorbers from collisional and photoionised gas in more detail in section 5.3.4.

In the upper panels of Figure 5.5 we separate the absorber sample by the type of central galaxy (star forming, green valley or quenched) and compare CGM absorber overdensity distributions. Absorbers in the CGM of star forming galaxies typically occur in gas that is more overdense than the absorbers around green valley and quenched galaxies, although the effect is small. For H I absorbers in the inner CGM, the median overdensity occurs $\sim$0.3 and $\sim$0.6 dex higher for star forming galaxies than green valley and quenched galaxies, respectively. For absorbers in the outer CGM, the median overdensity is similarly higher around star forming than green valley galaxies, but there is little difference between green valley and quenched galaxies.

For each ion, the green valley absorber distribution more closely resembles that of the quenched galaxies than the star forming galaxies. This supports our interpretation of the CDDF that the CGM of green valley galaxies shows signs of being quenched before star formation completely ends in the central galaxy. Although we do not show it here, the overdensity distributions are largely independent of galaxy stellar mass.

For the temperature distributions in the lower panels of Figure 5.5, we separate the absorber sample by the stellar mass of the central galaxy, since it turns out that the temperature distributions are mostly independent of star formation activity. The peak of the temperature distribution clearly increases with stellar mass, since galaxies of higher stellar mass typically live in higher mass halos with higher virial temperatures. Galaxies in the intermediate and high stellar mass bin produce absorbers which broadly follow the same temperature distributions.
In contrast, galaxies in the lowest stellar mass bin have proportionally more absorbers in the low temperature tail of the distribution.

The median temperatures for these distributions are indicated by ticks along the top axis separated into inner and outer CGM, with the long ticks representing the former and the short ones the latter. Absorbers from the inner CGM trace higher overdensities than those from the outer CGM. This trend is clearest for the H\textsubscript{i}, C\textsc{iv} and O\textsc{vi} absorbers. The trend with impact parameter is not present for the Mg\textsc{ii}, which was also the case with Ford et al. (2013). We have seen from the Mg\textsc{ii} CDDF that there are relatively few Mg\textsc{ii} absorbers. Similarly, there is a slight trend with temperature for the H\textsubscript{i} absorbers, with absorption in the inner CGM tracing preferentially cooler gas. For the metal lines there is very little difference between the median temperatures for inner and outer CGM.

To summarise, H\textsubscript{i} traces a wide range of physical conditions in terms of overdensity and temperature, while the distributions for the metal absorbers are narrower, with peaks that decrease in overdensity and increase in temperature with increasing excitation energy. Absorbers from the inner CGM trace denser gas than those from the outer CGM; in contrast there is little difference in the absorber temperatures between the inner and outer CGM, except in the case of H\textsubscript{i}. Separating the absorbers by the star formation activity of their central galaxies demonstrates that CGM absorbers from star forming galaxies typically arise from denser gas than those from green valley or quenched galaxies. The absorber temperature is largely set by the stellar mass of the central galaxy. While these trends are subtle, they suggest that the ensemble of CGM absorption lines hold some information about the phase of CGM gas.

5.3.3 CGM absorbers in cosmic phase space

We now combine the overdensity and temperature information into 2D phase space plots to gain a more complete picture of the absorber properties and examine their relationship with absorption strength. Figure 5.6 shows the temperature-overdensity phase space for the CGM absorbers, colour-coded by the absorber column density. Upper plots show H\textsubscript{i} (left), Mg\textsc{ii} (middle) and C\textsc{ii} (right) absorbers, while the lower plots show Si\textsc{iii} (left), C\textsc{iv} (middle) and O\textsc{vi} (right). In each case the upper panels show absorbers in the inner CGM ($r_\perp/r_{200}$ of 0.25 and 0.5) and the lower panels show those in the outer CGM ($r_\perp/r_{200}$ of 0.75, 1, and 1.25). The lower limits of the column density colour scale are the
Figure 5.6: Gas temperature against overdensity at the fitted position of CGM absorbers for the 6 species we consider, colour-coded by absorber column density. In each plot the top and bottom panels show absorbers in the inner and outer CGM, respectively. The upper plots show H\textsc{i} (left), Mg\textsc{ii} (middle) and C\textsc{ii} (right) absorbers, while the lower plots show Si\textsc{iii} (left), C\textsc{iv} (middle) and O\textsc{vi} (right). The lower colour scale limits are the completeness limits and are different for each ion. The background greyscale in each panel shows the global phase space distribution for all gas along all LOS (this is identical for each species).
completeness limits for each ion. In this plot we show all absorbers above the completeness limits, making no distinction between the different types of central galaxy. The phase space distribution for all gas along all the LOS is shown in greyscale. Compared with the phase space distribution for the entire original simulation at \( z = 0 \) (see Christiansen et al., 2020), the distribution from LOS gas is biased towards higher densities as they are selected to probe in and around the CGM of galaxies.

It is clear by comparing the upper and lower panels of Figure 5.6 that there are more identified absorbers for each metal ion in the inner CGM than in the outer CGM, and that high column density absorbers in particular are rarer in the outer CGM. The difference in number of absorbers is smaller for H\(_i\), however the outer CGM still has preferentially fewer strong H\(_i\) absorbers. This is consistent with previous work examining halos in the SIMBA simulation that demonstrate a decreasing radial profile for density (Sorini et al., 2020, 2022) and EW (Chapter 4). The similarity between the phase space distributions in the inner and outer CGM illustrates that these absorbers arise from broadly the same physical conditions, but that such conditions are rarer in the outer CGM. This is particularly true for Mg\(_{\text{II}}\) absorbers, which are almost non-existent in the outer CGM.

Following Davé et al. (2010), the vertical lines in Figure 5.6 represent the approximate overdensity threshold at the boundary of a virialised halo, based on Kitayama & Suto (1996):

\[
\delta_{\text{th}} = 6\pi^2 (1 + 0.4093(1/f_\Omega - 1)^{0.9052})
\]  

(5.3)

where \( f_\Omega \) is:

\[
f_\Omega = \frac{\Omega_m (1+z)^3}{\Omega_m(1+z)^3 + (1-\Omega_m - \Omega_\Lambda)(1+z)^2 + \Omega_\Lambda}
\]  

(5.4)

At \( z = 0 \), the overdensity threshold is \( \delta_{\text{th}} \sim 111 \) for our cosmology. The horizontal lines represent the temperature threshold \( T_{\text{th}} = 10^5 \text{K} \), which is the conventional definition for the Warm-Hot Intergalactic Medium (WHIM) gas (Cen & Ostriker, 1999); temperatures above this threshold cannot easily be reached without gravitational shock heating or feedback. Davé et al. (2010) use these thresholds to define four regions of cosmic phase space: condensed gas which is cool and dense; hot halo gas which is dense and shock heated; diffuse gas which is cool and mostly photoionised; and the WHIM, which is diffuse and hot. Only gas in the condensed or hot halo phases can be considered part of the CGM, as less
dense gas is typically not bound to any halo. It was explicitly verified by Sorini et al. (2022) that the cumulative mass fraction of the diffuse and WHIM phases identified based on the overdensity threshold given by equation (5.3) is very close to the value obtained by summing over the mass of gas elements outside the boundaries of haloes in SIMBA.

From Figure 5.5 we have seen that HI absorbers span a wide range of physical conditions; Figure 5.6 shows that these absorbers are spread across the four categories of phase space, with the high column density absorbers \( (N_{\text{HI}} > 10^{15}\text{cm}^{-2}) \) absorbers arising from cool, condensed gas. Absorbers from the hot halo, WHIM and diffuse gas phases are typically weak, close to or below the completeness limit.

Absorption from the low ionisation metal lines (Mg II, C II, and Si III) lies almost exclusively in the condensed phase; the temperature of these absorbers rarely rises above \( T_{\text{th}} \). The highest column density CII absorbers lie along the same locus in the condensed phase as the strong HI absorbers; this is not apparent for Mg II and Si III as these metal absorbers are weaker. For CIV, these absorbers lie mainly across the condensed and diffuse regions of phase space, with the bulk of absorbers arising from condensed gas. In contrast to the lower ions, the temperatures of CIV absorbers can rise above \( T_{\text{th}} \), and there is a contingent of these absorbers arising from both the WHIM and hot halo phases. There is however little trend in the column density of CIV absorbers; the strongest absorbers are only mildly biased towards higher densities. Finally in the case of OVI, these absorbers exist across the diffuse, WHIM and hot halo regions, with the highest column density absorbers arising from hot halo gas. In contrast to the other ions, OVI absorbers rarely arise from cool, condensed gas. The phase space distributions of Mg II, C II, CIV and OVI CGM absorbers presented here is consistent with measurements of halo metals in the Illustris-TNG simulations (Artale et al., 2022).

Figure 5.7 quantifies these trends, showing the fraction of absorption arising from each region of cosmic phase space for each ion. This is computed as the fraction of the total number of fitted line profiles in each phase (top panel) and as the fraction of total column density in each phase (bottom). The fractions are split into the inner and outer CGM, represented as solid and hatched bars, respectively; the inner CGM clearly dominates the absorber samples. The number fractions show the intuitive expected fractions from examining Figure 5.6, whereas the absorption fractions demonstrate how the particular regions of phase space would dominate the actual observed absorption. The total column density of the LOS is
Figure 5.7: Top: the number fraction of identified absorbers in each region of cosmic phase space, for each of the species we consider. Fractions are split into inner (solid) and outer (hatched) LOS. Bottom: as above, showing the fraction of total column density.
a robust measurement of the overall absorption, as we have verified that the total LOS equivalent width from profile fitting largely corresponds to the equivalent width from direct measurement. Hence the total absorption can be comfortably compared between works with different fitting procedures. However the number fractions is less comparable since arguably the precise number of absorbers is dependent on the details of the profile fitting procedure.

The number fractions of H\textsc{i} absorbers shows that 77\% arise from cool gas (39\% and 38\% in the condensed and diffuse phases, respectively), 9\% from hot halo gas and 14\% from the WHIM. In addition, the inner CGM represents 52\% of all H\textsc{i} absorbers and 69\% of the condensed H\textsc{i} absorbers. For the low metal ions, virtually all of the absorbers are condensed gas. These mostly arise from the inner CGM, with the fraction in the inner CGM decreasing with ionisation energy for these low metal ions. More generally, the fraction of absorbers in the condensed phase decreases with increasing ionisation energy. C\textsc{iv} in contrast has a significant fraction of absorbers (37\%) not in condensed gas. For O\textsc{vi}, the largest fraction of absorbers comes from the WHIM (48\%) while the hot halo and diffuse gas comprise 34\% and 16\% of the absorbers, respectively. The inner and outer CGM contribute roughly equally to the absorption from the WHIM (41\% and 59\%, respectively), whereas the hot halo absorption seen from the inner CGM is less common in the outer CGM.

Examining the fraction with respect to total absorption demonstrates that while our absorber sample represents a wide range of overdensities and temperatures, the majority of observed absorption arises from inner CGM gas in the condensed phase, except in the case of O\textsc{vi}. H\textsc{i} absorption is split into 83\% condensed in the inner CGM and 15\% condensed in the outer CGM, with only 2\% arising from the remaining phases. For the low ions Mg\textsc{ii}, C\textsc{ii} and Si\textsc{iii} the condensed gas fraction from the inner CGM is essentially 100\% of the total absorption (97\% for Si\textsc{iii}). For C\textsc{iv}, although 37\% of absorbers lie in the diffuse, hot halo and WHIM phases, these collectively contribute only 13\% of the total C\textsc{iv} absorption. Finally for O\textsc{vi} the proportion in the hot halo phase rises to 45\% when considering total absorption. This increase comes at the expense of all other phases, such that the WHIM and hot halo gas contribute equally to the majority of the absorption (88\%). Since gas in the WHIM phase as defined by Davé et al. (2010) is not strictly part of the CGM, we conclude that only about half (\sim 46\%) of the observed O\textsc{vi} absorption actually arises from CGM gas.

In summary, there are fewer absorbers in the outer CGM compared with the inner
CGM, regardless of absorber species. H I (and C IV) absorbers are distributed across all regions of phase space, but most of the overall absorption (98% and 87%) arises from condensed gas. Absorption from low ions (Mg II, C II, Si III) arises almost exclusively from condensed gas, both when considered in terms of number of absorbers and total overall absorption. In contrast, O VI absorbers arise predominantly from gas in the WHIM or hot gas phase, with less than half of absorbers arising from gas that is considered part of the CGM.

5.3.4 Sources of ionisation

Previous work using SIMBA to study the CGM has found that the choice of photoionising background makes a considerable difference to such measures of the CGM absorption structure (Chapter 4). Here we extend this analysis by examining the relative contributions of collisional ionisation versus photoionisation to the ionisation structure of the CGM. We do so for C IV and O VI as we have seen in §5.3.3 that these ions have significant fractions of absorbing gas above the canonical temperature limit for photoionisation ($T_{\text{photo}}$). Since the photoionising sources are unable to ionise above $\sim 10^5$K, we expect that any absorbers above $T_{\text{photo}}$ must arise from collisional ionisation; hence the hot gas absorption fraction is a reasonable proxy for the contribution to the total column density from collisional sources. Previous simulations have shown that high excitation ions trace collisionally ionised gas at high temperatures (Ford et al., 2013). We adopt temperature thresholds of $\log(T_{\text{photo}}/\text{K}) = 4.8$ for C IV (Mauerhofer et al., 2021) and $\log(T_{\text{photo}}/\text{K}) = 5$ for O VI (Böhringer, 1998; Oppenheimer & Davé, 2009).

Figure 5.8 shows the fraction of total column density arising from hot gas above $T_{\text{photo}}$ as a function of impact parameter, for C IV (dashed with squares) and O VI (dotted with triangles). We separate the absorbers by the stellar mass of their central galaxy: low (blue), intermediate (purple) and high (orange). Both ions show significant fractions of absorbers in collisionally ionised gas. O VI absorbers are at a minimum $\sim 80\%$ collisionally ionised, consistent with high O VI collisional fractions measured for this mass regime in the NIHAO simulations (Gutcke et al., 2017). Similarly, Oppenheimer & Davé (2009) found that while much of the IGM O VI absorption is photoionised, the strongest O VI absorbers (i.e. those who contribute most to the total O VI column density) are likely collisionally ionised and trace recent enrichment within galaxy halos. For C IV, the minimum collisional fraction is $\sim 20\%$. Large collisional fractions for the high excitation
Figure 5.8: The fraction of absorption arising from collisionally ionised gas as a function of impact parameter, for CIV (dashed lines with squares) and OVI (dotted lines with triangles). Line colours represent broad galaxy stellar mass bins.
ions naturally follows from our discussion of the high proportion of O\textsc{vi} absorbers which arise from hot halo gas or the WHIM. These results demonstrate that collisional ionisation is the dominant ionising source for these intermediate/high excitation ions.

We have tested the impact of the ionising background on our results by re-generating the C\textsc{iv} and O\textsc{vi} spectra using the Haardt & Madau (2012) photoionising background instead of the Faucher-Giguère (2020) background. The results are mostly unchanged - we find that the absorbers occupy the same regions of phase space and that the fraction of O\textsc{vi} from collisional sources is qualitatively the same. However for C\textsc{iv} there is a slight (∼5%) increase in the fraction of absorption from collisions. This arises since the Haardt & Madau (2012) background predicts lower fractions of C\textsc{iv} in the photoionised temperature regime, at the level of ∼0.15 dex at the photoionisation limit of 10^{4.8}\text{K} and δ = 10^2 (the precise decrease is sensitive to temperature and overdensity). This leads to (slightly) reduced C\textsc{iv} absorption from the ionising background and thus a higher contribution from collisional sources.

We break these results down further as a function of stellar mass and impact parameter. The contribution from collisions increases with stellar mass for both ions, which is a reflection of the increasing halo virial temperature for high mass galaxies (as discussed in §5.3.2). For the C\textsc{iv} absorbers there is little increase between the low and intermediate mass bins, but a substantial increase from of ∼30% in collisions between the intermediate and high mass bins. Galaxies in the intermediate mass bin (10^{10.5} M_\odot < M_\star < 10^{11} M_\odot) have mass-weighted average CGM temperatures down to 10^{5.5}\text{K} (see Figure 5.1) and their gas-phase metals roughly evenly distributed across the ISM and cool, warm and hot CGM gas phases (§4.2.2). Thus gas below the photoionisation limit for C\textsc{iv} is more common in low and intermediate mass halos, whereas the CGM of high mass galaxies (M_\star > 10^{11} M_\odot) have a mass-weighted average CGM temperature of no less than 10^6\text{K} and are more dominated by hot CGM gas (cit.), resulting in a large increase in collisional fraction. For the O\textsc{vi} absorbers the largest increase in collisional fraction (∼10%) occurs between the low and intermediate mass bins. In §5.3.2 we noted that the O\textsc{vi} temperature distribution around low mass galaxies has a tail of low temperature absorbers; here we see that this photoionised tail contributes at most ∼20% of the total absorption.

The collisional absorption fraction also slightly decreases with impact parameter, which reflects the downward radial trend in halo temperature as traced by X-
rays in Simba (Robson & Davé, 2020). The radial trend is present for the O\textsc{vi} absorbers in every mass bin; for C\textsc{iv} absorbers the trend is strongest for the highest mass bin, while absorber fraction profiles for low and intermediate mass galaxies is broadly flat within the noise. The radial decrease is consistent with results from the NIHAO simulations, which show that photoionisation is more prominent on the outskirts of halos (Roca-Fàbrega et al., 2019). To summarise, the fraction of collisionally ionised absorption is $\sim 25$–$55\%$ for C\textsc{iv} and $\sim 80$–$95\%$ for O\textsc{vi}, increases with galaxy stellar mass, and decreases with impact parameter.

### 5.3.5 Physical density versus column density

Naively, one expects that higher density gas gives rise to higher column density absorption. However, temperature and metallicity variations, along with geometrical effects, can complicate this relationship. Here we take a closer look at the relationship between physical overdensity of the 3D gas elements and the measured 2D column densities of our absorber sample, and examine any dependence on the type of central galaxy.

Figure 5.9 shows gas overdensity against absorber column density, colour-coded by $r_{200}$-normalised impact parameter. Results are split into absorbers arising around star forming (left), green valley (middle) and quenched (right) galaxies. The purple lines in each panel represent the running median for inner (dashed) and outer (dot-dashed) CGM absorbers; we show medians only where the column density bins contain at least 15 data points. The solid purple lines indicate the best fit to a power law for the H\textsc{i} absorber data between the completeness limit and $10^{15}$ cm$^{-2}$; the dotted purple lines show the best fit extrapolated to higher column densities.

First we examine the overall results for the whole absorber sample. For each absorber species, there is a positive trend of increasing overdensity with column density, as broadly expected. For the metal lines, the cloud of absorbers moves to lower overdensities with increasing excitation energy. The running medians provide an estimate for the overdensity for observed absorbers at a given column density. Nonetheless, there is a significant scatter around this relationship.

Of the absorber species we include, H\textsc{i} has the steepest increase of overdensity with column density. Between the completeness limit and $10^{15}$ cm$^{-2}$, the H\textsc{i}
Figure 5.9: Overdensity against column density for each of the species, split into star forming (left), green valley (middle) and quenched (right) galaxies. Points are colour-coded by the $r_{200}$-normalised impact parameter. Purple dashed and dot-dashed lines show the running medians for galaxies in the inner and outer CGM, respectively. The solid purple lines indicate the best power law fit to the HI absorber data between the completeness limit and $10^{15}$ cm$^{-2}$; the dotted purples lines show the best fit extrapolated to higher column densities.
absorbers follow a power law:

$$\log_{10} \delta = a \log_{10}(N_{\text{HI}}/\text{cm}^{-2}) + b$$

(5.5)

For star forming galaxies, the best fit parameters for the power law scaling are $a = 0.66$ and $b = -6.9$; for green valley galaxies these are $a = 0.7$ and $b = -7.5$; for quenched galaxies they are $a = 0.76$ and $b = -8.5$. These best fit parameters are reasonably similar to those found for H I for IGM absorbers by Davé et al. (1999) ($a = 0.7$ and $b = -8.5$), despite the many differences between these simulations. The H I absorbers in Simba trace higher overdensities for a given column density, perhaps simply owing to the higher densities found in the CGM compared with IGM gas, as well as changes in the assumed photo-ionising background.

In addition, the CGM in Simba has a deficit of H I absorbers at $N_{\text{HI}} \sim 10^{17}$ cm$^{-2}$, and a cloud of absorbers at higher column densities ($N_{\text{HI}} > 10^{17}$ cm$^{-2}$). Beyond $N_{\text{HI}} > 10^{15}$ cm$^{-2}$ the overdensity no longer increases uniformly with column density, as the absorption features are likely saturated. These absorbers exist on the logarithmic part of the curve of growth and so it becomes difficult for the Voigt profile fitter to disentangle increases in column density vs. linewidth. The cloud of high column densities comes from gas at all impact parameters, but preferentially comes from the inner CGM.

When we split the absorbers by the star formation rate of central galaxy, we see subtle differences in these absorber samples, as hinted by the star formation dependence of the power law best fit parameters. When compared with star forming galaxies, quenched galaxies have fewer absorbers overall and preferentially fewer at the highest overdensities and at the highest column densities, due to the heating of the galaxy halo via AGN feedback. The reduction in the number of absorbers around quenched galaxies is largest for low metal absorbers since AGN feedback heats gas out of the optimal cool, dense regime for low metal absorption (Christiansen et al., 2020). In general, absorbers around star forming galaxies above $\delta_{\text{th}}$ come from the inner CGM, whereas absorbers below this threshold come from the outer CGM. However for quenched galaxies, this clear distinction in overdensity between inner and outer CGM is not seen; this is clear from the running medians for the inner and outer CGM absorbers, which lie close together for quenched galaxies. These results suggest that the CGM around quenched galaxies is more uniform in nature.

Despite these differences, overall we find that absorbers from around galaxies of
different star formation activities do not come from fundamentally different types of gas, and it is not possible to cleanly separate our sample into different galaxy types by making cuts in overdensity-column density space. Although we do not show it here, we have also produced the converse version of Figure 5.9 (coloured by the star formation activity and split into inner and outer CGM), and similarly found that absorbers from different types of galaxies occupy similar regions of phase space. We have also examined the stellar mass dependence and found that the impact of increasing stellar mass largely resembles that of decreasing star formation rate—fewer absorbers for each species, but little difference in the distribution of absorbers in density space. These plots demonstrate a complex, degenerate relationship between star formation, stellar mass, absorption strength, and distance from the galaxy. We find that while the overall number and strength of absorption reduces when galaxies are quenched, conditions necessary for absorption do not change, and thus there is little change in the distribution of absorbers in phase space.

5.3.6 Typical physical conditions of absorbers

Having presented the physical gas conditions for our whole absorber sample, here we summarise these results by examining typical gas properties for each species. Figure 5.10 shows the column density-weighted median overdensity (top), temperature (middle) and metallicity (bottom) against ionisation energy for the six ions we consider. For H\textsubscript{i}, the metallicity is taken as the total metal mass fraction relative to solar abundance (0.0134), while for the individual ions it represents the solar-scaled abundance in the corresponding element. Column density-weighted medians represent the values at which 50% of the absorption occurs above (or below) this value. Points are colour coded by the \( r_{200} \)-normalised impact parameter. We show representative error bars for the \( r_{\perp}/r_{200} = 0.25 \) set of points, representing the 25-75 percentile range. The horizontal lines in the top and middle panels indicate \( \delta_{\text{th}} \) and \( T_{\text{th}} \), respectively.

First we examine the typical conditions of the H\textsubscript{i} absorbers. The typical overdensity of H\textsubscript{i} absorbers decreases with increasing impact parameter, dropping below \( \delta_{\text{th}} \) at 0.75\( r_{200} \), indicating that absorbers seen at these large impact parameters are not typically part of the CGM. The typical temperature is \( \log(T/\text{K}) \sim 4.5 - 4.7 \); there is a slight increase in temperature with increasing impact parameter, although in this case the increase is within the size of the
Figure 5.10: Top panel: The column density-weighted median overdensity against ionisation energy for the 6 ions considered in this work. Points are colour-coded by the $r_{200}$-normalised impact parameter. Representative error bars are plotted for the $r_{\perp}/r_{200} = 0.25$ set of points, representing the 25-75 percentile range. Middle panel: as above, for column density-weighted gas temperature. Bottom panel: as above, for column density-weighted gas metallicity. Horizontal dotted lines in the top and middle panels represent the $\delta_{\text{th}}$ and $T_{\text{th}}$ thresholds, respectively.
interquartile spread. Lastly, the typical metallicity of H\textsc{i} absorbers decreases with impact parameter; this reflects a decreasing metal enrichment of the CGM as a function of distance from the central galaxy.

Median overdensity of metal absorbers decreases with ionisation potential and with increasing impact parameter, consistent with Ford et al. (2013) (except in the case of Mg\textsc{ii} at high impact parameter; there are few Mg\textsc{ii} absorbers at this distance from the central galaxy). At high impact parameters, the halo gas is more diffuse and thus metals are ionised preferentially by photoionisation, rather than via collisions – this is reflected in the lower typical densities of metal absorbers at high impact parameter. For more detail on the relative contribution of the two sources of ionisation, see §5.3.4.

The median temperature of metal absorbers increases with ionisation potential. Mg\textsc{ii} and O\textsc{vi} have a weak dependence on impact parameter, with higher typical temperatures at lower impact parameters; the other metal ions have no dependence on impact parameter. C\textsc{ii}, Si\textsc{iii} and C\textsc{iv} require temperatures around $10^{4.5}$K, the canonical temperature threshold for photoionisation. O\textsc{vi} absorption arises from warm gas that constitutes the volume-filling ambient gas of the CGM (e.g. Ford et al., 2013; Oppenheimer et al., 2020a).

The median metallicity of metal absorbers increases slowly with ionisation potential and decreases with impact parameter. As with H\textsc{i}, this is a reflection of the underlying halo metallicity profile. However, the trend with impact parameter is much less strong than for H\textsc{i}, because much of this absorption arises in dense gas close to galaxies, at least for lower ionisation absorbers. At all impact parameters, the metal absorption arises from gas that is near the solar metallicity. This is even true for O\textsc{vi} at high impact parameters which arises predominantly in WHIM gas, potentially reflecting a bias for such absorption to arise in regions of relatively high enrichment even when the overall metallicity is lower (Oppenheimer & Davé, 2009).

## 5.4 Redshift dependence

Up until now, we have presented results for our sample of absorbers at $z = 0$; here we examine the redshift dependence of our results using our higher redshift galaxy samples. We generate a set of synthetic absorption spectra for the galaxies
Figure 5.11: Redshift dependence of the CDDF for each species considered in this work, shown at $z = 1, 0.5, 0.25$ and $z = 0$. The smaller panels show the fraction with respect to the CDDF at $z = 0$. Vertical error bars on show representative cosmic variance uncertainties, computed using the $z = 1$ sample; horizontal error bars indicate the width of the column density bins.

in our $z = 0.25, 0.5$ and 1 samples following an identical procedure to that for our $z = 0$ sample (beyond $z = 1$ these absorbers are redshifted out of the UV regime and as such different observational probes of the CGM are required). We fit Voigt profiles to these LOS as before to produce a sample of absorbers for each redshift.

Figure 5.11 shows the redshift evolution of the CDDF for each species. The larger panels show the CDDFs, while the smaller panels show the CDDF fraction of the higher redshift samples with respect to the $z = 0$ case. The horizontal dashed lines in the lower panels indicate the case where there is no redshift dependence. The vertical dashed lines indicate the $z = 0$ completeness limits; the completeness limits for the higher redshift samples are lower than these, owing to an increase in number of absorbers with redshift. As such, above these column densities the CDDFs for each redshift can be considered representative of a complete sample.

As in Figure 5.3, the horizontal error bars show the width of the column density bins, where we have increased the bin width at the high column density end to account for low numbers. Vertical error bars show representative cosmic variance uncertainties using all galaxies and all LOS. These are computed as in §5.3.1: the
and \( y \) sides of the simulation box are each split into quarters to form 16 cells across the face of the simulation; the absorbers are binned into the cells based on their \( x - y \) positions; the cosmic variance uncertainties are computed by jackknife resampling across the 16 cells. For the CDDF fraction panels, the vertical error bars are computed by adding the \( z = 1 \) errors in quadrature with the \( z = 0 \) errors.

In general, for each species the normalisation of the CDDF at intermediate column density decreases with redshift. This does not represent a reduction in the number of absorbers within the simulation but mostly owes to the increase in the comoving path length \( dX \); examining simple histograms of column density shows that the raw number and strength of absorbers increases with redshift. The increase is greatest for \( \text{H}\text{I} \), but is present for every species. At \( z \geq 0.5 \), we find that \( \text{O}\text{VI} \) absorbers exist outside of the narrow column density range seen at \( z = 0 \). However, when we account for the cosmic evolution of the path length (which increases by a factor of 4.6 between \( z = 0 \) and \( z = 1 \)), the CDDF decreases with redshift. At the high and low column ends of the CDDF, the large increase in the number of absorbers at \( z = 1 \) compensates for the path length, and here we find that the difference with respect to the \( z = 0 \) CDDF is reduced.

Finally, we turn to the redshift dependence of the physical conditions of the CGM absorbers. Figure 5.12 shows the redshift evolution of the typical gas conditions for our absorber sample. As in Figure 5.10, we plot the column density-weighted median overdensity (top), temperature (middle) and metallicity (bottom) against the ionisation energy of each species. Different coloured points represent the results for the different redshift samples. The vertical error bars represent the 25-75 percentile range for the \( z = 1 \) sample. The horizontal dashed lines in the top and middle panels indicate \( \delta_{\text{th}} \) and \( T_{\text{th}} \), respectively.

For each metal ion, the column density-weighted median overdensity of the gas increases with redshift; at low redshift the halo gas is more diffuse due to the evacuating impact of the AGN feedback (Sorini et al., 2022). The median temperatures however do not change significantly; we see a small rise in typical temperature of \( \text{H}\text{I} \) absorbers and a decrease for \( \text{O}\text{VI} \) absorbers. This occurs because the stronger photo-ionising background at \( z = 1 \) results in slightly higher temperatures in the diffuse IGM, while less hot IGM gas owing to AGN feedback at higher redshifts results in lower \( \text{O}\text{VI} \) temperatures.

Earlier in this Chapter we noted that the metal absorbers arise from a narrow range of gas temperature, which we now see does not significantly depend on
Figure 5.12: As in Fig. 5.10, showing the redshift dependence of the typical absorber conditions. The column density-weighted median absorber physical conditions for each species are plotted against ionisation energy: overdensity (top), temperature (middle) and metallicity (bottom). The different colours represent different redshifts. The horizontal lines in the top and middle panels indicate $\delta_{th}$ and $T_{th}$, respectively. The vertical error bars on the $z = 1$ data points represent the 25-75 percentile range.
redshift. In section 5.3.2 we showed that the absorber temperature is most sensitive to central galaxy stellar mass rather than star formation activity; the stellar mass of our galaxy sample does not change with redshift since we have selected our galaxies within the same stellar mass ranges at each redshift. The typical metallicity of the absorbers increases with decreasing redshift, owing to a general build-up of metals over cosmic time due to star formation (Finlator & Davé, 2008). The largest increase in metallicity is seen in the H\textsc{i} absorbers; where metal absorbers are seen they must arise from metal-enriched gas by necessity regardless of redshift, hence the increase in metallicity over cosmic time of the metal absorbers is less substantial. Thus the increase in typical metallicity of the metal lines reflects a further build-up of metals in gas that has already been enriched by $z = 1$, whereas the large increase in typical metallicity of the H\textsc{i} absorbers represents new metals that are added to the CGM at large since $z = 1$.

In summary, the CDDF normalisation decreases with redshift for all absorber species we examine, owing to the large increase in comoving path length rather than a reduction in number of absorbers within the volume. The increase in metallicity over cosmic time somewhat counteracts the decrease in absorber overdensity such that there is still significant metal absorber populations at $z = 0$. We note that these results are sensitive to our choice of galaxy sample selection with each redshift (a constant stellar mass criteria and redshift-dependent sSFR thresholds). This means that the quenched galaxies in our $z = 1$ sample will have quenched via a different combination of mechanisms to our $z = 0$ sample; likewise galaxies in our highest stellar mass bin at $z = 1$ are more rare at that epoch. A complementary approach would be to identify the progenitors of our $z = 0$ galaxy sample and examine the history of their CGM absorbers over the course of their individual evolutions; the progenitors would likely have higher absorber counts in their CGM owing to their lower masses at earlier epochs.

### 5.5 Contribution from satellites

A galaxy’s halo contains not only the CGM but also satellite galaxies. An interesting question is whether CGM’s absorption at a given impact parameter is associated with a satellite galaxy rather than the halo gas of the central (e.g. Agertz et al., 2009; Gauthier et al., 2010; Huang et al., 2016). This is sensitively dependent on how one defines such an association. In this section, we quantify the satellite contribution to CGM absorption, under various assumptions about
Figure 5.13: Profiles of median fraction of equivalent width arising from particles associated with satellite galaxies. Lines are colour coded by the particle distance used to decide association with satellites, defined as factors of the satellite half mass radius ($f_{1/2,*}$). Shading around the $\log f_{1/2,*} = 1$ lines show the 25-75 percentile range.
how far out the gas from a satellite is considered to be associated.

We quantify the contribution to CGM absorption from satellite galaxies by regenerating the synthetic absorption spectra for our galaxy sample using only gas elements that are associated with satellites. We define associated satellite gas using the galaxy half stellar mass radius ($R_{\text{half}}$). That is, for each satellite, we identify all gas particles within various factors of $R_{\text{half}}$, namely $\log R_{\text{half}} = 0, 0.5, 1, 1.5, 2$. For each factor of $R_{\text{half}}$ we output a new snapshot containing only particles associated with satellites within that distance, and generate new snapshot using the same lines of sight as in the main sample. A more theoretically-based approach to this problem would involve running a subhalo finder (e.g. Rockstar, Behroozi et al., 2013), and identifying all gas elements that are gravitationally bound to those subhalos; we leave this approach for future work, and focus here on our more observationally-accessible proxy. The contribution of satellite galaxies to the original spectra is computed as the median fraction of equivalent width which arises from these new snapshots, with respect to the original spectral sample. We restrict this test to LOS that had at least one absorption region detected by our profile fitting procedure.

Figure 5.13 shows profiles of median fraction of equivalent width arising from gas particles that are associated with satellite galaxies. Lines are colour-coded by the particle distance used to define association with satellites. The shaded region in each panel represents the 25-75 percentile range. The solid purple line in each panel represents the results for the satellite gas definition which we adopt going forward.

The satellite absorption fraction increases with increasing distance used to define satellite association, recovering essentially 100% of the original equivalent width where particles are included up to $10^{2.5}\times$ and $10^{2}\times$ the satellite $R_{\text{half}}$ for H\textsc{i} and the metals, respectively. There is little increase in satellite contribution between snapshots using 1 and 3 times the satellite $R_{\text{half}}$, and there is only a modest increase in satellite absorption fraction when increasing the radius from $1 \rightarrow 10R_{\text{half}}$, whereas above this the increase is substantial, suggesting that $10R_{\text{half}}$ is a reasonable and robust distance limit for gas particles to be associated with satellites. In Figure 5.13, the results for this definition are shown as the solid line. In this case, the satellite contribution to the overall H\textsc{i} absorption is around 3%, largely independent of radius. Thus most of the H\textsc{i} absorption is associated with ambient CGM gas, rather than satellites.
In contrast, for metals the contribution is always higher than this. While satellite
gas always provides a minority contribution to the overall absorption, it is evident
that the low ionisation lines have a larger contribution from satellites than higher
ionisation absorbers, and furthermore show a more strongly increasing trend with
radius. For instance, the satellite absorption fraction at $0.25r_{200}$ increases from
$\sim 8\%$ for O\textsc{vi} to $\sim 16\%$ for Mg\textsc{ii}, and the latter reaches $\sim 40\%$ in the outer
CGM. This effect occurs since Mg\textsc{ii} and C\textsc{ii} typically arise from cool, dense
gas (see section 5.3.3) that is unlikely to be in the warmer ambient medium of
the galaxy halo, unless that gas exists near a satellite galaxy. In contrast, high
ionisation absorption like O\textsc{vi} is spread more uniformly throughout the halo,
leading to little radial trend and mostly coincident association with satellites. In
short, CGM absorption predominantly comes from ambient gas, but particularly
for low ionisation absorption in the outer CGM, a substantial fraction is expected
to arise from gas close to a satellite galaxy.

5.6 Conclusions

We have examined a population of low redshift CGM absorbers and their physical
conditions in the Simba cosmological simulations. The underlying galaxy sample
population was chosen to evenly sample a range of stellar masses and star
formation activities at $z = 0, 0.25, 0.5$ and $z = 1$. For each galaxy in our sample
we have generated realistic line of sight synthetic absorption spectra for H\textsc{i} and
a selection of metal ions at a range of impact parameters, and fitted Voigt line
profiles to extract absorber column densities and physical conditions. We have
used this sample to examine CGM absorber counting statistics in the form of the
CDDF, the distribution of physical gas conditions that produces CGM absorption,
and the redshift dependence of our results. In addition, we have estimated the
contributions of collisional ionisation and of satellite galaxies to the resulting
absorption strengths. Our main results are as follows:

- The CGM CDDFs follow a power law at the high column density end, with a
turnover at the low column density end which defines the completeness limit
of our sample. H\textsc{i} is the most abundant absorber species in our sample, with
a large population with high column densities; Mg\textsc{ii} is the least abundant
absorber. The sample probes only a narrow range of O\textsc{vi} column densities
due to our spectrum generation approach.
• The normalisation of the CGM CDDFs (i.e. the number of absorbers at fixed redshift) increase with the sSFR of the central galaxies. The CDDFs from green valley galaxies resembles the quenched galaxy CDDF more closely than the star forming CDDF, suggesting that the CGM shows signs of quenching before the central galaxy itself is completely quenched. The CDDFs from high mass galaxies also have a lower normalisation owing to the hot gas content of these galaxy halos.

• $\text{H}^\text{i}$ absorbers trace gas across a wide range of overdensities and temperatures owing to the high oscillator strength of $\text{H}^\text{i}$; metal absorbers trace a narrower range of physical conditions. Median gas overdensity decreases and temperature increases with increasing ion excitation energy. In general, median gas overdensity increases with increasing galaxy sSFR, while median gas temperature increases with increasing stellar mass. The typical metallicity of CGM absorbing gas is around solar for all metal ions with only a weak radial trend, although $\text{H}^\text{i}$ can trace significantly lower metallicity gas at larger $r_\perp$.

• Absorbers from the inner CGM are more abundant than those from the outer CGM, in general have higher column densities and trace denser gas.

• $\text{H}^\text{i}$ absorption can arise from gas that is in any of the four regions of phase space (condensed, diffuse, hot gas, or WHIM). Condensed gas accounts for the majority ($\sim 100\%$) of all $\text{H}^\text{i}$ and low metal ion absorption, and $87\%$ of $\text{C}^\text{iv}$ absorption. $\text{C}^\text{iv}$, an intermediate ion, is the first of our selected metal ions to have a notable contribution from the other phase space regions ($37\%$ of absorbers). In contrast with the other ions, $\text{O}^\text{vi}$ absorbers predominantly arise from either hot gas or the WHIM (together $87\%$ of total absorption).

• The fraction of absorption from collisionally ionised gas is $\sim 25 - 55\%$ for $\text{C}^\text{iv}$ and $\sim 80 - 95\%$ for $\text{O}^\text{vi}$. The collisional contribution increases with galaxy stellar mass and decreases with impact parameter.

• In general more overdense gas leads to higher column density absorption, up to $\log(N/\text{cm}^{-2}) > 15$, above which the absorption features are saturated.

• Absorbers around star forming galaxies trace higher overdensities and column densities, and have a more pronounced radial trend in overdensity than quenched galaxies, with higher overdensities and column densities in general occurring in the inner CGM. However the differences between star
forming, green valley and quenched galaxies are small, and in general we find that absorbers around galaxies of different star formation activities do not probe fundamentally different phases of gas.

- When the excitation energy of the absorber species increases, the column density weighted median overdensity decreases, temperature increases, and metallicity increases to a lesser extent. Column density weighted overdensities and metallicities decrease with increasing impact parameter, however there is little trend with impact parameter and gas temperature.

- Although absorber counts increase with increasing redshift, this is counteracted by the increase in the comoving path length, resulting in a CDDF which decreases overall with increasing redshift.

- Satellite galaxies in the halos of central galaxies contribute $\sim 3\%$ of the overall H\textsc{i} absorption and $\sim 10\%$ of metal absorption, except for Mg\textsc{ii} absorption, of which $\sim 30\%$ comes from satellites. The fraction of Mg\textsc{ii} absorption from satellites increases with increasing impact parameter. Overall, the fraction of metal absorption from satellites decreases with increasing ion excitation energy. These results are based on a definition of satellite-associated gas being within 10 stellar half-light radii of the satellite, but are relatively insensitive to values below this. This suggests that the dominant component of CGM absorption arises from ambient halo gas.

These results quantify how UV absorbers trace the typical physical conditions in the CGM around galaxies of different types. It provides a guide towards interpreting low-redshift CGM absorption line observations, within the context of the Simba model. In future work we plan to compare this to the results from other simulations, to understand how the different feedback processes implemented across present-day galaxy formation simulations impact CGM absorption tracers.

Our results show that the cool CGM gas is well-traced by low-ionisation metal absorbers, although all such species do not arise in the same gas. This suggests that the common approach to obtaining physical conditions of CGM gas of fitting single-phase Cloudy models to low-ionisation absorbers may not be fully accurate. Also, the absorbing gas can be a highly biased tracer of the overall physical properties of CGM gas, so extracting information such as the total CGM mass (even within the cool phase) can be difficult. The UV absorbers also clearly do not trace the dominant hotter gas component around quenched and green
valley galaxies. In the future we plan to investigate more comprehensive methods for extracting the physical conditions of the CGM.

Understanding the CGM and the role it plays in baryon cycle is a frontier issue in current galaxy formation models. The work presented here is a step towards characterising the CGM via one particular technique, namely UV absorption. Combining this with other approaches such as resolved H\textsc{i} studies, X-ray absorption, and metal emission line mapping in the UV and X-rays will be critical for fully disentangling the complex multi-phase gas within the CGM of galaxies both today and back in time.

## 5.7 Voigt fitting accuracy

Here we provide a comparison of the results from our Voigt profile fitting routine to those retrieved directly from the synthetic spectra, both in terms of column density and equivalent width.

Figure 5.14 shows the log difference between the column densities derived from spectrum fitting $N_{\text{fit}}$ and column densities summed along the LOS $N_{\text{sum}}$, against $N_{\text{sum}}$. The column densities are totals within each absorption region, and the points are coloured by the $\chi^2$ of each region.
Figure 5.15: The total fitted equivalent width along the LOS against the total equivalent width summed within the ±600 km s$^{-1}$ window along the LOS. Points are coloured by the maximum $\chi^2_r$ of the fitted regions along the LOS.

The column densities presented here are totals within each detected absorption region. Points are coloured by the $\chi^2_r$ of each region; this dataset includes poor fits and column densities below the completeness limit. The horizontal dashed line at indicates where the fits match the summed column densities perfectly; the two measures of column density agree well in general. However, at the low (high) log $N$ end the fitting process has a tendency to underestimate (overestimate) the column density. In particular, the Voigt profile fitter gives an unreliable measure of H I column density for Lyman Limit Systems since these features are saturated and lie in the logarithmic part of the curve of growth.

Figure 5.15 shows the total fitted equivalent width along the LOS (EW$_{fit}$) against the equivalent width summed within the whole ±600 km s$^{-1}$ window (EW$_{sum}$). In contrast to Figure 5.14, here EW$_{fit}$ and EW$_{sum}$ are totals from the whole velocity window, rather than results for individual absorption regions. Here we include all LOS, including those with no detected absorbers (plotted at log(EW/A$_{fit}$) = -2). Points are coloured by the maximum $\chi^2_r$ of the fitted regions along the LOS. In general, EW$_{fit}$ and EW$_{sum}$ follow a 1:1 relation (indicated by the dashed black line). As with the column densities, Voigt fitting can overestimate the equivalent width of H I at the high EW end. More commonly, the total equivalent width
is underestimated due to one or more absorption regions not being detected and passed to the fitting routine.
Chapter 6

Mapping Circumgalactic Medium Observations to Theory Using Machine Learning

6.1 Introduction

Over recent years, there has been much effort to characterise the CGM via quasar absorption line studies (see reviews by Putman et al., 2012; Tumlinson et al., 2017; Péroux & Howk, 2020). Many of the studies probe the strong transitions that exist in the rest ultra-violet (UV) regime and which trace cool or warm gas. Such studies are motivated by the wish to understand the baryon cycle of gas flows in the CGM: accretion onto galaxies from the IGM and satellite galaxies; expulsion of gas via stellar winds and AGN feedback; recycling of previously ejected material back onto galaxies.

The physical conditions of the CGM are studied by retrieving kinematics, spatial distributions, metallicities, densities, and temperatures from the absorption features (e.g. Stocke et al., 2013; Savage et al., 2014; Werk et al., 2014; Lehner et al., 2014, 2018, 2019; Wotta et al., 2016, 2019; Keeney et al., 2017; Prochaska et al., 2017; Qu et al., 2022). To extract physical conditions, absorption systems are commonly fitted with Voigt profiles to model each absorption component and obtain column densities, linewidths and redshift-space positions. By running ionisation models (typically using CLOUDY, Ferland et al. 2017) and varying the
input physical parameters, a Bayesian search can be performed across parameter space for the physical conditions of each absorber component using the ensemble of absorption properties as constraints. In such models the clouds are often modelled as plane-parallel slabs of gas with an ionising flux incident on one face, making the (simplifying) assumption that each cloud is spatially isolated with single-valued properties (e.g. Churchill et al., 2003; Tripp et al., 2008; Werk et al., 2014; Fumagalli et al., 2016; Keeney et al., 2017; Prochaska et al., 2017).

The analysis and interpretation of CGM observations poses many challenges owing to the complex nature of the halo environment. The shapes of absorption profiles are sensitive to the underlying phase structure and likely contain contributions from different phases, for example due to the motion of gas within the halo and the clumpy gas structure. Even within individual absorber systems the metallicity of the absorbing gas can vary and multiple gas phases may be present (Lehner et al., 2019; Zahedy et al., 2019; Sankar et al., 2020; Chen et al., 2020; Haislmaier et al., 2021; Sameer et al., 2021). Detailed analysis of absorption systems can give relative abundances of different ions that constrain the physical conditions, but this requires high resolution spectroscopy. Studies that use this technique have moved away from the assumption of a single cloud, by modelling the high and low excitation ions separately (Zahedy et al., 2019, 2021; Haislmaier et al., 2021; Qu et al., 2022), or by modelling the absorption components as arising from multiple clouds (Cooper et al., 2021; Sameer et al., 2021; Nielsen et al., 2022). Interpreting the observational picture is further complicated due to the sensitivity of density and metallicity estimates to the shape of the UVB (Oppenheimer & Schaye, 2013; Acharya & Khaire, 2022; Gibson et al., 2022). Furthermore, particular ions are not necessarily produced by the same structures and processes at different redshifts due to the evolving UVB (Haardt & Madau, 2012; Faucher-Giguère, 2020).

Galaxy formation simulations provide a valuable theoretical perspective on these problems as they offer full particle data and physical properties for the gas that makes up the CGM, making it possible to directly interpret observations (see §4.1 and §5.1 for summary of the recent work on the CGM in cosmological simulations). Such simulations can be useful for examining the impact of different line analysis methods on the retrieved CGM gas conditions (e.g. Churchill et al., 2015; Liang et al., 2018). In a recent analysis of a sample of synthetic absorption lines from a cosmological simulation, Marra et al. (2021) tested the accuracy of the single cloud ionisation modelling method of retrieving physical gas conditions. The authors
find that while there is general agreement between intrinsic conditions and those derived from ionisation modelling, such methods capture the average properties of absorbing gas cells, consistent with observational tests by Sameer et al. (2021) comparing single-phase and multiphase modelling. Marra et al. (2022) followed up by testing the assumption of single spatially-isolated absorbing clouds in the CGM, showing that several distinct absorbing clouds may be present within a single absorption component. The distinct clouds may arise from gas of different phases that happen to be aligned kinematically. These results demonstrate that the CGM is a complex environment, with non-linear relationships between the underlying CGM conditions and the resulting absorption observables.

Machine learning (ML) algorithms have the capacity to learn complex, non-linear relationships and as such they have been widely applied to astrophysical problems (see review by Fluke & Jacobs, 2020). In this paper, we explore a novel approach for cosmological simulations to aid in interpreting CGM absorption observations using ML models. We present a framework for Random Forest (RF) mapping between synthetic CGM absorption observables from the SIMBA simulation (Davé et al., 2019) and the underlying absorber conditions from particle data. Such a mapping has the potential to be employed as a useful tool in retrieving physical conditions from real, multi-component absorption observations. This approach eliminates the need for simplifying assumptions about the structure and state of the gas, i.e. whether absorption arise from single or multiple gas phases. Instead the RF mappings implicitly assume the veracity of the SIMBA galaxy formation model and our choice of UVB (Faucher-Giguère, 2020) to produce its predictions.

As demonstrated earlier in this thesis, the SIMBA simulations reproduce well observational galaxy properties (Chapter 3) and the observed absorption properties of H I and selected metal lines in the CGM (Chapter 4), which arise from physically reasonable gaseous conditions (Chapter 5). Therefore SIMBA is a reasonable choice of simulation with which to explore the capabilities of ML methods to learn relationships in the CGM.

 Nonetheless, there is no guarantee SIMBA yields fully accurate and representative circumgalactic media. Indeed, CGM zoom simulations suggest that SIMBA’s resolution may be too poor to capture finer details of multi-phase gas, particularly for stronger absorbers (e.g. van de Voort et al., 2019; Suresh et al., 2019, though see Nelson et al. 2020). This drawback could be explored via comparing the results of this framework applied to other simulations. We leave this aspect for future work, and here focus on presenting the general framework and its results.
when applied to the Simba model.

In this chapter we train Random Forest models on the absorber sample presented in Chapter 5 to produce predictions for the underlying gas conditions in the CGM. This chapter is organised as follows. In §6.2 we describe the RF model and training process. In §6.3 we examine the accuracy of the RF models. In §6.4 we examine the feature importance of the RF models. In §6.5 we present the RF predictions in phase space. Finally in §6.6 we conclude and summarise.

6.2 Random Forest Methods

RFs (described in §1.4) are commonly used both for classification and regression problems. In this Chapter we use RF in its regression mode to deal with the continuous nature of the data. We use the Scikit-Learn (Pedregosa et al., 2011) module’s RF implementation, RandomForestRegressor.

6.2.1 Absorber sample

In this Chapter we use the sample of $z = 0$ absorbers from our investigation into the physical conditions of absorbing halo gas in Chapter 5 (see §5.2 for details of how this dataset is generated), except that we increase the available data volume by selecting and generating spectra for a further 12 galaxies in each $M_\star -$ sSFR bin. This doubles the underlying galaxy sample (except in the highest mass star forming and green valley bins, which have only 23 and 8 galaxies respectively) and roughly doubles the size of the absorber dataset. We probe H$\text{I}$ absorption and a selection of 5 widely-observed metal lines which trace a range of ionisation potentials (Mg$\text{II}$, C$\text{II}$, Si$\text{III}$, C$\text{IV}$ and O$\text{VI}$). Table 6.1 summarises the properties of the absorber sample for each ion.

6.2.2 Input and target features

For each of the ions in our selection, we train a RF model on the dataset of simulated CGM absorbers to predict their underlying physical gas conditions. We do this separately for each of the 6 ions we consider, such that the usefulness of this pipeline is not contingent on having line information simultaneously for
Table 6.1: Absorber sample properties for the RF models: the number of absorbers below the $\chi^2_r$ limit for each species; the column density completeness limit; the $\chi^2_r$ below which we recover 90% of the total EW; the median $\chi^2_r$ of all absorbers; the excitation energy of the species.

<table>
<thead>
<tr>
<th>Species</th>
<th>n</th>
<th>log($N_{\text{min}}$/cm$^{-2}$)</th>
<th>$\chi^2_{90%}$</th>
<th>Median $\chi^2_r$</th>
<th>$E$(eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H i</td>
<td>17750</td>
<td>12.7</td>
<td>3.5</td>
<td>0.7</td>
<td>13.60</td>
</tr>
<tr>
<td>Mg ii</td>
<td>5306</td>
<td>11.5</td>
<td>39.8</td>
<td>1.0</td>
<td>15.04</td>
</tr>
<tr>
<td>C ii</td>
<td>11062</td>
<td>12.8</td>
<td>15.8</td>
<td>1.3</td>
<td>24.38</td>
</tr>
<tr>
<td>Si iii</td>
<td>14119</td>
<td>11.7</td>
<td>35.5</td>
<td>1.9</td>
<td>33.49</td>
</tr>
<tr>
<td>C iv</td>
<td>17463</td>
<td>12.8</td>
<td>6.3</td>
<td>1.2</td>
<td>64.49</td>
</tr>
<tr>
<td>O vi</td>
<td>17463</td>
<td>13.2</td>
<td>4.0</td>
<td>1.2</td>
<td>138.12</td>
</tr>
</tbody>
</table>

all 6 ions. We exclude absorbers where the quality of the Voigt profile fit is low (i.e. the fit has a $\chi^2_r$ below the acceptable threshold for that ion) and the column density is below the completeness limit.

For each ion, we use the same set of input features and target predictors. The input features are chosen from among the properties of the CGM absorbers and their central galaxies. Included features which describe the absorbers themselves are: the column density ($N$), the equivalent width (EW), the linewidth ($b$), the velocity separation from the host galaxy (dv), and the impact parameter, expressed as a fraction of halo virial radius ($f_{r200} = r_\perp/r_{200}$). Properties of the central galaxy that are also included as input features are the stellar mass ($M_\star$), the specific star formation rate (sSFR), and the fraction of kinetic energy contained in rotation ($\kappa_{\text{rot}}$, Sales et al., 2012), which Kraljic et al. (2020b) found is a reasonable proxy for visual morphology.

From these 8 input features we predict 3 target gas predictors: the overdensity ($\delta = \rho/\bar{\rho}_m$), temperature ($T$) and metallicity ($Z$). Each of these is a column density-weighted average at the nearest LOS pixel to the absorber, computed at the time of spectral generation and binned along the LOS (see §5.3).

### 6.2.3 Training

Each of the features is transformed into log space; Jo & Kim (2019) showed that transforming quantities into log space improves the accuracy of machine learning predictions for astronomy problems, owing to the wide range of physical scales present in astronomical data. Exceptions to this are dv, $f_{r200}$ and $\kappa_{\text{rot}}$; dv and $\kappa_{\text{rot}}$ have nearly uniform intrinsic distributions, while $f_{r200}$ consists of 5 specific values
due to our choices of LOS. In addition, we standardize the input and output data by subtracting the mean of the distribution and scaling the variance to unity in each case. Where there are zeros in our dataset, we set them to a small non-zero value. For each ion, we divide the absorber dataset into 80% training data used to build the RF model, and 20% test data used to evaluate the performance of the model. Where multiple absorbers arise from the same LOS these can be separated into the training and test datasets; this mitigates over-fitting in the model due to galaxy or LOS properties.

We train the RF model separately for each target feature, as we find that this improves the accuracy of the prediction. We separately tune the hyperparameters of each RF model to optimize the model accuracy, using Scikit-Learn’s GridSearchCV method to perform an exhaustive grid search over hyperparameter space. The hyperparameters are the number of trees, the minimum number of data points required in order to split the data, and the minimum number of data points in each resulting split. For each set of hyperparameters, a k-fold cross validation is performed with $k = 5$, in which the training data is split into $k$ ‘folds’, and $k$ RF models are iteratively constructed using $k-1$ folds of the data; the overall score for each set of hyperparameters is the average of each of the $k$ RF models. The coefficient of determination $R^2$ is used internally to evaluate the performance of each model, given $n$ data points and true and predicted quantities $X_{true}$ and $X_{predicted}$:

$$R^2 = 1 - \frac{RSS}{TSS}$$ \hspace{1cm} (6.1)

where RSS is the residual sum of squares:

$$RSS = \sum_{i=1}^{n} (X^{i}_{true} - X^{i}_{predicted})^2$$ \hspace{1cm} (6.2)

and TSS is the total sum of squares:

$$TSS = \sum_{i=1}^{n} (X^{i}_{true} - \langle X_{true} \rangle)^2$$ \hspace{1cm} (6.3)

The mean squared error, $MSE = RSS/n$, is the cost function used to determine the best decision tree splits. An MSE of zero represents a perfectly accurate prediction.
Figure 6.1: Hexagonal joint histogram of the predicted H\textsubscript{i} physical conditions from the RF mapping and the true H\textsubscript{i} physical conditions, including only data in the test set. The number of data points in each bin is shown using colorbars. From left to right, the panels show overdensity, temperature and metallicity. The diagonal line represents the case where the RF model makes a perfect prediction. The accuracy of the predictions in each panel is summarised by the inset displaying the transverse scatter $\sigma_\perp$, the correlation coefficient $\rho_r$ and the mean square error MSE. The 1D histograms of the true and predicted values are shown on the top and side of each panel, respectively.

6.3 Predictive accuracy

Here we assess the performance of each of the RF models. Figure 6.1 shows the test data RF predictions for the H\textsubscript{i} absorber physical conditions ($\delta$, $T$ and $Z$) against the true values. The color scale of the hexagonal bins indicates the number of data points in each bin. The black diagonal dashed line represents the 1:1 case of a perfectly predicting model. In each panel the 1D histograms for the true and predicted values lie along the top and right, respectively. The accuracy of the model is summarised in each panel with three quantities: 1) the scatter $\sigma_\perp$, which we define perpendicular to the perfect 1:1 relation; 2) the Pearson correlation coefficient $\rho_r$, given by:

$$\rho_r = \frac{\text{cov}(X_{\text{true}}, X_{\text{predicted}})}{\sigma_{X_{\text{true}}} \sigma_{X_{\text{predicted}}}},$$

(6.4)

where $\sigma_{X_i}$ is the standard deviation of $X_i$; and 3) the MSE. High correlation is preferred, but does not necessarily indicate an accurate prediction as the outputs could have a systematic offset.

Beginning with the predictions for H\textsubscript{i} absorbers, density and temperature are well-predicted by the ML model. True values are highly correlated with the predictions, and the model predictions have low scatter and error. Density
and temperature are physically correlated with one another and have Gaussian distributions. Of the two, temperature ($\sigma_\perp = 0.2$ dex, MSE = 0.08) is predicted more accurately than overdensity ($\sigma_\perp = 0.3$ dex, MSE = 0.18). The RF models for HI density and temperature perform particularly well considering the models’ relative simplicity (compared with e.g. a NN-based model). Aside from transforming the features into log space and using the $k$-fold hyperparameter cross validation, the model has not been extensively tuned by hand. As such, these results represent a basic model which demonstrate the capability of RF models to predict gas conditions, which could be improved upon with further tuning. We have also explored alternative ML approaches such as NNs and CNNs, and found that such models do not offer a substantial improvement in terms of predictive accuracy and take considerably longer to run. This has lead us to favour the RF model for its simplicity, speed, and the degree of interpretability in the form of feature ‘importances’ (see §6.4).

The predictions for HI metallicities are less accurate ($\sigma_\perp = 0.5$ dex, MSE = 0.51). In general, points with $\log Z/Z_\odot < -1$ are overpredicted, while points with $\log Z/Z_\odot > -1$ are underpredicted. This points to the general tendency of our ML models to output a narrower predicted distribution than in the input dataset (this behaviour is also seen to a lesser extent in the density and temperature predictions). This means that the tails of the original distributions are not well captured in the ML model, perhaps as a result of sparse training data at the extremes. Perhaps the poor prediction is unsurprising since HI absorption is not Z-dependent, unlike metal lines which by necessity arise from metal-enriched gas. Therefore it was not obvious that any relationship between H\textsc{i} absorption and metallicity could have been learned from the data. The learned mapping in the metallicity RF model likely arises from the provided galaxy properties and H\textsc{i} absorption strength; we will explore the input feature importance later (§6.4).

The metal line absorber physical conditions are also reasonably well predicted. Figures 6.2 and 6.3 show the performance of the RF models for predicting C\textsc{ii} and C\textsc{iv} absorber conditions, using the same plot structure as above. The performance for Mg\textsc{ii}, Si\textsc{iii} and O\textsc{vi} absorbers are shown in §6.7. The RF models perform similarly well among all the metal lines, with the same tendency to predict a more concentrated distribution of values than in the original data. In general, for each metal line the predictions are less well correlated with the truth values than for H\textsc{i}; metal line absorber $\delta$, $T$ and $Z$ have median correlation coefficients of $\rho_r = 0.69, 0.7, 0.68$, respectively, compared with $\rho_r = 0.84, 0.88, 0.81$.
Figure 6.2: As in Figure 6.1, showing the predictions and true values for C\text{II} absorbers.

Figure 6.3: As in Figure 6.1, showing the predictions and true values for C\text{IV} absorbers.
However, the errors are in general lower for the metal line RF models, with median MSE = 0.11, 0.04, 0.06 and median $\sigma_\perp = 0.23, 0.15, 0.17$ for $\delta, T$ and $Z$, compared with MSE = 0.18, 0.08, 0.51 and $\sigma_\perp = 0.2, 0.3, 0.5$ for H I. The metal line RF models also have lower scatter, although we emphasise that this is partly due to the reduced range in physical conditions traced by the metal lines. Overall the RF models give reasonable predictions for the physical conditions, and again were not extensively tuned to achieve this.

An interesting feature of the original absorber dataset is bimodal metallicity distributions at log $Z/Z_\odot \sim -0.25$ and 0.25, which have not been reported in earlier SIMBA CGM work. The bimodality is apparent in every metal line apart from Mg II, and is broadly reproduced by the RF models. Populations of absorbers in the cool CGM of low redshift Lyman Limit Systems (LLSs) have also been observed to have bimodal metallicity distributions, with both metal-poor and metal-rich absorbers (albeit shifted to lower metallicities, Lehner et al., 2013, 2018, 2019; Wotta et al., 2016, 2019; Berg et al., 2022), suggesting multiple origins for the cool CGM gas, although the metallicities of the observed metal-poor absorbers are much lower than that seen in SIMBA. In future work we will investigate the origin of the bimodal absorber metallicity distribution in SIMBA.

### 6.4 Feature importance

In this section, we seek insights into the physical origin of the ML-probed correlations by assessing which input features are most useful in predicting the physical conditions.

#### 6.4.1 RF model importance

An advantage of the RF method over other ML algorithms (such as neural networks) is it allows some degree of interpretability in the form of the ‘importance’ of each feature, which arise ‘for free’ from the structure of the RF model. For an individual decision tree, a feature’s importance is computed from the number of times it is used to split the data and how close to the top of the tree the splits are. For an RF model, the importances are the normalised average over all decision trees. However, importance metrics are biased if the input features are highly correlated with one another (Strobl et al., 2007, 2008) and so they
Figure 6.4: Importance values in predicting HI physical conditions for each remaining input feature, against the input feature removed from the training data. From left to right, the target features are $\delta, T$ and $Z$.

should be treated with caution. Thus we prefer not to use the importance directly reported by RF, but instead compute it more empirically.

To do so, we determine each input feature’s importance by iteratively building the RF model and removing each of the features in turn, using the same optimized hyperparameters as in the full feature model. We then retrieve the importance of the remaining features for each model. This process determines whether a feature is genuinely important, or merely defined as such through a fluke of feature combinations. When the most important features are removed, identifying which features take its place as the most important gives an indication of what the RF model is learning.

Figure 6.4 shows the feature importance values for predicting HI absorber conditions, against the feature removed from the training data. For predicting overdensity, $N$ is most important feature. When column density is the removed feature, the most important feature is EW; $N$ and EW are correlated with one another and both are correlated with physical density (see Figure 5.9 of Chapter 5). When predicting temperature, $b$ is the most important feature since the linewidths of individual absorbers in the original spectrum are set in party by thermal Doppler broadening (with the additional effect of bulk gas motions). When predicting metallicity, the velocity separation is the most important feature. It is not intuitively obvious why this is the case; perhaps due to a dependence on halo velocity dispersion, which is correlated with $M_*$ and thus the metallicity of the host galaxy that is predominantly responsible for enriching its CGM.

Figures 6.5 and 6.6 likewise show the feature importance values for predicting CII and CIV absorber properties. We have examined feature importance for all
metals and found that these are representative cases. For the low ion C II the feature importance rankings for $\delta$ and $T$ are similar to that of H I. There is a slightly reduced relative importance of $N$ in predicting $\delta$ in favour of $b$ ($N$ and $b$ are correlated features due to their underlying dependence on $\delta$ and $T$). For both ions, the importance of $b$ in predicting $T$ is enhanced compared with H I.

In contrast with H I, the most important feature for predicting $Z$ for metal lines is sSFR; when sSFR is removed, the RF model learns from $M_\star$ and $\kappa_\text{rot}$ instead, indicating that the RF model predicts $Z$ from the galaxy properties. The picture is similar for the high ion C IV, except that in predicting $\delta$ the most important features are instead $f_{200}$ and sSFR. $f_{200}$ is perhaps less useful for C II since most of the low ion absorption arises from the inner CGM (see §5.3.5). Galaxies with high star formation have denser gas in their CGM (as discussed in Chapter 5), although this is also the case for H I and C II, so it is unclear why sSFR specifically is an important feature for C IV absorbers.
6.4.2 Change in predictive accuracy

Arguably the most meaningful measure of ‘importance’ to the model is in which features add the most useful information in terms of predictive accuracy. We assess this by iteratively removing each of the features in turn from the training data, and running the RF model as before (with the hyperparameters optimized for the full feature case). In contrast with §6.4.1, we now compute the scatter $\sigma_\perp$ for each new model; if the quality of the predictions are significantly degraded in the absence of a particular feature, this necessarily indicates that this feature encodes crucial information about the physical conditions.

Figure 6.7 shows the change in $\sigma_\perp$ resulting from the removal of each input feature in predicting the physical conditions, where a positive change indicates an increase in scatter. Each of the six panels shows the models for a different ion; the rows within each panel show the models for each of $\delta, T$ and $Z$. The upper plots show $\text{H} \text{I}$ (left), $\text{Mg} \text{II}$ (middle) and $\text{C} \text{II}$ (right) absorber models, while the lower plots show $\text{Si} \text{III}$ (left), $\text{C} \text{IV}$ (middle) and $\text{O} \text{VI}$ (right).

In contrast to the feature importance values, the largest increase in scatter when predicting $\text{H} \text{I}$ absorber overdensity comes from removing $d_v$ and $f_{r200}$. The loss of accuracy from removing $f_{r200}$ suggests that the RF model is learning the radial density profile; this is also the case for the metal line $\delta$ predictions. It is less clear why $d_v$ is necessary for an accurate prediction, since halo absorbers can appear
at any velocity separation depending on their kinematics. For all ions, removing $b$ causes an increase in scatter in $T$ predictions, confirming that the high feature importance of $b$ reflects the genuine physical relationship with temperature.

For H\textsubscript{i}, the metallicity predictions are degraded by removing $dv$, $f_{r200}$ or any galaxy property; interestingly the model accuracy improves when absorption-related features are removed. In other words, gas metallicity in the CGM of a given galaxy can be predicted with reasonable accuracy from only LOS and radial absorber position. The metal lines broadly show the same changes in scatter with removed features, although the changes to the model accuracy are more marginal.

### 6.5 Phase space

Having examined the predictive accuracy and inner workings of the RF models on individual properties, we now ask whether the RF models can reproduce the 2D ($\delta, T$) phase space structure of the absorbers. Although the RF models can separately predict $\delta$ and $T$, this does not guarantee that they reproduce the relationship between these quantities - particularly since the models are trained separately for $\delta$ and $T$, so each RF model has no knowledge of the other target quantities.

Figure 6.8 shows the predicted temperature against predicted overdensity for the 6 species we consider. The distributions for the truth and predicted data are shown along the top and right hand side of each panel. Note that the plot limits are different for each ion. The points are colour coded by the fractional distance from the true value in $\delta - T$ phase space:

$$
\sigma_{\text{phase}} = \sqrt{\left(\frac{\delta_{\text{true}} - \delta_{\text{predict}}}{\delta_{\text{true}}}\right)^2 + \left(\frac{T_{\text{true}} - T_{\text{predict}}}{T_{\text{true}}}\right)^2}
$$

(6.5)

The colour scale indicates that in general the predicted points lie near their truth values in phase space; the metal line absorbers in particular have a low displacement. The contours show the true distribution in phase space for the test dataset. The upper plots show H\textsubscript{i} (left), Mg\textsubscript{II} (middle) and C\textsubscript{II} (right) absorbers, while the lower plots show Si\textsubscript{III} (left), C\textsubscript{IV} (middle) and O\textsubscript{VI} (right). The original structure in phase space between overdensity and temperature is reproduced well by the RF models for each of the ions we consider, a success of the ML approach which (since there was no specific tuning of the model to achieve
Figure 6.8: Predicted temperature against predicted overdensity for each of the 6 ions we consider, coloured by overall phase space fractional error. The 1D truth (blue curve) and predicted (pink curve) distributions are shown along the top and right of each panel. The contours show the true distribution in phase space for the test dataset. The limits of the plots differ for each ion.

This arises because temperature information is encoded in the overdensity data and vice versa. This is an important test of the RF models which verifies that accurate predictions can be made for multiple physical properties per observation.

That said, although the RF models succeed in predicting the phase space structure, the predictions are in general too concentrated near the mean of the data. By comparing the predicted distribution with the contours from the original data, it is clear that the predictions ought to be more spread in phase space. This appears to be a generic feature of the RF models, which is also apparent in the 1D distributions of each feature - in general, the RF models produce predicted distributions that are too concentrated towards the mean. As mentioned in §6.3, this likely arises from sparse training data at the extremes. Thus, the predicted distributions do not capture the important information described by the intrinsic scatter in the original data, which biases the usefulness of these models for observational analysis. Therefore some additional step beyond the basic ML model is required in order to capture the full structure in phase space. Employing an oversampling technique (such as Synthetic Minority Over-Sampling Technique...
Figure 6.9: Predicted temperature against predicted overdensity for each of the 6 ions we consider, mapped to the shape of the truth 1D distributions using a quantile transformer. Points are coloured by overall phase space fractional error. The 1D distributions for the truth data (blue curve) and transformed predictions (pink curve) are shown along the top and right of each panel. The contours show the true distribution in phase space for the test dataset.

for Regression with Gaussian Noise, SMOGN) can boost under-sampled regions of phase space and as such mitigate issues that arise as a result of imbalanced training datasets (de Santi et al., 2022).

In order to extend the RF network to also properly capture the full 2D phase space distribution, we develop a new approach based on a normalising transform. By this, we mean that the predicted and truth data for each feature are mapped onto standard normal distributions, and then the predicted distribution is transformed back onto the shape of the truth data distribution.

To accomplish this we use the quantile transform non-parametric method implemented in Scikit-Learn (Pedregosa et al., 2011), QuantileTransformer. The method first maps the cumulative distribution of the data onto a standard Gaussian, and then computes the transformed values using a quantile function. The function also provides the inverse mapping that transforms a distribution back into the original coordinates. The inverse mapping for the truth data distribution (computed from the training dataset) is used to reconstruct the
Figure 6.10: Histograms showing the direct comparison between truth data (dark purple), the predictions (light purple), the transformed predictions (orange) and the predictions with added scatter (yellow) for H\textsubscript{i} overdensities and temperatures. Predicted distribution. In this way, we can reproduce the larger variance in the truth data without assuming a shape for the predicted data.

Figure 6.9 shows the the predicted temperature against predicted overdensity, using the above normalising transform approach to map the shape of the truth data. This can be seen in the 1D distributions along the top and right hand side, which in most cases closely follow the truth data distributions. Crucially, the transformed data also retains the phase space structure of the original predictions. In addition, transforming the predictions onto the shape of the truth data results in no loss of accuracy for the predictions; the MSE, $\rho$ and $\sigma_\perp$ for the transformed test datasets are very similar to those for the original test dataset predictions. As such, this is our preferred method for reproducing the scatter in the truth CGM conditions data.

We initially explored a simpler approach where we added scatter directly to the predicted data. The results in phase space for the additional scatter approach is shown in §6.8. We found that this approach substantially washed out non-Gaussian structure in the predicted distribution, such as the anti-correlation between $\delta$ and $T$ at $\delta > 10^3$. By instead using the normalising transform method, we are preserving these structures in phase space as much as possible.

Figure 6.10 directly compares the distributions for the truth data (dark purple), the predictions from the RF models (light purple), the transformed predictions (orange), and the predictions with additional scatter (yellow) for the H\textsubscript{i} absorber overdensities and temperatures. The predicted physical conditions from the
RF models are clearly too closely concentrated towards the mean. In contrast, applying the normalising transform approach results in a predicted dataset which closely matches the truth data. When additional scatter is instead added to the predictions, the resulting distribution also more closely matches the truth data, although not so precisely. Quantitatively, a two-sample Kolmogorov–Smirnov test with respect to the truth data gives $p$-values of $> 0.95$ for the transformed overdensity and temperature distributions, but $\sim 0.25$ for the distributions with added scatter.

There are pros and cons to including the normalising transform approach to ensure that the phase space scatter is well reproduced. If one wanted to compute distribution functions for physical quantities inferred from absorption line data, not including this post-processing step would result in the distribution functions improperly capturing the tails, which may be important for some applications. However, the additional step necessarily degrades the $\sigma_\perp$ and MSE of the predictions, albeit only marginally. The correlation coefficient does not change since we are only scaling the predictions. For example, in the case of HI overdensity, the $\sigma_\perp$ and MSE increase from $0.30 \to 0.31$ and $0.18 \to 0.20$ respectively after applying the normalising transform. For HI temperature, the $\sigma_\perp$ increases from $0.20 \to 0.21$, while the MSE does not increase (at this level of precision). Whether or not to employ the above method thus depends on the application.

### 6.6 Conclusions

We have produced machine-learnt mappings between CGM absorption observables and the underlying gas conditions for HI and selected metal lines using a Random Forest approach. RF models are preferred over other ML techniques for their relative simplicity and interpretability. These mappings represent a proof of concept for using ML models as part of an analysis pipeline for observational CGM data, which crucially does not make simplifying assumptions about the phase or composition of the absorbing gas. We identify a general tendency of the RF models to output a narrower predicted distribution than in the input data.

We demonstrate two methods of reproducing the scatter of the input data: first by adding random Gaussian noise to the predictions, and second by transforming the predictions to the shape of the truth data. Our main results are as follows:
- The RF models predict reasonable H I overdensities ($\sigma_\perp = 0.3$ dex, MSE = 0.18) and temperatures ($\sigma_\perp = 0.2$ dex, MSE = 0.08). The predictions of overdensity and temperature are highly correlated with their truth values. Metallicity is less well predicted ($\sigma_\perp = 0.5$ dex, MSE = 0.51); metallicity is not directly traced by HI, therefore the learned relationship likely arises from the correlation with density.

- The RF models also predict reasonable metal absorber conditions and perform to a similar accuracy among all metal lines, with median $\rho_r = 0.69, 0.7, 0.68$, median $\sigma_\perp = 0.23, 0.15, 0.17$ and median MSE = 0.11, 0.04, 0.06 for the overdensity, temperature and metallicity predictions, respectively. Lower $\sigma_\perp$ compared with H I predictions are partly due to the smaller dynamic ranges probed for metals.

- We report a bimodality in the absorber metallicity distributions for four of the five metal lines (C II, Si III, C IV and O VI), suggesting multiple origins for the CGM gas in the Simba model.

- In terms of feature importances, the RF models learn H I absorber overdensity from column density and equivalent width, temperature from the Doppler parameter, and metallicity from the LOS velocity separation. Low ion feature importances are similar to H I, except that metallicities are learned from sSFR and $\kappa_{\text{rot}}$. High ion feature importances are similar to the low ions, except that the overdensities are learned from radial distance and galaxy properties.

- In terms of predictive accuracy, the radial distance and LOS separation provide the most useful information for predicting H I overdensity; the radial distance is also most useful for the metal line overdensities. The Doppler parameter is again the most important feature for predicting temperature for all lines. The LOS separation provides the most useful information for predicting H I metallicity; the predictions for all lines are degraded by removing galaxy properties.

- The predictions for overdensity and temperature reproduce the phase space structure seen in the original data for all six ions, despite being trained for separately in the RF models. This verifies that accurate predictions can be made for multiple physical properties per observation.

- By mapping the predicted data distributions onto the shape of the input distributions using a quantile normalising transformer, we can reproduce
the intrinsic scatter in the CGM phase space conditions with no loss of predictive accuracy or phase space structure.

Although we have considered H I and the metal ions separately, future work on this topic could explore RF models using combinations of absorption lines to assess whether predictions may be improved by using information from multiple ion species. A shortcoming of the ML models presented here is the that the predictions are too concentrated towards the mean of the distribution. Further development would be needed on the pipeline in order to reproduce the scatter in the original data without losing information in phase space.

The motivation of this project is to develop a useful analysis tool for the astronomical community to aid in interpreting absorption observations of the CGM, assuming the galaxy formation model of the SIMBA simulations and the Faucher-Giguère (2020) UVB. The next phase of this work is to test the method by applying the ML mappings to real observational data and comparing to results derived from ionisation modelling. As such, the trained models produced for this work are available online\(^1\) and we encourage others to test the RF models on their own observational data.

A natural extension of this project will be to develop additional ML mappings using absorber data from other simulations such as EAGLE and IllustrisTNG to assess the impact of galaxy formation models on the predicted conditions for the CGM. Training the RF models on data from one simulation and testing on data from another would provide a robust test of the impact of galaxy formation model on our results. In addition, developing mappings using absorbers from the CAMELS project (Villaescusa-Navarro et al., 2021) would test the dependence of our results on both astrophysical and cosmological models.

### 6.7 Results for other metals

For completeness, here we present figures similar to Figure 6.1, for Mg II, Si III, and O VI. The general trends are already captured by the plots in the main text for C II and C IV. However, there are some interesting notable point. For instance, O VI shows significantly higher temperatures, as expected since it is a higher ionisation line. The metallicity bimodality is still slightly present for O VI,

\(^1\)https://github.com/sarahappleby/cgm_ml
Figure 6.11: As in Figure 6.1, showing the predictions and true values for Mg II absorbers.

Figure 6.12: As in Figure 6.1, showing the predictions and true values for Si III absorbers.

Figure 6.13: As in Figure 6.1, showing the predictions and true values for O VI absorbers.
Figure 6.14: Predicted temperature against predicted overdensity for each of the 6 ions we consider, with added random Gaussian noise to reproduce the original 1D distributions. Points are coloured by overall phase space fractional error. The 1D truth and predicted distributions are shown along the top and right of each panel. The contours show the true distribution in phase space for the test dataset. The widths of the random noise Gaussians for $\delta$ and $T$ are shown in the bottom right of each panel. Compared to the normalising transform results shown in Figure 6.9, this approach washes out features in phase space substantially more.

although at a much lower significance than for the lower ions. Si III has the largest scatter in the recovered $T$, and also shows some bias such that high-$T$ absorbers are under-predicted while low-$T$ ones are overpredicted. This may be because Si III seems to have absorbers spanning the widest range in temperatures from among the metal ions considered. In terms of the RF performance, however, these ions tell a similar story, which is encouraging since it means the RF methodology is widely applicable with similar efficacy across a range of commonly observed low-$z$ UV ions.
6.8 Direct Gaussian approach to adding phase space scatter

In §6.5 we described our normalising transform-based approach to adding scatter to the RF predictions in order to match the 2D truth distributions in phase space. A more straightforward approach is to directly add Gaussian scatter to the predicted $\delta$ and $T$ distributions to match the truth without first applying a normalising transformation. However, the results were less satisfactory.

Figure 6.14 shows the results, which can be compared to Figure 6.9. It is clear that the simpler approach causes features within the true phase space to be more washed out, and substantially degrades the predictive accuracy. Thus we prefer the normalising transform approach presented in §6.5.
Chapter 7

Conclusions and Future Work

7.1 Summary of Results

This thesis has focused on using the SIMBA simulations to examine gas properties inside and outwith galaxies, with the goal of understanding and constraining physical models of galaxy evolution.

7.1.1 The Impact of Quenching on Galaxy Profiles in SIMBA

In Chapter 3 I presented the galaxy size-mass relation within the SIMBA simulations and showed that SIMBA reproduces observations of star-forming galaxy sizes at \( z = 0 \) and their redshift evolution. The sizes of quenched galaxies however are too large at \( z = 0 \) owing to numerical heating. I have shown radial profiles of gas quantities in star forming galaxies and green valley galaxies (those currently being quenched) at \( z = 0 \). The peak of star formation within both galaxy types occurs at \( R \sim 0.5 - 1R_{\text{half}} \), with an exponential drop-off at large radii. In massive star forming galaxies and all green valley galaxies, there is a strong central depression in star formation and gas fraction, qualitatively as observed.

These results show that both inside-out and outside-in quenching processes play a role in suppressing star formation in SIMBA galaxies. The central depression of star formation is caused by the evacuation of star-forming gas by SIMBA’s implementation of X-ray black hole feedback. SIMBA’s satellite galaxies have
higher central sSFR and lower sSFR in the outskirts than in centrals, in qualitative agreement with observations. At $z = 2$ SIMBA does not show central depressions in massive star-forming galaxies, suggesting SIMBA’s X-ray feedback should be more active at high-$z$.

### 7.1.2 The Low Redshift Circumgalactic Medium in SIMBA

In Chapter 4 I examined the properties of the low-redshift CGM around star-forming and quenched galaxies in SIMBA. Halo baryon fractions are generally $\lesssim 50\%$ of the cosmic fraction due to stellar feedback at low masses and jet-mode black hole feedback at high masses. The halos of L* galaxies can contain as little as 10% of the cosmic baryon fraction in CGM gas. Mass and metal budgets of halos show that baryons and metals in the CGM of quenched galaxies are $\gtrsim 90\%$ hot gas, while the CGM of star-forming galaxies is multiphase. The hot CGM gas has low metallicity, while warm and cool CGM is more enriched, with a metallicity close to that of galactic gas.

I compare H I and metal line absorption in the CGM to observations from the COS-Halos and COS-Dwarfs surveys using a sample of matched galaxies. Equivalent widths, covering fractions and total path absorption of H I and selected metal lines (Mg II, Si III, C IV and O VI) around SIMBA star-forming galaxies are broadly consistent with COS-Halos and COS-Dwarfs observations to $\lesssim 0.4$ dex, depending on the ion and the assumed UVB. In particular, SIMBA reproduces the observed dichotomy of O VI absorption around star forming and quenched galaxies. However absorption around quenched galaxies is underpredicted in general. These results are surprisingly sensitive to the choice of UVB. Varying the feedback model shows that CGM metals primarily come from stellar feedback, while jet-mode black hole feedback reduces absorption, particularly for lower ions.

### 7.1.3 The Physical Nature of Circumgalactic Medium Absorbers in SIMBA

In Chapter 5, I examined the nature of the low-redshift CGM in the SIMBA cosmological simulations around galaxies with a range of stellar masses and star formation rates. I fitted Voigt profiles the synthetic CGM absorption spectra to obtain line properties, and estimate the density, temperature, and metallicity
of the absorbing gas. I find that CGM absorbers are most abundant around star forming galaxies with \( M_* < 10^{11} \) M\(_\odot\). The abundance of absorbers around green valley galaxies shows similar behaviour to those around quenched galaxies, suggesting that the CGM “quenches” before star formation ceases.

The CGM absorber overdensity depends on the sSFR of the galaxy, while the absorber temperature depends on galaxy stellar mass. While H\(_1\) absorbing gas exists across a broad range of cosmic phases (condensed gas, diffuse gas, hot halo gas and Warm-Hot Intergalactic Medium), I find that essentially all low-ionisation metal absorption arises from condensed gas, and O\( \text{VI} \) absorbers are split between hot halo gas and the WHIM. For the high ions, which trace hotter gas, \( \sim 25-55\% \) and \( \sim 80-95\% \) of the absorption arises from collisionally ionised CGM gas, for C\( \text{IV} \) and O\( \text{VI} \) respectively. Gas that belongs to satellite galaxies contributes just \( \sim 3\% \) of overall H\(_1\) absorption but as much as \( \sim 30\% \) of Mg\( \text{II} \) absorption.

### 7.1.4 Mapping Circumgalactic Medium Observations to Theory Using Machine Learning

In Chapter 6, I showed a proof of concept for using random forests, a form of machine learning, to predict physical conditions in the CGM from the absorption properties of H\(_1\) and metal lines. The random forest models predict reasonable overdensities and temperatures for H\(_1\) absorption, with less accurate predictions for H\(_1\) metallicity. The metal line physical conditions are also reasonably well predicted. Depending on the ion, accuracy in the predictions is linked with radial distance and LOS velocity separation information for predicting overdensities; Doppler parameter information for predicting temperatures; and LOS separation and galaxy properties for predicting metallicities. While the random forest models broadly reproduce the overdensity-temperature phase space structure, the predictions are in general too concentrated towards the mean of the distribution. As such, I presented a normalising transform method for reproducing the intrinsic scatter of the truth data.
7.2 Future Prospects

7.2.1 This thesis

There are a number of avenues of research that we can pursue in continuation of this PhD research. Firstly, further development is required on the ML framework set out in Chapter 6 in order for it to be a fully workable tool for the community. The two main improvements required are:

1. Higher predictive accuracy for the physical conditions, particularly for H1 metallicity. A reasonable expectation of an ML model is that it returns predictions with a correlation between true and predicted values higher than $\rho_r > 0.9$ and accuracy to within 0.1 dex. We will explore further tuning methods and alternative ML approaches to determine whether this is feasible with our dataset.

2. Reproduced scatter in the predictions such that the predictions are not biased towards the mean of the input distributions. We have demonstrated two methods of achieving this, with a normalising transform method and additional Gaussian noise. Other approaches may be possible, including supplying uniform distributions to the ML algorithm, or exploring alternative ML algorithms.

With an improved ML analysis pipeline, the true test of this approach will be to apply it to real observations of the CGM. We will determine the RF model’s performance in this regard with comparison to more established methods in the literature. An extension of this work will be to develop an analogous pipeline for other publically available simulation projects such as EAGLE and IllustrisTNG in order to examine how different galaxy formation models interpret CGM observables. These pipelines will also be made public to enable easy comparisons between models by the wider CGM community.

Secondly, a particle tracking analysis in the Simba simulation would answer a number of unanswered questions raised in this thesis:

1. The fate of the halo baryons that are lost to the IGM via jet mode AGN feedback (see §4.2.2)
2. The origins of metals and dust in the CGM, explaining the high metallicity of the hot and warm CGM phases (see §4.2.3) and the bimodal metallicity distribution of CGM absorbers (see §6.3)

3. The origin, evolution and survival of the cool gas in the halos of massive quenched galaxies (see §4.3)

Particle tracking in cosmological simulations is attractive as it offers a purely theoretical approach to studying the CGM, enabling direct measurement of the evolution of individual gas elements.

Thirdly, and following from the comparison to COS-Halos and COS-Dwarfs in Chapter 4, the same mock absorption survey analysis pipeline can be applied to any sample of CGM absorbers. This enables comparisons to a wide range of other observational datasets that probe different galaxy types and evolution stages. For example, a comparison with samples of Mg\textsuperscript{II} absorbers around massive quenched galaxies (e.g. COS-LRGs, Chen et al. 2018, 2019, Zahedy et al. 2019; the Red Dead Redemption Survey, Berg et al. 2019) would further quantify the extent of the problem of the lack of cool gas absorption around massive galaxies. A wide variety of targeted CGM absorber surveys that have been completed in recent years, probing the CGM of Milky Way analogs, AGN, starbursts and Lyman Limit Systems, as well as the connections between the ISM, CGM and IGM. At higher redshift, possibilities include the CaSBAH survey of NeVII absorbers (Burchett et al., 2019). Such comparison projects would represent further tests of SIMBA’s galaxy formation models. Of particular interest would be a wider comparison project compiling comparison results SIMBA, EAGLE and IllustrisTNG to build a broader picture of how current cosmological simulations perform at producing an accurate CGM.

Finally, a complementary avenue of comparisons lies in synthetic emission mapping of the CGM within the SIMBA simulation. Studying the CGM in emission remains the ideal as it provides direct spatially resolved information for the physical extent of the CGM, and (with multiwavelength studies) the extent of different gas phases. Mock observations in emission are possible within SIMBA using the PYGAD analysis package; in addition, SIMBA’s self-consistent dust model allows for the inclusion of dust attenuation. This approach provides a host of new comparisons by which to test the SIMBA model, for example IFU surveys using KCWI on Keck, MUSE on the VLT, and HARMONI on the ELT, and 21cm HI emission studies (e.g. the MHONGOOSE survey using the MeerKAT
SKA precursor de Blok et al., 2016). As with absorption line comparisons, we can produce predictions for the emission signatures of different galactic feedback prescriptions using the different SIMBA variant boxes. Furthermore, emission and absorption observations are not typically available for the same systems, as such it is unclear how emission and absorption measurements relate to each other, and whether they trace the same CGM gas. Cosmological simulations allow us to establish the connection between emission and absorption observations.

7.2.2 The field

The next generation of the SIMBA project will be the forthcoming KIARA simulations, which will feature a number of upgrades and improvements to the SIMBA sub-grid models, such as physically evolved winds (Huang et al., 2020), turbulent gas mixing (Rennehan, 2021), updated chemical enrichment (Hough et al, in prep.), and direct H$_2$ tracking for star formation (Jones et al, in prep.). CGM and ISM observables will continue to represent key predictions for the forthcoming simulations, and the tests of SIMBA presented in this thesis ought to be repeated for KIARA in order to quantify the impact of these processes. The new simulations will be run with the SWIFT gravity and hydrodynamics code (Borrow et al., 2018) using the SIMBA sub-grid models, which offers substantial computational speed improvements over the historical GIZMO implementation. The speed-up provided by SWIFT will allow for higher resolution volumes, additional physics modules, and more precise model calibration than was possible for the SIMBA simulations.

In a broader sense, cosmological simulations will be improved in the coming years by the widespread implementation of additional physical modules, chiefly turbulence, cosmic rays, magnetic fields and ray tracing. However, these mechanisms continue to represent a technical challenge due to the computational cost. Furthermore, the ideal for future simulations will be to move away from the approach of sub-grid modelling, in favour of directly simulating small scale processes, such as star formation and the growth of black holes. At present, such a direct approach has only been achieved within a cosmological simulation using intensive hyper refinement schemes of individual systems (Anglés-Alcázar et al., 2021).

On the observational side, the challenge is to empirically constrain the composition, physical state, origin and evolution of gaseous halos. Detailed observations
that resolve the CGM and ISM (both spatially and spectrally) across galaxy types and redshifts are required to achieve this. At present the CGM is typically probed with a single LOS absorption spectrum per galaxy, meaning such observations do not capture the full distribution of gas in the CGM; furthermore, gas that is not co-spatial may be observed at the same velocity position due to gas kinematics. Increasing the available LOS per galaxy, for example by using fainter quasars or other galaxies as background sources, and observing at high spectral resolution will help in this regard.

With optical transitions such as Mg II this is possible from the ground; however in the UV regime (which probes a wide range of gas temperatures and densities, see Tumlinson et al., 2017), space-based facilities are necessary. With high resolution UV spectroscopy it is possible to resolve the individual components of absorption features (e.g. Zahedy et al., 2019; Chen et al., 2020; Zahedy et al., 2021). Data from the COS and STIS instruments on HST (which formed the standard for the synthetic observations presented in this thesis) represent the best quality data available in the UV at present. In addition, spatially resolved emission mapping remains the ideal as it offers global quantities such as luminosity, mass, size, fluxes, angular momentum, and large scale structure. HST has been operational for the last 30 years, but when it reaches the end of its life there will be no comparable facilities for high resolution UV imaging and spectroscopy. A UV telescope for the next generation such as the Large Ultraviolet/Optical/Near Infrared Surveyor (LUVOIR, The LUVOIR Team, 2019) is needed in order to continue improving our understanding of the CGM. Ultimately multiwavelength data spanning radio, IR, optical UV and X-rays are needed across galaxy types and cosmic epochs in order to fully characterise all components of galaxy halos.
Bibliography


Appleby S., Davé R., Sorini D., Cui W., Christiansen J., 2023a, MNRAS,


Breiman L., 1996, Machine Learning, 24, 123

Breiman L., 2001, Machine Learning, 45, 5


205
Colella P., Woodward P. R., 1984, Journal of Computational Physics, 54, 174


Donahue M., Voit G. M., 2022, Phys. Rep., 973, 1


Fluke C. J., Jacobs C., 2020, WIREs Data Mining and Knowledge Discovery, 10, e1349
Fukushima K., 1980, Biological Cybernetics, 36, 193

208
Kwik R. J., Wang J., Barmby P., Holwerda B. W., 2022, Advances in Space Research, 70, 229


Saintonge A., Catinella B., 2022, ARA&A, 60, 319


Springel V., 2010a, ARA&A, 48, 391


220
Tripp T. M., et al., 2011, Science, 334, 952
Tumlinson J., et al., 2011, Science, 334, 948
Vogelsberger M., Marinacci F., Torrey P., Puchwein E., 2020, Nature Reviews Physics, 2, 42
de Santi N. S. M., Rodrigues N. V. N., Montero-Dorta A. D., Abramo L. R.,

van de Voort F., 2017, The Effect of Galactic Feedback on Gas Accretion and
Wind Recycling. p. 301, doi:10.1007/978-3-319-52512-9_13

van de Voort F., Schaye J., Booth C. M., Haas M. R., Dalla Vecchia C., 2011,
MNRAS, 414, 2458

van de Voort F., Springel V., Mandelker N., van den Bosch F. C., Pakmor R.,


van den Burgh S., 1991, Publications of the Astronomical Society of the Pacific,
103, 390